NASA

Technical Memorandum 83902

DEVELOPMENT OF A DIODE LASER HETERODYNE SPECTROMETER AND OBSERVATIONS OF SILICON MONOXIDE IN SUNSPOTS

David Alan Glenar

NOVEMBER 1981

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

DEVELOPMENT OF A DIODE LASER AND HETERODYNE SPECTROMETER

AND

OBSERVATIONS OF SILICON MONOXIDE IN SUNSPOTS

David Alan Glenar

Submitted in partial fulfillment of

the requirements for the

Degree of Doctor of Philosophy

in the Department of Physics and Astronomy

Pennsylvania State University

1981

ABSTRACT

A state-of-the-art, tunable diode laser infrared heterodyne spectrometer has been designed and constructed for ground-based observations throughout the 8-12 micron atmospheric window. The instrument has been optimized for use with presently available tunable diode lasers, and has been designed as a flexible field system for use with large reflecting telescopes. The instrument was aligned and calibrated using laboratory and astronomical sources.

Observations of SiO fundamental (v = 1-0) and hot band (v = 2-1) absorption features have been made in sunspots near 8 microns using the spectrometer. These measurements constitute the first reported detection of solar SiO, and the first such measurement using coherent detection. The data permit an unambiguous determination of the temperature-pressure relation in the upper layers of the umbral atmosphere, and support the sunspot model suggested by Stellmacher and Wiehr (Astron. and Astrophys., 7, 432, 1970).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Drs. Michael J. Mumma and Theodor Kostiuk for their guidance during the course of this investigation and to the following thesis committee members; Drs. John S. Nisbet, Lawrence W. Ramsey, Michael J. Mumma, Satoshi Matsushima, Douglas H. Sampson and T. K. McCubbin, Jr. Thanks also to Drs. Donald E. Jennings, Drake Deming, John Hillman, Lawrence Ramsey and David Buhl, who contributed to the research by many helpful discussions.

The author is indebted to Mr. James Faris for technical assistance and to Mr. Arnold Abbott and Ms. Jan Friskey for assistance during the observations. Thanks go to Martha, Jennifer and Mark Glenar for their continued support and tolerance.

This work was supported by the National Aeronautics and Space Administration under Grants NGT 39-009-800 and NGL 39-009-003.

CHAPTER I

INTRODUCTION

1.1 General Statement of the Problem

1.1.1 The Need for a General Purpose Tunable Diode Laser Heterodyne Spectrometer

Infrared heterodyne spectroscopy is a passive technique employing coherent optical detection for the study of spectral features in remote sources. It employs optical components such as mirrors and lenses normally associated with incoherent optics, but since it is a coherent process the principles of cm and mm wave receivers dictate the instrument design. Because of its coherent nature, it offers the same advantages as the longer wavelength techniques, notably ultra-high resolving power and high frequency precision. Heterodyne spectroscopy can be considered complementary to, rather than in competition with, Fourier transform spectroscopy (FTS), which is widely used for high resolution IR measurements of remote sources. The FTS technique permits measurements of large spectral intervals with frequency resolving powers $\nu/\Delta\nu < 10^5$. Infrared heterodyne spectroscopy covers a much smaller spectral interval in a single measurement, but can provide almost arbitrarily high frequency resolving power. This permits the recovery of fully resolved line shapes and reveals fine structure not accessible to the FTS technique.

Since the advent of high speed infrared photomixers, heterodyne spectroscopy using CO_2 gas lasers has become a powerful technique

for the study of terrestrial atmospheric species and astrophysical objects at high spatial and spectral resolution. Unlike coherent detection at radio frequencies, however, gas laser local oscillators only provide for operation at the fixed frequency of the laser transitions. Although a large number of transitions are available, the intermediate frequency coverage amounts to about 7 percent of the typical spacing between transitions, so these instruments only allow measurement over very small intervals throughout their overall operating wavelength range. Moreover, their operation is presently restricted to wavelengths between 8 and 12 microns.

The substitution of tunable diode lasers (TDL's) for gas lasers as local oscillators in heterodyne systems offers the promise for continuous tunability, allowing operation over the entire nominal wavelength range of the diode laser. The development of a reliable TDL heterodyne spectrometer with the sensitivity of gas laser instruments has yet to be realized, however, because of the many problems associated with the development of tunable lead-salt diode lasers. Instead of lasing in a single gaussian mode, the output power of these devices is typically divided among a mixture of high-order modes, and the output exhibits a variety of noise problems related to multi-mode emission. Also, because of their tunability, the output frequency is sensitive to mechanical vibrations, small temperature fluctuations, and small variations in the diode laser injection current normally used to fine tune the frequency.

Aside from these difficulties, tunable diode laser heterodyne

spectroscopy has been successfully demonstrated in measurements of laboratory gases, atmospheric constituents and continua of the moon and Mars. Meanwhile, improvements have been made in the performance and reliability of tunable lead-salt diode lasers, particularly with regard to power output.

In view of the success achieved with gas laser heterodyne systems, and the demonstrated potential of the diode laser heterodyne technique, there is an obvious demand for a general purpose ground based tunable diode laser heterodyne spectrometer, for both astrophysical observations and for measurements of terrestrial atmospheric constituents.

1.1.2 Measurement of SiO in Sunspots

One very interesting class of observational problems that is ideally suited for TDL infrared heterodyne spectroscopy is the detection of molecular absorption in sunspots. Since temperatures in sunspots are lower than that of the surrounding photosphere, the environment is suitable for the formation of a variety of molecular species, some of which have already been observed. The sunspot environment is of itself interesting since it is not well understood and the roles of magnetic fields, macroturbulence and heating from the overlying chromosphere in determining the environment have not been well established. In addition, the sunspot environment constitutes an excellent high-temperature source for the study of species not easily created in the laboratory.

In order to properly analyse absorption lines observed in the

sunspot spectrum, it is essential to know the vertical temperature-pressure distribution throughout the region where line formation takes place. Unfortunately, wide disagreement exists between empirical one-dimensional sunspot temperature/pressure models that have appeared in the literature. The disagreement is largest in the upper layers of the umbral atmosphere where the formation of most atomic and molecular lines occur.

For the deeper layers, models employ continuum observations at both visible and infrared wavelengths. One difficulty that exists with sunspot continuum observations is contamination by terrestrial lines and the uncertainty in the transmittance of the terrestrial atmosphere. Since atmospheric transmission depends on local conditions at the time of measurement, these effects are removed by recording spectra from the spot and photosphere in close time coincidence and ratioing to obtain the relative sunspot intensity. Also, at visible wavelengths the ratio of photospheric to sunspot surface intensities is large ($I_{ph}/I_{um} > 10$), so that observations are contaminated by scattered light from the surrounding photosphere. Each continuum measurement must include some method of correcting for this effect. Attempts to model the upper layers of the umbral atmosphere have been based on observations of atomic lines at visible wavelengths where the sunspot spectrum is very complicated. This makes analyses of individual line shapes exceedingly difficult because of blending with adjacent lines.

Most of the above-mentioned problems are not as serious at

infrared wavelengths. The ratio I_{ph}/I_{um} decreases at longer wavelengths and the continuum spectrum shows discrete, isolated, well defined absorption features. Few continuum measurements have been made at wavelengths longer than one micron, however, and observations of infrared absorption features are almost non-existent. All infrared line and continuum measurements are thus of great importance in the study of the sunspot environment.

The fundamental ($v = 1-0$) and hot band ($v = 2-1$) transitions of silicon monoxide (SiO) near 8 microns are of special interest. These transitions have been predicted in the sunspot spectrum but no detection has been reported in the literature. Also, SiO is a sensitive thermometer at the upper layers since it is rapidly dissociated at sunspot temperatures. By comparing the observed equivalent widths against calculated values based on assumed temperature - pressure profiles it should be possible to determine the temperatures at small optical depths and provide boundary conditions for improved models. The tunable diode laser heterodyne technique is ideally suited to these observations for several reasons. The 8 micron SiO bands occur at the frequencies of present commercially available diode lasers. Also, since it is a coherent detection technique, the field of view is diffraction limited, with a diameter of about 3 arsec at 8 microns using a 1 meter telescope aperture. This beam size is smaller than most umbral diameters so contamination from surrounding photospheric light is minimal, at least during periods of good seeing. Since the technique is capable of subdoppler resolving power, it permits line broadening and

residual intensity measurements, in addition to the usual equivalent width information. Finally, the high surface brightness of sunspot umbrae (T ~ 4000K) makes SiO observations a logical first choice among the list of astrophysically interesting problems for the TDL heterodyne technique.

1.2 Progress in IR Heterodyne Detection Using Diode Laser Local Oscillators

Prior to the development of milliwatt semiconductor tunable diode lasers and wideband HgCdTe photodetectors, sensitivity limitations in diode laser heterodyne detection work required the use of strong signal sources. Success in early heterodyne studies was achieved by reporting observed self-beating effects in a single source or by detecting optical beats from two IR local oscillators. Hinkley et al. (1968) observed beating effects between a current tunable PbSnTe diode laser and the P(20) transition of a CO_2 gas laser. The beat note at approximately 1250 MHz was observed using a copper doped germanium (Ge:Cu) detector, with 40mW of CO_2 laser power and approximately 100 μW single mode power from the diode laser. The instantaneous linewidth of the diode laser emission was found not to exceed some tens of KHz, although frequency modulation of the diode output was observed due to a 120 Hz ripple in the current supply. These authors also observed heterodyning with the P(18) CO_2 transition at IF frequencies below 3.5 GHz, with somewhat poorer amplitude and frequency stability.

Heterodyne measurements of the Lorentzian spectral distribution of output power from a liquid helium-cooled PbSnTe tunable diode

laser were performed by Hinkley and Freed (1969) in an attempt to verify theoretical predictions for the diode laser line shape. Because of the intrinsically small cavity dimensions and low output power levels of TDL's, diode laser linewidths are significantly greater than those of gas lasers operating in the same frequency range, so that the gas laser line can be considered a delta function by comparison. By combining both diode and CO_2 gas laser radiation at a liquid helium cooled Ge:Cu detector and current-tuning the diode output frequency across the extremely narrow CO_2 laser line ($\Delta\nu < 100$ Hz), the resulting beat note was detected and the entire diode laser line displayed on a spectrum analyzer. The experimentally determined linewidths were found to agree with theoretical predictions based on estimates of the diode "cold" - cavity bandwidth and the degree of population inversion. The authors were also able to demonstrate the inverse power variation of the diode laser linewidth, by increasing the diode current and adjusting the heat sink temperature to maintain a constant beat frequency.

The first successfully operating infrared heterodyne spectrometer using a tunable semiconductor diode laser local oscillator was constructed by Mumma et al. (1975) and used to make laboratory heterodyne measurements of N_2O line profiles in the ν_1 band (100-000). These authors also demonstrated heterodyne detection of thermal blackbody emission from the moon and Mars. The spectrometer employed a PbSe semi-conductor diode laser, a HgCdTe photodiode as a photomixer and a spectral line receiver consisting

of 8 contiguous 25 MHz filters to provide a coverage of 200 MHz at the IF. At the operating wavelength of 8.6 microns, this corresponds to a resolving power of about 1.4×10^6. N_2O heterodyne absorption measurements were performed using a cell filled with N_2O at 10 torr pressure placed in the beam of a 1300 K blackbody continuum source. The N_2O line width was about 170 MHz, which was fully resolved in the bank of filters.

Blackbody heterodyne measurements of the moon and Mars yielded measured signal to noise ratios of 6.7 and 5.4, respectively, for a single 25 MHz filter and roughly 8 minutes integration time. Much of the signal degradation was the result of insufficient local oscillator power. The maximum single mode power obtained in this experiment was ~100 μW, considerably below that required for shot noise limited operation (i.e., the point where the local oscillator shot noise dominates all other noise sources at the IF). From the measured photodiode quantum efficiency and diode laser single mode power, the overall transmission of atmosphere and system optics was estimated to lie between 10 and 20 percent.

Frerking and Muehlner (1977) developed a diode laser heterodyne spectrometer and obtained laboratory and terrestrial atmospheric O_3 absorption spectra in a 1 cm^{-1} region near 1011 cm^{-1}. This instrument employed a PbSnSe laser diode and HgCdTe photodiode mixer both mounted, together with the optical elements, in compact fashion on a common dewar operating at liquid nitrogen temperature (77K). A single IF amplifier and filter at the mixer output with a cutoff of 100 MHz provided a single IF channel with a width of 100 MHz,

yielding an IR resolving power of about 1.5 X 10^5. Spectral scans of the ν_3 band (001 ← 000) of atmospheric ozone in solar absorption were obtained by slowly current-tuning the diode between 1010.9 and 1011.8 cm^{-1} while recording the synchronously detected heterodyne signal on an X-Y plotter. Excellent agreement was observed between the heterodyne spectra and direct absorption laboratory spectra using an ozone absorption cell. The results also agreed nicely with synthetic O_3 spectra in this region computed by the same authors, and adjusted to match the instrument frequency resolution. The measured SNR was found to be about 25 for a 100 MHz IF bandwidth and 1 sec integration time. This is roughly a factor of 4800 smaller than the ideal value. The degradation was attributed to optics transmission losses, insufficient local oscillator power, amplifier noise and spurious IF noise.

The first use of a closed cycle cooler for continuous diode laser cooling and temperature control in a heterodyne system was reported by Ku and Spears (1977). These authors constructed a heterodyne radiometer using a tunable 10.6 μm PbSnSe diode laser as a local oscillator. The closed cycle cooler held the diode near 30K for the duration of the experiments. A heterodyne signal was obtained by combining the local oscillator beam and that from a 1173K blackbody source at a HgCdTe photomixer. The heterodyne detection electronics consisted of a 600 MHz low noise amplifier at the mixer, a spectrum analyzer, crystal detector and lock-in amplifier as part of the synchronous detection process. The instrument was tuned by varying both cooler temperature and the

diode current. Initial studies of noise at the IF suggested that self-beating effects in the laser output were the cause of extreme noise in the heterodyne signal over portions of the TDL tuning range. For quiet operation, signal-to-noise ratios as high as 150 were observed for a blackbody temperature of 1173K, a bandwidth B of 500 MHz and an integration time τ of 0.4 sec. In addition one of the laser modes was examined by heterodyning against a CO_2 gas laser, and an instantaneous linewidth of 100 kHz was measured, in qualitative agreement with earlier results of Hinkley and Freed (1969). The instrument was also used to measure several absorption lines of ethylene near 942 cm^{-1} in a gas absorption cell. The lines were first located in direct absorption and then observed in the heterodyne mode. The measured line widths were in good agreement with those expected for an IF resolution of $\sim$ 600 MHz.

A hybrid spectroscopic technique employing both direct absorption and heterodyne measurement options available using diode lasers has been employed by several workers for the study of molecular gases in the laboratory. By monitoring the beat frequency between a PbSnSe diode laser and various CO_2 gas laser lines, Worchesky and coworkers (1978) were able to establish precise line center positions of several D_2O absorption lines in near coincidence with CO_2 gas laser transitions. Both the diode and CO_2 laser beams were focussed onto a HgCdTe photomixer while monitoring the heterodyne signal to establish a frequency calibration scale. The remaining part of the diode laser radiation was passed through the D_2O absorption cell and onto a second detector where the

transmittance was simultaneously monitored. A total of nine lines in the ν_2 fundamental bands of D_2O were measured with an estimated error of ± 0.005 cm^{-1}. Instabilities in the diode laser output frequency were the largest contributor to this uncertainty.

Hillman et al. (1979) used a similar approach to study the 2sQ(1,1) transition in the $2\nu_2-\nu_2$ band of NH_3 near 949.5 cm^{-1}. The spectrometer employed a PbSSe tunable diode laser mounted in a closed cycle cooler and emitting near the 9.4 μm and 10.6 μm CO_2 laser bands. Examination of the TDL output on a spectrum analyzer revealed an anomalously large mode width of $\sim$ 15 MHz resulting from radiation losses in the diode stripe, diode injection current noise and cooler instabilities. The mode center position stabilized to within ±2 MHz over several minutes of time after the diode temperature had reached equilibrium. A small additional mode was also seen in the observed heterodyne spectrum about 60 MHz from the main mode. The presence of this type of satellite mode in the TDL output was thought to be the cause for regions of extreme excess noise of the type observed in heterodyne studies by Ku and Spears (1977). After an initial study of instrument performance, a number of scans of the 2sQ(1,1) line were made using a 30-cm cell of NH_3 at 1-Torr pressure placed in one of the divided TDL beams. Aside from small departures from linearity in the TDL tuning rate, data from 12 scans yielded a determination of the line center position to within $\pm$ 0.00007 cm^{-1} or ± 2.1 MHz. Spears and Freed (1973) examined the use of IR - microwave varacting for continuous heterodyne coverage between adjacent CO_2 transitions near 10.6 microns. This technique

also permits accurate wavelength calibration of a tunable diode laser using a CO_2 laser as a reference. Using a tunable microwave local oscillator pump and a HgCdTe detector, it is possible to detect optical beat frequencies much larger than the single conversion IF bandwidth of the detector. By injecting the output of a microwave pump with frequency f_p into the mixer, the optical beat frequency (f_2-f_1) of two IR sources is down converted to a final frequency $f = nf_p - (f_2 - f_1)$ which lies well below the nominal detector bandwidth. Spears and Freed (1973) measured a signal-to-noise exceeding 45 dB at 21.3 GHz with a 10 kHz noise bandwidth using two CO_2 lasers operating at different transitions and dellivering $\sim$ 0.5 mW each to the detector. The signal-to-noise ratio was limited mainly by the available microwave local oscillator power. Subsequent 3rd and 4th harmonic mixing studies (n = 3,4) resulted in the detection of beat frequencies between 40.3 GHz and 60.7 GHz.

The most recent, and probably the most successful attempt to employ TDL heterodyne spectroscopy for the study of atmospheric constituents was made during 1979 by Harward and Hoell (1980). The authors observed atmospheric NH_3, H_2O, O_3, CO_2 and HNO_3 absorption between 9 and 11 microns using a TDL radiometer and a small heliostat located at NASA Langely Research Center, Hampton, VA. The system employed a TDL local oscillator in a closed cycle cooler, the output of which was collimated using an f/1 Ge lens. Mixing of the LO and signal beams was accomplished with a HgCdTe photomixer. Coarse frequency calibration was performed using a monochromator,

and prior to each observation the final local oscillator position was set using reference gas line positions and a solid Ge etalon to provide a frequency scale. One interesting operational feature of the instrument was the use of a scanning etalon to select and track a single longitudinal diode laser mode.

The atmospheric spectra were recorded by varying the TDL current to sweep a single 100 MHz IF filter in frequency. Signal-to-noise ratios of 200 to 300 were obtained during these measurements. This study established the capability of the TDL heterodyne technique for terrestrial atmospheric studies between 9 and 11 microns.

1.3 Tunable Diode Laser Development

Two major classes of tunable semi-conductor diode lasers exist today - the first based on the III-V binary compound GaAs, where lasing typically occurs in the 0.6-5 micron spectral range. The Roman numerals refer to the number of valence electrons for the constituent atoms and defines the role - donor or acceptor - of the constituent. The state-of-the-art has advanced rapidly for this class of device largely because of their potential for optical communication applications. Present GaAs devices are very reliable and exhibit multi-milliwatt power output, high efficiency, beam uniformity and room temperature operation.

This section will be concerned exclusively with the second major diode laser class which is based on lead-salt, or IV-VI compounds. Interest in these devices stems from their potential for

wide tunability, their operation in the important 3-30 micron range (which includes the vibration-rotation bands of many molecular species), and their extremely high spectral purity (less than 10^{-4} cm^{-1} at $\nu \sim 1000$ cm^{-1}).

Butler et al. (1964) first reported lasing in the Pb-salt semi-conductor family using a PbTe (lead-telluride) diode laser operating at 6.6 microns. Lasing throughout the 2.7 to 33 micron region was later achieved after the development of ternary compounds employing basic binary constituents.

The nominal wavelength operating range of a Pb-salt diode laser is obtained during fabrication by the proper choice of material, which in turn determines the energy bandgap. Figure 1 shows the variation of energy bandgap E_g for various ternary Pb-salt compounds and the nominal operating wavelengths of these devices. Because of the effect known as band inversion, E_g approaches zero as x approaches 0.4 in the compound $Pb_{1-x}Sn_xTe$. In principle, this allows fabrication of very long wavelength devices. Several methods also exist for fine tuning the operating frequency of these devices. The output frequency can be smoothly varied by changing the applied pressure, the surrounding magnetic field, the injection current flowing through the device or the mount temperature. The latter two techniques have proven the most popular since they are easily performed under field operating conditions.

Early tunable diode lasers (Figure 2a) were broad area homojunction devices employing a P and N type material and a narrow active region. This region acts both as a barrier to confine

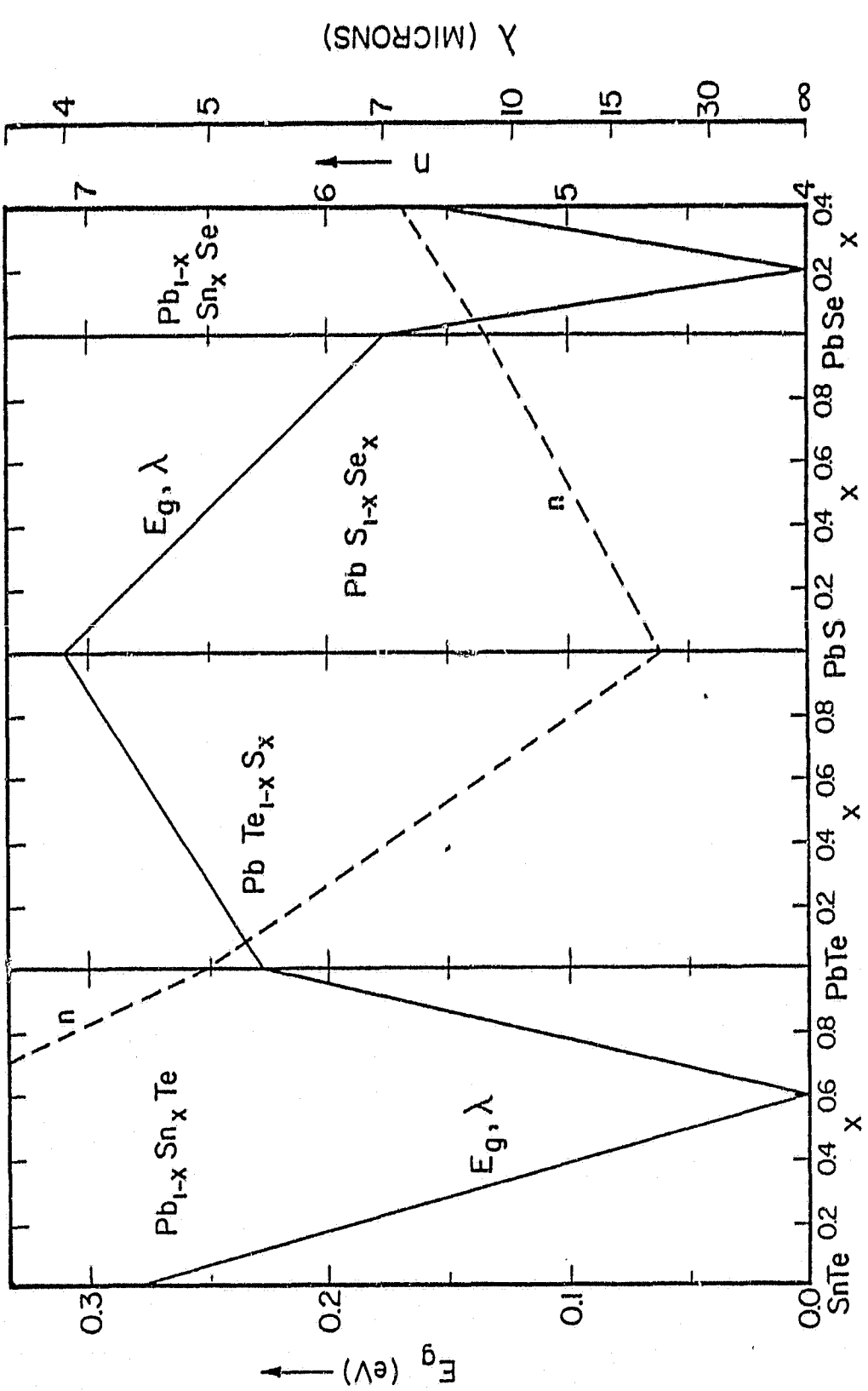


Figure 1. Bandgap (E_g) and refractive index (n) of $Pb_{1-x}Sn_xTe$, $PbTe_{1-x}S_x$ and $Pb_{1-x}Sn_xSe$ at 77K as function of composition (Preir, 1979).

Figure 2. Innovations in tunable lead-salt diode laser development.

minority carriers (the bandgap changes at the sides of the junction) and as a radiation waveguide (a change in the refractive index occurs at the boundaries). The front and back surfaces are cleaved parallel and sometimes polished to define a Fabry-Perot cavity. The sides are roughened to minimize reflections and the generation of unwanted high-order modes.

The operating spectral range of TDL's attained by temperature and bias current tuning can be specified, but the operating frequency range is much more restrictive because of mode hopping effects. The resonant mode frequencies in a device of cavity length L and index of refraction n are given by the Fabry-Perot condition

$$\nu_{res} = m/2nL \qquad cm^{-1} \qquad (1.1)$$

The cavity is usually many wavelengths long so "m" is typically a large integer. When the diode is fine tuned by varying the injection current, changes in the Joule heating of the material alter the temperature dependent index of refraction and hence the optical length. The effect of thermal expansion on the diode tuning is negligible by comparison. By subtracting Equation 1.1 from itself for adjacent "m", one obtains an expression for the frequency separation $\Delta\nu$ of adjacent modes,

$$\Delta\nu = \left[2nL \left(1 + \frac{\nu}{n} \frac{dn}{d\nu}\right)\right]^{-1} \qquad (1.2)$$

Typical dimensions for lead-salt diode lasers are 400 μm length and 200 μm in width and height, with typical output mode separations of $\sim 2\ cm^{-1}$. Changing the device temperature also tunes the energy bandgap E_g in the material, typically several times more rapidly than ν_{res} for Pb-salt TDL's. This differential motion changes the preferred laser mode when the diode is current tuned, sometimes causing a mode "hop", and very often results in large frequency gaps over which no lasing occurs. In general the gain envelope, as determined by E_g at the operating temperature, has a complicated shape and is wide enough to support several modes. This results in multi-mode output (Fig. 3) and necessitates provisions for external mode selection in many applications. The intent in fabricating a TDL is usually the generation of a single, powerful longitudinal mode having TE polarization. Most performance problems are related to poor optical confinement in the laser cavity which creates higher order unwanted modes. These modes share the available power between them, which causes an irregular output power pattern and limits the tunability. Early ternary devices had high threshold currents, lased only at the lowest temperatures and operated with parasitic bounce modes involving all sides of the laser chip.

A substantial improvement in parasitic mode suppression was realized using a stripe geometry for improved carrier confinement and mode guiding. In Figure 2b, the n layer formation is restricted to a narrow stripe on the p-type substrate using an SiO_2 diffusion mask, and a metallized contact is finally applied. The mode guiding cavity along the junction direction is bounded by the stripe edge,

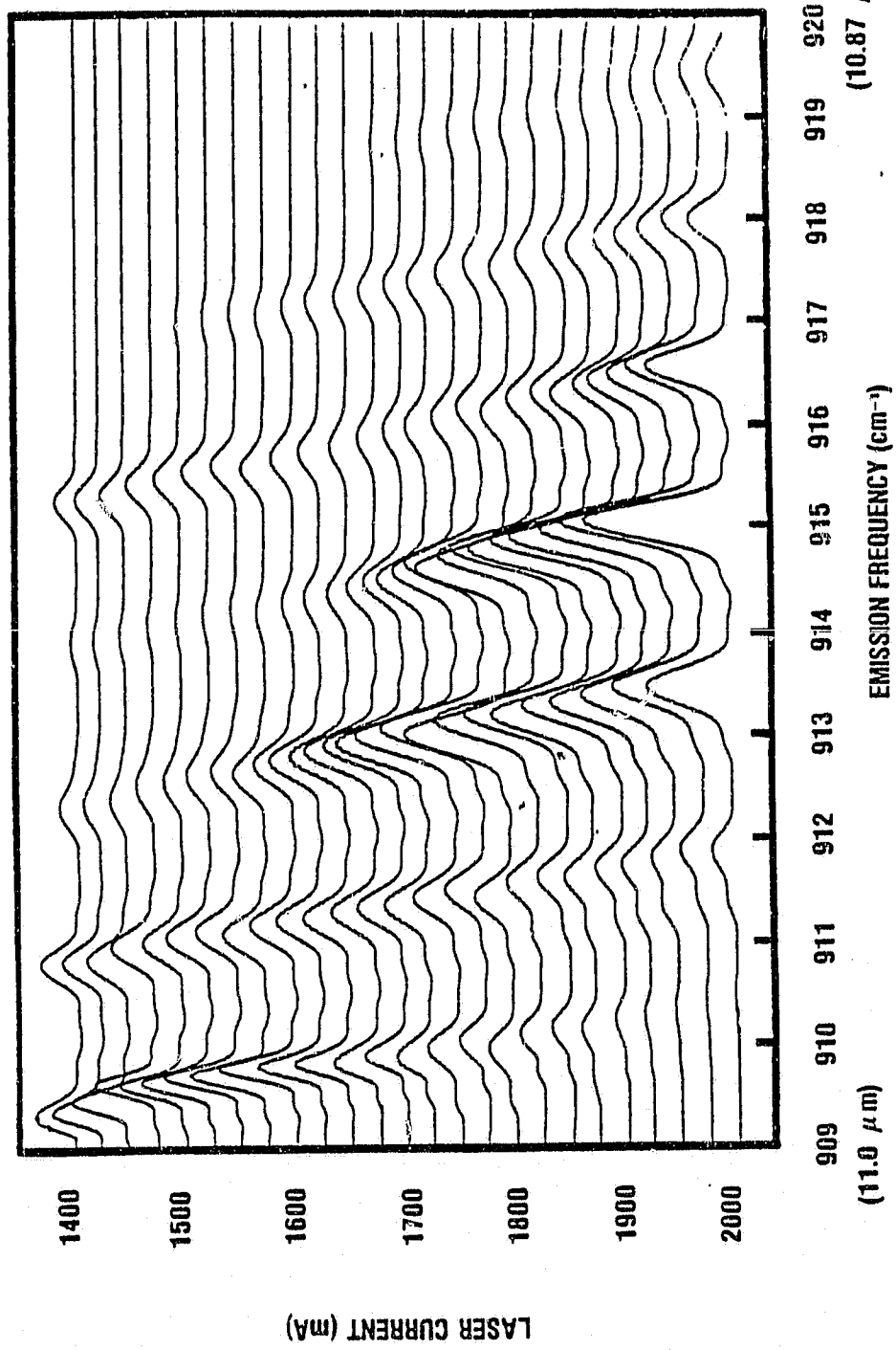


Figure 3. Variation of emission spectra with bias current for an 11 μm $Pb_{.97}Sn_{.03}Se$ diode laser at 15K. (Courtesy of Spectra-Physics Inc., Laser Analytics Division).

and sidebounce modes are effectively attenuated by losses in the bulk material. The first use of a stripe geometry was reported by Ralston et al. (1973), fabricated using $Pb_{0.88}Sn_{0.12}Te$. Sidebounce modes were effectively suppressed and the emission spectra of the stripe consisted of regularly spaced frequency modes corresponding to a well defined longitudinal cavity.

In addition to bounce-mode suppression, the performance of a TDL also depends on the degree of confinement of the injected carriers in the direction of the carrier drift, perpendicular to the junction. This is accomplished by the formation of heterostructures or heterojunctions. These confine the carriers and the emitted light with epitaxially grown regions of wider energy gap. This decreases the lasing threshold current and reduces the resistive heating, permitting operation at higher ambient temperatures. This result is very significant since TDL operation at or above liquid nitrogen temperature (77K) simplifies the diode cooling apparatus. Single heterojunction devices formed by vacuum deposition (Walpole et al., 1973) and by compositional interdiffusion (CID) and liquid-phase-epitaxy (LPE) (Linden et al., 1977) exhibited low threshold currents and high operating temperatures, showing improvements in internal quantum efficiency. These features inspired the development of double heterostructure lasers for high temperature operation. Groves et al. (1974) achieved high power ($\sim$1.2mW) single mode cw performance in a stripe geometry diode laser operating at 77K. This is shown in Figure 2c. The active region was formed by LPE growth of n-type PbSnTe through 50 μm wide stripe

openings in a MgF_2 mask. Diffusion into the PbTe substrate was prevented by doping the substrate with Tl, and the top border of the ~10 μm wide active region was defined by the interface with a final PbTe layer.

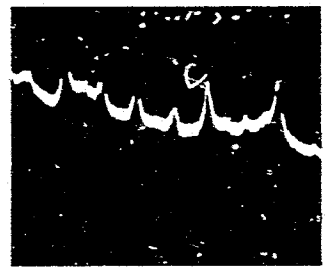
The fabrication of heterostructures is hampered by defects that form in the heterojunctions and in the surrounding material. These result from the difference in lattice parameters between adjacent layers. Fonstad et al. (1980) have successfully dealt with this mismatch problem using multi-layer ternary compounds having nearly identical lattice parameters. The threshold current (at 80K) of these devices is a factor of ~20 smaller than that for the PbSnTe/PbTe structure, and pulsed operation has been achieved to ~166K.

As discussed earlier, there is a tendency for a "mode-hop" to occur since the peak optical gain in the cavity tunes more rapidly with temperature than do the Fabry-Perot modes. This occurs even for a well defined cavity where bounce modes are suppressed and only the fundamental transverse mode exists. By replacing the Fabry-Perot end mirrors with a distributed feedback (DFB) cavity it is possible to greatly suppress this mode-hopping tendency (Walpole et al., 1976, 1977; Hsieh and Fonstad, 1979). This arrangement (Figure 2d) employs a weak periodic variation in the refractive index of the active region along the direction of wave propagation, with one end of the cavity beveled to suppress the generation of Fabry-Perot modes. In a DFB device the longitudinal modes are determined by the periodicity along the cavity and the single

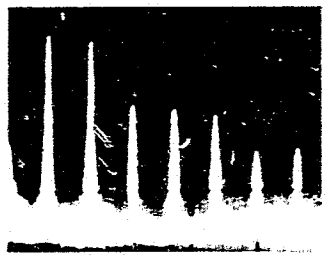
allowed mode will be the one closest to both the Bragg frequency and the center of the optical gain envelope. Hsieh and Fonstad (1979) achieved single mode emission continuously tunable over ~20 cm^{-1} near 12.8 microns. This level of performance was attained without incorporating some of the previously discussed innovations (e.g., lattice matched layers), so future efforts in DFB development should yield extremely wide continuous tunability.

