

THE BERKELEY INFRARED SPATIAL INTERFEROMETER: A HETERODYNE STELLAR INTERFEROMETER FOR THE MID-INFRARED

D. D. S. HALE, M. BESTER, W. C. DANCHI, W. FITELSON, S. HOSS,¹ E. A. LIPMAN,² J. D. MONNIER,³
P. G. TUTHILL,⁴ AND C. H. TOWNES

Space Sciences Laboratory, University of California, Berkeley, CA 94720-7450

Received 1999 August 17; accepted 2000 January 26

ABSTRACT

A detailed description is given of the Infrared Spatial Interferometer (ISI), developed at the Space Sciences Laboratory of the University of California at Berkeley, which is a high spatial resolution interferometer for mid-infrared wavelengths. The instrumentation, its capabilities and performance, data analysis, science program, and future plans are all discussed. The system's use of heterodyne detection, analogous to that of a modern radio interferometer, is also compared with the homodyne or direct methods more commonly encountered in the visible and infrared. The ISI has been operating productively on Mount Wilson for the past 10 years measuring materials immediately surrounding stars and their changes as well as some stellar diameters. The new spectral capabilities described here, a recent increase in baseline length, and the upcoming expansion to a closure-phase imaging array provide important additional types of measurements.

Subject headings: infrared: general — instrumentation: interferometers

1. INTRODUCTION

The field of high-resolution optical interferometry, pioneered by Michelson in the 1920s, has experienced a modern renaissance dating from the intensity interferometer proposed by Hanbury Brown & Twiss (1956) and built in the 1960s (Hanbury Brown, Davis, & Allen 1967). Although the intensity interferometer itself is a notable exception, almost all modern optical interferometers follow closely the general scheme of Michelson in which starlight from separate locations on the wave front is brought to a common location where the interference pattern can be recorded. This scheme, in which starlight interferes with itself, is known as homodyne or direct detection. The Berkeley Infrared Spatial Interferometer (or ISI), however, utilizes a heterodyne detection scheme, whereby starlight from each aperture is first mixed with a CO₂ laser local oscillator, downconverting the signal to microwave frequencies, followed by path-length matching and fringe detection in a correlator. Although heterodyne methods form the backbone of almost every modern astronomical interferometer from the submillimeter through the entire radio spectrum, the ISI is unique in its use of heterodyne detection in the infrared. The ISI (Danchi, Bester, & Townes 1988) and its prototype in 1974 (Johnson, Betz, & Townes 1974; cf. also Sutton et al. 1977) have so far been the only scientifically productive multiple-telescope interferometers in the mid-infrared spectral region.

Operating in the atmospheric window from 9 to 12 μm , the ISI utilizes Earth rotation and periodic discrete changes of the baseline to obtain a wide range of effective baselines and map the visibility functions of a variety of astronomical targets. In its current configuration (Bester, Danchi, &

Townes 1990; Townes et al. 1998), the ISI consists of two independent 1.65 m infrared telescopes, each mounted in a movable trailer equipped with a full heterodyne detection system. In addition to radiation collectors for interferometry, these telescopes can also function as front-end receivers for other types of instruments, such as a high-resolution spectrograph, thereby expanding the range of science that can be performed. Construction of a third telescope, which will be integrated into the ISI array to form an imaging interferometer, is now underway at the Space Sciences Laboratory of the University of California at Berkeley. A detailed review of the instrument, its operation and functional capabilities, is presented here with a brief summary of its scientific results to date. The paper is organized as follows: § 2 describes the various subsystems that make up the interferometer; § 3 discusses analysis of the data; § 4 compares performance such as signal-to-noise properties of heterodyne and direct detection interferometers for use at mid-infrared wavelengths; § 5 gives a brief overview of ISI scientific milestones to date; and finally § 6 summarizes the paper and outlines future enhancements of the array.

2. THE INSTRUMENT

This section describes the ISI interferometer array. There is first an overview of subsystems underlying the signal path from the sky to the data analysis stage. This is followed by detailed descriptions of the most important subsystems, including the optics, stellar acquisition and guiding, infrared heterodyne detection, narrowband filter bank, and computer control architecture.

2.1. System Overview

Although the basic principles of heterodyne interferometry are quite simple, in practice the ISI array is reliant on a large number of subsystems. A schematic showing those responsible for the basic signal train is given in Figure 1. Starlight gathered by each of two telescopes, with large optical elements, passes through a set of mirrors guiding it onto a beam splitter. As discussed in § 2.3 and illustrated in Figure 4 below, near-infrared radiation in each telescope is

¹ Present address: Mount Wilson Institute, Mount Wilson, CA 91023.

² Present address: Laboratory of Chemical Physics, National Institutes of Health, Bethesda, MD 20892-0520.

³ Present address: Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138.

⁴ Present address: Chatterdon Astronomy Department, University of Sydney, Sydney, NSW 2006, Australia.

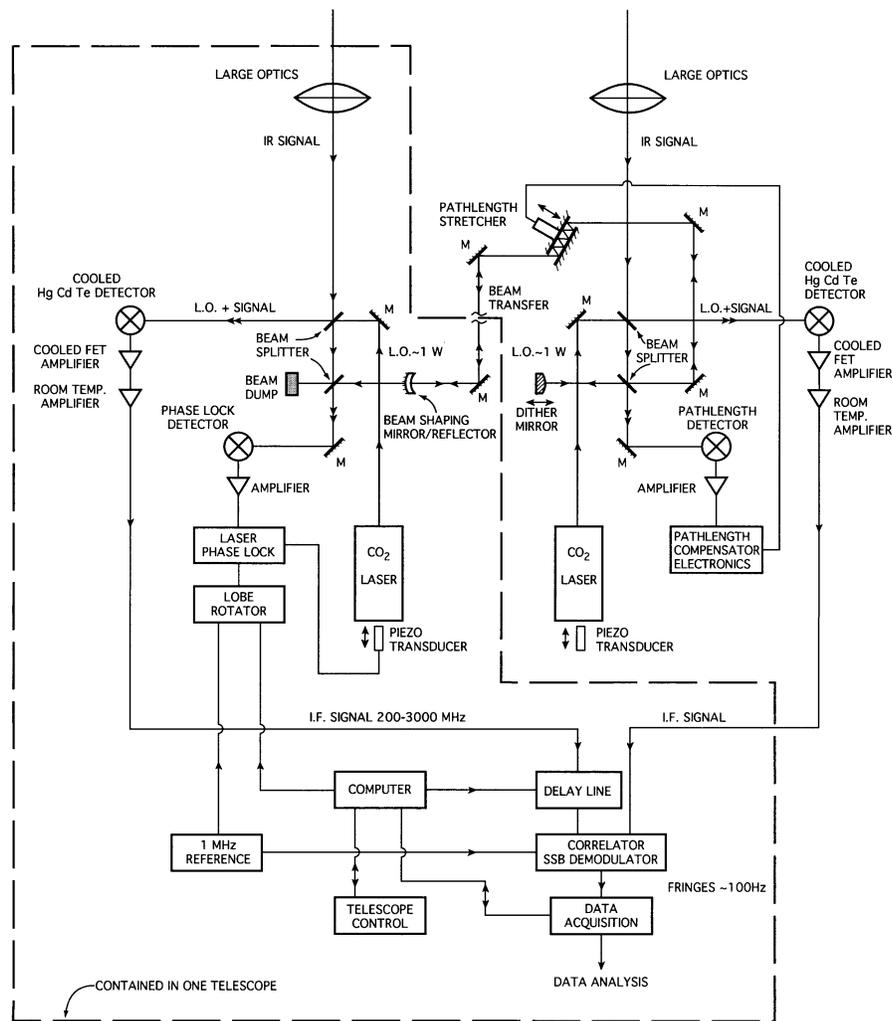


FIG. 1.—Conceptual block diagram of the ISI optics, circuitry, and heterodyne detection system, including laser phase lock and path length compensation systems. Telescope 1 and its trailer, with associated optics and circuitry, is enclosed within the dashed line. Telescope 2 is represented by equipment outside the dashed line. The trailer for telescope 1 is the primary operating location, with the central computer, delay line, correlator, and data acquisition equipment. Telescope 2 and its trailer contains path-length compensation equipment in addition to the necessary optics and detection hardware. When a third telescope is installed, the controls, delay lines, data acquisition, correlators, and other equipment not needed in each single telescope, will be housed in a central station.

split off and used to maintain the stellar image, which would otherwise move due to atmospheric turbulence, at a fixed focal location. At the beam splitter, the starlight interferes with laser light from the local oscillator (LO), which has itself passed through a separate set of beam-shaping optics. Signal and LO mixing occurs in a HgCdTe photodiode, resulting in an intermediate frequency (0.2–2.8 GHz) from each side band which can then be amplified, filtered, and routed to the delay line and correlator (for high spatial-resolution measurements), or to other instruments, depending on the research goals. This basic signal path is essentially identical for each telescope, or element of the array.

A separate set of subsystems depicted in Figure 1 functions to maintain a fixed phase relationship between the two CO₂ laser LOs. Although distribution of LO frequencies can be achieved in a straightforward way at radio wavelengths with electrical cables, or even a posteriori by synchronizing timing markers recorded with the data at widely separated locations (as for VLBI), a somewhat more complex approach is required to achieve this in the infrared. The relative frequency of the two lasers is controlled by a

subsystem dedicated to locking the separate lasers into a defined phase relationship. This entails transferring a portion of the laser beam, in this case from telescope 2 into telescope 1, as shown in Figure 1. Fluctuations in the path of this beam between the telescopes must be kept to a small fraction of the 11 μm laser wavelength. The mechanism for this and for the phase lock control are described in § 2.4. With phase-lock and transfer path-length controls, stellar interference fringes obtained at the correlator may be detected at any desired frequency, as described below in § 2.4. For convenience this frequency is usually chosen to be 100 Hz.

