INFRARED HETERODYNE SPECTROSCOPY IN ASTRONOMY*

Albert Betz
Department of Physics
University of California, Berkeley

SUMMARY

A heterodyne spectrometer has been constructed and applied to problems in infrared astronomical spectroscopy. The instrument offers distinct observational advantages for the detection and analysis of individual spectral lines at Doppler-limited resolution. Observations of carbon dioxide in planetary atmospheres and ammonia in circumstellar environments demonstrate the substantial role that infrared heterodyne techniques will play in the astronomical spectroscopy of the future.

INTRODUCTION

For astronomical observations which require Doppler-limited resolution and high absolute frequency accuracy, laser heterodyne spectroscopy offers to the infrared the technical advantages long enjoyed at radio frequencies. The vibrational-rotational spectra of planetary and circumstellar molecules can now be studied at the same high resolution commonly used for the pure rotational and other low-energy transitions falling in the microwave domain. In the infrared, the Doppler-limited resolution of heterodyne spectroscopy has made possible new types of observations. Examples are: studies of exact lineshapes of CO₂ absorption profiles in the atmospheres of Mars and Venus to derive the vertical pressure-temperature structure of the atmosphere, accurate measurements of wind velocities in the stratosphere and mesosphere of Venus from the Doppler shifts of narrow CO₂ lines, and investigations of the gas dynamics of circumstellar molecules by the detection of the 10 μm absorption lines of NH₃.

In heterodyne spectroscopy, whether radio or infrared, the source radiation collected through a telescope is mixed with a stable local oscillator signal in a nonlinear element (usually a device with a square-law response to the electric field), and the resulting difference frequencies are analyzed in a bank of radio frequency (RF) filters and power detectors. In a properly designed system, the frequency calibration and stability of the local oscillator (such as a klystron in the radio or a laser in the infrared) determines the absolute frequency calibration and ultimate frequency resolution capability of the entire spectrometer. In practice, the frequency widths of the individual RF filter channels are usually chosen to resolve adequately a Doppler-limited profile. The total number of available channels, and hence the total spectral range visible at one time, may be dictated either by economics or the intermediate

*This work supported in part by NASA Grant NGR 05-003-452.
frequency (IF) bandwidth available from the mixer. Doppler-limited spectroscopy usually requires resolving powers better than $10^6$, and so whereas channel widths of 300 kHz may be adequate in a 1 mm wavelength receiver, channels as wide as 30 MHz would be more practical for a 10 μm spectrometer. The following sections discuss the implementation and use of such a heterodyne spectrometer for the 10 μm band, with initial emphasis on the local oscillator and mixer characteristics of the instrument.

INSTRUMENT

Figure 1 shows a simplified schematic of the spectrometer. The converging beam from the telescope enters on the left, comes to a focus, and then passes through a coated NaCl beamsplitter. (The telescope used for all these observations is either the 1.5 or 0.75 meter apertures of the McMath Solar Telescope at Kitt Peak National Observatory in Tucson.) The vertically-polarized output from a CO$_2$ laser is first matched to the divergence of the signal beam and then partially reflected by the beamsplitter. By expanding the Gaussian laser beam before the beamsplitter and then selecting only the center, a nearly uniform intensity distribution can be obtained to match with the nearly uniform transverse intensity distribution of the telescope beam. In practice, the "lens" at the output of the laser is a pair of small mirrors which perform the equivalent beam-matching function illustrated in the diagram. The second germanium lens just before the detector is as shown, however, and is used at approximately $f/3$ in order to achieve a small spot size (∼60 μm) at an optimum point on the photodetector (∼150 μm diameter). Only about 0.5 to 0.8 mW of laser power is actually incident on the detector, an HgCdTe photodiode developed by David Spears of Lincoln Laboratory. With a low noise preamplifier following the mixer, useful bandwidths to 2 GHz can easily be obtained with these photodetectors, although the particular ones we have are normally rated to ∼1.5 GHz. The question of defining IF bandwidth in a real application will be discussed later. The amplified IF output is then directed into a second RF mixer which converts a selected 1280 MHz wide segment into a 64×20 MHz filter bank. The undesired image response of the 2nd mixer is rejected by filtering before the mixer. The direct RF to IF isolation of this mixer is measured to exceed 26 dB over the output band. (Future designs of this subsystem will employ a more complicated but versatile combination of 2nd and 3rd local oscillators.) The 2nd LO is also used to track out changes of source Doppler velocity (such as rotation of the Earth ∼3 MHz/hr) during the course of observations. For planetary observations a different filter bank of 40×5 MHz filters (200 MHz total) is used to provide finer resolution. The detected outputs from the filter-bank channels are simultaneously integrated in a following set of analog integrators and the net outputs periodically sampled by a computer. On weak sources, repeated integrations of 8 min interspaced by 2 min calibrations on a blackbody source may accumulate for several hours. Not shown in this figure are sky-chopper foreoptics at the telescope focus which alternate the field-of-view of the detector on and off the source at 150 Hz. The chopped signals from the filter bank are synchronously demodulated at this frequency just before integration.
Laser

