HYBRID CMOS SIPIN DETECTORS AS ASTRONOMICAL IMAGERS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Lance Michael Simms
November 2009

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Steven Kahn) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Sarah Church)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Malcolm Beasley)

Approved for the University Committee on Graduate Studies.

Abstract

Charge Coupled Devices (CCDs) have dominated optical and x-ray astronomy since their inception in 1969. Only recently, through improvements in design and fabrication methods, have imagers that use Complimentary Metal Oxide Semiconductor (CMOS) technology gained ground on CCDs in scientific imaging. We are now in the midst of an era where astronomers might begin to design optical telescope cameras that employ CMOS imagers.

The first three chapters of this dissertation are primarily composed of introductory material. In them, we discuss the potential advantages that CMOS imagers offer over CCDs in astronomical applications. We compare the two technologies in terms of the standard metrics used to evaluate and compare scientific imagers: dark current, read noise, linearity, etc. We also discuss novel features of CMOS devices and the benefits they offer to astronomy. In particular, we focus on a specific kind of hybrid CMOS sensor that uses Silicon PIN photodiodes to detect optical light in order to overcome deficiencies of commercial CMOS sensors.

The remaining four chapters focus on a specific type of hybrid CMOS Silicon PIN sensor: the Teledyne Hybrid Visible Silicon PIN Imager (HyViSI). In chapters four and five, results from testing HyViSI detectors in the laboratory and at the Kitt Peak 2.1m telescope are presented. We present our laboratory measurements of the standard detector metrics for a number of HyViSI devices, ranging from 1k×1k to 4k×4k format. We also include a description of the SIDECAR readout circuit that was used to control the detectors. We then show how they performed at the telescope in terms of photometry, astrometry, variability measurement, and telescope focusing and guiding.

Lastly, in the final two chapters we present results on detector artifacts such as pixel crosstalk, electronic crosstalk, and image persistence. One form of pixel crosstalk that has not been discussed elsewhere in the literature, which we refer to as Interpixel Charge Transfer (IPCT), is introduced. This effect has an extremely significant impact on x-ray astronomy. For persistence, a new theory and accompanying simulations are presented to explain latent images in the HyViSI.

In consideration of these artifacts and the overall measured performance, we argue that HyViSI sensors are ready for application in certain regimes of astronomy, such as telescope guiding, measurements of fast planetary transits, and x-ray imaging, but not for others, such as deep field imaging and large focal plane astronomical surveys.

Preface

More than four centuries have passed since Galileo first pointed a telescope at the sky and observed things that no human being had ever laid eyes upon. In that moment, he opened up an entirely new realm for exploration. We could see farther into the depths of space than ever before and find objects that lay hidden to our naked eyes.

When we look back in history, we see many cases like this. They are cases in which a new technology allowed us to venture into uncharted territory. The invention of the frigate allowed humans to cross seas and explore lands new to them, the submarine opened our eyes to the depths of the seas, the microscope made us aware of new micro-worlds.

Since Galileo's time, our exploration space in astronomy has continued to increase in its vastness. With the invention of photography, we became aware of very dim objects and structures in the cosmos, and a new search was possible. New detectors sensistive to wavelengths of light that we cannot even see revealed a host of new features on the sky. Through the use of this technology, there has been a veritable explosion in the areas in which we can explore.

However, the space we are able to explore in astronomy is still very limited in a fundamental way. We are forced to make theories and postulate about the nature of the universe based solely on the signals we receive here on earth. We make guesses about how massive or hot or distant things are, what composes seemingly empty space, or what causes unfathomably intense bursts of energy in other galaxies purely based upon the light we receive. In essence, we are stuck in exploration space purely because we are stuck here on earth. It is impossible for us to truly *grasp* the cosmos from this limiting vantage point. It is like trying to explore a forest while tied to a tree.

There is no denying that we have made–and will continue to make–progress in understanding the cosmos while we remain tethered to our home planet. Bigger, better, and newer instruments and technology will most certainly open new doors. But the surprises that await us down the road, when we leave our solar system and journey into interstellar space, seem far more likely to provide the kind of shock and revelation that Galileo brought when he first pointed that telescope at the sky.

Some researchers believe that money spent on space exploration is a waste and others argue that funding for fundamental astronomy is. In the end it boils down to differing opinions; not a logical

argument. Like anything else in this world, a good balance between the two seems the best path to take. After all, we cannot simply venture aimlessly into outer space in search of answers without a destination in mind or without properly preparing ourselves for the exotic and harsh environment it presents. That would be like going to a casino with a few quarters and hoping to win a sports car.

Acknowledgments

The commitments and sacrifices involved in undertaking my graduate degree and completing this dissertation were by no means trivial. After finishing my coursework at Stanford, the amount of "free time" I found during the regular work weeks dwindled to an almost nonexistent level. Luckily, the work in itself was extremely enjoyable, and the people I came to know and friendships I formed along the way made it an experience I would gladly repeat. I am extremely grateful to these acquaintances for all of the knowledge they imparted to me and help they offered. Many of them deserve special recognition, which I will attempt to give here in a somewhat chronological fashion.

The first person who stands out in my recollections of graduate school is my advisor, Steve Kahn. Right after accepting me as his student, Steve gave me a veritable "zoo" of opportunities to pursue for my research. Although there were certain ones which he thought would be more beneficial or worthwhile, he never pushed me in any direction that I did not wish to follow. And his incredible breadth of knowledge, along with his many experiences in the world of physics, meant I was always learning something from him no matter the topic of discussion. Despite his extremely busy and chaotic schedule, Steve always found the time to meet with me when necessary. And in circumstances where I needed more attention, he never failed to find an appropriate expert that could advise me.

Among these experts, I owe a very special thanks to Kirk Gilmore. After taking me in during my first year at Stanford, Kirk quickly appreciated my desire to work with detectors and astronomical instrumentation. Being the expert he is, he immediately opened many doors for me in this arena. Throughout the years that followed, he constantly taught me new things, including great laboratory expressions like "Don't King Kong it" — a reference to how much force should be applied to screws when tightening them. There were plenty of other things he did to make me laugh and appreciate the fun side of working on instruments. Plus, he always made sure there "was not too much on my plate".

I also owe special thanks to David Burke and Garrett Jernigan. I did my first real scientific research with David on the atmospheric turbulence at Cerro Pachon (the future site of the Large Synoptic Survey Telescope), and it is an experience I will never forget. He basically taught me how research is meant to be done; every step from organizing and analyzing data to searching

through scientific literature to efficiently presenting results. In a slightly different vein, Garrett first introduced me to the how the scientific community operates and what I needed to consider at each step of my graduate career and beyond. He has continually offered me a very valuable and unique perspective on these sorts of issues and brought my attention to ones that I would not have considered otherwise. And despite his desire to hold on to hardware that clearly belongs in a museum, I can also say he has taught me a great deal about the new and relevant technology in astronomy.

The infancy of my thesis work would not have been possible without two special individuals at Rochester Institute of Technology: Don Figer and Brandon Hanold. Don gave me a first look at laboratory detector work at the Space Telescope Science Institute and showed me a perfect example of how one can be successful in managing and operating a lab while also doing top-notch science. At RIT, he fostered my love for detectors and electronics further. Working in the lab there everyday with Brandon under Don's guidance and hastily preparing a camera system for the Kitt Peak 2.1m telescope was an incredibly rewarding experience. And when the time actually came to do real astronomy, their company made the long nights of observing quite enjoyable.

I also managed to see the other side of the astronomical instrumentation world thanks to Jim Beletic. Jim graciously offered me the opportunity to do an internship at Teledyne Scientific and Imaging, where the real magic of making the detectors happens. At TSI, my appreciation for detectors was greatly enriched by working directly with Yibin Bai, Markus Loose, and Raphael Ricardo. They exposed me to many interesting things and improved my understanding of the devices I had been working with for so long. I am also indebted to Mark Farriss, Rebecca Blockman, Greg Jacques, Richard Blank, Luis Gordillo, and Kevin Peralta for their help there.

Although they do not fit as nicely into the chronological ordering, there are several other people to whom I owe thanks. Andrew Rasmussen and John Peterson spent a great deal of time with me at the blackboard explaining things. Stuart Marshall revealed to me the secrets of Unix. Kirby Hnat and Richard Schickling taught me about cryogenic systems and showed me how to have fun in Tucson. And Dick Joyce, Skip Andree, and a number of other people at Kitt Peak National Observatory deserve thanks for helping us make our telescope observations a reality.

Of course, none of this would have been possible with some very special and dear people in my life that predate this chronology. In addition to being great parents in a general sense, my mother and father have given me every opportunity I could have asked for in life. My older brother has served as a great role model and given me great advice throughout grad school. And I also owe thanks to my sisters and sister-in-law for their constant encouragement throughout this period. I could not have done any of this without them.

Contents

List of Tables

List of Figures

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.

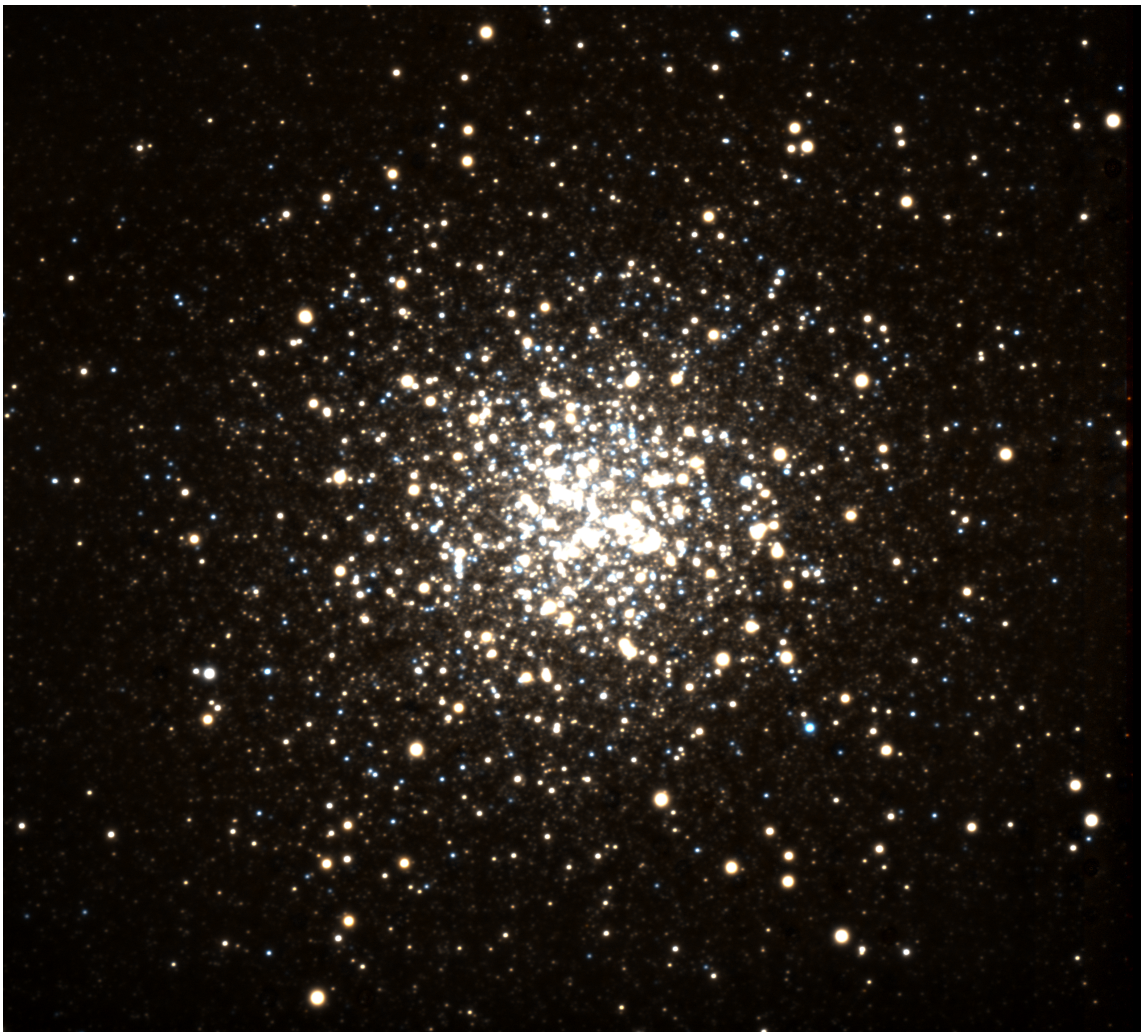


Figure 1.1: A multi-color image of M13, the Hercules Cluster, taken through the Kitt Peak 2.1m telescope. When viewed with the naked eye, M13 appears as a little fuzz in the sky. With only a small telescope, though, a plethora of stars comes into view. When photographic film or a CCD (and color filters) 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.

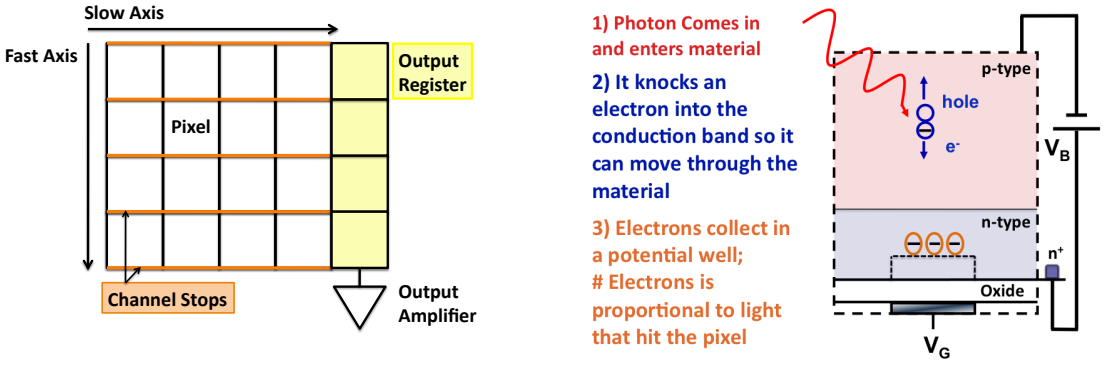


Figure 1.2: (left) A simple cartoon showing the basic features of a 4x4 CCD. The active, light-sensitive pixels are white. The pixels of the output register, shown in yellow, are usually masked from light. (right) A cross-sectional view of a buried n-channel CCD pixel showing how charge is generated 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.

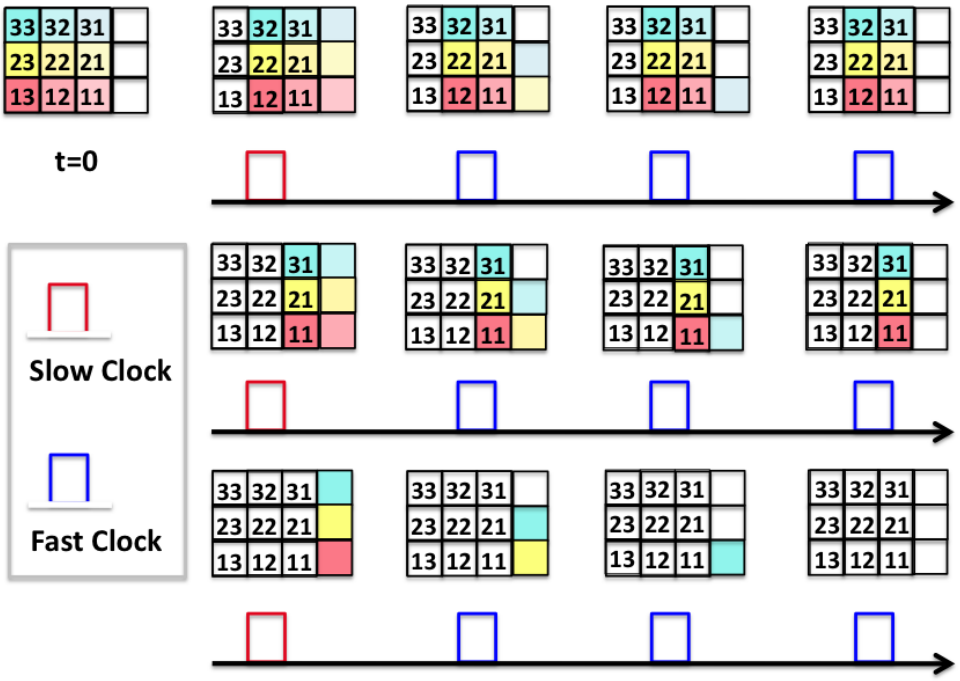


Figure 1.3: A simple cartoon showing the process of clocking in a CCD. The numbers of the pixels designate their physical location in the CCD. The colors have no relationship to wavelength or color of photons; they are solely meant to keep track of charge packets. Only one clock is shown for each fast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

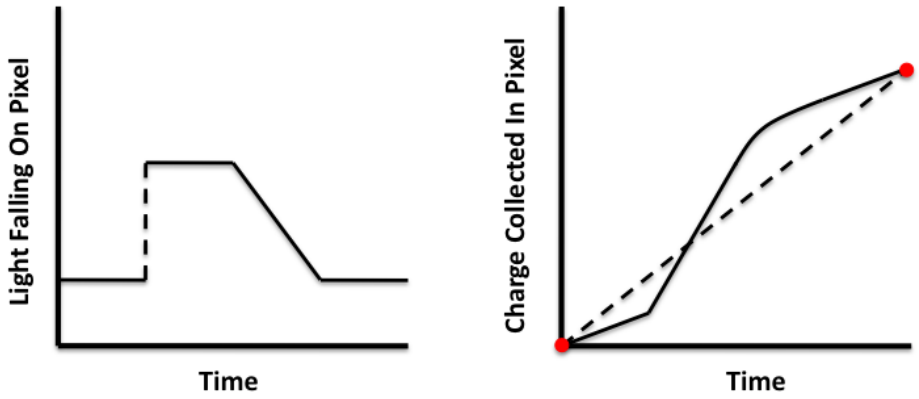


Figure 1.4: A simple cartoon showing a transient astronomical event such as a flaring M star and how the CCD perceives it. The left figure shows the number of photons hitting the pixel vs. time. The right figure shows the charge collected in the pixel as a function of time. The only data points available 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 × 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.

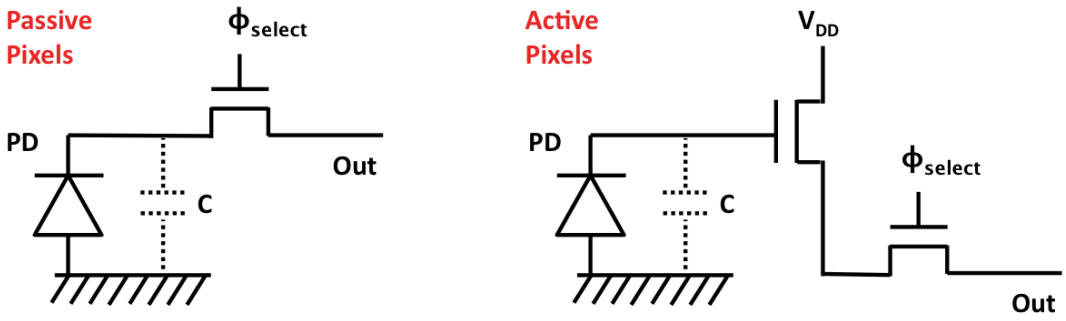


Figure 1.5: A simplified diagram that shows the main difference between the passive and active pixel CMOS architectures. In the passive version, the charge in the pixel is fed through the column bus directly to the output. In the active version, a source follower MOSFET buffers the voltage stored on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

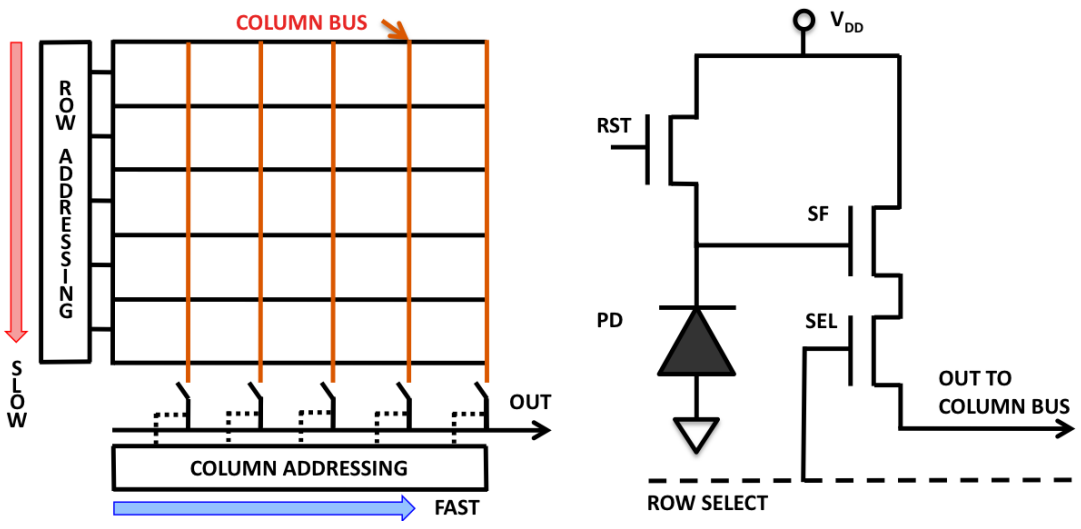


Figure 1.6: (Left) Diagram of a simple 5x5 CMOS multiplexer with one output, following Magnan [17]. The slow and fast read directions are indicated by the colored arrows and the column buses are indicated as the orange lines. (Right) 3T pixel schematic showing the reset transistor (RST), source follower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

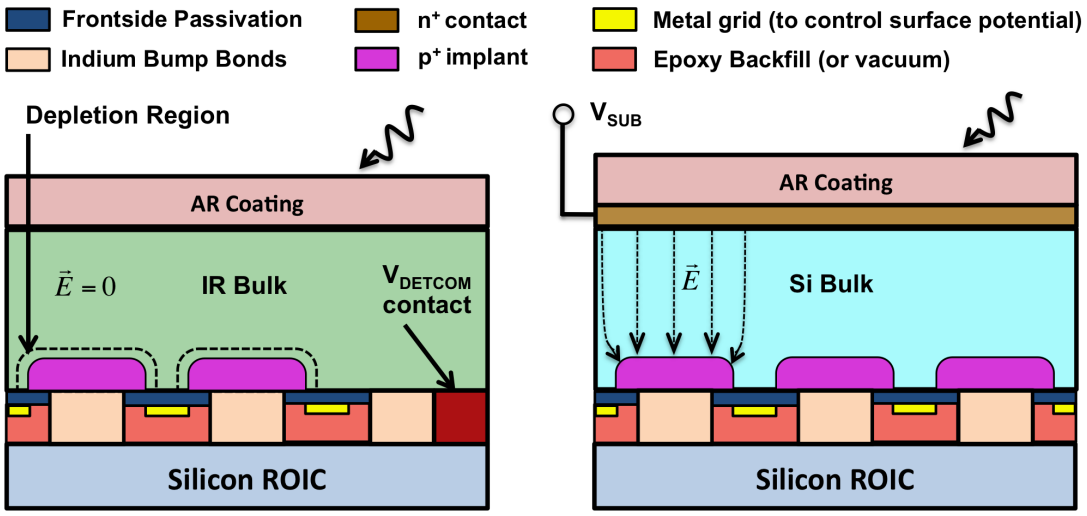


Figure 1.7: Diagrams of hybrid CMOS Arrays. AR stands for Anti-Reflection coating. (Left) A per-pixel depleted detector commonly used in infrared astronomy in which the depletion region (shown by the dashed line) only extends slightly above the pixel implant. The bulk is held at a constant voltage V_{DETCOM} 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 (∼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.

Chapter 2

Overview of Silicon PIN Dectectors

2.1 PIN Diodes

2.1.1 General Discussion of PIN Diodes

A PIN diode is very similar to a PN diode (a pn junction), except that an intrinsic layer[1], sometimes referred to as the bulk of the diode, is placed in between the p and n type materials. PIN diodes are more commonly used in photodetectors than PN diodes since the intrinsic region presents a larger volume in which photons can produce electron-hole pairs and so the thickness of this region can be adapted to increase quantum efficiency. The thickness of this region also gives them a lower capacitance than a typical PN diode. With a forward bias, the PIN diode behaves like a variable resistor for high-frequency signals. With a reverse bias, it acts as a parallel plate capacitor [53].

The most commonly used materials in PIN diodes are Silicon (Si) and Galium Arsenide (GaAs). Since the focus of this thesis is on Si PIN photodiodes, we will refer to Silicon as the relevant material unless otherwise noted. The bulk region of Si PIN detectors is ideally composed of a very high resistivity (ρ) intrinsic silicon. However, in practice this region is doped with slightly n-type, n^- silicon, or slightly p-type, p^- silicon. One usually calls the former a PνN diode and the latter a PπN diode. The main difference between the two is the sign of the space charge density in the bulk region. For the case of the PνN diode, the *ionized*[2] donors will yield a positive charge density while the PπN will have a negative charge density. It follows that the resultant electric field and potential in each will also be different.

A diagram of a PνN diode is shown in the upper portion of Figure 2.1. W is the total width of the intrinsic (or slightly doped) region and A is the cross-sectional area. W_P is the width of the

[1] Intrinsic silicon has the property that, in thermal equilibrium, the number of conduction band electrons per unit volume, n_c is equal to the number of valence band holes, p_v.

[2] The term *ionized* can be confusing for the case of semiconductors. When a donor atom loses its extra electron, it will contribute a net positive charge. When an acceptor atom loses a hole, it will contribute a net negative charge.

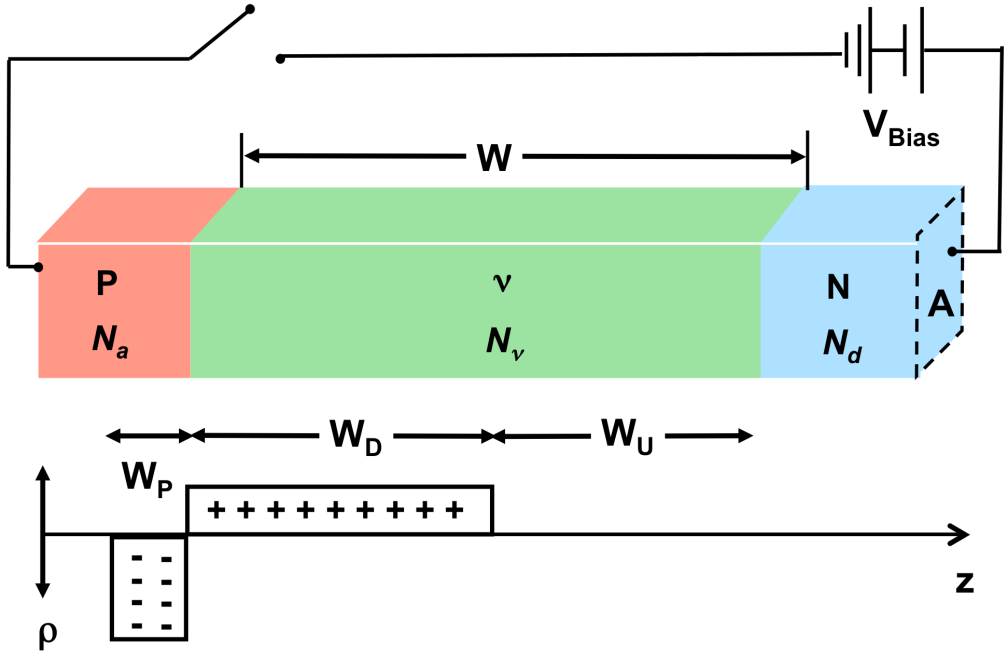


Figure 2.1: The top diagram shows the dimensions and composition of a PIN diode with donor densities of N_d in the N region and N_ν in the ν region, and an acceptor density of N_a in the P region. The cross sectional area is A and the width of the intrinsic region is W. The widths of the depleted and undepleted portions of the lightly doped n^- region are W_D and W_U, respectively. The width of the depletion region in the P material is W_P. Below is a plot of the space charge density due to the ionized donors/acceptors.

depletion region in the P contact, W_D is the width of the depletion region in the intrinsic layer, and $W_U = W - W_D$ is the width of the undepleted, or *diffusion* region in the intrinsic layer. Below the diagram is a rough sketch of the charge density for the case when the switch is open. The open switch is one of three main regimes under which the PIN diode can be operated, as follows.

1. **Open Voltage** : Switch is open

   - Charge carriers will diffuse until equilibrium is reached. In the dark, the voltage across the diode will take a nonzero value , V_{bi}, called the *built-in* voltage.
   - The field is zero near the contacts, so charge conservation demands that $W_P N_A = W_D N_\nu$. Thus, the depletion region extends further into the lightly-doped material than it does into the p material.
   - If illuminated, the voltage will change in proportion to the photon flux. This is known as the *photovoltaic effect*.

2. **Reverse Bias** : Switch is closed, $V_{Bias} > 0$

   - W_D will increase with increasing $|V_{Bias}|$.
   - When $|V_{Bias}|$ reaches a critical value, V_{PT}, called the the *punch-through* voltage, $W_D = W$, and the diode is said to be *fully depleted* of charge carriers. For voltages greater than this, the bulk will be *overdepleted*.
   - An electric field exists in the space charge region. This field will sweep carriers generated by thermal or photon excitation to the edges of the depletion region.
   - For $|V| < V_{PT}$, a diffusion region exists near the νN interface where there is no electric field. Charges created here will wander some typical distance D_B. Most will recombine; some will make it to the edges of the depletion region where they will be collected.
   - At thermal equilibrium, only a small reverse current will flow.

3. **Forward Bias** : Switch is closed, $V_{Bias} < 0$

   - W_D will shrink with increasing $|V_{Bias}|$.
   - At thermal equilibrium, a large current will flow across the diode as carriers are injected. This current is almost entirely due to diffusion of these carriers across the p^+-n junction; not drift. Holes from the p side diffuse to the n side and become minority carriers and electrons from the n side do the same thing on their way to the p side. The current is limited by the recombination of the two species when they become minority carriers on the opposite side.

In photodetection applications, PIN diodes are nearly always operated in reverse bias. However, the case of switching from a forward or zero bias to a reverse bias can cause non-equilibrium effects and difficulty in operating them as photodetectors and cannot be ignored completely. Both cases will be considered in detail in later sections.

2.1.2 PIN Diode Circuit Equivalent Model

One can model the PIN diode as a simple circuit consisting of capacitors and resistors [54]. The circuit equivalent at zero bias is shown in Figure 2.2. The undepleted portion of the intrinsic region is modeled as a capacitor, C_i, in parallel with a resistor, R_i. $C_i = \epsilon A/W_U$ is usually referred to as the diffusion capacitance and accounts for the charge stored in the undepleted portion of the bulk region. $R_i = \rho W_U/A$ is the resistance arising from the high resistivity silicon that has not been depleted in the bulk. In series with this combination, the depleted portion of the bulk contributes a capacitance $C_j = \epsilon A/W_D$ and the N and P contacts have a resistance R_C.

While the value of R_C remains nearly constant for varying values of V_{Bias}, C_J, C_i, and R_i are very dependent on it. As noted in the previous section, when a reverse bias is applied, W_D will

increase and W_U will decrease. Inspection of C_i shows that it will diverge as the bulk becomes fully depleted. However, at the same time the resistance R_i is going to zero. So when $W_D = W$, the parallel component of the circuit can be modeled as a short and the total capacitance of the circuit approaches a constant value of C_j. This is the value that is usually quoted for the capacitance of a photodetector since they are usually operated in full depletion.

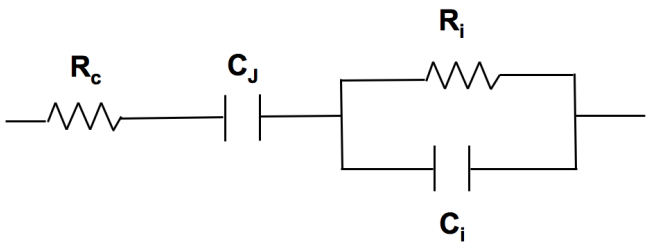


Figure 2.2: Circuit equivalent of a PIN diode. The capacitance, C_i, and resistance, R_i of the intrinsic (or lightly doped region) are in parallel and this portion of the circuit is in series with the junction capacitance, C_j and the contact resistance, R_c.

2.1.3 Punch Trough Voltage

If we apply a large enough reverse bias voltage V_{Bias} to the N side of the diode and fully deplete the bulk of charge carriers, there will be a non-zero electric field extending all the way across it due to the ionized donor atoms. The voltage at which this happens is referred to as the *punch-through* voltage. At a positive voltage less than this, there will be a *diffusion* region near the νN interface in which charge carriers experience no field.

The electric field in the fully depleted region can be approximated by neglecting any transverse fields (i.e. considering this a 1-d problem along the z direction) and applying Gauss's law,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon}, \tag{2.1}$$

where ρ is the charge density and ϵ is the relative permittivity of silicon. If we consider the charge distribution as arising purely from the ionized donors, then $\rho(z) = N_\nu(z)$ is a constant in the depletion region and zero in the diffusion region. Integration of Gauss's law then yields the electric field:

$$E(z) = \frac{zN_\nu q}{\epsilon}, \tag{2.2}$$

and integrating this field yields the electric potential (neglecting sign):

$$\phi(z) = \frac{z^2 N_\nu q}{2\epsilon}. \tag{2.3}$$

Assuming that the punch-through voltage has been reached and the bulk has been fully depleted so

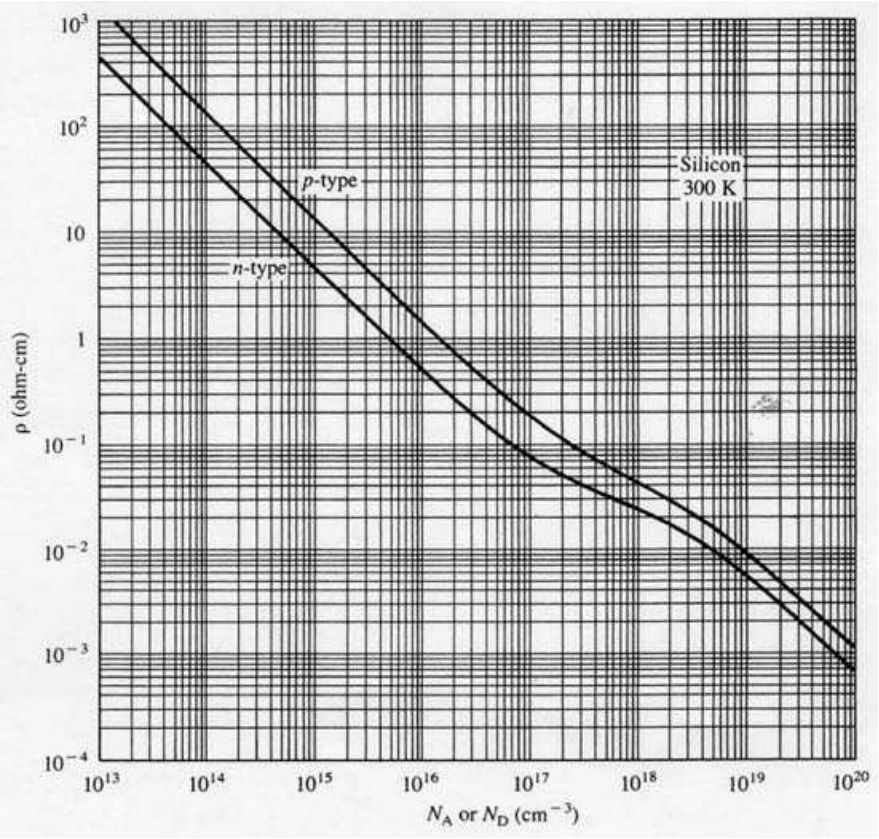


Figure 2.3: A plot taken from [55] showing the resistivity of silicon as a function of doping density.

that $z = W$, we have,

$$V_{PT} = \frac{W^2 N_\nu q}{2\epsilon}. \tag{2.4}$$

Equation 2.4 allows us to calculate the punch-through voltage and reveals the $\sqrt{V}$ dependence of the depletion region width on the reverse bias. However, it does not describe the shape or magnitude of the field in regions where the donors have not been ionized or at the PI and IN interfaces. The latter regions are important as the fields here may dominate charge transport.

Furthermore, this calculation neglects the P and N regions entirely, considering them essentially as metal electrodes. As shown by Figure 2.3, the resistivity of silicon declines to about 10^{-3} Ω·cm at very high doping densities. However, this is still 3 orders of magnitude higher than the resistivity of typical conductors such as copper or silver. The voltage necessary to deplete these regions of their carriers may not be negligible in certain cases.

To describe the fields in the undepleted region and understand the distribution and dynamics of the charge carriers near the interface, it is necessary to solve the equations numerically.

2.2 HyViSI Detector

The detectors studied in this thesis are Hybrid Visible Silicon (HyViSI) CMOS arrays. They consist of two layers: an array of Silicon PIN didoes that serve as the photodetectors and a source follower per detector (SFD) CMOS HxRG multiplexer (where the "x" stands for 1, 2, or 4, depending on the format) that acts as the readout circuit. The two layers are manufactured separately, allowing for independent optimization of photocollection and readout. While many aspects of the optimizations will be covered, the reader is referred to Bai et al. [37, 56, 25] for the exhaustive details.

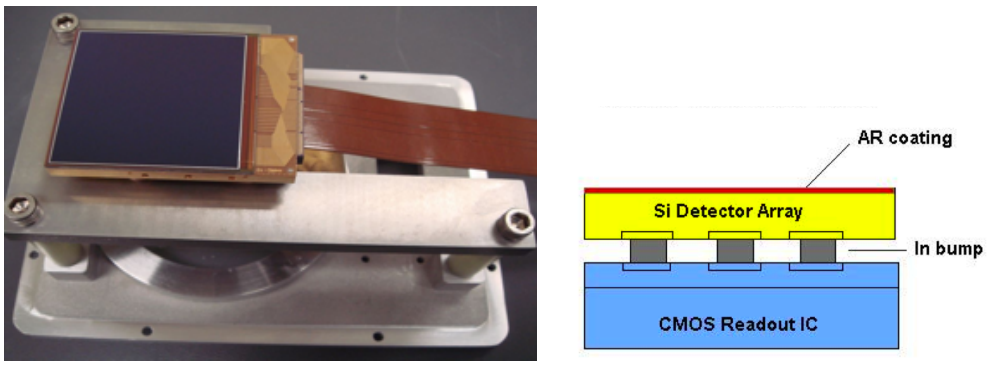


Figure 2.4: (Left) Picture of H4RG HyViSI detector. (Right) Cross section view of HyViSI detectors taken from [37].

Once fabricated, the two layers are precisely aligned so that the pixels of the detector lie directly on top of the pixels in the MUX, as shown in the right of Figure 2.4. They are then "cold welded" together with a pressure of up to several hundred kilograms [23]. After the cold welding, the two layers will be effectively "glued" together by the indium bumps, which also serve as the conductive path between a pixel in the photodector and the mux. As one might imagine, this process is very difficult. Many devices suffer from a large amount of "broken" pixels (the word "broken" will be clarified in Section 4.2). This low yield factor is one of the great challenges that must be met before SiPIN detector arrays become a viable alternative to more conventional imagers. For most devices tested in this thesis work—all 1k×1k H1RGs and 2k×2k H2RGs, but not the 4k×4k H4RG—over 99.9% of the pixels were functional and behave effectively as PIN diodes.

2.2.1 HyViSI PIN Diode

The diodes in the HyViSI detectors possess a bulk region with a width of $W = 100$ μm that is slightly n-type. Thus, when the bulk is depleted the ionized donors in it give rise to a positive space charge density. This slightly doped intrinsic region is sandwiched between highly doped p^+ and n^+ regions, as shown in Figure 2.1. For the purpose of this section, the cross sectional area is assumed

to be that of an entire 10 μm pixel, i.e. $10\mu\text{m} \times 10\mu\text{m} = 100\mu\text{m}^2$. However, the p^+ region does not actually occupy the entire pixel and this fact will be taken into account in later analyses of the detector.

On the n^+ side a substrate voltage called V_{SUB} is supplied with either a power supply or battery. At the p^+ site the voltage is V_{Node}: the same voltage seen at the gate of the unit cell source follower in the multiplexer, the two being coupled by the indium bump bond between them. V_{Node} can be held at V_{RESET} by holding the reset switch or allowed to rise as photocharges swept out of the bulk integrate. If $V_{Node} > V_{SUB}$, the diode is forward biased and if $V_{Node} < V_{SUB}$, it is reverse biased.

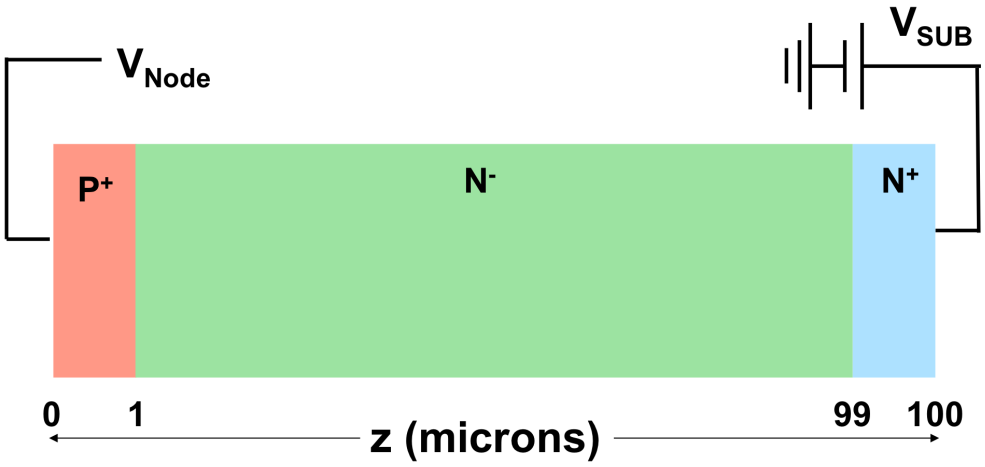


Figure 2.5: Diagram showing the dimensions and composition of the PIN diodes in the HyViSI detector (not to scale). For simplicity, the entire diode is assumed to be 100 μm long. The bulk is slightly n-type material. At the p^+ side, the node voltage V_{node} integrates as holes are collected. The reverse bias is held by V_{SUB} on the n^+ side with a power supply or battery.

In reverse bias, with a thickness of 100 μm and 11.8 as the dielectric constant of silicon, Equation 2.4 yields a punch through voltage of 7.7 Volts. The PIN diodes in the HyViSI devices are typically operated at voltages higher than this to ensure that the bulk is overdepleted. This ensures that the photo-generated charges are swept out of the bulk before recombining, and the higher the value of V_{SUB} is, the less lateral diffusion of charge there will be. However, values of V_{SUB} between zero and V_{PT} are useful in determining characteristics of the detector such as lifetime and diffusion lengths of the charge carriers.

2.2.1.1 Simulation Results

Using the method outlined in Appendix C and guesses for the doping densities in the HyViSI PIN diodes, we have calculated the electric field, carrier densities, and charge densities, for various values

of V_{SUB}. From these quantities we can easily obtain the potential and energy-band diagrams and investigate the band bending near the surfaces, a matter which will be handled in a later section. As a first approximation, the voltage at the integrating node, V_{Node}, is held at ground and the PI and IN interfaces are assumed to be diffused layers with Gaussian-type distributions, giving a doping profile similar to the one modeled in Kurata [57]. The first of these approximations will be refined in later sections to account for the voltage increase due to integration of photo-generated charge at the p^+ nodes.

The simulation is performed with a nonuniform mesh consisting of 4000 points, with closely spaced grid points near the junctions ($\Delta x_i = 1$ nm) and coarsely spaced ones (a maximum of $\Delta x_i \approx 0.26$ μm) in the bulk. The results at steady-state, when the generation and recombination terms are equal at each grid point, are shown in Figure 2.6 for the case where the bulk is not fully depleted and in Figure 2.7 for the case in which the bulk is depleted or overdepleted. There are several interesting features to note in each of these cases.

0-10 Volts: The electric field plots in the top of Figure 2.6 show good agreement with the result of Equation 2.4. For voltages less than $V_{PT} \sim 8$ volts, we see that a region with zero electric field exists in the bulk silicon. In these regions there is a substantial non-zero charge carrier density. We expect holes generated in this region to recombine before making it to the boundary where the amplitude of the electric field starts to increase. These diffusion regions are undesired in photo-detection operation since many of the electron-hole pairs generated by impinging photons will recombine and will not contribute to the signal.

Another interesting feature is the large electric field amplitude at the Pν and νN interfaces. This is to be expected from the large mismatches in doping concentrations at these boundaries and the fact that we demand the electric field vanish at the contacts. In reality, the electric field may have nonzero values due to the fact that the p+ and n+ contacts are not perfect conductors and are in contact with regions that may be hosts to numerous charge traps. Also, in normal operation the p+ region will actually be accumulating minority carrier holes from the bulk, so we expect to modify the boundary condition at $z = 0$ to account for this.

Lastly, at 10 volts we see that a non-zero electric field exists throughout the diode. This indicates that the punch-through voltage has been reached. However, near the n+ we see that the electron carrier density still has a large non-zero value of about 10^{12} cm-3. This indicates that the diode has not reached an overdepleted state; a further increase in V_{SUB} is necessary to reach this regime.

15-35 Volts: After the punch-through voltage has been reached, the diode is said to be in an overdepleted state. Increasing V_{SUB} further reduces the carrier and charge densities near the n+ν boundary. It also increases the magnitude of the electric field throughout the diode. The latter is especially important for increasing CCE and reducing crosstalk.

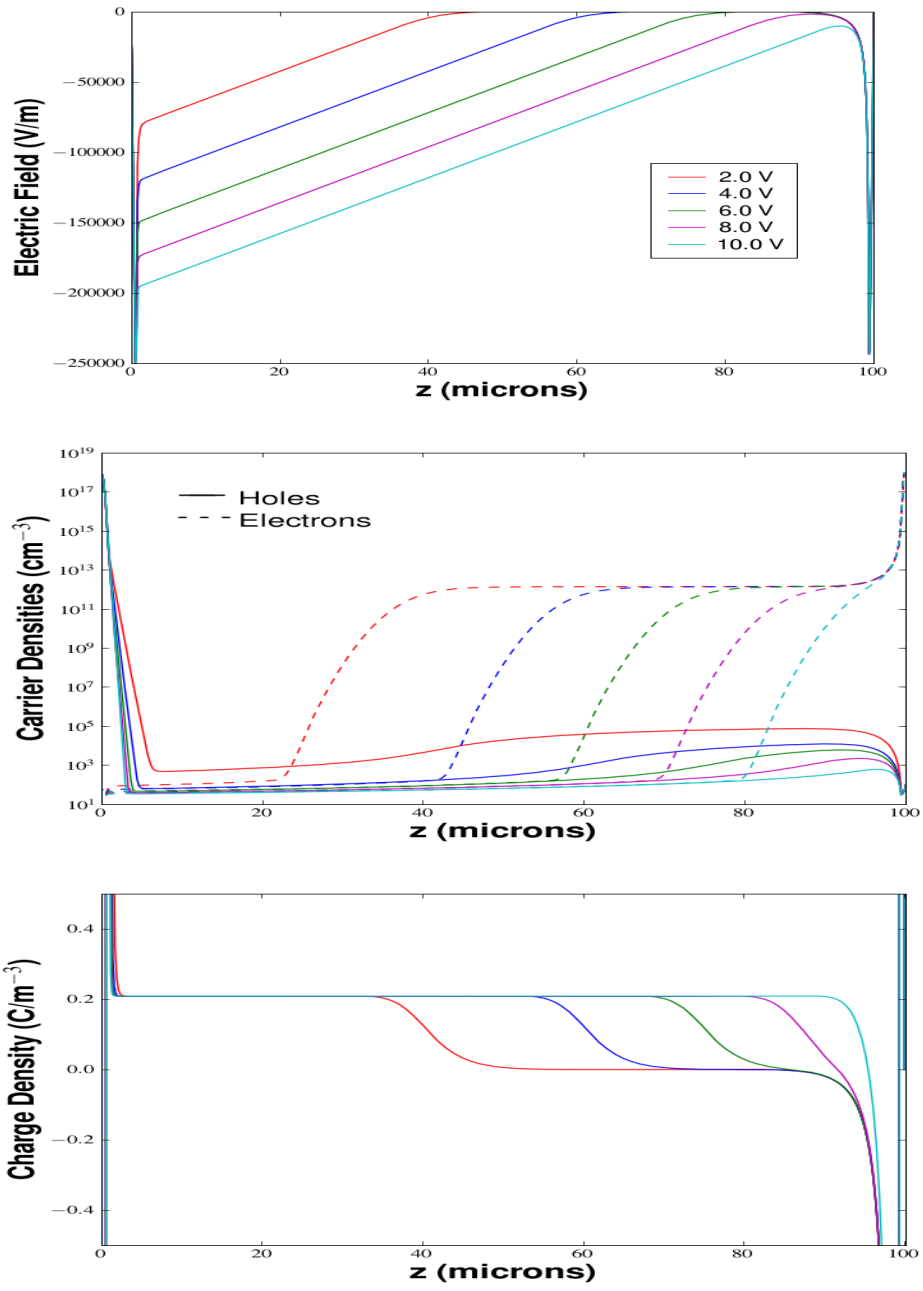


Figure 2.6: Simulated electric field (top), carrier densities (middle), and charge density (bottom) in the HyViSI PIN diodes. For most voltages shown the bulk is not fully depleted. All quantities were obtained through the numerical simulations described in Appendix C.

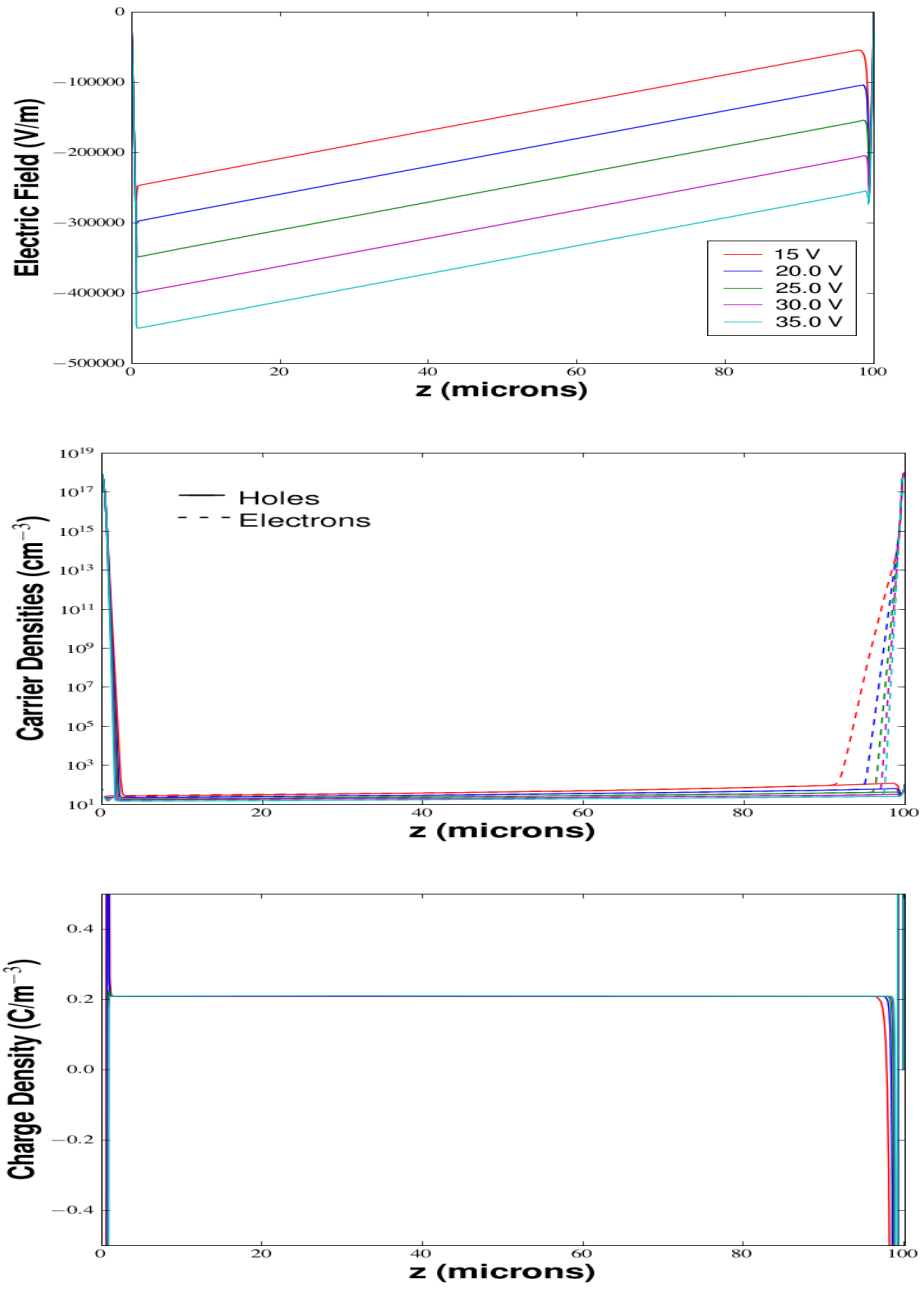


Figure 2.7: Simulated electric field (top), carrier densities (middle), and charge density (bottom) in the HyViSI PIN diodes. For most voltages shown the bulk is overdepleted. All quantities were obtained through the numerical simulations described in Appendix C.

2.2.2 The HyViSI Pixel

As seen in Figure 2.4, a full HyViSI pixel consists of **1)** the PIN diode in the detector layer and **2)** the FETs and silicon layer in the ROIC, as well as the indium bump bond and SiO_2 layer that separate the two. In presenting the aspects of the HyViSI pixel, we will first consider the capacitances associated with its various components and then study how signal is generated and eliminated.

2.2.2.1 Pixel Capacitance

To this point we have only considered the PIN diode as a two-terminal device with its ends held at fixed potentials. However, in the bulk of the HyViSI detectors, the p^+ side is actually in contact with an indium bump that connects to the input of a readout node. As mentioned earlier, this node can be the pixel of a CCD or CTIA, DI, or SFD CMOS array. In our case, the readout node is the gate of a source follower in the unit cell of a source follower per detector (SFD) CMOS array. A mock schematic of the SFD pixel is shown in Figure 2.8.

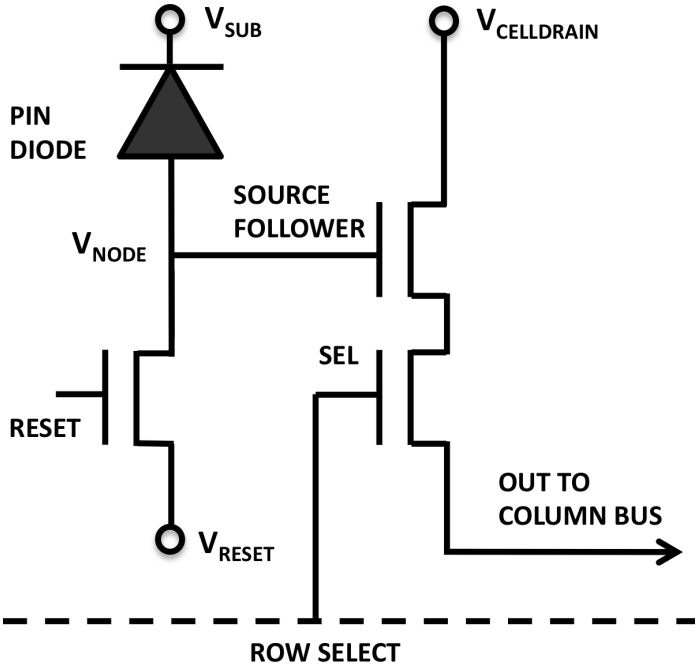


Figure 2.8: A mock schematic of a full HyViSI pixel showing the PIN photodiode in the Silicon detector layer and transistors in the ROIC. The voltages V_{SUB} and V_{RESET} are used to bias the photodiode. $V_{CELLDRAIN}$ is the drain voltage of the source follower. V_{NODE} is the effective signal that is measured through the amplification stages present in the pixel source follower and the output source follower, if it used. The select transistor allows the pixel to be addressed so that its output is placed on the column bus and eventually fed to the detector output.

The operation of this configuration is quite simple if we model the PIN diode in the bulk as a parallel plate capacitor. There are many problems with this simple model, especially in the case where the bulk region is not depleted. But it will be presented here, along with some refinements, as a starting ground for understanding how the pixels are reset and how signal is generated in the detector. The basic sequence with this model assumed is shown in Figure 2.9.

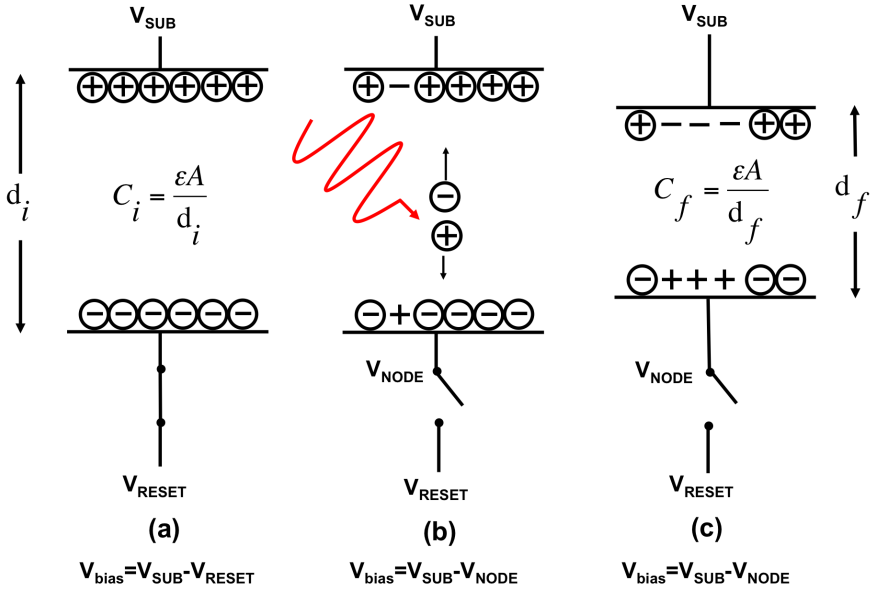


Figure 2.9: A HyViSI PIN diode pixel modeled as a capacitor. In (a) the reset switch is closed and the capacitor charges. In (b) the switch is closed and the pixel is allowed to integrate photocurrent. As V_{bias} shrinks (c), the plates of the capacitor get closer together and the capacitance grows.

Before starting an exposure, the reset switch in the multiplexer unit cell is closed, as shown in (a) of Figure 2.9. This allows the capacitor to fill up with negative (positive) charge on the bottom (top) and brings V_{Bias} to a maximum. To make the model a little more tangible we can envision electrons entering from the source of the reset transistor and filling the holes in the p+ implant and electrons exiting the n+ node through the upper contact, leaving vacant holes there. The capacitance in this case is determined purely by the geometry:

$$C = \frac{\epsilon A}{d_i}, \tag{2.5}$$

where A is the area of the pixel, ϵ is the dielectric constant of silicon, and d_i is the initial distance between the plates. We assume d_i to be the full 100 μm thickness of the detector substrate at this stage. The total charge stored on the capacitor is thus $Q_i = C_i V_{Bias}$. Ideally, we should have

$$V_{Bias} = V_{SUB} - V_{RESET} \tag{2.6}$$

However, in practice each pixel has its own small voltage offset. These offsets are of no great consequence, though, since they can be removed by subtracting a bias frame or subtracting successive reads of the detector.

After we are done resetting, the switch is opened as in (b) of Figure 2.9. Any minority carrier holes

generated, either thermally or through photo-excitation, in the bulk n material will be accelerated downward by the electric field and swept towards the p^+ implant. Once at the p^+ implant they become majority carriers. Since the probability for them to recombine with electrons in the implant is very low, they will collect there at a rate that is proportional to the number of photons impinging on the bulk. There will also be holes collected there as a result of leakage currents at the surface as well as generation-recombination (G-R) and diffusion currents in the bulk. As the holes accumulate the capacitor discharges. V_{NODE} rises and V_{Bias} decreases. The voltage V_{NODE} is what we attempt to measure to determine our signal.

As more and more carriers are collected, the depletion width of the diode will effectively decrease. In our capacitor model, this corresponds to a decrease in the distance between plates, as in (c) of Figure 2.9, and in turn, an increased capacitance. This changing capacitance can lead to non-linearity in the response of hybrid detectors, as shown by the following equation:

$$dQ = CdV + VdC. \tag{2.7}$$

In a real (ideal) parallel plate capacitor the second term would be zero since $dC = 0$. But for the PIN Diode, the capacitance changes as

$$dC = -\frac{\epsilon A}{\mathrm{d}^2} d\mathrm{d}, \tag{2.8}$$

or

$$\frac{dC}{C} = -\frac{d\mathrm{d}}{\mathrm{d}}. \tag{2.9}$$

For many hybrid CMOS detectors that use simple pn junctions as photodiodes, dC/C can be quite large. However, for the HyViSI PIN diodes this change is quite small.

The reason why dC/C is small has to do with the fact that the total capacitance of a pixel is actually the sum of several different capacitances in the detector and ROIC:

$$C_{TOT} = C_{PIN} + C_{SF} + C_{Stray} + C_{IP}. \tag{2.10}$$

where

- C_{PIN} is the depletion capacitance of the silicon PIN diode, which we estimate with Equation 2.5.

- C_{SF} is the capacitance of the source follower transistor in the multiplexer to ground.

- C_{IP} is the interpixel capacitance.

- C_{Stray} accounts for stray capacitances in the detector or ROIC.

For HyViSI PIN diodes, we find that $C_{PIN} \sim$ 0.02-0.35 fF, depending on how much of the 10 or 18 μm pixel might be occupied by the p^+ implant, while measurements show that $C_{TOT} \sim 14$ fF.

2.2.2.2 Photocurrent and Signal Generation

As the name implies, the purpose of a photodetector is to detect photons. The HyViSI accomplishes this by turning electron-hole pairs, which are produced at a rate G_e in the silicon layer, into a photocurrent, I_{PHOTO}. As shown in Figure 2.10, I_{PHOTO} is accompanied by two unwanted currents: the dark current, I_{DARK}, and persistence current, $I_{PERSIST}$. All three of these currents integrate on the capacitance C_{TOT}. The latter two will be covered in detail in later chapters; here we are only concerned with I_{PHOTO}.

We begin by assuming there is some flux of photons, Φ_o, incident on the n^+ side of the detector (originating from the right in Figure 2.5). For simplicity we will assume the photons are monochromatic with wavelength λ, and that they all have the same absorption depth, $\alpha(\lambda)$. Assuming Φ_o accounts for any photons reflected at the surface and the quantum efficiency is $\eta(\lambda)$, the rate of electron-hole production at a given depth in the detector is:

$$G_e(\mathrm{z}) = \Phi_o \alpha \eta \exp(-\alpha \mathrm{z}), \tag{2.11}$$

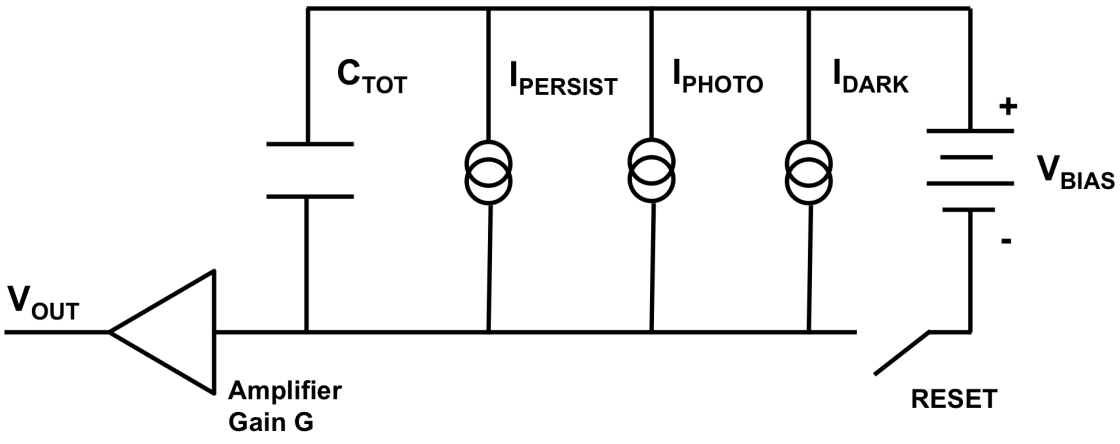


Figure 2.10: A circuit equivalent for the HyViSI pixel with source follower readout (following Figure 8.11 in McCaughrean [41]). The capacitance C_{TOT} is the one listed in Equation 2.10. I_{PHOTO} is the photocurrent, I_{DARK} is the dark current, and $I_{PERSIST}$ is current generated from persistent charge. V_{BIAS} is the bias applied to the diode at reset and V_{OUT} is the voltage sensed through the amplification, G, that results from the source followers in the signal path.

where we have taken the back surface as z=0 and the direction moving into the diode as the positive z direction. Gärtner uses this expression to solve for the total current density, J_{tot}, through a one-dimensional PIN diode [58]. However, he makes the crucial assumption that the electric field through the diode is constant. It is clear from Figures 2.6 and 2.7 that the electric field in the HyViSI PIN diodes is not constant, which renders the analysis invalid for them.

It is especially important to consider cases where the diodes are not fully depleted since there will be regions completely free of electric field, and carriers generated in these regions will contribute to a lateral current in the x, y directions. For bulk regions where there is a nonzero photon flux and electric field, the drift current is given by

$$J_{drft} = -q \int_0^{W_D} \left[G_e(\mathrm{z}) + \Delta p_{diff}(\mathrm{z})\right] E(\mathrm{z}) \mu_p d\mathrm{z}, \tag{2.12}$$

where μ_p is the mobility of holes, $E(\mathrm{z})$ is the electric field along the length of the diode, and W_D is the width of the depletion region. Δp_{diff} represents any of the free holes that have diffused into the pixel field, either from a photogeneration site directly above the depletion region or from the field free region of a neighboring pixel.

The diffusion of holes from an illuminated pixel to its neighbors can happen quite easily when the detector is not overdepleted since the accumulation of photocurrent actually causes the depletion region in the illuminated pixel to collapse. When holes drift to the front surface and recombine with the ionized acceptors, W_P in Figure 2.1 shrinks; the same happens for W_D when electrons drift to the back surface. This leads to an increase in the undepleted width ΔW_U^{lum} for the illuminated pixel. In the neighboring pixels, no such reduction of W_D has occurred. Hence, any of the diffusing carriers in the newly formed ΔW_U^{lum} that migrate laterally will find themselves directly in the depletion region of a neighboring pixel and contribute to its drift current through the term Δp_{diff}. Eventually, the depletion regions of the neighboring pixels will collapse as well, allowing holes to diffuse and be collected by pixels beyond. An example of this, which is known as "blooming" in imager speak, is shown in Figure 2.11.

