Chapter 8
Detectors

> Honestly, I cannot congratulate you upon it. Detection is, or ought to be, an exact science, and should be treated in the same cold and unemotional manner. You have attempted to tinge it with romanticism, which produces much the same effect as if you worked a love-story or an elopement into the fifth proposition of Euclid.
>
> "But romance was there," I remonstrated.
>
> – Arthur Conan Doyle, *The Sign of the Four*, 1890

Astronomical detection, even more than the work of Sherlock Holmes, is an exact science. Watson, though, has an equally important point: no astronomer, not even the coldest and most unemotional, is immune to that pleasant, even romantic, thrill that comes when the detector *does* work, and the Universe *does* seem to be speaking.

An astronomical detector receives photons from a source and produces a corresponding ***signal***. The signal characterizes the incoming photons: it may measure their rate of arrival, their energy distribution, or perhaps their wave phase or polarization. Although detecting the signal may be an exact science, its characterization of the source is rarely exact. Photons never pass directly from source to detector without some mediation. They traverse both space and the Earth's atmosphere, and in both places emissions and absorptions may modify the photon stream. A telescope and other elements of the observing system, like correcting lenses, mirrors, filters, optical fibers, and spectrograph gratings, collect and direct the photons, but also alter them. Only in the end does the detector do its work. Figure 8.1 illustrates this two-stage process of signal generation: background, atmosphere, telescope, and instruments first modify light from the source; then a detector detects.

An astronomer must understand both mediation and detection if she is to extract meaning from measurement. This chapter describes only the second step in the measurement process, detection. We first outline the qualities an astronomer will generally find important in any detector. Then we examine a few important detectors in detail: the CCD, a few photo-emissive devices, the infrared array, and the bolometer.

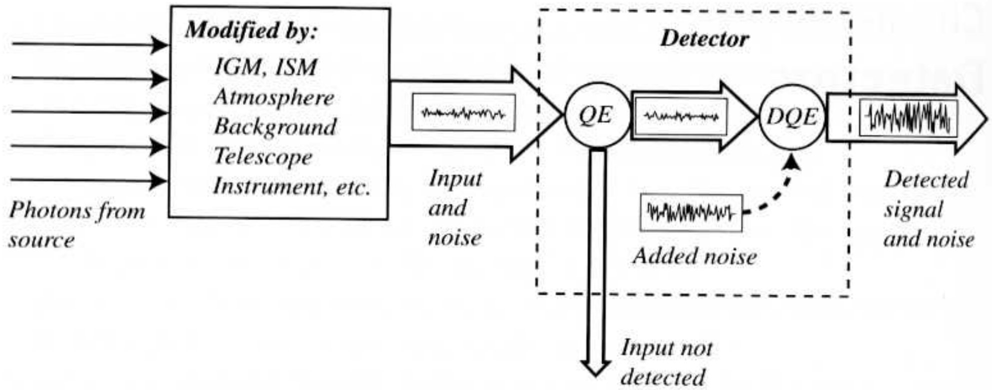

Fig. 8.1 Mediation and detection of a light signal (IGM = intergalactic medium, ISM = interstellar medium). The detection step may fail to record some of the mediated signal, and may introduce additional noise to the part of the signal that is recorded.

8.1 Detector characterization

Why does an astronomer choose one detector instead of another? Why did optical astronomers in the 1980s largely abandon photography, the then-dominant detector for imaging, in favor of solid-state arrays? Why are these same arrays useless for other purposes, such as measuring very rapid changes in brightness? Is there a *perfect* detector? We begin an answer with a list of several critical characteristics of any detector.

8.1.1 Detection modes

We can distinguish three distinct modes for detecting light.

Photon detectors produce a signal that depends on an individual photon altering the quantum-mechanical state of one or more detector electrons. For example, in the last chapter, we saw how a change in electron energy in a photoconductor or photodiode can produce a change in the macroscopic electrical properties like conductivity, voltage, or current. Other changes in quantum state might produce chemical reactions (as in photography) or a pulse of free electrons, as in vacuum photomultipliers. Photon detectors are particularly suited to shorter wavelengths (infrared and shorter), where the energies of individual photons are large compared to the thermal energies of the electrons in the detector.

Thermal detectors absorb the energy of the incoming photon stream and convert it into heat. In these devices the signal is the temperature change in the body of the detector. Although thermal detectors are in principle useful at all wavelengths, in practice, thermal detectors, especially a class called ***bolometers***, have been fundamentally important in the infrared and microwave regions, as well as very useful in the gamma and X-ray regions.

Wave detectors produce signals in response to the oscillating electric or magnetic field of the incoming electromagnetic waves, usually by measuring the interference effect the incoming fields have on a wave produced by a local oscillator. In principle, these detectors, unlike photon and thermal detectors, can gauge the phase, intensity, and polarization of the detected wave. Wave detectors are especially useful in the radio and microwave parts of the spectrum.

8.1.2 Efficiency and yield

Thou shalt not waste photons.

– Anonymous, *c.* 1980

A good detector is efficient. We construct costly telescopes to gather as many photons as possible, and it seems perverse if a detector does not use a large fraction of these expensive photons to construct its signal.

Photography, for example, is relatively inefficient. The photographic detector, the emulsion, consists of a large number of tiny crystals, or ***grains***, of silver halide (usually AgBr) suspended in a transparent gelatin matrix. Photons can interact with a grain to eventually turn the entire grain into elemental silver. The more silver grains present in the emulsion after it has been processed, the stronger is the signal.

Why is the process inefficient? Some photons reflect from the surface of the emulsion and are not detected. Some pass right through the emulsion, while others are absorbed in its inactive parts without contributing to the signal. Nevertheless, silver halide grains absorb something like 40–90% of the incident photons. These absorbed photons produce photoelectrons that can induce a chemical change by reducing a silver ion to a neutral atom. The corresponding neutral bromine atom (the hole produced by photo-absorption) can vanish, either combining with the gelatin or with another bromine to form a molecule that escapes the crystal. Most holes do not vanish, however, and most photoelectrons recombine with holes before they can neutralize a silver ion. Some neutral silver atoms are created, but most are re-ionized by holes before the grain can be developed. Finally, it is only after three to six silver atoms drift and clump together at a spot on the grain that the crystal becomes developable. In the end, very few of the incident photons actually have an effect in photography. The process is inefficient.

The ***quantum efficiency***, ***QE***, is a common measure of detector efficiency. It is usually defined as the fraction of photons incident on the detector that actually contribute to the signal.

$$\mathrm{QE} = \frac{N_{\mathrm{detect}}}{N_{\mathrm{in}}} \tag{8.1}$$

In a perfect detector, every incident photon would be absorbed in a fashion that contributed equally to the signal, and the detector would have a QE of 100%.

Photographic emulsions have QE valuess in the range 0.5–5%.[1] Solid-state devices – like silicon photodiodes, superconducting tunnel junction (STJ) diodes, or metal-oxide-semiconductor (MOS) capacitors – have QE values in the 20%–95% range. Astronomers prefer these devices, in part, because of their high quantum efficiencies.

The quantum efficiency of a particular device is not always easy to measure, since (as in photography) the chain of events from incident photon to detection may be difficult to describe and quantify. ***Absorptive quantum efficiency*** is physically more straightforward, but somewhat less informative. It is defined as the photon flux absorbed in the detector divided by the total flux incident on its surface:

$$\eta = \frac{N_{\rm abs}}{N_{\rm in}}$$

Because absorbed photons are not necessarily detected, QE $\leq \eta$.

The ***quantum yield*** of a photon detector is the number of detection "events" per incident photon. For example, in silicon photoconductors, the detection event is the production of an electron–hole pair. If an incident photon has energy less than about 5 eV, it can produce at most one electron–hole pair, so the quantum yield is 1. For higher energy photons, a larger number of pairs are produced, around one e–h pair per 3.65 eV of photon energy. What happens in detail is that the first electron produced has so much kinetic energy that it can collide with the lattice to produce phonons that generate additional pairs. A 10-angstrom X-ray, therefore, will yield (on average) 34 photoelectrons. An STJ-based detector, you will recall, is particularly attractive because of its very large, wavelength-sensitive quantum yield.

8.1.3 Noise

> There are two kinds of light – the glow that illuminates, and the glare that obscures.
>
> – James Thurber (1894–1961)

Although efficiency in a detector is important, what really matters in evaluating a measurement is its uncertainty. The uncertainty in the output signal produced by a detector is often called the ***noise***, and we are familiar with the use of the ***signal-to-noise ratio***, ***SNR***, as an indication of the quality of a measurement. It

[1] Quantum efficiency is a bit of a slippery concept in photography. For example, once a grain has formed a stable clump of three–six silver atoms, absorbed photons can make no further contribution to the signal, even though they create additional silver atoms. The entire grain is either developed or not developed depending only on the presence or absence of the minimum number of atoms. In photography, QE is thus a strong function of signal level – the highest efficiencies only apply if the density of developed grains is relatively low.

would seem that a perfect detector would produce a signal with zero noise. This is not the case.