All our observational work to date has been accomplished with fixed-frequency CO₂ lasers, which are step-tunable to discrete vibrational-rotational transitions of various CO₂ isotopes over the 9 to 12 µm band. The particular model now in use is based on the Lincoln Laboratory design of Charles Freed in using a semi-sealed discharge tube, invar stabilizing rods, and grating tuning control with zero-order output coupling. This design maximizes the Q of the 12 µm laser cavity and permits laser oscillation on J-values up to 45 on the stronger isotopes of CO₂. Output powers up to several watts are still obtained on the stronger lines. Aside from continuous tunability, this laser offers all the features desired in a local oscillator: more than adequate power in a clean single mode, freedom from AM and FM perturbations, reasonable physical size and efficiency, and accurate absolute frequency calibration for all discrete oscillation frequencies. In fact, the lack of continuous tunability and hence uncertainty in the oscillation frequency may be viewed as a positive feature for many spectroscopic applications, in that the laser frequency is auto-calibrated either to better than 1 part in 10⁷ accuracy with the discharge impedance sensing technique, or to better than 1 part in 10⁹ absolute accuracy with the saturated resonance fluorescence technique. Consequently, a separate involved calibration mechanism is not necessary, as would be the case with tunable diode lasers, for example. The completeness of spectral coverage in the 9 to 12 µm region can be estimated by counting the total number of laser lines available from isotopic CO₂ and ¹⁸N₂¹⁶O lasers in the sequence and "hot" bands as well as the more conventional laser bands. The average spacing between lines is ~3 GHz, which is also close to the IF output bandwidth available from state-of-the-art infrared photomixers. This is not to say that a laser line is always available to convert a target frequency into the center of the available IF band, but that, more often than not, a laser line can be found in the 9 to 12 µm range so that practical observations may be attempted. An added influencing factor aiding the use of a fixed-frequency LO is that the Doppler frequency of the target line will vary with the orbital motion of the Earth, and shifts as large as ±3 GHz (±30 km/s) can be expected at 10 µm, depending on the direction to the source. Even for the least favorably placed stellar sources, during at least one month of the year a specified line will usually be observable with a CO₂ or N₂O laser. The more favorable planetary sources, the planets Mars and Venus with rich CO₂ atmospheres, are trivially observable from a Doppler viewpoint, as the Doppler shifts from motion with respect to the Earth do not exceed 1.5 GHz (15 km/s). In this latter application, where CO₂ itself is the target molecule, the CO₂ laser is of course ideally suited as the local oscillator. In those other cases where the target line is not a laser transition, such as the NH₃ lines seen in stars, then the offset frequency between the target line and the laser must be determined by laboratory spectroscopy. Aside from the heterodyne measurements done in our own laboratory on specific lines of NH₃, OCS, C₂H₂, and SiH₄, both heterodyne and a number of other laser-related techniques have been applied to the 10 µm spectroscopy of astrophysically important molecules with ~3 MHz accuracy.⁵⁻¹¹
The photomixer is a reversed-biased HgCdTe photodiode with an $\sim 2 \times 10^{-4} \text{cm}^{-2}$ active area and an effective AC quantum efficiency of $\sim 35\%$ at $10.6 \mu\text{m}$. The LO power level required for quantum-noise (shot-noise) limited operation depends, of course, on the noise temperature of the first IF amplifier stage. With conventional bipolar amplifiers featuring an $\sim 250 \text{ K}$ noise temperature over the 100 to 1500 MHz band, laser power levels of about 0.5 to 0.8 mW are required on the detector to raise the laser-induced shot noise from the photomixer to several times the amplifier noise level. LO power levels much above 1 mW on the detector risk inducing saturation and a concomitant decrease in effective quantum efficiency. An operational definition of the detector bandwidth for quantum-noise-limited performance can be stated as the IF frequency where the LO induced shot noise and amplifier noise contributions are equal. Operational bandwidths exceeding the detector's $-3 \text{ dB}$ bandwidth are thus possible. For the Lincoln Laboratory photomixers that we use, the output bandwidth is $\sim 1500 \text{ MHz}$ for operation with a bipolar preamplifier with a 250 K noise temperature. Obviously, a reduction in amplifier noise at high intermediate frequencies can extend the usable bandwidth of the photodetector. Any drop in IF system gain with detector rolloff can be compensated in later stages of amplification where the noise temperature factor is not critical. The current availability of low-noise GaAs field effect transistors permits an optimized preamplifier to be constructed to match the photomixer response characteristics. An amplifier based on designs of the Radio Astronomy Laboratory at Berkeley is now in use on the spectrometer. It features a reactively-tuned 2-stage design with about 20 dB gain and good input/output match (VSWR $\leq 2.0$) from 500 to 2000 MHz. The noise temperature minimum is adjusted to fall at $\sim 1600 \text{ MHz}$ so as to provide the best performance at a frequency where the photomixer response is already falling off. When operated at room temperature, the amplifier has a noise temperature of 60 K at 1600 MHz, 80 K at 2000 MHz, and 90 K at 1000 MHz. Below 1 GHz, the noise temperature rises rapidly to $\sim 250 \text{ K}$ at 500 MHz. When the preamplifier is cooled to 77 K, as depicted in Figure 1, the noise temperatures fall to about half of their respective room temperature values. Even for a warm amplifier, however, the detector/preamplifier system operates in the quantum noise limit to beyond 2 GHz, the limit of our measurement system. Although this relatively slight increase of 30% in operating bandwidth may appear small and not worth the additional effort required, in practice it significantly improves the spectrometer's usefulness, in that wide spectral lines can be centered at a higher intermediate frequency and thus avoid the lineshape interpretation problems associated with side-band foldover about the local oscillator frequency. In addition, several lines of circumstellar NH$_3$ have been detected with the 2 GHz system which are unreachable with only a 1.5 GHz IF. For astronomical observations, more benefits will come from increased output bandwidths in photomixers than from improvements in quantum efficiency, although, of course, both are desired. But, even as improved detectors with 3 GHz bandwidths ($-3 \text{ dB}$) are fabricated in the near future, cooled FET preamps will still be desired to further extend operational bandwidths. Currently, a cooled preamp with an 0.1 to 4.0 GHz response is under development at Berkeley. This amplifier will be integrated with the photodiode package to eliminate intervening cabling and the attendant problems associated with VSWR variations.
Another obvious advantage of a low noise preamp is the reduction in LO drive level required to raise the laser-induced shot noise above the preamplifier noise contribution. Cooled FET preamps should allow receivers using tunable diode lasers of modest single mode power (\(\sim 100 \, \mu W\)) to achieve quantum-noise-limited performance over IF bandwidths approaching the detector's \(\sim 3.\, \text{dB}\) response.

