Experimental Astronomy (2005) 19: 111–134
DOI: 10.1007/s10686-005-9013-2

CMOS DETECTOR TECHNOLOGY

ALAN HOFFMAN[1], MARKUS LOOSE[2] and VYSHNAVI SUNTHARALINGAM[3]
[1]*Raytheon Vision Systems;* [2]*Rockwell Scientific Company;* [3]*MIT Lincoln Laboratory*

(Received 15 September 2005; accepted 6 December 2005)

Abstract. An entry level overview of state-of-the-art CMOS detector technology is presented. Operating principles and system architecture are explained in comparison to the well-established CCD technology, followed by a discussion of important benefits of modern CMOS-based detector arrays. A number of unique CMOS features including different shutter modes and scanning concepts are described. In addition, sub-field stitching is presented as a technique for producing very large imagers. After a brief introduction to the concept of monolithic CMOS sensors, hybrid detectors technology is introduced. A comparison of noise reduction methods for CMOS hybrids is presented. The final sections review CMOS fabrication processes for monolithic and vertically integrated image sensors.

Keywords: APS, active pixel sensor, CCD, CMOS, focal plane array, HgCdTe, hybrid, image sensor, InSb, three-dimensionally stacked circuits, vertical integration

1. Introduction

The idea of using complementary metal oxide semiconductor (CMOS) technology to build two-dimensional arrays of photosensitive pixels is over 20 years old. However, the first CMOS arrays were designed exclusively for use in hybrid infrared detector arrays and were therefore too exotic and too expensive for the general public to notice. In the mid 1980s, the sensor domain for visible light had just been taken over by the newly developed charge-coupled device (CCD) technology, and CMOS with its large design rules was not a serious alternative. CMOS technology was superior for the infrared where silicon is not sensitive, and where the pixels were big and the large feature size of CMOS did not matter. As a consequence, large hybrid arrays of up to 1K $\times$ 1K resolution became available by the mid 1990s.

At about the same time, CMOS process technology had advanced to sufficiently small dimensions to enable building monolithic CMOS sensors with small pixels and good performance. It took many more years, however, before CMOS sensors became capable of achieving the high performance level of CCDs. Today, CMOS image sensors make up a significant share of the commercial sensor market and can be found in a variety of consumer devices. In addition, CMOS hybrid technology has progressed further into multimegapixel arrays for both infrared and visible applications.

2. General CMOS sensor concept

CMOS image sensors use the same highly integrated circuit technology as microprocessors and memory chips. For that reason, any desired circuit can be implemented, and there are virtually no limits to the functionality and complexity of CMOS sensors. On the other hand, most of the imagers follow the same basic concept and have a similar architecture. The following sections introduce common principles applicable to the majority of today's CMOS sensors.

2.1. Basic principle

The goal of any image sensor is to detect light as efficiently as possible. CCDs and CMOS-based imagers use the photoelectric effect of a semiconductor to convert photons into electrical charges. Unlike CCDs, however, CMOS sensors are not limited to silicon as the detector material. Instead, a number of different materials can be hybridized to a CMOS readout integrated circuit (ROIC) to provide sensitivity for ultraviolet, visible, or infrared light, as discussed in Section 4. The following description of the basic CMOS concept is applicable to both monolithic CMOS detectors and hybrid CMOS detectors.

Figure 1 compares the principle of CMOS sensors to that of CCDs. Both detector technologies use a photodiode to generate and separate the charges in the pixel.

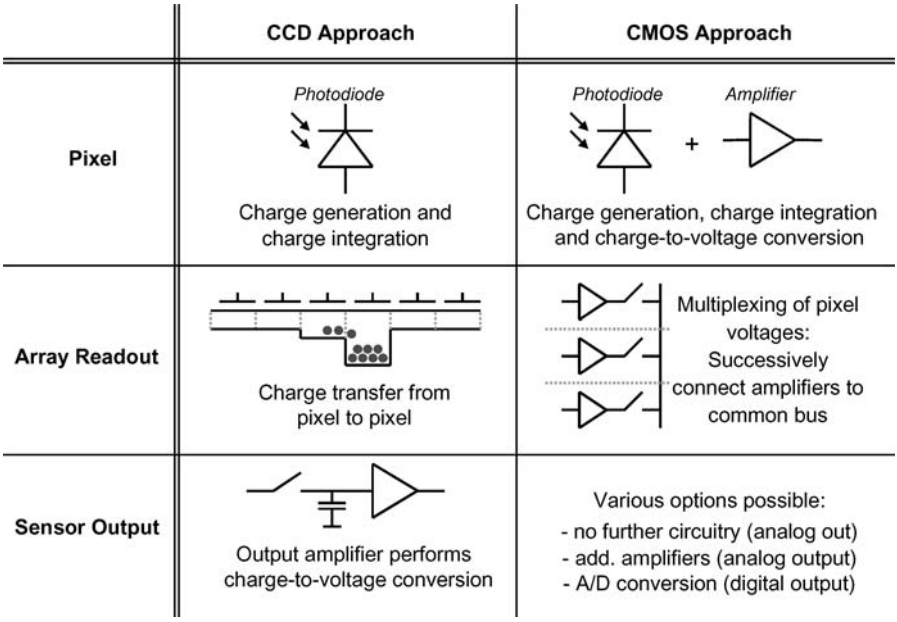

Figure 1. Comparison between the CCD-based and the CMOS-based image sensor approach.

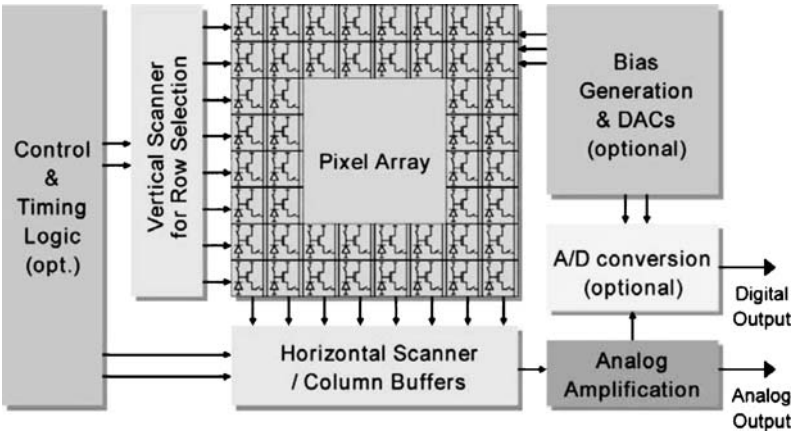

Figure 2. Block diagram of a generic CMOS sensor.

Beyond that, however, the two sensor schemes differ significantly. During readout, CCDs shift the collected charge from pixel to pixel all the way to the perimeter. Eventually, all charges are sequentially pushed to one common location (floating diffusion), and a single amplifier generates the corresponding output voltages. On the other hand, CMOS detectors have an independent amplifier in each pixel, also called an active pixel sensor (APS). The amplifier converts the integrated charge into a voltage and thus eliminates the need to transfer charge from pixel to pixel. Instead, the voltages are multiplexed onto a common bus line using integrated CMOS switches. Analog and digital sensor outputs are possible by implementing either a video output amplifier or an analog-to-digital (A/D) converter on the chip.

2.2. Architecture of CMOS image sensors

In addition to the basic concept of active pixels, a number of common features can be found in most CMOS-based imagers. As shown in Figure 2, two different scanners surround the actual pixel array: a vertical scanner to control the row selection, and a horizontal scanner to amplify and multiplex the analog pixel signals. While most CMOS imagers include comparable vertical scanners, they differ quite substantially in the architecture of the horizontal scanner. Low-speed astronomy detectors typically use a row of analog switches controlled by a simple digital shift register. Faster image sensors require additional circuitry in each column like sample/hold stages or column buffers. In some cases, even the A/D conversion is integrated into the horizontal scanner as a part of the column structure.