For heterodyne and other high signal to noise applications of TDL's, the level of excess noise present in the output is an important consideration. Several broad classes of diode excess noise related to multi-mode emission exist. These extend in frequency from several Hz to over 1 GHz and have been studied by various authors (Ku and Spears, 1977; Jennings and Hallman, 1977; Eng et al., 1979; Harward and Hoell, 1979). The noise sources include broadband noise, noise generated by mechanical vibrations associated with closed cycle cooler operation, uncontrolled feedback from optics external to the laser cavity, and the effects of mode competition within the diode. The various classes of high frequency TDL noise as documented by Harward and Sidney (1980) are shown in Figure 4. Most restrictive in heterodyne applications are the spiking resonance symptom (b) which seems to be associated with the heterodyne beating of spurious frequency modes, and broadband excess noise (e) covering the entire IF spectrum. Because of these effects, only a small fraction of the nominal diode tuning range is suitable for heterodyne work.

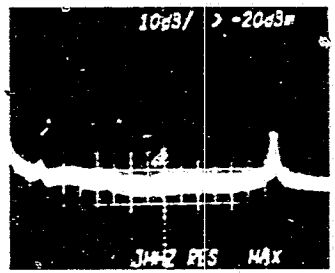
a) low frequency noise
(0-90 MHz)



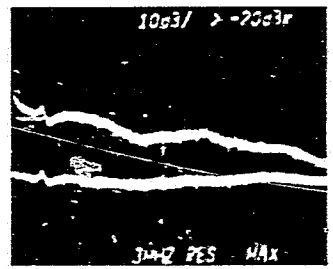
b) harmonic spiking
resonance
(0-1200 MHz)



c) optical feedback
induced harmonic
resonance
(0-500 MHz)



d) shot noise induced
spiking resonance
or
self-pulsations
(0-1200 MHz)



e) broadband noise
(0-1200 MHz)

Figure 4. Categories of diode laser excess noise (Harward and Sidney, 1980).

1.4 Previous sunspot Measurements and Models

Most previous identifications of molecular absorption features in the sunspot spectrum have led to statements concerning the temperatures and pressures in sunspots. At the very least, attempts to observe molecules in spots imply some assumption about the ambient temperature. The converse is also largely true. Modelers of the temperature and pressure distribution in sunspots employ data on the size of molecular features along with continuum intensity measurements at several wavelengths. These data provide boundary conditions which, together with the rules for model stellar atmosphere construction, allow the computation of the vertical temperature and pressure structure. A review of sunspot models, then, must also include a description of sunspot observations.

Wohl (1971) made one of the first synoptic studies of umbral features. The observations were made in 3 large umbrae during June and July of 1969 and covered the region from 400 to 800nm. The instrument used was a photoelectric spectograph with a frequency resolving power of ~300,000. Previous published molecular concentration studies were used by Wohl as a ledger in searching for spectral lines. Because of photospheric scattered light and the effects of line blending, many of the lines could not be confirmed. Following a search for some 15,000 features, the following species were found to be present based on agreement with their expected positions and strengths: C_2, CH, CN, CaH, MgH, TiO, CoH, HF, NH, and NiH. Tentative detections of the following species were also reported: AlH, BO, BeH, AlF, BF, BH, BaO, CuH, FeO, LaO, MgO, ND,

ScO, VO and ZrO. Because of a lack of good molecular constants, it was not possible to accurately predict the frequencies and intensities of selected lines. This limited the number of identified molecules. It was also not possible to observe many weaker features since the instrument signal-to-noise ratio limited the detectable equivalent width to $> 1 \times 10^{-4}$nm.

Two of the early one-dimensional sunspot models, employing both infrared continuum measurements and observations of medium and strong atomic lines in the visible spectrum, were published by Henoux (1968, 1969). Observations of umbral continuum intensities in a large spot (r >10 arc sec) were made at seven wavelengths between 0.86 and 2.2 microns using a 9 meter spectrograph and 50 cm telescope. Defining $\tau_{0.5}$ as the 0.5 micron continuum optical depth, Henoux's (1968) model for $\tau_{0.5} > 0.3$ was computed from these measurements by inverting the radiative transfer equation. Additional atomic line observations measured between 470 and 640nm at a dispersion of 0.02 nm/mm were reported by Henoux in 1969. These enabled him to extend the model to roughly $\tau_{0.5} \sim 10^{-4}$. The temperature structure for $\tau_{0.5} < 0.3$ was adjusted to fit wings of observed Na D lines which change very rapidly with temperature. After obtaining a plausible temperature profile, the measured equivalent widths for unblended lines of FeII, CrII, TiII, FeI, CrI and TiI were used to check for agreement with the model by varying the microturbulence parameter. A microturbulent velocity of 1.5 km/sec yielded good agreement with observations. All of the line intensity measurements could then be corrected for stray light by

computing expected visible continuum intensities, based on the more reliable infrared values.

Only a few detections of sunspot molecules have been made at infrared wavelengths. (Hall and coworkers at Kitt Peak). The 4 micron SiO sunspot spectrum was examined by Hall in 1969 but the results remained unpublished (Hall, 1981). Hall and Noyes (1969) did report measurements of several umbral HF and comparison CO lines in both umbra and photosphere near 2.4 microns. These observations were made at the McMath solar telescope using a vacuum spectrograph with a resolution of 150,000. Although the intent was to estimate the solar fluorine abundance, obtaining this quantity from the observed absorption spectrum requires a fairly accurate temperature model, since hydrogen fluoride dissociates rapidly at sunspot temperatures. Effective temperatures for each spot were obtained by measuring the ratio ϕ_u of spot to adjacent photosphere intensity and comparing with absolute photospheric continuum estimates and model limb darkening relations. The temperatures deduced for two large spots using this approach was 4000 ± 200K and for the small spot, 4400 ± 250K. The authors generated sunspot models consistent with the observed effective temperatures using the SAO stellar atmospheres and line synthesis program. The observed CO lines have widely different lower level energies so they were used as selection criteria for the choice of the best temperature model. After adopting a suitable model, the fluorine abundance was varied to obtain agreement with the measured equivalent widths. Hall and Noyes deduced a fluorine abundance (A_F) of $\log_{10} A_F = 4.56 \pm 0.33$ on a

scale where $\log_{10} A_H = 12$.

Stellmacher and Wiehr (1970) tested several previously published models, including Henoux (1969) using photographic and photoelectric measurements of 4 magnetically non-split lines, which were mostly free of blends. The authors then constructed a new model for the upper layers. Lines of Fe λ406.5402 nm, 543.4527 nm, Fe^+ λ722.4464 nm and Ti λ571.3895 nm were measured and continuum intensity measurements were made in line-free windows at λ461.5 nm, 542.7 nm, 615.7 nm, 703.5 nm and 778.7nm for comparison purposes. Parasitic light corrections for these measurements were estimated using the ratio of observed lines of Fe λ614.925 nm, which is weakened in spots, to V λ615.015, which becomes stronger.

Stellmacher and Wiehr computed wavelength and center to limb variations of the continuum as well as profiles of 3 of the measured non-split lines using published sunspot models and assuming LTE. Best agreement with both the continuum and line measurements occurred for the Henoux model, which was adopted for $\log \tau_{0.5} > 0$. The authors also cited the previous range of sunspot microturbulent velocity determinations (1.1 - 2.9 km/sec, Bray and Loughhead (1964); 1.8 km/sec photosphere, 1.2 km/sec umbra, Bruckner (1965)), and quote a measured value of < 1.0 km/sec by fitting the model to the Fe λ406.54 line. A value of ~0.5 km/sec, which is considerably smaller than most previous solar microturbulent velocity determinations, was deduced from comparison with the Ti and Fe λ543.45 nm features.

An attempt was made by Webber (1971) to construct a one

dimensional sunspot model using a large body of visible absorption features obtained from 4th and 5th order spectrograms taken at Mt. Wilson. Since the effects of scattered light varied greatly during the measurements the emphasis was on a rotational analysis of the lines within each spectrogram. Equivalent widths of several hundred observed lines of TiO, MgH and CaH were measured as part of Webber's rotational analysis. The CaH analysis was complicated by the presence of excessive line blending and terrestrial atmospheric absorption. Different rotational temperatures were deduced for each species, allowing an assessment of the vertical temperature structure when Zwaans (1968) sunspot model was used to approximate the depth of line formation for each molecule. The final temperature profile was obtained by applying a small perturbation to the sunspot model of Henoux (1968). Some disagreement seems to exist between Webber's proposed model and the effective temperatures obtained from the very data used to construct the model. Also, the use of two separate previously published models in the analysis (Zwaan, 1968; Henoux, 1968) suggests a possible lack of self-consistency in the results.

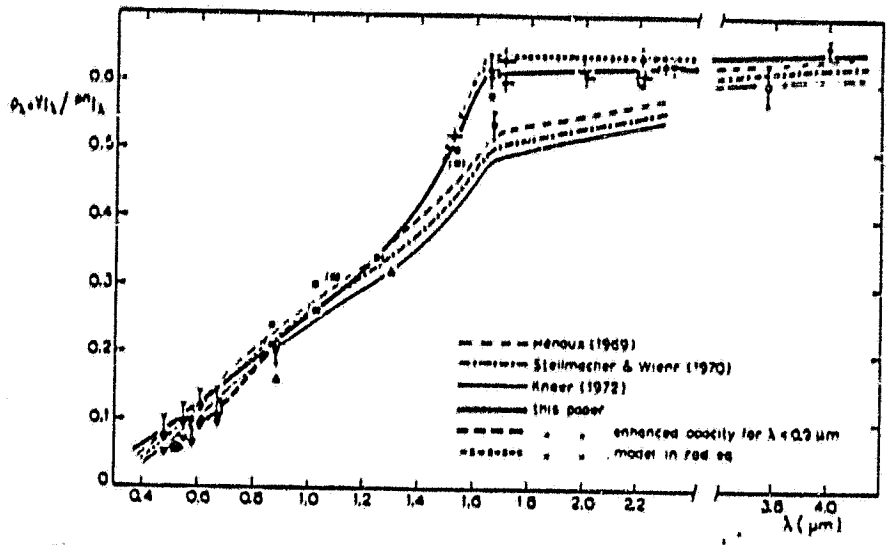
Kneer (1972) published a sunspot model based on both photoelectric measurements of the visible sunspot continuum at two wavelengths, and high resolution photographic spectra of 3 non-split FeI lines (λ512.37, 543.45 and 557.61nm) in 3 large umbral regions. These observations were corrected for stray light by extrapolating aureole intensities measured photoelectrically. After adjusting a hydrostatic equilibrium model to force the computed lines and

continuum intensities to agree with the observations, the author produced a model in satisfactory agreement with the results of Stellmacher and Wiehr (1970). Since the wings of the observed FeI lines are only weakly pressure dependent, the model pressures were uncertain by a factor of ~2, so small departures from hydrostatic equilibrium could not be ruled out. This study suggested a microturbulent velocity of < 1 km/sec, in rough agreement with Stellmacher and Wiehr (1970), and an estimate of ~1 km/sec for umbral macroturbulence.

Some of the supporting measurements used by the previously cited authors were summarized by Zwaan (1974) in an attempt to; (i) explain discrepancies between published models and (ii) construct a semi-empirical continuum model atmosphere which reproduces measured continuum intensities between 0.48 and 2.3 microns. This study emphasized IR continuum measurements between 1.5 and 4 microns. Supporting observations were taken from Henoux (1968), Hall (1970), Coupiac and Koutchmy (1972), and Maltby and Staveland (1971). The overlapping measurements of Henoux (1968) and Hall (1970) were basically in agreement, lending support to the model in the deepest layers, near the 1.7 micron H^- opacity minimum. Additional visible and near IR photometric measurements were used to determine the run of temperature in the shallower layers. The model was fit to the visible data using a wavelength dependent opacity enhancement factor, since line blanketing is a dominant source of error at visible wavelengths.

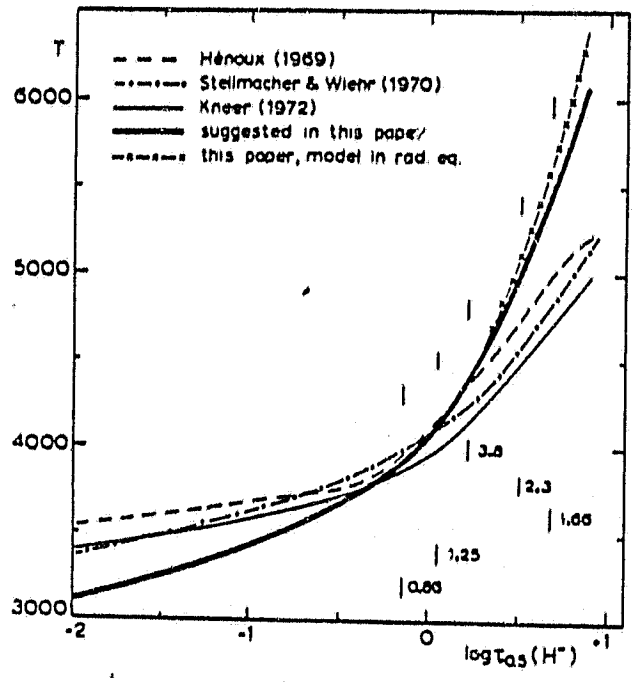
Figure 5 shows Zwaan's observational summary and the variation

a)



Observed and computed relative umbral continuum intensities near the center of the disk as a function of wavelength.

b)



Temperature vs monochromatic optical depth at $\lambda = 0.5$ μm for some umbral model atmospheres. The positions of unity optical depth at various wavelengths are shown by vertical bars, the wavelengths are indicated in μm.

Figure 5. Summary of sunspot continuum measurements and published umbral models prior to 1974 (Zwaan, 1974).

of surface brightness with λ, calculated using several previously cited models. Figure 5b shows the models themselves and the discrepancies, which are minimized near log $\tau_{0.5}$ = -0.3 to 0.0. Zwaan's result is consistent with radiative equilibrium for $\tau_{0.5}$ < 1.7 and departs to smaller lapse rates for $\tau_{0.5}$ > 2. This model is considered reliable between 0.5 < $\tau_{0.5}$ < 2 but has some limitations. The run of temperature is given out to $\tau_{0.5} \sim .01$, but the data give no check on the model for $\tau_{0.5}$ < 0.5. The author also points out the existence of a temperature inversion somewhere at small optical depths because of lateral radiative influx from the photosphere and heating from the chromosphere above the sunspot fluxtube. The models of Henoux (1969) and Stellmacher and Wiehr (1970), on the other hand, have some experimental justification at smaller optical depths, since they are based on fairly strong atomic line observations. The elevated temperatures in these models probably show the effects of heating from adjacent regions.

Zwaan's (1974) model temperatures were also criticised as being too high for τ > 1.5, since the Harvard Smithsonian Reference Atmosphere (Gingerich et al., 1971) used to convert the measured umbral intensities to brightness temperatures, was also considered to be too high (Vernazza et al., 1974). In response to this, Zwaan revised the deeper layer temperature structure, lowering the lapse rate to just above the adiabatic value (Zwaan, 1975). It should be pointed out that this change does not alter the validity of the IR observations used by Zwaan; only the conversion of the observations to umbral temperatures are affected.

Hall and Noyes (1972) observed a number of rotation - vibration lines of $H^{35}Cl$ in the sunspot spectrum as part of an effort to map the 2400 - 3000 cm^{-1} region. The measurements were made at the Kitt Peak McMath solar telescope using an IR spectrometer with a resolving power $\nu/\Delta\nu > 10^5$ and a signal to noise ratio > 100. The frequency uncertainty for the measurements was $\sim .005$ cm^{-1}. The authors detected 10, v=1-0 transitions and 4, v=2-1 transitions for $H^{35}Cl$. Equivalent width measurements were made for all except 4 of the v=1-0 transitions, which were blended with adjacent lines. The results yielded a solar chlorine abundance $\log_{10} N(Cl) = 5.5 \pm 0.4$ for both isotopes, on a scale where $\log_{10} N(H)=12$.

Gaur and coworkers published a number of separate studies investigating the presence of selected molecules in the sunspot environment. By performing dissociation equilibrium calculations using Zwaans (1965) sunspot model and using published estimates for elemental abundances and dissociation coefficients, Gaur et al. (1973) produced a search list containing the expected concentrations of a number of molecules in sunspots. The average formation depth for molecules included in these equilibrium calculations was also determined. The molecules considered include those already observed plus many not yet observed at the time of publication (1973). Table 1 is a reproduction of the search list published by Gaur et al. The list does not include CoH and NiH discovered by Wohl (1971), since Gaur did not have dissociation constants for these molecules.

Gaur (1976) also computed the equivalent widths for selected $\Delta v=1$, R branch transitions of NO near 5.3 microns in sunspots. The

TABLE 1

Survey of Molecules Expected in Sunspots

(Gaur et al., 1973)

Sl. No.	Molecule [a]	log f [b]	$\bar{z}$ [b] Km	$T(\bar{z})$
1	VO	14·64	43·5	3277 K
2	TiO*	15·97	45·8	3288
3	CaO	15·19	47·6	3294
4	YO	16·11	52·6	3320
5	PO	16·33	56·5	3338
6	AlO	13·11	61·2	3359
7	H_2O*	20·28	62·7	3365
8	CO_2	16·70	69·6	3398
9	SO	16·76	71·0	3405
10	BO	15·17	74·9	3424
11	H_2S	16·94	81·3	3462
12	ZrO*	14·90	81·4	3465
13	O_3	16·93	81·8	3467
14	MgO*	16·30	83·9	3475
15	S_2	13·16	84·5	3480
16	HCl	17·68	84·7	3483
17	OH*	20·58	86·1	3490
18	H_2	23·88	86·2	3495
19	CaH*	15·35	87·5	3500
20	NaH	14·02	89·3	3508
21	SiO	19·78	90·3	3517
22	AlH*	16·37	91·2	3522
23	TiO_2	15·18	94·5	3538
24	PN	15·72	95·1	3547
25	SH	18·63	97·1	3560
26	N_2	20·20	101·9	3588
27	HF*	17·19	104·5	3605
28	NH_3	15·07	105·0	3608
29	NO	17·90	106·2	3615
30	PH	16·31	107·7	3622
31	MgH*	17·94	113·6	3672
32	CO*	21·38	116·0	3675
33	NS	15·09	117·6	3683
34	NH_2	16·55	122·5	3713
35	HCO	16·90	130·7	3773
36	SiH*	18·06	139·9	3835
37	NH*	18·47	144·8	3880
38	HCN	16·30	146·2	3892
39	CH_2	16·51	152·3	3957
40	CS	16·37	153·0	3970
41	CN*	17·90	154·3	3015
42	CH*	18·11	203·0	5910
43	C_2*	16·15	206·8	5985

[a] The asterisked species are those discovered prior to 1974.

[b] The abundance I is the total number per cm^2 above optical depth $\tau_{0.5} = 2.73$. $\bar{z}$ is the average formation depth.

computations were done for J ~ J_{max} = 27 using the sunspot models of Stellmacher and Wiehr (1970) and Henoux (1969). The dependence of the computed equivalent widths (W) on solar disk position was also examined by calculating W at solar zenith angles (μ) of 1, 0.7, 0.5 and 0.3. For each model, W was found to increase with μ but systematically larger W's were computed using Henoux's model. The computed values ranged from $10 < W < 30 \times 10^{-4}$nm.

A similar study (Gaur et al.,1978) was performed for SiO by calculating equivalent widths in the 8 micron sunspot spectrum for v=1-0 and v=2-1 P and R branch transitions near J_{max}=43. These calculations were done with Zwaans (1974) sunspot model using published values for the SiO dissociation energy and equilibrium constant. The $\tau_{0.5}$ formation depth for all of the computed lines was about 0.063 and the equivalent widths ranged from 40-95 $\times 10^{-4}$nm as D_o was varied from 8.1 to 8.33 eV. The computed widths were also not very sensitive to the lower level vibrational quantum number. The line formation depth and temperature sensitivity for dissociation, however, show the usefulness of SiO measurements as a means of determining the temperature structure at higher layers. This study by Gaur et al. formed the incentive for the SiO observations discussed in Chapter V.

CHAPTER II

THEORY OF INFRARED HETERODYNE DETECTION

2.1 The Basic Heterodyne Detection Process

The theory of infrared heterodyne detection is well understood and has been reviewed by a number of authors (Blaney, 1975; Cohen, 1975; Abbas et al., 1976; Betz, 1977). The intent here is to provide a brief but concise review of these fundamentals as they apply to an operating heterodyne system.

Figure 6 shows the basic IR heterodyne process in schematic form employing an infrared local oscillator and a monochromatic IR signal source. Coherent detection occurs at a high speed photodetector/mixer where the output current varies as the square of the total electric field incident on the detector. If $\vec{E}_{LO}(x,y,t)$ and $\vec{E}_S(x,y,t)$ are the complex local oscillator and signal electric field contributions at a detector of active area A_D, the intermediate frequency (IF) current $i_D(t)$ from the detector will be

$$i_D(t) = C \iint_{A_D} \vec{E}^*(x,y,t) \cdot \vec{E}(x,y,t)\, dxdy \qquad (2.1)$$

where

$$\vec{E}(x,y,t) = \vec{E}_{LO}(x,y)\, \exp(-i\omega_{LO}t) + \vec{E}_S(x,y)\, \exp(-i\omega_S t)$$

and C is a constant. The surface integral in Equation 2.1 reveals

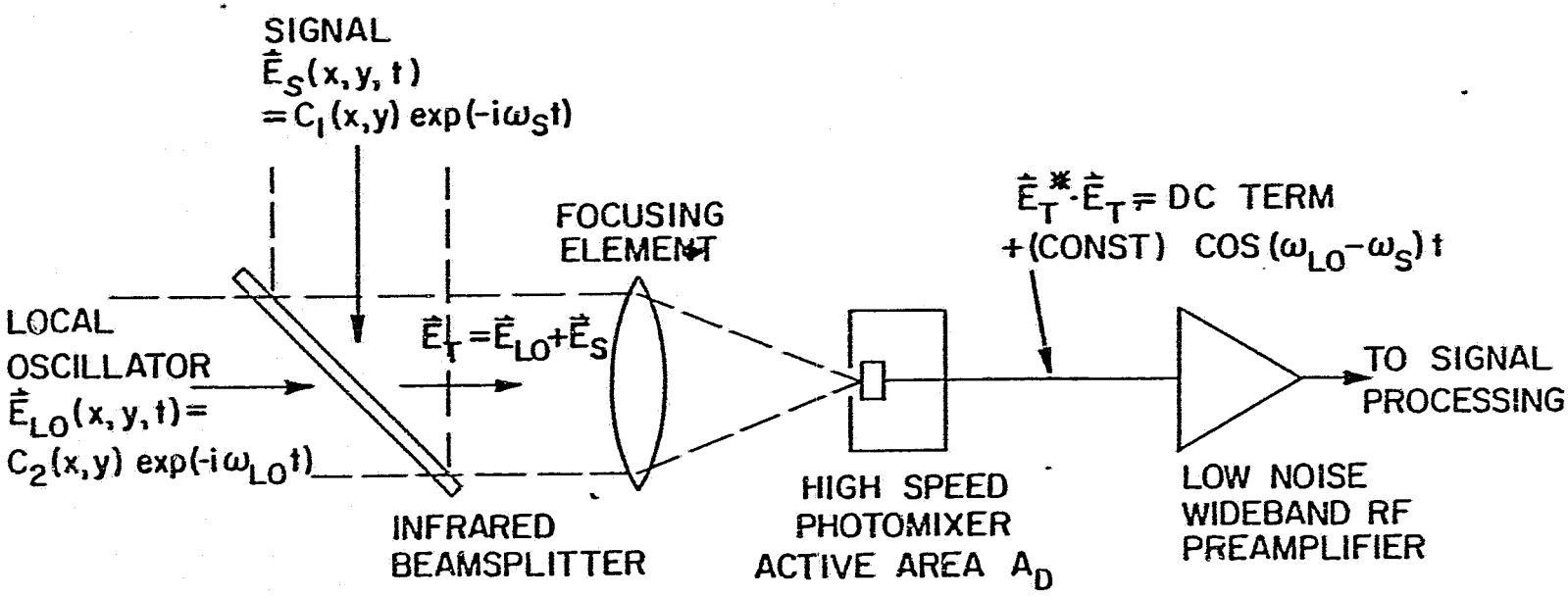


Figure 6. The basic heterodyne detection process.

the importance of proper alignment in an operating system so that phase fronts from the local oscillator and signal beams are properly matched across the detector surface. Also, the dot product shows that only a portion of the randomly polarized signal combines with the laser local oscillator to produce coherent detection. For a randomly polarized signal source, this means a factor of two reduction in the available signal power (Section 2.3). After separating the time dependence in Equation 2.1 and taking the real part of the result, one obtains

$$i_D(t) = C\,A_D\left[|E_{LO}|^2 + |E_S|^2\right] + 2C\cos(\omega_{LO} - \omega_S)t\,|E_{LO}||E_S| \iint_{A_D} (\vec{\epsilon}_{LO} \cdot \vec{\epsilon}_S)\,dxdy \qquad (2.2)$$

In Equation 2.2, $\vec{\epsilon}_{LO}$ and $\vec{\epsilon}_S$ are the unit vectors for $\vec{E}_{LO}$ and $\vec{E}_S$. The constant C can be eliminated by examining the DC term in Equation 2.2, which represents the detector's incoherent response to the total power incident on the active area. For a detector with constant quantum efficiency η, the optical power to IF-current conversion constant is $\eta e/h\nu$, e being the electronic charge. This implies

$$|E_{LO}| = \left[\frac{\eta e\,P_{LO}}{h\nu\,A_D\,C}\right]^{1/2} \qquad (2.3a)$$

and

$$|E_s| = \left[\frac{\eta e P_S}{h\nu A_D C} \right]^{1/2} \qquad (2.3b)$$

so that in terms of the LO and source powers, Equation 2.2 can be written

$$i_D(t) = \frac{\eta e}{h\nu} \left[P_{LO} + P_s + 2(P_{LO}P_S)^{1/2} S(t) \right] \qquad (2.4)$$

with

$$S(t) = \frac{1}{A_D} \cos(\omega_{LO}-\omega_S)t \iint_{A_D} (\vec{\varepsilon}_{LO} \cdot \vec{\varepsilon}_S)\, dxdy$$

The first two terms in Equation 2.4 represent the photocurrent i_{DC} arising from the total IR power on the detector, coherent or otherwise. For a randomly polarized signal source, the integral of $\vec{\varepsilon}_{LO} \cdot \vec{\varepsilon}_S$ equals $1/\sqrt{2}$, and the intermediate frequency signal term $i_{IF}(t)$ in the expression for the detector current $i_D(t)$ becomes

$$i_{IF}(t) = \frac{\sqrt{2}\eta e}{h\nu} (P_{LO}P_S)^{1/2} \cos(\omega_{LO}-\omega_S)t \qquad (2.5)$$

with a mean square value of

$$\langle i_{IF}^2 \rangle = (\eta e/h\nu)^2 P_{LO}P_S \qquad (2.6)$$

after the integration over time.

2.2 Characteristics of the Source Radiation

In reality, the source relation for nearly all natural IR sources (e.g. the terrestrial atmosphere, planets, nebulae) will also be incoherent in nature and distributed over a broad band of wavelengths. The usual problem entails observations of absorption or emission features against a blackbody continuum at a temperature T, in which case the background flux can be regarded as constant over the IF bandwidth and equal to the Planck function $B_\lambda(T)$ at that temperature. Even when observations of non-thermal sources are involved, the frequency width of the emitting feature is usually greater than a typical IF resolution element B_{IF}, so it is useful to define a brightness temperature T_B for any source. The intermediate signal frequency $\omega_{LO}-\omega_S$ in Equation 2.5 is then the center frequency of the IF bandwidth element of width B_{IF}, and the signal power P_S is the power per unit bandwidth integrated over B_{IF}. For small B_{IF}, the single sideband power P_S (watts) becomes

$$P_S = \frac{2h\nu}{\lambda^2} \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]^{-1} B_{IF} \; A \; \Omega \qquad (2.7)$$

The terms A and Ω in Equation 2.7 refer to the aperture area and field of view of the collecting optics, and are required in addition to bandwidth B_{IF} to convert $B_\nu(T)$ in $W\,m^{-2}\,Hz^{-1}\,ster^{-1}$ to received

power. For a coherent receiver, the field of view (Ω) of the collecting optics is just the diffraction limited field of view for a collecting aperture of area A. The antenna theorem (Siegman, 1966) imposes an additional constraint relating A, Ω and the wavelength λ, given by,

$$A\,\Omega \sim \lambda^2 \tag{2.8}$$

The product $A\Omega$ is called the etendue, and is a constant of the entire system when all components are properly aligned and matched. This quantity, applied to the active photodetector area A_D, sets the precision required in the alignment to guarantee that the signal and local oscillator beams are properly phase matched across the detector. Equating λ^2 with $A\Omega$, Equation 2.7 becomes

$$P_S = 2h\nu \left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1} B_{IF} \tag{2.9}$$

so that the mean square IF current is

$$\langle i_{IF}^2 \rangle = 2\,\frac{(ne)^2}{h\nu}\,P_{LO}\left[\exp\left(\frac{h\nu}{kt}\right)-1\right]^{-1} B_{IF} \tag{2.10}$$

2.3 Factors Influencing the Signal-to-Noise Ratio

Since noise sources must be taken into account before the instrument performance can be estimated, of great interest in a

heterodyne system is the signal-to-noise ratio (SNR). The dominant noise source arises from statistical fluctuations in the arrival of photons, primarily from the local oscillator, and is called shot noise or sometimes quantum noise (Oliver, 1965). The mean square shot noise current is determined by the detector DC photocurrent (i_{DC}) and IF bandwidth and is given by

$$\langle i_{SH} \rangle^2 = 2e\, i_{DC}\, B_{IF} = \frac{2\eta e^2}{h\nu} (P_{LO} + P_S)\, B_{IF} \qquad (2.11)$$

Normally $P_{LO} >> P_S$, so the shot noise current is local oscillator dominated and is simply called LO shot noise.

If sufficient local oscillator power is available so that the LO shot noise dominates all other noise sources, the signal-to-noise ratio will be maximized and the expression for the predetection signal-to-noise ratio (SNR) takes a particularly simple form

$$SNR = \frac{\langle i_{IF}^2 \rangle}{\langle i_{SH}^2 \rangle} = \eta \left[\exp\left(\frac{h\nu}{kT} \right) - 1 \right]^{-1} \qquad (2.12)$$

This case is referred to as shot noise limited operation and shows that the signal-to-noise ratio becomes independent of local oscillator power at sufficiently large P_{LO}. The dependence of signal-to-noise on IF bandwidth and integration time appears when the IF output is square-law detected and time integrated. For an integration time of τ seconds and an IF bandwidth B_{IF} in Hz, the signal to noise ratio improves as $(B\tau)^{1/2}$ so that the post-detection

value is given by

$$SNR = \eta \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]^{-1} (B_{IF}\tau)^{1/2} \qquad (2.13)$$

for shot noise limited single sideband operation.

In cases where the available LO power is limited, it is necessary to consider noise sources of thermal origin which occur as a result of the physical temperatures of both the mixer and the preamplifier. For a matched system with characteristic impedance R_A, the mean square noise current $\langle i^2_A \rangle$ contributed from these items is related to the sum of their equivalent noise temperatures by

$$\langle i^2_A \rangle = \frac{P_A + P_M}{R_A} = \frac{4k}{R_A} (T_A + T_M) B_{IF} \qquad (2.14)$$

Most mixers currently used for IR heterodyne work are mercury cadmium telluride (HgCdTe) and operate at liquid nitrogen temperatures ($\sim$77K) in order to reduce their contribution to the total noise current i_A.

When the amplifier and mixer noise contributions are taken into account, the post detection signal-to-noise ratio has the more general form,

$$SNR = \frac{\langle i_{IF}^2 \rangle (B_{IF}\tau)^{1/2}}{\langle i_{SH}^2 \rangle + \langle i_A^2 \rangle} \qquad (2.15)$$

Using the above expressions for $\langle i_{IF}^2 \rangle$, $\langle i_{SH}^2 \rangle$ and $\langle i_A^2 \rangle$, multiplying by 2 for double sideband operation, and making the approximation $P_{LO} \sim P_{LO} + P_S$, the signal-to-noise ratio becomes

$$SNR = \frac{\frac{(ne)^2}{h\nu} \; P_{LO} \left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1} (B_{IF}\tau)^{1/2}}{(ne^2/h\nu) \; P_{LO} + \left[\frac{2k(T_A + T_M)}{R_A}\right]} \qquad (2.16)$$

which reduces to the shot noise limited case (Equation 2.13) when $P_{LO} > (h\nu/ne^2)(2k/R_A)(T_A + T_M)$. In reality, the best anti-reflection coated HgCdTe detectors have quantum efficiencies of about 0.8 at low frequencies, decreasing to 0.2 - 0.3 above 1 GHz. The dynamic impedance R_A for these devices is usually greater than 100 Ω. The amplifier noise temperature T_A is determined largely by the first amplifier stage after the mixer. Wide-band low noise uhf amplifiers have measured noise figures typically less than 2.0 db throughout the 0.1 - 2.5 GHz IF range considered here. This corresponds to a noise temperature T_A of < 170K. Using these representative values, Equation 2.16 shows that laser local oscillator powers of 100-200μW are required for near shot-noise limited operation at 8 microns. Lead-salt tunable diode lasers exhibit comparable single mode powers, and under optimum conditions can provide operation near the shot noise limit.

In practice, the performance of an operating heterodyne system

is also degraded by losses in optical components, imperfect matching of the signal and local oscillator beams and as a result of the synchronous detection process used in separating the signal from the noisy background. The losses associated with tunable diode lasers are the most severe because of their diverse noise properties and their marginal power output.

Equation 2.13 may be modified to include all sources of system performance degradation (including η) by inserting the term delta (Δ) in the denominator, viz.,

$$SNR = \frac{2\,(B_{IF}\,\tau)^{1/2}}{\Delta\,[\exp(h\nu/kT)-1]} \qquad (2.17)$$

The term Δ is an efficiency parameter introduced by Abbas et al. (1976) which represents the degradation in SNR from all sources of system noise and loss of heterodyne signal. By definition, Δ =1 for ideal shot-noise limited, single sideband operation with no polarization or chopping loss, and increases as losses are introduced. In practice, a factor of 4 in degradation always exists since only one polarization component of the source radiation, that aligned with the LO field, is used while heterodyning, and the signal is chopped with a 50-50 duty cycle. A total Δ as small as ~ 7 has been achieved by the GSFC CO_2 laser heterodyne system now operating at the Kitt Peak McMath solar telescope (Kostiuk, 1981). For diode laser systems, however, the Δ is usually higher, as discussed in Chapter IV.

An alternate way of expressing the heterodyne sensitivity is in

terms of noise equivalent power (NEP). This is defined as the source power per unit frequency which produces unity signal-to-noise at the output. For a system operating at frequency ν, the noise equivalent power is

$$NEP = \frac{10^{-7}\, h\nu\, \Delta}{(B_{IF}\, \tau)^{1/2}} \quad W\ Hz^{-1} \tag{2.18}$$

The factor 10^{-7} in Equation 2.18 converts cgs units to watts.

