Chapter 1

Introduction

1.1 A Very Brief History of Visible Detectors in Astronomy: The Pursuit of More Photons

The night sky is a very dim place. When we look up with our naked eye–even on a clear, moonless night–we see only a minute fraction of the objects within our galaxy. Sure, these things are inherently hard to see since they are very far away and the abundant light sources on earth create a bright background in our sky. But our failure to see them is mainly because our eyes are bad detectors when it comes to faint sources. They simply cannot collect and hold on to the photons that are showering down on earth from these objects. Our eyes are meant to sense changes in our environment and help us coordinate movement. They refresh the image we see about every 40 milliseconds so that we can observe a new picture of the world around us.

Telescopes help. A mirror or lens with a large diameter, D is able to collect many more photons than our eyes can. The number of photons is proportional to the collecting area, so the telescope gives us a D^2/d^2 boost in the number of photons if d is the diameter of our pupils. With proper focusing, we can direct the collected photons onto our retinas and begin to bring faint objects out of the darkness. When Galileo pointed a telescope towards the heavens in the early 17th century, he was able to see things in the night sky that no human had ever seen before. Not only was he able to see surface detail on Jupiter because of the magnification the telescope provided; he could actually see moons that were hidden to his naked eye because of their faintness. In that century and the one following it, bigger and bigger telescopes brought fainter and fainter things into view: galaxies, nebulae, clusters of stars, and comets and moons within our solar system. But still, one must have been inclined to ask: *if these things were hidden from our naked eye, what else might be hidden from our telescopes*?

A fraction of an answer to that question came in the early 1850s when photographic plates were first put at the focal plane of a telescope. The emulsion based film used at the time allowed photons to etch their mark over longer periods of time than the short $t_{eye} = 40$ milliseconds our eyes can do. The number of photons collected from a celestial object should be proportional to the collecting time, t, so a long exposure photograph taken over time t_{photo} gave another boost in sensitivity. Photographic film does not collect these photons very efficiently, so the boost was not quite t_{photo}/t_{eye}. Nevertheless, new features and patterns in our sky were discovered because of the gain in sensitivity. Huge clouds of gas and dust were found to linger in regions of sky that were previously believed to be empty and black. Dramatic colors that the insensitive cones of our eyes cannot perceive even with a giant telescope were seen in the shells of planetary nebula, in galaxies, and in the stellar populations of clusters, like the Hercules cluster shown in Figure 1.1.

In the following decades, vast improvements were made to the early photographic techniques in astronomy. Innovations in photographic plates made them more sensitive, less messy (they originally used wet emulsions and solutions), and more transportable (in the very early days, horse-drawn darkrooms needed to be hauled to the observing sites). Improvements in the granularity of the photo-sensitive substances they used also increased their spatial resolution. And the use of negatives made it possible to do objective *photometry*[1] and *astrometry*[2] that did not directly rely on the use of a human's eyes.

As significant as these improvements were, photographic plates and film are rather poor detectors for astronomical purposes. Even the most efficient films capture less than 10% of the incident light, and they all exhibit a highly nonlinear response to flux. They have produced beautiful wide-field images of spiral galaxies and diffuse nebulae and still beat modern imagers in the area of sky the can cover in one exposure. However, as scientific instruments they fall short.

Between the 1930s and 1970s, several other visible imaging devices came into the arena. The photomultiplier tube (PMT) was used to measure brightnesses of extremely faint objects with great precision because of its high gain. It also allowed astronomers to do high speed photometry on objects that vary in brightness on short timescales. Several types of vidicon instruments were installed on telescopes in order to try and produce images of the sky. The silicon photodiode was also used in a few astronomical instruments. While each of these instruments were useful in their own regard, none of them had the same sort of impact that the telescope or photograph did.

Many consider that a revolution in astronomy akin to the inception of the telescope and photograph came in 1969 when the Charge Coupled Device (CCD) was born. The CCD was able to produce two-dimensional maps of the sky in one exposure and make extremely precise photometric and astrometric measurements with very little noise. It did away with the nonlinearity and poor quantum efficiency of photographs, and the ease with which it could be interfaced to a computer made it incredibly more efficient for data analysis. Very long exposures taken with CCDs revealed

[1] *Photometry* is the measurement of the apparent brightness of an object or set of objects. It can be done in a relative fashion or with respect to some standard photometric system.

[2] *Astrometry* is the precise measurement of the positions of objects in the night sky.

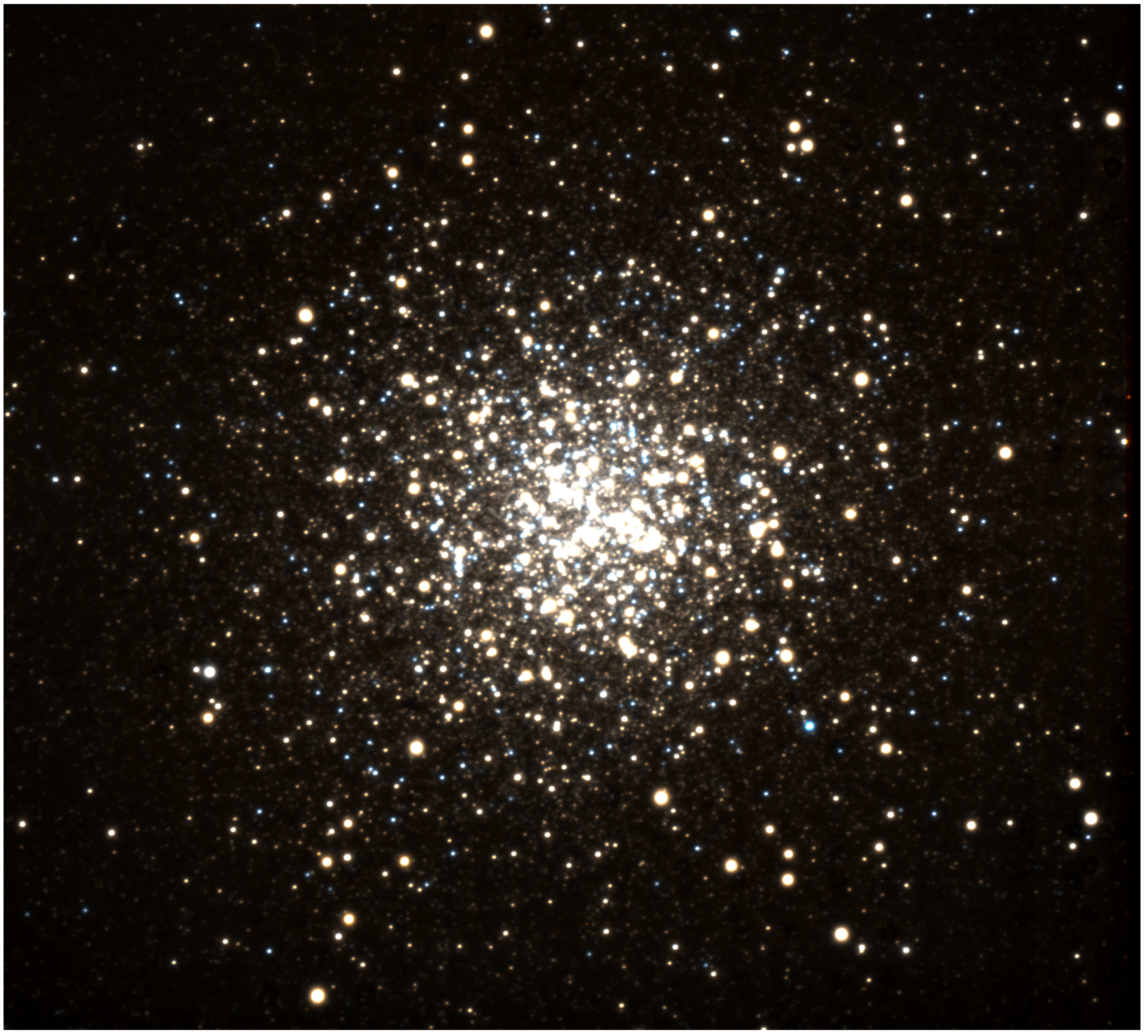

Figure 1.1: A multi-color image of M13, the Hercules Cluster, taken through the Kitt Peak 2.1m telescope. When viewed with the naked eye, M13 appears as a little fuzz in the sky. With only a small telescope, though, a plethora of stars comes into view. When photographic film or a CCD (and color filters) is placed at the telescope focus, the number of the stars visible grows dramatically and their colors are revealed. This particular image was not generated with photographic film or a CCD, but with one of the detectors studied for this thesis work: a HyViSI H4RG.

