Charge coupled CMOS and hybrid detector arrays

James Janesick♠
Sarnoff Corporation, 4952 Warner Ave., Suite 300, Huntington Beach, CA. 92649
Headquarters: CN5300, 201 Washington Road Princeton, NJ 08543-5300

ABSTRACT

Over a decade has passed since complementary metal oxide semiconductor (CMOS) imaging detectors made their move into the charge-coupled device (CCD) arena. Low cost, low power, on-chip system integration, high-speed operation and tolerance to high-energy radiation sources are unique features that make CMOS detectors popular. However, it remains unclear if CMOS arrays can compete with the CCD in high performance applications (e.g., scientific). This paper compares fundamental performance parameters common to both CMOS and CCD imagers, and lists specific CMOS performance deficiencies that prevent the technology from high end use. In this paper we will present custom CMOS pixel designs and related fabrication processes that solve most deficiencies. We will also discuss "hybrid" imaging arrays that marry the advantages of CCD and CMOS, producing sensors with superior performance in comparison to CCD and CMOS bulk monolithic sensors. CCD to CMOS, CMOS to CMOS and CMOS SOI hybrids are reviewed.

Keywords: charge coupled devices, pinned photo diode, active CMOS pixels, hybrid imagers, SOI hybrids

1. CCD AND CMOS IMAGER DEFICIENCIES

For over 20 years, charge-coupled devices (CCDs) have dominated most digital imaging applications and markets. Very few performance deficiencies are associated with current state-of-the-art CCDs.[1] The combination of near-ideal performance and implementation has made the CCD a highly successful imaging sensor, however, specific fundamental weaknesses are associated with the technology. First, the CCD is difficult to apply in high-energy radiation environments. Second, the CCD requires a significant amount of support electronics for operation. Third, high-speed readout is difficult for the CCD because signal charge, for the most part, must be readout in a serial fashion as opposed to parallel or random access pixel acquisition.

Complementary metal oxide semiconductor (CMOS) imagers are displacing CCDs in some imaging applications, and this trend is expected to continue. Coincidently, CCD deficiencies listed above are the main strengths for CMOS sensors. CMOS technology is considerably more tolerant to high-energy radiation environments. CMOS inherently allows on-chip system integration designs that reduce camera size, power, weight and cost while increasing reliability and lifetime. Lastly, CMOS arrays read pixels in a parallel, random access fashion, allowing high-speed operation and low noise performance. However, current CMOS technology is lacking in nearly every performance category compared to the CCD.

CMOS deficiencies are best discussed by reviewing the four operational tasks of generating an image: charge generation, charge collection, charge transfer, and charge measurement. Charge generation is the sensor's ability to intercept incoming photons and generate signal charges through the photoelectric effect. An ideal sensor would have 100% quantum efficiency (QE) at all wavelengths. To achieve high response, sensor manufacturers must minimize three QE loss mechanisms: absorption, reflection, and transmission. Absorption loss is associated with optically dead structures, typically located above but also within the pixel. Reflection and transmission losses are inherent to the physical properties of silicon. CMOS arrays experience greater absorption loss than CCDs because MOSFETs incorporated in each pixel for readout are for the most part are optically dead. CMOS sensors require several metal layers to interconnect the MOSFETs. The metal bus lines are stacked and interleaved above the pixels, producing an optical tunnel through which incoming photons must pass. The metal stack is typically several microns high and creates a host of undesired

♠ e-mail: CMOSCCD@AOL.COM, Tel: 714-377-6223

Focal Plane Arrays for Space Telescopes, edited by Thomas J. Grycewicz, Craig R. McCreight,
Proceedings of SPIE Vol. 5167 (SPIE, Bellingham, WA, 2004)
0277-786X/04/$15 · doi: 10.1117/12.511864

optical effects including light scattering, vignetting, non-symmetric pixel cross talk, and diffraction problems. Light shields above the pixel can somewhat control the problem with a reduction in pixel fill factor and lower QE performance. Response nonuniformity across a pixel is problematic for applications requiring a symmetric point spread pixel response (e.g., wave front and star tracker sensors). Figure 1 shows a collection of QE curves taken from various detectors including a CMOS array with micro lens.[2] Note that the backside illuminated CCD achieves the highest response and the CMOS the lowest. Figure 2 displays the results of scanning a focused spot (< 2 μm) in two dimensions across a single 9 μm CMOS pixel.[3] The spot step size utilized was 0.3 μm producing approximately 900 data points in the image. MOSFET and metal traces contained within the pixel absorb incoming photons resulting in a non-uniform response.

Charge collection, the second operational task of an imager, determines the sensor's ability to reproduce an image. Charge collection efficiency (CCE) is a critical parameter because it defines the spatial resolution of the detector. Ideally, signal electrons generated in the silicon should remain in the target pixel. Thermal diffusion and weak electric fields within a pixel's active volume cause signal electrons to wander into neighboring pixels, creating cross talk and related modulation-transfer-function (MTF) loss, and poor color performance. Therefore, it is important that the charge-collecting electric fields within the photo region of the pixel be sufficiently deep to match the photon absorption depth. Pixel cross talk is most conspicuous for near-IR and soft x-ray photons that penetrate deep into the sensor, where weak electric fields exist. Also, pixel cross talk increases dramatically as pixel size is reduced. Figure 3 compares 430 nm and 630 nm flat field responses for a CMOS test pixel array that incorporates 4 μm pixels. Very little charge diffusion and cross talk is seen within the 5 x 5 pixel light shielded region when stimulated with 430 nm light. At this wavelength, photons are absorbed at the immediate surface of the pixel where collecting fields are strongest. On the other hand, the 630 nm response shows significant cross talk because the photons penetrate deeper than the pixel's collecting fields. The test array also contains different light shield arrangements (labeled Shield 1, Shield 2, etc.) to direct incoming photons into the sweet spot of the pixel. The shields do not significantly lower the amount of cross talk measured at long wavelengths, only a reduction in QE performance is seen.

CCD manufacturers have minimized charge diffusion by using high-resistivity epitaxial silicon wafers and high-voltage clocking to exploit the fact that the electric field depth varies as a function of the square root of resistivity and applied voltage. Electric fields typically extend to 7 to 10 μm, allowing visible and near-IR spectral coverage (400 to 1100 nm). In contrast, CMOS arrays show relatively poor CCE performance because standard foundry processes use low-resistivity silicon wafers (typically < 10 ohm-cm) and low voltage drive (< 3.3 V). Low-resistivity material is necessary to prevent cosmic-ray/radiation-event-triggered CMOS circuit latch-up and ground-bounce problems associated with support CMOS electronic circuitry. Also, low-voltage operation inherent to CMOS, reduces electric field depth. This problem is becoming severe because operating voltages decrease proportionally to feature size. For example, state-of-the-art 0.18-μm CMOS processes operate at only 1.8 V. Design rules and operating voltages will decrease as CMOS technology advances. To compound the problem, the depletion region related to the photo region decreases as signal charge collects. Figure 4 plots depletion depth for a photo diode CMOS pixel as a function of silicon resistivity and photo diode bias. Two sets of curves are shown for two different n-p junction depths (1 and 2 μm). Note that 60 ohm-cm silicon resistivity and a 2 μm deep junction are required to cover the visible spectrum (i.e., 400 to 700 nm). As discussed below, employing a deep n-well process, offered by some CMOS foundries, can make the n-p diode junction very deep. Junction depths of 3-4 μm are achieved using high-energy (2 MeV) implants. However, deep n-well processes are only compatible for large pixel designs (> 8μm). The diffusion cross talk problem can also be improved by using thinner epitaxial silicon at the expense of losing red sensitivity. For example, Figure 5 shows MTF data for a 12 μm CMOS pixel.[3] As expected, MTF degrades for longer wavelengths because of charge diffusion. However, the MTF stabilizes for wavelengths greater than 700 nm as photons penetrate beyond the silicon epitaxial interface into the substrate region where photoelectrons recombine.

The third function, charge transfer, is critical to CCD operation. CMOS pixels are directly addressable and are not as sensitive to charge transfer problems. However, charge transfer CMOS pixels must deal with a serious problem called image lag. Although image lag can be experienced for all CMOS pixels, the photo gate (PG) pixel inherently shows the effect (refer to Figure 8 for the PG pixel). This is because the gap between the PG and transfer gates is implanted with n^+ dopant to conductively short the region for charge transfer. The n^+ region contains a free supply of electrons that mix with signal electrons, which cause a charge transfer problem. Figure 6a illustrates the physics behind the problem, showing how charge is transferred from the PG to the sense node. The transfer gate potential initially defines the

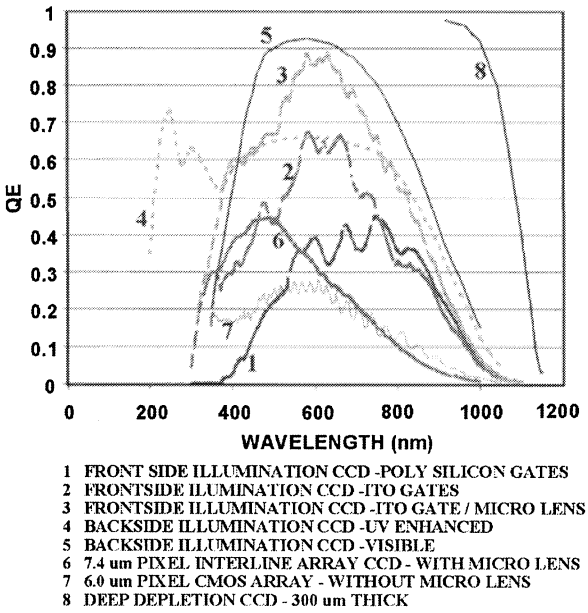

Figure 1: Quantum efficiency curves for various CCD and CMOS sensors. Highest response is obtained by a backside-illuminated CCD array. The CMOS sensor exhibits lowest sensitivity because of a low fill factor and thin epitaxial silicon used.[2]