CHAPTER III

DEVELOPMENT OF THE SPECTROMETER

3.1 Requirements for a High Sensitivity, 8-12 Micron TDL Heterodyne Spectrometer

In order to produce state-of-the-art performance from a tunable diode laser heterodyne instrument, the overall design must incorporate features to reduce or eliminate the problems discussed in Chapter I. These problems are mostly related to, but not limited to, the multitude of noise characteristics associated with TDL's and the problems of environment and frequency control for these devices. To minimize these difficulties and produce a flexible field system for general astrophysical use, the instrument should have the following features;

(i) A cryogenically cooled, vibrationally isolated environment for the TDL local oscillator, capable of continuous long-term cooling. This minimizes device degradation resulting from repeated thermal cyclings to room temperature, which tends to occur with He dewar systems.

(ii) High precision temperature control ($\lesssim$ 0.3 mK) of the TDL for periods of one hour or more, as required for extended integration times on weak sources.

(iii) Provisions in the optical layout for matching the local oscillator beam to a wide range of signal beam focal lengths. This feature is essential if the instrument is to be moved to more than one field location.

(iv) Inclusion of a mode selector grating to eliminate all but one of the Fabry-Perot modes emitted simultaneously by the TDL.

(v) Incorporation of diffraction limited reflecting optics wherever possible for improved instrument throughput and to avoid the need for frequent realignment as the operating wavelength is changed. This problem is encountered with refractive optical elements.

(vi) A dual-beam synchronous detection scheme capable of using any combination of remote source, remote reference (sky), local source and local reference in the signal and reference beams. Synchronous detection is essential for the extremely low heterodyne signal levels, and the choice of signal and reference must be flexible to accomodate a wide variety of observational approaches.

(vii) Multiplex and time integration capability to extend the weak signal detection limit.

3.2 Environment and Frequency Control for the Diode Laser Local Oscillator

The local oscillator presently installed in the instrument and used for nearly all of the observational work was a compositionally inter-diffused (CID), PbSe tunable diode laser (TDL #8300-14), manufactured by Spectra Physics, Inc., Laser Analytics Div. Referring to Figure 1, the composition of this diode lies at the right on the vertical axis defined by the binary material Pbse. The output from this laser is generally multi-mode with a total power that varies between 300μW and 1 mW, depending on operating

frequency. The power in a single mode, as required for heterodyne work, is generally less than 300 μW, however. The diode operates from about 1170 cm^{-1} at a diode current of 0.20 A and mount temperature of 12 K, to nearly 1260 cm^{-1} at 2.0 A and 60 K.

Proper mounting of the device poses some special challenges. Since the TDL frequency can be tuned by adjusting the current and temperature, these must be precisely controlled to achieve the required frequency stability. The devices are also acoustically sensitive and their output is easily amplitude modulated by vibrations of any kind.

To overcome these problems, the TDL was mounted on a vibrationally and thermally isolated platform enclosed in an evacuated shroud as shown in Figure 7. This platform is located at the second stage of a two-stage helium closed-cycle cooler capable of operation at $\leq$ 10K. A similar mounting scheme is discussed by Jennings and Hillman (1977) and has been used successfully to stabilize a diode for direct absorption spectroscopic studies. The laser and its copper mounting fixture are maintained at cryogenic temperatures by coupling the assembly through a flexible braid to the second stage of the cooler. This braid, together with a metal bellows on the vacuum shroud, isolates the diode mount assembly from mechanical shock associated with cooler operation. The entire assembly is enclosed in a 50 K radiation shield attached to the refrigerator first stage for increased immunity to room temperature fluctuations and to eliminate room temperature loading of the cooling capacity. A high-precision temperature controller is used to

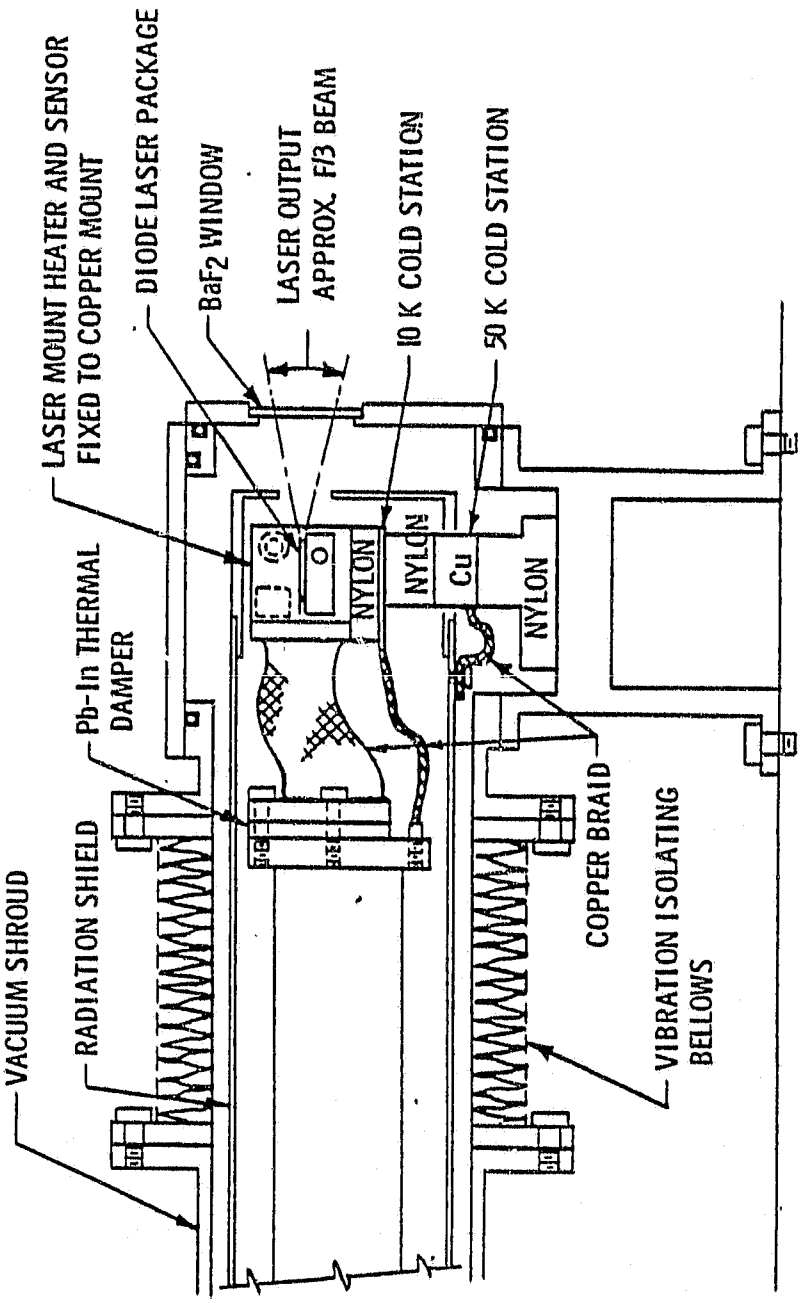


Figure 7. Method of mounting and shielding the tunable diode laser in a two-stage closed cycle cooler.

preset the diode mount temperature anywhere between the cooling limit of ∿9 K, and the maximum recommended diode temperature of about 70 K. A temperature feedback loop, consisting of a sensor and heater at the diode mount, is used to stabilize the temperature to within ± 0.3 mK over a 1 hour period, and has been demonstated to be repeatable from day to day. For diode #8300-14, this corresponds to a frequency stability over 1 hour of about 20 MHz below 1200 cm^{-1} and about 50 MHz between 1200 and 1240 cm^{-1}. Better stability is achieved over shorter time periods.

At constant temperature, the diode output frequency can be fine-tuned by varying the injection current, or the frequency may be preset to some fixed value for multi-channel integration using an RF spectral line receiver. The PbSnTe device now in use exhibits a current tuning rate of about 1.0 cm^{-1}/A near 1180 cm^{-1}, increasing to ∿3.0 cm^{-1}/A at 1230 cm^{-1}. These rates are slow enough so that fluctuations in the frequency due to current supply instabilities are negligible.

Photographs of the refrigerator vacuum shroud, vibration isolation assembly and the diode mounting arrangement are shown in Figure 8. The top panel (a) shows the vibration isolating bellows which externally isolates the cooler second stage from the diode mount. Panel (b) shows the diode laser package mounted to the copper heat sink. The scale at the left in (b) is in inch units.

3.3 Beam Matching and Imaging

The basic requirement for the spectrometer is to combine the local oscillator and signal beams using an infrared beamsplitter or

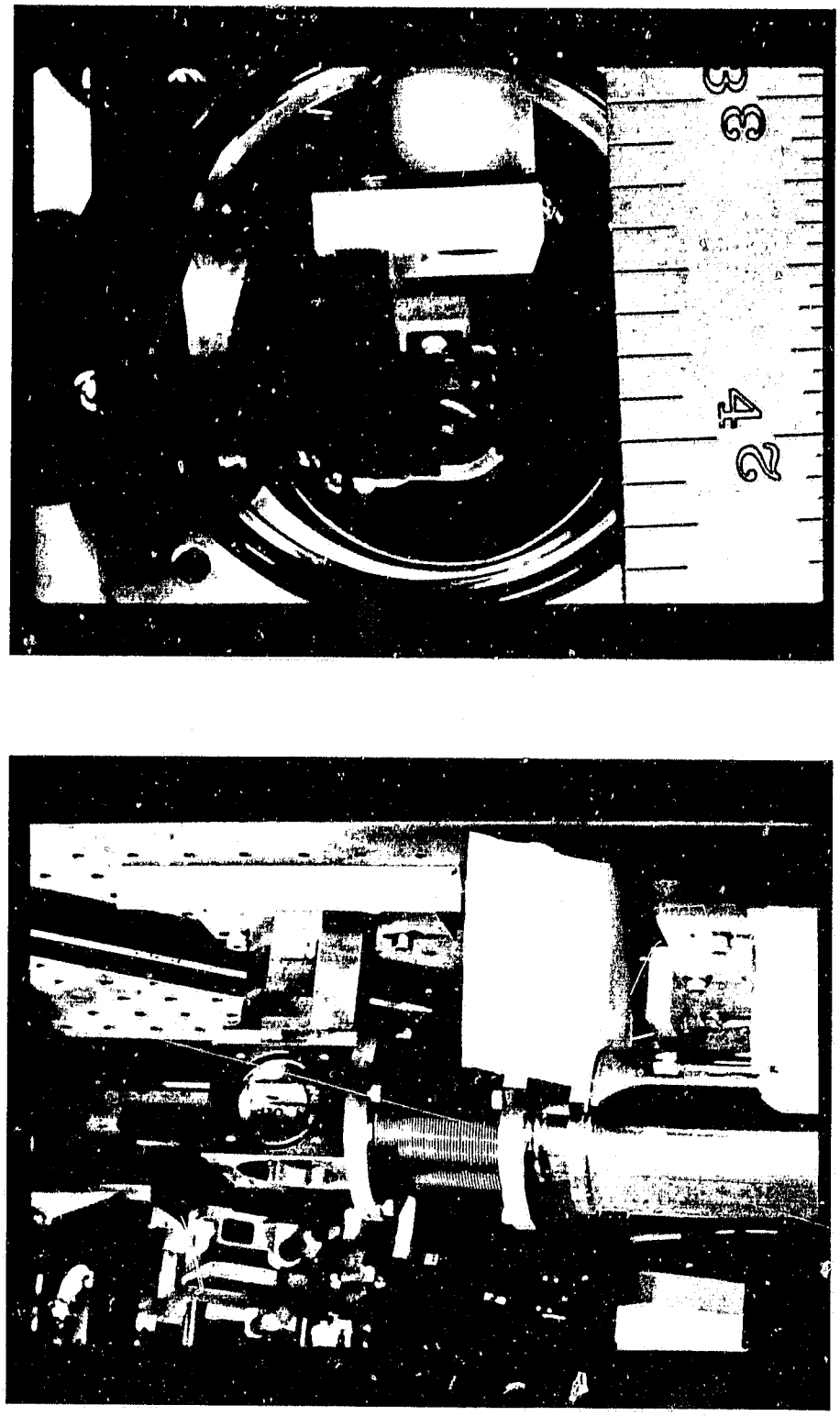


Figure 8. Photographs of the diode laser mounting in the cryogenic closed cycle cooler

some other technique, and to image both together at the active area of a high speed photo-mixer which generates the heterodyne signal. To accomplish this the system uses all reflecting optics, with the exception of the 50 percent transmitting ZnSe beam combiner and the refrigerator and detector BaF_2 windows. The use of off-axis parabolic mirrors (OAP's) as a substitute for lenses minimizes transmission loss and dispersion effects and improves the quality of focal images throughout the system. In Figure 9, the local oscillator path originates with the diode laser output at A, expands at f/3 to f/6, is converted to a parallel beam by the OAP at B, and propagates from B to the IR beamsplitter at F. The signal beam follows a similar route, beginning at the instrument focus (I) and combining with the local oscillator beam at F. The two beams propagate coaxially to the OAP at point G and are imaged together on the active area of a high speed mercury-cadmium-telluride detector at H. This portion of the beam combining process alone poses many problems, each requiring careful consideration. Off-axis parabolas are difficult to align and produce diffraction limited images only in a very small region near the geometric focus. For example, the present configuration uses a 2.5 cm beam diameter. The local oscillator emitting aperture has dimensions of order 50 microns and the detector aperture is 50-100 microns in diameter. The optimum focal lengths in this case are roughly 15 cm and 10 cm for the TDL and detector mirrors, respectively. Alignment tolerances for these focal lengths are very small and adjustment errors of order 1 mm in either the off-axis or axial directions for either mirror will

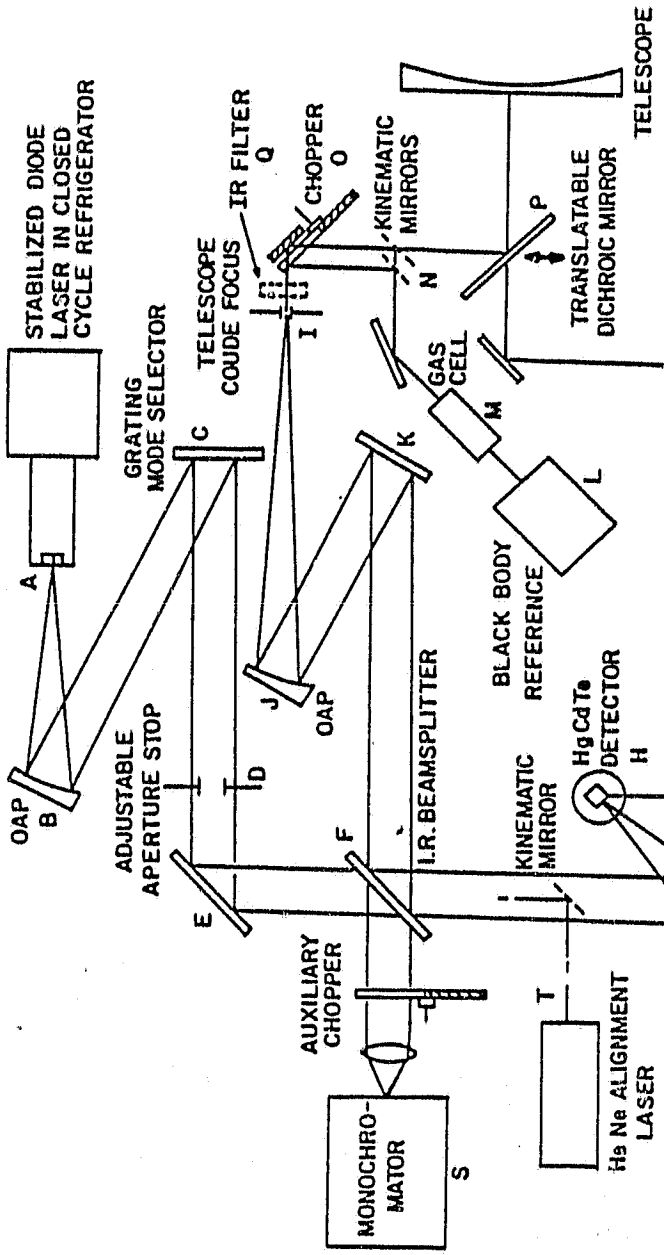


Figure 9. Schematic diagram of the diode laser heterodyne spectrometer optical front end.

seriously degrade performance. The situation is less critical for the long focal length signal beam mirror at J. This is chosen to match the f number of the incoming signal beam which varies between f/30 and f/120 at the Coude focus of large reflecting telescopes.

The signal and reference beams must also be parallel and coaxial after they are combined at the beamsplitter to avoid serious loss of signal. Cohen (1975) has performed calculations of heterodyne signal at 10 microns as a function of phase front misalignment between the local oscillator and reference beams. This study shows that the heterodyne efficiency for matched Airy beams will be reduced by 80% for as little as 4^{o} phasefront misalignment between beams. With an f/4 collecting system, this condition is reached when one beam is off-axis by no more than several millimeters.

Under normal conditions the diode will simultaneously emit several Fabry-Perot modes, separated in frequency by $\sim 2\ cm^{-1}$. In order to establish the zero-signal level and perform meaningful spectroscopic measurements, only one of these principal modes can be allowed to propagate to the detector. To reject the undesired modes, an IR grating has been installed in the LO beam at C. The grating must allow only one frequency mode to be imaged on the active area of the detector. In first order, the reciprocal dispersion $d\nu/d\beta$ for a grating with element spacing δ and operating angle β is

$$d\nu/d\beta = \nu^2 \delta \cos \beta \qquad (3.1)$$

for ν in cm^{-1}. Using $\beta \sim \pi/2$, $\nu \sim 1200$ cm^{-1} and a mode separation of ~ 2 cm^{-1}, typical of the TDL used, this leads to mode spacing of

$$\Delta\beta \sim 2 \times 10^{-6}/\delta \tag{3.2}$$

produced by the grating. In order to resolve adjacent frequency modes, the angular mode spacing $\Delta\beta$ must equal or exceed the angular width of the grating principle maxima $\Delta\theta$. For a grating illuminated by a beam of width D, this is given by

$$\Delta\theta = \left[\frac{2}{\nu D \cos\beta}\right] + \alpha \tag{3.3}$$

The added term α in Equation 3.3 is the angular width of the diode emitting region, and is non-negligible for the dimensions considered here. For D=2.5 cm, f=15 cm and assuming a 50 micron emitting aperture,

$$\Delta\theta \sim 1.3 \times 10^{-3} \text{ rad} \tag{3.4}$$

Since $\Delta\beta$ in Equation 3.2 must equal or exceed $\Delta\theta$ the grating element density $(1/\delta)$ must be

$$(1/\delta) > 650 \text{ cm}^{-1} = 65 \text{ mm}^{-1} \qquad (3.5)$$

The requirement guarantees that the modes are resolved but says nothing about their separation. In order to have only one mode on the detector, the linear mode separation in the focal plane of the detector must be larger than the detector diameter. For a detector mirror at G with focal length 10 cm, the linear mode spacing ΔX for a 65 line/mm grating is

$$\Delta X = f \Delta\beta \sim 150 \text{ } \mu\text{m} \qquad (3.6)$$

This distance is marginal since it is comparable to the size of the detector aperture. A commercially available 150 line/mm grating was used which provided a mode spacing of about 350 microns in the detector focal plane, isolating single frequency modes on the detector. The TDL is mounted so that the polarization of the $\vec{E}$ field vector is horizontal (in the plane of Figure 9) and orthogonal to the grating rulings. In this configuration the grating efficiency in first order exceeds 85% at wavelengths between 8 and 12 microns, and is suitable regardless of the diode used.

Since the instantaneous heterodyne signal is extremely small and masked by the local oscillator noise, a synchronous detection

process is required which compares the heterodyne signal plus noise with the noise in the absence of signal. In Figure 9, this is accomplished using a rotating blade chopper in combination with a stationary flat mirror (O) which provides signal and reference beams from the source in alternate half-cycles of the chopping cycle.

Maximum flexibility for a wide variety of measurements is achieved by means of dual kinematic mirrors at N. These allow a calibrated blackbody source L to be placed in either the signal or the reference beam for sensitivity measurements and instrument calibration. The blackbody source is also used for laboratory studies of subject gases by placing a gas cell in the blackbody beam. When observing remote sources, any remaining telescope or synchronous detection offsets can be removed by changing the signal and reference beam paths using a translatable dichroic mirror at P. The transmitted visible beam from this mirror also provides a conjugate image for telescope guiding.

When heterodyning against the Sun or reference blackbody, the synchronously detected source shot noise is appreciable and can be comparable in size to the heterodyne signal. This source of noise is suppressed by use of an infrared prefilter (Q) to attenuate the visible and near-infrared wavelengths. Source shot noise is created since the photocurrent generated in the detector by the source is due to the detector's DC response over all wavelengths. The response of HgCdTe detectors can be quite high, even at near infrared wavelengths where the power per unit frequency is very large. An order of magnitude estimate of this effect can be made by

comparing the mean square heterodyne signal current from Equation 2.10 to the shot noise current from Equation 2.11, with P_{LO} omitted. Apart from terms of order unity the relative size of source and signal terms is $\sim P_s/P_{LO}$. For a source temperature $\sim$ 5000 K, a system etendue of $\sim 6.4 \times 10^{-7}$ and assuming a detector quantum efficiency of $\sim$0.3 for $\lambda > 2.0$ μm, the incoherent source power is ∿30 μW, comparable in size to single mode output powers of the diode laser, typically 100-200 μW. Although a large fraction of the bolometric solar flux is attenuated by the atmosphere, the heterodyne signal for a practical system is typically much smaller than that given in Equation 2.10 so that the ratio is still a valid indicator of the size of this effect. The IR prefilter installed at Q passes wavelengths between $\sim$ 8 and 12 microns. Since the power in this wavelength interval is ∿1% of the total bolometric flux, the IR filter removes essentially all of the synchronously detected source shot noise. Of course, the bolometric flux varies as T^4 so an IR prefilter is not required for observations of sources with brightness temperatures significantly less than 1000 K.

Mention should also be made of the adjustable aperture stop (D) located in the local oscillator signal beam. Off-axis beam components have been identified as being related to both narrow band and broadband noise features observed at IF frequencies while heterodyning. The aperture stop is effective in suppressing these, as discussed in Chapter V.

In order to perform remote measurements, the optical system has been matched to the 48-inch telescope at the NASA/Goddard optical

facility near Greenbelt, Md. The incoming beam from the telescope is approximately f/30 and results in a diffraction limited spot size (full width at half maximum) of about 300 microns at the coude focus (I) when the operating wavelength is near 8 microns. The telescope beam is collimated using a 30 inch focal length for the signal beam parabolic mirror (J) and is combined with the laser local oscillator beam as previously described.

3.4 Methods of IF Signal Processing

Two methods for processing the IF output of the HgCdTe detector (H) are now being employed. The first method, called sweep integration, uses a single channel which is swept in frequency by current tuning the TDL. Variations of this approach have been used in previous attempts to perform diode laser heterodyne spectroscopy of gas cells and remote sources (Frerking and Muehlner, 1977; Ku and Spears, 1977; Harward and Hoell, 1980). In Figure 10, the total IF signal from the IR detector is amplified 50 dB using a low-noise, wideband (0-1.5 GHz) amplifier. A portion of this output is again amplified and band limited using a 100 MHz low pass filter. The filter output is square-law detected and the DC signal is delivered to a lock-in amplifier and signal averager. Using a high precision TDL current supply, the diode laser can be slowly tuned to scan the full region of a single TDL mode where diode excess noise is minimal. The signal averager is triggered simultaneously with the beginning of each new sweep and permits stacking of both source and reference blackbody scans. As a general guideline, in order to avoid undue smearing of the line shape, the diode frequency scan

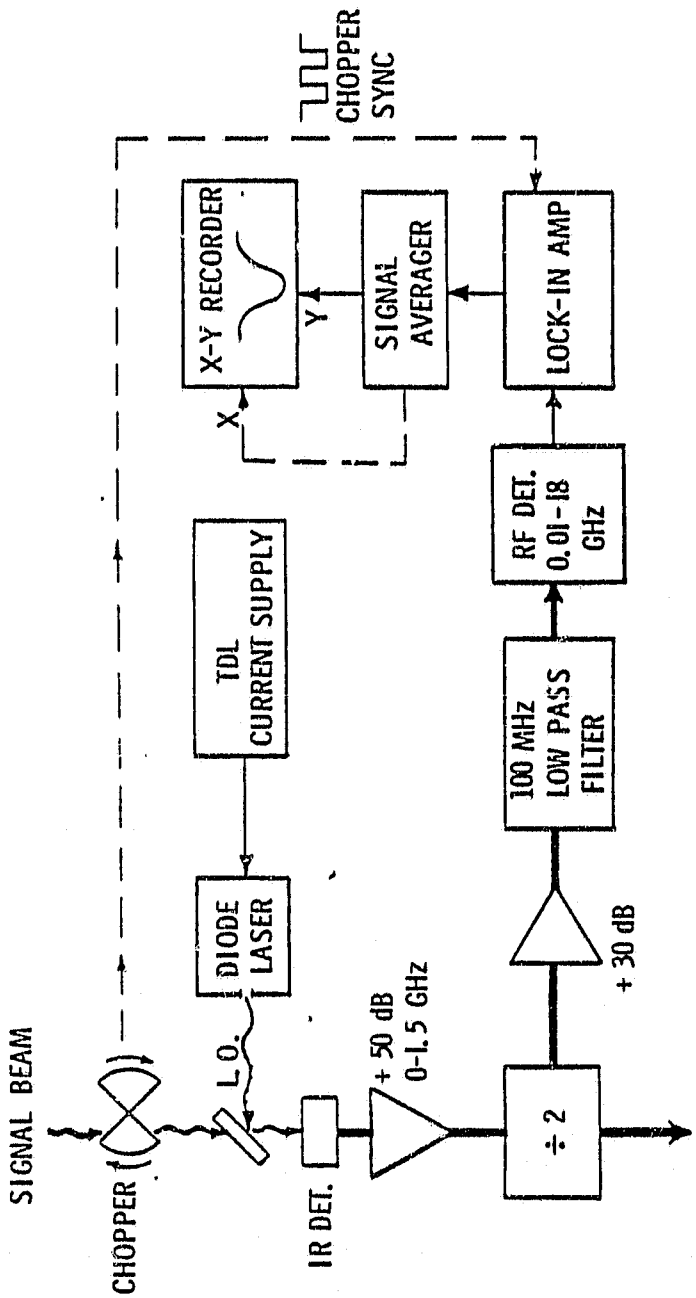


Figure 10. Single channel, sweep-integration heterodyne detection electronics.

rate $d\nu/dt$ has been governed by the IF filter width B and the time constant τ of the lock-in amplifier according to

$$d\nu/dt < 0.2\ B/\tau \qquad (3.7)$$

For B = 100 MHz, τ = 400 msec and a typical single mode tuning range of 0.3 cm^{-1}, this corresponds to individual scan times equal to or greater than 3 minutes. Since it is sometimes possible to tune over a fraction of a wavenumber, this method of acquiring data is advantageous for observations of highly broadened features, e.g. wings of tropospheric lines in solar absorption, or velocity broadened features in astrophysical objects.

The second approach employs a 64-channel RF spectral line receiver (Buhl and Mumma, 1980) to provide multiplexing and extended time integration capability. This is much more useful for measurements of doppler broadened features and for sources with low brightness temperatures. A block diagram of this system is shown in Figure 11.

In this mode of operation, the TDL local oscillator is pre-set to some fixed frequency in near coincidence with the spectral feature of interest, and the IF output of the detector-preamplifier is delivered to a variable attenuator at the beginning of the RF chain. The IF output is split into 64 contiguous frequency channels, each 25 MHz wide, covering 100-1700 MHz. After the signal

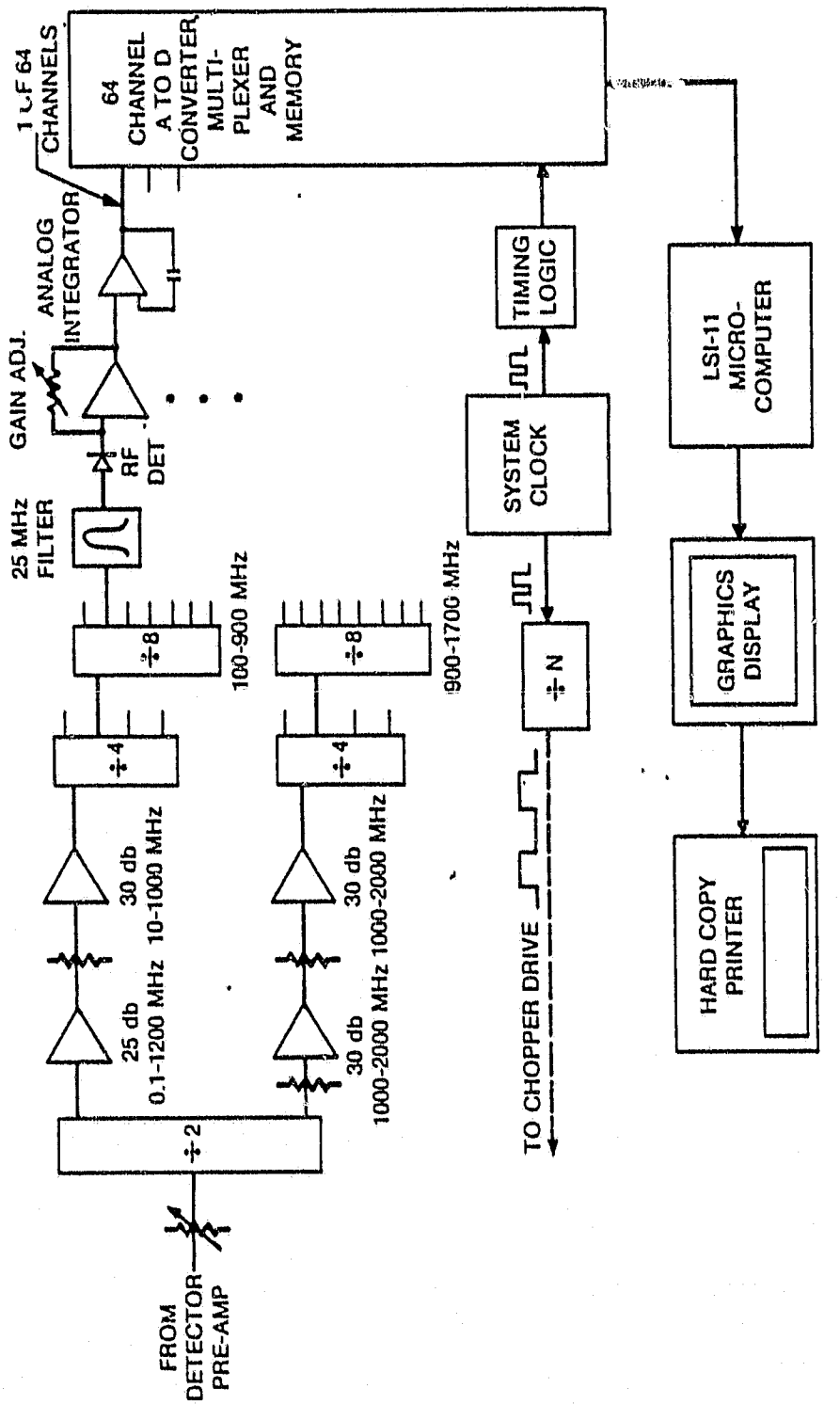


Figure 11. Diagram showing principle parts of the RF spectral line receiver.

is divided and passed through an individual filter, it is detected and amplified. At this point the receiver noise has been reduced to a DC value plus a noise spectrum of several kHz to allow the chopped signal to propagate through. Gain pots on the front of the filter bank allow the channels to be equalized. The signal is then integrated in an op-amp circuit for half of a 16 msec chopping cycle. At the end of the half-cycle, the individual channels are digitized by an A/D converter for each bank and added to the signal buffer memory. The integrators are reset (1.2 μsec) and the integration started for the reference half of the chopping cycle. This time the data are stored in the reference half of the memory. When the run is finished the quantity (signal - reference) / (reference - zero) for each channel is automatically stored on a computer disk. The zero is determined by a short integration with the RF turned off, yielding individual channel offsets, defined to be "zero". These data can then be manipulated and hard-copied by an on-line minicomputer system. Use of a spectral line receiver improves the sensitivity of the spectrometer since the entire IF spectrum is recorded simultaneously rather than sequentially, and long integration times are thus possible. In addition, a 25 MHz filter element corresponds to a resolving power $\nu/\Delta\nu$ of about 1.5×10^6 at 8 microns. This is about a factor of 8 larger than that used in the sweep integration mode and very clearly sub-doppler for temperatures found in planetary atmospheres.

CHAPTER IV

INSTRUMENT DIAGNOSTICS AND TEST RESULTS

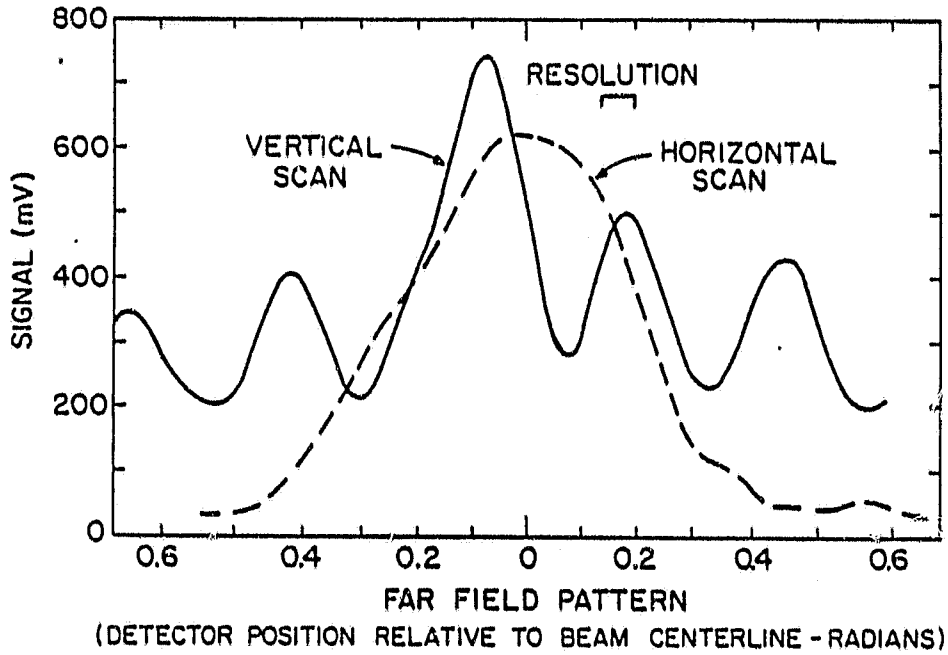
4.1 Critical Aspects of System Performance

Since lead salt diode laser development is still largely in the experimental phase, these devices possess a number of undesirable operating characteristics. When used as local oscillators, most problems are related to inadequate power output, limited tunability and a high degree of IF excess noise. Each TDL has its own "personality" and must be thoroughly diagnosed to ascertain how it will affect the instrument performance. The following sections describe the procedures necessary to optimally match the diode output to the instrument optics, to suppress the IF noise, to determine the diode operating frequencies and to evaluate the overall instrument sensitivity.