OPERATIONAL CHARACTERISTICS

The sensitivity of the spectrometer on the telescope may be computed directly from the throughputs of the various components. At a wavelength of 10.5 \(\mu\)m, 92% of the incident signal flux is transmitted through the terrestrial atmosphere, 77% through the 5 mirrors of the solar telescope, 50% average through the sky-chopper, and 75% through the table optics. Of the remaining flux, 35% produces an effective IF signal current in the photomixer. Furthermore, only signal radiation in the polarization of the laser is detected, which may be regarded as an additional 50% loss for unpolarized sources. The net system detection efficiency of only 5%, however, is still adequate for useful astronomical work, as will be seen in the next section.

Spectra are generated by integrating for 4 minutes with the source in the positive beam of the sky-chopper and then for 4 minutes with the source in the negative beam. The difference between these integrations is the total signal with integrator offsets cancelling. Laboratory blackbodies at ambient and an elevated temperature (228°C) are then interposed in the two beams for the calibration cycle and the process repeated, except that each integration is for 1 minute. The final spectra is produced by normalizing the signal by the blackbody integration. Previous calibrations of telescope and atmospheric transmission allow an absolute signal level to be inferred. It is important that the laser power on the photodetector remain constant until the calibration cycle is completed. Otherwise, changes in the photodiode impedance will not allow gain variations from standing waves between mixer and preamp to be completely normalized by the calibration cycle, and a distorted spectra may result.