a whole host of new faint sources that were beyond the sensitivity reach of photographic plates and eluded the tiny field of view inherent to photomultiplier tubes. Since that time, many improvements and modifications have been made to the CCD, and it is still the preferred imager in astronomy. In the following section, we provide a brief overview of the CCD and its place in astronomy.

1.2 The CCD: Astronomy's Champion Workhorse

1.2.1 CCD Operation

Charge Coupled Device (CCD) imaging arrays come in many variations and sizes. However, all of them share one common structure: at heart they are simply an array of picture elements (aka pixels) spread over a two-dimensional grid. One axis of this grid, let us call it x, is the *slow* axis aligned with the rows of the array. The other axis, y, is the *fast* axis aligned with the columns. The terms slow and fast refer to the speed at which electric charge is shifted into a neighboring pixel along the respective axis. The name *Charge Coupled Device* describes how the shifting takes place: it is coupled from one pixel to its neighbor.

The axes of a CCD are not quite symmetric. While charge can be shifted along any row in the x direction, it can only be shifted along one special column in the y direction.[3] This column is often referred to as the *output register*. An electrical barrier called a *channel stop*, made of heavily doped silicon, prevents transfer along all of the other columns. Shifting the charge is the mechanism used to move it toward an output amplifier where it can be sensed as a voltage and turned into a digital number. The diagram in Figure 1.2 shows the essential features of a simple 4×4 CCD as viewed from above.

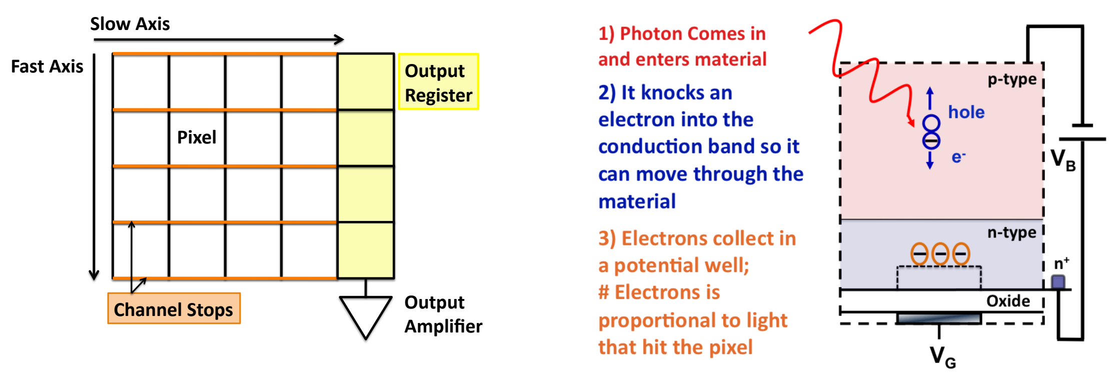

Figure 1.2: (left) A simple cartoon showing the basic features of a 4x4 CCD. The active, light-sensitive pixels are white. The pixels of the output register, shown in yellow, are usually masked from light. (right) A cross-sectional view of a buried n-channel CCD pixel showing how charge is generated and collected. A gate voltage, V_G, is applied to the metal gate below the pixel and a bias voltage V_B is applied across the doped layers. The dashed line is meant to represent the boundaries of the pixel. Not shown are the highly doped areas that form the channel stops. The collection well where the electrons accumulate lies above the dotted line, away from the surface.

[3] This is not true for all CCDs. Recent monolithic CCDs commonly implement multiple outputs, usually 2-4 [1], in which case there will be more than one fast column. Also, in Orthogonal Transfer Arrays, charge can be shifted in both directions [2].

Explaining in exact detail how a CCD creates an image is beyond the scope of this thesis, partly because there are so many different ways in which it is done. CCDs come in many overlapping flavors: *interline* or *frame transfer*, *two phase*, *three phase*, or *four phase*, *surface channel* or *buried channel*, *backside illuminated* or *frontside illuminated*, *p-channel* or *n-channel* to name a few. With each there are differences in where the charge is generated, where it is collected, and how it is shifted. Fortunately, there are many excellent treatises on this subject, including references [3], [4] and [5]. Here we will only consider a backside illuminated buried n-channel CCD–a kind which is frequently used for astronomy–to highlight key features of how a CCD works.

Taking an exposure with a CCD can be divided into two distinct and separate processes: **I) Exposing** and **II) Read-out**.[4] Examining these two processes in more detail will highlight some characteristics of the CCD and the physics upon which it is based.

I) Exposing The bulk light-sensitive material of a CCD is made of a semiconductor, usually silicon. The semiconductor is not fabricated to be pure typically, but rather is doped with impurity atoms in order to enhance charge conduction and collection. In the case of the buried n-channel CCD the bulk material is p-type, meaning it is doped with acceptor atoms, and this bulk rests on top of a channel made of material that is n-type, meaning it is doped with donor atoms. For the buried channel CCD to operate, these layers are placed above an oxide, and in each pixel a metal gate is attached below this oxide, as illustrated in Figure 1.2. This configuration of a **M**etal, **O**xide, and **S**emiconductor stacked on top of each other is referred to as a MOS capacitor. Each pixel contains one of these MOS capacitors. Following the illustration in Figure 1.2, exposing a CCD goes as follows:

1) Photons are allowed to shine on the bulk of the semiconductor. Illuminating from the backside prevents photons from being blocked by any metallic structures used to apply voltages to the pixels. An anti-reflection coating assists in letting photons through the surface.

2) When photons with energy greater than the band gap of the semiconductor shine on the bulk material, electron-hole pairs are created. An electron freed from its host atom, which we refer to as a *photoelectron*, will be promoted to the conduction band so that it is free to roam in the crystal and will leave behind a hole in the valence band, as shown in Figure 1.2. The goal of the CCD is to capture the photoelectron in the pixel closest to the spatial location where it originated and keep it there until it is time for read-out.

3) In order to keep a photoelectron from wandering too far away from where it was produced, an electric field is maintained across the bulk material. This electric field is the result of a voltage V_G, applied to a metal gate below the oxide, and a bias voltage V_B, applied to a highly

[4]There are some exceptions to this. Most notable is *drift-scanning*, a technique frequently employed in astronomy, in which read-out and exposure take place simultaneously as the scene being imaged is allowed to drift across the detector (the direction of image motion and charge shift are the same).

doped n-type implant above the oxide. In addition to creating a field, these two voltages are used to deplete the pn junction of charge carriers and create a potential well in which the photoelectrons can collect. V_G and V_B can be adjusted to change the depth and location of the potential well minimum.

