CHARGE-COUPLLED DEVICES IN ASTRONOMY

Craig D. Mackay

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, England

1. INTRODUCTION

Barely a dozen years have passed since the first two-dimensional charge-coupled devices (CCDs) were made. Since then they have become one of the most commonly used detectors for ground-based astronomy. CCDs are so widely used in astronomy because for many applications they represent a substantial improvement on previous detector systems. In many respects, CCDs are within a small factor of being perfect detectors. Because they give such dramatic improvements over earlier technologies, they have been introduced rapidly, sometimes with an inadequate appreciation of their true virtues and vices. CCDs are capable of giving the astronomer data of the highest quality only when they are used with a good understanding of the realities of CCDs in general and of the particular CCD system used.

This review attempts to provide the general observer with a guide through the jargon that surrounds these instruments so that he can gain some insight into the principles of their operation. We then look at the problems of real CCDs in practice and examine how these should be characterized in ways that can give a clear guide to the application of a particular system to a specific observing program. This review mainly covers aspects of CCD systems that are currently in use for production astronomy. No attempt is made to cover possible future applications of CCD technology to X-ray or infrared wavelengths, since these are very immature technologies and are a long way from being common user instruments. However, we do look briefly at other applications for the future at the end of the review.

0066-4146/86/0915-0255$02.00

© Annual Reviews Inc. • Provided by the NASA Astrophysics Data System
2. PRINCIPLES OF CCD OPERATION

When a photon is absorbed in silicon, an electron-hole pair is created. In a CCD, covering electrodes create potential wells in the silicon that collect those electrons that do not recombine with the lattice. Red photons on average pass farther into the bulk of the silicon before absorption than blue photons. The variation of typical absorption depth with wavelength is given in Figure 1. For many scientific applications CCDs are cooled to reduce dark current. This cooling also reduces red response because the band gap increases [from 1.10 eV at 300 K to 1.15 eV at 100 K (MacFarlane et al. 1958)] and also because the absorption of the lower energy photons requires the presence of higher energy phonons in the silicon lattice. These are rarer at lower temperatures.

The surface of the CCD is covered with electrodes biased so that ele-

![Figure 1a](image)  

*Figure 1a* The absorption coefficient of photons in silicon. Surface reflection losses are neglected.
Quantum efficiencies for a variety of CCDs: (1) A typical P8600 CCD, unthinned, illuminated through the electrodes; (2) a P8600 CCD overcoated with a layer of laser dyes to extend performance into the near-UV (Cullum et al. 1986); (3a, 3b) predicted response from a thinned CCD with backside treatment plus two different antireflection coatings (J. R. Janesick & J. McCarthy, private communication); (4) the distribution of the best responses for silicon detectors (devices that are not CCDs but that also depend on photon absorption in silicon) from a variety of manufacturers data sheets.

Electronic carriers preferentially diffuse toward the electrodes with the highest applied potential. (The holes diffuse away into the bulk of the silicon and are effectively lost.) The voltages on the electrodes relative to the substrate voltage create potential gradients over a few microns that sweep out any free carriers. This is the depletion region. Carriers generated outside the depletion region diffuse slowly and will often recombine with the lattice before they come under the influence of the depletion-region field gradient. Carriers created near the edge of the depletion region can diffuse sideways and may then be collected by adjacent electrodes. This causes device resolution to be reduced in the red, where photons are absorbed deeper in the device on average.

The structure of a three-phase CCD is shown in Figure 2. The silicon is covered with three sets of electrode strips, each isolated from the silicon substrate and from one another. One of the three is biased more positively than the other two, and it is under this one that carriers are accumulated. The carriers are restrained from moving along the length of the electrode by channel stops. These are narrow regions of heavily doped p-type material normally in depletion. Their negative charge repels electrons and therefore
prevents carrier movement across the stops. This defines the pixel extent in that direction.

At any time the charge accumulated under one electrode may be transferred to the area below an adjacent electrode if the adjacent electrode potential is raised while the first is lowered (Figure 2a). The transfer may be in either direction, and all the charges stored over the two-dimensional imaging area are moved simultaneously in that direction. This process may be repeated in order to transfer the accumulated two-dimensional charge distribution over many pixels.

Most CCDs are manufactured for use at domestic TV frame rates, and their internal organization reflects this. TV applications require that light falls continuously onto the sensitive surface (no shutter), and thus the device organization must allow the accumulated charge image to be moved out quickly (to minimize image blurring) and then read out slowly to give the continuous data stream that constitutes the TV image.

Figure 2a  The charge-transfer mechanism employed in three-phase CCDs.
Figure 2b  The basic layout of a three-phase two-dimensional CCD. The sequence 1, 2, 3 on each set of electrodes indicates the normal direction of charge transfer in the parallel and serial registers.

Examples of frame-transfer devices are three-phase and virtual phase CCDs made by Texas Instruments (United States) and devices made by RCA (United States), EEV/GEC (UK), Thompson CSF (France), and Mullard/Philips (Netherlands). These divide the silicon into two areas of comparable size that may be clocked to transfer the charge image either separately or together (Figure 3). During the frame flyback interval the charge image accumulated in the image area is rapidly moved into the store area. The store area is generally similar to the image area except that it is covered with an opaque screen. The contents of the store area are clocked out at a steady rate. Each row of store-area pixels is transferred into an output register. The output or serial register is a one-dimensional CCD structure with each element connected at the end of each column to the two-dimensional or parallel register. One end of the output register is connected to an output amplifier. The charges in the output register are transferred one at a time to the output amplifier in exactly the way that the charges are transferred across the parallel part of the device. Most of the devices structured with separate image and store regions are available without the opaque screen over the store area. The store-area pixels are then just as light sensitive as those in the image area, and so users, such
as astronomers, who can use a mechanical shutter will benefit from a doubling of the number of sensitive pixels.

