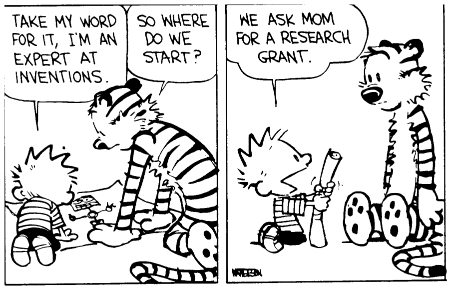
2. Infrared Detector Technology



Outline

This chapter explains the physical and technological background of the near infrared detectors used in astronomy. It discusses the advantages of array detectors over single element devices and examines the differences in operation between the new generation of infrared arrays and visible CCD imagers.

The design and operation of infrared arrays in general, and the NICMOS device used at COAST in particular, are considered along with the parameters for evaluating infrared array operation. The detailed design and performance of the COAST camera is described in the next chapter.

2.1 Detecting photons

This thesis is concerned partly with a camera designed to detect light in the near infrared, particularly the spectral region with wavelengths of $\lambda = 1 - 2.5\ \mu m$. At these wavelengths most instruments and detectors respond to the arriving radiation as photons in the same way as visible detectors. At longer wavelengths, $\lambda > 20\mu m$, the low energy photons cannot be detected directly. Instead the physical effects of the photons on some material are used, typically the photons heat a material or change its resistance.

In the infrared array we are concerned with, the basic form of a photon detector is a semiconductor. Here, an incoming photon causes an electron to be excited out of a bound state so that it can move freely in the conduction band. In semiconductor physics the net positive charge left behind when an electron is removed from an atom is referred to as a hole and is treated as if it were a virtual positively charged electron. The energy needed to remove an electron is called the bandgap energy and depends on the composition of the semiconductor.

The longest wavelength to which the detector is sensitive is the cut-off wavelength ($\lambda_{cut-off}$). This is related to the band gap energy by a simple relationship.

$$\lambda_{cut-off} = \frac{hc}{E_{bandgap}} \qquad \text{or, in common units} \qquad \lambda(\mu m) = \frac{1.24}{E_{bandgap}(eV)}$$

The bandgap energy is temperature dependent, generally at lower temperatures the bandgap increases and so the cut-off wavelength decreases.
The table below gives the bandgap energies, cut-off wavelengths and typical operating temperatures for a number of common detector materials, adapted from Joyce(1992).

Material	$E_{bandgap}$ (eV)	$\lambda_{cut-off}$ (μm)	Temperature (K)
Silicon (Si)	1.18	1.05	150 - 300
Indium Gallium Arsenide (InGaAs)	0.7	1.7	77 - 200
Platinum Silicide (PtSi)	0.25	5.0	40-60
Indium Antimonide (InSb)	0.23	5.4	20-40
Mercury Cadmium Telluride (HgCdTe)	0.25 - 0.5	2.4 - 4.8	60-77

Figure 2.1: Infrared detector materials

Silicon is used in almost all devices operating at visible wavelengths. The last three materials are the most commonly used in infrared array detectors. The composition of HgCdTe can be tuned to adjust the bandgap and so choose the wavelength sensitivity. Note that Mercury Cadmium Telluride is often referred to as MCT or CMT.

2.2 Measuring the Signal

Once an electron-hole pair has been created by an incoming photon, it must be detected and recorded. An electric field in the region where the photon was absorbed separates the oppositely charged electron and hole and prevents them from immediately recombining. The electron must be stored and measured to detect the photon. The main schemes for detecting the photon generated current are the photovoltaic and photoconductive effects.

2.2.1 Photoconductive effect

Photoconductive devices place an external electric field across the semiconductor. Electrons and holes are dragged in opposite directions by this field and the resulting current is sensed by a change in voltage across an external resistor. This scheme suffers from the drawback that any changes in the intrinsic resistance of the semiconductor or the resistor are indistinguishable from the photon signal. This makes the devices very sensitive to slight temperature changes or fluctuations in the bias voltage producing the electric field. In general photoconductive detectors are only used at longer wavelengths where photovoltaic devices are not possible.

2.2.2 Photovoltaic effect

Most imaging devices in the visible and near infrared use the photovoltaic effect. A detector is produced in the semiconductor by implanting ions of another material. These impurities create a diode junction which is biased to produce an electric field. An electron-hole pair produced by the absorbed photons is separated by the field and the electrons are accumulate on the diode. At the end of the integration the voltage across the diode is measured. If the capacitance of the junction is known, this can be used to determine the number of electrons detected.