4.2 Far-field and Near-field TDL Output Scans

Optimum matching of the signal and local oscillator fields on the detector is essential since only the scalar product $\vec{E}_S \cdot \vec{E}_{LO}$ on the detector active area is effective in producing heterodyne signal. Since the TDL output power pattern is often unpredictable and irregular, far-field and near-field scans of the output, shown in Figure 12, constitute the first phase of the LO alignment.

Far-field scans were made as a means of centering the diode power distribution on the off-axis parabolic mirror (B) in Figure 9. This is necessary to insure that the maximum amount of local



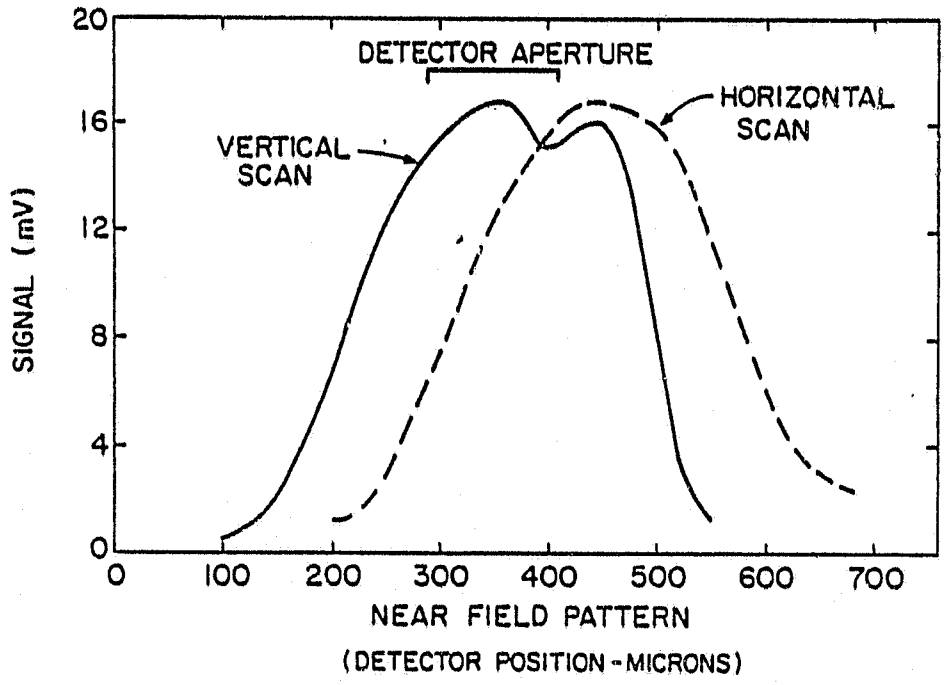


Figure 12. Far-field (a) and near-field (b) scans of the output power pattern from TDL #8300-14, operating near 1185 cm^{-1} (T_d = 19 K, I_{LO} = 0.83 A).

oscillator power is delivered to the detector and that the LO beam is roughly co-axial with the signal beam, as required for optimum phasing of the signal and LO fields in the detector focal plane. The far-field beam scans were performed using a 1 mm pyroelectric detector mounted to a 2-axis horizontal/vertical micrometer drive, with the vertical axis motor driven. In Figure 12a the detector was positioned about 3 cm from the diode at (A) (less than 2 cm from the output window) yielding an angular resolution of about 2°. To improve the signal-to-noise ratio, the laser was chopped and the detector output and sync signal were fed to a lock-in amplifier. The result could then be displayed on a chart recorder. The horizontal scan appears roughly gaussian, suggesting a single emitting mode, but the multiple lobe distribution in the vertical scan is characteristic of high order mode operation. The near field pattern was examined by scanning the diode image in two orthogonal directions in the HgCdTe detector focal plane (H), with the grating (C) positioned for central order reflection. The image resolution is degraded since it is defined by the detector size of ~100 microns plus the ~40 micron blur circle diameter of the detector off-axis parabolic mirror (G).

Both scans are in qualitative agreement regarding the diode output geometry. The horizontal scan shows a gaussian profile for both patterns, at least to within the resolution. In the vertical direction, however, there is evidence of two emitting sources spaced 50-100 microns apart, which are just barely resolved. These give rise to the fringes observed in the vertical far-field intensity

distribution. The near-field image in Figure 12b was obtained using a 10.2 cm focal length mirror, equal to that at the detector end, resulting in a 1:1 reconstruction of the diode image. For longer focal length LO mirrors, which seem to yield improved performance in the heterodyne mode, some demagnification of the diode image occurs and these peaks are completely unresolved. In these cases the diode image is positioned by adjusting for a peak in the LO induced detector photocurrent.

The phase relationship between these adjacent peaks and the effects of aperture stopping on heterodyne signal and noise levels are not well understood, since these effects vary widely with operating conditions and with the device under test. For the TDL diagnosed here, the criteria for aperture stopping depend on the mode of heterodyne operation considered - the single-channel sweep average mode or use of the 64-channel spectral line receiver. It appears that off-axis contributions to the LO power are often the source of narrow band noise features observable in the IF, similar to the spiking resonance symptom in Figure 4b. Aperture stopping provides some discrimination against this out also reduces the overall heterodyne signal level. Some IF noise of this type can be tolerated using the spectral line receiver since the noise spikes generally occupy only one or a few channels and the isolated bad channels can be artifically removed from the final data. The single channel scheme is much more sensitive to this source, since no IF discrimination is possible. Attempts made to use RF bandpass filters for IF selection have been unsuccessful since the frequencies

of these spikes change as the diode is tuned. When this type of noise is present, aperture stopping to a diameter of ∿3 cm generally gives best results for the spectral line receiver, but a much smaller diameter (∿1 cm) is generally required for a single IF channel. This often defeats the sweep integration mode, since only a small fraction of the LO power is imaged on the detector.

4.3 Measurement of Broadband IF Excess Noise

Another dominant source of TDL excess noise is broadband and generally covers the entire 0-2 GHz IF spectral range. This is generated along with the other characteristic noise types at many diode current settings. Figure 13 shows the effect of the mirror (B) focal length on broadband noise, examined for two different regions of diode operation; one near 1180 cm^{-1} at a mount temperature of ∿ 19 K and the other near 1250 cm^{-1} at a temperature of ∿ 42 K. These regions represent the practical extremes of operation for this TDL. Curves II and V show quiet diode operation, the level being largely due to local oscillator shot noise. Since the vertical scale is expressed in dB, the separation i-ii, and also iv-v, indicates the ratio of noise power with excess noise relative to noise power with the diode blocked. This effect rolls off at high intermediate frequencies for all of the mirrors tried, but is minimized for the longer focal lengths. The LO shot noise contribution (ii-iii, and v-vi) is a measure of LO power on the detector and hence heterodyne signal, assuming proper matching. This also decreases at large focal lengths, simply because mirror

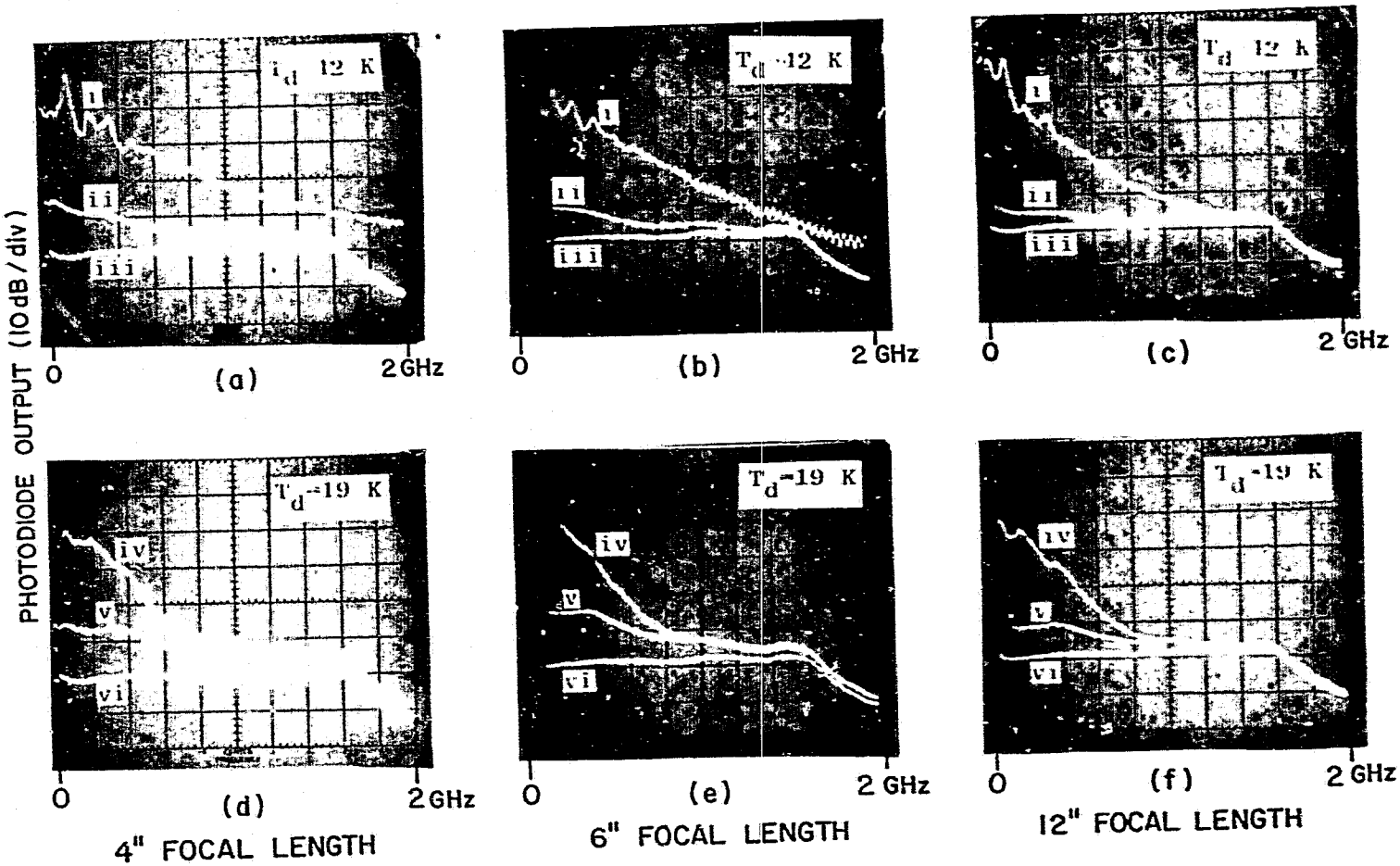


Figure 13. Measurements of the TDL IF noise frequency spectrum as a function of LO mirror focal length and diode laser current and temperature. TDL current settings: (i) 0.69 A, (ii) 0.62 A, (iii-vi) LO blocked, (iv) 0.80 A, (v) 0.83 A.

(B) no longer captures an appreciable portion of the LO power. In practice the best all around performance in the heterodyne mode seems to occur using the 6" focal length mirror. In this case, the nominal instrument beam diameter collects at about f/4, and captures the entire central lobe in the vertical scan of Figure 12a, producing the best matching between LO and signal beams.

4.4 System Sensitivity Test Using the 64-Channel Spectral Line Receiver

Figure 14 shows a blackbody heterodyne measurement using the 64-channel spectral line receiver for an LO power of about 300uW, and a high quantum efficiency HgCdTe photodiode as mixer. A number of different photomixers were used during the instrument development and observational phases. This detector was one part of a 12 element array. Figure 14 shows the overall IF response of mixer, preamplifier and cables, typical of LO performance near 1180 cm^{-1} where TDL power is generally high and excess noise minimal. The decreased response at large intermtdiate frequencies results from the roll-off of the preamplifer (3 dB B.W. $\sim$ 1500 MHz) and from the HgCdTe detector. The standing waves are caused by an impedance mismatch between the IR detector and low noise preamplifier. It was possible to suppress these by reducing the number of RF fittings and the length of cable between the detector and amplifier.

The vertical scale gives the ratio of heterodyne signal to reference which are obtained from the chopped signal and reference beams. These values can be compared with the expression for SNR from Equation 2.13 (multiplying by 2 for double sideband operation)

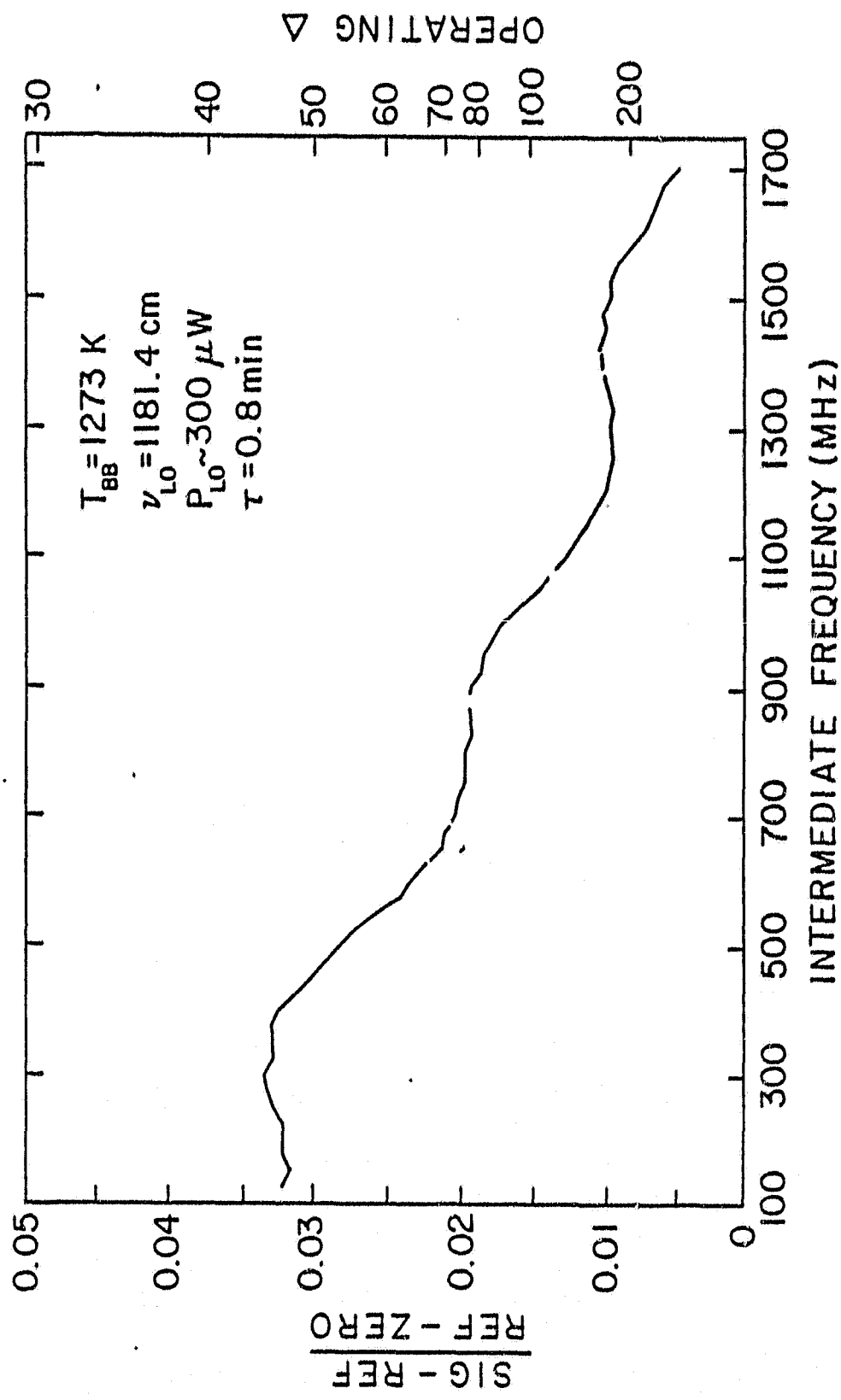


Figure 14. Heterodyne measurement of 1273 K blackbody reference source using 64 channel RF spectral line receiver.

to give the system Δ for these operating conditions. Values for Δ are shown in the right hand column. The smallest Δ, indicating highest efficiency, is about 43 near 300 MHz in the IF. By analogy with the terminology used at millimeter and longer wavelengths, this number might be considered to correspond to a system noise temperature of ∿ 50,000 K. It should be noted, however, that the convention for expressing system efficiency in terms of noise temperature is not followed for coherent detection near 10 microns. For the highest continuum temperatures of interest using this technique (i.e. the solar photosphere at ∿ 5000 K) hν/kT is ∿ 0.35, so the Rayleigh-Jeans approximation no longer applies.

4.5 Survey of the TDL Fabry-Perot Mode Frequencies

The frequency coverage of the spectrometer is limited to regions near the diode laser Fabry - Perot modes, and the continuous tunability is determined by the mode tuning range. In order to determine where the powerful modes exist and how widely they tune, a survey was performed for diode #8300-14, showing the locations of all available modes between 1180 and 1240 cm^{-1} (Figure 15). This study is a prerequisite to precise frequency calibration using standard calibration lines of reference gases.

The survey was made by first fixing the diode current at 0.75 A and then scanning the mode distribution in frequency with a monochromator (S). This was performed at a number of mount temperatures over the range of 15K to 45K. While mapping the mode distribution, the mode selector grating (C) was placed in central

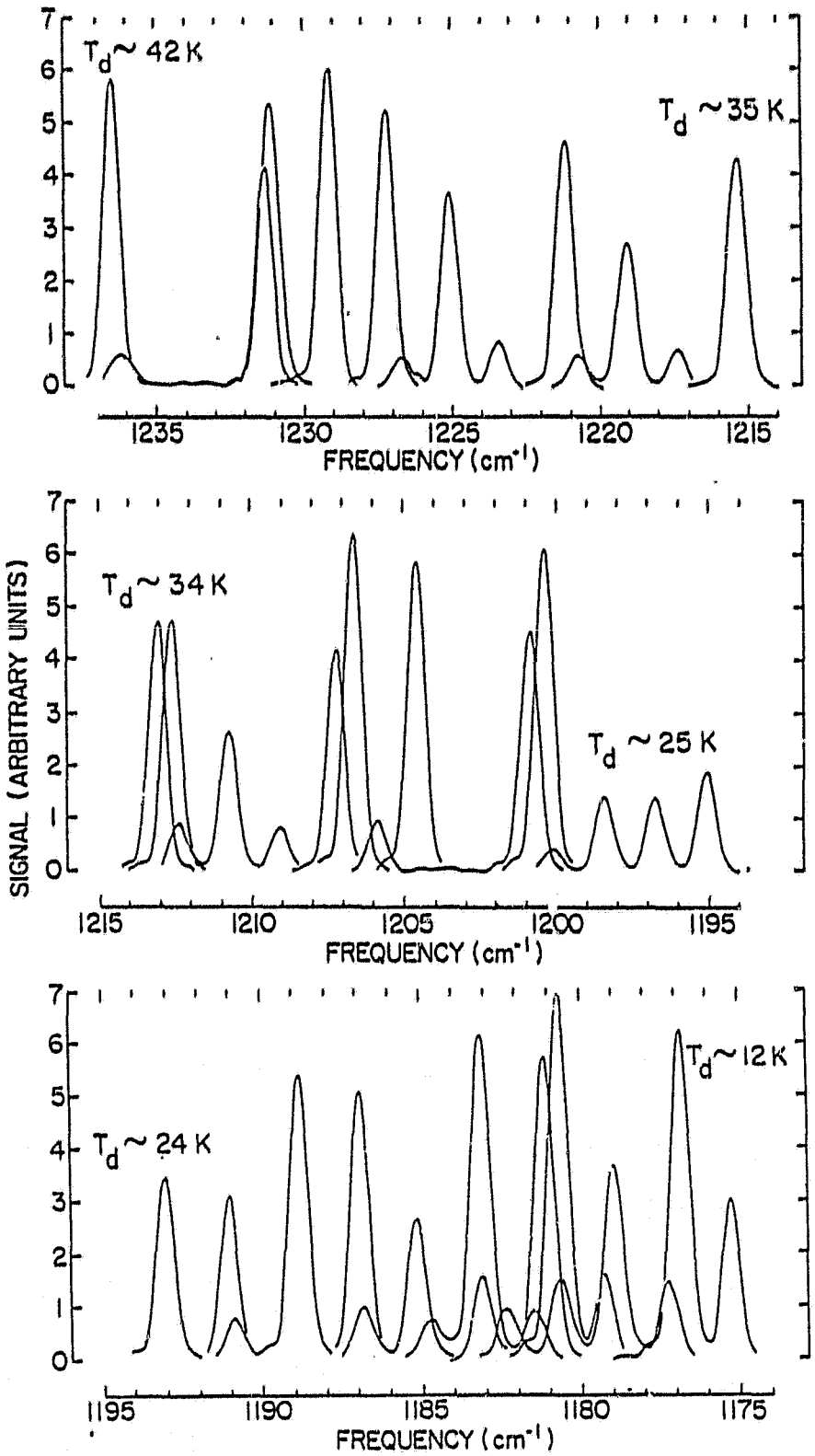


Figure 15. Monochromator mode survey of TDL #8300-14 between 1174 and 1237 cm^{-1}, obtained by varying the diode mount temperature T_d (I_{LD} = 0.75 A).

order and acted only as a reflecting element. The LO beam was mechanically chopped using a tuning fork chopper and the sync signal was delivered to a look-in amplifier along with that from a pyroelectric detector located at the monochromator exit slit. This is also the arrangement used for making individual frequency determinations between measurements during normal operation. Absolute frequency calibration of the monochromator to ± 0.1 cm^{-1} was achieved by performing a third order polynomial fit to a series of HeNe orders, near the first order grating position at 3 microns.

A few of the sample temperature settings are indicated in Figure 15 at the frequencies where they produce principal modes. The mode locations coincide with the position of the gain envelope while their number and size give an indication of the envelope shape. The Fabry-Perot frequency separation for this device is roughly 2 cm^{-1}, with a typical mode tuning range of 0.2-0.5 cm^{-1}. Examples of continuous tunability can be seen near 1181, 1201, 1207, 1213 and 1231 cm^{-1} where the mode tunes smoothly over the interval between two different temperature settings. The diode thus lases over 15-30% of its nominal tuning range. As previously discussed, however, not even this fraction is usually suitable for heterodyne work, since some of it is dominated by large amounts of IF excess noise.

In addition to indicating the mode locations, the survey shows the regions in which diode output power is generally high, and also the locations where single-mode operation occurs. Although the existence of two or more simultaneous Fabry-Perot modes does not

imply excess noise, the largest single mode power, and hence largest heterodyne signals are obtained during single mode operation.

No clear evidence exists for appreciable output power degradation of diode #8300-14 after ~2 years of intermittent use, at least in terms of output power. Measurements of long term changes in performance are difficult, however, since several different HgCdTe detectors have been used during the development phase, with a wide range of DC and heterodyne responsivities. Even with the same detector, some degradation with time occurs, probably at a faster rate than for the TDL.

4.6 Precise Frequency Calibration Using Reference Gases

Although the monochromator is suitable for absolute frequency calibration to within ± 0.1 cm^{-1}, the IF channel width in the spectral line receiver is ~ 0.001 cm^{-1}, which corresponds roughly to the frequency stability of the LO over resonable integration periods (~1/2 hour). In order to preset the LO for heterodyne measurements of selected lines, it is thus essential to determine the LO frequency, at least to within a few milli-wavenumbers. Several laboratory gases have been used for this purpose by recording the positions of absorption lines that have been identified and characterized in published line position studies. The advantages and disadvantages of each standard gas are listed below. In any case, the initial 0.1 cm^{-1} uncertainty is small enough to unambiguously identify reference gas features.

N_2O - line positions known extremely well near 1200 cm^{-1} ($\Delta\nu < 0.001\ cm^{-1}$) but reference lines do not always exist over the continuous tuning range of a single mode.

SO_2 - has an abundance of identified lines, but no synoptic observational line position studies for SO_2 have been made in this part of the spectrum. The theoretical line positions can be in error more than 0.01 cm^{-1}, although the usual discrepancy near 1180 cm^{-1} is $\sim$.003 cm^{-1}, when calibrated against N_2O reference lines.

H_2O_2 - abundance of lines between 1180-1240 cm^{-1} and positions generally good to $\sim$.003 cm^{-1}, but the liquid is difficult to handle and the vapor reacts with BaF_2 and ZnSe cell windows.

CH_4 - generally useful but, like N_2O, reference lines are infrequent in some regions. Line position uncertainties are typically < 0.005 cm^{-1}.

In cases where no N_2O line exists near the frequency of interest, SO_2, H_2O_2, and CH_4 have been found to be useful as transfer frequency standards. When lines of these molecules can be

calibrated against N_2O, they provide nearly the same frequency positional accuracy as N_2O itself.

Figure 16 shows an example of a frequency calibration spectrum typical of those performed prior to making heterodyne observations. Here, N_2O, SO_2 and a 0.0625 cm^{-1} germanium (Ge) calibration etalon were used together to achieve a .002 cm^{-1} uncertainty in LO position. The precision was limited largely by the current resettability of the TDL output frequency. The calibration spectrum was made by alternately placing N_2O and SO_2 gas cells in the LO beam while mechanically chopping the LO with a tuning fork chopper and plotting the synchronously detected HgCdTe DC output on a chart recorder. The voltage ramp used to sweep the diode current was also applied to the chart recorder x-axis drive so that the x-position was repeatable with diode current and hence frequency. For this measurement the grating was used in first order to select only the principal mode of interest.

The frequency error at any diode current setting is determined by the degree of nonlinearity in the diode tuning rate and the amount of offset from an N_2O line. The quoted SO_2 line positions are the theoretical values, but the internal consistency for this example is quite good ($\leq$ .003 cm^{-1} over the entire page). The fringes from a Ge etalon like the one used here may also show appreciable drift due to small changes in room temperature during the measurement period. Over the time between adjacent fringe maxima ($\sim$20 sec), a uniform drift in one direction has been found to introduce a 10 to 15 percent change in the absolute frequency scale

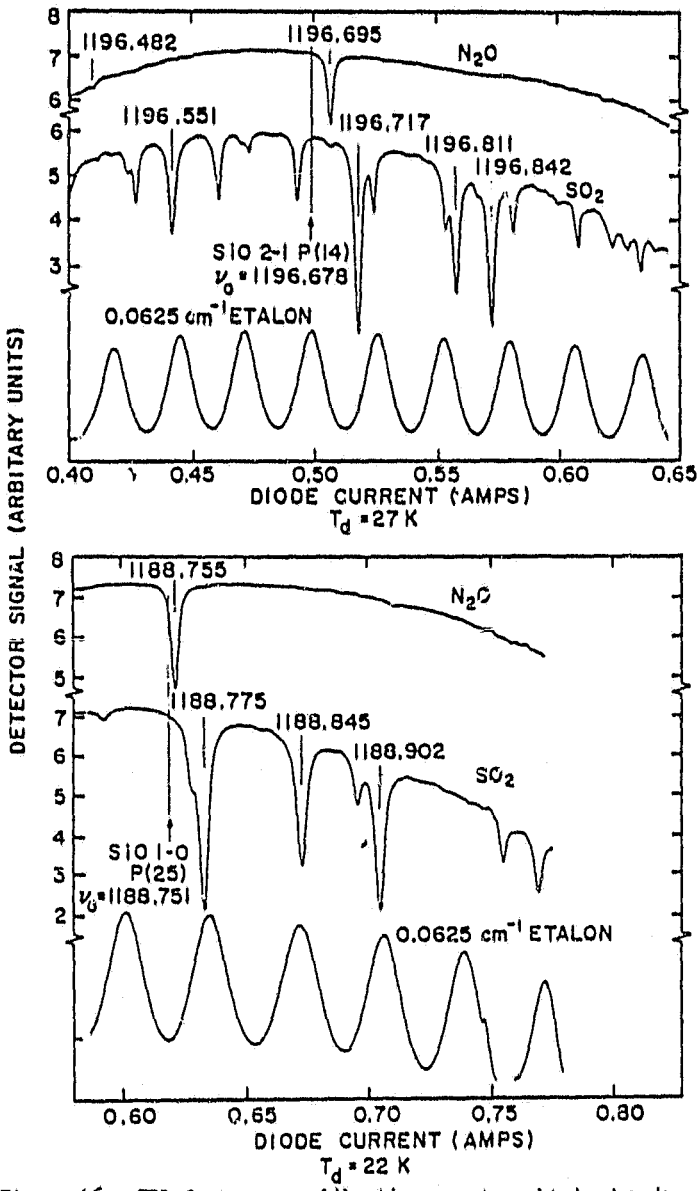


Figure 16. TDL frequency calibration spectra obtained using N_2O and SO_2 as frequency standards (frequencies in cm^{-1}.)

set by the fringe spacing.

Also shown on the calibration spectra are the positions of two SiO absorption features, the frequencies of which have been independently determined by gas phase laboratory measurements (Lovas et al., 1981). The spectra of Figure 16 were used to preset the TDL current prior to the observations of these SiO lines in the sunspot spectrum. The near coincidence of these SiO lines with N_2O permits an independent frequency determination that is in principle as accurate as that for N_2O, i.e. 0.001 cm^{-1}. In practice, LO frequency drift degrades the accuracy of the measurements.

4.7 Heterodyne Detection of Laboratory N_2O Features

When using the RF spectral line receiver it is necessary to remove the system IF frequency response shown in Figure 14. This is done by ratioing the source spectrum to that of a reference blackbody. Examples of spectra corrected in this way are shown in Figure 17, which shows the heterodyne detection of some laboratory N_2O lines against a 1273 K blackbody source. The LO frequency for this measurement was determined by locating the N_2O feature in direct absorption and offsetting the LO a predetermined amount. A total of eight 48 second integrations were then made with an N_2O cell in the reference blackbody beam, followed by an equal integration time on the blackbody with the cell removed. Since the frequency response of the system is present in both signal and reference scans, it can effectively be removed by ratioing the signal to the reference, leaving only the gas cell absorption

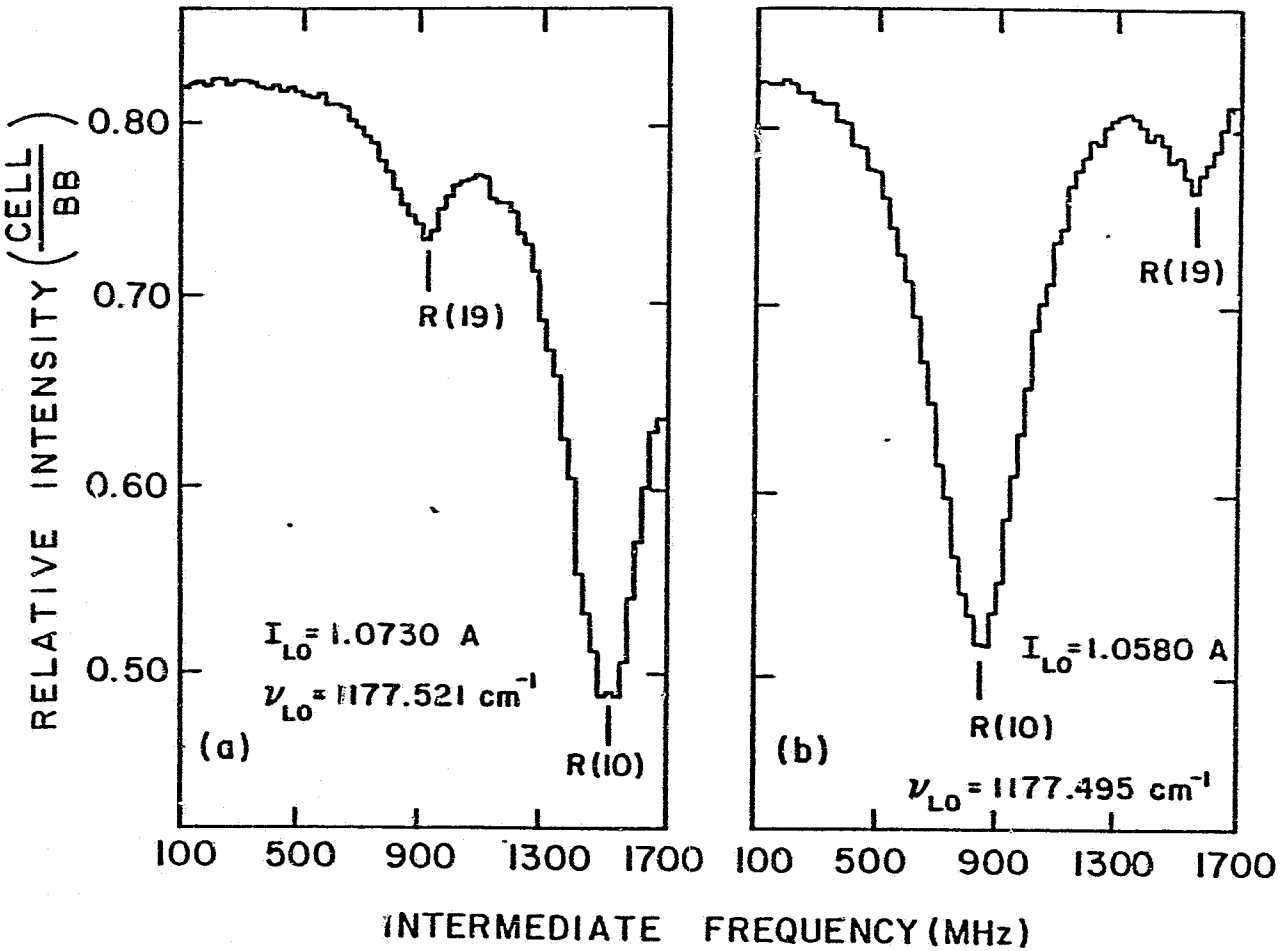


Figure 17. Heterodyne measurements of strong and weak N_2O lines with line center positions of 1177.468 and 1177.550 cm^{-1}.

spectrum. The strong and weak features are the $2\nu_2$, R(10) and $3\nu_2-\nu_2$ R(19) lines with rest frequencies of 1177.468 and 1177.550 cm^{-1}. For N_2O heterodyne measurements the LO frequency can be determined to within the line position uncertainty, which is less than 30 MHz, or 0.001 cm^{-1} at 8 microns.

In both (a) and (b), the LO frequency lies between the two line centers placing the observed features in opposite sidebands, which then appear folded over each other in this double sideband spectrum. The strong feature resides in the lower frequency sideband and the weak feature in the upper frequency sideband. In going from (a) to (b) the LO frequency has been decreased, the line with the lower rest frequency moving closer to the LO line position and the weaker line moving in the opposite direction. This figure illustrates the importance of continuous LO tunability, not only to accommodate large frequency changes, but also to enable determination of the sideband in which a given feature resides. The former degree of freedom is not readily available with gas laser systems.

4.8 Comments Concerning the Remote Detection Tests

The remote detection capability of the spectrometer has been tested in both the sweep average and spectral line receiver modes at the NASA/GSFC 48" reflecting telescope, and also at the NASA 3-meter IR telescope at Mauna Kea, Hawaii. A number of observations were made to "shakedown" engineering problems that occur in the remote observation mode and verify the field-worthiness of the spectrometer. These efforts included: measurements of absorption

from terrestrial N_2O, CH_4, O_3 and H_2O; continuum measurements of Mars and Venus (the first such measurements of Venus); the first continuum measurements of the infrared object IRC+10216 using this technique. These studies are described in more detail in Appendices B and C.