The holes which are collected by the electric field accumulate as majority carriers in the p^+ implant and are stored as signal charge, $Q(t)$:

$$Q(t) = -qA \int_0^t \left(\int_0^{W_D} \left[\Phi_o \alpha \eta \exp(-\alpha \mathrm{z}) + \Delta p_{diff}(\mathrm{z})\right] E(\mathrm{z}) \mu_p d\mathrm{z} \right) d\mathrm{t}. \tag{2.13}$$

The integration of Q does not go on indefinitely; this equation is only valid until the potential well in the p^+ region has filled up to the full well, $Q = Q_{FW}$ (see Section 6.1.4.3 for a discussion of diffusion at the front surface). Because of the field free region, there will also be a large number of excess minority carriers in the bulk. The number of minority carrier holes present in a given pixel will depend on the number generated by incident photons, Δp_{lum}, the net difference between holes diffusing in

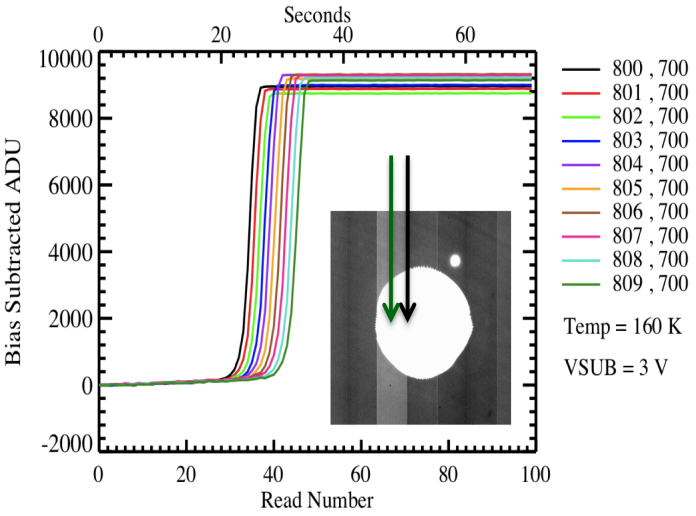


Figure 2.11: This plot shows the signal vs. time (after subtraction of bias offset) for a set of pixels that have received holes from strong lateral diffusion in the undepleted region of H1RG-022. At first they integrate only a small flux from the sky, Φ_o^{sky}. Once the depletion regions in their right neighbor pixels have collapsed they see a huge jump in signal due to the hole diffusion, Δp_{diff}. The raw image from which these ramps were taken is shown in the inset.

and diffusing out, Δp_{diff}^{net}, and their lifetime, τ_p, as they recombine. Unfortunately, as seen in Figure 2.11, the output of the detector is railed by the 3.3 V upper limit of the CMOS multiplexer, making it difficult to fold in the lifetime of the carriers.[3] The exact doping and geometry in the detector material is not known either, which presents another difficulty. But it is still worthwhile to make a semi-quantitative analysis of the expected distribution of holes in the underdepleted state. The sum of collected holes and free minority holes should get smaller with distance away from the center of illumination, and we can designate four separate radial regions based upon the constituent sources of holes:

$$p = \begin{cases} Q_{FW} + \Delta p_{lum} - \Delta p_{diff}^{net} & 0 < r < r_1 : \quad \text{Saturated by Illumination} \\ Q_{FW} + \Delta p_{diff}^{net} & r_1 < r < r_2 : \quad \text{Saturated by Diffusion} \\ Q(t) + \Delta p_{diff}^{net} & r_2 < r < r_3 : \quad \text{Integrating Diffused Holes} \\ 0 & r_3 < r \quad : \quad \text{Outside Diffusion Envelope} \end{cases} \tag{2.14}$$

If we are considering point sources of light, r_1 should be about the 1.5-3 $\times$ the FWHM of the star. Outside of r_1, the pixel wells fill up purely because of the holes that diffuse to their depletion regions. At some radius r_2, the wells have just begun to collect diffused holes (one can see r_2 moves outward with time in Figure 2.11). And outside of a radius r_3, the diffusion has not yet reached the pixels, so they are integrating only dark current. It should be noted that **these regions are only relevant for the underdepleted case** (when overdepleted, there is very little measurable diffusion). Still, they will be extremely important in analyzing image persistence in Section 7.2.

[3] In a detector where the output did not rail at some voltage, V_{rail}, we would be able to see a gradient in the signal vs. position. This gradient would allow us to solve the diffusion equation for the diffusion coefficient, D_p, and lifetime, τ_p. But because $V_{rail} = 3.3V$, we are only sensitive to the small range of signal and the gradient is masked.

Chapter 3

Hybrid Imager Features

In contrast with CCDs, the nondestructive nature of the HxRG readout makes it possible to do *Up-The-Ramp sampling* (UTR) of the detector, and the multiplexer in the ROIC makes it possible to randomly select and read pixels. With these features we can define at least three distinct ways of operating the detector.

The first, and most conventional way of using the detector, is taking *full frame* exposures – ones in which we clock through every pixel in the detector one or more times. The second is taking *window mode* exposures – ones in which we clock through only the pixels in a sub-region of the detector one or more times. And the third is *guide mode* exposures – ones in which we alternate between full frame reads for the purpose of collecting science data and window mode reads that allow us to reset or read a small window on the detector. The full-frame and guide modes are well suited for long exposures of faint objects and the *window mode* is more appropriate for high speed photometric measurements or telescope guiding. Each will be discussed in turn.

3.1 Full Frame Mode

A diagram that illustrates a *full frame* Up-the-Ramp (UTR) sequence is shown in Figure 3.1. The *ramp*, synonymous with the *exposure*, consists of a series of N *frames*, the basic unit in which the entire array of pixels is clocked through. The clocking scheme is similar to the one employed in CCDs, with a fast direction along the rows controlled by a horizontal clock (HCLK) and a slow direction along the columns controlled by a vertical clock (VCLK). The duration t_f of the *frame* is determined by the pixel time t_p—which corresponds to the time between edges of the HCLK signal and can be as low as 2 microseconds for the native *slow mode* of the HxRG detectors—and the number of outputs N_o used on the multiplexer. Following the nomenclature used for astronomical images, if NAXIS1 is the number of columns and NAXIS2 is the number of rows in the frame, then $t_f = t_p * \text{NAXIS1} * \text{NAXIS2}/N_o + \text{NAXIS2} * t_{rs}$, where t_{rs} is a time associated with moving to the next

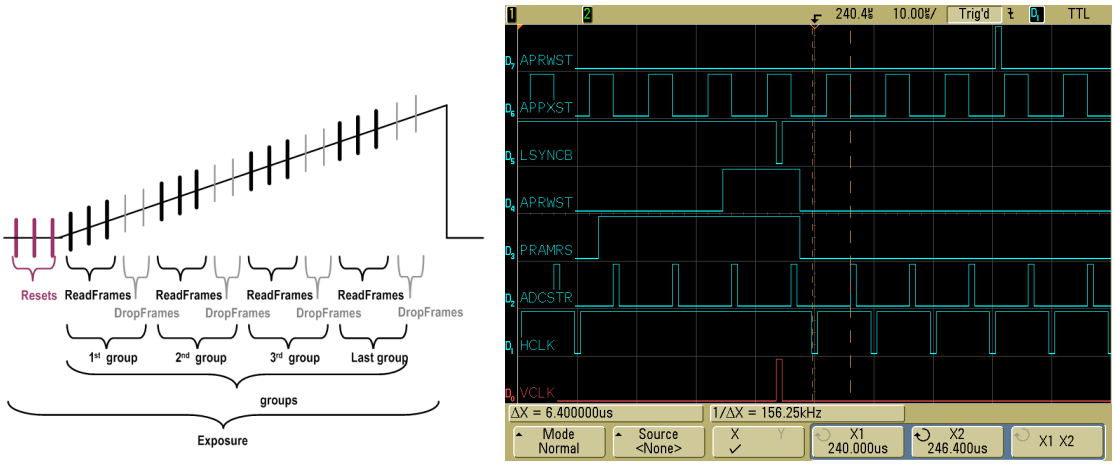


Figure 3.1: (Left) An Illustration of the ramp sequence, courtesy of Markus Loose. The number of resets, drops, groups, and reads are programmable in software, allowing control over the exposure time. (Right) An oscilloscope trace of several clock lines on the SIDECAR ASIC. The delay time t_{rs} is visible.

row of NAXIS2/N_0 pixels in the array. t_f is also the time between reading a pixel in one *frame* and reading that same pixel again in the following *frame*. As a representative example, the SIDECAR ASIC we used had a maximum of 36 inputs, so we chose $N_o = 32$ when using the H4RG HyViSI. We used a 10 μs pixel time, giving $t_f = 5.24288s + 4096 * t_{rs}$. Eventually, we implemented a counter on the SIDECAR that allowed us to precisely measure a time of $t_f = 5.453$ s and $t_{rs} = 52$ μs. This was also verified on an oscilloscope, the trace of which is shown in Figure 3.1. It is worth noting that t_{rs} was limited by the assembly code used to operate the SIDECAR and not the multiplexer itself. The time for one VCLK and a slight delay before and after to allow the output voltage to settle will place a smaller lower limit on t_{rs}, but this was not tested.

The illustration in Figure 3.1 shows three different types of *frames*: *reset frames*, *drop frames* and *read frames*. During the *reset frames*, the array is clocked so that the gate of the reset FET in the ROIC pixel is enabled and the pixel is held at the voltage V_{RESET} supplied by the control electronics. This allows the charge on the capacitive node in the silicon to dissipate so that it is ready for a subsequent integration. In the *read frames*, the column select and row enable signals for all of the pixels are enabled in turn, bringing the voltage at a pixel to its corresponding output. This voltage is converted to a digital number and sent out to the Data Acquisition System (DAQ) for recording. The unique feature of the *drop frames* is that no data is output from the control electronics to the computer. This allows the user to take a very long exposure without dealing with overwhelming amounts of data. It should be mentioned that data from the *reset frames* can be output to the external DAQ and the pixels can be clocked during the *drop frames*. In fact, as will be discussed later, clocking the pixels during *drop frames* is absolutely essential if one wishes to avoid

temporal drifts and instabilities in the pixel voltages.

3.1.1 Up the Ramp Terminology

In a fashion similar to the scheme used in the NICMOS and JWST specifications [59], we refer to a ramp having a **cadence** of N_{rs} *reset frames*, N_d *drop frames*, N_g *group frames*, and N_{rd} *read frames*, as a N_{rs}-N_{rd}-N_d-N_g ramp. Such a ramp has a total of $N_g \times N_{rd}$ data frames. With these definitions, the exposure time of a ramp in *full frame* is given by

$$t_e = N_g * (N_d + N_{rd}) * t_f, \tag{3.1}$$

with t_f being the frame time given above. To avoid confusion in the following sections, we will use this syntax and reserve the terms *ramp* or *exposure* for such a sequence. We will use the terms *read* or *frame* to describe a *read frame* in these ramps. And the term *image* will be used to refer to the data produced after reduction of these *ramps*, which is described Section 5.2.

3.2 Window Mode

The HxRGs contains a serial register that can be configured to allow random access to a contiguous $X \times Y$ subset of pixels, or window, through one output of the detector. X and Y, the number of columns and rows in the window, respectively, must be greater than 1 pixel and less than NAXIS1 pixels. Programming the addresses of the window limits, X_{start}, X_{stop}, Y_{start}, and Y_{stop}, can be done on the fly. This allows multiple windowed regions of the detector to be read out in a ping-pong like fashion and also the possibility of feedback control loops in which the window coordinates are continually adjusted to track an object. Other applications of window mode, such as telescope focusing, are presented in Simms et al. [39] and more will be discussed in Chapter 5.

Our basic unit in a *window mode* observing sequence is a correlated double sample (CDS) in which we clocked through all $X * Y$ pixels in a window 3 times, resetting each pixel once and then reading the pixel voltages twice.[1] The clocking is similar to full frame mode, with the fast clocking along the rows and the slow clock along the columns. With a pixel time of t_p, one row takes $X * t_p$ seconds with an additional t_{rs} seconds of overhead in shifting to the next row. To complete all rows thus requires $t_w = Y * (X * t_p + t_{rs})$ seconds and this is the total time for a reset or a read in the window sequence. An effective exposure time for a CDS in *window mode* is given by

$$t_e = 2t_w + 2t_{rs}, \tag{3.2}$$

[1] Note that there is nothing preventing one from obtaining UTR data in window mode. UTR windows were used for a variety of purposes as well, but the CDS sequence is more relevant for applications like telescope guiding where the pixels are *shot-noise* limited (see Section 4.4.2.2).

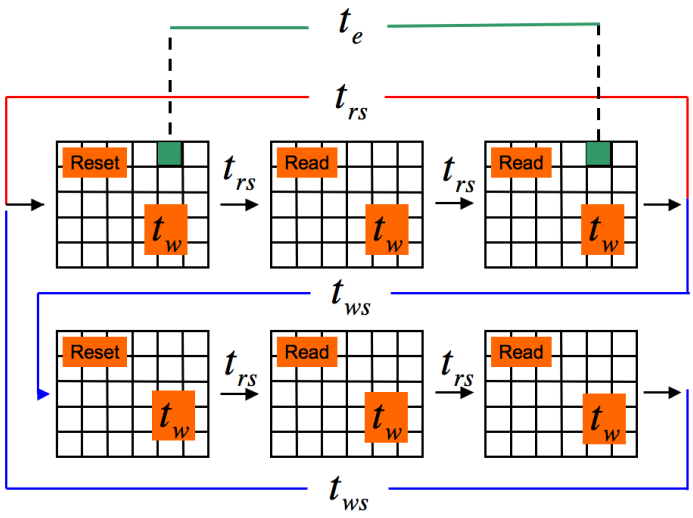


Figure 3.2: An illustration showing the sequencing and times (described in text) involved with *window mode*. The red line at top shows the path taken in *single window* mode and the blue line shows that taken in *multiple window* mode when $N = 2$ windows are used.

which is the time between the reset of a pixel and sampling it in the second read. For our observations, we used a pixel time of $t_p = 10$ μs and minimized the overhead to attain $t_{rs} = 18$ μs. t_p can be be decreased in order to decrease the time to complete one CDS sequence and up the sampling rate with a penalty in noise.

After a CDS sequence, either another CDS is repeated on the same window or the serial register is programmed with a new set of window coordinates. We refer to the former as *single window* mode and the latter as *multiple window* mode. With the microcode used on the SIDECAR ASIC, the operation of writing new window coordinates for X_{start}, X_{stop}, Y_{start}, and Y_{stop} takes $t_{ws} = 150$ μs; adjusting only two of these four would take ~ 75 μs. A diagram showing the times associated with window mode is shown in Figure 3.2. It should be apparent from the diagram that if N windows are used in *multiple window* mode, the time it takes to complete a full cycle and return to the first window is given by:

$$t_c^{mw} = 3Nt_w + 2Nt_{rs} + Nt_{ws} \tag{3.3}$$

And if *single window* mode is used this time becomes:

$$t_c^{sw} = 3(t_w + t_{rs}) \tag{3.4}$$

It is also important to know the time between exposures in different windows for the purposes of measuring temporal correlations. In *multiple window* mode, the time it takes between a CDS in

window n and a CDS in window m is given by

$$t_{nw} = 3(m-n)t_w + 2(m-n)t_{rs} + (m-n)t_{ws} \tag{3.5}$$

The full cycle is repeated M times in one full observing sequence, giving a total time of $M * t_c^{mw}$ in *multiple window* mode and $M * t_c^{sw}$ in *single window* mode.

3.2.1 Variation of Multiple Window Readout Sequences

In the simple implementation just described, the flux measurements of individual windows are made serially. A given window is reset and a series of reads is performed before moving onto the next window and repeating. We refer to this method as a **Staggered Reset-Read (SRR)**.[2] Another type of sequencing is to reset each window in sequence and then read them in sequence a number of times afterwards so that the exposure time is given by an expression similar to Equation 3.3 instead of Equation 3.2. This method we call **Staggered Reset-Staggered Read (SRSR)**. The SRR sequence yields a serial set of flux measurements in time, but offers the benefit that very bright stars are read immediately after reset to prevent saturation of the pixels. The SRSR sequence offers a more parallel set of flux measurements in time–the windows are integrating the same wavefront for some portion of the sequence–and is useful for dim stars where the signal to noise can be improved with the longer integration. There are, of course, other possibilities, one of which is interleaving the SRR and SRSR sequences. For our measurements, only these two were implemented.

3.3 Guide Mode

The HxRG multiplexers can be operated in a special "guide mode" in which the *full frame mode* and *window mode* sequences are interleaved. In this mode, a subwindow of the array that contains a guide star or set of saturated pixels can be continuously read or reset while the rest of pixels in the array integrate charge. If the former is the case, the centroid of the guide star can be calculated in the control electronics or DAQ for values of Δx and Δy that can be fed to a Telescope Control System (TCS) in order to make adjustments in the telescope pointing. At any time during the integration the full array can be read as well, allowing for UTR sampling. Essentially, the guide mode is just an alternation between full frame and window mode. Further discussion of guide mode will be saved for Section 5.5.

[2]The name **Staggered Reset-Read** implies only one read of the window after it is reset. However, many reads of the window can be performed before moving to the next window. In fact, at least two should be taken to allow for a correlated double sample. Using one **read** is only for nomenclatural convenience.

3.4 Reference Pixels

The H1RG, H2RG, and H4RG all contain a ring of reference pixels that surround the science pixels. There are 4 columns of reference pixels on the left and right hand side of the array and 4 rows of reference pixels on the top and bottom sides of the array. Ideally, the bottom and top reference rows can be used to remove offsets common to an entire output of the detector and the left and right reference columns can be used to remove noise common to a particular row since they are being read only fractions of a millisecond before the science pixels in that same row.

Figure 3.3 shows the average signal of 3 reference pixels in one row and the signal of a science pixel in that same row recorded during a dark exposure with H1RG-022. Both signals show an initial droop during the first 40 reads or so: an effect that is due to some type of electrical settling immediately following a reset of the pixels. This effect is similar to the reset anomaly, seen elsewhere in H2RG hybrid detectors [59]. The size of the drop can can be alleviated with tuning of the biases at a given temperature, but if it is present in data that was collected, the reference pixels provide a means of removing it.

As seen in Figure 3.3, the peak to peak amplitude of the reference pixels is about 4 times greater than that of the science pixel. This is most likely due to a difference in capacitance between the two types of pixels. It is noted in the specifications that parasitic capacitance is present on the reference pixels, increasing their capacitance, C_{ref}, by about 50% relative to the capacitance of the science pixels, C_{sci}. From the plot, it appears that the ratio of the two might be greater for this particular device. Figure 3.4 shows a subtraction of the average of 3 reference pixels in a row from two different science pixels in that same row. The signal in the initial reads is seen to increase quickly due to the

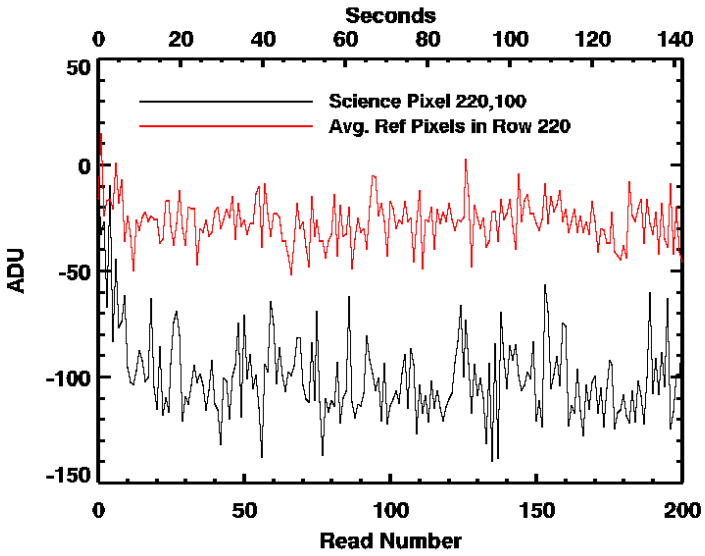


Figure 3.3: Signal vs. time in an UTR exposure for the average of 3 reference pixels in row 220 (black) and 1 science pixel (red) in the same row. The first sample has been subtracted so both ramps start from zero. Note the initial dip in signal for the first 40 reads. Since this exposure was taken in the dark, after this drop, the pixels are integrating only dark current at a very slow rate.

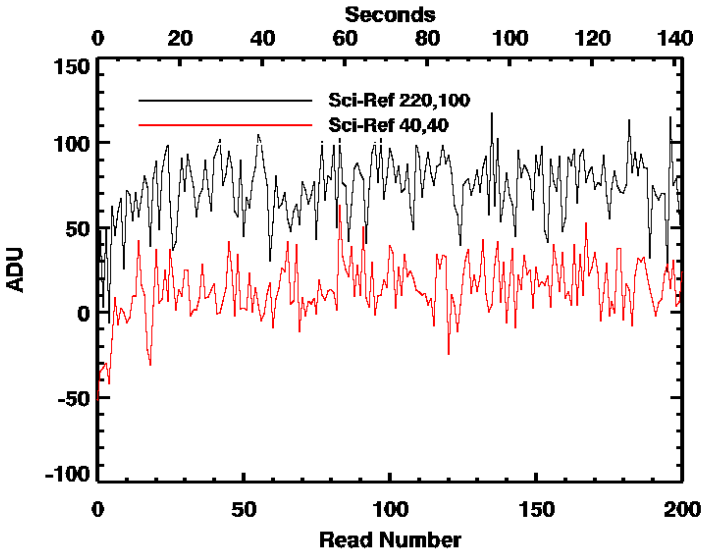


Figure 3.4: Difference in signal between science pixels and the average of 3 of the reference pixels in that same row vs. time in a UTR exposure. Because of a mismatch in capacitance, the signals do not subtract evenly. Instead of an initial droop, the subtraction makes the science pixels appear to rise in signal.

difference in amplitude between the science pixels and reference pixels.

To account for the discrepancy between the capacitance of the science and reference pixels, the coefficient C_{FAC} is introduced in the relation

$$C_{sci} = C_{FAC} * C_{ref}. \tag{3.6}$$

With the assumption that the sum of the dark and photo-generated signal in the science pixels, $S_{sci}(t)$, should be linear with respect to time, we seek to find the value of C_{FAC} that maximizes the linearity of the ideal signal, $S(t)$, with the following relation:

$$S(t,i,j) = S_{sci}(t,i,j) - C_{FAC} * S_{ref}(t,j), \tag{3.7}$$

where i and j are the column and row of the pixel, respectively, and $S_{ref}(t,j)$ is some estimate of the signal in the reference pixels in the row j. $S(t)$ is the signal we would expect to see in the absence of electrical noise in the detector and control electronics. It is the sum of the dark-generated signal, $D(t)$, and the photo-generated signal, $I(t)$.

To find C_{FAC}, we use exposures with low illumination so that the electrical noise is not dominated by shot noise and choose a region of pixels on the detector that contains a minimal amount of defects. A value of C_{FAC} is chosen and for that value a line is fit to the points given by Equation 3.7 for all of the pixels in the region. For $S_{ref}(t,j)$ the average of the 4 reference pixels on the left side of the detector and the average of the 4 reference pixels on the right side are taken separately across the rows and then smoothed with a Savitsky-Golay filter to eliminate spikes in the noise. The left

average is used for the pixels on the left half of the detector and the right average for the right half. The error of the fits (see section 5.2.4 for how this was calculated) for all of the pixels is then averaged, and this process is repeated for a range of C_{FAC} values until the minimum is found.

The results for the HyViSIs H1RG-022 and H2RG-32-147 are shown in Figure 3.5 and listed in Table 3.1. Measurements made on other regions of these detectors show that C_{FAC} does not vary significantly with pixel location, but does depend on the bias voltages V_{RESET} and D_{SUB}. The value was stable over multiple nights of using the detectors with these voltages held constant.

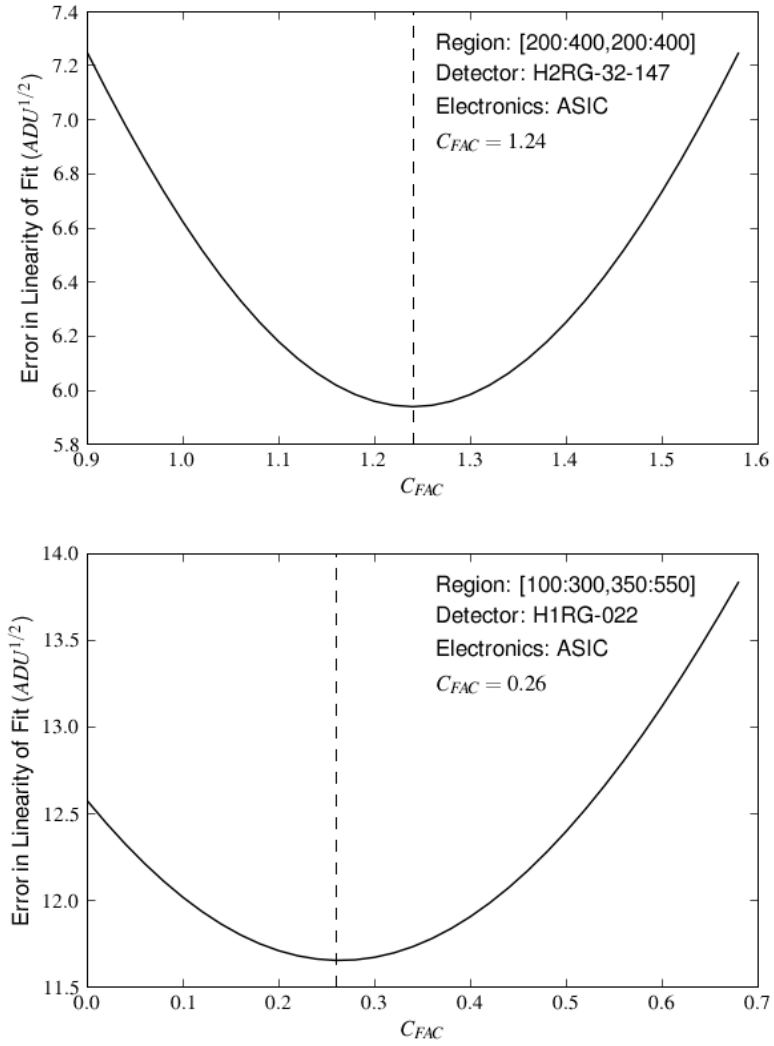


Figure 3.5: The error in the fit of a straight line to the data points obtained with Equation 3.7 vs. the capacitance factor C_{FAC}. The pixels, detector, and electronics used are indicated on the figures.

Table 3.1: C_{FAC} values along with the detector biases V_{RESET} and D_{SUB}. The average error in the linear fit of Equation 3.7 is also indicated.

Detector	Electronics	D_{SUB} (V)	V_{RESET} (V)	C_{FAC}	Error ($\sqrt{ADU}$)
H1RG-022	SIDECAR	0.375	0.094	0.26	5.97
H2RG-32-147	SIDECAR	0.305	0.300	1.24	11.65

Figure 3.6 shows the same subtraction as the one in Figure 3.4 except that the reference pixel average has been multiplied by the factor $C_{FAC} = 0.27$. No initial dip or rise is seen with the factor taken into account. Rather, the slope of the expected dark current signal is well matched throughout the ramps indicating that an adequate value of C_{FAC} has been applied.

This treatment was not applied to H4RG-007 because the dark current shot noise was dominant over the electrical noise. Also, it should be noted that this method was not applied to exposures taken with the ARC electronics. The reason for this is that crosstalk on the video boards causes a large coupling between the signal measured in the reference pixels and the science pixels.

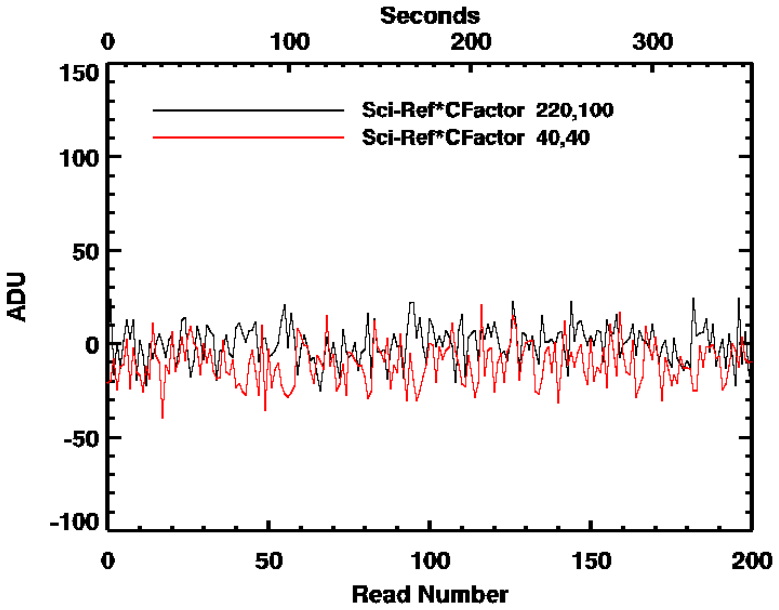


Figure 3.6: Difference in signal between science pixels and the average of 3 of the reference pixels in that same row multiplied by a factor C_{FAC} that takes into account the difference in capacitance between the two types of pixels.

3.5 Readout Electronics: SIDECAR ASIC

Interfacing focal plane arrays (FPAs) to a data acquisition system (DAQ) in astronomy usually requires a middle-man set of electronics that are referred to as the "readout" or "control" electronics. The readout electronics handle tasks such as converting the analog video signals to digital numbers, filtering noise from the signals, providing power or bias voltages to the FPA, and generating the clocking signals necessary to take an exposure with the array. In many cases this set of electronics is equipped with a microprocessor or microcontroller that can store a set of instructions in its internal memory. Different sets of instructions can be loaded based upon the desired mode of operation for the FPA (i.e. binning pixels, reading a subset of the array, etc.). And with a given set of instructions loaded, writing individual registers allows for fine tuning of parameters such as the frame rate and exposure time.

In most cases these readout electronic systems are rather bulky. Large racks are often needed to hold them and the power supplies that they need to operate. In some cases the readout electronics draw enough electrical current from the supplies to necessitate a fan to prevent overheating. They typically consist of multiple circuit boards with discrete chips for each function, i.e. voltage regulators, DACs, ADCs, memory etc. Often times, to make slight adjustments, additional components such as resistors or capacitors must be inserted or soldered to the boards. Since these electronics are usually bolted to a telescope or flying on a satellite in astronomical applications, weight and size can be an issue, and certainly one would rather avoid swapping out components.

To step away from the bulkiness and large power requirement of traditional electronics, Teledyne Scientific has produced a multi-purpose control Application-Specific Integrated Circuit (ASIC) called the SIDECAR (**S**ystem for **I**mage **D**igitization, **E**nhancement, **C**ontrol **A**nd **R**etrieval). A block diagram of the SIDECAR ASIC and photographs of the 22 x 14.5 mm^2 die mounted in two different packages is shown in Figure 3.7. As the diagram indicates, the chip contains all of the functionality needed to control and readout a detector: clocks, biases, ADCs, etc. And in addition to the functions shown, the chip provides pre-amplification and amplification stages as well as array processors that permit data processing function such as co-adding channels or subtracting offsets.

In addition to the compact science mission packaging schemes shown in Figure 3.7, Teledyne offers the SIDECAR in a development kit intended for laboratory use. The development kit, which was used for most measurements included in this thesis, is shown in Figure 3.8. In this configuration, the ASIC is placed on a board that has numerous test points that allow one to probe its various output signals, and LEDs that show clock activity. This board connects to a smaller board called the JADE card that handles communication with a Windows PC via a USB interface. The development kit as a whole is essentially a plug and play device. The user only needs to write the assembly code instructions for the microcontroller in the ASIC and the software that the PC will use to extract the data through the USB bus. For the latter, a library of Microsoft COM functions is provided so that typical astronomy applications such as IDL can be used for data retrieval and configuration.

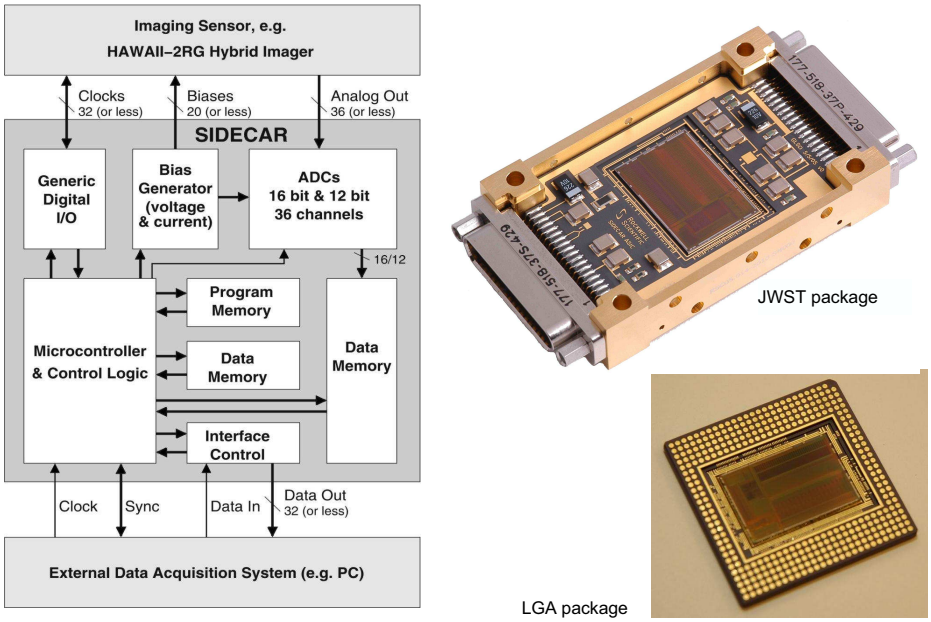


Figure 3.7: A block diagram of the SIDECAR ASIC along with two packaging options available for the chip (taken from Ref. [60]).

At the time of writing, Teleldyne offers the SIDECAR Development Kit in two flavors: warm and cold. The warm kit operates at room temperature, and so it must be placed outside the cryogenic system enclosing the detector to which it is attached. This implies that there is a significant length of cable over which the analog signals must travel between the ASIC and the detector. The cold, or *cryogenic*, kit can operate at cryogenic temperatures, which means that it can be placed directly beside the detector so that only digital signals going to and from the DAQ must travel over long distances. This has the benefit of reducing noise pickup during the analog transmission and the lower temperature of the chip inherently reduces the noise and leakage currents in the ASIC itself. Both warm and cryogenic kits were tested for this thesis work, but in this section only the warm kit is discussed. Results for the cryogenic kit are included in later chapters.

In the following subsections we will not provide a comprehensive description of the chip; only a brief review of some of the features and its performance in the various modes of operation with the HyViSI devices. For a detailed description of the SIDECAR, the reader is referred to reference [60].

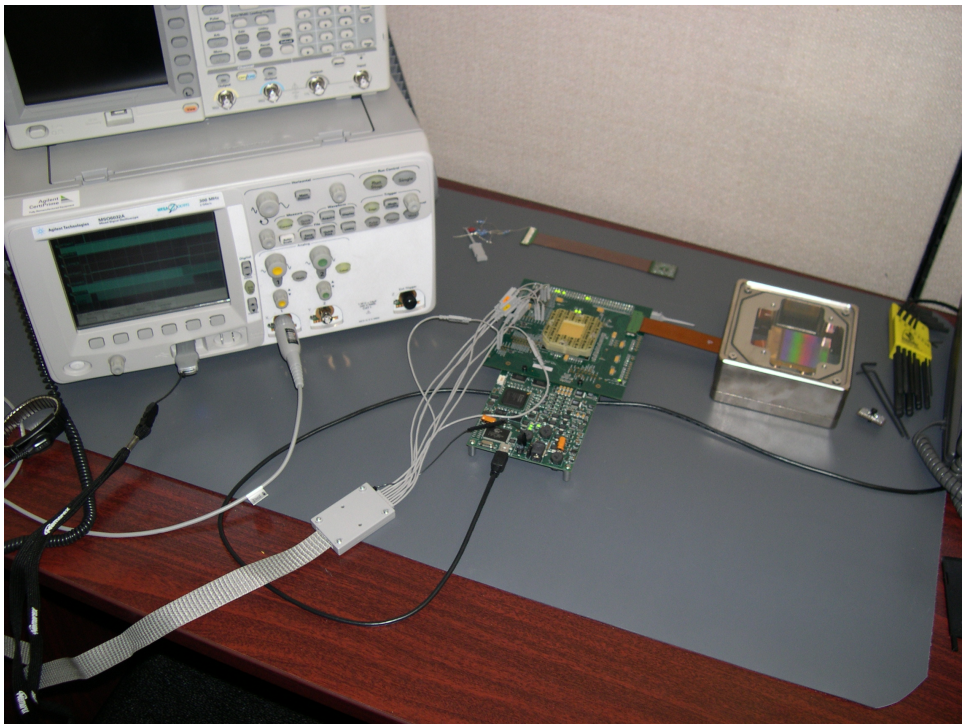


Figure 3.8: A photograph of the SIDECAR development kit attached to an H4RG multiplexer and a USB cable going to a laptop computer not shown. Multiple clocks and analog signals are being traced on the oscilloscope.

3.5.1 Pre-Amplification Stage

Before analog-to-digital conversion takes place, the video signals from the detector are first fed into the SIDECAR pre-amps. In the pre-amp stage, signals can be routed to different channels or shared among them via an internal mux, offsets can be added to them with a DAC, and they can be amplified and filtered. The amplification sub-block of the pre-amps, displayed in Figure 3.9, shows that capacitive feedback is used for gain selection with the capacitors C_{FB} and the inputs are capacitively coupled to the amplifier through C_{IN}. Capacitive feedback has the advantage that is does not lower gain since it does not trade gain for bandwidth [61].

During the amplification stage, the switches $S3$ and $S6$ are open. This leaves the nodes of the capacitors C_{FB} floating at a potential set by a reset transistor switch. As mentioned in Ref. [60], inevitable leakage currents in the silicon will cause these nodes to drift with a time constant that depends highly on temperature. When the SIDECAR chip is placed inside the dewar and cooled along with the detector, the leakage currents are very small and this does not present a problem. However, when the SIDECAR is held at room temperature, the pre-amp drifts are very noticeable after fractions of a second. This drift and the noise associated with resetting the capacitors take

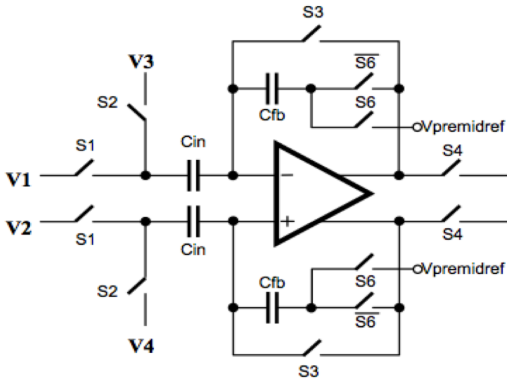


Figure 3.9: A schematic of the amplification stage for a given channel in the SIDECAR. Note that capacitors are used for feedback rather than resistors.

different forms based upon the implementation of the signals available in the pre-amp stage.

Pre-amp reset once per frame From the standpoint of assembly coding, an easy-to-implement conversion sequence is to reset C_{FB} once per frame. However, for frame times as short as 0.25 seconds, the leakage currents can cause the output voltage going into the ADCs to swing by as much as 25 mV. This drift over the frame, which will continue for longer frame times, is shown in the top of Figure 3.10. It is evident then that more frequent resets of the pre-amp capacitors are needed at room temperature.

Pre-amp reset once per row: No kTC Removal An alternative to resetting once per frame is to reset before every row conversion. The row may consist of the number of pixels in a row per output of the detector or the number of pixels in the row of a sub-window. The problem with this is that after each reset, a random amount of charge will be left on C_{FB}. The noise associated with this random charge should be proportional to kTC_{FB}, where k is Boltzmann's constant and T is the temperature. The pattern associated with this noise, shown in the middle frame of Figure 3.10, consists of horizontal bands across the frame. And since it is uncorrelated from frame to frame, CDS subtraction of two consecutive frames will boost this noise by $\sqrt{2}$.

Pre-amp reset once per row: With kTC Removal In order to deal with the horizontal banding caused by kTC noise, an intrinsic analog CDS scheme is used in the pre-amp and ADC blocks. In each ADC conversion, both the video signal from the pre-amp and the voltage on the capacitors are sampled. The ADC digitizes the sampled signal using the latter as a reference so that the kTC offset subtracts out. The last frame in Figure 3.10 shows that the noise is essentially white when this method is used. The implementation of the kTC removal scheme involves toggling one signal in the pre-amp and slightly changing the internal bias voltages from those used in the normal reset schemes; it does not require additional time for conversions.

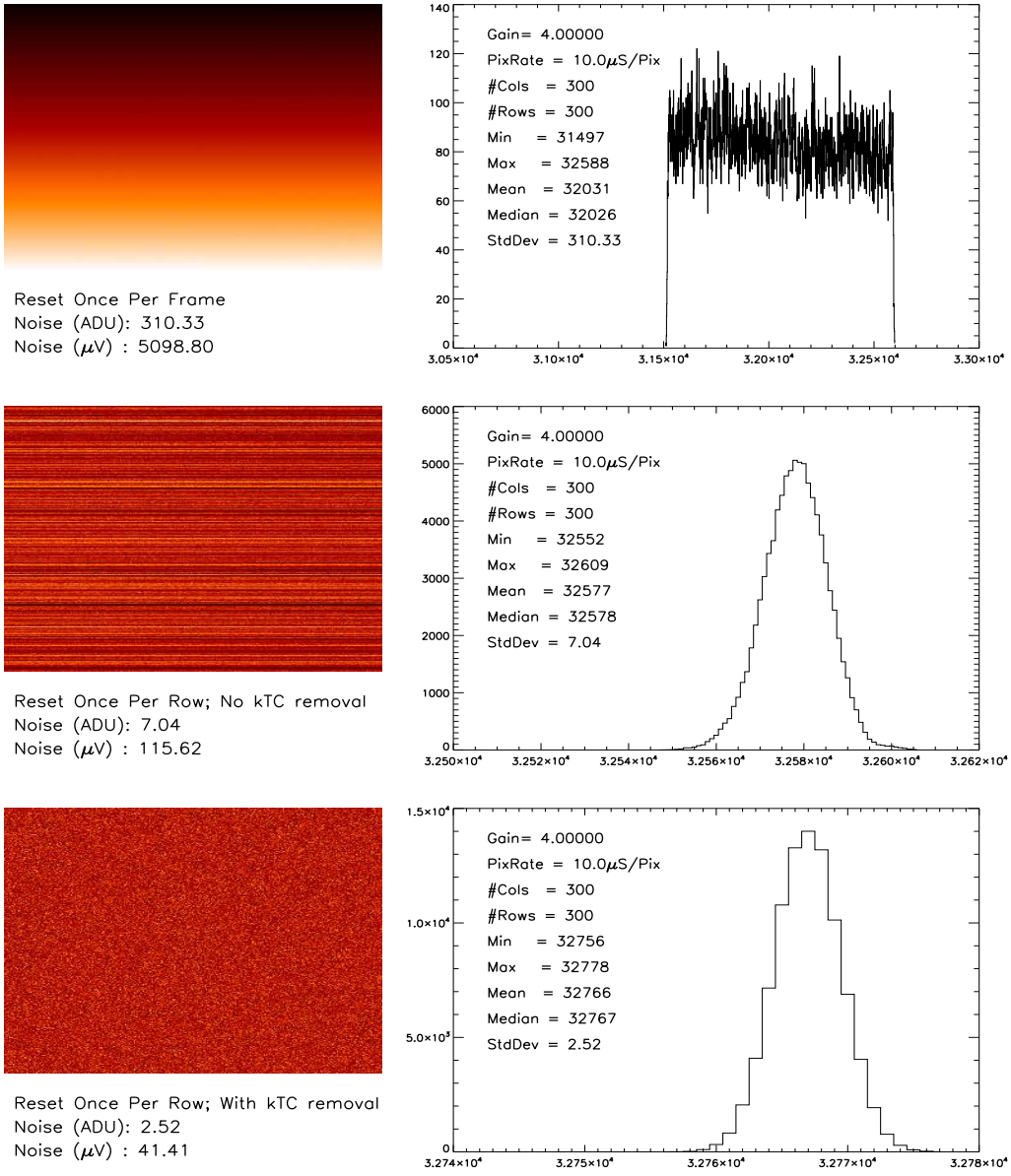


Figure 3.10: Different reset schemes for the SIDECAR pre-amps. The methods used to obtain each are described in the text. All measurements were made with the input voltages $V1$-$V4$ connected to the internal ground.

3.5.2 Conversion Gain

In all measurements described in this thesis, the 16 bit Successive Approximation (SAR) ADCs were used. For the SAR ADC, the conversion from microvolts to ADU is given approximately as

$$Counts(ADU) = 32768 * \frac{V_{In} - V_{Ref}}{VRP - VRN} * G + 32768, \tag{3.8}$$

where V_{In} is the input voltage to the pre-amp, V_{Ref} is one of the selectable reference voltages provided on the chip, VRP and VRN are ADC reference voltages, and G is the gain of the pre-amp. G is configurable in 3 dB increments from -3 dB to 27 dB. V_{In} and V_{Ref} correspond to the input voltages $V1$ and $V2$ in Figure 3.9. The selection $V_{REF} = V_{REFMAIN}$ was used for all measurements in this section aside from the ones where both inputs were set to ground. As will be discussed, when the detector outputs are being sampled, $V_{REFMAIN}$ results in a much higher noise than that obtained by using the signal V_{REFOUT} from the detector. Routing V_{REFOUT} from the detector to $InPCommon$ or $InNCommon$ on the SIDECAR is absolutely essential for low noise performance.

To obtain a conversion from voltage to ADU and measure the actual values of G for our configuration, a voltage was supplied to the pre-amp inputs with an Agilent E3647A Dual Output 30 volt power supply. This voltage was increased from 0-3.3 V by 0.1 V increments and at each increment a set of 20,000 digitizations were recorded. The mean of these digitizations were taken to yield an average ADU value at that voltage. The conversion from μV to ADU at a gain of 1 is shown in Figure 3.11. The conversion measured at other gains is listed in Table 3.2.

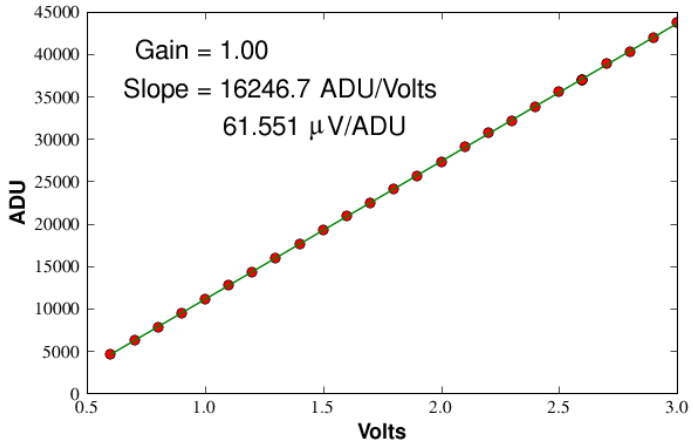


Figure 3.11: Average digital ADU values (shown as red dots) vs. the input voltage at which they were measured. The inverse slope of the best-fit line (shown in green) to the points yields the conversion in μV/ADU.

Table 3.2 shows that as the gain setting in the pre-amp is increased, the measured gain moves farther from the expected value, with the former being less than the latter. This is due to the use of low noise resistors and single-stage buffers in the system (low noise settings result in diminished

gain). However, the measured values are very repeatable, so with proper calibration this discrepancy can be taken into account and should not be a performance-limiting issue.

The last two columns in the table labeled *Ground Noise* show the digitization noise when the pre-amps are configured to measure the internal ground signal on the chip. In agreement with Figure 5 in Loose et al. [60], the noise is ADC limited below a $G = 4$ and above that it is limited by the pre-amp. The two columns labeled *Supply Noise* contain the digitization noise measured when reading the voltage from the Agilent power supply. The RMS output noise of the power supply was measured with an oscilloscope to be about 340 μV, so this asymptotic behavior of the noise with increasing gain is to be expected.

Table 3.2: Measured pre-amp gains for the SIDECAR ASIC.

Gain	Gain (dB)	Measured Gain (μV/ADU)	Measured Gain	Supply Noise (ADU)	Supply Noise (μV)	Ground Noise (ADU)	Ground Noise (μV)
0.71	-3.01	86.07	0.71	4.79	412.3	2.70	232.4
1.00	0.00	61.53	1.00	6.81	383.9	2.72	166.7
1.41	3.01	43.56	1.41	8.35	363.7	2.78	121.1
2.00	6.02	30.77	2.00	11.67	359.1	2.80	86.2
2.83	9.03	22.96	2.83	16.22	352.4	2.75	59.7
4.00	12.04	15.46	3.97	22.52	348.2	2.81	43.4
5.65	15.05	11.01	5.58	31.10	342.4	2.97	32.7
8.00	18.06	8.01	7.67	43.03	344.6	3.22	25.8
11.31	21.07	5.75	10.71	59.90	344.2	3.62	20.8
16.00	24.08	4.15	14.84	83.82	347.2	4.29	17.8
22.62	27.09	3.00	20.48	115.01	345.9	5.25	15.8

3.5.3 Averaging Multiple Channels

The input routing multiplexer of the pre-amp and math capabilities of the array processor (AP) allow one input signal to be shared and digitized on multiple channels and then averaged before it is written to the dual port memory and read out by the DAQ. This might be advantageous if buffer size or memory overflow is an issue in the readout system. And this feature is particularly useful for the HxRG multiplexers as the number of outputs is configurable. For instance, the detector can be run in four output mode, with each output being sampled and averaged between eight channels on the SIDECAR. And in window mode only one output of the detector is used, so there is no reason not to take advantage of multiple channels on the SIDECAR other than power consumption.

Each math operation and read/write in the array processors requires at least one clock cycle.

One would expect that doing math between A/D conversions would therefore result in an overall slower pixel conversion time. However, the array processor clock can be configured to run at a faster rate than the ADC clock so that no time is lost between successive A/D conversions.[3] Example clock rates and conversion times are given in Table 3.3. The redundancy in the table is meant to illustrate that no decrease in pixel rate is suffered in the channel averaging process.

The basic process for averaging N channels is as follows: The first channel writes its value to dual port memory. After a certain delay, the second channel reads this value from memory, adds its own A/D value to it, and writes it back to the same address. This process continues with the N channels until all have been coadded. Then a bit shift (or a multiplication followed by a bit shift for the case of $N = 6$) is used to achieve the averaging. Finally, the value is written back to memory and stored until it is extracted by the DAQ.

Table 3.3: SIDECAR clocking scheme for averaging multiple channels.

Channels	System Clock (MHz)	AP Clock (MHz)	ADC Clock (MHz)	AP Cycles / Pixel (Used/Total)	ADC Cycles / Pixel (Total)	Pixel Time (μs)
1	10	1.00	1.00	2/10	10	10
2	10	1.00	1.00	5/10	10	10
4	10	5.00	1.00	13/50	10	10
6	10	5.00	1.00	17/50	10	10
8	10	5.00	1.00	24/50	10	10

Figure 3.12 shows the reduction in noise when multiple channels are averaged. The decrease goes nearly as the theoretically predicted $1/\sqrt{N}$ channels, indicating that the noise in the ADCs is uncorrelated.

Application to H1RG HyViSI Detector As an example of the usefulness of averaging multiple channels and selecting the proper conversion gain on the SIDECAR, temporal read noise measurements were made on the 1024x1024 Hybrid Visible Silicon (HyViSI) detector H1RG-022. To make these measurements we ran the detector in window mode and obtained a set of correlated double sample (CDS or *reset-read-read*) frames. The window selected included the first 100 rows and 100 columns of pixels. The first 4 rows and first 4 columns of the window were reference pixels that have a capacitance of roughly $C_{Ref} = 35\ fF$ and all of the others were science pixels that have a capacitance of $C_{Sci} = 14\ fF$. After obtaining a stack of CDS frames, we look at the temporal variation of each pixel and measure its standard deviation. Averaging the standard deviations yields the temporal read noise.

[3]The only caveat is that the ADC clock must be divided relative to the AP clock. Since both clocks are derived from the microcontroller system clock, the ADC clock must run more slowly than the system clock if more than 2 channels are averaged. See Table 3.3 for the relative rates.

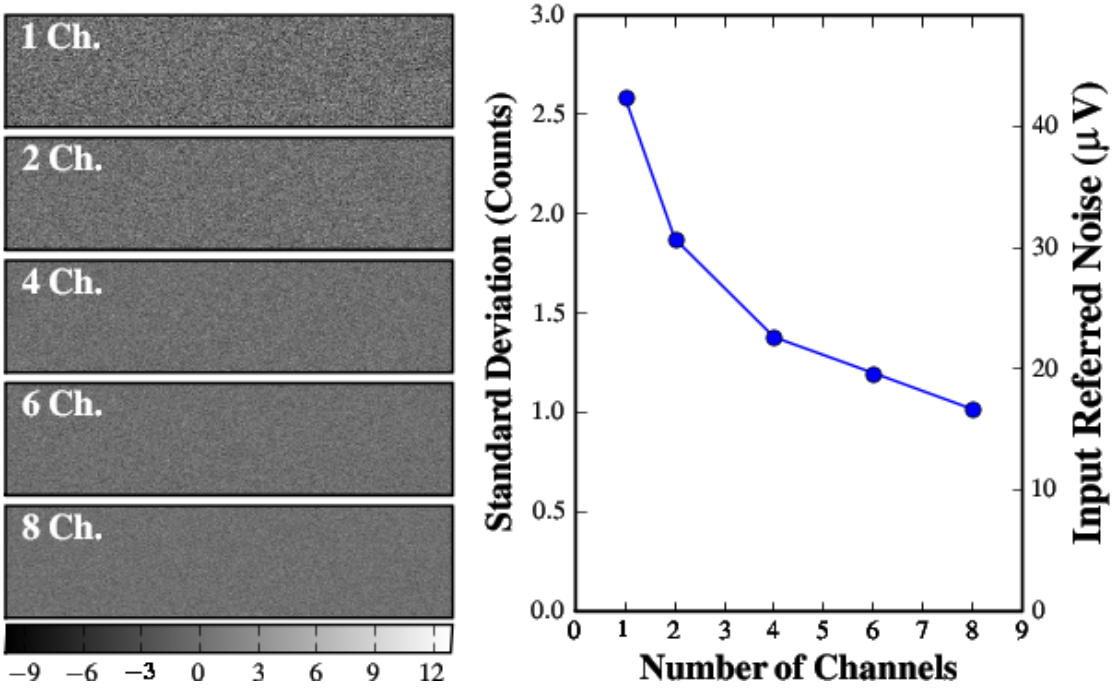


Figure 3.12: (left) Digitizations of noise in the SIDECAR for averaging 1-8 channels at a pre-amp gain of 4. The colorbar shows the ADU values after subtraction of the mean value of 32768. (right) The RMS value of the noise as a function of the number of channels averaged.

As illustrated in Figure 3.13, the difference in pixel capacitances makes a significant difference in the read noise, especially when measured in electrons. In astronomical applications, quantifying the read noise in electrons is the most useful since this is the quantity that can be directly used, along with quantum efficiency, to estimate the number of photons that hit a given pixel. We see that for small capacitances, the larger gain yields a much smaller read noise in electrons. For the eight channel average noise of the reference pixels, both measurements yield a noise of about 3 ADU, indicating that the ADC contribution to the noise might be dominating rather than the contribution from the detector noise.

The lower read noise at higher gain is very desirable, suggesting that the SIDECAR pre-amps should be run at a high gain with HyViSI detectors. However, one must also consider the trade-off between gain and the usable dynamic range of the detector. For instance, in the case of H1RG-018 the well depth is approximately $110,000$ e^-. If the SIDECAR pre-amps are set at a gain of 4 the conversion gain is about 1.4 e^-/ADU and the full ADU range needed to cover the entire well will be 78,500 ADU. However, the 16 bit ADC allows a maximum of 65,536 values, so some portion of the voltage range will be lost. If the entire well depth is to be used, a gain slightly below this must be chosen.

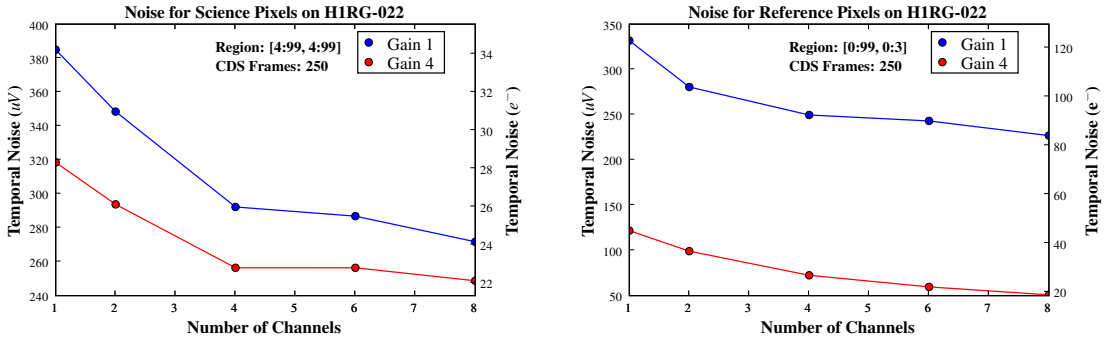


Figure 3.13: (Left) Read noise of H1RG-022 science pixels measured with multiple channels averaged. (Right) Read noise of H1RG-022 reference pixels measured with multiple channels averaged. The capacitance of the science pixels is roughly 2.5 times as large as that of the reference pixels. For this particular measurement, the SIDECAR was configured so that $G = 1$ yielded 64 uV/ADU and $G = 4$ yielded 16 μV/ADU.

3.5.4 Noise Performance vs. Pixel Time

Among other projects, the SIDECAR is being considered for use in the guider cameras of the **L**arge **S**ynoptic **S**urvey **T**elescope (LSST) and the **S**uper**N**ova **A**cceleration **P**robe (SNAP). An important parameter for both of these systems will be the frame rate for imaging a guide star, which is dependent on the pixel conversion time. For 100 Hz adjustments of the telescope pointing using centroids from multiple guide stars, a pixel conversion time smaller than 10 μs may be required. And for science imaging applications, reducing the pixel time has the added benefit of increasing dynamic range since bright stars can be imaged before they saturate the pixels. The performance of the SIDECAR ADC at faster A/D conversion times will thus directly affect both of these applications.

Increasing the pixel conversion rate means that voltages in the system, i.e. the analog output of the detector, the digital clock signals, the DAC voltages used in the SAR ADC, etc., must change more quickly. The relationship $I = C\ dV/dt$ tells us that changing these voltages more quickly requires increases in electrical current. In general, increasing the speed at which signals are converted to digital output should be accommodated by an increase in drive currents in the system and an attempt to reduce source impedances and load capacitances to account for this. In the case of the SIDECAR, we need to increase the bias currents in the pre-amp and SAR ADC.

For most ground based astronomy missions, boosting the currents is perfectly fine since there is usually plenty of electrical power for the electrical and thermal systems. However, for space based missions it might be desirable to keep the currents and heat dissipation as low as possible since electrical power from solar panels is not an abundant resource. For this reason, the minimum ADC and pre-amp currents needed to operate the SIDECAR at each clock speed were sought. Values for which the noise histogram remains approximately Gaussian in shape up to a gain of $G = 4$ are listed

Table 3.4: Operating parameters and noise performance at different A/D sampling rates on the SIDECAR ASIC. Note that the noise can be kept at a nearly constant level below 167 kHz by increasing the bias currents.

Sampling Rate	100 kHz	125 kHz	167 kHz	250 kHz
$I_{PreAmpBias}$ (μA)	6.40	6.40	16.00	17.60
$I_{PreAmpCasc}$ (μA)	1.60	1.60	1.60	1.60
I_{NBIAS1} (μA)	11.30	14.50	17.70	48.00
I_{NBIAS2} (μA)	11.30	14.50	17.70	48.00
I_{NFB1} (μA)	5.00	5.00	13.00	34.00
I_{NFB2} (μA)	5.00	5.00	13.00	34.00
$I_{VRP\ DAC}$ (mA)	0.343	0.343	0.343	0.654
$I_{VRN\ DAC}$ (mA)	60.00	60.00	60.00	60.00
RMS Noise (ADU)	2.65	2.65	2.76	3.13
RMS Noise uV	41.0	41.0	42.7	48.4

in Table 3.4. It should be note that these currents minimized the noise when the SIDECAR inputs were connected to ground. They do not necessarily offer the best performance when reading out the detector.

If the drive currents fall significantly below these values, certain instabilities in the digital output of the SIDECAR will result. A description of these instabilities, along with the current that is most likely lacking is included in the following section. In general, increasing the currents I_{NBIAS1}, I_{NBIAS2}, I_{NFB1}, and I_{NFB2} will decrease the ADC noise (Markus Loose, private communication). However, if brought too high they can also cause instabilities. It should also be mentioned that the values listed in Table 3.4 work for gains less than or equal to $G = 4$. If a higher gain is used, one or more of the currents may need to be increased.

3.5.5 Reference Voltages for SIDECAR when Connected to an HxRG

The most important point to consider when hooking an HxRG detector up to the SIDECAR, by far, is that using the signal V_{REFOUT} from the detector as a reference in A/D conversions offers the best performance. In some versions of the cabling that we used, V_{REFOUT} was routed to one of the upper input channels (IN32) of the SIDECAR. In this configuration, it must be measured against a reference voltage and digitized separately, which only allows it to be used in digital subtraction. In other versions, V_{REFOUT} was routed to $InPCommon$ or $InNCommon$, where it can be routed to all 36 channels and used as a reference for measurement of the detector outputs. This is preferred since it provides a truly differential measurement and eliminates common mode noise before digitization.

As a measure of the significance of using V_{REFOUT} versus an internal reference voltage on the

SIDECAR, in a particular configuration of the detector, using $V_{REFMAIN}$ will result in 40-50 e^- read noise. When V_{REFOUT} is used instead, with no other changes being made, the read noise reduces to 10 e^-. Care should therefore be taken when designing the cables from the detector to the SIDECAR to ensure that V_{REFOUT} connects to $InPCommon$ or $InNCommon$.

Additional Notes

- Noise histogram is very sensitive to DAC buffer current of VRP and VRN.
- The most important settings in determining the RMS and the shape of the noise histogram seem to be $I_{PreAmpBias}$. If $I_{PreAmpBias}$ is too low, the noise will be high and at a certain threshold, the histogram will take a pitchfork shape like that in middle panel of Figure 3.14.
- When we increase the pre-amp gains, we should also increase I_{NBIAS1}, I_{NBIAS2}, I_{NFB1} and I_{NFB2}. Otherwise, the noise distribution becomes jagged, indicating a loss of the LSB.
- If I_{NBIAS1} and I_{NBIAS2} are too small, the histogram will appear jagged, with the odd ADU values having smaller values than their even neighbors as in the top panel of Figure 3.14.
- If I_{NFB1} and I_{NFB2} are too small, the distribution will be less strongly peaked. Increasing these currents will make the peak tighter. If these currents are too small, a gap might form in the distribution.
- Large capacitance loads will cause ringing somewhere in the ADC. The resultant histogram will have at least three separate peaks. Increasing the drive currents will bring these peaks closer together and eventually they will merge.
- The DAC buffer current for VRP should be adjusted for different gains in order to keep a nicely peaked histogram.
- If I_{NBIAS1}, I_{NBIAS2} are set too high, speckles will appear in the digitizations, indicating voltage spikes somewhere in the system, as in the bottom panel of Figure 3.14.

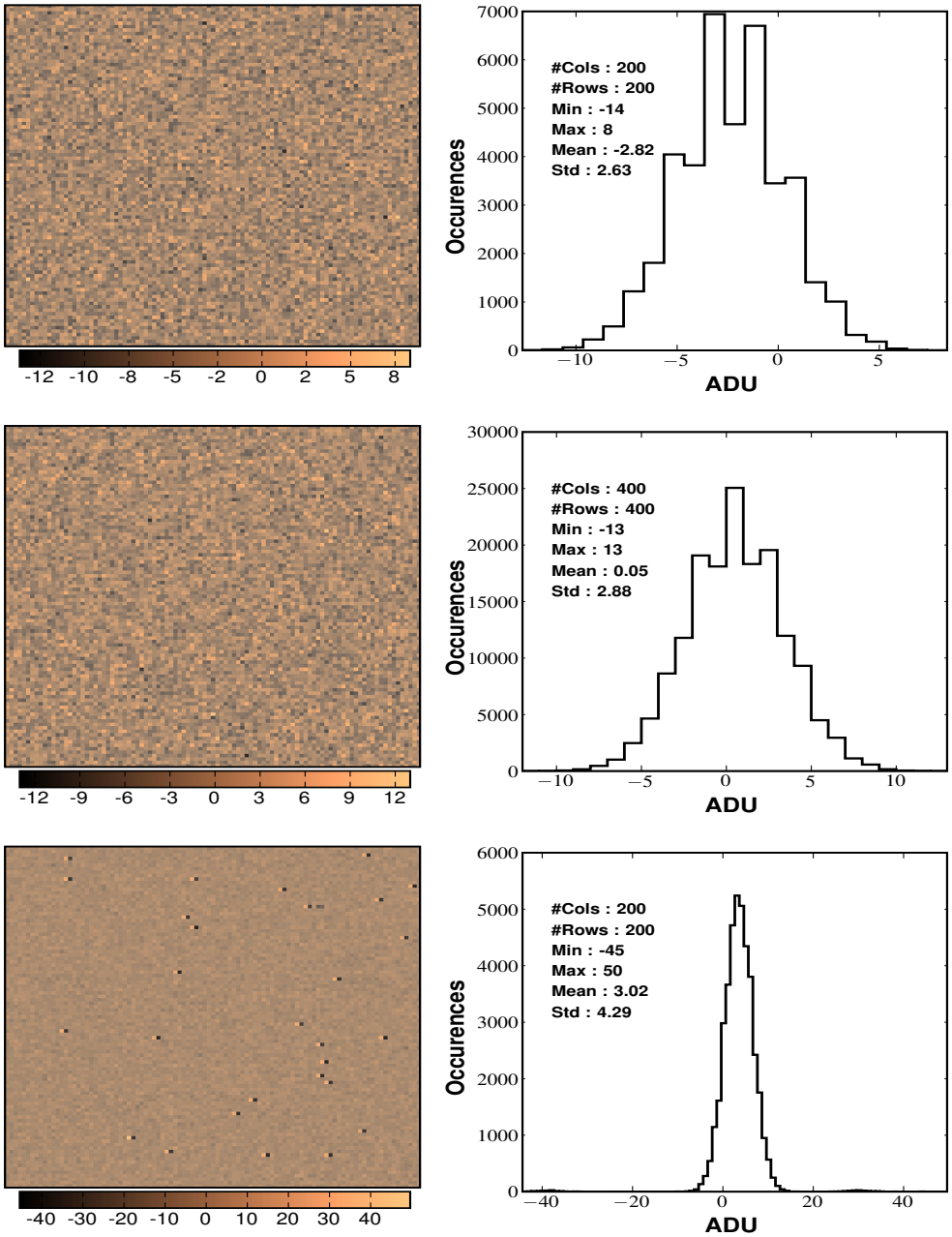


Figure 3.14: (Top) Digitizations taken at 8 μs when $I_{NBIAS1} = I_{NBIAS2} = 8.1$ μA were too low. (Middle) Digitizations taken at 6 μs when $I_{PreAmpBias} = 1.6$ μA was too low. (Bottom) Digitizations taken at 4 μs when $I_{NBIAS1} = I_{NBIAS2} = 53$ μA were too high. All digitizations were taken with a pre-amp gain of $G = 4$.