2.3 Array detectors

2.3.1 Array advantages

Until the mid 1980s, most infrared cameras used single element detectors. Images were built up by scanning the telescope over a region of the sky, Gillett(1987). Today most infrared cameras use arrays operating in the same way as visible systems. These cameras have many individual detectors observing the scene simultaneously to produce an image of the object. This has a number of advantages over single element detectors which are described below.

2.3.1.1 Observing efficiency

The most obvious advantage of an array is observing efficiency. A system with N pixels will observe a region to the same signal to noise in only 1/N the time taken by a single element. In reality the advantage is even greater if the time taken to move the telescope in scanning system is considered. As a simple example, a 1 hour observation with a NICMOS array of 256x256 pixels would take 15 years with a single element detector !

2.3.1.2 Image Resolution

Although in theory a scanning system can achieve any spatial resolution simply by moving the telescope in smaller increments, there comes a point where drifts in telescope pointing and mechanical stability limit the accuracy of the image. An image produced by an array has the relative positions of the pixels fixed with great stability and so object separations in an image can be measured to a high precision. The resolution of arrays can be further increased by micro-scanning. This technique involves moving the array by a fraction of a pixel between exposures and interpolating the result to produce a single higher resolution image. In a similar way, a larger area can be imaged by taking exposures of a number of overlapping fields. The individual frames are combined together in a mosaic. This is especially useful with the relatively small sizes of current infrared arrays when compared to visible detectors.

2.3.1.3 Background Subtraction

The sky is brighter at infrared than visible wavelengths. In addition, the sky brightness changes on timescales of minutes to hours largely due to absorption from OH molecules, Rieke(1993a). In order to image the faintest infrared objects it is usually necessary to subtract the sky background from the raw image of the source. A single element detector takes a considerable time to scan a region of the sky during which the background may change. With an array camera all the pixels receive the same total background signal and so the sky contribution can be subtracted with much greater accuracy.

2.3.1.4 Disadvantages

There are a number of disadvantages to array based detectors. In an array based camera the complexity of the camera electronics and the associated data analysis software is much greater than for a single element detector, although they are simpler than systems which have been developed for visible CCD cameras. These usually have a much larger number of pixels. In addition the limited market for infrared arrays and the fact that this is often a military market means that they are much more expensive than visible CCDs.

One often unrealised advantage of single element detectors is useful if the source signal level is very low. It is possible to integrate a larger part of the scan on the sky as each measurement. In this way it is possible to trade resolution for signal to noise. With an array camera, the signal in each pixel must reach the limiting signal to noise. A similar scheme is commonly used with CCD based cameras where the nature of the readout allows the signal in adjacent pixels to be combined before being read.

2.3.2 Array design

Before describing the details of infrared array architecture I will define some of the terms and explain some acronyms. The term *Focal Plane Array* (FPA) is used to refer to systems with a large number of pixels simultaneously integrating a signal from an image projected onto the array. *Charge Coupled Devices* (CCDs) are a particular kind of focal plane array made from a single structure where surface electrodes are used to couple the pixels together so that the charge can be transferred across the surface of the array to the output. These devices are so commonly used in the visible region that the term CCD is often incorrectly used to refer to any focal plane imager. The charge transfer mechanism of a CCD is independent of the detection of the photons and infrared arrays can use CCDs as a method of transferring charge. The term *Direct Read-Out* array (DRO) is used for infrared devices constructed from a large number of individual pixels, all connected to a separate multiplexor and readout circuit. The most common infrared arrays used in astronomy are of this type.

I shall assume that the reader is familiar with the basic principles of Charge Coupled Devices (CCDs) used in visible astronomy cameras. A detailed review of their use in astronomy is contained in Mackay(1986). The structure, performance and operation of infrared arrays will be compared with CCD cameras. Although based on the Rockwell NICMOS array, this discussion is in general enough terms to apply to most infrared arrays.