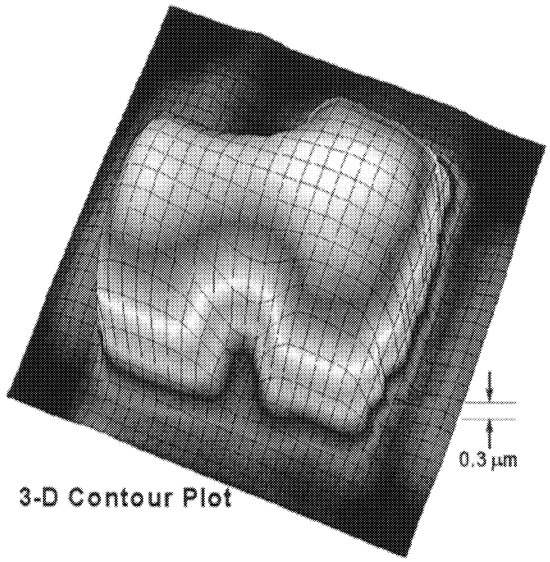

Figure 2: CMOS 9 um pixel response spot scanned with a He-Ne laser showing nonuniform sensitivity caused by MOSFETs and metal traces photon absorption.[3]

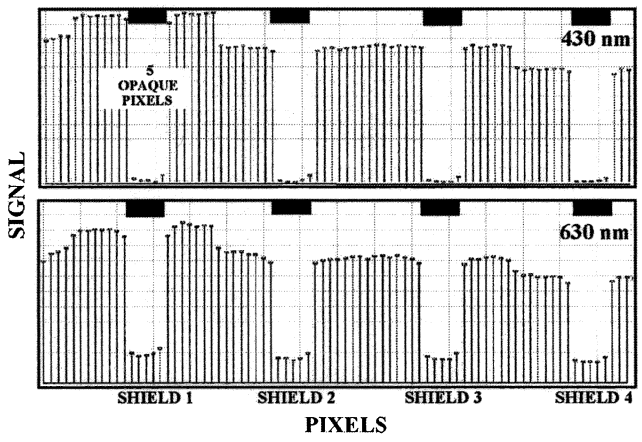

Figure 3: CMOS pixel cross talk measurements taken at 430 and 630 nm wavelengths. The signal measured within the five opaque pixels indicates charge diffusion from neighboring pixels. Different light shields with different fill factors show very little cross talk improvement at long wavelengths.

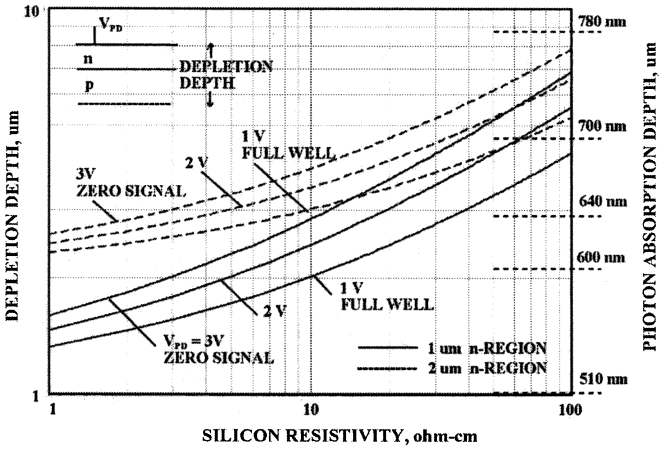

Figure 4: Depletion depth for a photo diode pixel as a function of silicon resistivity under zero charge and full well conditions. Analysis is for photo diode n-regions with 1 and 2 um depths. Photon absorption length for selected wavelengths is also shown. Conventional CMOS pixels ultilize < 10 ohm-cm silicon and 1 um n-well processing that limits depletion depth to less than 2 um under full well conditions.

potential of the n^+ region when it is clocked high. During this time, electrons from the n^+ region thermally jump over the transfer gate potential onto the sense node, charging it negatively. The illustration shows the amount of n^+ discharge for times up to 100 ms. The n^+ region leaves a potential pocket that acts as a charge trap, greatly degrading charge transfer performance for low light level imagery. Figure 6b shows the output voltage of a test PG pixel, showing the long time constant involved with the discharge process. Image lag is most noticeable when a dark scene follows a bright scene where a residual image is seen. The figure also shows a 1000 e- square wave response demonstrating how the sense node charges and discharges in response to a light source that is turned on and off. Image lag is unacceptable for most applications and can be eliminated by driving the PG and transfer gates with a clock voltage greater than V_{DD} (referred to as "hard reset" described below). However, this mode of operation results in high reset noise depending on the capacitance associated with the PG. As we will discuss below, it is best to eliminate image lag by removing the n^+ region and transfer charge another way.

The last imaging operation involves measuring the signal charge contained in each pixel. Readout for CMOS and CCD imagers is the same, and in theory, can achieve similar noise floors (i.e., a few electrons). Both imagers use a sense node capacitor to convert charge to voltage and a source follower MOSFET amplifier to buffer this output voltage. However, CMOS designers must contend with other noise sources, which are not normally a problem for CCDs. For example, the photo diode pixel discussed below with its three-transistor readout architecture, is limited by reset noise, a large noise component that is generated when the pixels are reset. In contrast, the serial readout nature of the CCD allows reset noise to be removed by correlated double sampling (CDS). CMOS sensors must also contend with numerous electrical ground-bounce noise problems generated by on-chip timing logic and ADC circuitry. These system noise sources are difficult to control and often limit the sensor's noise floor above the reset noise level. Optimizing on-chip ADC circuitry without quantizing or bit-weighting noise can also represent a problem. For example, Figure 7a shows a photon transfer curve taken from a VGA CMOS array with an on board 10-bit ADC.[4] The undesired response is caused by a subtle bit-weighting ADC problem. Figure 7b is a transfer curve for the ADC showing that the bit-weighting problem occurs every 16 digital numbers (DN). Bit-weighting problems like this limit digital conversion to no more than 12-bits of resolution for monolithic CMOS sensors. Thermal dark-current is also considerably higher for CMOS arrays, exhibiting levels of 1000-2000 pA/cm^2 at 300 K. Also, CMOS and CCD arrays show thermally generated dark spikes, or hot pixels, which create "salt and pepper" in an image.[1] The problem is most pronounced at high operating temperatures or when long exposures are taken.

2. CONVENTIONAL CMOS PIXELS

Four popular CMOS pixel architectures are illustrated in Figure 8: photodiode (PD) pixel, pinned-photodiode (PPD) pixel, photogate (PG) pixel, and charge-share (CS) pixel. The PD is considered a passive pixel because signal charge is collected and read from the same region of silicon (i.e., charge is not transferred). PPD and PG architectures are called active pixels because charge is transferred from a collecting region to a readout region. The CS pixel is semi-active in that charge is transferred but shared between the collecting and readout regions. Note that the PPD, PG, and CS pixels are extensions of the PD pixel. That is, the three MOSFETS used for PD readout are also required to read out charge-transfer pixels.

Most CMOS imagers are based on the PD pixel because of design simplicity and conformity to conventional CMOS processes. As Figure 8 shows, the PD pixel is comprised of four major components. The PD region is responsible for generating and collecting signal charge. Three MOSFETs are required to measure the charge on the diode: (a) a source-follower MOSFET, which converts signal charge to an output voltage, (b) a reset MOSFET, which resets the photodiode before charge is integrated on the diode, and (c) a row-select MOSFET, which selects a line for scanned readout. The PD pixel has many desirable qualities, including high UV QE, good fill factor, low cross talk for visible wavelengths, high full well, and anti-blooming. High reset read noise is a weakness because the architecture is incompatible with CDS signal processing. As discussed below, the PD pixel uses "rolling shutter" readout, which produces a reset noise level of 20-70 e- depending on the size of the pixel. Reset noise can be reduced by various techniques, however, these methods typically fall short of what true CDS can achieve and are usually plagued with image lag problems.[5,6] PD pixels exhibit dark-current generation rates that range from 0.2 to 2nA/cm2, considerably higher than CCDs.

The PPD pixel is a popular charge transfer CMOS pixel. The pixel relies on "charge coupling," invented by CCDs. The charge-coupling process is noiseless and perfectly transfers charge in silicon without loss (down to the single electron !).

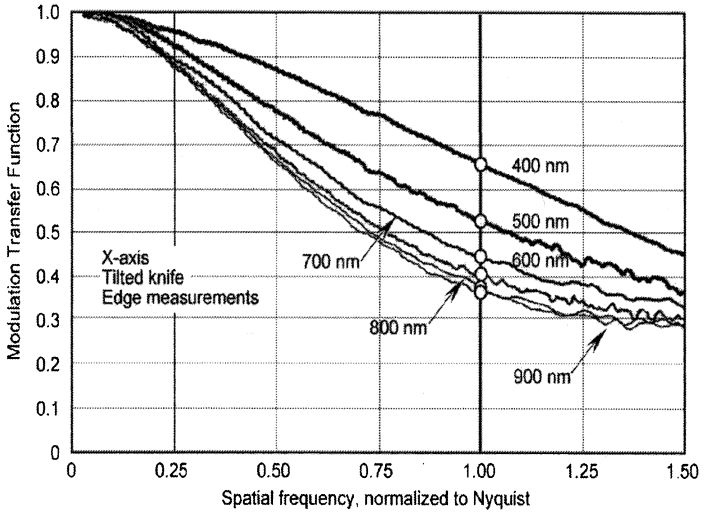

Figure 5: MTF measurements taken from a 12 um CMOS pixel array.[3] MTF degradation for longer wavelengths is caused by charge diffusion between pixels. Note that MTF does not degrade further for wavelengths >700 nm. This is because photons penetrate beyond the epitaxial silicon interface where photo electrons recombine.

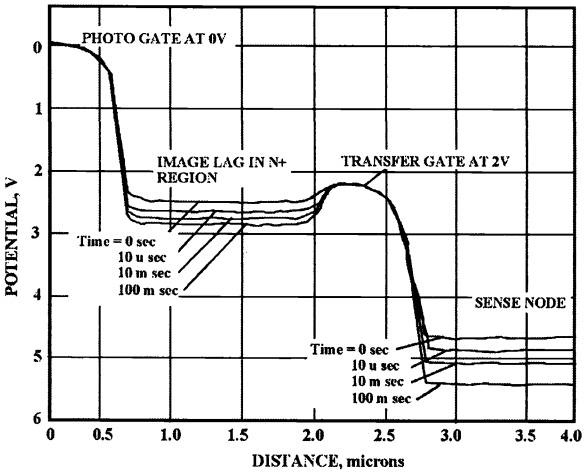

Figure 6a: PISCES modeling illustration showing the n+ region responsible for image lag in the PG pixel illustrated in Figure 8. Image lag is a common, but serious problem for CMOS sensors.