It is important to note that the thicker the bulk material is, the greater V_B must be to steer and collect the charge to the proper pixel. Deep depletion CCDs with thicknesses on the order of 300 μm are currently being developed where V_B is brought as high as 115 volts [1] in order to promote proper charge collection. The advantage of using a thicker CCD is that more of the incoming photons, particularly the ones with longer wavelength, will be turned into photoelectrons and counted as signal.

The CCD itself has no mechanism to halt charge production while light is falling on it. To stop the CCD from exposing, it is necessary to block it with a mechanical shutter or mask of some kind. Otherwise, a smearing effect will occur as the charge packets are shifted along during read-out. In some cases the mechanical element used to block the light can limit the speed with which consecutive exposures are taken.

II) Read-Out Once the exposure is finished and the array has been shuttered, the charge contained in each pixel is ready to be read-out. We will not delve into the exact details of how the charge is moved from its original pixel to the CCD output. For our purposes, it will suffice to say that the voltages on a set of metal electrodes connected to each pixel are changed in a synchronized fashion in order to create movable potential wells that carry the charge along toward the output. This process is referred to as *clocking*.

Each axis of the CCD has its own set of clocks. The *fast* axis has fast clocks that shift the charge along the output register and the *slow* axis has slow clocks that shifts the charge along the rows. In a full clocking sequence, the fast clock goes through however many cycles are necessary to empty the output register and then a slow clock cycle moves a new column into the output register. This process, shown in Figure 1.3, is then repeated until all of the pixels have been emptied of their charge.

Figure 1.3 illustrates several important points about the CCD that may be construed as disadvantageous in certain contexts. For one, reading a CCD is a *serial* process. Each pixel must wait in line to reach the output. Two, the readout is a *destructive* readout in the sense that, in reading the array, the two dimensional map of charge that contains the image must be removed from the pixels. In other words, reading the array means resetting it. And three, pixels cannot be randomly addressed. If we want to know how much charge is in pixel 33, say, we cannot directly tap into that pixel without first tapping into all of the other pixels in the array.

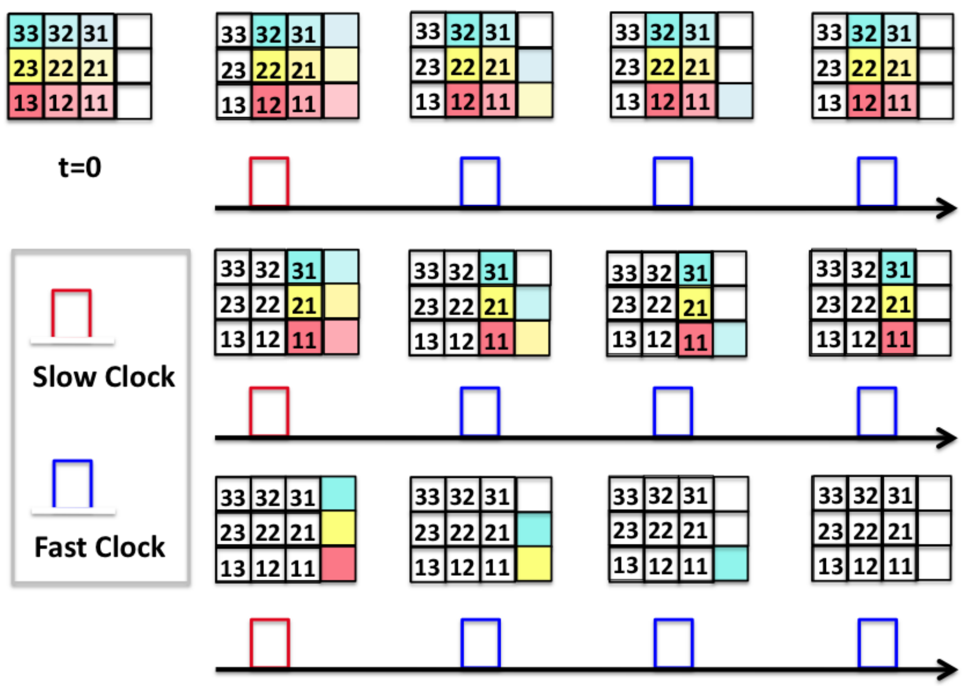

Figure 1.3: A simple cartoon showing the process of clocking in a CCD. The numbers of the pixels designate their physical location in the CCD. The colors have no relationship to wavelength or color of photons; they are solely meant to keep track of charge packets. Only one clock is shown for each fast and slow cycle in this illustration. Time moves from left to right and then down.

1.2.2 Drawbacks of the CCD in Astronomy

The previous section illustrates some of the short-comings of CCDs. There are several other areas where CCD imagers exhibit limitations in some astronomical applications.

1.2.2.1 Destructive Readout

The destructive nature of the CCD read-out implies that only one data point can be collected for each pixel in a given exposure.[5] During very long astronomical exposures (the ones needed to image faint objects can exceed an hour) the observer is essentially blind to what is occurring in the sky temporally.

Figure 1.4 illustrates a hypothetical transient event in the sky such as a supernova or flaring M star that happens to be a part of the field being observed in a long exposure. The light emitted by the object as a function of time is shown on the left. Because the read-out of the CCD is destructive the only data points that are collected are the red dots at the initial and final times of the exposure.

[5]An additional point can be had by taking a bias frame before the given exposure, but the bias frame contains no information relevant to the illumination sources being observed.

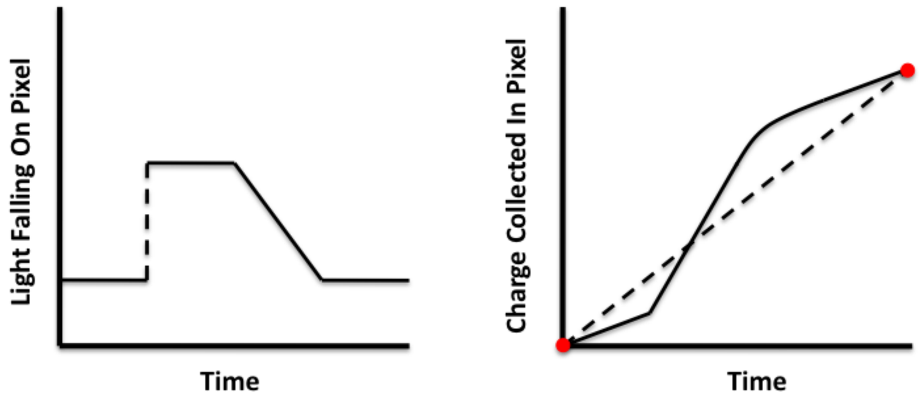

Figure 1.4: A simple cartoon showing a transient astronomical event such as a flaring M star and how the CCD perceives it. The left figure shows the number of photons hitting the pixel vs. time. The right figure shows the charge collected in the pixel as a function of time. The only data points available to the observer are the red dots, leaving the actual light curve of the object a mystery.

With a CCD, **no information on when the event occurred or its temporal signature** is available. Only the integrated flux is obtained.

1.2.2.2 Serial Output of Pixel Values

Perhaps the greatest limitation of the serial CCD read-out sequence is that pixels cannot be randomly accessed. As an extreme example, if we wish to read a small, rectangular subset of 15x15 pixels, or *window*, in the corner furthest away from the output amplifier of a 1024x1024 array, we must read out the other 1,048,351 pixels as well. For repeated exposures, the duty cycle for the 15x15 window will therefore be very low. There are ways of running the CCD in which the undesired pixels are read out at a faster rate than the pixels of interest, but this requires complicated clocking patterns and results in higher noise due to a lower Charge Transfer Efficiency (CTE). Ultimately, the horizontal clock frequency, f_H in units of pixels/second, will be limited by how fast the CCD can be "erased"[3].