2.3.2.1 Charge Coupled Devices

Optical astronomers benefit from a series of lucky coincidences. Silicon is an ideal material for manufacturing semiconductor devices since it has an insulating oxide form and can be doped both positive and negative without disturbing the crystal structure. It is also sensitive to light across the whole visible region and in the infrared out to the atmospheric cut-off at $\lambda = 1\ \mu m$. Most modern electronic visible cameras use silicon based charge coupled devices as the detector. These are built from a single piece of silicon using normal integrated circuit fabrication techniques and so are relatively cheap and easy to manufacture. Since almost identical devices are used in a wide range of domestic video cameras there is a huge research and development investment in these components.

The operation of a CCD is quite simple. A photon absorbed by the detector creates an electron-hole pair in the silicon semiconductor. A series of electrodes laid across the surface are biased to create an electric field which gathers and stores the electrons. The holes diffuse back into the bulk of the silicon and recombine. The pattern of electric fields produced by these electrodes defines the pixels. To read out the image, the potential on adjacent electrodes is changed to adjust the field and drag the charge across the detector. A single amplifier in one corner of the device is used to measure the charge from each pixel in turn. The simple structure of the CCD means that large format devices of good quality are easy to produce. The process of integrating the signal charge directly in the potential well is highly linear and the simple signal path gives low noise.

An obvious form of infrared camera would be to construct a CCD from some material other than silicon in order to use a different wavelength band. Unfortunately, the intrinsic advantages of silicon and the huge effort that has gone into developing silicon integrated circuits for other applications is unlikely to be duplicated for any material suitable as an infrared detector. Little success has been achieved in producing CCDs in infrared materials. Although some trial devices have been made from InSb and HgCdTe, these are not suitable for demanding astronomical applications, Roberts(1980).

2.3.2.2 Hybrid Arrays

Another solution to the problem of an infrared array is to build the detector from two different materials. An infrared sensitive semiconductor could be used to detect arriving photons and a silicon based device used for the readout circuitry. The resulting device is known as a *hybrid* array, Bailey(1987).

Early hybrid devices used a conventional CCD as the readout circuit; this has all the CCD advantages of low noise, high storage capacity and linear response. Unfortunately, in order to inject the charge from the infrared detector into the CCD, a type must be used where the charge storage region is directly accessible. These *surface-channel* devices are much noisier than the *buried-channel* CCDs used in astronomical cameras.

An alternative scheme uses a Direct Read-Out (DRO) array. This has the infrared detector layer bonded to a silicon integrated circuit containing an array of pixels each with their own individual storage site and amplifier. A series of switches in the silicon called a multiplexor provides a signal path from each pixel to the output. These devices have a number of interesting features which are discussed in section 2.3.4 .

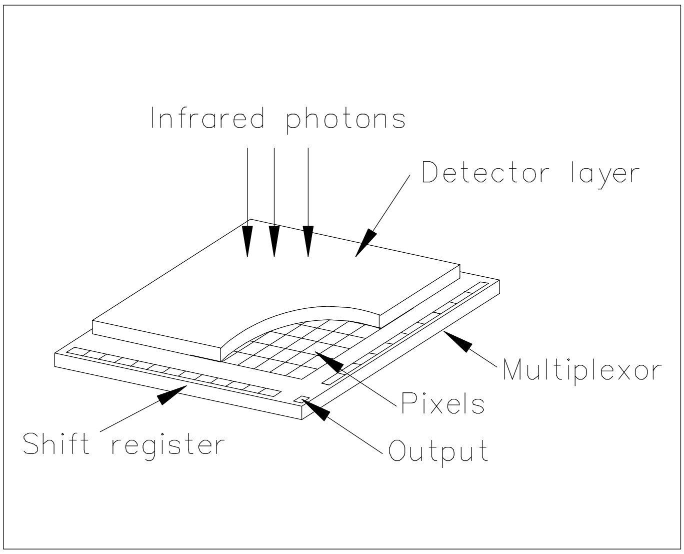


Figure 2.2: Structure of a hybrid array

A major difficulty in producing hybrid arrays is that the silicon readout and the detector material have different material properties. Since the devices are manufactured at room temperature but operated at 20 - 80 K, any difference in thermal expansion between the two layers can cause the device to break apart. This effect limits the maximum size of the array that can be produced. The new generation of 1024x1024 pixel arrays are only possible by reducing the physical size of each pixel. Unfortunately the reduced space available for the circuitry in each pixel reduces the well capacity, and the smaller spacing increases cross-talk, Hodapp(1994). Alternative techniques such as growing the infrared detectors directly on silicon substrates will be needed to reliably produce larger arrays, Irvine(1992).

