CHAPTER 22
VISIBLE ARRAY DETECTORS

Timothy J. Tredwell
Eastman Kodak Company
Sensor Systems Division
Imager Systems Development Laboratory
Rochester, New York

22.1 GLOSSARY

- $A$ area of the pixel
- $C_{PD}$ total capacitance of the floating diffusion in a CCD output
- $C_g$ gate capacitance per unit area
- $C_r$ readout line capacitance
- $J_D$ total dark current per unit area
- $J_s$ surface generation current
- $L_e$ diffusion length of electrons in silicon
- $L_p$ diffusion length of holes in silicon
- $N_A$ $p$-type dopant concentration in silicon
- $N_D$ $n$-type dopant concentration in silicon
- $N_e$ total number of electrons collected in a pixel
- $N(0)$ number of photons entering the silicon
- $N(x)$ number of photons remaining a distance $x$ below the surface
- $n_i$ intrinsic carrier concentration in silicon
- $P(x)$ probability that an electron-hole pair generated a distance $x$ from the surface will be collected before recombination
- $q$ electron charge
- $s_s$ surface recombination velocity
- $T_{int}$ integration time of light in an image sensor
- $T(\lambda)$ transmission of light
- $V_{bi}$ built-in voltage for a silicon $pn$ junction
- $W(V)$ width of the depletion layer at a given bias voltage in the MOS capacitor or junction-photodiode
22.2 IMAGING DETECTORS

\[ \alpha(\lambda) \] absorption coefficient of light

\[ \varepsilon_s \] silicon dielectric constant

\[ \mu_e \] electron mobility in silicon

\[ \mu_p \] hole mobility in silicon

\[ \tau_o \] depletion-layer lifetime

\[ \tau_p \] minority carrier hole lifetime in silicon

\[ \Phi_s \] electrostatic potential at the silicon-silicon dioxide, also called surface potential

22.2 INTRODUCTION

Since the invention of the image sensor in 1964, solid state image sensors have advanced in resolution, sensitivity, and image quality to the point where they have replaced other methods of converting visible light to electronic signals in nearly all imaging applications. There are two types of image sensors: area image sensors, which are used in cameras, and linear sensors, which are used in scanning applications. Cameras using area image sensors dominate the camcorder, video and broadcast, machine vision, scientific, and medical fields. Area image sensors for camcorder applications are typically 400,000 picture elements in resolution, 60 dB in dynamic range, and have noise levels of a few tens of electrons. Area image sensors for scientific applications may have resolutions of over six million elements, dynamic ranges exceeding 80 dB, and noise levels approaching the single electron level. Production of area image sensors exceeded eight million devices in 1993. Scanners employing linear solid-state image sensors dominate facsimile, document scanner, digital copier, and film scanner applications. Linear sensors range from 2000-element monochrome arrays with 40 dB of dynamic range used in facsimile applications to 8000 or more element trilinear arrays with 80 dB of dynamic range for high-performance color scanning applications. Production of linear CCD image sensors also exceeded eight million devices in 1993.

The steps involved in image sensing consist of (1) converting the incoming photons to charge at picture element (pixel), and (2) transferring that charge to an output amplifier and converting the charge to a voltage or current signal which can be sensed by circuits external to the sensor. The image-sensing elements will be described first, followed by readout elements. Sensor architectures for area and linear sensors will then be described.

22.3 IMAGE SENSING ELEMENTS

There are four basic types of structures which are used for image sensing: the junction photodiode, the photocapacitor, the pinned \((p \cdot n p)\) photodiode, and the photoconductor. The first three are generally fabricated in single-crystal silicon as part of the image sensor; the photoconductor is usually fabricated from amorphous silicon deposited over the image sensor. The photoconversion process begins with the absorption

* For some scientific applications in which high quantum efficiency and fill factor are essential, the silicon wafer will be thinned to 10 \(\mu m\) or less in thickness and illuminated from the backside. The frontside contains an area charge coupled device, which is used to collect the photogenerated carriers.
Absorption of light at a particular wavelength is given by:

\[ N(x, \lambda) = N(0)e^{-\alpha(\lambda)x} \]  

where \( N(x) \) is the number of photons remaining a distance \( x \) below the surface, \( N(0) \) is the number of photons entering the silicon, and \( \alpha(\lambda) \) is the absorption coefficient. The absorption coefficient is shown as a function of wavelength \( \lambda \) for single-crystal silicon, doped polycrystalline silicon, and hydrogenated amorphous silicon in Fig. 1. The absorption depth is defined as the inverse of the absorption coefficient \( d(\lambda) = 1/\alpha(\lambda) \). In single-crystal silicon the absorption depth is 0.4 μm in the blue (450 nm), 1.5 μm in the green (550 nm), and 3.0 μm in the red (640 nm). In the infrared, the absorption depth increases to 10.5 μm at 800 nm. Beyond 1100 nm, the absorption is virtually zero because the photon energy is less than the 1.1-eV silicon bandgap.

Junction Photodiode

The junction photodiode is one of the most common image-sensing elements. The physical structure and band diagram of the junction photodiode are shown in Fig. 2a for a p-type substrate. The n-type region is formed by ion implantation or diffusion of phosphorous or arsenic to a depth of 2000 to 10,000 Å into the p-type silicon. The n-type dopant region is usually graded, with the highest concentration at the surface. The gradient in n-type dopant concentration results in a gradient in electrostatic potential which accelerates photogenerated carriers (holes in the n-type region) away from the surface. This reduces loss of photogenerated carriers to surface recombination. The photodiode is typically operated with a reverse bias \( V \) of 1 to 5 volts. A depletion layer is formed between the n- and p-type regions. The width of the depletion layer \( W(V) \) for an abrupt \( n + p \) junction is given by:

\[ W(V) = \sqrt{\frac{2\varepsilon(V + V_b)}{qN_A}} \]  

FIGURE 2 Cross-sectional diagrams and band diagrams for (a) junction photodiode; (b) surface-channel MOS capacitor; (c) buried-channel MOS capacitor; (d) pinned, or hole-accumulated, diode; (e) amorphous silicon photoconductor.
where \( q \) is the electronic charge, \( \varepsilon_s \) is the silicon dielectric constant, \( N_A \) is the \( p \)-type dopant concentration, and \( V_{bi} \) is the built-in voltage given by

\[
V_{bi} = \frac{kT}{q} \ln \frac{N_A}{n_i}
\]

where \( n_i \) is the intrinsic carrier concentration. For silicon doped at \( 1 \times 10^{16} \text{ cm}^{-3} \), the depletion width at 5-V reverse bias is 0.8 \( \mu \text{m} \).

If the diode is illuminated, some of the photons will be absorbed in the \( n \)-region, some in the depletion layer, and the remainder in the \( p \)-type substrate. The quantum efficiency \( \eta(\lambda) \) is the ratio of the charge collected to the number of photons incident on the diode (i.e., 100-percent quantum efficiency refers to one electron-hole pair collected for every incident photon). The quantum efficiency depends on three factors: transmission \( T(\lambda) \) of light through the overlying layers into the silicon, absorption of light in the silicon, and the probability \( P(x) \) that an electron-hole pair generated a distance \( x \) from the surface will be collected before recombination:

\[
\eta(\lambda) = T(\lambda) \int_{x=0}^{x=\infty} (1 - \exp^{-n(\lambda)x})P(x) \, dx
\]

The transmission of light through the overlying layers into the silicon can be calculated using standard multilayer interference models.

The collection of the photogenerated charge takes place by two processes: drift and diffusion. Drift is the movement of electrons and holes due to an electric field. Even for small electric fields, transport of carriers by drift will dominate diffusion. This is the case in the depletion region. Outside the depletion region, such as in the \( p \)-type substrate, there is no electric field, and carrier transport occurs by diffusion. For example, a photon absorbed in the \( p \)-type region will excite an electron from the valence band to the conduction band. The electron will move in a three-dimensional random walk until it recombines or until it encounters the edge of the depletion region, where it is swept across the junction by the electric field. The probability that an electron a distance \( x' \) from the junction can diffuse to the junction before recombining is given by:

\[
P(x') = \exp^{-x'/L_e} \quad \text{where} \quad L_e = \sqrt{\frac{kT}{q} \mu_e \tau_e}
\]

in which \( L_e \) is the diffusion length, \( \mu_e \) is the electron mobility, and \( \tau_e \) the electron lifetime (typically \( \sim 1 \mu \text{s} \)). For a 1-\( \mu \text{s} \) lifetime, the electron diffusion length in \( p \)-type silicon is 50 \( \mu \text{m} \). Electrons generated from photons absorbed less than the diffusion length from the junction have a high probability of collection.\(^6\)\(^\dagger\) Similarly, photons absorbed in the \( n \)-type

---

\( ^6 \) In image sensors, the collection of photogenerated carriers can be complicated by a variety of factors. The doping concentration may not be uniform on either the \( n \)- or \( p \)-sides. On the \( n \)-side, the dopant concentration is designed to decrease from the surface to the junction, building in a potential gradient for holes away from the surface and toward the junction, preventing surface recombination. On the \( p \)-side, the dopant concentration may not be uniform owing to the use of epitaxial layers or wells diffused into the silicon. Additionally, the carrier lifetime may not be uniform. Impurity gettering, a process used in many image sensors to remove metallic contaminants during processing, will leave a region of crystalline defects in the silicon starting 20 to 50 \( \mu \text{m} \) beneath the silicon surface. The defects result in a very short electron lifetime in this region. Finally, diffusion takes place in three dimensions; a carrier absorbed beneath a given pixel in an array will diffuse laterally by the same amount it diffuses vertically. This could cause it to be collected in adjacent pixels.

layer create electron-hole pairs. The holes must travel by diffusion to the junction to be collected. The probability that a hole a distance $x'$ from the junction can diffuse to the junction is given by

$$P(x') = e^{-x'/L_p}$$

where $L_p = \sqrt{\frac{kT}{q\mu_p\tau_p}}$.