One area of astronomy where the serial readout is a significant hindrance is high-speed photometry. Consider the case where one would like to do temporal photometric measurements of a variable star (similar arguments will also apply for planetary transits, supernovae, etc.). The star is expected to fluctuate in brightness because of its intrinsic variability, but it will also fluctuate because of atmospheric scintillations. In some cases, the atmospheric effects might dominate. To decouple the two effects, a reference star in the same isoplanatic field must be measured simultaneously.[6]

[6] The *isoplanatic* field is the sky field over which perturbations in the wavefront are more or less the same. Since the light from all stars in this field effectively travels through the same column of air, the scintillation patterns for the stars within it should be similar.

If a CCD is being used to record the reference *and* variable stars, the maximum frame rate will be a function of the separation of the two stars. The limiting factor may, in fact, turn out to be the speed of the shutter in front of the chip, but we will ignore this here.

1.2.2.3 Slow Read-Out Speed

The serial format and charge coupling mechanism used in CCD read-out along with the limited number of outputs on the chip generally result in a slow frame rate. The electrons along the CCD channel must overcome a frictional force as they are shifted between gates, which sets a fundamental limit on the speed at which they can be transferred. In fact, past attempts to boost CCD speed led manufacturers to explore alternative semiconductors such as GaAs purely because the increased electron mobility allows a faster transit time between gates [6]. Adding more outputs is prohibitive because it requires considerable power for driving the electrodes and substantial signal generation and processing effort [7]. Additionally, interline and frame-transfer CCDs that utilize "electronic shuttering" to boost speed suffer from image smear and are generally only effective to a minimum exposure time of 20 μs [8].

There does exist a certain class of "Ultrahigh-frame CCD imagers" that can operate at speeds of up to a million frames per second [9] with a read noise of ~ 15 e^- rms. And many CCD sensors designed for adaptive optics operate at less than 2 e^- rms [10] with frame rates greater than 1000 fps. However, all of these CCDs come in extremely small formats ($< 264 \times 264$ pixels) and are not intended for use in large focal plane arrays for astronomy. In fact, the electronic circuity surrounding the very small, light sensitive silicon detector dominates the packaging and completely precludes the capability of butting them together to form a multi-detector focal plane array.

Modern deep-depletion CCDs, which are most similar to the hybrid CMOS SiPIN arrays, are limited to about 70 kpixels/sec. For a large format megapixel array, this yields frame rates on the order of tens of seconds [1]. The clocking speed can be increased, but usually results in an unacceptable increase in the read noise and the charge transfer inefficiency.

1.2.2.4 High Power Consumption

The act of moving charge in a CCD is an energy intensive process. As charges are moved from pixel to pixel, they must overcome resistive forces and potential barriers, and energy must be expended in order for them to charge and discharge the capacitance of the gate electrodes in each pixel. The clock voltages in a CCD typically exceed 12 volts and a typical CCD might require 25 W to operate [3].

For some ground based astronomy missions the large power requirements may not be an issue. However, for instruments aboard satellites, the added power can be a drain on solar panels.

1.2.2.5 Susceptibility to Radiation Damage

CCDs are inherently vulnerable to damage by high energy radiation in the form of both photons and charged particles. According to Janesick, radiation damage is the “Achilles Heel“ of the CCD because thousand of transfers are required to move the signal charge to the output, so image lag occurs when the silicon is damaged [11]. A damaged pixel in the column nearest the output register can corrupt its entire row, leading to a decrease in the CTE of the device. In addition, high energy protons and neutrons can induce “dark current spikes“ that are not easily subtracted because they depend on input flux.

Radiation damage is a critical concern in space based astronomy missions. In low earth orbit, the Van Allen Belts provide a significant hazard, and in high earth orbit, instruments aboard a satellite are constantly bombarded by charged particles from the sun. The CCDs aboard the Hubble Space Telescope (HST) experienced large increases in dark current, an increase in the number of hot pixels, and degradation of CTE as a result of high energy irradiation [12]. The CCDs aboard the Galileo spacecraft had dark spikes so large after the mission was completed than an exposure taken at 17°C saturated within a minute [3].

Most, if not all, CCDs in space-based applications to date have been *backside thinned* n-channel CCDs that show this vulnerability to high energy radiation. Deep depletion CCDs, on the other hand, show promise for space flight because of their radiation hardness. According to Bebek et al. [13], the primary reason they are radiation hard is that they are p-channel devices. The dominant trap in the p-channel is the divacancy (VV) trap, which is less likely to be formed by irradiation than the corresponding phosphorus-vacancy (PV) trap in an n-channel. Also, the VV trap lies further from the mid-band (0.20 eV above the valence band) than the PV traps (0.45 eV), making it less efficient in producing dark current.

1.2.2.6 Poor Quantum Efficiency at Red Wavelengths

This problem is inherent to the typical thinned CCD. The absorption length of light quickly increases with wavelength, so photons at the red end of the visible spectrum have a small chance of interacting with the silicon if it does not have sufficient thickness. Deep depletion CCDs do not suffer from this problem.

1.2.2.7 Requirement of Mechanical Shutter

Because the CCD has no electrical mechanism to cease photogenerated electrons from being created by incoming light, a mechanical element must be placed over the CCD to block light when it is being reset. In many applications in astronomy, this does not pose a problem, as small shutters are straightforward to design and build. However, for the future generation of extremely large cameras such as the Large Synoptic Survey Telescope (LSST), very large, fast shutters must be used with the array. This can present a challenge when real estate inside the camera body is needed for elements such as filters, lenses, and cabling, etc.

1.2.3 Where the CCD Wins in Astronomy

There are several reasons why CCDs have reigned supreme in astronomy since their inception. Here are listed the most important ones:

1.2.3.1 Low Read Noise

If the CCD is clocked and biased with optimal electrode voltages and no charge traps exist in the path of a charge packet on its way to the output amplifier, the charge transfer is fundamentally a noiseless process. The noise floor on the source follower output amplifier, limited by $1/f$ flicker noise, can be as low as 1 electron [14]. In astronomy, this low read noise is essential for maximizing the signal to noise ratio for hard-to-detect, faint sources such as galaxies near the edge of the visible universe.

1.2.3.2 Low Dark Current

Low dark current is essential for imaging faint sources that starve the detector of photons. A very long exposure time is required to image dim and distant galaxies, the pinnacle example being the Hubble Ultra Deep Field, where the exposure time was one million seconds [15]. The dark current of the Hubble WFPC2 CCD was as low as 5.7 $\times 10^{-5}$ e^-/s/pix at -88° C, allowing this incredibly long exposure to be taken without the potential wells filling up significantly [16]. Even commercial CCD cameras can achieve a dark current of less than 1 e^-/s/pixel at temperatures achievable with thermoelectric cooling elements [5].

1.2.3.3 Linearity

CCD pixels have a very *linear* response to light, with respect to both flux and fluence.[7] A given light source will produce twice as much signal in twice as much time, and a signal twice as bright as another will produce twice as much signal in the same period of time. The high linearity allows for high dynamic range imaging (75 dB for frontside illuminated devices and 90 dB for backside illuminated [17]). This linearity is absolutely essential in making photometric measurements across a range of astronomical magnitudes. While nonlinearities can be corrected for with proper calibration, this is cumbersome and wastes time when calibration exposures must be collected during observation.