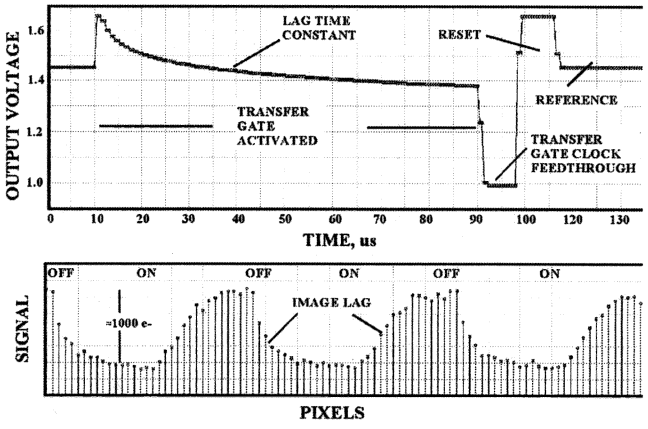

Figure 6b: Image time constant during charge transfer for a PG pixel. The time constant results from charge leaving the n+ region while the transfer gate is in the high state. The 1 000 e- video line trace shows image lag characteristics in response to a light source that is switched on and off. The exponential response seen is due to the n+ time constant.

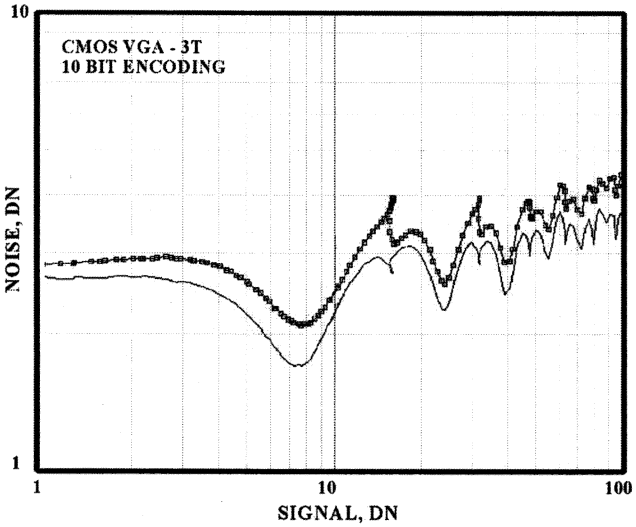

Figure 7a: Photon transfer generated by a VGA CMOS sensor that exhibits an ADC bit-weighting problem. The read noise changes as a function of signal level because of ADC quantizing noise. ADC noise is seen as streaks across the image. In general, the problem is limiting digital encoding for CMOS arrays to no more that 12 bits.

The transfer gate, when activated high, transfers signal charge from the pinned region to the sense node without image lag. The global reset gate shown is used for an exposure control and other important purposes discussed below. For certain readout modes, the PPD can employ true CDS signal processing to eliminate reset noise. Compared to the PD pixel, low charge capacity is a disadvantage limiting use to large pixel applications. This is because the PPD potential is set approximately mid way between the potential of the transfer- gate when low and the potential of the sense node after reset. The PPD requires two custom implants, the same way PPD CCD pixels are processed. The implants' dose and alignment are very critical in delivering good charge transfer between the PPD and sense node regions without image lag.

The PG pixel shown in Figure 8 is fabricated using standard CMOS fabrication processes, which is an advantage over PPD technology. Charge transfer to the sense node takes place by clocking the photo and transfer gates, similar to CCD operation. However, the pixel exhibits low QE because of PG absorption loss. In addition, image lag represents a serious problem for the pixel (refer to Figure 6). For these reasons, the PG pixel is not popular in the commercial community.

The CS pixel shown in Figure 8 is similar to the PPD pixel except that photo region is a simple photo diode without pinning or full depletion. Like the PG, this arrangement makes the pixel vulnerable to image lag because of undepleted n^+ material. The problem is eliminated, similar to the PG pixel, by performing a hard reset with the penalty of high reset noise associated with the capacitance of the diode. Also, as its name implies, signal charge is not totally transferred but is shared between the PD and sense node regions. This trait typically results in a noise floor greater than 100 e-, and therefore, the CS pixel is not often applied.

Three popular readout schemes are used to read CMOS pixels. These readout architectures are referred to as "rolling shutter", "snap" and "progressive scan." The readout modes have a significant influence on performance especially read noise. The rolling shutter readout mode is employed by PD pixels. Charge integration time, or the time period where the PD is allowed to discharge to photon input, is governed by leaving the reset switch off a specified number of line periods before a row is selected for readout. Rolling shutter exhibits image smear and motion artifacts because charge integration takes place over the entire readout period (1/30 sec for standard video rates). Rolling shutter inherently exhibits high read noise because of its incompatibility with CDS signal processing. With rolling shutter, the reference and video levels are not referenced to the same reset pulse (i.e., a reset occurs between the two samples). The resultant read noise is kTC reset noise limited.

Snap readout is used by charge transfer CMOS pixels (i.e., PPD, PG and CS). The readout scheme is initiated by resetting all pixels using a global reset gate or by resetting with the sense node, transfer and PGs together. Next, charge is integrated within the photo region for a given exposure period. During this time, the sense node is reset to eliminate dark current or signal charge build up there. Then, signal charge transfers to the sense node for all pixels at one time. After transfer, readout commences line-by-line using CDS-like processing (i.e., sample 1, reset, sample 2). Therefore, the resultant noise for snap readout is also kTC reset noise limited. Charge generated and collected by the sense node during readout causes image smear, especially for moving scenes (referred to as "sense node smear"). Metal light shields above the pixel prevent photons from interacting with the sense node region. However, photoelectrons generated below the pixel's depletion region can diffuse to the sense node. The smear problem is most conspicuous for very short near-IR exposures. Charge diffusion is minimized by a shielding implant beneath the sense node and transfer gate, as shown in Figure 8. The implant gradient produces a self-induced electric field that directs signal charge back into the photo region. The sense node and read MOSFETs are placed in a p-well that naturally reflects signal carriers. The sense node must also be protected from PPD charge that leaks over the transfer gate during readout. "Sense node blooming" as it is called, occurs when the exposure time is less than the readout time, as the sensor is still illuminated during readout. Sense node blooming is controlled by biasing the global reset gate shown in Figure 8 at a potential lower than the transfer gate potential, allowing excess charge to escape to a drain region.

Progressive scan is a readout scheme also used by charge transfer pixels. A mechanical shutter or pulsed light source is usually employed for this readout mode. After exposure, charge is read line-by-line using true CDS readout (i.e., reset, sample-1, charge transfer, sample-2). The read mode allows the read noise to be limited by the source follower MOSFET. Low noise is achieved, at levels equivalent to the CCD running slow scan (i.e., a few noise electrons as demonstrated below). Progressive scan also offers low dark current and sense node smear because the sense node is continuously reset until a line of pixels is read.

3. SCIENTIFIC CMOS PIXELS

The table below presents general performance parameters for hypothetical 8 μm PD, PPD, PG, CS CMOS pixels and a scientific backside illuminated CCD. Note that CMOS pixels fall considerably short in delivering performance compared to the CCD primarily in the areas of read noise, QE, MTF and dark current.

PERFORMANCE PARAMETER	PD PIXEL	PPD PIXEL	PG PIXEL	CS PIXEL	CCD
Dark Current, pA/cm² (300K)	200-2000	50-500	200-2000	200-2000	3 -50 MPP
Dark Current FPN, % rms	10-30 %	10-30 %	10-30 %	10-30 %	10-30 %
Sensitivity, V/e-	15	25	25	15	5
Read Noise, e- rms unless noted Progressive Scan	60 Rolling Shutter	5	70	150	2
Charge Capacity, e- (Hard Reset)	100,000	50,000	75,000	100,000	80,000
Dynamic Range	1667	10000	1070	666	40,000
Fill Factor, %	75	65	65	65	100
Interacting QE, % (400 nm)	35	35	4	35	90
Pixel FPN, % rms	1	1	1	1	2
MTF, % (900 nm) (Nyquist)	10-25	10 -25	10-25	10-25	50
Nonlinearity, % (Full Well)	7	7	7	7	1

V_{DD} = 3.3 V for CMOS

3.1 Backside Illuminated - Deep Depletion - PD Pixel

The advanced CMOS PD pixel shown in Figure 9 is custom designed for backside illumination and full depletion. A negatively biased substrate forces the PD depletion region to the backside of the pixel. To accommodate this biasing feature, the pixel uses a deep n-well process. The deep n-well is connected to a shallow n-well which together act as the PD. The n-well completely encircles the p-well to isolate the region from the substrate. Without this isolation a large leakage current would flow between the p-well and biased substrate. The grounded p-well is used by the three readout MOSFETs shown. A frontside p^+ ohmic contact provides the necessary substrate bias. The poly ring shown controls surface dark current generated around the diode region.[7]

The PISCES modeling illustration in Figure 10 shows equal potential contours and the resultant depletion depth created by the pixel shown in Figure 9. A depletion depth of 6 μm is realized assuming 10 ohm-cm epi-silicon, PD bias of V_{PD}= 3 V and substrate bias of V_{SUB} = - 10 V. Thinning the chip to approximately 7 to 8 μm is necessary to achieve full depletion. As a side note, it is important that the depletion edge does not completely extend to the back surface or else high dark current is generated.[1] Proper depletion depth is set by varying the substrate bias. Higher resistivity and thicker

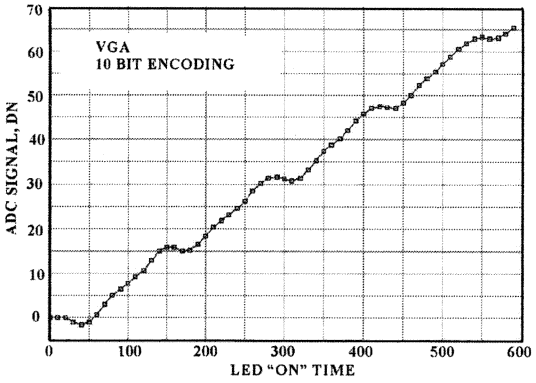

Figure 7b: An ADC linearity transfer curve for the CMOS VGA sensor characterized in Figure 7a. The response shows a bit-weighting problem that occurs every 16 DN. The problem results in missing codes, which produce quantizing noise.