2.2. Telescope Array and Optics

Each ISI telescope is of a Pfund-type design, mounted in a semitrailer for ease of mobility and baseline selection. In this design (see Fig. 2) a 2 m flat mirror is installed in an altitude-azimuth (alt-az) mount and is used to direct starlight onto a vertically-mounted, stationary 1.65 m parabolic mirror. Each telescope independently tracks a star by steering its flat mirror with DC servomotors controlled by computers running the VxWorks operating system. (A Sun

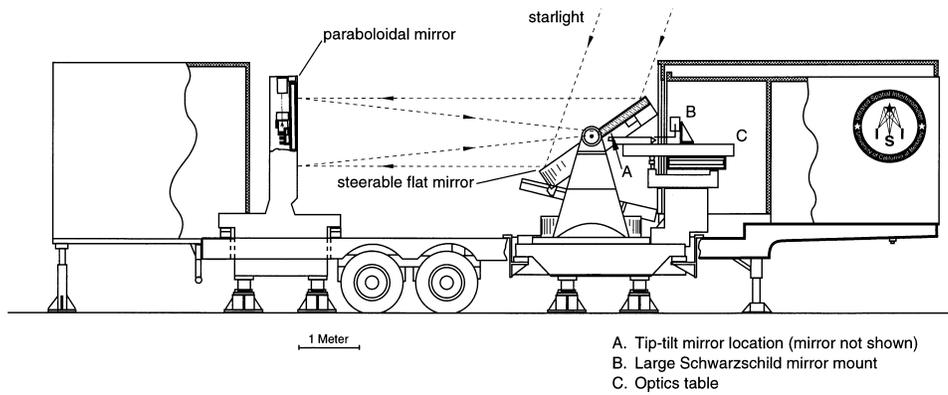


FIG. 2.—ISI Trailer and Optics Schematic. Light from a celestial source is reflected off the steerable flat mirror onto the stationary paraboloid mirror shown at left, focusing radiation back through a hole in the flat. Behind the flat mirror (right) is the optics room which contains the guider camera, beam combining optics, and detection system. Observers operate the telescope from a small control room area at the far right of the trailer.

Microsystems workstation using the Solaris 2.6 variant of the Unix operating system provides a graphical user interface for this and all other telescope control functions.) Starlight is reflected by the flat mirror onto the $f/3.14$ parabolic mirror, and is then focused by the parabola back through a hole in the flat mirror into a Schwarzschild optical system. The prime focus of the parabolic mirror is located a few inches behind the front surface of the flat mirror. The Pfund-type design is convenient in allowing quite large but compact and stable optics to be installed inside a movable trailer. However, while the trailers can be pointed either north or south, in any one position the telescopes see only 180° in azimuth.

The Schwarzschild optical system (see Fig. 3) consists of a pair of spherical front-surface mirrors which provide a less divergent beam to optical components which follow it, and allow easy matching of the incoming stellar wavefront with that of the laser local oscillator. The Schwarzschild design allows use of two spherical mirrors with no resulting aberration, except for small amounts of coma off-axis. After the $f/89$ beam emerging from the Schwarzschild system reaches a secondary focus, it is split by a dichroic beam splitter (see Fig. 4) that reflects well the mid-infrared wavelengths (between ~ 9 and $13 \mu\text{m}$), and transmits well the near-infrared wavelengths (~ 1 to $3 \mu\text{m}$). These shorter wavelengths are sent to the guider camera, while the reflected mid-infrared radiation is sent to a beam splitter where it is combined with the beam from a CO_2 laser local oscillator and then sent to a liquid nitrogen cooled HgCdTe photodiode signal detector (see Figs. 4 and 7 below). The infrared heterodyne detection system is described in more detail in §§ 2.4 and 5.

The two large mirrors of each telescope are made of Zerodur, a glass ceramic compound having a low thermal

expansion coefficient, and are coated with a $\sim 1000 \text{ \AA}$ layer of pure (99.999%) aluminum. They are supported in their cells by hanging roller chains, and each is located in place by means of three adjustable defining points. The steerable flat mirrors are additionally supported by a three-segment air-bag system. The air pressure in these bags is automatically adjusted and maintained by a regulator which senses the flat mirror altitude so that pressure on the defining points does not appreciably distort the mirror. Although the telescope mounts which contain each large mirror are moved between different baseline sites by the trailers, when put in place for observations they are mechanically separated from the trailer frame and rest upon their own set of kinematic supports which are placed on concrete pads approximately 18 inches thick and set in hard granite soil. This insures that all optics are free from mechanical contact with the trailer frame and isolated from frame vibrations due to observers inside the control room area or other causes. In order to make a telescope mobile, the large mirrors and some other elements are clamped in position, and the trailer frame is raised on its wheel axles, lifting the mirror mounts from their kinematic supports. The telescopes can then be towed, by tractor, to a new location and also along standard highways to new sites. Four or five days are required to perform a baseline change maneuver with

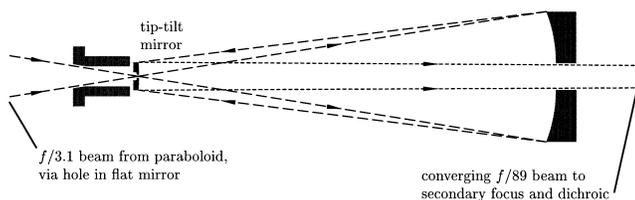


FIG. 3.—Schwarzschild optical system and telescope beam. The focus of the ISI telescope primary mirror lies at the center of the first small mirror, which is mounted on a tip-tilt stage for rapid compensation of pointing fluctuations due to “seeing.”

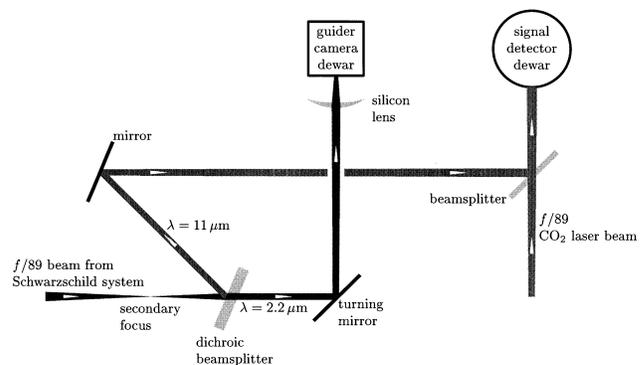


FIG. 4.—Dichroic beam splitter, camera, and signal detector. The $f/89$ starlight beam from the Schwarzschild optical system (Fig. 3) converges to a focus just before the dichroic beam splitter. The near IR light is sent to the guider camera, and the mid-IR light is combined with the CO_2 laser light on a beam splitter in front of the signal detector dewar. The beam convergence at the secondary focus is exaggerated for clarity. Positions and sizes of various components shown in this schematic illustration can be seen to scale in Figs 7 and 8.

attendant realignment of various optical components and accurate measurement of the new baseline vector.

The ISI is located at Mount Wilson Observatory in the San Gabriel mountain range northeast of Los Angeles, California. This site (elevation 1742 m, 34°13' north latitude) was chosen because of its excellent atmospheric ("seeing") conditions, which result from a combination of local topography, atmospheric characteristics, and proximity to the Pacific Ocean. The ISI telescopes can be placed in a number of different positions which are determined by the layout of concrete support pads at the observatory (Fig. 5). To date, five of the available configurations (Table 1) have been used, providing a range of nonredundant baselines from 4 to 32 m. Additional pads, scheduled for installation by early 2000, will allow the baselines of up to approximately 80 m (see Fig. 6).

2.3. Stellar Acquisition and Guiding

Many objects of interest in the mid-infrared are enshrouded by dust, and thus have little flux in the visible

portion of the spectrum. The ISI therefore uses a near-infrared guiding camera to detect the stars adequately, and also a tip-tilt correction system to keep the telescopes well pointed for interferometric observations. The near-infrared (2 μm) portion of the stellar image in an ISI telescope is passed through a dichroic beamsplitter onto the focal plane array of the infrared camera. The image from this camera is acquired by a computer (the guider PC), which determines the position of the star in the focal plane. Using this information, the guider PC adjusts both the main telescope drive and a small tip-tilt mirror in order to stabilize the image. Adjustment of the telescope drive takes place every few seconds, and corrects small errors in tracking. The tip-tilt mirror is driven by a rapid (up to 73 correction cycles per second) feedback loop which compensates for atmospheric fluctuations as well as any small and rapid errors in the telescope drive.

The guider camera, an Amber Engineering (now Raytheon Systems, Inc.) AE4128, consists of a 128 × 128 InSb focal plane array packaged in a compact liquid nitrogen dewar. The camera is supplied with clocking and readout electronics as a turnkey system, and is able to run at a very high frame rate (over 200 frames per second). The AE4128 was designed for surveillance and industrial inspection, and as a result, its dark current and read noise are higher than optimal for astronomical applications. However, it is a relatively low-cost, off-the-shelf product which is highly reliable. It can presently be used on stars which are brighter than K-band magnitude 3.5. Possible improvements of the camera's readout electronics are under consideration; it can perhaps be increased in sensitivity by a few magnitudes.

When properly calibrated, the ISI telescopes are able to use input coordinates to place the image of the target source within ~ 15" of the center of the guider camera's 40" field of view. After this initial acquisition, the guider PC takes

TABLE 1
ISI BASELINES AND RESOLUTION

Pad-Pair	Baseline (m)	Maximum Resolution $\frac{\lambda}{2L}$ for $\lambda = 11 \mu\text{m}$ (arcsec)
1-4	4	0.284
4-5	9.6	0.118
1-5	13	0.087
0-5	16	0.071
1-6	32	0.035

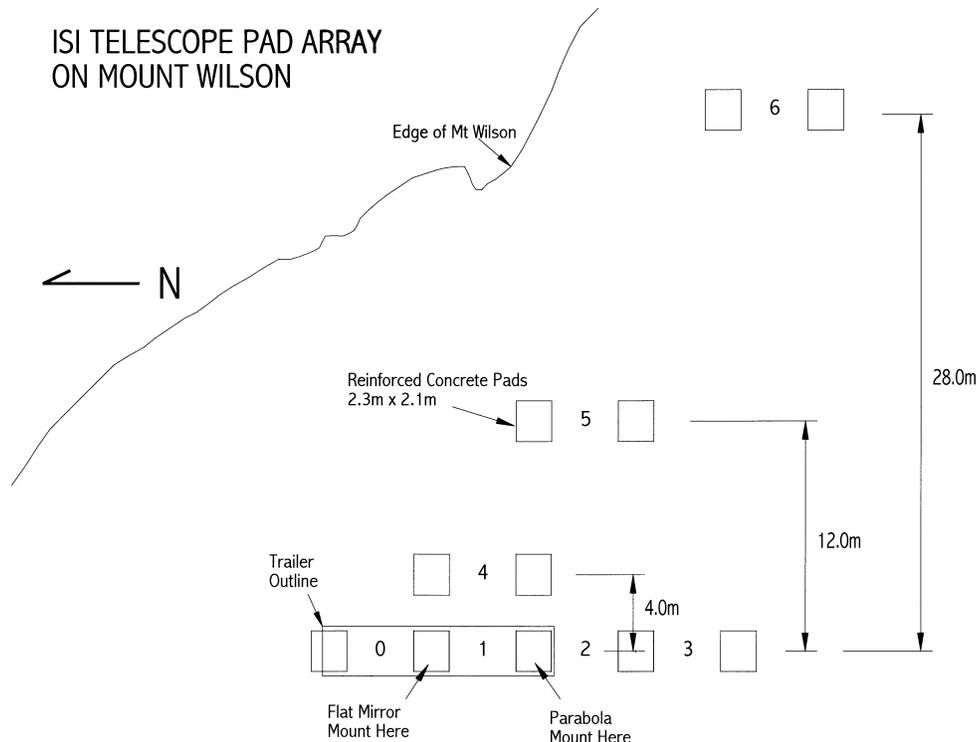


FIG. 5.—Present site layout of the ISI telescope pad array on Mt. Wilson, showing the relative placement of concrete pads, upon which rest the large telescope optics. The numbered pad-pairs indicate the seven different possible locations for the telescope.