You will recall that there is an uncertainly *inherent* in measuring the strength of any incident light ray. For a photon-counting device, this uncertainty arises from the Poisson statistics[2] of photon arrivals, and is just

$$\sigma = \sqrt{N}$$

where N is the number of photons actually counted. A perfect detector, with QE = 1, faithfully counts all incident photons and will therefore produce

$$(\mathrm{SNR})_{\mathrm{perfect}} = \frac{N_{\mathrm{out}}}{\sigma_{\mathrm{out}}} = \frac{N_{\mathrm{in}}}{\sigma_{\mathrm{in}}} = \sqrt{N_{\mathrm{in}}}$$

Real detectors will differ from this perfect detector by either counting fewer photons (reducing the output noise, but also reducing both the output signal and the output SNR) or by exhibiting additional noise sources (also reducing the SNR). The ***detective quantum efficiency (DQE)*** describes this departure of a real detector from perfection. If a detector is given an input of N_{in} photons and has an output with signal-to-noise ratio $(\mathrm{SNR})_{\mathrm{out}}$, then the DQE is defined as a ratio:

$$\mathrm{DQE} = \frac{(\mathrm{SNR})^2_{\mathrm{out}}}{(\mathrm{SNR})^2_{\mathrm{perfect}}} = \frac{N_{\mathrm{out}}}{N_{\mathrm{in}}} \qquad (8.2)$$

Here N_{out} is a fictitious number of photons, the number that a perfect detector would have to count to produce a signal-to-noise ratio equal to $(\mathrm{SNR})_{\mathrm{out}}$. The DQE gives a much better indication of the quality of a detector than does the raw QE, since it measures how much a particular detector degrades the information content of the incoming stream of photons. For a perfect detector, DQE = QE = 1. For any detector, it should be clear from Equation (8.2) that DQE $\leq$ QE. If two detectors are identical in all other characteristics, then you should choose the detector with the higher DQE. If a parameter (wavelength of the observation, for example) affects both the input signal and the DQE, then you should choose a value that maximizes the value

$$(\mathrm{Signal})_{\mathrm{in}}\sqrt{(\mathrm{DQE})} = (\mathrm{SNR})_{\mathrm{out}}$$

Returning to the example of the photographic emulsion, the noise in an image is experienced as ***granularity***: the microscopic structure of, say, a star image consists in an integral number of developed grains. Statistically, counting grains in an image is a Poisson process, and has an uncertainty and a SNR of $\sqrt{N_{\mathrm{grains}}}$.

[2] Although we have been treating the photon-counting process as if it were perfectly described by Poisson statistics, both theory and experiment show this is not the case. Photon arrivals are not statistically independent – real photons tend to clump together slightly more than Poisson would predict. This makes little practical difference in the computation of uncertainties.

Since it takes something like 10–20 absorbed photons to produce one developed grain, the photographic process clearly degrades SNR. In addition, grains are not uniformly distributed in the emulsion, and some grains not activated by photons will nevertheless get developed to produce a background "fog." Both of these effects contribute noise, and thus reduce the DQE. A typical emulsion might have $\eta = 0.5$, QE = 0.04 and DQE = 0.02. Many solid-state detectors do not degrade the information in the input to anything like the degree that photography does, and their DQE values are close to their QE values – in the range 20–90%.

The DQE generally is a function of the input level. Suppose, for example, a certain QE = 1 detector produces a background level of 100 electrons per second. You observe two sources. The first is bright. You observe it for 1 second, long enough to collect 10,000 photoelectrons (so $\mathrm{SNR_{in}} = 100$). For this first source, $\mathrm{SNR_{out}} = 10,000/\sqrt{(10,100 + 10^2)} = 98$, and DQE = 0.96. The second source is 100 times fainter. You observe it for 100 seconds, and also collect 10,000 photoelectrons. For the second source, $\mathrm{SNR_{out}} = 10,000/\sqrt{20,000 + 10,000} = 57.8$, and DQE = 0.33.

8.1.4 Spectral response and discrimination

The QE of a detector is generally a function of the wavelength of the input photons. Some wonderful detectors are useless or have low QE at some wavelengths. Silicon devices, for example, cannot respond to photons with $\lambda > 1.1$ μm since these photons have energies below the silicon band-gap energy. The precise relationship between efficiency and wavelength for a particular detector is an essential characteristic.

One can imagine an ideal detector that measures both the intensity and the wavelength distribution of the incoming beam. An STJ diode, operated in a pulse-counting mode, for example, ***discriminates*** among photons of different wavelength.

8.1.5 Linearity

In an ideal detector, the output signal is directly proportional to the input illumination. Departures from this strict linearity are common. Some of these are not very problematic if the functional relation between input and output is well known and well behaved. For example, in the range of useful exposures, the density of a developed photograph is directly proportional to the logarithm of the input flux. Figure 8.2 illustrates two very typical departures from linearity. At lower light levels, a detector may not respond at all – it behaves as if there were an input ***threshold*** below which it cannot provide meaningful information. At the other extreme, at very large inputs, a detector can ***saturate***, and an upper threshold limits its maximum possible response. Further increases in input will not move the output signal above the saturation level.

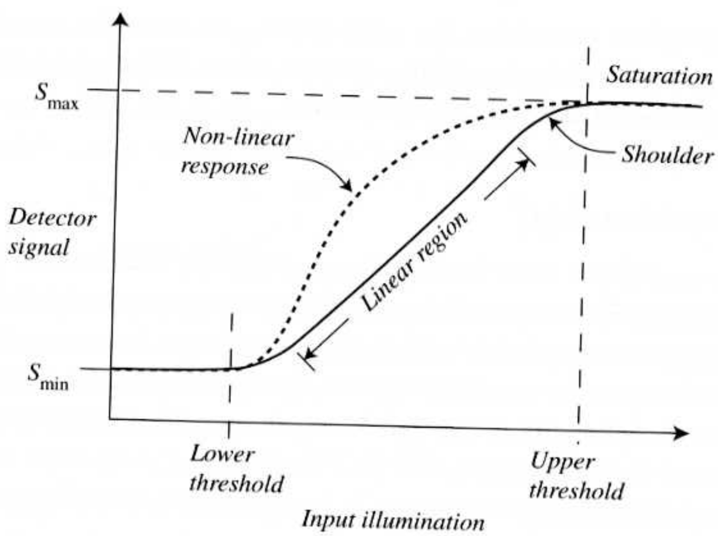

Fig. 8.2 Linear and non-linear regions in a typical detector response curve. The dashed response is completely non-linear.

8.1.6 Stability

The environment of a detector will change over time, perhaps because of variation in temperature, atmospheric conditions, or orientation with respect to gravity or to local magnetic fields. The detector itself may age because of chemical or mechanical deterioration, electrical damage, or radiation and particle exposure. Unrecognized changes can introduce systematic effects and increase uncertainties.

Two general approaches cope with detector instability. The first is to avoid or minimize anticipated changes: e.g. use thermostatic controls to maintain a constant temperature, keep the detector in a vacuum, shield it from radiation, use fiber-optic feeds so that the detector remains motionless. Basically, employ whatever strategies seem reasonable to isolate the detector from the environment. The second approach is to recognize that some changes are unavoidable and calibrate the detector to correct for the instability. For example, if the response of a detector deteriorates with age, make repeated observations of the same standard source so you can compute a correction that compensates for the deterioration.

Hysteresis is a form of detector instability in which the detector response depends on its illumination history. Human vision, for example, exhibits the phenomenum of positive and negative afterimage. Some solid-state detectors can continue to report ghost signals from bright objects long after the source has been removed.

8.1.7 Response time

How quickly can the detector make and report a measurement, then make and report the next measurement? The minimum time required is an important parameter. Readout procedures for large CCDs, for example, can limit their response time to a hundred seconds or more, while STJs and photo-emissive devices have sub-millisecond response times.

8.1.8 Dynamic range

What is the maximum range in output signal that the detector will produce in response to input? From Figure 8.2, you might surmise (correctly) that the upper and lower detection thresholds limit the dynamic range. However, other details of the detection process can influence the dynamic range. For example, if the signal is recorded digitally as a 16-bit binary integer, then the smallest possible signal is 1, and the largest is 65, 535 ($= 2^{16} - 1$). Thus, even if the range set by saturation is larger, the dynamic range is limited by data recording to 1:65, 535.

8.1.9 Physical size and pixel number

The physical size of the detector can be very important. To measure the light from a single star in the telescope focal plane, for example, it will be advantageous to match the detector size with the image size produced by the telescope: if the detector is too small, it will not intercept all the light from the source; if it is too large, it will intercept unwanted background light and probably produce a higher level of detector noise. For some detectors, physical size is related to other properties like dynamic range and response time.

A ***single-channel*** detector measures one signal at a time, while a ***multi-channel*** detector measures several at once. An astronomer might use a simple two-channel detector, for example, to simultaneously measure the brightness of a source and the brightness of the nearby background sky. A ***linear array*** (a string of closely packed detectors arranged in a straight line) might be a good configuration for sensing the output of a spectrograph. A ***two-dimensional array*** of detectors can record all parts of an astronomical image simultaneously.

Clearly, the physical size of each detector of an array determines how closely spaced its elements, or ***pixels*** (for *picture element*) can be. Sometimes there must be some inactive area between the sensitive parts of the pixels, sometimes not. Large arrays are more easily manufactured for some types of detectors (e.g. MOS capacitors) than for others (e.g. bolometers and wave detectors). There is an obvious advantage in field of view for detectors with a large number of pixels.

Astronomers currently employ mosaics of solid-state arrays of up to one billion pixels, with the largest individual arrays (CCDs of up to 100 megapixels in size) finding application in the X-ray through optical regions. Somewhat smaller arrays (1–4 megapixel) are in use at near-infrared (NIR) and mid-infrared (MIR) wavelengths. Focal-plane arrays of hundreds of pixels are used on some far-infrared (FIR) and sub-millimeter telescopes. Radio detectors are almost always single-pixel or few-pixel devices. At the beginning of the CCD era, photographic plates had a clear advantage in pixel number: for a very moderate cost, a photographic plate had a very large area (tens of centimeters on a side), and thus, in effect, contained up to 10^9 pixels. Mosaics of CCD arrays, although quite expensive, now match the size of medium-sized photographic plates.

8.1.10 Image degradation

Astronomers go to extremes to improve the resolution of the image produced by a telescope – minimize aberrations, launch the telescope into space, and create active and adaptive optics systems. Two-dimensional detectors like arrays should preserve that resolution, but in practice can often degrade it. ***Sampling theory*** was originally developed to understand electronic communications in media such as radio broadcasting and music reproduction. The Nyquist theorem states that the sampling frequency of a waveform should be greater than two times the highest frequency present in the wave. Extending this theorem to the spatial domain means that to preserve maximum detail, pixel-to-pixel spacing should be less than the ***Nyquist spacing***. The Nyquist spacing is one-half the full width at half-maximum (FWHM) of the point-spread function of the telescope. If pixel spacing is larger than the Nyquist value, the resulting ***under-sampling*** of the image degrades resolution.

Other effects can degrade resolution. Signal can drift or bleed from its pixel of origin into a neighboring pixel, or photons can scatter within the array before they are detected.