4.9 Summary of Instrument Performance and Specifications

Table 2 lists the specifications for the TDL heterodyne spectrometer, as measured using TDL #8300-14. These specifications are only representative, and will improve dramatically with the substitution of better diode lasers and photodetectors. Some of the parameters are fixed by the instrument design, while others are diode dependent and vary widely, depending on the choice of operating frequency.

TABLE 2

Specifications of the Diode Laser Heterodyne Spectrometer

Instrument:		
Nominal Operating Frequency Range -	800 - 1250	cm^{-1}
Intermediate Frequency Bandwidth -	± 0.0567	cm^{-1}
Frequency Resolution ($\nu/\Delta\nu$) -	$\sim 10^6$	
Signal Beam f/number Range -	20 - 120	
TDL Temperature Control Range -	9 - 100	K
TDL Injection Current Range -	0 - 2	A
Ambient Operating Temperature -	< 32	C
Diode Laser: (Spectra-Physics Inc., Laser Analytics Div., Diode #8300-14)		
Operating Frequency Range -	1170 - 1260	cm^{-1}
Typical Single Mode Tuning Range -	0.2 - 0.6	cm^{-1}
Observed Mode Separation -	2	cm^{-1}
Single Mode Output Power (typical) -	60 - 300	μW
Safe Operating Temperature Range -	< 70	K
Overall:		
NEP at 8 Microns: Best -	1.1×10^{-18}	$W\ Hz^{-1}$
Typical -	3×10^{-18}	$W\ Hz^{-1}$
Operating Frequency Uncertainty -	0.001 - 0.010	cm^{-1}
Frequency Drift: ≤ 10 min -	0.002	cm^{-1}
≤ 1 hr -	0.005	cm^{-1}

CHAPTER V

HETERODYNE OBSERVATIONS OF SiO IN SUNSPOTS

5.1 Description of the Observing Program

Table 3 lists the dates and relevant data on the series of observations of SiO absorption features in sunspot umbrae, made in March, 1981. Measurements were made of 5, v=1-0 fundamental and v=2-1 hot band absorption transitions near 1200 cm^{-1} (8.3 microns) using the diode laser heterodyne spectrometer described in Chapters III and IV. The spectrometer was located at the coude focus of the 48 inch reflecting telescope at NASA Goddard Space Flight Center, Greenbelt, MD. The SiO lines were observed at subdoppler resolution using the 64-channel spectral line receiver.

The observational conditions were widely variable during March, and large differences in air mass, haze and extent of cloud cover prevailed during the observations. This variability complicates any comparison of observed lines based on their relative absorption. Individual umbrae ranged in size from about 5 to 18 arsec, with values of μ between 0.75 and 1.00. As is customary, $\mu = \cos\theta$ with θ the angular distance from solar disk center. With the exception of the large spot visible near the center of the disk on March 25 and 26, all of the observed sunspot umbrae were 10 arcsec or less. The combination of poor observing conditions and small spot size made the observations susceptible to photospheric scattered light and it was necessary to correct for this effect as described

TABLE 3

SiO Observational Data

Transition	Dates	ν_o (cm^{-1}) [a]	μ	Airmass	Integration Time (sec)	Comments [b]
1-0 R(4)	Mar. 31	1236.705	0.75	1.40	770	Small spot group - small umbrae
"	"	"	"	1.25	"	"
1-0 P(20)	Mar. 22	1198.764	0.82	1.40	670	Medium spot, haze, clouds
"	"	"	"	1.30	770	"
"	"	"	1.00	"	380	"
1-0 P(26)	Mar. 19	1188.752	0.99	1.60	"	Medium spot, haze
"	Mar. 26	"	0.80	1.40	580	Spot group - large umbrae
2-1 P(14)	Mar. 21	1196.678	0.91	1.60	770	Medium isolated spot
"	Mar. 25	"	0.95	1.55	"	Large isolated spot, haze
P(25)	Mar. 13	1178.805	0.83	1.40	"	Medium spot, haze
"	"	"	"	1.45	"	"
"	Mar. 17	"	"	1.60	670	Medium spot - telescope wind shake

[a] Lovas, Maki and Olson (1981). Frequency uncertainty is $\pm$ 0.002 cm^{-1}.

[b] Small umbrae, 4 - 7 arcsec dia.; Medium, 8 - 12 arcsec; Large, 13 - 18 arcsec.

in Section 6.2.

5.2 Details of the Observations

5.2.1 Choice of Observed SiO Lines

As shown in Figure 18, the observed lines are allowed electric dipole transitions at frequencies near 1200 cm^{-1} with a change of +1 in the vibrational quantum number v and a change of ±1 in the rotational quantum number J. Absorption transitions for which J changes by +1 are called R-branch transitions and those for which J decreases by 1 are called P-branch transitions. A transition is typically identified using the notation $v'v'' \overset{P}{R}(J'')$, where the double primes refer to the lower state and primes refer to the upper state.

The SiO transition frequencies used in setting the local oscillator frequency were those reported by Lovas et al. (1981). These authors have performed gas phase measurements of $^{28}Si^{16}O$ at the National Bureau of Standards using a tunable diode laser and have combined these results with microwave measurements to yield a set of emperical Dunham ro-vibrational constants. Their results provide transition frequencies accurate to within ± .002 cm^{-1} for the 5 SiO lines of interest. The Dunham coefficients obtained by Lovas et al. and the expression used to obtain the SiO level energies from the coefficients are shown at the right in Figure 18.

The choice of sunspot SiO lines was determined by accidental coincidences with the frequencies of the diode laser modes, and also by avoiding the regions of heavy water vapor absorption. Figure 19

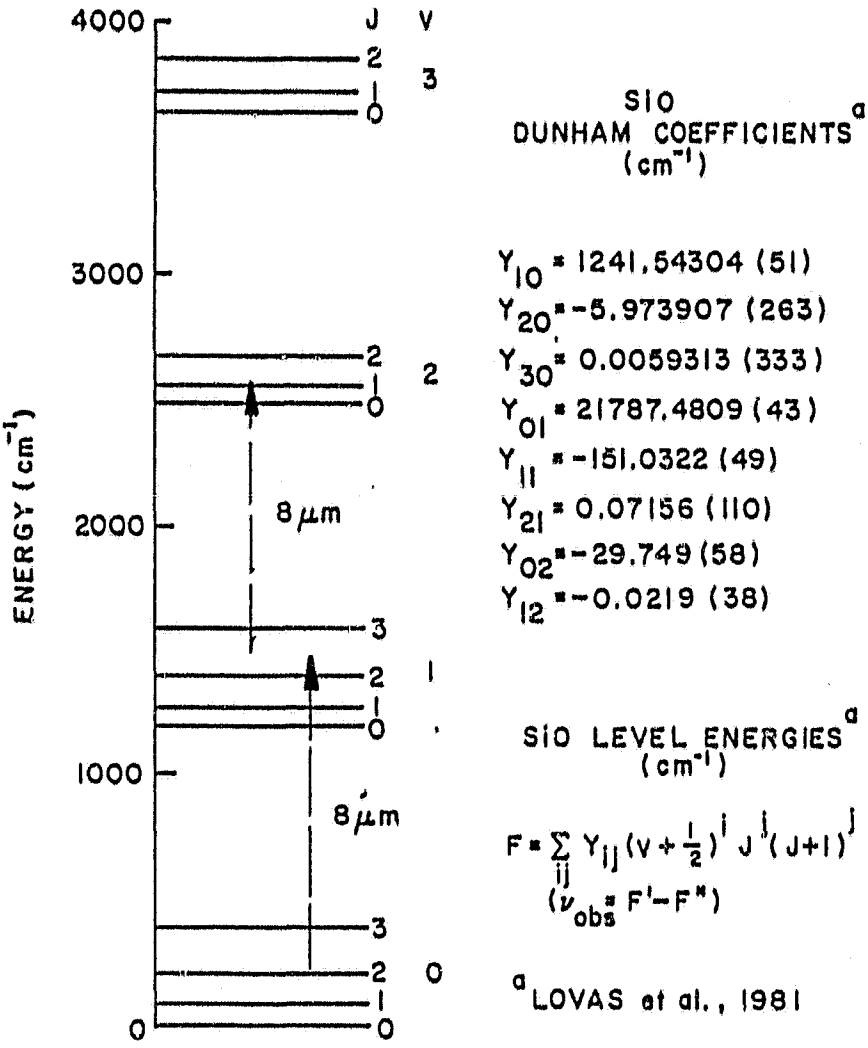


Figure 18. SiO energy level diagram and Dunham coefficients for determining level and transition energies.

shows high resolution ground based FTS measurements by Goldman and Blatherwick (1980) of terrestrial atmospheric absorption near the frequencies of the sunspot SiO lines. These measurements were made at low solar zenith angle (about 5 airmasses) to enhance weak terrestrial atmospheric features, particularly those of trace constituents. The positions of each of the 5 SiO lines are indicated on Figure 19 to show their proximity to major terrestrial absorption features. For some frequencies, tropospheric H_2O is clearly the dominant observer and observations near the wings of H_2O features are very sensitive to the amount of tropospheric water vapor. The 1-0 P(20) line in Figure 19c, for example, was sensitive to atmospheric haze and one of the P(20) measurements was terminated early because of a gradual decline in solar continuum signal during the measurement. For the remaining lines, atmospheric absorption is not as severe as Figure 19 would suggest, since the heterodyne observations were made at small airmass (<1.6).

5.2.2 Observational Method

The synchronous detection scheme and observational sequence for the SiO measurements was chosen to: remove the terrestrial absorption spectrum to reveal only solar features; have separate access to both sunspot and photospheric absorption spectra for purposes of identifying terrestrial features and; obtain absolute sunspot flux estimates. Separate measurements of the umbra and of the surrounding photosphere were made by placing a blackened metal screen in the reference beam to provide a room temperature reference

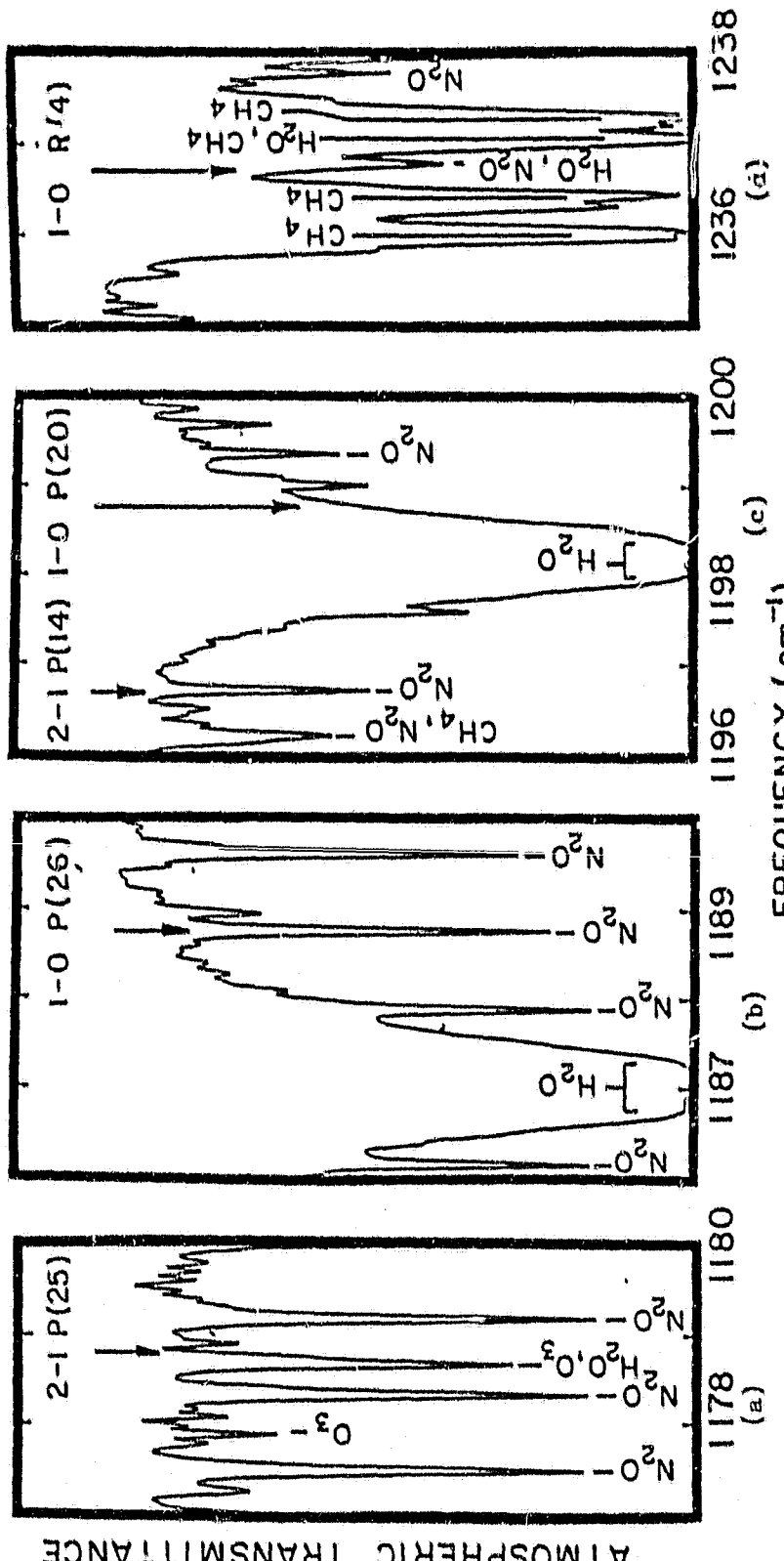


Figure 19. Locations of observed solar SiO features in the terrestrial absorption spectrum.

and by chopping source against screen. The sequence for each SiO observation consisted of the following:

(i) N consecutive pairs of measurements of sunspot against screen followed by photosphere against screen. This interlacing of umbral and photospheric scans minimizes the effects of time variations of atmospheric transmittance and changes in LO performance on the composite ratio spectra.

(ii) M consecutive measurements of blackbody against card followed in some cases by an N_2O cell inserted in the blackbody beam.

(iii) Post - measurement check of LO frequency drift by checking the repeatability of diode current settings at the locations of reference N_2O or SO_2 lines in direct absorption. In the case of the P(14) measurement, the reference N_2O line is close enough to examine in the heterodyne mode and step (iii) is not necessary.

Frequency calibration prior to each measurement was accomplished in two stages, using the method discussed in Section 4.6. A monochromator was used to preset the LO frequency to within ~ 0.1 cm^{-1} and either N_2O or SO_2 was used as a frequency standard to reduce the uncertainty to several milli-wavenumbers or less. Prior to each measurement, the SiO rest frequencies were corrected for solar rotation at the time of observation and position on the solar disk. The TDL local oscillator was then adjusted to locate the feature at the desired intermediate frequency. The 1-0 P(26) and 2-1 P(14) SiO transitions were close enough to reference N_2O line

positions to enable heterodyne measurements of both features at the same LO setting. Since the N_2O line positions for $J'' < 60$ are known to within ± 0.001 cm^{-1} in this frequency range, independent measurements of these SiO rest frequencies could be made to within ± 0.003 cm^{-1}, the additional uncertainty resulting from small amounts of local oscillator frequency drift during the measurement.

Figure 20 shows a sample data set for the detection of the 2-1 P(14) line ($\tau \sim 770$ sec). This example is instructive since the SiO feature was strong enough to be clearly seen in the un-ratioed spectrum (b) and since an N_2O absorption feature at $\nu = 1195.695$ cm^{-1} resides in the opposite sideband. Terrestrial N_2O can be seen in the solar spectra (a) and (b), as confirmed in (c) where an N_2O gas cell has been placed in front of a local blackbody source. Although the average of the photosphere integrations in (a) and that for the sunspot in (b) represent independent measurements, both show highly correlated noise features. These result since both measurements are ratioed by the same blackbody average, taken about 1/2 hour later. Over this period of time, small variations in the diode laser intermediate frequency noise spectrum occur; enough to introduce large amounts of anomalous noise in the ratio spectrum. The sequence of collecting data was chosen to enhance the detectability of sunspot SiO features at the expense of signal-to-noise in the terrestrial absorption spectrum. Because of the tight interlacing of sunspot and photosphere scans, and since the relative umbral intensity is independent of blackbody, this noise virtually disappears in the ratio spectrum (d) which shows the

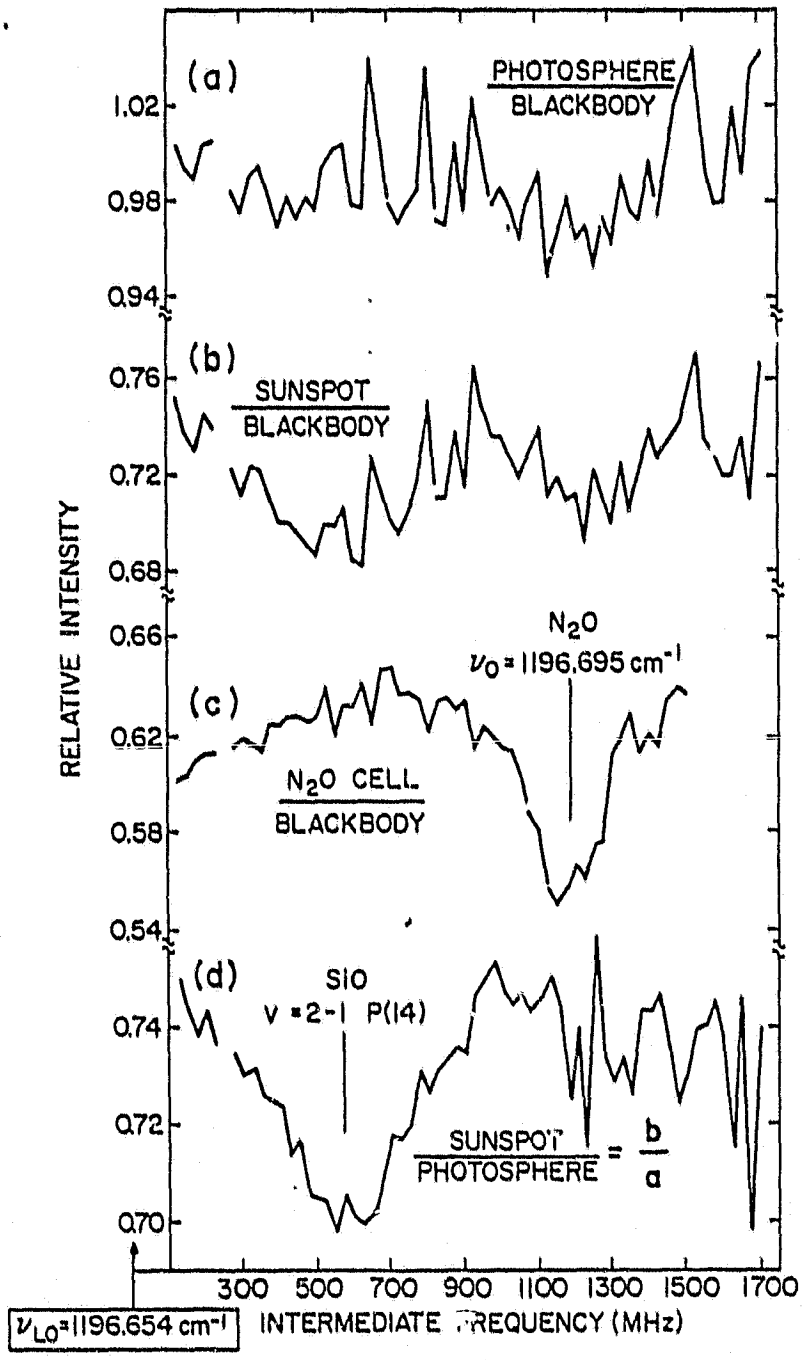


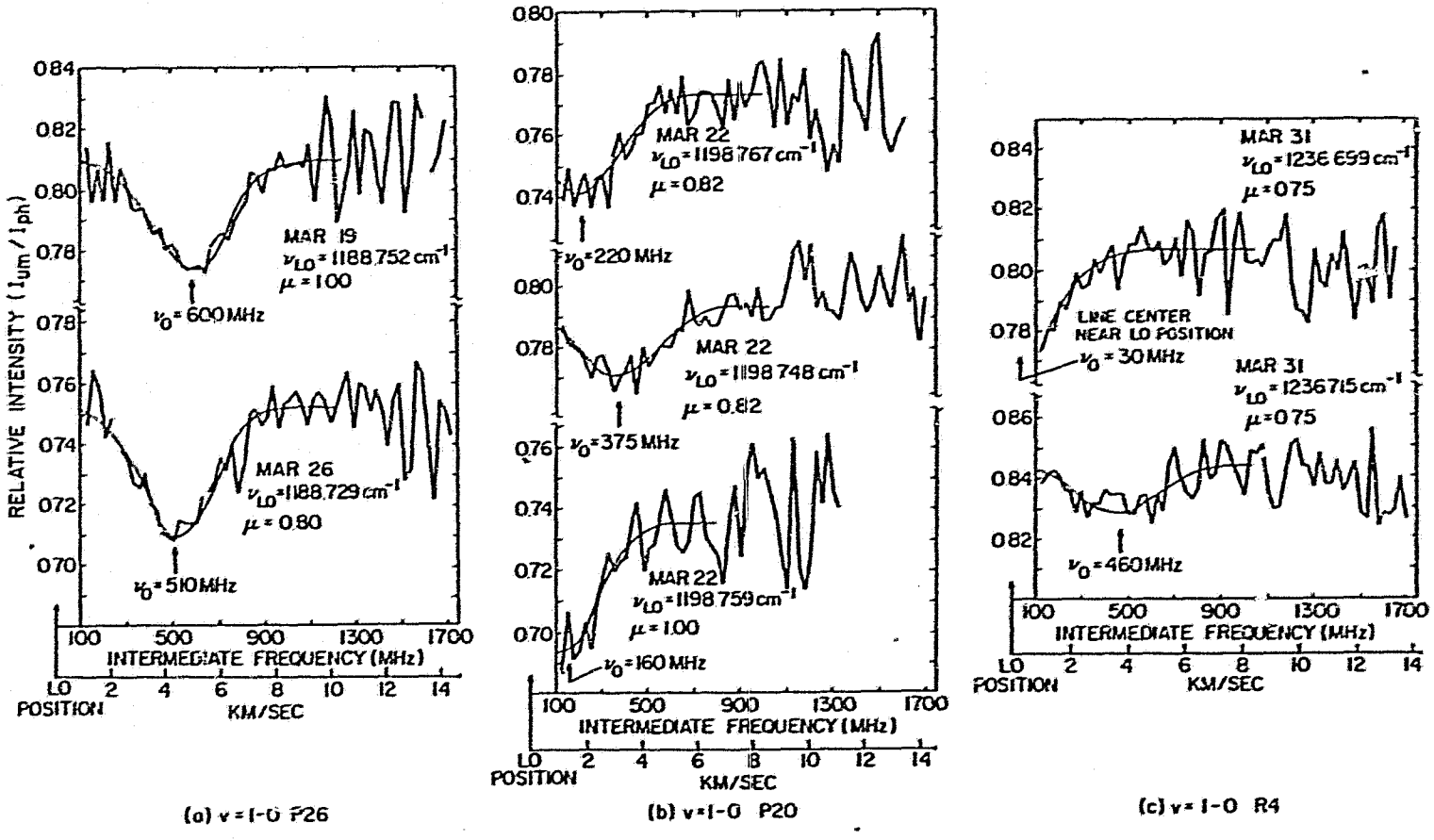
Figure 20. Sample sunspot SiO absorption measurement using the TDL heterodyne spectrometer.

SiO feature.

5.3 Method of Fitting Observed Double Sideband Features

The sunspot SiO observations using the TDL heterodyne spectrometer are shown in Figure 21. The ordinate for each spectrum is the umbral intensity ratioed to that of the surrounding photosphere which, to first order, removes the frequency dependence of terrestrial absorption. Continuum signal-to-noise ratios for these spectra were typically 300 to 400 near the zero of IF, decreasing to less than ~50 at high frequencies because of decreased response from the photomixer and IF amplifiers (Figure 14). For this reason the absorption features were positioned at IF frequencies below ~700 MHz.

From each SiO observation it is possible to extract 4 quantities; (i) the continuum intensity and brightness temperature, (ii) the line equivalent width, (iii) the line full width at half-maximum (FWHM) and (iv) the residual intensity at line center. The data in Figure 21, however, show the average of the line profiles in one sideband with the continuum in the other. Also, in cases where the line center is near the zero of IF, part of the line shape extends into the opposite sideband, resulting in a distorted double sideband spectrum. In order to extract the unfolded line shape from the data, and estimate quantities (i)-(iv), the line fitting routine described in Appendix A was used to generate synthetic double sideband gaussian line shapes. These profiles were plotted over the data, and good visual agreement was



(a) v = 1-0 P26 (b) v = 1-0 P20 (c) v = 1-0 R4

Figure 21a. Observed v = 1-0 SiO absorption features.

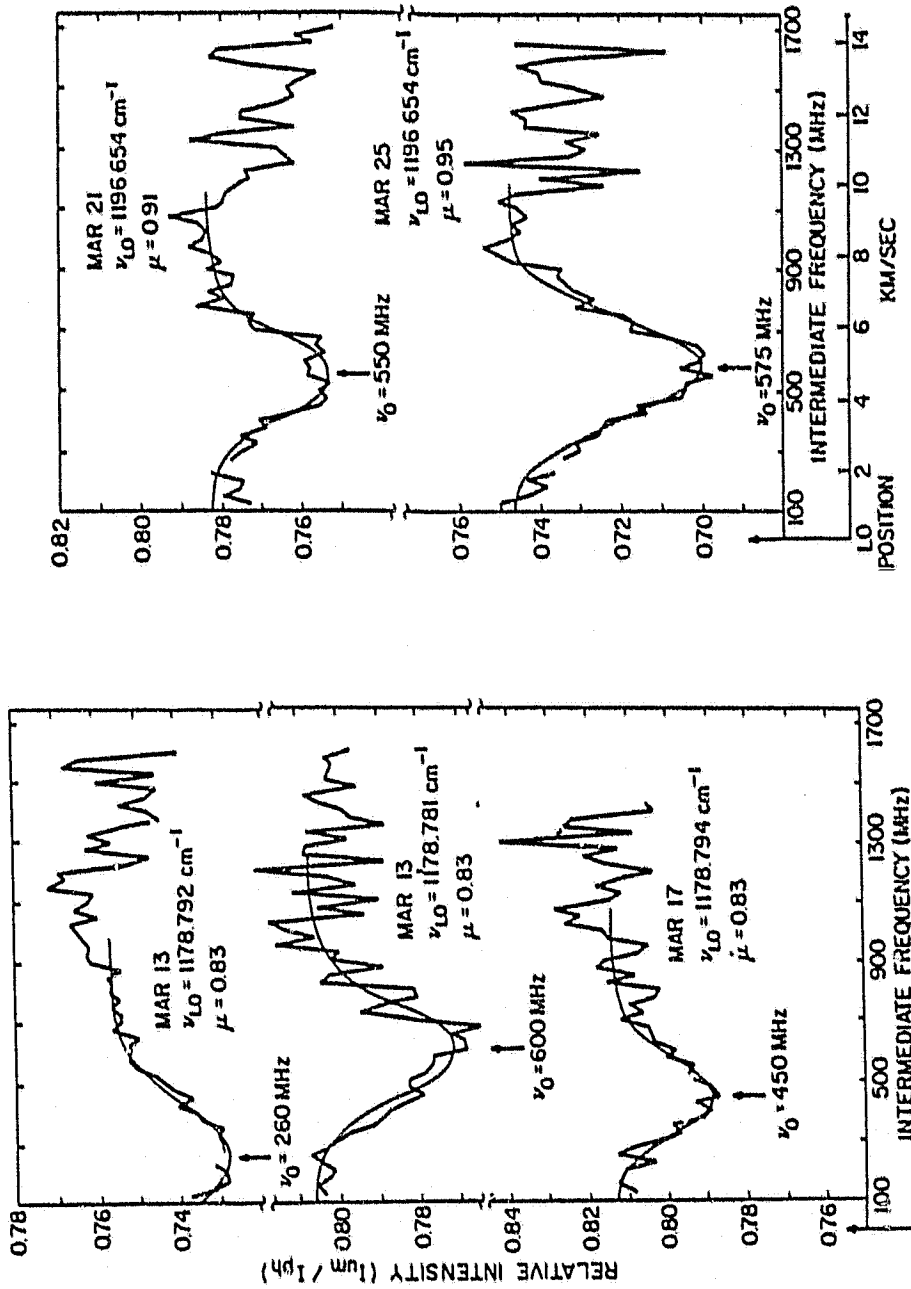


Figure 21b. Observed v = 2-1 SiO absorption features.

obtained by adjusting the line center frequency, line depth and microturbulence. The doppler width for the SiO lines is roughly constant over a reasonable range of temperatures in the region of line formation ($\Delta\nu_D/\nu$ = 6.0 X 10^{-6} at 3500 K and 6.5 X 10^{-6} at 4000K).

5.4 Results of the Measurements

Table 4 lists the measured frequencies, sunspot continua and line shape parameters deduced from the observations in Figure 21. The umbral continuum intensities measured relative to the photosphere are designated ϕ'_c for consistency with the notation in Chapter VI. The corresponding line center intensities are $\phi'(\nu_o)$, and the measured equivalent widths are designated W'. Limb darkening corrections for these measurements are negligible, except for the R(4) observation. Even in this case, the correction is minor and has been treated in Section 6.3. The measured line halfwidths in Table 4 refer to the full width at half maximum.

TABLE 4

SiO Measurements

Transition	Measured[a] ν_o (cm^{-1})	Continuum Intensity ϕ_c'	Line Depth $(1 - \phi'(\nu_o)/\phi_c')$	W' ($cm^{-1} \times 10^3$)	Line Halfwidth $\Delta\nu/\nu \times 10^5$
1-0 R(4)	_____	0.807	0.056	0.83	1.01
"	_____	0.844	0.037	0.56	1.04
P(20)	_____	0.773	0.082	1.18	1.02
"	_____	0.793	0.061	0.98	1.13
"	_____	0.735	0.102	1.28	0.88
P(26)	_____	0.810	0.089	1.36	1.08
"	1188.751(002)	0.753	0.113	1.58	1.00
2-1 P(14)	1196.675(003)	0.783	0.075	0.94	0.88
"	1196.675(003)	0.747	0.130	1.83	0.99
P(25)	_____	0.757	0.078	1.25	1.15
"	_____	0.807	0.089	1.29	1.03
"	_____	0.814	0.059	0.74	0.90

[a] SiO frequencies measured relative to N_2O features using N_2O line position data by Olson et al. (1981).

CHAPTER VI

ANALYSIS

6.1 The Line Synthesis Program (SPCTRM)

6.1.1 Outline of the Program

In order to use the SiO measurements to test the model temperature-pressure profiles, it is necessary to have a means of computing line and continuum intensities based on the input model. Calculations of SiO line and continuum intensities were made by a modified version of a synthetic spectrum program, called SPCTRM, described by Deming (1978). Figure 22 shows an outline of the sequence of steps used by SPCTRM in computing the sunspot surface intensities and line equivalent widths. The program accepts as input a variety of atomic and molecular data, along with a set of temperature/pressure points which define the model atmosphere. From this it computes the atomic, molecular and electron number densities, partition functions and opacities. The radiative transfer equation is then integrated through the atmosphere to yield the emergent line and continuum intensities and the line equivalent widths. The synthetic line shape information can then be compared with the observations.

6.1.2 Computing the Number Densities of each Species

The program uses subroutines from ATLAS5 (Kurucz, 1970) to obtain the equilibrium number densities for all species (i.e., atoms, ions, electrons and molecules) after reading the abundance of

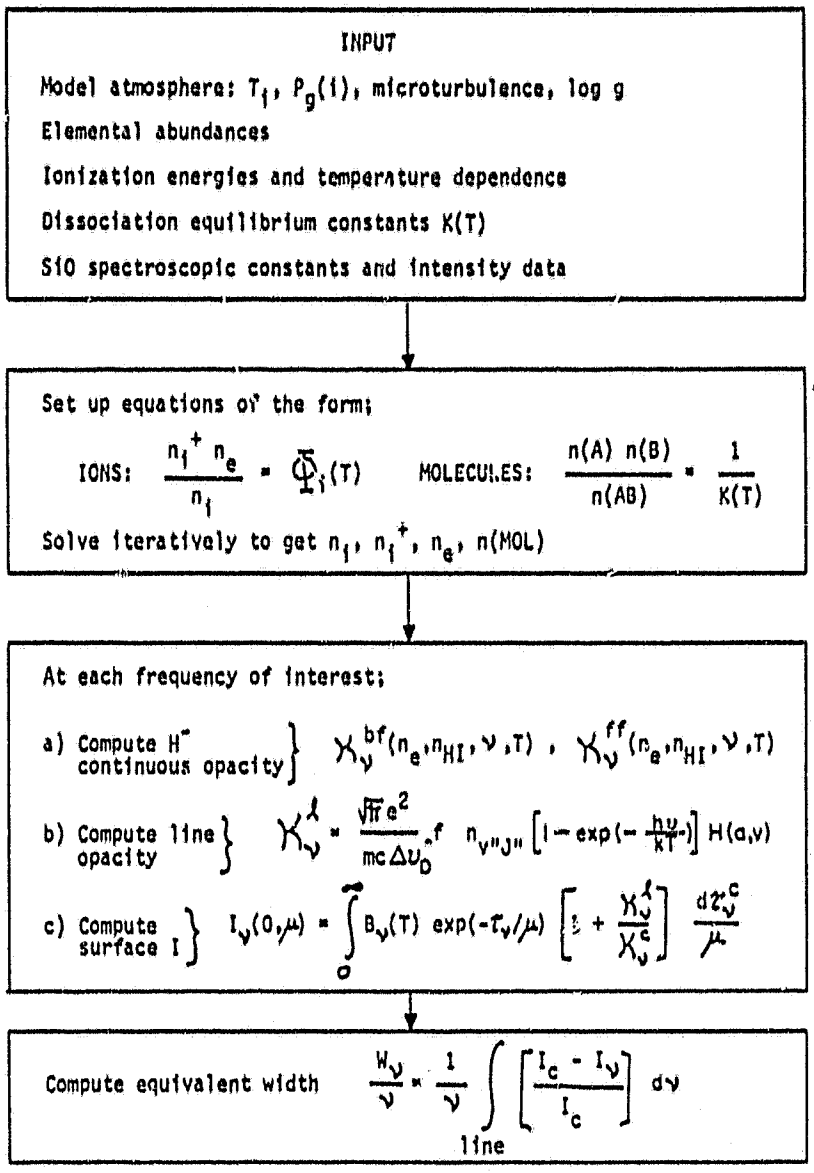


Figure 22. Flowchart for the line synthesis program SPCTRM.

all elements relative to hydrogen. These elements are taken to be independent of depth in the atmosphere.