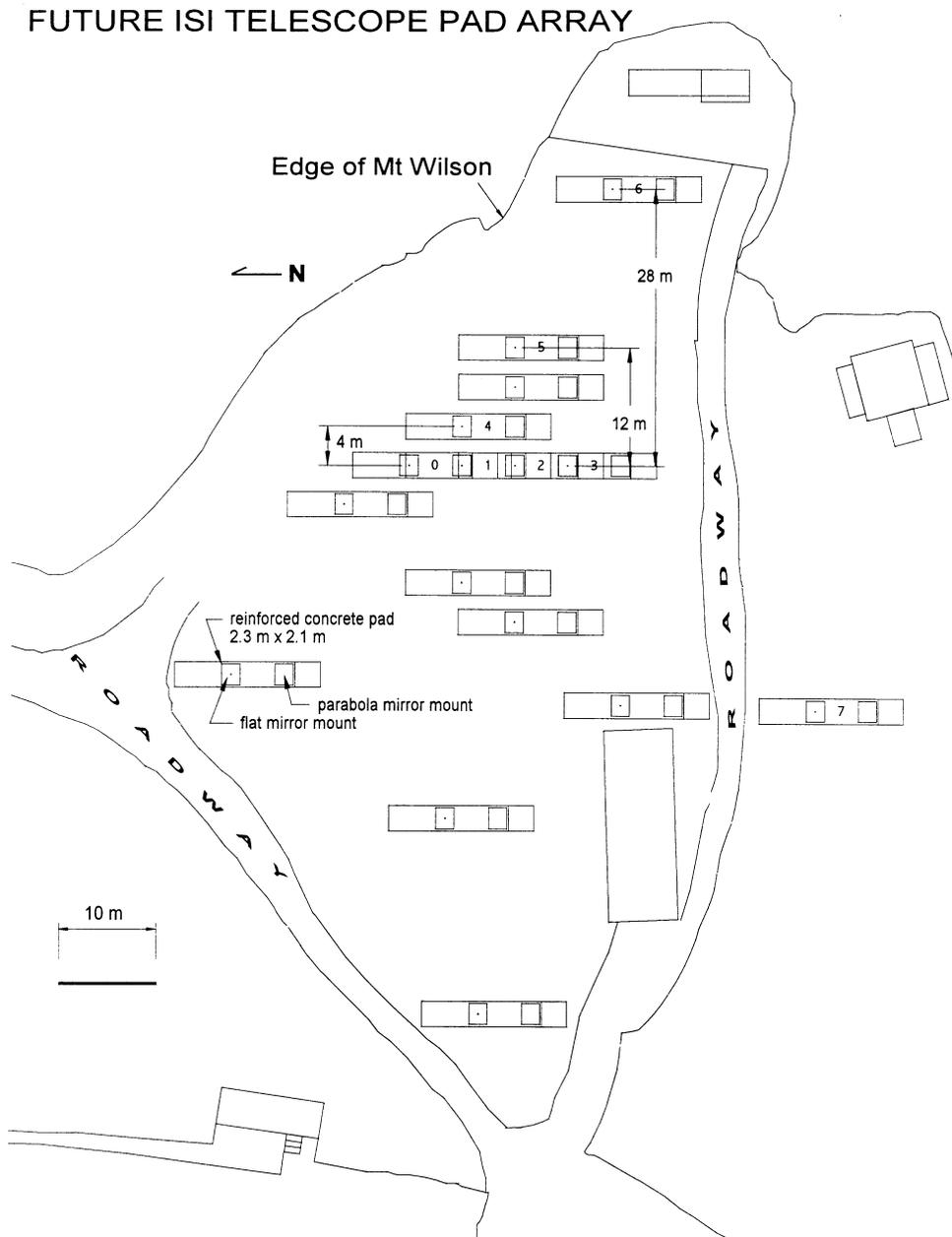


FIG. 6.—A planned future ISI telescope pad array layout. Pad 7 has been installed for the 1999 observing season, providing a 56 m baseline when the other telescope is placed on pad 6. The additional, unnumbered pad-pairs are to be constructed when the third telescope is delivered to Mt. Wilson, and will provide a variety of baselines up to about 80 m. Rectangular structures not labeled represent buildings which are not part of the ISI.

over and automatically points the telescope to lock the star in the center of the field. The tip-tilt mirror, located in the Schwarzschild optical system (Fig. 3), has a throw in all directions which is equivalent to slightly more than 2" on the sky—enough to correct for atmospheric fluctuations on all but the worst nights at Mount Wilson. The tip-tilt feedback loop can run as fast as 73 Hz, so long as the target is bright enough in the near-infrared to allow a correspondingly short integration time for the camera. For dimmer stars, slower rates can be used since substantial atmospheric fluctuations on Mount Wilson for the near-infrared wavelengths are typically not faster than 25 Hz, and are even slower in the mid-infrared. The addition of tip-tilt correction to the ISI guiding system has increased the stellar signals by as much as 40% under poor seeing conditions.

Detailed descriptions of the design and performance of the ISI guiding system can be found in Lipman (1998) and Lipman et al. (1998).

2.4. Infrared Heterodyne Detection System

The ISI, as is common with modern radio telescope arrays, uses a technique known as heterodyne detection for receiving infrared radiation from celestial sources. With this technique, a mid-infrared stellar signal ($\nu_{\text{IR}} \approx 30 \text{ THz}$) is downconverted into a radio-frequency (RF) signal, of single sideband width about 2.6 GHz, while preserving the amplitude and phase information of the original source signal, except for fluctuations due to quantum effects. The arrangement which is described here for the two telescopes cur-

rently used in the ISI array can be applied to arrays with arbitrary numbers of receivers. The stellar signal is combined with light from a laser "local oscillator" (LO), the frequency of which lies at the center of the band to be detected. The combined signals are incident upon a nonlinear (square law) photodiode detector, whose output signal is at the difference frequency between the two inputs, referred to as the "intermediate frequency," or IF. The IF encompasses a range of frequencies; those greater than that of the LO are part of the upper sideband while those less than the LO are in the lower sideband. (For a thorough discussion of heterodyne frequency conversion, see Thompson et al. 1986). In the case of the ISI, a stable CO_2 laser is used as the LO. Its radiation is combined with radiation collected by the telescope using a beam splitter, and focused onto a HgCdTe photodiode (cf. Fig. 4). Both sidebands are detected within the range of frequencies for which the photodiode and IF circuitry are responsive, approximately 200 to 2800 MHz, and transmitted along coaxial RF cables.

The operational frequency of the laser local oscillators, and hence of the ISI telescopes, can be adjusted to any one of a series of discrete CO_2 frequencies between approximately 9 and 12 μm . Each laser can use a variety of CO_2 isotopes for operation on many lines and can, with modest additional effort, be made to oscillate with other gases. The lasers are similar to the Freed design (Freed 1968), and operate with a meter-long semifocal cavity having a concave mirror at one end with a radius of curvature of twice the cavity length, and a grating at the other in a Littrow configuration (acting as a plane mirror where the beam waist is located). Oscillation on undesirable modes is prevented or reduced by apertures inside the cavity in front of the grating and the concave mirror. Different laser lines can be easily selected by setting the tilt of the 80 groove mm^{-1} original ruled diffraction grating. The ISI most commonly operates on the P(20) rotational-vibrational line of $^{13}\text{CO}_2$ at 11.15 μm , chosen because of the excellent atmospheric transparency at that wavelength.

Passive stabilization of the oscillation frequency is integral to the design. Four invar rods define the cavity length and the gas discharge operates in a sealed tube surrounded by a jacket of circulating and thermally stabilized water. Active frequency control is achieved by means of a two-stage piezoelectric tuner onto which the concave reflecting mirror is affixed. For the two lasers to be locked in phase, this tuner must operate at frequencies well above 10 kHz to compensate for acoustic noise and other sources of frequency fluctuations. Lasers in the two telescopes are synchronized so that their relative phases are kept constant, as is discussed below.

The optical table of each telescope is depicted in Figures 7 and 8, which show the complete light paths and detection systems. As already mentioned, light collected by the telescope enters through a hole in the first Schwarzschild mirror from the left (compare Figs. 3 and 4), and reaches the dichroic beamsplitter after passing through the Schwarzschild system. The dichroic transmits the near-infrared wavelengths from the star to the IR guider camera (*top center*) and reflects the mid-infrared starlight down toward the optical table. The mid-infrared starlight proceeds to a 97% reflective beam splitter, which combines it with 3% of the laser local oscillator beam.

Light from the CO_2 laser (located beneath the optics table, and not shown) is brought up to the table from the right, passes through several optical elements which produce an f/89 beam, and reaches the same beam splitter as the starlight, in front of the signal Dewar. The combined beams propagate to a mirror which sends the beam up into the signal detection dewar through an antireflection coated (AR) f/1.5 aspheric ZnSe lens and an AR coated ZnSe window tilted at 5° from normal incidence to the optical axis. (This window is tilted so that reflected LO power will not back-propagate along any of the normal optical paths, causing spurious signals from glints off optical elements.) The radiation is focused by the lens on a liquid nitrogen cooled HgCdTe photodiode, which produces the IF signal.