2.3.3 Structure of Hybrid Arrays

This section discusses the detailed structure and manufacture of hybrid infrared arrays. The emphasis is on the Rockwell NICMOS device but the principles should apply to many similar devices.

2.3.3.1 Detector layer

The incoming photon is absorbed in the detector material and so generates an electron. To detect this electron it must be produced near to the connection to the readout circuit. This means that the arrays are backside illuminated in the sense that the photons must travel through the thickness of the detector material before being absorbed. To increase the chance of a photon reaching a region where it can be detected, the detector material must be in a thin layer. In the NICMOS device a layer of detector material is grown epitaxially on an inert substrate made from single crystal synthetic sapphire. In the case of InSb arrays, a wafer of bulk InSb is thinned by etching and polishing.

Diodes are implanted in the detector material to produce the local field where the photon will be absorbed. Narrow insulating regions are used to form barriers between these diodes to define the pixels. These barriers however produce a dead space between pixels where any arriving photon will be lost.

2.3.3.2 Interconnecting Bonds

The connection between the detector and the readout circuit is made by columns of metallic indium deposited on each pixel. When the two parts of the array are pressed together these columns cold-weld and form an electrical connection along which the photon generated current flows. The gap between the readout circuit and detector wafer is then usually filled with epoxy to improve the mechanical strength of the device. Failures of the indium bonds are common and cause isolated dead pixels in the array.

2.3.3.3 Pixel unit cell

In a hybrid array each pixel is a separate detector and has its own charge storage site and readout circuitry. This group of components which is repeated for every pixel in the array is called the unit cell. Electrons generated in the infrared detector layer travel down the indium bonds and are integrated on a small capacitor in the unit cell. Generally the capacitor is pre-charged to some fixed voltage before an exposure and the arriving photo-electrons decrease the potential. This *Capacitive Discharge Mode* (CDM) prevents the charge from a saturated pixel leaking into its neighbours and smearing the image which is a common problem in CCD based devices.

The signal capacitor is connected to the input of an insulated gate FET. The very high resistance of the CMOS technology means that almost no leakage of the signal charge occurs. Since the signal is not affected by the readout process, this is called a *non-destructive* readout.

2.3.3.4 Multiplexor

The multiplexor is a series of FET switches in the readout circuitry which connects each pixel in turn to the output amplifier. At any time only one pixel is connected to the output. It is not generally possible to randomly access any individual pixel directly, instead the X-Y address of the pixel currently being accessed is set by incrementing values in row and column shift registers. It is possible to quickly skip over pixels without reading the signal, this allows sub-apertures of the array to be read quickly. Some array designs also allow the shift registers to run backwards, making it easier to scan over a small region quickly.

In the NICMOS3 multiplexor only the current pixel can be reset, in other devices a single instruction resets the entire array.

2.3.4 Differences between IR-Arrays and CCDs

In use these infrared arrays can be thought of as analogous to CCDs. The operation of the data capture and reduction systems at the telescope are often identical. However there are a few important differences in the structure which lead to some interesting effects in their operation.

2.3.4.1 Non destructive reads

The process of accessing and reading a pixel does not change the signal integrated in the pixel, this allows the same image to be read a number of times. By combining each of the separate reads it is possible to reduce the effects of readout noise. A study of this suggests that the noise can theoretically be reduced by a factor of $\sqrt{N}$ with the array read N times, Fowler(1991).

2.3.4.2 On chip binning

With CCD cameras it is possible to combine the charge held in adjacent pixels together on the chip before readout. This both increases the readout rate and reduces the effect of readout noise, at the expense of reduced spatial resolution. Since in a direct read-out array the charge stored in each pixel is entirely separate, and can only be accessed through the amplifier in that pixel, it is not possible to do this.

2.3.4.3 Noise

The readout noise of DRO arrays is generally greater than for CCDs. This is due to the more complicated signal path and the extra amplifier for each pixel. Typically astronomical infrared systems produce a noise of 10-100 electrons/pixel rather than the 2-3 electrons possible with modern CCD cameras.

