Introduction to Infrared Observing Methods

Ron Probst
January 2009

The article below, by my NOAO colleague Dick Joyce, remains an excellent introduction to the challenges posed for 1-5 micron infrared observations through the Earth’s atmosphere, and ways used to overcome them.

NEWFIRM users should also see the section “How to observe in a pipeline-friendly manner” in the Introduction to the Quick Reduce Pipeline document. This can be accessed through the Web page sidebar in the Data Reduction Pipelines section of the listing.
OBSERVING WITH INFRARED ARRAYS

RICHARD R. JOYCE
Kitt Peak National Observatory†, P.O. Box 26732, Tucson, AZ 85726-6732

ABSTRACT Within the past 5 or 6 years, two-dimensional infrared arrays with the requisite sensitivity for astronomical observations have come into use. Although these arrays differ significantly in architecture and operation from the CCDs used in the visible, the general observing and data reduction techniques are similar. This paper will provide a brief overview of the infrared spectral region, construction and operation of infrared arrays, and techniques for observing in both imaging and spectroscopic applications.

INTRODUCTION

Although the field of infrared astronomy can be dated from William Herschel’s 1800 observation of solar infrared radiation, it is only within the past 25 years that infrared detectors have been developed with the requisite sensitivity for serious astronomical observations (Gillett et al. 1977; Joyce 1983), and within the past 5 or 6 years that two-dimensional infrared arrays have become available. An excellent review of the rapid progress in this field is the proceedings of the workshop on Astrophysics with Infrared Arrays (Elston 1991). Even the earliest versions of these arrays represented a revolutionary advance in capability over the single detectors previously available, and significant advances in performance and format size are still being made. Near-infrared arrays 256 pixels square are presently in use at several observatories, and 640 X 488 arrays are anticipated in the near future. Although these arrays and the associated instrumentation differ significantly from the CCDs used in the visible, most of the observational techniques are similar, and we have noted at NOAO that observers with previous CCD experience have had very little difficulty in adapting to the use of infrared array instrumentation.

While the historical infrared was defined by the limits of human vision (approximately 0.7 μm), a more modern demarcation is at a wavelength of 1 μm, which is near the limit of the CCDs (and, yes, photomultiplier tubes and photographic plates) used in modern astronomy. Further subdivision into the near-infrared (1 - 5 μm), mid-infrared (5 - 25 μm), and far-infrared (25 - 350 μm) can be

†Operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.
defined somewhat arbitrarily by the types of detectors and observing techniques required. The majority of array experience to date has been limited to the near-infrared, a fact which will be reflected in this paper, although the efforts of some groups who have made successful observations with arrays in the 10 μm region deserves mention (McCaughean and Gezari 1990; Gezari and Yusef-Zadeh 1990).

In contrast to the visible region, the transmission spectrum of the earth’s atmosphere has strong absorption bands at some wavelengths, due primarily to H₂O and CO₂. The intervening, relatively transparent regions have been designated as photometric bands identified by letters, in an analogous fashion to the visible UBVRI bands. These are the J (1.1 - 1.4 μm), H (1.5 - 1.8 μm), K (2.0 - 2.4 μm), L (3.0 - 4.2 μm), M (4.5 - 5.2 μm), N (8 - 13 μm), and Q (17 - 22 μm) bands. This nomenclature is used throughout the literature and will undoubtedly slip into this paper as well. Much of the early work in establishing the infrared photometric bands is due to Johnson (1966).

WHY INFRARED?

Although a friendly rivalry often exists between visible and infrared astronomers, the two fields are, of course, complementary. Stars are the most visible manifestation of luminous energy in the universe, and most of what is known about stellar physical properties and evolution has come from observations in the visible. However, many objects (planets, asteroids, comets, and hypothetical brown dwarfs) are too cool to radiate significant energy in the visible. The Infrared Astronomical Satellite (IRAS) found some active galactic nuclei to be among the most luminous objects in the universe, radiating far more energy in the infrared than in the visible. Obviously, infrared observations of these objects are a prerequisite to any understanding of the physical processes responsible for this behavior. Even in visible stars, infrared spectroscopy probes a different physics, that of molecular vibrations, from the atomic electronic transitions that constitute the majority of spectral features in the visible.

However, one of the most compelling general motivations for infrared observations is that our view of the universe is extremely limited at visible wavelengths by dust. Dust is astrophysically very important, as it is the medium through which elements produced by one generation of stars is passed to the next. Many late-type giant stars pass through a stage of extremely high mass loss near the end of their life. The earliest stages of star formation occur within dense molecular clouds. In both cases, the dust shells around the objects of interest absorb virtually all of the visible radiation. Even the nucleus of our own galaxy, approximately 100 times closer than that of M31, is undetectable at visible wavelengths because of absorption by the intervening dust in the plane of our galaxy (Fig. 1).

Due to the small size of these dust particles, their absorption efficiency decreases very rapidly with increasing wavelength. The extinction coefficient at 2.2 μm is only about 10% of that at 0.5 μm. As a result, the estimated 15 magnitudes (10²⁰) visual extinction of the galactic center is reduced to only 1.5 magnitudes (4) at 2.2 μm. An image at this wavelength (Fig 2) clearly shows the central bulge, as well as the plane of our galaxy. Another example, the star-forming region NGC 2024, is almost completely hidden at a wavelength of 0.7 μm by a dense dust lane,
Fig. 1. SRC I-band (0.7 µm) photograph of the region of the Galactic Center, covering a field roughly 27 arcmin square.