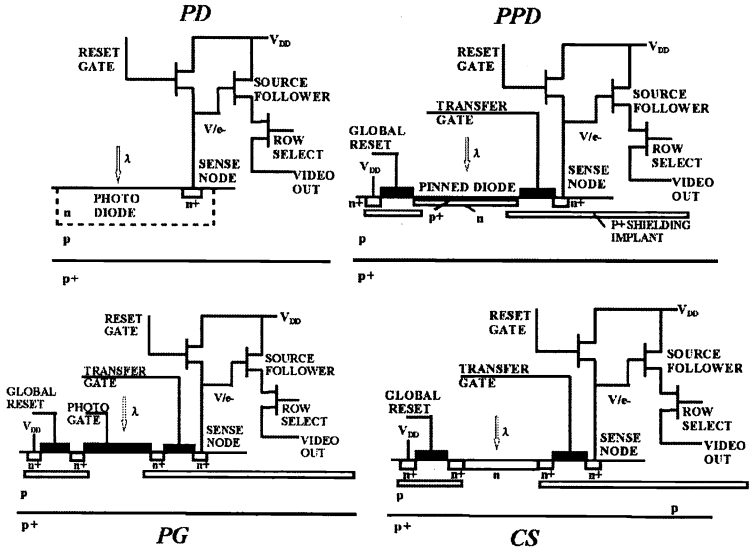

Figure 8: Conventional photo diode (PD), pinned photo diode (PPD), photo gate (PG) and charge shared (CS) CMOS pixels. The PD pixel architecture is the most popular pixel used in CMOS sensors. The PG and CS pixels exhibit serious performance deficiencies as explained in text. The PPD pixel is the highest performing CMOS pixel used in high-end still camera applications. The pixel requires custom implant processing to form the PPD region.

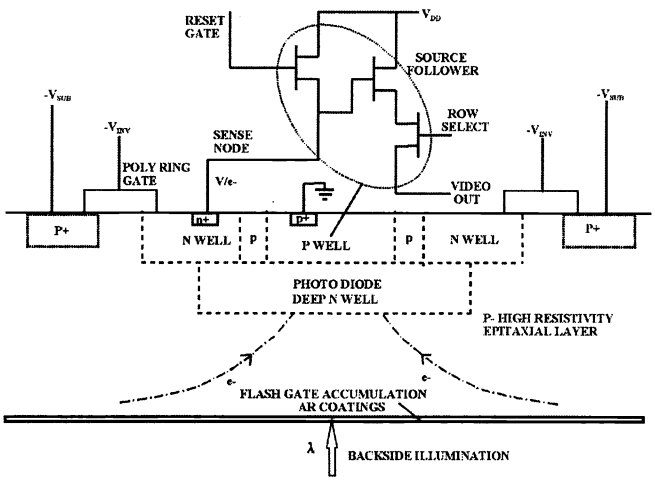

Figure 9: An advanced backside illuminated, fully depleted, PD CMOS pixel. The grounded MOSFET p-well is isolated from the substrate region by a deep n-well. Negative substrate bias extends the PD depletion region to the rear of the device for high MTF performance.

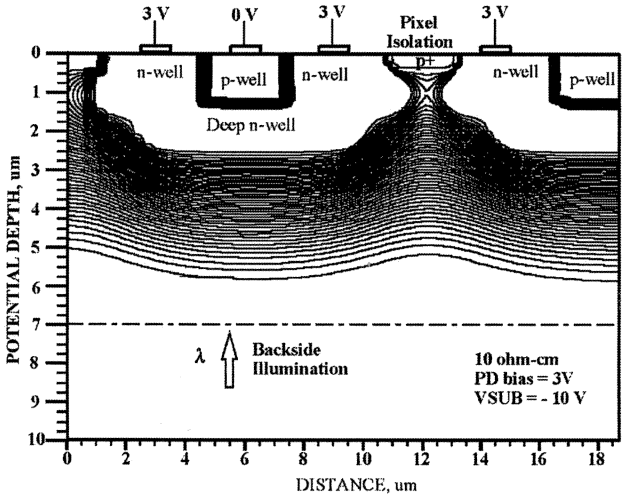

Figure 10: PISCES modeling output for the advanced CMOS pixel illustrated in Figure 9. The plot shows equal potential contours and a 6 um depletion depth achieved for 10 ohm-cm epitaxial silicon and the bias voltage conditions indicated.

silicon extends the depletion depth to meet application requirements. For example, Figure 11 show a depletion depth of 15 μm assuming 100 ohm-cm silicon and V_{SUB}= -10 V. In this case, thinning the chip to 15 μm will provide full depletion. Figure 12 plots the net depletion depth as a function of silicon resistivity for different bias conditions assuming a diode junction depth of 2 μm. Also shown is the photon absorption depth for various wavelengths. Silicon resistivity > 500 ohm-cm is ideal for deep depletion and high near IR QE performance. Figure 12 also plots diode depletion capacitance, which decreases with silicon resistivity and applied bias. This characteristic is advantageous for the pixel because depletion capacitance can be tailored to meet sensitivity (V/e-) and reset noise requirements. MTF and reset noise both decrease with substrate bias.

It should be mentioned that the PD pixel shown in Figure 9 is vulnerable to image lag if not properly reset, as is the case for all CMOS pixels. Image lag occurs when the reset gate is only clocked high to the pixels supply voltage, V_{DD} (referred to as "soft reset"). During the reset time, the potential under the reset gate is less than gate voltage by the MOSFETs' threshold and body effect voltage, Vth (i.e, V_{DD} - V_{th}). In this state, free charge from the sense node can thermally jump over the reset gate barrier to the V_{DD} drain region (the same effect described in Figure 6). As charge escapes, the sense node voltage increases which results in a signal variation that is measured as image lag at the output of the source follower. Image lag is eliminated by driving the gate of reset switch with a clock voltage greater than V_{DD} + V_{th} (referred to as "hard reset"). This clocking action forces both the sense node and potential under the reset gate to V_{DD}, thereby eliminating the field that would cause electrons to leave the sense node. This mode of operation is also advantageous because it increases the pixel's well capacity by allowing the sense node to swing over a greater range (typically by a factor of two) and increases diode depletion depth. For these reasons, the advanced pixel shown in Figure 9 assumes hard reset operation. Figure 13 is a reset transfer curve showing the relationship between the reset gate clock voltage and V_{DD} at the onset of image lag. Image lag is measured by using a point source test target and looking for deferred charge. Deferred charge that follows a point image decreases as the reset clock voltage increases. The problem is eliminated when the sense node assumes V_{DD} potential or when the gate voltage equals V_{DD} +V_{th}. Several images are averaged to clearly show the deferred tail, as demonstrated in Figure 13.

3.2 Back Illuminated - Deep Depletion - Charge-Coupled - Photo gate Pixel

Figure 14 shows a custom designed charge transfer PG pixel that is backside illuminated and substrate biased. During charge integration, substrate bias is applied for deep depletion under the PG. V_{DD} is disengaged at this time to prevent excess current through the MOSFETs. After signal charge collects, the substrate bias is removed, V_{DD} is activated and readout takes place. Figure 15 plots depletion depth during charge integration as a function of silicon resistivity with various PG and substrate bias settings. In comparison to Figure 12, depletion depth for the PG pixel is not as deep as a PD pixel assuming the same bias conditions to the substrate, PD and PG. This is because the n-region of the PD also adds to net depletion depth whereas PG depletion starts at the immediate surface of the pixel. Also plotted in Figure 15 is the PG surface potential at the Si-SiO_2 interface, showing that the voltage across the gate oxide is essentially independent of substrate bias. This is an important characteristic for breakdown reasons because the substrate voltage is the highest voltage applied to the chip.

Figure 14 shows that the n^+ region between the photo and transfer gates is eliminated because of the image lag problem discussed in Figure 6. This modification allows charge coupling and complete charge transfer. Several techniques are available using standard CMOS processes to perform charge coupling. For example, the PG pixel shown in Figure 16 was fabricated with double poly overlapping gates similar to how multi-phased CCDs are fabricated.[1] Figure 16 also presents a square wave light response taken by the double poly pixel. The line trace shows very little image lag in comparison to Figure 6b. For single poly gate processing, charge coupling can also be incorporated by leaving a small gap between the PG and transfer gate for isolation reasons. PISCES modeling indicates that potential barriers, common to gapped regions that interfere with charge transfer, can be avoided if high resistivity silicon is employed (high resistivity silicon and high voltage clocking increases fringing field strength between the gates that override barrier problems). It is also possible to use a metal gate within the gap, that is connected to the PG or transfer gate, to control the problem. Both these pixel types are currently being fabricated for test.

When charge coupling is incorporated, true CDS can be performed to eliminate reset noise. Figure 17 shows a photon transfer curve generated by a PG test pixel that exhibits a noise level of a few electrons. Without CDS, the reset noise floor exhibited by the pixel is 70 e-. The pixel's source follower amplifier limits the noise floor seen in Figure 17 by the amount of flicker noise present. Figure 18 shows spectral noise characteristics taken from a 0.44 (W) x 0.36 (L) μm

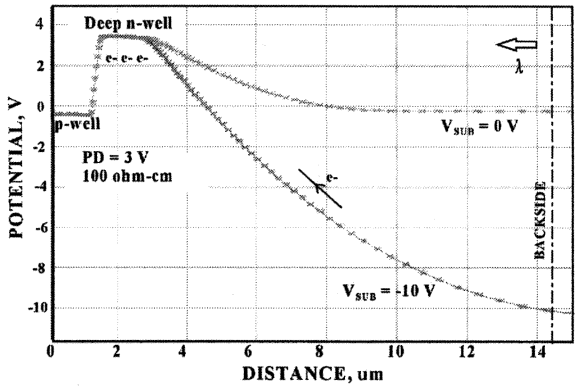

Figure 11: PISCES output showing depletion depth obtained for the pixel shown in Figure 9 at two substrate biased conditions. The 100 ohm-cm silicon modeled generates a depletion depth of greater than 15 um for V_{SUB} = -10 V bias.