The silicon abundance for the SiO calculations was taken to be the photospheric abundance as determined by Lambert and Warner (1968). These authors obtained a value of $\log_{10} N(Si) = 7.55 \pm 0.10$ on a scale where $\log_{10} N(H) = 12.00$, using observations of selected SiI lines and intermediate coupling calculations for the SiI spectrum. The results are consistent with an earlier estimate of log N(Si) = 7.50, derived by Goldberg, Muller and Aller (1960).

To determine the number densities for all species, ATLAS5 sets up constraint equations involving the number conservation for each element, plus two additional equations; one for charge conservation and one for the total density. This results in a 2-dimensional matrix of non-linear equations with a large number of cross terms, which are iteratively solved until a self-consistent solution for the neutral atomic densities are obtained. The density of the other species (i.e., ions, molecules and electrons) can then be computed in straightforward fashion.

At photospheric temperatures and below, H is largely neutral and the main contributors to the ambient electron density are the less abundant heavier elements (metals) because of their low ionization potentials. The ionization equilibrium for each heavy element (i) that donates to the electron density is given by a Saha relation of the form

$$n^+_i \, n_e \, / \, n_i = \Phi_i \, (T) \qquad (6.1)$$

where Φ_i (T) is known function of T involving the ionization energy of species i. All of the heavy elements thus form a coupled set of equations, whose simultaneous solution yields the electron density. Partition functions used in solving the Saha equations for most of the atomic species (including H^-) were tabulated by Drawin and Felenbok (1965) and Moore (1949).

The molecular concentrations are obtained from the atomic number densities through a set of temperature dependent equilibrium constants and dissociation energies. For SiO, it is essential to have the best available estimates for the dissociation energy D_o and the equilibrium constant $K_n(T)$. The best current estimate for D_o is given by Hulser and Herzberg (1979), with a value of 8.24 ± 0.13 eV. Values for the SiO equilibrium constant $K_n(T)$, defined by

$$n(SiO) = K_n(T)\ n(Si)\ n(O) \qquad (6.2)$$

were computed from the corresponding pressure equilibrium constants $K_p(T)$ over the temperature range 2800K-4000K. These data were supplied by S. Abramowitz of the National Bureau of Standards (Abramowitz, 1981). Table 5 lists the SiO line strength and dissociation data required by ATLAS5.

A set of coupled equations of this form also exist for oxygen since it is involved in a number of reactions of this type. Those included in the model dissociation equilibrium are (in order of decreasing importance); CO, SiO, H_2O, OH, SO, PO, O_2, MgO, AlO, NaO

TABLE 5

(a) SiO Line Strength Data

Transition	$E''(cm^{-1})$ [a]	$A_{v'v''}(sec^{-1})$ [b]	$f_{v'v''}(X\ 10^6)$	$f_{v'J'v''J''}(X\ 10^6)$ [a]	g	log gf
1-0 R(4)	633.76	5.11	5.07	2.83	9	-4.59
P(20)	923.44	"	"	2.41	41	-4.01
P(26)	1127.67	"	"	2.40	53	-3.90
2-1 P(14)	1999.90	9.96	9.92	4.71	29	-3.87
P(25)	2316.35	"	"	"	51	-3.62

(b) SiO Equilibrium Constants $K_p(T)$, $K_n(T)$ [c]

T(K)	log $K_p(T)$	log $K_n(T)$	T(K)	log $K_p(T)$	log $K_n(T)$	T(K)	log $K_p(T)$	log $K_n(T)$
3000	7.168	-11.221	3400	5.492	-12.843	3800	4.163	-14.123
3100	6.709	-11.666	3500	5.132	-13.190	3900	3.873	-14.402
3200	6.278	-12.083	3600	4.792	-13.518	4000	3.598	-14.666
3300	5.874	-12.474	3700	4.469	-13.829			

[a] Computed from constants by Lovas et al., 1981.
[b] Hedelund and Lambert, 1972.
[c] Values of $K_n(T)$ obtained from $K_p(T)$ provided by Abramowitz (1981).

and ClO. For other molecular species the equilibrium constants were obtained by interpolation from a compendium of high temperature gas thermodynamic properties by McBride et al. (1963).

6.1.3 Opacity Calculations

Continuous opacity in the umbral atmosphere at the wavelengths of interest is due entirely to bound-free and free-free absorption by the negative hydrogen ion H^-. Vernazza et al. (1976) have examined the source function for the H^- bound-free transition in the photosphere at a variety of wavelengths and over a range of optical depths which extend into the lower chromosphere. They find that wherever H^- is an important absorber, no significant difference exists between continuum intensities computed using LTE and non-LTE analyses. In LTE the H^- bound-free opacity κ_ν^{bf} is

$$\kappa_\nu^{bf} = \frac{n(H^-)}{\rho} \; a_\nu \; [1 - \exp(-h\nu/kT)] \qquad (6.3)$$

where

$$n(H^-) = \left(\frac{h^2}{2\pi \, mkT} \right)^{3/2} \frac{n_{HI} n_e}{2} \exp(8762/T)$$

and a_ν is the cross-section taken from Gingerich (1964). The H^- free-free absorption coefficient κ_ν^{ff} is

$$\kappa_\nu^{ff} = \frac{n_{HI}}{\rho} \; n_e \; F_\nu(T) \tag{6.4}$$

The term $F_\nu(T)$ in Equation 6.4 is an analytic fit to the quantum mechanical dipole-length calculations of the H^- free-free opacity by Stilly and Callaway (1970).

$$F_\nu(T) = \left(\frac{1}{\nu}\right)[1.37 \times 10^{-25} + (4.375 \times 10^{-10} - 2.60 \times 10^{-7} T)/\nu]$$

Included in $F_\nu(T)$ is the correction for stimulated emission, which is important at the temperatures and frequencies of interest here.

The effect of scattering can be shown to be completely negligible compared to absorption processes. From Vernazza et al. (1976), the Thompson scattering absorption coefficient is

$$\sigma_{el} = 6.65 \times 10^{-25} n_e \tag{6.5}$$

and the absorption coefficient for Rayleigh scattering is

$$\sigma_{Ray} = R(\lambda) \; n_{HI} \tag{6.6}$$

The cross section $R(\lambda)$ has been computed by Gavrila (1967) for

$\lambda < 2.3$ microns and is a rapidly decreasing function of wavelength. The H^- free-free opacity, on the other hand, is rapidly increasing with wavelength. As an upper limit to the scattering contribution, the ratio $(\sigma_{el} + \sigma_{Ray})/\kappa_\nu^{ff}$ has been evaluated at $\lambda \sim 2.3$ microns and found to be $< 10^{-2}$ for densities and temperatures typical of the umbral atmosphere. The ratio at ~ 8 microns should be several orders of magnitude smaller.

The SiO line opacities κ_ν^ℓ at each layer are obtained from the transition oscillator strengths $f_{v'J'v''J''}$ and the lower level number densities for SiO ($n_{v''J''}$) obtained from the total SiO and the usual Boltzmann factors. The frequency dependence is given by a normalized Voigt profile H(a,v) with $v=(\nu-\nu_o)/\Delta\nu_D$ and the pressure broadening parameter a. The expression for κ_ν^ℓ is

$$K_\nu^\ell = \frac{\sqrt{\pi}\, e^2}{mc\Delta\nu_D} f_{v'J'v''J''}[1 - \exp(-h\nu/kT)]\, n_{v''J''}(SiO)\, H(a,v) \qquad (6.7)$$

The transition oscillator strengths $f_{v'J'v''J''}$ in Equation 6.7 were determined using the experimental band transition probabilities $A_{v'v''}$ as listed by Hedelund and Lambert (1972). These values are 5.11 sec^{-1} and 9.96 sec^{-1} for v'-v'' = 1-0 and 2-1 respectively (Table 5). The transition probabilities are converted to band oscillator strengths $f_{v'v''}$ by the standard relation (Mihalas, 1970; Hedelund and Lambert, 1972)

$$f_{v'v''} = (mc^3/8\pi^2e^2)\ (A_{v'v''}/\ \bar{\nu}^2_{v'v''}) \qquad (6.8)$$

In Equation 6.8, $\bar{\nu}_{v'v''}$ is the average transition frequency (cm^{-1}) over the band. The band center (Q branch frequency) was used for $\bar{\nu}_{v'v''}$ in each case. All other symbols have the usual meanings. The absorption band f values were then converted to absorption transition f values $f_{v'J'v''J''}$ by assigning a lower level multiplicity according to whether the transition belongs to the R branch or the P branch. Following Beer, Lambert and Sneden (1974),

$$f_{v'J'v''J''} = f_{v'v''}\ (\lambda_{v'v''}/\lambda_{v'J'v''J''}) \times [|m|\ F(v'v''J'J'')/(2J'' + 1)] \qquad (6.9)$$

The ratio $|m|/(2J'' + 1)$ is called the Honl-London factor for the transition with $m = J'' + 1$ for an R-branch transition and $m = -J''$ for a P-branch transition, $2J'' + 1$ being the total lower level multiplicity. The term $F(v'v''J'J'')$ is the Herman-Wallis factor, which lies between 0.98 and 1.09 for $J, J'' \leq 78$. A value of F=1 has been used in this analysis.

At sunspot temperatures (3500-4000K) the SiO vibrational population ratio n(v=1)/n(v=0) is ~0.6. The larger v=2-1 oscillator strengths compensate for the population difference, however, so the 1-0 and 2-1 line intensity ratios are of order unity for equal J. The assumption of vibrational equilibrium for SiO was also checked and found to be valid for the sunspot environment. Computations

performed by Nuth and Donn (1981) for collisions of SiO with atomic hydrogen show that for all temperatures $\lesssim$ 3000 K the kinetic and vibrational temperatures are very nearly equal for H pressures > 10^{-5} atmospheres. For comparison, SiO formation in sunspots occurs at layers where pressures exceed 10^{-3} atmospheres.

It is also of interest to examine the terms that make up the total level multiplicities for the v=1-0 and v=2-1 bands of SiO since these directly determine the lower level populations and hence the observed intensities. For a heteronuclear diatomic molecule, where interchange symmetry does not apply, the nuclear spin contribution g_n to the total statistical weight of a level is $(2I_a+1)\ (2I_b+1)$, I_a and I_b being the nuclear spin quantum numbers. The nuclear spin part $g_n=1$ for SiO since $I(^{16}O) = I(^{28}SI) = 0$ (Steinfeld, 1974). Also, SiO has a simple $^1\Sigma^+$ electronic ground state configuration. This means that both the electronic orbital and spin angular momentum quantum numbers Λ and Σ for the electronic ground state are zero. As a result the electronic contribution $g_E=1$, and the total level multiplicity, including rotation, $g = g_R g_N g_E = g_R = 2J'' + 1$ for SiO, as expected.

6.1.4 Computing the Surface Intensities and Equivalent Widths

In differential form the radiative transfer equation at a point in the atmosphere is

$$\mu\ dI_\nu/dz = j_\nu\ \rho - I_\nu k_\nu\ \rho \qquad (6.10)$$

with; dz a length element in the vertical direction

μ cos θ, where θ zenith angle on the sun

j_ν the mass emission coefficient in erg gm^{-1} $sec^{-1}Hz^{-1}$ $ster^{-1}$

ρ the density in gm

I_ν the intensity in erg cm^{-2} sec^{-1} Hz^{-1} $ster^{-1}$

k_ν the total absorption coefficient, including continous opacity $\kappa_\nu^{\,c}$, line opacity $\kappa_\nu^{\,\ell}$ and scattering σ_ν.

$k_\nu = \kappa_\nu^{\,c} + \kappa_\nu^{\,\ell} + \sigma_\nu$

Since scattering is negligible, only the absorption opacities contribute to k_ν. Defining an element of optical depth $d\tau_\nu = -k_\nu \rho dz$, and the source function $S_\nu = j_\nu/k_\nu$, equation 6.10 can be rewritten

$$\mu \, dI_\nu / d\tau_\nu = - S_\nu + I_\nu \qquad (6.11)$$

The optical depth scale is a direct measure of the absorption along a path which in turn depends on the local densities and temperature. for this reason, it is a more useful indicator than geometric depth in the atmosphere. The source function S_ν is in general very complicated since it involves the coupled radiation field and the bound and continuum states for all particles in the stellar gas. Under the assumption of L.T.E, however, absorption and emission are related by Kirchoff's formula

$$j_\nu = (\kappa_\nu^{\,c} + \kappa_\nu^{\,\ell}) \, B_\nu(T) \qquad (6.12)$$

so that

$$S_\nu(T) = B_\nu(T) \tag{6.13}$$

Using these definitions and assumptions, Equation 6.2 can be integrated over optical depth to get the surface intensity;

$$I_\nu\ (o,\mu) = \int_o^\infty B_\nu(T)\ \exp\ (-\tau_\nu\ /\ \mu)\ (1 + \kappa_\nu{}^{\ell}\ /\kappa_\nu{}^{c})\ d\tau_\nu{}^{c} \tag{6.14}$$

with

$$\tau_\nu = \int_o^{\tau_\nu{}^{c}} (1 + \kappa_\nu{}^{\ell}\ /\kappa_\nu{}^{c})\ d\tau_\nu{}^{c}$$

Finally, the equivalent width W_ν for each SiO line profile can be evaluated by SPCTRM for direct comparison with the observations.

$$W_\nu = \int_{\text{Line}} [1 - I_\nu(o,\mu)/I^{c}(o,\mu)]\ d\nu \tag{6.15}$$

In Equation 6.15, the integral is taken over the line profile, and $I^{c}(o,\mu)$ is defined as the surface continuum intensity in the vicinity of the line.

6.2 Method of Correcting Measured Equivalent Widths

Before the observed SiO equivalent widths can be used to predict the umbral temperatures, they must be corrected for the effects of scattered light. Because of marginal observing conditions during most of the SiO observations, the effects of scattered light from the surrounding photosphere were found to be significant. This was particularly true for the R(4) measurement, with umbral diameters less than $\sim$ 7 arcsec, and for the Mar. 17 P(25) measurement with noticeable telescope wind shake. The results suggest an effective heterodyne field of view comparable to or greater than the smaller umbral diameters, which is almost entirely the result of observing conditions. By comparison, the 8 micron diffraction limited field of view (HPBW) for a 48 inch telescope is $\sim$ 2 arcsec. Contamination from photospheric light is suggested in the observational data of Tables 3 and 4, which show a correlation between spot size and measured umbral intensity. The small spot R(4) observations on Mar. 31, for example, show intensities between 0.81-0.85. On the other hand, the observations of the 2-1 P(14) and 1-0 P(26) features on Mar. 25 and 26 and the third 1-0 P(20) observation on Mar. 22 were made in larger umbral regions and show intensities between 0.73-0.76. This suggests contamination by photospheric light in the smaller features. There is also no

evidence to indicate that even the large umbrae fill the heterodyne field of view and hence that even these lower intensity ratios are reliable approximations to the actual umbral continuum intensity. It is also not clear whether the relation between spot size and measured intensity is totally due to photospheric contamination, since changes in continuum brightness temperature of several hundred Kelvin between measurements of large and small spots have been reported at 2.3 microns, the temperature increasing as the spot size decreases (Hall and Noyes, 1969). The 2.3 micron continuum, however, is formed where most models indicate a steep temperature gradient, and it does not follow that the same intrinsic temperature fluctuations will be observed at higher layers where the 8 micron continuum is formed, and moreover where SiO line formation occurs.

Figure 23 shows temperature as a function of both optical depth and atmospheric pressure P_g for three competing sunspot temperature models, along with previous sunspot continuum observations used as boundary conditions for these models. The supporting observations are previous sunspot continuum measurements in the 1-4 micron range by Henoux (1968) and Hall (1970), and visible sunspot measurements by Maltby and Staveland (1971) corrected for stray light effects using the Mercury transit at the time of observation. These measurements and models are discussed in Section 1.4. All of these earlier relative intensity measurements were converted to brightness temperature using an empirical photospheric temperature model by Vernazza et al. (1976) which shows disk center brightness temperatures of 5090 K to 5290 K at 8.3 microns. Supporting

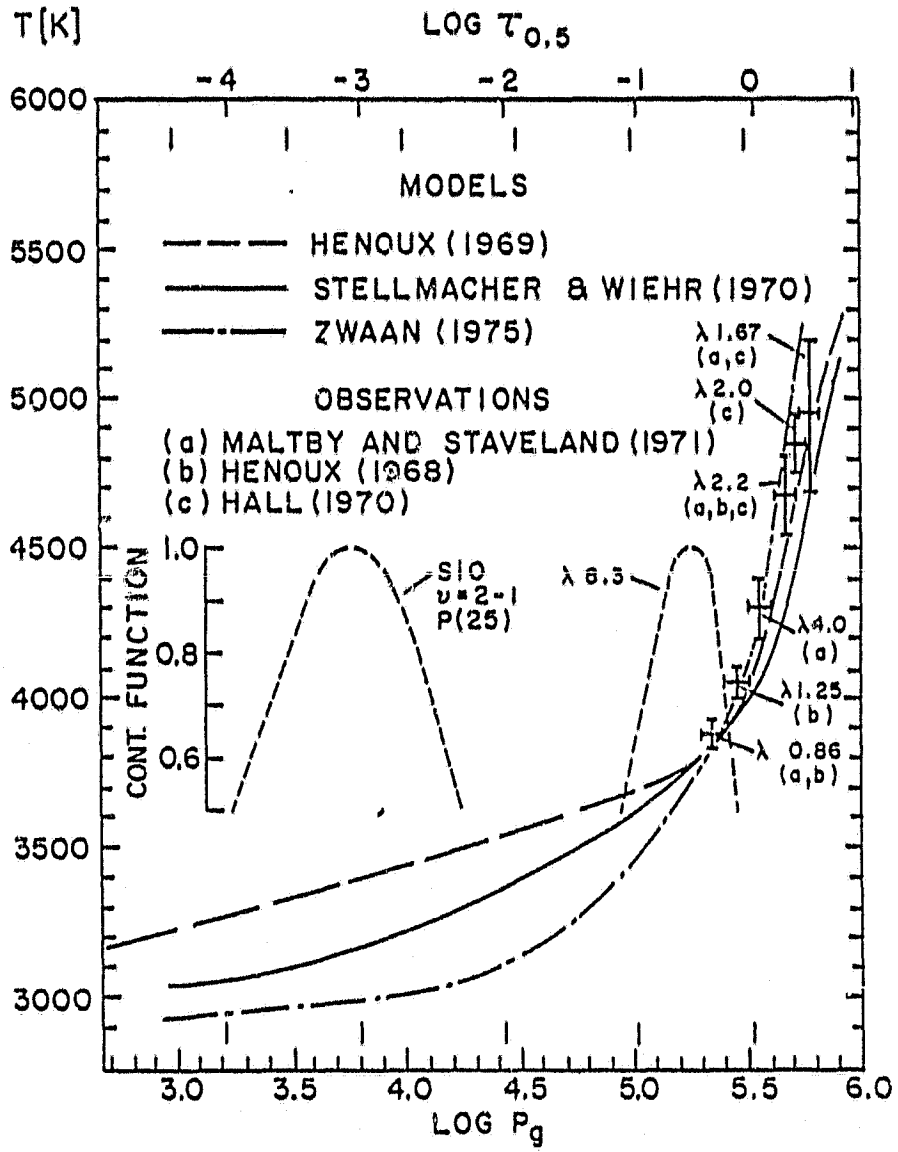


Figure 23. Comparison of sunspot continuum observations with several published temperature models. The SiO line center and continuum contribution functions are also shown.

observations for the photospheric temperatures were high altitude ground based measurements by Kondratyev et al. (1965) and Koutchmy and Peyturaux (1968).

The contribution function for the 1-0 P(26) line center has been included in Figure 23 to indicate the region where the stronger SiO lines are formed and also to define the upper boundary for which SiO observations provide temperature information. The sunspot models rapidly diverge at layers shallower than log P_g = 5.4, since no calibrated long wavelength continuum data are available, and wide disagreement exists between them in the region of SiO line formation. The SiO temperature sensitivity through dissociation is the key factor in removing the temperature ambiguity in these layers orf the atmosphere.

In contrast, the 8.3 micron continuum contribution function is located close to previous reliable continuum measurements. These indicate what brightness temperatures might be expected at 8 microns and allow the data to be corrected for photospheric scattered light. B?r?rasic agreement occurs between all of the observations and models for values of $\log_{10} P_g$ between 5.3 and 5.5. The region near $\log_{10} P_g \sim 5.4$ represents a particularly good choice for a reference layer since the observational uncertainty is small and the temperature models, fitted to these and other observations, are in good agreement. A temperature of 3890 K will thus be adopted for the layer $\log_{10} P_g = 5.4$. The divergence of the models at shallower layers suggests temperatures between 3550 and 3800 K where the 8.3 micron continuum is formed. The highest conceivable temperature

near the peak of the continuum contribution function is about 3800 K, which would occur for a temperature gradient much smaller than that predicted by any of the published models. For comparison the lowest continuum intensity observed using the TDL heterodyne system, ($\phi'_c \sim 0.74$ in Table 3) corresponds to a brightness temperature of ~4000 K, so it is evident that photospheric scattered light poses a problem.

The scattered light correction procedure may be treated as a "filling-factor" problem where the size of the umbra is insufficent to fill the heterodyne field of view. The heterodyne field of view is defined to have unit area on the solar disk and the spot umbra inside the field of view has fractional area A. The umbral filling factor A has been determined for each observation by comparing the measured umbral intensity to the expected intensity near the region of 8.3 micron continuum formation in Figure 23. If $\phi(\nu)$, ϕ_c are the true umbral line and continuum intensities relative to the photosphere, and $\phi'(\nu)$, ϕ'_c are the measured values contaminated by photospheric light, then the fractional line depth $D(\nu) = 1-(\phi(\nu)/\phi_c)$ and the line equivalent width W can be found from the measured values (Appendix E) by

$$D(\nu) = \frac{\phi'_c}{\phi_c A} D'(\nu) \qquad \text{(E5)}$$

and

$$W = \frac{\phi'_c}{\phi_c A} W' \qquad \text{(E8)}$$

with the fractional area A given by

$$A = (1 - \phi'_c) / (1 - \phi_c) \qquad \text{(E4)}$$

Adopting T = 3550 - 3800 K as a plausible temperature range for the region of 8.3 micron continuum formation, and including the ∿ 200 K uncertainty in the photospheric brightness temperature, this leads to values of 0.62 - 0.70 for the umbral intensity ϕ_c.

Table 6 lists the values of the filling factor A and the line depth and equivalent width for each of the measurements corrected to lower intensities. The range of values in Table 6 correspond to the range of values of acceptable sunspot continuum intensities.

6.3 Selection of an Empirical Temperature Model

The adopted emperical umbral temperature model must satisfy two principal criteria. When used to calculate umbral surface intensities, it must yield agreement with the previous continuum intensity measurements. The second requirement is that it must reproduce, within the errors, the measured line equivalent widths, for reasonable choices of microturbulent velocity ξ_M. This establishes the temperatures where SiO line formation occurs.

TABLE 6

Observational Results Corrected for Stray Light

Transition	Filling Factor A		Multiplier $S = \Phi_c'/\Phi_c A$		Line Depth $S(1-\Phi_{\nu_c}'/\Phi_c')$		log (S W /ν)	
	max	min	min	max	min	max	min	max
1-0 R(4)	0.67	0.48	1.8	2.6	0.10	0.14	-5.93	-5.78
"	0.54	0.39	2.3	3.3	0.08	0.12	-5.99	-5.84
1-0 P(20)	0.78	0.57	1.5	2.1	0.12	0.17	-5.84	-5.69
"	0.71	0.52	1.6	2.3	0.10	0.14	-5.88	-5.73
"	0.91	0.66	1.2	1.7	0.12	0.17	-5.89	-5.74
1-0 P(26)	0.66	0.48	1.8	2.6	0.16	0.23	-5.68	-5.53
"	0.85	0.62	1.3	1.9	0.15	0.21	-5.75	-5.60
2-1 P(14)	0.75	0.54	1.6	2.2	0.12	0.17	-5.91	-5.76
"	0.87	0.63	1.3	1.8	0.17	0.23	-5.71	-5.56
2-1 P(25)	0.84	0.61	1.3	1.9	0.10	0.15	-5.84	-5.69
"	0.67	0.48	1.8	2.6	0.16	0.23	-5.70	-5.55
"	0.64	0.47	1.9	2.7	0.11	0.16	-5.92	-5.77

Figure 24 shows a set of synthetic line profiles computed by SPCTRM for different assumed microtubulent velocities, using three published umbral temperature models as input (Henoux, 1969; Stellmacher and Wiehr, 1970; Zwaan, 1975). The rapid decrease in SiO absorption because of dissociation is apparent in going from the coolest model (Zwaan, 1975) to the warmest (Henoux, 1969). Line saturation occurs using Zwaans (1975) model atmosphere since unity optical depth at line center has moved out to layers where the temperature gradient is very small (log P_g = 3.5 to 4.0).

To check the value of microturbulent velocity which reproduces the data, the observed line widths were averaged and compared with the profiles in Figure 22. From Table 4, the measured FWHM values in dimensionless units ($\Delta\nu/\nu$) averaged 1.01 X 10^{-5} with a standard deviation of about 10%. This average is probably high, since there was evidence for frequency drift of the local oscillator during several of the measurements. This could amount to an increase of up to 0.3 X 10^{-5} for individual measurements that make up this average. The value of 1.01 X 10^{-5} then, is considered an upper limit for the observed FWHM values. A more reasonable estimate would probably be 0.90 X 10^{-5}, which excludes any of the line shapes of Figure 24 (predicted using Zwaan's model) and also any large values of microturbulent velocity V_T. Reasonable values of V_T appear to be 0.5 km/sec or less by comparison with the models of Henoux (1969) and Stellmacher and Wiehr (1970).

Because the spectral line receiver produces a folded line shape, and because of dilution by photospheric light, it was not

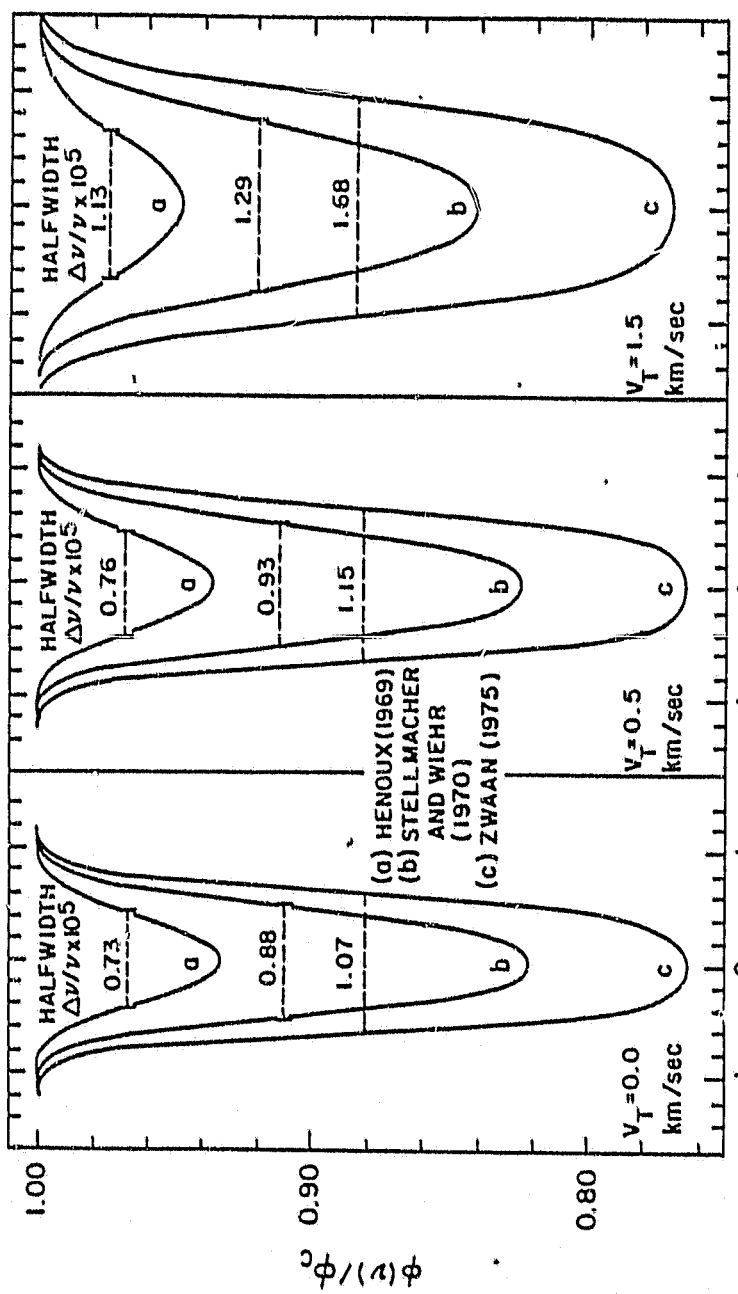


Figure 24. Profiles for the SiO F(20) line computed using SPCTRH. Calculations were performed for published models of (a) Henoux (1969), (b) Stellmacher and Wiehr (1970) and (c) Zwaan (1975) for three different microturbulent velocities.

possible to "fit" the line profiles computed by SPCTRM directly to the measurements. The selection of a temperature model was accomplished by comparing the measured equivalent widths with those computed by SPCTRM for each input model atmosphere. This was most conveniently done by computing a curve of growth (COG) for each model. The advantages of a model curve of growth for the analysis of stellar atmospheres is discussed by Mihalas (1970). For a given model atmosphere, the line shape and hence equivalent width of the line is completely determined once the transition frequency ν, lower level excitation energy χ and line strength gf are specified. For medium and strong features, some estimate of the amount and type of broadening must also be made. At a glance, the curve of growth shows the dependence of W on lower state energy, the temperature sensitivity of W through the choice of models, the character of the observed lines - whether weak or strong, and how the uncertainties introduced by assumptions regarding broadening and the effects of scattered light affect confidence in the choice of an appropriate model.

Figure 25 shows calculated curves of growth for the competing sunspot models. The predicted equivalent widths are independent of broadening for weak lines (small gf values) but depend on broadening for stronger features. Observed equivalent widths corrected using the method of Section 6.2 are also shown and agree nicely with the Stellmacher and Wiehr (1970) model for small values of microturbulent velocity. This model appears to be the best choice for agreement with the SiO observations. Zwaans model yields

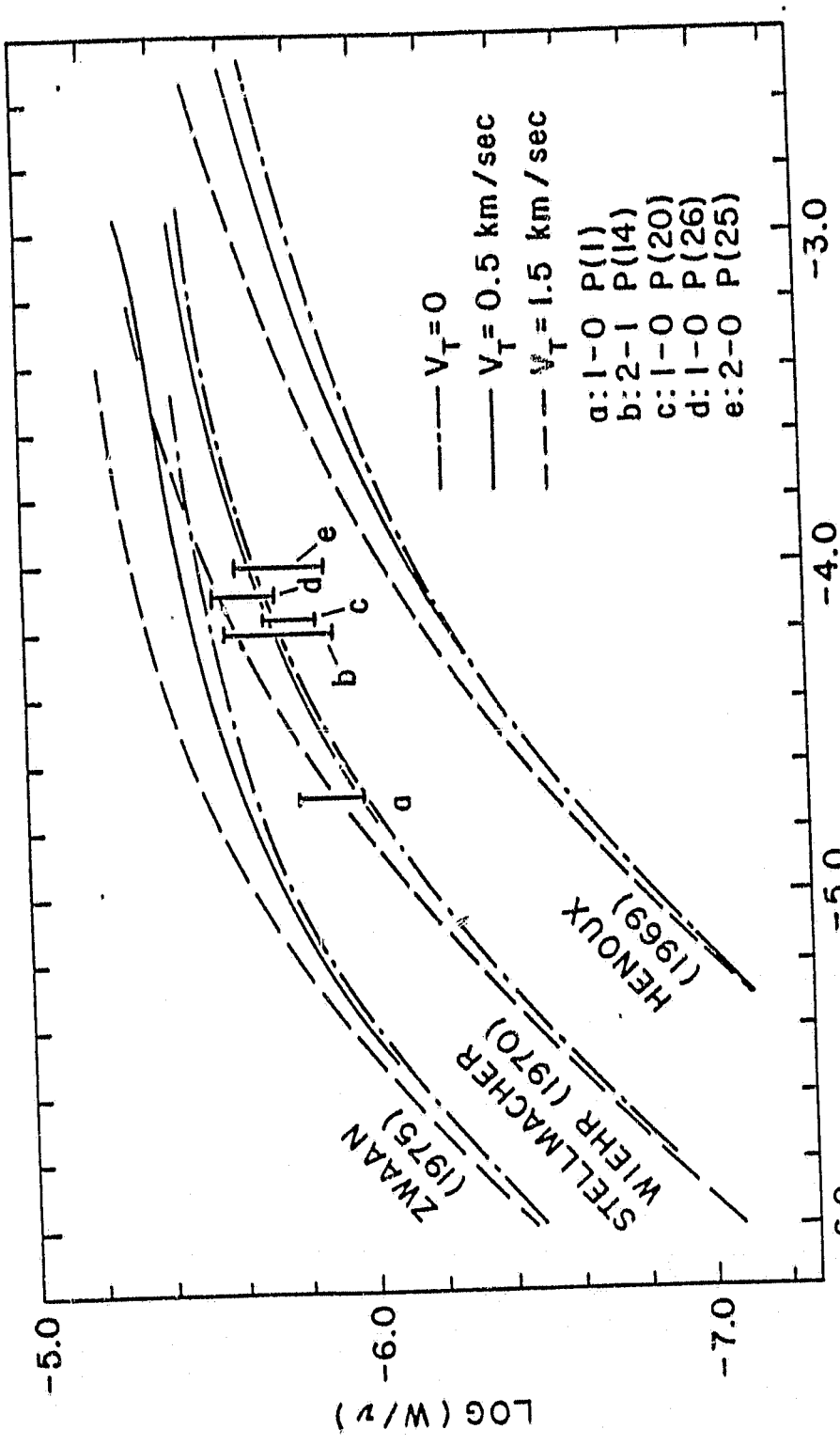


Figure 25. Curve of growth comparison of model equivalent widths with the observed values corrected for photospheric light.

unacceptably large values of W, which might be expected since it is entirely extrapolated for $\log_{10} P_g \lesssim 5.2$. On the other hand, Henoux's model predicts equivalent widths which are consistently too small. The reason for this discrepancy is not clear since the procedure used by Henoux in his analysis of visible Na D lines (Section 1.4) should be fairly temperature sensitive.

The temperature uncertainty in the region of SiO line formation can be inferred from the curves of growth for the Zwaan and Henoux models and from the corresponding temperature profiles in Figure 23. These models, which lie above and below the observations in Figure 25 differ by roughly 0.7 in log (W/ν) where observations b-e are located. From Figure 23, the same models differ in temperature by ∿ 400 K throughout the region of SiO line formation. The choice of acceptable curves of growth which agree with the data are limited to a range of < 0.2 in log (W/ν), so to first order, the temperature in this region is determined to < (2/7) 400K = .115K.