Chapter 4

Laboratory Testing of SiPIN Detectors

Understanding and calibrating astronomical detectors in a laboratory environment before they are used to make measurements on the sky is essential. Not only does it help weed out broken parts or components that might be susceptible to damage when used continuously as part of an astronomical survey; characterizing a detector is crucial to understanding the signature it will impart on the data. For instance, the point spread function of the detector needs to be understood in order to make shape measurements of galaxies for the purpose of weak lensing studies.

In this chapter, standard characterization tests for an imaging detector are described in theory and measurements for HyViSI devices are presented. These standard tests include pixel operability, conversion gain and nodal capacitance, read noise, dark current, quantum efficiency, linearity, and well depth. Although pixel crosstalk and persistence are typically included in this list, these subjects are saved for later chapters as they entail a great deal of complexity. Before delving into these matters, we begin by describing the test systems that were used to carry out these tests.

4.1 Description of Laboratory Setups and Devices Tested

Most of the laboratory work done for this thesis was performed at the Rochester Imaging Detector Laboratory (RIDL) at the Rochester Institute of Technology (RIT). The RIDL system (see Figure 4.1) consists of software and hardware similar to that described in Figer et al. [62]. The modular architecture of the system allows for rapid acquisition and reduction of large datasets over a broad range of experimental conditions. Minimal effort is required to change between different detectors and different types of detectors, and the system can be transported for operation on a telescope.

The RIDL system includes a 16 inch diameter dewar (Universal Cryogenics, Tucson, AZ) with a

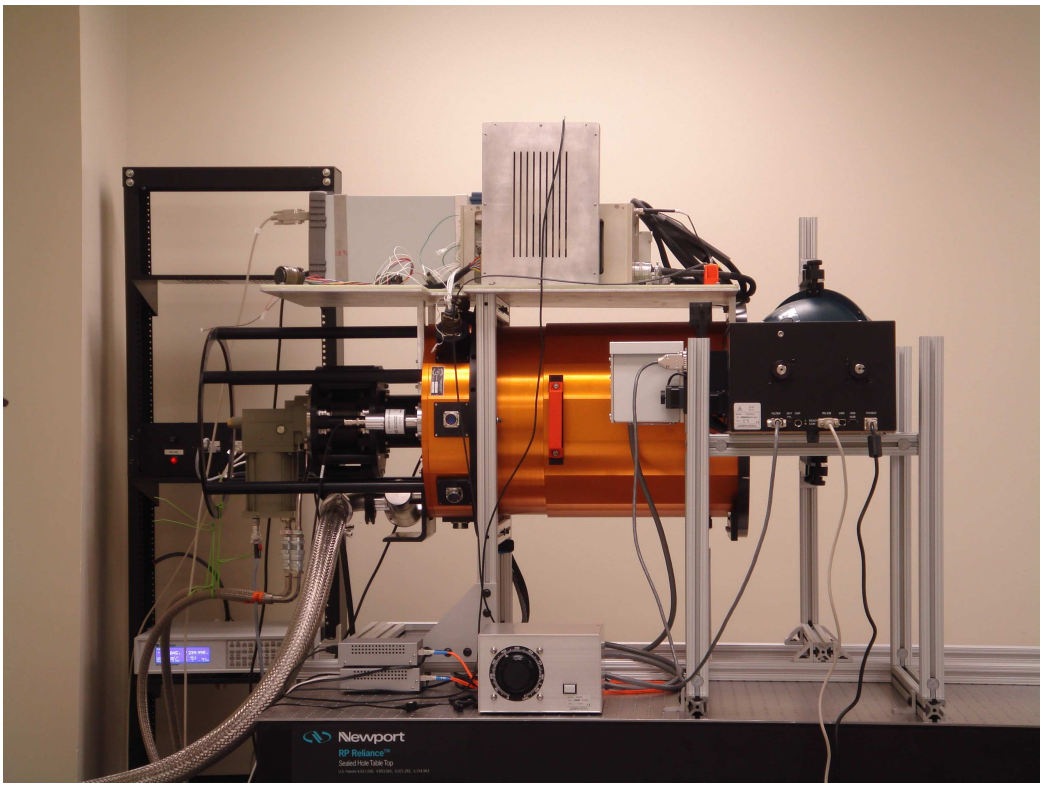


Figure 4.1: RIDL system. An orange dewar houses two filter wheels and a detector enclosure. A helium cryo-cooler cools the system. The picture shows an integrating sphere and monochromator near the front of the dewar. Post-processing electronics are mounted on a plate attached to the top side of the dewar. Off camera are three computers (two four-way and one eight-way CPU) with 4 GB, 12 GB, and 16 GB of RAM and 12 TB of RAID5 storage.

110 mm diameter CaF2 window, two cryogenic filter wheels, and a detector enclosure. The system is cooled with a two-stage cooler (CTI Model 1050, Brooks Automation, Chelmsford, MA), and the detector is thermally stabilized with a 10-channel temperature controller (Lakeshore Cryotronics, Westerville, OH). The detector enclosure provides thermal and electrical feedthroughs and a light-tight cavity for the detector. The filter wheels can accommodate eight filters and/or radiation sources. We used two sets of readout electronics with this setup: 1) the Generation III electronics from Astronomical Research Cameras, Inc. (San Diego, CA), and 2) the room temperature SIDECAR ASIC from Teledyne Scientific & Imaging, LLC (Thousand Oaks, CA). A variety of programmable gains were used on each set of electronics. Unless otherwise noted, the data were recorded using a 5 us pixel time for the ARC electronics and 10 us pixel time for the SIDECAR. These correspond to 200 kHz and 100 kHz, respectively.

Additional measurements were performed at Teledyne Imaging Sensors (TIS) with a small custom dewar. The only electronic components inside the dewar, aside from standard silicon diode temperature sensors, were an H2RG detector and a cryogenic SIDECAR ASIC. Although the cryogenic SIDECAR, powered by a linear power supply instead of USB power, provided better noise performance than the room temperature SIDECAR kit used at RIDL, this setup was far more prohibitive in terms of the available tests that could be performed. There was no entry window into the dewar, so no light could be shined from outside. An attempt to place an LED inside of the dewar failed because its power supply created approximately 200 microvolts of noise, and limited testing time precluded the option of placing a light source inside to project a well sized spot of light for persistence measurements. However, an Fe^{55} source was made available for x-ray characterization, and very low noise dark current and single pixel reset measurements were performed.

The last setup from which measurements were obtained was a temporary one in the LSST laboratory at Stanford University. This system was composed of one of the LSST test dewars (also from Universal Cryogenics) on loan from Purdue University and an H2RG detector borrowed from Teledyne. Its primary purpose was to serve as a testbed for the LSST guider system, with a laser projection system delivering a small FWHM beam through a 6.5" pure fused silica window to the detector in order to simulate a guide star. Only a small fraction of the data from this detector and setup will be discussed–primarily to highlight a reduction in the Interpixel Charge Transfer (IPCT) discussed in Section 6.1.2—since ample time was not available to do rigorous analysis.

A list of the detectors that were tested in these various settings is shown in Table 4.1. The table includes the time period over which they were tested and where the testing was done. With the exception of H2RG-001 and H1RG-022, the dates correspond roughly to the newest technology available at the time, and thus, improvements in the HyViSI processing. For instance, a new type of surface treatment was used with H2RG-148 that significantly reduced the charge loss from IPCT seen in the previous devices. Other metrics were fairly consistent over all the detectors (aside from the H4RG), and all showed the same characteristics in image persistence.

HyViSI Detector	Pixel Size (μm)	Locations Tested	Dates of Testing
H1RG-022	18	RIDL, KPNO	3/07-7/08
H2RG-32-147	18	RIDL, KPNO	11/07-4/08
H2RG-001	18	TIS	11/08-12/08
H2RG-148	18	Stanford	7/09-8/09
H4RG-10-007	10	RIDL, KPNO	3/07-9/07

Table 4.1: A list of the various detectors studied in the laboratory for this thesis work. KPNO indicates that the device was also tested at the Kitt Peak 2.1m telescope.

4.2 Pixel Classification and Operability

The percentage of good pixels on a detector is called *pixel operability* [63]. It is a key figure of merit for both ground and space based astronomy missions. When a focal plane array is constructed, it is desirable to start out with the lowest fraction of bad pixels possible since some of the good ones will inevitably fail after prolonged use or bombardment with radiation, especially in the environment of space [64, 12]. Operability is usually specified as a detector requirement in the initial planning stages of a mission. For instance, the James Webb Space Telescope requires 98% operability for its science arrays [65].

There are several distinct types of pixels on the HyViSI detectors that are not suitable for scientific measurements. These pixels cannot be used to estimate the signal from a source without applying some correction, and in some instances, they cannot be used at all. They are categorized as dead, hot, and open pixels. In analyzing science data, these pixels are masked and not used. All of the other pixels on the detector are of suitable quality to be used in the science data analysis.

4.2.1 Dead or Railed Pixels

The first type of unusable pixels are "dead" pixels that do not increase in signal, no matter how much light they see. Most of them are railed at the high end of the detector voltage range, suggesting they are shorted to one of the high bias voltages. A substantial fraction of them also fall at the center of very high dark current clusters, sometimes referred to as "volcanoes". These pixels are easy to detect since they do not show an increase in signal over time. In order to find them, we take differences between a read r and the first read $r = 1$ from median flat field images and flag pixels that had a difference in signal, $I_{Diff}(x, y)$, below a certain threshold T. I.e.

$$I_{Diff}(x, y) = I(x, y, r) - I(x, y, 0) < T, \tag{4.1}$$

for all r. T depends on the gain of the pre-amps in the SIDECAR ASIC, but is typically set at $3\sigma_r$, where σ_r is the read noise of the detector at that gain.

For H2RG-32-147, nearly all of the dead pixels are found in volcanoes. For H4RG-10-007, the number includes an entire row of 4096 pixels that is presumably a bad line in the ROIC.

4.2.2 Hot Pixels

The hot pixels are found by (1) looking for very high pixel signal slopes, $\Delta I/\Delta t$, in UTR exposures and (2) looking for pixels that have a value of I greater than 75% of the full A/D range in the first read of median dark exposures. In the latter case, the dark current is so extreme that the pixel voltage reaches the upper rail almost immediately after reset, and its slope is flattened before the first read. For (1), pixels are flagged if they have a dark current greater than 10 e^-/s/pix.

4.2.3 Open Pixels

Open pixels are pixels that have a value that falls significantly below that of their nearest neighbors in well-illuminated images. I.e.

$$I(x, y, r) \ll I(x \pm 1, y \pm 1, r). \tag{4.2}$$

These pixels are presumed to be open in the sense that the indium bump bond does not connect the silicon substrate to the ROIC. Their spatial distribution over the detector is not uniform, and it has been mapped to a set of suspected opens by the manufacturer: Teledyne Scientific and Imaging. The fraction of open pixels on each of the HyViSI detectors tested are listed in Table 4.2.

Figure 4.2 illustrates how the open pixels appear in a typical image. The image, generated by subtracting the first read from the last read of a flat field exposure, shows isolated black squares and black squares surrounded by neighbors with enhanced signal. We associate the former with hot pixels since we have verified that their signal quickly reaches saturation such that they appear dark in a difference image. The open pixels are the black ones with bright neighbors. They see a very reduced rate of signal increase relative to *both* the background pixels and neighbors.

While the open pixels appear to be dead in the stretch shown in the figure, they do have increasing signal versus read number, but the slope may be anywhere in the range of 5-70% of the mean slope for regularly behaving pixels. Also, their neighbors have an elevated signal with respect to the regularly behaving pixels. This effect cannot be due to conventional interpixel capacitance because the open pixels have both a consistently low *raw voltage* relative to the neighbors *and* a low slope in voltage vs. time. Also, in dark exposures, they integrate a dark current that is consistent with the mean dark current on the array. The low raw signal and dark current make sense since the reset transistor can still communicate with the pixel in the multiplexer even if it is not connected to the detector material via the indium bump bond, and the pixel can still integrate leakage currents from the ROIC.

The most likely explanation for this behavior is that the p^+ implant of an open pixel in the detector PIN diode has some varying degree of impedance to the silicon in the multiplexer. The actual potential in the implant may be very high—perhaps close to V_{SUB}—so that lateral diffusion is taking place at the front surface of the detector (see Section 6.1.4.3) and causing the potential in the neighboring pixels to increase. But because of the high impedance, there is some voltage drop between the p^+ implant and the p^+ silicon in the multiplexer for the open pixels.

4.2.4 Volcanoes

In certain areas of the HyViSI detectors there are groups of hot pixels clustered together. When the voltage in these pixel reaches the upper rail, it appears that the charge spills over into the neighboring pixels. When these neighboring pixels have sufficient charge, the spilling proceeds to

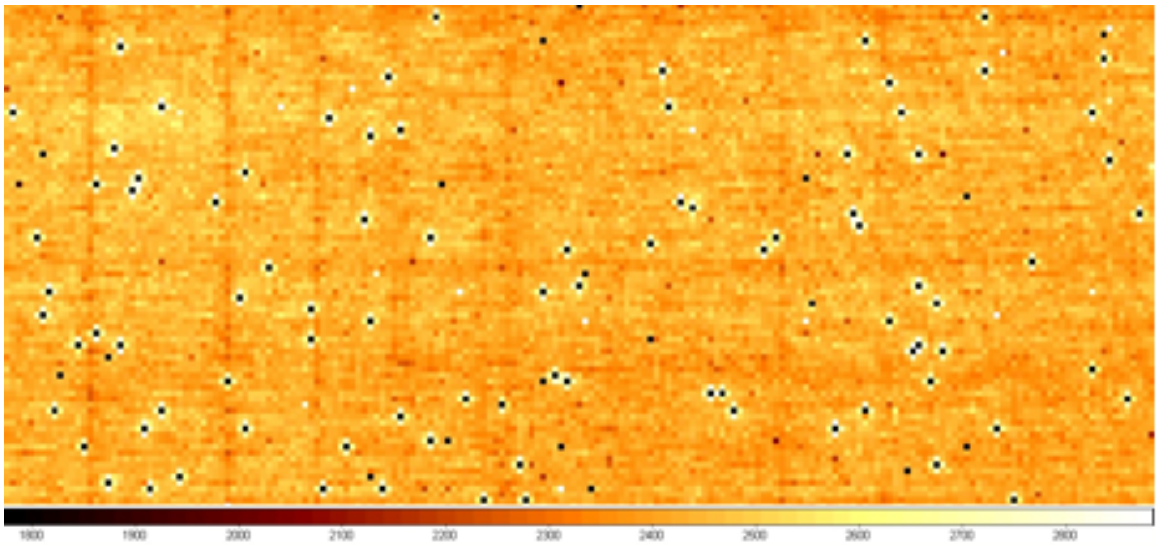


Figure 4.2: Difference image of portion of H4RG-10-007 array under flat field illumination at 1000 nm. Notice that there are two populations of dark pixels. One is surrounded by neighbors with normal response and another is surrounded by pixels with elevated apparent response. The former are identified as hot pixels which have a difference near zero and the latter are identified as open pixels.

their neighbors, and so on. For this reason, they have been dubbed "volcanoes". The pixels in these clusters are classified as either dead or hot, so their numbers are included in those categories.

4.2.5 Summary of Pixels

The summary of the unusable pixels for the principle detectors tested in this thesis is tabulated in Table 4.2. H4RG-10-007 has an inordinately large number of high dark current pixels in addition to a high mean pixel dark current. This problem was linked to an anti-blooming diode in the multiplexer and has been removed in the new pixel architecture (Yibin Bai, Private Communication).

Table 4.2: Pixel Type Fractions for HyViSI Detectors Tested

Detector	Dead		Open		Hot		Total	
	#	%	#	%	#	%	#	%
H4RG-10-007	6341	0.0378%	76,959	0.4587%	210,063	1.2520%	293,363	1.7486%
H2RG-32-147	197	0.0047%	293	0.0070%	1528	0.0364%	2018	0.0481%
H1RG-022	7	0.0007%	52	0.0050%	44	0.0042%	103	0.0098%

4.3 Conversion Gain and Nodal Capacitance

One of the most important properties of any detector is the net conversion gain between the digital number (DN or ADU) recorded for a pixel and the amount of charge present in the collecting node of that pixel.[1] We shall call this gain G_{NET} since it represents the product of all gain stages in the signal path between the physical charge in the pixel and the DN recorded in the DAQ. Nearly all of the other detector properties such as read noise, quantum efficiency (QE), etc. rely on this measurement. In addition, when combined with the QE, it can provide an estimate of the apparent magnitude of astronomical sources when photometric standards are not available.

4.3.1 Contributions to the Gain

G_{NET}, measured in e^-/ADU, is the product of several gain stages:

$$G_{NET} = G_{PIXEL} * G_{UC} * G_{OUT} * G_{AMP} * G_{A/D} \tag{4.3}$$

The pixel gain, G_{PIXEL} (e^-/V), accounts for the voltage change per unit charge, also known as the inverse of the capacitance. It is linear over small signal ranges but becomes nonlinear when the pixel is near capacity. The detector readout has two source follower FETs between each pixel and the output pad. One is in each unit cell, and it induces a gain of G_{UC} (V/V). The other, the output FET, introduces a similar gain, referred to as G_{OUT} (V/V). The output FET may or may not be included in the signal path, but for this discussion it is assumed to be. The processing electronics have stages to amplify the signal, and this amplification is included in the term G_{AMP} (V/V). Finally, $G_{A/D}$ (V/ADU), represents the conversion between volts and analog to digital units (ADUs). In the case of the SIDECAR and ARC electronics, we can express the product of the latter two gains as the electronics gain, $G_{ELEC} = G_{AMP} * G_{A/D}$.

In the following sections we shall describe each of these gains in a little more detail, along with the methods by which they are measured.

4.3.2 Electronics Gain – G_{AMP} & $G_{A/D}$

Measuring the conversion gain of the SIDECAR ASIC control electronics was described in Section 3.5.2. The same technique, applicable to any A/D converter and amplification stage, was also applied to the ARC controller.

Both sets of electronics allow the signal to bypass the amplification stage and go directly to the A/D. Making the measurement in this configuration will yield $G_{A/D}$ (V/ADU). Including the amplification stage and repeating the measurement will yield $G_{ELEC} = G_{AMP} * G_{A/D}$, from which

[1] The term *node* may be used interchangeably with pixel in this case.

G_{AMP} (V/V) can be inferred. We did not bypass the preamp of the SIDECAR while collecting science data because it aids in filtering and buffering the signal to the A/D.

4.3.3 Unit Cell Source Follower Gain – G_{UC}

Each unit cell of the HxRG multiplexers contains a FET source follower (SF) that buffers the voltage present at its gate to the detector output. In the case of the HyViSI devices, this voltage, V_{node}, is generated by the charge contained in the p^+ implant of the photodiode. The voltage at the source of the FET will follow the voltage at the gate amplified by a gain, G_{UC}, that is less than unity. The gain of the FETs is sometimes referred to as the *electronic gain* of the detector.

As shown in Figure 4.3, the gate of the unit cell SF can be held at the potential V_{reset} while the reset switch in the pixel is closed. If the output FET at the right of the figure is bypassed by closing the switch controlled by BUFDISABLE, we can directly measure a change in voltage at the output, $\Delta V_{OUT\ NOSF}$, induced by a change of ΔV_{RESET}. The two are related by:

$$\Delta V_{OUT\ NOSF} = \Delta V_{RESET} * G_{UC} * G_{ELEC}. \tag{4.4}$$

To obtain G_{UC} we program a set of voltages for V_{RESET} using the control electronics and measure the corresponding values of $V_{OUT\ NOSF}$. When we plot the quantities, the slope gives us G_{UC}.

4.3.4 Output Source Follower Gain – G_{OUT}

The method for measuring G_{OUT} is nearly identical to the one described for measuring G_{UC} except that we close the switch controlled by BUFDISABLE, thereby placing the output SF in the signal

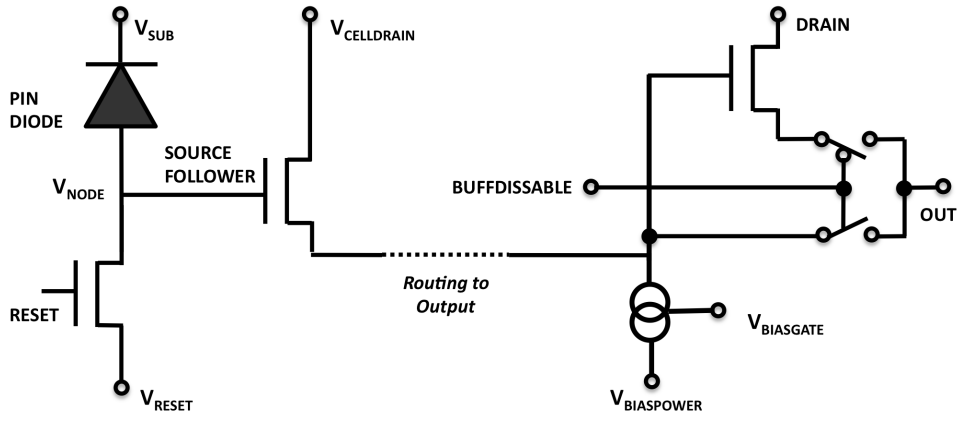


Figure 4.3: A diagram showing the unit cell, control lines, and routing at the output of the detector. Two source followers are present in the signal path with gains G_{UC} and G_{OUT}.

path. With both FETs contributing gain, Equation 4.4 becomes

$$\Delta V_{OUT\ SF} = \Delta V_{RESET} * G_{UC} * G_{SF} * G_{ELEC}. \tag{4.5}$$

Since G_{UC} is known, we can solve for G_{SF}. In Figure 4.4, we show the DN (given by $DN = V_{OUT\ SF}/G_{ELEC}$) vs. V_{RESET} for H1RG-022. This data was taken with both FETs active so the slope of the line is $G_{UC} * G_{SF}$. It should be noted that both of these gains are heavily dependent on the current available to their respective FETs. The gain and linearity of the pixel source follower will increase with decreasing $V_{BIASGATE}$, since decreasing this voltage increases the drain current of the unit cell FET. For the output FETs, these quantities will depend similarly on the current supplied to them through either a pull-up resistor or current source external to the detector.

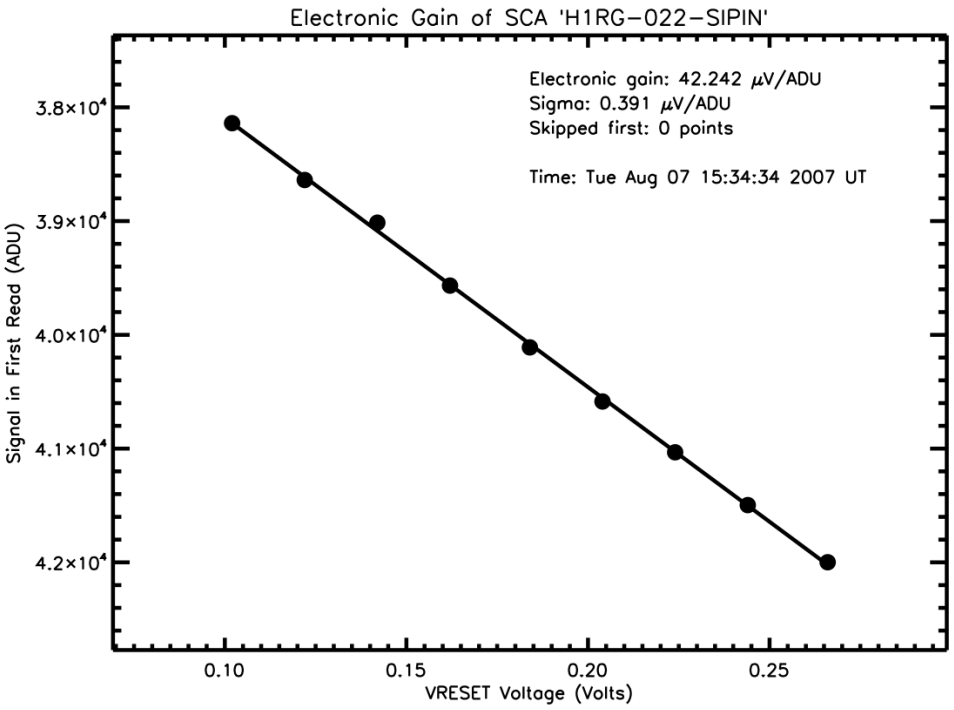


Figure 4.4: DN vs. V_{RESET} measured while the pixel reset is held down. The slope of this line is used to determine the gain of the unit cell SF and the output SF.

4.3.5 Net Conversion Gain – G_{NET}

Several methods can be used to obtain the number of electrons needed to change the ADU count by one. They include noise squared vs. signal (photon transfer), Fe^{55} calibration, and the capacitance comparison method. For Hybrid CMOS arrays, it has been shown that the noise squared vs. signal method overestimates the nodal capacitance and, in turn, the conversion gain [52]. To avoid this, we use the Fe^{55} method to estimate the conversion gain and the average pixel capacitance.

4.3.5.1 Fe^{55} Test

Fe^{55} is a radioactive material that is commonly used in astronomy to calibrate detectors in both space and ground based missions. An Fe^{55} atom produces soft x-ray photons when it decays into an Mn atom, and the energy spectrum of these x-rays and their interaction behavior in silicon is well known [3]. The five most prominent emission lines and the number of electrons they will produce upon interacting in a silicon substrate (assuming that 3.65 e^-/eV are produced) are listed in Table 4.3.

Table 4.3: Fe^{55} lines along with the number of electrons they generate in Si. The process that generates the photon, where $\rightarrow$ indicates an electron moving from one orbital shell to another, is also indicated.

Line	Energy (keV)	Number of e^-	Process
K_α	5.9	1620	L $\rightarrow$ K Shell with Auger Process
K_α escape	4.2	1133	L $\rightarrow$ K Shell
K_β	6.5	1778	M $\rightarrow$ K Shell with Auger Process
K_β escape	4.8	1291	M $\rightarrow$ K Shell
Si	1.7	487	Photon Escapes

The K_α line is significantly stronger than the other lines. It is emitted 7 times more frequently than the K_β line, which is the next in order of strength. The Fe^{55} test is sometimes referred to as the "acid test" for imagers [11] because a detector that has poor charge collection efficiency (CCE), read noise, CTE, QE, or any combination of these will not be able to distinguish between these separate peaks. Instead, the distribution of counts due to Fe^{55} hits will appear broadened spectrum that peaks around the K_α line.

Fe^{55} sources are fairly easy to come by, small in size, and relatively inexpensive, making them a good calibration tool for astronomy. One position in the filter wheel contained inside the RIDL dewar held an Fe^{55} source to calibrate each of the HyViSI devices. With the source available at any time, we were able to use it before each telescope observing run and verify the conversion gain of the detector being used. In the dewar used at Teledyne Scientific, the source had to be inserted

and removed by hand, but resulted in approximately the same configuration as the one in the RIDL system.

To measure the conversion gain with the Fe^{55} source we obtain a large number of exposures while the Fe^{55} source is located an inch above the detector. In collecting the data, we switch between CDS window mode with a small window size and full frame up the ramp mode. The former method better avoids double hits in a single read, which makes data analysis easier. It also provides better time resolution on the signal in each pixel. However, it severely diminishes the collection area, and thus, the effective hit rate in comparison to full frame mode.

Once all the data have been collected, they are analyzed to find single pixel events according to the method described in Section 6.1.3. We histogram the data and assume the peak in ADU corresponds to 1620 e^-. An example histogram for H2RG-32-147 is shown in Figure 4.5.

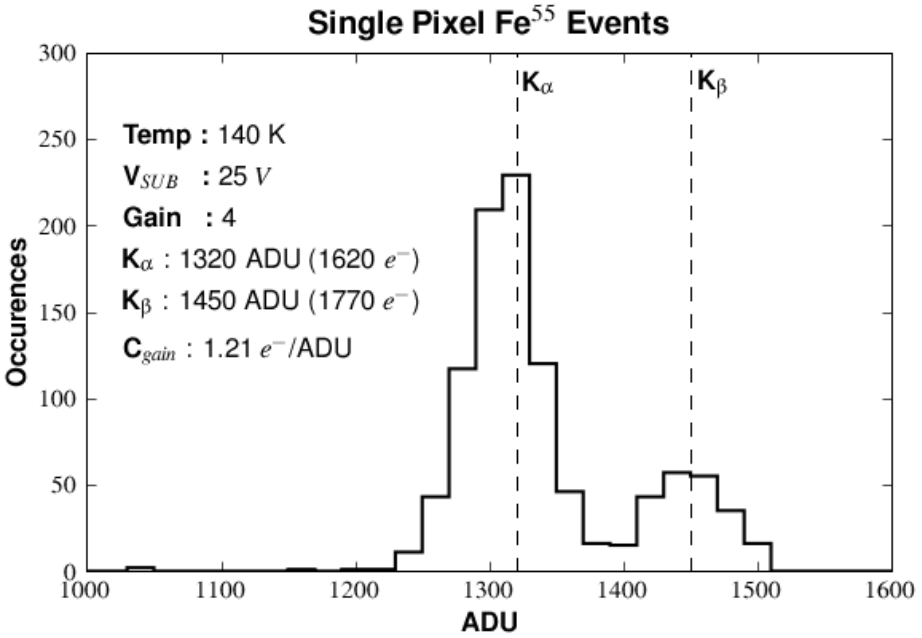


Figure 4.5: Histogram of single pixel events while the detector H2RG-001 was exposed to the Iron 55 source. The data was collected in window mode with a frame time of 1.6 seconds. Operating parameters are listed. The K_α and K_β peaks are clearly discernible, indicating good energy resolution.

4.3.6 Nodal Capacitance – G_{PIXEL}

G_{PIXEL} is effectively the nodal capacitance, C_{NODE}. It is a very important quantity that is often a steering factor in the fabrication of the detector. Unfortunately, it is a double edged sword in a way. If C_{NODE} is too small the pixel will not be able to hold much charge, and this will result in a small

dynamic range. In other words, the pixels will saturate very quickly. Due to the dependence on the area of the pixel, nodal capacitances are getting smaller as advances in CMOS processing yield smaller pixel dimensions. If C_{NODE} is too large, on the other hand, an electron added to the pixel will change the voltage of the node by a negligible amount. It will thus take a significant amount of electrons to increase the DN by one, resulting in poor sensitivity. Ultimately, a compromise is sought that will result in both good dynamic range and sensitivity.

With all of the quantities G_{NET}, G_{UC}, G_{SF}, $G_{A/D}$, and G_{AMP} in hand, Equation 4.3 can be inverted to find G_{PIXEL}. What we actually calculate is not just the capacitance of the PIN photodiode, but rather the sum of capacitances referred to as C_{total} in Section 2.2.2.

4.3.7 Results for HyViSI Detectors

Table 4.4 shows an example of the individually measured gains for H4RG-10-007. The product $G_{UC} * G_{OUT}$ was found to be very low relative to the value of 0.90 that is typically measured in hybrid CMOS arrays. For the other HyViSI devices, only G_{NET} was measured specifically. Values for G_{PIXEL} of these devices were estimated using $G_{UC} * G_{OUT}$=0.9 and are listed along with the well depth in Section 4.7.

Table 4.4: Measured gains using the ARC electronics and the H4RG-10-007 Si PIN detector

H4RG LEACH SFE					
G_{NET}	G_{AMP}	$G_{A/D}$	$G_{UC} * G_{OUT}$	$1/G_{PIXEL}$	Unit Cell Capacitance
(e^-/ADU)	(V/V)	(μV/ADU)	(V/V)	(μV/e^-)	(fF)
2.32	1.81	42.97	0.736	25.21	6.347
0.63	6.62	11.08	0.725	24.45	6.544

4.3.7.1 Dependence of Conversion Gain on Temperature

The drain current of the pixel source follower decreases with decreasing temperature if the voltage to its current source, $V_{BIASGATE}$, is held constant. This decrease in current will translate into a loss of gain–less μV/e^- and thus, less ADU/e^-–for the pixel, as well as a decreased transconductance. The impact of the decreased transconductance on noise will be discussed later in Section 4.4.1.2. The impact on the conversion gain G_{NET} is shown in Figure 4.6. As expected, the conversion gain decreases with decreasing temperature. To prevent this from happening, the voltage $V_{BIASGATE}$ must be adjusted accordingly.

Figure 4.6 also shows an unexpected and not yet understood effect that was observed several times during temperature cycling and testing of H2RG-001. On four separate occasions, a very large

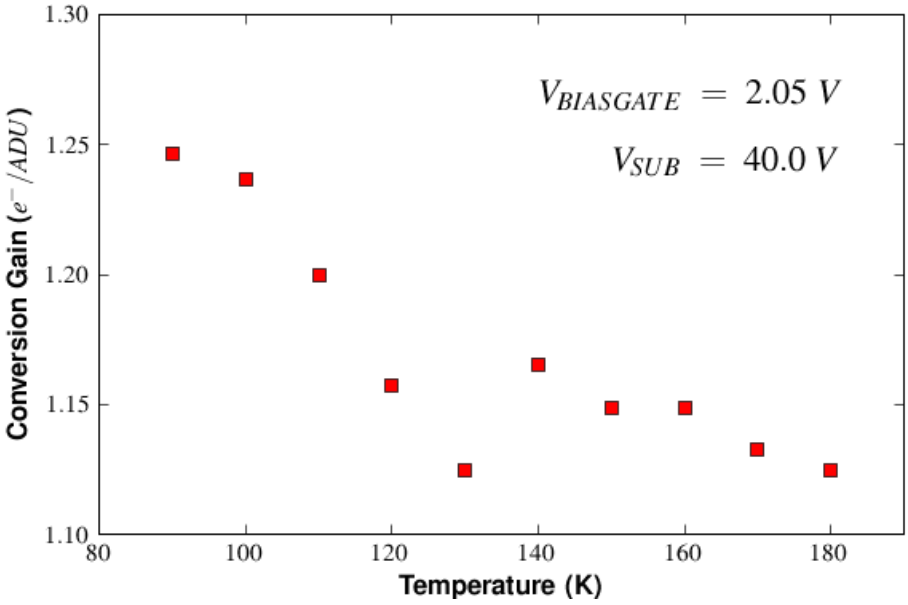


Figure 4.6: With the voltage to the pixel source follower current source $V_{BIASGATE}$ held constant, the net conversion gain is seen to decrease with decreasing temperature. The discontinuity at 140 K is related to an unexplained shift in signal for all detector outputs that occurred several times when using H2RG-001.

voltage spike was seen after the detector was warmed to 130 K or 140 K. In each case, the voltage spike occurred during up-the-ramp integrations and lasted less than 10 seconds. After the spike, the average output signal increased by about 120 mV for all science pixels on the array and by about 100 mV for the reference pixels. A decreased read noise was observed after the spike, along with a rise in G_{NET}. The effect was only observed when the substrate voltage was greater than 10 volts, which might seem to suggest that it is related to a shift in the equilibrium state of carriers in the bulk. However, because it was also observed in the reference pixels, which do not see the substrate voltage, this cannot be the case. The most plausible culprit seems to be the protection diode circuitry for V_{SUB}. It is not known whether this circuitry is located in the cryogenic SIDECAR readout electronics or on the pads of the readout multiplexer.

4.4 Read Noise

Any image produced by a detector can be divided into two components: *signal* and *noise*. The signal is the part that we are interested in; it represents the imprint of the incoming light on the detector. The noise is the part that we wish to get rid of or minimize. Some of the noise is inherent in the light itself and physically inevitable. Photons have a shot noise associated with them, which means the more photons that fall on the detector in a given time, the larger the spread in their number will be. The other portion of noise is related to the way in which the photons are converted into a measurable signal, and the electronics used to detect the signal, including everything from the uncertainty of charge on a capacitor to fluctuations in the number of electrons that actually make it through the drain of a transistor. This is the portion that we seek to eliminate with improvements in detector technology. In astronomy, a low read noise is absolutely critical for faint source detection.

4.4.1 Sources of Read Noise in Hybrid CMOS Detectors

Noise sources in CMOS detectors have been studied and modeled extensively. It is beyond the scope of this thesis to treat these sources in detail. However, basic descriptions of the most dominant noise sources in Hybrid CMOS detectors are given in the following sections and the reader is pointed to the references that provide exhaustive detail.

4.4.1.1 kTC Noise

At the simplest level, resetting the pixels in a hybrid CMOS detector can be viewed as filling or draining a capacitor of charge through a resistive path [66]. The resistive path is the channel of the reset MOSFET transistor and the capacitor is the capacitance of the pixel node, as shown in Figure 4.7. During reset, the transistor is in its "On" state and current flows through the channel against

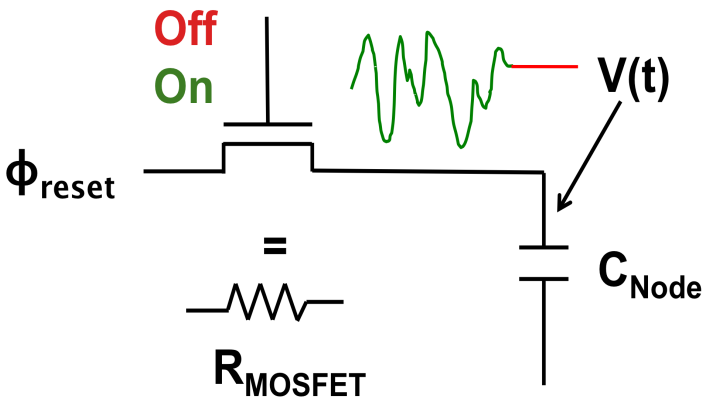


Figure 4.7: A very simple illustration of the process that generates reset noise in CMOS pixels. While the reset switch is closed, thermal Johnson noise causes the voltage V_{Node} to oscillate, as shown by the green line. When it is opened, the impedance of the switch is ideally high enough to prevent any current from flowing, leaving the voltage at the level shown by the red line.

an effective resistance R_{MOSFET}. The voltage, and thus the charge, at the pixel node fluctuates due to thermal Johnson noise inherent in the current flow. When the switch is finally opened and the transistor is turned "Off", the high impedance of the channel fixes the node voltage at whatever level it was at during the oscillations, ϕ_{reset}. And since this circuit is essentially a low pass filter with an equivalent noise bandwidth of $B = 1/(4R_{MOSFET}C_{Node})$, the RMS of the distribution of voltage and charge that will be measured in a large sample can be shown to be [3]

$$\sigma_{kTC}(V) = \sqrt{\frac{kT}{C_{Node}}} \qquad \sigma_{kTC}(e^-) = \sqrt{\frac{kTC_{Node}}{q}}, \tag{4.6}$$

where k is Boltzmann's constant.

With a capacitance of $C_{Node} = 14f$F, the reset noise is very substantial in HyViSI detectors. At 180 K, it is about 40 e^- and at 130 K, it is about 33 e^-. At a level of 30+ e^-, it may well dominate all other sources. Removing it with a correlated double sample (CDS) or any one of the methods described below is essential in scientific applications.

4.4.1.2 Source Follower Noise

Similar to the output amplifier in a CCD, the transistors in the CMOS multiplexer contribute noise to the signal measurement during a read. Janesick provides an excellent, thorough description of the noise sources associated with MOSFETs: white noise, flicker or $1/f$ noise, shot noise, contact noise, and popcorn noise [3]. While all the FETs in the multiplexer may contribute noise, several authors point out that the main noise contributer to read noise in the CMOS signal path is the pixel source follower [67, 68, 69]. The spectrum of the pixel source follower noise is generally dominated by a "white" and "pink" component.

White Noise White noise is random and has a flat power spectral density. The exact expressions in the references [67, 68, 69] for the white thermal noise voltage at the source of the transistor varies depending on the geometry considered, but they share in common the form:

$$\bar{V_n^2} \propto \frac{kT}{Cg_m}, \tag{4.7}$$

where k is Boltzmann's Constant, T is the temperature, C is a term representing the effective capacitance of the regions downstream of the transistor source, and g_m is the transconductance of the transistor. As Moore points out [43], since the transconductance goes like the square root of the drain current, a high drain current should be used for the pixel source follower to achieve low noise. For the HxRG multiplexers, the drain current is controlled by the bias voltage $V_{BIASGATE}$. It is indeed found that by lowering this voltage (which increases the drain current) the signal amplitude increases and the noise decreases. And since the current decreases with decreasing temperature, it

is important to adjust $V_{BIASGATE}$ when changing the detector temperature.

Pink Noise Pink noise is characterized by a power spectral density that grows with decreasing frequency f, with a dependence $\propto 1/f$. It is often called "one over f noise" or "flicker noise". For MOSFETS, its origin is believed to lie in the trapping of current carriers while they flow through the transistor channel [3]. When large numbers of electrons are trapped and de-trapped, the current is modulated, resulting in a noise that shows the $1/f$ shape. Large area devices are well described by classical $1/f$ noise models that assume large numbers of carriers, but for smaller devices, these models break down because the number of mobile charge carriers is small and behavior of individual charge carriers becomes visible and significant [70]. In this *Random Telegraph Signal (RTS)* regime, the presence of even one individual trap may be observed as noise that looks like a toggling between an "on" and "off" state in the output signal. Expressions for $1/f$ and RTS noise are given in [3, 14, 70].

4.4.1.3 Dark Current and Photon Shot Noise

It is well known that both dark current and photons exhibit *Poisson* noise, commonly referred to as shot noise. Simply stated in the context of detectors, if the dark current or photo current produces an average of N electrons in a time t, then the variance in the number of electrons produced will also be N in that same time for a set of measurements. Consequently, the noise will be higher for high luminosity measurements, where $N_{photons}$ is high, and high temperatures, where N_{dark} is high.

4.4.1.4 Bias Coupling Noise

Noise in the voltages and currents used to bias the detector can couple to the pixel node and output bus, creating another source of noise in measurements of pixel values. For instance, an oscillating $V_{BIASGATE}$ will cause the drain current in the pixel source followers to shift as well, which translates to an oscillating output voltage, independent of what pixel is being sampled. Moore finds evidence that the row enable FET and reset FET couple to the pixel node voltage [43]. Noise in either of these transistors will translate into noise in the pixel.

4.4.1.5 Output Crosstalk Noise

When multiple outputs are being used on the HyViSI detectors, a large signal on one output will couple to the other outputs. Evidence for this will be shown in Section 6.2. In addition to systematically raising the signal on the other outputs, noise on the high voltage being transmitted as well as the shot noise on the electrical current that carries the high signal will present itself on the other outputs.

4.4.2 Noise Reduction Techniques

The HxRG reference pixels and separate reference output can both be used to eliminate common mode noise. The application of the reference pixels, which were described in Section 3.4, are discussed later in Section 5.2.2 and the separate reference output is covered below. Aside from these, the unit cell and readout multiplexer of the HyViSI possess no circuitry to remove noise in the analog domain. However, in the digital domain, correlated kTC reset noise can be eliminated and uncorrelated noise can be diminished by reading the pixel multiple times and properly manipulating the values. Techniques to manipulate the pixel values are well treated in Fowler et al. [71]. The three most common of these: Correlated Double Sampling, Fowler Sampling, and Slope Fitting, are discussed below. All of the techniques described are hinged on the non-destructive readout of the hybrid pixels.

4.4.2.1 Common Mode Referencing

Perhaps the single most useful signal delivered by the multiplexer is the reference output V_{REFOUT}. This is an independent output channel derived from a single pixel, which is connected to either D_{SUB} or V_{RESET}, and is read out in parallel with the other pixels. Any common mode noise introduced by the power supply, picked up as interference, etc. can be eliminated by using this voltage as the reference for the video outputs. As a demonstration of its importance, with all other aspects of our test system configuration being equal at $T = 100$ K, the RMS read noise drops from 40 e^- CDS when referencing the detector outputs against $V_{REFMAIN}$ to 10 e^- CDS when referencing against V_{REFOUT}.

Great care should be taken to ensure that $REFOUT$ is wired correctly to the control electronics for differential analog measurement. For instance, if it is wired to inputs $InP32 - InP36$ on the SIDECAR, one is forced to sample both it and the video signal against an internal SIDECAR reference, and subtract the two signals digitally. Since the electronics noise in the two channels is uncorrelated, an additional noise factor of $\sqrt{2}$ will be introduced. On the other hand, if V_{REFOUT} is wired to $InPCommon$, the input routing multiplexer can be used to feed it to the negative side of the preamp in every channel, allowing a truly analog differential measurement.

4.4.2.2 Correlated Double Sampling

Correlated Double Sampling (CDS) is the most easily understood and implemented multiple sampling technique. CDS is a widespread technique in astronomical imaging, and it should be emphasized that the CDS technique described here is a *digital* one and not an *analog* CDS like the one used in the output amplifier of a CCD. To obtain a digital CDS, after resetting a pixel at i, j, the pixel is read once at time t_1 and then again at $t_1 + \Delta t$, yielding the values $S(i, j, t = t_1)$ and

$S(i, j, t = t_1 + \Delta t)$. The signal is then calculated as:

$$S(i,j) = \frac{S(i,j,t=t_1+\Delta t) - S(i,j,t=t_1)}{\Delta t}. \tag{4.8}$$

Note that no restriction is placed on the sampling pattern. The pixel at i, j can be read twice in succession before clocking to the next pixel or all pixels in the frame can be read before returning to i, j; the difference is accounted for by Δt. Figure 4.8 can be used to visualize the second of these sequences if $N_{Fowler} = 1$ and $t_{FowlerExp} = \Delta t$.

CDS eliminates kTC noise and has been speculated to improve spatial uniformity [71]. It is straightforward to deduce that the $1/f$ component noise component will depend on the integration time. If the detector is read noise limited, CDS will be bested by a technique that uses more than two samples. But interestingly, if the detector is background limited[2], CDS will yield the best estimate of the signal. Garnett and Forrest provide an elegant proof of this in [72].

4.4.2.3 Fowler Sampling

Multiple Correlated sampling, or Fowler sampling, named for its pioneer Al Fowler, is a technique that was first implemented in infrared arrays to reduce read noise [71]. In a Fowler sampling sequence, N_{Fowler} reads of the detector are performed immediately after the pixels are reset. Then, after some integration time $\Delta t = t_{FowlerExp} + t_{FowlerPair}$, N_{Fowler} more reads are taken. The signal estimate is given by:

$$S(i,j) = \frac{\sum_{r=N_{Fowler}+1}^{2N_{Fowler}} S(i,j,r) - \sum_{r=1}^{N_{Fowler}} S(i,j,r)}{N_{Fowler}\Delta t}. \tag{4.9}$$

Figure 4.8 shows a Fowler sequence with $N_{Fowler} = 5$. Often times, such a sequence is referred to as *Fowler 5*, or Fowler sampling with *5 Fowler Pairs*, since each of the reads in the first set is matched with another in the second.

If the noise in each read σ_r is white, then the effective noise for the signal estimate given in Equation 4.9 will be [73]:

$$\sigma_t = \sqrt{\frac{2}{N_{Fowler}}}\sigma_r. \tag{4.10}$$

This equation can also be applied to the CDS case, where $N_{Fowler} = 1$. Garnett and Forrest show that Fowler Sampling achieves its best performance at a duty cycle of 2/3, meaning that 2/3 of the total observing time is spent sampling the pixels [72]. In the context of Figure 4.8 this means $t_{FowlerPair} = t_{FowlerExp}$. Even with the optimum duty cycle, though, Fowler Sampling yields a slightly worse signal-to-noise (SNR) ratio then Slope Fitting (by about ∼6%). Also, if used in a

[2]Signal to noise calculations are often divided into two regimes. *Read-noise limited* is the case where the detector noise is dominated by the readout process. *Background limited* or *shot-noise limited* is when the detector noise is dominated by photon shot noise.

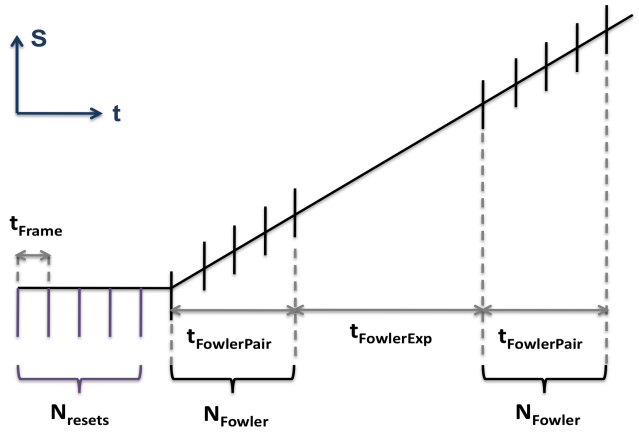


Figure 4.8: A diagram showing an up the ramp exposure with Fowler Sampling. Before the exposure, $N_{resets} = 5$ resets are performed. At the front of the integration, $N_{Fowler} = 5$ samples are taken. After a wait time of $t_{FowlerExp}$, another 5 samples are taken. It should be noted that during the wait time, the detector should still be clocked to avoid thermal instability.

manner in which only the final value $S(i, j)$ is saved to disk, Fowler Sampling offers no means to reject cosmic ray events. In fact, Offenberg et al. [73] show that the SNR is reduced to zero at the locations hit by cosmic rays. If all of the samples are saved, however, an estimate for S can still be obtained by only considering certain reads.

It is worth mentioning that the real benefit of the Fowler method is the simplicity of the algorithm and the small amount of data storage required. The astute reader might also gather that a reduction in electrical power can be gained by leaving the detector idle during the period of time $t_{FowlerExp}$, and this would be correct for an ideal detector. However, when the HyViSI pixels are not clocked during this period, a large offset in signal is observed for the second set of reads even with no illumination. The sign and amplitude of the offset are not understood and the latter can vary greatly depending on the operating conditions. For this reason, the pixels must be clocked in the same fashion as they are during the sampling period and no power will be saved.

4.4.2.4 Slope Fitting

As an alternative to simply averaging reads and subtracting pairs, a straight line can be fit to the signal as a function of time. The slope of the line yields an estimate of the instantaneous photocurrent in the pixels, and thus the flux F. This method is referred to as *Slope Fitting* or *Sampling-Up-The-Ramp (SUR)* It reduces both $1/f$ and white noise, and is very useful in detecting large jumps or dips in signal due to cosmic rays, voltage spikes, and "noise bursts" [74]. It is computationally intensive, however, and difficult to implement in FPGAs and control electronics circuitry. As mentioned in the previous section, it does offer slightly better performance then Fowler Sampling in the read-noise limited case. If N equally spaced samples are used for the fit with equal weighting of each, then the effective noise will be [75]:

$$\sigma_t = \sqrt{\frac{12}{N}}\sigma_r. \tag{4.11}$$

Note that the duty cycle and integration time of a Fowler Sampling scheme must be taken into account when comparing Equations 4.11 and 4.10. The exact implementation of this method will be covered in Section 5.2.4 in the context of astronomical data reduction.

4.4.3 HyViSI Measurements

There are several standard practices for measuring read noise in hybrid detectors [76]. In the *temporal* method, a stack of CDS frames are collected and the RMS value for each pixel across the frames is computed to form a final two-dimensional image. The pixel values thus indicate how the signal in a given pixel varies over time, and the mean of the RMS values is the figure that is reported as an estimate of the variation. In the *spatial* method, also referred to as the *pixel-to-pixel* method, two CDS frames are subtracted from each other and the resultant frame is divided by $\sqrt{2}$ to account for the statistical increase in noise due to subtraction. The resulting **read noise map**, an example of which is shown on the left in Figure 4.9, is then binned into a histogram after significant outliers have been rejected. The histogram is then fit with a Gaussian to yield a standard deviation and RMS value for the distribution.

As mentioned previously, spatial read noise measurements for H1RG-022, H2RG-32-147, and H4RG-10-007 were dominated by noise from the reference voltage on the JADE card and showed excessively high noise (20-40 e^-). The spatial read noise values obtained for H2RG-001 were the best obtained and ranged from 7-13 e^- RMS. As Figure 4.9 shows, the noise floor is dominated by a pattern that shows up as strong banding and has a power spectrum characteristic of $1/f$ noise.

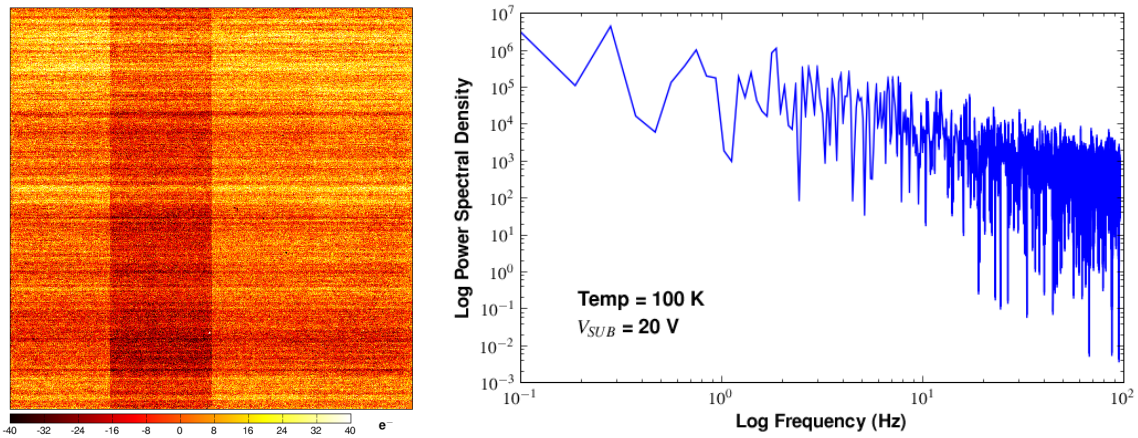


Figure 4.9: (Left) Read noise map of H2RG-001 at 100 Kelvin operated with four outputs. Each of the four outputs was averaged on eight channels in the SIDECAR, and the preamps in the SIDECAR were only reset once per frame. (Right) The power spectrum shows that the strong banding is dominated by $1/f$ noise, which sets the noise floor.

Note that the horizontal structure is not the same as the row-to-row reset noise of the SIDECAR preamps described in Section 3.5.1. If it were the latter, each of the four channels for the detector would have its own independent banding pattern. In fact, for these measurements the SIDECAR preamps were reset only once per frame and each of the four detector outputs were averaged on eight SIDECAR channels. Numerous attempts to reduce the $1/f$ banding with grounding measures and voltage and current adjustments on the SIDECAR were unsuccessful. However, after subtraction of the reference pixel columns, the banding is greatly reduced. And subtracting the mean of the reference pixels for each channel eliminates the channel to channel offsets, as shown in the read noise map of Figure 4.10.

Figure 4.10 also shows the measured CDS read noise for H2RG-001 as a function of temperature with the backside contact voltage at 25 volts. The measurements were made with the detector operating in window mode with a frame time of $t_{frame} = 10.6$s and a window size of 300x300 pixels. Interestingly, the noise decreases with increasing temperature until it hits a minimum at 130-140 K and then increases thereafter. One possible explanation for this is that the transconductance of the pixel source follower was decreasing for $T < 130K$ as a result of a constant $V_{BIASGATE}$. The constant $V_{BIASGATE}$ causes a decrease in the drain current, which in turn increases the noise. For temperatures above 140 K, the increase in noise is assumed to be a result of increased thermal noise, although the slope is not linear.

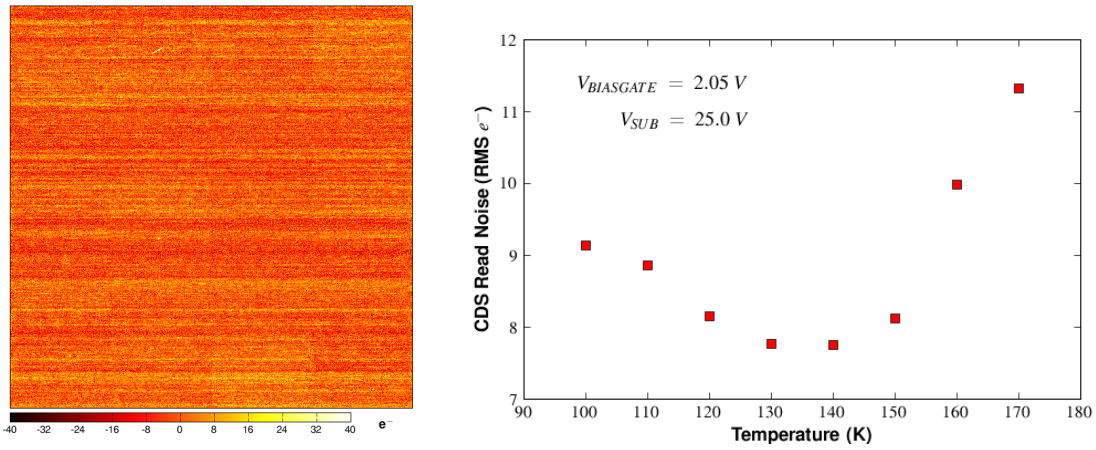


Figure 4.10: (Left) Read noise map of H2RG at 100 Kelvin operated with four outputs after reference pixel correction. (Right) Read noise vs. temperature for H2RG-001. The increase of read noise at low temperatures is due to the decrease in transconductance of the unit cell source follower and the increase at high temperatures is a result of increasing thermal white noise.

The read noise improves slightly at 40 volts, as shown in Figure 4.11, and there is evidence that it will continue to decrease with increasing V_{SUB} [40]. In addition to the temperature dependence for CDS ($N_{FP} = 1$) frames, the plot in Figure 4.11 shows the dependence of the noise on the number

of Fowler Pairs N_{FP} used to estimate the signal. By comparing with the dashed line, one can see that the noise does not follow the $1/\sqrt{N_{FP}}$ dependence expected from uncorrelated "white" noise. Dorn et al. observed a similar behavior [49]. In this case, the reason that the noise does not drop off like $1/\sqrt{N_{FP}}$ is that the "pink" noise $(1/f)$ maintains a non-zero amplitude even after reference pixel subtraction and averaging the multiple samples.

The noise measurements in Figure 4.11 represent the best noise performance obtained with H2RG-001. A limited amount of long, 100 read dark exposures were collected at temperatures between 100-140 K and a substrate voltage of 20 volts. For these data, the CDS noise is slightly larger (10-14 e^-), but the noise with 30 Fowler pairs reduces to about 2.5 e^-. In certain cases, the read noise bottoms out at this value, with further samples giving no further reduction. In others, the noise actually increases with more samples, most likely because of the $1/f$ contribution. Again, it should be emphasized that the noise is expected to decrease with decreasing temperature if $V_{BIASGATE}$ is lowered accordingly.

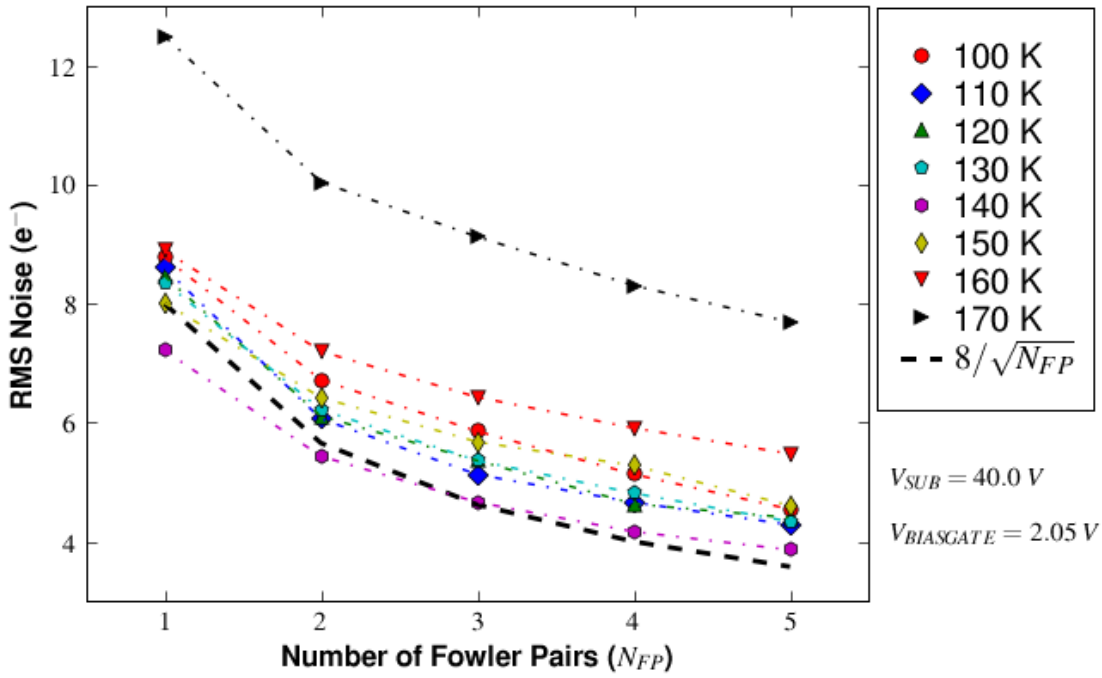


Figure 4.11: Read noise vs. number of Fowler Pairs for a 300x300 window with $t_{frame} = 1.6$ s on H2RG-001 at temperatures from 100-170 K. The substrate voltage $V_{SUB} = 40$ V was the highest tested and $V_{BIASGATE}$ was again held constant at 2.05 V. The noise does not fall off as $1/\sqrt{N_{FP}}$ (the black dashed line shows 8 $e^-/\sqrt{N_{FP}}$ for comparison) because of the presence of a $1/f$ contribution.

4.5 Dark Current

Dark current is perhaps the greatest bane of astronomical detectors. It decreases the dynamic range of the detector, increases the read noise, and limits exposure times. For practical purposes, any photocurrent signal that produces less charge per unit time than the charge imparted by dark current will be undetectable. In an ideal detector, dark current would not exist, or more accurately, would not make any contribution to the measured signal. Unfortunately, in all known detectors it is coupled with the signal one is trying to measure.

After a hybrid detector has been fabricated, the only "knobs" one has to control the dark current are temperature and bias voltages. Increasing the reset voltage will decrease the dark current in some cases, but doing so will also limit the dynamic range of the pixels. Cooling the detector will drastically reduce the dark current, however, this comes with the cost of decreased quantum efficiency. For this reason, every attempt is made in the design and fabrication of the detector to make the dark current as low as possible.

4.5.1 Sources of Dark Current

Dark current is the sum of any thermal leakage currents that will cause the detector to integrate charge even in the absence of light. These leakage currents can arise anywhere in the detector and even the readout circuitry may potentially make contributions to it (the ROIC in the HyViSI exhibits dark current when no detector has been bump bonded to it). The most prominent sources, however, are the detector bulk and the surfaces at the various interfaces.

Since a thorough treatment of dark current in semiconductor imaging arrays is beyond the scope of this thesis, several good references are worth noting. Janesick [3] provides an excellent treatise on dark current in CCDs, much of which is relevant for silicon hybrids. McCaughrean [41] presents a similarly well formed description of dark current in infrared hybrid detectors.

4.5.2 Estimating Dark Current

To find an estimate of the dark current over the whole detector at a particular temperature we first adjust the temperature inside the dewar and let it settle to equilibrium. Once settled, we block all light to the detector and take a series of multi-read exposures. The number of reads taken is varied to give results for a range of different exposure times (the cadence is also varied in some cases).

For a set of exposures which all consist of the same number of reads, we take the median value of each pixel over all the exposures in order to reject cosmic ray events and eliminate any spurious electrical signals that may occur. A slope is then fitted to each pixel, yielding an estimate of the number of electrons collected in the pixel due to thermally generated carriers vs. time.

A typical histogram for the final set of dark slopes for H1RG-022 is shown in Figure 4.12. The median, mean, and mode of the distribution all provide a different way of evaluating the dark current.

The **mean dark current** takes into account all of the outliers such as hot, open, and dead pixels. It therefore tends to be the highest of the three estimates. The **median dark current** rejects outliers, but still includes the contributions from regions of the detector that have a higher dark current than the majority of pixels. For instance, the pixels near the edge of H1RG-022 show a slightly higher dark current than the ones near its middle. The **mode of the dark current** gives an estimate for the most commonly occurring dark current, i.e. the peak of the distribution. One must be careful to specify which estimate is being used when comparing dark current values for a given detector. This is especially true in astronomy, where fractions of an electron per second per pixel are significant.

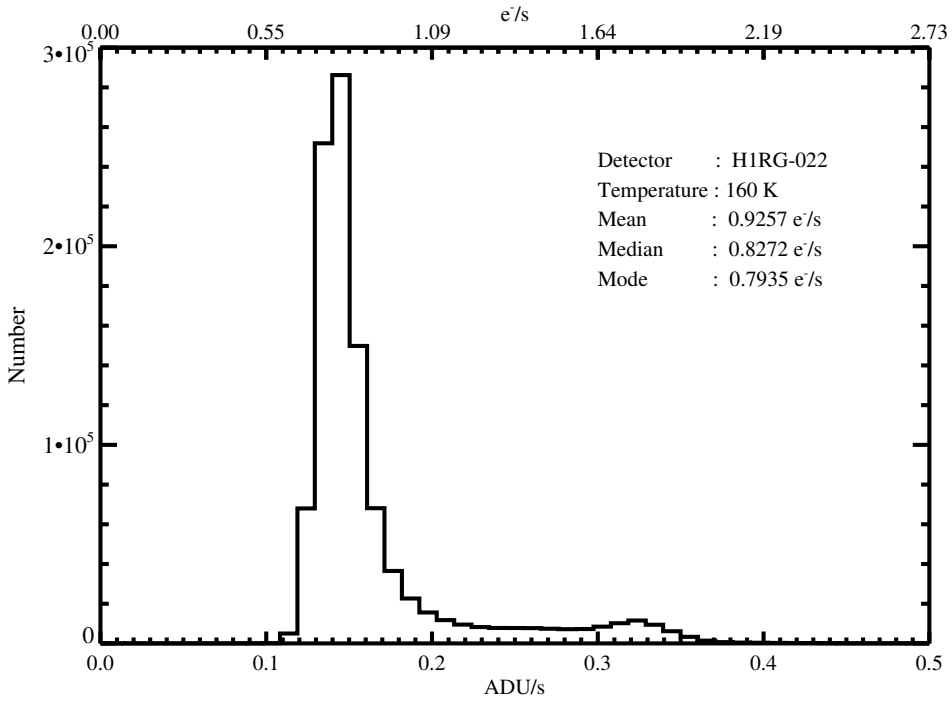


Figure 4.12: A representative dark current histogram generated from a dark exposure slope-fit for H1RG-022. The secondary peak around 0.32 e^-/s is due to slightly higher dark current around the periphery of the detector.

4.5.2.1 Units of Measurement

One often sees dark current expressed as the number of electrons generated per pixel per second. Because pixels can take any shape or size, a different unit must be used in order to compare different detectors: one that does not depend on the area or geometry of the pixel. For this reason, dark current is usually expressed in terms of electrical current per unit area: a *current density*.