8.2 The CCD

> One morning in October 1969, I was challenged to create a new kind of computer memory. That afternoon, I got together with George Smith and brainstormed for an hour or so. . . .When we had the shops at Bell Labs make up the device, it worked exactly as expected, much to the surprise of our colleagues.
>
> – Willard Boyle, Canada Science and Technology Museum, 2008

When Boyle and Smith (1971) invented the first ***charge-coupled devices*** at Bell Laboratories in 1969 they quickly recognized the CCD's potential as multi-pixel light detector instead of a computer memory. By 1976, astronomers had recorded the first CCD images of celestial objects.[3] Since that time, the CCD has become a standard component in applications that include scanners, copiers, mass-market still and video cameras, surveillance and medical imagers, industrial robotics, and military weapon systems. This large market has diluted the research and development costs for astronomy. The consequent rapid evolution

[3] The first CCD images reported from a professional telescope were of the planets Jupiter, Saturn, and Uranus, taken in 1976 by Bradford Smith and James Janesick with the LPL 61-inch telescope outside Tucson, Arizona.

of the scientific CCD has profoundly revolutionized the practice of optical observational astronomy. This section gives a basic introduction to the principles of operation of the CDD and its characteristics as a detector.

8.2.1 General operation

Recall how an MOS capacitor stores photoelectrons in a potential well. A CCD is an array of MOS capacitors (one capacitor per pixel) equipped with circuitry to read out the charge stored in each pixel after a timed exposure. This read-out scheme (called "charge-coupling") moves charges from one pixel to a neighboring pixel; pixel-by-pixel shifting is what makes the array a CCD, rather than something else.

The basic ideas behind the array operation are simple. Imagine a matrix of MOS capacitors placed behind a shutter in the focal plane of a telescope. To take a picture, we first make sure all the capacitor wells are empty, open the shutter for the exposure time, then close the shutter. While the shutter is open, each pixel accumulates photoelectrons at a rate proportional to the rate of photon arrival on the pixel. At the end of the exposure, the array stores an electronic record of the image.

Figure 8.3 describes how the CCD changes this stored pattern of electrons into a useful form – numbers in a computer. In Figure 8.3a we show the major components of the detector. There is the light-sensitive matrix of MOS capacitors: in this case an array three columns wide by three rows tall. A column of pixels in the light-sensitive array is called a ***parallel register***, so the entire light-sensitive array is known collectively as the parallel registers. There is one additional row, called the ***serial register***, located at the lower edge of the array and shielded from light. The serial register has one pixel for each column of parallel registers (in this case, three pixels). Both the serial and parallel register structures are fabricated onto a single chip of silicon crystal.

Reading the array requires two different charge-shifting operations. The first (Figure 8.3b) shifts pixel content down the columns of the parallel registers by one pixel. In this example, electrons originally stored in row 3 shift to the serial register, electrons in row 2 move to row 3, electrons in row 1 move to row 2. Just before this first shift is initiated, the serial register is cleared of any charges that may have accumulated before or during the exposure.

The second operation now reads the newly filled serial register by shifting its contents to the right by one pixel (Figure 8.3c1). The electrons in the rightmost pixel shift into a new structure – a series of ***output amplifiers*** – that ultimately converts the charge to a voltage. This voltage is in turn converted to a binary number by the next structure, the ***analog-to-digital converter (ADC)***, and the number is then stored in some form of computer memory. The CCD continues this shift-and-read of the serial register, one pixel at a time (Figures 8.3c2 and 8.3c3) until all serial register pixels have been read.

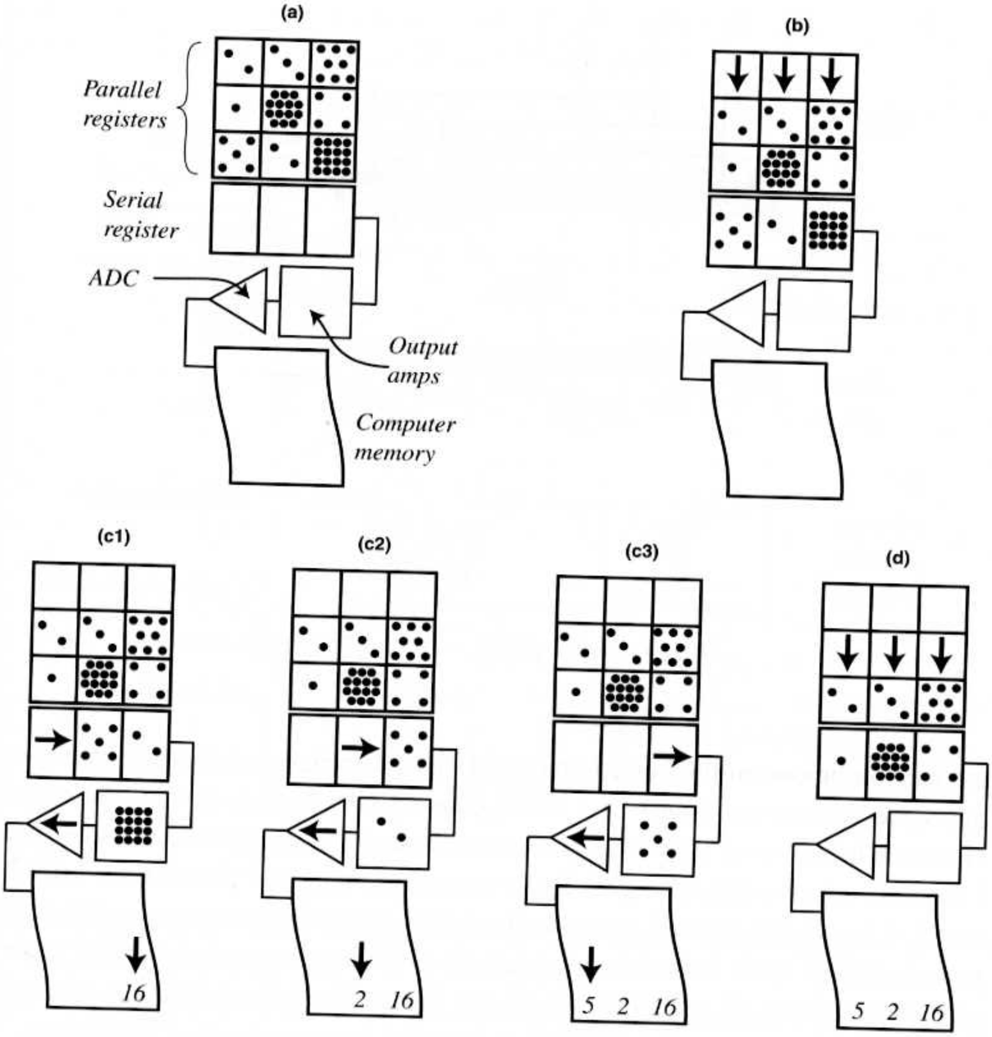

Fig. 8.3 CCD components and readout. (a) The accumulated photo-electrons in a 3 × 3 array of capacitors – the parallel register. (b) Shift of the bottom row into the serial register, all remaining rows shift down in the parallel register. (c) Read of the serial register one column at a time. (d) Next row shifts down into the empty parallel register.

Now the whole operation repeats for the next row: there is another shift of the parallel registers to refill the serial register with the next row (Figure 8.3d); the serial register is in turn read out to memory. The process continues (parallel shift, serial shifts, and reads) until the entire array has been read to memory. The first stage of the output amplifier is usually fabricated onto the same silicon chip as the registers. The subsequent amplifiers and the ADC are usually located in a separate electronics unit.

How does the CCD persuade the electrons stored in one capacitor to move to the neighboring capacitor? Many strategies are possible, all of which depend upon manipulating the depth and location of the potential well that stores the electrons. A parallel or serial register is like a bucket brigade. The bucket (potential well) is passed down the line of pixels, so that its contents (electrons) can be dumped out at the end. Figure 8.4 illustrates one strategy for moving the well. The depth of a potential well depends on the voltage applied to the metal, and is greatest at the Si–SiO_2 junction, closest to the metal layer. (See, however

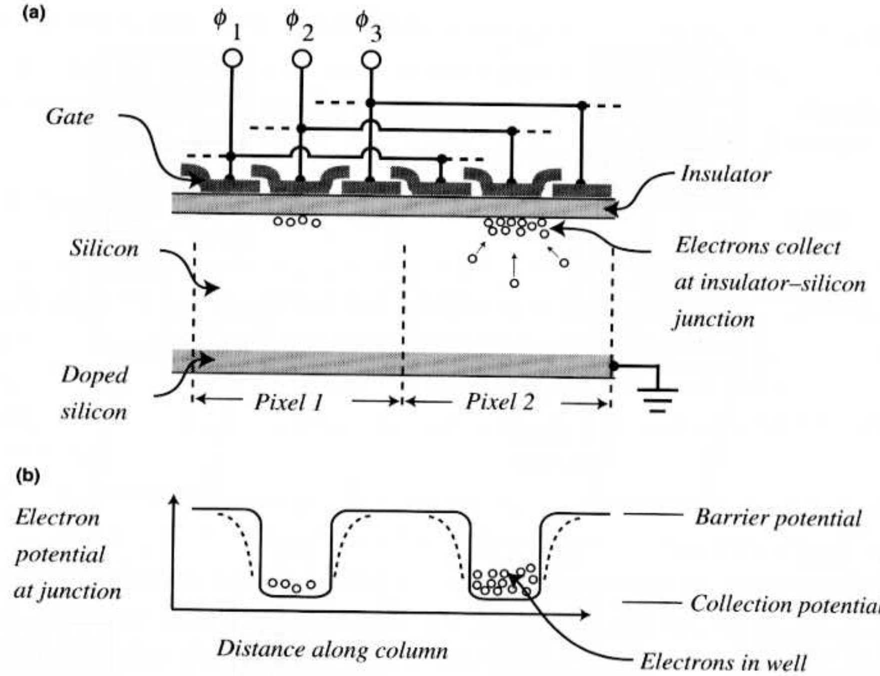

Fig. 8.4 Gate structure in a three-phase CCD. Two pixels are shown in cross-section. Collection and barrier potentials on the gates isolate the pixels from each other during an exposure. Overlapping gates produce a gradient in the barrier region (dashed curve in lower figure) that enhances collection.

the section below on the buried-channel CCD.) The figure shows two pixels in the same register (column) of a ***three-phase CCD***. In this device, the metal electrode is separated into three ***gates***, and these are interconnected so that gate 1 of every pixel connects to gate 1 of every other pixel, and likewise for gates 2 and 3. Thus, a single pixel can simultaneously have three separate voltages or ***phases*** applied to its front side, producing a corresponding variation in the depth of the potential well, as illustrated in the figure. The interconnection of gates insures that the pattern of well depth is identical in every pixel of the register.