Most CMOS sensors include additional circuitry for bias generation, timing control, and A/D conversion. The latter can be found more and more in modern sensors that use deep submicron process technologies. By integrating all the support electronics into the same silicon as the pixel array, complete cameras can be built on

a single chip. This approach is very attractive to the commercial image sensor market because of size and cost constraints. On the other hand, high-performance scientific sensors typically do not push towards the highest integration level. Here, very high resolutions at very low power levels require a simpler detector architecture, and most of the support circuitry is provided by external electronics.

2.3. Common CMOS properties

A high level of flexibility and a number of unique features characterize CMOS detector technology. This section summarizes some of the important aspects, in particular the properties that set CMOS detectors apart from CCD sensors.

In terms of manufacturing, CMOS imagers in general benefit from the large availability of foundries worldwide. Using the same foundry resources as microchips guarantees cost-efficient production and highly mature process technology. Design rules as advanced as 0.13 μm are being used for the latest generation of image sensors.

A significant advantage of CMOS sensor technology is its high level of flexibility. Small and simple pixels with three or four transistors are being used to achieve basic light detection at high resolutions. Larger and more complicated pixels with hundreds of transistors provide A/D conversion or other image processing capabilities directly at the pixel level. Furthermore, additional on-chip circuitry can support many analog and digital signal processing functions to reduce the requirements on system power or transmission bandwidth. Fortunately, CMOS imagers operate at very low power levels and therefore can tolerate the increased power consumption of most on-chip functions.

Typically, CMOS-based detectors do not need a mechanical shutter. Instead, integration time is controlled electronically. Two of the main electronic shutter concepts, the snapshot shutter and the rolling shutter, are explained in Section 2.4. An interesting feature is that pixels can be read out without destroying the integrated detector signal (nondestructive read). This allows, each pixel can be read multiple times, thereby reducing read noise by averaging multiple reads. Unlike CCDs, the detector array can be scanned in a number of different ways, including random access to any pixel at any time. Section 2.3.1 illustrates some of the special scanning techniques.

Because every pixel includes active components, the matching of device properties from pixel to pixel is important. Any mismatch can lead to pixel-to-pixel nonuniformities, called fixed pattern noise (FPN) or photoresponse nonuniformity (PRNU). Many sensors resolve the issue by providing on-chip correlated double sampling (CDS). This reduces the FPN and, at the same time, eliminates the temporal kTC noise. However, in some cases, the correction for the nonidealities has to be performed outside the sensor, thus adding complexity to the system.

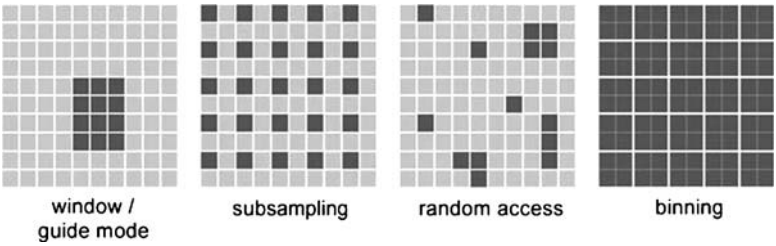

Figure 3. Different scanning schemes available in CMOS sensors.

2.3.1. *Scan modes*

Similar to a memory chip, CMOS sensors are capable of accessing pixels in random order. It is merely a matter of the surrounding scanner logic to define the available scanning methods. Figure 3 shows four examples of common scanning schemes beyond the standard full frame read mode. Because of a reduced number of effective pixels, all special scanning modes provide the advantage of faster frame rates compared to the full frame mode.

The first example is the window mode, in which a rectangular shaped subsection is being read out. The location and the size of the window are usually programmable parameters. If the window can be reset and read without disturbing the other pixels in the detector array, the window mode can be used for simultaneous full field science exposure and fast guide window operation for telescopes. The second special scanning technique is called subsampling. In this method pixels are skipped during the frame readout. As a result, the complete scene is captured at a faster frame rate, yet at a reduced resolution. The third possible readout or reset scheme is random access. Every pixel is read when desired, and no predefined sequence has to exist. This can be helpful in selectively resetting saturated pixels. Fourth, CMOS sensors can combine multiple pixels into one larger pixel. This process is similar to the binning technique known from CCDs. However, CMOS devices typically do not achieve the same improvements in signal-to-noise ratio from binning as do CCD devices, and the scheme is therefore not as commonly used.

2.4. SNAPSHOT VS. ROLLING SHUTTER

Two main concepts of electronic shutters have been established for CMOS image sensors: the snapshot shutter and the rolling shutter. Neither requires a mechanical shutter, but they differ with respect to their specific implications. Figure 4 illustrates both shutter concepts by means of a timing diagram of five consecutive rows. In the case of the snapshot shutter (left side), all rows start and stop integrating at the same time. Consequently, a complete picture is captured simultaneously by the whole array, a fact that is very important for fast moving objects. Depending on the design and the complexity of the pixel, the next exposure has to wait until all pixels

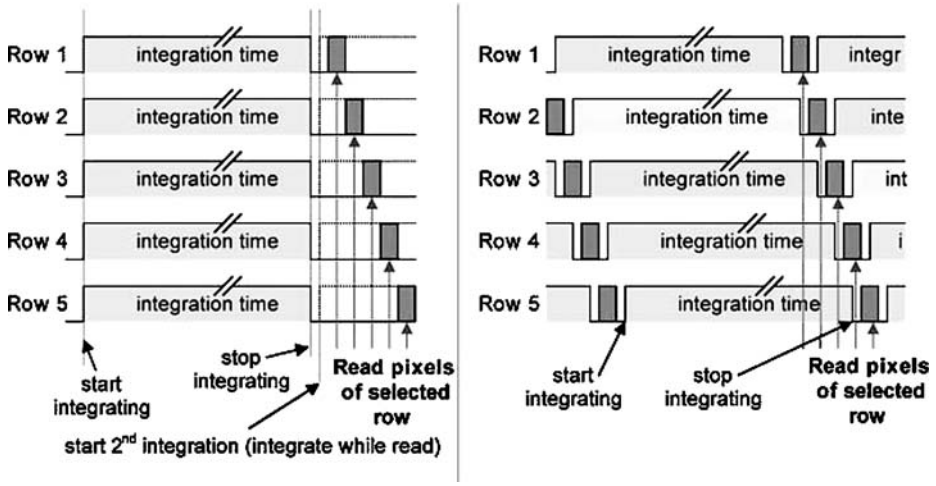

Figure 4. (left) Snapshot shutter vs. *(right)* rolling shutter.

are read (integrate, then read), or the next exposure can start while the previous values are being read (integrate while read).

The rolling shutter approach, on the other hand, does not globally start and stop the exposure. Instead, the start and the stop time of the actual integration is shifted every row by one row time with respect to the previous row. This can be seen on the right side of Figure 4. Basically, as soon as the exposure of one row has finished, the row is read out and then prepared for the next integration period. Thus, the pixel does not require any sample and hold circuit. This results in smaller pixels and typically higher performance. The rolling shutter is beneficial for all applications that observe static or slow moving objects.

2.5. CMOS-BASED DETECTOR SYSTEMS

As described in Section 2.2, CMOS image sensors require additional support electronics for controlling and biasing the detector system and for digitizing the analog signals. Three main approaches are used for implementation of the support electronics, as shown in Figure 5, (a) a single chip solution like in Kozlowski et al. (2005), (b) a system with analog image sensor and discrete external electronics, (c) a dual chip approach with analog image sensor and application specific integrated circuit (ASIC).