[7] *Flux* is a measure of the energy of the light falling on a given area per unit time. *Fluence* is the product of flux $\times$ time.

1.2.3.4 Charge Transfer Efficiency

Charge transfer in the CCD has some advantages. It allows for binning of multiple pixels on the chip (treating a 2x2 region as one pixel), special applications like charge shifting for tip/tilt correction and drift scanning, and it allows the low noise amplifier at the output to be built without space constrictions. However, charge transfer in the CCD also has certain disadvantages. In addition to requiring high power, slow operation, and a mechanical mechanism to block light from the pixels, a "hot" pixel or defect can corrupt an entire column of the CCD. As long as a row is not corrupted by charge traps, transferring charge along it can be done with a Charge Transfer Efficiency (CTE) of 0.99999 [18]. This means that nearly every electron collected in a given pixel is accounted for when it is converted to a voltage on the output amplifier. As will be seen, though, CTE is not an important metric for comparing CCDs, which require transfer of charge and operate in the *charge domain*, to devices that relay a voltage from the pixel to an output and operate in the *voltage domain*. The consideration of devices that work in the voltage domain naturally brings us to the discussion of CMOS imagers.

1.3 CMOS: Motivation for a New Detector

The title of this section is a little misleading. CMOS detectors have been around since 1967, just as long as CCDs have. However, early CMOS imagers had *passive pixels*, and their performance relative to CCD detectors was second rate [19]. In the passive pixel architecture, shown in Figure 1.5, a photodiode converts light to charge and a simple switch connects the pixel signal charge to the column bus capacitance when it is selected for readout [20]. The performance suffers because the large capacitance of the column bus (one that increases with the dimensions of the bus) reduces signal to noise and slows down charge transfer. In fact, many people referred to CMOS as the "poor man's CCD"! But with the advent of *active pixel sensors* (APS) in 1997, CMOS imagers began to gain ground on CCD detectors. In active pixel sensors, an "active" transistor within the pixel unit cell buffers the charge in the pixel to the output. The active pixel provides lower noise readout, improved scalability to large array formats, and higher readout speed compared to the passive pixel devices [21]. Nearly all CMOS imagers utilize the APS architecture, and passive pixels will not be considered further.

It is far beyond the scope of this dissertation to discuss the plethora of different CMOS imager technologies available. In the following sections a very broad overview of CMOS sensors will be given, mainly to illustrate the differences between CCD and CMOS, and to highlight the great potential of CMOS sensors to simplify and improve astronomical focal plane arrays. An ample set of references will be listed along the way to direct the reader towards more thorough descriptions.

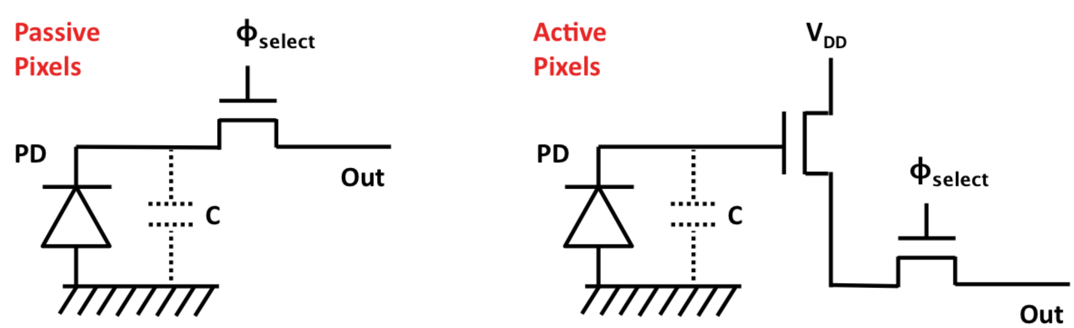

Figure 1.5: A simplified diagram that shows the main difference between the passive and active pixel CMOS architectures. In the passive version, the charge in the pixel is fed through the column bus directly to the output. In the active version, a source follower MOSFET buffers the voltage stored on the capacitance of the pixel.

1.3.1 Overview of CMOS Imager Operation

The manner in which photons are converted into charge carriers in a CMOS detector is no different from a CCD. Each pixel contains a photodiode (usually just a simple reverse biased p-n junction) that separates electron hole-pairs, which in turn creates a photocurrent. The way in which the photocurrent is measured varies among CMOS devices. The three simplest and most widely used input circuits are the source follower per detector (SFD), the capacitive transimpedance amplifier (CTIA) and the direct injection (DI) [22]. The DI and CTIA architectures are well suited for high flux applications and will not be discussed in this dissertation. For further description, the reader is referred to Hoffman et al. [22] and Beletic et al. [23]. The SFD architecture is preferred in astronomical applications since it offers the lowest noise. In this architecture, the photocurrent integrates in the pixel–mainly on the capacitance of the photodiode and source follower–to be measured as an accumulated quantity of charge. Unless otherwise mentioned, all references will be made in regard to SFD CMOS detectors.

The APS pixel diagram on the right side of Figure 1.5 dramatically oversimplifies a CMOS pixel. However, it does illustrate one of the key features of CMOS detectors: a pixel is sampled *by activating a switch* (or set of switches) that connects it to an output. For a two-dimensional CMOS imaging array, two switches are used to address a given pixel: one to select a column and one to select a row. Because the signals from multiple pixels are read out through one or more outputs[8] with the proper choice of addressing, the CMOS array is also referred to as a *multiplexer*. Any of the pixels can be randomly accessed at any given time by "dialing in the proper address" on a set of addressing shift registers at the periphery of the array, as shown in Figure 1.6. For conventional readout of the full pixel array, the switches are toggled in a serial fashion. Usually a given row is selected and the column buses are sequentially connected to the output before moving onto the next row, giving rise to a *slow* axis and *fast* axis, just as in the case of a CCD. A similar clocking pattern that connects only a subset of rows and columns can be used to yield a *window* of pixels on the detector.

Figures 1.5 and 1.6 illustrate another key difference between the readout in CCD and CMOS detectors. Because the charge integrated by the photodiode in the CMOS detector is buffered to the output by a transistor, reading the pixel is **non-destructive**. That is, sampling the pixel does not upset the charge distribution on the photodiode or "reset" the pixel. Unlike a CCD, where *charge* must be shifted from the pixel to the output of the detector in order to be sampled, in a CMOS detector a *voltage* is simply relayed to the output. An important consequence of this is that **the exposure and readout can occur simultaneously**.

A more realistic rendition of a conventional "3T" (3 transistors in each pixel) CMOS pixel is presented on the right in Figure 1.6. In the 3T architecture, a source follower FET (SF) buffers the pixel voltage, a row select FET (SEL) connects the buffered voltages of all the pixels in a given row

[8]For a good discussion of multiple outputs in the context of CMOS detectors (and accompanying diagrams), the reader should consult Moore [24].

to their respective column buses, and a reset FET (RST) resets the pixel by restoring the reverse bias on the photodiode. For high quality scientific imaging, the 3T pixel architecture is rarely used in *monolithic* CMOS devices since the sensitivity is limited by the capacitance of the photodiode, source follower gate, and the reset transistor source terminal [25]. However, it is still used in many *hybrid* CMOS imagers, where the pixel capacitance is often dominated by the photodiode in a separate detector layer. The distinction between monolithic and hybrid CMOS detectors will now be discussed.