In order to make the conversion between these two commonly used units, we can use the following equation for the dark current, DC:

$$DC(q/\mathrm{s/cm^2}) = \frac{q}{A} * DC(\mathrm{e^-/s/pix}), \tag{4.12}$$

where $q = 1.602 \times 10^{-19}$ Coulombs/e$^-$ and A is the area of the pixel expressed in cm^2. For the 18 μm pixels of the H2RG and H1RG detectors, the factor on the right hand side turns out to be about $q/A = 5 \times 10^{-14}$ (C·pix/e$^-$·cm^2). Thus, a dark current of 1 e$^-$/s/pix corresponds to roughly 50 fA/cm^2.

4.5.3 HyViSI Dark Currents

Dark current has been measured for several different HyViSI detectors. Figure 4.13 shows the dark current density measured for H1RG-018, H1RG-022, H2RG-32-147, and H4RG-10-007.[3] While the values below 160K are acceptable for many astronomical applications, the dark current is still rel-

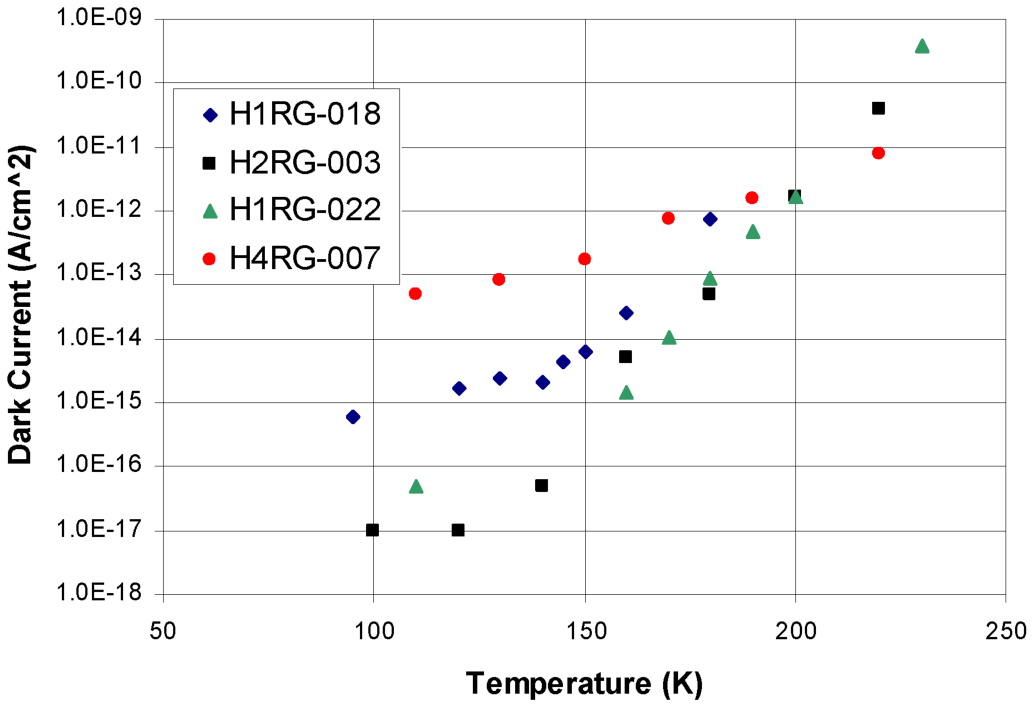


Figure 4.13: Dark current versus temperature for several HyViSI devices. H4RG-10-007 has 10 μm square pixels. All other devices have 18 μm square pixels.

[3]Results for H1RG-018 were obtained by Don Figer at the Independent Detector Testing Laboratory (IDTL).

atively high. For instance, the HyViSI detectors have to be cooled to 200K to achieve 1pA/cm^2, whereas this level of dark current is attainable at 60° C in CCDs [77]. The exact source of this dark current is not fully understood. Further tests, such as using identical detectors with different thicknesses, testing bare multiplexers without detector layers bonded to them at cryogenic temperatures, and comparing the results for devices with different treatments of the Si-SiO_2 frontside passivation (the processing techniques for each unique device are proprietary to TIS), should help pinpoint where it is generated.

4.5.4 Reset Anomaly in HyViSI

The term *reset anomaly* is used to describe an oddity that has been observed in most, if not all, infrared hybrid detectors. The effect shows itself as a large non-linearity in signal immediately following reset. The non-linearity can be fit with an exponential function that has a time constant ranging from seconds to hours, and although not fully understood, is usually attributed to RC charging effects somewhere in the detector or multiplexer [59].

HyViSI detectors show similar "anomalous" nonlinearities under certain operating conditions. One large nonlinearity is tied purely to a low-voltage pixel reset. We refer to this as the HyViSI reset anomaly. Another is tied to clocking inactivity in the array. If the clocks are stopped during an up-the-ramp integration or while the detector is idle, the pixels exhibit a drop in signal followed by a nonlinear return to the value before the clocking ceased. These two effects show similarities and can be easily mistaken for one another. As shown in the following sections, though, the nonlinearity induced by a low value of V_{RESET} and that caused by not clocking are two different phenomena. It should also be mentioned that additional nonlinearities arise from persistence and after forward biasing the photodiodes. These topics will be covered in a later chapter.

4.5.4.1 Nonlinearity After Reset

HyViSI detectors show a large nonlinearity in the early reads after the pixels are reset to a voltage V_{RESET} below about 150mV. The ramp for each pixel, an example of which is shown in Figure 4.14, can be fit with a 4 parameter function of the form:

$$S(i,j,t) = S_o(i,j) + A_{i,j}(1 - exp^{-t/\tau_{i,j}}) + B_{i,j}t \tag{4.13}$$

The parameters A, B, S_o, and τ vary greatly over the array. This is partly because the fit is not very good for most pixels; it shows a large ξ^2 value for the majority of pixels. $B_{i,j}$ should represent the dark current in equilibrium, but in most cases it overestimates the dark current by an order of magnitude. Nevertheless, the values A and τ are useful indicators of the voltage swing (or the equivalent change in pixel carrier density) and how long the effect takes to subside.

The low V_{RESET} reset anomaly only occurs in the science pixels. The reference pixels are not

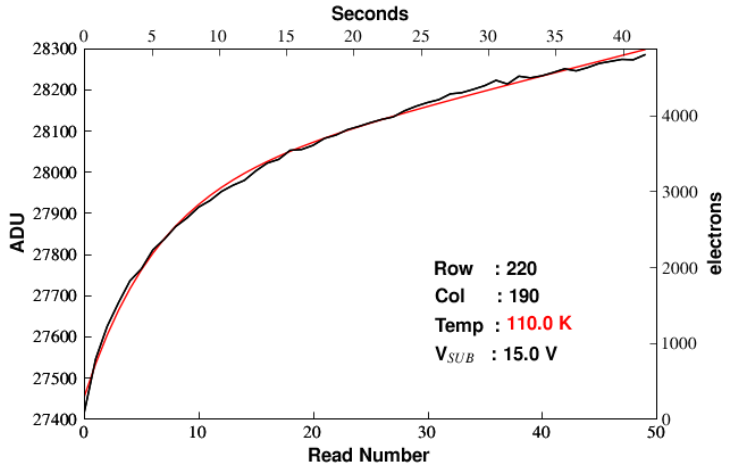


Figure 4.14: The black curve shows the signal of pixel (190, 220) in an up the ramp dark exposure. The non-linearity occurs for values of $V_{RESET} < 150mV$. The red curve is the fit from Equation 4.13 with $A = 442$, $B = 6.0$, $\tau = 5.3$, and $S_o = 27453$.

affected. Further, it shows a definite spatial correlation with the ROIC outputs, as can be seen in Figure 4.15. This suggests the effect occurs in the detector bulk or the interface between the detector and multiplexer. One possibility is that the setting V_{RESET} sufficiently low causes the n type surface above the SiO_2 layer to invert and collect holes during the actual reset. When the reset is finished and the pixels begin to integrate, these holes then diffuse back to the p^+ implants and cause the rise in signal. The spatial pattern would then be explained by variations in the surface potential as a function of distance away from the output buses.

4.5.4.2 Nonlinearity After Inactivity or Change in Mode of Operation

The HyViSI pixels show another unexpected, nonlinear behavior after periods of inactivity, after resets under certain operating conditions, or during a fast sequence of resets and reads of the detector. In any of these cases, the reference pixels and the science pixels exhibit the effect, indicating that it arises in the ROIC. It is therefore likely an electrical effect and not due to leakage currents, but it is included here because it is an effect that can easily be mistaken for elevated dark current.

By inactivity, we mean that the pixels are not being clocked. This may occur unintentionally between exposures because of a bug in assembly code or it may occur intentionally as part of an observing strategy. An example of the latter would be a pause in clocking during the wait period of a Fowler Sampling ramp designed to save power. The problem is that if the clocking of the pixels cease after some time t_1 and then start again at t_2 during a ramp, the pixel values will show a large drop between the two times. The drop ranges from 5-10 mV in the cases we have observed.

We also see the effect when we operate the detector in window mode and take a series of exposures without doing idle resets between the exposures. And it shows up strongly when the value of $V_{BIASGATE}$ is changed between exposures. The effect has a strong temperature dependence, which might be attributed to the dependence of the drain current in the unit cell source follower.

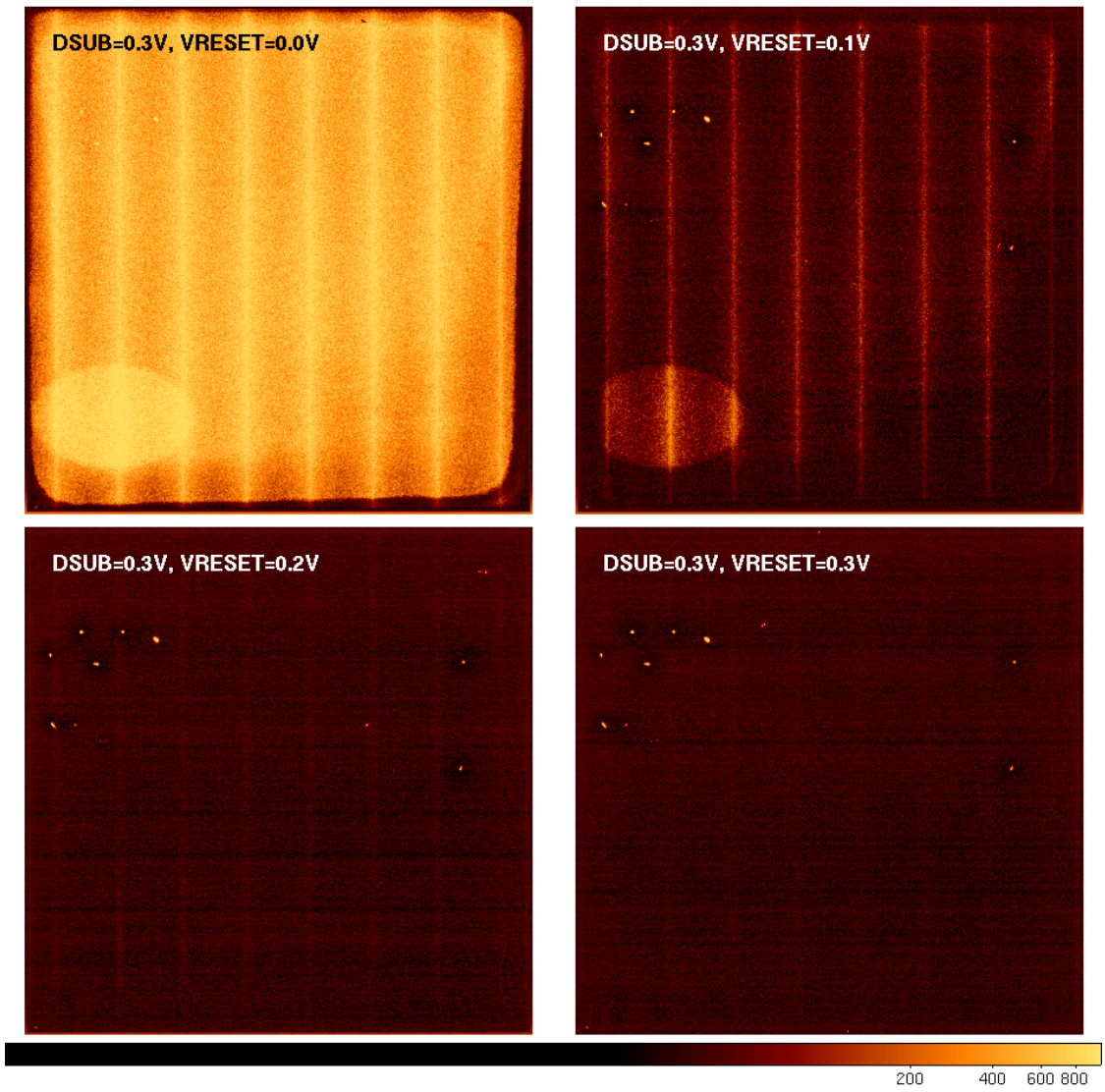


Figure 4.15: Images formed by subtracting the first read from the last read of 50 read ramps taken by H1RG-022 at 100K. A log scale is used for the stretch. The images clearly show that for values of V_{RESET} below 150 mV, the dark current signal is very large. They also show that a spatial pattern that suggests that pixel voltages are coupled to the output column buses.

4.6 Quantum Efficiency

There are several different definitions of Quantum Efficiency (QE) available in the literature. But for most intents and purposes, the QE of a detector can simply be thought of as the fraction of incoming photons that are converted to electron-hole pairs and collected as signal. When characterizing a detector, one measures the QE as a function of wavelength in a certain wavelength range, and this serves as a good indicator of how sensitive the detector is to these wavelengths.

4.6.1 PIN Diode Quantum Efficiency

The QE in the PIN diode detector layer of the HyViSI is determined by a number of different factors. Before the photons have a chance to be absorbed in the detector bulk, they must make it past the back surface of the detector. An anti-reflection coating is applied to this surface to minimize the fraction of photons that are reflected, R. Photons with wavelength λ that make it inside the detector then travel for some characteristic *absorption length*, $\alpha(\lambda)$, before they are converted into an electron-hole pair.[4] If we assume that the detector is fully depleted so that W_D is equal to the detector thickness, and that all photons absorbed in the depletion region are collected as signal, then we can express the QE as:

$$QE = (1 - R)\left[1 - \exp(-\alpha W_D)\right] \tag{4.14}$$

This equation, which is similar to the one in Sze [53] except that the diffusion of the minority carriers has been neglected, reveals the benefit of the thick $W_D = 100$ μm silicon in the HyViSI.

4.6.2 HyViSI Detective Quantum Efficiency (DQE)

DQE is the realized S/N compared to that of an ideal detector. It is often measured in the background-limited case so that it is most closely related to the photon capture process in the bulk material of the detector, as opposed to being related to read noise effects in the post-capture electronics. DQE can vary with wavelength, temperature, and individual pixel properties. To measure it, we illuminate the detector with a monochromatic flat field produced by an integrating sphere and monochromator. The light is monitored by a calibrated silicon photodiode located at a port on the integrating sphere. A similar calibrated diode is placed at the location of the detector in order to transfer the flux measured at the integrating sphere to the focal plane. Once this wavelength-dependent calibration is made, the detector is then placed at the focal plane and the experiment is repeated.

The results of our measurements for H4RG-10-007 and others obtained for H2RG-003 are shown in Figure 4.16. For QE measurements of an H2RG HyViSI below 800 nm, the reader is referred to

[4] $\alpha(\lambda)$ is defined as the depth in the material at which 1-1/e of the incident photons of wavelength λ have been absorbed.

Dorn et al. [78]. As can be seen in Figure 4.17, because the QE measurement involves illuminating the detector with a narrow wavelength range, interference fringes are observed in the illumination pattern. The fringes are strongest near 1 μm and show up similarly when a y band filter is used on broadband light. Removing them with "generic" flat fields is not trivial because the pattern depends on the angles of the incident rays. The best results are obtained when the flats are taken with a very similar illumination source to the one present during the science exposures (see Section 5.2.5).

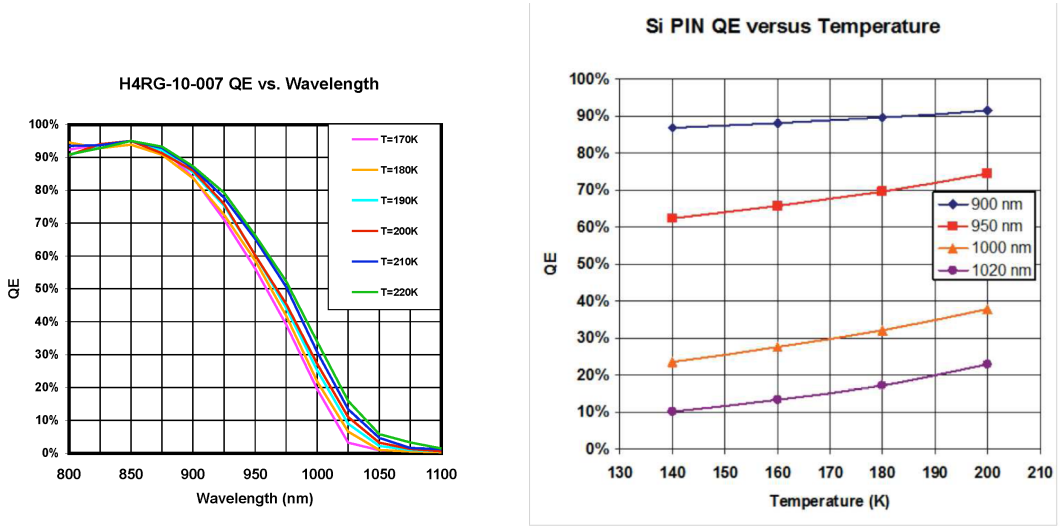


Figure 4.16: H4RG-10-007 relative QE versus wavelength (left) and H2RG-003 relative QE versus temperature (right) near the long wavelength cutoff. The results show that QE increases with temperature and are consistent with a silicon detector having 100 μm thickness.

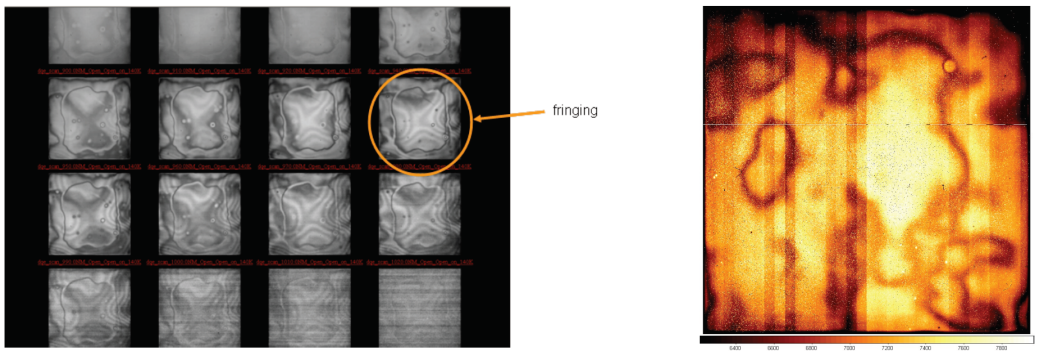


Figure 4.17: Fringing seen in monochromatic flat field images near 1 μm obtained with the H2RG-003 device (left). The fringing indicates thickness variation of a few microns. Fringing for the H4RG-10-007 device (right). The columnar striping is an artifact of electronic readout offsets and is not a QE variation of the detector.

4.7 Linearity/Well Depth

Linearity is one of the most important detector properties for astronomical purposes. Accurate photometry can only be performed if a detector has a stable, well-known response to varying brightnesses and integration times. While the term *linearity* implies the response should have a linear relationship to both of these quantities, this is often not the case. Infrared detectors, for instance, show a decreasing response in signal as they get closer to saturation: the nodal sensitivity actually decreases as the capacitance grows larger.

While it is generally possible to correct for nonlinear behavior by properly characterizing the pixel response, this adds one more step to the already elaborate data reduction process employed in astronomy. One would prefer to start with a detector that has an intrinsically linear response to light. And for this reason, linearity is usually listed as a standard detector specification along with the read noise, quantum efficiency, and dark current.

Linearity, or an equivalent nonlinearity, is defined in several different ways throughout the literature [41, 3, 79]. In the end, all of the definitions convey the deviation of an overall conversion gain G_{NET} (e^-/ADU or e^-/DN) from some average value. The deviation might arise as the signal integrates for a given brightness or it may occur at a given signal level when the brightness is varied. Usually what is quoted is the ratio of the deviation in G_{NET} to its average value, expressed as a percentage. A detailed linearity curve will show this deviation as a function of signal level, and for a number of flux levels.

Any one of the factors in Equation 4.3 can contribute to nonlinear behavior. Janesick divides nonlinearity into two distinct categories: V/V nonlinearity and V/e^- nonlinearity [79], based upon which of these gain factors is contributing. If the capacitance of the pixel G_{PIXEL} is changing, the nonlinearity is said to be V/e^-. If any of the other four gains, G_{UC}, G_{OUT}, G_{AMP}, or $G_{A/D}$ is changing, then the nonlinearity is said to be V/V. Usually V/V nonlinearity is attributed to the pixel source follower amplifier (G_{UC}), but for HxRG multiplexers, the output source follower (G_{OUT}) is also a suspect as it relies on an external current source.

For any detector that integrates photocharge on a pixel capacitance, the signal response will eventually become nonlinear as the pixel "well" nears its maximum capacity. The maximum capacity of the pixel, in units of electrons, is referred to as the *well depth*. Photoelectrons or holes generated above a pixel that has reached its well depth will either bloom into neighboring pixels or, for certain architectures, be dissipated by an anti-blooming diode or drained by the pixel reset FET (if it is held in a soft reset mode).

4.7.1 Sources of Nonlinearity in HyViSI Detectors

It was shown in Section 2.2 that the fractional change in nodal capacitance for the HyViSI due to the change in spacing between the diode "capacitor plates" is negligible when compared to the other

capacitances. For this reason, V/e$^-$ nonlinearity is not expected to be a primary contributor in the HyViSI pixels. This expectation is confirmed in photon transfer curves obtained with H2RG-32-147 and H1RG-022, examples of which are shown in Figure 4.18. If V/e$^-$ nonlinearity was present, the fixed pattern noise should deviate from a slope of 1 (in a log-log plot). This is not observed. Instead, the shot noise deviates from a slope of 1/2, which indicates that V/V nonlinearity is the culprit.

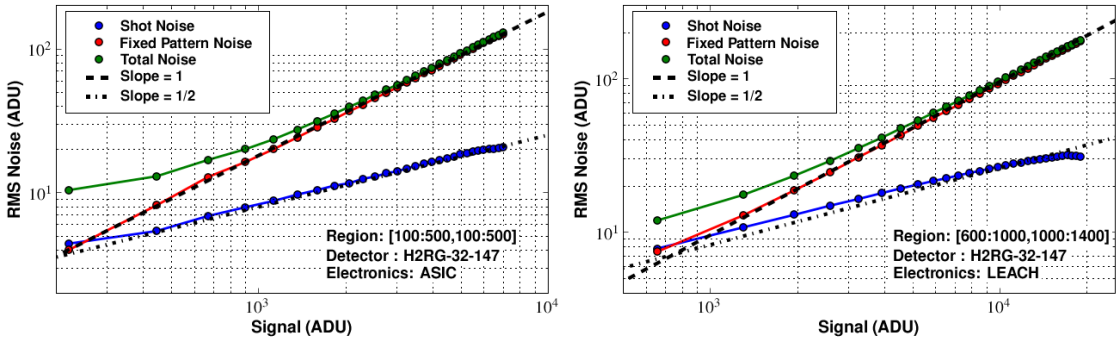


Figure 4.18: Photon Transfer Curves for H2RG-32-147 using the SIDECAR ASIC (left) and ARC (right) electronics. The curves were generated from flat field images obtained at at the Kitt Peak 2.1m telescope. The total noise is shown by the green curve. After subtracting read noise in quadrature, total noise curve is broken down into shot noise (blue) and fixed pattern noise (red), following Janesick [79]. Also shown are lines with slope 1/2 and 1, which should correspond to shot and fixed pattern noise, respectively. The deviation in shot noise from slope 1/2 indicates V/V nonlinearity.

The two primary suspects for the V/V nonlinearity are the pixel source follower, through a changing G_{UC}, and the output source follower, through a changing G_{OUT}. Measurements made with a known voltage source input to the SIDECAR and ARC electronics show that the gains G_{AMP} and $G_{A/D}$ vary by less than 0.5% over the full voltage range of the A/D converter, so these terms can be safely ignored. The control electronics may still contribute to nonlinearity in cases where the output buffer SF is used, though, since they provide it a current source. And if an external voltage on the control electronics (as opposed to one coming from the multiplexer) is used as a reference for the pixel voltages, any oscillations, sagging, etc. on the reference voltage will result in nonlinear behavior. These sources will be elaborated upon further in the next section.

4.7.2 Measurements of Nonlinearity in HyViSI Detectors

Numerous linearity measurements have been reported for HyViSI detectors. Dorn et al. report a 5% nonlinearity for an H2RG HyViSI over a 90,000 e$^-$ well [49] and Figer et al. report a 1-2% linearity for a similar H2RG device [80]. Simms et al. report a very high nonlinearity for an H4RG HyViSI of about 10% over its full 55,000 e$^-$ well. The large disparity among the numbers suggests that the

nonlinear behavior is not an intrinsic property of the imager, but instead depends on the operating conditions such as bias voltages and currents.

Unfortunately, a thorough study has not been performed to determine the configuration that minimizes nonlinearity in HyViSI imagers. From first principles, though, several bias voltages and additional sources should be relevant:

1) **The biasing to the pixel source follower FET** is especially critical to the linearity of the output and may be the primary source of the discrepancy between the reported values. Because the drain current changes over the signal range, the transconductance g_m will also vary. This, in turn, will induce a change in the output impedance [81], resulting in nonlinearity. Since the output impedance goes like $1/g_m$, the situation is improved for higher transconductance. And for the case of the HyViSI, this means that a lower $V_{BIASGATE}$ should result in better linearity because of a higher drain current and higher transconductance.

2) **The biasing to the output source follower FETs** is also very critical if they are included in the signal path. When operated in buffered mode, the HxRG detectors require an external current source or pull-up resistor as a load for the output FETs, and the more this load deviates from an ideal current source, the more the response will deviate from a linear one.

3) **Charge injection for low V_{RESET}** will cause the pixels to have an exponential ramp until their voltage reaches about 150 mV. Unlike the previous two sources, this one can be removed with a dark current subtraction.

4) **A changing reference voltage on the control electronics** will cause nonlinearity if it is used for differential measurements of the analog outputs. For instance, if $V_{REFMAIN}$ on the SIDECAR ASIC oscillates independently from the bias voltages to the multiplexer, the measured pixel voltages will appear to oscillate. This source can be eliminated by using the reference output V_{REFOUT}.

5) **Output coupling** may result in nonlinearity for all of the pixels selected at a given time. The large signal from an over-saturated pixel in one channel will cause an increase in the signals from all of the pixels in the other outputs.

5) **Coupling of the pixel node voltage to the supply voltage, reset FET gate, etc.** will cause the measured pixel voltages to deviate from the "true" voltage that would be generated purely by the charge stored on the nodal capacitance. Moore found significant coupling of the nodal capacitance to the unit cell supply, row enable gate, and reset gate [24]. While similar measurements were not carried out on the HxRG multiplexers as part of this dissertation, it is physically reasonable to assume that similar couplings exist. The reference pixels on the HxRG devices provide a means to subtract some of these couplings.

Linearity measurements for H4RG-10-007, H1RG-022, and H2RG-32-147 were made using flat field UTR exposures in which both the brightness of the source and the integration time were varied. All of these exposures were taken with the output buffer FETs in the signal path since the electronics were separated from the HxRGs by about 18" of cable. The ARC electronics uses a pull-up resistor and the SIDECAR ASIC uses a current source in the preamp circuitry to provide the drain current for the output FETs. The current provided by both should be between 300-600 μA. Measurements on H2RG-001 were made without the output FET in the signal path since a cold SIDECAR was located very close to the detector. However, the LED used to illuminate the detector resulted in very nonuniform illumination, and as a result only very small regions could be used to measure similar signal levels.

To measure the linearity from a given exposure, the mean signal for a small region of pixels was calculated for each nondestructive read (after bias subtraction), yielding an average signal $\overline{S_r}$ as a function of exposure time t_r. To generate the plots shown in Figure 4.19, the slopes $\overline{S_r}/t_r$ are normalized to $\overline{S_1}/t_1$ and plotted at S_r.

Following the discrepancy in measurements made by different authors, the plots in Figure 4.19 show dramatic differences, even for the same detector. This further supports the theory that nonlinearity in HyViSI devices is due mainly to points 1-6 listed above rather than a changing nodal capacitance. The linearity measured for H2RG-32-147 with the SIDECAR ASIC is particularly poor compared to that measured with the ARC electronics. While $V_{BIASGATE} = 2.29$ V is significantly higher for the former, Figer et al. measured 1-2% nonlinearity with $V_{BIASGATE} = 2.4$ V, suggesting that this voltage is not responsible for the difference. A similar argument can be made for $V_{BIASPWR}$. The remaining possible sources are the current to the output source follower and reference voltage used to measure the analog video outputs of the detector. If $V_{REFMAIN}$ on the SIDECAR was the source of the nonlinearity, subtracting the reference pixel voltages should reduce the nonlinearity uniformly across the signal range. Since this is not the case, it can be concluded that the current source to the output source follower is primarily responsible for the difference between the second and third plots.

The output buffer FET is eliminated from the measurement for H2RG-001 shown in the first plot. Although the signal range of the detector is limited to half of the full well–using a higher gain and V_{REFOUT} resulted in the signal clipping the upper rail of the ADC– the linearity is good to $< 1\%$ after reference pixel correction in this case. Extrapolating the curve suggests that it will remain within 1% over the full well, which further bolsters the argument that the output source follower or the current source it uses is primarily responsible for the nonlinearity in the HyViSI detectors. Further experiments should be undertaken to verify this.

The plots in Figure 4.19 do not indicate exactly what flux level was used to illuminate the detectors. However, measurements made between $500 - 10,000$ e^-/s all yield similar results. Measurements for very low flux levels need to be performed to verify linearity at very small signal levels.

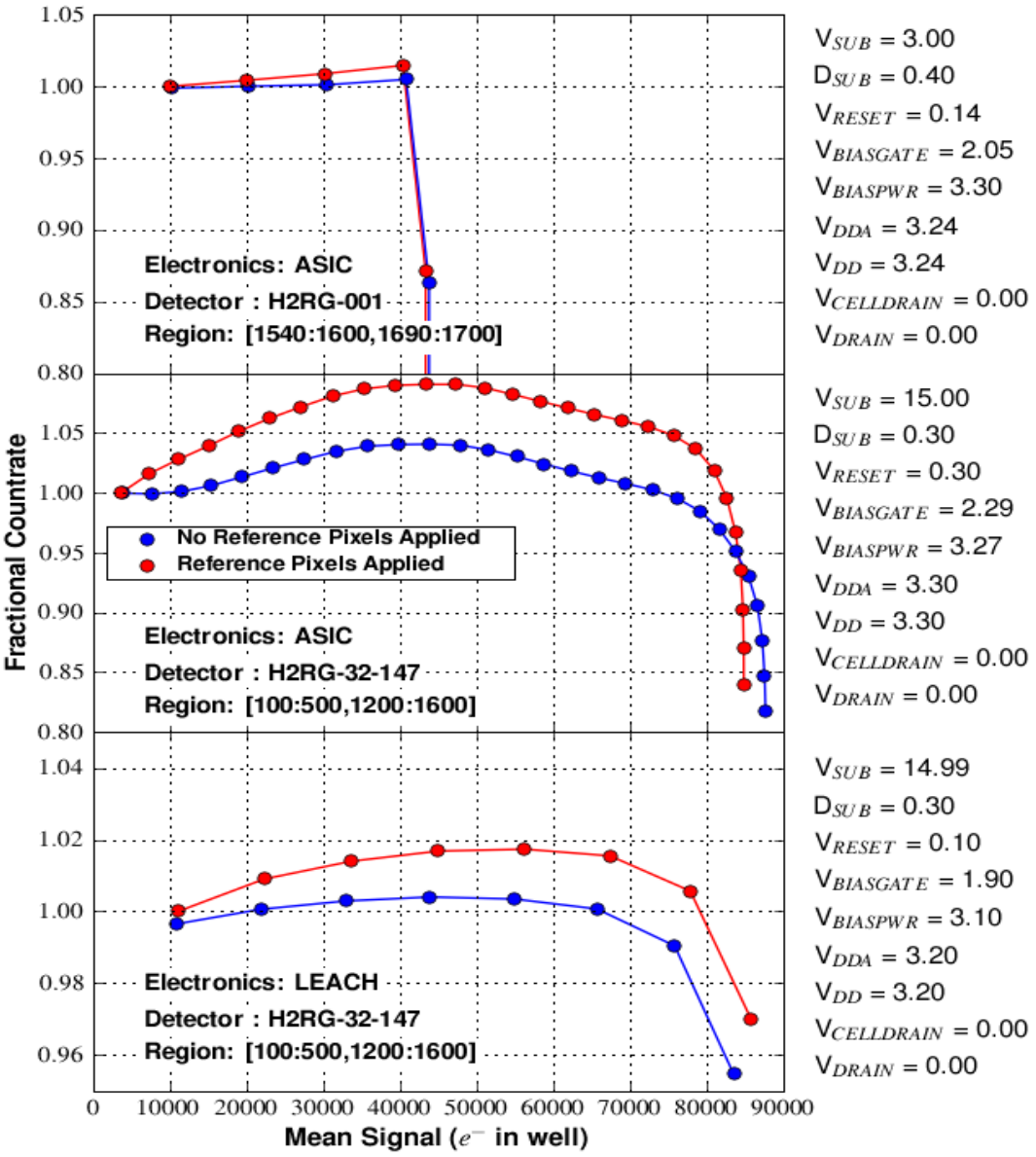


Figure 4.19: Linearity of H2RG-001 and H2RG-32-147 as a function of signal accumulated in pixel. H2RG-32-147 measurements were made with the output source follower in the signal path; those for H2RG-001 were made without it. The region for H2RG-001 was very small because the illumination was quite nonuniform over the detector and the linearity cuts off at 45,000 e^- because the pixel voltages were out of range for the SIDECAR ADC. The bias voltages used for the measurements are listed to the side of each plot in units of volts. The output source follower and $V_{BIASGATE}$ appear to have the largest effect on the linearity.

4.7.3 HyViSI Well Depths

A large pixel well depth is highly desirable in astronomy as it allows for large dynamic range imaging. If the wells are small, bright stars may fill them before the pixels are read, preventing an estimate of their luminosity. In theory, the well depth, WD, should be determined by the doping of the active collecting volume for the photogenerated charge. For the case where holes are collected:

$$WD = V * N_A, \tag{4.15}$$

where V is the volume of the collecting region and N_A is the acceptor density. However, this prediction is rarely, if ever, met in CCD or CMOS detectors [3]. The HyViSI pixels are no exception. If the p^+ implant in the pixel has a depth of 1 μm, a size of 9-18 μm on a side, and an acceptor concentration of $N_A = 10^{14} - 10^{19}/\text{cm}^3$, the well depth is expected to be $WD = 10^7 - 10^{10}$ e^-. This is far in excess of the typical measured values on the order of 10^5 e^-.

In addition to the pixel structure, the well depth should depend on the reset voltage V_{RESET}. The maximum depth is achieved when $V_{RESET} = 0$ V is at its minimum value and the reverse bias of the PIN diode is at its maximum. However, the exponential dark current mentioned in Section 4.5.4.1 makes values of $V_{RESET} < 150$ mV impractical.

The well depth can be measured from the same exposures used to generate the linearity curves in the previous section. Piecewise slopes $(S_{r+1} - S_r)/(t_{r+1} - t_r)$ are calculated between consecutive reads and normalized to the one calculated from a slope fit (see Section 5.2.4). The pixels are deemed saturated at a signal where their normalized slopes fall below 0.9. The mean of the saturation signal level is then taken to be the average well depth. Results for the HyViSI sensors H1RG-022, H2RG-32-147, and H4RG-10-007 are listed in Table 4.5. The low well depth for H1RG-022 relative to that of H2RG-32-147 is not understood since the pixel layout should be similar in the two devices.

Table 4.5: Well Depths for several HyViSI detectors along with pixel size and the value of V_{RESET} at which they were measured. The pixel responsivity, G_{PIXEL}^{-1}, is discussed in Section 4.3.

	H1RG-022	H2RG-32-147	H4RG-10-007
Pixel Size (μm)	18	18	10
V_{RESET} (mV)	90	100	100
Well Depth (e^-)	55,000	85,000	55,000
G_{PIXEL}^{-1} (μV/e^-)	10.53	10.61	25.21

Chapter 5

Silicon PIN Detectors in Astronomy

5.1 Observations with the Kitt Peak 2.1m Telescope

We observed on the 2.1m telescope at the KPNO observatory in Tucson, Arizona (110° 58′ 0″ W, 32° 13′ 0″ N) during three separate runs in 4/07, 11/07, and 12/07. On each run, a different detector was tested: H4RG-10-007 on the first, H1RG-022 on the second, and H2RG-32-147 on the third. A picture of the RIDL dewar and electronics mounted to the telescope is shown in Figure 5.1. The dewar was placed at the Cassegrain focus, with the telescope delivering an f7.6 beam to the focal plane. Since the pixel size and dimensions of each detector were slightly different, the plate scale and field of view (FOV) varied from run to run. Table 5.1 summarizes these details. It should also be noted that the detector read noise during observations (20-30 e^- CDS) was significantly higher than what we have obtained in recent laboratory measurements (8-10 e^- CDS).

Table 5.1: Details for observing runs at Kitt Peak 2.1m Telescope.

	Run 1: H4RG-10-007	Run 2: H1RG-022	Run 3: H2RG-32-147
Dates	4/24-4/30	11/13-11/19	12/12-12/19
Plate Scale	0.126″/pix	0.227″/pix	0.227″/pix
Field Of View	8.5′×8.5′	3.8′×3.8′	7.5′×7.5′
Photometric Nights	6	5	4
Electronics	SIDECAR	SIDECAR and ARC	SIDECAR
Guide Mode	No	No	Yes

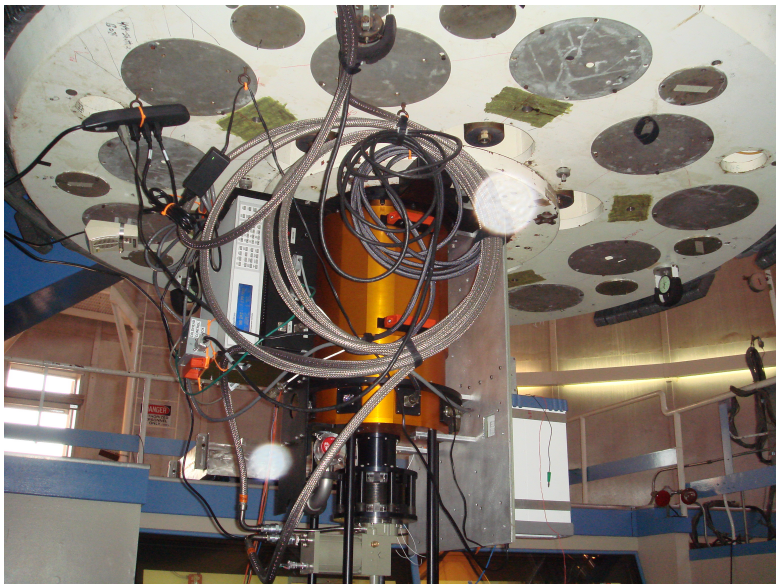


Figure 5.1: A photograph of the RIDL dewar mounted to the Kitt Peak 2.1m telescope.

The detector was housed inside the RIDL dewar and positioned behind a filter wheel with six positions controlled by Phytron motors. One position was left open, one had an Fe^{55} source, one was a blank that prevented light from reaching the detector, and the other three were occupied by g,i, and y filters described in Table 5.2. To accommodate the RIDL dewar, the telescope guider camera was removed from the telescope. The telescope was set to track at the sidereal rate and a rotator was used to correct for rotations of the field of view. The only exception to this was when H2RG-32-147 was operated in guide mode to simultaneously guide the telescope and take long exposures. In all, nearly 2 Terabytes of calibration and science data were collected in single window, multiple window, full frame, and guide mode. The majority of targets were chosen primarily to verify that HyViSI detectors can indeed function as astrometric and photometric instruments. Open clusters and bright stars from the SAO and GSC catalogs were used to test guide mode and measure persistence. And lastly, select Messier and NGC objects were targeted for tri-color imaging.

Table 5.2: Filter Characteristics

Filter Name	Peak (nm)	Cuton (nm)	Cutoff (nm)	Transmission (%)
g	476	401.22	557.87	96.39
i	742	667.61	815.77	96.14
y	1003	970.54	1036.01	86.28

5.2 Data Reduction and Calibration

As mentioned in Chapter 1, optical astronomers and astrophysicists are most familiar with CCDs. Even if they do not understand all of the steps in the signal chain or the technical details of how their sensor turns photons into a digital number, they at least know that certain steps must be performed to remove from the data as much of the CCD signature as possible. This process of cleaning the data and removing the instrumental signature is referred to as *data reduction*. Its real aim is to *reduce* all extraneous components of the data as much as possible so that what is left is pure signal.

It is beyond the scope of this thesis to present an adequate treatment of data reduction. Many excellent references are devoted to the subject. McLean gives a superb overview of the theory behind data reduction for non-destructive detectors [4]. Bushouse et al. provide a very detailed description of the data pipeline used to reduce exposures from the NICMOS infrared camera on Hubble [44]. Here we present only the basic theory and the steps needed to turn a set of raw astronomical images into a finished product that can be used for scientific study.

5.2.1 Data Reduction Theory

Because of the non-destructive readout of hybrid CMOS detectors, the reduction of the data is quite different from that of a CCD. In fact, it is more akin to data reduction for IR detectors. For a CCD, the data output of a single exposure is a two-dimensional set of pixel intensities $I(x, y)$. In contrast, for an up-the-ramp exposure as described in Section 3.1, the data output is in general a three-dimensional *data cube* with pixel intensities $I(x, y, r)$. The third dimension, r, is the read number and corresponds to the time at which the pixel at x, y was sampled. If the detector is reset and clocked in a uniform fashion with a frame time of t_{frame} then the times at which a pixel is sampled relative to the time it was reset are given by $t(r) = t_{frame} * r$.[1] The total number of reads R can vary from 1 to any greater integer, but an exposure with $R = 1$ is rarely used in scientific applications because it is dominated by reset noise and pixel to pixel transistor current offsets.

Each sample of the datacube represents the measurement of an analog voltage V_{out} that has made its way from the pixel to the output of the detector. The voltage is proportional to the amount of charge integrated on the pixel capacitance from dark current, photo-current, and other sources such as carriers emitted from traps or released by minimum ionizing particles or radiation passing through the detector. Each of these carrier generating processes contains noise. This noise adds to the other noise contributors such as the Johnson noise of electrical currents passing through each of the transistors in the signal path and the ones described in Section 4.4. The voltages of the pixels will also each have their own equilibrium offset from ground due to subtleties such as the precise

[1] Note that for a given reset scheme $t(r) = t_{frame} * r + t_{offset}$. If the reset is applied pixel by pixel as the rows are clocked, then $t_{offset} = 0$. If the reset is applied to a whole row at a time, then $t_{offset} = t_{pix} * n_{col}$, where t_{pix} is the pixel time and n_{col} is the pixel column number relative to the first pixel in the row of the output channel to which it belongs.

values of N_A and N_D in the pixel implants, traps in the pixel source follower channel, proximity to current carrying buses, etc. The term *bias* is often used to describe this offset, although care must be taken to distinguish this from the case where it is used to describe *bias* voltages. In general, we can write the measured voltage as:

$$V_{out}(x,y,r) = V_{lum}(x,y,r) + V_{dark}(x,y,r) + V_{spurious}(x,y,r) + V_{bias}(x,y) + V_{noise\ det.}(x,y,r) \quad (5.1)$$

where

- V_{lum} is the contribution from signal charge integrated due to light falling on the pixel. This is the signal of interest in astronomy.

- V_{dark} is the contribution from charge integrated due to thermal leakage currents in or around the pixel.

- $V_{spurious}$ is the contribution from charge integrated as a result of trap emission or capture, persistence, overflow from neighboring pixels, x-ray and cosmic ray events, etc.

- V_{bias} is the equilibrium offset voltage after the pixel has been reset.

- $V_{noise\ det.}$ represents all sources of noise between the pixel and the detector output. This term includes coupling to neighboring pixels as well as inductive or capacitive pickup from other current carrying buses on the detector.

The first three terms in Equation 5.1 are related to the charge that has accumulated in the pixel, $Q(x,y,r)$, by the gains described in Section 4.3. That is, $V(x,y,r) = Q(x,y,r)/(G_{PIXEL} * G_{UC} * G_{OUT})$. The last two terms cannot be as easily related to measurable quantities. For instance, V_{noise} might include pickup from a ground loop on the detector, which could vary wildly in different configurations and is extremely difficult to pinpoint.

The pixel intensity that is stored in the datacube is a sample of the analog voltage recorded by the control electronics. The electronics and sampling process itself imparts a noise $V_{noise\ elec.}$ on the signal (see Section 3.5), and in addition, amplifies $V_{noise\ det.}$ by the electronics gain G_{AMP}. Since the noise component from the detector and the one from the control electronics are, at least in principle, uncorrelated, they add in quadrature:

$$V_{noise} = \sqrt{G_{AMP}^2 V_{noise\ det.}^2 + V_{noise\ elec.}^2} \quad (5.2)$$

This expression effectively replaces $V_{noise\ det.}$ in Equation 5.1. With all voltage contributions taken

into account, the ADC converts the sum to a digital number, I, in Analog-Data-Units (ADU):

$$I(x,y,r) = G_{A/D}\Big[G_{AMP}\big(V_{lum}(x,y,r) + V_{dark}(x,y,r) + V_{spurious}(x,y,r) + V_{bias}(x,y)\big) + \sqrt{G^2_{AMP}V^2_{noise\ det.}(x,y,r) + V^2_{noise\ elec.}}\Big] \tag{5.3}$$

Note that here we are using $G_{A/D}$ in units of ADU/μV.

After conversion the pixel values are transmitted to a computer for storage. In astronomical applications, the most commonly used storage format is the Flexible Image Transport System (FITS). For the data collected for this thesis with the SIDECAR and ARC electronics, the pixel values were stored as datacubes of 16-bit unsigned integers in FITS format.

The digital pixel values are often called the *raw data*, and it is with these values that data reduction is performed. A fundamental assumption is that each of the terms in Equation 5.3 are independent of one another. For instance, the assumption is made that $V_{dark}(x,y,r)$ is the same whether or not light is falling on the detector. With this being the case, an exposure taken in the dark can be subtracted from an illuminated exposure to remove the dark current component along with any well behaved, time-dependent noise sources. Another assumption is that certain noise components are Gaussian distributed. Thus, they can be beaten down by taking multiple exposures with the same V_{lum}. While these techniques and many others apply to both CCD and CMOS sensors, the fact that a typical CMOS exposure contains multiple samples of a time-dependent signal gives rise to some important differences. The data volume is in general much larger, saturated pixels can still yield flux estimates, and the signal can be measured as a rate of change of light falling on the detector or as a total of the integrated light. The following sections present a short list of data reduction techniques in the context of hybrid CMOS detectors.

5.2.2 Reference Pixel Correction

The reference pixels of the HxRG multiplexers provide an excellent means for reducing common mode noise in the detector. A thorough description of them is given in Section 3.4. For each read in the datacube, the average of the left set of reference pixels ($i = 0, 1, 2, 3$) and the average of the right set of reference pixels ($i = N-3, N-2, N-1, N$) is taken across the rows to yield two one dimensional column vectors:

$$S_{ref\ l}(j,r) = \sum_{i=0}^{3} \frac{S(i,j,r)}{4} \qquad S_{ref\ r}(j,r) = \sum_{i=N-3}^{N} \frac{S(i,j,r)}{4}, \tag{5.4}$$

where N is the number of pixels in a detector row. A Savitsky-Golay filter is then used to smooth the columns and yield the two vectors $\hat{S}_{ref\ l}$ and $\hat{S}_{ref\ r}$. Finally, the smooth vectors are multiplied

by the factor C_{FAC} and subtracted from the science pixels belonging to their respective half:

$$S(i,j,r) = S_{sci}(i,j,r) - C_{FAC} * \hat{S}_{ref} \tag{5.5}$$

The subtraction of the reference pixels will effectively boost the read noise by $\sqrt{2}$. In some cases, this number will be small compared to bias voltage drifts or $1/f$ noise. In other cases the reference pixels will not show any drift at all. The stability of the common mode noise has a very complex dependence on the state and history of all operating conditions: temperature, bias voltages, substrate voltage, clocking patterns, etc. Subtraction of the reference pixels should therefore be made on a case by case basis.

5.2.3 Dark Subtraction

To remove the contribution from dark current, $V_{dark}(x,y,r)$, we first obtain a series of dark ramps having the same cadence as the ramps we wish to reduce (see Section 3.1.1 for a definition of *cadence*). We then take the median of each pixel value at each read across the exposures to yield a median dark datacube with pixel values $I^{med}_{dark}(x,y,r)$. The median filter efficiently rejects cosmic rays and voltage spikes, but a mean filter with outlier rejection can be applied instead to improve the signal estimation. It is extremely important that the bias read ($r = 1$) of each exposure be subtracted from each of the subsequent reads in that same exposure to remove kTC noise **before the median filter is applied**. In other words, for the jth dark exposure:

$$I^j_{dark}(x,y,r) = I^j_{dark}(x,y,r) - I^j_{dark}(x,y,r=1) \tag{5.6}$$

This can usually be done with the stored values in computer memory before the median is applied as long as the datacubes are sufficiently small. Note that the frame $I^{med}_{dark}(x,y,r=1) = 0$ and can be discarded from the median dark datacube to save disk space.

From each illuminated (object) ramp through a given filter we subtract the median dark ramp:

$$I_{obj}(x,y,r) = I_{lum}(x,y,r) - I^{med}_{dark}(x,y,r), \tag{5.7}$$

where I_{lum} is the pixel value from the illuminated image. The bias read $I_{obj}(x,y,r=1)$ will remain the same, but the dark current contribution to each subsequent read will be removed. The resulting frames in this ramp should, in theory, have pixel values, I_{obj}, that correspond to luminance from the source being observed.

5.2.4 Slope Fitting

The next order of business is to estimate the photocurrent for each pixel ramp in the datacube of interest: I_{obj}. As discussed in Section 4.4.2.4, fitting a slope to the ramps is a good technique for

this since it beats down the read noise and aids in cosmic ray rejection. Plus, it is assumed that this data reduction is being performed on a computer dedicated for analysis, so processing complexity and memory should not be a critical issue.

The slopefit should only be applied to the ADU values for which the pixel ramp is linear. According to the discussion in Section 4.7, the ramps will take on a nonlinear behavior when the pixel well is near capacity, and this transition will take place at some upper limit in the *raw* signal, I_{lum}^{max}. Nominally, a value $I_{lum}^{max}(x,y)$ for each pixel would be used, but a mean value can also be used without a noticeable reduction in the quality of the slopefits. One may also prescribe some lower value $I_{lum}^{min}(x,y)$ below which the signal is not considered for a slopefit, perhaps because of some nonlinearity in the early reads. Once these limits have been prescribed, we consider the ADU values in between to be in the *linear regime*.

In order to find the slope that minimizes the error for the points in the linear regime, i.e. where I_{lum} lies between I_{lum}^{min} and I_{lum}^{max}, we fit a line to the corresponding values of I_{obj} using the technique described in Numerical Recipes [82]. Namely, for a pixel that has $N_{rd} = r_{max} - r_{min}$ values of $I_{lum}(r)$ between reads r_{min} and r_{max}, and corresponding dark subtracted values $I(r) = I_{obj}(r)$,

$$b = \frac{\sum_{r_{min}}^{r_{max}} t(r)^2 \sum_{r_{min}}^{r_{max}} I(r) - \sum_{r_{min}}^{r_{max}} t(r) \sum_{r_{min}}^{r_{max}} t(r)I(r)}{N_{rd} \sum_{r_{min}}^{r_{max}} t(r)^2 - \left(\sum_{r_{min}}^{r_{max}} t(r)\right)^2}$$

$$m = \frac{N_{rd} \sum_{r_{min}}^{r_{max}} t(r)I(r) - \sum_{r_{min}}^{r_{max}} t(r) \sum_{r_{min}}^{r_{max}} I(r)}{N_{rd} \sum_{r_{min}}^{r_{max}} t(r)^2 - \left(\sum_{r_{min}}^{r_{max}} t(r)\right)^2}. \tag{5.8}$$

b is an approximation to the bias offset in ADU, and m is the number of ADU/s attributed to the source of illumination. It should be apparent that only the difference in times $t(r)$ and $t(r+1)$ matters, so it suffices to use the average time for the rth read. With a slope obtained and the proper conversion gain in e^-/ADU, m can be converted to units of e^-/s.

Bright objects will induce saturation very quickly, and in some cases there will be too few or no values of $I_{lum}(r)$ in the linear regime. These cases require alternative approaches. If a pixel is saturated in the second read (the bias read is considered the first read) and $I_{lum}(1) < I_{lum}^{max} - 0.5I_{lum}^{FR}$, where $I_{lum}^{FR} = I_{lum}^{max} - I_{lum}^{min}$, the slope is approximated by the CDS value $m = (I(2) - I(1))/(t(2) - t(1))$. And if a pixel is saturated in the second read and $I(1) > I_{lum}^{max} - 0.5I_{lum}^{FR}$, the best one can do is approximate the slope by $m = I(1)/t(1)$.

Once a slope has been fit to the points in the linear regime, cosmic rays are detected as large deviations from the fit using the same method as the one in the NICMOS reduction code [44]. In this method, the difference between the data points and the fit is first computed as

$$D(r) = I(r) - (m \cdot t(r) + b). \tag{5.9}$$

Then the difference between adjacent points in this difference is taken,

$$DD(r) = D(r) - D(r-1), \tag{5.10}$$

along with its standard deviation, σ_{DD}. Cosmic rays are flagged as points where $DD(r) > T_r\sigma_{DD}$, where T_r is some threshold. For our HyViSI data, T_r was typically set around 3. The idea behind this scheme is that the cosmic rays particles will release significantly more charge in the pixel over some small time interval than the integrating photocurrent and show up as a large negative to positive spike in D_i. Using a separate σ_{DD} for each individual pixel seems to be quite effective and well suited for treating the variation in pixel sensitivity across the array.

Ideally, if a cosmic ray is detected in a read $r = r_{cr}$ late in the ramp, $r_{cr} > r_{min} + N_{rd}/2$, then the slope is refit using points from $r = r_{min}$ to $r = r_{cr}$. And if it is detected early in the ramp, $r_{cr} < r_{min} + N_{rd}/2$, the slope would be ideally refit using points $r = r_{cr}$ to $r = N_{rd}$. However, the latter results in extremely large error for nearly all of the HyViSI devices tested for this thesis due to IPCT (see Section 6.1.2), and so the beginning of the ramp must be used instead. After refitting, the number of points used in the fit after the cosmic ray rejection should be included as an extension to the image for purposes of error analysis.

An added benefit of UTR sampling is that the energy deposited by a high energy particle can be well approximated by examining the signal $I_{r_{cr}} - I_{r_{cr}-1}$. These values can be recorded and stored as a separate image. They are potentially interesting for measuring the angular distributions, frequency, and morphology of such events.

5.2.5 Flat Fielding

After every pixel has had a slope fit to it for the object exposure, we are left with an NAXIS1×NAXIS2 array of slopes, $m_{obj}(x, y)$, that we take to be the image. To account for the pixel to pixel variations in sensitivity—this can be due to small differences in the gain between individual pixels, dust on the detector, vignetting, etc.—we must apply a flat field.

Flat field exposures are recorded while the detector is uniformly illuminated through the appropriate filter. The light level should be great enough so that vignetted pixels see a reasonable photocurrent and the cadence should be chosen so that the pixels do not oversaturate. As mentioned in Section 4.6.2, the illumination setup used for the flats should be as close to that for the object exposures as possible in order to account for fringing at wavelengths near 1 μm. If the incoming rays are not oriented similarly in the two, the fringes will appear in different spatial locations. As an example of this, during our telescope observations we collected both "dome" flats where we exposed the detector to a uniformly lit white spot on the interior of the telescope dome and "sky" flats where we exposed the detector to the sky at twilight. The "sky" flats, which are clearly more representative of the illumination delivered by the parallel rays coming from the night sky, were far

more successful in removing the fringes than the dome "flats", where the rays entering the telescope optical system cannot be considered parallel due to the proximity of the source to the primary mirror of the telescope.

After recording a sufficient amount of flat field exposures, a median flat is generated in the same way that the median dark was formed (i.e. the bias reads are subtracted from the individual exposures and then the median is obtained). Then, the median dark is subtracted and slopes are fit to the pixel ramps in the median flat according to the method described in Section 5.2.4, which yields a 2-d image that we call $m_{flat}(x, y)$. For simplicity, at this point $m_{flat}(x, y)$ is normalized so that the values range from 0 to 1. Finally, the flat is applied to the object image by dividing the slopes:

$$m_{fin}(x, y) = \frac{m_{obj}(x, y)}{m_{flat}(x, y)}, \tag{5.11}$$

to yield the final slopes $m_{fin}(x, y)$.

One important consideration to keep in mind when recording darks and flats with the HyViSI detectors is that persistence in flat field images will show up in dark current measurements. Therefore, it is very important that the dark exposures are either recorded before the flat fields or sufficient time is given for any persistence to subside before recording the darks.

5.2.6 Combining Dithers

Dithering is a very useful technique for eliminating the impact of bad pixels or regions of defects on the detector. In the majority of cases, for our *full frame* exposures we used a dithering technique to provide a number of samples of each field, each of them being slightly offset from the others in terms of location on the detector. The dither pattern was a 3×3 box where each telescope pointing was offset from the previous one by 20″ (∼160 pixels on the H4RG and ∼80 pixels on the H1RG or H2RG). A full exposure was taken at each of these pointings, yielding 9 or more slopefitted images of the field in each filter depending on whether or not the dithering sequence was repeated.

The flat-fielded slope images, $m_{fin}(x, y)$, from each filter are aligned using several bright stars and then combined into a mosaic. To combine the data from the multiple pixel values at a spatial location (x,y) in mosaic image coordinates, we use both a median filter and the mean of the pixels that were not flagged as bad or rejected as 3σ outliers. The mean provides a slightly better reduction in noise, and the mean mosaics are the images that we use for analysis in the following sections.

5.2.7 Telescope Calibration with Multiple Windows

High speed photometric measurements of the Delta Scuti star BE Lyn (V∼8.8) were made during the observing run with H2RG-32-147. On the night of December 17, 2007, BE Lyn and the nearby reference star GSC 03425-00544 (V∼11.1) were simultaneously imaged at a rate of 4 Hz over the course of several hours. The two stars are separated by approximately 2′. A separate 61 × 61 guide

window was used for each star, in a sequence where the first and second windows were reset in series, read once in series, and read once again in series to provide a Correlated Double Sample (CDS). This sequence was executed repeatedly and was only interrupted when it was necessary to write the data to a FITS file. In addition to providing temporal flux measurements, from which the period of BE Lyn was measured in i band (see Section 5.3.4), this data inadvertently provided a detailed look at the telescope tracking.

The star centroids calculated from this data show that there are significant tracking errors for both right ascension and declination. In fact, the errors were so large that it was necessary to manually offset the telescope while the observations were being made. This was possible since the data was being displayed in real time on a computer monitor. The errors are quantified as *drifts* between the telescope tracking rate and the sidereal rate. For periods of about 20 minutes or less, the drift in DEC can be fit with a straight line; any longer than this and a curvature becomes evident, especially as the telescope crosses the meridian. The drift in RA has both a long period linear component and a short period sinusoidal one. The drifts Δ_{DEC} and Δ_{RA}, measured in arcseconds, can be reasonably well fit over time-spans less than 20 minutes with the functions:

$$\begin{aligned} \Delta_{DEC} &= m_{DEC} \ * \ t \\ \Delta_{RA} &= m_{RA} \ * \ t + A\sin(2\pi f_{osc} t), \end{aligned} \tag{5.12}$$

where m_{DEC} and m_{RA} are the linear slopes of the drifts, and A and f_{osc} are the amplitude and frequency of the short period oscillation in RA, respectively.

For 12 piecewise fits over the 4 hour time span, the average slopes were found using least square fits to the data. The magnitude of m_{RA} varied between $0.110''$/minute and $0.950''$/minute, and was greatest right before crossing the meridian. The sign of m_{RA} was always negative, indicating that the telescope was lagging the movement of the sky. The sign of m_{DEC} changed several times over the 4 hour period. Its magnitude ranged from $0.028\ ''$/minute to $0.39\ ''$/minute, and was also largest while the telescope crossed the meridian.

After removing the linear component in the RA data, a discrete Fourier transform of its autocorrelation was performed for each piecewise fit. The frequency was determined to be $f_{osc} = 0.00833$ Hz, corresponding to a period of 120 seconds. This frequency was extremely consistent throughout the data and so was the value of the amplitude, $A = 0.5104''$. Although the residuals in the fit show that the oscillations are not purely sinusoidal (more time is spent in the leading portion of the curve than the lagging portion), the amplitude and frequency provide an estimate on the reduction in image quality. A fit to the data for a 16 minute period is shown in Figure 5.2. It should also be noted that autocorrelations in the DEC drift show power at f_{osc}, indicating that the two coordinate motions are coupled. However, the amplitude is much lower, at about the level of the high frequency components induced by atmospheric effects.

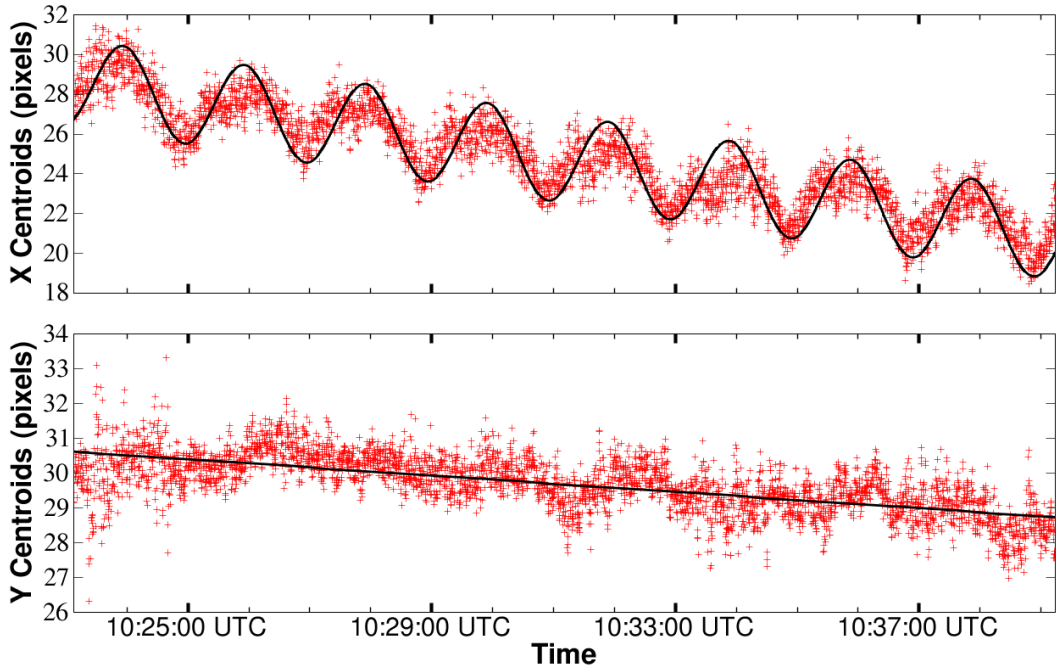


Figure 5.2: The telescope drifted significantly over a 20 minute period and showed sizable oscillations without the assistance of a guider. The centroids from images taken at 4 Hz are shown as plus marks. The black lines show the fits to the drifts from Equation 5.12 with $m_{RA} = -0.108''$/minute, $m_{DEC} = -0.026''$/minute, $f_{osc} = 0.00833$ Hz, and $A = 0.5104''$.

5.2.8 Expected Point Spread Function

The tracking error functions in Equation 5.12 can be used directly to predict the image quality for unguided and guided exposures. The full data set collected during the observing runs with H1RG-022 and H2RG-2-147 shows that the system point spread function (PSF) is Gaussian and seeing-limited for intermediate length exposures (long enough to average out atmospheric turbulence and short enough to avoid degradation from tracking error). For long exposures, the center of the Gaussian profile is assumed to drift at the rates m_{DEC} and m_{RA} in DEC and RA, respectively, and oscillate in RA at a frequency of f_{osc}. For an exposure of length t_{exp}, the resulting point spread function will thus be:

$$I(x,y) = \int_0^{t_{exp}} I_o \exp\left(\frac{-(x - m_{RA} * t - A\sin(2\pi f_{osc} t))^2}{2(\epsilon/2.35)^2} + \frac{-(y - m_{DEC} * t)^2}{2(\epsilon/2.35)^2}\right) dt, \qquad (5.13)$$

where ϵ is the Full Width at Half Max (FWHM) due to the average seeing in arcseconds, and I_o is the peak intensity of the stellar image. Numerical solutions to this integral were obtained for representative values of ϵ, t_{exp}, m_{RA}, m_{DEC} in order to predict the image quality in both guided

and unguided operation.

To quantitatively describe the PSF for each solution, a 21 × 21 box around the centroid of the intensity distribution is used to compute the second moments I_{xx}, I_{yy}, and I_{xy}, where:

$$I_{ij} = \frac{\sum_{n,m} x_i(n,m) x_j(n,m) I(n,m)}{\sum_{n,m} I(n,m)}. \tag{5.14}$$

From these, the ellipticity vector is calculated from its components, e_1 and e_2, according to the prescription in Bacon et al. [83]:

$$e_1 = \frac{I_{xx} - I_{yy}}{I_{xx} + I_{yy}}, \qquad e_2 = \frac{I_{xy}}{I_{xx} + I_{yy}}. \tag{5.15}$$

The magnitude of ellipticity, e, and its angle, θ, are given by

$$e = (e_1^2 + e_2^2)^{1/2}, \qquad \tan 2\theta = e_2/e_1. \tag{5.16}$$

In addition to these measurements, the radial profiles around the centroid are fit with one-dimensional Gaussian and Moffat functions. The FWHM from the fits, along with the second moments, measure the overall extent of the PSF, while the ellipticity measures the asymmetry in RA and DEC. Together these quantities adequately describe the PSF.

5.2.8.1 Unguided Operation

It is fairly straightforward to see from Equation 5.12 that if the telescope is not guided, the tracking errors will quickly produce a PSF that is broadened beyond the seeing disc and elongated along the direction of the drift. However, it is not trivial to quantitatively describe the effect the drift will have on its shape. For two representative drift rates observed during the photometric measurements of BE Lyn, the ellipticity and FWHM obtained from a Gaussian fit were calculated for a range of atmospheric seeing values, ϵ, assuming $t_{exp} = 1800$ s. The results are shown in the top two curves of Figure 5.3.