In which $L_p$ is the diffusion length, $\mu_p$ is the hole mobility, and $\tau_p$ is the hole lifetime. For a 1-μs lifetime, the hole diffusion length in n-type silicon is 30 μm. Since the n-type region in an $n+p$ junction is less than 1 μm thick, the holes have no difficulty diffusing through the n-type region to the junction unless the n-type region is so heavily damaged or so heavily doped that the hole lifetime is very short. More typically, loss of quantum efficiency on the n-side results from recombination at the surface.

The quantum efficiency of a junction photodiode is illustrated as a function of wavelength in Fig. 3. In the ultraviolet, the light is absorbed very near to the surface and some of the photogenerated charge can be lost to surface recombination. In the 420- to 700-nm region, most of the light is absorbed close to the junction and is easily collected. The structure in the quantum efficiency illustrated in Fig. 3 results from the multilayer reflections of light in the oxide layer which overlies the photodiodes used in image sensors. The positions of the minima and maxima depend on the thicknesses and induces of refraction of the layers overlying the silicon. Beyond 800 nm, some of the photons are absorbed sufficiently deep in the silicon that the carriers recombine before they can diffuse to the junction; this results in the decrease in quantum efficiency at longer wavelengths.

In most image sensing applications, the junction diode at each picture element is used not only to collect the photogenerated carriers but also to store the carriers until they can be read out. In an imaging array, each photodiode would be reset to a reverse bias $V_r$ by a MOS gate. The capacitance of the diode of area $A$ for an abrupt $n+p$ junction is given by $C(V) = \varepsilon e A / W(V)$, where $W(V)$ is the depletion width.

When a photogenerated carrier is collected by the junction, it is stored on the junction until it is read out. Storage of a carrier will cause the voltage on the junction to decrease by $q/C(V)$. When the photogenerated charge is removed from the junction during readout, the junction voltage is restored to its original value.

If so much photogenerated charge is stored on the photodiode that the voltage drops to zero before the charge can be read out, additional charge cannot be collected and will diffuse into the $p$-type substrate. This condition is referred to as saturation. The diffusion of excess charge into neighboring photosites is called blooming.
One of the difficulties encountered in using the junction photodiode in image sensor applications is image lag. The combination of the capacitance of the photodiode and the channel resistance of the MOS transfer gate used to read out the diode give rise to a time constant for transferring the photogenerated charge from the diode to the readout structure. As a result, not all the charge can be completely drained from the diode during the short reset times typically used in imaging applications. The remaining charge is drained in successive readouts, causing an afterimage. This effect is called image lag.3

MOS Capacitor

The MOS capacitor consists of the silicon substrate (taken to be p-type in this section), a thin layer of silicon dioxide (typically 200 to 1000 Å thick), and an electrode (typically polycrystalline silicon doped heavily n-type with phosphorous). The physical structure and the band diagrams of the surface channel MOS capacitor are shown in Fig. 2b for a p-type substrate. If the gate of the MOS capacitor is biased positive, a depletion layer is created in the p-type silicon substrate. The depth of the depletion layer depends on the substrate doping, gate voltage, and oxide thickness. The calculation of the depth of the depletion layer is somewhat more complex than the photodiode* and depends on the electrostatic potential at the surface, called the surface potential $\Phi_s$. For values typical of image sensors ($N_d = 1 \times 10^{19}$, 5-V gate bias, 500-Å oxide thickness) the depletion layer is 2.4 $\mu$m deep. On the edges of the photocapacitor (see Fig. 2b) is a heavily doped p-type region overlaid by a thick (2000 to 5000-Å) oxide layer. This is called the field or channel stop region. Because of the heavier p-type doping, the field is not depleted by the voltage on the gate. The channel stops confine the electrons to the channel region.

If the MOS capacitor is illuminated, a fraction of the light will be reflected, a fraction will be absorbed in the polysilicon, and the rest will be transmitted into the silicon substrate. The absorption coefficient of heavily doped polysilicon is shown as a function of wavelength in Fig. 1. The absorption coefficient is $4 \times 10^4$ cm$^{-1}$ at 450 nm and $1.2 \times 10^4$ cm$^{-1}$ at 550 nm. For a 3000-Å-thick polysilicon electrode, less than 30 percent of the blue light and less than 70 percent of the green light is transmitted through the polysilicon. The photons entering the silicon are absorbed, either in the depletion layer of the MOS capacitor or in the undepleted p-type silicon beneath the depletion layer. Those photogenerated electrons created in the undepleted p-type region move by diffusion until they are captured by the depletion layer or until they recombine. The electrons are held at the silicon-SiO$_2$ interface, where they remain until they are read out. The quantum efficiency as a function of wavelength is illustrated in Fig. 3. The efficiency is low in the blue owing to absorption in the polysilicon. The structure in the quantum efficiency as a function of wavelength is due to multilayer interference in the polysilicon-oxide-silicon stack.

Charge is stored in the MOS capacitor at the silicon-silicon dioxide interface as a layer of sheet-charge only a few hundred angstroms thick. As additional photogenerated charge is added, the surface potential decreases. If sufficient photogenerated charge is added, the surface potential becomes zero and no additional charge can be stored. This condition is saturation.

The capacitance per unit area on which photogenerated charge is stored is the parallel capacitance of the oxide and the depletion layer:

$$C^{-1} = \left( \frac{1}{\varepsilon_{ox}} + \frac{W(\Phi_s)}{\varepsilon_a} \right)$$

To determine the depletion depth, the surface potential $\Phi_s$ must be calculated. $\Phi_s$ depends on the voltage on the gate of the MOS capacitor, the oxide thickness, substrate doping, and weakly on temperature. See Sze, Physics of Semiconductor Devices.
where the depletion width is $W(\Phi_s)$.

$$W(\Phi_s) = \frac{2e_s \Phi_s}{qN_A}$$  \hspace{1cm} (8)

In nearly all cases, the oxide capacitance is the dominant term. As a result, the storage capacity of the MOS capacitor is significantly larger than the junction diode in which the charge is stored only on the depletion capacitance.

A variant of the surface-channel photocapacitor is the buried-channel photocapacitor. The structure and band diagram are shown in Fig. 2c. In this device, a lightly-doped $n$-type region is diffused or implanted into the silicon surface early in the fabrication process. This $n$-type region is sufficiently lightly doped that it is fully depleted. The $n$-type dopant in the buried-channel results in a band diagram with a potential minimum, or well, for electrons just below the surface. This well is separated from the surface by a few thousand angstroms in distance and about 1 V in potential. When the buried-channel photocapacitor is illuminated, the electrons collect in the buried channel and do not contact the surface. The primary purpose of the buried channel is to prevent electrons from being trapped by interface states at the silicon-silicon dioxide interface.

The photocapacitor has several advantages over the junction photodiode. These include higher storage capacity per unit area, zero lag readout, and generally lower dark current. The principal disadvantage is the low quantum efficiency in the blue. In some applications, a transparent electrode, such as indium-tin-oxide (ITO) may be substituted to improve the blue response. ITO has very low absorption over the visible (420 to 750 nm) and can be deposited sufficiently conductive for use as a gate electrode in an image sensor. Materials and processing complexity have prevented incorporation of ITO into commercial image sensors in volume until recently.

Another approach to achieving high quantum efficiency is the thinned backside-illuminated charge-coupled device (CCD). In this approach, the CCD (which is an array of MOS capacitors) is fabricated on the frontside of a silicon wafer. The wafer is then thinned from the backside to 10 μm or less in thickness. The backside is passivated to prevent surface recombination. Photons entering the backside are absorbed in the silicon beneath the MOS capacitors (usually buried channel). The photogenerated carriers diffuse to the capacitors, where they are held until they are read out. This device has quantum efficiency similar to the photodiode in the visible. However, because the silicon is thin, some of the photons at wavelengths beyond 700 nm will not be absorbed and so the quantum efficiency falls off beyond 700 nm. Owing to their extreme complexity in process and their extremely fragile design, backside-illuminated image sensors are limited to special scientific applications, especially astronomy.

### Pinned Photodiode

The third type of photosensitive element is the pinned ($p^+ n p$) photodiode. This is sometimes called the hole accumulation diode, or HAD. This element combines the best features of the photodiode and photocapacitor, offering the high blue response of the photodiode with the high charge capacity, zero lag, and low dark current of the buried-channel photocapacitor. The pinned photodiode consists of a very shallow (<2000 Å) $P^+$ layer overlying an $n$-type buried-channel region. The structure and band

---

diagrams are shown in Fig. 2d. The $p^+$ surface layer, which contacts the $p^+$ channel stop region on the sides, holds the electrostatic potential at the surface at 0 V. When the photodiode is illuminated, the photogenerated electrons are held in the $n$-type buried-channel region just below the surface.