Interline-transfer CCDs such as those made by Fairchild (US) are structured so that individual columns of imaging pixels simultaneously transfer their charges into transfer registers that are light shielded. These in turn may be clocked so as to pass their charge to a shielded output register. As frame-transfer devices are the most widely used devices for astronomy at present, we shall concentrate on them. The Texas Instruments 800 × 800 devices made for the Hubble Space Telescope Wide Field/Planetary Camera and the relatively new Tektronix devices are designed specifically for non-TV applications and so are organized with a single imaging area. The Tektronix devices (Figure 4; Blouke et al. 1985) are connected to serial output registers on both sides of the device. This allows the image to be read out into either register. For these devices the registers differ only in the structure of their output amplifiers, with one optimized for slow readout rates and the other for faster rates.

The simpler CCD structures consist of electrodes that are deposited directly onto an insulating layer on the top of a slice of uniformly doped

![Figure 3](image_url)  
*Figure 3* The frame-transfer three-phase CCD.
silicon. The charge is stored and transferred in contact with the surface. This is a surface-channel device. However, the structure of the surface layer is capable of adversely affecting charge-transfer efficiency (CTE). It is easy for charge to become trapped at sites (often associated with defects in the lattice) that have a slow release time. This can give significant image smear. Furthermore, surface defects inject spurious charges into the silicon that give increased dark current.

Most manufacturers avoid these problems by growing a more highly doped n-type layer onto the p-type substrate on top of the silicon between the substrate and the insulating layers under the polysilicon electrodes. The potential profile created gives a minimum that is away from the surface, and it is at this level that the charge is stored and transferred. This is the buried-channel structure. Note that the same considerations affect the design of the on-chip output amplifier. With a buried-channel design amplifier, performance (particularly with respect to noise) is greatly improved, since surface defects can dominate amplifier noise.

3. A TYPICAL FRAME-TRANSFER CCD

Many astronomers believe that CCDs are complicated, almost magical devices that are beyond their comprehension. In fact, CCDs are standard
integrated circuits made with the widely used CMOS technology. They also have an extremely simple structure, with the large numbers of pixels obtained by orthogonal strips of polysilicon electrodes and implanted channel stops. To demonstrate this simplicity, we look in detail at the P8600 CCD (Figure 5) made by EEV Ltd. (UK) and examine the voltages or waveforms that have to be applied to each electrode to give good performance.

The P8600 CCD is built from a slice of silicon fairly heavily doped so that carriers generated diffuse only a very short distance before recombining with the silicon crystal lattice. A thin layer is then grown on top of this to follow the lattice structure (an epitaxial layer) with a much lower doping level and so a much longer diffusion scale. This ensures that only carriers created within the epitaxial layer and not in the substrate will be collected by the electrodes. If carriers are allowed to diffuse up from deep in the device (as happens with bulk CCDs that have no epitaxial layer), the resolution of the device is degraded and cosmic-ray events (see below) become a much greater problem.

The light-sensitive area of the P8600 (Figure 5) is defined by 1732 strips

![Figure 5](image)

Figure 5  The full schematic of the P8600 series CCDs marketed by EEV Ltd., England. The numbers in circles indicate the device pin numbers. The detailed function of each electrode is described in the text.
of electrodes of polysilicon (silicon deposited from the vapor phase during processing). These are connected in threes and are biased (one high at about 0 V, two low at −10 V relative to some arbitrary level) to define one collection site or pixel. This gives 578 rows of pixels. Implanted channel stops define the boundaries of the pixels in the other direction to give 385 columns of pixels, each of area 22 μm². The image- and store-section clock lines are connected in phase to give one contiguous detecting area of 385 × 578 pixels. The substrate voltage is between −2 and −6 V.

The bottom of each column is connected to one electrode of the output register. The output register similarly consists of polysilicon electrodes connected in threes biased with one high (0 V) and two low (−10 V). Both the parallel-imaging area and the serial output register have input structures that are normally biased so as to stop spurious charge from getting into the register. The input drains of the parallel and serial structures (pins 1 and 22, respectively) are biased high (+12 V), whereas the input gates (pin 30 and pins 20 and 21, respectively) are biased low (−12 V).

With these fixed potentials, the CCD will store a two-dimensional charge distribution in proportion to the number of photons that fall on each pixel. Note that the output register is also light sensitive, as is the output transistor. At the end of the exposure the shutter is closed to avoid image smear during the readout process, and the two-dimensional charge distribution is moved down one row toward the output register. This is done by gradually (perhaps over a few microseconds) decreasing the voltage on the electrode held high while increasing the voltage on the next electrode nearer the output register (see Figure 2a). This is repeated twice to complete the charge transfer toward the output register by one row. When the electrode nearest to the output register goes low, the charges are transferred to the set of electrodes in the output register that are high. If the charge is now read out along the serial output register in exactly the same way with similar waveforms, the full resolution of the CCD will be obtained. Alternatively, it is possible to add the charge from consecutive rows into the output register by repeating the parallel charge-transfer operation before reading out the serial register. The pixels of the output register are considerably larger than those in the imaging area in order to give them greater charge capacity and so reduce the risk of saturation.

In order to understand the readout operation, we need to look at the layout of the on-chip output amplifier. Figure 5 shows a diode at the output end of the serial output register. This is reverse biased so that it is effectively a simple capacitor (about 0.1 pF), and it is across this capacitor that the charge from each pixel of the output register will be dumped. This capacitor is connected to a transistor with two gates, one of which is biased
high (reset bias gate, +10 V, pin 12) and the other which is biased low (reset gate, −10 V, pin 11). The latter gate turns the transistor off, so that the reset transistor drain (+5 V, pin 13) is effectively disconnected from the output node capacitor. This capacitor is also connected to the output transistor, which has its drain high (+12 V, pin 14) and its source (pin 15) supplied by a constant-current source set to 2–2.5 mA. The output transistor source is also connected to the signal-processing electronics and has a gain of about 0.7. This means that the output source voltage changes at 0.7 times the rate of voltage change across the node capacitor.

The output amplifier allows the charge present on the node capacitor to be measured. The node capacitor causes the charges dumped across it to be integrated. The level with reference to which this integration occurs may be preset by switching the reset gate high (0 V) and so turning on the reset transistor (making the reset transistor source drain impedance low and setting the potential on the node capacitor to that of the reset drain). The reset transistor is turned off after a few microseconds, and the next charge packet may then be transferred onto the node capacitor. The P8600 has an identical dummy output amplifier with no charge-input structure. It is used to provide a balanced output for TV scan-rate applications and is best left disconnected for slow-scan astronomical use.