The plots show that the ellipticity can easily exceed 0.2 when the blurring due to seeing is not large enough to mask the sinusoidal oscillation and drift of the telescope. This effect is more pronounced when $m_{RA} > m_{DEC}$ since the centroid motion along the x axis of the detector is far greater than the motion along the y axis during the exposure. It should also be noted that at a given value of ϵ the ellipticity will be an increasing function of t_{exp}, as the smearing along the direction of drift increases, while in the transverse direction it remains constant. For short exposures ($t_{exp} < 1$ m), a Gaussian shape is preserved for typical seeing values.

5.2.8.2 Guided Operation

During our last observing run we implemented guide mode with H2RG-32-147 (results are shown in Section 5.5) and issued offsets to the telescope to correct for tracking error. In our configuration, one set of IDL scripts was used to collect data from the SIDECAR ASIC through its USB interface and another set of IDL scripts communicated to the Telescope Control System (TCS) via a RS-232 serial port connection. Centroids were calculated from the guide window data in IDL as the other detector pixels were integrating up the ramp, and an offset was issued to the telescope if it appeared that the telescope was drifting.

The TCS for the 2.1m accepts offsets as small as 0.1 arcseconds. However, due to telescope motor hysteresis and other mechanical effects, the minimum offset which can be issued with precision is not this small. While a precise value is not known, it is definitely less than 1 arcsecond. After some initial measurements, we found that the best results were obtained when we issued offsets only if the centroid of the star had moved by $0.3''$, which corresponds to about 1.3 pixels with the $0.232''$plate scale. On average we found that an offset in RA was needed every 20-50 seconds. The declination adjustment varied between once per 1-5 minutes depending on the altitude of the pointing.

To predict the quality of the PSF with the adjustments described, Equation 5.12 is not used directly. Instead, it is broken up into a total of $t_{exp}/t_{min\ adj.}$ separate integrals, where $t_{min\ adj.} = 20$ s is the interval of time between the most frequent adjustments in RA. For each integration, the center in x of the Gaussian is set back to zero. The center in y is set back to zero only at the end of a time interval $t_{max\ adj.} = \alpha t_{min\ adj.}$, where α represents the ratio of RA/DEC adjustment frequencies. This number ranged between 3 and 15 depending on the pointing. The phase of the sinusoidal motion is preserved between successive integrations to best represent the observed behavior of the telescope motion after small adjustments.

The results of calculations for which $t_{min\ adj.} = 22$ s, $\alpha = 12$, and $t_{exp} = 1800$ s are shown in the bottom two curves of Figure 5.3. The linear drift rates are the same as the two cases represented in the unguided calculation. The ellipticity and FWHM are greatly diminished as expected. However, the asymmetric centroid motion along the x and y axes yields a nonzero ellipticity. For small linear drift rates, the sinusoidal motion along the x-axis is actually the greatest contributor. This can be understood by considering the times immediately after the telescope has been corrected in RA. Since the phase of the oscillatory motion is preserved after adjustment, the pointing of the telescope following a correction will in some cases swing in the opposite direction of the linear motion and contribute to a greater blurring in the x direction.

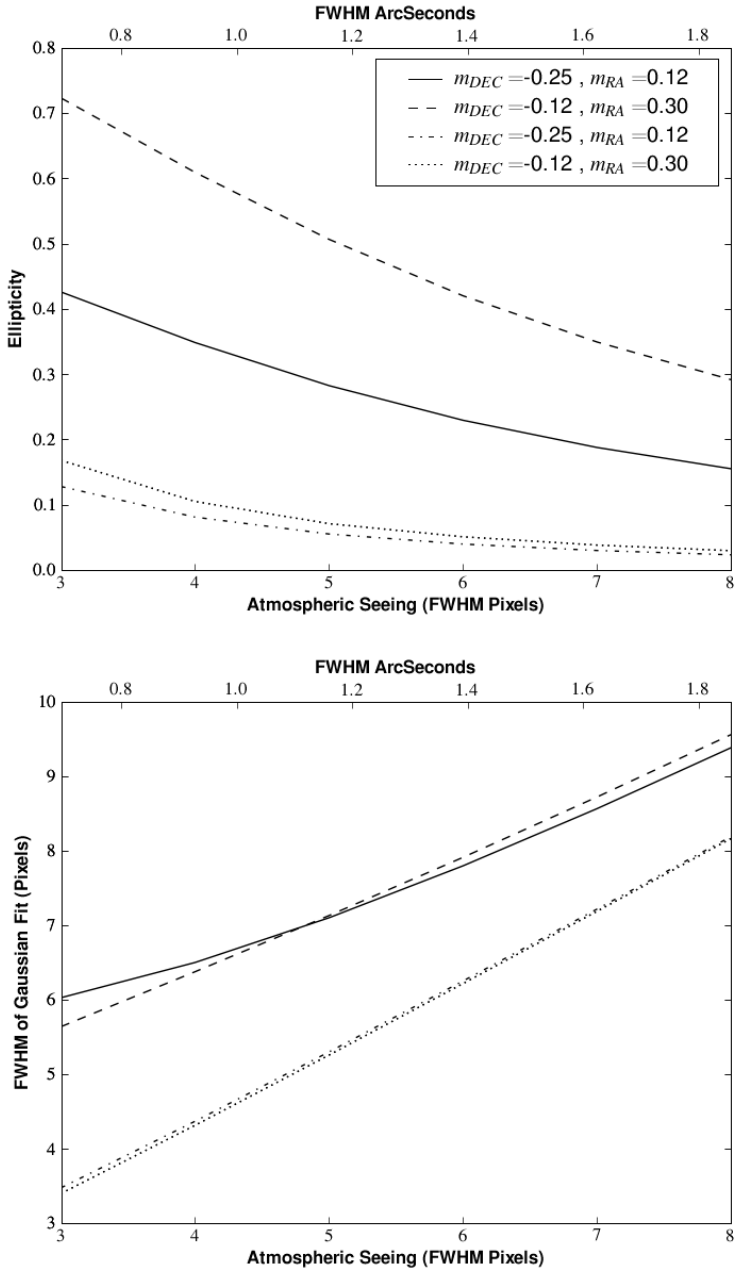


Figure 5.3: Expected ellipticity (top) and FWHM (bottom) derived from a Gaussian fit with a drifting center for a 30 minute exposure. The drift rates m_{DEC} and m_{RA}, expressed in $''$/minute are shown in the legend. The top two curves in each plot represent the results for the cases where the telescope is allowed to drift. The bottom two show the calculation results for the cases where the drift is corrected (by re-centering the Gaussian) whenever the pointing has exceeded 1.3$''$in RA or DEC, as was the case during our guiding operation.

5.3 Photometry

In an astronomical context, photometry is the measurement of celestial source brigthnesses.[2] While the traditional practice of expressing these brightnesses in *magnitudes* is still favored in astronomy, for astrophysical purposes, the end goal of *absolute photometry* is to relate them to the amount of energy emitted by a particular object per unit time per unit area, A.K.A. the flux. The practical way to do this with a particular instrument is to observe a well measured standard star and use *relative photometry* to convert between the instrumental measurements and an absolute scale. Even with this *apparent* flux, though, the distance to the object must be known to obtain the *absolute* flux. In general, the brightnesses are measured over a wavelength bandpass that is relatively large in comparison to the wavelength itself, e.g. through color filters for visible wavelengths. When the bandpasses become sufficiently small and a number of them are used, one enters the realm of spectrophotometry. Photometry is undoubtedly one of the most important areas of astronomy; our understanding of open and globular clusters, variable stars (which also provide distance estimates), supernovae, and a whole host of other phenonomena would be nonexistent without it. As such, it is essential that an astronomical detector be capable of delivering accurate photometric measurements.

Since the CCD entered into the arena of astronomy, it has been the "undisputed leader" in photometry [84]. The CCD is superior to photographic emulsions because of its high quantum efficiency and large dynamic range, and highly more efficient than an accurate photomultiplier tube because of its large field of view. In the past, the poor quantum efficiency and fill factor of CMOS detectors made them inferior to CCDs as well. However, with the advent of hybrid and back illuminated CMOS detectors, a comparitive study between photometric performance of CCD and CMOS detectors is now a legitimate one. In fact, if CMOS detectors can match the photometric accuracy of CCDs, they will be inherently better in the fast time domain because of the windowing capability and fast readout speed.

To assess the potential of the HyViSI detectors as photometric instruments, several studies were carried out. These included aperture photometry on well calibrated photometric sources (Landolt stars), crowded-field photometry of globular clusters, and high speed window photometry of variable and multiple sources. Each of these illustrates the ability of the detector to operate in a different photometric regime.

Photometric measurements of Landolt stars are well suited for testing the ability of the detector to sense brightness differences within a particular wavelength passband as well as determining how well the detector can be calibrated to obtain absolute magnitudes within that band. Crowded-field photometry of the stars in M13 is good for determining the spectral responsivity of the detector by comparing measurements through separate passbands since the relationship of color index vs. magnitude has been well studied in the cluster. And lastly, fast measurements in window mode

[2] Technically, there is a distinction between *photometry*, which applies to the visible wavelength range, and *radiometry*, which applies to the entire spectrum. In practice, astronomers use the former term to describe both.

provide an indication of how well the detector can sense changes in illumination. In the following sections, each will be described in turn.

5.3.1 Aperture Photometry

Landolt Standard stars provide a good set of basis measurements for comparison as they have been repeatedly observed through different filter sets with different calibrated instruments. With H4RG-10-007, instrumental magnitudes of well isolated standard stars in the Landolt Equatorial Fields PG 1530 and SA 109 were obtained. Unfortunately, these standards have been observed very frequently through UBVRI filters of the Johnson-Kron-Cousins system, but not through g,i, and y filters that are more closely matched to our set. It is possible, however, to make a transformation between pairs of filters in these two systems under the assumption that the variation of the spectral energy distribution $E(\lambda)$ of the stars is sufficiently continuous over the intervals considered to allow a Taylor expansion in λ.

Such a transformation is done, as in Verdoes et al. [85], by solving the following equations:

$$m_l^{H4RG} = M_j^{CAL} Z - KX + CT \times (M_k^{CAL} - M_j^{CAL}), \tag{5.17}$$

where M_j^{CAL} is the reference magnitude in the filters $j = V$, I, and Y, m_l^{H4RG} is the instrumental magnitude that we measured through the filters $l = g$, i, and y, X is the airmass, K is the atmospheric extinction coefficient, CT is a color coefficient and $M_k^{CAL} - M_j^{CAL}$ is the color defined by filter j and $k = B$, R, and H from the reference measurements. Astronomers frequently use these equations in order to compare measurements made at different telescopes and to calibrate for slight differences between filter sets. They are typically solved using a population of several hundred standard stars. Due to limited time and data, we have only seven.

The transformations require a calculation of an average airmass over the duration of our exposures in each filter band. For this we used the algorithm suggested by Stetson [86] :

$$X_{avg} = (X_{beg} + 4X_{mid} + X_{end})/6, \tag{5.18}$$

where X_{beg} is the airmass at the beginning of the first exposure, X_{mid} is the airmass midway through the exposure (for us it was the fifth dither in the sequence) and X_{end} was the airmass at the end of the dither sequence. The average airmasses through which we observed ranged from 1.12 to 1.26, so loss due to atmospheric extinction was not very significant.

Once the equations are solved, they can be used to predict the magnitudes we should expect to observe for g, i, and y filters at zero airmass based upon standard magnitudes in V, I, and Y and the corresponding colors. The fitting coefficients and errors are shown in Table 5.4 and a plot that illustrates the goodness of the fit is shown in Figure 5.4. The errors in the fit are not unreasonable. They are similar to the ones found in the transformations made in the Sloan Digital Sky Survey [87].

The error in g is substantially higher than it is in i and y. This may be attributed to a number of factors including observing conditions, wavelength dependent lateral diffusion in the detector, and nonlinearity. The exact reason for the large discrepancy is still being investigated. However, the data and the fit indicate that H4RG-10-007 is capable of doing absolute photometry.

From a preliminary aperture photometry analysis of open cluster data obtained with H1RG-022 and H2RG-32-147, the photometric errors for these detectors are smaller than those for H4RG-10-007. This improvement in performance is expected since the dark current shot noise is significantly less in the H1RG and H2RG. Also, the reset voltages were tuned properly for observations with these detectors to prevent the nonlinear behavior we observed in the early reads of the ramps in H4RG-10-007. In the newest generation of H4RG, a large reduction in dark current is observed, suggesting that the photometric accuracy will be improved [88].

Table 5.3: Landolt Star Magnitudes. Our measured magnitudes are denoted by lowercase m. M are magnitudes taken from catalogs; $+$ are taken from Landolt (1973)[89] , $\diamond$ are taken from Landolt (1992) [90], and $\star$ from Gullixson (1995) [91]. All M_Y are from Persson (2002) [92].

Star	M_Y^{CAL}	ΔM_Y^{CAL}	$M_Y^{CAL} Error$	$M_Y^{CAL} - M_H^{CAL}$	m_y^{H4RG}	Δm_y^{H4RG}	$m_y^{H4RG} Error$
109-956	12.516	0.000	0.029	1.038	15.362	0.000	0.001
109-954	10.254	2.262	0.028	1.064	13.079	2.283	0.001
109-949	11.384	1.132	0.029	0.595	14.250	1.112	0.001
	M_V^{CAL}	ΔM_V^{CAL}	$M_V^{CAL} Error$	$M_B^{CAL} - M_V^{CAL}$	m_g^{H4RG}	Δm_g^{H4RG}	$m_g^{H4RG} Error$
109-959 *	12.790	0.000	0.029	0.780	12.404	0.000	0.003
109-956 $^\diamond$	14.639	-1.849	0.011	1.283	14.407	-2.003	0.009
109-954 $^\diamond$	12.436	0.354	0.009	1.296	12.187	0.217	0.003
109-949 $^\diamond$	12.828	-0.038	0.006	0.806	12.408	-0.004	0.002
1530-057 $^\diamond$	14.21	0.000	0.000	0.151	12.756	0.000	0.003
1530-057A$^\diamond$	13.71	0.500	0.000	0.829	12.514	0.242	0.003
1530-057B$^\diamond$	12.84	1.37	0.000	0.745	11.595	1.161	0.002
	M_I^{CAL}	ΔM_I^{CAL}	$M_I^{CAL} Error$	$M_R^{CAL} - M_I^{CAL}$	m_i^{H4RG}	Δm_i^{H4RG}	$m_i^{H4RG} Error$
109-959 $^+$	11.572	0.000	0.009	0.671	12.197	0.000	0.003
109-956 $^\diamond$	13.114	-1.542	0.016	0.743	13.886	-1.689	0.007
109-954 $^\diamond$	10.940	0.632	0.003	0.731	11.391	0.806	0.002
109-949 $^\diamond$	11.708	-0.136	0.003	0.517	12.422	-0.225	0.003
1530-057 $^\diamond$	14.011	0.000	0.000	0.036	12.967	0.000	0.003
1530-057A$^\diamond$	12.842	1.169	0.000	0.412	11.967	1.000	0.002
1530-057B$^\diamond$	12.041	1.970	0.000	0.376	11.160	1.807	0.003

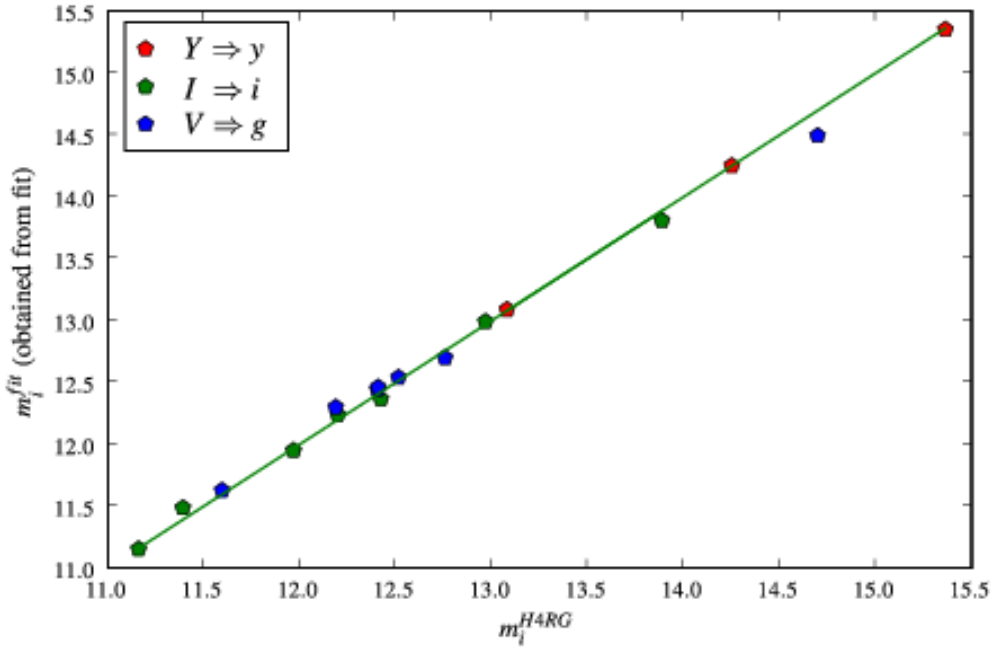


Figure 5.4: The results from the best-fit solution to the transformation equations. The horizontal axis shows the magnitudes that we measured in our images. The vertical axis shows the magnitudes that are expected from the fits in Table 5.4 based upon the standard magnitudes. The green line shows where $m_i^{H4RG} = m_i^{fit}$.

Table 5.4: Transformation equations obtained from the data in Table 5.3. The superscripts make explicit the fact that the m_l are the magnitudes that are calculated after Equation 5.17 has been solved.

Transformation	LSF Error	σ
$m_y^{fit} = M_Y^{CAL} + 2.173 - 0.068(M_Y^{CAL} - M_H^{CAL})$	0.0068	0.0096
$m_i^{fit} = M_I^{CAL} - 17.03 + 14.08X + 0.388(M_R^{CAL} - M_I^{CAL})$	0.0724	0.0591
$m_g^{fit} = M_V^{CAL} - 10.57 + 8.025X + 0.502(M_B^{CAL} - M_V^{CAL})$	0.1261	0.1030

Table 5.5: Parameters from M13 exposures.

Filter	Airmass	Exp. Time	Cadence (s)	FWHM ($''$)	Avg. Background (counts/sec)	σ Background (counts/sec)	Stars Found
g	1.003909	163.59	1-0-30-1	1.64	2.01	0.38	9104
i	1.011084	81.79	1-0-15-1	1.14	3.05	0.34	15,812
y	1.015034	163.59	1-0-30-1	1.32	1.80	0.28	2439

5.3.2 Crowded Field Photometry

We measured instrumental magnitudes of stars in M13: the Hercules cluster, a color image of which can be seen in Figure 1.1. We observed the cluster in a range of RA from 250.404-250.580 and DEC from 36.368-36.511 on two separate nights : 4/26/07 and 4/28/07. The raw data for M13 on 4/26/07 consists of 9 dithered up-the-ramp exposures for each of the g, i, and y filters, and is described in Table 5.5. Photometric analysis was performed on the slopefitted mosaics of M13 using the DAOPHOT algorithms DAOPHOT, GETPSF, SUBSTAR, NSTAR, and ALLSTAR in IRAF through the PyRAF interface. For details on parameters, refer to Appendix A.1, and for a description of DAOPHOT, refer to Davis [93]. The basic purpose of DAOPHOT is to identify point sources and measure the brightness of those sources alone. It does this by allowing the user to create a semi-analytic model that represents the point-spread function (PSF) of a star in the image and goes on to fit each star with that model. This technique is necessary for the case in which the field is "crowded", i.e. the images of the stars are so close that they overlap and a pixel receives light from more than one source. The result of the DAOPHOT algorithms is an instrumental magnitude for each star.

In order to determine the zero-points of the instrumental magnitudes obtained with the DAOPHOT package, we obtained g and i magnitudes for the stars SDSS J16420106+362401.0 and SDSS J1646154.09+362348.8 from the Sloan Digital Sky Survey and compared them with ours. We find, roughly, that $M_g^{CAL} = m_g^{H4RG} - 0.25$ and $M_i^{CAL} = m_i^{H4RG} - 0.62$. For y, we use the first equation in Table 5.4 and set the color term to zero since no stars could be found with y band magnitudes for reference in the field.

With the magnitudes adjusted for the zero point offset, color magnitude diagrams were created for the $g - i$ filter pair and the $g - y$ filter pair. These diagrams, shown in Figures 5.5 and 5.6, are qualitatively similar to those obtained from previous photometric studies for B and V filters. They show well the features of the red giant and blue straggler populations in the cluster. This is a good indication that the H4RG is capable of doing relative photometry between these bandpasses and verifies that the spectral responsivity is good out to the 1 μm region. Similar results will hold for HyViSI detectors in general.

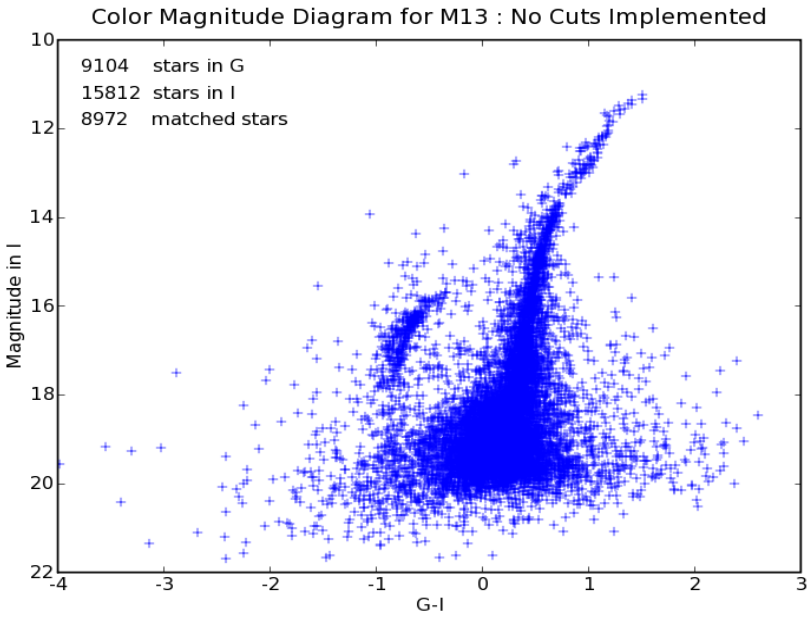


Figure 5.5: A Color Magnitude diagram for the M13 cluster in g and i bands.

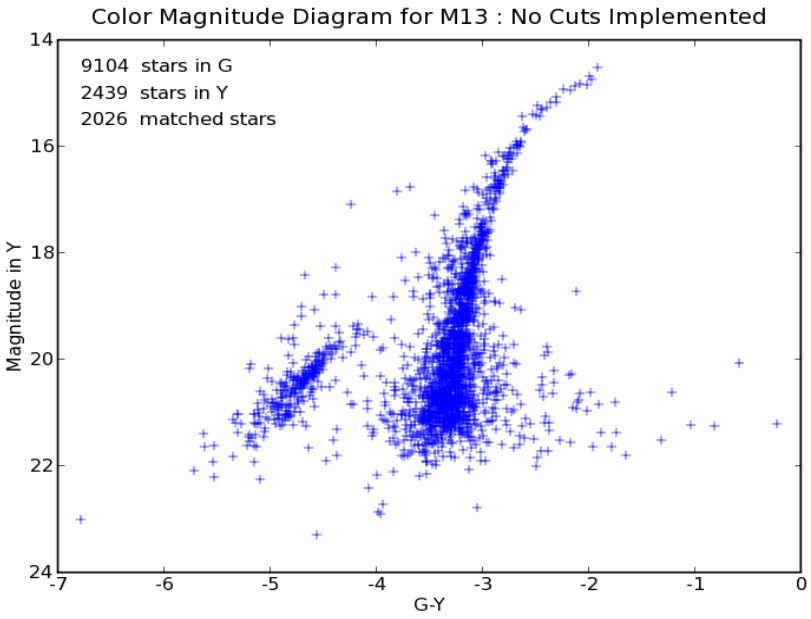


Figure 5.6: A Color Magnitude diagram for the M13 cluster in g and y bands.

5.3.3 High Speed Photometry with Guide Windows

Random, non-destructive access to pixels in a CMOS sensor open up intriguing possibilities for high speed astronomical photometry. The guide window capability of the HxRG allows one to skip around the detector and measure the flux of different stars without being forced to read pixels that are not in regions of interest. The time required to read out two separate windows will be independent of their separation in the sky, which is not a luxury of a CCD detector.

5.3.4 Variability Measurements of BE Lyn

As a preliminary test of photometry using multiple guide windows, the Delta Scuti star BE Lyn (HD 79889) was imaged with H2RG-32-147 over a period of several hours (see Section 5.2.7 for details). Delta Scuti stars are short period (0.05-0.25 days) variable stars that are believed to change brightness due to radial pulsation. BE Lyn is a particularly interesting example since its period has been measured numerous times and appears to be changing over time [94], [95], [96], [97], [98]. It has even been suggested that BE Lyn might possibly have a companion [99]. A brief timeline of the measurements and predictions surrounding the star is as follows:

1987 Oja detects variability and reports a period of $P = 0.0958697$ days for the oscillation in brightness [94].

1991 Yanying et al. measure a change in the period relative to previous measurements. After using Observation vs. Calculation curves (O-C), they placed the average period at $P = 0.095869547$ days and fit their data with an equation of the form [95]:

$$T_{max} = HDJ2446506.00774 + 0^d.095869547 \times E + 0.5 \times G \times E^2, \tag{5.19}$$

where T_{max} is the time of maximum, $T_o = HDJ2446506.00774$ is an initial reference time, E is the cycle number, and $G = -2.1 \times 10^{-12}$ days per cycle^{-2} is the rate of period change. The value of G they reported (see Table 5.6) indicated the period was decreasing with time.

1992 Qingquan et al. fit their own data and reported that the period is actually increasing with time [96].

1994 Zhongli et al. place the average period at 0.09586963 days and concur the observations of Yang et al. that the period is increasing with time [97].

1995 Kiss and Szatmary perform photoelectric photometry and fit the BE Lyn O-C curve supposing a cyclic period variation due to an orbital companion [99].

2003 Derekas et al. make additional measurements and reject the presence of long-term light-curve shape changes. They place an upper limit on the rate of change of the period as

$\frac{1}{P}\frac{dP}{dt} = (-5 \pm 1.9) \times 10^{-8}$ year^{-1} and admit no unambiguous conclusion can be drawn at that point in time [98].

Table 5.6: Variability Parameters for the Delta Scuti Star BE Lyn as measured by various authors. See references for additional measurements.

Author	Year Reported	P (days)	T_o (Julian Day)	G (days per cycle^{-2})
Oja [94]	1987	0.0958697	None	None
Yanying [95]	1991	0.095869547	HDJ2446506.00774	-2.1 $\times 10^{-11}$
Qingquan [96]	1992	0.09586938	HDJ2446506.0079	1.1 $\times 10^{-11}$
Zhongli [97]	1994	0.09586963	HDJ2449018.2684	8.8 $\times 10^{-12}$
Derekas [98]	2003	0.095869521	HDJ2449018.2681	None (See text above)

As this timeline indicates, the period of BE Lyn changes in a fashion that is not understood through conventional physical models. However, the period is extremely long in comparison to the sampling period attainable with the guide windows and provides a good target for testing the capability of the HyViSI to perform high speed photometry. An investigation was thus undertaken with H2RG-32-147 in order to examine the usefulness of reference correction with multiple guide windows and the stability of the photometric response over a long time span.

5.3.4.1 Observations with Kitt Peak 2.1m and Reduction

Observations of BE Lyn were made through the i filter with H2RG-32-147 from 09:45 UTC until 13:00 UTC on the night of Dec 17, 2007. The staggered-reset-staggered-read-staggered-read (SRSRSR) multiple window mode was chosen (see Section 3.2.1) for these observations in order to allow alternation between the target star and a reference star whilst providing similar atmospheric noise for both integrations. The window size for both the target BE Lyn and the reference star GSC 03425-00544 was 61 $\times$ 61 pixels, which yielded an integration time of roughly 0.08 seconds and a period of about 0.241 seconds ($\sim$ 4 Hz). Unfortunately, high cirrus clouds were present and the seeing was not nominal; it varied from 1.4-1.8 arcseconds over the night. Out of the approximately 100,000 windows collected (50,000 for each star), nearly 15,000 could not be used because of cloud cover or an extremely diffuse PSF.

After data collection, for each CDS window frame centroids were measured using the IDL FIND utility and then aperture photometry was performed with IDL APER. All relevant information was recorded to a MYSQL database for later analysis. To apply a correction with the reference star, the measured flux of the target was divided by the measured flux of the reference star on a point by point basis. After correction, a median filter was applied to the light curve of the target in order to remove outliers, reduce noise, and fill in the many missing data points.

5.3.4.2 Results

Figure 5.7 shows the instrument corrected magnitudes returned by APER for BE Lyn and the reference star during the 3 hour observation period. The variable nature of BE Lyn is unambiguously evident when compared to the reference star. The brightness of the reference star is relatively flat, but becomes broader near the end of the night and in areas where cloud cover obscured the field of view. Meanwhile, BE Lyn undergoes a ~ 0.25 change in magnitude. The outlying points and gaps in the data for the reference star coincide with those of BE Lyn: a fact which is more easily observed in Figure 5.8, which shows the magnitude of BE Lyn after the flux has been normalized by the reference star. While the spread of data points grows greatly in the normalization (the read noise and shot noise are effectively multiplied in the division), the outliers are no longer evident, indicating the high magnitude outliers are caused by cloud cover or uniform atmospheric effects.

The missing data points and the fact that BE Lyn was only observed for 1.25 periods makes it difficult to precisely measure its period with this data set. Using an autocorrelation technique as well as a period searching program similar to the one described by Oja [94], our measurements place the period between and $0.0958286 < P < 0.0961084$ days. A value in this range is consistent with previously measured periods. Previous O-C curves yield errors of 10^{-3} days for predictions of the

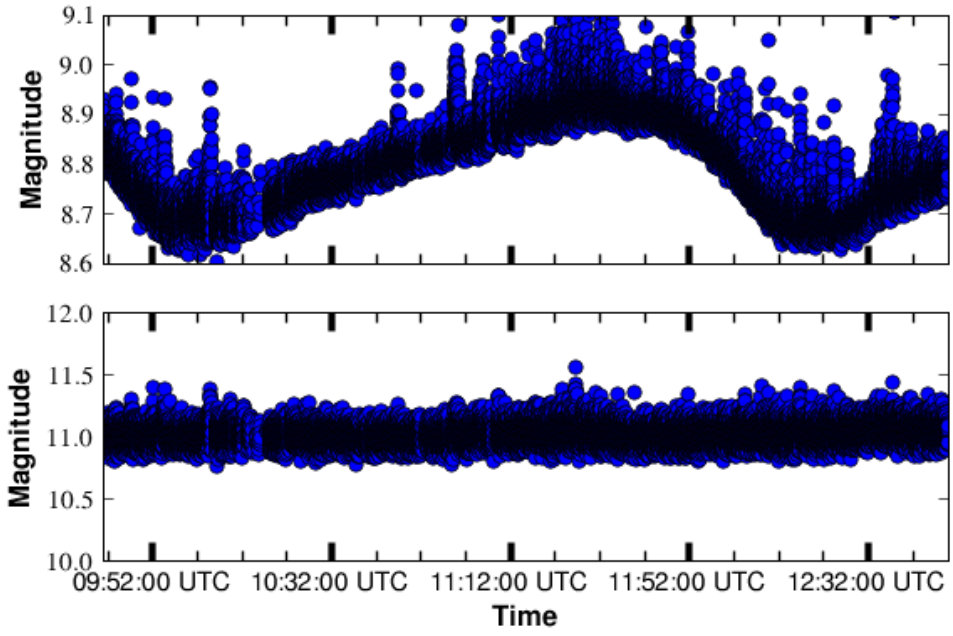


Figure 5.7: (Top) Magnitude vs. time plot for the Delta Scuti star BE Lyn over a 3 hour period. (Bottom) The same plot for the reference star GSC 03425-00544.

period, so the error in the present measurements are within reason.

The overall amplitude of oscillation in brightness for BE Lyn is measured to be about 0.21 magnitudes. No previous i band measurements could be found for this star, but the V band amplitude is reported as 0.39 magnitudes. The reference star fluctuations are 0.065 magnitudes RMS without any rejection of outliers. With the improved read noise obtained in recent laboratory measurements (20-30 e^- read noise RMS was measured at the telescope; 8-10 e^- in lab tests following the observations), the read noise contribution to the overall noise fluctuations in magnitude will be significantly diminished. Finger et al. have performed very similar measurements with a HgCdTe H2RG array on a magnitude 5 star and obtained an error of 0.026 magnitudes RMS with a read noise of 8.2 e^- [100]. These results are very promising for high speed photometry with HyViSI sensors, placing them in good favor for observing fast variables like Delta Scuti stars, extra-solar planet transits and occultations, or other known or unknown "bumps in the night".

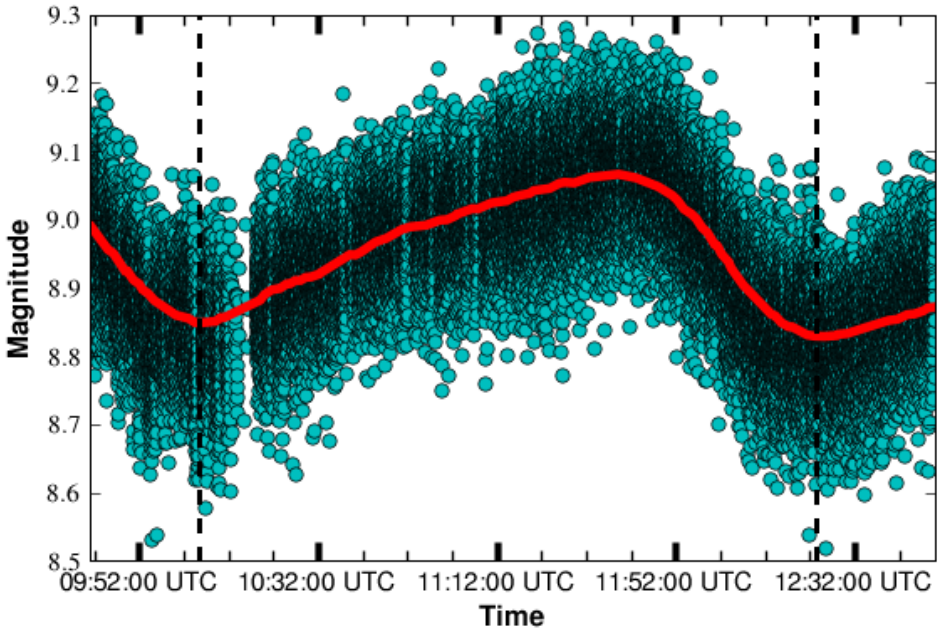


Figure 5.8: Magnitude vs. time plot for BE Lyn after normalization by the reference star. The cyan points show the raw data after normalization and the red line through them is a median filter that was applied to the data to reject outliers and interpolate the missing data points. The two dashed black vertical lines indicate the accepted period of oscillation, which is 0.0958697 days or 02:18:03 hours.

5.4 Astrometry

Astrometry is the precise measurement of positions and motions of celestial bodies. Because astronomical distances possess a large uncertainty, these measurements are usually confined to the two-dimensional projection of the sky as seen by observers on earth, and positions of astronomical objects are tabulated in catalogs as a set of two sky *angles* measured with respect to a set of reference points.[3] In modern astronomy, standard practice is to record coordinate information to the FITS image header at the time of observation using the world coordinate system (WCS), which greatly facilitates analysis of astrometric data and its transaction between astronomers [101]. With proper WCS information, positions in one image can be easily compared with those in another, regardless of the type of detector used or the epoch at which they were recorded.

As described by Monet, the apparent position of an object with respect to nearby reference stars is

$$Position = Constant + Proper\ Motion + Parallax + Perturbations \tag{5.20}$$

in each coordinate [102]. For one particular observation, the telescope, detector, and atmosphere must be considered as additional terms on the right hand side of this equation. As will be discussed, though, these terms can be removed with proper calibration techniques and using multiple exposures. Each term on the right side of Equation 5.20, aside from the constant one, is significant in astronomy and astrophysics. Measuring parallax is extremely important because it yields the most unambiguous distance measurements possible. In fact, all other distance estimates rely on parallax as the "bottom rung" of the cosmic distance ladder. Proper motion and perturbations of objects within our galaxy yield a great deal of information about the dynamics of the Milky Way as a whole in addition to smaller gravitationally bound systems such as binary stars. Also, extrasolar planets can be detected with precise, milli-arcsecond (mas) astrometric measurements. Astrometry is thus a very important facet of astronomy, and will become even more so as large scale surveys attempt to "map" the sky in finer detail.

While systematic errors in the relative positions of objects being measured by a given detector can be corrected for with proper calibration, *the accuracy* with which the detector can measure the individual positions is limited at a physical level. The pixel size (or more precisely, the plate scale in $''$/pixel), quantum yield, read noise, linearity, pixel and output crosstalk, and dark current all determine how well centroids can be measured on the array, and thus, what the astrometric error will be. Theoretical calculations and Monte Carlo simulations can be used to make predictions on the error, but in the end they must be compared to real data for validation.

In the following sections, the relative astrometric errors measured with HyViSI sensors at the

[3] The most common system is the *equatorial coordinate system*. Object positions are projected onto the *celestial sphere* and the angular coordinates of right ascension (α or RA) and declination (δ or DEC) serve as analogs to longitude and latitude, respectively. The zero position for RA is measured with respect to the First Point of Aries and the zero position for DEC is measured with respect to the celestial equator.

Kitt Peak 2.1m telescope are presented. The plate scale for the 18 micron pixel H1RG-022 and H2RG-32-147 detectors was 0.227″/pixel and that for H4RG-10-007 was 0.126″/pixel. In theory, the astrometric error measured with these detectors in terms of pixel units can be used to predict angular errors for other optical configurations.

5.4.1 Sources of Astrometric Error

There are a number of sources that contribute to astrometric error. The telescope optics and detector can cause distortions of the field in the final image. However, both of these are typically stable (although the distortion of the primary mirror may vary over its full range of motion) and can be corrected by applying a distortion map to the x-y pixel positions in the image. Atmospheric turbulence and read noise can lead to nonuniform, random errors in centroid positions across the focal plane. Both of these contribute greatly to error in a single, short exposure, but they can be diminished with long integration times or averaging multiple exposures. Finally, the optical system (e.g. the point spread function) and the aforementioned detector characteristics impose a base level error for the centroid positions in a given frame.

Following the methodology of Zacharias [103], the variance in a transformation of x,y centroid positions in a given frame into the average x,y positions in a reference coordinate map can be written as

$$\sigma_{astrm}^2 = \sigma_{atm}^2 + \sigma_b^2, \tag{5.21}$$

where σ_{atm} is the noise induced the atmosphere and σ_b is the base level noise level in an individual exposure. The factor of $\sqrt{2}$ is removed because the transformation is done between an individual frame and a reference grid rather than two individual frames. It is important to note that this is for the shot-noise limited case. If faint stars are included and the detector is read-noise limited, an additional term $\sim \sigma_{wn}/N_{reads}$ should be included, where σ_{wn} is the white noise and N_{reads} is the number of reads used to estimate the signal. With sufficiently bright stars, though, this term can be neglected. Since the dependence of σ_{atm} on integration time goes as $t^{-1/2}$, Equation 5.21 can be written as

$$\sigma_{astrm}^2 = \sigma_a^2 t^{-1} + \sigma_b^2, \tag{5.22}$$

where σ_a is the noise inherent in a semi-instantaneous realization of the atmosphere.

Equation 5.22 predicts that the astrometric noise will go down as $t^{-1/2}$ and level out at σ_b for very long integration times. In this prediction, it is assumed that the telescope is being properly guided. For the case where the telescope is not being guided, an additional term $\sigma_{tracking}$ must be added:

$$\sigma_{astrm}^2 = \sigma_a^2 t^{-1} + \sigma_b^2 + \sigma_{tracking}^2 \tag{5.23}$$

It will be shown that $\sigma_{tracking}$ becomes significant for the unguided Kitt Peak observations after a

time of about 50 seconds.

5.4.2 Astrometric Reduction

Astrometric analysis of the Kitt Peak 2.1m data obtained with H1RG-022, H2RG-32-147, and H4RG-10-007 was conducted primarily with utilities in the IRAF IMCOORDS package. These utilities were accessed through the pyraf command language. The analysis carried out for a set of exposures (the final slopefitted images) on a given field consists of the following basic steps:

1) Use the WCSTools **scat** program to obtain coordinates, magnitudes, and proper motions for all stars in the field detected in the USNO CCD Astrograph Catalog (UCAC2) [104]. The UCAC2 catalog has a limiting magnitude of R=16 and a standard error of 70 mas.

2) Detect source positions in each exposure using the STARFIND utility in IRAF. STARFIND calculates centroids by convolving each source with a Gaussian of a specified FWHM, so the IMEXAMINE routine is used on each image to provide a good estimate of this. Center of mass centroids were also calculated for comparison, and found to produce very similar results.

3) Match x-y positions of stars in the image to RA/DEC entries in catalog using the IRAF utility CCXYMATCH. After matching the stars, a six parameter astrometric fit is performed. Two of the parameters account for rotation, two for scale, and two for linear shift. The fit yields an improved set of world coordinate system (WCS) coordinates for the image. With the new WCS, the x-y positions of each star are converted to RA and DEC, and these new celestial coordinates are written to a file.

4) After the RA and DEC values have been computed for all the images of the field, the average RA and DEC positions for each star are calculated. These average celestial coordinates constitute a *reference map* to measure the astrometric error. The reference map removes any systematic errors generated by the 70 mas error from the UCAC2 catalog and proper motions of the stars since the time at which they were observed.

5) Measure astrometric error for a given frame by comparing it to the reference map. The astrometric error, σ_{astrm}, is then given for each coordinate as:

$$\sigma_{astrm}^{DEC} = \sqrt{\sum_{i=0}^{N}(\delta_i - \delta_i^{Ref})^2/N} \qquad \sigma_{astrm}^{RA} = \sqrt{\sum_{i=0}^{N}(\alpha_i - \alpha_i^{Ref})^2/N}, \tag{5.24}$$

where δ_i and α_i are the declination and right ascension of the ith star in the individual frame, respectively, δ_i^{Ref} and α_i^{Ref} are the corresponding coordinates as measured in the reference map, and N is the total number of stars found in the frame.

The reduction was performed for a number of observed fields. In some fields, multiple exposure times were sampled; in others the same exposure time was used for a variety of pointings (as part of a dither sequence described in Section 5.2.6). The *g,i,y* filter set was used and the temperature of the detector did not leave the range of 170-180 K. The seeing conditions varied from 0.8-1.8 arcseconds, and an expected degradation of the astrometric accuracy is observed with the poor seeing conditions.

5.4.3 Astrometric Results with H2RG-32-147 and H1RG-022

As was discussed in Section 5.2.7, most of the Kitt Peak observations were done without telescope guiding. For short exposure times, the tracking induced errors are expected to be negligible in comparison to the read noise and atmospheric errors. For long exposures, the read noise and atmosphere should average out, but the tracking errors are expected to produce significant error in the right ascension because of the oscillations described. This is precisely what is seen in Figure 5.9. The RA and DEC errors both go down approximately as $t^{-1/2}$ for exposure times of less than 50 seconds and reach a minimum around 7-9 mas, but then begin to rise thereafter. The rise is worse for the RA because of the large amplitude oscillations of the telescope pointing along this axis.

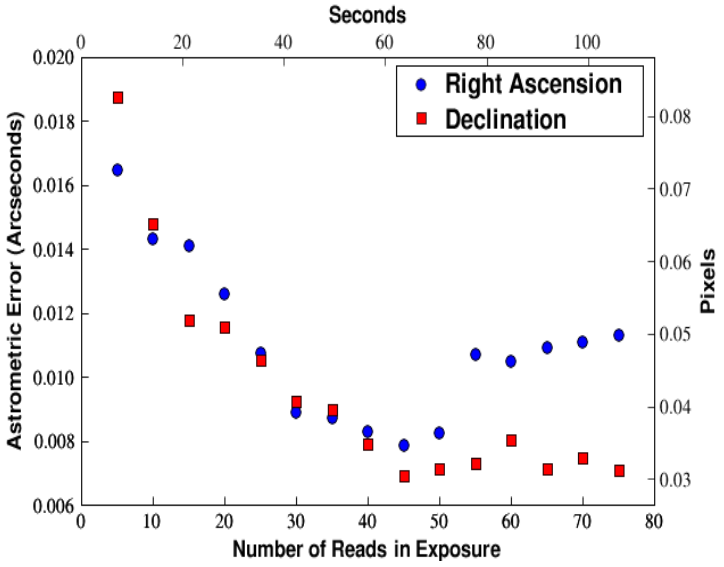


Figure 5.9: Astrometric error vs. exposure time for the open cluster NGC 956 observed with H2RG-32-147 in UTR mode. Each data point represents the average RMS deviations of 15-70 measured centroid positions (depending on filter and exposure time) from the reference map positions for a given exposure time. The average is taken across images from exposures in *g,i,y*. The large error in RA at long exposure times is due to the unguided operation of the telescope.

The situation improves in the case where H2RG-32-147 was simultaneously taking long exposures and guiding the telescope (details are explained in Section 3.3). For these exposures, the oscillations in RA are greatly reduced and the drift in DEC is not detectable. Table 5.7 shows a comparison for two 180 second exposures, one taken with H1RG-022 with no guiding and one taken with H2RG-32-147 with guiding. The astrometric error in DEC is comparable for both cases: about 6.5 mas. The error in the RA is larger without guiding, but achieves a similar value to the DEC value when the telescope was guided.

Table 5.7: Comparison of the measured astrometric error for two 180 second exposures. NGC 2419 was observed with H1RG-022 while the telescope was not being guided and SA0 116737 was observed with H2RG-32-147, which simultaneously guided the telescope in guide mode. The exposure time, t_{exp}, is listed along with the error in RA and DEC in units of pixels (pix) and milliarcseconds (mas). Both detectors have 18 micron pixels and the plate scale is 0.227″/pixel.

Field	RA	DEC	t_{exp} (s)	σ_{RA} (mas)	σ_{RA} (pix)	σ_{DEC} (mas)	σ_{DEC} (pix)	Guided
NGC 2419	$07^h38^m08.51^s$	38°52′54.9″	204.0	9.02	0.040	6.75	0.030	No
SA0 116737	$08^h25^m01.06^s$	09°25′33.8″	196.8	6.30	0.028	6.23	0.027	Yes

From these results, it can be concluded that the base level astrometric error for the configurations with H1RG-022 and H2RG-32-147 mounted to the Kitt Peak telescope (0.227″/pixel) is about 6.3 mas. This amounts to 0.028 pixels (slightly greater than $1/40^{th}$ of a pixel). The 0.028 pixel accuracy can, in principle, be used to predict the astrometric accuracy for a different optical system.

5.4.4 Astrometric Results with H4RG-10-007

Results obtained with H4RG-10-007 showed very poor astrometric accuracy. This is primarily because the large tracking errors were not known at the time and the shortest exposure time used on a field suitable for astrometry was 81 seconds. As shown in the previous section, tracking errors are clearly a problem for exposure times of this length. Other factors include a large CDS read noise of 25-30 e^-, electrical pickup from the readout electronics, and a large signal reset anomaly brought about because of a V_{RESET} value of 90 mV.

The best results with this device were about 9 mas in both RA and DEC for observations of M13. Since the pixel pitch is 10 μm pixels for this device and the plate scale was 0.126″/pixel, this translates to about 0.07 pixels. With an identical H4RG model, Dorland et al. measured accuracy of $1/30^{th}$ of a pixel at 180 seconds exposure time and predict $1/40^{th}$ of a pixel for longer, guided exposures [40]. With proper guiding, biasing, and read noise reduction to a level of $10e^-$, the astrometric error for H4RG-10-007 is expected to improve to this level.

5.5 Telescope Guiding in Guide Mode

Modern telescopes use very precise mechanical motors to track celestial objects across the sky while their location changes due to the rotation of the Earth.[4] The job is slightly easier for equatorial-mount telescopes that only need to adjust in right ascension (RA) than it is for altitude-azimuth telescopes that must adjust in right ascension *and* declination (DEC) to compensate for the sidereal motion of the heavens. However, even for equatorial-mount telescopes with the most sophisticated motors available it is an imperfect process. The majority of telescopes must keep their pointing accurate to less than an arcsecond in order to prevent "smearing" of stars and galaxies while they are being imaged.

The widespread solution in astronomy is to use a "guide star" to correct for any errors in the telescope tracking, the idea being that if the sidereal motion is being properly accounted for, a given star should stay in the same physical location (x_o, y_o) on the focal plane of the telescope. If any movement in the location $(\Delta x, \Delta y)$ is detected (by a CCD or other imaging detector), tip/tilt corrections are made in the pointing of the telescope in order to bring the star back to its original location and make $\Delta x = \Delta y = 0$. Usually a sufficiently bright guide star is used to track offsets while a very dim object is being imaged in a long exposure.

If a CCD is being used as a "science sensor" to make the long duration exposure, it cannot be simultaneously used to guide. Accessing the pixels that contain the guide star would result in a destructive read of the entire array (the charge is shifted out of the pixels). Thus, a separate "guide sensor" is required to measure Δx and Δy. In contrast, because of the non-destructive readout and random access capabilities inherent in CMOS detector architecture, a CMOS device can act as both the guide and science sensor simultaneously. In particular, the Teledyne HxRG multiplexer has been designed with a special "guide mode" to accomplish this [22, 100]. Such capability is of great advantage for large focal plane arrays that consist of many detectors since any one of them can be used to track a guide star while it simultaneously participates in the science exposure.

Another benefit of the guide mode operation is that saturated pixels can be reset while the rest of the array integrates light. Resetting the saturated pixels prevents the dark current from hot pixels and photocharges generated by bright stars from blooming into neighboring pixels and eliminates output crosstalk. This is very advantageous for long exposures in which both bright stars and very dim objects are present.

5.5.1 Purpose of Experiment

Laboratory measurements and calculations can provide some indication of the impact that interpixel capacitance (IPC), read noise, nonlinearity, and image persistence will have on the ability of a HyViSI detector to guide a telescope. For instance, convolving the pixel impulse response with a

[4] This motion is referred to as *sidereal motion*.

simulation of the telescope point spread function and folding in read and shot noise will provide an estimate of the expected signal to noise and centroid accuracy for a stable point source. However, it is not immediately clear what subtle effects atmospheric turbulence and scintillation will induce when latent images are present (see Section 7.2). For instance, if atmospheric effects displace the stellar image by a few pixels for several seconds, the latent image that forms might cause errant centroid values for subsequent measurements after the star has returned to its original location on the detector. Reproducing such effects would require very elaborate simulations or a complicated laboratory setup. And in fact, analytical models and simulations used to fit latent images based upon the flux and fluence of the offending image result in large errors (see Section 7.5.2), so this might not be a valid option to pursue.

To directly measure the impact of these second order effects and see whether or not the HyViSI can accurately track a star, we used H2RG-32-147 to guide the Kitt Peak 2.1m telescope. Operating in guide mode, it was able to simultaneously obtain high dynamic range exposures of the full field and track a guide star over long periods of time. For a description of the experimental setup used for guiding, the reader is referred to Section 5.2.8.2. In the following sections, we show measurements made from long exposures with and without guiding and discuss the results.

5.5.2 Results without Guide Mode

To attach the RIDL dewar to the Kitt Peak 2.1m telescope, it was necessary to remove the CCD guider from the instrument rack. As a consequence, the default configuration for our observing runs had the TCS tracking the sky at the sidereal rate with no guiding offsets being issued to account for errors in tracking.

Recognizing that tracking errors would be an issue, the typical cadence we used for our science exposures was a set of 9 or more dithers with short ($<$ 1 minute) exposures in each filter. The tracking errors during these short exposures are very minute, so they do not smear the image, and we only issued offsets between exposures. In software we account for both the controlled dither offsets of the telescope and any errors in the pointing by aligning the dithered images with bright stars.

However, in some cases we took long exposures without any guiding. The slopefit from such an exposure, taken in i band, is shown on the left in Figure 5.10. No dark subtraction or flat field has been applied so hot pixels, defects, and non-uniformity are present in the image. As exemplified by this figure, tracking errors clearly show up in long exposures taken with H1RG-022 and H2RG-32-147. Because the telescope leads or lags the sidereal motion, the light from the stars is smeared out over the pixels so that they have a "jelly-bean" like appearance. We will provide a quantitative description of this in Section 5.5.4.

The bright star in this exposure is SAO 54817, a star with magnitude 6.55 in the V-band and 4.717 in the J-band. The flux falling on the detector was thus somewhere between 400-900 mW/m^2

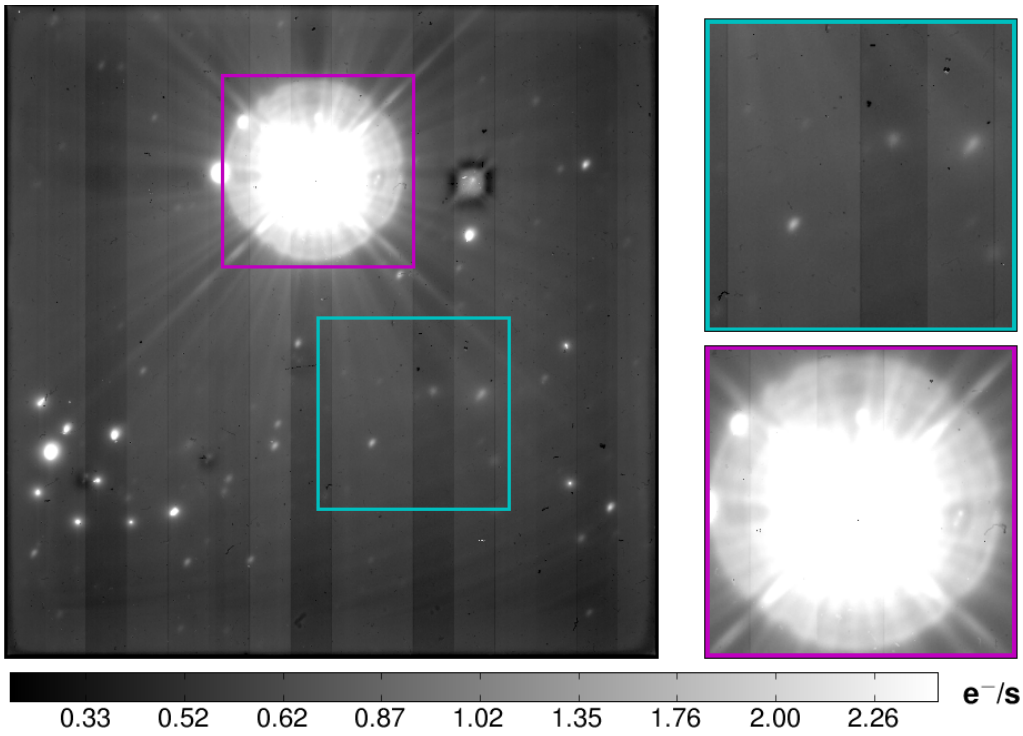


Figure 5.10: (Left) Full frame slopefit to a 1770 s exposure taken with H1RG-022 with no guiding. (Right) A zoomed view of the box enclosing the very saturated star SAO 54817 is shown at bottom right. The light from this star has consumed a circular region of pixels with a radius of 50 pixels. The box at upper right shows an area away from the guide window that contains three stars of irregular shape. Both boxes are 300 × 300 pixels.

in i band. By the end of the exposure, the light from this star has saturated a circular region out to a radius of 50 pixels. For the dimmer stars, image blur due to the tracking error is clearly evident, as evidenced in the top right image of Figure 5.10.

5.5.3 Results with Guide Mode

An example guide exposure is shown in Figure 5.11. It is a slopefit (no dark subtracted or flat field to remove hot pixels and defects) to a 2597 s i band exposure of an $8' \times 8'$ field centered around the star IRAS 09595+2513, which was used as the guide star. The magnitude in V of this star was not found in any catalogs, but is listed as 5.807 in J and 4.594 in K. It has colors of a late M star, so V-K$\sim$6.2, and thus V$\sim$10.8 [105]. Based upon these colors, in i band we expect a flux of 20-170 mW/m^2 was falling on the detector.

For this exposure, 220 reads of the full frame were recorded and in between them 200 CDS frames of the window were read out. The window was 35 × 37 pixels and the time for a CDS pair was

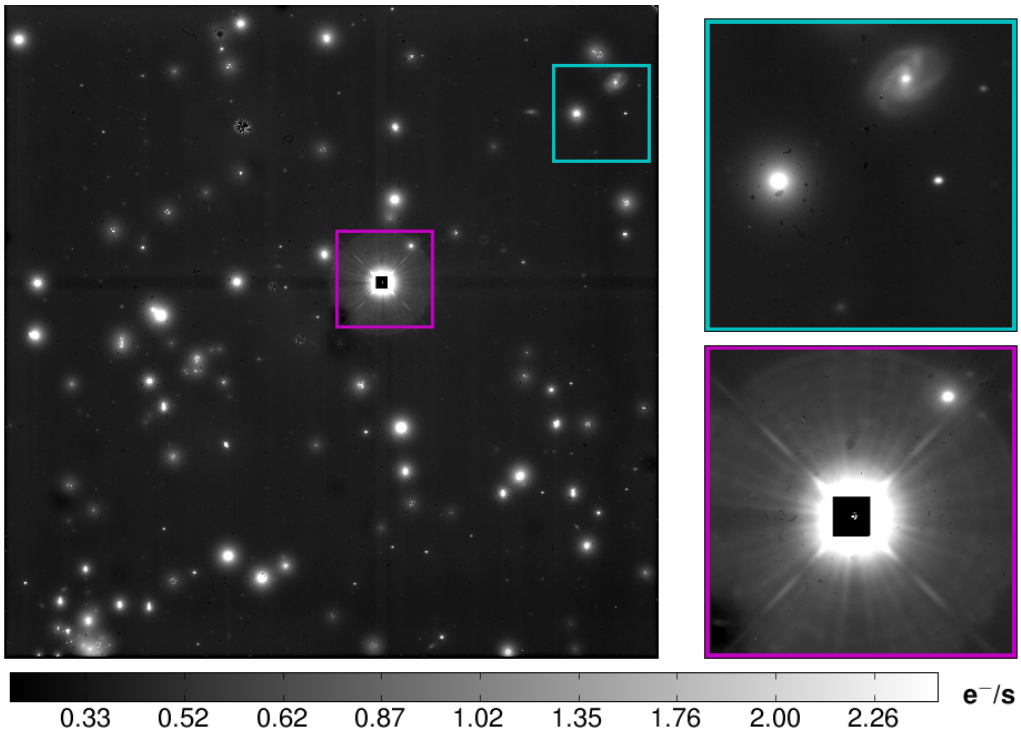


Figure 5.11: (Left) Full frame slopefit to a 2597 s exposure taken with H2RG-32-147 while it was operating in guide mode to guide the telescope. No flat field or dark subtraction was applied. (Right) The box blown up at the bottom right shows the guide window that has been reset constantly during the long exposure at the center of the bright star IRAS 09595+2513. The box at upper right shows an area away from the guide window that contains two faint galaxies. Both boxes are 300 × 300 pixels.

$t_{sw} = 40.9$ ms, which corresponds to approximately 25 Hz sampling. Using the flux estimate above, at this sampling rate the maximum fluence was 0.8-6.8 mJ/m^2. All of the 200 frames were coadded and the centroid of the star was calculated from the final sum using a 9 × 9 box at the middle of the full window, which yielded an effective guiding rate of 0.12 Hz. This is slow in comparison to typical guiding rates; atmospheric disturbances average out on the order of seconds and any centroid displacement after that length of time is attributed to tracking error. However, we did not issue offsets faster than this because offsets less than $0.3''$ were found to be inaccurate, resulting in increased blurring of the image (see Section 5.2.8.2).

5.5.4 Comparison between Exposures with and without Guide Mode

Figure 5.12 shows a close up-image of a star taken from each of the long exposures described in the previous sections along with a quantitative comparison of the stars in terms of their ellipticity and

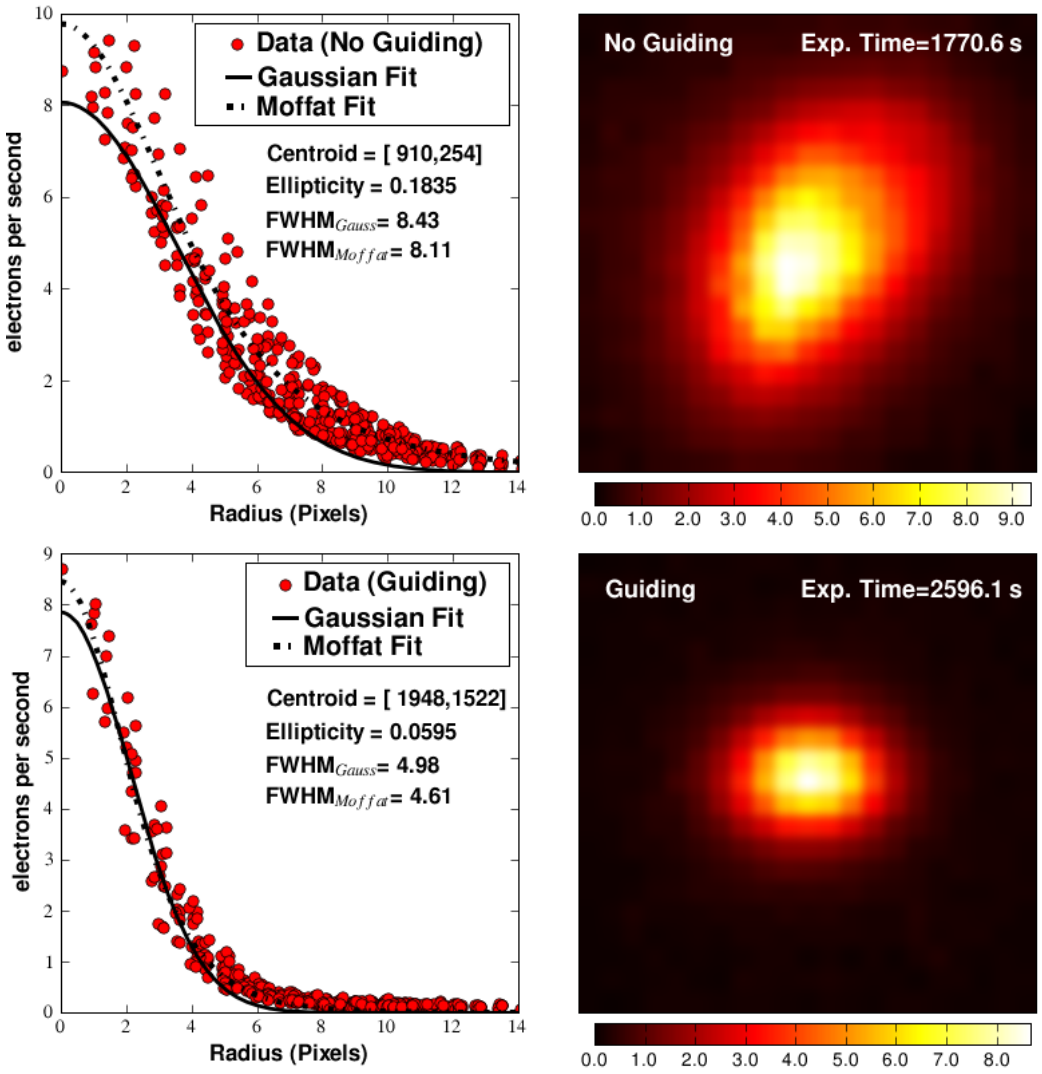


Figure 5.12: The top two figures show a stellar profile from a 1770 s exposure taken with H1RG-022 without telescope guiding. The bottom two show the radial profiles from a 2597 s exposure taken with H2RG-32-147 while it was being used in guide mode to provide offsets to the telescope.

FWHM. We consider these here along with the difference in limiting magnitude in each exposure and the saturation of pixels surrounding the central star.

Ellipticity The stars in the guide mode slopefit do have a slight average ellipticity of 0.06 along the x direction. We note that this is the direction that coincides with right ascension: the direction that was corrected more frequently. However, the ellipticity is 3 times times smaller than it was

when the telescope was not being guided. In the latter case the ellipticity of the stars was clearly induced by tracking errors, particularly the oscillations discussed in Section 5.2.8. It is clear that the guiding succeeded in diminishing the ellipticity of the stars.

Full Width at Half Maximum As shown in Figure 5.12, the FWHM is significantly smaller when the telescope was guided. It is just slightly larger than the seeing of $1.0''$ (4.3 pixels) recorded for the night. The seeing on the night when we imaged SA0 54817 was actually slightly better at approximately $0.9''$. We thus conclude that the smaller image blur is a result of guiding rather than atmospheric conditions.

Limiting Magnitude In terms of faint sources, in the guide mode slopefit we are able to detect galaxies and point sources in the Sloan Digital Sky Survey catalog down to a magnitude of 23.1 in the i-band at the 3σ level. This is much fainter than the limiting magnitude in i-band of approximately 21.5 that we obtained by matching sources in the USNOA-2.0 catalog with those in our exposure with no guiding.

There is, however, a discrepancy in exposure times. To properly account for this difference, we can extrapolate for the limiting magnitude with no guiding based on theory using a standard equation for signal to noise prediction (see Appendix B). The background sky fluxes B and dark current D in the exposures are very similar; the real difference is the radius subtended by the faint stars in the exposures, r_{source}. Figure 5.13 shows the expected limiting magnitude at a certain exposure time for the guiding case where $r_{source} \sim 7.2$ pixels and $B \sim 6.8$ e^-/s, as well as the non-guiding case where $r_{source} \sim 17$ pixels and $B \sim 10.6$ e^-/s.

The behavior is easy to understand intuitively. When the telescope is not tracking a star properly, its light gets distributed into a large area of pixels rather than being concentrated in a small region. The light from the star then tends to blend in with the background sky light and be confused with

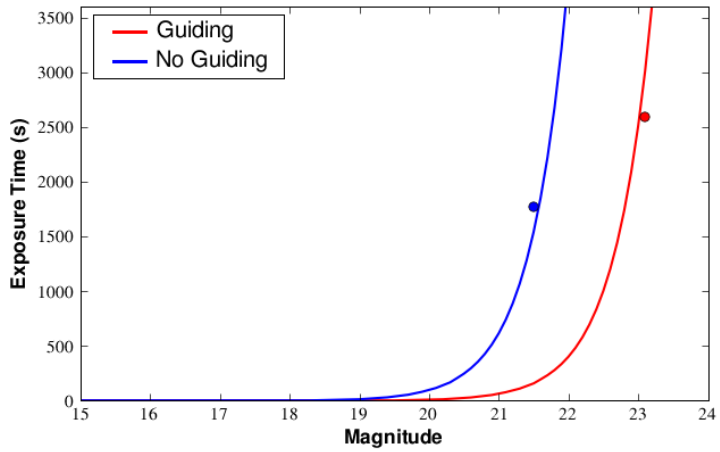


Figure 5.13: Theoretical exposure time required to reach limiting magnitude with a signal to noise ratio of 3 for the cases where the telescope is being guided by H2RG-32-147 (right curve) and where it is simply trying to track stars at the sidereal rate with H1RG-022 (left curve). The data points are also plotted as circles. The difference is due almost entirely to the disparity in PSFs.

dark current, especially near the edge of the star. Taking into account the disparity in r_{source}, we see in Figure 5.13 that for an hour long exposure the guide mode will go about 1.4 magnitudes fainter.