The quantum efficiency of the pinned photodiode is nearly identical to that of the photodiode shown in Fig. 2. Because there is no overlying polysilicon electrode, the blue response is very high, similar to that of the photodiode. Because the buried-channel region can be completely emptied, the pinned photodiode does not have the lag of a normal junction ($n^+p$ or $p^+n$) photodiode. The pinned photodiode is becoming the most widely used image-sensing element in interline area CCDs used for camcorders and for industrial and medical cameras. The pinned photodiode is also used in some linear image sensors, particularly where low image lag is critical.

Photoconductor

The last type of photosensitive device is the photoconductor. The most common material for the photoconductor is hydrogenated amorphous silicon, although other material systems have been explored. Amorphous silicon photoconductors have been used for two types of image sensors: area image sensors, where it is deposited on top of an area array to improve fill factor (i.e., the proportion of the picture element which is photosensitive), and contact linear sensors, where it is deposited on large ceramic or glass substrates to fabricate very long line or linear arrays.

The structure of an amorphous silicon photoconductor on a CCD image sensor and the corresponding band diagrams are shown in Fig. 2e. The hydrogenated amorphous silicon photoconductor consists of a back electrode, an undoped amorphous silicon layer approximately 1 μm thick, and a transparent top electrode. Additional doped amorphous silicon or silicon nitride layers may be added to the amorphous silicon front or back surfaces to improve performance. Photons absorbed in the amorphous silicon generate electron-hole pairs. Photogenerated electrons and holes are quickly swept to the back and front electrodes, respectively, because of the high electric field in the photoconductor. When the amorphous silicon is used as part of an area image sensor, the electrons on the back electrode can be transferred into the readout element; when used as part of a contact linear array, the voltage change can be amplified and read out through a multiplexer.

The quantum efficiency of an amorphous silicon photoconductor is shown in Fig. 3. Owing to the wider bandgap of the amorphous silicon, photons of wavelength greater than about 650 nm are not absorbed. The advantage of the amorphous silicon is the high quantum efficiency across the visible wavelengths and the ability to fabricate devices either on top of area CCDs for higher fill factor or to process on glass or ceramic for very large linear sensors. The disadvantages include charge trapping at defects in the amorphous silicon and low carrier mobility, both of which lead to field-dependent nonlinear response and image lag when used in an image sensor. Recent improvements in material and device technology have mitigated many of these disadvantages at the expense of process and device complexity.

Antiblooming in Charge-sensing Elements

Blooming in image sensors occurs when the charge generated in an image-sensing element exceeds its capacity. If no method is provided to remove this excess charge, it will be injected either onto the readout element (CCD or MOS readout line) or into the substrate. If the excess charge is injected onto the readout element, it will usually appear as a bright line in the image. If it is injected into the substrate, the charge can diffuse in a circular pattern and be collected by neighboring elements.
There are two basic types of antiblooming circuits: lateral and vertical.\textsuperscript{8,9} In lateral antiblooming, illustrated in Fig. 4a and Fig. 4b, an MOS antiblooming gate and an antiblooming drain are provided adjacent to each image-sensing element. Excess charge on the sensing element overflows the antiblooming gate onto the antiblooming drain. The antiblooming drains of all elements on the array are connected and the current sunk in a bias supply.

In a vertical antiblooming structure, illustrated in Fig. 4c and 4d for a pinned photodiode sensing element, the image-sensing element is fabricated in a shallow, lightly doped $p$-well. A large (10- to 30-V) bias is applied to the $n$-type silicon substrate, causing the $p$-well underneath the photodiode to become completely depleted. Once the charge on the diode exceeds its capacity, the excess charge flows over the saddle-point in the $p$-well, \textsuperscript{8} Other types of antiblooming are occasionally used in image-sensing arrays. One of these is charge pumping, in which an MOS gate is repeatedly clocked in order to cause excess charge held underneath it to recombine at interface states at the silicon-silicon dioxide interface. In charge pumping, the MOS gate is pulsed sufficiently negative to cause accumulation, resulting in the interface states filling with holes. When the gate is pulsed out of accumulation, excess electrons can recombine with the holes.
under the diode, and into the substrate. The substrate is connected to a bias supply which sinks the blooming current. The vertical antiblooming structure has the advantage of requiring no additional silicon area, so that antiblooming is achieved with no reduction in fill factor. Lateral antiblooming, on the other hand, requires additional silicon area in each pixel, reducing fill factor. The vertical antiblooming has the disadvantage of lower quantum efficiency at wavelengths longer than about 500 nm. Because any light absorbed below the photodiode or photocapacitor is drained into the substrate, the quantum efficiency of photodiodes or photocapacitors with vertical antiblooming is reduced.

**Dark Current in Photosensing Elements**

Signal in photosensing elements is the result of collection of electron-hole pairs generated by the absorption of light. However, charge is generated at each photosensing element even in the absence of light. This generation is called dark current and is a result of thermal generation of electron-hole pairs. The thermal generation occurs at defects, such as impurities or crystalline defects, in the bulk of the silicon and at surface states at the silicon-silicon dioxide interface.

There are four sources of dark current: (1) diffusion current, which is the thermal generation of carriers in the undepleted n- and p-type regions; (2) depletion layer generation current, which occurs in the depletion layer of a diode or MOS capacitor; (3) surface generation current, which is the generation of electron-hole pairs at interface states at the Si–SiO₂ interface; and (4) leakage, which refers to generation at extended defects such as impurity clusters or stacking faults, particularly in the presence of a large electric field. These sources of dark current are illustrated for a photodiode and a photocapacitor in Fig. 5. The generation of the electron-hole pairs in both diffusion current and depletion-layer generation-recombination current occurs almost exclusively at impurity sites. Impurities with energy levels near midgap, such as gold, copper, and iron, are

![FIGURE 5](image-url)
particularly effective in the thermal generation of charge. The depletion-layer generation current is given by:

\[ J_g (V) = \frac{q n_i W(V)}{\tau_0} \]

(9)

where \( W(V) \) is the width of the depletion layer at a given bias voltage in the MOS capacitor or junction-photodiode, \( n_i \) is the intrinsic carrier concentration, and \( \tau_0 \) is the depletion-layer lifetime, for carriers. Typical values of \( \tau_0 \) in clean MOS processes would be 1 to 10 ms and values of \( J_g \) would be 30 to 300 pA/cm².

The diffusion current generation current for a \( p \)-type region is given by:

\[ J_{\text{diff}} = \frac{q n_i^2 L_e}{\tau_e} \frac{e q}{kT} \]

(10)

Since the intrinsic carrier concentration \( n_i \) depends on temperature as \( e^{\frac{E_g}{2kT}} \), depletion-layer generation current and diffusion current will have different temperature dependences. Depletion-layer generation current will increase as \( e^{\frac{E_g}{2kT}} \), which corresponds to a doubling in dark current for every 9 to 11°C increase in temperature near room temperature. Diffusion current will increase as \( e^{\frac{E_g}{kT}} \) which corresponds to a doubling every 4.5 to 5.5°C increase in temperature.

The surface generation current \( J_s \) occurs almost exclusively at regions where the depletion layer intersects the Si-SiO₂ interface, such as the surface region between the \( n^- \) and \( p^- \)-regions around a photodiode or in the depleted surface under a MOS capacitor. It is given by:

\[ J_s = \frac{q n_i s_o}{2} \]

(11)

where \( s_o \) is the surface recombination velocity. A typical value of \( J_s \) for a MOS surface would be 100 pA/cm²."
high electric field, only a very slight dependence on temperature. In an image sensor, they are visible as “bright spots” in a few isolated pixels against the otherwise low level of background thermal generation.\textsuperscript{10}

Values for the dark current vary widely because of variation in the amount of impurities in the silicon; the dark current can range from 0.01 nA/cm\textsuperscript{2} in very high quality image sensors to >10 nA/cm\textsuperscript{2} in sensors with significant metallic contamination. The total number of electrons $N_e$ collected in a pixel is:

$$N_e = J A T_{\text{int}}/q$$ \text{(12)}

where $J$ is the total dark current per unit area, $A$ is the area of the pixel, and $T_{\text{int}}$ is the integration time. For a 1/2-in format CCD such as would be used in a camcorder, typical values would be a dark current of 0.5 nA/cm\textsuperscript{2}, a pixel area of 100 $\mu$m\textsuperscript{2}, and an integration time of 1/30 s; the number of thermally generated electrons would be 105 electrons per pixel. Image sensors developed for scientific purposes might have a dark current 5 to 10 times lower at room temperature. These same devices might also be operated below room temperature to reduce the thermal generation to levels below one electron per pixel.

There are two types of noise associated with the charge generated by the dark current: shot noise and pattern noise. The shot noise due to the dark current is the square root of the number of dark electrons in a pixel. Pattern noise is due to pixel-to-pixel variations in the dark current and is often highly correlated between neighboring pixels. A numerical value for the dark pattern noise is typically obtained by using the standard deviation of the pixel values in the dark from a large region of an imager, where the values are obtained by averaging over many frames to eliminate shot noise and other temporal noise sources.\textsuperscript{*}

\textbf{22.4 READOUT ELEMENTS}

The readout element transfers the charge from the image-sensing element (photodiode, photocapacitor, or photoconductor) to the output of the image sensor. In a linear sensor, the readout is only in one direction. In an area sensor, both $x$ and $y$ readout is required. There are two basic types of readout elements: charge-coupled devices, or CCDs, and $x$-$y$ addressed photodiode arrays (typically called MOS arrays owing to the MOS transistors used in the addressing of the pixels). CCDs are by far the most widely used readout elements owing to their very low noise. Nearly all consumer camcorders, facsimile machines, scanners, copiers, and professional and scientific cameras utilize CCDs. Applications of MOS arrays are largely restricted to those where addressing of an individual pixel or subarray is required.