The single chip system is the most integrated concept since all required electronics are part of the sensor chip itself. Single chip camera systems can be very small and do not consume much power. However, they are difficult and expensive to design. Also, they can show undesired side effects like transistor glow or higher power consumption close to the pixel. Both effects potentially increase the detector

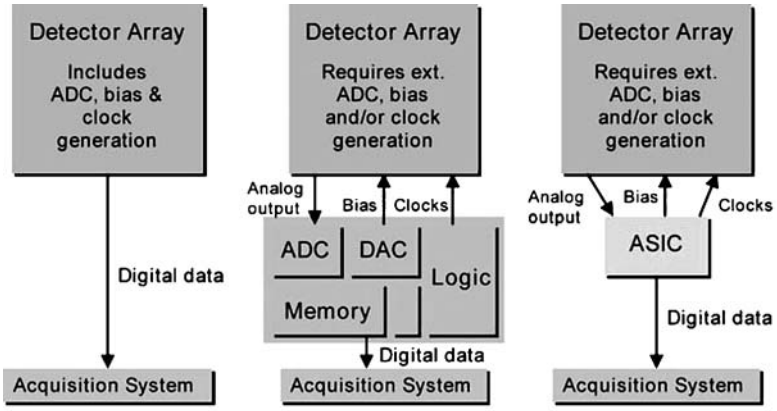

Figure 5. Three concepts for implementing support electronics in CMOS-based detector systems: *(left)* single chip, *(center)* discrete electronics, and *(right)* dual chip.

dark current. Therefore, single chip solutions are typically not being used for high-performance scientific applications in order to maintain lowest dark current and highest sensitivity.

Conventionally, scientific detectors, and in particular astronomy camera systems, have been using the second approach comprising of an analog sensor chip and discrete external electronics. Some of the advantages include modularity, i.e., one controller can be used with many different detectors, and a less expensive detector design. On the negative side, discrete systems dissipate more power and are much larger.

To combine the benefits of the two concepts, a third approach is being pursued. Here, the detector system is composed of two separate chips. The first one is an analog image sensor, like the one used with discrete electronics controllers. The second chip is a programmable mixed-signal ASIC that effectively combines all functions of the discrete controller in a single chip Loose et al. (2003). The two chips together form a small and lightweight, yet flexible and high-performance, camera system.

2.6. STITCHING

The maximum size of a CMOS chip in modern deep submicron technology is limited to about 22 mm $\times$ 22 mm. This is the maximum reticle size that can be exposed in a single step. To build larger sensor arrays, a special process called *stitching* has to be applied. To do this, the large CMOS sensor is divided into smaller subblocks. These blocks have to be small enough that they all fit into the limited reticle space. Later, during the production of the CMOS wafers, the complete sensor chips are being stitched together from the building blocks in the reticle. Each block can be exposed multiple times within the same sensor, thus creating arrays much larger than the reticle itself.

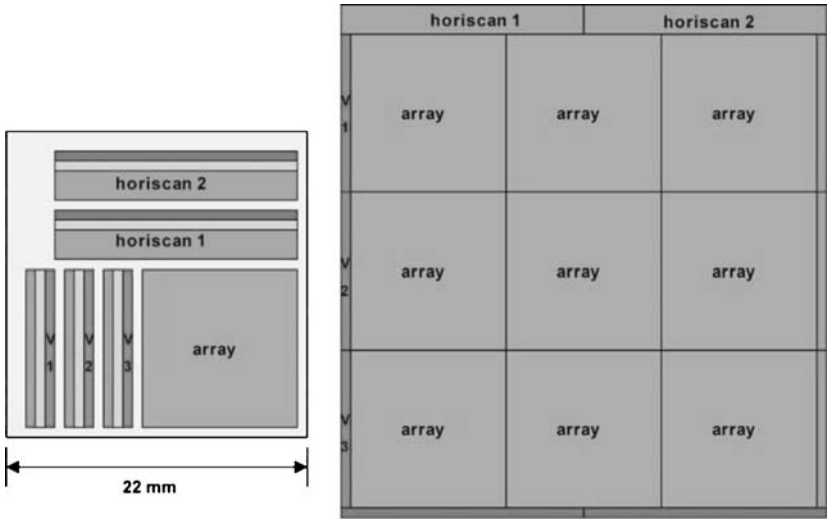

Figure 6. (left) Mask reticle and *(right)* stitched CMOS readout array.

An example of a stitched FPA is shown in Figure 6. The left part of the picture includes the reticle with six independent subblocks. The right half illustrates a completely assembled CMOS readout chip, built from the individual subblocks in the reticle. The ultimate limit for this approach is only the wafer size itself, i.e., 15 to 30 cm depending on the process technology. However, yield limitations will make it nearly impossible to manufacture wafer-scale arrays, and mosaics of smaller independent detectors are typically used to build very large focal plane arrays (FPAs).

3. Monolithic CMOS sensors

Monolithic CMOS sensors are image sensor chips that combine the photodetectors and the readout circuitry on the same piece of silicon. Compared to hybrid FPAs, monolithic arrays are less expensive to manufacture, but the sensitivity is limited to the visible and near infrared wavelength range. Over the course of the last decade, monolithic CMOS technology has made significant improvements. In recent years, a large number of consumer products, including cell phones, digital cameras, and camcorders, have been replacing CCDs with CMOS sensors for the sake of smaller, cheaper, and lower-power systems. A monolithic CMOS sensor is also called an APS.

Although a wide variety of pixel designs exist, most sensors use one of the two circuits shown in Figure 7. The circuit on the left side consists of three transistors (3T): a reset field-effect transistor (FET), a selection switch, and a source follower

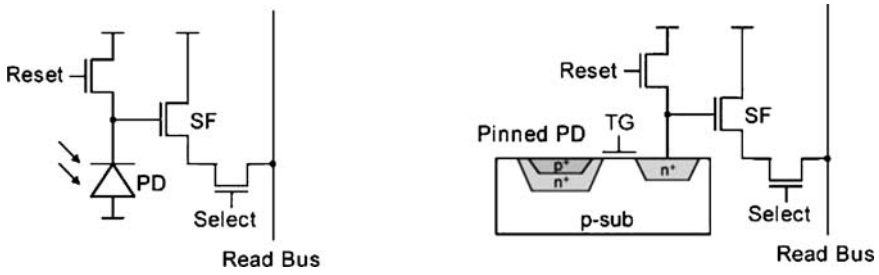

Figure 7. Diagram of two common monolithic CMOS pixel circuits: *(left)* three transistors with regular photodiode and *(right)* four transistors with pinned photodiode.

for driving the signal onto the column bus. The right side of Figure 7 displays a pixel with four transistors (4T) and a pinned photodiode. Three of the FETs have the same function as in the 3T pixel. The fourth transistor works as a transfer gate that moves charge from the photodiode to a floating diffusion. Usually, both pixels operate in rolling shutter mode. The 4T pixel is capable of performing CDS to eliminate the reset noise (kTC noise) and the pixel offsets. The 3T pixel can only be used with noncorrelated double sampling (DS), which is sufficient to reduce the pixel-to-pixel offsets but does not eliminate the temporal noise. However, temporal noise can be addressed by other methods like soft reset or tapered reset.

In a monolithic CMOS sensor, the photodiodes share the pixel area with the transistors. For that reason, the fill factor is always less than 100%. In addition, most CMOS imagers are front side illuminated. This limits the sensitivity in the red because of a relatively shallow absorption material. A typical quantum efficiency (QE) plot for a monolithic CMOS sensor with microlenses can be seen in Figure 8 Joshi et al. (2005). The curve corresponds to a 3T pixel with roughly 50% fill factor. For comparison, the same figure shows two QE plots of silicon PIN detectors. Because they are hybrid FPAs with a dedicated detector layer, the thicker material and 100% fill factor result in much improved red response.

4. Hybrid CMOS technology

The hybrid CMOS revolution began in the mid 1970s when the indium bump interconnect technology was invented. By 1990, CMOS processing of silicon was standard for the commercial analog and digital integrated circuit industry. The CMOS process has continuously improved because of demands of the computer industry, allowing higher-density circuits and large array formats. At the same time, indium bump hybridization has also improved to the point where 16 million pixels per array can be reliably connected.