Setting the correct voltages on three separate gates implements both charge-shifting and pixel isolation. For example, during an exposure, phase 2, the voltage on the central metal electrode, can be set to a large positive value (say 15 V), producing what is known as the ***collection potential*** in the semiconductor. The other two phases are set to a smaller positive voltage (say 5 V), which produces the ***barrier potential***. The barrier potential maintains the depletion region in the silicon, but prevents electrons from drifting across pixel boundaries. Photoelectrons generated in the barrier region of the silicon will diffuse into the nearest deep well under the collection phase and remain there. Each isolated pixel thus stores only charges generated within its boundaries.

To illustrate how the three gates might be used for charge shifting, assume again that the pixels are isolated during an exposure with collection under phase 2 (φ_2 = +15 V) and a barrier under the other phases ($\varphi_1 = \varphi_3$ = +5 V).

Figure 8.5 illustrates the three voltage changes that will shift charges by one pixel.

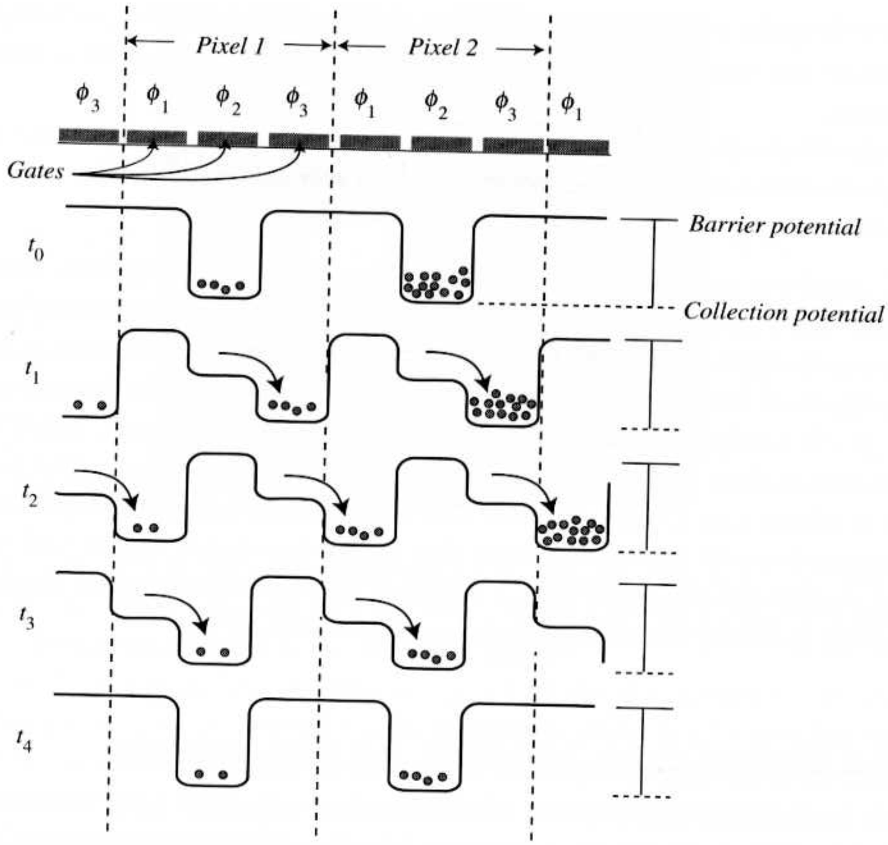

Fig. 8.5 Shifting potential wells in a three-phase CCD. See Figure 8.4 for the corresponding physical structure. Two pixels in the same register (either parallel or serial) are illustrated here. At the end of the shift, electrons stored in pixel 1 have shifted to pixel 2.

1. At time t_1, gate voltages change so that $\varphi_3 = 15$ V and $\varphi_2 = 10$ V. The electrons under φ_2 will diffuse to the right, and collect under φ_3.
2. At time t_2, after a delay that is long enough for all electrons to diffuse to the new location of the deep well, voltages change again, so that $\varphi_1 = 15$ V, $\varphi_3 = 10$ V and $\varphi_2 = 5$V. Stored electrons drain from phase 3 of the original pixel to phase 1 of the neighboring pixel.
3. A third cycling of gate voltages ($\varphi_1 = 10$ V, $\varphi_2 = 15$ V and $\varphi_3 = 5$ V) brings the electrons to the middle of the pixels at time t_3, and the one-pixel shift is complete.

The values of the barrier and collection potentials are somewhat arbitrary, but there are usually some fairly well-defined optimal values. These values, along with the properties of the insulator layer, determine required values of the ***clock voltages*** (the input values for φ_1, φ_2, and φ_3). An electronic system called the ***CCD controller*** or ***CCD sequencer*** sets the clock voltages and manages the very precise timing of their changes. The controller, usually built around a simple microprocessor, is generally housed in the same electronics box as the ADC and output amplifiers. Alternatively, the controller can be a program on a general-purpose computer. Besides manipulating the clock voltages, the controller also performs and coordinates several other functions, generally including:

- clearing the appropriate registers before an exposure and or a read;
- opening and closing the shutter;

- controlling the sequence of reads of the parallel and serial registers, including the patterns for special reads (see the discussions of on-chip binning and windowing below);
- controlling the parameters of the output amplifiers and the ADC (in particular, setting two constants called the ***bias level*** and the ***CCD gain*** discussed below);
- communicating with the computer that stores the data.

Two-phase and four-phase readout schemes are also sometimes used in CCDs. Most modern consumer digital cameras utilize arrays of (complementary-metal-oxide-semiconductor) ***CMOS*** capacitors, in which individual output amplifiers are fabricated onto the front side of each pixel. This design means that the pixels can be read out in parallel, rather than one at a time. The CMOS detectors are less expensive than CCDs of the same size, consume less power, and read out very rapidly (around 70 megapixels per second). They have not seen much use in astronomy, since they suffer from much higher read noise, dark current, pixel-to-pixel charge diffusion, and (usually) lower QE; however, they are gradually becoming more competitive with CCDs.

8.2.2 Channel stops, blooming, full well, and gain

The barrier potential prevents electrons from migrating from one pixel to another along a column in the parallel registers. What about migration along a row? In a classical CCD, shifts along a row are never needed, except in the serial register. The CCDs prevent charge migration along a row in the parallel registers by implanting (by heavily diffusing a dopant) a very highly conductive strip of silicon between columns. These ***channel stops*** held, say, at electrical ground, produce a permanent, extra-high barrier potential for stored electrons. Think of a pixel as a square bucket that holds water (or electrons). Two sides of the bucket, those that separate it from the adjacent columns, are maintained by the channel stop and are permanently tall and thin. The other two sides, the ones that separate it from its neighbors on the same column, are not as tall, and can be lowered or moved by "clocking" the gate voltages.

Consider what might happen if a pixel in an array fills with electrons during an exposure. As additional photoelectrons are generated in this saturated pixel, they will be able to spill over the barrier potential into the adjacent wells along their column, but cannot cross the channel stop. This spilling of charge along a column is called ***blooming*** (see Figure 8.6). Bloomed images are both unattractive and harmful: detection of photons in a pixel with a filled well becomes very non-linear; moreover, blooming from a bright source can ruin the images of other objects that happen to lie on the same CCD column. Nevertheless, in order to optimize the exposure of fainter sources of interest, astronomers will routinely tolerate saturated and bloomed images in the same field.

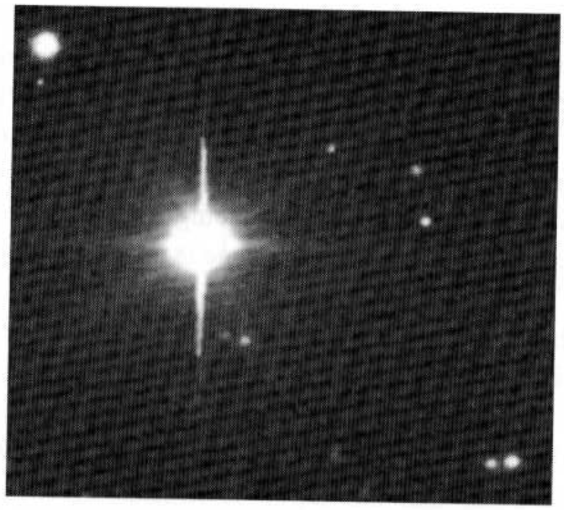

Fig. 8.6 Blooming on a CCD image: the saturated vertical columns are the bloom. The other linear spikes on the bright star image result from diffraction by the vanes supporting the telescope's secondary mirror.

There are designs for ***anti-blooming CCDs***. Recent designs utilize special clocking during the exposure in a buried-channel CCD (see below) to temporarily trap excess electrons at the oxide interface.

The maximum number of electrons that can be stored in a single pixel without their energies exceeding the barrier potential is called the CCD's ***full well***. The size of the full well depends on both the physical dimensions of the pixel, design of the gates, and the difference between the collecting and barrier potentials. Typical pixels in astronomical CCDs are 8–30 μm on a side and have full-well sizes in the range 25,000 to 500,000 electrons.

The final output from a scientific CCD is an array of numbers reported by the ADC to the storage computer. The number for a particular pixel is usually called its ***pixel content***, and is measured in ***ADU***s (analog-to-digital units). Pixel contents are proportional to the voltage the ADC receives from the output amplifier. The ***gain*** of the CCD is the number of electrons that need to be added to a pixel in order to increase the output contents for that pixel by one ADU.