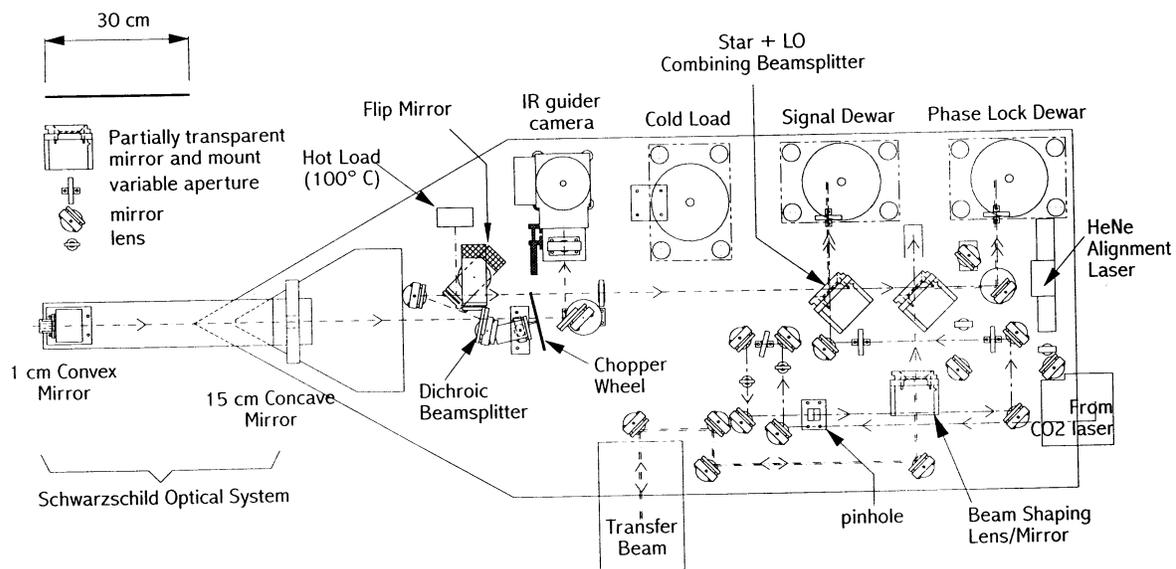


FIG. 7.—Telescope 1 optics table, shown to scale. Starlight from the telescope enters at left, the laser LO beam enters at right, and they are combined on a beam splitter in front of the signal Dewar (*top*), which contains a detector. A portion of the laser LO from telescope 2 (the transfer beam) is seen entering and leaving at bottom. The two laser beams are combined to produce a phase lock signal at a detector in the phase lock Dewar. The hot source and cold load provide standard amounts of radiation for measurement calibration.

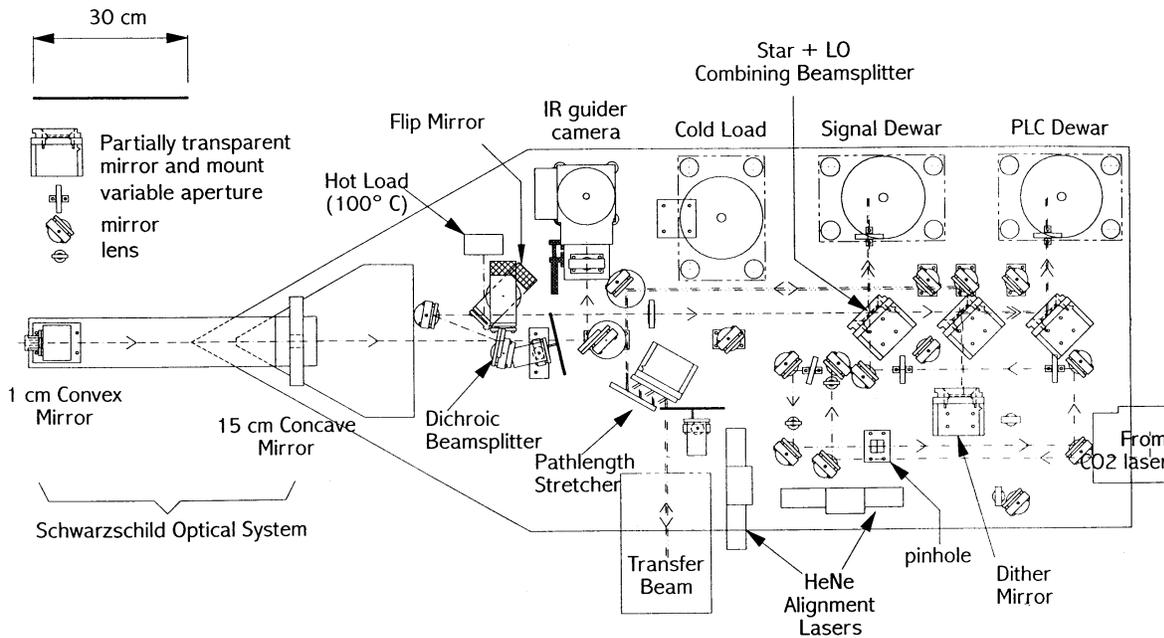


FIG. 8.—Telescope 2 optics table, shown to scale. This table contains additional optics for performing path-length compensation (see text for detailed description). A portion of the laser LO zigzags through a pair of mirrors, which move small amounts to keep the path length constant, and exits the table at bottom. The free-space, round-trip beam reenters along the same path, and the two beams are recombined on a detector in the path-length compensation (PLC) Dewar to form a signal for maintaining a constant phase of the wave, or a constant path length.

Also visible on the table are additional optics for alignment purposes, a flip-mirror for injecting a hot-load signal used for testing and calibration, and a stable cold-load reference used when the beam from the telescope is chopped during IR power measurement.

Because each telescope uses a separate laser LO to create the heterodyne IF signal, a precise phase relationship between the two LOs must be maintained at all times during interferometric observations, even though the two oscillators differ in frequency by 1 MHz, as will be explained below. To accomplish this, a feedback loop is established by sending a portion of the laser beam from telescope 2 to telescope 1 and combining it with a portion of the beam from the laser of telescope 1. The combined beams are focused onto a detector, producing a beat frequency between the two lasers. This detection subsystem thus operates in a manner similar to that used for detecting starlight. The beam frequency between the two lasers is compared with a frequency near 1 MHz produced by a synthesizer referenced to a very stable crystal oscillator. An output (phase error) signal, which is proportional to the relative phase between the beat frequency and the synthesizer frequency, is fed back into the piezoelectric tuner of the second laser, producing a frequency shift in the cavity proportional to the error signal, so that the second laser is phase-locked to the first one with an rms phase error of less than about 5° . Thus the difference in frequency and phase between the two lasers follows precisely the 1 MHz synthesizer, and relative phase and frequency drifts or fluctuations are canceled out.

Additional relative phase errors may arise from the fact that the laser beam between telescopes is transmitted through a relatively long path of open air which may vary in length, and these errors must be minimized. To correct for variations of this optical path length, a portion of the transfer beam is returned back along exactly the same path.

The returning beam is combined with a portion of the original beam which acts as a reference (top right of Fig. 8). The pathlength of the reference beam is modulated by a dither mirror, shown in Fig. 8, with an amplitude of $\sim 1 \mu\text{m}$ at a frequency of $\sim 300 \text{ Hz}$ and the combined beams are detected in the same way as that described for starlight. This detector puts out a sinusoidal signal at the frequency of modulation of the reference path, with zero amplitude at this frequency when the two beams are at the null of their interference. Thus the amplitude of this detector output is proportional to the pathlength variation of the beam propagating between the telescopes. An error signal is generated from this output by synchronous detection at the 300 Hz reference modulation frequency, with a bandwidth of about 10 Hz. This error signal is fed into a high voltage op-amp to control the length of a large piezoelectric tube onto which one of two parallel mirrors is attached. The transfer beam is reflected several times between these two mirrors before leaving telescope 2 (see "path-length stretcher" in Fig. 8). This arrangement keeps the effective pathlength constant to a small fraction of the $10 \mu\text{m}$ wavelength, giving a phase error at the signal mixing beam splitters of less than a few degrees. The entire system of laser phase-locking and pathlength compensation can be seen conceptually in Figure 1.

The IF signal from each telescope is amplified, filtered, and then delivered to the correlator section of the instrument where the two signals are multiplied together by a square-law detector to produce a difference frequency that represents modulation due to interference between the two signals. However, immediately before the two signals are correlated to produce fringes, the difference in geometrical delay of the radiation between the two interferometer elements is offset by appropriately delaying one of the signals. A direct-detection interferometer must control a light beam by mirrors which are adjustable in position to provide a

delay accurate to a fraction of the radiation's wavelength divided by the fractional bandwidth used. The narrow bandwidth of a heterodyne interferometer such as the ISI needs only to delay a radio-frequency signal with an accuracy of a fraction of the shortest IF wavelength. Hence the delay line consists of a set of coaxial cables that are switched into place using computer-controlled RF relays. The cable lengths provide increments of 1 cm (binary coded in lengths of 2^n cm, where n is the delay line bit selected), allowing tracking of the "white light" fringe. Before correlation, each IF signal (one of them having been delayed) is divided, using an RF power splitter, into two equal parts. One part of each signal is sent to the correlator where the signals from both telescopes are multiplied together to produce an electronic interference signal; the second part is used to measure the amount of infrared power in each telescope with standard RF power detectors and lock-in amplifiers. The multiplication is achieved by sending each IF signal into a four-port hybrid ("magic T") the outputs of which are the sum and difference signals, i.e., if the amplitude of the field from one telescope is A and that from the other is B , then one output is $(A + B)$, the other $(A - B)$. The interference fringes are detected by back-diodes, which produce $(A + B)^2$ and $(A - B)^2$ at the outputs, respectively. These signals, which are at a frequency of about 1 MHz due to the frequency difference between the two LOs, are differenced by a video op-amp whose output is hence proportional to the product of the two signals, $4AB$.