The charges stored in the output register are transferred toward the node capacitor in exactly the same way as charges are transferred across the parallel-imaging area. The last electrode of the output register is the output gate (pin 11), and it is biased at −10 V, just above the low level (−12 V) of the serial clock electrodes. Charges are then transferred through this output gate to the node capacitor when the electrode next to the output gate goes below −10 V. This arrangement stops charges on the node capacitor from escaping back into the last elements of the output register. The output register consists of 400 elements, 15 more than the number of columns in the parallel-imaging area. It has 11 elements at the beginning and 4 at the end not connected to the imaging area. Properly used, these elements can provide a dark reference level. Indeed, if the parallel or the output registers are clocked for more than the actual numbers of rows or columns, respectively, the additional pixels should also indicate this true dark level. Overclocking the output register is most commonly done for this purpose. The serial clocking and signal processing may be organized so that the charges from several contiguous output register pixels may be summed across the output node capacitor. We see later that this pixel-binning procedure can improve signal-to-noise levels under the right conditions.

Some users have the impression that CCDs are very fragile devices that are easily destroyed. Mechanically, they are very robust and can be exposed
to high light levels even when cooled without risk of permanent damage. Electrically, it is possible to damage CCDs. They are CMOS integrated circuits, and their gate connections are prone to damage by static discharge. Proper CMOS handling procedures must be used, but if these are used, then the device should be safe. Once installed, it is possible to damage CCDs by forward biasing (and so drawing current) through the junctions that form the drains and sources on the device. For the P8600 devices, the drain and source voltages must always be greater than the substrate voltage. Finally, if any of the on-chip transistors are allowed to conduct too much current, they can be destroyed. This can be avoided with simple current-limiting circuitry. In practice, with a working CCD system designed to cope with fault conditions that could damage the CCD and an established procedure for powering the system on or off, there should be little real risk to the CCD.

4. THE PRACTICE AND CHARACTERIZATION OF CCDs

In reality, CCDs are not as predictable as the previous section might seem to imply. Often the potentials and drive waveforms have to be carefully adjusted to give optimum performance, although some CCDs (such as the P8600) work well over quite a range of drive conditions. Several excellent reviews of the more subtle aspects of the care and feeding of CCDs have been published. Djorgowski (1984) emphasizes the astronomical applications, whereas Janesick et al. (1985a) give an excellent, very detailed account of many aspects of CCD performance. A wealth of technical detail is given by Wright (1982). In this section we look at some of the practical problems often encountered with CCDs and CCD systems.

4.1 Photon Detection in Silicon

In a device such as the P8600 described in Section 3, photons pass through the polysilicon electrodes into the silicon. The redder photons pass farther, on average, into the silicon, since it becomes progressively more transparent to longer wavelength photons. Carriers are created once the photons are absorbed, and these are collected under the electrodes, provided that they are generated in the depleted region of the epitaxial layer. However, the polysilicon electrodes are essentially opaque to light with wavelengths shorter than 400 nm, and they only become progressively more transparent up to 600 nm. Peak quantum efficiencies above 40% can be obtained with this structure. Other techniques have to be used to give sensitivity at shorter wavelengths.

A fairly simple technique of obtaining some blue response is to overcoat
the device with a thin layer of a chemical that absorbs blue or UV photons and re-emits visible or red photons. Organic phosphor films such as coronene have been used for the CCDs flying in the Hubble Space Telescope. Recently, Cullum et al. (1986) used laser dye mixtures to give a useful blue response (Figure 1b). Although effective, none of these coatings can give very high quantum efficiency, since the re-emitted photons are radiated in all directions.

Another method is to use a different basic CCD structure, one pioneered by Texas Instruments and known as the virtual phase CCD (Janesick et al. 1981). Here only one electrode is used, the other half of the pixel being left clear of polysilicon. Instead of multiple electrodes, different levels of surface doping create potential steps that direct charge flow as the single-electrode voltages are swung above and below the potentials of the uncovered (virtual) electrode. Such a structure provides some blue response, yet the efficiency is still some way from the extremely high efficiency with which blue photons are absorbed by silicon.

In order to get direct access to the silicon, it is possible to thin the supporting substrate of the CCD down to the boundary with the epitaxial layer. The thinned device may be freely supported (as in the Texas Instruments 800 × 800 CCDs) or stuck onto a glass window (as with some RCA CCDs) and the light allowed to fall directly on the exposed silicon surface. Although the quantum efficiencies obtained in the red were excellent, it was found that the response at wavelengths shorter than 400 nm was very poor despite the high absorption of blue photons by the silicon. Indeed, it was to overcome this problem that coronene was used on the Hubble Space Telescope CCDs to improve the UV response of their thinned devices. The reasons for the poor UV response are complex (Janesick et al. 1985a). Essentially it is because the thinned surface becomes charged, so that the potential rises through the surface and then falls into the depleted region of the device. Carriers created by blue and UV photons are generated very close to the surface of the silicon and often do not have enough energy to escape from the surface potential well. Eventually they recombine with the lattice. Many attempts have been made to understand this phenomenon and then to overcome it. It was discovered that by flooding some Texas Instruments (TI) devices (where the silicon surface was exposed) with UV light and then cooling them immediately, the blue/UV response was greatly increased and device uniformity improved. Apparently the UV photons ionized molecules that adhered to the surface of the device and, by depositing charge on the surface, distorted the potential through the surface so as to force surface carriers into the bulk of the thinned device. Running a small discharge arc was found to have a similar effect. More recently J. R. Janesick (Jet Propulsion Laboratory internal memos) has used thin (50
Å) coverings of various metals to modify the surface potential and to improve the blue response.

If the surface is not adequately charged, then incident photons can create charges that contribute to the surface charge of a thinned device. This in turn affects the local quantum efficiency and can appear to modulate the device sensitivity in a manner dependent upon the amount of light already incident on the device. This effect can cause ghost images to persist after an image has been read out, since the local device sensitivity will depend in a nonlinear fashion on the past exposure of the device to light. Such an occurrence is called quantum efficiency hysteresis.