\textbf{Charge-Coupled Device (CCD)}

\textbf{CCD Operation.} The CCD works by moving packets of charge physically at or near surface of the silicon from the image-sensing element to an output, where the charge packet is converted into a voltage. The CCD is formed by an array of overlapping MOS capacitors.\textsuperscript{11-13} There are a number of different types of CCD, including four-phase, three-phase, two-phase, virtual (single-) phase, and ripple-clocked CCDs. The number of phases refers to the number of separately clocked elements within a single stage of the

\textsuperscript{*} A histogram of the dark current values of the pixels is rarely gaussian. In most cases, it exhibits an extended tail at high dark current values due to pixels with crystalline defects. The histogram may also exhibit quantization due to pixels with integral numbers of impurities (i.e., 1, 2, or 3 gold atoms); the quantization may even be used as a signature of the impurity present. See, for example, McColgin et al., “Effects of Deliberate Metal Contamination on CCD Image Sensors,” Materials Research Society Symposium Proceedings, 262:769 (1992) and “Dark Current Quantization in CCD Image Sensors,” Proc. 1992 International Electron Device Meeting, Washington, D.C., 113–116 (1992).
CCD. These will be described later. For understanding the principle of charge transfer in a CCD, consider the four-phase CCD illustrated in Fig. 6a.

The four-phase CCD physically comprises a silicon substrate (assumed to be p-type for purposes of illustration), a gate oxide 300 to 1000 Å thick, and overlapping polysilicon electrodes 1000 to 5000 Å thick which have been heavily doped with phosphorus to lower their resistance. For a four-phase CCD, two levels of electrode are required. The first level is deposited, then defined photolithographically to form two phases ($f_1$ and $f_3$). A thin (500-Å) oxide is grown over the first level of polysilicon to insulate it from the second level. The second level of polysilicon is then deposited, doped with phosphorus, and defined to form the other two phases ($f_2$ and $f_4$). An electron micrograph of a four-phase CCD with six-micron long gates is shown in Fig. 6b.

The process of charge transfer in a four-phase CCD is illustrated in Fig. 6c. In order to hold a packet of electrons, two adjacent gates ($f_2$ and $f_3$, for example) would be held at a high positive potential (+5 V) while the other two phases would be held at a low potential (−0 V). A depletion layer, or well, is formed under $f_2$ and $f_3$, allowing electrons to be held at or below the surface. The other two phases, $f_1$ and $f_4$, serve as potential barriers, keeping the charge packet under $f_2$ and $f_3$. To transfer the electrons through the silicon, the electrode ahead of the charge packet ($f_4$) is clocked positive and the electrode behind ($f_2$) is clocked negative. The electrons move along the silicon surface following

**FIGURE 6** (a) Cross-sectional diagram of unit cell of a four-phase CCD; (b) scanning electron micrograph of a unit cell of a four-phase CCD; (c) illustration of charge transfer along a four-phase CCD, showing both transfer along the register and at the sensor output.
the positive potential. This is repeated through all four phases, during which the charge packet is moved forward one pixel.

The CCD may be either surface-channel or buried-channel. In a surface-channel CCD (Figs. 2b and 6a) the electrons are held at the silicon surface. Surface-channel CCDS are rarely used owing to the trapping of electrons at interface states at the silicon surface. At the silicon-SiO₂ interface there is a density of states of about \(1 \times 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}\). These states can trap electrons from one charge packet and reemit the electrons into a later charge packet. This results in transfer efficiency. In the buried-channel CCD (Figs. 2c and 6d), a lightly doped n-type layer is formed in the silicon. This n-type layer results in a potential well for electrons just below the surface rather than at the surface (Fig. 2b). As a result, the electrons remain separated from the interface and are not trapped by interface states. This results in the ability to transfer charge from one stage to another with very high efficiency. Virtually all CCDs for image-sensing applications utilize buried-channel CCDs. The clock voltages used to drive CCD gates typically swing by 5 to 8 V between high and low levels.

For buried-channel CCDs, transfer rates of up to 20 megapixels per second can usually be achieved without special design considerations. Above this rate, special attention must be paid to optimizing the electric field along the direction of transfer to speed up charge transfer. Transfer rates exceeding 50 MPix/s have been achieved in optimized CCD designs.¹⁴

**CCD Output.** At the output of the CCD is a circuit to convert the charge packets into a voltage signal. By far the most common type of output circuit is the floating diffusion with source follower amplifier. The floating diffusion output is shown schematically in Fig. 7a and a photograph of the end of a CCD shift register with the first stage of the amplifier is shown in Fig. 7b. The floating diffusion output consists of an output gate (OG), a floating diffusion, a reset gate (RG), and a reset drain. The floating diffusion is an n⁺-type region in between the output gate and the reset gate. The floating diffusion is connected to the gate of a source-follower amplifier. The output gate is held at a fixed dc voltage, creating a barrier potential over which the packet of electrons can be transferred onto the floating diffusion when the last phase of the CCD register is clocked to its low (−0 V) voltage.

The sequence of events in the CCD output is shown in Fig. 6c. As the charge packet is transferred along the CCD, it arrives at the last phase (ϕ4) before the output gate (time \(t_2\) in Fig. 6c). When ϕ4 is clocked low (time \(t_3\)), the packet of electrons is transferred over the output gate onto the floating diffusion (time \(t_0\)). The voltage of the floating diffusion changes by an amount

\[ V = Nq/C \]  

where \(N\) is the number of electrons and \(C\) is the total capacitance of the floating diffusion itself, the interconnect to the source-follower amplifier and the input capacitance of the amplifier. In typical CCDs, this capacitance would be in the 10-1F to 50-1F range. Because this capacitance is so small, there is a large change in voltage for a small change in charge.

The charge-to-voltage conversion is an important parameter in CCD design; for a 10-1F capacitance the charge-to-voltage conversion is 16 microvolts per electron. After the change in voltage has been sensed by the amplifier, the charge packet must be removed from the floating diffusion before the next packet arrives. This is achieved by clocking the reset gate positive (time \(t_1\) in Fig. 6c), allowing the charge packet to flow from the floating diffusion to the reset drain. The reset drain is held at a constant positive voltage, typically −10 V. The reset drain is then turned off and the floating diffusion is prepared to accept the next packet of electrons.

The change in voltage from the floating diffusion is typically buffered on-chip by a source-follower (Fig. 7c). A two-stage source follower is most often used. The first stage utilizes very small-dimensioned FET transistors in order to minimize the input capacitance. The second uses much larger FET transistors in order to achieve sufficient drive current to
overcome off-chip capacitances such as package and lead capacitances on the circuit board. The source-follower amplifier is typically designed to have a bandwidth on the order of ten times the CCD pixel rate. For high-pixel-rate applications (>10 Mpix/s), three-stage source-follower amplifiers are employed.

For special purpose applications, such as CCD signal processing, nondestructive readout is required. Floating gate output amplifiers are used in these applications. These amplifiers are similar to the floating diffusion except that the floating diffusion is replaced by a MOS gate which is connected to the MOS amplifier. Other outputs, such as buried-channel JFET structures, have seen limited use in very low noise applications.15

Types of CCDs. In addition to the four-phase CCD described here, there are a number of other types of CCD shift registers, including three-phase, two-phase, and virtual phase. The different types are illustrated in Fig. 8. The four-phase CCD (Fig. 8a) was described previously. The three-phase (Fig. 8b) consists of three different layers of polysilicon electrodes. The charge is normally held under one of the three; during transfer, the gate ahead of the charge packet is clocked positive and the gate holding the charge is clocked negative in order to transfer the charge one gate ahead. The three-phase CCD has the advantage of a shorter unit cell than the four-phase but at the expense of additional processing complexity (i.e., a third polysilicon layer).

In the two-phase CCD, each phase has a barrier and a well region. The barrier is
VISIBLE ARRAY DETECTORS

FIGURE 8 Types of CCD registers: (a) four-phase; (b) three-phase; (c) pseudo-two-phase; (d) true two-phase; (e) virtual phase. In the pseudo-two-phase CCD, each phase consists of two polysilicon gates, one of which is offset in potential from the other by an implanted threshold adjustment. In the true two-phase CCD, each phase consists of a single polysilicon gate in which the implanted threshold adjust region is formed underneath a portion of the gate.

formed by doping the barrier region slightly less n-type than the well region, making the electrostatic potential in the barrier region a few volts lower than the well region for the same gate voltage. Electrons will flow over the barrier region and be held in the well region. There are two methods of fabricating a two-phase CCD. In the first method (Fig. 8c), two separate gates are used for each phase; one gate receives a threshold adjust ion implantation in order to create the barrier region. The two gates, however, are connected and thus driven at the same voltage. In the other method (Fig. 8d), the barrier and well regions are created under a single polysilicon gate. The two-phase CCD is very commonly used for three reasons. First, it requires only two clocks (\( f_1 \) and \( f_2 \)), simplifying drive circuitry. Second, the clocks are complementary. This reduces clock feedthrough. Third, the two-phase has a higher horizontal density, especially the two-phase with barrier and well regions under the same gate.