Some mention should be made about the R(4) observation since several aspects of this measurement were different from the others. The excitation energy for R(4) is considerably smaller than for the other lines, making this measurement particularly interesting. However, only a few small sunspots were visible at the date of observation, and these were located a larger distance from disk center than the others ($\mu = 0.75$). The position of W for the R(4) line is somewhat higher on the ordinate of Figure 23 than might be expected based on the slope d(log W)/d(log gf) of the Stellmacher and Wiehr (1970) curve of growth. One might expect the R(4) result

to be located below the curve, since any increase in temperature in the smaller spots should increase the SiO dissociation, decreasing the equivalent width. One possible source of this discrepancy is the μ dependence, which has been taken into account in Figure 23. This affects the equivalent width directly and also indirectly since limb darkening alters the measured continuum intensities, affecting the scattered light correction. These effects were checked by computing the Stellmacher and Wiehr (1970) and HSRA (1971) models over the range from μ=0.6 to 1.0. It was found that relative to disk center both ϕ_0 and log W increased by ∿0.02. These numbers were used to normalize the observations to disk center, resulting in a reduction of ∿14%. This correction is in the right direction, but the measured range of W for R(4) remains anomalously high.

VII

SUMMARY AND CONCLUSIONS

7.1 Summary and Conclusions from the Sunspot SiO Analysis

The measurements discussed in Chapters V and VI constitute the first reported detection of SiO in the solar atmosphere and the first such measurement using coherent detection. Although there are observational uncertainties in the measured equivalent widths, the absence of blending from adjacent lines in the IR and the high temperature sensitivity of the SiO abundance through dissociation permit an unambiguous temperature determination over an extended region in the umbral atmosphere. By combining the TDL heterodyne SiO and continuum observations with previous IR and visible continuum measurements, the vertical temperature structure in the umbral atmosphere has been bounded with an accuracy of $\lesssim$ 100 K over the region log P_g = 3.5-5.5. This permits an unambiguous selection of a temperature model, provided it is well behaved. The published umbral model of Stellmacher and Wiehr (1970) yields good agreement with the observed SiO equivalent widths, so this model has been adopted as the best choice consistent with the TDL heterodyne results. From the temperature sensitivity of the SiO abundance, the lack of large amounts of scatter in equivalent widths observed from spots of different sizes suggests a remarkable consistency of temperature in the region of SiO line formation ($\Delta T < 80$ K).

From a comparison of the measured and computed line widths it

does not appear necessary to invoke any broadening processes other than microturbulence to explain the observations. In fact, the lowest reasonable estimate suggested by Stellmacher and Wiehr (1970) of $V_T \sim 0.5$ km/sec yields line widths comparable to the observations when their model is used as input to the line synthesis routine SPCTRM. Comparing the TDL heterodyne results with Figure 24 suggests that V_T may be even smaller than 0.5 km/sec in the region of SiO line formation (log P_g = 3.5 to 4.0). The absence of evidence for pressure broadening of SiO in the umbral atmosphere is of interest since no laboratory measurements have been made of SiO broadening by foreign gases. An upper limit for broadening of SiO by the stellar gas can be obtained by setting V_T=0 and adjusting the damping constant in the model until agreement with the observations is obtained. This was done for the 2-1 P(14) line using the Stellmacher and Wiehr (1970) model and yielded a value of 1×10^{-4} cm^{-1}/Torr as an upper limit for the broadening coefficient for SiO in the umbral atmosphere.

Regarding future sunspot observations, a number of di- and tri-atomic molecules expected in fair abundance in sunspots have not yet been observed and some of these have rotation-vibration bands in the 8-12 micron region, accessible to the TDL heterodyne technique. From Table 1, those species are; VO, YO, PO, AlO, CO_2, SO, H_2S, H_2, NaH, NH_3, NS, HCN and CS. These species were predicted by Gaur et al.(1973) to be formed at both larger and smaller depths than SiO so that similar observations and analyses of these species should yield

a thorough characterization of the umbral upper atmosphere.

7.2 Additional TDL Heterodyne Observational Problems of General Interest

The sensitivity of the diode laser heterodyne technique and its practicality as a multi-disciplinary observational tool continue to improve. These advances are largely the result of recent improvements in diode laser fabrication and an increased understanding by the user of the noise properties of these devices. With each improvement in heterodyne sensitivity, new observational problems can be considered which lend themselves to the high spatial and spectral resolution as well as the inherent tunability.

Besides the obvious advantages of diode laser heterodyne spectroscopy for the study of molecular lines in the sunspot environment and for sub-doppler observations of telluric atmospheric constituents in solar absorption, the method has great potential for further astrophysical work. Several interesting problem areas should be mentioned here;

1) Anomalous emission features in the solar spectrum, near 12.33 and 12.22 μ have been discovered by Murcray et al. (1981), and have been conjectured to originate in the solar chromosphere or corona. At the observational resolution of 0.02 cm^{-1}, the features were unresolved and exhibited a peak intensity ~10% above the continuum. The fully resolved line shapes obtainable using the heterodyne technique will permit accurate measurements of the kinetic and brightness temperatures for these features. In addition, accurate determination of the line center positions should

provide information on the dynamics in the regions where the features are formed.

2 is demonstrated by the observations of Mars, Venus and IRC+10216, TDL heterodyne spectroscopy is evolving to the point where remote sources other than the Sun can be detected with signal-to-noise adequate for spectroscopic purposes. The measured A of 43, which occurred using the RF spectral line receiver is within a factor of 7 of the best performance from the NASA CO_2 heterodyne system. Part of this difference (a factor of ~2) is accounted for by the reflection characteristics of the ZnSe beamsplitter, which differs between the two systems.

3) Several bright IR stars, e.g. IRC + 10216 (already observed using the TDL heterodyne spectrometer), VY Canis Majoris, W Hydra and α Orionis have also been considered for study after some additional improvement in system sensitivity. A variety of molecular absorption and emission lines have been observed in these and similar objects at both IR and millimeter wavelengths. The list of such detections has grown too large to discuss in this context. Of great interest is observed SiO maser emission in a number of these IR sources. Infrared measurements of vibrational transitions between the levels of interest may answer some long standing questions regarding the mechanisms for pumping these masers.

4) Molecular hydrogen emission has been detected in the Orion region (Beckwith et al., 1978; Beck et al., 1979) and observations of selected H_2 lines at high spatial and spectral resolution should reveal new information about this complex region.

Such measurements will be possible for the H_2 line at 12.28 microns when the system sensitivity is improved and will become possible for H_2 lines at 17.03 and 28.22 microns when long wavelength infrared mixers and TDL's become available.

5) Infrared measurements of some compact HII regions reveal large fluxes in the fine structure lines [ArIII], [SIV], and [NeII] at 8.99, 10.51 and 12.81 microns (Willner, 1977). For a 3 meter ?r?rtelescope aperture, estimates yeild brightness temperatures in these lines as large as 300K, making them attractive sources for diode laser work. Low resolution observations of these lines have also been made in planetary nebula (Aitken et al., 1979) and suggest line brightnesses adequate for high resolution observations with signal to noise ratios larger than $\sim$ 50.

7.3 Future Directions in TDL Heterodyne Instrument Development

In order to make diode laser heterodyne detection fully competitive with (or superior to) the overall performances of gas laser instruments, advances are necessary in the areas of TDL and photomixer development as well as in the overall instrument design.

In order to successfully perform most of the observations above, it is essential to have TDL's with improved (i) power output, (ii) mode purity, (iii) excess noise suppression, (iv) tunability and (v) operating temperature range. One additional long-standing proolem - that of device deterioration by slow contact degradation - has been largely solved. This results in substantially increased lifetime over devices of just a few years

ago. Items (ii) and (iii) are closely related since many of the diode noise properties are coupled to beat effects and switching between competing frequency modes. The elimination of multi-moding in stripe geometry lasers also greatly enhances single mode tunability by restricting operation to only the lowest order transverse mode. As a result, success in improving (ii) should also largely eliminate problems associated with (iii) and (iv).

The HgCdTe IR photomixers incorporated into the system have been provided by D. Spears of MIT Lincoln Labs (Spears, 1977). These devices exhibit high quantium efficiencies in shot-noise limited applications, but have been optimized for use with CO_2 gas laser heterodyne systems which routinely deliver 1mW to the detector. Because of the inherently low power outputs of present TDL's, detectors are required which are optimized for smaller local oscillator powers. Such devices are just now becoming available (Spears, 1981) and should improve the heterodyne efficiency in cases where the available TDL power is split among several usable modes.

One major engineering problem associated with the spectrometer has been the inability to stabilize the diode laser frequency over reasonably long integration times. A related problem is the frequency uncertainty that arises when it's necessary to extrapolate the diode frequency away from precise reference gas line positions. In practice, the frequency stability of the TDL should be less than the single IF channel width of 25 MHz over a period of 1-2 hours, which is the maximum practical integration time for most sources. Present commercially available temperature controllers employing a

heater-sensor servo loop are not capable of consistently satisfying this requirement, particularly when the diode tuning rate exceeds $\sim$ 2.0 cm^{-1}/K. Also, because of non-linearities in the diode tuning rate and because of thermal drift of the Ge etalon fringes, local oscillator frequencies normally cannot be accurately extrapolated more than $\sim$0.01 cm^{-1} from a reference gas line. A reliable method of predicting and holding the TDL output frequency to <0.001 cm^{-1} is necessary for reliable observational work and long integration times. Poultney et al. (1980) have achieved TDL frequency control to $\sim 10^{-4} cm^{-1}$, using reference gas features and a 0.01 cm^{-1} air gap etalon. The TDL frequency was determined relative to the etalon fringes using an interpolation scheme based on a TDL fm analysis of the etalon fringe pattern. Another approach (Savage and Augeri, 1980) uses a servo-controlled Fabry-Perot etalon in which the transmitted LO power is synchronously detected and used to fine tune the TDL current controller. A number of other schemes employing reference gas lines and/or variations of these methods might be employed to achieve the desired frequency precision and long-term stability.

Another design problem results from the inefficiency introduced by the zinc selenide beamsplitter. In the present spectrometer design, half of the useful signal and LO oscillator power is lost at the beamsplitter because of its 50/50 reflection/transmission characteristics. Two plausible schemes for recovering this lost signal are: (i) introduction of a second HgCdTe detector, the output of which is combined in the proper phase with the first

detector using a broad-band IF coupler, (ii) Replacement of the beamsplitter with an all-reflective beam combiner. This second scheme has the potential of being easy to implement, and its inclusion into the spectrometer optical train is presently being considered.

APPENDIX A:

TDL HETERODYNE DETECTION OF ATMOSPHERIC N_2O, CH_4, O_3 AND H_2O

A number of observations of terrestrial features have been made in solar occulation while testing the instrument at the NASA/GSFC 48" reflecting telescope, near Greenbelt, MD. Figures 26 and 27 show heterodyne detection of atmospheric absorption features measured in the sweep integration mode. Figure 26 shows P and R branch lines in the $2\nu_2$ fundamental band of N_2O. The features in Figure 27 belong to the ν_1 fundamental band of ozone, based on line identification studies and position measurements by Goldman and Blatherwick (1980) and Barbe et al. (1981). Each solar measurement was followed by a reference line measurement using an N_2O cell in front of a 1273 K blackbody. A calibrated etalon and a calibrated blackbody continuum spectrum (not shown) were measured for each line allowing quantitative analysis of the line shape. Because the strong N_2O reference lines have separation larger than 0.5 cm^{-1}, the 0.1 cm^{-1} frequency calibration using the infrared monochromator makes it possible to identify each feature and obtain its frequency using N_2O line position standards. A recent compilation of N_2O absorption line standards, accurate to < 10 MHz (Olson et al., 1981) has been used for this purpose.

Both N_2O scans in Figure 26 were obtained in early June 1980 at nearly the same air mass. However, a substantial amount of water vapor was present on the day the R(19) line was obtained and absorption in the wing of a highly broadened H_2O line centered at

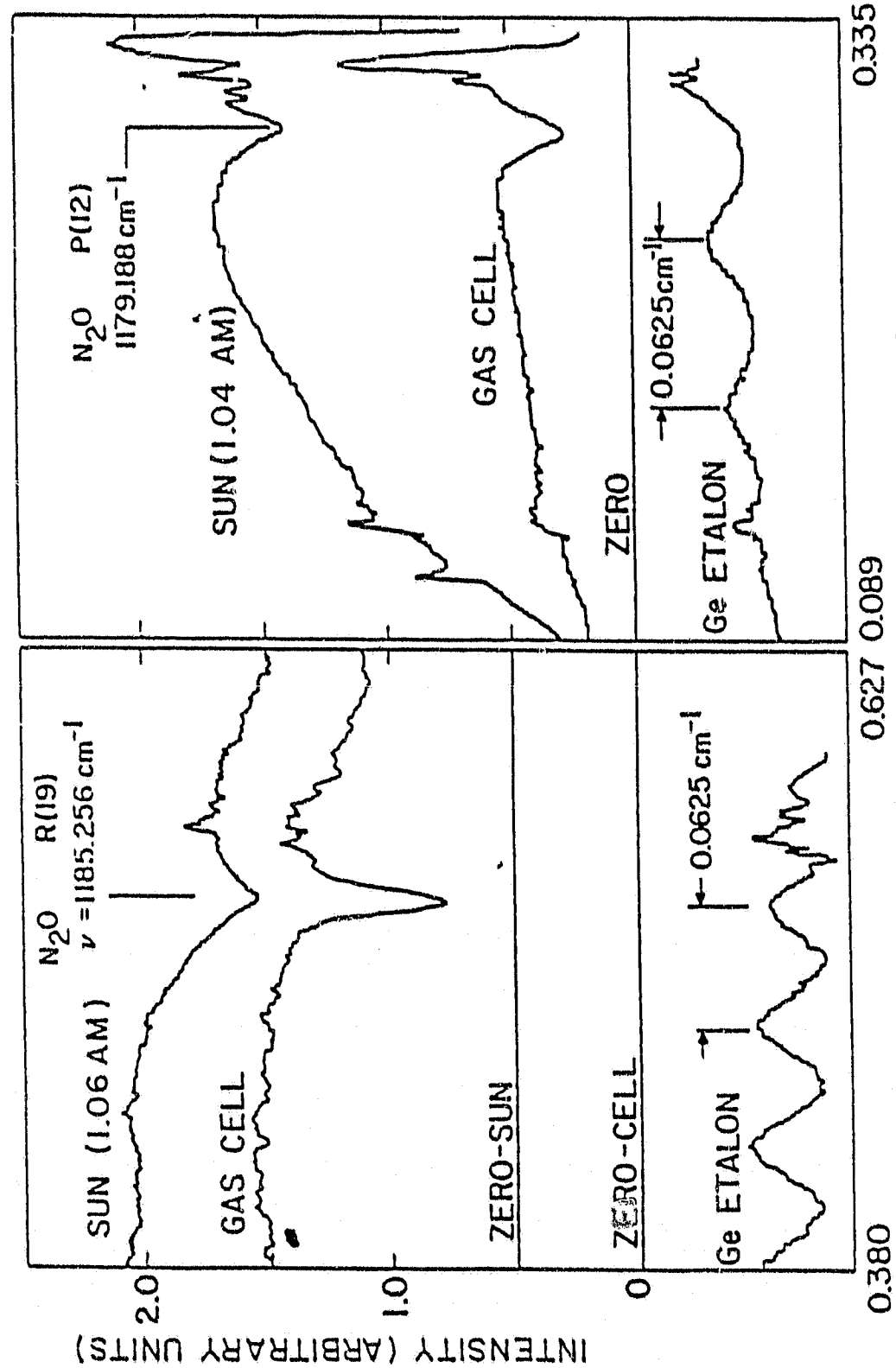


Figure 26. Diode laser heterodyne measurements of lines in the $2\nu_2$ band of atmospheric H_2O using a swept 100 MHz single channel receiver.

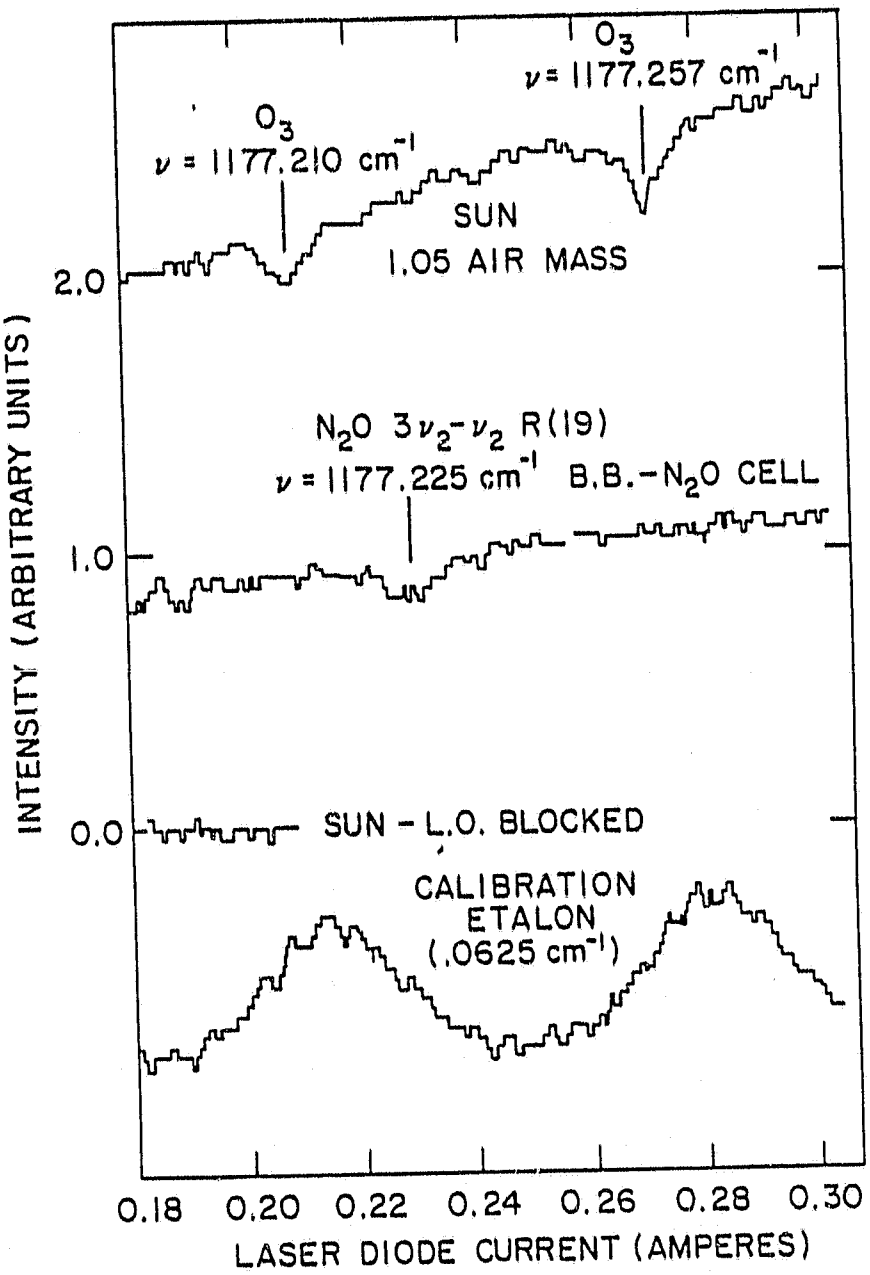


Figure 27. Heterodyne measurement of atmospheric ozone lines in near coincidence with laboratory N_2O using the sweep integration mode.

1187.0 cm^{-1} accounts for the reduced solar continuum signal during that measurement. The R(19) result was obtained by stacking 3 scans, each with a time constant of 1.25 sec for a total integration time (τ) of 3.75 seconds. Only a single scan with τ = 1.25 sec was required for the P(12) measurement because of the larger solar continuum signal and lower TDL noise level.

Most of the noise in Figure 26 results from the choice of amplitude resolution used when storing the scan average in the signal averager. The excess noise, however, results from mode competition in the diode laser which produces large audio frequency fluctations. In practice it is sometimes possible to adjust the diode current and mount temperature to move the interesting feature into a region where noise is not observed. For the P(12) line the double side-band SNR for the line wing is as high as 200 for an IF bandwidth of 100 MHz, integration time of 1.25 sec and blackbody temperature of 1273 K. The expected SNR for an ideal system under these same conditions is about 16,000, making the measured system delta (Δ) about 80 in this region of the scan. A delta as low as 60 has been observed in the single channel sweep mode, although not in close proximity to an absorption feature in any of the subject gases used here.

Figure 28 shows atmospheric methane and ozone observed in solar absorption after interfacing the spectrometer to the 64-channel spectral line receiver. The detection of CH_4 in (a) was confirmed by heterodyning a CH_4 cell in front of the reference blackbody immediately after the atmospheric measurement. Line position standards for methane (Reisfeld et al., 1978) were used to identify

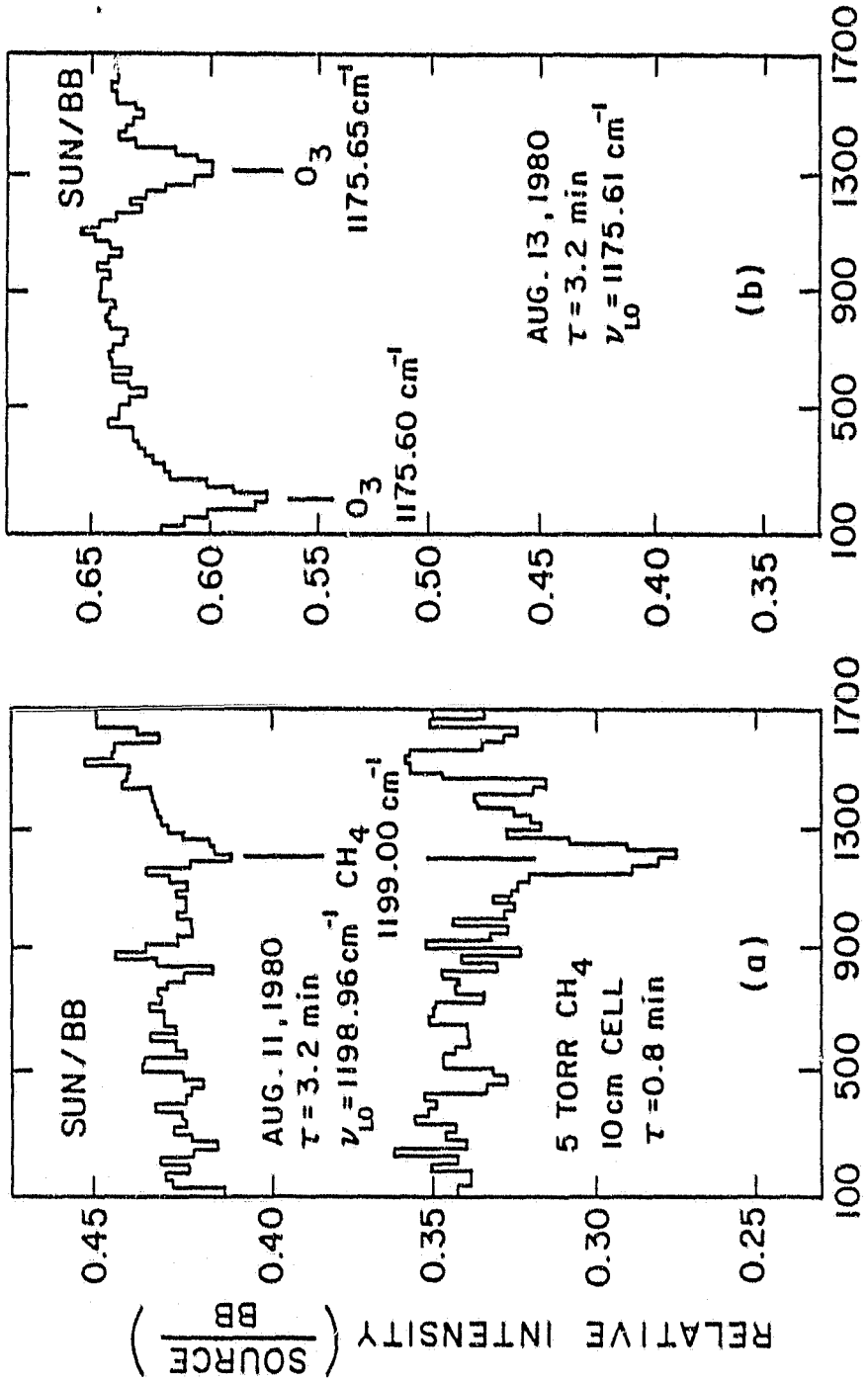


Figure 28. Spectral line receiver measurements of atmospheric CH_4 and O_3, obtained on separate days using different diode laser modes.

the feature, after frequency calibration to within ± 0.1 cm^{-1} using a monochromator. The poor signal to noise in this measurement results from both excess noise in the diode mode and low atmospheric transmission (~25%) at this frequency, as given by the ratio of observed solar signal to reference blackbody signal. The CH_4 feature lies well within the wing of a highly absorbing H_2O line at 1198.2 cm^{-1}. This would account for the heavy absorption, since the atmospheric temperature and relative humidity were both high at the time of measurement.

Absorption from H_2O at 1174.5 cm^{-1} also limits the signal observed in (b). This shows atmospheric O_3 line identifications that were made using SO_2 as a frequency standard. The LO frequencies for these measurements were determined by direct absorption using a 10 cm cell filled with 5 torr of SO_2. By noting the LO current settings at several SO_2 line centers and using line position data for the ν_1 band of SO_2 (Allario et al., 1975) the LO frequency could be found by interpolation. Based on SO_2 abundance estimates (Aikin, 1979), no observable SO_2 absorption is expected in the solar transmittance spectrum. Both of the features in (b) were identified as belonging to the ν_1 fundamental band of O_3 using published ozone line positions by Barbe et al. (1981).

APPENDIX B:

HETERODYNE MEASUREMENTS OF MARS AND VENUS

Observations of Mars and Venus using the TDL heterodyne spectrometer were made during November, 1980 at the coude focus of the NASA 3-meter IR telescope at Mauna Kea, Hawaii. A photo of the spectrometer in the 3-meter coude room is shown in Figure 29. The beam of this telescope is approximately f/120, yielding a 0.5 arcsec/mm field at the coude focus. The 8 micron "seeing" at Mauna Kea is about 1-2 arcsec under good conditions, and because of the altitude ($\sim$4.3 km) the amount of precipital water vapor is very small (typically 1-2 mm). Since H_2O is the dominant absorber at the frequencies of interest, this altitude is of great advantage for IR observations.

The planetary measurements (Figure 30) were made using the 64-channel spectral line receiver and the LO was adjusted to operate using a single powerful Fabry-Perot mode. The mode frequency was measured with the monochromator to be 1181.4 cm^{-1}. Examination of 0.06 cm^{-1} resolution ground-based atmospheric transmittance data by Goldman and Blatherwick (1980) shows that this frequency lies between terrestrial N_2O features at 1180.9 and 1181.8 cm^{-1}. No other major terrestrial absorption features exist near the observing frequency so nothing was expected to appear in the spectra, given the observed signal to noise ratio. The reference 1300 K blackbody measurement was made prior to the planetary measurements and shows an operating system Δ of $\sim$ 45 near 400 MHz IF. The blackbody measurement is shown

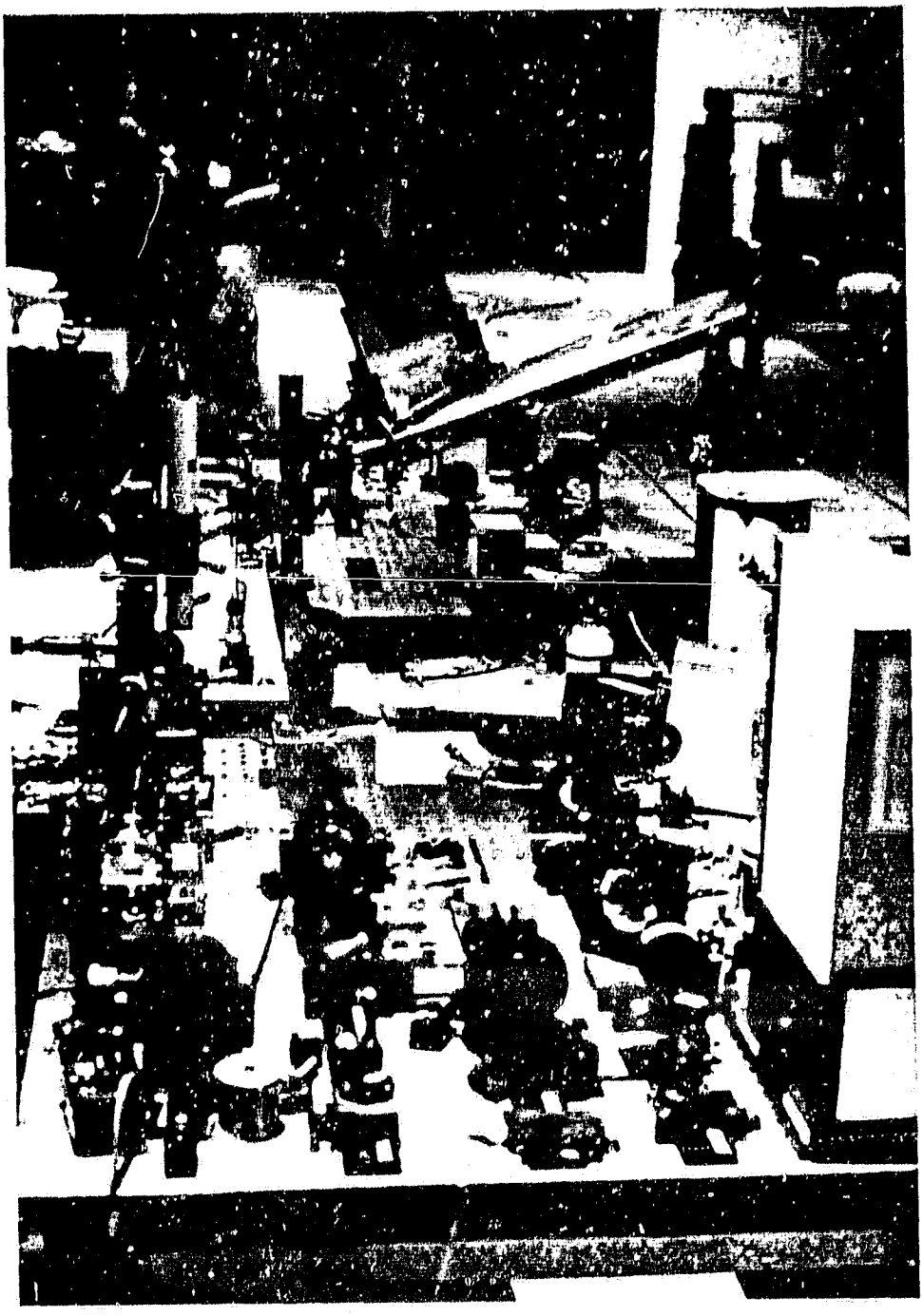


Figure 29. Photograph of the TDL heterodyne spectrometer in the coude room of the NASA 3-meter infrared telescope at Mauna Kea, Hawaii

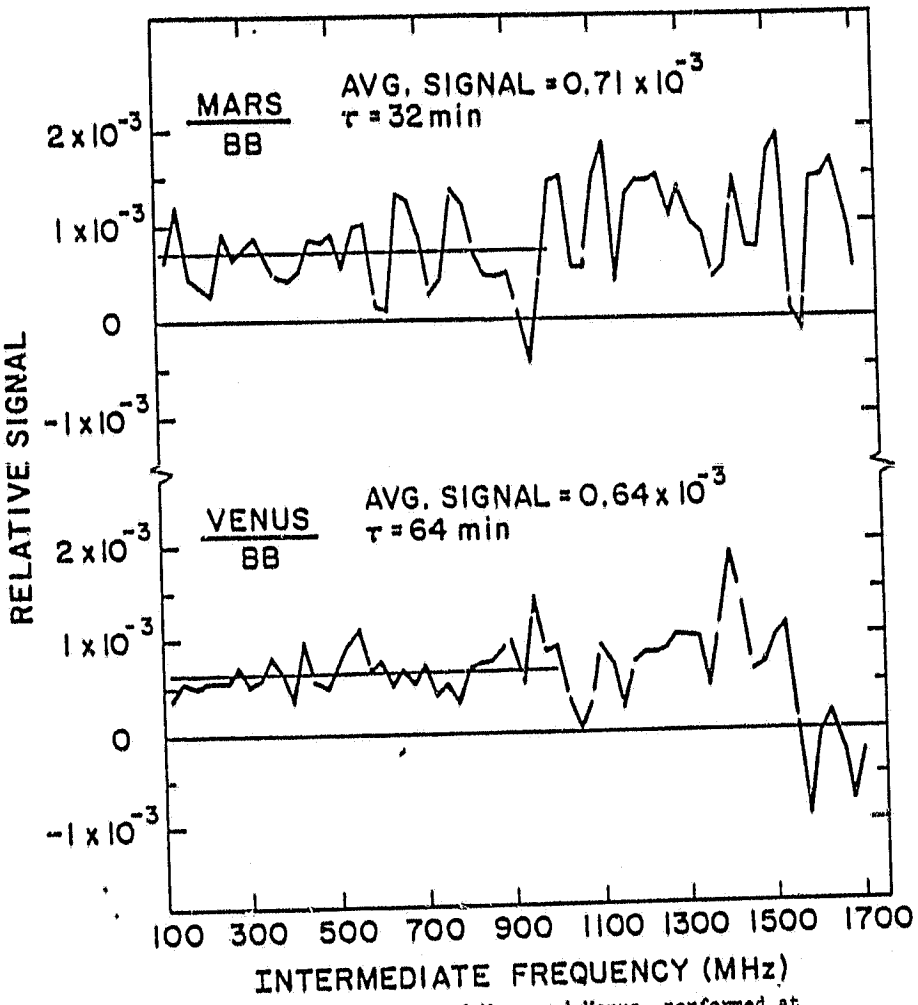


Figure 30. Heterodyne measurements of Mars and Venus, performed at Mauna Kea.

in Section 4.4, Figure 14. The synchronous detection scheme involved the use of a wobbling secondary with a ~1 arcmin throw, in place of the rotating blade chopper. Either beam could be centered on the planet with the other on the sky. This capability allows the cancellation of systematic differences between the two beams that are comparable in size to the planetary signal.

The Mars result (top) consisted of 10 co-added integrations at disk center made by alternating beams, for a total integration time of ~64 minutes. The Mars measurement was made during daylight hours since Mars transit occurred at about 2 p.m. local time. This made tracking a major problem since the high background light level made the disk of Mars almost invisible in the visual field. Tracking difficulties were not quite as severe for the Venus measurement, because of the planets higher visible albedo and the high solar zenith angle at the time of measurement.

Adopting a disk center continuum temperature of 260 K for Mars and 240 K for Venus, the computed fluxes of these sources when ratioed to the blackbody should be 3.1×10^{-3} and 1.8×10^{-3}, respectively. The Venus continuum signal of 0.64×10^{-3} then, suggests a total transmittance for atmosphere and telescope of ~0.4. Compared to Venus, the ratio Mars/BB should have been ~1.2×10^{-3}. The observed value of 0.8×10^{-3} is presumably the result of guiding difficulties. The integration time for Mars was only about half that for Venus, which accounts for the poorer signal to noise observed in the Mars measurement.