The two IF signals sent to the correlator are multiplied together, as described above, to produce an electronic fringe signal. Diurnal motion produces a continual change in the relative distance from the star to the two telescopes, and hence of the phase relationship between the starlight at the two telescopes. This causes the fringe signal to emerge from the correlator at a frequency (known as the natural fringe frequency) that varies as a function of position of the source in the sky. Since from an instrumental point of view it is undesirable to measure a signal which constantly varies in frequency, the ISI performs what is known in radio astronomy as "lobe rotation," whereby the effects of diurnal motion can be removed. This is accomplished by having a real-time computer determine the natural fringe frequency once a second while tracking a star, and then changing the frequency of one of the laser oscillators with respect to the other by an amount calculated to maintain the correlator output frequency at 100 Hz (see Fig. 10 below). Besides a correction for the natural fringe frequency, an additional reference frequency of 1 MHz is added to one of the laser LOs so that the initially detected beat signal is well removed from potentially noisy audio frequencies. The lobe rotator (or phase shifter) operates on the same laser for which the frequency is determined by the phase-lock loop. The output of the multiplier, $4AB$, is itself multiplied by (or demodulated by) the 1 MHz reference signal with single sideband detection to generate the low-frequency difference which is the fringe signal. In the absence of atmospheric pathlength variations between the source and the telescopes, the fringe signal is a constant 100 Hz sinusoid. The fringe signal is sampled at 500 Hz by a 15 bit analog-to-digital (A/D) converter and stored on disk for later analysis.

To measure the infrared power and take into account sky brightness, the telescopes are "position-switched," generally every 15 s. In this mode the telescopes alternate between the source and two positions $5''$ on either side of the source.

During these observations, the beam from the telescope is chopped at a high rate (150 Hz) against a cold-load reference which has approximately the same brightness at $10 \mu\text{m}$ as the sky. Total infrared power from each telescope is determined by synchronous detection at the 150 Hz reference frequency of the chopper, using lock-in amplifiers with an averaging time constant of about 0.1 s. The infrared power of the source itself is thus determined by subtracting the lock-in amplifier output voltages for the on-source and off-source positions. To detect very weak fringe signals, chopping and position switching can be omitted, which increases the fringe power by approximately a factor of 4.

A visibility measurement is made by integrating the power spectrum of the recorded fringe signal (see Fig. 10 below), and then normalizing it with an infrared power measurement obtained by position switching and chopping. The ratio of the two is proportional to the square of the visibility as normally defined (see eq. [1] below). For calibration of this proportionality, measurement of a point source is used. A complete track in the (u, v) plane can in principle be mapped out over the course of one evening by following a star through its diurnal motion. However, a star is generally not tracked closer to the horizon than 15° – 20° since atmospheric fluctuations through the resulting atmospheric pathlength tend to become severe. This results in nearly continuous measurements of the visibility for any fixed location of the two telescopes, providing resolutions varying from the maximum allowed by the total telescope separation to less than half of this value.

2.5. Filter-Bank System

A new area of astrophysical study has recently become available to the ISI with the completed integration of a filter-bank system. This system allows visibility measurements to be made on spectral lines between 9 and $12 \mu\text{m}$. Heavy mass loss on the AGB promotes circumstellar formation of a variety of polyatomic molecules (e.g., silane, ammonia, ethylene). The location of molecular formation can be determined by measuring the angular size of the absorption region in the line core. This can be derived from a comparison of visibilities made on the line frequency where some of the dust radiation has been absorbed, and at a frequency where there is no molecular line or absorption. Although other molecular measurements have been made at millimeter and submillimeter wavelengths, except for the case of maser radiation where very high resolution has been obtained at radio frequencies, the measurements at $10 \mu\text{m}$ made by the ISI provide substantially higher spatial resolution on molecules near stars than what has been practical with millimeter wave spectroscopy. Distribution of ions around high temperature stars can also be measured, using mid-infrared fine structure lines.

When the filter-bank system is in operation, before going into the correlator the IF signal from each telescope is diverted through a dual-channel bank of 32 high-resolution RF filters (one bank for each telescope). Each filter has a 60 MHz bandwidth, corresponding to a spectral resolution, $\lambda/\Delta\lambda$, of 450,000 at $11 \mu\text{m}$, fine enough to resolve absorption line cores as narrow as $\sim 0.6 \text{ km s}^{-1}$. The filterbank spans a frequency range of 270 to 2190 MHz, most of the IF bandwidth of the ISI.

Appropriate gas mixtures and tuning of the laser oscillators are used in conjunction with careful planning of observing times to account for Doppler shifts due to the

Earth's motion relative to the Local Standard of Rest (LSR). This ensures that the desired spectral features fall within the bandpass of the filter bank. Any combination of the 32 filters can be selected, with that selection being identical for both channels. Switching of filters in and out of the IF chain is performed by RF solid-state switches under computer control. The filter bank is programmed to chop between a narrow filter bandpass centered on the desired spectral feature and a broad bandpass covering the continuum; this bandpass chopping is done on the timescale of atmospheric fluctuations (~ 4 Hz) for optimal signal to noise ratio (SNR) and precise calibration. More detailed instrumental description and analysis of the first results can be found in Monnier (1999).

Since CO_2 lasers, even with all practically available isotopes of CO_2 , oscillate only on a selection of discrete lines, not all spectral lines in the $10 \mu\text{m}$ region can be observed. It is hoped that in time appropriate tunable solid-state lasers will become available.

2.6. Control Architecture

The various subsystems of the ISI are controlled by an interconnected set of computers as shown in Figure 9. These computers are connected by an ethernet backbone and are overseen by a Sun Microsystems workstation (host computer) running the Solaris operating system. Custom software resident on the host computer provides a graphical user interface giving the observer control over data acquisition, real-time analysis of incoming data, and all telescope and detection system functions.

The real-time system consists of three single-board computers (Motorola MVME-147) running the VXWorks operating system. One such computer is resident in each telescope (e.g., VME1) and is used to control and read out instruments such as lock-in amplifiers, laser distance interferometers, and analog-to-digital converters. These computers also calculate parameters for the servo system (desired encoder counts) at one second intervals and pass them to the servo computer via a serial line. An additional real-time computer (VME0) controls interferometer functions such as calculation of lobe rotation frequencies, delay settings, and filter-bank bandpasses. This computer also

samples the output of the correlator with a fast analog-to-digital converter. Data taken with the real-time computers are archived on the disk drive of the host computer.

Also in each telescope are Intel 486 based PC computers running the MSDOS operating system (e.g., Isicam1). These computers control the guider cameras and tip-tilt actuators and also are connected to the servo computer via a custom interface box connected to a serial port. The interface box converts from ASCII command strings to encoder counts used by the servo microprocessors. The PCs also connect to the host computer by ethernet through PC-NFS.

Synchronization with external events is achieved using timing signals broadcast by radio station WWVB from the National Institute of Standards and Technology. Time codes from a WWVB receiver are distributed to timing boards in the real-time computer's backplane. These boards provide appropriate interrupts for the real-time computers.

3. DATA ANALYSIS

The primary data stream digitized by the ISI comprises three signals: the infrared power signals in the two telescopes IR_1 and IR_2 , and the fringe signal F . As described in § 2.4, the frequency of the fringe signal is held at 100 Hz by the lobe rotator. However, in practice, atmospherically induced pathlength fluctuations above each telescope cause the fringe power to be smeared over some range of frequencies, which can be as large as 40 Hz on a night of very poor seeing or as narrow as 0.001 Hz during excellent seeing (cf. Fig. 10). The fringe power P_F is extracted from the power spectrum of the fringe signal F by integrating over some frequency range (typically between 2 and 10 Hz unless seeing is very poor) centered on 100 Hz, and subtracting the noise power determined by measurement of it in adjacent frequency regions.

An interferometer characteristically measures fringe visibility, normally defined as

$$V \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (1)$$

where I_{\max} and I_{\min} are the maximum and minimum intensities when equal intensities of light from the star into two

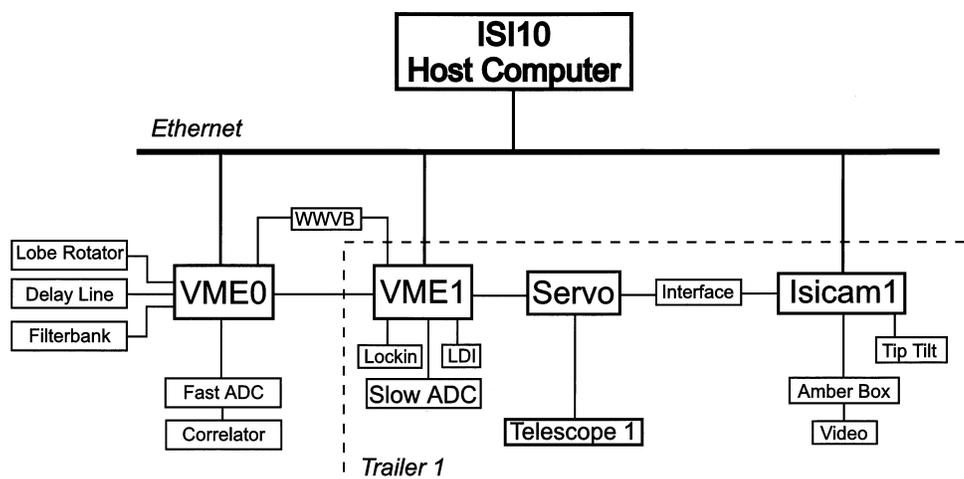


FIG. 9.—Schematic of the ISI Control Circuitry and its connections to subsystems in telescope 1. Isicam is the near IR Amber camera which images a star being tracked. LDI represents one of the He Ne laser distance interferometers used to measure changes in optical path lengths within the telescope. WWVB indicates a timing system linked to WWVB of the National Institute of Standards and Technology. ADC represents an analog to digital data conversion system.