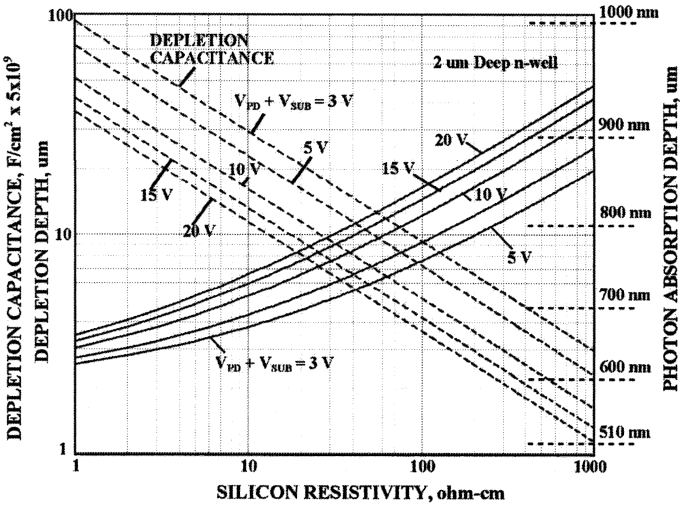

Figure12: Depletion depth as a function of silicon resistivity for different PD and substrate bias conditions. The plot shows that 100 ohm-cm silicon matches 15 um thinning for full depletion using a 2 um deep n-well and V_{SUB} = -10 V bias. Depletion capacitance is also plotted to show that pixel sensitivity (V/e-) and reset noise are both functions of substrate voltage.

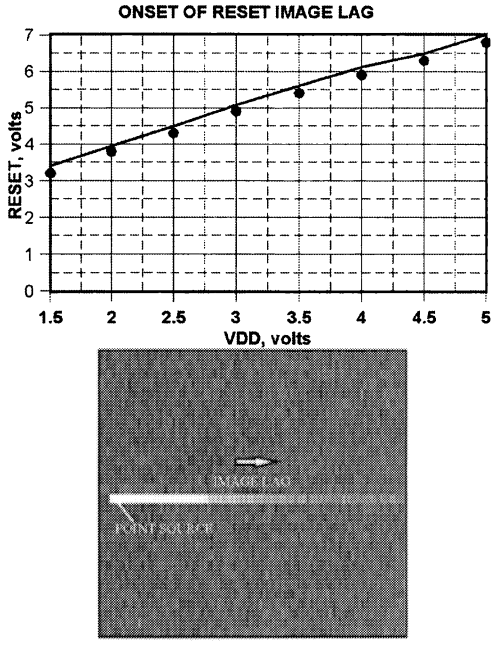

Figure 13: Reset transfer curve showing the relationship between the reset gate clock level and V_{DD} at the onset of image lag. The image below shows image lag generated by a point source image. Image lag is eliminated when the reset gate is clocked above $V_{DD} + V_{th}$ as explained in text.

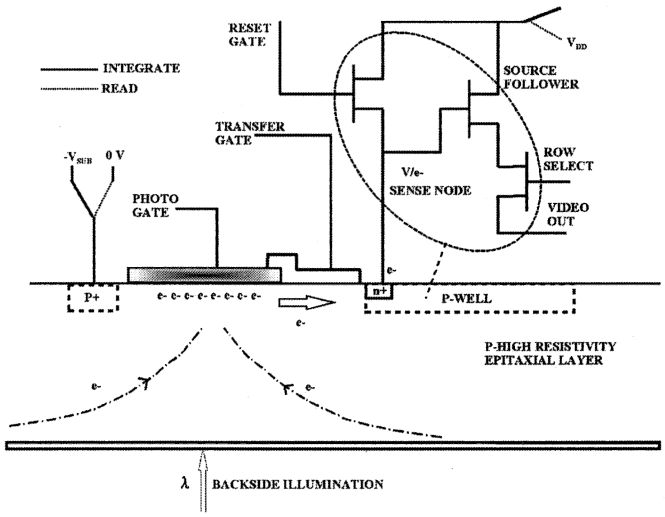

Figure 14: An advanced backside illuminated, fully depleted, charge coupled, PG pixel. The pixel uses double-over-lapping poly gates to perform charge coupling because of image lag. Substrate bias is applied during charge integration for deep depletion operation.

CMOS source follower amplifier. Note that the noise power drops an order of magnitude for each decade of frequency measured, characteristic of flicker noise. The white noise level for the MOSFET and bias conditions is approximately 15nV/root-Hz. The corresponding 1/f corner frequency is approximately f_c =10 MHz (i.e., where 1/f noise and white noise power are equal). The noise bumps seen in the plot exhibit random telegraph signal (RTS) behavior, characteristic of ultra small MOSFETs. Figure 19 shows an RTS noise signature taken from a 0.32(W) x 0.55(L) MOSFET under high bias current conditions (I_D =10 μA). The high frequency noise seen riding on the RTS is primarily associated with source follower white noise. Figure 20 plots read noise as a function of sample time for this MOSFET using CDS processing. A 4 e- noise floor is achieved for a 30 μs sample-to-sample time for a gate voltage of 1.2 V (2 uA bias).

CMOS read noise varies significantly from source follower to source follower, and therefore, from pixel to pixel. The 1/f variance is related to trap density and location within the current carrying channel of the MOSFET. CMOS noise sensitivity is greater than CCD source follower amplifiers, which are physically larger. For example, slow scan scientific CCDs typically show < 20 % variance of the average noise floor. These fundamental noise characteristics exhibited by CMOS amplifiers could represent a serious drawback in achieving reliable low-level noise performance.

Read noise generated by the source follower MOSFET amplifier is estimated by,[1]

$$R = \frac{1}{S_V A_{SF} (1 - e^{-t_s/\tau_D})} \left[\int_0^\infty W^2 \left(1 + \frac{f_c}{f} \right) + \frac{1}{1 + (2 \pi f \tau_D \pi)^2} (2 + 2 \cos(2 \pi f t_s)) \, df \right]^{1/2} \quad (1)$$

where R is the read noise (rms e-). The first term under the integral represents the noise frequencies generated by the source follower amplifier. The second term is the low-pass filter transfer function that filters white noise (single pole filter). The third term is the CDS transfer function. Here S_V is the sense node sensitivity (V/e-), A_{SF} is the source follower voltage gain (V/V), W is the white noise spectral voltage ($V/Hz^{1/2}$), f_c is the 1/f noise corner frequency (Hz), τ_D is the dominant time constant of the signal processor (sec) and t_s is CDS sample-to-sample time (sec). Figure 20 plots theoretical read noise for the MOSFET parameters indicated and using Eq. (1). Read noise is lowest at approximately $t_s = 2\tau_D$. When $t_s > 2 \tau_D$ the read noise gradually increases as more 1/f noise enters the pass band. When $t_s < 2\tau_D$ the noise also increases because signal strength diminishes (i.e., the first term of Eq. (1) in front of the integral).

Inverting the surface under the PG pixel can significantly reduce dark current. For example, Figure 21 shows a dark current reduction of 1 nA/cm^2 to 80 pA/cm^2 by inverting the PG to minus one volt. However, inverting the PG collapses the depletion region, which provides no charge capacity (i.e., the surface voltage assumes substrate potential). The measurement in Figure 21 was performed by integrating dark charge directly on the sense node (i.e., the transfer gate is set high). To circumvent the full well problem, a custom phosphorus implant is required under the PG, similar to how multi pinned phase (MPP) CCDs are fabricated.[1] This arrangement allows inversion with some charge capacity. However, the inverted state cannot be maintained if the PG voltage is ever greater than the substrate voltage. Therefore, the substrate-biased pixels as shown above cannot be inverted.

4. HYBRID ARRAYS

4.1 CCD to CMOS Hybrid

The CCD has shown textbook performance with near perfect photon interaction and collection of signal carriers. CMOS technology has not demonstrated such high performance levels. On the other hand, CMOS arrays allow fast signal acquisition through parallel signal processing with low noise, low power and electronic-circuit compactness. Imaging groups are taking both technologies and integrating CCD pixel and CMOS readout integrated circuit (ROIC) chips into a hybrid array. The CCD and CMOS chips are joined together with standard indium bump bonds. Backside or frontside illumination is possible.

CCD to CMOS hybrid architectures vary significantly. For example, a single CCD output port connected to a single CMOS ROIC is employed for slow scan low noise scientific applications.[8] Also, a ROIC can also be mated to the output of each CCD vertical register at the top and bottom of the chip. For example, Figure 22 shows the output region of a CCD array with this design in mind. Indium bumps are located at the output of each CCD source follower amplifier. For low power dissipation, source followers are removed from the design with bumps applied to the sense nodes for direct

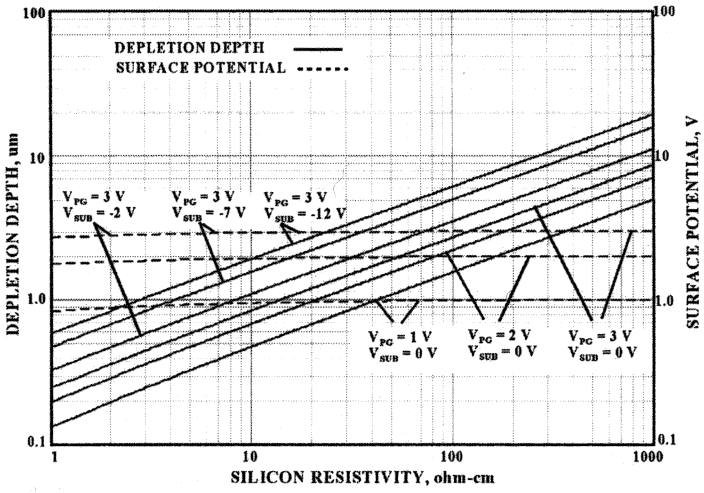

Figure 15: Depletion depth as a function of silicon resistivity for various PG and substrate bias conditions. The plot shows that >100 ohm-cm silicon is necessary to match 10 um thinning for full depletion operation.