Fig. 2. Image of the same field as Fig. 1, at a wavelength of 2.2 µm.
but is resolved into a cluster of young stars at a wavelength of 2.2 μm (Fig 3). Actually, there is a missing piece here; the bright source in the center of the IR image was discovered by Grasdalen (1974) using a single-detector IR photometer, but until the image was made, that was the only known source in the star-forming region. The importance of the two-dimensional array is that it provides an unbiased coverage of the entire cluster, which is a prerequisite for attempting to model the initial parameters of the star-forming episode.

![Image of star-forming region](image.jpg)

Fig. 3. Images of the region NGC 2024, at wavelengths of 0.7μm [left] and 2.2μm [right]. (Courtesy R. Probst, KPNO).

PHYSICS AND ARCHITECTURE OF IR ARRAYS

The physical operating principles of the IR detectors used in arrays are similar to those utilized in CCDs, but the construction and operation differ in several significant aspects. This section will describe the layout and operation of a typical IR array used for astronomy, pointing out the similarities and differences. Fortunately, most of the operational differences can be hidden by clever operating software, making the user interface to CCDs and IR array systems virtually identical.
A CCD is basically an analog shift register formed by a string of closely spaced MOS capacitors. By a happy coincidence, the Si from which CCDs are constructed absorbs light which is sufficiently energetic (hν > 1.1 eV; λ < 1.1 μm) to excite electrons from the valence band to the conduction band. By proper layout, the absorption efficiency of radiation and the collection efficiency of the resultant photoelectrons can be optimised. The collected electrons are then "read out" using the shift register capability of the CCD, which has been optimized for transfer efficiency and low noise. The result is a high quantum efficiency, linear, low noise detector array.

Infrared detectors operating on the same principle are made from semiconductors with a smaller energy gap between the valence and conduction bands; a number of these are listed in Table I. The photosensitivity of PtSi results from an energy gap (Schottky barrier) formed at the junction of dissimilar materials; the energy gap in HgCdTe is dependent on the Hg/Cd ratio, although the value given in Table I is currently receiving the most attention for astronomical use. InSb and the doped Si materials have been used as single-element detectors for the past 15 years (Gillett et al. 1977).

<table>
<thead>
<tr>
<th>MATERIAL</th>
<th>Δ (eV)</th>
<th>λc (μm)</th>
<th>TEMPERATURE (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si (CCD)</td>
<td>1.1</td>
<td>1.0</td>
<td>150 - 170</td>
</tr>
<tr>
<td>Ge</td>
<td>0.7</td>
<td>1.7</td>
<td>77 - 100</td>
</tr>
<tr>
<td>HgCdTe</td>
<td>0.5</td>
<td>2.5</td>
<td>60 - 77</td>
</tr>
<tr>
<td>PtSi</td>
<td>0.25</td>
<td>5.0</td>
<td>40 - 60</td>
</tr>
<tr>
<td>InSb</td>
<td>0.23</td>
<td>5.4</td>
<td>30 - 40</td>
</tr>
<tr>
<td>Si:In</td>
<td>0.18</td>
<td>7</td>
<td>30 - 40</td>
</tr>
<tr>
<td>Si:Ga</td>
<td>0.07</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Si:As</td>
<td>0.05</td>
<td>27</td>
<td>10</td>
</tr>
</tbody>
</table>

The operating temperatures listed in Table I are significant in both the operation of the detectors and the design of the associated instrument. Electrons can be excited into the conduction band not only by photons, but by thermal energy, resulting in undesirable dark current. This is of the approximate form i_d ~ exp(-Δ/kT), where Δ is the energy gap of the material. Si CCDs are usually operated near 150 K to reduce the dark current to negligible levels. The IR detectors must therefore be operated at lower temperatures in inverse proportion to their long wavelength response. An ancillary consequence is the sensitivity to thermal background radiation from the ambient environment, which increases very quickly for wavelengths longer than 1.5 μm. It follows that the detector must not only be cooled for optimum operation, but it must also be extremely well baffled by cold (typically 77
Observing with Infrared Arrays

K) surfaces to admit only the solid angle subtended by the telescope optics. In addition, instrumental optics (mirrors, filters, gratings) must also be cooled. The net result is that IR detectors are usually built into their associated instrumentation, rather than being self-contained modules which can be mounted on a variety of instruments.

Although it would seem logical to construct a two-dimensional array along the lines of an optical CCD utilizing a material such as InSb, such attempts (even with small linear arrays) never produced detectors suitable for the signal levels encountered in astronomical applications. The reasons are both technical and physical. IR detector materials have not been made to function adequately as shift registers and amplifiers, particularly at the temperatures required for efficient detector operation. On the other hand, silicon is a very well-understood material with a mature fabrication technology. The high-efficiency, low noise CCDs available today are the result of many thousands of man-years of effort in silicon technology, unlikely to be duplicated for other materials.