Thinning can have other undesirable effects even if the thinned surface is properly charged. Silicon has a high refractive index, and so there are inevitably significant reflections from any silicon-glass or silicon-air/vacuum interface. In the red, silicon becomes progressively more transparent, and multiple internal reflections can occur. These can modulate the local device sensitivity, giving a distinctive fringe pattern under monochromatic illumination. Such patterns can be fatal for far-red spectroscopic programs, where each pixel is illuminated with essentially monochromatic light. Nearly 100% sensitivity modulation can happen. Problems can occur even in direct broadband imaging, since the components of the night-sky emission (Broadfoot & Kendal 1968) are practically monochromatic and can therefore produce lower level fringes on the image. Antireflection coatings can help but will never eliminate this effect, and therefore it is generally easier to get better uniformity in the red by using an unthinned device.

4.2 Unwanted Signal Generation in CCDs

Carriers are created in silicon not only by light but also by thermal effects. These give rise to dark current, the signal generated in the absence of any input light source. Its magnitude depends critically on the quality of the silicon used and on the processing procedures used to manufacture the device. The boundary between the silicon and the insulation under the electrodes is such a substantial source of dark current that buried-channel devices are essential if dark current is to be minimized. In buried-channel devices, this component may be effectively eliminated by appropriate substrate bias (Saks 1980). A second source of dark current is defects in bulk of the silicon. Such dark current is present in both buried- and surface-channel devices. The dark current \( I \) follows the following well-known diode law:

\[
I = A \exp \left( -\frac{B}{kT} \right).
\]

where \( A \) and \( B \) are constants, \( k \) is Boltzmann’s constant, and \( T \) is the
device temperature. Cooling the CCD, therefore, dramatically reduces dark current, and at typical operational temperatures, a 10°C change in device temperature gives a factor of three change in dark current. In practice, three-phase CCDs in astronomy are cooled to temperatures between −160°C and −60°C, depending upon the application. Under some circumstances (such as to maximize far-red response) it may be necessary to operate with an elevated device temperature. In principle, it is possible to subtract off the dark current, provided this can be determined precisely. Such a determination requires accurate dark exposures to be made and the device temperature to be controlled precisely. In addition, the effects of dark current can be more serious than might be expected from the mean dark-current figure, since devices are often very nonuniform in their dark current. “Hot” pixels can give dark-current spikes on an image that are hard to subtract out.

Another important source of unwanted signals in CCDs is provided by cosmic-ray events. High-energy particles generate secondary muons, which deposit around 80 electrons μm⁻¹ in silicon. With a collection depth of 10–20 μm, a cosmic-ray event is seen on a CCD image as having a signal of perhaps a few thousand electrons, usually concentrated in a small number of pixels (Leach & Gursky 1980). In nonepitaxial devices, where the collection depth is greater and the events are more diffuse, the removal of these electrons can be rather tricky (Goad 1980). The cosmic-ray event rate is a subject that has received considerable attention, since the values predicted (Leighton 1957, p. 689) are several times lower according to several authors than those observed. Recently, however, experiments done by the present author have managed to reduce the “cosmic-ray event rate” by a factor of 5 to 1.5–2 events cm⁻² min⁻¹. This reduction was achieved by replacing the glass front window of the CCD dewar with fused-silica windows. It appears that several forms of UV- and blue-transmitting glass, such as the Schott (Germany) UBK7 and GG385 glasses, emit X-rays at a low level, and these are efficiently detected by the CCDs. These glasses appear to emit at levels of around 10 events cm⁻² min⁻¹.

A final source of unwanted signals in all CCDs (Janesick et al. 1985a) is device luminescence, which can be caused by operating the device near the limits of the acceptable voltage range. It is also caused by partial shorts between electrodes in the parallel or serial registers. A low-impedance path can act like a light-emitting diode and saturate the device in only a few minutes. For this reason, it is important to select devices carefully and operate them properly for use with long integration times. Alternatively, device potentials may be changed to eliminate luminescence during long integration and reset to a different value for readout.
4.3 Collection and Transfer of Charge in CCDs

Photons absorbed deep in a front side–illuminated (unthinned) CCD create carriers that have an increased chance of recombination before being collected by the electrodes. This reduces the effective overall detective quantum efficiency of the device. Furthermore, the collection of charges from these depths permits charge to diffuse sideways, which compromises the resolution of the device. Generally, in these devices blue resolution is significantly better than far-red resolution. For thinned devices, the thickness is generally less than the pixel size, so resolution is fairly constant.

Once the two-dimensional charge distribution has been accumulated under the CCD electrodes, it has to be moved out by the clock waveforms applied to the electrodes. There are several things that can stop the charge from being transferred out properly. Problems will occur if the potential well under the electrodes is nearly filled. The clock potentials become affected and charge transfer can be degraded. If the wells are indeed full, then charge can spread up and down the columns containing the saturated pixels. Other problems relate to specific device faults in individual columns or at specific elements of the output register. A break in one of the electrode strips can effectively stop or delay charge from being transferred past that point. Defects in the bulk substrate silicon can propagate during manufacture into the CCD structure. Charges can be caught there (trapped states), and charge transfer past them can be delayed in a rather nonlinear manner. In practice, it is very difficult to work with data from columns containing trapped states. Even in devices that are free from such defects, charge-transfer problems are common. It is often difficult to design clock drivers to give the highest charge-transfer efficiency (CTE), since the structure of the overlapping electrodes may create small pockets in the channel potential distribution that can trap small amounts of charge. In some devices the connections between the parallel and serial registers are inadequate, so that charge is inefficiently transferred from one register to the other.

The low-level charge transfer inefficiency (sometimes called the deferred-charge problem) can be greatly reduced in importance by precharging the device before readout. If a constant amount of charge is added to each pixel, then the CTE of the device is substantially improved. This added signal of a few tens to a few hundreds of electrons per pixel is sometimes called a fat zero or skinny zero (Baum et al. 1981). It is often generated with a short flash from a light-emitting diode, and of course it implies the use of a separate "dark" exposure so that this added charge pedestal can be subtracted from the final image. The principal disadvantages of this
procedure are (a) the added charge increases the effective readout noise of the frame, and (b) devices that need this treatment usually have residual charge-transfer problems that continue to affect performance at the lowest light levels.