The virtual phase CCD\(^{16,17} \) (Fig. 8e) consists of one gate with both barrier and well
regions within it and a second region, the virtual phase, in which a shallow, high-dose ion implantation is used to create a heavily doped surface region which pins the surface potential at 0 V. The virtual phase has both barrier and well regions within it and acts the same as a polysilicon gate held at a constant voltage. Charge is first transferred from the clocked phase into the virtual phase, then the charge is transferred from the virtual phase into the next clocked phase. The virtual phase CCD has a number of advantages, including higher quantum efficiency than two-polysilicon-level area CCD sensors and simpler clocking. The disadvantages of the virtual phase include the need for larger voltage swings on the single-clock and high-clock feedthrough into the output due to the lack of complementary clocks.

**CCD Characteristics.** The four major performance parameters of a CCD shift register and output amplifier are charge-handling capacity, charge-transfer efficiency, charge-to-voltage conversion ratio, and noise. The charge capacity is the number of electrons which can be held and transferred in the CCD shift register. As charge is added to a shift register, a point is reached at which excess charge cannot be held; the excess charge either overflows into adjacent pixels or overflows into the bulk beneath the CCD or, in the case of a buried-channel CCD, overflows the barrier to the Si-SiO₂ surface. The charge capacity is a function of the device design, device layout, and CCD process. Figure 9a shows the electrostatic potential and charge distribution in a buried-channel CCD at three levels of charge: approximately one-quarter of saturation, at saturation, and beyond saturation. Below saturation, the electrons fill the center of the buried channel in the region of largest potential, separated from the surface by approximately 0.2 μm in distance and about 500 mV in potential. As additional charge is added, the electron distribution spreads towards the surface and the potential barrier to the surface drops. Beyond saturation, the electrons contact the surface directly, resulting in charge-transfer inefficiency and blooming to neighboring pixels. The charge capacity of most CCD shift registers is of the order 1 × 10¹² electrons/cm² of area in which the charge is held. In most CCD cells, the charge is held in only a fraction of the total cell area both along the CCD register and across the register. Typically, area CCD image sensors are designed with CCD charge capacities in the range of 50,000 to 200,000 electrons. Linear CCD image sensors, because of the larger amount of silicon area available for the CCD shift register, often are designed for 100,000 to 1,000,000 electrons.

The second major performance parameter for CCD shift registers is charge transfer efficiency. The transfer of charge from one stage to the next is neither instantaneous nor complete, limiting both transfer rate and the total number of stages in the CCD. There are two intrinsic mechanisms governing charge transfer: drift and diffusion. Drift is the movement of charge in the presence of an electric field. There are two origins of the electric field seen by an electron during charge transfer: the self-induced field resulting from the other electrons under the gate and the externally induced, or fringing, field. During the early stages of charge transfer, the self-induced field is large and is the dominant factor. After the charge concentration under the gate has decreased to a low value, the remainder of the charge transfer will be governed either by diffusion or by drift due to externally induced, or fringing, fields.

For a surface-channel CCD, the self-induced fields can be estimated from the formula:

\[
E = -\frac{q}{C} \frac{dN}{dx}
\]  

where \( C \) is the gate capacitance per unit area and \( N(x) \) is the density of electrons as a function of distance \( x \) from the edge of the electrode. In the early stages of charge transfer, both \( N \) and \( dN/dx \) are large. As the transfer proceeds, both \( N \) and \( dN/dx \) become small and the charge transfer is governed by fringing fields and diffusion.

Fringing fields are due to the two-dimensional nature of the electrostatic potential. If
the charge is being transferred from gate 1 to gate 2, the potential will change smoothly between the two gates. The effect of the potential from one gate will typically extend 1 to 3 \( \mu \text{m} \) into a neighboring gate. The charge within the range of the fringing field will move by drift to the neighboring gate. Charge out of the range of the fringing fields will move by diffusion.

Figure 9b shows an example of charge transfer calculated for charge transfer from one 8-\( \mu \text{m} \)-long CCD gate into an adjacent gate. The charge density is shown as a function of distance along the CCD at several times after the start of charge transfer. At the start of the transfer, all the charge is under the first gate. At 0.01 ns into the transfer, the charge has moved from the edge of the first gate into the second gate, as a result of self-induced...
drift. By 0.2 ns, approximately half the charge has been transferred. By 0.7 ns, over 90 percent of the charge has moved into the second gate, leaving a residual in the first. At this point, the self-induced drift is sufficiently small that drift due to fringing fields and diffusion are the dominant mechanisms. Figure 9c shows the charge transfer inefficiency as a function of time for this example. Two slopes are evident in the transfer inefficiency. In the first 0.5 ns, the charge is transferred rapidly owing to the self-induced drift. For times longer than 0.7 ns, the transfer is due to fringing fields in this example.

For CCDs with longer gates or lower fringing fields than the example above, the final ~10% of the charge must transfer by diffusion. Charge transfer by diffusion follows an exponential time dependence.

\[ N(t) = N(0)e^{-t/\tau_{\text{diff}}} \]  

(15)

where the time constant \( \tau_{\text{diff}} \) for diffusion is

\[ \tau_{\text{diff}} = \frac{4L_g^2}{\pi^2D} \]  

(16)

where \( L_g \) is one gate length and \( D = KT\mu_e/q \).

For electrons at room temperature, \( D = 25.8 \text{ cm}^2/\text{s} \). For an 8-\( \mu \text{m} \) long gate, the diffusion time constant is \( \approx 10 \text{ ns} \). To achieve a transfer inefficiency below \( 2 \times 10^{-5} \) per transfer, 11 time-constants, or 110 ns in this example, are required. For this reason CCDs are typically designed with gate lengths shorter than ~8 \( \mu \text{m} \) and are also designed to build in fringing fields.

In the simplest case for very low levels of transfer inefficiency, the total transfer inefficiency in a CCD register is the product of the number of stages \( N \) in the register and
the transfer inefficiency per stage. Each stage will require two or more transfers. Virtual phase and two-phase require two transfers per stage, three-phase requires three transfers, and four-phase requires four transfers. The inefficiency per stage, however, will likely depend in a complicated manner on the amount of charge in the charge packet, the amount of charge in preceding charge packets, the voltages, and frequency of operation.

In addition to the intrinsic sources of transfer inefficiency, there are a variety of extrinsic sources. These include surface and bulk traps and potential wells and barriers. The traps and the potential obstacles hold back an amount of charge from a charge packet. The charge is reemitted to later charge packets. The inefficiency due to traps depends on whether the traps have been filled by preceding charge packets and the emission time constants of the traps, and so is not modeled in a simple manner.

The intrinsic sources of noise in CCD shift registers include dark current and output amplifier noise. In addition, other sources of noise not intrinsic to the CCD itself include noise due to clock feedthrough from the CCD clocks to the output signal and noise in external electronics. The generation of dark current in CCD shift registers is the same as described at the end of Sec. 22.3 for photosensing elements. Associated with the dark current are both shot noise and pattern noise. The magnitude of the pattern noise in a CCD shift register is reduced over that of a single element since the charge packet averages the dark current over many pixels as it is transferred to the output.

The noise associated with the CCD output consists of the Johnson or thermal noise, the $1/f$ noise of the output amplifier, and the kTC noise associated with resetting the floating diffusion. The kTC noise is due to uncertainty in the amount of charge remaining on the floating diffusion following reset owing to thermal fluctuations in the reset gate. The rms noise $\sigma_n$ in the number of electrons caused by kTC noise is given by

$$\sigma_n = \sqrt{\frac{kT}{q} C_{FD}}$$

where $C_{FD}$ is one total floating diffusion capacitance. For a 10-fF floating diffusion capacitance, the rms noise is $\sim 40$ electrons.

The kTC noise may be eliminated entirely by use of a signal-processing technique called correlated double sampling (CDS). In correlated double sampling, the output level is sampled before the charge packet is transferred onto the floating diffusion and again after transfer of the charge packet (times $t_2$ and $t_0$ respectively in Fig. 6c). The two values are subtracted either by an analog circuit or by digital subtraction. Any uncertainty in the voltage level of the floating diffusion following reset is subtracted and thus the kTC noise eliminated. The output amplifier low-frequency noise, called $1/f$ noise because of the inverse frequency ($1/f$) shape of the noise power spectrum, is also reduced but not eliminated by correlated double sampling. The total noise of a CCD output amplifier is in the range of 7 to 40 rms electrons per pixel depending on the amplifier design and the operating speed. Values of a single rms electron or less have been obtained for very slow pixel rates under cooled conditions.18

**MOS Readout**

The other major category of readout structures is the MOS readout. The individual light-sensing elements (photodiodes, photocapacitors, or photoconductors) at each pixel are connected to a readout line by means of a transfer gate. Each pixel along the readout line is addressed separately by addressing circuitry. When a particular pixel is addressed, the transfer gate is turned on and the charge transferred from the pixel to the readout

---

18 The charge in the preceding packets will affect the filling of both bulk and interface traps. See, for example, Sequin and Thompsett, pp. 70–108.
An amplifier at the end of the readout line senses the change in voltage or current resulting from the charge transfer.