The logical solution was the hybrid array, in which the detector array and readout array are fabricated separately, carefully aligned, and then joined. Small bumps of In metal are grown on each pixel to serve as the electrical and mechanical interface between the detector and readout sections. The architecture of such an array, the Santa Barbara Research Center (SBRC) InSb detector, is illustrated in Fig 4. The detector consists of a substrate of n-type InSb, with p-type InSb regions at the locations of each pixel; the resultant diode junction, when biased, forms a potential well which collects the photoelectrons. An additional gate structure controls the surface potential to reduce dark current and crosstalk between adjacent pixels. Finally, bumps of In metal are grown on each pixel.

The multiplexer is more complicated, consisting of row and column address decoders, source follower and reset switch unit cells, and an output amplifier. A more complete description is given by Fowler et al. (1987). For the purpose of this discussion, two points are emphasized. First, the multiplexer is fabricated from silicon, so the processing and operation benefit from well-understood technology. Secondly, the multiplexers used for the vast majority (if not all) the arrays used for astronomy are NOT CCDs, but rather consist of individual amplifiers for each pixel. One likely reason is that the In bump connection would have required the use of a surface-channel CCD, which is generally much noisier than the buried-channel CCD used as a detector.

Once the multiplexer is fabricated, bumps of In are grown on each input, as with the detector array. The detector and multiplexer arrays are tested separately, and those that pass this screening are deemed suitable for hybridizing. A multiplexer and detector array are carefully aligned, with the In bumps facing each other, and then pressed together with sufficient force to "cold weld" the In metal to form a solid mechanical and electrical bond between the two arrays. A schematic cross section is shown in Fig. 5.

Several advantages accrue from the hybrid layout. The gate structure on the array and the individual amplifier on the multiplexer mean that each pixel is effectively isolated from its neighbor. This eliminates a very bright source from
Fig. 4. Photomicrograph of an InSb detector array before hybridization, showing the diode and gate structures and the In bumps. (Courtesy Santa Barbara Research Center).

Fig. 5. Schematic diagram of an InSb backside-illuminated hybrid array, showing the In bump interconnect. (Courtesy Santa Barbara Research Center).
"blooming" into adjacent pixels, as well as the "charge trails" seen with CCDs, since there is no charge transfer from one pixel to another. In addition, the same multiplexer can be mated to a variety of detector materials, as long as the physical dimensions are the same. It is possible to read the charge on each pixel nondestructively (without resetting the bias), allowing multiple-read algorithms for reducing the read noise. Finally, the placement of the multiplexer behind the array (so-called backside illumination) results in essentially 100% fill factor, critical for astronomical applications.

There are, of course, some compensating difficulties. In a backside illuminated array, the photons enter the detector material on the side opposite the junction where the photoelectrons are collected. The thickness of the array is therefore very critical; if too thin, the absorption coefficient drops rapidly (i.e., the detector becomes transparent to radiation); if too thick, photoelectrons will be created so far from the junction that they may recombine before being collected. Either condition results in a dramatic decrease in the array quantum efficiency. For InSb, the optimum thickness is about 20 μm, sufficiently small that the detector must be mechanically thinned after hybridization with the readout. Even a small thickness variation can result in a significant gradient in quantum efficiency across the array. For arrays with dissimilar detector and multiplexer materials (such as InSb), differential thermal contraction may stress fracture the detector array or result in shorts between adjacent In bumps; this sets a practical upper limit of about 10 mm on the physical size of the array. Detectors fabricated primarily from Si (PtSi or doped Si) do not suffer this limitation.

ARRAY OPERATION

Only a brief description of the operation of the array will be given here; for more detailed information, see Fowler and Gatley (1991). Each pixel of the detector array may be considered a capacitor which, when biased to a preset voltage, acts as a potential well which attracts and collects electrons produced by photons absorbed in the material (as well as dark current). The readout circuit measures the voltage on the diode, which is related to the charge by the capacitance of the junction. This readout process is nondestructive in that it can be carried out at any time without affecting the charge on the pixel.

Fig. 6 is a schematic representation of the voltage on a pixel as a function of time. At the beginning of an integration, the voltage is set to a preselected bias level. This process is noted on the Figure by the intervals labeled as "bias". When the reset switch is opened, the voltage may shift slightly to a new value and then increase linearly with time as charge is collected by the diode junction, until the pixel is either rebiased or the charge "fills up" the potential well and saturates the pixel. The offset after biasing is proportional to √kTC (referred to as "KTC noise"), and is effectively an uncertainty in the actual charge on the diode capacitance C when the bias switch is opened. This is a random uncertainty, so the offset may vary in amplitude and sign from one reset cycle to the next.

A number of readout techniques are illustrated in Figure 6; a dot represents a point in time when the pixel is addressed and the voltage read by the multiplexer amplifier. Keep in mind that the multiplexer design permits nondestructive readout
of the array at any time during an integration, but each readout requires a finite time (typically 8 μs/pixel) and has an associated noise level (defined as the read noise).

READOUT TECHNIQUES

Fig. 6. Representation of the voltage on a single pixel as a function of time, to illustrate the different techniques of nondestructive readout described in the text. The dots represent times when the array is sampled, for "video", "delta-reset" (Δ-R), "double correlated sampling" (DCS), "triple correlated sampling" (TCS), "slope sampling", and "multiple correlated sampling" (MCS) techniques.