The final problem we must address in this section is the methods generally used to measure CTE. It is quite common to find that the CTE of a device is quoted as some number such as 0.99998 per transfer. This implies that after 1000 transfers, 98% of the charge originating in one pixel is still in that pixel, the remaining 2% occurring in adjacent pixels. However, it is clear from the first part of this section that CTE is something that is dependent on a number of other factors. These factors include the temperature of the device, sometimes a serious problem (see Figure 25 of Janesick et al. 1985a) and sometimes (as with the P8600 devices from EEV/GEC) only a very slow function of temperature, at least down to −185°C. Tests using a calibration chart of fine bright lines on a dark background sometimes give results that depend critically on both the bright-line signal level and on the background dark level (which can provide the fat zero some devices need to improve CTE). It is also usually true that the CTE in the parallel-transfer direction is different from that in the serial register direction. Finally, we will find in any real device that some columns transfer charge better than others.

Techniques used to measure CTE often rely on illuminating the CCD to a high level and then seeing what is left beyond the last row and column. Often these measurements are lumped together to provide a single measure of CTE that is totally meaningless.

All these considerations make it clear that we need better methods of measuring CTE. Janesick et al. (1985a) particularly emphasize the value of X rays in determining CTE. X rays of a few keV energy are absorbed efficiently in silicon, and they give rise to one electron for every 3.65 eV of X-ray energy. The physics of the absorption process are complicated (Janesick et al. 1985c). A small cloud of electrons (a few hundred, or a 1000-Å diameter) is created at a depth of typically 20 microns. With thinned devices or devices with a small collection depth, X-ray capture will often lead to essentially all of the charge being in a single pixel. It is very instructive to examine the charge levels in pixels surrounding the event pixel, since they give a good guide to the CTE along that row and column and at that signal level.

Generally, however, the most useful CTE measurements come either from laboratory simulations of the kind of observations of interest to the astronomer or else directly from the observations themselves. CTE problems are at their most acute in spectroscopic applications, where the signal levels are at their lowest. Laboratory tests are easier at higher signal
levels but are usually only a poor guide to low-signal-level performance. A test chart consisting of a coarse grid of narrow bright lines on a dark background might look excellent at high signal levels such as 100–200 electrons per pixel, but at levels of 10–20 electrons per pixel (which might be close to the readout noise level of the device) it is difficult to estimate the line profile. It is often very helpful to compare the observed positions of the grid at two signal levels, since a displacement at the lower level of a pixel or two in the directions away from the output amplifier can be the first indication of CTE problems. Indeed, in spectroscopic applications this is the most critical test problem, since when sky subtraction is to be performed, the object spectrum can provide a local signal pedestal to ensure sky lines on the object are well transferred, whereas sky lines away from the object (where there is much less continuum) will be more affected by CTE problems.

4.4 Readout Procedures

The structure of the CCD permits the charges in any number of rows to be added together into the serial output register and then for any number of output register pixels to be added together onto the CCD output gate (provided saturation does not occur).

The charge packet \( Q \) from a pixel or group of pixels that have been added or binned in the device is transferred across the parallel register and into the serial register, then along the serial register and through the output gate. There it is deposited on the gate of the output field effect transistor (see Figure 5). This output gate has a capacitance associated with it \( (C_{\text{node}}) \), and so a voltage change occurs \( (V = Q/C_{\text{node}}) \). The smaller the node capacitance, the larger the value of \( V \) obtained for a given charge packet \( Q \). For the P8600 device, \( C_{\text{node}} \) is about 0.1 pF, and one electron corresponds to 1.6 \( \mu \)V.

CCD architecture differs from device to device more in the design of the on-chip output amplifier than in other areas. Nevertheless, the same basic principles hold, and the EEV/GEC P8600 device described in Section 3 is quite typical of the simpler designs. The output amplifier and a basic signal-processing chain are shown in Figure 6. A constant-current source provides a high dynamic impedance (and so maximum gain) for the output transistor. Typical gains are 0.6–0.7, giving an output sensitivity of about 1 electron = 1 \( \mu \)V. This gain is increased with increasing source current, but eventually the output transistor will run in surface-channel mode and consequently will greatly degrade the noise performance. The noise performance of the system should be set only by the noise performance of the output transistor and not degraded significantly by other aspects of the electronic system. All transistors have intrinsic internal noise sources...
Figure 6  A typical analog signal-processing chain used with slow-scan CCD detector systems. The functions of the various parts are described in the text.
that may be set equivalent to an input noise voltage. If we know the node capacitance, the equivalent input noise voltage may be interpreted as equivalent to an input charge noise level in electrons root mean square. The overall system noise level referred to the node capacitance is usually called the readout noise. It is always present, even in the absence of any input signal, and represents an irreducible minimum in the overall system noise performance.

The charges passed through the output gate are integrated by the node capacitance. In order to minimize variations in the working point of the output transistor, the node capacitor is usually recharged to a fixed potential before the next charge packet is transferred to the node capacitor. This recharging is achieved by briefly pulsing the gate of the reset transistor high to turn the reset transistor on and to connect the node capacitor to a fixed potential via the low impedance of the reset transistor. Once the reset transistor is returned to its high-impedance or "off" state, the voltage across the node capacitor is established with an uncertainty of $(kT/C)^{1/2}$. This uncertainty is $400 C^{1/2}_{\text{node}}$ electrons rms at room temperature and $250 C^{1/2}_{\text{node}}$ electrons rms at 120 K (with $C_{\text{node}}$ given in picofarads). This noise level of 80–125 electrons is very much greater than the noise of the output transistor itself. In addition, the output transistor has an intrinsic noise spectrum, as shown in Figure 7. At low frequencies it increases as $1/f$, so that the broadband noise is dominated by the low-frequency noise characteristics. The high-frequency noise level is mainly thermal noise from the channel resistance with a white (flat) spectrum.