The infrared transmission of the terrestrial atmosphere can be determined from spectra of the Moon or the planet Mercury. The Sun may also be used as a blackbody for spectral calibration of the terrestrial atmosphere if accuracies no better than \(\sim 1\%\) are required. Our repeated observations of the intrinsic solar spectrum show numerous unidentified absorption lines throughout the 10 \(\mu\)m spectrum. The lines are on the order of 1% deep and 300 to 400 MHz wide, and vary in strength and detail from place to place on the Sun. Possible contributors include the fundamental vibrational-rotational bands of FeO and other metallic oxides; however, the current state of laboratory spectroscopy on these molecules is relatively so poor that positive identifications may not readily be achieved.
ASTRONOMICAL APPLICATIONS

Planets

The first use of infrared heterodyne for astronomical spectroscopy was to measure the lineshapes of individual $^{13}\text{C}^{16}\text{O}_2$ lines in the atmosphere of Mars. Figure 2 shows a subsequent observation with an improved version of that early spectrometer. The P(16) line of $^{12}\text{C}^{16}\text{O}_2$ in the 10 μm laser band is seen in absorption in the Martian atmosphere. Each channel is 5 MHz wide and overlapping segments of 40 channels were combined to reconstruct the wing of half the line. The equivalent integration time for the spectrum is 40 minutes. The intent of this and similar observations of other CO$_2$ lines was to model the vertical pressure-temperature structure of the atmosphere of Mars. The data were fitted with a simple 4 parameter model, and the solid curves show the effects of different surface pressure estimates to the fit of the data. The ability of heterodyne spectroscopy to resolve the lineshape permits measurements of atmospheric surface pressure to better than 0.5 millibar (≈0.4 torr) accuracy. Two additional unexpected features (at the time) are also seen in this profile. The small arrow on the right points to the absorption line of the P(23) transition of $^{12}\text{C}^{16}\text{O}_2^{18}\text{O}$ in the 10 μm laser band. On the left, the central core of the $^{12}\text{C}^{16}\text{O}_2$ absorption line is seen to be in emission. The width of this Gaussian re-emission component is only ≈35 MHz (FWHM), which implies a gas kinetic temperature of only ≈170 K at the altitude of line formation. The intensity of the emission, however, is much stronger than that expected from CO$_2$ in local thermodynamic equilibrium at this temperature and suggests a nonthermal excitation mechanism for the phenomenon. The detection of similar and stronger re-emission cores in $^{12}\text{C}^{16}\text{O}_2$ absorption lines on Venus strengthens the interpretation that the nonthermal excitation of CO$_2$ comes from direct solar pumping of short wavelength CO$_2$ bands. The emitting CO$_2$ lies high in the mesospheres of the planets (75 km on Mars, 120 km on Venus) so that collisional de-excitation cannot quench the subsequent 10 μm "fluorescence". Figure 3 illustrates the appearance of the re-emission core of the P(16) line for three different positions on Venus. Sunlight illuminates the planet from the left in this diagram and is strongest at position 1, close to the local noon on the planet. The strength of the 10 μm emission line is clearly seen to be proportional to the incident solar intensity. No emission line is seen from the dark half of the planet at the right; only the relatively flat residual continuum radiation at the center of the very broad $^{12}\text{C}^{16}\text{O}_2$ absorption line is detected. On Venus, the column density of $^{12}\text{C}^{16}\text{O}_2$ is so high that the absorption lines of $^{12}\text{C}^{16}\text{O}_2$ are "saturation-broadened" to be much wider than our photomixer bandwidth. The complete profiles of weaker lines of $^{13}\text{C}^{16}\text{O}_2$ can be readily seen in absorption, however. For Venus, the
Doppler-shifts of both the mesospheric $^{12}\text{C}^{16}\text{O}_2$ emission components (120 km) and the stratospheric $^{13}\text{C}^{16}\text{O}_2$ absorption lines (80 km altitude) have been monitored for several years in order to determine the average wind circulation patterns at these altitudes. The signal-to-noise ratio and absolute laser frequency calibration of the heterodyne spectra permit emission and absorption line shifts to be measured to $\sim$2 and 15 m/sec accuracy, respectively. The results indicate an average retrograde circulation of 90 m/sec at stratospheric altitudes and a symmetric subsolar to anti-solar flow as fast as 130 m/sec in the mesosphere. A complete report on four years of observations will be submitted for publication following planned observations in late Spring, 1980.