APPENDIX C:

HETERODYNE MEASUREMENTS OF THE INFRARED OBJECT IRC+10216

Measurements of IRC+10216 were also made at Mauna Kea during January 1981. As Figure 31 shows, these measurements were made on January 14 and 15 using diode laser modes at 1219.30 and 1227.12 cm^{-1}. These frequencies were established by first coarse calibrating using the monochromator, and then offsetting from N_2O reference lines at 1219.260 and 1227.099 cm^{-1}. A 1300K reference blackbody was used for these measurements.

As with the planetary run, successive measurements on the source were made by alternating beams from the wobbling secondary to remove thermal background offsets that exist between the two beams. In order to successfully track and remain centered on the source, the brightest part of which is confined to a region less than ~1 arcsec, the instrument was periodically used as a broadband detector by spiralling the telescope over a 1-3' arcsec field, while sensing for a peak in the synchronously detected direct signal from the source. This procedure was repeated after each pair of source integrations (~ once every 10 minutes) and was necessary to correct for slow drifts in telescope pointing.

One motivation for making these measurements was the possibility of detecting SiO absorption in this source. For this reason the choice of LO frequencies was restricted to close coincidence with expected positions of low J, v = 1-0 and v = 2-1 lines in the source. The mode used on January 14 exhibited excessive amounts of excess

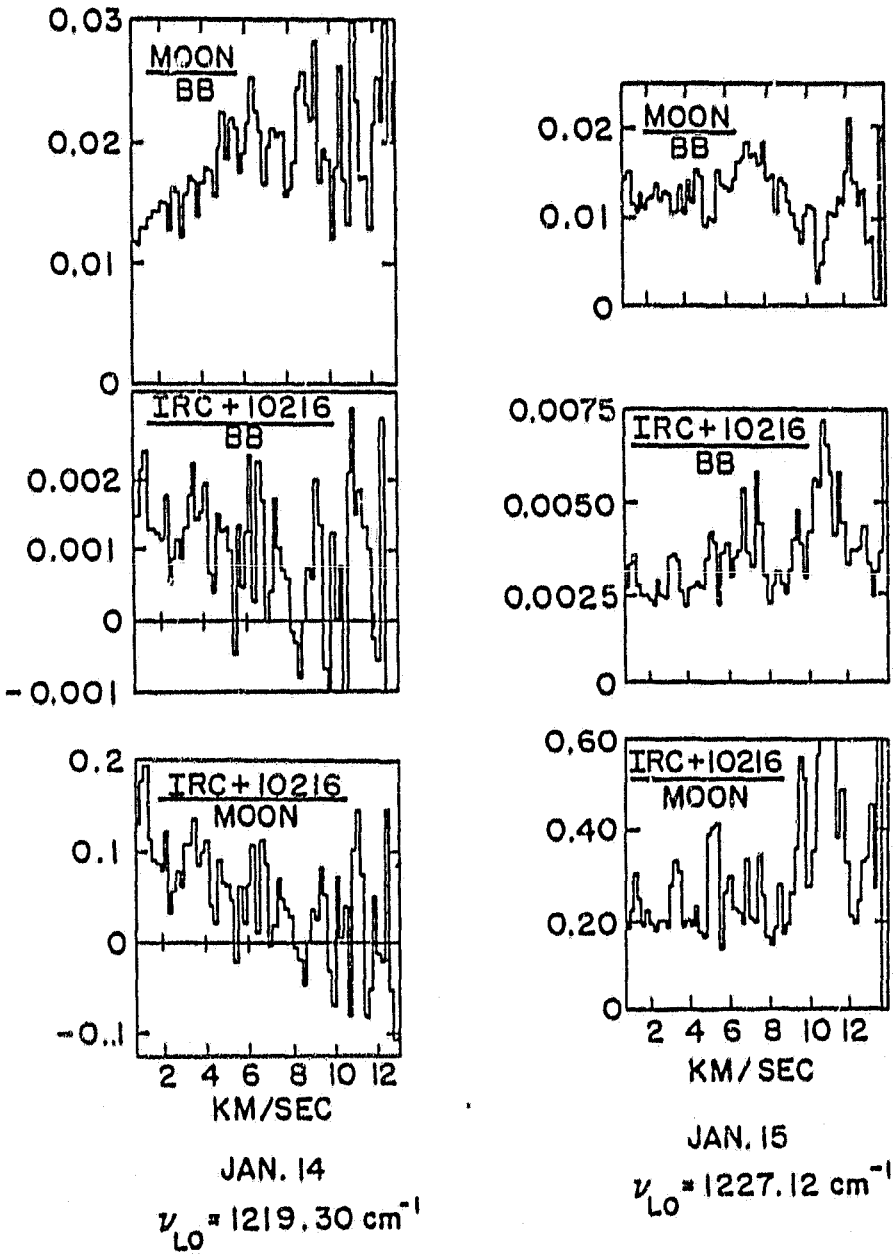


Figure 31. Heterodyne measurements of the infrared object IRC+10216, performed at Mauna Kea.

noise at high IF, causing a steep roll-off in the IR response. Ideally any IF response characteristics will disappear in the ratio spectra. During the time intervals between source, moon and BB reference measurements, however, small changes in LO performance occurred. Because of the low signal levels being detected, these changes introduced large anomalies in the ratio spectra. For this diode laser mode the general IF noise level showed time variability as exhibited in both moon calibration runs and in several sky versus sky comparisons.

The January 15 data appear considerably better, partly because the TDL mode in this case was well behaved, and partly because of improved seeing - an important requirement for observations of point sources. The heterodyne calibration measurements were made by placing one beam of the wobbling secondary on the lunar limb. Since the lunar phase was near first quarter, the subsolar point was also near the limb and the lunar continuum temperature was expected to be about 390K. This would yield a moon/BB ratio of 0.032. The measured ratio of $\sim$ 0.012 - 0.015 suggest an atmosphere - telescope transmission factor of about 0.4, in agreement with the November, 1980 results. The ratio IRC+10216/moon was measured to be $\sim$ 0.20. This result is not inconsistent with a value for IRC+10216/moon of 0.38 measured by Kostiuk and coworkers (1981) with the Goddard 10 micron CO_2 heterodyne spectrometer at Kitt Peak (Kostiuk, 1981). The moon was full at the time of these measurements (January 2, 1980) and because of the lower temperature at the lunar limb, larger values of the star/moon flux ratio would be expected. The diode laser

measurement suggests an 8 micron brightness temperature of 290K for IRC+10216.

APPENDIX D:

DETAILS OF THE VOIGT PROFILE DOUBLE-SIDEBAND FITTING ROUTINE

The spectral line receiver produces a folded double sideband spectrum which complicates any analysis of an absorption feature in one of the sidebands. For two special cases the unfolding process in simplified: If an isolated narrow feature is observed centered at some distance out in the IF, e.g., a doppler-broadened line at $\geq 5\Delta\nu_D$ from the local oscillator frequency, then the effect of folding is to double the continuum intensity relative to the line depth and reduce the contrast in the line by a factor of 2. The single sideband lineshape $(\phi_c-\phi_\nu)/\phi_c$ can then be retrieved by doubling to get $2(\phi_c-\phi_\nu)/\phi_c$. The other special case arises when the local oscillator is placed exactly at the line center position. Then, for a symmetric line, the folding of each half of the line around the LO position doubles the line contrast so the measured $(\phi_c-\phi_\nu)/\phi_c$ is the single sideband value.

In all other cases, either because of excessive broadening or close proximity of the line center to the LO position, folding of the line wing in the opposite sideband causes a distortion of the line shape and a visibility intermediate to that of the two special cases above. This is often the case with the observed SiO features.

In order to recover the line shape and extract reliable values for the widths, equivalent widths and residual line intensities, a computer program was written which folds a Voigt function about some selected point and overlays the folded curve on the observed

double-sideband feature. Figure 32 and the following discussion illustrate the details of the folding and fitting process.

Letting ϕ_c and $\phi(\nu_o)$ be the continuum and line center intensities, the single sideband intensity ${}^{ssb}\phi(\Delta\nu)$ at a point in an absorption line a distance $\Delta\nu$ from line center is

$$ {}^{ssb}\phi(\Delta\nu) = \phi_c - [\phi_c - {}^{ssb}\phi(\nu_o)] \quad [H(a,\nu)/H(a,0)] \qquad (D1) $$

In Equation D1, $H(a,\nu)$ is the Voigt function, with the customary definitions for a and ν.

$$ a = \gamma/\Delta\nu_D = \text{ratio of collisional broadening to doppler broadening} $$

Here $\Delta\nu_D$ is the doppler half-width for the absorption line. Since a is dimensionless, γ has the units of $\Delta\nu_D$ and represents the collisional broadening contribution. Also, letting ν_o be the line center frequency

$$ \nu = \Delta\nu/\Delta\nu_D = (\nu - \nu_o)/\Delta\nu_D \qquad (D2) $$

As usual, ν is the independent variable which gives the distance from line center $\Delta\nu$, in units of the doppler width. Of course $\nu = 0$ when $\nu = \nu_o$. Placing $H(a,0)$ in the denominator of Equation D1 normalizes the height of the Voigt function at line center to $(\phi_c - {}^{ssb}\phi(\nu_o))$ so that ${}^{ssb}\phi(\Delta\nu=c) = {}^{ssb}\phi(\nu_o)$ as required. The double sideband

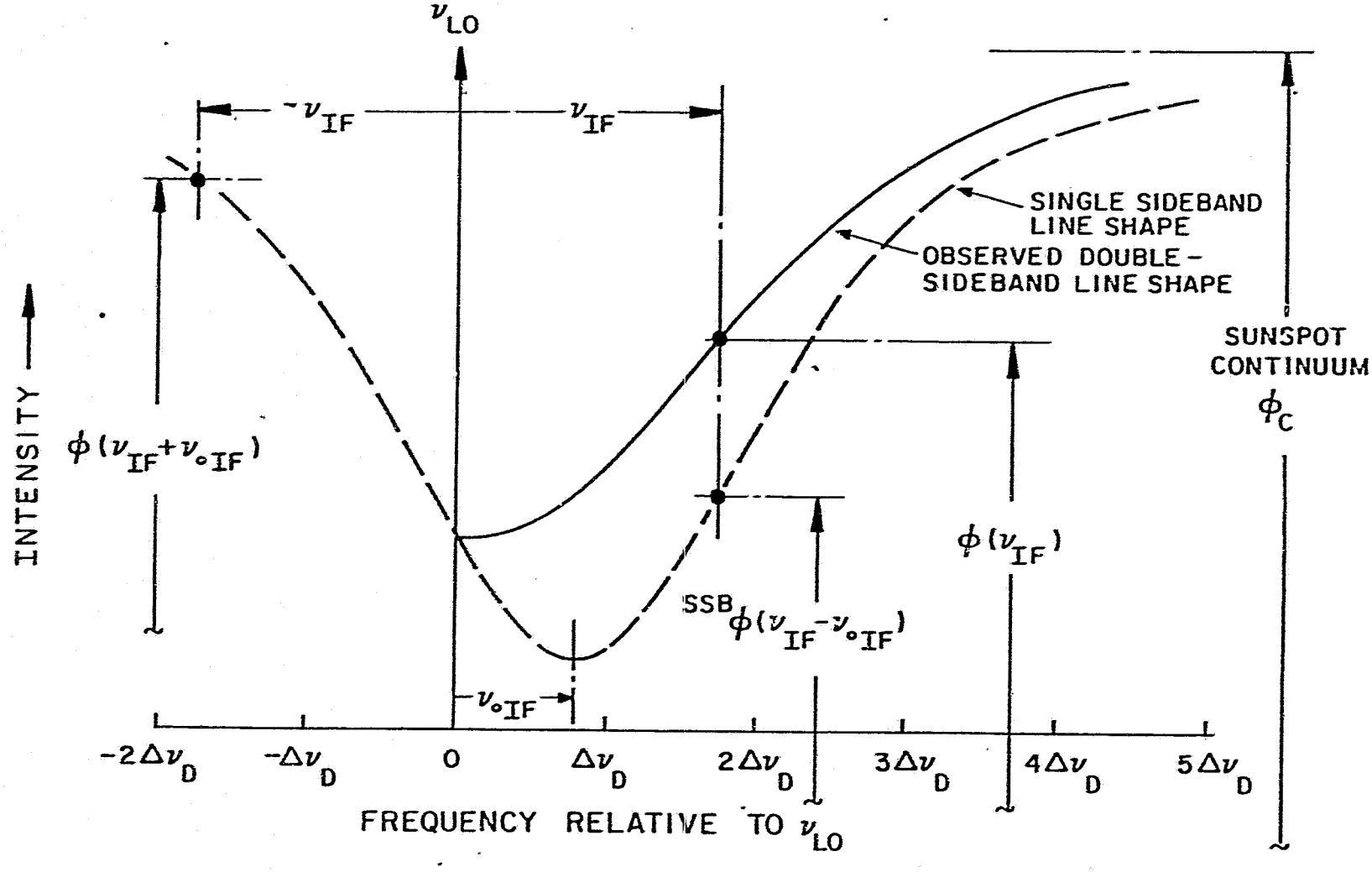


Figure 32. Effect of double sideband folding at the intermediate frequency on a line absorption profile.

intensity at an intermediate frequency ν_{IF} is the average of the single-sideband values at $(\nu_{cIF} + \nu_{IF})$ and $(\nu_{cIF} - \nu_{IF})$. From Figure 32 this can be written

$$\phi(\nu_{IF}) = 1/2 \left[{}^{ssb}\phi(\nu_{IF} - \nu_{cIF}) + {}^{ssb}\phi(\nu_{IF} + \nu_{cIF}) \right] \qquad (D3)$$

where

$${}^{ssb}\phi(\nu_{IF} \pm \nu_{cIF}) = \phi_c - {}^{ssb}\phi(\nu_{cIF})\, H'(a,v_{\pm})$$

In Equation D3 we have defined

$$H'(a,v_{\pm}) = H(a,v_{\pm})/H(a,0)$$

with

$$v_{\pm} = (\nu_{IF} \pm \nu_{cIF})/\Delta\nu_D$$

Simplifying, Equation D3 becomes

$$\phi(\nu_{IF}) = \phi_c - 1/2\ (\phi_c - {}^{sso}\phi(\nu_{cIF})\ [H'(a,v_-) + H'(a,v_+) \qquad (D4)$$

It is also important to mention the treatment of the doppler term $\Delta\nu_D$ in this analysis. For a molecule of mass M (in amu) at temperature T, the thermal doppler (1/e) width for absorption (or emission) at frequency ν_c, hereafter called $\Delta\nu_{th}$, is

$$\Delta\nu_{th} = (2\ k\ T/M\ m_H\ c^2)\ \nu_c \qquad (D5)$$

with k = Boltzmann's constant and $M\ m_H$ the molecular weight ($M\ m_H$=44 for SiO). For sunspot temperatures (T = 3500-4000 K) $\Delta\nu_{th}$ for SiO is rather well defined, and lies between .0038 and .0041 cm^{-1}. The effect of microturbulence broadening is taken into account by adding an assumed microturbulent velocity ξ_M in quadrature to the thermal doppler velocity to obtain an effective doppler width $\Delta\nu_D$. The total particle velocity ξ_D is related to $\Delta\nu_D$ by

$$\Delta\nu_D = \xi_D\nu_c/c \qquad (D6)$$

and can be expressed as the sum of the thermal part (ξ_{th}) and the microturbulence contribution (ξ_M) by

$$\xi_D = (\xi_{th}^2 + \xi_M^2)^{1/2} \qquad (D7)$$

Since $\xi_{th} = (c/\nu_o)^2 \Delta\nu^2_{th}$, Equation D7 becomes

$$\Delta\nu_D = \nu_o/c \ [(c/\nu_o)^2 \Delta\nu^2_{th} + \xi_M^2]^{1/2} \qquad (D8)$$

From Equation D6, the thermal term in brackets is only a function of temperature. Since ξ_M is usually given in km/sec, the first term can be expressed in $(km/sec)^2$ so that for SiO, Equation D8 becomes

$$\Delta\nu_D = (10^5 \nu_o/c) \ [3.76 \times 10^{-4} T + \xi_M^2]^{1/2} \qquad (D9)$$

The procedure used in fitting the observed lines was to adjust ν_{oIF}, $ssb_{\phi(\nu_{oIF})}$, ϕ_c, a and ξ_M until good agreement was obtained between the observations and the folded Voigt lineshape (Equation D4), which is overlaid with the data. The line width and equivalent width could then be extracted from the unfolded line shape.

APPENDIX E:

EFFECTS OF SCATTERED PHOTOSPHERIC LIGHT ON MEASURED LINE SHAPES

For observations of small spots or in cases where the heterodyne field of view does not reside in the spot umbra for the entire integration period, some dilution of the umbral spectrum occurs because of contamination by surrounding photospheric light. This situation is equivalent to a filling factor problem where the source size is smaller than the field of view. For a beam area of unit size with a fraction A of the beam filled by the spot, the measured umbral intensity $I'(\nu)$ will be

$$I'(\nu) = A\ I(\nu) + (1-A)\ I_{ph} \qquad (E1)$$

Here $I(\nu)$ is the corresponding intensity for A=1, (i.e. when the umbra completely fills the beam). The measured quantity ($\phi'\nu$) is the ratio of spot to adjacent photosphere intensity $I'(\nu)/I_{ph}$ so that most useful form of Equation E1 is obtained by dividing each term by I_{ph} to get

$$\phi'(\nu) = A\ \phi(\nu) + (1-A) \qquad (E2)$$

Likewise for the continuum measurement

$$\phi'_c = A\ \phi_c = (1-A) \qquad (E3)$$

For each measurement, the effective filling factor A can be deduced from the measured continuum ϕ'_c after a suitable value for the "true" relative umbral intensity ϕ_c has been determined. Rearranging Equation E3 gives

$$A = (1-\phi'_c)/(1-\phi_c) \tag{E4}$$

For each measurement it is then possible to get an upper limit $D(\nu)$ to the line depth $D'(\nu)$ at any point on the line. Since

$$D'(\nu) = (\phi'_c - \phi'(\nu))/\phi'_c = A/\phi'_c(\phi_c-\phi(\nu))$$

solving for $D(\nu)$ gives

$$D(\nu) = (\phi_c - \phi(\nu))/\phi_c = (\phi'_c/\phi_c A)\, D'(\nu) \tag{E5}$$

The equivalent width W' can be adjusted in a similar way. From the definition of equivalent width,

$$W\, I_c = \int_{\text{Line}} (I_c - I(\nu))\, d\nu \tag{E6}$$

Dividing first by I_c and then by I_{ph}, Equation E6 becomes

$$W = \int_{\text{Line}} (1-\phi(n))/\phi_c\, d\nu = \int_{\text{Line}} D(\nu)\, d\nu \tag{E7}$$

From Equation E5, this is equivalent to

$$W = (\phi'_c/\phi_c A) \int_{\text{Line}} D'(\nu)\, d\nu$$

or

$$W = (\phi'_c/\phi_c A)\, W' \qquad \text{(E8)}$$

Equations E5 and E8 allow the measured line intensities to be adjusted, once a suitable value for umbral intensity ϕ_c is assumed.

REFERENCES

Abbas, M. M., M. J. Mumma, T. Kostiuk, and D. Buhl, Sensitivity limits of an infrared heterodyne spectrometer for astrophysical applications, Appl. Opt. 15, 427 (1976).

Abramowitz, S., Private communication, 1981.

Allario, F., C. H. Blair and J. F. Butler, High-resolution spectral measurements of SO_2 from 1176.0 to 1265.8 cm^{-1} using a single PbSe laser with magnetic and current tuning, IEEE J. Quant. Elec. QE-11, 205.

Allen, C. W., Astrophysical Quantities, 3rd Ed. The Athlone Press, London, 1973.

Barbe, A., C. Secroun, P. Jouve, A. Goldman and D. G. Murcray, High resolution infrared atmospheric spectra of ozone in the 10μm, region - analysis of ν_1 amd ν_3 bands - assignment of the $\nu_1+\nu_2-\nu_2$ band, to be published, 1981.

Beck, S. C., J. H. Lacy and T. R. Geballe, Detection of the 12.28 micron rotational line of molecular hydrogen in the Orion molecular cloud, Astrophys. Jour. (Letters), 234, L213, 1979.

Beckwith, S., S. E. Persson, G. Neugebauer and E. E. Becklin, Observations of molecular hydrogen emission from the Orion nebula, Astrophys. Jour. (Letters), 223, 464, 1978.

Beer, R, D. L. Lambert and C. Sneden, The silicon monoxide radical and the atmosphere of α-Orionis, Pub. Astron. Soc. Pacific, 86, 806-812, 1974.

Betz, A. L., Infrared heterodyne spectroscopy of carbon dioxide in planetary atmospheres, PhD. Thesis, University of California at Berkeley, March 1977.

Blaney, T. G., Signal-to-noise ratio and other characteristics of heterodyne radiation receivers, Space Science Reviews, 17, 691, 1975.

Bray, R. J., and R. E. Loughhead, Sunspots, Chapman and Hall, London, 1964.

Bruckner, G., I.A.U. Symposium No. 22, Rottach-Egern, 293, 1965.

Buhl, D. and M. J. Mumma, RF spectrometers for heterodyne receivers, NASA Heterodyne Systems and Technology Conference, NASA CP-2139, Williamsburg, VA, April 1980.

Butler, J. F., A. R. Calawa, R. J. Phelan, Jr., T. C. Harman, A. J. Strauss and R. H. Rediker, Appl. Phys. Lett., 5, 75, 1964.

Cohen, S. C., Heterodyne detection: phase front alignment, beam spot size and detection uniformity, Applied Optics, 14, 1953, 1975.

Coupiac, P. and Koutchmy, S., Photometry of a sunspot at λ=3.75 μ, Astron. Astrophys., 16, 272, 1972.

Deming, D., Mixing and the strong-cyanogen phenomenon, Ap.J., 222 (1), 246-262, 1978.

Drawin, H. R. and P. Felenbok, Data for Plasmas in Local Thermodynamic Equilibrium, Gautier-Villars, Paris, 503 pp., 1965.

Eng, R. S., A. W. Mantz and T. R. Todd, Low-frequency noise characteristics of Pb-salt semiconductor lasers, Appl. Opt., 18, 1088 1979.

Finn, G. D., and D. Mugglestone, Tables of the Line Broadening Function H(a,v), Monthly Not. Royal Astron. Soc., 129, 221, 1965.

Fonstad, C. G., D. Kasemset, H. H. Hsieh and S. Rotter, New PbSnTe heterojunction laser diode structures with improved performance, NASA Heterodyne Systems and Technology Conf., I., NASA CP-2138, 23, 1980.

Frerking, M. A. and D. J. Muehlner, Infrared heterodyne spectroscopy of atmospheric ozone, Applied Optics, 16(3), 526, 1977.

Gaur, V. P., M. C. Pande and B. M. Tripathi, Molecular abundances in sunspots, Bull. Astron. Inst. Czech, 24 (3), 138, 1973.

Gaur, V. P. Vibration rotation bands of NO in sunspots, Solar Physics, 46, 121-123, 1976.

Gaur, V. P., M. C. Pande and B. M. Tripathi, Vibration rotation bands of SiO in sunspots, Solar Physics, 56, 67-69, 1978.

Gavrila, M., Elastic scattering of photons by a hydrogen atom, Phys. Rev., 163, 147, 1967.

Gingerich, O., Proceedings First Harvard-Smithsonian Conference on Stellar Atmospheres, Smithsonian Astrophys. Obs. Spec. Rep. No. 167, 17-41, 1964.

Gingerich, O., R. W. Noyes and W. Kalkofen, The Harvard-Smithsonian reference atmosphere, Solar Physics, 18, 347-365, 1971.

Glenar, D., T. Kostiuk, D. E. Jennings and M. J. Mumma, Development and performance of a laser heterodyne spectrometer using tunable semiconductor lasers as local oscillators, NASA Heterodyne Systems and Technology Conference, NASA CP-2138, Williamsburg, VA, 1980, p. 199-208.

Glenar, D. A., T. Kostiuk, D. E. Jennings, and D. Buhl and M. J. Mumma, A tunable diode laser heterodyne spectrometer for remote observations near 8 microns, accepted for publication in Applied Optics, May, 1981.

Goldberg, L., E. A. Muller and L. H. Aller, The abundances of the elements in the solar atmosphere, Ap.J. Suppl. 5, 1, 1960.

Goldman, A. and R. D. Blatherwick, Analysis of high resolution solar spectra in the 2.5 to 15μm region, Tech. Description of Project and Results, Dec. 1977 to May 1980, Department of Physics, Univ. of Denver, 1980.

Groves, S. H., K. W. Nill, and A. J. Strauss, Double Heterostructure $Pb_{1-x}Sn_xTe$ - PbTe lasers with cw operation at 77K, Appl. Lett., 25, (6), 331, 1974.

Hall, D. N. B. : PhD Thesis, 1970, Kitt Peak National Obs. Contribution No. 556.

Hall, D. N. B. and R. W. Noyes, Observation of hydrogen fluoride in sunspots and the determination of the solar fluorine abundance, Astrophys. Letters, 4, 143-148, 1969.

Hall, D. N. B. and R. W. Noyes, The identification of the [illegible] and 2-1 bands of HCl in the infrared sunspot spectrum, Ap.J., 175, L95-L97, 1972.

Harward, C. N. and J. M. Hoell, Optical feedback effects on the performance of $Pb_{1-x}Sn_xSe$ semiconductor lasers, Appl. Opt., 18, 3978, 1979.

Harward, C. N. and J. M. Hoell, Jr., Atmospheric solar absorption measurements in the 9-11 micron region using a diode laser heterodyne spectrometer, Heterodyne Systems and Technology, I, NASA Conf. Pub. 2138, 1980.

Harward, C. N. and B. D. Sidney, Excess noise in $Pb_{1-x}Sn_xSe$ semiconductor lasers, NASA Heterodyne Systems and Technology Conference, NASA CP-2138, Williamsburg, VA, April 1980.

Henoux, J. C., Un modele thermique de l'ombre d'une tache base sur la spectrophotometrie du continuum infrarouge, Annals D' Astrophysique, 31, 511-522, 1968.

Henoux, J. C., Un modele de l'ombre des taches, Astron & Astrophys, 2, 288-308, 1969.

Hedelund, J. and D. L. Lambert, Transition probabilities for the vibration-rotation bands of silicon monoxide, Astrophys. Lett., 11, 71-75, 1972.

Hillman, J. J., D. E. Jennings and J. L. Faris, Dicde Laser - CO_2 laser heterodyne spectrometer: measurement of 2sQ(1,1) in $2\nu_2-\nu_2$ of NH_3, Applied Optics, 18 (11), 1808, 1979.

Hinkley, E. D. and C. Freed, Direct observation of the Lorentzian line shape as limited by quantum phase noise in a laser above threshold, Phys. Rev. Letters, 23 (6), 277, 1969.

Hinkley, E. D., T. C. Harman and C. Freed, optical heterodyne detection at 10.6 μm of the beat frequency between a tunable PbSnTe diode laser and a CO_2 gas laser, Appl. Phys. Letters. 13(2), 49, 1968.

Hsieh, H. H. and C. G. Fonstad, LPE-Grown PbSnTe distributed feedback laser diodes with broad continuous tuning range in single mode output, Proc. 1979 International Electron Devices Meeting, IEEE, 126-129, 1979.

Huber, K. P. and G. Herzberg, Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules, Van Nostrand Reinhold Company, 1979.

Jennings, D. E. and J. J. Hillman, Shock isolator for diode laser operation on a close-cycle refrigerator, Rev. Sci. Instrum., 43, 1568, 1977.

Kneer, Linienprofile in sonnenflecken, Astron. & Astrophys., 18, 39-46,

Kondratyev, K. Y., S. D. Andreev, I. Y. Badinov, V. S. Grishechkin and L. V. Popova, Atmospheric optics investigations on Mt. Elbrus, Appl. Optics, 4, 1069, 1965.

Kostiuk, T., Private communication, 1981.

Koutchmy, S. and R. Peyturaux, C. R. Acad. Sci. Paris, 267, 905, 1968.

Ku, R. T. and D. L. Spears, High-sensitivity heterodyne radiometer using a tunable dicde laser local oscillator, Solid State Device Research, 1, 1, 1977.

Kurucz, R. L., ATLAS: A computer program for calculating model stellar atmospheres, Research in Space Sciences, SAO Special Report No. 309, 1970.

Lambert, D. L. and B. Warner, The abundances of the elements in the solar photosphere - III. Silicon, Mon. Not. R. Astr. Soc., 138, 213-227, 1966.

Linden, K. J., K. W. Nill and J. F. Butler, Single heterostructure lasers of $PbS_{1-x}Se_x$ and $Pb_{1-x}Sn_xSe$ with wide tunability, IEEE Journ. Quant. Elec., QE-13 (8), 720, 1977.

Lo, W., The elimination of contact degradation in lead-salt diode lasers, Jap. J. Appl. Phys., 18 (Suppl. 18-1), 367 (1979).

Lovas, F. J., A. G. Maki and W. B. Olson, The infrared spectrum of SiO near 1240 cm^{-1} and its relation to the circumstellar SiO maser, Journ. of Molec. Spec., 87, 449-458, 1981.

Maltby, P. and L. Staveland, Sunspot intensity observations during the 9 May 1970 Mercury transit, Solar Physics, 18, 443-449, 1971.

McBride, B. J., S. Heimel, J. G. Ehlers and S. Gordon, Thermodynamic properties to 6000°K for 210 substances involving the first 18 elements, NASA SP-3001, 1966.

Mihalas, D., Stellar Atmospheres, W. H. Freeman and Co. Ed., 1970

Moore, C. E. Atomic energy levels. NBS Circ. 467, U. S. Govt. Printing Office, Washington, DC, Vol. I, 1949; Vol. II, 1952; Vol. III, 1958.

Mumma, M., T. Kostiuk, S. Cohen, D. Buhl and P. C. von Thuna, Infrared heterodyne spectroscopy of astronomical and laboratory sources at 8.5μm, NATURE, 253, (5492), 514-516, 1975.

Mumma, M., Private communication, 1980.

Murcray, F. J., A. Goldman, F. H. Murcray, C. M. Bradford and D. G. Murcray, Observation of new emission lines in the infrared solar spectrum, near 12.33, 12.22 and 7.38μ, preprint, 1981.

Nuth, J. A. and B. Donn, Vibrational disequilibrium in low pressure clouds, to be published in Ap.J., August, 1981.

Oliver, B. M., Thermal and quantum noise, IEEE Proceedings, 53, 436, 1965.

Olson, W. B., A. G. Maki and W. J. Lafferty, Nitrous oxide (N_2O) spectra for accurate infrared wavelength calibration of diode lasers, J. Phys. Chem. Ref. Data, to be published, 1981.

Poultney, S., D. Chen, G. Steinberg, F. Wu, A. Pires, M. Miller and M. McNally, Diagnostics and control of wavenumber stability and purity of tunable diode lasers relevant to their use as local oscillators in heterodyne systems, NASA Heterodyne Systems and Technology Conference, NASA CP-2138, Williamsburg, VA, April 1980.

Preier, H., Recent advances in lead-chalcogenide diode lasers, Appl. Phys., 20, 189-206, 1979.

Ralston, R. W., I. Melngailis, W. T. Lindley and A. R. Calawa, Stripe-geometry $Pb_{1-x}Sn_xTe$ diode lasers, IEEE J. Quantum Electronics, QE-9, 350, 1973.

Ralston, R. W., J. N. Walpole, A. R. Calawa, T. C. Harman and J. P. McVittie, High cw output power in stripe-geometry PbS diode lasers, J. Appl. Phys., 45, 1323, 1974.

Reisfeld, M. J., H. Flicker and J. D. King, private communication, 1978.

Savage, M. G. and R. C. Augeri, Fine wavelength ID for tunable laser local oscillators, NASA Heterodyne Systems and Technology Conference, NASA CP-2138, Williamsburg, VA, April 1980.

Schlossberg, H. R., and P. L. Kelley, Using tunable lasers, Physics Today, 25 (7), 36, 1972.

Siegman, A. E., The antenna properties of optical heterodyne receivers, Proc. IEEE, 54 (10), 1350, 1966.

Spears, D. L. and C. Freed, HgCdTe varactor photodiode detection of CW CO_2 laser beats beyond 60 GHz, Appl. Physics letters, 23, 445, 1973.

Spears, D. L., Planar HgCdTe quadrantal heterodyne arrays with GHz response at 10.6 μm, Infrared Phys., 17, 5 1977.

Spears, D., Private communication, 1981.

Steinfeld, J. I., Molecules and Radiation: An Introduction to Molecular Spectroscopy, the MIT Press, 1974.

Stellmacher, G. and E. Wiehr, Magnetically non-split lines in sunspots, Astron & Astrophys., 7, 432-442, 1970.

Stilley, J. L. and J. Callaway, Free-free absorption coefficient of the negative hydrogen ion, Ap.J., 160, 245, 1970.

Vernazza, J. E., E. H. Avrett, and R. Loeser, Center for Astrophysics Preprint Series No. 215, 1974.

Vernazza, J. E., E. H. Avrett and R. Loeser, Structure of the solar chromosphere II. the underlying photosphere and temperature-minimum region, Ap.J. Suppl. Series, 30, 1-60, 1976.

Walpole, N. J., A. R. Calawa, S. R. Chinn, S. H. Groves and T. C. Harman, Distributed feedback $Pb_{1-x}Sn_xTe$ double heterstructure lasers, Appl. Phys. Lett., 29, 307-309, 1976.

Walpole, N. J. A. R. Calawa, S. R. Chinn, S. H. Groves and T. C. Harman, Distributed feedback $Pb_{1-x}Sn_xTe$ double heterostructure lasers, Appl. Phys. Lett., 29, 307-309, 1976.

Walpole, N. J., A. R. Calawa, R. W. Ralston, T. C. Harman and J. P. McVittie, Single heterojunction $Pb_{1-x}Sn_xTe$ diode lasers, Appl. Phys. Lett., 23 (11), 1, 1973.

Webber, J. C., The molecular spectrum of sunspots, Solar Physics, 16, 340-361, 1971.

Wohl, H., On molecules in sunspots, Solar Physics, 16, 362-372, 1971.

Worchesky, T. L., K. J. Ritter, J. P. Sattler and W. A. Reissler, Heterodyne measurements of infrared absorption frequencies of D_2O, Optics Letters, 2 (3), 70, 1978.

Zwaan, C., Rech. Astron. Obs. Utrecht, 17 (4), 1965.

Zwaan, C., The structure of sunspots, Ann. Rev. Astron. Astrophys., 6, 135, 1968.

Zwaan, C., Structure of sunspots II: A continuum model atmosphere for dark umbral cores, Solar Physics, 37, 99-111, 1974.

Zwaan, C., Structure of sunspots III: A minimum-gradient model atmosphere for umbrae, Solar Physics, 45, 115-118, 1975.