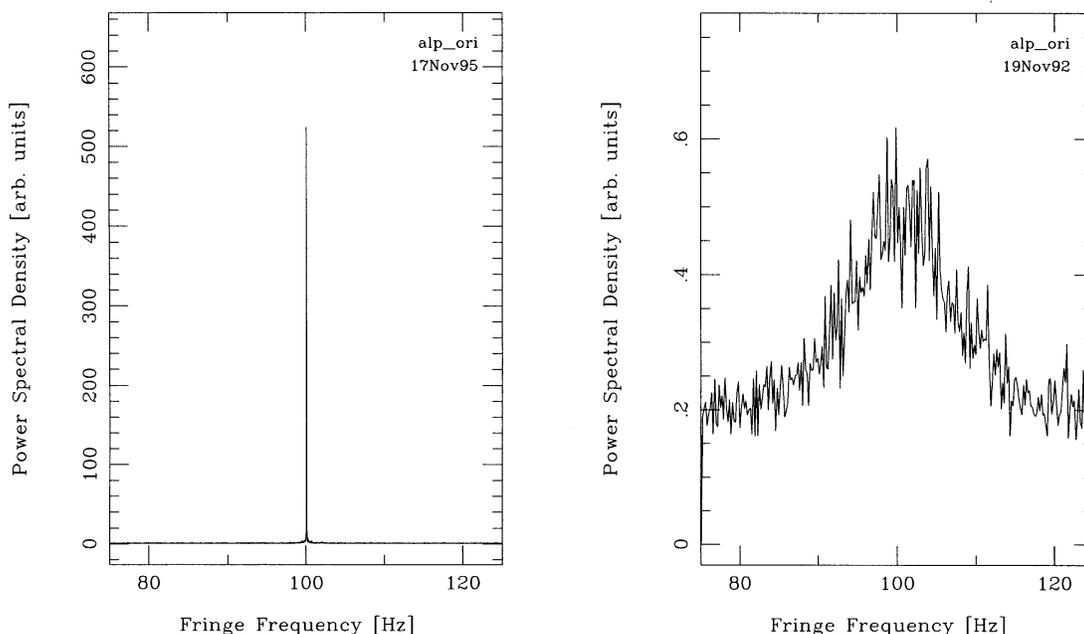


FIG. 10.—Fringe power spectrum output from the correlator, under different seeing conditions. The natural fringe frequency is removed and the fringes are modulated at 100 Hz by the phase shifter or “lobe rotation” portion of the instrument. Measuring the strength of the fringes, or “fringe contrast” is then a simple matter of integrating the correlator power spectrum shown here. The extremely narrow spectrum on the left is indicative of excellent atmospheric seeing conditions. It results from approximately 500 s of observation and its spectrum is approximately 0.002 Hz wide, indicating very little phase modulation. Poor atmospheric seeing causes phase modulation of the signal components, resulting in broadened spectra such as that shown at right. This spectrum is the result of approximately 250 s of observation, and its unusually large width indicates large phase fluctuations in times as short as 0.02 s.

telescopes interfere constructively and destructively respectively. This is equivalent to the fraction of the total power through one telescope which can interfere completely with that from the other. Because the signal in each telescope is heterodyne detected by matching an incoming signal with the pure wavefront of a local oscillator, only a pure plane wave is detected and hence interference between the two telescope signals can be complete. However, since the intensities from the two telescopes are not necessarily equal, the fringe visibility measured by the ISI is the square root of the ratio of the total fringe power P_F to the product of the total IR power in each telescope, or $(P_F/IR_1 \times IR_2)^{1/2}$. This ratio is proportional to the visibility as defined above, and is calibrated by observations on a star which is small enough to approximate a point source, for which the visibility is unity by definition. Calibration is needed any time there is a major change in the system such as a baseline change, and for surety is usually checked every second or third night of observation.

The phase of a fringe may also be measured by making a Fourier analysis of the fringe signal, and accurate phase information can allow good astrometric measurements (cf. Sutton, Subramaniam, & Townes 1982). This phase in a two-telescope system gives no useful information for imaging a source since there are unknown differences between the optical pathlengths to the two telescopes. However, it can give information about atmospheric pathlength fluctuations or the angular separation of two celestial objects. And three telescopes with three baselines do give phase information which can be useful in construction of an image, as will be noted below.

4. SCIENCE WITH THE ISI

Several publications (Lipman et al. 2000; Hale et al. 1997; Lopez et al. 1997; Monnier et al. 1997; Bester et al. 1996;

Danchi et al. 1994) have already detailed recent measurements and analyses performed by the ISI, so only a brief summary of results is given here. The ISI has been used primarily to study the spatial distribution of dust around late-type stars by measuring visibility curves at 11.15 μm and applying spherically symmetric radiative transfer models. Models are used to estimate dust shell inner radii, temperature, and optical depth at 11 μm by fitting the measured visibility data. For most of the approximately 20 stars observed, the ISI has resolved the inner radii of dust shells (Danchi et al. 1994), and in a few cases resolved and measured the size of the central star (Bester et al. 1996) with baselines presently in use. The latter measurements provide clear-cut examples of the limb darkening of stars, which makes their size at visible wavelengths appear to be less than the actual size, as measured more precisely by mid-IR interferometry (Bester et al. 1996). The recent increase to 56 m in baselines available allows still more accurate and extensive measurements of stellar sizes. With continual monitoring, changes have been observed in the measured visibilities due to variations in stellar luminosity, and also due to movements and changes in the dust surrounding these stars. For a substantial fraction of the red giant stars observed, dust shells are created episodically with periods ranging between about 10 and 100 yr—timescales that are not consistent with the stellar luminosity changes, which normally have periods of 1–2 yr. For some stars, evidence for multiple dust shells are seen as the result of these episodes (Hale et al. 1997; Monnier et al. 1997). Stars which do not expel gas episodically have visibility curves consistent with dust created by a uniform outflow that varies in temperature with the stellar luminosity. Dust features have also been observed to exhibit spatial motion on time scales as short as one year (Hale et al. 1997), providing stellar distance estimates. In some cases, the emission of dust and its

distribution is quite complex and asymmetric (Lopez et al. 1997; Monnier et al. 1999). The complexity of dust shells and their variation with time make it important to have repeated and careful measurements on at least a few prominent stars of a given type, and with phase closure to detect asymmetric distributions.

5. QUALITY OF PERFORMANCE: PRECISION AND FEASIBILITY

The quality of an interferometer may be measured by a number of parameters—these include the precision of visibility measurements, sensitivity in measuring weak sources, precision in measuring fringe phases, range of baselines available, and generally flexibility. Some of these factors will be discussed in this section.

Although complete interference can occur if the wavefronts are both perfectly planar as they strike the two telescopes, they are likely to be distorted over a given telescope aperture by imperfect seeing, which can give misleadingly low values of visibility. Heterodyne detection selects and detects only the components of the wavefront of the stellar radiation that are in phase with the wave front of the laser local oscillator (cf. Kingston 1978), and thus tends to prevent this difficulty. A similar result can be achieved when direct, rather than heterodyne, detection is used by spatial filtering with a glass fiber to obtain a single geometric mode. However, this usually entails some loss of signal.

The fundamental noise power for a heterodyne detector (which detects only one polarization, i.e., that of the LO) is equivalent to an average of one quantum per second per unit bandwidth in the same polarization as the LO. For an ideal photodiode, the average noise power is thus

$$N_h = hv \sqrt{\frac{2\Delta\nu}{t}}, \quad (2)$$

where hv is the quantum energy, $\Delta\nu$ is the single sideband bandwidth in Hz, and t the postdetection averaging time in seconds (cf. Teich 1970; Kingston 1978; Townes 1984). Since heterodyne detection has the ability to measure the phase of a wave, and phase is complementary to energy or number of quanta, this noise is an inescapable result of quantum mechanics and the uncertainty principle (cf. Serber & Townes 1960; Kimble & Wells 1987). The ratio of uncertainties in number of photons to those in phase of the wave can be changed while still satisfying the uncertainty principle (Serber & Townes 1960; Kimble & Wells 1987), but such possibilities are not very practical for heterodyne detectors and are not considered here.

The fundamental noise for an ideal direct detector, which detects both polarizations, is due to thermal radiation striking the detector, which for an ideal detector produces a noise power fluctuation of

$$N_d = hv \sqrt{\frac{2\Delta\nu}{t} \frac{1 - \varepsilon}{e^{hv/kT} - 1}}, \quad (3)$$

where $\Delta\nu$ is bandwidth, t the averaging time, T the temperature of optics and atmosphere through which the signal is received, and ε the fractional transmission of radiation reaching the telescope. Direct detection does not determine the phase of a wave, and hence noise due to the uncertainty principle is not present; in principle, noise is due only to fluctuations in number of quanta in the radiation received.

These fluctuations are also present in heterodyne detection, but at IR frequencies are generally much smaller than the uncertainty principle noise, and hence are omitted from expression (2).

For a useful signal, equivalent power of noise fluctuations must be substantially less than the signal power P_v from a source observed. The signal to noise ratio for stellar flux in the two cases, assuming ideal detectors with 100% quantum efficiency, is

$$\left(\frac{S}{N}\right)_h = \frac{P_v}{hv} \sqrt{2\Delta\nu t} \quad (4)$$

and

$$\left(\frac{S}{N}\right)_d = \frac{P_v}{hv} \sqrt{\frac{2\Delta\nu t (e^{hv/kT} - 1)}{1 - \varepsilon}}, \quad (5)$$

where P_v is the power in each polarization per unit bandwidth (Hz). Thus, even when direct detection has bandwidths as narrow as heterodyne detection, it appears to have a substantial advantage by a factor:

$$\sqrt{\frac{e^{hv/kT} - 1}{1 - \varepsilon}}, \quad (6)$$

and indeed it does under some circumstances. For a wavelength of 10 μm , room temperature T of 293 K, and transmission $\varepsilon = 0.9$ (characteristic of atmospheric transmission at 10 μm), this factor can be as large as 37. For visible or near-IR radiation the sensitivity advantage of direct detection is usually overwhelming, even though direct detection interferometers have complex enough optics that the net transmission of atmosphere and optics is usually substantially less than 0.9—sometimes as low as 0.10. However, there are a number of other considerations, which for wavelengths as large as 10 μm can in some cases give a substantial advantage to heterodyne detection, and in other cases to direct detection. These will be discussed below.

The signal to noise for measurement of fringe power in an interferometer with heterodyne detection is (Townes 1984; cf. also Johnson 1974, for detailed discussions of detection with a photoconductor)

$$\left(\frac{S}{N}\right)_{\text{Fringe}} \equiv \left(\frac{\text{vis} \times P_v}{hv}\right)^2 (t_0 t)^{1/2} \Delta\nu, \quad (7)$$

where $P_v \equiv$ the power per unit bandwidth of source, as in equation (4), $hv \equiv$ the quantum energy, $\text{vis} \equiv$ the visibility or fraction of the source power which provides interference, $t_0 \equiv$ the total observing time, $t \equiv$ the length of time atmospheric fluctuations do not change the fringe phase by more than approximately 1 rad, and $\Delta\nu \equiv$ the single sideband IF bandwidth, as in equation (4). The signal-to-noise ratio (S/N) for fringe amplitude, and thus for visibility determination, is proportional to the square root of that for fringe power.