The simplest technique, known as the "video" readout, uses a single read just before the array is reset. This technique is used where high frame rates are required, as in video cameras or for astronomical applications such as high background or speckle imaging. Almost as fast is the "Δ-reset" technique, where the array is read during the reset interval and again just before the following reset; the integration time is the interval between the opening of the reset transistor and the second read. The obvious disadvantage is that the KTC offset appears as a noise
Observing with Infrared Arrays

factor, although non-uniformity is removed. A technique used widely in astronomy is that of "double correlated sampling" in which the array is read immediately after and before the reset; the integration time is the interval between the two reads, and the signal is the voltage difference. This technique has the advantage of removing the KTC noise, but at the cost of \(\sqrt{2}\) in read noise, since two independent readouts are made. A related technique, applicable for very long integrations, is "triple correlated sampling", where the array is also read during the reset intervals; this can remove any very low frequency drift ("1/f" noise) in the electronics, but at a further cost in read noise. Another method is to sample the voltage many times at equal time intervals during the integration; one may either fit a linear slope to the results or average the adjacent differences. This technique is not widely used, but it does reduce the effective read noise under low background/signal conditions, where many reads can be taken; one promising application is in spaceborne instrumentation, where frequent particle events produce discontinuous jumps in the pixel charge. The adjacent differencing technique can identify and remove such events. A variant of the double correlated sampling technique involves reading the array many \((N)\) times at the beginning and end of each integration (Fowler and Gatley 1990). This results in a substantial reduction (by close to \(\sqrt{N}\)) in the read noise, and is used at a number of observatories around the world.

These sampling techniques presuppose a linear relation between charge and voltage on the pixel. This is certainly not the case if the pixel is allowed to saturate, but even within the "normal" operating limits, there is a slight voltage dependence of the diode capacitance. The effects of this are small if one avoids near-saturation operation; alternatively, one may construct a linearity curve which can be used to correct data on-line.

OBSERVATIONAL TECHNIQUES

From the above description, one can see that there are significant differences in the ways that most IR arrays and visible CCDs operate. For the most part, though, these differences are hidden by the system "microcode" which is not apparent to the user of the instrument. At NOAO, the user level software is written so that most of the observing commands used by the observer are the same for IR and CCD arrays. Although the process of obtaining data is thus virtually identical for the two types of arrays, the observing and reduction techniques differ in some significant ways, as a result of the following differences in the arrays and observing environments:

1. By comparison to visible CCDs, IR arrays generally have higher dark current, which is not linear with integration time. Typical values are a few to \(\sim 100\) e/s-pixel. Thus, when dark frames are required for reduction (more on that later), integration times specific to that purpose must be used.

2. IR arrays have higher read noise (defined as the noise in a short, dark integration, where photon shot noise is negligible) than CCDs. There are a number of reasons for this, notably higher capacitance (especially with InSb) and crosstalk in the readout circuit. Multiple readout techniques (Fowler and Gatley 1990) have been successful in reducing the read noise in the SBRC InSb array from 300 to 80 e; values as low as 15 e have been obtained with the Hughes PtSi array.
3. The radiation background from the sky and observing environment is much higher in the IR than in the visible. This fact, more than any other, is responsible for the differences in observing techniques in the two wavelength regimes, and will be discussed in detail below. For most observing applications, photon shot noise from the background is much greater than the read noise of the array, making the latter a less critical parameter.

4. At present, IR arrays are smaller in format than CCDs, and therefore cover a much smaller area of the sky in imaging applications. Coverage of large spatial areas requires "mosaicking" of a number of individual images.

![Graph showing the KPNO sky brightness as a function of wavelength from 0.4 to 4 μm.](image)

**Fig. 7.** Plot of the KPNO sky brightness as a function of wavelength from 0.4 to 4 μm.

Figure 7 is a plot of the measured sky brightness at KPNO from 0.4 to 4 μm, in units of photons-s⁻¹-cm⁻²-μm⁻¹-arcsec⁻². The visual points represent measurements by Pilachowski et al. (1989; see also Massey et al. 1990) in the B, V, and I filters; the data from 1 - 4 μm were obtained by the author with the cryogenic spectrometer utilizing a 58X62 InSb array. Also plotted is the flux from a 300 K blackbody in the same units. To provide a familiar astronomical benchmark, curves of constant magnitude-arcsec⁻² are also plotted. The background arises from three
separate factors. Short of about 2.3 \( \mu \text{m} \) most of the sky emission arises from molecules (primarily OH and O\(_2\)) excited by solar radiation during the day. In spectroscopy, this "airglow" results in a rich spectrum of lines superposed on that of the astronomical source. Because of its atmospheric origin, the airglow background is often temporally and spatially variable, especially during twilight. Longward of 2.3\( \mu \text{m} \), thermal emission from the telescope optics and obscurations, such as the spider and central obscuration, produces a background which rises very rapidly with increasing wavelength; the emissivity (ratio of observed background to that from a true blackbody) is typically 0.1 - 0.15 on a telescope specifically engineered for IR observing. Finally, thermal emission from optically thick telluric absorption lines contribute background in this thermal region. In spectroscopy, this means that absorption lines in an astronomical spectrum correspond to emission lines in the sky background. Although apparently challenging, the same reduction techniques used with CCDs, slightly modified, work quite well with IR data.