The subsequent signal-processing chain is essential to minimizing the $(kT/C)^{1/2}$ and the low-frequency noise contributions. In Figure 6 we see that the signal from the source of the output transistor is amplified by a low-noise amplifier A1. Amplifier A4 is an integrator. It is connected to A3 for a precisely determined fixed period and then connected to ground while the charge is transferred to the output node capacitor. Then the integrator is connected to A2 for an identical period of time. Since A2 and A3 give identical output signals of opposite polarity for identical times, the integrator accurately measures the difference in the signal present at the output of amplifier A1 immediately before and immediately after the pixel charge is transferred out. This technique is known as double-correlated sampling (B. Atwood, private communication, 1985), and it effectively suppresses the $(kT/C)^{1/2}$ noise and the components of the output transistor noise much slower than the overall signal-processing time (perhaps 20–60 μs) and much faster than the individual integration periods (5–30 μs). The output of the integrator A4 is connected to a fast analog-to-digital converter (ADC).

If it is intended to work at the lowest signal levels, then the gain of the
signal chain should be chosen to make the root mean square (rms) readout noise equal to a few bits or data numbers (DN) of the ADC. The maximum signal in one pixel without reaching saturation varies from device to device but can be as high as 500,000 electrons in some P8600 devices. With an overall system readout noise of about 5 electrons rms, the vast dynamic range of 100,000:1 is often greater than can be managed by a 16-bit ADC. Significant deviations from linearity are seen near the top end of the signal range, but at lower levels it is likely to be errors in the analog signal chain that compromise linearity before the CCD itself becomes a problem.

It is important to be able to calibrate readout noise in a manner that does not depend on measurements of the properties of the various components of the signal-processing chain. The most powerful method is the variance method (Figure 8), recently described as the photon-transfer technique by Janesick et al. (1985d). For any system there are several different noise regimes as the signal level varies. At the lowest level, the

![Figure 7](image)

*Figure 7*  The noise spectrum of a P8600 CCD for two samples of the device. The solid line is noise with a $1/f$ spectrum plus a white-noise floor of 25 nV Hz$^{-1/2}$. The corner frequency is 13 kHz.
readout noise dominates (Figure 8, regime 1). However, as increasing signal charges are read out, the shot noise of the signal charge is added in quadrature with the intrinsic readout noise. Eventually the readout noise becomes negligible, and the signal-to-noise ratio goes in proportion to the square root of the signal in electrons (not data numbers; see Figure 8, regime 2). At the highest levels, the signal-to-noise ratio will rise more slowly with signal (Figure 8, regime 3). This is usually attributed to the onset of nonlinearity near saturation, although in some devices there can be a significant signal range that it is affected in this way, which implies that a more subtle explanation is probably needed.

In the above, the signal-to-noise ratio relates to a measure of repeatability in identical images as a function of signal level. We cannot use the standard deviation in the measured intensities from many different pixels, even with totally uniform illumination, because errors caused by device nonuniformities will dominate the apparent noise at the higher signal levels. So we must always use the ratios of identical images to establish the variance diagram.

From the variance diagram we can directly determine the conversion between electrons and data numbers, since the noise in regime 2 is equal to the square root of the signal level in electrons. The intercept of the plot with the noise axis at zero signal establishes the readout noise in DN and

![Diagram](image)

*Figure 8* The rms noise output from a P8600-based CCD system as a function of signal level. This diagram is sometimes called the variance plot or the photon-transfer curve.
therefore in electrons. No assumptions need be made about any part of the signal-processing chain, since these are measurements that are made purely on the actual data output to the host computer system. This is the great power of the variance method. The astronomer can learn much about the CCD system he is using by constructing parts on the variance diagram by taking pairs of flat-field exposures at several different signal levels when direct imaging. In spectroscopy, the continuum calibration lamp often gives a considerable variation in signal level, so several points in the variance curve may be determined from a single exposure, and a smaller number of exposure pairs will be satisfactory. The astronomer at the telescope should be encouraged to try these tests, since from them he can check the quality of the complete observing system, including the flat-fielding and data-reduction techniques used.

A useful cross-check on these calibrations can be made using X-ray sources in the laboratory, since we know that X rays of a few keV are detected efficiently in silicon, generating 1 electron for every 3.65 eV of X-ray energy. Although many of these X rays will be absorbed deep within the silicon, and thus the generated charge will be incompletely collected by the electrodes, a few events will be fully collected by the single pixel they illuminate. The peak signal level in single-pixel events can give another independent estimate of the electrons per data number conversion factor.

4.5 Device Uniformity

The only uniform CCD is a dead CCD. When we take an exposure with a CCD of a perfectly flat, uniform scene, we generally find that we get an image that is distinctly nonuniform. Individual pixels can be particularly high or low relative to the mean if they are defective. Other faults in the device fabrication can cause bright or dark columns (wholly or partly bad), trapped states that appear as short sections of a column that look defective. Furthermore, flat-field images at different signal levels will often show that some of these defects are nonlinear at any signal level. Generally, it is wise to abandon any attempts to extract data from significantly nonlinear pixels and columns, and ratioing of flat-field exposures taken at different signal levels is the best way to identify these areas.

Fortunately, most modern CCDs on telescopes are relatively free from these cosmetic defects, and thus we might expect to be able to use calibration images taken of uniformly illuminated sources to correct for the residual device uniformities. These nonuniformities are largely caused by variations in the collection areas of pixels or columns of pixels in the device. CCDs are made with modern integrated circuit technology, but, like all modern technologies, they are not perfect. The computer-aided design facilities that make the masks for IC manufacture as well as the
rest of the fabrication process generate small errors in geometry that give rise to small but nonnegligible variations in pixel size. Periodic patterning in the flat-field image is often the end result. Given the excellent linearity of the CCD, all one has to do is to take a uniformly illuminated image to calibrate out the nonuniformities. In practice, however, this process is much more complicated, principally because the nonuniformities are significantly color dependent. At different wavelengths the photons are typically absorbed at different depths within the device. Devices illuminated through the electrodes may have slight nonuniformities in electrode thicknesses across the device that affect response uniformity in a color-dependent way. Thinned devices in the red will be transparent enough for the electrodes on the other surface to affect the proportion of light reflected back into the device, whereas in the blue the response of thinned devices is affected by the subtleties of the surface charge distribution.