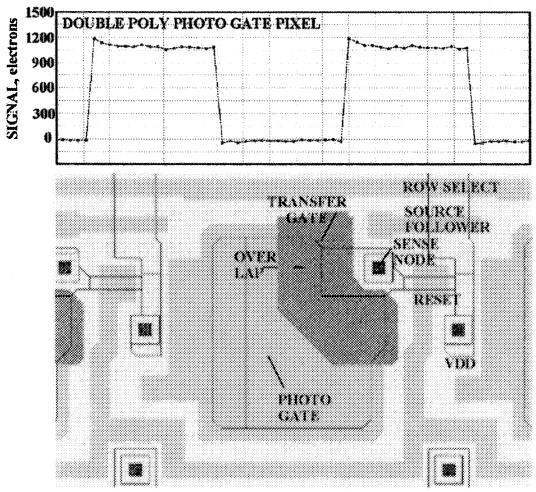

Figure 16: A double poly gate, charge coupled, CMOS PG pixel. The corresponding video output for the pixel is in reaction to a changing light source. The response shows very little image lag in comparison to Figure 6 b.

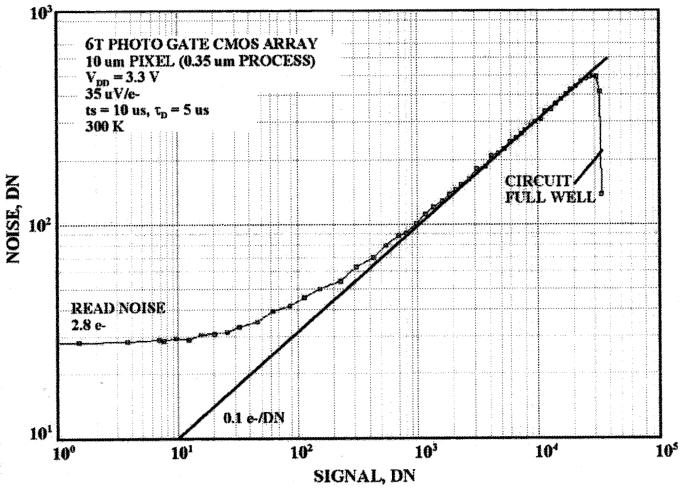

Figure 17: Photon transfer generated by a PG pixel read by true CDS processing. The read noise floor of 2.8 e- is limited by source follower 1/f noise, which varies significantly with bias current.

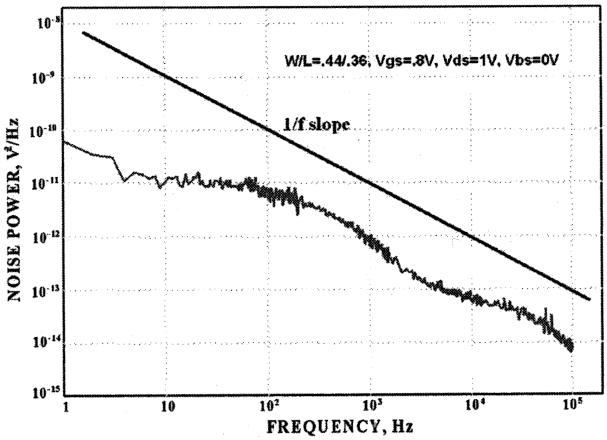

Figure 18: Spectral noise characteristics for a small source follower CMOS MOSFET. The noise bumps in the plot are characteristic of random telegraph noise generated by electron traps in the channel of the MOSFET.

charge readout. In this case, a CMOS ROIC array provides charge-to-voltage amplifiers to process the signal charge packets. This parallel output design offers high-speed / low noise performance compared to CCD monolithics. For example, a 1024 x 1024 pixel hybrid can be read at 20 frames/sec assuming a horizontal row scan time of 50 us. A single ADC is employed to encode a net pixel rate of 20 Mpixels/sec. Faster frame rates are possible using multiple ADCs to shorten the horizontal scan time. For example, four ADCs working in parallel would increase the frame rate to 80 frames/sec. Low noise is achieved for the hybrid because the CDS bandwidth is considerably smaller compared to a CCD that must work with fewer channels. This CCD limitation is seen by calculating the read noise as a function of frame rate using Eq. 1 and assuming,

$$t_s = \frac{N_{CDS}}{N_L\ N_P\ f_R} - t_{OH} \tag{2}$$

where N_{CDS} is the number of CDS processing channels, N_L is the number of lines/frame, N_P is the number of pixels/line, f_R is the frame rate (frames/sec) and t_{OH} is the overhead time required to process and digitize pixels (sec). Note that a CCD to CMOS hybrid with CDS processors employed in each column reduces Eq. 2 to,

$$t_s = \frac{1}{N_L f_R} - t_{OH} \tag{3}$$

Figure 23 plots read noise described by Eqs. 1 and 2 as a function of frames/sec for different CCD array sizes and the number of readout ports utilized. The CCD analyzed assumes $W(f) = 10nV/Hz^{1/2}$, $f_c = 10^5$ Hz, $S_V = 4 \times 10^{-6}$ V/e-, $\tau_D = 0.5t_s$ and $t_{OH} = 30$ ns. In comparison, the noise generated by a CCD to CMOS hybrid is independent of frame rate because the sample time t_s can be fixed by employing more ADCs as required. For a fixed sample-to-sample time of 1 us, a few noise electrons can be achieved assuming that system noise does not increase with frame rate.

There are some drawbacks to the CCD to CMOS hybrid. For example, horizontal and vertical CCD clocks are not compatible with low-voltage CMOS fabrication processes. Therefore, custom high-voltage electronic circuitry is required between the two chips. Also, CCD to CMOS technology is susceptible to high-energy radiation damage, a problem also common to CCD monolithics. For very large pixels, indium bumps can be applied to every pixel to avoid transferring charge from pixel to pixel. Each pixel incorporates a photo gate and transfer gate to transfer signal charge to a bump bonded sense node. This architecture produces ultra high frame rates. For example, a net pixel rate of 1 Gpixel/sec is achieved for a 100 x 100 array with each pixel being read at 100,000 pixels/sec.

4.2 CMOS to CMOS Hybrid

Figure 24 shows CMOS to CMOS hybrid arrangement where a backside illuminated CMOS pixel array is bump bonded to a CMOS ROIC. Similar to the CCD to CMOS hybrid, the two chips are fabricated independently allowing the CMOS pixel array to be custom fabricated using processes not compatible with the CMOS ROIC (e.g., silicon resistivity). For example, the backside illuminated - deep depletion pixels shown in Figures 9 and 14 are well suited for the hybrid. The pixel array is thinned and passively accumulated (e.g., flash gate[1]) after full hybridization and functionally tested. Similar to the CCD to CMOS hybrid, the ROIC can be bump bonded to the CMOS pixel array on the sides of the chip. However, because the CMOS pixel array is back illuminated and the pixels are actively isolated with source follower amplifiers, it is also possible to bump bond from *inside* the array. With this feature, very large hybrids can be fabricated by configuring N x N sub arrays together into a monolithic mosaic. The sub arrays are read in parallel fashion. For example, Figure 25 illustrates how 512 x 512 sub arrays are formatted into a large hybrid. Note that within each 512 subarray, the source follower amplifiers are bussed together along each column and end with a bump bond. The segmented vertically integrated design can deliver ultra high-resolution at high frame rates. Figure 26 plots pixel rate as a function of array size assuming 512 x 512 subarrays as a function of frame rate. For example, an 8192 x 8192 hybrid, based on 256 - 512 x 512 subarrays, can deliver 2 Gpixels/sec assuming 30 frames/sec operation. Such an array will need to be "stitched" because CMOS lithography is limited to fabricating arrays of 20 to 40 mm depending on the manufacturer's alignment capabilities. For a 10-μm pixel, this limits arrays to 2k to 4k pixel formats, less support circuitry. As is true for CCDs, many high-end applications demand larger formats. Some CMOS foundries offer stitching capability, which means

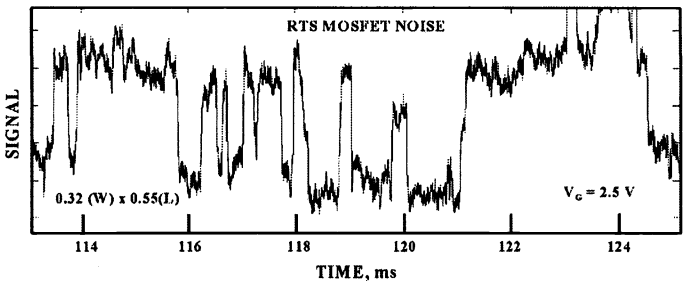

Figure19: Random Telegraph Signal (RTS) generated by a 0.32 (W) x 0.55 (L) MOSFET source follower amplifier.

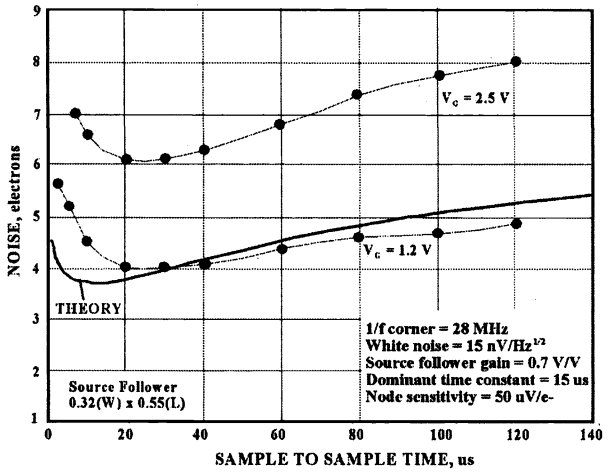

Figure 20: CDS read noise taken from CMOS source follower amplifier as a function of CDS sample time. The theoretical curve is derived from Eq. 1 using the parameters indicated.