1.3.2 Monolithic CMOS Imagers

In a monolithic CMOS imager, the detector array and accompanying readout integrated circuit (ROIC) are produced in the same substrate, and both are thinned [26]. Standard monolithic CMOS imagers can be manufactured in the same foundries that produce standard microchips for computers and high-end electronics, which guarantees cost-efficient production and highly mature process technologies [22]. For this reason, they are extremely popular in commercial applications such as cell-phone and digital cameras. However, the standard CMOS processes cannot produce the type of high-performance imagers required in astronomy, and so custom manufacturing schemes similar to those used by CCDs are required [11]. Still, custom CMOS is unconstrained by CCD process requirements, so it is relatively inexpensive compared to large scientific CCDs [27].

A large appeal of monolithic CMOS imagers is that performance improving and power saving

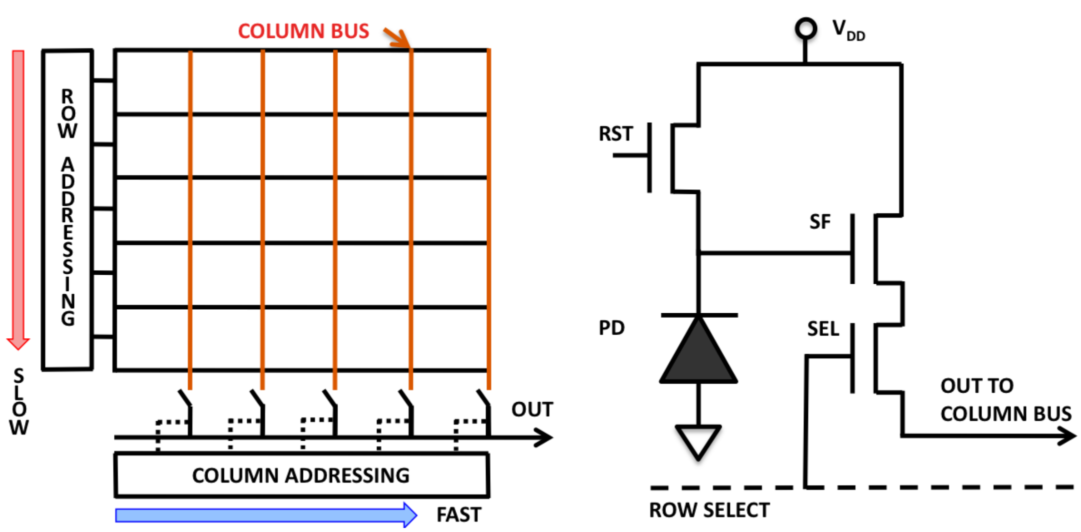

Figure 1.6: (Left) Diagram of a simple 5x5 CMOS multiplexer with one output, following Magnan [17]. The slow and fast read directions are indicated by the colored arrows and the column buses are indicated as the orange lines. (Right) 3T pixel schematic showing the reset transistor (RST), source follower transistor (SF), row select transistor (SEL), photodiode (PD), and the supply voltage V_{DD}.

signal processing operations can be done *in the pixel*. Adding just one additional transistor (referred to as the transfer gate) to the pixel allows separation of the photodetection and photoconversion regions. This "4T" arrangement, called a pinned photodiode, allows correlated double sampling in the pixel and greatly reduces the readout noise by increasing the sensitivity (μV/e$^-$) [28]. 4T pixels achieve readout noise levels comparable to CCDs (primarily because charge is transferred onto a sense node before conversion as it is in a CCD). Adding additional transistors to the pixel allows for additional in-pixel functionality such as threshold detection, A/D conversion, anti-blooming, and contrast stretch [29, 30, 31].

The problem with adding additional transistors–or having any transistors in the pixel at all, for that matter–is that it reduces the fill factor of the pixel.[9] This is highly undesirable in low light-level applications like astronomy, where every photon counts. The solution in commercial applications is to insert tiny microlenses above each pixel that direct the incoming light into the photosensitive regions. The microlenses do increase the effective fill factor of the pixel, but they can also lead to blurring, increased spatial non-uniformity, vignetting, and poor angular response [32]. Another solution being pursued is to illuminate the monolithic device from the *backside* instead of the *frontside*. In a backside illuminated device, the incoming photons do not pass through the metal lines and transistors in order to reach the photodiodes. Rather, the photons must pass through the photodiodes to reach the non-light-sensitive components. This is the same approach that was taken for CCDs to increase the response for short optical wavelengths and offers some promise. At the time of writing, though, the vast majority of CMOS devices are frontside illuminated and only a select few vendors are able to manufacture backside devices successfully [22, 25, 33]. Plus, the thickness of these devices is typically on the order of 10 μm, so the quantum efficiency at red and near infrared wavelengths is subpar.

To summarize, monolithic CMOS detectors are on the brink of rivaling CCDs in high light-level applications. Fow low light-level applications, the **two main deficiencies** are **1) low fill factor** and **2) poor quantum efficiency**. Two of the most prevalent solutions used to compensate for these deficiencies are using microlenses or backside illumination. The first of these is not a viable alternative for astronomy and the second provides only a limited increase in quantum efficiency. An alternative solution involves mating a separate array of detectors to the monolithic device. This process of *hybridizing* the CMOS imager to a separate detector layer will be covered in the next section.

1.3.3 Hybrid CMOS Imagers

In a hybrid CMOS array, two separate layers are joined together to form an imager. One layer is a pixelated array of photodetectors that serves to convert photons into charge carriers. This layer is usually referred to simply as *the detector*. The other is a monolithic CMOS device, or

[9] Fill factor is the fraction of area of the pixel that is light sensitive.

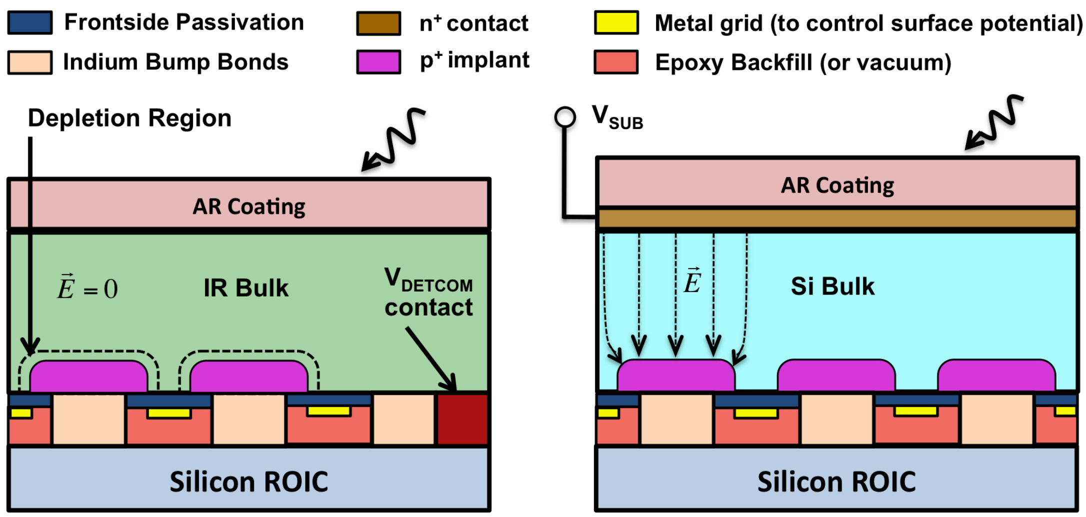

Figure 1.7: Diagrams of hybrid CMOS Arrays. AR stands for Anti-Reflection coating. (Left) A per-pixel depleted detector commonly used in infrared astronomy in which the depletion region (shown by the dashed line) only extends slightly above the pixel implant. The bulk is held at a constant voltage V_{DETCOM} administered through the silicon multiplexer. (Right) A fully depleted detector in which the backside contact voltage V_{SUB} fully depletes the bulk of charge carriers and creates an electric field that extends throughout. In both device types, the manufacturers may place an epoxy backfill in the region between the detector and the ROIC to prevent stress buildup and place a metal underneath the frontside passivation to control the surface potential.