In all devices, including unthinned devices, multipath interference effects in the device windows, covering electrodes, etc., will modulate the sensitivity. With broadband illumination these effects may be negligible, but with narrowband illumination (as in spectroscopy) these effects may become serious. This phenomenon is often referred to as fringing. Thinned devices in the red, where the silicon is most transparent, can give a monochromatic flat field with nearly 100% modulation. No amount of flat fielding can help alleviate this problem.

When an astronomer takes a broadband sky-limited image, he will generally need to compensate for detector nonuniformity. Obtaining the proper flat-field calibration image is not easy. Often the inside of the telescope dome is illuminated to ensure that the telescope beam and all subsequent optics are filled in the same way as when observing the sky. With narrow filter bandwidths it is possible to match the color of the calibration source to that of the sky fairly well. If we use color correction filters, it is possible to make a tungsten lamp have roughly the correct color balance, but generally with broadband sources it is not practical to arrange that the source of flat-field illumination mimics the spectrum of the night sky. This is because the source is a combination of a hot thermal continuum plus a number of essentially monochromatic night-sky lines (Broadfoot & Kendal 1968). Furthermore, it will be fine to try to get as close as possible to the color of the sky for faint objects (where the sky light dominates), but if the light from the object dominates, the flat-field image will be wrong for that object, with colors different from that of the sky.

We must conclude, therefore, that flat fielding is at best a compromise and in no way a substitute for intrinsic device uniformity. Other methods have been used to give improved flat fields. Several workers use many of the images obtained during one night or through an observing run.
estimating the mode of the data value distribution through the set of images on a pixel-by-pixel basis, we can construct an average flat field that would show, for example, low-level sensitivity fringing and do a better job than straightforward flat fielding with a continuum source. In practice, the sky color changes throughout the night, as do the intensities and the relative intensities of the night-sky emission lines, and thus the method is not without its limitations. In practice, CCDs have intrinsic non-uniformities in response of a few percent rms. With careful flat fielding it is possible to reduce these to 0.3–0.5% rms, but it is difficult to achieve better accuracy than this with these methods. In evaluating what these limits are, it is vital that we do not confuse this true residual nonuniformity with the repeatability that can be easily obtained with CCDs. It is quite easy to make a series of exposures that give repeatability in measured flux, for example, of 0.1% rms or better. Unfortunately, the excellent repeatability neglects to include the fact that the relative fluxes of two objects will be much less well established because of residual flat-field errors.

There is one method that has been demonstrated to give flat fielding to better than 0.1% rms. This is the technique of drift scanning (Wright & Mackay 1981). The normal way of using a CCD is to open the shutter so that light falling onto the CCD creates a charge distribution in the CCD exactly in proportion to the incident light flux. The shutter is closed and the image read out. In the drift scanning method, the shutter is opened and the CCD is slowly clocked (e.g. one row every second) while light continues to fall on it. In order to avoid image smear, the CCD itself is moved mechanically, so that the light from one sky pixel continues to fall onto the charge already accumulated from that sky pixel as it is slowly moved across the device. Eventually, this sky pixel is read out, and the charge accumulated in it is detected with the mean sensitivity of all the pixels in that column. This is also true of all subsequent pixels read out in that column, provided that the corresponding sky pixels were tracked across the entire length of the device. The columns of the output image are extremely uniform along their length, although column-to-column nonuniformities remain as before. However, individual column sensitivities may be determined by computing the mode of the data values down each column and using that single number to rescale the fluxes in each column. This method has the advantage in that it uses the actual sky level during the observation to flat field the data. It has yielded sky nonuniformities of less than 0.1% rms and has given data for the faintest objects yet observed from the ground (Hall & Mackay 1984). Nevertheless, it still gives a less than perfect answer for objects of brightness comparable to that of the sky when these objects have colors different from that of the
sky. Drift scanning is a technique that can help when conventional flat-fielding methods prove to be inadequate, but this point will become harder to reach as intrinsic device uniformity improves. One other situation where the technique should help us is when narrowband images (e.g. in one emission line) are being obtained and sensitivity fringing is a problem. Drift scanning should average out these fringes with good accuracy and permit the higher quantum efficiencies of thinned devices to be exploited despite fringing effects that would otherwise give unacceptable nonuniform images.

One aspect of device uniformity that has been largely ignored is that of intrapixel nonuniformity. Wright (1982) measured P8600 type devices and found variations within a single pixel of 29% at 540 nm, 15% at 685 nm, and 17% at 943 nm. At long wavelengths, variations in reflectance are important, whereas at short wavelengths it is variations in device absorption that dominate. When an object is imaged onto a small number of pixels, the exact location of the star on the pixel grid will affect the mean detected flux. It is quite easy to get variations of a few percent simply by displacing the peak of an image by a half-pixel. Similar, though less severe, effects are probably present in thinned CCDs, since these devices are often transparent enough for the electrodes to be seen through the device under a microscope as being colored. Despite these variations (Wright 1982), no attempt has been made to look seriously at this problem insofar as it will affect precision photometric work.

Finally, we note that although most CCDs are very flat (since they are bonded to a flat window or package structure), some devices (e.g. the Texas Instruments 800 × 800 devices made for the Hubble Space Telescope Wide Field/Planetary Camera) are thin membranes suspended by their edges. They show surface nonuniformities of typically 70 µm peak to peak, so with 15-µm pixels it is not possible to focus the device uniquely in a fast beam such as is often produced by a spectrograph camera.

4.6 Hardware-Induced Problems

Although most users of CCDs accept that the devices themselves are not perfect, many are surprised to find that most of the electronic hardware systems used to drive CCDs themselves have problems that can have a significant effect on the data taken at a telescope. Ground-loop problems between dewar and driver electronics, within the driver electronics, or between the driver electronics and the computer system can cause interference with the readout electronics, giving patterns that are synchronized with the line frequency. Electrical noise from motors, light dimmers, or computer parts can be picked up by the detector system, since they need to run with an overall system noise of only a few microvolts rms. Electronic
misalignment can give greatly different numbers of odd and even numbers from the analog-to-digital converter, making the least significant bit unreliable. Electronic drift can cause background (black or bias) levels to change during the readout of one frame or over a longer period. This problem can be insured against by reading out more pixels in each row than are physically there. This process gives extra-dark pixels at the end of each row of data, which may be used to monitor variations of black level and of readout noise. It is very important that the astronomer always be on the lookout for problems of these sorts at the telescope and during data reduction. All such problems will compromise to a greater or lesser degree the quality of the data.