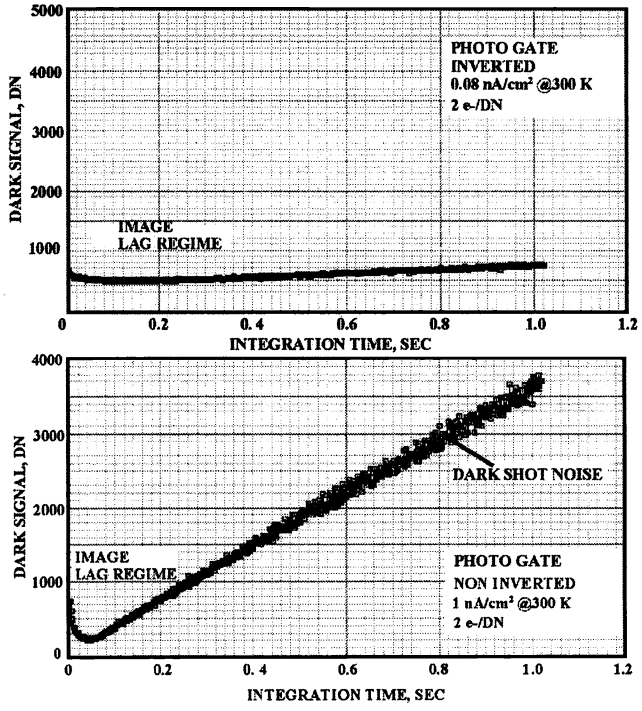

Figure 21: Dark current generated by a PG sensor under inverted and noninverted conditions. A reduction of 12.5 times is seen. Image lag is seen for low dark current levels caused by the problem discussed in Figure 6a.

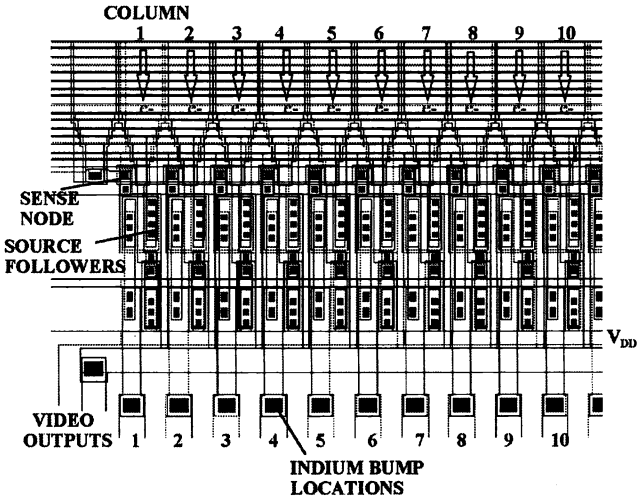

Figure 22: A CCD output region showing bump locations at the output of source followers contained in each vertical register. Bumps can interface to a CMOS ROIC chip forming a CCD to CMOS hybrid array. Bumps can also be directly applied to the sense nodes without source follower amplifiers for low power consumption.

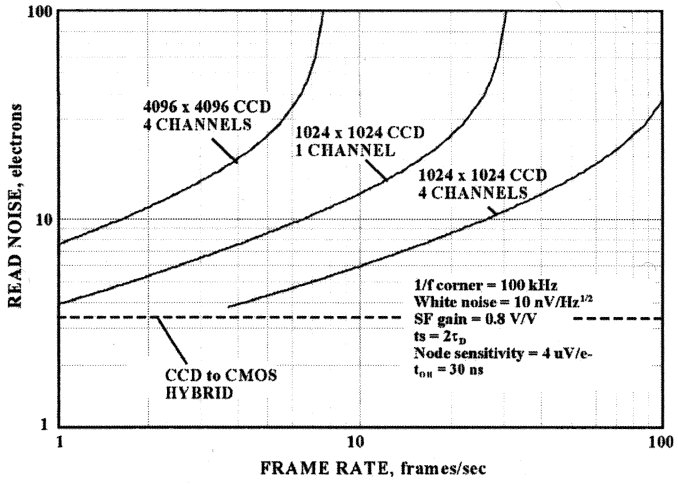

Figure 23: CCD read noise as a function of frame rate for different array sizes and output channels utilized. The increase in noise with frame rate is caused by greater bandwidth requirements and increased source follower white noise. In theory, the read noise for a CCD to CMOS hybrid sensor can be designed independent of frame rate by providing addition parallel channels.

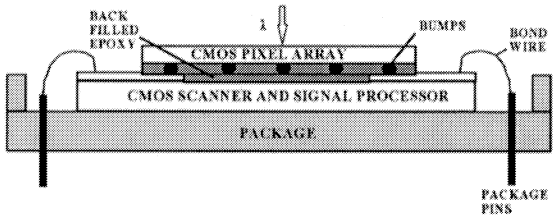

Figure 24: CMOS to CMOS hybrid sensor cross-section. The CMOS pixel array is backside illuminated. The hybrid is thinned and packaged after hybridization.

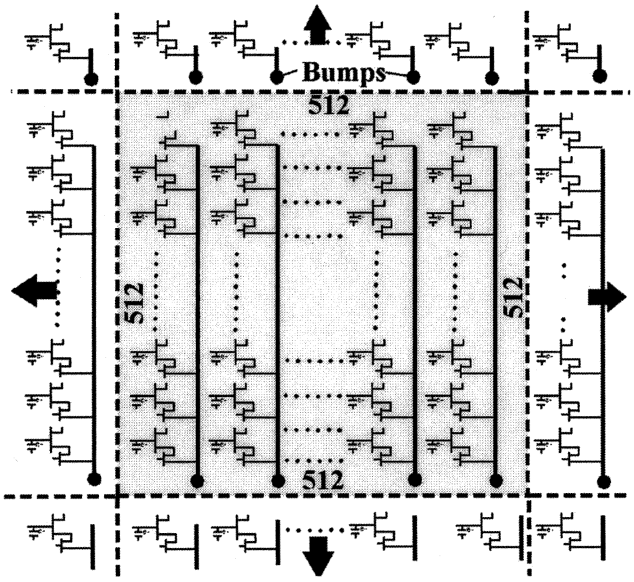

Figure 25: N x N - 512 x 512 pixel sub arrays mosaiced into a hybrid sensor. Each sub array is read out in parallel. Also, within each sub array each column is read out in parallel. All connections to the subarrays are made through indium bump bonds connected to a CMOS ROIC chip as shown in Figure 24.

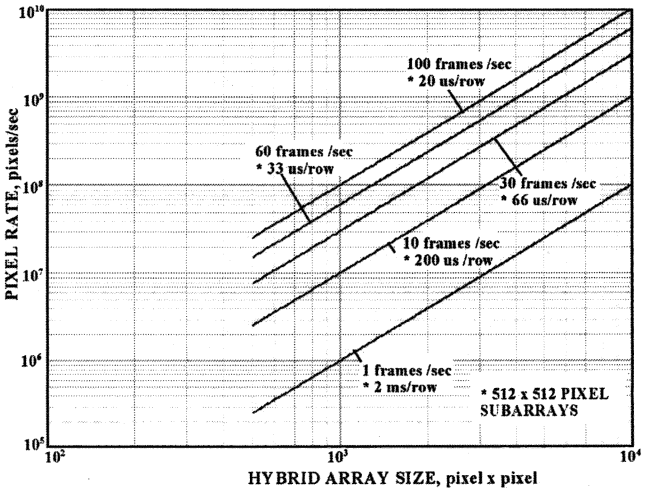

Figure 26: Pixel rate as a function of hybrid size based on 512 x 512 pixel sub arrays. The row time indicated is the time given to scan 512 columns of the sub array by the ROIC chip.

multiple blocks of subarrays can be fabricated to form very large arrays, theoretically limited in size by the dimensions of the wafer (20 cm is common).

The pixel and ROIC chips are built on different substrates that provide good electrical isolation. The hybrid solves many grounding concerns experienced by commercial bulk CMOS arrays (e.g., logic ground bounce noise). The hybrid also solves electronic horizontal and vertical scanning problems associated with large monolythic CMOS arrays. For example, Figure 27 shows dark responses taken from a 640(H) x 480(V) VGA CMOS array that demonstrates a slow scan problem.[4] The top image shows a uniform dark response when the device is clocked at 625 kpixels/sec (1 ms/line) at room temperature. The middle image shows the response when the pixel rate is reduced to 47 kpixels/sec (13.6 ms/line). The nonuniformity pattern seen at this rate is caused by a leakage problem associated with the CDS sample-and-hold capacitors (i.e., the 1 pF capacitors employed have difficulty "holding" the pixel signal level). The lower line trace shows that the leakage problem exponentially increases with increasing operating temperature. Note, as the sampled pixels are scanned from left to right, the amount of leakage increases until the CDS circuitry saturates. By maintaining small subarrays that make up a CMOS to CMOS hybrid, leakage current effects can be minimized to allow slow scan readout and long exposure periods. In addition, long metal bus line problems associated with V_{DD} and clocked lines is better controlled by sub array hybrid architecture. Minimum sub array dimensions are limited by ROIC real estate requirements that drive and read the individual sub arrays.

Unlike CCD type detectors, CMOS pixel arrays read pixels directly without transferring charge from pixel to pixel. This feature is advantageous for imagers that work in high-energy radiation environments. CCDs are very sensitive to the problem because high-energy photons and particles induce charge traps, which degrade CTE performance.[1] Nevertheless, CMOS arrays, including the CMOS to CMOS hybrid are sensitive to other radiation damage effects. For example, hot pixels (dark spikes) induced by high-energy particles are common to both CCDs and CMOS detectors. Figure 28 shows the effect of 63 MeV protons on photon transfer curves taken from a Sarnoff CMOS VGA sensor. The two total noise curves shown include dark current fix pattern noise (FPN) produced by dark spikes. The corresponding shot noise curves are generated by taking two identical frames for each data point and subtracting them pixel by pixel to remove FPN. The subtraction process "despikes" the image leaving only shot and read noise. Note that read noise climbs to 250 e- after proton irradiation but settles to 64 e- after despiking. Figure 29 shows a dark current histogram before and after proton irradiation for the same device showing the dark spike increase. Cooling the detector is the most direct solution to the dark spike problem.