5.5.5 Saturated Pixels

When we implemented guide mode with H2RG-32-147, resetting the pixels in the guide window did not prevent the surrounding pixels from reaching saturation. However, the reason that the surrounding pixels saturated was not because the accumulated charge from the bright star bloomed into its neighbors. Rather, the diffraction of light from IRAS 09595+2513 into the pixels immediately surrounding the guide window is significant enough to generate ~ 50 e^-/s and fill the pixel wells after about 1500 seconds. Further out, the rate of carrier generation is smaller, but still enough to saturate the pixels in 2500 seconds.

We note that the operation of guide mode will indeed prevent charge blooming if the guide window is large enough to contain all of the pixels integrating photocharge from the star. We have also used it successfully in resetting a hot pixel so that it does not spill into its neighbors. One additional benefit is that resetting the brightest portion of the stellar image prevents crosstalk between detector outputs. For N_{out} detector outputs and $N_{pix/row}$ pixels per row, this effect manifests itself as N_{out} repeated images of the star evenly spaced across the columns at intervals of $N_{pix/row}/N_{out}$ pixels (for certain read directions). More details can be found in Finger et al. [100] and in Section 6.2.2.

5.5.6 Summary of Results

The results from the long exposures with and without the implementation of guide mode are shown in Table 5.8. As is expected for a guided camera system, the ellipticity and FWHM are reduced and the limiting magnitude is boosted when we use H2RG-32-147 to guide the Kitt Peak 2.1m telescope.

Table 5.8: Quantitative comparison of long exposures obtained when the telescope was being guided by H2RG-32-147 and simply tracking at the sidereal rate.

Parameter	H1RG-022: No Guiding	H2RG-32-147: Guide Mode
Ellipticity ($\|\vec{e}\|$)	0.18	0.06
FWHM$_{Gauss}$ (pixels)/($''$)	8.70 (2.02)	5.90 (1.37)
FWHM$_{Moffat}$ (pixels)/($''$)	7.24 (1.67)	4.60 (1.07)
Atmospheric Seeing (pixels)/($''$)	3.9 (0.90)	4.31 (1.00)
M_{Lim} in 3600s exposure	21.9	23.3

5.5.7 Discussion

The ellipticity and FWHM of the stars in both guided and unguided mode agree very well with the values predicted using Equation 5.13 along with the nightly seeing conditions and average drift rates for m_{DEC} and m_{RA}. The measurements indicate that image persistence brought about by high flux/low fluence integrations of the guide star in window mode and pixel nonlinearity do not produce significant errors in the centroid accuracy. Based upon the brightest star used to guide, SAO 81129 (J$\sim$3.98) observed in y band, this is true for fluxes < 2 W/m^2 and fluences < 80 mJ/m^2. For future star tracking applications, it should be sufficient to consider the standard parameters of read noise, dark current shot noise, interpixel capacitance, etc. when predicting performance, at least for fluxes and fluences below these limits. It is expected that brighter stars and longer integration times can still be used, but to address the situation concretely, additional measurements need to be carried out. In the meantime, simulations are being attempted to address the issue (see Section 7.5.2.1).

Although it was not discussed in the previous sections, one very important aspect of using the HxRGs in guide mode is the behavior of the outermost rows and columns in the guide window. Regardless of the state of the pixels outside the window, the first column in the window will generally have a slightly higher signal due to settling effects after switching rows in the clocking sequence. This offset should subtract out in a CDS, but the column will suffer from increased noise in the CDS. There is also a large dependence of the outer rows and columns on the state of the pixels outside the window due to IPC, as shown in Figure 5.14. If the rows and columns immediately outside of the window have a large voltage, they will pull the outer rows and columns of the window up in signal. The coupling manifests itself as both an offset and, in cases where the signal is sufficiently high, an increased apparent photocurrent for the outer rows and columns of the window. An easy way to deal with this problem is to make the guide window slightly larger than the region that will be used for the centroid calculation so that the outer rows and columns can be ignored.

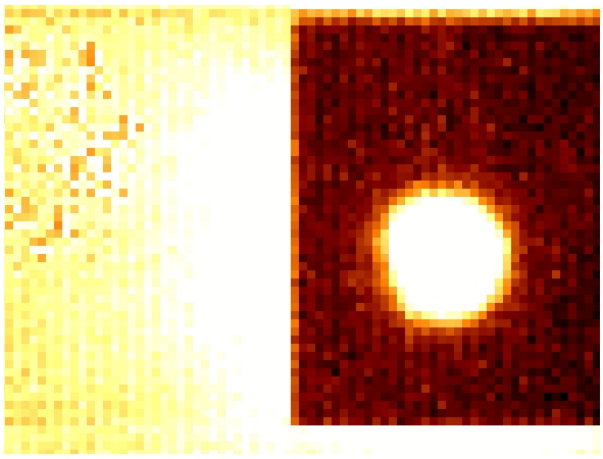


Figure 5.14: An image taken from a full frame read of the detector while it was operating in guide mode. When the pixels inside the guide window (dark pixels) are continuously reset while the pixels outside integrate a large photocurrent (bright pixels), the outer rows and columns of the guide window see an increased signal due to IPC. Making the window larger than the light envelope of the star prevents the outer columns and rows from contaminating the centroid calculation.

5.6 Near Infrared Response

Conventional back thinned CCD detectors in astronomy have a weak red response and virtually no response at wavelengths above 800 nanometers. Since the interaction depth of near infrared photons is about 25 nm at 750nm (and grows rapidly with increasing wavelength), thinned CCDs simply lack the thickness to absorb them efficiently. As such, the Near-Infrared (NIR) regime above 750 nm has been the territory of non-silicon based hybrid CMOS infrared detectors.[5] From a practical standpoint, camera systems using these detectors are more involved since they require temperatures below 80 K (20-80 K for HgCdTe and 30 K for InSb [18]). HyViSI detectors extend into the NIR without requiring such heavy cooling.

The y filter provides a perfect opportunity to test the HyViSIs in between the traditionally classified "optical" and "infrared" bands [92], and the Orion Nebula provides a perfect target since it emits strongly in both bands. The image shown in Figure 5.15 is a three color RGB mosaic of the heart of the Orion Nebula (M42) taken with H1RG-022 through the g (blue), i (green), and y (red) filter set. It was the product of combining the slopefits from 45 dithered exposures in each filter. Each exposure in g and i was 5.1 seconds long (15 reads UTR) and the ones in y were 10.2 seconds (30 reads UTR). The scales used in combining the final mosaics in each filter were set to match the quantum efficiencies measured by Dorn et al. [49] – 0.82 (g): 1.00 (i): 0.15 (y). In other words, 10 e^-/s in the i band has the same bitmap value as 1.5 e^-/s in y and 8.2 e^-/s in g.

The purpose of presenting this image is not to introduce new quantitative or scientific analysis of this region. The Orion Nebula has been studied extensively in all wavelengths, particularly the visible and infrared [41]. Rather, the goal is to verify that the quantum efficiencies reported from HyViSI laboratory measurements – remembering that techniques like photon transfer have resulted in QE $> 100\%$ in some cases – are accurate. Qualitatively, the results are exactly what is expected from measurements made in previous infrared and visible studies: the bright O and B stars of the Trapezium cluster show up prominently in all three filters, the HII region glows strongly in g and i and hardly at all in y , and the stars obscured by dust and gas hardly show any signal in g but glow strongly in the near infrared y band. This provides confirmation that the quantum efficiencies measured in the laboratory are valid and the HyViSI performs as expected at 1 micron. Again, the uniqueness of this is that optical and near infrared observations were made with the same silicon based detector operating at 160K.

The bright horizontal banding and ghost images that appear on every output in the rows with a very bright star are artifacts of the detector and ARC electronics. Since the H1RG was operated in 16 output mode, there are 15 ghost stars for each bright star in an individual exposure. The mosaic processing helps to reduce their amplitude and blend them together. The ghosts and banding are examples of output crosstalk, which will be considered in depth in the next chapter.

[5] There seems to be some disagreement between the exact definition of the various infrared bands in the literature. For the purpose of this study, the NIR band will be considered to be 750-1400 nm.

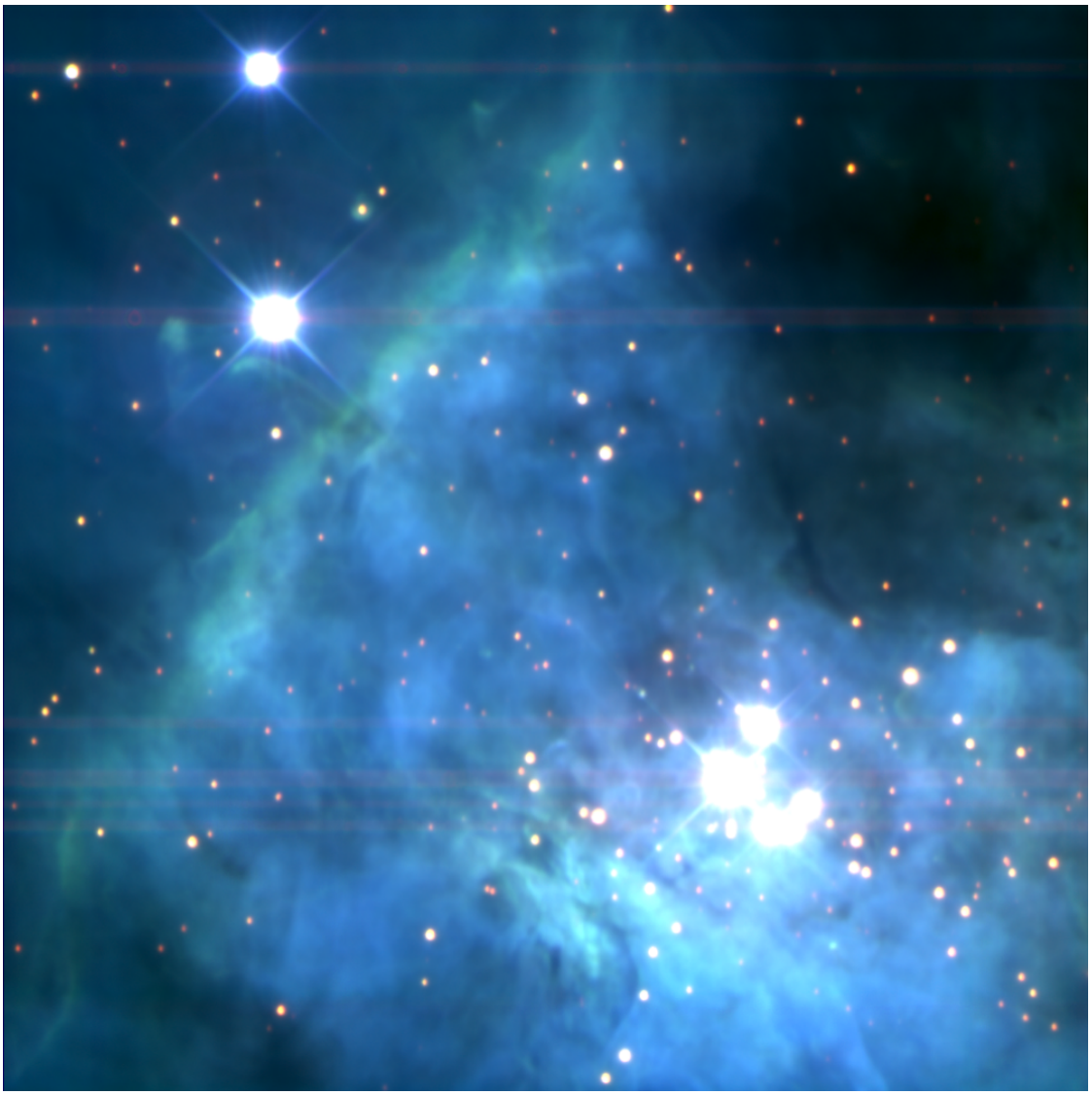


Figure 5.15: Tri-color image of the Orion Nebula (M42) taken in g, i, and y filters. The y band is shown in red, g in blue, and i in green. Other exposure details are listed in the text. The vast majority of stars are not visible in g and i because of obscuring gas and dust, but show up in y because the infrared wavelengths penetrate the dust.

Chapter 6

Pixel and Electronic Crosstalk

In an ideal imager, each pixel is an independent photon detector. Whether current *from the pixel* or charge *in the pixel* is measured, the electrical content and activity in one pixel does not affect its neighbors (or any other pixels across the detector) and vice versa. In a real detector, this is not the case. Electrical signals couple to one another through direct means such as charge spilling from one pixel to its neighbor or indirectly through capacitance or inductance between current carrying buses.

In the case of hybrid CMOS detectors, there is *crosstalk* between neighboring pixels during charge collection, storage, and possibly readout. This is referred to as *pixel crosstalk*. There is also coupling of the signals from pixels that are readout simultaneously through separate channels on the detector and direct coupling between detector bias currents or voltages and pixel signals. These effects are referred to as *electrical crosstalk*. Both pixel and electrical crosstalk will be discussed in the following sections.

6.1 Pixel Crosstalk

During charge collection, electrons and holes in the detector bulk can diffuse laterally into neighboring pixels before they arrive at a pixel well. After collection, when the charge is stored in the collecting node, capacitive coupling between neighboring nodes will result in additional crosstalk. We report here on a further coupling mechanism in which charge carriers appear to be moving between pixels. Because the charge is actually being transferred from one pixel to another, we refer to this as **Interpixel Charge Transfer** or **IPCT**. In this chapter we begin by giving a basic overview of these different crosstalk mechanisms. We then present data from cosmic ray events, Fe^{55} measurements, and single pixel reset tests that illustrate how pixel crosstalk manifests itself in HyViSI detectors.

6.1.1 Mechanisms of Pixel Crosstalk

The term pixel crosstalk applies to a number of separate phenomena that may be occurring in the detector or ROIC simultaneously. Crosstalk can occur between the time photocharges are generated and stored since carriers generated directly above a pixel location can diffuse to neighboring pixels. It can also occur after the charge has been collected via capacitive coupling of neighboring pixels, direct charge transfer through a conductive channel, or thermionic diffusion of carriers over an interpixel potential barrier. The last of these occurs only under rare circumstances and will be saved for discussion until Section 6.1.4.3, where it is relevant. The other forms of crosstalk will be briefly summarized here as a basis for understanding the experimental results that follow.

6.1.1.1 Lateral Charge Diffusion in HyViSI Bulk

After an electron-hole pair is generated in the bulk of the HyViSI, the hole must travel to the collection node before it is counted as signal charge. On its way, it will undergo some lateral diffusion in the bulk material. If there is both an undepleted and depleted region in the bulk or if the detector layer is sufficiently thick, predicting where, on average, the hole will end up is a complicated problem. Pavlov et al. [106] treat this problem in great detail for the case of x-rays in thick CCD detectors, and many of the results are applicable for a SiPIN diode detector. If the detector is an overdepleted state, as the HyViSI should be for astronomical applications, the problem becomes much simpler. O'Connor et al. [107] provide the following expression for the mean diffusion radius of the holes at the collecting surface by assuming the bulk is high resistivity ($\rho > 10,000\Omega$-cm) and overdepleted:

$$\sigma_D = d\sqrt{\frac{2kT}{qV}}, \tag{6.1}$$

where d is the detector thickness, k is Boltzmann's constant, T is the temperature, q is the electron charge, and V is the bias voltage across the detector. While this expression is the result of a simplified treatment, it reveals the two primary "knobs" we can use to reduce the charge diffusion: the detector temperature and the backside voltage V_{SUB}. Minimizing the lateral diffusion is essential for reducing the PSF to levels suitable for astronomy.

6.1.1.2 Interpixel Capacitance (IPC)

It has been suggested by Moore et al. [24] that interpixel capacitance plays the primary role in coupling the voltages on neighboring pixels in SiPIN detectors. In this model, a pixel at i, j is described as a capacitor $C[i, j]$ that receives a signal $Q[i, j]$ that corresponds to the photocurrent entering the pixel. The detector is modeled as a discrete linear shift-invariant (LSI) system that outputs an array of voltages

$$V[i, j] = Q[i, j] * h_c[i, j], \tag{6.2}$$

where * is the 2-d discrete convolution operator and $h_c[i,j]$ is the impulse response of the collection array. Ideally, $h_c[i,j] = \delta[i,j]/C_{node}$, since the discrete delta function represents no coupling between neighbors. However, in this model there are coupling capacitors between pixels so that $h_c[i,j]$ has contributions from neighbors, i.e. pixels with $i \pm 1$ and $j \pm 1$.

The level of coupling is usually measured with the parameter α, which is the fraction of charge that appears in a neighboring pixel due to IPC. For instance, if only symmetric coupling to the four nearest neighbors is considered, the center pixel loses 4α of its charge and α of that charge will be measured in any one of the four neighbors. For the asymmetric coupling that is always observed in HxRG detectors, each neighbor will have a unique value $\alpha_{i,j}$.

Interpixel capacitance could potentially occur in the ROIC or the detector substrate, or even between these two layers at the level of the indium bump bonds. There is very strong evidence that it does not take place in the ROIC, though. If the coupling took place in the ROIC, we would expect to see the effect between science pixels at the edge of the detector and the neighboring reference pixels. However, in all cases: cosmic ray hits, hot pixel volcanos, single pixel reset experiments, etc., we do not see the reference pixel signal affected by the signal on the neighboring science pixels. We are therefore forced to conclude that the coupling occurs in the bulk of the detector or between the ROIC and detector layer.

The coupling in the IPC model adequately describes the degradation in modulation transfer and point spread functions of the detector. These, in turn, result in an overestimation of the QE of the detector. However, in this model it is assumed that $h[i,j] \geq 0$. Physically, this means that charge entering a given pixel does not leave the pixel or get destroyed through recombination. A different mechanism is thus needed to describe the observed behavior in the HyViSI devices, where charge appears to be leave and enter pixels.

6.1.1.3 Interpixel Charge Transfer (IPCT)

IPCT manifests itself in the HyViSI detectors in a number of situations. The place where it shows up most strongly is in cosmic ray hits. After the initial spike in signal attributed to a deposit of charge by the cosmic ray, we see a decay in the signal of the central pixels and a rise in the signal of one or more of their neighbors. It also shows up in Fe^{55} gain tests, where we see the pixels hit by an x-ray gain and lose signal in a similar fashion.

The precise details of how IPCT occurs are not yet known, in part because we do not know the relevant details of the detector (e.g. the pixel doping, passivant thickness, etc.). However, our initial speculation that it arises at the surface interface between the SiO_2 passivant and the lightly doped bulk appears to be right. In the latest of the HyViSI sensors we tested—one for which this surface interface was treated differently—the IPCT was dramatically reduced. Nevertheless, in the following sections we will present the empirical results that show the signature of IPCT as it appears in the majority of the devices we tested.

6.1.2 Measurement via Cosmic Rays

A striking example of Interpixel Charge Transfer is seen when "cosmic rays" make their way into the detector. The term "cosmic rays" is put in quotes because, as Groom points out [108], in ground based detector systems many of these events are generated by particles that do not have a cosmic origin. The genuine cosmic rays consist primarily of relativistic muons produced by secondary meson decay, and at higher elevation, small fraction of protons, helium nuclei, electrons, positrons, and photons [109]. The other, non-cosmic events are attributed to local sources on the ground such as gamma rays from radioactive decay or beta particles. Smith et al. [110] show evidence that Compton recoil electrons from gamma rays passing through fully depleted CCDs are the primary source of events that deviate from a straight "muon track" or a "spot" generated by a normally incident particle. To validate this, we have performed simulations with the GEANT4 particle physics simulation package. The simulations not only show that Compton recoil electrons indeed give rise to "worm" tracks in thick silicon detectors, but that impact ionization from electrons liberated by a muon passing through the detector can cause secondary tracks ("Delta Rays") or large spots at random locations along the straight track. The variety of event types seen in the simulations matches well to the data collected with the HyViSI detectors.

While the physical interactions that comprise the cosmic ray events differ greatly, they all share one thing in common: they take place on a timescale ($t_{event} \sim 10^{-9}$ s) far shorter then the frame time of the detector, even for the case where a very small window of pixels is being read out ($t_{frame} \sim 10^{-3}$ s).[1] Thus, the expected signature of one of these events should be a large step in signal vs. time for each affected pixel at the read r following the time when the event took place. After the step, assuming the collecting well is not full, the node should continue to collect photo or dark carriers at the same rate it did before the particle passed through the detector. This sort of behavior, shown in the upper left signal vs. time plot of Figure 6.1, is observed in cosmic ray hits that occur during dark and illuminated exposures at temperatures at or below about 130 K.

At temperatures above 130K, however, the HyViSI detectors quickly depart from this ideal behavior. The other plots in Figure 6.1 show that for these temperatures, the signal decreases after the cosmic ray has passed through the detector. For reasons that will be made clear later in this chapter, it is assumed that the decrease in signal is due to a net charge loss in the pixel. As can be seen in the figure, the total charge lost in a given period of time, Q_{lost}, increases with increasing temperature while the time it takes for the loss decreases. This behavior begins to change around 180 K. At this point it appears the dark current overrides the charge loss and causes the pixel to begin integrating holes more quickly. While the plots in Figure 6.1 show the behavior at $V_{SUB} = 5$ V, the charge loss occurs at all voltages tested, all the way up to 40V. There are some subtle effects

[1] The incident particles are traveling through the silicon at relativistic speeds, so the event time should be dominated by the time it takes for liberated holes to drift to the integrating node, t_{col}. Assuming an average electric field of $E = 10^3$ V/cm and hole mobility of $\mu_p = 10^4$ cm^2/s/V in the high purity silicon, $t_{col} = 1$ ns for a hole that starts at rest and has to traverse the entire depletion region (100 μm).

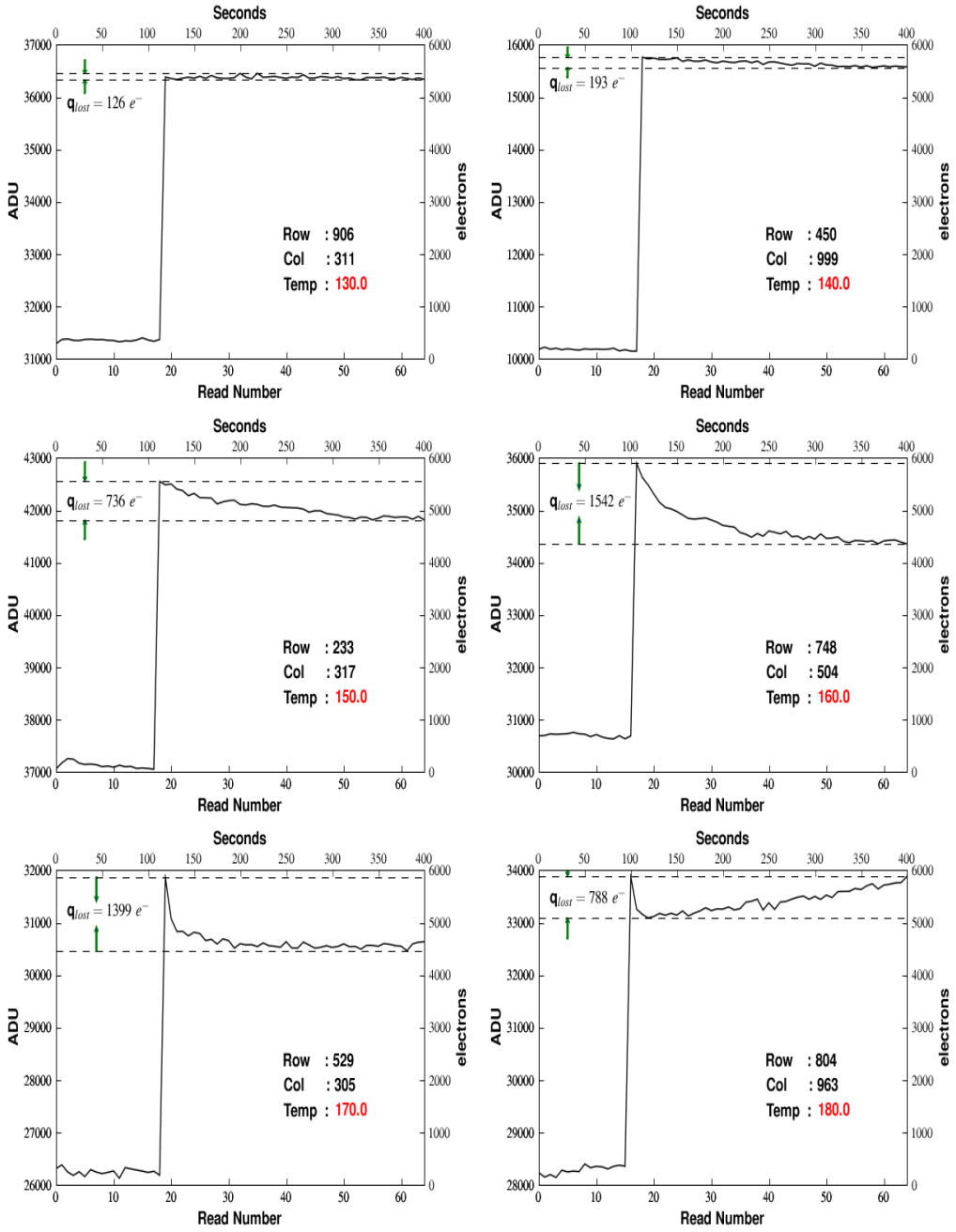


Figure 6.1: Cosmic ray events in H1RG-022 found in 400 second dark exposures at temperatures from 130-180 K and a backside voltage of $V_{SUB} = 5$ V. The plots show the signal as a function of time in up the ramp integrations. The right-hand scale of the y-axes runs from 0-6000 e^- for all plots. A substantial charge loss, q_{lost} in the pixels is clearly evident at 150 K and even at 180 K where the dark current is high.

for voltages where the detector is not fully depleted, and these will be discussed in a later section. In this section, the focus will be on the temperature dependence, which is the most dominant factor.

In the range of 130-180 K, the decay of signal after the cosmic ray hit displays an exponential behavior. The rate of decay is proportional to the amount of charge the pixel has yet to lose before the total loss reaches the value Q_{lost}. In other words, if the pixel has lost the equivalent of q_{lost} holes at time t after the cosmic ray hit and will eventually lose Q_{lost} holes before it starts to integrate positive signal again, then the rate of hole loss can be expressed as

$$\frac{dQ(t)}{dt} \propto -(Q_{lost} - q_{lost}). \tag{6.3}$$

As stated earlier, the constant of proportionality clearly depends on temperature. However, there is some evidence that at a given temperature the constant of proportionality is not actually a constant, but rather depends on the conditions in the neighboring pixels as well as the physical location of interaction in the detector. While important for determining the mechanism that causes the charge loss, this deviation will be ignored and treated as a second order effect for the moment. To first order, if a time constant of e_n is chosen, then the charge lost in a time t after the cosmic ray hit will be

$$q_{lost}(t) = Q_{lost}(1 - e^{-e_n t}), \tag{6.4}$$

and the charge measured in a read of the detector, Q_{meas}, at time t will be given by:

$$Q_{meas}(t) = Q_{dep}(t = 0) - Q_{lost}(1 - e^{-e_n t}), \tag{6.5}$$

where $Q_{dep}(t = 0)$ is the number of holes initially deposited in the pixel. Note that because each pixel is sampled at a fixed interval, $Q_{meas} = Q_{dep}$ only if the cosmic ray interacts immediately before the pixel is read. In all other cases the first measurement of the pixel will yield a result $Q_{meas} < Q_{dep}$, and the charge deposit will be underestimated. The impact this has on x-ray energy measurement will be considered in the next section.

At a given temperature, there is also a near-linear relationship between the amount of charge deposited in the pixel and the amount of charge it loses, i.e.

$$Q_{lost} \propto Q_{dep} \tag{6.6}$$

Like the time constant, the deviation from linearity appears to be dependent on conditions in the neighboring pixels. It also depends on the depth at which the charge carriers were generated, as evidenced by long muon tracks in exposures where the detector bulk is not fully depleted. For the vast majority of decays, it is observed that

$$Q_{lost} < m(T) \cdot Q_{dep}, \tag{6.7}$$

where $m(T)$ is a temperature dependent slope that places an upper bound on the amount of charge the pixel will lose before is starts to integrate once again.

It is important to note that this effect is not particular to any one detector tested. **The charge loss is clearly observable in H1RG-018, H1RG-022, H2RG-001, H2RG-32-147, and H4RG-10-007**. In an attempt to reduce the effect in recent designs—and also for the purposes of reducing the magnitude of IPC in HyViSIs—a different surface treatment was applied to improve the Si-SiO_2 interface at the front surface of the detector (Bai, private communication). In H2RG-148, which is the most recently fabricated detector that we tested, the effect is significantly reduced.

6.1.2.1 Description of Cosmic Ray Experiment and Analysis

To measure the amount of charge lost, cosmic ray events were characterized from a large set of 1060 second up the ramp dark integrations obtained with H2RG-001. The exposures each consisted of 100 reads and were taken in 4 output mode, yielding a frame time of $t_{frame} = 10.6$ s. They were recorded at temperatures from 100 K to 180 K at 10 K intervals. Once all temperatures were probed, the bias voltage was adjusted so that data was recorded at V_{SUB} =5, 10, and 15 volts. Ample time was given for the detector to settle whenever the temperature or V_{SUB} was adjusted.

Detection of the events is done with a robust method that uses both the temporal information contained in the time axis of the datacube and the two dimensional spatial information in each frame. First, a median dark is formed for each temperature to be used for subtraction of the average dark current in each exposure. For a given exposure, starting at the lower left hand pixel of the array, the pixels ramps are sequentially fitted with a line as discussed in Section 5.2.4. If a discontinuity in the line above a certain threshold is detected, it is flagged as a cosmic ray. For the surrounding pixels that have not yet been evaluated, the difference in signal from the read before, r, and after, $r + 1$, the hit are taken. This difference is referred to as *the step*. If the step is greater than $3\sigma_{rn}$, then that pixel is added to the event and the search is continued with its nearest neighbors. After a pixel has been flagged or slope fitted, it is masked with a zero so that it is not counted as part of an event found later in the array.

Once all of the contiguous pixels with a difference of $3\sigma_{rn}$ have been evaluated for a given event, the entire collection is categorized by global parameters such as its morphology (similar to the methodology described in Groom [108], we classify them as straight muon track, worm, isolated spot, Gaussian spot, delta rays, and unclassified), its minima and maxima in signal increase, etc. Then, for each pixel in the event, the step in signal from r to $r + 1$ is recorded as Q_{dep}, and the decay after the step is analyzed to find a minimum and maximum value. After a clear minimum has been found, Q_{lost} is recorded as the difference in signal between the step and the minimum after the step, and the decay is fit with Equation 6.4 to yield an estimate of e_n. All of the recorded values for the individual pixels and the global event are then recorded to a database and the search proceeds to look for the next event.

6.1.2.2 Characterization of Decays

The results of the charge decay characterization are conveyed in Figures 6.2 and 6.3. The events in the scatter plot of Figure 6.2 are ones that occurred between 200 and 500 seconds after the start of the integration and had a minimum Q_{dep} of 500 e^-. The range was chosen so that a good estimate of the slope both before and after the cosmic ray hit can be obtained, and to allow significant decay in signal for the lower temperature range. At each temperature, a line with slope $m(T)$ and intercept of $b = 0$ (the charge lost for zero charge deposited must necessarily be zero) is fit to the points to indicate the proportionality described by Equation 6.6. The plot clearly shows that Q_{lost} increases with Q_{dep} at all temperatures. It also illustrates a large scatter at low values of $Qdep$

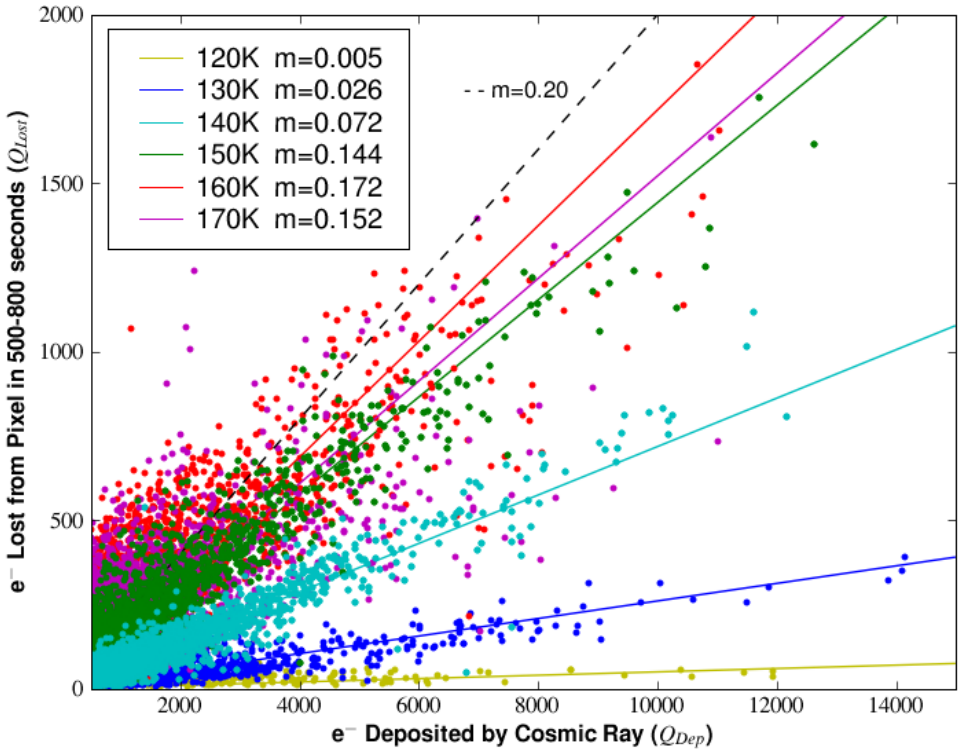


Figure 6.2: Scatter plot showing the amount of charge lost by a pixel vs. the amount of charge initially recorded in the pixel after a cosmic ray hit. The hits are chosen such that the time between the hit and the last measurement in the ramp is about 500-800 seconds. The lower limit to the x-axis is 500 e^-. The temperature and slope of each line is indicated in the legend and the black dotted line shows the slope of $m = 0.2$. Note that, although covered by the foreground plots, the scatters of the points at temperatures above 160 are very large at lower x values.

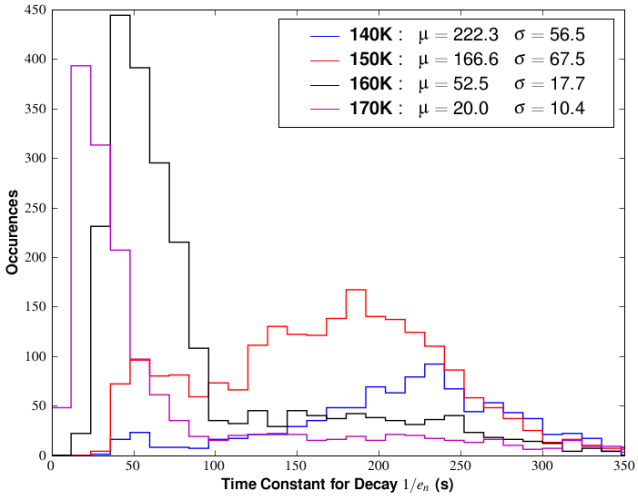


Figure 6.3: Histograms of the time constant, $1/e_n$, for the decay of signal after impact by a cosmic ray at temperatures between 140 and 170 Kelvin. The decays were measured in 1060 second dark exposures for cosmic ray hits that occurred in a time interval 200-500 seconds after the start of the exposure. The mean time constant, μ, and standard deviation, σ, are indicated in the legend for each temperature.

for temperatures above 160 K. The most reasonable explanation for the scatter is that the delay between the hit and the first time the pixel is read results in a large discrepancy between $Q_{dep}(t = 0)$ and Q_{meas}. This explanation is supported by analysis of Fe^{55} events, which will be presented in the following section, and darks taken in window mode with a frame time of $t = 1.6$ s. Although the amount of cosmic rays observed in window mode is very limited due to the small amount of surface area in the window, the value of $m(T)$ at 170 K is observed to increase to about $m = 0.20$ and the scatter is reduced. The dashed line in the plot shows this slope.

For temperatures below 140 K, the decay time may be considerably larger than 800 seconds, so there is some question as to whether the ratio $m(T)$ accounts for all of the charge that will be lost before the pixel begins to integrate in the positive direction. For example, if the exposure was allowed to continue for 1 hour, it is possible that at 130 K the ratio of Q_{lost}/Q_{dep} would increase to the maximum value of $m = 0.20$ observed at higher temperatures. However, as shown in the legend of Figure 6.3, the mean time constant μ_{e_n} at 140 K is short enough to allow at least 80% of the charge to decay in 500 seconds. And the histograms in the plot show that above 140 K, the time constants rapidly decrease with increase in temperature, so there is more than enough time for the full decay to take place in 500-800 seconds. Thus, it can be concluded that the decrease in $m(T)$ for lower temperatures is not simply a result of cutting off the integration before the decay has ceased, but rather that **less charge is effectively lost by the pixels at lower temperatures**.

6.1.2.3 Effect on Dark Rate and Spatial Dependence

There are several other notable trends in the data that may provide some insight into the physical origin of the charge decay. One very interesting trend seen in the decays at temperatures where the

dark current is greater than about 0.5 e^-/s (almost exclusively for temperatures above 150 K) is that the dark rate changes after the decay. In some cases it increases and in other cases it decreases after the nonlinear decay subsides and the signal rate once again becomes linear. The change is typically no more than ± 0.2 e^-/s.

Another noteworthy point is that the pixels tend to lose the least amount of charge near the edges of the detector. Although only a handful of these events are present, about nine out of ten of them do not show signal decay, even for large values of Q_{dep}. Pixels that were less than 150 pixels away from the perimeter of the detector were intentionally left out of the plots in Figure 6.2 for this reason.

Finally, by examining the decays for long muon tracks when the detector is not fully depleted, it appears that the charge loss is dependent on the depth of interaction in the detector. Pixels at the narrow end of the track–where the muon crossed the front side–lose almost no signal at all or gain signal. The pixels at the wide end of the track–where the muon crossed the back side and generated electrons in the field free region–see a decay that exhibits the average values of m and e_n. This is depicted in Figure 6.4, which shows the ramps for two pixels at opposing ends of a muon track. Knowing which side of the detector the muon entered, and assuming that it ionizes electrons throughout its passage in the detector, one can approximate the depth at which it passed over a given pixel. Taking $z = 0$ as the front side of the detector, if the muon entered the detector at (x_o, y_o) and left a track of length L_{track}, the depth at which it passed over the pixel at (x, y) is given

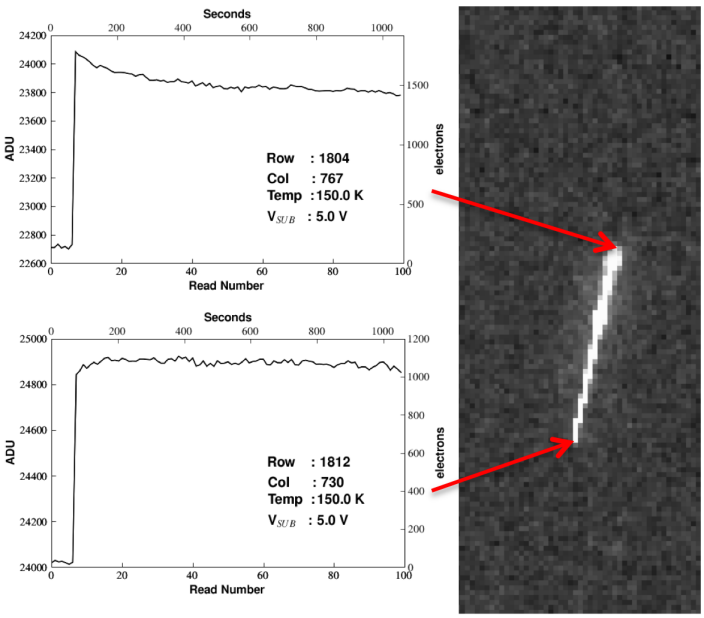


Figure 6.4: The pixels affected by long muon tracks (image at right) show a dependence of charge loss on distance along the track, which is directly related to the depth of the muon in the director when it passed through. The image was taken with an underdepleted detector ($V_{SUB} = 5$ V), so the depth can be mapped to the width of the track. (Top left) Pixels near the back side show the typical loss and decay rate.
(Bottom left) Pixels near the front side (closest to the p^+ implants) show very little signal loss, and in some cases, a gain in signal.

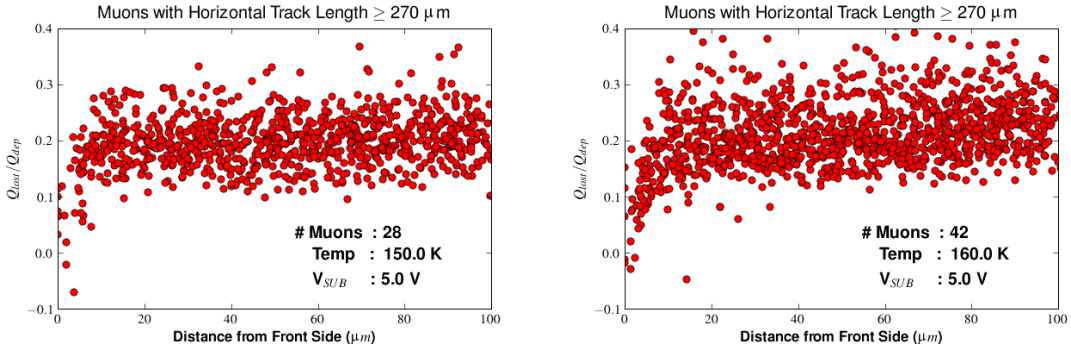


Figure 6.5: Both plots show the ratio of Q_{lost}/Q_{dep} against the calculated depth of interaction (see text) for extended muon tracks that had a clear disparity in width at opposing ends of the track. A negative ratio indicates that the pixels actually *gained* charge, evidenced by a nonlinear increase in signal after the muon hit. The number of muons sampled, along with the temperature and substrate voltages are listed in the plots.

by

$$z = \frac{\sqrt{(x - x_o)^2 + (y - y_o)^2}}{L_{track}} \times 100\mu\text{m}. \tag{6.8}$$

As part of the search algorithm described above, this calculation was carried out on all of the cosmic ray events classified as muons for which a clear disparity in width at opposing ends of the track existed. Figure 6.5 shows the results for the muon tracks that had a track length greater than 14 pixels (252 microns) at temperatures of 150 and 160 K. The figures clearly show that the charge loss is diminished for pixels at locations near to where the muon passed through the front side of the detector ($z = 0$). A negative ratio m corresponds to the pixels gaining a positive charge Q_{lost} as opposed to losing Q_{lost}, which indicates that some of them gain holes instead of losing them. At a depth of ~ 10 μm, the ratio is seen to level off around $m = 0.20$. This might suggest that lateral diffusion of holes into the gaps between the p^+ implants plays a strong role in the charge loss and a threshold height $z_{threshold}$ exists where the lateral diffusion brings the charge carriers all the way into the neighboring pixels. But interestingly, the data at 160 K shows a slight positive slope all the way up to $z = 100$ μm.

When the detector is fully or over-depleted ($V_{SUB} > \sim 7$ V), there is no diffusion region at the back side of the detector and in many cases it is no longer possible to discern which end of the muon track is which based upon the width. However, the same trend of the pixels at one end losing very little charge and at the other end losing the average still exists. The effect is still strong all the way up to 20 volts; higher voltages were not examined.

6.1.2.4 Transfer of Charge Between Pixels

There is evidence that the charge lost in the decays is not simply leaving the detector through the backside contact or recombining in the detector, but that it is being electrically transferred to neighboring pixels. When a decay occurs in a particular pixel, often times one or more neighboring pixels see an opposing rise in signal. Without exception, the overall signal in the pixel that decays is larger than the signal in the pixel that rises. And in nearly all cases observed, the pixels that lose holes lie at the center of the event and the ones that gain holes lie near the boundary. An example is shown in Figure 6.6. Note that this response cannot be attributed to interpixel capacitance. If the coupling was governed by IPC, all neighboring pixels should see signal shifts in the same direction [24].

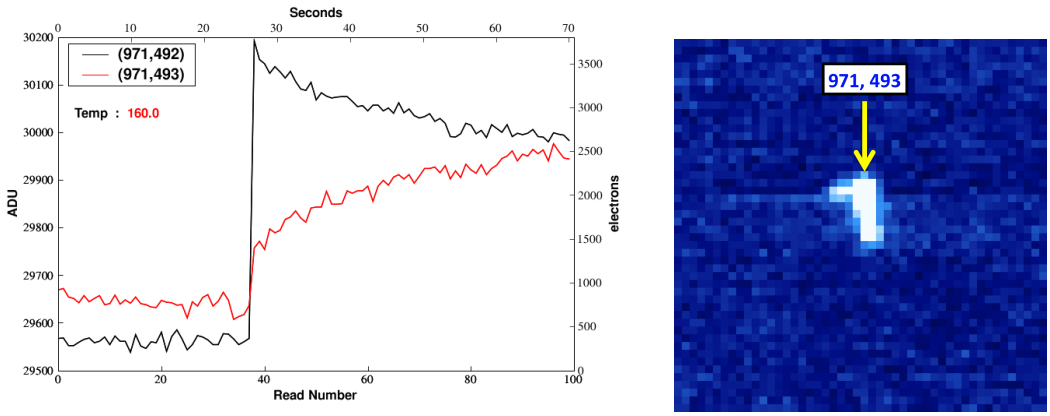


Figure 6.6: (Left) The ramps of two pixels hit by a cosmic ray. The plot in black shows the pixel that absorbed and then lost the most charge. The plot in red shows its nearest neighbor, which appears to be the recipient of some of the lost charge. (Right) Difference image of read before and after cosmic ray hit from an illuminated exposure with $V_{SUB} = 3.0$ V. The arrow points to the pixel at (971, 493) and the one below it is (971, 492).

One might imagine that this coupling mechanism is taking place in the ROIC. For instance, the electronics might be coupling the analog output of neighboring pixels due to settling effects. However, in cases where the cosmic ray hits at the edge of the detector, an example of which is shown in Figure 6.7, no evidence of any charge sharing between the science pixels that integrate the charge dumped by the cosmic ray and their neighboring reference pixels is seen. If the charge sharing was taking place in the multiplexer then coupling between the science and reference pixels should be observed. We therefore conclude that **IPCT must occur in the photosensitive detector layer or in the oxide layer**.

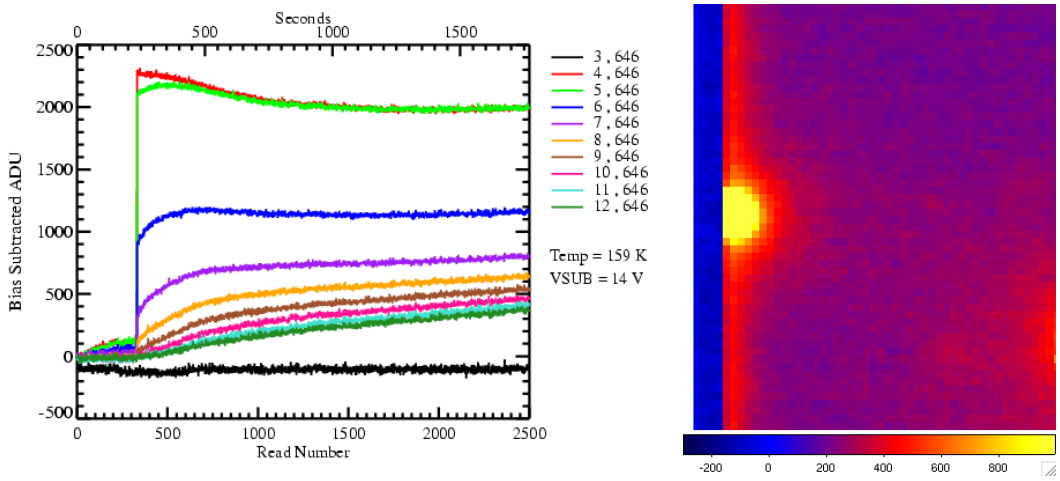


Figure 6.7: (Left) The ramps of nine pixels in order of increasing x value during an exposure where a large event occurred. The pixel with x=3 is a reference pixel. It shows the same behavior as all other reference pixels in nearby rows and columns but does not show any coupling to the science pixel after the accumulation of charge from the large cosmic ray event. (Right) The last minus first read of the integration from which the ramps were taken. The image shows that this event was a very energetic one, probably induced by an x-ray or gamma ray.

6.1.2.5 Physical Explanations

Based upon the temperature dependence of m and e_n, the depth dependence of Q_{lost}, and the more subtle effects just described, we suggest the following explanation for what is physically happening in the detector. A fraction of holes generated by the cosmic ray are trapped at the Si-SiO_2 interface. Holes that were generated high above the pixel implant at an x, y location offset from the pixel reach these traps more easily than those generated very near to the implant because the vertical electric field gives them a high velocity, making it more difficult for the horizontal field near the implant to "steer" the hole directly to the implant. The trapped holes cause a rise in the signal for p^+ implant to which they are closest. When emitted from the traps, the holes may drift away from this implant if there is a conductive channel near the surface. Based upon Gauss's law and the relationship between the electric field and potential (see Section 7.5), the potential (signal) at the implant will decrease as the hole moves further away. At the same time, the pixel towards which it is moving will see an increase in potential (signal).

The ease with which this happens will depend on the number of available hole traps ($E_T < E_F$), which will go like $\exp([E_T - E_F]/kT)/(1 + \exp([E_T - E_F]/kT)$, and the trap capture coefficient, which goes like $T^{1/2}$. The trapping rate efficiency will thus increase with increasing temperature, and the detrapping time will get smaller. This model is further supported by the fact that IPCT was greatly reduced in H2RG-148 after the Si-SiO_2 surface was treated differently during fabrication.

6.1.3 Measurement via Fe^{55}

Ionizing radiation provides a means to measure the coupling between neighboring pixels in a detector. Figer et al. [65] use cosmic ray events in up the ramp dark exposures to measure the compounded effect of charge diffusion and incomplete settling on pixel crosstalk. And as mentioned in Section 6.1.2.3, muon tracks can be used to directly measure charge diffusion since the entry and exit point of the particle are known. However, since cosmic rays come from a whole host of sources [108] and sample a large, nonuniform energy spectrum, it is not possible to determine the depth at which the particle interacted in the detector or cross check the number of electron hole pairs generated against the energy of the incident particle.

An x-ray source like Fe^{55} solves these problems since the energy spectrum, and thus, the absorption depth and number of electron-hole pairs generated in the silicon is well known (approximately 30 μm for the Mn x-ray photon).[2] The following sections describe a method for precisely measuring pixel crosstalk with an x-ray source such as Fe^{55}. Section 6.1.3.3 shows results that are typically associated with diffusive crosstalk and IPC and Section 6.1.3.4 addresses the role IPCT plays in Fe^{55} measurements.

6.1.3.1 Data Collection

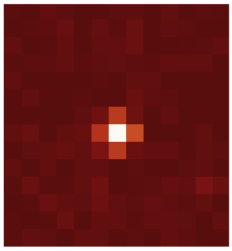
The experimental setup used to collect the Fe^{55} data is the same as the one described in Section 4.3. The data collected for H2RG-32-147, H1RG-022, and H4RG-10-007 in the RIDL dewar were not used to measure crosstalk because they had an RMS read noise of σ_{rn} = 20-50 e^- per pixel. The signal to noise ratio for these Fe^{55} events thus prevents us from distinguishing the K_α and K_β peaks. For H2RG-001, the read noise was greatly improved due to the changes described in Section 4.4. σ_{rn} ranged between 7-10 e^-, which is less than the Fano noise of σ_{Fano} =13 e^-, and this allows us to separate the peaks. With H2RG-001, several days worth of data were collected at temperatures ranging from 90-180 K in 10 K increments, and ample time was taken for the detector to stabilize after each temperature change. Lastly, backside voltages of V_{SUB} = 25 V and V_{SUB} = 40 V were used.

6.1.3.2 Data Analysis

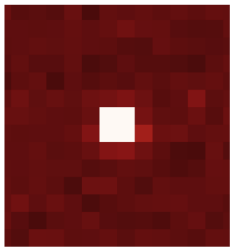
Analysis of the Fe^{55} data consists of a *guess and check* method to determine pixel crosstalk. We *guess* the amount of crosstalk between nearest neighbors and *check* to see if the energy spectrum is well represented. At the estimate that most closely matches the actual physical crosstalk, the Fe^{55} peaks should be best resolved. If the estimate is far off, the peaks will blend together as one.

The crux of this method is the assumption that a small fraction of the Fe^{55} events in the data are *single pixel events*. That is, the initial electron cloud is centered directly above, and close enough

[2] The absorption depth of a given particle in a given material is defined as the depth in a material at which 1-1/e of the incident particles in a large distribution have been absorbed.



-8	-11	-12	5	-5
-5	10	65	7	8
2	61	999	77	-12
-4	17	66	5	-2
-4	-3	-10	-7	-13



17	22	29	16	0
13	173	227	23	7
31	354	427	55	15
12	17	66	18	3
-1	2	8	7	0

Figure 6.8: Representative examples of single (left) and multiple (right) events generated from exposure to Fe^{55} source. The tables to the right of the images show the 5×5 set of raw ADU values surrounding the pixel with the maximum value (after bias subtraction).

to, the pixel collection node that all of the 1620 ± 12.7 holes generated by a K_α x-ray or 1770 ± 13.3 holes generated by a K_β x-ray are collected by that pixel. Figure 6.8 illustrates the difference between a single pixel event and a multiple pixel event. In the single pixel event, to first order crosstalk occurs only between the pixel that collected the charge and its eight nearest neighbors. In the multiple pixel event, this number grows because more than one pixel has collected charge.

Identifying single pixel events once the crosstalk percentages are known is a very straightforward process. If the crosstalk is not known, one must iterate through potential values using the guess and check method described above. The steps in going from a three dimensional datacube to obtaining the crosstalk values are described below.

1) Form 2-d hit images For a datacube with reads $r = 0, 1, ..., R$, where R is the last read, $R-1$ difference images are formed by subtracting consecutive frames. The pixel values in the r^{th} difference image yield the amount of charge deposited by the x-rays that hit between the reads $r-1$ and r. One might argue that averaging the pixel value over the reads before the hit and averaging the pixel values after the hit and using this difference will yield better signal to noise. However, as will be shown in Section 6.1.3.4, at temperatures above 120 K the signal of a hit pixel decays in the reads following the hit. Averaging the values after the hit for these temperatures therefore increases the uncertainty in the amount of charge deposited by the x-ray.

2) Identify contiguous sets of pixels Each 2-d image is first masked so that all bad pixels have their values set to 0. The images are then searched for contiguous sets of pixels that have a value greater than $3\sigma_{rn}$. For each contiguous set, the total number of pixels that meet this criteria along with the largest pixel value and the values of its eight nearest neighbors are recorded to a MYSQL database. To be clear, this means that even if only two pixels in a contiguous set are above $3\sigma_{rn}$, nine pixel values will be stored. This allows for measurement of crosstalk below the $3\sigma_{rn}$ level. Also recorded are the location of the peak pixel, its read number, and other tracking information such as bias voltage, temperature, and the name of the file in which it was found.

3) Iterate over potential crosstalk values Once all events are recorded, trial crosstalk values can be used to search the database for single pixel events. A lower threshold for the maximum pixel value, obtained by inspecting a handful of events, is also used in the search to ensure that only K_α and K_β events are returned. Queries are only made for entries that had nine or less contiguous pixels since more than nine pixels indicates a double hit. For each query, the returned pixel values are ordered by their position in the 3×3 box surrounding the central pixel, as this is how they were stored.

The crosstalk values, or coupling constants, $\alpha_{x,y}$, are defined relative the to intensity of the central pixel, $S_{0,0}$. The pixel immediately to the left of center will have a value $S_{-1,0} = \alpha_{-1,0} S_{0,0}$, the pixel in the upper right corner will have a value $S_{1,1} = \alpha_{1,1} S_{0,0}$, and so on. For notational convenience, the left and right coupling constants are expressed as $\alpha_{\pm 1,0}$, the top and bottom as $\alpha_{0,\pm 1}$, and the diagonals as $\alpha_{\pm 1,\pm 1}$.

The queries are made by demanding that, in addition to the constraints on $S_{0,0}$ and the number of pixels discussed above, the coupling constants fall within certain ranges α^{min} to α^{max}:

$$\begin{aligned} \alpha^{min}_{\pm 1,0} < \alpha_{\pm 1,0} < \alpha^{max}_{\pm 1,0} \\ \alpha^{min}_{0,\pm 1} < \alpha_{0,\pm 1} < \alpha^{max}_{0,\pm 1}, \end{aligned} \tag{6.9}$$

At each α range, the returned single pixel ADU values are binned and an attempt to detect the K_α and K_β is made. If the attempt fails, the range is ruled out of consideration. If the attempt succeeds, a minimum $S^{min}_{0,0}$ between the two peaks at $S^{K_{alpha}}_{0,0}$ and $S^{K_{beta}}_{0,0}$ is determined. The widths of the two peaks is then approximated by taking the second moment of the distributions from the range of ADU values extending a distance of $\Delta K_\alpha = S^{min}_{0,0} - S^{K_\alpha}_{0,0}$ to the left and right of the K_α peak, and a similarly defined range around the K_β peak. The parameter finally used to compare different α values is the ratio of the peak heights to the peak widths. Note that the bin size must be set appropriately small ($\sim \sigma_{rn}/2$) to ensure the growth of the peak does not overwhelm the growth of the width.

Starting at $\alpha^{min} = 0$ and $\alpha^{max} = 0.045$, which corresponds to a range of slightly more than $\pm 2\sigma_{rn}$, the values are incremented by 0.01 in parallel. Once a clear minimum in the peak height to width ratios is found, for example $\alpha_{min} = 0.05$ and $\alpha_{max} = 0.095$, finer adjustments in the α values are made and the range is adjusted. Also, $\alpha_{\pm 1,0}$ and $\alpha_{0,\pm 1}$ are varied independently if a clear asymmetry is present in the two directions. Some example values from the iteration process performed on the Fe^{55} data taken with H2RG-001 at 160 K and $V_{SUB} = 25$ V are shown in Figure 6.9. It is again emphasized that the three distributions in the figure are drawn from the same parent distribution of events. They only differ because of the constraints listed in Equation 6.9. The upper left plot is a distribution that constrains the crosstalk to be too small, the lower left one constrains it to be too large, and the right one constrains the crosstalk to most closely match the actual value.

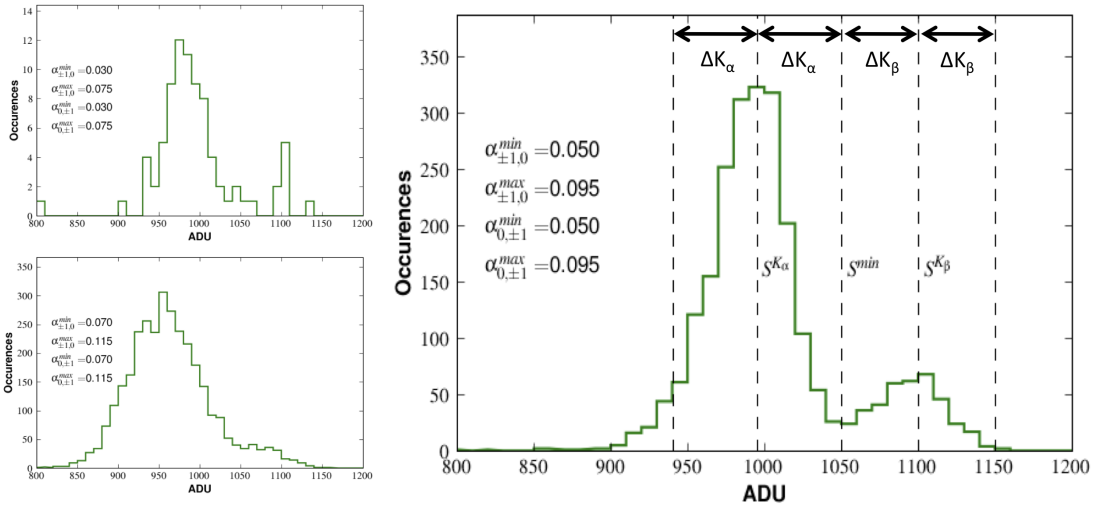


Figure 6.9: Example distributions obtained with the guess and check method for crosstalk values too low (upper left), too high (bottom left), and properly centered (right). In the right diagram, the quantities $S^{K_\alpha}_{0,0}$, $S^{min}_{0,0}$, and $S^{K_\beta}_{0,0}$ are shown with the $0,0$ omitted. Also shown are the widths used to calculate the seconds moments ΔK_α and ΔK_α as well as the crosstalk values, α, that were used to find single pixel events.

6.1.3.3 Results

The results for measurements made between 100 K and 170 K with the detector operating in Full Frame mode ($t_{frame} = 10.6$ s) are shown in Table 6.1 and Figure 6.10. The most noticeable feature in the plot is the large jump in crosstalk values at 130 K. While not fully understood, this is believed to be the same phenomenon mentioned in Section 4.3.7.1, which affects the conversion gain and voltage offsets of the pixels (presumably as a result of an offset on the output of the pixel source follower).

Aside from the large jump, the coupling values are relatively stable as the temperature increases. The slight increase in coupling with increase in temperature above 130 K may be a result of lateral charge diffusion, which goes like $\sqrt{T}$. Ideally, lateral charge diffusion should be negligible since the data points represent single pixel events and it is assumed that none of the holes generated in the initial charge cloud are collected by the surrounding pixels. In other words, all of the crosstalk is assumed to be due to IPC or any other form of electric crosstalk. However, even if the holes from the charge cloud do not make their way to the neighboring pixels, which would cause an increase in their signal and a resulting increase in α, some may make their way to the surface surrounding the collecting implant, causing a deficiency in the charge measured in the center pixel. The latter would also cause an increase in α.

The asymmetry between the vertical and horizontal neighbors is not fully understood, but has

been measured in previous studies of infrared and visible hybrid devices [111, 52]. No known asymmetry exists in the layout of pixels in the detector material, so this would lead one to the conclusion that it must be caused by the multiplexer. However, Moore found no substantial evidence of asymmetry—or coupling at all for that matter—in a bare multiplexer [24]. This suggests the higher voltages present at the gates of the pixel source followers brought about by the use of the backside contact voltage (and corresponding current draw) are partly responsible. The horizontal neighbors could possibly be coupled more strongly by either the read control line or reset control line shared by pixels in a common row. There is a discrepancy between the left and right α values that is constant with temperature, suggesting it is not a result of the Fe^{55} source being off-center. All of these discrepancies require further investigation.

In addition to the coupling values, the analysis provides a ratio for the number of single pixel K_α and K_β events to the total number. Table 6.1 shows that this number ranges from 0.016-0.027 over the temperature range tested. While this number may seem small, the following argument shows that it is reasonable. In order for the holes in the charge cloud to be completely absorbed by a single pixel, it must be generated sufficiently deep in the detector to prohibit loss of carriers by lateral diffusion. The absorption depth of the Mn x-rays is 30 μm, so the fraction f_D that make it to a depth D in the detector (measured from the illuminated backside) is

$$f_D = e^{-D/30\ \mu m}. \tag{6.10}$$

The charge cloud should also be generated directly overhead the p^+ collecting node or else diffusion to neighboring pixels or loss to traps in the gaps will take place. The fraction of these events will go like

$$f_A = A_{implant}/A_{pixel}, \tag{6.11}$$

where $A_{implant}$ is the area of the highly doped implant and A_{pixel} is the total area of pixel. The fraction of single pixel events measured at depth D should be the product of these two factors:

$$f_{sp} = f_A f_D = \frac{A_{implant} \cdot e^{-D/30\ \mu m}}{A_{pixel}} = \frac{l_{implant}^2 \cdot e^{-D/30\ \mu m}}{l_{pixel}^2} \tag{6.12}$$

The rightmost expression assumes both the pixel and implant have a square shape with length l. While the pixel is known to be square with $l_{pixel} = 18$ μm, the actual geometry of the implant and its size are proprietary and not known. But assuming that single pixel events are generated at a depth between $D = 90 - 99$ μm and using $\overline{f_{sp}} = 0.0227$, we obtain 12.1 μm $< l_{implant} <$ 14.0 μm. This is a reasonable value for the dimensions of the implant, which suggests that the fraction of single events measured is a realistic number.

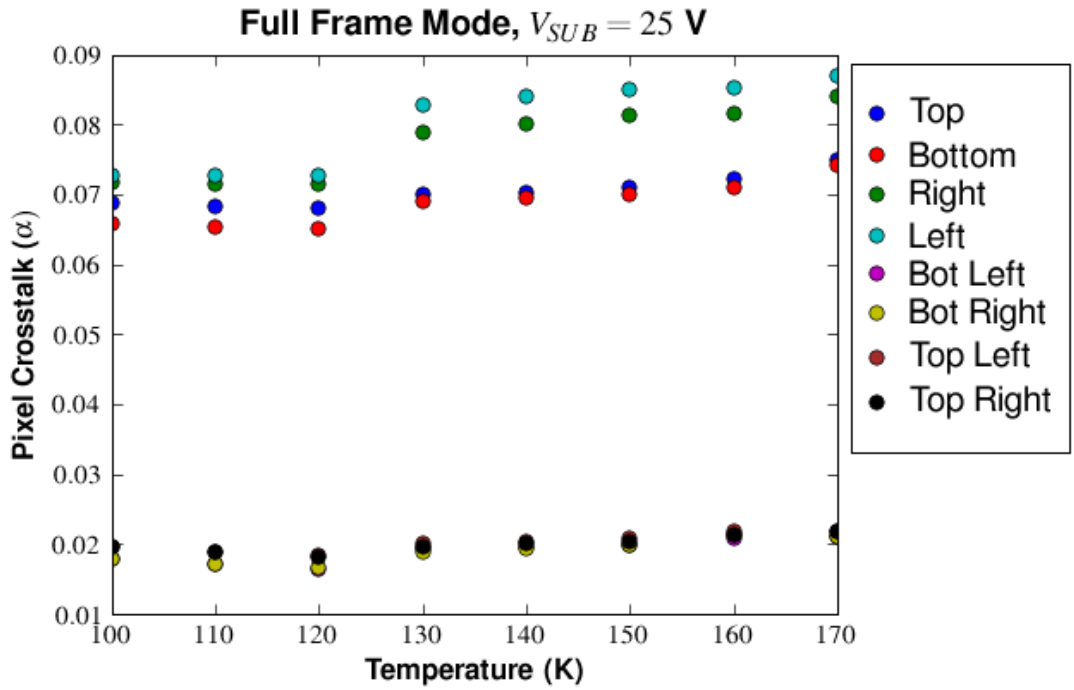


Figure 6.10: Crosstalk values vs. temperature for H2RG-001 measured from single pixel Fe^{55} events. The details of the bump and disparity between the nearest neighbors are explained in the text.