The hybrid CMOS approach separates the problems of (1) converting electromagnetic radiation into an electronic signal and (2) amplifying the electronic signal,

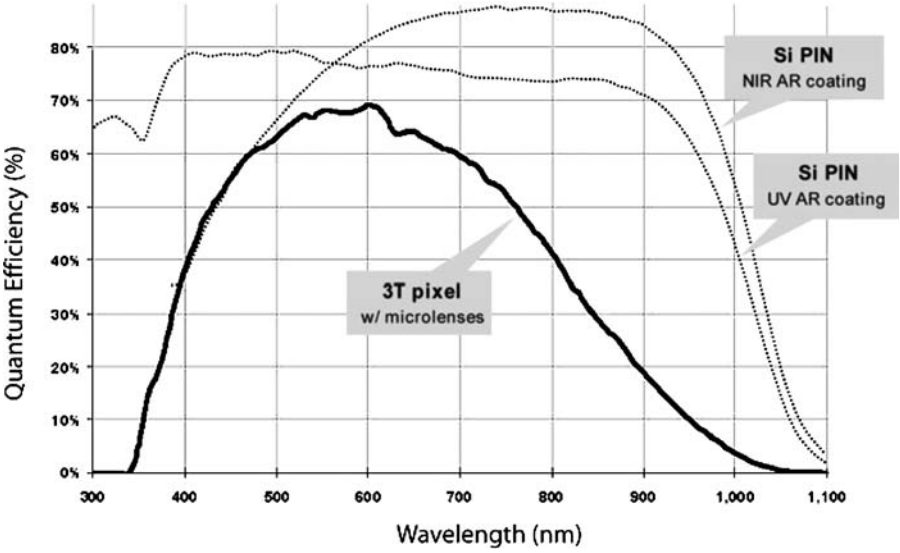

Figure 8. Spectral QE of a monolithic CMOS image sensor with three transistors per pixel.

processing it, and multiplexing the signals from many detector elements onto a single output line. In a hybrid CMOS device, the first function is performed by a detector array that is typically electrically connected to a CMOS integrated circuit chip using an array of indium bumps, one per detector element. The CMOS chip, often called a ROIC (readout integrated circuit), performs the second function. A detector hybridized to a CMOS ROIC may be called a hybrid array, a sensor chip assembly (SCA), or a focal plane (FPA).

Because the detection and electronics functions are separated in a hybrid CMOS SCA, the design and processing of detectors and ROICs can each be optimized for its own function. In addition, detectors and ROICs can often be interchanged, that is, a given ROIC design might be hybridized to different detector arrays depending on the wavelength of interest. Similarly, a given detector array might be hybridized to different ROICs depending on the desired frame rate, well capacity, of other signal processing functions.

4.1. DETECTORS

Some of the detector materials available in large array formats and compatible with the CMOS hybrid approach are shown in Table I. The spectral range of each material is listed along with the approximate operating temperature to achieve much less than 1-electron/second dark current. All of these detector materials have been demonstrated in formats of 1K $\times$ 1K or greater, all have essentially 100% fill factor, and all have very nearly 100% internal QE if the detector is properly

TABLE I
Detector materials available for large-format CMOS hybrids (SCAs)

Detector material	Spectral range[a] (μm)	Operating temperature[b] (K)
Si PIN	0.4–1.0	~200
InGaAs[c]	0.9[d]–1.7	~130
HgCdTe[c]		
1.7 μm	0.9[d]–1.7	~140
2.5 μm	0.9[d]–2.5	~90
5.2 μm	0.9[d]–5.2	~50
10.0 μm	5.0–10.0	~25
InSb	0.4–5.2	~35
Si:As IBC (BIB)	5.0–28.0	~7

(a) Long wave cutoff is defined as 50% QE point.

(b) Approximate detector temperatures for dark currents <<1 electron/second.

(c) Requires special packaging due to thermal contraction mismatch between detector and ROIC.

(d) Spectral range can be extended into visible range by removing substrate.

designed and fabricated. All of these detectors, with the exception of Si:As, are generally produced as p-on-n diode structures and are interchangeable with regard to ROIC hybridization.

One of the design issues facing the hybrid CMOS structure has been how to deal with differences in thermal contraction of the detector and ROIC as the SCA is cooled. For example, if the detector contracts much more than the ROIC, the indium bump connections could be torn apart upon cooling. The problem becomes more severe for larger arrays and for lower operating temperatures. Fortunately, for a thinned detector such as InSb, the membrane-like detector (10 μm thick) is able to stretch to match the contraction of the ROIC. For thick detectors, such as HgCdTe and InGaAs, detector manufacturers have devised mechanical structures to force the ROIC to contract the same amount as the detector material, thus nearly eliminating stress at the indium bump interface. For silicon-based detectors, such as Si PIN and Si:As, there is little differential contraction compared to a silicon CMOS ROIC, and thermal cycle reliability is not a concern.

4.2. ROICs

CMOS technology allows complex input amplifier circuits to be designed to temporally integrate the detector signal. Figure 9 depicts schematic examples of the three most widely used, and simplest, input circuits—the source follower per detector (SFD), the capacitive transimpedance amplifier (CTIA), and the direct injection (DI).

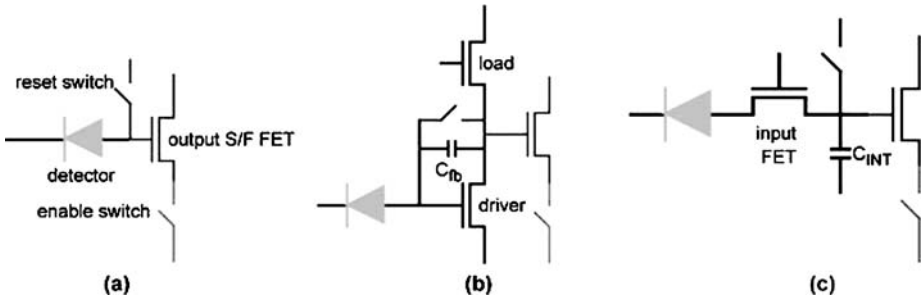

Figure 9. Circuit schematics for the three most commonly used input circuits for CMOS ROICs: *(a)* the SFD, *(b)* the CTIA, and *(c)* the DI.

Table II provides a description of the advantages and disadvantages of each circuit. The SFD is most commonly used in large-format hybrid astronomy arrays as well as commercial monolithic CMOS cameras. It is simple, has low noise, low power, and low FET glow. The CTIA is more complex and higher power but is extremely linear. The DI circuit is used in higher-flux situations and has the same low power and FET glow as an SFD. The gain (volts out / charge in) of the input circuits are determined by charge integration capacitors: for an SFD, it is the detector capacitance; for CTIA and DI, charge is integrated on ROIC capacitors. ROIC capacitors are determined by the layout of the ROIC and can be varied according to the application.

TABLE II
Comparison of attributes of the three most common input circuits

Circuit	Advantages	Disadvantages	Comments
Source follower per detector (SFD), . . . "self-integrator"	Simple Low noise Low FET glow Low power	Gain fixed by detector and ROIC input capacitance Detector bias changes during integration Some nonlinearity	Most common circuit in IR astronomy
Capacitance transimpedance amplifier (CTIA)	Very linear Gain determined by ROIC design (C_{fb}) Detector bias remains constant	More complex circuit FET glow Higher power	Very high gains demonstrated
Direct injection (DI)	Large well capacity Gain determined by ROIC design (C_{INT}) Detector bias remains constant Low FET glow Low power	Poor performance at low flux	Standard circuit for high flux

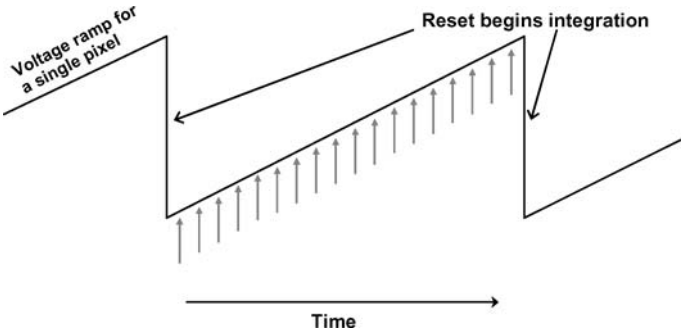

Figure 10. Illustration of multiple sampling of a detector during a long integration. The sawtooth line represents the voltage on the output of a detector input circuit as a function of time. The upwardly rising ramp represents the integration of photocurrent; the sudden downward voltage is the reset that begins the next integration. Arrows show regularly spaced points in time where the integrated signal is sampled.