2.3.4.4 Non-linearity

In an ideal detector the output signal is directly proportional to the number of photons integrated and so the device is said to be linear. Over the normal operating range CCDs are highly linear but direct read-out arrays are not. This is due to the capacitive discharge operation. The capacitor starts the integration at a positive voltage which is discharged by the arriving photo-electrons. The voltage on the capacitor decreases as the signal is integrated and so the detector bias is reduced. A consequence of this is that the signal measured from a pixel is not proportional to the true number of photons integrated by that pixel. However, the relationship between the output and received signal is monotonic and can be calibrated. This system also has the useful property that a totally saturated pixel contains no charge and so cannot leak signal charge into the neighbouring sites. The difference between a CCD and DRO is shown by the analogy in Figure . The buckets represent the storage site in the pixel, the tap is the arriving signal and the drips represent the dark current. The model is adapted from McCaughrean(1989).

1, A CCD starts an integration empty and is filled with charge by the photons arriving during the exposure. The dark current acts in the same sense as the arriving signal. During the integration the charge level in the pixel increases linearly with the number of photons arriving. Under very high illumination conditions the pixel can become overfilled. The electric field produced by the surface electrodes cannot hold the charge and some flows into the field in the adjacent pixels. The effect is to cause a smearing of the image.

2, The DRO array begins an exposure with each pixel pre-filled to a fixed level. During the integration the level decrease due to infrared generated photons and dark current. The reduced voltage on the capacitor also reduces the detector bias voltage. Since this reduces the detector's collecting efficiency, the measured signal rate decreases as the pixel fills. This is shown by the sloping sides of the bucket. Eventually the pixel is completely emptied and any more arriving signal has no further effect. There is also no effect of dark current on a totally saturated pixel since there is no longer a bias voltage on the detector. As the pixels are totally isolated from each other, this saturated pixel cannot affect adjoining sites.

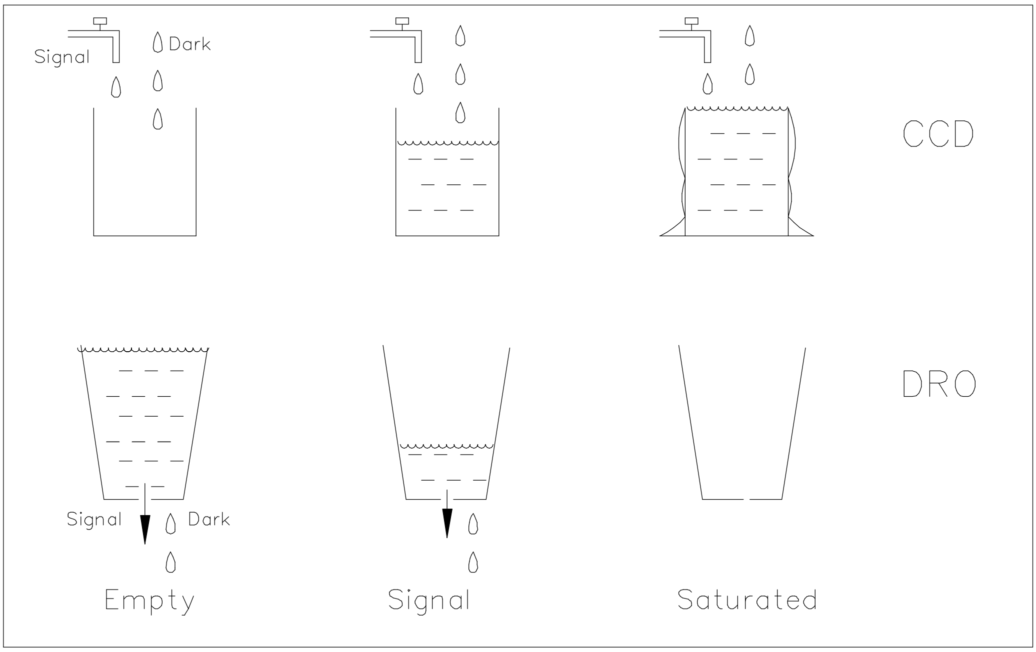


Figure 2.3: Bucket analogy, linearity of CCD and DRO.

2.4 Array Parameters

A number of parameters have been defined for measuring and comparing the performance of visible and infrared photo-electric devices. Most of these are based on the concept of Noise Equivalent Power (NEP). This is the power received from the source which gives a signal to noise ratio of one. It is more commonly used in the form of normalised detectivity (D*), which is the reciprocal of the NEP normalised to unit detector area and unit bandwidth. For definitions of the terms see Ballingall(1990) and Wolfe(1989).