The most significant factor in the sky background is its value in relation to the signal from astronomical sources. In the V band, the sky brightness is roughly +22 mag-arcsec\(^{-2}\), comparable to (or a factor of 10 or so brighter) than the limiting magnitude of many observations. However, in the K band, the sky is typically +12 mag-arcsec\(^{-2}\); at 5 \( \mu \text{m} \), about 0 mag-arcsec\(^{-2}\). For those hardy souls observing at 10 \( \mu \text{m} \), the sky is roughly equal to the brightest astronomical source known! The predominant consideration for IR observing is the extraction of information which is a very small fraction (as small as \( 10^{-4} \)) of the sky level. Therein lies the primary observational difference between the visible and IR.

For illustration, consider an imaging observation of integration time \( t_0 \); the collected charge on a pixel \((i,j)\) can be written:

\[
N_{\text{obs}}(i,j) = N_d(i,j,t_0) + N_s F(i,j) + N_\alpha(i,j) F(i,j)
\]

where \( N_d(i,j,t_0) = \text{dark current} \)

\( N_s = \text{sky signal} \)

\( N_\alpha(i,j) = \text{astronomical signal} \)

Note that the sky signal is assumed to be intrinsically flat, but it and the astronomical signal are modulated by the function \( F(i,j) \), which incorporates pixel-to-pixel variations in quantum efficiency and electrical gain, as well as instrumental and telescope illumination functions. This function is generally referred to as the "flatfield", or "flat", since it is generally obtained by observation of a presumably uniformly illuminated source. More on this later.

If one can determine the dark current and the function \( F(i,j) \), the desired result \( N_\alpha(i,j) \) can be obtained:

\[
\frac{N_{\text{obs}}(i,j) - N_d(i,j,t_0)}{F(i,j)} = N_\alpha(i,j) + N_s
\]
The astronomical image is thus superposed on a constant sky level, just as it really is. The sky level can be either determined and subtracted from the image, or extracted using an IRAF routine such as APHOT.

A potential problem lies with the accuracy in determining the function F(i,j). Depending on the method and temporal correlation with the observations, small changes in the illumination function with telescope position or instabilities in the array itself can result in a flatfield which does not give a precisely flat sky in the reduced image. Variations on the order of 1% are not uncommon in flatfield images obtained at various times during a night. For most visible observations, the resultant uncertainty in the sky is near the sensitivity limit of the detector. In the IR, however, uncertainties of this size in the reduced sky can be larger than the signals which one is attempting to measure; this is particularly serious in measuring extended, low-surface brightness objects such as galaxies.

Let us move off the object field to a patch of blank sky and make another observation at the same integration time. The collected charge on each pixel in this case is:

\[ N_{\text{sky}}(i,j) = N_{\text{d}}(i,j,t_o) + N'_s \cdot F(i,j) \]

where the sky background \(N'_s\) may differ slightly from that of the object observation, which was taken at a different time. Subtracting the sky image from the object field image and solving for \(N_o(i,j)\),

\[ \frac{N_{\text{obs}}(i,j) - N_{\text{sky}}(i,j)}{F(i,j)} = N_o(i,j) + (N_s - N'_s) \]

Two points deserve notice. First, there is no need for an explicit dark observation, as the dark current is contained implicitly in the sky image. Secondly, the flat function F(i,j) is divided into the residual sky difference \((N_s - N'_s)\), which is almost always much smaller than the sky level itself. Therefore, any errors introduced by flatfielding the sky will be reduced by that ratio, which on a stable night may be .01 or less.

**SKYS AND FLATS**

A number of questions present themselves at this point. In the descriptive coverage given above, the sky and flat functions were treated rather cavalierly as well-determined observables. In fact, a truly blank sky field in proximity to the object is a virtual impossibility, particularly near the galactic plane. In addition, the process of observing a dedicated sky field and subtracting that image from an object field image extracts penalties in both observing time efficiency and noise. Finally, how does one extract the flatfield function F(i,j), which depends on a number of factors, some of which may change with time or telescope position.
Sky -- Pointlike Sources

Many observations are of regions containing pointlike or small objects, such as stars, distant galaxies, or QSOs. In imaging observations, most of the array is, in fact, measuring sky, at no extra cost to the observer. If one makes several exposures of the region of interest, moving the telescope a small amount between each, then each pixel on the array will be measuring the sky in many (probably most) of the exposures. If all of these images are averaged using a median, rather than a mean algorithm (i.e., the value assigned to a pixel in the average image is the median of the values of that pixel in the input images), the result will be an image of the sky. The field must be sufficiently sparse that a given pixel has flux from an astronomical source falling on it in fewer than 50% of the images. Note that transient events such as cosmic rays are also eliminated by this process. The resultant image will be referred to here as a local sky, since it is obtained from the same observations used for the source region. Figure 8 illustrates the median averaging of 5 images to yield a sky frame.