4.3. Noise reduction techniques for CMOS hybrids

There are three general types of noise in CMOS hybrids: temporal, fixed pattern, and random telegraph signal (RTS). Temporal noise may be divided into white or uncorrelated noise and 1/f noise. This noise may be from the detector and/or ROIC, depending on the design. Fixed pattern noise is caused by residual nonuniformity after calibration, which in turn may be caused by 1/f noise or drift in key system parameters such as focal plane temperature. RTS is randomly occurring charge trapping/detrapping events and is highly dependent on CMOS process and design parameters.

White temporal noise has usually been the dominant type of noise in CMOS devices used astronomy. A common technique to reduce this noise is to sample each detector element multiple times during a long integration time, as illustrated in Figure 10.

There are two common methods of analyzing multiply sampled data. One is Fowler sampling, where the first N samples and the last N samples are each averaged and the two averages are then subtracted. Noise is reduced by root N if sample-to-sample noise is uncorrelated. This technique also performs a CDS function, reducing noise introduced by resetting the integration capacitor (kTC and other noise sources). The second technique is to perform a least squares fit of all samples to a straight line or polynomial.

Figure 11 compares the signal-to-noise ratio (SNR) for Fowler sampling with the sample-up-the-ramp (SUTR) noise reduction technique for the theoretical case where all samples are uncorrelated. This example uses a total of 64 samples scaled so that Fowler-1 (CDS) SNR is set to 1. The SNR of the Fowler sampling is a function of sample pairs with 32 being the maximum in this case. A peak in SNR where the

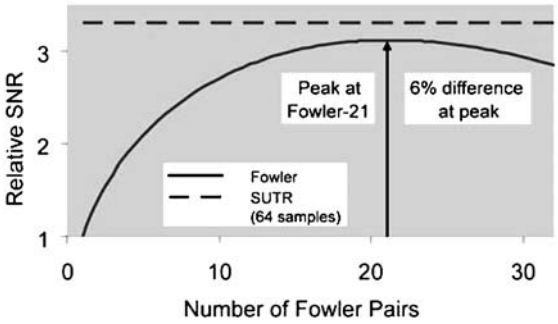

Figure 11. Example of SNR improvement with multiple uncorrelated samples. SNR as a function of Fowler sample pairs is shown in comparison to SUTR. In this example, there are a total of 64 samples spaced over equal time intervals. SNR is scaled for SNR = 1 for Fowler-1 (CDS). The peak in Fowler sampling occurs when the number of Fowler pairs is one third of the total number of samples and the middle third of the samples are ignored. This peak is 6% below the SUTR SNR value.

number of Fowler pairs is one third of the total number of samples (take the last third of the samples and subtract from the first third while ignoring the middle third) and is a general result. The Fowler SNR peak is 6% below the SUTR peak. Detector dark current and photocurrent are integrated, and therefore later samples are correlated to earlier ones. In the limit of background-limited performance (BLIP), Fowler-1 (simple CDS) yields the best SNR, which is 9% better than SUTR (Garnett and Forrest, 1993). Noise as low as 2 electrons rms has been achieved with cryogenic (30 K) CMOS ROICs in a Fowler-16 mode McMurtry.

5. CMOS fabrication processes

The attraction of CMOS-based process technologies for imaging applications is that they can bring control electronics and a sophisticated range of signal processing tools in very close proximity to the focal plane. Still, CCD fabrication methods have been steadily optimized over the past 35 years to produce large-format devices with broad spectral response, low noise readout, and good yield. In this section we compare the process technologies used for CCD- and CMOS-based imagers and describe why the newest advancements in CMOS processes offer some attractions for high-performance imaging.

5.1. CCD vs. CMOS fabrication process comparison

CCD imagers are generally fabricated on 125-mm silicon substrates in facilities that are "trailing" in feature-size advancements relative to the most aggressive CMOS foundries. The circuits fabricated in CMOS-based process technologies benefit from improvements encouraged by a wide-scale demand for fast, compact, low-power electronics. The concurrent economics of scaling wafer size up and feature size

down propel frequent migrations to new technology "nodes," which correspond to logic speed improvements. Many CMOS foundries are routinely processing 200-mm wafers, with the newest facilities now handling 300-mm wafers, and most research laboratories using 150-mm wafers. Commercial ventures benefit from wafer size increase because the cost per chip generally goes down, but even research efforts are eventually forced to migrate up in wafer size because of the availability of high-quality substrates and advanced deposition, etch, and lithography tools.

Manufacturers of CCDs may eventually feel this pull as well, though presently the very high resistivity float-zone silicon substrates (5000–7000 Ω-cm) required for deep depletion and good spectroscopic performance at long wavelengths are only available up to 200-mm diameter, and the commonly used full-wafer lithography tools are not yet available for the larger wafer sizes. Since float-zone silicon is vulnerable to the generation of plastic slip and dislocations and requires special processing techniques that limit thermally induced stresses during high-temperature processes Gregory et al. (1996), migrating a science-grade CCD process technology to an even larger wafer size would require reassessment of all thermal processing as well as design/layout methods to minimize stresses induced by multiple layers of patterned polysilicon crossing channel stops.

In its highly evolved present state, the common implementation of a CCD imager structure does produce a satisfactory yield of millions of active elements in a single device. The buried-channel structure, multiple oxidation cycles, and composite gate dielectric are all engineered to have superior charge transfer efficiency, noiseless operation (until readout), and resistance to interpoly and substrate shorts. However, CCDs generally require clocks of at least 10 V to maintain sufficient charge capacity in the well. These 10-V clocks are incompatible with the 1- to 5-V levels necessary for low-power on-chip circuitry. Shallower CCD buried channels and thinner gate dielectrics can produce imagers that operate at lower voltages Suntharalingam et al. (2000), and such improvements have the greatest benefit if accompanied by on-chip access to digital logic gates as well.

Most mixed-signal CMOS technologies easily support implementation of single chip designs with multiple operating voltages, for example, to manage dynamic range and power. A further aid to the circuit designer is the presence of several levels of interconnect metal (typically five to seven), which permits great flexibility in routing methodology. The highly nonplanar structure of a CCD hampers reliable patterning and isolation of even one or two levels of metal interconnect. An extreme case of this is seen in Figure 12, which presents an early optical micrographic view of the four-poly pixel array and a scanning electron cross section through the pile-up of layers that cross the channel stop. Each reentrant edge seen is a vulnerable spot for interpoly shorts and a trap for hidden metal shorts.

The structure of CMOS devices can be appreciated in the scanning electron micrographs of Figure 13, which provide a detailed view of a five-transistor SRAM cell and a cross section through three interconnect metal layers with planarized interlevel dielectrics.

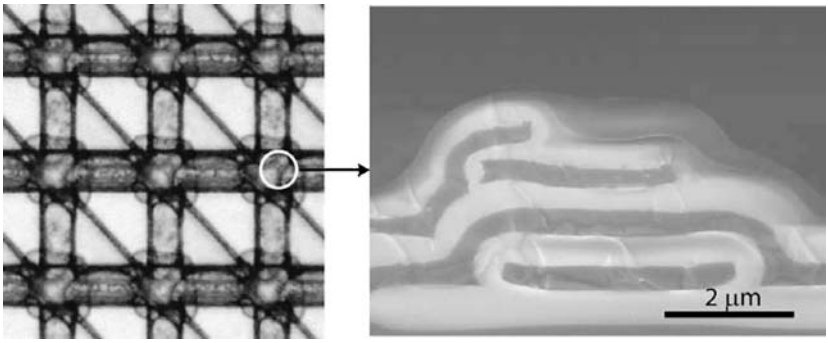

Figure 12. Optical photomicrograph and cross section through channel stop region of four-poly orthogonal-transfer CCD (OTCCD).