These terms are generally more applicable to defining the performance of devices used as thermal detectors or bulk detector material and are not very useful for studying the performance of the imaging arrays used in astronomy. For array devices it is more useful to define parameters which are more closely linked to the way in which the device is actually operated.

2.4.1 Number of pixels

The advantages of large format arrays over single elements have been discussed in section 2.3.1.1. Currently, visible CCD devices of 2048 pixels square are in general use and formats twice this size are available. However, near infrared hybrid arrays are only a few generations behind with 256 square arrays common and 1024 square becoming available during 1995, Fowler(1994). The overall size of hybrid arrays is currently limited by the problem of thermal mismatch between the different materials used in the detector and the readout circuitry as described in section 2.3.2.2.

2.4.2 Filling factor

The filling factor is the proportion of the area of the detector where the arrival of a photon can lead to a detected signal. In a device containing a number of pixels there is usually a dead-zone between each pixel which defines the extent of the pixel and is insensitive to light. Early devices built up from many discrete components had significant dead space between each detector element. Large format arrays have implanted barriers between pixels which account for 10% of the area of the detector. The size of these barriers is a compromise between the need to prevent cross-talk between adjacent pixels and maximising the filling factor.

Poor filling factor obviously leads to a proportional loss in the efficiency of detecting photons, but there are other effects. In an undersampled image, the gaps between pixels cause a change in shape of the stellar profile which can lead to photometric errors. If an image plane fringe detecting scheme were used for COAST there would be an apparent loss in visibility due to dead space between the detectors.

2.4.3 Quantum efficiency

The probability that a photon which arrives at the detector will be recorded is the quantum efficiency (QE). This factor depends partly on the chance of a photon generating an electron-hole pair in the semiconductor, which depends on the material used and the temperature at which it is operating. It also depends on the probability that the resulting electron will be transferred to the charge storage structure and be detected, this depends on the design of the storage diode junction and the readout circuitry. As mentioned above, hybrid arrays are back-side illuminated so that the photons pass through a base material before reaching the sensitive region. Any reflection or transmission losses in the substrate or the detector material also lead to a loss in efficiency. In the NICMOS design, the substrate is sapphire which although usefully transparent from 1 - 5.5 μm, has a high refractive index. As the surface is not anti-reflection coated, around 10% of the photons are lost from reflections at the front of the device.

2.4.4 Well capacity

The maximum number of electrons each pixel can store is called the well capacity. This, together with the read noise, sets the dynamic range of the device. Each pixel must integrate a number of electrons significantly greater than the square of the read noise. For mid to long wavelength infrared, the well capacity becomes more important since the high background flux means it may be difficult to read out the device quickly enough to prevent the pixel saturating.

2.4.5 Uniformity and Linearity

An ideal detector produces a signal which is linear with the amount of light integrated. Direct readout hybrid arrays which integrate the signal on a capacitor and then read the voltage across the capacitor are inherently non-linear as described in section 2.3.4.4. However the relationship is monotonic and so can be calibrated, Hoffman(1987).

2.4.6 Read noise

Each pixel accumulates the signal on a capacitor during the exposure and at the end of the integration the charge in each pixel is measured. This measurement of the integrated signal is subject to a fixed uncertainty which is independent of the size of the signal. This noise from the system in the absence of any input signal is called the read noise. The read noise together with the source brightness determines the exposure time. In order to be limited by the shot noise on the arriving photons rather than the system noise, a number of photons greater than the square of the read noise must be integrated.

2.4.7 Dark current

Thermal excitation processes in the semiconductor can generate electron-hole pairs in the same way as real photons. This extra signal rate is exponentially dependant on detector temperature and is the main reason that astronomical detectors are operated at cryogenic temperatures. The dark current is also dependant on the wavelength response of the detector. Mid and long wave infrared materials must have a narrower bandgap and so less thermal energy is needed to produce a signal. In addition to the real dark current produced by physical processes in the semiconductor, there is also an extra signal from warm components in the detector field of view, out-of-band filter leakage and luminescence from circuitry on the detector chip. In planning real astronomical observations it is more useful to consider all these processes and define dark current as the signal received from anything other than the sky and astronomical object.

2.5 Conclusion

Infrared astronomy is no longer an art requiring experience of building and operating unstable and fickle detectors. The technology has advanced to the point that cameras are as simple to operate as visible systems and the data reduction process almost as straightforward. This chapter has described the terminology and operation of infrared detectors, the next chapter will study the detailed design and performance of the camera system for COAST.