For example, suppose a particular CCD has a full well of 200,000 electrons, and is equipped with a 16-bit ADC. The ADC is limited to digital outputs between 0 and 65,535 (= $2^{16} - 1$). A reasonable value for the gain might be 200,000/65,535 = 3.05 electrons/ADU. A smaller gain would mean that the CCD is better able to report small differences in pixel content, but would reach ***digital saturation*** before reaching the electronic full well. One might do this intentionally to avoid the non-linear shoulder in Figure 8.2. At a larger gain, the CCD would reach full well before the output could reach the maximum possible digital signal, so dynamic range would be reduced.

8.2.3 Readout time, read noise, and bias

To maximize DQE, the amplifier and ADC of an astronomical CCD should introduce the smallest possible noise to the output signal. A technique called

correlated double sampling (***CDS***) is capable of very low noise operation – only a few electrons per pixel. The noise added by the CDS circuit depends crucially on how quickly it does its job – the faster, the noisier. Another consideration – the time needed for the analog-to-digital conversion – also limits the read time per pixel. Practical times correspond to a pixel sample frequency of 10 to 200 kHz, with higher frequencies producing higher noise. Except for low frequencies, noise added by the amplifier stage is proportional to the square root of the frequency.

The basis of charge-coupled readout is the one-pixel-at-a-time movement of the array contents through a single amplifier, and this is a bottleneck. A low-noise CDS stage in a scientific CCD must read out slowly, and the larger the array, the longer the read time. An important difference between scientific-grade CCDs and the commercial-grade CCDs and CMOS arrays in camcorders is the readout rate – to obtain real-motion video images, an array must read out about 30 times a second. The large read noise that results is usually not objectionable in a consumer camera because of the high input level. In contrast, the astronomical input signal is usually painfully low, and a low-noise, ***slow-scan*** CCD for astronomy may require many tens of seconds to read a single image.

There are some cases in astronomy where the large read noise of a rapid scan CCD is not objectionable, and in which time resolution is very important – observations of occultations of bright stars or rapid changes in solar features, for example. Also note that a rapid scan is not a problem if no data are being digitized. Thus, reading an array to clear it before an exposure can be done very quickly.

For the usual astronomical tasks, though, it is mainly lengthy readout time that puts a practical limit on the number of pixels in a CCD. (Time spent reading the detector is time wasted at the telescope!) Two strategies can speed read times. The first uses multiple amplifiers on a single array. Imagine, as in Figure 8.7a, an array with an amplifier at each corner. The CCD has two serial registers, at the top and bottom. The controller clocks the readout to split the parallel registers – they read out to both ends simultaneously – and does the same with each serial register. Each amplifier reads one quarter of the array, so the total read time is reduced by the same factor. The image can then be re-assembled in software. Multi-amplifier astronomical CCDs up to 9000×9000 pixels now (2010) exist.

A second strategy is to build a mosaic of several very closely spaced but electrically independent CCDs. Figure 8.7b shows an eight-element mosaic read by 16 amplifiers. An early device similar to this, the *Mosaic Imager*, was placed in service at the Kitt Peak National Observatory in 1998. It contained eight 2048×4096 CCDs arranged to form an 8196×8196 pixel (64 megapixel) detector that is 12 cm (5 inches) on a side. Gaps between the individual CCDs are about 0.6 mm (40 pixels). A relatively simple combination of shifted multiple exposures will fill in those parts of an image masked by the gaps on a single

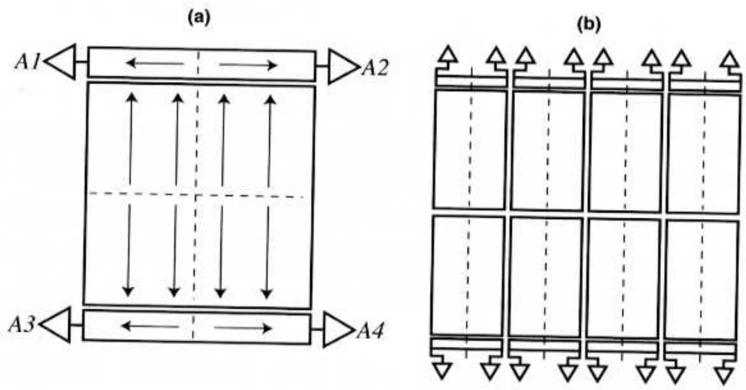

Fig. 8.7 Large-format CCD strategies. (a) A large monolithic detector with multiple serial registers and amplifiers (four, in this case). Read time is reduced by a factor equal to the number of amplifiers, and the total CTE is improved. (b) A mosaic of eight arrays butted to form a single large-area detector.

exposure. Mosaics have become so important that some modern CCDs are manufactured to be "almost-four-side-buttable" – so that the width of the gaps in a mosaic need be only to 30–100 pixels on all sides. At the present time (2009), there are several 50–120 megapixel mosaic arrays in service, and several observatories are about to introduce mosaics of over 100 devices and up to 3 gigapixels (see chapter 4 of Howell, 2006). These huge arrays expect to have not-very-objectionable readout times in the 20–60 second range. A major problem with these large-format arrays, in fact, may turn out to be simple data storage: an observer can expect to generate terabytes of image data in a few nights.

8.2.4 Dark current, cooling, and vacuum enclosures

At room temperature, a CCD is a problematic detector for astronomy. The energy of thermal agitation generates electron–hole pairs in the depletion zone and the resulting steady flow of electrons into the CCD potential wells is called ***dark current***. Dark current is bad for three reasons:

1. It adds some number of electrons, N_D, to whatever photoelectrons are produced in a pixel. You must make careful calibrations to subtract N_D from the total.
2. Dark current adds not only a background *level*, N_D, but also introduces an associated uncertainty or *noise* to any signal. Since the capture of dark-current electrons into the pixel wells is a random counting process, it is governed by Poisson statistics. The noise associated with N_D dark electrons should be $\sqrt{N_D}$. This noise is more insidious than the background level, since it can never be removed. Dark current *always* degrades SNR.
3. At room temperature, the dark current can saturate a scientific CCD in seconds, which makes it impossible to record faint objects. Not good.

Lower the temperature of the CCD, and you reduce dark current. The Fermi distribution provides an estimate for the rate at which dark charges accumulate in a semiconductor pixel:

$$\frac{dN_D}{dt} = A_0 T^{\frac{3}{2}} e^{-\frac{E_G}{2kT}}$$

Here T is the temperature in kelvins, A_0 is a constant that depends on pixel size and structure, and E_G is the band-gap energy. A large fraction of dark current in a pixel arises at the Si–SiO_2 interface of the capacitor, where discontinuities in the crystal structure produce many energy states that fall within the forbidden band. Electrons in these interface states have small effective band gaps, and hence produce a large dark current.

A common method for cooling a CCD is to connect the detector to a ***cryogen*** – a very cold material with a large thermal mass. A very popular cryogen is a bath of ***liquid nitrogen*** (***LN_2***), a chemically inert substance that boils at 77 K = –196 °C. Since it is generally a good idea to keep the CCD at a somewhat warmer temperature (around –100 °C), the thermal link between detector and bath is often equipped with a heater and thermostat.

A cold CCD produces difficulties. The CCD and the LN_2 reservoir must be sealed in a vacuum chamber for two reasons. First, a CCD at –100 °C in open air will immediately develop a coating of frost and other volatiles. Second, the vacuum thermally insulates the LN_2 reservoir from the environment, and prevents the supply of cryogen from boiling away too rapidly. Filling the CCD chamber with an inert gas like argon is a somewhat inferior alternative. Vacuum containers, called ***Dewars***, can be complicated devices (see Figure 8.8), but are quite common in observatories. At a minimum, the dewar must provide a transparent window for the input, a method for feeding electrical signals though the vacuum seal, a system for adding cryogen, and a method for periodically renewing the vacuum.

Another option for more modest cooling is dry ice (solid CO_2), which is less expensive than LN_2. Dry ice sublimates at –76 °C = 197 K.

Compact and relatively inexpensive thermoelectric (***Peltier junction***) coolers instead of cryogens require very small dewar sizes. These solid-state coolers can maintain a detector in the –30 to –50°C range, where the dark current of an ordinary CCD is still quite high, but where the dark current from an MPP CCD (see below) is acceptable for many astronomical applications. Such coolers are considerably more convenient to use than cryogens.

At the other extreme, superconducting junctions, many small band-gap detectors for the infrared, and most bolometers, require temperatures below what liquid nitrogen provides. ***Liquid helium***, which boils at 4.2 K, is an expensive cryogen that is difficult to handle. Liquid ^{3}He boils at 3.2 K, but is even more difficult and expensive. To avoid the expense of evaporating helium into the air, one option is a ***closed-cycle refrigerator*** that compresses and expands helium fluid in a cycle. If they employ two or three stages, these systems can cool detectors to the 10–60 K range. Special closed systems using helium-3 evaporation can bring small samples to temperatures in the 0.3–3.2 K range.

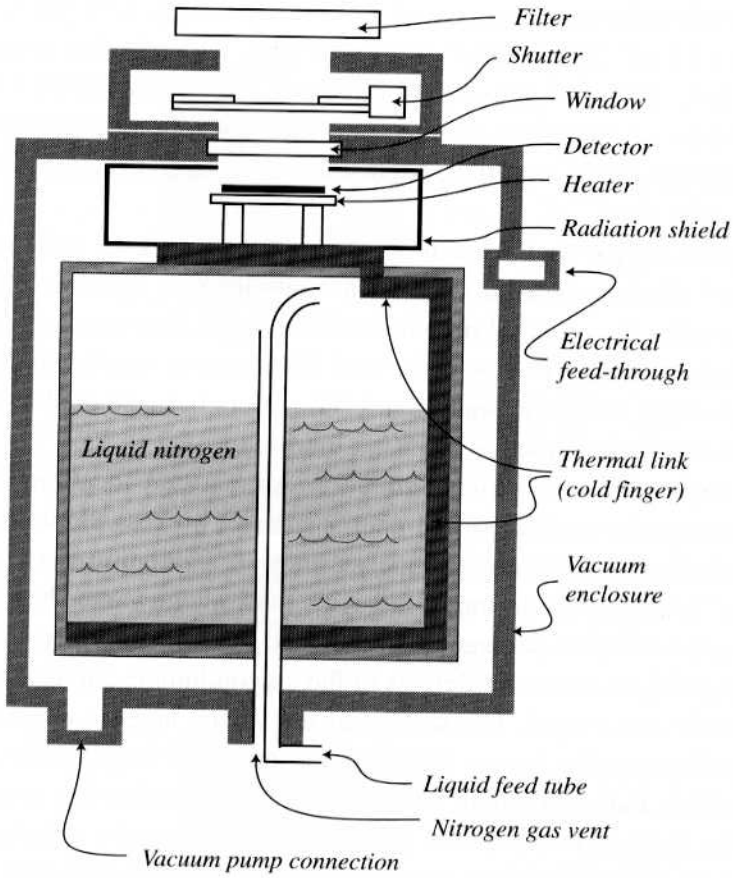

Fig. 8.8 A simple dewar for cooling a detector using liquid nitrogen. This design is common for devices that "look upward," and prevents cryogen from spilling out of the reservoir as the dewar is tilted at moderate angles.