Manufacturers of CCDs are choosing to implement some of these CMOS-developed processes when appropriate, but one remaining comparison is architecturally, rather than process, driven—the repetitive transfer of charge packets through the silicon lattice in a CCD causes the device to be vulnerable to space-radiation-induced traps. CMOS-based active pixel imagers require few to no charge transfers and can readily adopt a number of design and process hardening methods to yield science-grade imaging sensors that are suitable for long-term space-based applications.

5.2. CMOS TECHNOLOGY FEATURES AND GENERAL LIMITATIONS OF MONOLITHIC IMAGERS

CMOS device size and voltage scaling are driven by logic applications that demand ever-faster computation with reduced power dissipation. In each generation of

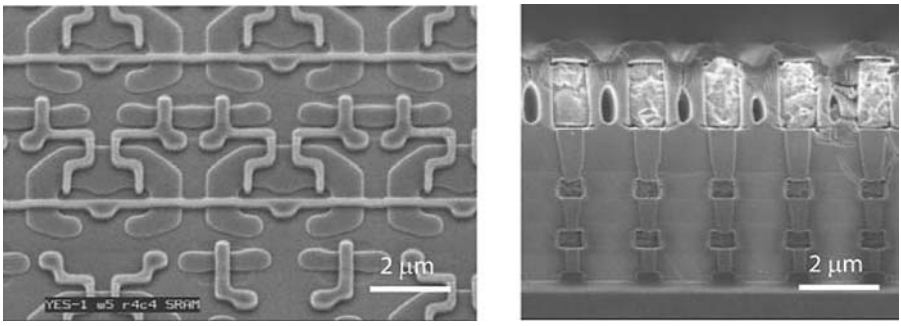

Figure 13. (left) Silicon-on-insulator (CMOS) 180-nm SRAM cell and *(right)* three levels of metal stacked with vias to poly.

TABLE III
Summary of CMOS process technology features

Feature	0.35–0.60-μm gate length technology	0.18–0.25-μm gate length technology
Operating voltage	3.3–5 V	1.8–2.5 V
Field isolation	LOCOS	Shallow trench isolation
Gate oxide	70–125 Å	32–50 Å
Gate dielectric	Silicon dioxide	Nitrided silicon dioxides
Junction profile	Graded junction	Shallow junction
Thermal budget	Furnace anneal	Rapid thermal processing
Spacer etch	Oxide spacer	Silicon nitride spacer
Device	Polycide/poly	Self-aligned silicide
Planarization	Spin-on-glass and reflow	Chemical mechanical planarization

fabrication technologies a number of process features change, as detailed in Table III. For example, the local oxidation process (LOCOS) used for device isolation in both CCD and larger-geometry CMOS technologies is replaced by a more compact shallow trench isolation (STI) process that permits much higher circuit layout densities and better junction isolation.

To produce the same transistor channel inversion charge the lower operating voltage requires higher gate capacitance, and so the gate dielectric is made thinner and nitrogen may be introduced to inhibit dopant penetration from the gate into the silicon channel. For high transistor on-state current with well-behaved off-state current, the shallow device junctions have sharply defined doping profiles fixed by rapid thermal activation of the extension and source/drain implants.

However, many of the process modules that become standard for highly scaled CMOS, such as STI, thin gate dielectric, and silicided junctions, introduce additional sources of device and junction leakage, and the optically opaque silicides can block photoabsorption in the silicon. While anisotropic physical profiles are obtained with now-common plasma (dry) etch processes, the surface damage created may not be repaired within the strict thermal budget. Even low levels of surface damage or stress-induced stacking faults like those seen in Figure 14 can create leakage that contributes to dark current.

To support the fabrication of low-cost CMOS image sensors for consumer applications, many CMOS logic foundries now offer special steps that protect the sensitive pixel transistors from some of the most damaging exposures. The process flow shown in Figure 15 Wuu et al. (2000) illustrates how pixel transistors are coated with an additional oxide layer, which is etched back and patterned to permit silicide formation only on the gate polysilicon and on the source/drain regions of peripheral devices.

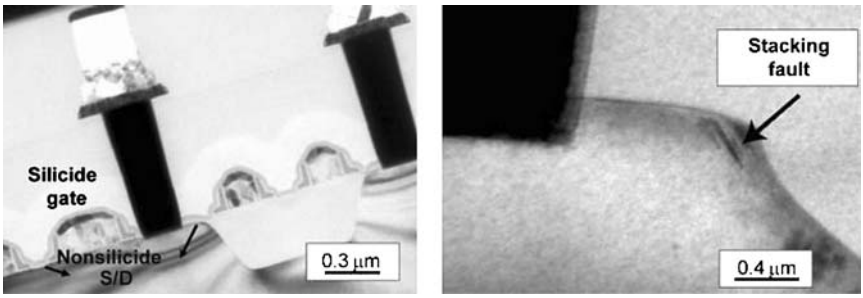

Figure 14. Cross-sectional transmission electron micrograph of pixel. From Wuu et al. (2000).

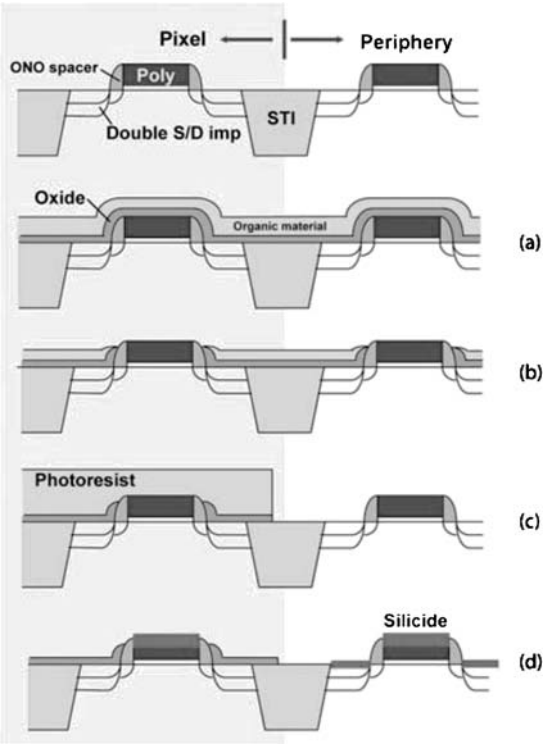

Figure 15. CMOS pixel process flow: *(a)* deposit oxide and spin coat organic material, *(b)* etch back and remove oxide, *(c)* remove organic material and pattern oxide (photo/etch), and *(d)* form silicide on peripheral devices. Adapted from Wuu et al. (2000).

6. Emerging technologies

For monolithic active pixel architectures the pixel area must be shared between the photodiode and the access and amplification transistors, as seen in the plan view and schematic cross-sectional view of Figure 16. The photo-collection junction is typically formed at the drain-substrate or well-substrate interface. It is clearly seen

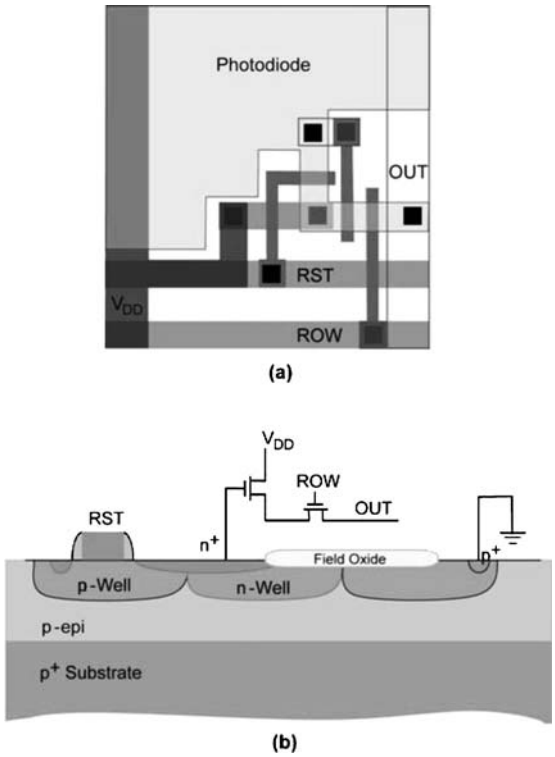

Figure 16. Example of three-transistor pixel layout *(a)* in plan view and *(b)* in cross-section.

in these views not only that the fill factor of the pixel is less than 100% but also that the fabrication of the photojunction and access transistors must be co-optimized, with some necessary compromises on layout density, absorption layer thickness, and device isolation. Foundries targeting consumer applications typically apply microlenses to compensate for the loss of optical fill factor.