4.3 Silicon on Insulator Hybrid

Development of CMOS imagers fabricated on SOI (silicon on insulator) wafers is in its infancy but is already achieving high performance compared to conventional bulk technology (10). SOI wafers include two silicon layers that are separated by an oxide insulator. Considered hybrid, the top and bottom silicon layers are referred to as the device and handle layers, respectively (refer to Figure 30). The insulating layer is called the box layer. Each silicon layer can perform different functions with separate ground returns. The device layer incorporates the three read pixel MOSFETs, whereas the handle layer is used for the photo region. Fabricating CMOS imagers on SOI greatly improves performance. For example, CMOS circuitry is isolated from photo regions, allowing high resistivity silicon to be employed. In addition, CMOS circuit ground returns for SOI are isolated, eliminating substrate bounce and transient coupling problems. This allows higher operating speeds and lower noise. SOI also makes it easier to passivate surfaces for low dark-current generation because of its planar structure. This advantage, in turn, makes the device much more resistant to high-energy radiation environments. Lower FPN and lower power consumption have also been reported for SOI.[9]

5. SUMMARY

Compared to the CCD, CMOS performance is currently preventing the technology from scientific and high end use. Custom CMOS pixel designs and fabrication processes are required to improve performance. Discussions in this paper show that high CMOS performance will come from several fronts. For example, low noise will be achieved by pixels that employ true CDS processing. Charge coupling will eliminate image lag for charge transfer pixels. Surface inversion will control dark current generation. High MTF will be obtained from depleting the entire volume of the pixel through substrate bias and high resistivity silicon. High QE will come from backside illumination. Off-chip ADCs may need to be utilized for low quantizing noise and high dynamic range requirements (i.e., 16-bit encoding is often required scientifically). Development work in these areas is actively taking place. For example, the custom CMOS pixels shown

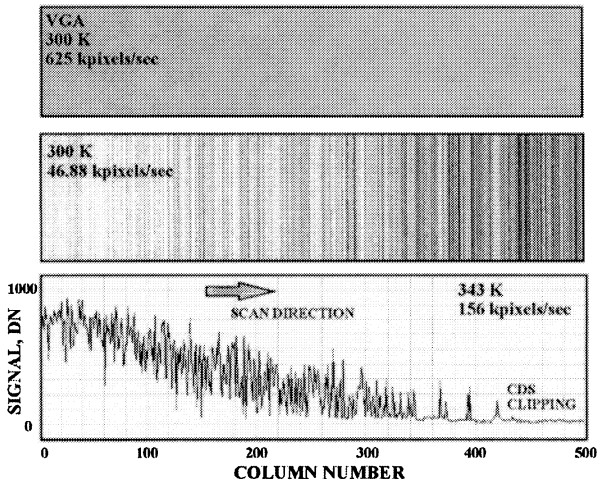

Figure 27: CDS sample-and-hold capacitor leakage problem that becomes noticeable when the horizontal scan rate is reduced and operating temperature is increased. CDS leakage concerns worsen as array size and horizontal scan time increase. Large CMOS to CMOS hybrids limit scan time by sub array size.

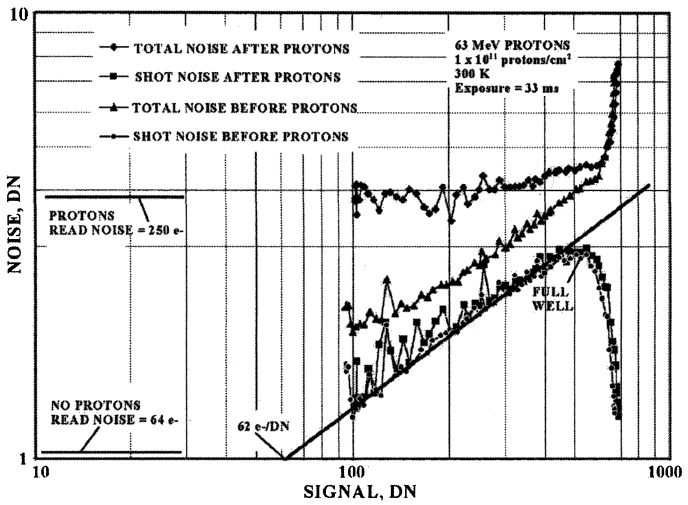

Figure 28: Photon transfer curves taken before and after proton irradiation. Dark spikes are the primary damage effect for CMOS sensors that work in high-energy radiation environments. The plots show that read noise increases to 250 e- as a result of proton induced dark spikes. Dark spike non uniformity noise is eliminated by subtracting a dark frame, pixel by pixel, from the image as performed above. Except for a small amount of dark shot noise that remains after the differencing process, the read noise floor is nearly the same before irradiation (approximately 64 e-).

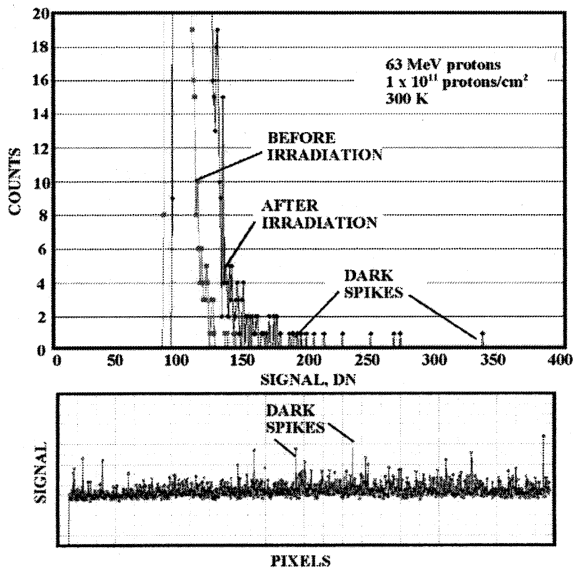

Figure 29: Dark current histogram showing dark spike count before and after proton irradiation. The line trace shows hot pixels induced by protons for a VGA CMOS sensor.

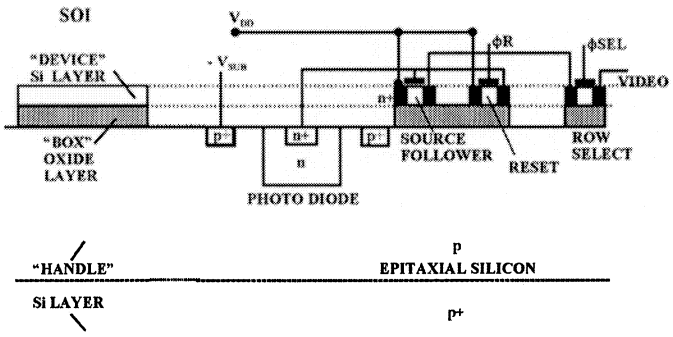

Figure 30: CMOS silicon-on-insulator (SOI) hybrid photo diode pixel. The "device" and "handle" layers are electrically isolated, to allow substrate bias and deep depletion operation.

in Figures 9 and 14 will be characterized this year.[10] However, improved performance also comes with a high price tag relative to commercial arrays that use standard processes. It is difficult to find CMOS foundries that entertain custom development efforts unless large sums of money are exchanged ($300k - $400k appears to be the threshold). Eight-inch reticle and wafer costs are higher for CMOS fabrication compared to CCD. For example, CCD foundries offer 6-inch custom lot processing for approximately $100k assuming a 24-wafer lot size. Nevertheless, it will be interesting to see where Scientific CMOS development takes us in the future.

ACKNOWLEDGEMENTS

Thanks goes to Jim Andrews (Sarnoff) and Taner Dosluoglu (Dialog) for initial discussions behind deep n-well pixel technology. Appreciation goes to my boss John Tower for his continued support to advance scientific CMOS arrays at Sarnoff. Special thanks go to Terry Lomhein and Tracy Dutton for Figures 2 and 5 and their insight behind CMOS MTF characteristics. I also wish to express my gratitude for the opportunity to collaborate with Mark Muzilla, Benjamin Muto and EugeneDines of DRS Technologies on advanced CMOS pixel designs.

REFERENCES

1. J. Janesick, *Scientific Charge-Coupled Devices*, SPIE Press, Bellingham, Washington, 2001.

2. J. Janesick and G. Putnam, Developments and Applications of High Performance CCD and CMOS Imaging Arrays, Annu. Rev. Nucl. Part. Sci. 2003. 53:x:x

3. T. Dutton, J. Kang, T. Lomheim, R. Boucher, R. Shima and M. Nelson, Measurement and Analysis of Pixel Geometric and Diffusion Modulation Transfer Function Components in Photodiode Active Pixel Sensors, 2003 IEEE Workshop on CCDs and Advanced Image Sensors, Elmau, Germany May, 2003.

4. Lux Transfer; CMOS vs CCD, J. Janesick, Proc. SPIE, vol. 4669A, paper 45, San Jose CA., 2002.

5. B. Pain, T. Cunningham, B. Hancock, G. Yang, S. Seshadri, M. Ortiz, Reset Noise Suppression in Two-dimensional CMOS Photodiode Pixels Through Column-based Feedback-reset, Electron Devices Meeting, EDM '02. Digest. International, Dec 8-11, 2002.

6. B. Fowler, M. Godfrey, J. Balicki, and J. Canfield, Low Noise Readout using Active Reset for CMOS APS, SPIE (see http://qss.stanford.edu/~godfrey/imaging/active_reset_spie_00.pdf)

7. H. Cheng and Y. King, Ultra-low Dark Current CMOS Image Sensor Cell Using n / sup+ / Ring Reset, Electron Device Letters, IEEE , Volume: 23 Issue: 9 , Sept. 2002, pgs 538 -540.

8. Refer to: http://www.nasatech.com/Briefs/Sept99/NPO20542.html

9. Zhang, M. Chan, H. Wang, Ko, P.K., Building Hybrid Active Pixels for CMOS Imager on SOI Substrate SOI Conference, 1999. Proceedings. 1999 IEEE International , 4-7 Oct. 1999, pgs 102 -103.

10. To be published: J. Janesick and E. Dines, SPIE, Electronic Imaging 2004, San Jose, Ca.