8.2.5 Charge-transfer efficiency

The charge-coupled readout works perfectly only if all the electrons in a well shift from pixel to pixel. Disaster results if significant numbers of electrons are left behind by a shift. Images will appear streaked, and photometry becomes inaccurate. Signal loss because of charge-transfer inefficiency is greatest from the pixels furthest from the amplifier. The fraction of electrons in a pixel that are successfully moved during a one-pixel transfer is the ***charge-transfer efficiency***, or ***CTE***. Although one transfer will require three clock cycles and sub-pixel transfers in a three-phase device, CTE is always computed for a full pixel transfer. In a single-amplifier CCD, p is the actual number of full pixel transfers needed to read a particular charge packet. If the rows and columns of the parallel registers are numbered from the corner nearest the amplifier, then $p = R + C$, where R and C are the row and column numbers of the pixel in question. The fraction of the original charge packet that remains after p transfers (the total transfer efficiency, or TTE) is just

$$\mathrm{TTE} = (\mathrm{CTE})^p$$

The CTE needs to be very close to one. For example, suppose a 350 × 350 pixel array has a CTE of "three nines" (CTE = 0.999), which in this context is *not* very close to 1. Then $p = 350 + 350 = 700$, so TTE = $(0.999)^{700} = 0.49$; this device will lose over half the charge from the most distant pixel in the array before bringing it to the amplifier. Multi-megapixel arrays require CTE values approaching six nines.

What limits CTE? One issue is time – when CCD gate voltages change during a read, electrons need time to diffuse into the new location of the potential well. Usually, the required time is shorter than the time needed for the CDS and amplifiers to complete a low-noise read. However, at very low temperatures, electron velocities can be so small that CTE suffers because of slow diffusion, and so operation below about −100 °C is inadvisable.

Charge ***traps*** are a more serious limitation. A trap is any location that will not release electrons during the normal charge-transfer process. Some traps result from imperfections in the gates, channel stops, or the insulation of a pixel – flaws that deform the potentials during a read cycle to create unwanted barriers. Other traps are due to radiation damage, to unintended impurity atoms (usually metals like iron or gold), to structural defects in the silicon lattice, and to some effects not completely understood. The surface of the silicon layer in contact with the insulator will invariably have a large number of charge traps; these are such a serious problem that all modern CCDs are designed so that the potential well excludes the front surface (see the next section). Some traps affect only a few electrons each. If scattered throughout the entire body of a CCD, they produce a small decrease in the overall CTE. Other traps can render a pixel non-functional, so that it will not transfer charge in a meaningful way. This compromises the entire column upstream from the trap. Devices with a "bad column" or two are still very useful, but place additional demands on the observing technique.

Manufacturing defects can also cause a complete failure of charge transfer. The usual problems are short circuits and open circuits in the gate structure, or shorts between a gate and the semiconductor. Any of these can render a single pixel, a partial or complete column, or an entire device unreadable. The expense of a particular CCD is directly related to the ***manufacturing yield*** – if many devices in a production run need to be discarded, the cost of a single good device must rise. In the early days of CCD manufacture, yields of acceptable devices of a few percent were not uncommon.

8.2.6 The buried-channel CCD

The simple MOS/MIS (metal-insulator-semiconductor) capacitor we have been discussing up until now has its minimum electron potential (i.e. the bottom of the collection well) at the Si–SiO_2 interface. A CCD made of these capacitors is a ***surface-channel*** device, since charge transfer will require movement of electrons close to the interface. The high density of trapping

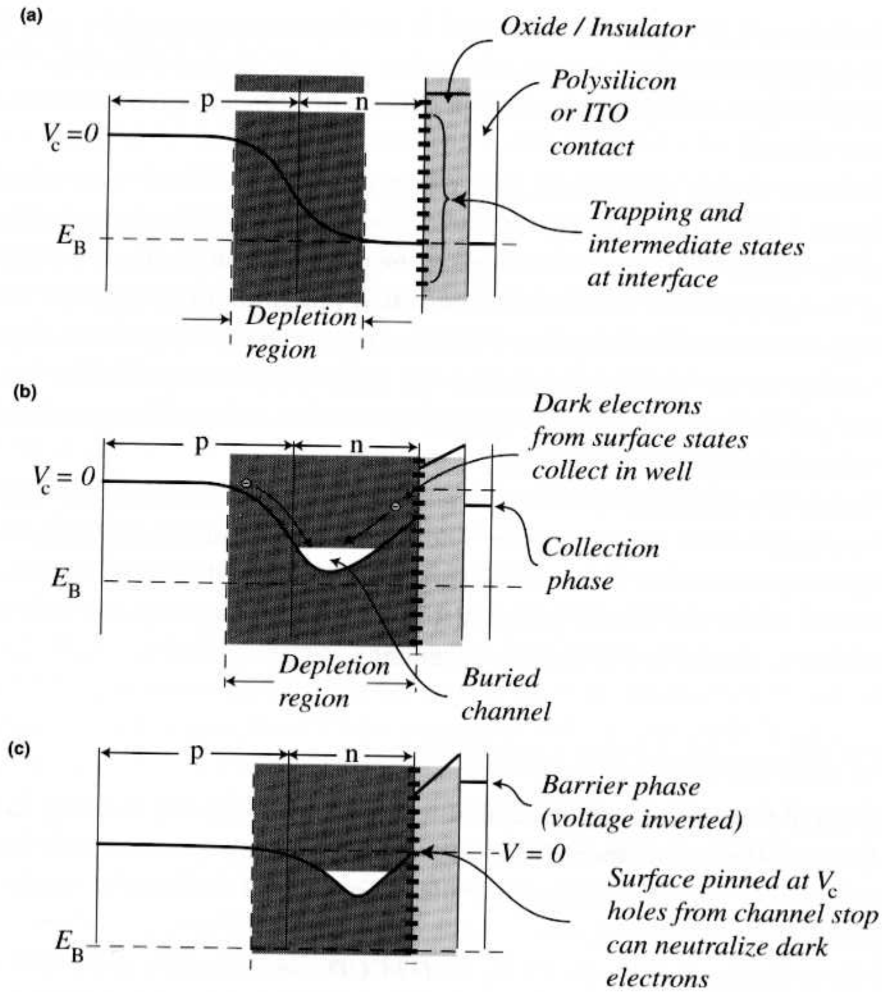

Fig. 8.9 A buried channel in a p–n junction capacitor. (a) There is no buried channel in the electron potential when the normal collection phase voltage is applied. If the gate voltage is reduced, as in (b), electrons collect away from the interface. (c) Inverting the voltage on the barrier-phase electrodes pins the surface potential to the channel-stop value and allows a current of holes to flow to neutralize dark-current electrons.

states at the interface makes it impossible to achieve acceptable charge-transfer efficiency in a surface-channel CCD. (Values of only 0.99 are typical.) All modern scientific CCDs are designed so that the transfer channel is located several hundred nanometers below the interface. In these ***buried-channel CCDs*** (***BCCDs***), all electrons collect in a region safely removed from the surface traps, and all charge transfers take place within the unperturbed interior of the semiconductor lattice.

Manufacturers can produce a buried channel by constructing a p–n junction near the semiconductor surface. Figure 8.9 illustrates the basic principle. Figure 8.9a shows the potential energy for electrons in an MOS or MIS device in which the semiconductor consists of a thin n-type region (perhaps 300–800 nm thick) layered on top of a much thicker p-type region. Within the semiconductor, the potential exhibits the basic pattern for a junction diode – there is a high-resistivity region depleted of majority charge carriers near the junction, and a potential difference, E_B, across the depletion zone. In Figure 8.9a we connect the p side to electrical ground, and set the gate voltage a relatively large positive voltage near

E_B. In this state, photoelectrons created in the depletion zone will be swept into the broad channel in the n region, where they can still interact with surface traps. Making the gate voltage even more positive will deepen the well and create a surface channel.

To create the buried channel, the voltage on the gate is made *more negative*. In Figure 8.9b, the gate voltage has been lowered so that electrons are repelled from the surface. This alters the shape of the potential and produces a minimum in the n-region, which is called the ***collection potential***. (The required voltage on the gate is the ***collection phase***.) Note two important features: First, electrons that collect in the well do not contact the surface. (This is good.) Second, the capacity of the well is reduced compared to a surface-channel device made from the same material. (This is not so good).

Figure 8.9c illustrates the electron potential under the barrier phase. Here the gate voltage is even more negative. The potential minimum in the semiconductor, although somewhat closer to the surface, is still buried. As a result, electrons generated under the barrier phase also avoid the surface traps as they move internally to the nearest collection potential.

8.2.7 Alternative CCD readout designs

You should be aware of several alternative methods for reading out the CCD that offer some specialized advantages. Consult Howell (2006) or the manufacturers and observatory websites (e.g. pan-STARRS, e2v, Kodak; see Appendix I) for further details.

The ***orthogonal-transfer CCD***, or ***OTCCD*** (see Tonry *et al.*, 1997) has a gate structure that permits charge-coupled shifting of pixel contents either along the row or along the column, on either the entire array or on subsections. Orthogonal-transfer CCDs can make small image shifts to compensate for tip–tilt seeing-disk motion during an exposure, and are being used for the 1 gigapixel mosaic of the pan-STARRS project.