This technique of sky determination has a number of positive qualities. Because the sky is calculated from the object observations, good temporal and spatial correlation is achieved, at no additional cost in observing time. As mentioned above, the detector dark current is implicitly included in the sky frame. Finally, since the sky frame is an average of a number of observations, its noise will be reduced from that of a single observation, and the noise factor introduced by sky subtraction is minimized.

It may seem very inefficient to be required to make several observations of a field to obtain the sky level, particularly from the viewpoint of observers who are accustomed to taking long (> 30 min) exposures with CCDs in the visible. However, the very conditions that make the determination of the sky important in the IR, namely the high sky brightness, will also limit the possible integration time. Broadband imaging in the K (2.2 μm) band, for example, is typically limited to 1 - 2 min integration times. Multiple exposures will thus be required in almost all cases to obtain the total observation time necessary for detection of faint sources.

Sky -- Extended Sources

When observing an object which is significantly extended on the scale of the array or extremely rich in stars, the above technique cannot be used, since small telescope motions are usually unable to ensure that each pixel is observing a true "sky" level in a majority of the exposures. In this case, there is no alternative to moving the telescope off of the source region to obtain dedicated sky observations. To improve the temporal correlation, most observers prefer to intersperse sky and object field observations (say, in the sequence "object, sky, object, sky, object..."). As with the local sky, it is necessary to make each sky observation at a slightly different telescope position, so that the median averaging technique will remove any stars in the field. Even though the object observations are not used to compute the sky (which will be referred to as a remote sky), it is still good practice to move the telescope slightly between each observation, since this is useful in removing array systematics (dead or flaky pixels) from the reduced image.
Fig. 8. Raw images of Cyg X-3 at 2.2µm obtained at the KPNO 4-m. Integration time was 20 s, at a plate scale of 0.4 arcsec/pixel. The telescope was moved about 10 arcsec between frames. The image at lower right is the sky frame obtained by median averaging of the other five images.
The obvious penalty is the requirement to spend a significant fraction of time observing "empty" sky. (Of course, a really high-class observer will be prepared with QSO fields next to large galaxies, so that observations of the former can be used as sky for the latter). Unfortunately, there is no easy way around this requirement, since it is precisely the extended, low-surface brightness objects for which accurate determination of the sky is so important. Those who are tempted to short cuts are advised by an experienced observational astronomer (M. J. Rieke, private comm.):

"And if you try to cheat, and you don't take the proper number of sky frames, then you get what you deserve."

Spectroscopy

Although the procedures above were covered in the context of imaging observations, the techniques work equally well for spectroscopy. In longslit spectroscopy, the sky is inherently flat only along the spatial axis, while the spectral axis displays the rich structure that makes IR observations so exciting. As with imaging, the proper steps are sky subtraction, to remove as much of the sky emission (and dark current) as possible, followed by flatfield division to restore the inherent uniformity of the residual sky in the spatial axis. In addition, flatfielding also allows correlation of spectra taken at different spatial positions along the slit.

The techniques for sky determination are the same as for imaging, except the telescope offsets between observations must be along the slit, so that multiple observations of the object at different positions on the slit will be obtained. Median averaging of these images will result in a sky spectrum (Fig. 9). Again, for an extended object, dedicated sky observations must be made; sampling several spatial positions off-object will ensure that no "invisible" star spectrum unexpectedly appears in the sky frame! As with imaging, moving the (extended) object along the slit will minimize the effects of bad pixels on the reduced spectrum.

Flats -- Local

In theory, the only contributions to a sky frame are the sky and the dark current. Subtraction of a dark frame of the same integration time used for the sky will yield a result proportional to the function F(i,j). This is true, of course, only for imaging observations. Furthermore, since the flatfield is divided into the sky-subtracted image, its signal-to-noise must be as high as possible. Imaging observations through broadband (> 0.1 m) filters or at wavelengths > 3 m will almost always be limited in integration time by the background, so the resultant sky frame will be the average of a significant number of photon shot noise limited frames. In addition, the dark current will be negligible, except for scattered noisy pixels, and subtraction of several averaged "dark" observations will not increase the noise significantly. If desired, the result can be divided by the mean value of the pixels to yield a flatfield with an average gain of unity.

Obtaining the flatfield function in this way (referred to as a local flat) demonstrates the power of the median averaging of multiple observations. For broadband imaging of pointlike sources, the source observations can be also used to generate both the sky and flatfield frames necessary for reduction. This is not only very time efficient, but results in excellent quality in the reduced images.
Fig. 9. Raw spectroscopic images of Nova Her 1991 obtained at the KPNO 1.3-m on 31 May 1991 UT. Spectral coverage approximately 1.42 - 1.77 μm, at a resolution of 0.0058μm/pixel. Integration time 50 s. Note the intense emission from OH airglow. The object was moved along the slit approximately 15 arcsec between frames. The image at lower right is the sky frame obtained by median averaging of the other five images.
Flats -- Dome

In narrowband imaging, where the signal-to-noise in an averaged sky frame may be marginal, or in spectroscopy, where the sky is not flat in the spectral dimension, it is necessary to determine the flatfield function by dedicated observation of a presumed uniform source. Most commonly, this is a screen (often referred to as the Great White Spot) mounted on the inside of the dome, which is illuminated by lamps on the top ring of the telescope. Because of the brightness of the lamps, one may average a large number of images of short integration time (typically a few seconds) to obtain high signal-to-noise. A number of dark frames at the same integration times as the flat observations (separate for each filter used) are subtracted to yield the flatfield.