Stars

In 1978, the first application of the spectrometer to stellar spectroscopy was the detection of several absorption lines in the 10 $\mu$m $v_2$ band of ammonia in the gas cloud surrounding a supergiant star called IRC + 10216. Since then, circumstellar ammonia has been detected around a number of supergiant stars of various types, some of which also emit peculiar microwave maser emission from circumstellar OH, H$_2$O, and SiO molecules. Ammonia is relatively abundant in these sources and is an excellent indicator of the gas dynamics throughout the circumstellar region. The absolute frequency accuracy and resolution of the infrared heterodyne spectrometer permits direct comparisons with similarly accurate microwave data available from radio astronomy on these maser stars. For the non-maser star IRC + 10216, Figure 4 shows an absorption line of $^{14}\text{NH}_3$ in the ground $(0,0)$ rotational state, a level well-populated in the colder (<200 K) regions of the outer circumstellar envelope. The integration time on this line was 80 minutes. Laboratory spectroscopy on this transition placed it only 753 MHz below the P(20) laser line of $^{12}\text{C}^{18}\text{O}_2$ in the 10 $\mu$m band. Because of the abundance of ground-state ammonia, this particular laser line has turned out to be the "workhorse" LO frequency for the initial detection of NH$_3$ in stellar sources. The Doppler velocity of the circumstellar gas with respect to a local standard of rest (LSR velocity), as shown at the bottom of the figure, can be readily determined from the measured frequency offset of the observed line from the laboratory rest frequency of 28.533534 THz. In addition, a knowledge of the intrinsic stellar velocity from other types of observations can be used to infer the expansion velocity of the circumstellar gas with respect to the stellar core, as indicated at the top of the figure. (At a transition frequency of 28.5 THz ($\lambda=10.5$ $\mu$m), each 20 MHz filter channel represents a Doppler-velocity shift of 0.2 km/sec.) Future stellar observations will emphasize studies of NH$_3$ in additional types of stars, and undoubtedly the detection of other interesting molecules in the stronger sources.
PROSPECTS

Although the 10 μm band is currently the most favorable spectral region for infrared heterodyne spectroscopy, if practical considerations of atmospheric transparency and front-end technology are considered, the 5 μm band also appears promising for stellar spectroscopy because of the astrophysically important bands of CO near 4.7 μm. HgCdTe photomixers function at this wavelength, and isotopic CO lasers appear an obvious choice for the LO. In addition, more complete spectral coverage at 5 μm may come from frequency-doubled CO₂ and N₂O lasers. Developmental work on a 5 μm heterodyne spectrometer is now starting at Berkeley. This emphasis on gas lasers as local oscillators is not meant to exclude other viable technologies, however. Following the work at other institutions, a smaller effort is now underway in our own laboratory to evaluate tunable diode lasers for laboratory spectroscopy and as possible local oscillators at frequencies unobtainable with gas lasers.

At a wavelength of 5 μm, Doppler widths are twice as broad as at 10 μm for a given velocity range. Thus, an IF bandwidth of 3 GHz will be needed to get the same velocity coverage provided by 1.5 GHz at 10 μm. To support the increased IF requirements both at 5 and 10 μm, another 64x20 MHz filterbank is now under construction. This will expand our simultaneous frequency coverage to 2.56 GHz and greatly improve the efficient use of observing time on a large telescope. The success of observations on a 1.5 m telescope shows us that in the future we can effectively use a 3 m instrument, such as the NASA 3 m infrared telescope facility in Hawaii, and thereby hope to gain a four-fold increase in signal-to-noise ratio on unresolved stellar sources. On such a large telescope, good spectroscopy of NH₃ and other circumstellar molecules will be possible on well over 100 stars.
REFERENCES


12. Spears, D.L.: "Planar HgCdTe Quadrantal Heterodyne Arrays with GHz Response at 10.6 \mu m", Infrared Physics, 17, pp. 5-8, Jan. 1977.


Figure 1.- Infrared heterodyne spectrometer.

Figure 2.- Heterodyne measurements of CO$_2$ in Martian atmosphere.
Figure 3.- Heterodyne measurements of CO$_2$ in Venusian atmosphere.

Figure 4.- Heterodyne measurements of NH$_3$ in star IRC + 10216 showing Doppler shift.