Frame-transfer CCDs permit a very short time interval between successive frames. They recognize that it is the amplifier stage that limits the readout rate of a scientific CCD, so rapidly read an acquired frame into an inactive (shielded) set of parallel registers. The device then reads the shielded frame slowly through the amplifier while the next frame is being acquired.

Low-light-level CCDs or ***L3CCDs*** have additional extra-large, deep-well MOS capacitors in a "charge multiplication" extension of the serial register. The device clocks charges from the serial register into these capacitors at a very high voltage, so that the energy of a transferred electron can produce an additional electron–hole pair when it enters a multiplication capacitor. Several hundred multiplication transfers typically produce multiplication gains of 100–1000 before amplification, so read noise is insignificant, permitting rapid readout (1–10 MHz) and true photon-counting at low light levels.

8.2.8 The MPP CCD

Interface states at the Si–SiO_2 junction remain the major source of dark current in a simple BCCD. Thermal electrons can reach the conduction band by "hopping" from one interface state to another across the forbidden gap. You can eliminate this electron hopping by ***pinning*** a phase, as in Figure 8.9c. To pin the phase, you set the voltage on the gate to so negative a value that the potential at the interface ***inverts***, that is, it reaches the same potential as the back side of the p region, which is also the same, V_c, as the potential of the conductive channel stops. Any further reduction in the gate voltage has little effect on the interface potential, since the surface is now held at ground by holes that flood in from the channel stops. The abundance of holes means that thermal electrons are neutralized before they can hop through the interface states. Dark current in a pinned phase is reduced by several orders of magnitude.

A ***partially inverted*** three-phase CCD operates with one non-inverted phase (the collection phase, as in Figure 8.9b), and with the other two phases pinned and serving as the barrier phases, as in Figure 8.9c. Dark current in such a device is about one third of what it would be in a completely non-inverted mode. If all three phases are pinned, the CCD is a ***multi-pinned-phase (MPP)*** device, and dark current less than 1% the rate in non-inverted mode. The obvious difficulty with MPP operation is that there is no collection phase – the buried channel runs the entire length of a column. Multi-pinned-phase devices therefore require additional doping under one of the phases to make a permanent collection potential. This is possible because the value of E_B in Figure 8.9 depends on the density of dopants in the semiconductor. In an MPP device, for example, the surface under phase 2 might invert with the collection phase set at –5 V, while the other two (barrier) phases require –7 V for inversion.

With their remarkably low dark currents, MPP CCDs can operate at room temperature for several minutes without saturation. In recent designs, dark rates below 0.1 electron per second are routine at –40 °C, a temperature attainable with inexpensive thermoelectric coolers. An MPP CCD controlled by a standard personal computer is a formidable and inexpensive astronomical detector within the financial means of many small observatories, both professional and amateur. As a result, modern observers using telescope apertures below 1 meter are making quantitative astronomical measurements of a kind that would have been impossible at the very best observatories in the world in 1975.

The full-well capacity of an MPP device is a factor of two or three less than a partially inverted BCCD. Modern MPP devices nevertheless have respectable full wells. Appendix I gives the specifications for a few devices currently on the market. If the larger full well is more important than the reduced dark current, the proper selection of clock voltages makes it possible to run a device designed for MPP operation in a partially inverted mode.

8.2.9 Surface issues

We need to address the very practical question of getting light into the depletion region of the CCD pixels.

Frontside options. The most direct approach sends light through the metal gates. Since even very thin layers of most metals like copper or aluminum are poor transmitters, the "metal" layer of the CCD is usually made of highly doped ***polysilicon***: silicon in a glass-like, amorphous state – a random jumble of microscopic crystals. A thin (about 0.5 micron) layer of doped polysilicon is both relatively transparent as well as a good electrical conductor, but it does, however, absorb green, blue and (especially) ultraviolet light. Other conductive materials, like doped ***indium tin oxide*** (***ITO***) have better transparency properties than polysilicon; ITO electrodes are becoming common, but are somewhat harder to fabricate.

There are two general strategies for further improving the short wavelength QE of a front-illuminated CCD. The first is somehow to make the gate structure more transparent. The second is to change the wavelength of the incoming light to one at which the gates are more transparent.

Open-electrode architecture improves transparency with a gate structure that leaves part of the pixel uncovered. For example, the collection-phase electrode might be oversized and shaped like a hollow rectangle. It is even possible to fabricate pixel-sized ***microlenses*** over the frontside to redirect much of the incoming light to the uncovered area of each pixel.

A related approach is the ***virtual-phase CCD***, where a single gate covers half of the pixel, and a four-step potential profile is constructed by implanting dopants in the semiconductor. Changing the voltage on the single gate can produce pixel-to-pixel charge transfer similar to a four-phase CCD. Virtual-phase CCDs have even better blue QEs than open-electrode devices, especially if equipped with microlenses, but are more difficult to fabricate and generally have relatively poor CTE values.

A different strategy applies a thin coating of ***phosphor*** on top of the gates. The useful phosphors are organic molecules that absorb a short-wavelength photon to move to an excited state, then de-excite by emitting one or more longer-wavelength photons. Lumigen (or lumogen), for example, is a commercial compound that absorbs light shortward of 420 nm, and is otherwise transparent. Upon de-excitation, it emits photons at around 530 nm, which can easily penetrate polysilicon gates. Since phosphors emit in all directions, they will slightly degrade image resolution at short wavelengths. Another drawback is that some phosphors tend to evaporate in a vacuum, especially at high temperatures.

Backthinning. A completely different solution sends the light in through the back (from the bottom of Figure 8.10) of the device, avoiding the gates completely. This ***backside illumination*** has the advantage that green, blue, and ultraviolet, which would be absorbed by a polysilicon or ITO layer, will pass

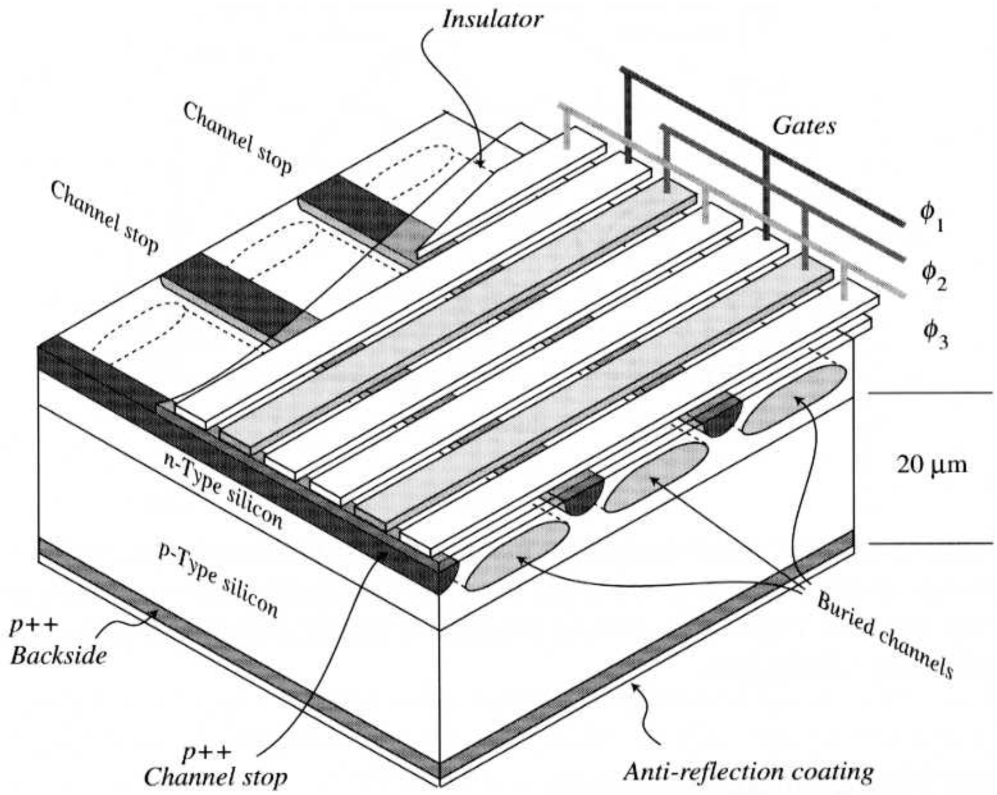

Fig. 8.10 Schematic of a thinned, three-phase CCD. In a conventional CCD, insulated gate electrodes usually overlap, while in an open architecture, gaps more closely follow the pixel pattern. This drawing is of a backthinned device. A front-illuminated device would have a much thicker silicon layer, with the AR coating above the gates.

directly into the silicon. Since these photons have a short absorption depth, they create photoelectrons mainly near the back face of the device. This is a serious problem. In order for the electrons to be able to diffuse from the back face into the depletion zone without recombining, the semiconductor layer needs to be very thin (10–20 μm). "Thinning" the silicon will in turn reduce its ability to absorb NIR photons, which have a large absorption depth. The final geometry needs to be something of a compromise. Nevertheless, astronomers have generally embraced backthinned CCDs, since they detect a considerably larger fraction of incident photons of all wavelengths than does any frontside-illuminated device (see Figure 8.11). Their main drawback is that they are difficult to manufacture and therefore expensive, if available at all.

If red and near-infrared QE is very important, the ***deep-depleted CCD*** offers some improvement over the normal backthinned device. Because the depth of the light-sensitive depletion zone is inversely proportional to the dopant concentration, use of a lightly doped (high resistivity) silicon layer means that the total layer thickness of the CCD can be increased to about 50 μm. The thicker detector has greater long-wavelength sensitivity, and is mechanically easier to fabricate. However, achieving the required resistivity can be difficult, and cosmetic quality to date has been inferior to thin devices.