Readout Integrated Circuit (ROIC), that allows the electrical signal in each pixel to be measured. As described in the previous section, the ROIC generates clocks and biases necessary to operate the detector, handles the amplification and multiplexing of signals, and may provide additional processing capabilities. The two layers are joined together via a set of bump bonds, as shown in Figure 1.7. In discussing the hybrid imager, the term *detector* will be used solely to refer to the detector layer and the terms *multiplexer* and *ROIC* will be reserved for the CMOS readout layer. The terms *device*, *imager*, and *sensor chip assembly* (SCA) will be used to refer to the integrated device.[10]

Hybrid CMOS is by no means a new technology. In fact, hybrid CMOS focal plane arrays have dominated infrared astronomy since their inception in the 1980s [34, 4, 35]. As Rieke points out, the near non-existence of monolithic infrared imaging arrays is primarily due to the fact that the materials sensitive to infrared wavelengths (silicon cuts off at a wavelength around $\lambda_c = 1$ micron) generally do not have the properties needed for high-performance electronics, such as the easily formed rugged oxide of silicon that allows for robust insulating layers [36]. The hybrid architecture circumnavigates this problem by allowing the ROIC to be made out of silicon and the detector to be

[10]The term *HyViSI* will also be used when referring to an assembled HxRG ROIC and PIN diode array.

made of a different semiconductor material. The detector material can be specifically chosen based upon the wavelength of interest. For instance, infrared detector vendors fabricate near infrared InSb detector arrays with $\lambda_c = 5.5$ microns as well as mid infrared HgCdTe arrays with $\lambda_c = 1.24 - 14$ microns and Si:As arrays with $\lambda_c = 25$ microns. All of these detectors can be bump bonded to the same silicon readout circuit [18].

While mating an infrared sensitive material to a silicon ROIC has been done routinely for several decades in infrared astronomy, it turns out that mating a silicon detector to a silicon ROIC has not been a common practice in optical astronomy. Rockwell Scientific has been fabricating Silicon PIN detectors detectors[11] since 1998 [37] and Raytheon Vision Systems has been fabricating them since 1988 [38], but the first time they were tested on a telescope was 2007 [39, 40]. These silicon PIN arrays seek to provide 100% fill factor and high quantum efficiency in the ultraviolet and near infrared while still offering the advantages of the CMOS readout circuitry.

Before discussing the advantages of hybrid CMOS in detail, another important distinction between the infrared and optical hybrid CMOS sensors needs to be made. It has to do with the extent of the depletion region in the detectors. In general, the infrared arrays are **per-pixel depleted**. That is, each pixel in them has its own small depletion region that extends around the collecting implant, as shown on the left in Figure 1.7. The bulk of the detector is conductive and free of an electric field, so it is usually made thin in order to prevent carriers from wandering too far from the pixel over which they were generated. The nodal capacitance in these detectors is usually dominated by the depletion capacitance, which changes as the pixel integrates charge. The changing capacitance presents a challenge since it makes the response of the pixel nonlinear over the full well capacity [41]. Another challenging area in these detectors is the interface of the bulk material lying directly above the surface passivation layer, as traps in this region can lead to increases in dark current and image persistence. Solomon and Moore provide a good treatise on these issues and others associated with per-pixel depleted arrays [42, 43].

In contrast to the per-pixel depletion arrangement, Si PIN arrays are typically operated as **fully depleted detectors**. The high purity, high resistivity silicon used in them allows the bulk to be fully depleted with modest voltages (5-10 volts) even for large thicknesses (80-200 microns). As shown in the right of Figure 1.7, full depletion means the electric field extends all the way to the back surface of the detector array. The electric field, generated by applying a voltage V_{SUB} to an n^+ contact at the backside of the detector, inhibits lateral diffusion and the chance of recombination, which in turn improves Charge Collection Efficiency (CCE) and the detector point spread function. Because the detector can be made thick without compromising CCE, fully depleted detectors offer a solution to the quantum efficiency and fill factor problems associated with monolithic CMOS devices. The greater thickness also extends the detector response into the near-infrared

[11] PIN detector arrays consist of P-I-N, or **P** type-**I**ntrinsic-**N** type photodiodes. They are discussed in depth in section 2.1.

1.3.4 Advantages of CMOS Arrays for Astronomy

The advantages of CMOS detectors for astronomy are well matched to the disadvantages of CCDs. They are listed here in the order followed in Section 1.2.2. In some cases the advantages apply to hybrid CMOS imagers and not the monolithic type.

1.3.4.1 Non-Destructive Readout

The non-destructive readout in CMOS imagers allows each of the pixels in the array to be sampled multiple times during a long integration. Not only does this enable a reduction in read noise; it provides temporal information on astronomical sources. For instance, the charge versus time profile in Figure 1.4 may be sampled at even intervals from the start of the exposure to the end. As long as the pixel is not saturated or railed by the analog-to-digital converter, an estimate of the object flux can be made at each of these points.

Multiple samples during an integration may also yield flux estimates for objects that saturate during the middle of the integration. This enables high dynamic range imaging in a single exposure. Cosmic ray rejection is also greatly facilitated with non-destructive reads of the detector (see reference [44]).

1.3.4.2 Random Access to Pixel Values

Random access to pixels has several unique applications in astronomy. For high speed photometric measurements of a fast variable source, a small window of pixels can be rapidly read out at high frame rates unattainable when reading the full pixel array. Alternating between this window and another one containing a reference source rejects common mode brightness fluctuations due to atmospheric turbulence or cloud cover.

Another area where the windowing capability of CMOS sensors shows extreme promise is in telescope guiding. Because the pixels can be randomly accessed, a star can be imaged at the same frame rate anywhere on the detector. More importantly, to first order **the frame rate for the window is independent of the size of the full array**. The LSST focal plane is a wonderful example of how useful this feature is. Guide sensors will be placed in each of the four corners of the LSST focal plane, and a large collecting area is needed in each of the sensors to ensure a high probability that a bright enough star will be available to guide [45]. With a CCD, the frame rate depends on the size of the sensor, and large format, high speed CCD arrays are not available. Thus, LSST will use 2k×2k hybrid CMOS SiPIN arrays as the guide sensors, and a large collecting area will be achieved without penalty in speed.

If Hybrid CMOS SiPIN sensors were used for the science sensors in LSST, the windowing capability could be taken one step further. As will be discussed in Section 3.3, a windowed readout of the array can be interleaved with a full frame readout of the array in a *guide mode*. If all of the science sensors were CMOS, any one of them could be used to guide the telescope while simultaneously participating in the full exposure.

Lastly, the windowing capability is very appealing for adaptive optics applications that demand high frame rates ($\sim$1000 Hz). As mentioned in Section 1.2.2.3, CCDs used for adaptive optics come in very small formats. Large format CMOS detectors can offer the same speed as these specialized CCDs with the benefit of having a large collecting area from which to choose a star.