Typically, the pixels would be addressed serially along the line. The first pixel would be addressed, causing the charge from the image-sensing element to be transferred onto the readout line. The voltage change or current would be sensed, the readout line reset to its original voltage if necessary, and the next pixel addressed. This is different from a CCD. In the CCD, charge from all pixels is transferred into the CCD register simultaneously. Individual pixels or groups of pixels cannot easily be addressed in a random fashion by the CCD, but this random addressing can be accomplished readily by the MOS readout.

There are several types of MOS readout devices. These include the CID, the AMI, and the CMD in addition to the normal MOS array. The CID has no readout line. Each pixel consists of two overlapping gates, one controlled by a row address and the other by column address. When neither row nor column of a particular pixel is being addressed, the photogenerated charge is held under both gates and can be transferred between them. When a row of a pixel is addressed, the charge transfers onto the column gate. Then both row and column are addressed, the charge is injected into the substrate, and the current sensed. The CID is not widely used in visible imaging applications because the charge conversion sensitivity is very poor and noise very high compared to the CCD or to the other MOS architectures.

In the amplified MOS imager (AMI), the image-sensing element at each pixel consists of a phototransistor rather than a simple photodiode. The photogenerated charge is stored on the gate of the MOS transistor. When a particular pixel is addressed, the photogenerated charge modulates the transistor current. This current amplification at each pixel helps to overcome many of the noise and speed limitations of conventional MOS arrays.

MOS readout differs in an important way from CCD readout. In MOS readout, the charge is transferred from a single pixel onto a readout line and the change in voltage or current in the readout line is sensed. In a CCD, the charge packets are kept intact while being transferred physically to a low-capacitance output. The lower sensitivity of a simple MOS array can be illustrated as follows. The change in voltage on the readout line is given by

\[ V = \frac{Nq}{C} \]

where \( N \) is the number of electrons, \( q \) the electron charge, and \( C \) the readout line capacitance. Because the readout line covers the full length of the array, its capacitance is in the picofarad range (typically 2 to 10 pF depending on design and process). This compares to the 10-femtofarad capacitance for the CCD output. As a result, the voltage swings on the readout line are very small (16 nV/electron for a 10-pF readout line capacitance). This leads to a high sensitivity to clock noise due to capacitive feedthrough of the row and column address clocks onto the readout line. The feedthrough may be many times larger than the signal in most MOS sensors. Once the charge has been transferred onto the readout line, it is sensed either by a current-sensitive amplifier or by a voltage-sensitive amplifier, followed by a reset of the readout line to its original voltage.

CCD readout has the advantages of very high sensitivity and low noise. However, CCD readouts are limited in charge-handling capacity, while MOS readouts are capable of carrying very large amounts of charge and so are not as limited on the high end of the dynamic range. However, because the MOS readout line has much higher capacitance than the CCD, the sensitivity is lower and the noise is higher. Another difference is in the readout architecture. The CCD readout is essentially a serial readout device; it is not suited to random readout or partial-array readout. The MOS array, however, can be addressed in a manner similar to a memory, making it well-suited to pixel or partial-array addressing.

### 22.5 SENSOR ARCHITECTURES

Solid-state image sensors are classified into two basic groupings: linear and area. Linear sensors include single-line arrays, multilinear arrays for color scanning, and time delay and
integrate (TDI) arrays for low-light-level scanning. Area sensor architectures include the frame transfer CCD, the interline transfer CCD, and various forms of MOS $x$–$y$ addressed arrays.

**Linear Image Sensor Arrays**

Linear sensors are used almost exclusively in scanning systems for scanning of documents, film, and three-dimensional still objects. There are two basic classes of scanning systems: contact scanners and reduction scanners. These are illustrated in Figs. 10a and 10b. In reduction scanners (Fig. 10a), the sensor is smaller than the document to be scanned; lenses are used to image the document onto the sensor. In contact scanners (Fig. 10b), the sensor is the same width as the item to be scanned, usually a document. Relay optics is used between the sensor and the document. Selfoc lenses (Fig. 10c and 10d) and roof-mirror-lens arrays are the two types of relay optics used most frequently.

There are three basic architectures for linear sensing arrays: MOS line arrays, CCD linear and multilinear sensors, and time-delay and integrate (TDI) sensors. These architectures are illustrated in Fig. 11. The MOS array is used most often in contact scanning applications where material or processing problems make CCD arrays impractical. These applications include arrays fabricated from polysilicon or amorphous silicon on nonsilicon substrates, arrays covering large distances, or arrays requiring special processing (such as logarithmic amplification) at each pixel. The CCD linear and multilinear arrays are used most often in reduction scanning where wide dynamic range and small pixel size is required. However, contact arrays are also often realized by butting multiple CCD arrays end-to-end. TDI arrays are used in very low light level scanning applications where integration over many lines is required to achieve adequate signal-to-noise.

The MOS linear array (Fig. 11a) consists of individual photosensing elements, an
architectures for linear image sensors: (a) MOS line array consisting of photodiodes, preamplifier, MOS switches addressed by an address register, and a readout line with amplifier; (b) linear CCD image sensor consisting of photodiodes, transfer gate, and CCD readout; (c) linear CCD image sensor with two CCD output registers, one for the odd diodes and the other for the even diodes, for higher horizontal pitch; (d) staggered linear CCD image sensor with two rows of photodiodes offset by one-half pixel to increase horizontal sampling.

The photosensing element is usually a photodiode, although photoconductors are used in contact scanning arrays fabricated from amorphous silicon. Since the charge generated in the diode is generally very small (in the hundreds to thousands of electrons), a simple amplifier is usually placed at each pixel to drive the high-capacitance readout line. The use of amplification at each pixel can allow some signal-processing functions, such as logarithmic amplification, clipping, triggering and latching, etc., to be performed at the pixel. A MOS switch placed after the amplifier allows each pixel to be addressed in sequence; the switch is driven by an address register. At the end of the readout line is an amplifier which may buffer and/or amplify the signal. The MOS array has the advantage of process simplicity and the ability to perform signal processing at each pixel; it has the disadvantage of low signal level (because of the large readout line capacitance) and pattern noise introduced by feedthrough from the switching transistor.

The linear CCD image sensor is the most often used architecture for scanning applications owing to its low noise, high sensitivity, wide dynamic range, and small pixel pitch. Figure 11b shows the simplest type of linear CCD, in which a single row of photodiodes is connected to a single CCD register via a transfer gate. In operation, the
signal is integrated on the image-sensing element (generally a photodiode) for a line time. The horizontal CCD is then stopped, the transfer gate opened, and the charge transferred from all the photodiodes simultaneously to the CCD. The transfer time is typically a few microseconds. The transfer gate is then closed, integration resumed for the next line, and the CCD clocked to read out the charge packets. Many arrays also feature antiblooming for situations where the light level may not be controlled, as well as electronic shuttering, which allows an integration time on the photodiodes to be less than the readout time of the CCD.

For linear-sensing applications requiring a higher pixel density, a double-sided readout is often used (Fig. 11c). In this architecture, a CCD array is placed on either side of the line of photodiodes and charge transfer from the diodes alternates between the top and bottom CCDs. This architecture uses lower horizontal clock rates and a higher pixel pitch, since the diode pitch can usually be made smaller than the CCD pitch. The charge packets from the two arrays may be multiplexed into one output if desired. The disadvantage of this architecture is slight differences between even and odd pixels, due to slight differences in the two outputs (or slight differences due to multiplexing the two registers).

Another architecture which is used to further decrease the sampling pitch is the
staggered linear array (Fig. 11d). In the staggered array, two rows of photodiodes are offset by a half pixel. The two rows are read out by CCD arrays. The first array is delayed (usually in a digital line store) and then combined with the second to form a double-density scan.

Multilinear arrays (Fig. 11d) have recently been developed for color scanning applications. In this architecture, several (usually three) linear arrays are combined on the same silicon die separated by a distance equivalent to an integral number of scan lines. Color filters (either integral or in close proximity) are aligned over the arrays. External digital line delays are used to realign the three arrays. Separate electronic shuttering may be provided for each array in order to adjust for differences in intensity in each of the bands.

The third major class of line arrays is time-delay and integrate, or TDI, arrays.\textsuperscript{22,23} The TDI architecture is shown in Fig. 11f. TDI arrays are used when inadequate signal-to-noise ratio from a single-line array requires averaging over multiple lines. Applications for TDI arrays include high-speed document scanners and space-based imaging systems. Instead of a single row of photodiodes, the TDI array utilizes CCD stages in the vertical dimension which are clocked synchronously with the movement of the document to be scanned. The signal level in a TDI array increases linearly with the number of stages. The noise level, however, increases most as the square root of the number of stages.

### Area Image Sensor Arrays

There are three major classes of area image sensor architectures: MOS diode arrays, frame-transfer CCDs, and interline-transfer CCDs. Within each there are a number of variations. CCDs have come to dominate the majority of applications owing to their higher sensitivity. However, MOS arrays are still used for specialized applications where addressability or high readout rate is important.

Historically, the physical dimensions of the active imaging areas of CCD arrays for consumer and commercial applications are specified by the size of the vidicon tube which it replaces. The common format sizes include $\frac{1}{4}$-in, $\frac{1}{3}$-in, $\frac{1}{2}$-in, and 1-in. In most cases, the aspect ratio is 4:3, reflecting the television standard. Table 1 lists the formats and the corresponding dimensions of the imaging area of the array.