6.1. VERTICALLY INTEGRATED ACTIVE PIXEL IMAGER

Without further limiting the pixel fill factor, monolithic architectures also require that addressing and signal processing circuitry be placed at the periphery of the array. In contrast, a three-dimensionally (3-D) stacked circuit construct, such as shown in Figure 17, relieves many of the limitations inherent to monolithic structures. Active-pixel focal plane architectures are well suited for 3-D interconnection because signal integration, amplification, and readout can be in close proximity to the photodetection elements while still achieving 100% optical fill factor. The further capability to perform complex signal processing behind every pixel can dramatically reduce total image sensor power and bandwidth requirements.

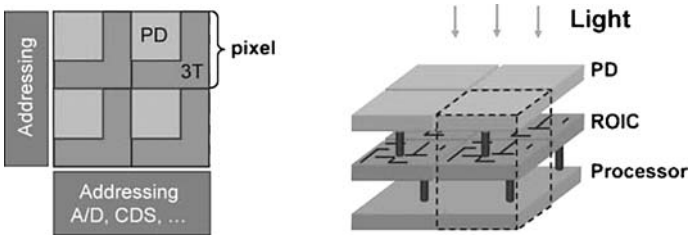

Figure 17. Advantages of vertical integration: *(left)* conventional monolithic APS compared with *(right)* 3-D pixel.

Vertically hybridized flip-chip imagers already offer independently optimized photodetector and readout multiplexer designs that can achieve scientific-grade image sensor performance Bai et al. (2004). However, these bump-bonded approaches are limited to two circuit layers, to large pixel sizes ($\geq$18 μm), and do not permit post-integration hydrogen-passivation anneals, which are critical for dark current suppression. In Figure 18 a feature-size comparison is made among three methods to vertically interconnect circuit layers: (a) bump bond, (b) insulated through-silicon vias, and (c) Lincoln Laboratory's SOI based via. The Lincoln integration method is extendable to three or more circuit layers and is capable of achieving far smaller pixel sizes than possible with bump bonding.

The process technology for the method in Figure 18(c) has recently been used to demonstrate a four-side abuttable 3-D integrated 1024 $\times$ 1024, 8-μm pixel visible image sensor fabricated with oxide-to-oxide wafer bonding and 2-μm-square 3-D vias in every pixel. The 150-mm wafer technology integrates a low-leakage, deep-depletion, 100% fill factor photodiode layer to a 3.3-V, 0.35-μm gate length fully depleted (FD) SOI CMOS readout circuit layer Suntharalingam et al. (2005).

A cross-sectional scanning electron micrograph (SEM) through several 8-μm pixels of a functional active pixel imager is shown in Figure 19. The oxide-oxide bond between the two tiers is imperceptible. A 3-D via connects tier-2 FDSOI CMOS metal 3 to tier-1 (diode) metal 1, and a metal cap (back metal 1) covers the 3-D via plug. The 50-nm-thick SOI transistor features can be seen near the top of the SEM. The dominant misalignment ($\sim$1 μm) is created by the wafer-to-wafer bonding step; newer tools and methods in development are expected to further reduce this misalignment.

Each sensor contained over 3.8 million transistors and over one million 3-D vias. We measured pixel operability in excess of 99.9% with the principal yield detractor arising from column or row dropouts, i.e., not 3-D vias. The devices have successfully been processed through diode wafer thinning for back-illumination operation.

The high degree of pixel functionality can be seen in Figure 20, which presents an image acquired by projecting a 35-mm slide onto either front-illuminated or back-illuminated processed imaging devices (different chips).

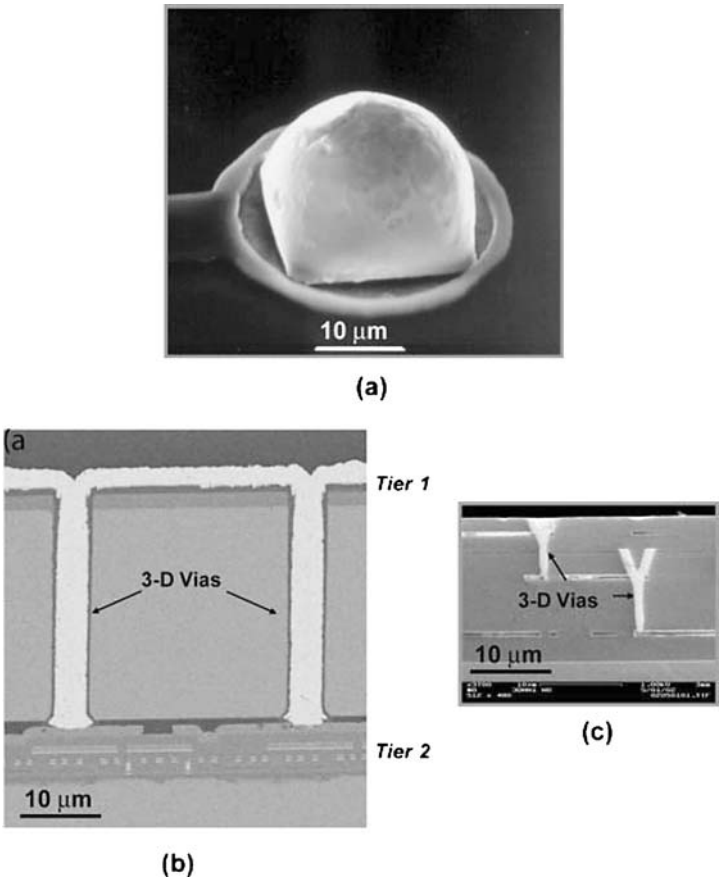

Figure 18. Approaches to 3-D integration: *(a)* Bump bond used to flip-chip interconnect two circuit layers, *(b)* two-layer stack with insulated vias through thinned bulk Si, and *(c)* two-layer stack using Lincoln Laboratory's SOI-based vias. Illustrations are to scale. Photo in (b) courtesy of RTI.

6.2. DIGITAL FOCAL PLANE ARCHITECTURES

Conventional analog focal plane architectures reach a readout bottleneck when the system implementation demands wide-area coverage, high frame rates, and high bit precision. By digitizing the signal while the photoelectrons are being collected, rather than after charge accumulation, the need for large charge storage capacitors and highly linear analog electronics can be eliminated. Additionally, the power dissipation and noise problems, which result from communicating analog signals across an imager at high data rates, are greatly reduced.