1.3.4.3 Fast Read-Out Spead

Operating in the *voltage domain* makes CMOS imagers inherently faster than CCDs. There is no need to worry about CTE in a CMOS device (at least the ones that do not employ pinned photodiodes). Also, true electronic shuttering of a CMOS device can be done in a 2 to 10 μs range [8], so resetting the pixels generally does not limit frame speed. Instead, the speed limitation is set by the settling times called for by the capacitance of the column bus, output lines, and input stages of the external acquisition electronics. Adding additional outputs in a CMOS imager does not create significant power demands and so presents an easy way to boost the full frame speed. Megapixel monolithic CMOS arrays can be operated at 1000 fps [7] and hybrid SiPIN CMOS arrays can achieve speeds of 150 fps [25]. An RMS readout noise is not reported for these speeds.

The fast frame times achievable with CMOS arrays open exciting possibilities for high speed measurements of pulsars, rotating radio transients (RRATs), and other yet unknown fast variable sources. And as alluded to in the previous section, adaptive optics in astronomy routinely demands these high speeds.

1.3.4.4 Low Power Consumption

Operation of monolithic CMOS sensors usually requires no more than one voltage source at 3.3 volts. Hybrid CMOS SiPIN arrays need an additional, higher voltage applied to deplete the bulk, but since it is a reverse bias, a negligible current is drawn. Thus, both types consume a low power and dissipate a small amount of heat that can generally be expressed in milliwatts (1-200 mW) [17, 25].

1.3.4.5 Radiation Tolerance

Both Hybrid and Monolithic CMOS arrays are intrinsically more tolerant to high energy radiation than CCD detectors. This radiation is one of the most pressing problems for devices operating in the harsh environments of low or high-altitude orbit, where there is a significant background of high-energy protons ($>$ 1MeV) and neutrons, heavy ions, high-energy gamma-rays, x-rays, and electrons. Part of the reason that CMOS detectors are not as prone to radiation damage has to do with the fact that they do not require charge transfer across the pixels. A damaged pixel does not compromise the other pixels in its row as it does in a CCD.

In addition to the lack of degradation from CTI, the pixels in hybrid SiPIN CMOS arrays should show less vulnerability to irradiation. There are several types of damage that occur in CMOS pixels, and they are generally considered to be a result of either *ionization damage* or *bulk damage* [46]. Ionizing damage leads to build-up of trapped oxide charges and unfilled traps in SiO_2 layers as well as an increase in interface trap density due to Si-O bond deformation and breakage, along with release of impurities within the SiO_2. Typically SiO_2 passivation layers in CMOS have about the same thickness as the gate oxides in CCDs (on the order of a micron), so increases in dark current due to damage in or near the SiO_2 should be comparable in the two. But in the case of bulk damage, which is caused primarily by protons and neutrons, the hybrid SiPIN CMOS should win over a thinned CCD. This is because the tendency of heavy irradiation is to push Silicon from its initial doping towards a slightly p-type quasi-intrinsic (π) material [47]. SiPIN detectors are tailored to have n-type quasi-intrinsic Si at fabrication, whereas the epitaxial layer and channel of a CCD are significantly doped. Thus, effects of the bulk irradiation will cause more of a change for the CCD dark current and voltages necessary for clocking than for the hybrid CMOS. Along with the bulk material, the CMOS multiplexer itself is radiation tolerant. The CMOS structures of the readout multiplexer are inherently radiation hard to levels greater than those required for any astronomical missions ($>$ 100 krads) [48].

1.3.4.6 Good Quantum Efficiency at Red Wavelengths

While monolithic CMOS detectors show very poor quantum efficiency (QE), hybrid CMOS SiPIN detectors excel in this category. Hybrid CMOS SiPIN imagers outperform many CCDs above 500 nm and have a response that extends into the near infrared [39, 49]. The QE decreases at all wavelengths for decreasing temperature due to phonon absorption length. Yet it still remains relatively high for temperatures in the range of 120-160 K, below the onset of high dark current and other deleterious effects that will be discussed later.

1.3.4.7 Electronic Shuttering Capability

The reset transistor in CMOS pixels support a frame refresh without mechanical shuttering. This is incredibly useful in astronomy, where mechanical shutters have been a "perennial problem" [27]. Large shutters will be especially problematic with the increasingly large focal planes being planned for extremely large telescopes. In some high speed applications, special methods must be implemented in electronic shuttering to avoid motion artifacts. This should not be a problem in astronomy, however.

1.3.5 Disadvantages of CMOS Arrays in Astronomy

With all of the advantages of CMOS imagers in astronomy, one might wonder why they are not used in populating the focal planes on every science grade telescope. The answer lies in the fact that the advantages just listed are secondary in importance to the small number of disadvantages.

1.3.5.1 High Read Noise

As previously mentioned, monolithic CMOS arrays with 4T pixels are now achieving very low read noise, but are not candidates for astronomy because they waste a great deal of incoming photons. The hybrid CMOS arrays that are contenders in astronomy because of their good fill factor and QE typically have a correlated double sample (CDS) read noise of 8-10 e^- RMS, far above the single or sub-electron read noise delivered by CCDs. With multiple non-destructive reads, the noise for a single pixel can be reduced to about 2-4 e^-, which is a significant improvement. But in optical astronomy, these few extra electrons of noise can mean the difference between detecting a source and missing it altogether, which is one of the main reasons CCDs are preferred.

1.3.5.2 High Dark Current

Leakage currents in monolithic and hybrid CMOS imagers have been a very big problem and area of study for manufacturers. Specialized processing techniques and pixel architectures are being implemented to try and decrease dark current to bring it to levels comparable to CCDs, but these increase cost and decrease yield [50, 51]. In hybrid CMOS SiPIN sensors, acceptable dark currents for certain astronomy applications (0.001-0.01 e^-/s/pix) can be obtained, but the operation temperature must be brought rather low ($<$ 160 K).

1.3.5.3 Linearity

Linearity is not as much of an issue in fully depleted hybrid CMOS detectors as it is in per pixel depleted detectors because the change in depletion region width is relatively small in comparison to the full depletion width. Improper bias voltages can lead to exponential signal behavior in hybrid CMOS SiPIN devices, but if tuned properly, the integration of photocurrent over time is linear over more than 90% of the full well.

1.3.5.4 Persistence

Hybrid CMOS sensors are prone to an effect called *persistence* in which previously well-illuminated pixels show a recurrence of signal after reset. The recurring signal, or *latent image*, can last from seconds to hours depending on the mode of operation, temperature, bias voltages, history, etc. Persistence is very troublesome in the context of astronomical surveys since regions afflicted by bright stars are rendered unusable for some time thereafter. These regions cannot be used to accurately measure flux until the persistence has subsided, so they are essentially wasted pixels during that time.

1.3.5.5 Interpixel Coupling

In addition to diffusive crosstalk that occurs while charge carriers are being collected–an effect that is common to CCDs–hybrid CMOS pixels show coupling in the form of Interpixel Capacitance (IPC) after charge collection [43, 52]. Interpixel capacitance attenuates Poisson noise, increases the detector point spread function, and causes single pixel x-ray events to appear as being spread over multiple pixels. It is a *deterministic* mechanism, so it can be removed from astronomical images with proper calibration (a deconvolution with the detector impulse response). This adds a layer of complexity to data reduction, though.

In Teledyne HyViSI detectors, another mechanism of interpixel coupling is observed in which pixels appear to transfer charge to each other. This effect, which will be referred to as Interpixel Charge Transfer (IPCT), occurs only at temperatures greater than about 130 K. It leads to underestimates of x-ray energies at high temperatures ($>$ 160 K) and long frame times. It does not appear to pose any more threat to optical observations than persistence, but must absolutely be taken into account in x-ray applications. It is worth mentioning here that IPCT has been significantly reduced with improved surface treatments in new detectors. This will be covered in more detail in Section 6.1.2.