<table>
<thead>
<tr>
<th>Optical format</th>
<th>1-in</th>
<th>$\frac{1}{4}$-in</th>
<th>$\frac{1}{3}$-in</th>
<th>$\frac{1}{2}$-in</th>
<th>1-in</th>
<th>1-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active area (4:3 aspect ratio)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height (mm)</td>
<td>9.6</td>
<td>6.6</td>
<td>4.8</td>
<td>3.3</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>Width (mm)</td>
<td>12.8</td>
<td>8.8</td>
<td>6.4</td>
<td>4.4</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Diagonal (mm)</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>5.5</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pixel dimensions (484 lines × 768 pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (μm)</td>
</tr>
<tr>
<td>Width (μm)</td>
</tr>
</tbody>
</table>

\textit{The format name is given in inches based on historic image tube formats. The pixel dimensions are based on a 484 × 768 pixel array.}
The standards for the number of vertical lines in arrays for consumer and professional video are usually based on the corresponding television standards, including NTSC, PAL, and the various HDTV standards. NTSC has 484 active lines, PAL has 575, the Japanese HDTV standard has 1035, and the European HDTV standard has 1150. In all cases the lines are interlaced; i.e., odd lines are read out in one field and even lines in the next. Historically, for sensors for NTSC television, the number of pixels horizontally in a line has been associated with multiples of the color subcarrier frequency; common horizontal pixel counts include 384, 576, and 768. Sensors for the Japanese HDTV standard typically have 1920 horizontal pixels. The pixels are rectangular rather than square. Table 1 lists approximate pixel dimensions in micrometers for a few common formats for a 484 (V) × 768 (H) pixel image sensor.

For industrial, scientific, graphics electronic photography, digital television, and multimedia applications, however, the nonsquare pixels and interlaced readout of sensors based on television standards are significant disadvantages. Image sensors designed for these industrial and commercial applications typically have square pixels and progressive scan readout. In interlaced NTSC readout, for example, the first field consists of lines 1, 3, 5 · · · 483 and is read out in the first 1/60 second of a frame. The second field consists of lines 2, 4, 6 · · · 484 and is read out in the second 1/60 second of a frame. The resulting temporal and spatial displacement between the two fields is undesirable for these applications. In addition, digital compression of interlaced scan moving images, especially with motion estimation, is difficult and also introduces artifacts.

In progressive scan readout each line is read out sequentially. There is no even or odd field, only a single frame. As a result there are no temporal and spatial sampling displacement differences. However, for a given resolution and frame rate, the readout rate of a progressive scan image sensor is double that of an interlaced scan. In addition, for these applications, the number of pixels is often based on powers of 2: such as 512 × 512 or 1024 × 1024—facilitating memory mapping and image processing.

**MOS Area Array Image Sensors.** The architecture of MOS area arrays is illustrated in Fig. 12.24 It consists of the imaging array, vertical and horizontal address registers, and output amplifiers. The pixel of a MOS array consists of an image-sensing element (photodiode, photocapacitor, phototransistor, or photoconductor), a row-address gate, and a vertical readout line. The row-address gate is bussed horizontally across the array and is driven from a row-address register on the side(s) of the array. At the start of a line, a single row is addressed, causing the charge from all the photodiodes in a row to be transferred onto the vertical readout line. Horizontal address gates are placed at the bottom of the vertical readout line. Buffer amplifiers to drive the horizontal readout line may also be placed at the end of the vertical readout line. The horizontal address register then serially addresses each vertical readout line, sequentially turning on the horizontal address gates. After the horizontal addressing is completed, the readout lines may be reset and precharged and the next row addressed.

There is a wide range of variations on this basic architecture. An example is the charge modulation device, or CMD,25,26 array in which a phototransistor is placed at each pixel site in order to achieve amplification and to achieve high currents to drive the capacitance of the vertical and horizontal readout lines. Other examples include arrays with sophisticated charge collection circuits at the end of each vertical readout line.

**Frame Transfer CCD Image Sensors.** CCD area arrays fall into two categories: frame transfer and interline transfer. The simplest form of frame transfer CCD is the full-frame CCD, shown in Fig. 13a. A photograph of a few pixels of a frame transfer CCD is shown in Fig. 14a. The array consists of a single image area composed of vertical CCDs and a single horizontal register with an output amplifier at its end. In this architecture, the pixel consists of a single stage of a vertical CCD. This type of device requires an external shutter. When the shutter is opened, the entire surface of the sensor is exposed and the
charge is collected in the CCD potential wells at each pixel. After the shutter is closed, the sensor is read out a row at a time by clocking a row of the vertical register into the horizontal register, then clocking the horizontal register to read out the row through the output amplifier. For higher readout rates dual horizontal CCDs are used in parallel. The full-frame CCD has the advantage of progressive scan readout high fill factor, very low noise, and wide dynamic range. However, it requires an external shutter. It is most often used in still electronic photography, scientific, industrial, and graphics applications.

For motion imaging applications, a shutter is not practical. In order to overcome the need for a shutter, frame transfer CCDs incorporate a storage area in addition to imaging area. For interlaced video applications, this storage area is sufficiently large to hold a field (242 lines in NTSC television). The image area consists of vertical CCDs. For interlaced NTSC video, there are 242 pixels vertically in the image area. Interlace is achieved by changing the gates under which integration is performed in the even and odd fields. For example, for a four-phase CCD, integration would be performed under phases 1 and 2 in one field and 3 and 4 in another, effectively shifting the sampling area by half a pixel in each field. The storage area consists also of 242 pixels vertically. The device operation is as follows. The image area integrates for a field time and the photogenerated charge held in the vertical CCDs. The vertical CCDs are then clocked in order to rapidly transfer the charge from the image area into the storage area. Because the sensor is still under illumination, this transfer time must be much shorter than the integration time. This transfer typically requires 0.2 to 0.5 ms. The storage area is then read out a row at a time by transferring a row into the horizontal register and clocking the horizontal register. While this readout is occurring, the image area is integrating the next field. The most significant disadvantage of frame transfer CCDs is the image smear caused by illumination during the transfer from the image to the storage area. This smear can be on the order of 3 percent.

FIGURE 12 MOS photodiode array, consisting of vertical and horizontal address registers and readout line.
In both frame transfer and full-frame devices, the light must pass through the polysilicon electrodes before being absorbed in the silicon. Owing to the high absorption of short wavelengths in the polysilicon, the quantum efficiency in the blue is only about 20 percent and the quantum efficiency in the green is about 50 percent. Figure 14b shows the quantum efficiency for a full-frame image sensor with polysilicon gates. Three approaches have been used to improve the efficiency: the virtual phase CCD, transparent electrodes, and backside illumination. In the virtual phase CCD (see “Types of CCDs” in Sec. 22.4), the second polysilicon electrode is replaced with a very shallow highly doped $p^+$ layer, very similar to the pinned photodiode (Fig. 2d). Since there is no electrode over this phase to absorb the light, higher quantum efficiency is achieved, particularly at wavelengths less than 500 nm. Backside illumination provides nearly unity quantum efficiency but requires that the sensor be thinned to less than 10 μm. Owing to its cost, backside thinning is employed only in sensors for very specialized scientific or aerospace applications. Indium-tin-oxide, or ITO, is the most commonly used transparent electrode. Usually it is substituted for the second level of polysilicon. Figure 14b also shows the quantum efficiency of a full-frame device in which ITO has been substituted for one of the polysilicon gates.

**Interline Transfer CCD Image Sensors.** The interline transfer CCD is fundamentally
different from the frame-transfer CCD in that, in addition to the vertical CCD, the pixel also contains a separate image-sensing element (photodiode, pinned photodiode, photocapacitor, or photoconductor) and a transfer region between the photodiode and the vertical CCD. Interline CCDs with photodiodes for sensing elements are considered first. Figure 13b illustrates the architecture of an interline transfer CCD and Fig. 15 illustrates a pixel of the CCD. The CCD and the transfer region between the diode and CCD are covered with a light shield (typically aluminum, although metal silicides are also used). The light shield prevents any light from entering the vertical CCD registers, allowing them to be read out while the sensor is illuminated. When the sensor is illuminated, the photogenerated charge is held on the photodiode. During the vertical retrace interval at the end of a field, the photogenerated charge is transferred into the vertical CCD by clocking the CCD gate over the transfer region. Once the charge has been transferred from the diodes into the vertical CCDs, the diodes resume integrating and the vertical
FIGURE 15 (a) Diagram of a pixel of an interline transfer CCD; (b) photograph of a pixel from an interline transfer CCD.
CCDs are clocked in order to transfer a row at a time from the image area into the horizontal CCD.

For consumer and most commercial applications, interlaced interline CCDs are used to be consistent with television standards. In interlaced interline CCDs there is one vertical CCD stage for every two photodiodes. During the retrace time before the first field the charge from the odd rows of photodiodes is transferred into the vertical CCDs; the charge is then transferred out a row at a time into the horizontal CCD. During the retrace time before the second field the charge from the even rows of photodiodes is transferred into the vertical CCDs; once again, the charge is transferred out a row at a time into the horizontal CCD. In NTSC television the fields are approximately 1/60 second long; in PAL the field time is 1/50 second.