The S/N for a direct detection interferometer involves somewhat more complex considerations. This is because the relative path lengths for the signals from the two telescopes must be accurately tracked, and this tracking usually involves radiation of a different wavelength and bandwidth from that used for visibility measurements. If the relative path lengths can be tracked accurately, then the theoretical S/N for fringes is comparable to that for power, but reduced by the visibility. It can in principle be made large by long-

term averaging. However, the path lengths themselves fluctuate quite rapidly on timescales of roughly 0.01–1.0 s due to atmospheric seeing, and there must be enough sensitivity to determine the phase of interference between light from the two telescope sources during a time as short as these fluctuations. If the S/N approaches unity or less for these short times, no measurements of fringe intensity can be made. Long-term observations for averaging and improving the S/N are then also not possible.

Equations (4) and (5) assume essentially perfect detectors and optics. Actually, detectors normally have quantum efficiencies in the range 0.2–0.8, which reduces sensitivity by these factors. In addition, transmission of signals through the optical systems involves losses, which for stellar interferometers typically reduces signals by factors between 0.05 and 0.8, depending on the number of mirrors or other optical components involved.

Heterodyne detection sensitivity actually obtained on interference fringes is rather close to the theoretical limit given above. For fringe amplitude, the ISI is within a factor of about 4 of this limit, with much of this factor due to present detector quantum efficiencies being between about 0.25 and 0.40 rather than unity. This closeness to the theoretical limit is in part because heterodyne detection conveniently eliminates unwanted radiation outside a chosen bandwidth and partly because only a very small source at a long distance can produce interference, so more local and extraneous radiation does not provide a false interference signal. However, in contrast to fringe observation, the detection of power radiated into a single telescope suffers from variations in stray radiation as does direct detection, discussed below.

For bandwidths narrower than about 1 cm^{-1} (3×10^{10} Hz), the dominant noise in direct detection is typically associated with detector dark current and readout fluctuations (or “read” noise), combined with some stray radiation. For narrow bandwidths, the highest sensitivity normally achieved in direct detection groundbased astronomical systems is $\sim 2 \times 10^{-15}$ W for a 1 s averaging time (J. H. Lacy 1998, private communication) time.⁵ This equals the theoretical noise for heterodyne detection with a bandwidth of about 5×10^9 Hz. Of course, heterodyne detection also does not give perfect theoretical performance, largely because the quantum efficiency of detectors is not 100%. However, its noise does decrease with decreasing bandwidth in accordance with theory rather than reaching a lower limit, as is characteristic of direct detection. Hence, for bandwidths appreciably narrower than about 5×10^9 Hz, heterodyne detection is typically the more sensitive since its noise continues to decrease with decreasing bandwidth. This is important for certain applications such as measuring spectral lines (which are often as narrow as 5×10^7 Hz).

For continuum radiation, direct detection has the advantage of being able to use broad bandwidths. The single-sideband bandwidths for efficient heterodyne detectors at $10 \mu\text{m}$ are presently $\sim 0.1 \text{ cm}^{-1}$, or 3×10^9 Hz. However, quantum-well detectors have been made with substantially larger bandwidths, and those as large as $\sim 0.5 \text{ cm}^{-1}$ can be envisioned. Direct detection can in principle include essentially all the mid-IR radiation transmitted by the atmo-

sphere, or a bandwidth as large as 600 cm^{-1} . However, without further division of the radiation such a broad band of wavelengths results in a rather broad range of resolutions, with varying visibilities. Hence a range of not more than 10%, which is $1 \mu\text{m}$ in wavelength range or 100 cm^{-1} , is the maximum considered here. A bandwidth of $10\text{--}20 \text{ cm}^{-1}$ would be normal, given a 1%–2% range of resolution. Such bandwidths provide a substantial apparent advantage over the narrower band heterodyne detection. However, for broad bandwidths, fluctuations in radiation from the sky and optics usually dominate the noise rather than fundamental quantum fluctuations or detector noise. For a $1 \mu\text{m}$ bandwidth, experience shows that non-fundamental noise prevents the detection of power below about 5×10^{-14} W for an averaging time of 1 s.⁶ This is about 40 times larger than what would be obtained with the theoretical limit given above, and for this case much of the expected gain from broad bandwidths is negated. In principle, the sensitivity may possibly come closer to theoretical values, but this represents a long-standing challenge to experimentalists.

Interferometers frequently use multiple telescopes to obtain many baselines simultaneously; for example the VLA radio interferometer uses 27 telescopes (or 351 baselines). This requires sending individual signals from each telescope to 26 different interference measurements or correlators. For heterodyne detection, such a system involves no further loss in S/N because after detection the signal can be amplified and divided without introduction of any significant noise, as is done in radio interferometry. However, a direct detection interferometer would normally divide the signal of each telescope into 26 equal parts. This would reduce the signal by the same factor and negate much of the theoretical signal to noise advantage noted above. In addition, bringing the signals together from separate telescopes for correlation is also more tractable in the case of heterodyne detection, requiring only electrical cables rather than the evacuated light pipes and multiple optical components which are needed for direct detection. These considerations lead us to conclude that a many-telescope multiple baseline system for the mid-IR region is probably simpler, cheaper, and more flexible with heterodyne detection than with direct detection.

A prominent task for direct detection interferometry is to accurately track the relative delay between any two beams which are to interfere. As noted above, this is generally done at a wavelength different from that used for fringe measurement and with a relatively broad bandwidth. However, a useful signal for delay compensation must be obtained in a time as short as atmospheric fluctuations, hence in about 0.01 s. If this signal is detected well, this short time has no direct effect on the sensitivity of measurement, and the delay compensation is helpful in decreasing noise because the phase of interference does not then fluctuate, as it may for heterodyne detection. However, if the signal is lost visibil-

⁶ For 10% bandwidth at $11.7 \mu\text{m}$ and 1 s averaging time, the MIRAC2 camera used on the UKIRT telescope has a sensitivity (1σ uncertainty) of 5×10^{-13} W (Hoffman et al. 1998). For $1 \mu\text{m}$ spectral width and 1 s averaging time, specification for the MICS mid-IR camera on the UKIRT telescope is 1.2×10^{-13} W, and for $0.1 \mu\text{m}$ spectral width, it is 1.3×10^{-14} W for a 1 s averaging time (Miyata et al. 1999). When used as a $10 \mu\text{m}$ camera on the Keck telescope, the LWS has a sensitivity 5.9×10^{-14} W for a $1 \mu\text{m}$ bandwidth and 1 s averaging time (Keck website LWS Instrument Document).

⁵ Also personal experience of C. H. Townes.

ities cannot be measured. In contrast, the narrow bandwidth and IF delays of heterodyne detection allow the use of calculated delays without direct tracking, and hence the time for averaging a signal can be arbitrarily long regardless of signal strength. This is another basic convenience of narrow bandwidths. For an averaging time of one hour instead of 0.01 s, this can boost the relative advantage of heterodyne interferometry is detecting fringe power by as much as 600 (cf. eq. [7]), or in detecting fringe amplitude by a factor of $\sqrt{600} \approx 25$. The averaging time can also be increased to many hours by many nights of observing, as is often done in radio astronomy. Direct detection can avoid this disadvantage, however, if there is a bright star close enough to be within the same isoplanatic patch as the object being measured, since then the bright star can be used for delay line tracking. And there are some other techniques, not yet generally used, which can ameliorate this sometimes large disadvantage of direct detection. These might involve use of multiple wavelength bands, each possibly as large as atmospheric dispersion allows, to track the path-length variations, or multiple tracking units, each involving very short distances.

If a relatively wide bandwidth and single baseline are used, and if there is a strong guide star within the same isoplanatic patch as the object observed, then for wavelengths as short as 10 μm direct detection has a large advantage. However, if many baselines are used and there is no strong guide star, or if bandwidths less than about 1 cm^{-1} are used, heterodyne detection has a substantial advantage. Both techniques are useful, each has its own optimum functions, and achievement of ideal performance in either one is challenging.

The measurement of fringe intensity with the ISI has for some time come close to theoretical sensitivity expectations because averaging of the fringe power can be carried out reliably over a long period of time and there are no important spurious sources of noise or false correlation signals. Measurement of infrared power has, however, the usual problems of spurious signals from room-temperature radiation, which are well known in any mid-IR measurement. The ISI has also had a particular type of spurious signal peculiar to a heterodyne detection system. This, and its suppression, are described below.

The local oscillator power on the ISI detectors exceeds that of the brightest 10 μm sources in the northern sky by about a factor of 10^8 , and it produces most of the current and the RF (quantum) noise in the heterodyne detector. Some of the local oscillator radiation is scattered back towards the telescope and then scattered back again into the detector. In this path it encounters the signal chopper which hence chops the scattered local oscillator power synchronously with the stellar signal. The scattered radiation field can be either in or out of phase with that of the local oscillator, and this will change from one to the other as its pathlength changes by a half wavelength due to temperature variations or other changes. It can thus produce an unpredictable signal of infrared power (not of fringe power) varying on time scales between a few seconds and many minutes, which is difficult to isolate from the stellar signal to be detected. Because the scattered wave is coherent with the local oscillator, even with an intensity only 10^{-10} of the local oscillator it can give a fractional change in power at the detector of 2×10^{-5} , and hence a false signal of appreciable magnitude. To correct for this varying signal, it was

noted that because of coherence between the local oscillator and scattered power, as the phase of the scattered power changes their combination also produces a variation in the detector current, and this variation is synchronous with variations in the spurious signal. This varying current has hence been converted into a voltage of appropriate magnitude and subtracted from the apparent incoming signal. The subtraction can rather completely cancel the spurious signals due to scattering and its use has recently allowed an improvement in the precision of IR power measurements by a factor of 5–10.

Sensitivity actually obtained by the ISI with a detector of quantum efficiency 40%, and bandwidth 2×10^9 Hz, is within a factor of 2.5 of what is predicted by theoretical equations such as (4) and (7). The missing factor is partly associated with net transmission efficiency from star to detector, which is estimated as 60%. This leaves only a factor of 1.5 unaccounted for, possibly due to further imperfections in optics or detection.