Table 6.1: Pixel crosstalk parameters for H2RG-001. The discrepancy in the total event numbers is a result of uneven sampling at different temperatures. The large conversion gain at T=170 K is due to IPCT (discussed in Section 6.1.2).

Temperature	Total Events	Single Pixel Events	Fraction	Conversion Gain
100	475,286	12,785	0.0269	1.141
110	546,693	13,298	0.0243	1.125
120	541,769	11,446	0.0211	1.080
130	510,669	12,147	0.0237	1.149
140	491,390	11,521	0.0234	1.157
150	448,881	10,186	0.0227	1.149
160	313,070	4,917	0.0157	1.165
170	111,782	2,357	0.0211	1.280

6.1.3.4 Charge Loss in Fe^{55} Measurements

One of the areas where Interpixel Charge Transfer (IPCT) shows most impact is in the measurement of conversion gain through the use of Fe^{55}. Unlike the trails of electron-hole pairs left by minimum ionizing particles and wandering recoil electrons generated by stray gamma ray events, the 5.9 keV x-ray photons emitted from the Fe^{55} source generate 1620 e^- contained within a diameter of only 0.4 μm [3], so it is not immediately obvious that the charge decay will exhibit the same characteristics. But as Figure 6.2 shows, for charge deposits in the range of 1000-2000 e^-, there will be a significant amount of holes–in most cases greater than the CDS read noise of the detector–lost by the pixels that integrated the deposit. And as with the cosmic ray events, the effect will be more pronounced at lower substrate voltages, slow frame rates, and higher temperatures, especially at temperatures above 140 K. It will now be shown that the charge decay can have a detrimental impact on x-ray energy measurements.

As an illustrative example of the decay after Fe^{55} hits, Figure 6.11 shows the ramps of two pixels from an exposure where H2RG-32-147 was exposed to the Fe^{55} source. The decay in signal after a single hit or multiple hits is evident in the ramps, and it effectively appears as a decrease in the signal for that pixel. The dark current rebounds after a few time constants $1/e_n$. However, as with cosmic rays, in most cases it does not return to the same rate as before the hit, and the direction of change seems to depend on the conditions in neighboring pixels.

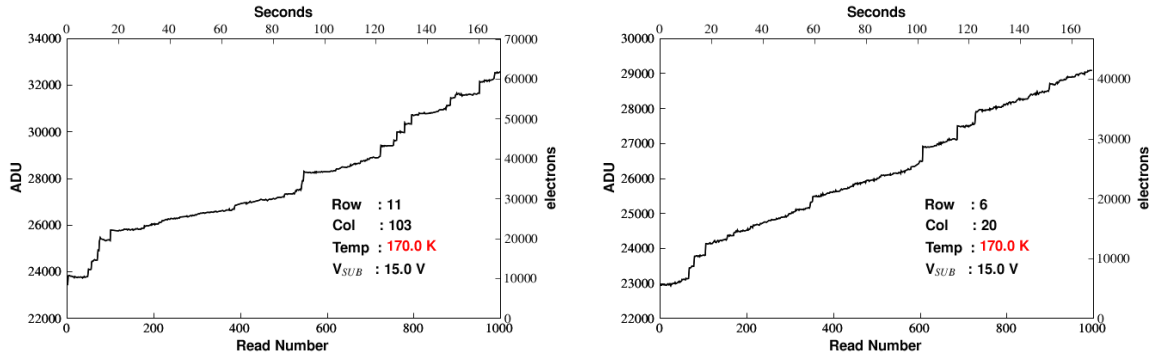


Figure 6.11: Two ramps taken with H2RG-32-147 in window mode while it was exposed to the Fe^{55} source. After a hit (or multiple hits) the pixel loses a small fraction of signal, resulting in an apparent decrease in signal. The charge in the pixel before the source follower voltage is read may be slightly greater than the amount of charge present when it is actually read, resulting in an incomplete sample of the electrons deposited by the x-rays.

As discussed in the previous section, because the sampling rate of a given pixel is not infinitesimally small, at certain temperatures the pixel will have lost a significant amount of charge between the time the x-ray hit, t_{hit}, and the time the pixel is read, t_{read}. This is a source of systematic error since there is uncertainty in the quantity $\Delta t = t_{read} - t_{hit}$. The events occur with a uniform

probability in time, $P(t)$, and Δt can vary between $\Delta t = 0$, where the pixel is read before any charge has been lost and, $\Delta t = t_{frame}$, where the pixel is read nearly an entire frame time after it has been hit. The latter case will lead to the largest measurement error.

The amount of "missing charge" and its dependence on temperature can be estimated by using the empirical results from Section 6.1.2. If the pixel will lose a total amount of holes Q_{lost} during the exponential decay (note that this cannot be stated as $t \rightarrow \infty$ since the dark current takes on a different value after the decay in some cases), then the amount of charge it will lose in Δt is

$$q_{lost} = Q_{lost}(1 - e^{-e_n(T)\Delta t}), \tag{6.13}$$

And as shown by Figure 6.2, the total amount of holes lost during the decay can be roughly approximated by a linearly relationship to the amount of holes deposited by the x-ray:[3]

$$Q_{lost} = Q_{dep} \cdot m(T) \tag{6.14}$$

So the relationship between the measured charge and the charge deposited is

$$Q_{meas} = Q_{dep}[1 - m(T)(1 - e^{-e_n(T)\Delta t})]. \tag{6.15}$$

This equation shows that measured signal charge decreases with Δt and increasing temperature since $m(T)$ is an increasing function of T. Figure serves to 6.12 illustrates this. There are two important consequences of the charge loss:

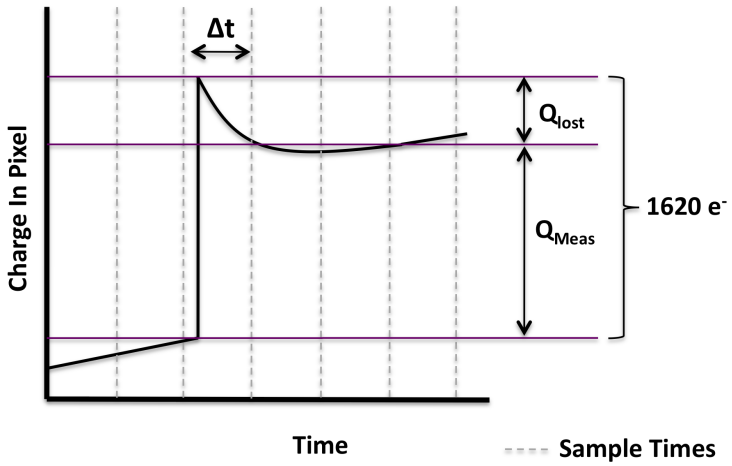


Figure 6.12: A pictorial plot that shows the behavior of pixel signal vs. time before and after an Fe^{55} x-ray has struck at temperatures higher than about 130 K. The bottom horizontal line shows the signal before the event and the top one shows the signal after the hit. The middle horizontal line shows the signal a time Δt after the hit, when it is sampled. The quantity of charge Q_{lost} is lost before the read, so that Q_{Meas} is measured.

[3]For other events generated by ionizing radiation, the relationship between the charge lost and the charge deposited for a given pixel is dependent on the position of the pixel in the event as a whole and the substrate voltage.

1) The K_α and K_β lines will be measured at a lower ADU value, resulting in an underestimate of the x-ray energies and an overestimate of the conversion gain (more e^- per ADU). Because the events occur with a uniform probability of $1/\Delta t$ in time, the mean value of the charge lost when measuring with a sample interval Δt will be:

$$\langle q_{lost} \rangle = Q_{lost} \left(1 - \int_0^{\Delta t} \frac{e^{-e_n t}}{\Delta t} dt \right) = Q_{lost} \left(\frac{e^{-e_n \Delta t} - 1}{e_n \Delta t} + 1 \right) \tag{6.16}$$

The peaks will shift by this amount relative to a measurement where $\Delta t \to 0$, which represents the case when the interval between successive samples of the pixels is very small.

2) The width of both peaks will increase due to the uncertainty in the interval of time between when the x-ray hit and when the measurement takes place. The variance in the measured amount of charge lost will be

$$\langle (q_{lost} - \langle q_{lost} \rangle)^2 \rangle = \int_0^{\Delta t} \frac{1}{\Delta t} \left(Q_{lost} e^{-e_n t} - \langle q_{lost} \rangle \right)^2 dt = Q_{lost} \left(-\frac{1 - e^{-2e_n \Delta t}}{2 e_n \Delta t} - \frac{(1 - e^{-e_n \Delta t})^2}{e_n^2 \Delta t^2} \right) \tag{6.17}$$

The widths of the K_α and K_β peaks, $\sigma_{K_{\alpha,\beta}}$, should be increased by this uncertainty relative to the read noise, σ_{rn}, and Fano noise widths, σ_{Fano}, in quadrature:

$$\sigma_{K_{\alpha,\beta}} = \sqrt{\sigma_{rn}^2 + \sigma_{Fano}^2 + \sigma_{IPCT}^2}, \tag{6.18}$$

where $\sigma_{IPCT} = \sqrt{\langle (q_{lost} - \langle q_{lost} \rangle)^2 \rangle}$, and the $IPCT$ subscript indicates the noise is due to charge loss/transfer. Although the peak widths will not be symmetric around the mean (due to the fact that the decay is more likely to be sampled near the bottom), an approximation of the full width at half maximum (FWHM) can be made with FWHM $= 2.354\sigma$. When the peaks begin to encroach one another for high T and long Δt, an obvious consequence will be that they are no longer resolvable as separate peaks.

Equations 6.16 and 6.17 are plotted in Figure 6.13, and agree very well with the data. Below 160 K, a degradation of the Fe^{55} spectrum is not seen for the lowest sampling rate of $t_{frame} = 10.6$ s used to collect the data. This is expected since the time constants at these temperatures for the charge loss are much greater then the frame time. The peaks shift only slightly and the widths are dominated by the read and Fano noise. However, the effect begins to degrade the spectrum at 160 K where the time constant approaches 50 seconds and is seen quite dramatically at 170 K, as illustrated by Figure 6.14. At this temperature, if a Mn K_α x-ray deposits all of its charge in one pixel and a time $\Delta t = 10.6$ s elapses before the pixel is read (this is the time for an H2RG frame when reading out of four outputs), it will have lost approximately 133 e^- (114 ADU). This is more than thirteen times as large as the read noise. On average the measured charge will be 92 e^- (80

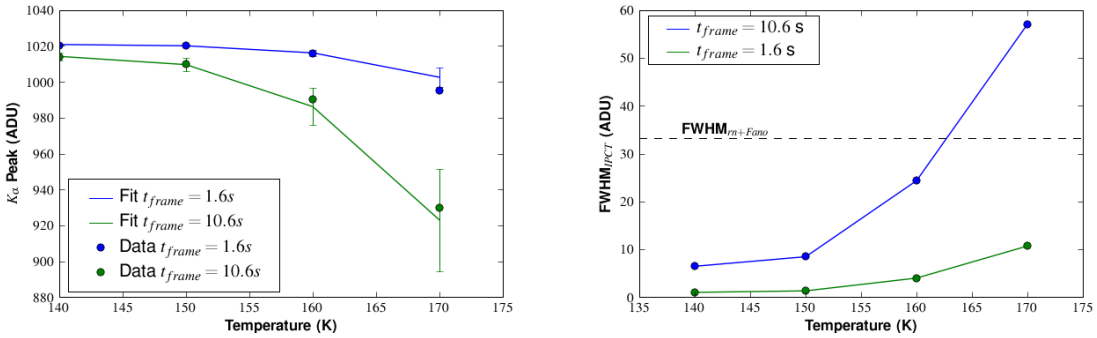


Figure 6.13: (Left) K_α ADU value for the center pixel in single pixel Fe^{55} hits from measurements taken in full frame mode with a frame time of 10.6 seconds and window mode with a frame time of 1.6 seconds. The data is shown by the circles and the predicted peak shift (assuming $Q_{dep}(t = 0) = 1022$ e^-) from the fit in Equation 6.16 is shown by the solid lines. The error bars are determined by the uncertainty in the decay time constant e_n. (Right) The contribution of IPCT to the FWHM of the K_α peak as determined by Equation 6.18. The FWHM due to the pixel read noise and Fano noise alone is shown by the black dashed line. The conversion gain for the data is roughly 1.16 e^-/ADU.

ADU) less than the charge deposited, and the peak will be shifted by this amount. The observed peak shift is 104 e^- (90 ADU), and the two values agree within the error due to the uncertainty in the time constant. Also, as shown by the dashed line in Figure 6.13, at 170 K and $t_{frame} = 10.6$ s the FWHM of each peak is dominated by the uncertainty in the charge lost. The calculation shows that the FWHM of the peaks will be about 65 ADU as opposed to 33 ADU in the case when the read and Fano noise dominate. Since they are only separated by about 100 ADU, the peaks should no longer be discernible, which is exactly what is observed in the measured spectrum.

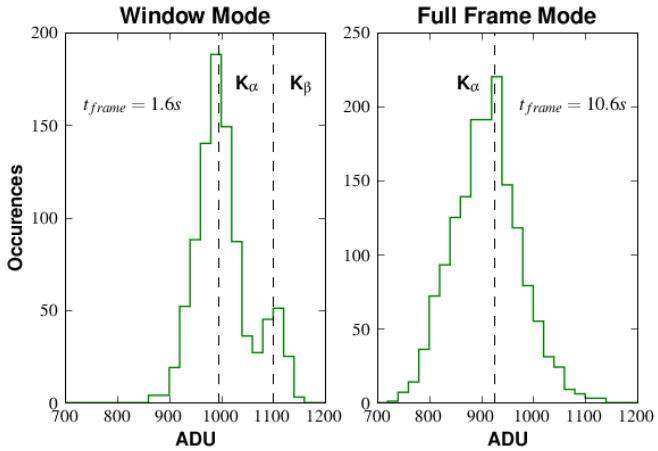


Figure 6.14: Two histograms showing the effect of IPCT on the Fe^{55} spectra taken with a fast frame rate (left) and a slow frame rate (right). With $t_{frame} = 1.6$ s, both peaks are detected. But with $t_{frame} = 10.6$ s, the K_β peak is not detected. All other operating conditions used in collecting the data for the two plots (e.g. V_{SUB}, clocking speed, bias voltages, etc.) are the same.

6.1.3.5 Application to Soft X-Ray Measurements

HyViSI detectors are implemented, and have been proposed for future use, in x-ray astronomy and spectroscopy [25]. The calculations and data in the preceding section show that if the HyViSI detector is to be used in measuring x-ray energies below 10 keV, sufficiently low temperatures (< 160 K) and short frame times should be used. Otherwise, the energies must be estimated using Equation 6.16 to account for charge loss. Even with this calibration, for high temperatures and long frame times detection of closely spaced spectral lines such as K_α and K_β will be impossible. The operating conditions for the detector must therefore be carefully chosen. Note that new surface treatments have dramatically reduced this effect, so a preliminary inspection of each HyViSI device should be taken to see whether or not it is present at the level just described.

6.1.4 Measurement via Single Pixel Reset

Using the guide mode feature of the HxRG multiplexers allows one to reset or read one pixel of the array while the other pixels of the array are integrating. In an ideal detector, resetting the single pixel would not affect the neighboring pixels. However, in the HyViSI we observe that the neighboring pixels are indeed affected, either through capacitive coupling, charge injection, or oxide charging. Lateral charge diffusion before collection is eliminated as a contributor to the crosstalk since charge carriers do not traverse from the backside of the detector. The act of resetting the single pixel in guide mode also creates a persistence signal after the readout is switched back to full frame mode, which will be important for the discussions in Chapter 7. The following sections describe the experimental setup used for the Single Pixel Reset (SPR) scheme and the results. Some of the results are straightforward and have been discussed previously in the literature. Others indicate unexpected effects that have not been discussed.

6.1.4.1 Experimental Sequence

To identify which mechanisms are dominant in the post-collection, electrical pixel crosstalk, we ran experiments that used the single pixel reset mode. Our measurements were made in the dark, in contrast to those made by Finger et al. [52], where the detector was uniformly illuminated. The exposure sequence used for the experiments was the following:

1) Reset the full array to a value V_{RstFF}.

2) Reset the single pixel at location i, j to a value V_{RstWin} a total of N_{Rsts}= 25, 50, 75, or 100 times.

3) Read the full frame in 32 output mode.

4) Repeat steps 2-3 a total of 10 times.

After these 4 steps, we performed the sequence again, but this time without applying the reset in step 2. Instead of applying the reset, the single pixel was simply clocked the same number of times–25, 50, 75, or 100–as it was in the previous sequence. Thus, the exposure time of the sequence with the reset voltage applied in step 2 is the same as the one without the reset voltage applied. This ensures that the integration time for the pixels is the same in each sequence and allows us to subtract corresponding reads in successive exposures.

We performed these sequences at 10 K increments from T =100-180 K. At each temperature, we began with V_{RstWin}=0.1 V and then increased this voltage in the sequence V_{RstWin}=0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6 V while holding V_{RstFF} at a constant voltage. We first performed this sequence at V_{RstFF}=0.30 V and V_{SUB}=5 V while moving up in temperature. Once we reached T =180 K, we increased V_{SUB} to 15 V and moved through the temperature range in a decreasing fashion. At T =100 K we increased V_{SUB} to 25 V and proceeded to increase the temperature. Lastly, once T =180 K was reached, we increased the full frame reset voltage to V_{RstFF}=0.70 V and repeated the experiment moving down in temperature.

To analyze the data, we subtracted the exposures in which the pixel was not held under reset from the ones in which it was. Since the concern is with the offset of the pixels from their value when they are all uniformly reset to the same voltage and not the integration of dark current, the bias value were used for comparison. In other words, if the signal measured after reset in the center pixel and its left neighbor are $S_{FF}(0,0)$ and $S_{FF}(-1,0)$ without the single pixel reset, respectively, and $S_{SPR}(0,0)$ and $S_{SPR}(0,-1)$ are the corresponding values with the single pixel reset, then the crosstalk value for the left neighbor will be

$$\alpha_{0,-1} = \frac{S_{SPR}(-1,0) - S_{FF}(-1,0)}{S_{SPR}(0,0) - S_{FF}(0,0)}, \tag{6.19}$$

Expressions for the other neighbors are made simply by swapping indices.

6.1.4.2 Expected Results

When the window reset voltage V_{RstWin} is amply greater than the full frame reset voltage V_{RstFF} (300 mV$< V_{RstWin} - V_{RstFF} <$ 800 mV), the measured crosstalk showed the trend expected from interpixel capacitance with a slight asymmetry between the fast and slow directions. As can be seen in Figure 6.15, the crosstalk values to diagonal neighbors are nearly identical to the ones obtained from single pixel Fe^{55} events. The crosstalk to the four closest neighbors, 6.9% for the fast read direction and 4.7% for the slow read direction, is significantly less, though. The measurements made by Finger et al. [52] for HyViSI devices were 8.5% in the slow and 9.3% in the fast: also significantly higher.

The source of discrepancy between these SPR results and those of Finger et al. can be attributed in part to lateral charge diffusion prior to collection at the integrating node since an illumination

source was used for their measurements and not in ours. In fact, with illumination the photogenerated holes should actually be drawn more strongly to the pixel that is reset since it presents the minimum in potential (all surrounding pixels are allowed to integrate, thereby increasing their potential and decreasing the vertical electric field). This will decrease the denominator in Equation 6.19 and make the α values larger. A logical guess might be that the discrepancy with the Fe^{55} values can also be attributed to lateral diffusion or carriers that are not swept by the metallurgical junction to the collecting node. However, the next section will show that—for the top and bottom neighbors at least—the results are affected to a great degree by an overall shift in signal on the entire column of the pixel undergoing SPR.

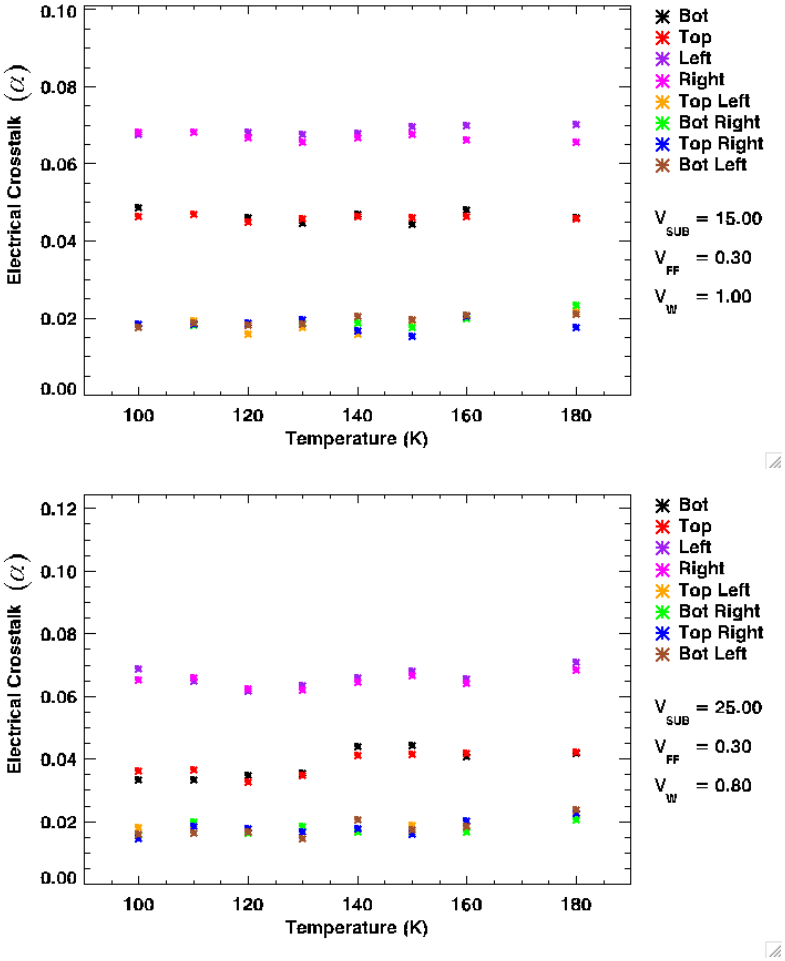


Figure 6.15: Pixel crosstalk observed in single pixel reset. The full frame is reset at $V_{FF} = V_{RstFF}$ and the center pixel is reset continuously at $V_W = V_{RstWin}$ while the other pixels integrate. The backside voltage is maintained at V_{SUB}.

6.1.4.3 Unanticipated Results

Column Depression The left image in Figure 6.16 shows an unexpected artifact in the SPR tests. The entire column that the single pixel belongs to suffers from a lowered voltage. The slump in voltage ranges from a few millivolts to tens of millivolts depending on the difference between V_{RstWin} and V_{RstFF} as well as the temperature. To the best of our knowledge, this phenomenon and the effect it has on IPC measurements has not been discussed in the literature. As the right plot in Figure 6.16 shows, though, its impact can be substantial.

Bezawada et al. found an increase in the DC offset along the rows of windowed pixels in an infrared H1RG [76]. It was noted that the offset is only caused by resetting the window; not reading it. However, our measurements with the H2RG HyViSI show that reading the window also causes offsets along the columns and rows. To test the possibility that the offset was a result of not properly clocking the window, extra HCLKs and VCLKS were added to the clocking sequence. The change in clocking did not remove or reduce the offsets.

The offsets are not entirely a surprise; it is a well known fact that changing clocking patterns in CMOS detectors can cause such behavior. But the role this plays in the asymmetry of the pixel crosstalk in SPR measurements has not been explicitly considered. We find that the α values for the top and bottom neighbors can swing from 0.03 all the way to 0.25 as a result of the column offset, depending on the biases applied and operating conditions.

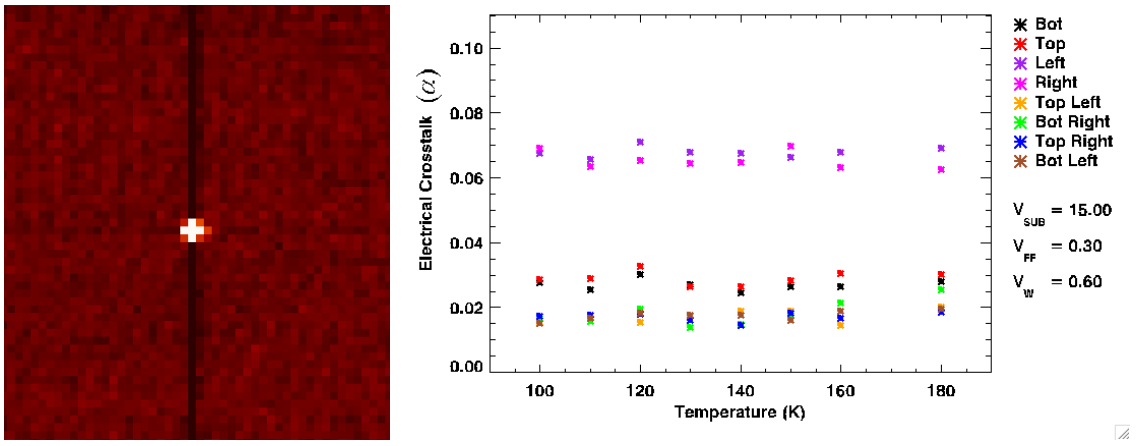


Figure 6.16: In SPR the coupling with the column neighbors is consistently lower than that for the row neighbors. The left image shows that the entire column of the pixel sees a slump in signal when $V_{RstWin} = 1.0$ V and $V_{RstFF} = 0.15$ V. A dark exposure in which the single pixel was not reset has been subtracted. For this case the column is about 11 mV below the rest of the array. The right plot shows that the top and bottom neighbors can have very low α values in SPR.

Front Surface Diffusion Another unexpected finding is that for reset voltages above 0.8 V and temperatures above 150 K, a significant outward diffusion of charge occurs from the pixel under reset to the surrounding ones. The diffusion, shown in Figure 6.17, generates a rise in signal for pixels well beyond the nearest neighbors at temperatures above 160 K. The signal in these nearest pixels see a quasi-exponential rise with time. Further out, the pixels integrate at the rate expected from dark current and then begin to rise at the quasi-exponential rate when the diffused charge has reached them. The radius to which the carriers diffuse in a time t and the slope in the signal rise increases as V_{SUB} decreases. For temperatures below 140 K, the signals do not change significantly from their bias value.

The strong temperature and backside bias dependence of this diffusion suggests it is caused by thermionic emission of holes over the interpixel potential barrier. For temperatures above 140 K, we assume the hole current has the form

$$J_p = A_o * (T - T_{th})^2 e^{\frac{V_{RESET} - V_{th} - f(V_{SUB})}{kT}}, \tag{6.20}$$

where A_o is a constant, $T_{th} \sim 140$ K is the threshold temperature below which the current is not detectable in the surrounding pixels and $V_{th} \sim 0.8$ V is a threshold voltage which represents the barrier height between two neighboring pixels. From the limited data collected in the SPR experiment, the function $f(V_{SUB})$ is not well determined. A closed analytical form may not exist since it depends on the potential profile between the back surface and the surface above the interpixel gap, which is inadequately treated without numerical integrations. However, it is clear that $f(V_{SUB})$ is an increasing function of V_{SUB} since it effectively strengthens the barrier between the pixels. Invoking the continuity equation, the neighboring pixels will see an increase in signal if the hole

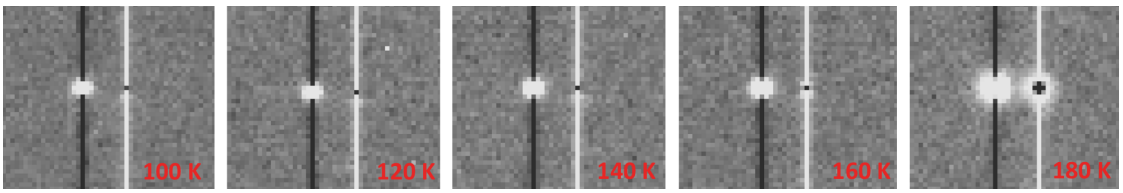


Figure 6.17: These images show charge diffusion and persistence in SPR at temperatures from 100-180 K. Each image is the product of subtracting the last read of an exposure where the pixel to the right was selected as a guide window and read 100 times between full frame reads (no resets applied) from the last read of an exposure where the pixel to the left was selected as a guide window and reset to $V_{RstWin} = 1.6$ V 100 times between full frame reads. The former was taken approximately 45 seconds after the latter, so the signal in the pixel to the right is a form of persistence. At 160 K and above, the core-halo structure (see Section 7.2) around the pixel is evident. The colors in the linear stretch—white corresponds to +12.8 mV and black to -9.8 mV—are inverted for the core-halo structure since it has been subtracted from another exposure.

current in to the pixel, J_p^{in}, is greater than the hole current out of the pixel, J_p^{out}. This explains how some of the pixels initially see a rise in signal and then a decline in signal some time later.

Figures 6.18 and 6.19 further support the theory that a diffusive hole current is leaving the center pixel. These figures show the relationship between the programmed reset voltage of the single pixel and the ADU value at which it is measured during a full frame read. If carriers were not leaving the pixel, the data points at a given value of V_{RstWin} should all lie on top of each other.[4] But because of the diffusion, there is some dispersion. When less resets are applied to the single pixel (25, for instance), the amount of excess carriers that escape the pixel during the period where it is not being reset is sufficient to bring the signal down within range of the ADC. When more resets are applied, there is an ample supply of excess carriers to keep the ADC railed despite the diffusion. The anomalous data point at $V_{RESET} = 1.6$ V and $T = 100$ K is a result of the dewar warming up during testing.

Although the effect this diffusion has in SPR tests is quite dramatic, the role it will play in astronomical applications may very well be insignificant. The pixel under reset is very near saturation before the leakage takes place. If a star is bright enough to saturate pixels at its center, doing photometry or astrometry on it will be fruitless regardless of whether or not there is a diffusion of holes to the surrounding pixels. And if it is a medium brightness star, the pixel ramps can be examined only for the time span before the saturation takes place to obtain the photocurrent. Also, the SPR experiments described consisted only of tests up to $V_{SUB} = 25$ V. It is possible that increasing V_{SUB} to a higher value could diminish the effect further. More experiments need to be done to confirm this.

Lastly, it should be noted that controlling the inter-pixel barrier height can be effectively accomplished through the use of gates placed in between the indium bump bonds. The voltage on the gate can be changed to increase or decrease the barrier, by accumulating or inverting the front surface of the detector, respectively. In the RVS SiPIN detectors, the gate is typically set at 5 V, which should result in a substantial decrease in the inter-pixel diffusion [38].

[4]Note that the disparity between the 100-130 K and 140-180 K ranges is due to the unexplained offset mentioned in Section 4.3.7.1.

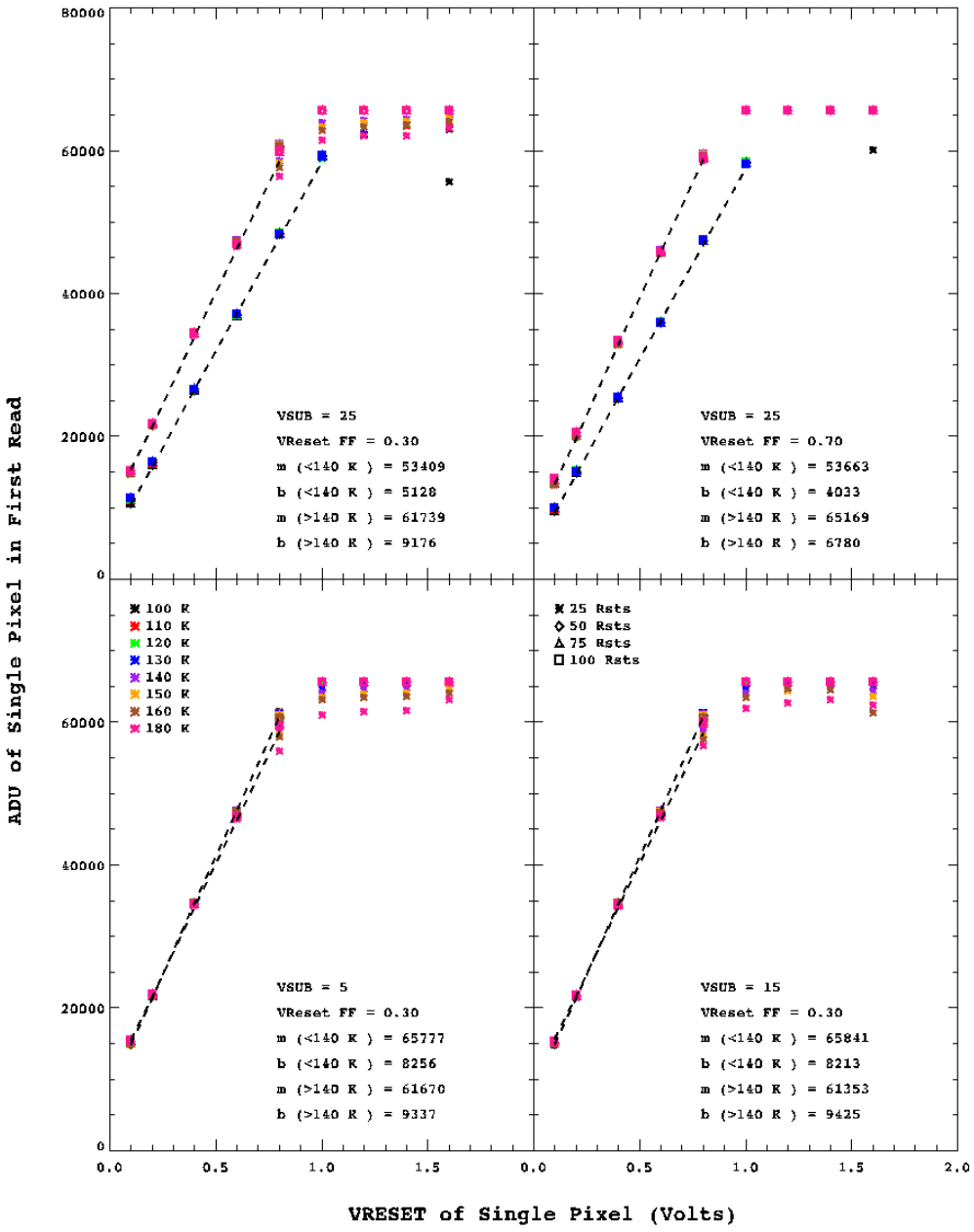


Figure 6.18: The output signal in ADU vs. the value of VRESET applied for a single pixel subject to single pixel reset in guide mode. Four cases with different values of VSUB and VRESET for the full array are shown. The legend shows the different temperatures as different colors and the different symbol shapes indicate the number of resets applied to the single pixel between full frame reads of the array. The dashed lines show the fitted slope, m, and intercept, b in terms of ADU/Volts. These numbers are also indicated in the diagram.

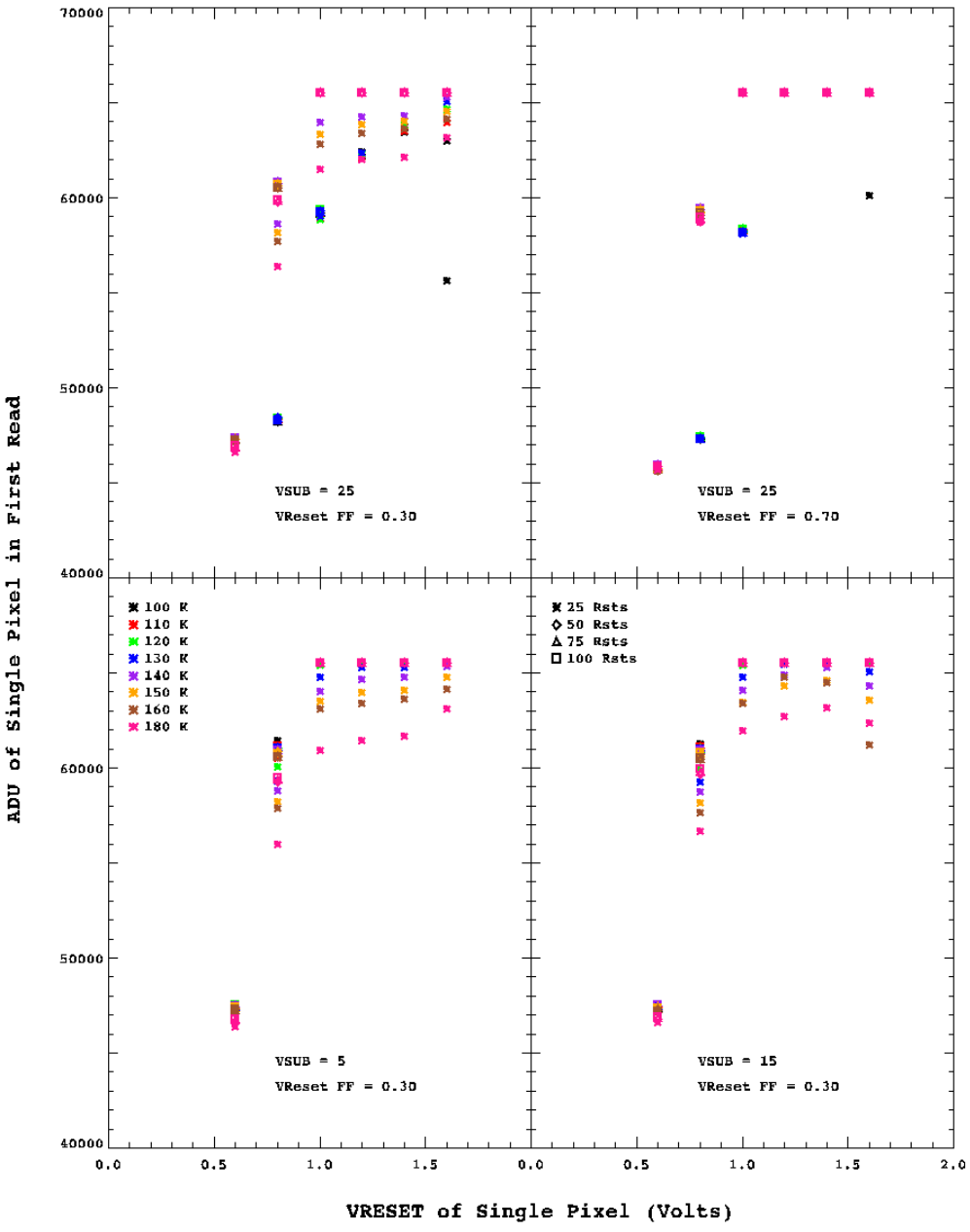


Figure 6.19: The same plots as in Figure 6.18 with only a small region shown for high values of VRESET. The legend shows the different temperatures as different colors and the different symbol shapes indicate the number of resets applied to the single pixel between full frame reads of the array.

6.2 Electronic Crosstalk

Electronic crosstalk can occur in several different ways in HyViSI detectors and hybrid detectors in general. Pixels may couple to supply buses. They may also couple to the other pixels in their respective column. And if multiple outputs are used, the signals from the pixels that are simultaneously selected can affect one another. Each of the last two cases will be examined in turn. For a good discussion of supply bus coupling, see Moore [24].

6.2.1 Column Bleeding

In many of the astronomical exposures recorded with the HyViSIs, a bleeding of signal is observed in the columns above pixels saturated by bright stars. Figure 6.20 shows the effect in a very pronounced form, as this is a mosaic where the same bright star saturated pixels in all of the individual exposures and the read noise has been beaten down after slope fitting the pixels and taking the mean. The

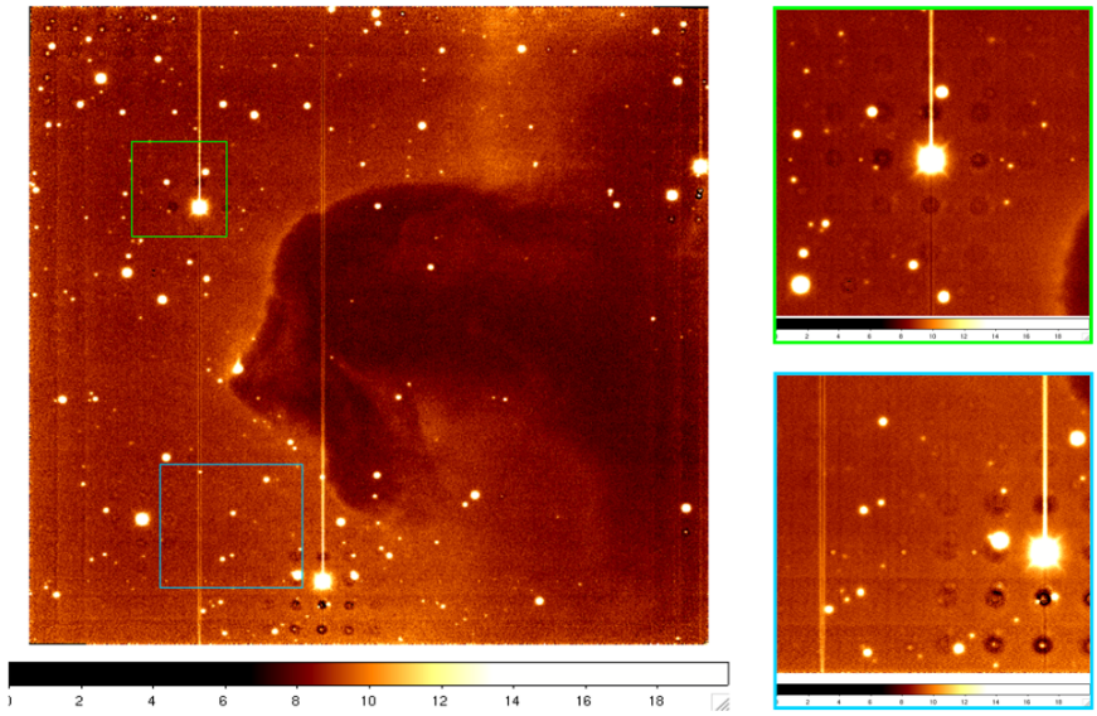


Figure 6.20: If a pixel has sufficiently high voltage, it will bleed into its column for a certain amount of time after it is selected. This is a mosaic of the Horsehead Nebula that shows the effect. The dark region in the center columns above the bright stars occurs because the image that went into the mosaic were *slopefits* with pixel values that represent the change in voltage with respect to time. The raw images show elevated voltages for all the pixels in the region. The black circles are due to persistence effects that will be described in the next chapter.

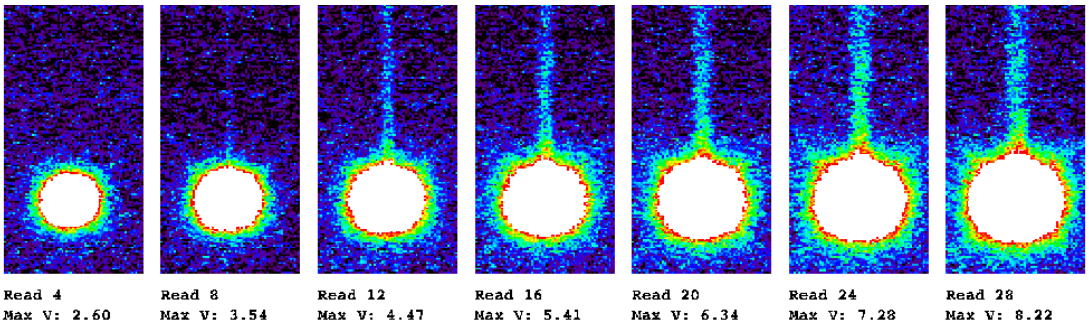


Figure 6.21: The column crosstalk does not depend on the instantaneous photocurrent, but rather occurs after a certain threshold voltage. Each image is a snippet from a 30 read exposure of a star that produces 15,000 e^-/s in the center pixel. The sky value has been subtracted and the min/max of the scale are -10/100 ADU. The raw voltage in the pixel, listed at the bottom of the snippet, is just an extrapolation from the SIDECAR conversion of ADU to μV from Equation 3.11. Although inaccurate for values above 3.5 volts or so, this shows the onset of the column bleeding begins when the pixel nears 3.3 Volts.

same bleeding effect is also observed for the dark current *volcanoes* described in Section 4.2.4.

As evidenced by the readout sequence in Figure 6.20, the effect is not initiated until the pixel has reached a voltage of about 3.3 volts.[5] This voltage is not arbitrary. It corresponds to the digital (V_{DD}) and analog supply (V_{DDA}) voltages as well as the source node voltage for the internal current source of the pixel source followers, $V_{BIASPOWER}$. If the voltage at the p^+ integrating node rises above this, the source of the unit cell FET can no longer follow the gate.

After the threshold has been reached by a particular pixel at i, j the pixels in the columns immediately above it ($j+1$, $j+2$, ..., $j+\Delta j$) experience an offset in signal of about +3-6 millivolts. The strength of the offset decays with an exponential signature as a function of Δj, i.e.

$$S(i, j+\Delta j)_{after\ offset} \sim S(i, j+\Delta j)_{before\ offset} + 4.5\ \mathrm{mV} * exp(-\Delta j/J), \tag{6.21}$$

where J is 50 pixels or so. If the saturated pixel is near the top of the detector, the decay will wrap around to the bottom of the detector. This increase is a constant offset to the signal that does not get larger with time; the affected pixels continue to integrate at their previous rate after it has occurred.

[5] Although there is no lack of data that shows the effect, the detector output becomes nonlinear at about V = 3.0 V. The exact onset voltage of the nonlinearity depends on the temperature and $V_{BIASGATE}$. Thus, 3.3 V is an estimate.

6.2.1.1 Explanations and Mitigation

The behavior described rules out the possibility that the column bleeding occurs in the silicon detector layer. It is not unreasonable to guess that saturated charge carriers might bloom preferentially along the columns in the detector, but there is no way they would wrap around from the very top row to the very bottom row. And if it was somehow an artifact of the control electronics, we should expect the response to be identical along the rows. It must therefore be generated in the multiplexer.

Pixels in the same column share a common vertical read bus in the multiplexer, A.K.A. the column bus. When a given row is selected, all of the pixels in that row are connected to their respective column buses through the CMOS *row select switch*. When a given pixel is selected, its column bus is connected to the horizontal read bus through another CMOS switch, the *column select switch*, and the column bus then carries an electrical current to the output. If there is sufficient capacitance between the input and output of the row select switch, then it is possible that a transient voltage will be present at the output after it has been closed and the next row has been selected. Once the column bus is again selected for the subsequent row, the transient voltage may couple to it and cause an offset in the signal. Unfortunately, the details of the multiplexer are not available for study to pinpoint the exact location of this stray capacitance.

While this effect is observed in all of the H1RG and H2RG devices tested, it is not observed in the H4RG. The reason for this is that the H4RG pixel has an anti-blooming diode (in parallel with the photodiode) that clamps the gate voltage at V_{DDA}, preventing the threshold voltage from being reached. Also, the effect appears to subside at lower values of $V_{BIASGATE}$. It is present in all of the data recorded with the SIDECAR, which operated with $V_{BIASGATE} = 2.29$ V. However, it is not present in any of the data recorded with the ARC electronics, which had $V_{BIASGATE} = 1.90$ V. In the latter case, though, strong bleeding along the rows occurred instead. A dependence on $V_{BIASGATE}$ should be expected since this is the voltage that governs the amount of current flowing through the buses and to the pixel drain. However, a quantitative relationship has not been established. A future experiment should be undertaken in which the bleeding of saturated pixels or the dark current volcanoes (which can be done in the dark) is observed as a function of $V_{BIASGATE}$, V_{DDA}, and $V_{BIASPOWER}$ to better understand the effect and how to prevent it from occurring.

6.2.2 Output Coupling

Another type of electrical coupling that occurs in the HxRG multiplexers is inter-channel, or output, coupling. When N outputs of the detector are being used, N pixels in a given row are being selected concurrently. If one of these N pixels is transmitting a large signal to its output, the other $N-1$ pixels may see an elevated signal, or *ghost*. An example of this is shown in Figure 6.22. The location of the ghosts will depend on the location of the saturated or near saturated pixel with respect to the outer edges of the channel as well as the read direction in each channel.

For the data taken with H2RG-32-147 and H1RG-022 at the Kitt Peak telescope the ghost signal appeared solely as a positive offset. However, for data recorded with H2RG-001, negative offsets were also observed, indicating two distinct types of behavior. Finger et al. observed the latter type in a HgCdTe H2RG [100]. As was done in their studies, the control electronics have been tested separately and eliminated as a possible source of the channel coupling. The mechanism that generates the ghosts in the HgCdTe devices is believed to be the same as that for the HyViSI. The primary difference in the purely positive coupling and positive/negative coupling is believed to due to the use/absence of the output source follower in the signal path.

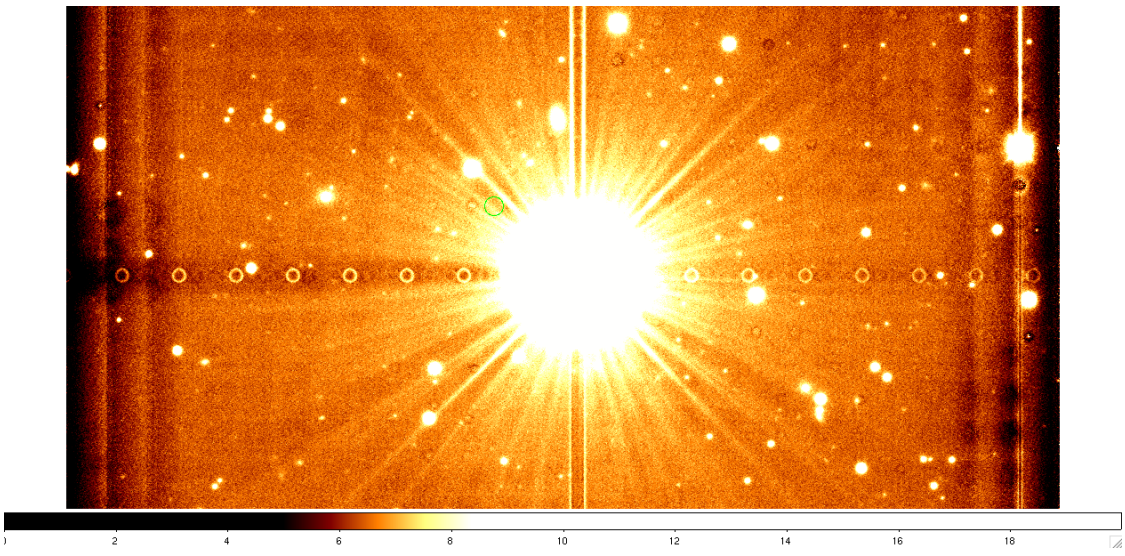


Figure 6.22: Saturated stars show up as ghost images in the other channels of the multiplexer. This is a mosaic of *slopefit* images of the bright star SAO 117637. The donut shape with the bright annulus and dark core is due to the slope fit. In a raw image, all pixels in the ghost show up with a positive offset. Also, in a raw image all 32 channels show a ghost image. The mean filtering of the mosaic removed some of the ghosts, but not all of them.

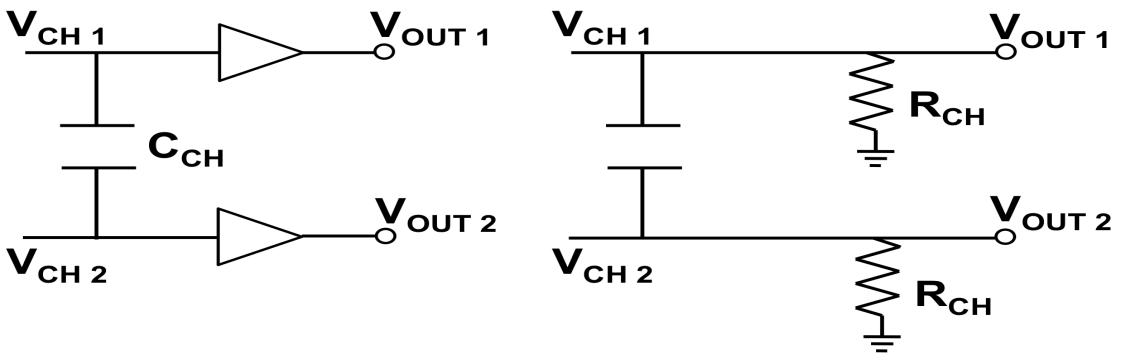


Figure 6.23: Circuit diagrams to explain output coupling through inter-channel capacitance. (Left) For the case where the output source follower is used, only positive signal coupling is observed, suggesting displacement current through a coupling capacitance. (Right) For the case where the output source follower is not used, the ghost signal has sign corresponding to the derivative of the the signal in the offending pixels, suggesting an RC differentiator.

6.2.2.1 Coupling with Output Source Follower

With the output buffer placed in the signal path for each channel, the signal from the pixel sees a very high impedance. If a coupling capacitance exists between the buses, a displacement current, i_{dis}, between them may be responsible for the signal in one channel showing up in the others. No quantitative or qualitative information on the architecture of the channel routing is available (e.g. the pitch of the wires or their orientation). But a simple two channel model like the one on the left in Figure 6.23 may be used to express the voltage change on an affected output as $dV_{OUT\ 2}/dt = i_{dis}/C_{CH}$, where $CH\ 1$ is the line that carries the high voltage level. Instead of attempting to calculate i_{dis} and C_{CH}, a coupling constant, α, is used to relate the voltage changes:

$$\Delta V_{OUT\ 2} = \alpha \Delta V_{OUT\ 1}. \tag{6.22}$$

The average measured value is $\alpha = 3.5 \times 10^{-3}$, an extremely small coupling. Thus, the output coupling with the source follower enabled only presents itself for signals near full well or after multiple reads have been averaged to reduce the read noise.

The coupling grows even weaker as the channel separation grows. Note the rings to the very right in Figure 6.22 are substantially fainter than the ones closest to the star. It should be noted that the current source supplied to the output buffer by the SIDECAR and ARC electronics may play a part in the coupling. While coupling of the amplifier channels on these boards were tested by providing an input voltage from a power supply, the extremely low output impedance of the supply placed little or no demand on the current sources in these tests. A higher impedance source should be used to check for channel coupling in the electronics.

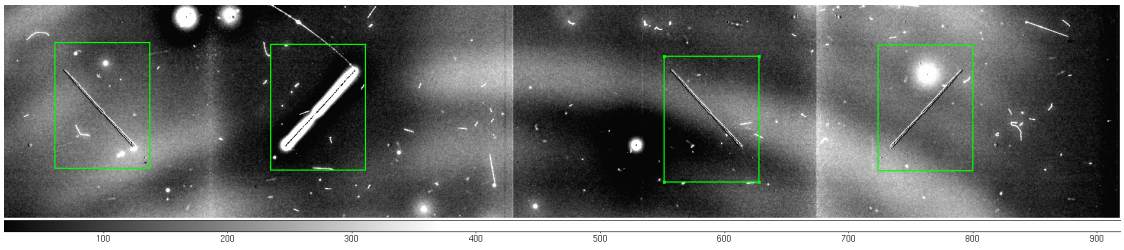


Figure 6.24: A last read minus first read image taken from a 100 read exposure with H2RG-001 operating in 4 output mode. The read direction in the channels is represented as ←→←→. A scratch in the detector creates high leakage current for the large region shown in the second green box to the right. In the other green boxes the ghost images show positive coupling while the voltage on the second output increases during the clocking sequence and negative coupling when the voltage is decreasing.

6.2.2.2 Coupling without Output Source Follower

The output source follower was disabled for noise tests performed on H2RG-001. These data show a negative coupling as well as a positive one. Figure 6.24 shows an example where a high leakage current area in H2RG-001 (read out through output 2) couples to the other 3 outputs being used. The read directions for the channels are ← → ← →. As can be seen in the figure, if the signal in channel 2 is increasing during the clocking sequence, the signals in the other outputs couples positively. And if it is decreasing during the clocking sequence, it couples negatively

As Finger et al. note, this is the behavior exhibited in a simple RC differentiator circuit [100]. Such a circuit is shown in the right of Figure 6.23, with C_{CH} representing the coupling capacitance and R_{CH} representing some resistance to ground. R_{CH} might be present in the multiplexer itself or at the preamplifier stage. Every time a new pixel is selected, the signals at the output swing by ΔV_{OUT} from the previous pixel voltage to a new pixel voltage. If CH 1 is carrying the high voltage, then the expected change in voltage on CH 2 will be:

$$\Delta V_{OUT\ 2} = R_{CH} C_{CH} \frac{\Delta V_{OUT\ 1}}{\Delta t} \tag{6.23}$$

where Δt is some fraction of the sampling time for a given pixel. Finger et al. show that the coupling indeed worsens when the pixel time is shortened [100] and measure a coupling of about 0.001 at a pixel time of 10 μs in a HgCdTe H2RG. In HyViSI H2RG-001, we measure a coupling in the range 0.0005-0.0018 (0.05%-0.18%) with a pixel time of 10 μs and $V_{BIASGATE}$ =2.05 V.

The coupling without the output source follower diminishes when $V_{BIASGATE}$ is decreased. This makes sense as the increased drain to source current of the pixel source follower decreases the settling time for each output. Of course, for applications that do not require fast full frame readout, one may operate the sensor with only one output to avoid this effect altogether.

Chapter 7

Persistence in HyViSI Detectors

This chapter addresses one of the most prominent effects observed in HyViSI sensors: image persistence. Persistence in the HyViSI presents itself in a rich and complex way, as the following sections will reveal. But before delving into the observations, it is necessary to first present the theory and mathematics behind charge traps, which are commonly used to explain "persistent" effects in semiconductors. After this framework has been established, the data will be presented and an attempt to explain it with a somewhat complicated model will be made.

7.1 Trap Theory

Charge traps are one of the many nuisances in semiconductor optical detectors (and all semiconductors for that matter) and have been extensively studied for several decades. In CCDs charge trapping can degrade the charge transfer efficiency of the detector and render entire rows useless. In Hybrid CMOS detectors, we find that the traps give rise to latent images, signal decays after cosmic ray events, and high dark current after forward biasing the photodiodes. Further, the emission and capture of charge carriers by the traps may be one of the most important mechanisms for generating cross-talk between pixels.

A *trap* is actually nothing more than a *deep impurity* in the semiconductor crystal. These impurities have states which lie in the band gap far away from the conduction and valence band edges, typically requiring an energy of $\sim 5kT$ to ionize. They are thus very efficient centers for recombination in the crystal. The term *deep* distinguishes them from *shallow* impurities, which are the kind of impurities that are purposely implanted in the semiconductor in the doping process. *Shallow* impurities typically lie at about $\sim kT$ away from the band edges, so a large fraction of them are ionized purely by thermal energy.

If we were to look at a silicon crystal doped with impurities at the subatomic level, we would find that traps are constantly being filled and vacated. The idea is somewhat abstract, especially

since holes are fictitious particles. But we imagine the electrons and holes are trapped in energy space somewhere in the bandgap (below E_c and above E_v) and that these traps also correspond to physical locations in the lattice. Electrons are captured at a rate c_n and emitted at e_n while holes are being captured at c_p and emitted at e_p. At equilibrium, all these processes are balanced so that the number densities of holes in the valence band, p_v, and electrons in the conduction band, n_c, remain essentially constant when measured by a macroscopic means, i.e. $dp_v/dt = dn_c/dt = 0$.

The processes of capture and emission for holes and electrons can be summarized as follows (with the corresponding coefficient listed):

- **Electron Capture -** c_n **:** An electron drops from the conduction band into a trap.
- **Electron Emission -** e_n **:** An electron jumps to the conduction band from a trap.
- **Hole Capture -** c_p **:** An electron in a trap drops into a hole in the valence band.
- **Hole Emission -** e_p **:** An electron in the valence band jumps into a trap and leaves a hole in the valence band.

The processes of *generation* and *recombination* of carriers can occur with combinations of these. Janesick provides two useful statements [3] to help understand how these processes occur **1) Electron and hole capture are required for recombination** and **2) Hole and electron emissions must occur nearly simultaneously through the trap in order to generate a dark carrier**.

If the temperature of the semiconductor crystal is changed rapidly or light is shined on it, the system will be brought away from equilibrium. The non-equilibrium situation was examined in detail by Shockley and Read [112], who formulated the mathematics that describe the role of traps in recombination and generation in a semiconductor. Among the numerous applications of their theory, McNutt and Meyer [113] used the statistics of traps and their temperature dependence to identify impurities in buried channel CCDs and Solomon [42] used Shockley-Read traps to model persistence in InSb hybrid infrared detectors. The reader is referred to these references for a detailed discussion of trapping statistics.

For this discussion, the notation in [113] will be adopted. N_{TT} will denote the total bulk state concentration of traps, n_T will denote the concentration of those traps filled with electrons, and p_T will denote the concentration of those traps filled with holes. It follows that

$$N_{TT} = n_T + p_T. \tag{7.1}$$

The dynamics of the free carriers in the system is governed by

$$\frac{dn}{dt} = \frac{dn_c}{dt} = e_n n_T - c_n n p_T \tag{7.2}$$

$$\frac{dp}{dt} = \frac{dp_v}{dt} = e_p p_T - c_p p n_T, \tag{7.3}$$

where $n = n_c$ is the concentration of electrons in the conduction band and $p = p_v$ is the concentration of holes in the valence band. The subscripts in n_c and p_v will be left off for notational convenience. In a conserved system where no free carriers are introduced from outside, trapped electrons can either be stolen from free electrons in the conduction band or gained through the donation of a free hole to the valence band. A similar argument for trapped holes leads to the equations:

$$\frac{dn_T}{dt} = c_n n p_T - e_n n_T - c_p p n_T + e_p p_T \tag{7.4}$$

$$\frac{dp_T}{dt} = -c_n n p_T + e_n n_T + c_p p n_T - e_p p_T \tag{7.5}$$

It is worth nothing that the number of conserved systems in the detector is very limited. In fact, the only system that will be treated in this way is the *entire detector layer* while it is not exposed to light and the pixels are integrating dark current. The number of free holes introduced through leakage currents and electrons lost to the external load supplying V_{SUB} will be negligible. Carriers emitted by traps become free carriers and may relocate inside the detector, but they are assumed not to escape.

In a non-conserved system, the equations for dp/dt and dn/dt have additional terms. Take the p^+ implant in the detector pixel, for instance. The number of free holes in the implant increases with time ($dp/dt > 0$) as they are introduced by photogenerated carriers from the depletion region or surface/bulk thermal leakage currents. At reset, electrons are injected into it through the reset transistor channel in the multiplexer to recombine with the free holes ($dp/dt < 0$), which brings its potential to V_{RESET}. If traps outside the implant emit holes, these holes may be added to the implant as free carriers as well ($dp/dt > 0$), and the migration of these holes will cause a rise in signal. In this last case, even though the sum of free carriers and trapped charge in the detector may be conserved, their redistribution in space can result in an increased electric potential at the collecting node of the pixel. In fact, Solomon attributed latent images in InSb arrays to holes being emitted from traps outside of the p^+ implant and collected inside of it [42]. The detector was considered a conserved system in which the holes emitted by traps translated directly into free carriers that were swept into the p^+ implant and counted as signal. The implications of this model will now be examined and later it will be shown that it cannot account for persistence in HyViSI devices.

7.1.1 Hole Capture and Release from Shallow Traps

Solomon attributed persistence in InSb hybrid CMOS detectors entirely to holes being trapped in the surface interface between the bulk InSb and SiO_2 passivant during illumination (see Figure 1.7). When the holes are emitted from the traps in subsequent exposures, they either drift or diffuse to the collecting node and cause a rise in signal. Thus, persistence is described entirely through the c_p term in Equation 7.5, which accounts for holes being captured during illumination, and the e_p term

in Equation 7.3, which accounts for holes being emitted in subsequent exposures and collected as signal charge.

According to this model, if the detector pixels are illuminated with sufficient flux or fluence and holes are introduced at the rate p_γ, then at steady state

$$\frac{e_p}{c_p} = p_\gamma e^{F-E_T/kT}, \tag{7.6}$$

where F is the Fermi Level (or quasi-Fermi level in the non-equilibrium case) and E_T is the trap energy. If E_T lies just above the valence band then the capturing processes will dominate emission and there will be a surplus of trapped holes relative to when the detector was not illuminated. The capture cannot continue indefinitely, of course. It goes like $p_\gamma c_p n_T$ and cuts off when the number of electron-filled traps, n_T, goes to zero. When the light source is removed or the array has been shuttered, the rates will change according to

$$\frac{e_p}{c_p} = N_V e^{E_V-E_T/kT}, \tag{7.7}$$

where N_V is the effective density of states in the valence band and E_V is the valence band energy. The trapped holes will thus be re-emitted in a subsequent exposure with a time constant of e_p^{-1} and generate a signal, even in the absence of illumination. And since the traps are *shallow* (they are located much closer to E_V then F), the emission occurs at a very quick rate en route to equilibrium. These emitted holes are assumed to be the source of latent images.

The signal in a subsequent exposure will depend on the excess of holes trapped during illumination, Δp_T, the time since the source of illumination was removed, t_o, and the integration time of the exposure, i.e. the time since reset, Δt. From the above equations, we can derive the following formula for the signal imparted by the release of holes, $S(t_o, \Delta t)$:

$$S(t_o, \Delta t) = \Delta p_T e^{-t_o/\tau}(1 - e^{-\Delta t/\tau}), \tag{7.8}$$

where $\tau = e_p^{-1}$ is the time constant associated with the release of the holes from the traps. This expression represents the case in which there is only one trap species with one time constant. However, we might also imagine that a whole slough of impurities are present, and that each has its own associated time constant τ_i and number of traps N_{T_i}. Equation 7.8 then becomes a linear combination over those species:

$$S(t_o, \Delta t) = \sum_i^N \Delta p_i e^{-t_o/\tau_i}(1 - e^{-\Delta t/\tau_i}), \tag{7.9}$$

If we differentiate this equation with respect to the integration time, we find that the rate of change

of the signal (the instantaneous dark current, D) goes as

$$D(T) = \frac{dS(t_o, \Delta t)}{d\Delta t} = \sum_i^N \frac{\Delta p_i}{\tau_i} e^{-(t_o+\Delta t)/\tau_i}. \tag{7.10}$$

We see that the total time, $T = t_o + \Delta t$, is what actually determines the rate of signal change in any subsequent exposure. And further, this rate should decrease monotonically with time after the source has been removed. One way of interpreting this is that the traps are being depopulated regardless of whether the detector is set to integrate or is being held in reset. We can measure τ by performing integrations and fitting $S(T)$ with the function in Equation 7.8, or some linear combination of these functions which represent different trap species, each having their own density and time constants. The expressions tell us that if we wait a time $t > 3\tau_l$, where τ_l is the longest time constant involved, dS/dt should be essentially zero. Another way of saying this is that if we wait long enough — whether we integrate, reset, or leave the detector idle while waiting — the trapped holes should all be released and not interfere with the signal in the next exposure. Once $dS(t_o, \Delta t)/d\Delta t \to 0$, it should remain at zero until we populate more traps with illumination. Since this **does not happen in the HyViSI detectors**, another model for the persistence needs to be introduced.