Because the vertical CCDs in an interline CCD are covered by a light shield, very little stray light is absorbed in the CCD. However, due to light scattering under the light shield and lateral diffusion of photogenerated electrons, some charge can reach the vertical registers as they are read out during a field. This results in smear. For consumer applications the level of smear is not noticeable. However, for especially demanding applications such as television broadcast cameras, a field storage area is added to the bottom of the image area. This architecture is called the frame interline transfer, or FIT CCD (Fig. 13c). Following transfer of the photogenerated charge from the diodes into the vertical CCDs, the vertical CCDs are clocked to rapidly shift the charge from the image area into the storage area. This transfer typically takes less than 0.5 ms, reducing smear 30-fold. One line at a time is transferred from the storage area to the horizontal register and the horizontal register readout.

For still electronic photography, scientific, computer-related, graphics and professional applications, interlaced video is not desirable. For these applications a progressive scan architecture is utilized. In a progressive scan interline CCD there is a full CCD stage for every photodiode. Following integration, the photogenerated charge from all the photodiodes is transferred into the vertical CCDs. The photodiodes resume integration and the charge packets from the vertical registers are transferred into the horizontal a row at a time. The progressive scan interline CCD requires twice the vertical CCD density and is therefore more complicated to fabricate. However, progressive scan readout provides many advantages in image quality for both motion and still imaging and is emerging as a likely standard for future digital HDTV systems as well as multimedia.

Nearly all interline CCDs used in camcorder, broadcast camera, or commercial applications utilize vertical antiblooming in order to prevent blooming when the sensor is illuminated beyond saturation. A cross-section diagram of an interline CCD with vertical antiblooming is shown in Fig. 15b. The device illustrated uses a pinned photodiode photosensing element. The CCD is built on an n-type silicon substrate. A p-well is formed about 2 μm deep in the n-type silicon. An n-type buried-channel is then formed followed by a p+ surface layer. The n-type substrate is reverse biased with respect to the p-well. When the photodiode is illuminated above saturation, the excess electrons spill out of the n-type buried channel and into the substrate. Owing to the vertical overflow, however, photogenerated carriers from photons absorbed below the p-well are drawn into the substrate and are not collected by the diode. Thus the quantum efficiency of these devices falls rapidly at wavelengths beyond 550 nm. Figure 16 shows the quantum efficiency of an interline CCD with vertical antiblooming as a function of wavelength. The internal quantum efficiency of the photodiode is nearly 100 percent to about 550 nm, after which it decreases. However, since the photodiode only occupies about 20 percent of the pixel, and this aperture is typically reduced further by the light shield in order to eliminate optical scattering into the vertical CCD, the actual efficiency of the device is only about 15 percent. The photoresponse is linear at low signal levels but becomes nonlinear at charge levels near saturation.

Two major approaches are used to improve the fill factor (and therefore the quantum efficiency) of interline CCDs. These are the use of microlens arrays to focus the light incident on a pixel onto the photodiode and vertical integration of image sensors with
amorphous silicon photoconductors to achieve a high fill factor. Figure 17 shows a microlens array on top of an interline CCD. The lenses are formed by coating a spacer layer on the wafer of CCD devices followed by a lens-forming layer. The lens forming layer is patterned and then reflowed to form the lens arrays. The quantum efficiency of an interline CCD with and without a microlens array is shown in Fig. 18. A 3-fold improvement in quantum efficiency is achieved because light from nearly the entire pixel area is focused onto the photodiode.

The second approach to improving fill factor in interline CCD devices is to vertically integrate the device by use of an amorphous silicon photoconductor. The structure of the photoconductor and its band diagram are illustrated in Fig. 2e. The amorphous silicon is

---

**FIGURE 16** Quantum efficiency as a function of wavelength for an interline transfer CCD with and without microlens array.

**FIGURE 17** Scanning electron micrograph of microlens array on top of an interline transfer CCD.
about 1 μm thick; owing to its high absorption coefficient in the visible, it can achieve nearly 100 percent internal quantum efficiency. The back contact of the photoconductor contacts the diode in the interline CCD. Light absorbed in the amorphous silicon generates electron-hole pairs. The amorphous silicon is biased such as to create a high field which sweeps out the electrons to the back contact, where they can be stored on the diode until they are transferred into the CCD. Because the photoconductor is fabricated on top of the CCD pixel, it can have nearly 100 percent fill factor. In addition, because of the wide bandgap of the amorphous silicon, its dark current (due to thermal generation of carriers) is often lower than the single-crystal silicon. However, the amorphous silicon photoconductors suffer difficulties related to charge trapping. Owing to impurities and dangling or strained bonds, there is a high density of traps in amorphous silicon. These can trap charge carriers and reemit them at a later time, causing image lag. Because the trapping and detrapping is field-dependent, it can also result in nonlinear response.

**CCD Performance.** In all CCDs, both interline and frame transfer, the signal output is linear with the illumination except at levels approaching saturation. This linear response is in marked contrast to image tubes, which exhibit highly nonlinear response. For interline CCDs, the total charge capacity ranges from 100,000 electrons for larger cells (such as the 13.6 (V) × 11.6 (H)-μm cell typical for a 3-in format) to 20,000 electrons or less for smaller cells (such as the 6.8 × 5.8-μm cell in a 1/2-in format 484 × 768-pixel CCD).

The principal noise sources in both interline and full-frame image sensors are dark current pattern and shot noise and output amplifier noise. To illustrate the contributions, modern CCDs have dark current levels at 40°C of less than 1 nA/cm². For a 3/2-in format sensor, this would correspond to about 320 electrons per pixel dark level. The corresponding shot noise would be 18 rms electrons and the pattern noise would typically be at a similar level. However, because of the nonrandom nature of pattern noise, its appearance is considerably more noticeable. The output amplifier noise would also be at about the 15-rms electron level. Thus, the overall noise for this example would be in the 30-electron range and the dynamic range would be over 2000 for a charge capacity of 90,000 electrons. For scientific applications where the sensor can be cooled and the readout performed at a lower frequency, noise levels less than 5 electrons can be achieved and even subelectron noise has been reported.

**Color Imaging**

Silicon based CCDs are monochrome in nature. That is, they have no natural ability to determine the varying amounts of red, green, and blue (RGB) illumination presented to the photodetectors. There are three techniques to extract color information.
1. **Color Sequential** (Fig. 19)—A color image can be created using a CCD by taking three successive exposures while switching in optical filters having the desired RGB transmittances. This approach is normally used only to provide still images of stationary scenes. The resulting image is then reconstructed off-chip. The advantage to this technique is that resolution of each color can remain that of the CCD itself. The disadvantage is that three exposures are required, reducing frame times by more than a factor of three. Color
misregistration can also occur due to subject or camera motion. The filter switching assembly also adds to the mechanical complexity of the system.

2. **Three-chip Color** (Fig. 19) — Three-chip color systems use an optical system to split the scene into three separate color images. A dichroic prism beam splitter is normally used to provide RGB images. Color images can then be detected by synchronizing the outputs of the three CCDs. The disadvantage to such a system is that the optical complexity is very high and registration between sensors is difficult.

3. **Integral color Filter Arrays (CFA)** (Fig. 19) — Instead of performing the color filtering off-chip, filters of the appropriate characteristics can be fabricated above individual photosites. This approach can be performed during device fabrication using dyed (e.g., cyan, magenta, yellow) photoresists in various patterns. The major problem with this approach is that each pixel is sensitive to only one color. Off-chip processing is required to “fill in” the missing color information between pixels.

In order to minimize size, weight, and cost, most consumer color camcorders use a CCD sensor with an integral CFA. The photosites are covered with individual color filters—for example, a red, green, and blue striped filter, or a green, magenta, cyan, and yellow mosaic filter. Some popular CFA patterns are shown in Fig. 20. Because each photosite can sense only one color, the color sampling is not coincident. For example, a blue pixel might be seeing a white line, while nearby red and green pixels are seeing a dark line in the scene. As a result, high-frequency luminance edges can be aliased into bright color bands. These color bands depend not only on the color filter pattern used, but also on the optical prefilter and CFA interpolation algorithm. The color bands are caused by aliasing, which is a property of any sampled system. Aliasing occurs when the frequency of the input signal is greater than the Nyquist limit of one-half the sampling frequency. If the input frequency is well below the Nyquist limit, there are many samples per cycle. This allows the input to be reconstructed perfectly, if a proper reconstruction filter is used. When the input frequency is greater than the Nyquist limit, there are less than two samples per cycle. The sample values now define a new curve, which has a frequency lower than the input frequency. In effect, the high frequency takes on the alias of a lower frequency.

---

![Figure 20: Common color filter array patterns.](image)

Note: Color R = Red, G = Green, B = Blue
Y = Yellow, M = Magenta, C = Cyan, W = White

**FIGURE 20** Common color filter array patterns.
FIGURE 21 Birefringent blur filter used to reduce aliasing in single-chip color image sensors.

Aliasing is a particular problem with color sensors, since the sampling phase is different for the different color photosites. Therefore, the aliased signal has different phases for different colors. This creates the color bands.

The color aliasing is reduced by using an optical anti-aliasing or “blur” filter, positioned in front of the color CCD sensor. Blur filters are typically made of birefringent quartz, with the crystal axis oriented at a 45° angle, as shown in Fig. 21. In this orientation, the birefringent quartz exhibits the double refraction effect. An unpolarized input ray emerges as two polarized output rays. The output ray separation is proportional to the filter’s thickness. A 1.5-mm-thick plate will give a separation of about 9 μm. Figure 21 shows a simple “two-spot” filter. More complex filters use three or more pieces of quartz cemented in a stack.

22.6 REFERENCES