Anti-reflection coatings. An anti-reflection (AR) coating is most effective for light of a particular wavelength, so a CCD designer must choose the coating

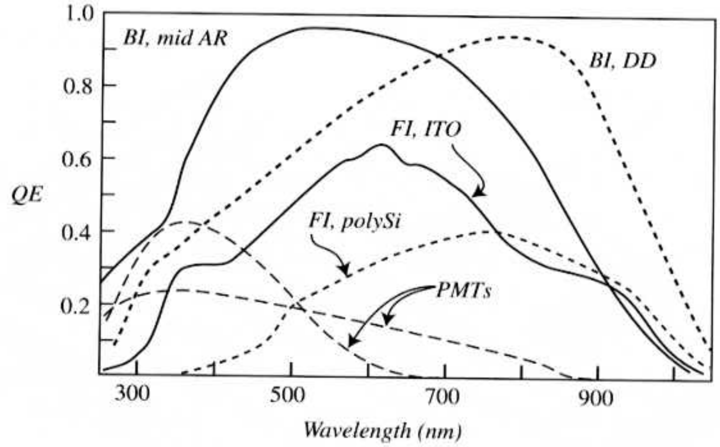

Fig. 8.11 Efficiencies of light detection for various illumination strategies in a CCD, and photocathode choices in a PMT. Curves are representative of the extremes. The abbreviations BI and FI indicate back- and front-illuminated CCDs. The figure shows the QE curves for a normal thinned device with a mid-band AR coating, a deep-depletion (DD) CCD with a near-infrared coating, ITO and polysilicon front-illuminated CCDs. The two photomultiplier tubes (PMTs) are very high efficiency bi-alkali photocathodes with different spectral sensitivities.

with the intended use of the detector in mind. Often CCD manufacturers offer a choice of coatings to enhance either the short-wavelength, mid-wavelength or NIR response. Figure 8.11 shows a selection of the QE characteristics of a few different CCD designs.

8.3 Photo-emissive devices

Researchers have developed the simple vacuum photodiode described in the last chapter from a detector of very limited capability (with poor QE in the red, very low signal levels, mechanical fragility, and single-channel operation) into devices that compete with or enhance CCDs in special circumstances. In this section we examine three astronomical detectors that depend upon the vacuum photoelectric effect.

8.3.1 The photomultiplier tube

One disadvantage of the simple vacuum photodiode described in the last chapter (Figure 7.24) is low signal level. The ***photomultiplier tube (PMT)*** is a vacuum device that increases this signal by several orders of magnitude. Figure 8.12 illustrates its operation. In the figure, a voltage supply holds a semi-transparent photocathode on the inside of the entrance window at large negative voltage, usually around one or two kilovolts. A photon hits the cathode and ejects a single electron. In the vacuum, this electron accelerates towards the more positive potential of a nearby electrode called a ***dynode***, which is coated with a material (e.g. Cs_3Sb, CsKSb, BeO, GaP) that can easily release electrons to the vacuum if hit by an energetic particle. Because the original photoelectron impacts the dynode with 100 eV or so of kinetic energy, it usually ejects several secondary electrons. The number of secondary electrons is a statistical quantity whose mean value, δ, usually lies between 2 and 10. The group of electrons

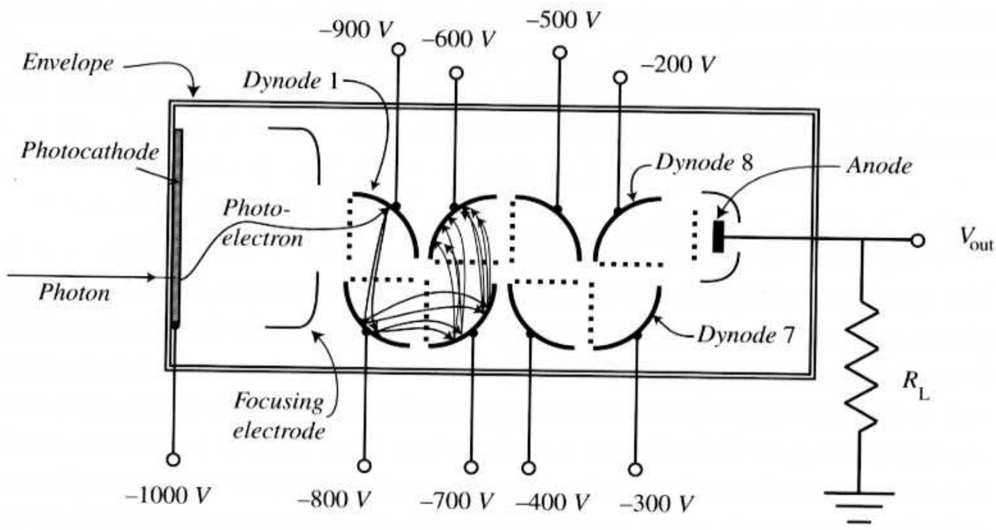

Fig. 8.12 A simple photomultiplier tube. The potential of the first dynode accelerates a single photoelectron emitted from the cathode. Its impact releases several secondary electrons, which accelerate and hit dynode 2, releasing another generation of secondaries. After (in this case) eight stages of amplification, a large pulse of electrons flows through the anode and load resistor to ground.

ejected from the first dynode then accelerates to the second dynode, where each first-dynode electron produces δ second-dynode electrons. The process continues through n dynodes, until the greatly multiplied pulse of electrons lands on the anode of the PMT. If each dynode is equivalent, the total number of electrons in a pulse generated by a single photoelectron is

$$N = A\delta^n$$

where the factor A accounts for inefficiencies in redirecting and collecting primary and secondary electrons.

In the figure, the signal is the average DC voltage measured across a load resistor. However, for weak sources, the large pulses of electrons that arrive at the anode are easily counted electronically, and the PMT can operate in a ***pulse-counting mode***: each pulse is generated by the arrival of a *single* photon at the cathode. In this mode, the QE of the PMT depends on the QE of the photocathode, which can be as high as 40–50% for some materials (see Figure 8.11).

The single-channel PMT was the detector of choice for precise astronomical brightness measurements from 1945 until the advent of CCDs in the early 1980s. The spectral responses of the available PMT photocathode materials defined, in part, some of the now-standard photometric band-passes (the U, B, and V bands in Table 1.2, for example). Since photomultipliers have few advantages over CCDs, they have become rare at observatories. One important advantage of the PMT, however, is response time. The temporal spread of a single pulse at the anode limits the shortest interval over which a PMT can sense a meaningful change in signal. Pulse widths are so narrow (5–10 nanosecond) for many PMTs that they can, in principle, detect signal changes as rapid as a few milliseconds. The response time of a CCD, in contrast, is several tens of seconds for a standard slow-scan device, with quicker response possible with increased noise. Arrays of STJs, although still in the development phase, have the potential for response times similar to PMTs.

8.3.2 The microchannel plate

The upper part of Figure 8.13 shows an important variation on the PMT. Take a glass capillary with a diameter between 5 and 25 μm, and a length around 40 times its diameter. Coat the inside surface of this tube with a semiconductor that has good secondary electron-emitting properties, and connect the ends of the channel coating to the voltages as shown. You have created a ***microchannel***. Place this microchannel assembly in an evacuated chamber between a photocathode and an anode, and it can serve in place of the dynode chain of a PMT. A

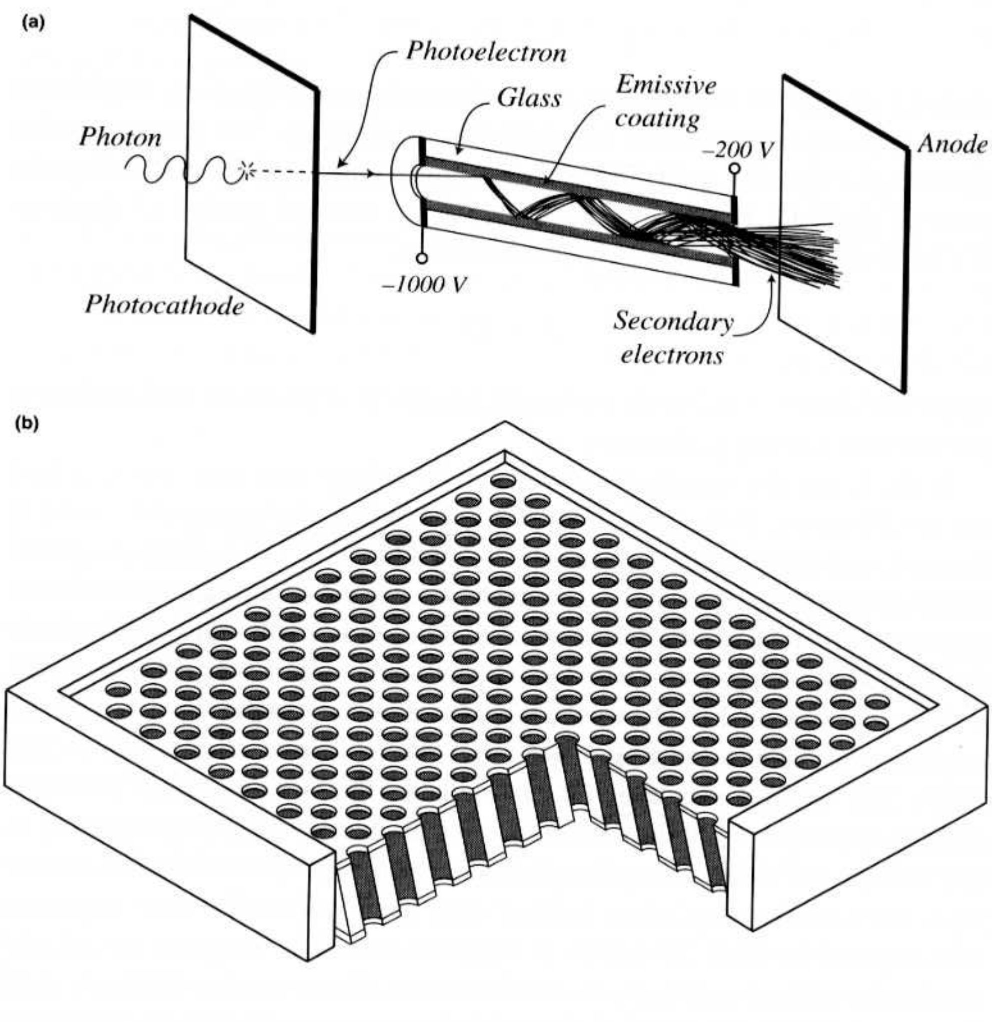

Fig. 8.13 (a) A single microchannel – a small glass tube whose interior is coated with dynode-type material. A large potential drop along the tube insures that an electron impact at one end will produce a burst of secondary electrons at the other. (b) A closely packed array of channels forming an MCP.