Modest improvements in detectors and associated technology are to be expected, but present sensitivity can provide measurements of fringe power with useful signal to noise on sources with intensities about 50 janskys at 10 microns wavelength after an hour of observation, assuming good seeing conditions. This intensity is about 100 times, or 5 mag, less than that of a number of the more intense stellar sources. There is no clear limit to the number of hours of observation, so that still weaker sources can be studied.

6. ATMOSPHERIC FLUCTUATIONS AND COMPENSATION

Atmospheric “seeing,” or more properly, relative fluctuations in optical path lengths through the atmosphere to the two apertures of an interferometer, are a primary source of difficulties and error in interferometric visibility and phase measurements. They change the fringe phase rapidly, and thus broaden out the spectral range of the fringe rather than allowing a pure frequency (cf. Fig. 10). Each ISI telescope has some compensation built into it since HeNe laser interferometers were installed to continually monitor the path length within each telescope from the flat mirror to the parabola and then back to the optics table and detector. Differences between changes in these two paths can be put into an effective path-length delay by appropriately modulating the phase of the interference signal as it is being spectrally analyzed. This correction is relatively small since it comes from a small total path, but it can be significant, and is particularly important if there is some mechanical motion of the optics. However, most of the fluctuations come from the atmosphere above the telescope.

Atmospheric path-length fluctuations have been measured with the ISI system in several ways and their characteristics found to agree only roughly with Kolmogorov turbulence theory, or the Kolmogorov-Taylor approximation, which have generally been assumed to be a good first-order model. The size of relative pathlength fluctuations typically increases with separation of optical paths and with decreasing frequency of fluctuation somewhat more slowly than what Kolmogorov-Taylor approximation suggests (Bester et al. 1992). This is partly due to the fact that a substantial fraction, typically about 50%, of the fluctuations are within about 30 m of the ground and are hence associated with ground structure and convection, which are not included in the Kolmogorov-Taylor approximation. The importance of this low-lying turbulence and its approx-

imate height have been demonstrated by the correlation of fluctuations within the telescope, measured by the HeNe laser interferometers, with the fringe phase fluctuations associated with paths through the entire atmosphere (Treuhaft et al. 1995). Because of this correlation, it is at least sometimes practical to compensate for about 50% of the slower fluctuations, and thus improve measurements by narrowing the spectral width of a fringe. A series of local and fast temperature measurements near the telescope's optical path and extending some tens of meters into the atmosphere are planned in order to examine the possibility of further compensation of "seeing," or path-length fluctuations.

7. CONCLUSIONS AND FUTURE DIRECTIONS: TOWARD AN IMAGING ARRAY

The Berkeley Infrared Spatial Interferometer utilizes infrared-optimized mobile telescopes configured as a heterodyne interferometer in order to explore high-resolution structure of a range of celestial targets. As the only scientifically productive mid-infrared long-baseline interferometer operational for over a decade, the ISI is interesting both for its innovative technology and for its exploration of previously inaccessible realms of angular size and wavelength. Although fundamental signal to noise calculations appear to strongly favor homodyne or direct over heterodyne detection schemes, there are a number of factors, including automatic selection of spherical wave-front components of incoming signals which can help insure precision in measurement of visibility, nonfundamental but every-present sources of noise, the possibility of longer averaging times, reliability of data calibration, relaxed specifications on IF signal path hardware, greater flexibility and S/N advantages for large arrays, and capability for ultrahigh-resolution spectral line studies which tend to favor heterodyne detection. The two methods can be seen to be largely complementary, with different strengths and weaknesses, and roles for both can be expected in this spectral region.

Continuing improvements in sensitivity and performance, many of which have been described in this report, bring new astrophysical goals within reach of the instrument. Recently initiated programs for studies of young stellar objects and hot massive stars are now underway. In addition, periodic monitoring of stars already observed is crucial to revealing secular changes due to evolution of the dusty envelopes surrounding evolved stars. The 1999 observing season has seen the commissioning of a 56 m baseline (between pads 6 and 7 in Fig. 6), which roughly doubles the resolving power of the instrument and enables accurate direct measurement of stellar diameters for a handful of the largest M giants and supergiants. Definitive sizes of these stars have been elusive; measurements at

shorter wavelengths are complicated by the presence of pronounced molecular blanketing or radiation due to various layers in the stellar photosphere, circumstellar scattering, nonuniform temperatures over the stellar disk, limb-darkening, and other processes. All of these difficulties should be greatly alleviated in the mid-infrared, where the low opacity of the material in the extended circumstellar environment should lead to a clean estimate of the true photospheric diameter if a narrow spectral region free of molecular lines is used. The star should, in addition, represent a disk of relatively uniform brightness at mid-IR wavelengths.

First results from narrow spectral band studies with the filterbank described in § 2.5 have been obtained. This unique combination of high spectral resolution (60 MHz bandwidth) and subarcsecond spatial resolution has enabled studies of the distribution of individual molecular species formed around stars (Monnier 1999).

It is expected that a third telescope, currently under construction, will be integrated into the ISI, forming a three-telescope imaging array in 2001. The more rapid accumulation of multiple baseline measurements which this will allow is made particularly important by the rather rapid changes in stellar dust shell radiation which have been found as a result of variation in stellar luminosity, dust cloud motions, and new dust formation. In the interim, prior to full integration, the third telescope will be utilized to enhance the Fourier coverage by switching between different pairs of telescopes, with each such baseline yielding a separate track in the (u, v) plane. Full integration of the third telescope, in addition to providing three simultaneous baselines, will also allow measurement of one closure phase. This observable, computed by finding the differences of the Fourier phases around a closed triangle of baselines, gives the interferometer a new datum with which to constrain stellar images. The addition of phase measurement yields information on source symmetry which is lacking in visibility amplitude measurements alone. Thus the ISI will be able to determine noncentro-symmetric brightness distributions, and provide mid-IR imaging by interferometry, following aperture synthesis methods taken from radio astronomy.

The authors are grateful for and acknowledge the contributions of a number of individuals during the course of this work, particularly E. C. Sutton, M. A. Johnson, A. L. Betz, B. Sadoulet, C. G. Degiacomi, T. R. Geballe, R. Fulton, R. H. Weitzman, A. Arthur, T. McDonald, and P. Bratt.

This work has been supported in part by grants from the Office of Naval Research (FDN00014-96-1-0737), the National Science Foundation (AST-9221105, AST-9321289, AST-9321384, and AST-9500525), the Perkins Fund, the France-Berkeley Fund, and the Space Telescope Science Institute (STSI-G0-07416.96A).

REFERENCES

- Bester, M., Danchi, W. C., Hale, D., Townes, C. H., Degiacomi, C. G., Mékarnia, D., & Geballe, T. R. 1996, *ApJ*, 463, 336
 Beser, M., Danchi, W. C., Degiacomi, C. G., Greenhill, L. J., & Townes, C. H. 1992, *ApJ*, 392, 357
 Bester, M., Danchi, W. C., & Townes, C. H. 1990, *Proc. SPIE*, 1237, 40
 Danchi, W. C., Bester, M., Degiacomi, C. G., Greenhill, L. J. & Townes, C. H. 1994, *AJ*, 107 (4), 1469
 Danchi, W. C., Bester, M., & Townes, C. H. 1988, in *ESO Conf. Workshop Proc.*, 29, 867
 Freed, C. 1968, *J. Quant. Electronics*, QE-4, 404
 Hale, D. D. S., et al. 1997, *ApJ*, 490, 407
 Hanbury Brown, R., Davis, J., & Allen, L. R. 1967, *MNRAS*, 137, 375
 Hanbury Brown, R., & Twiss, R. Q. 1956, *Nature*, 178, 1046
 Hoffman, W. F., Hora, J. L., Fazio, G. G., Deutsch, L. K., & Dayal, A. 1998, *Proc. SPIE*, 3354, 647
 Johnson, M. A. 1974, Ph.D. dissertation, University of California at Berkeley
 Johnson, M. A., Betz, A. L., & Townes, C. H. 1974, *Phys. Rev. Lett.*, 33, 1617
 Kimble, H. J., & Wells, D. F. 1987, *J. Opt. Soc. Am. B*, 4, 1450
 Kingston, R. H. 1978, in *Optical Sciences*, Vol. 10, ed. D. L. MacAdam (New York: Springer), 28

- Lipman, E. A. 1998, Ph.D. thesis, University of California at Berkeley
- Lipman, E. A., Bester, M., Danchi, W. C., & Townes, C. H. 1998, Proc. SPIE, 3350, 933
- Lipman, E. A., Hale, D. D. S., Monnier, J. D., Tuthill, P. G., Danchi, W. C., & Townes, C. H. 2000, ApJ, 532, 467
- Lopez, B., et al. 1997, ApJ, 488, 807
- Miyata, T., Kataza, H., Okamoto, Y., Junabe, T., Onaka, T., Yamashita, T., Nakamura, K., & Shibai, H. 1999, PASP, 111, 750
- Monnier, J. D. 1999, Ph.D. thesis, University of California at Berkeley
- Monnier, J. D., et al. 1997, ApJ, 481, 420
- Monnier, J. D., Tuthill, P. G., Lopez, B., Cruzalebes, P., Danchi, W. C., & Haniff, C. A. 1999, ApJ, 512, 351
- Serber, R., & Townes, C. H. 1960, Quantum Electronics, ed. C. H. Townes, Columbia Univ. Press, 223
- Sutton, E. C., Storey, J. W. V., Betz, A. L., Townes, C. H., & Spars, D. L. 1977, ApJ, 217 L97
- Sutton, E. C., Subramanian, S., & Townes, C. H. 1982, A&A, 110, 324
- Teich, M. C. 1970, Semiconductors and Semimetals, ed. R. K. Willardson & A. C. Beer (New York: Academic), 361
- Thompson, A. R., Moran, J. M., & Swenson, G. W. 1986, Interferometry and Synthesis in Radio Astronomy (New York: Wiley)
- Townes, C. H. 1984, J. Astrophys. Astron., 5, 111
- Townes, C. H., et al. 1998, Proc. SPIE, 3350, 908
- Treuhaft, R. N., Lowe, S. T., Bester, M., Danchi, W. C., & Townes, C. H. 1995, ApJ, 453, 522