5. CCDs IN ASTRONOMICAL APPLICATIONS

In the above sections we have looked in detail at most of the properties of CCDs from the technical point of view. In this section we discuss briefly the principal trade-offs that must be examined when deciding whether to use a CCD system or some other detector; in addition, we look at what aspects of CCD system performance are likely to be most critical for different sorts of applications.

In making such evaluations, it is critical that we judge on the basis of the actual performance of the detector systems being considered, not on the performance claimed by their creators or keepers, who are often very proud of their offspring. One should talk to recent users to find out their experience of the systems, particularly to users who have actually tried to reduce data from the system.

Most applications can be divided into high-light-level applications, where the CCD system noise level is photon-shot noise limited, and low-light-level applications, where the CCD system is readout noise limited. At the higher light levels such as are found in broadband, direct-imaging CCD systems are generally the system of choice because of their much higher detective quantum efficiencies (DQE) than photographic or image-intensifier-based systems. Indeed, an interesting series of measurements by Cromwell (1985) suggests that nearly all intensifiers that use phosphor screens internally have only half the DQE that is measured for the first photocathode alone. The best thinned CCDs can have a peak DQE nearly 10 times that of some photon-counting systems. Nevertheless, raw DQE is not necessarily enough, even in broadband direct imaging, where there is ample light. A star image may cover 6–12 pixels, so that the readout noise per star is $6^{1/2} - 12^{1/2}$ times greater than that for a single pixel, and it is invariably the single-pixel readout noise that is quoted as an example.
RCA CCDs are often used for broadband images, and they can have a readout noise of 70 electrons rms per pixel and therefore 170–240 electrons per star image. This means that the system DQE is half that of the CCD quantum efficiency for star fluxes of 30,000–60,000 photons per exposure and lower still for the much lower flux levels often encountered. Readout noise is very important, and lower readout noise levels will often more than compensate for lower DQE figures.

At low light levels readout noise is always critical for CCD systems, particularly when readout noise per resolution element (rather than per pixel) is used. A good account of the trade-offs between various forms of imaging systems is given by Lemonier et al. (1985). In photon-counting systems at the lowest light levels, dark current is often significant in long exposure times and has the same effect on system performance as readout noise has on CCD systems.

In broadband direct imaging, much light is gathered to enable exposures to go to very faint levels against the sky or to enable good photometric accuracy to be attained. In either case the limits will often be set by residual flat-field errors, and it is this aspect of the system that has to be looked at most carefully. For spectroscopic applications it is the readout noise that will limit the overall signal-to-noise ratio in principle, and it is often said that CCDs will take over all spectroscopic applications as readout noise levels come down. This is not the whole story, since the on-chip output amplifier improvements that reduce readout noise do nothing to help change transfer efficiency. If the transfer efficiency is too poor to allow accurate sky subtraction, then fat zero–added signals have to be used (see Section 4.3 above), which can easily increase overall system noise levels unacceptably.

6. OTHER APPLICATIONS OF CCDs NOW AND IN THE FUTURE

This review has emphasized the current use of CCDs by astronomers. Most of the CCDs now in use are operated in slow-scan mode, with the photons from the sky being recorded directly in silicon. CCD systems are also at various stages of development for other applications. Recently, Bone (1985) has described the use of a CCD readout at pixel rates of several megahertz for the detection of photon events from a high-gain image intensifier. The system has a performance rather like that of TV vidicon readout systems. Similar systems have been developed in Australia (Gorham et al. 1982) and California. CCD systems are also being developed for auto-guider use. Here, the device is read out in a restricted...
format (e.g. $50 \times 50$ or $100 \times 100$ pixels) to reduce readout times so as to give a rapid response. Exposure times are short, and thus devices need be cooled much less to give negligible dark time. However, the readout noise is still present on each frame, and it is difficult to equal the performance of intensified TV vidicon–based systems.

The response of CCDs to X rays has been mentioned above. For X-ray imaging, the ability in principle to measure not only the position of an X-ray but also its energy is very exciting. In order to achieve good charge-collection efficiency (Janesick et al. 1985b), it is necessary to have devices with a good collection depth (at least 30 $\mu$m). Lumb et al. (1985) have produced deep-depletion devices that have good properties as X-ray detectors. A CCD X-ray imager is likely to be flown on the AXAF satellite.

Considerable effort is underway to produce infrared CCDs, principally for military purposes. However, the military requirements are rather different from those of ground-based astronomers, and at the time of writing, infrared CCDs are probably some way away for everyday use. Although there are many technical problems in manufacturing infrared CCDs, one of the principal ones at present is that the manufacturers with the good IR detector technologies do not seem to have the best CCD technologies available, particularly with respect to CTE.

Astronomers have been very fortunate with visible-light CCDs, since devices intended for domestic TV frame-rate applications have proved excellent for cooled slow-scan applications. There are signs that the next generations of TV sensors may be less well suited to astronomical application, since the demands for color sensors favor interline transfer and X–Y-addressable architectures, whereas it is the frame-transfer architecture that is most suited to slow-scan operation at the lowest signal levels. These trends may require astronomers and other scientific users to actively fund the CCD development programs they desire in the future. An excellent article by Collet (1985) describes the priorities of the sensor manufacturers. However, there are still manufacturers who are prepared to produce CCD sensors for the scientific market, and they have made great progress in the last few years in improving CCD performance. The CCDs now available commercially are already excellent. We can confidently expect devices with nonuniformities less than 1% peak to peak and with good charge transfer efficiency. Readout noise figures will come down to the 1–2 electrons rms level, and the quantum efficiencies should approach those of the best silicon-target vidicons for thinned devices with implanted surface treatment (Figure 1b). The best CCD electronics driver systems are also excellent, and it is clear that more and more observers will use CCDs as the detectors of choice in the future.
Literature Cited


© Annual Reviews Inc. • Provided by the NASA Astrophysics Data System