DATA REDUCTION

The process of reducing the data obtained through the techniques described above is relatively straightforward with an image processing package such as IRAF, although the large number of images (relative to visual CCD observations) can make the process tedious. Experienced IRAF pundits can create "scripts" to minimize the tedium of repetitive tasks.

Figure 10 illustrates the procedure in "flow chart" format. Multiple observations of the object field, with small telescope motions between each ("jittering"), are taken. The integration time should be sufficiently long to achieve background limited operation, while avoiding saturation by bright sources or the background. In the case of extended sources, dedicated sky observations are interspersed with the object observations. Median averaging of the object or sky images then yields a local or remote sky frame, which is subtracted from each of the object images to yield sky-subtracted images. The appropriate flatfield (local for high background imaging, dome flat for low background imaging or spectroscopy) is then divided into each of the sky-subtracted images, to give sky-subtracted, flattened images.

The reduction to this point is demonstrated in Figure 11, which compares the reduction of a single 2.2 μm frame of Cyg X-3 using the traditional "dark subtract, flat divide, constant subtraction" and the "sky subtract, flat divide" technique used in the IR. Because Cyg X-3 is relatively bright (K ~ 12) and local sky and flat generation were employed, the two techniques yield surprisingly similar results. The difference is more evident in Fig. 12, which compares the two reduction techniques in the case of spectroscopy. The object in this case is Nova Her 1991 (H ~ 9 on 31 May 1991) over the 1.42 - 1.77 μm range, where the OH emission from the sky is very evident. In one technique, the raw frame (12a) is dark subtracted and divided by the dome flat (12c) to yield a frame where the sky spectrum is ideally uniform in the column direction (12e). Using the IRAF "background" package, selected spatial regions off the object are used to generate a sky spectrum for the entire image, which is then subtracted (12g). In the other, the raw frame (12a) and generated sky frame (12b; c.f., Fig. 9) are subtracted (12d), flattened (12f), and background subtracted (12h). Note that the sky subtraction removes virtually all of the sky, so the flatfielding and background removal seem almost superfluous (but keep in mind that pixel-to-pixel gain variations along the spectrum are removed by the flattening). Comparison of the final results (12g and 12h) clearly shows the advantages of the
Fig. 10. Representation of image reduction, illustrating the use of local or remote sky and local or dome flat techniques, as described in the text. Operations in parentheses (MEDIAN, -, /) represent median averaging, subtraction, or division.
Fig. 11. Comparison of reduction techniques for a single frame of the Cyg X-3 observations shown in Fig. 8. On the left, the raw frame (a) is dark subtracted and divided by a local flat to produce (c); a constant representing the sky level is subtracted to yield (e). On the right, the sky frame obtained in Fig. 8 is subtracted from the raw frame (b) to yield the sky-subtracted frame (d), which is then divided by the local flat to yield (f).
Fig. 12. Comparison of reduction techniques for a single frame of the Nova Her 1991 observations shown in Fig. 9. On the left, the raw frame (a) is dark subtracted and then divided by a dome flat (c) to yield (e). Note that the column gradient evident in both the raw and flat is substantially eliminated in the result. Removal of the sky background by the IRAF "background" task results in (g). On the right, the sky frame (b) obtained in Fig. 9 is subtracted from the raw frame (a) to yield the sky-subtracted frame (d). Note that the sky is virtually eliminated in this single step. Flat-division and removal of the residual sky background yield (f) and (h), respectively.
sky-subtraction technique. Not unexpectedly, the major source of error is the dome flat, which is unable to duplicate exactly the array illumination during the astronomical observation.

A point worth mentioning is that a spectroscopic dome flat will also introduce the spectrum of the light source into the result. This is of no consequence, as long as calibration spectra of standard sources are also reduced in the same manner.

The final averaging of the sky-subtracted, flattened images is illustrated in Figure 13. The images are copied into larger format images, and shifted so that the new images are registered with respect to the stars; note that this results in systematic array features (bad pixels, etc.) now appearing in different positions in the shifted images. The shifted images are now median averaged to yield the final result (Fig. 14). This process not only averages the statistical noise consistent with the total integration time on the source, but removes fixed defects and transient events such as cosmic rays.

Treatment of spectroscopic images is similar. Spectra of extended objects can be copied from the reduced spectral frames into new images, with the spectra shifted spatially until they are coaligned. In the case of point objects, the IRAF "apextract" and "apusum" routines can extract one-dimensional spectra for averaging. As with images, median averaging will remove bad pixels from the result, but one must be careful in the case of spectroscopy, since variations in seeing and guiding will generally result in different amplitudes for the input spectra, and a straight median average will be ineffective. One must employ either a weighted median average or a "clipped average" in which values deviating from the mean by more than a specified amount are rejected. As with all of these techniques described here, the method which works best may vary with the type and quality of the data, and a cookbook approach should be regarded as a guideline. Experience is, as always, the best teacher.