Using hybridization methods, Lincoln Laboratory has demonstrated 32 × 32-pixel focal planes based on Geiger-mode avalanche photodiodes (APDs), which can detect a single photon and produce a digital logic pulse directly from the

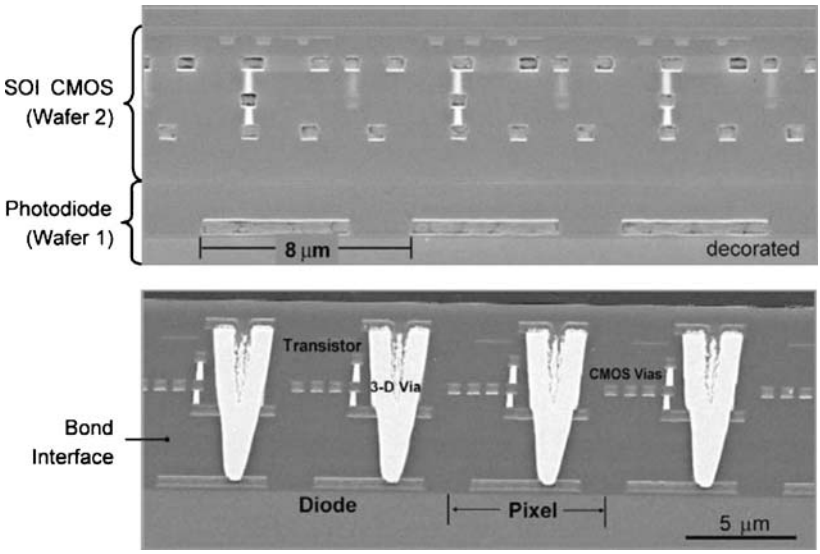

Figure 19. Cross sections through 3-D imager.

detector Aull et al. (2004). As shown in Figure 21, the detector is connected directly to a CMOS inverter, the output of which serves as a stop signal to a digital timing circuit, or in the case of intensity imaging as an increment signal to a counter. This digital-domain design yields power savings, noiseless readout, and quantum-limited sensitivity. Multiple operating voltages for arming the APD and operating the low-power digital circuitry can be readily accommodated. By using the SOI-based, oxide-bonded, micron-scale 3-D interconnection technology previously described, the pixel pitch of the array can be further reduced and the timing resolution improved, to produce a scalable architecture.

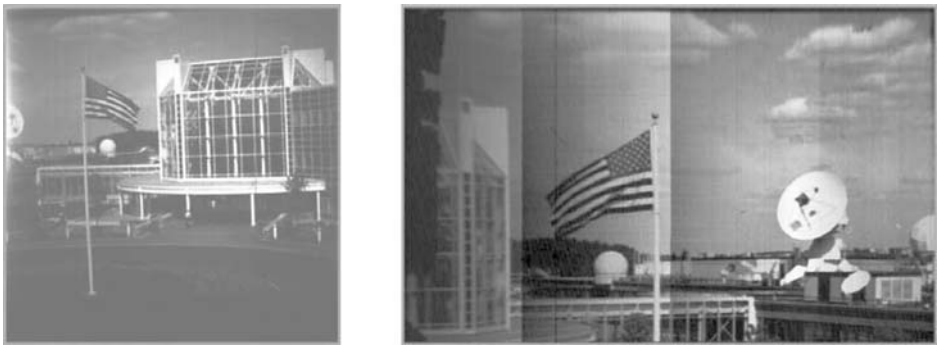

Figure 20. Four-side abuttable vertically integrated imaging tile: *(left)* front illuminated and *(right)* back illuminated (partial frame of top 700 rows).

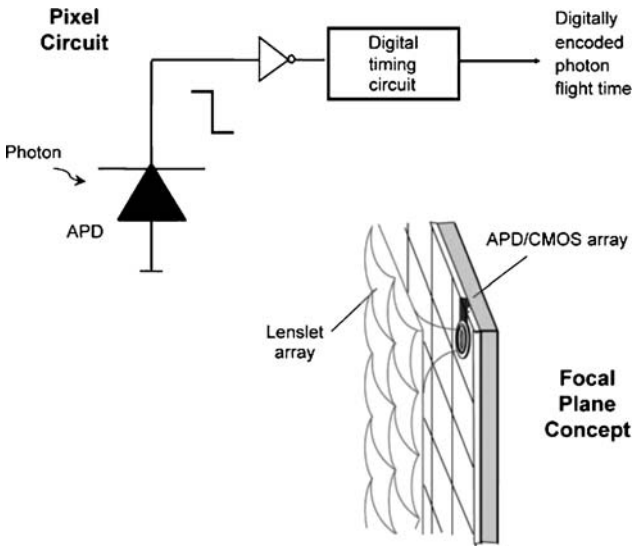

Figure 21. Geiger-mode imager illustrating photon-to-digital conversion.

Acknowledgments

We are most grateful to all who assisted in the preparation of this review paper by sharing their data and insights, either from prior publications or personal communications. Karen Challberg of MIT Lincoln Laboratory provided extremely responsive and meticulous editorial assistance during the assembly of this paper, and for that we are most appreciative.

References

Aull, B. F., Loomis, A. H., Young, D. J., Stern, A., Felton, B. J., Daniels, P. J., Landers, D. J., Retherford, L., Rathman, D. D., Heinrichs, R. M., Marino, R. M., Fouche, D. G., Albota, M. A., Hatch, R. E., Rowe, G. S., Kocher, D. G., Mooney, J. G., O'Brien, M. E., Player, B. E., Willard, B. C., Liau, Z.-L. and Zayhowski, J. J.: 2004, *Three-dimensional imaging with arrays of Geiger-mode avalanche photodiodes*, Proc. SPIE, vol. **5353**, pp. 105–116.

Bai, Y., Bernd, S. G., Hosack, J. R., Farris, M. C., Montroy, J. T. and Bajaj, J.: 2004, *Hybrid CMOS focal plane array with extended UV and NIR response for space applications*, Proc. SPIE, vol. **5167**, pp 83–93.

Garnett, J. D. and Forrest, W. J.: 1993, *Multiply sampled read limited and background limited noise performance*, Proc. SPIE, vol. **1946**, pp. 395–404.

Gregory, J. A., Burke, B. E., Cooper, M. J., Mountain, R. W. and Kosicki, B. B.: 1996, *Fabrication of large-area CCD detectors on high-purity float zone silicon*, Nucl. Instrum. Methods Phys. Res. A, vol. **377**, pp. 325–333.

Joshi, A., Chiaverini, D., Jung, K., Douence, V., Wijeratne, G., Williams, G. M. and Loose, M.: 2005, *Scalable architecture for high-resolution video-rate CMOS imaging system on chip*, IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, pp. 181–184.

Kozlowski, L., Rossi, G., Blanquart, L., Marchesini, R., Huang, Y., Chow, G. and Richardson, J.: 2005, *A progressive 1920 × 1080 imaging system-on-chip for HDTV cameras*, IEEE International Solid-State Circuits Conference, Digest of Technical Papers, pp. 358–359.

Loose, M., Lewyn, L., Durmus, H., Garnett, J. D., Hall, D. N., Joshi, A. B., Kozlowski, L. J. and Ovsiannikov, I.: 2003, *SIDECAR low-power control ASIC for focal plane arrays including A/D conversion and bias generation*, Proc. SPIE, vol. **4841**, pp. 782–794.

McMurtry, C. W.: University of Rochester, private communication. Data shown in Hoffman, A. et al., 2005, 1024 × 1024 *Si:As IBC detector arrays for mid-infrared astronomy*, SDW 2005 Proceedings, Springer, Netherlands.

Suntharalingam, V., Berger, R., Burns, J. A., Chen, C. K., Keast, C. L., Knecht, J. M., Lambert, R. D., Newcomb, K. L., O'Mara, D. M., Rathman, D. D., Shaver, D. C., Soares, A. M., Stevenson, C. N., Tyrrell, B. M., Warner, K., Wheeler, B. D., Yost, D. R. W. and Young, D. J.: 2005, *Megapixel CMOS image sensor fabricated in three-dimensional integrated circuit technology*, IEEE International Solid-State Circuits Conference, Digest of Technical Papers, pp. 356–357.

Suntharalingam, V., Burke, B., Cooper, M., Yost, D., Gouker, P., Anthony, M., Whittingham, H., Sage, J., Rabe, S., Chen, C., Knecht, J., Cann, S., Wyatt, P. and Keast, C.: 2000, *Monolithic 3.3 V CCD/SOI-CMOS imager technology*, International Electron Devices Meeting Technical Digest, pp. 697–700.

Wuu, S. G., Yaung, D. N., Tseng, C. H., Chien, H. C., Wang, C. S., Fang, T. K., Wang, C. C., Sodini, C. G., Hsiao, Y. K., Chang, C. K. and Chang, B. J.: 2000, *High performance 0.25-μm CMOS color imager technology with nonsilicide source/drain pixel*, International Electron Devices Meeting Technical Digest, pp. 705–708.

On their free day, the participants paid an emotional visit to the burial place of deceased astronomical detectors. (Courtesy: Sandro D'Odorico).