MOSAICKING
Because of the relatively small size of available IR arrays, a single frame is often unable cover the desired field of view. This is particularly relevant from the point of serendipity, since there has never existed any previous panoramic detector in the IR, and exciting discoveries can (and are) made by simply imaging suspected "interesting" regions. In either case, the technique of combining a number of images into a spatial mosaic is common, and IRAF (and probably other image analysis packages) now contains routines specifically for this process.

The observational and reduction techniques are straightforward. Individual images are taken, with intervening telescope motions somewhat smaller than the field of view, so that there is sufficient spatial overlap for later registration. Taking two or three images at each "position", or significantly increasing the spatial overlap of adjacent images can prevent loss of information due to array defects. The individual images are then reduced as described above, using local or remote sky and flatfield as the observing conditions dictate. The mosaicking occurs in two steps. The images are first copied into a single large image, registered with respect to each other as on the sky. Stars common to spatially adjacent images are then identified.
Fig. 13. Six images of Cyg X-3, reduced as in Fig. 11, copied into larger images and shifted to co-register the stars in all images. Note that the array defects now appear in different locations in the new images.
and marked, with the coordinates being written to a database, which is then used to shift adjacent images so the common regions overlap. The routine also determines the mean sky level in each overlap region and applies a small offset to match the sky levels; this corrects for small variations in the background over the time period of the observations. An example of such a mosaic is the 2.2μm image of the galactic center (Fig. 2), constructed from 25 individual 180 s observations, and covering a field 27 arcmin on a side.

Fig. 14 Median average of the six images shown in Fig. 13. Note the substantial reduction in sky noise resulting from the averaging, as well as the removal of the array defects.

SUMMARY

Although the IR arrays described in this paper may seem small in comparison to the visible CCDs available today, they are continuing to evolve in both size and performance. The speed of this progress may be judged by a comparison of results presented at two recent conferences on IR arrays (Wynn-Williams and Becklin 1987; Elston 1991). The progress up the learning curve has undoubtedly been facilitated by the experience gained in the optimization of visible CCDs for astronomy.

There is clearly not space here to detail the fields of research opened by the advent of IR arrays; the conference proceedings referred to above are a good starting point for interested readers. It is significant, though, that the IR arrays presently in use have reached the requisite sensitivity for imaging and spectroscopic observations of many "interesting" extragalactic visual objects. They are thus used not only to observe sources totally inaccessible in the visual (star-forming regions, stars with
dense circumstellar shells, etc.), but to complement visual observations. These complementary visual/IR observations not only provide the advantages of extended spectral coverage, but also the ability to probe different regions within an object as a result of the low extinction in the IR. Examples include high-redshift galaxies and QSOs, in which rest-frame transitions of Lyα, MgII, and CIV are observed in the visible and star-formation and metallicity indicators such as Hα, Hβ, OIII[3727], and OIII[5007] may be observed in the IR. Imaging of active or irregular galaxies in the near-IR may reveal underlying morphology such as bar or ring structures hidden by the luminous blue stars, HII regions, and dust which dominate the visual appearance (Gatley, DePoy, and Fowler 1988).

While IR arrays in 2048 X 2048 format may be a long time coming, present arrays are sufficiently large (256 X 256, with 640 X 488 anticipated) that large-scale instrumentation similar to that used in the visible is being developed and implemented at a number of observatories. NOAO has recently put into service a wide-field (6 arcmin) camera utilizing four 256 X 256 arrays to provide simultaneous imaging in the J, H, K, and L bands. A number of infrared echelle spectrographs with resolving power λ/Δλ ~ 10^4 - 10^5 are under construction or going into service. The next generation of instruments may utilize several arrays in a line to achieve extended spectral coverage, or (at least) a 2 X 2 array mosaic in a wide-field camera. While the higher sky background will probably never permit ground-based IR arrays to achieve the limiting performance of CCDs in the visible, space-borne IR instrumentation such as the proposed HST NICMOS and that on the cooled SIRTF can reap the benefits of a significantly reduced background flux.

REFERENCES


DISCUSSION

M. Sitko: Are these arrays big enough to take advantage of the 2D format for doing echelle spectroscopy?

D. Joyce: Since the echelle grating operates in lower orders in the IR, the free spectral range is larger in wavelength units. In an echelle spectrometer with \( R = 50000 \), a 256x256 array would cover only \( \sim 7\% \) of the free spectral range. Cross-dispersion, while possible, is thus not scientifically practical.

M. Sitko: Can these arrays be butted together?

D. Joyce: In most, if not all, of the current arrays, the readout extends beyond the detector array for wirebond connections to the chip carrier. A readout architecture with the connection pads confined to two opposite sides would permit a linear mosaic of two or more detector arrays.

M. Magalhaes: Could you comment on the sensitivity of the infrared arrays you described to polarized light?

D. Joyce: I know of no such sensitivity. Single element In Sb detectors have been used extensively for polarimetry with a linear polarizer analyzer without demonstrating intrinsic dependence on the polarizer orientation. The half-wave plate/linear polarizer combination which we have used with In Sb arrays would hide any such dependence, but there is no reason to suspect it, since the detector material is the same. We have no experience in polarimetry with other detector materials.

J. Roth: How do you get an IR "dark" frame when objects at room temperature are not "dark," especially at longer wavelengths?

D. Joyce: The dark slide is inside the dewar, cooled to same temperature as the filters, around 77K.