7.1.2 Hole and Electron Capture from Deep Level Traps

If the traps are deep traps[1] with energies E_T that lie close to the Fermi level F, the situation changes from the one just described. Thermal emission rates are proportional to a Boltzmann factor, $\exp(-\Delta E/kT)$, where ΔE is the depth of the trap (free energy) from the band edge to which the carrier is emitted [114]. The electron and hole emissions will be governed by

$$e_n = A_n e^{-(E_C - E_T)/kT} \tag{7.11}$$

$$e_p = A_p e^{-(E_T - E_V)/kT}, \tag{7.12}$$

where A_n and A_p are coefficients related to the density of states at the conduction and valence band edges, respectively, as well as the capture cross sections (both coefficients go like T^2). Thus, if E_T lies far from both E_C and E_V the emission rates should be small and capturing events should dominate.

In the case of the SiPIN detector, if the bulk material and surface interface between the bulk and SiO_2 passivant is fully depleted and in thermal equilibrium (in the dark), then $n \approx p \approx 0$. If deep level traps exist in these regions, they cannot capture carriers in an equilibrium state simply because

[1] As Miller et al. point out [114], the names applied to deep defect states–traps, recombination centers, generation centers, deep levels, deep impurities, and so on–can be quite confusing. For the argument presented here, it will imply that the trap energy lies only a few kT away from the Fermi level.

carriers are not available. But with illumination, reset, or forward bias, both of these regions see an increase in free carriers that are available for trapping. As in Solomon's model, under illumination the number of hole filled traps increases at the rate $p_\gamma c_p n_T$, and when the array is finally shuttered there will be an excess of hole filled traps Δp_T. However, because the traps are deep in the forbidden energy gap, the holes will not be efficiently emitted from the traps. Instead, they will be "stuck" at whatever atom plays host to the trap until an electron is available to return the trap to its initial charge state.

The electrons necessary to repopulate the hole filled traps can be provided by closing the reset switch. The function of the reset is to provide the carriers necessary to restore the potential at the integrating node to V_{RESET}. Ideally, the reset would provide an ample amount of electrons in a given pixel Δn_T to repopulate the traps in that pixel at the rate $nc_n p_T$. But if some of these electrons migrate to other regions of the detector instead of being captured by the traps, the full restoration will not happen. Electrons that quickly escape and drift all the way to the back surface of the detector will cause a dramatic rise in the potential[2] and those that diffuse horizontally may cause an decrease in signal for the neighboring pixels. Once all of the electrons in the pixel have either recombined or left the collecting node and a quasi-equilibrium state has been reached, $dS(t_o, \Delta t)/d\Delta t \rightarrow 0$. However, the fraction of hole filled traps that remain at the time of the next reset will cause the same effect, giving rise to a nonzero value for $dS(t_o, \Delta t)/d\Delta t \rightarrow 0$ in the following exposure, but with a diminished level. This is the crucial difference between the shallow trap model and the deep trap model, one that explains how the persistence can be "recharged" after a reset. In Section 7.5, the deep trap model will be considered in more depth to account for other features of the persistence in HyViSI detectors.

[2] Actually, applying Poisson's equation and integrating the electric field shows that even the electrons that simply migrate a small distance to the trapping sites will also cause an increase in the potential. But because the electrons that drift all the way to the back surface travel further in the electric field, they generate a larger change in potential.

7.2 Latent Images and Persistence

Hybrid CMOS detectors are prone to an effect called "persistence" in which regions of pixels that were previously exposed to significant brightness seem to "re-emit" that brightness after they have been reset. The "re-emission of brightness" forms a *latent image* of the object that emitted the light. CCDs show a similar phenomenon referred to as Residual Bulk Image (RBI). Both phenomena are usually attributed to the release of charge from *traps* somewhere in the detector, [42, 24, 113] usually near the surface of the bulk material closest to the collection node.

In the HyViSI detectors, a different type of this "persistence" effect is seen. Figure 7.1 shows a very dramatic case of this in a raw frame taken from an exposure of Saturn. The white shape at top is a saturated image of Saturn and five of its moons. Below it is a latent image from the previous exposure taken before the telescope pointing was offset to move the light from Saturn across the detector.

What is unique is that we do not only see regions of positive brightness in the subsequent exposures; we also see dark pixels surrounding the bright ones. Of course, in this sense brightness and darkness are simply terms that relate the signal level to the background far away from the affected area since the persistence shows itself even in exposures taken when the detector is in the

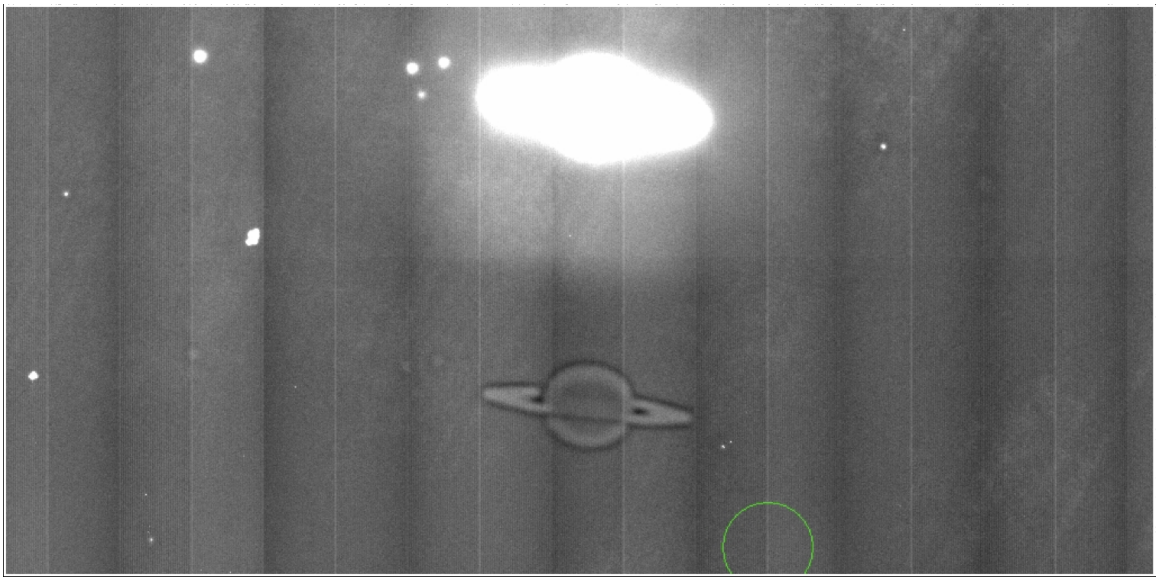


Figure 7.1: An extreme example of persistence in HyViSI detectors. This is a raw read (no subtraction or manipulation of pixel values) from an up the ramp exposure of Saturn. Five of its moons are visible: Titan, Dione, Tethys, Enceladus, and Rhea (from left to right). It is tempting to wonder whether the persistence might serve a useful purpose since it provides contrast information from a previous, completely saturated exposure that contained none.

dark. Because there is no inversion of the signal taking place, the regions of brightness are due to pixels that see a rapid increase in voltage with time. And because the pixels are collecting minority carriers (holes generated in the n-type substrate), this indicates that holes are accumulating in the p^+ implants of those pixels or electrons are leaving the implants. The regions of darkness are due to pixels that see a decrease in voltage with time, indicating that holes are leaving those p^+ sites, or rather that they are being populated by electrons.

Latent images present themselves in two distinct and fundamentally different ways in the HyViSI detectors: **1) persistent charge emission** and **2) slow-decay offset** or **Laser Burn-In**. The first of these gives rise to ephemeral latent images that possess the aforementioned dark and bright regions and vanish on the order of minutes or hours. The after-image of Saturn in Figure 7.1 is an example of persistent charge emission. The second type can last for months or years. These "burned-in" images occur for very large flux, and are hypothesized to be due to stressing of the source follower gate oxide in the multiplexer unit cell. They will be discussed in Section 7.3.

7.2.1 Persistent Charge Emission

As mentioned in Section 7.1.1, numerous studies have used Shockley-Read trapping physics to explain persistence in hybrid detectors. Under these assumptions, the time constant associated with a given trap determines the rate at which the traps are depopulated, and consequently, the rate at which the pixel signal changes. The data collected in our experiments show that there are multiple time constants involved with latent image decays, and in fact, traps that seem to be empty can be recharged by a reset of the detector. This behavior has been noted by several authors in the study of changes in the flatband voltage of Si-SiO_2 MOS capacitors [115]. They designated these rechargeable centers as "Anomalous Positive Charge" centers (APC), and showed they had a duration on the order of 60 minutes. HyViSI detectors exhibit a similar type of "rechargeable", long-term shift in operating voltages. However, there is evidence to indicate that the shift is due to diffusion of trapped charge carriers rather than an anomalous positive charge. In the next section, an introductory look at the signature of the latent images is given.

7.2.1.1 Spatio-Temporal Dependence of Persistence

Figures 7.2 and 7.3 are meant to convey the spatial and temporal behavior of persistent charge after saturation from a bright star. The first of these is a time sequenced set of 40 second up-the-ramp exposures taken with H2RG-32-147. In the very upper left plot of Figure 7.2, the radial profiles from sequential reads of the star in a saturating exposure are shown. The plots show the pixel values after the first read has been subtracted, so the fact that the radial profile close to the end of the exposure has a large divot near the center indicates those pixels have accumulated enough charge to rail the detector output, and presumably saturate the well. The image at the very upper right was taken from the last read of the same exposure and is simply a two dimensional representation

of the latter. In the other seven plots and images taken from the seven subsequent exposures, the telescope has been moved and the star is no longer in the picture. No star is present, so only the sky background light should generate signal in the image. However, the detector somehow "remembers" that a bright star was there previously, even though 15 resets were applied between each exposure. Its memory of the star shows itself as a distorted image that not only has a peak or "core", but a surrounding valley or "halo". The latent image deteriorates with time until after several minutes time, the detector "forgets" the star was there. It should be noted that all data shown in Figure 7.2 were taken with the detector exposed to the sky, so a non-negligible photon flux causes the signal to rise even after the persistence has subsided.

The spatial structure of the persistent image is especially interesting. Figure 7.3, a radial profile from a dark exposure taken immediately following a 35.4 second I band exposure, encapsulates some of the interesting features. Again, the numbers take negative values because the first read value has been subtracted from each pixel. The radial center $r = 0$ in the figure coincides with the centroid of a saturated star in the I band exposure. At this radius, the persistence signal has its maximum value I_{max}. At a radius of R_{core}, the pixels have seen *no net change* in signal. The may have risen and fallen back or vice versa, but they have returned to the voltage at which they started. The radius R_{min} coincides with the location where the pixels have seen the greatest *net decrease* I_{min}. And further out, R_{Halo} is the radius at which the pixels have again seen no net change. Beyond this radius, the pixel signals have not yet been affected by the persistent "cloud". Also shown in the plot is the positive core region shaded in blue and the negative halo in yellow.

The pixel values are reasonably well fit with the following empirical expression (shown in green):

$$I(r, t_0, \Delta t) = I_{max}(t_0, \Delta t) * \cos(\alpha(t_0, \Delta t) r^{\beta(t_0, \Delta t)}) e^{-r^2/\gamma(t_0, \Delta t)} \tag{7.13}$$

where t_0 is the time since the array was shuttered (or since the illuminating or *offending* source was removed), Δt is the time since the array was reset, and α, β, and γ are, in general, increasing functions of t_0 and Δt. It is important to note that no particular physical significance has been found in this expression; it is purely empirical. Moreover, the fit only works under certain limiting circumstances such as short exposure times, non-saturated pixels, and minimal PSF broadening due to atmospheric blur or tracking error. However, it has utility in some cases, as it may be used to fit and remove a persistent image from a subsequent exposure. It also provides insight into the structure of the persistent shape and how it evolves in time.

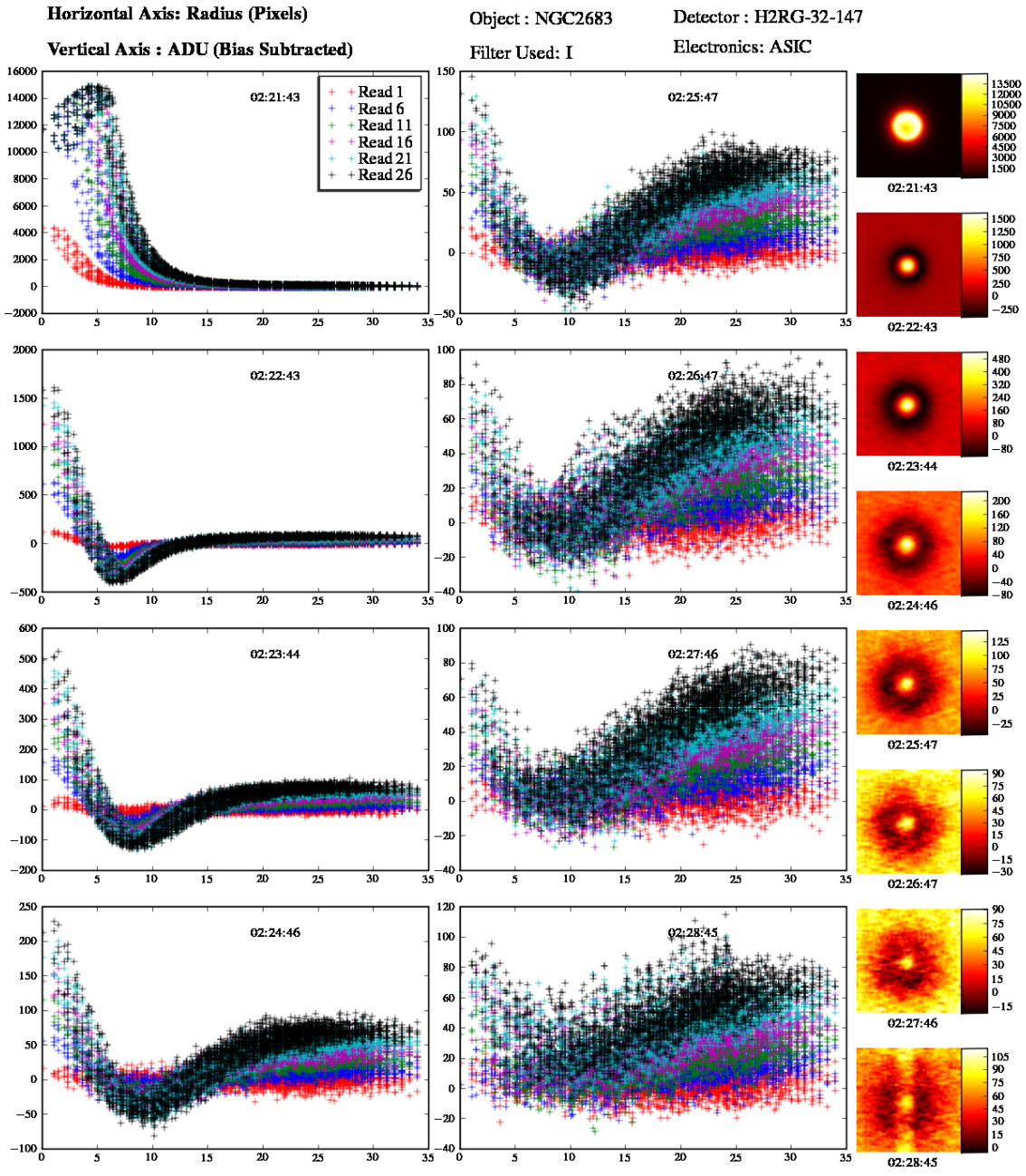


Figure 7.2: Radial profiles and image snippets taken from 30 read up the ramp exposures illustrating the evolution of persistence (starting at top left) in HyViSI pixels. The top left plot and upper image are of an actual star. In the other seven plots and images, no star is present, but persistent charge emission generates signal in the pixels.

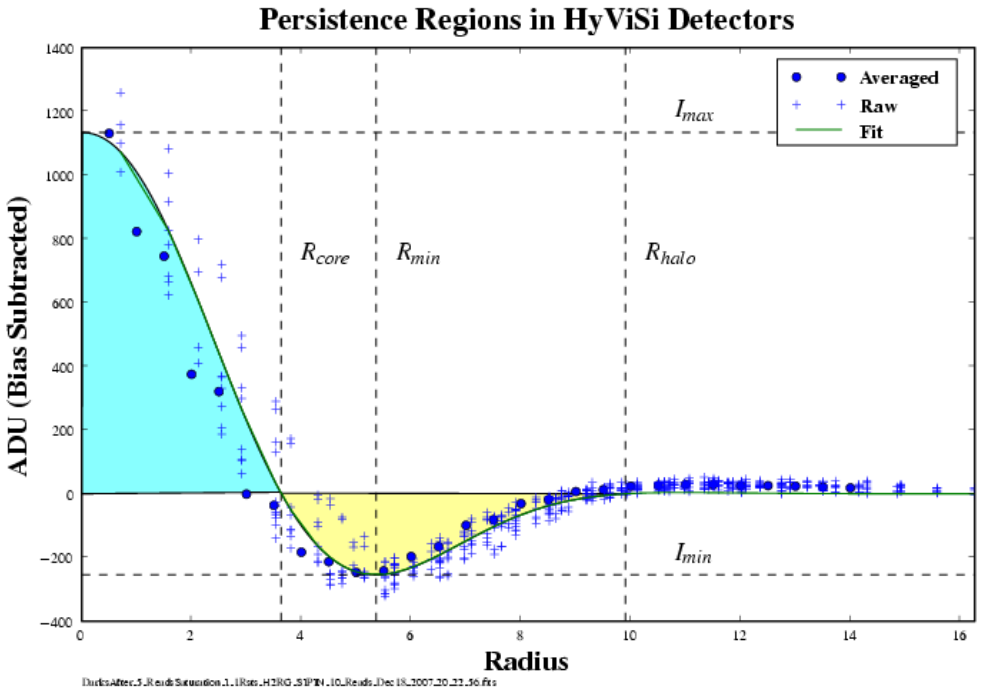


Figure 7.3: Radial profile for a region from a dark exposure. The radius is measured in pixels. A bright star was imaged in the same region in an I band exposure taken immediately preceding the dark. The blue shaded area represents the persistence core and the yellow shaded region represents the halo in this "core-halo" persistence. Several other coordinates are noted.

Although Equation 7.13 and the two figures shown do not address certain subtleties with the persistence (dependence on flux and fluence, wavelength dependence, etc), they do convey several key features.

1. **$I_{max}(t_0, \Delta t)$ is an increasing function of Δt**
   In an up-the-ramp integration it has a preliminary exponential dependence on the exposure time as it rises to some maximum value I_{Full} and goes like

$$I_{max}(t_0, \Delta t) = I_{Full} * (1 - exp^{-t/\tau_{rise}}), \tag{7.14}$$

   where τ_{rise} is the time constant for the rise. τ_{rise} depends on temperature, biasing, operating conditions, etc., but is typically on the order of seconds or minutes. As shown in Figure 7.4, for very long exposures, I_{max} will tend to decrease after it has reached I_{Full} in a quasi-linear fashion. This last aspect makes it very difficult to fit latent images in long exposures.

2. **$I_{Full}(t_0, \Delta t)$ is a decreasing function of t_0**

It has an exponential dependence on t_0 as it decays to zero:

$$I_{Full}(t_0, \Delta t) = I_{Full}(t_0 = 0) * (exp^{-t/\tau_{decay}}), \tag{7.15}$$

where τ_{decay} is the time constant for the decay. In some cases, it is on the order minutes; in others, hours. The decay time has a dependence on the number of resets and reads performed since $t_0 = 0$.

3. $I_{Full}(t_0, \Delta t)$ **is proportional to the incident flux/fluence in the offending exposure**. It shows a linear relationship to the fluence in the stimulus image in cases where the detector pixels were at or below saturation in the offending exposure. Above saturation, the relationship becomes nonlinear and eventually I_{Full} reaches its own saturation threshold.

4. $\cos(\alpha(t_0, \Delta t) r^{\beta(t_0, \Delta t)})$ **accounts for the oscillation from the positive core to the negative halo**
The fact that α and β are both increasing functions of the two time variables means that the halo is moving outward with time. This also means that R_{Halo} increases faster than R_{Min} and R_{Min} increases faster than R_{Core}.

5. $e^{-r^2/\gamma(t_0, \Delta t)}$ **indicates the structure is diffusing radially outward**
It effectively represents the fact that the overall core-halo structure is diminishing in amplitude and getting broader with time. The diffusive behavior of the structure suggests that its dynamics are, at least in part, governed by the diffusion of charge carriers.

Each of these 5 points, along with some subtleties in the core-halo behavior, such as the existence of multiple time constants, will be discussed in further detail in the sections that follow.

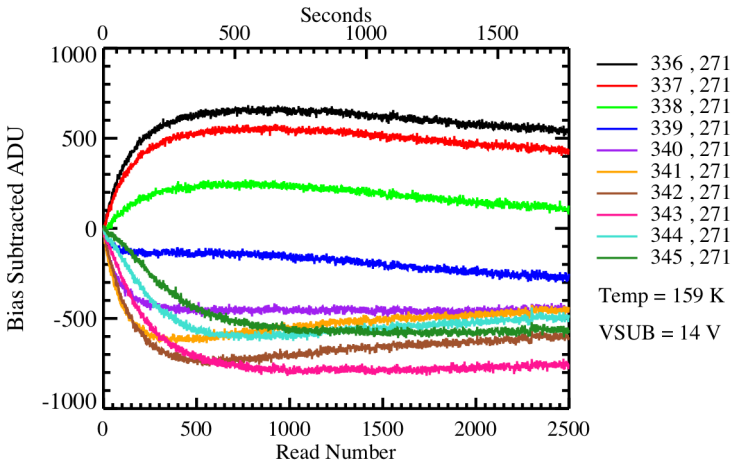


Figure 7.4: Persistence signal in a very long up the ramp exposure after saturation from a bright star. The signal vs. time is shown at different radii from the center of the core-halo image (pixel 336, 271). out toward the halo. Note that the pixels near the core initially see a rapid gain in signal followed by a slow, quasi-linear loss.

7.2.2 Dependence on Flux and Fluence

As Smith et al. note, it is a common misconception that persistence is caused by saturation or that it is a result of improper pixel resets [116]. For both SiPIN and per-pixel depleted infrared detectors, both of these statements are fallacious. In fact, persistence seems to be a process that is as intrinsic to the inner workings of a semiconductor detector as the photoelectric effect itself is. In this section, experimental results are shown that indicate latent images form well below the saturation threshold in HyViSI detectors. The dependence of the persistent signal on the flux and fluence of the *offending* stimulus will now be described.

7.2.2.1 Experimental Description

To probe a large range of brightnesses in a short time, the open clusters NGC 9256 and NGC 2395 were observed with H2RG-32-147 in full frame mode. The detector was operated at a temperature of 170 K and a substrate voltage of V_{SUB} =15 V. No idle resets were performed in between exposures. To measure persistence, the following procedure was applied:

1) An UTR exposure of the cluster with N_{Reads} reads was taken. Each read took approximately 1.41 seconds.

2) The filter wheel was moved to the blank position so the detector was no longer illuminated.

3) A set of 20 dark exposures was taken. Each dark exposure consisted of 1 reset and 10 reads. The read time was 1.41 seconds. The integration time for the 10 reads was 14.4 seconds and approximately 24 seconds elapsed (the IDL scripts took some time to write the file and header) between the start of consecutive exposures.

This process was repeated in each filter band (g, i, and y) for N_{Reads}=5, 10, 15, ..., 40. The range of brightnesses and exposure times provided an ample sampling over the detector well depth and beyond saturation.

7.2.2.2 Core Maximum and Halo Minimum

Figures 7.5 and 7.6 show the maximum and minimum persistence signal, respectively, attained in the first dark exposure of step 3) as a function of the maximum stimulus signal in step 1). Each data point represents a core (halo) value for a particular star in a particular image. For instance, a star that fills up 25% of the full well in the center pixel for N_{Reads}=5 will have a corresponding data point at about 50% of full well for N_{Reads}=10, 75% for N_{Reads}=15, etc. An estimate for stars that saturate the detector quickly in the stimulus images is made by extrapolating the photocurrent slope before saturation.

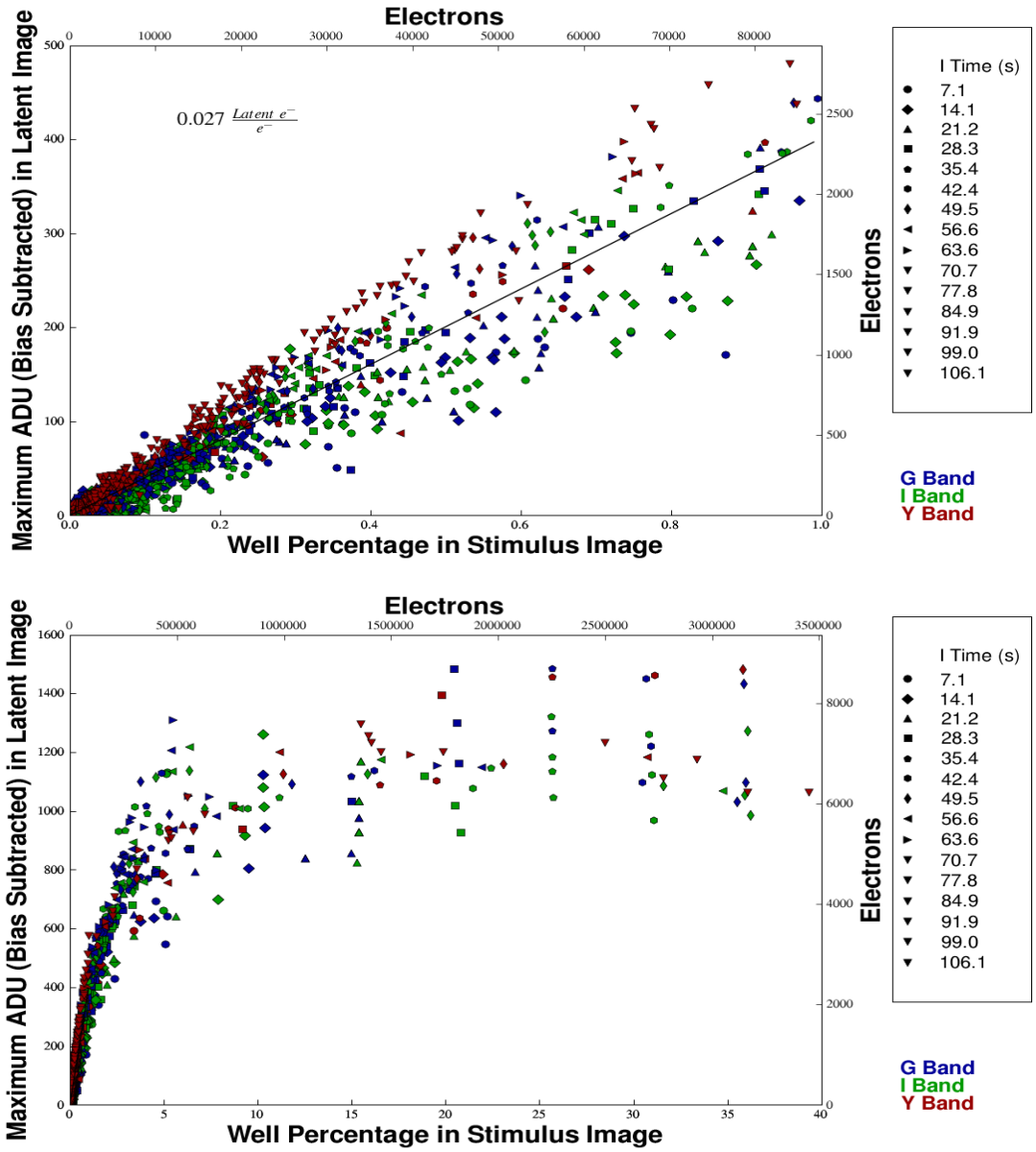


Figure 7.5: These plots show the relationship between persistence signal in the *core* and the stimulus that caused it. On the x axis is the maximum signal attained by the illuminated stimulus (fluence). Values beyond full well are extrapolated by fitting a slope to the pixel before it is saturated. On the y axis is the maximum persistence signal attained in a 14.4 second dark exposure following the illuminated stimulus. Each data point in these plots corresponds to a particular star in a particular image. The integration time for the point is listed on the legend at the right.

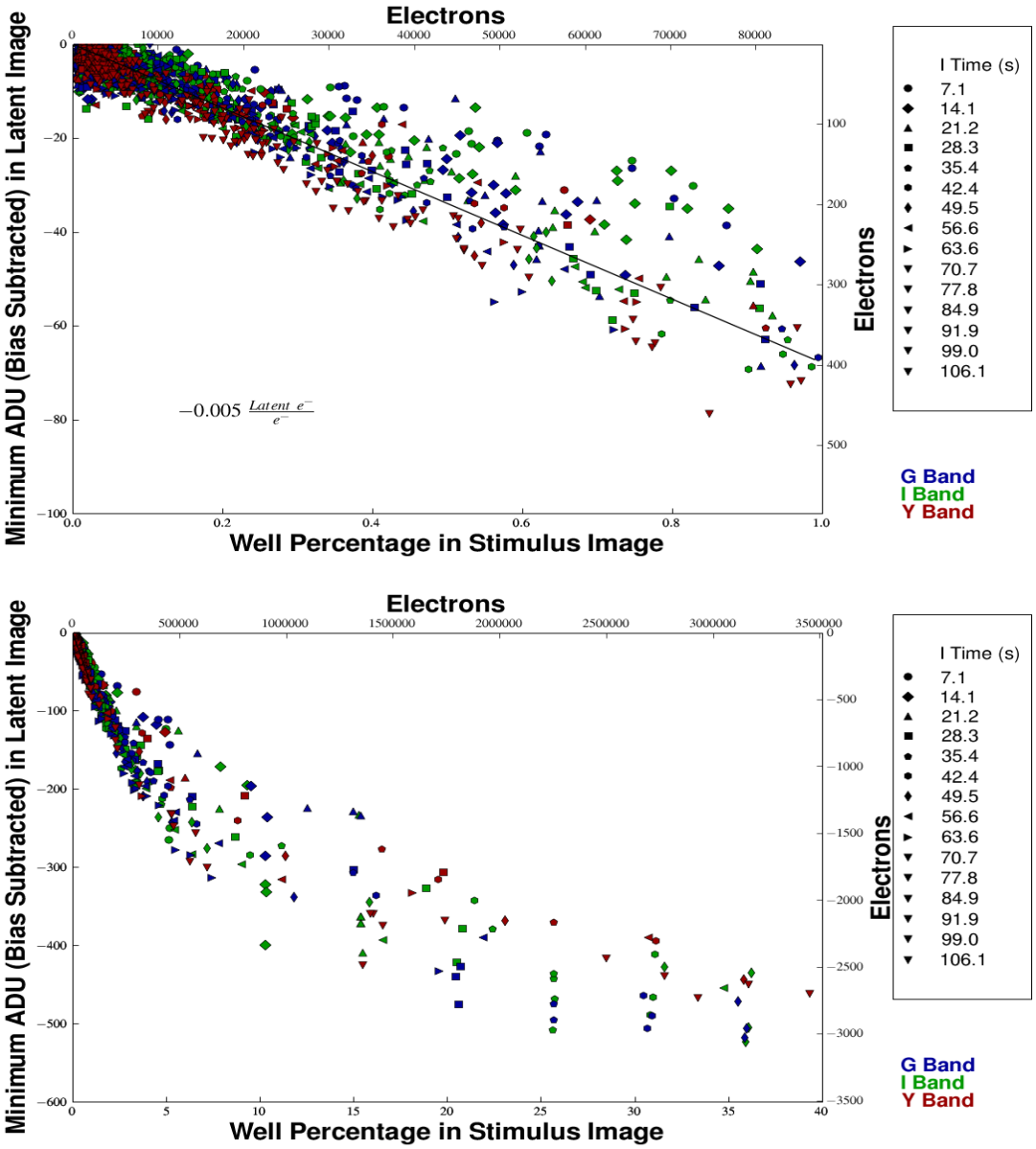


Figure 7.6: These plots show the relationship between persistence signal in the *halo* and the stimulus that caused it. On the x axis is the maximum signal attained by the illuminated stimulus (fluence). Values beyond full well are extrapolated by fitting a slope to the pixel before it is saturated. On the y axis is the minimum persistence signal in the negative persistence *halo* attained in a 14.4 second dark exposure following the illuminated stimulus. Each data point in these plots corresponds to a particular star in a particular image. The integration time for the point is listed on the legend at the right.

Note that the persistence response for both the core and halo is fairly linear with respect to stimulus from 0 to 100% of the full well (FW). This is proof that the detector pixels need not be near saturation to exhibit latent images. A linear fit yields about 0.027 latent image electrons per stimulus electron at the peak of the core and about 0.005 at the halo minimum. Again, it must be emphasized that the core pixels are seeing an increase in potential and the halo pixels are seeing a decrease in potential. The conversion from potential to electrons is made using the same conversion gain obtained from Fe^{55} calibration. After FW, the response becomes nonlinear and at about 10×FW, a saturation of the persistence signal is observed.

Also noteworthy is a slight flux/integration time dependence. In the plot of the maximum signal, for instance, the data points that lie above the line are mostly for exposures longer than 21 seconds and those below the line are for exposures shorter than 21 seconds. For a given well percentage (fluence), the exposure time needed for a star to reach that well percentage is inversely proportional to its magnitude (flux). Therefore, the short integration stars below the line have a higher flux while the longer integration stars above it have a lower flux. The correlation to the filter band is expected since most of the stars in these open clusters have intrinsic blue colors, meaning they are bright in the g band and faint in the y band.

The larger persistence signal for longer integration times might indicate that the accumulated holes are being trapped at the collecting node while the detector is integrating. To say this with certainty an additional experiment needs to be performed. In this experiment, a pulse of light of duration Δt would be shined on the detector at different times relative to the start and stop of an exposure of duration t_{exp}, where $\Delta t << t_{exp}$. If the holes are being trapped after collection, the persistence would be much greater when the pulse occurs near the beginning of the integration than when it occurs at the end.

7.2.2.3 Decay Time

The dark current signal due to persistence decays with a timescale that depends strongly on temperature and a number of other factors, which will be considered in the next section. While it can potentially last several hours, for the experiment described in Section 7.2.2.1, the heightened dark current typically decayed to its equilibrium value in 2-4 minutes. The decay appears to be exponential, and is better fit with two time constants (each with a separate coefficient) than one:

$$D_{avg}(t) = D_1 e^{-t/\tau_1} + D_2 e^{-t/\tau_2} + D_{equil}, \tag{7.16}$$

$D_{avg}(t)$ is the instantaneous dark current at time t, and D_1 and D_2 are components related to the time constants τ_1 and τ_2, respectively. D_{equil} represents the dark current after the persistence has subsided. Solomon fit persistence in InSb arrays with the same functional form[3] and attributed the

[3]The actual fit was to the number of holes released from traps as a function of time. This can easily be obtained by integrating Equation 7.16 from $t = 0$ to $t = \infty$.

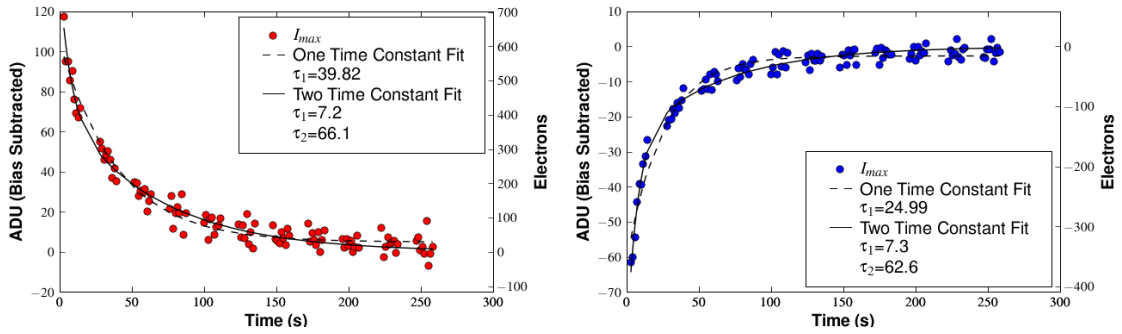


Figure 7.7: (Left) The decay of dark current for I_{Max} at the core center. In the core, the signal voltage increases with time. (Right) The decay of dark current for I_{min} at the radius R_{halo}. In the halo, the signal voltage decreases with time.

two time constants to two energy levels of a single trapping state [42].

Figure 7.7 shows the two time constant fit to the instantaneous dark rate observed after the persistence stimulus. Both the dark signal vs. time at the center of the core-halo structure and that at the radius R_{halo} are shown. The time constants are very similar for this example, and Figure 7.8 shows that this trend holds for the majority of the latent images that were successfully fitted.

The fact that the time constants for the rise of I_{max} and fall of I_{min} are so close in value strongly suggests that charge is being transferred from the core to the halo. If the rising signal in the core and decaying signal in the halo were the result of one species of trap emitting electrons and another species emitting holes, one would expect the emission time constants to be different since they depend on the trap energies and cross sections. It is thus reasoned that the core-halo persistence is a result of charge carriers being exchanged between the pixels in the core and pixels in the halo.

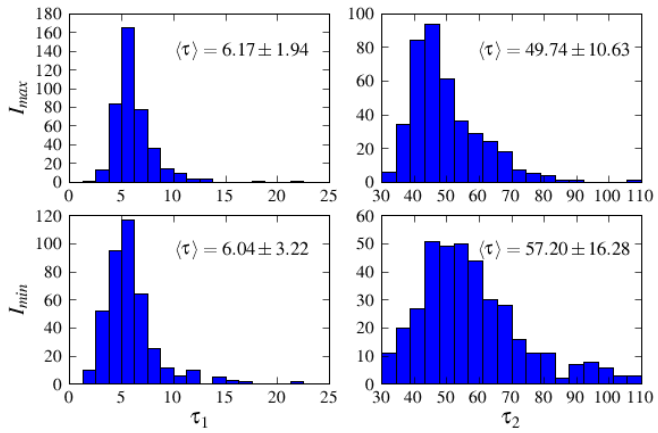


Figure 7.8: Histograms of the time constants for persistence decays of I_{max} at the center of the latent image and I_{min} at the radius R_{halo}. τ_1 accounts for the fast portion of the decay and τ_2 accounts for the slower portion. The time constants are in units of seconds.

7.2.3 Dependence on Detector Activity

In some cases latent images in the HyViSIs can last far longer than 3-5 minutes. This very long duration persistence is linked to the environmental and operating conditions of the detector, i.e. the temperature, whether the detector is exposed to light or darkness, whether the pixels are left to integrate or are being held under reset, the bias voltages, etc. While no study was performed to explicitly explore the dependence of the latent images on this parameter space, the dominant culprit appears to be a lack of resets on the pixels. The general statement can be made that *more than one reset must be performed to remove latent images in the HyViSIs.* If only one reset of the frame occurs after stimulus and an hour long dark exposure is taken in which latent images are present, the next exposure will show the latent images as well (unless they are extremely weak).

Figure 7.9 shows an example of a latent image that endured for more than an hour. The dark exposures represented in the plot were taken by H2RG-32-147 in window mode with a 150 × 150 window (V_{SUB} = 15 V, T=170 K). Interestingly, at the end of the first dark exposure, shown in the left third of the plot, the dark current in the center of the latent image **decreases to zero**. The

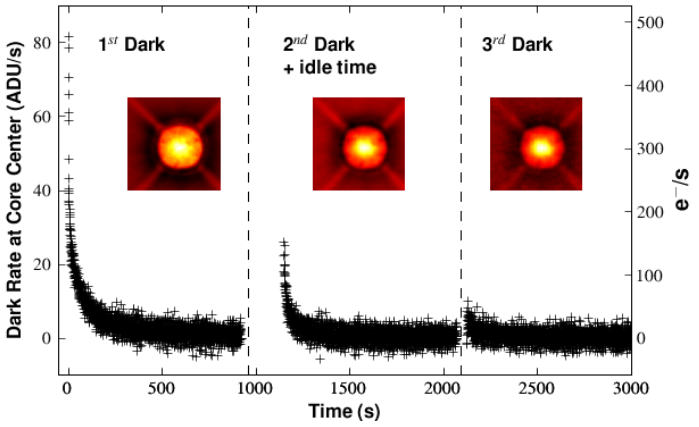


Figure 7.9: An example of a latent image that persisted for more than an hour. The plus marks show the instantaneous dark current measured at the center of the latent image (i.e. $\Delta I_{max}/\Delta t$). The persistence dark current appears do be gone at the end of the first exposure, but reappears in the second.

dark rate in most of the core pixels has done so also. Meanwhile, the pixels at the periphery of the halo continue to show a negative dark current (decrease in voltage) and the outer edge of the halo continues to move radially outward. Surprisingly, after a small amount of idle time following this exposure and one reset, a high dark current in the core pixels reappears in the next exposure. And after an additional reset, it appears again in an exposure taken more than 2000 seconds after the initial stimulus. The latent image thus endures far longer when only two resets are applied than it did in the experiment of Section 7.2.2.1 when resets were being applied every 24 seconds.

Emission of holes from shallow traps cannot be used to explain this behavior. If the increase in core signal was a result of holes in shallow traps ($E_T - E_V << 3kT$) at the surface or in the bulk being emitted at a rate e_p, the emission would continue until all of the traps are

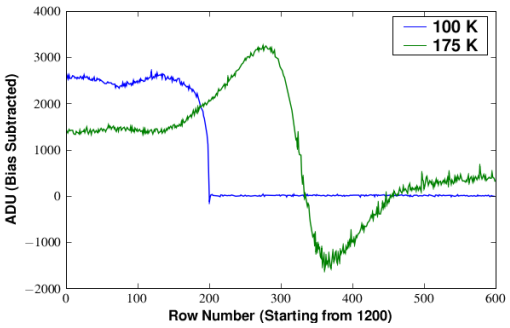
empty, regardless of whether or not the pixels are reset or left to integrate (see Section 7.1.1 for an explanation of why this is so). If conventional trap theory is to be used to explain this behavior, it suggests that the traps are deep traps, and that sometime after reset, these deep traps are eventually "starved" of carriers with which to recombine. Additional resets are required to inject the carriers necessary to "feed" these traps and restore the detector to the equilibrium state before the stimulus occurred. A qualititative explanation of how this may occur is pursued in Section 7.5.

7.2.4 Dependence on Temperature

The measurements made for this dissertation indicate that persistent charge emission takes place at temperatures from 90 K-200 K. Below this range, it is likely that persistence still occurs; 90 K was simply the lowest temperature that was probed. Above it, the thermal dark current is so high that it completely swamps out the persistent dark current. A thorough study of the relationship between the temporal and spatial characteristics of the latent images and temperature was not carried out due to a lack of time and resources. Nearly all of the latent images in our data were recorded between 160-170 K. However, from the limited sample of data that was collected, several important qualitative aspects stand out.

Figure 7.10 shows the most noteworthy feature of the core-halo persistence at temperatures below 110 K: that is, the halo is almost non-existent. In the image at bottom right, which shows the last minus first read of an exposure taken immediately after an LED light source was removed from the H2RG-001 at 100 K, the halo is nearly undetectable because its amplitude is much smaller than the range of pixel values in the stretch. In the plot at left, which shows a plot of the pixel values along the column highlighted by the cyan line in the image, one can see that the halo is a small dip that occupies 1-2 pixels at the edge of the latent image. The pixels in the halo lose only 1-2 mV of signal in 14 seconds at this temperature. Meanwhile, the pixels in the core gain about 35 mV of signal, which is comparable to the amount of signal gained in 14 seconds by the core pixels in a 160 K latent image. Thus, the behavior of the halo pixels is extremely different in these two temperature extremes, while that for the core pixels is somewhat similar.

Although the core pixels share a common persistent dark rate at these temperature extremes, Figure 7.10 reveals a difference in the way they act. The core pixels near the center of the latent image (which has a radius of about 600 pixels at 100 K and a radius of about 800 pixels at 180 K) form a rather flat pedestal in signal at any time, t, for both temperatures. In other words, they are all seeing the same dark rate, $D(t)$, over the exposure time. At 100 K there is a very small difference in the dark rate at any location in the core, and it looks essentially flat. But at 175 K, near the core-halo boundary, the pixels in the halo see a much greater $D(t)$ than the ones near the center, resulting in the peak in the left plot, and the pixels immediately outside see a negative $D(t)$, resulting in the trough. One cannot help but reason that the formation of the peak and trough is due to the horizontal exchange of charge carriers somewhere in the detector.



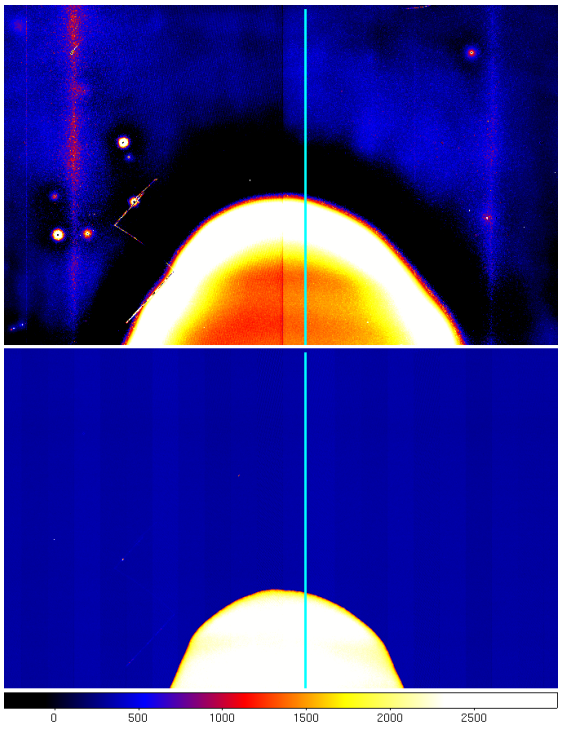


Figure 7.10: (Right) The images at right were taken from an exposure after an LED illuminated and saturated the detector H2RG-001. The first read has been subtracted from both. The top was taken at 175 K and the bottom at 100 K. (Above) A column plot showing the bias subtracted signal vs. row number. The column used to make the plot is indicated by the cyan line in the images at right. At 100 K, the halo occupies only 1-2 pixels.

Nearly all properties of silicon—the carrier mobilities and diffusion coefficients, intrinsic carrier concentration, resistivity, etc.—are dependent on temperature. But if we believe that charge transport through the detector material is responsible for the halo and peak near the core edge, then the most likely candidate variable is the diffusion coefficient for electrons, D_n. Electrons injected into a pixel at reset do not see a large inter-pixel potential barrier if the front surface between p^+ implants is in an accumulated state, so they are free to diffuse. For pixels near the center of the large core, there is no gradient in the concentration of these electrons, and thus no diffusion to cause a discrepancy in the dark rates. A change in signal level for these pixels still occurs as electrons diffuse toward the back surface, though. Near the edge of the core, where the pixels outside have not been filled with electrons, there is such a gradient, and the diffusion adds an additional component to the signal change. At low temperatures, the diffusion of these electrons may occur too slowly to cause an appreciable decrease in the halo or let it reach an appreciable distance. At higher temperatures (the diffusion length should go as $T^{3/4}$ [117]), the opposite is true. The core-edge pixels see a larger signal rise because of the electrons diffusing horizontally out of them and the halo pixels that receive these electrons see a decrease. This theory will be explored further in Section 7.5.

7.2.5 Dependence on V_{SUB}

The latent images show very interesting spatial signatures as V_{SUB} is varied, especially below 8 V, when the detector is not fully depleted. As can be seen in Figure 7.11, the spatial distribution of the photoholes collected during the stimulus image are equally interesting. This figure encapsulates a qualitative summary of a very long imaging process performed at the telescope with H1RG-022 and warrants a thorough description.

7.2.5.1 Description of Latent Image Exposure Sequence

Each of the image snippets in Figure 7.11 shows the fourth exposure out of a four exposure sequence at a particular value of V_{SUB}, listed in green. They were all recorded with H1RG-022. In this fourth exposure, the star being imaged is in the upper left quadrant. In the third exposure the star was imaged in the upper right quadrant, in the second it was imaged in the lower right quadrant, and in the first it was imaged in the lower left quadrant. For values of $V_{SUB} > 1.5$ V, latent images are evident in all three of these quadrants. The latent image in the upper right quadrant has the "freshest" persistence since it was stimulated 71 seconds prior to the fourth exposure and the one in the lower left is the "least fresh" since it was stimulated 213 seconds prior. The image snippets were made by subtracting the last read of a 100 read exposure from the first (bias subtraction). No dark current subtraction or flat field was applied. The telescope was dithered by about two arcminutes between exposures to bring the star to its new location. Idle resets were performed at $V_{RESET} = 0.094$ V in between exposures and the temperature was 170 K.[4] The detector was operated in 16 output mode with a frame time of $t_{frame} = 0.71$ s.

7.2.5.2 Charge Collection

It is immediately clear from Figure 7.11 that charge collection in the pixels depends strongly on the backside voltage. The diffusive behavior in the 1.5-3.0 volt range agrees well with the model presented in Section 2.2.2.2 in that it shows fairly radially symmetric diffusion when the depletion regions collapse. For $V_{SUB} < 1.5$V and $V_{SUB} > 3.0$V, on the other hand, there are substantial deviations from radial symmetry that are difficult to overlook.

For $V_{SUB} \leq$ 1.5 V a clear pattern following the columns in the multiplexer is evident. The pixels along the odd multiples of 64 (OMC) see a rapid decrease in voltage of $\sim 50 - 60$ mV in about 10 seconds. After the drop, they either begin to integrate or hold steady. The pixels along the even multiples (EMC) see a quasi-exponential increase of $\sim 120 - 190$ mV over the full 71 seconds. For pixels between two OMCs, the change in voltage over the 71 seconds increases monotonically from

[4] If resets were not performed while the telescope was moved, the star would leave a latent image *trail* in its path along the detector. These trails are present throughout the telescope data from the instances when no resets were performed while the telescope was being slewed to a new field.

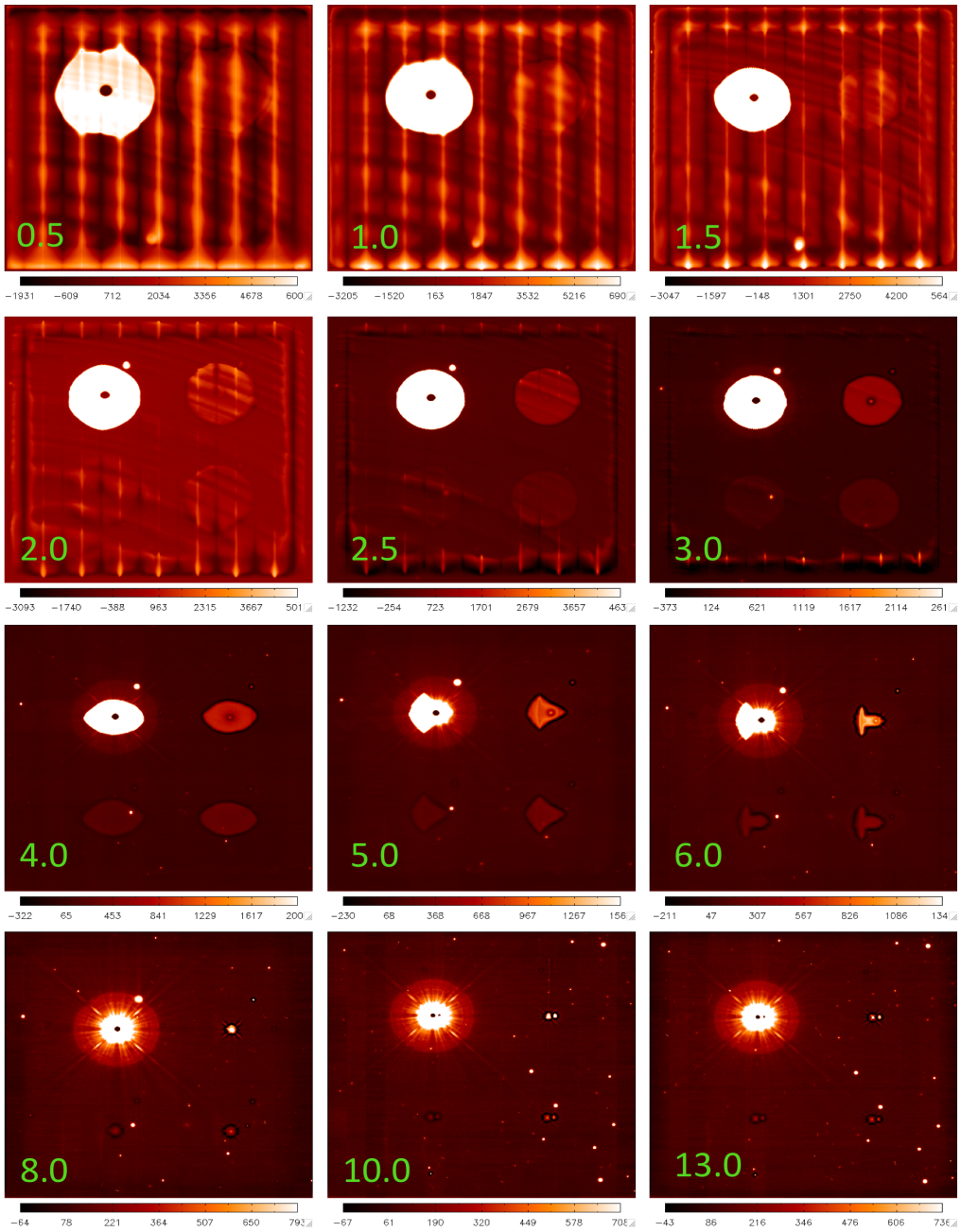


Figure 7.11: Latent images for saturating stars at different values of V_{SUB} taken through i band. The value of V_{SUB} is listed in each image snippet in green. See the text for a description of the imaging sequence.

−60mV to 190 mV and then decreases back down to −60mV along the row. This behavior suggests a horizontal gradient in the detector front surface potential, which might arise because of a voltage differential in the multiplexer. The physical layout of the multiplexer is unknown, so determining the specific cause is difficult. Some possibilities are coupling to current carrying lines that are located at different depths beneath the plane of the detector or to the currents going through the 16 outputs (the schematic in the manual indicates that the individual column buses carry the signals off chip before they are actually routed to the outputs, though, so the latter may not be realistic; also, the pattern is present even when the detector is run through one output).

Charge carriers generated by the star diffuse radially outward, but advance more quickly along the EMC and less quickly along the OMC. While this behavior is interesting, understanding it is of little value for astronomical application. Operating the detector at such low bias voltages would yield extremely poor performance.

For 1.5 V $< V_{SUB} \leq$ 3.0 V the observations match very well with the expectations outlined in Section 2.2.2.2. After the depletion regions of the illuminated pixels collapse, the carriers diffuse radially outward and destroy the depletion regions in a new annulus of pixels, creating a domino effect of blooming.

For 3.0 $< V_{SUB} <$ 8.0 V a clear asymmetry in the diffusion is evident. At 4.0 V, the carriers diffuse more strongly in the x direction, resulting in a highly elliptical star. At 5.0 and 6.0 V, they diffuse more strongly along the x direction as well. Oddly, though, another asymmetry is observed between the $-/+x$ directions at these voltages. Once carriers diffusing in the $-x$ direction arrive at one of the OMCs, they show a strong tendency to follow the OMC vertically. Note that this is opposite to the behavior for $V_{SUB} <$ 1.5 V, where the charge followed the EMCs. In the $+x$ direction, the charge cloud tightens as it moves. The most logical explanation for this is that the front-side potential is lower along the OMCs than it is along the EMCs at these bias voltages. The holes thus tend to sink into these wells more easily.

It is not at all clear why the disparity in surface potential is so strong when V_{SUB} is between 5-6 V or why such a large asymmetry sets in at 4 V. Understanding this would require a two or three dimensional simulation on top of intimate knowledge of the detector and multiplexer layout. A simulation of this nature would be useful, though, since *these asymmetries might very well offer a clue into the $x-y$ asymmetry exhibited in pixel crosstalk and interpixel capacitance* (see Sections 6.1.4 and 6.1.3). With or without comprehension of this phenomenon, the "point spread functions" at these voltages demonstrate that the detector should be operated at $V_{SUB} \geq 8$ V for astronomical applications.

For $V_{SUB} >$ 8.0 V the detector achieves more ideal performance as it enters the overdepleted state. There is no sign of the depletion regions collapsing and allowing photocharges to invade

neighboring pixels. Bright stars still generate significant photocurrent outside the FWHM, but this is a result of diffracted light and not diffusion of carriers from field free regions.

7.2.5.3 Charge Persistence

Perhaps the most striking attribute of latent images generated when the detector is not fully depleted is *the presence of two core-halo structures* (CHS). This is most easily seen in the $V_{SUB} = 3.0$ V image of Figure 7.11. One CHS, call it CHS$_{inner}$, emerges near the center of the image, with $r_{core} \sim FWHM$ of the star in the previous exposure and $r_{halo} \sim 2 - 3 \times FWHM$. Another, call it CHS$_{outer}$, emerges at the outskirts of the charge distribution created by the star in the previous exposure. In between these two regions, the pixels show roughly the same increase in signal due to persistence, making it one large core, effectively.

Figure 7.12 provides a key into understanding how this happens. The observations to this point imply that persistent signal arises because of carriers trapped near the p$^+$-n junction, i.e. near the front surface (the way in which these trapped carriers actually *generate* persistence will be discussed in Section 7.5). If the number of trapped carriers is proportional to both the number of stored majority holes in the collecting node **and** the excess minority holes in the bulk, then the trapped carrier distribution should look similar to the one shown in Figure 7.12. One can see that the locations of the CHS$_{inner}$ and CHS$_{outer}$ are seen to coincide with the steep gradients in hole concentrations at r_1 and r_2, respectively. It was already shown in the previous section that diffusion is a likely candidate for the charge transport that generates the CHS. The existence of CHSs in regions where there is a strong gradient dp/dr in carrier concentration provides further support for this. And when the field free diffusion regions vanish in overdepletion, r_2 is drawn in and merges with r_1 since $\Delta p_{diff} = 0$. This explains why CHS$_{outer}$ vanishes as V_{SUB} is increased.

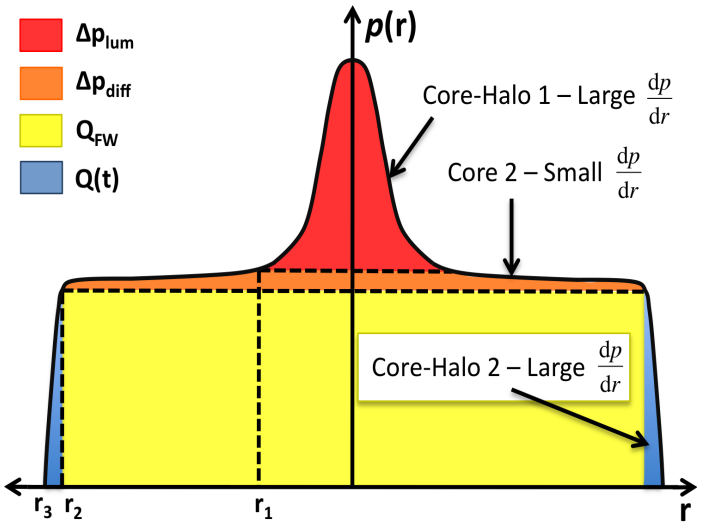


Figure 7.12: Number of stored and free holes versus radius when light from a saturating star is incident on the undepleted HyViSI. Following Section 2.2.2.2, the stored holes, Q_{FW} and $Q(t)$ have accumulated in the p$^+$ implant while the free holes, Δp_{lum} and Δp_{diff} exist as excess minority carriers in the field free regions near the metallurgical p$^+$-n junction. At reset, the large gradients in these holes translates to large gradients in injected carriers, and core-halo persistence occurs at these gradients.

7.3 Semi-Permanent Offsets: Laser Burn-In

There is another type of "after-image" that can be generated in HyViSI devices, which is of a fundamentally different nature than the short-term persistence just described. We refer to this effect as **Laser Burn-In** (LBI) because it has been accidentally discovered by several groups who used a laser to illuminate the detector in laboratory experiments. Of course, any light source that produces a beam of comparable intensity to a laser should create such an after-image. Figure 7.13, which shows two offsets created by imaging Mars through the 2.1m telescope with H2RG-32-147, proves this is indeed the case.

The unique thing to note about Figure 7.13 is that it is a reset frame of the detector, meaning that the pixel values are read while the reset switches are held closed in the pixels of the multiplexer. This is very convincing evidence that the origin of the offset lies in the ROIC and not the detector. Bai et al. [25] provide confirmation of this, attributing the offset to "damage" to the gate oxide of the pixel source follower, which has a minimum rated breakdown voltage of ~ 5.2 V (damage is placed in quotes because the offset will eventually anneal after a very long time, as will be shown). In infrared hybrid CMOS detectors, the effect is not observed because the small reverse bias used (< 1 V) does not allow the sense node to get this high. But with the large backside voltage needed for astronomy (15-50 V), the pixel source followers in HyViSI sensors are susceptible to this damage.

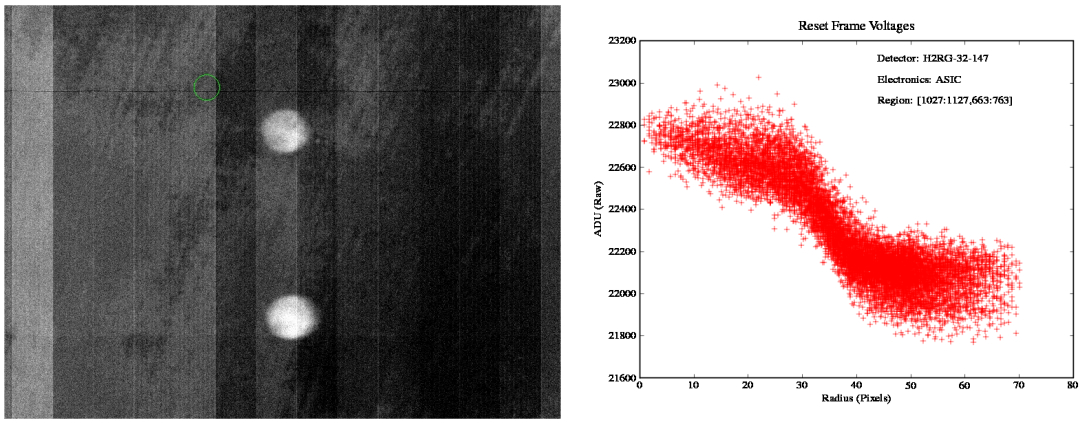


Figure 7.13: (Left) Image of raw ADU values in a reset frame showing the offset induced by imaging Mars through the Kitt Peak 2.1m telescope. (Right) The radial profile shows the elevation in ADU values relative to the background. The peak to valley of ~ 700 ADU corresponds to about 45 mV.

The conditions under which the LBI occurs, referred to by Bai et al. [25] as *supersaturation*, are a complicated function of actual breakdown voltage, detector bias, temperature, input light level, integration time, and the time between resets of the sense node. The empirical evidence indicates that the biggest of these factors is the amount of flux falling on the detector. For instance, a short, 10 second exposure with a 5 pixel FWHM laser beam at a high flux of 3.54 kW/m^2 and $V_{SUB} = 7$ V

was enough to create a 50 mV offset at room temperature in a HyViSI H2RG. But for long duration exposures of bright star fields, where the fluence of a given star was 50-100 × full well with a low flux of < 10 W/m^2, no offset was created at $V_{SUB} = 15$ V. The reason for this is that there are leakage paths (such as the thermionic diffusion discussed in Section 6.1.4.3) to remove the accumulating holes and prevent the voltage at the sense node from getting as high as 5.2 V in the low flux case. It is only when the amount of photocurrent generated above the pixel is sufficiently large that it cannot be effectively removed through the leakage paths where the voltage can rise above 5.2 V. This happens with high flux conditions.

The exact relationship between the operating conditions and the resulting offset is not known, nor is the maximum voltage offset that can be created.[5] In the frame shown in Figure 7.13, the offset is approximately 45 mV. In other laboratory LBI, offsets as large as 120 mV have been induced. It is possible the upper limit for the offset may extend even further. Fortunately, **the LBI offset is purely DC and has no measurable effect on photocurrent estimates using CDS, Fowler Sampling, or slopefitting**, which are the methods used to measure signal in astronomical applications. Unfortunately, the offset reduces the practical well depth in the affected pixels, and hence, their dynamic range. For instance, immediately after observing Mars with H2RG-32-147, the stressed pixels started at a level 45 mV (800 ADU at a gain of 1 on the SIDECAR) above the level measured prior to burn-in, but still saturated at the same level as the one measured prior to burn-in. Thus, their effective well depth is reduced by about 4500 e$^-$. Several tests were run with adjusted ADC ranges for the readout electronics to confirm that the detector output is saturated and not the ADC.

It was originally speculated that the LBI was a form of permanent damage to the sensor. As Figure 7.14 shows, though, the offset does anneal over time. This can take anywhere from days to several months. For the particular case shown in Figure 7.14, the spot decayed over six months while the detector was kept at room temperature and used somewhat infrequently. The Mars burn-in showed no signs of a reduction over three months time with the detector cooled between 100-200 K and used regularly. While three to six months may seem like an extremely long time, previous studies have shown that in Si-SiO_2 interfaces, traps in SiO_2 can exchange charge with a Si substrate on time scales ranging from less than 1 μs to many years [118]. In fact, many electrically programmable memory devices rely on such a longevity of trapped charge in oxides to store data [119]. There is some indication that constantly resetting the detector can aid in disintegrating the spot over a much quicker time, though. A 120 mV offset burned into H2RG-148 with a 1kW/m^2 laser disappeared over 10 days. The detector was held at 160K and $V_{SUB} = 15$ V for the majority of this time with idle resets being constantly applied to the detector while it was not taking exposures.

Regardless of the time required for the LBI to vanish, the loss in well depth is obviously a concern for astronomical imaging since the largest possible dynamic range is desired. To avoid

[5] There is some degree of apprehension involved with the LBI since it could potentially compromise these very expensive devices.

damage, Teledyne Imaging Sensors recommends either 1) biasing the detector with $V_{SUB} < 5$ V or 2) avoiding oversaturation. The first of these is highly impractical for astronomical applications, since the PSF at such a low bias voltage is quite large and takes on the strange shapes shown in the previous section. Also, the dark current volcanoes occupy a much larger radius, rendering many pixels unusable when the detector is not fully depleted. The second of these might be a potential solution, although it will truly be a challenge to implement. It requires knowing in advance the position of bright stars and planets and using the guide mode feature of the detector to reset the pixels subject to the most concentrated light. Resetting the pixels will prevent supersaturation and the ensuing damage, with the side effect of creating offsets along the columns and rows of the guide window (see Section 6.1.4.3).

In certain optical systems, the size of the telescope aperture, filters used, plate scale, etc. might preclude the LBI from presenting a threat at all. For instance, in our three weeks of 2.1m observations, in which we observed many stars between 1^{st} and 5^{th} magnitude, the only source that caused the offset was Mars. In fact, imaging the planet through g, i, and y did not create an offset. It was only when no filter was used that the damage was done. An accurate calculation of the flux that Mars produced at the detector when no filter was present based upon its apparent magnitude is not possible. The detector was saturated before the first read, so estimating the flux from photocurrent is not an option, and the transmission of the mirrors and detector window are not known, so using the magnitude to obtain a flux at the detector would be equally uncertain. It is likely that it was somewhere on the order of 1kW/m^2 like the lasers that produced LBI in the laboratory.

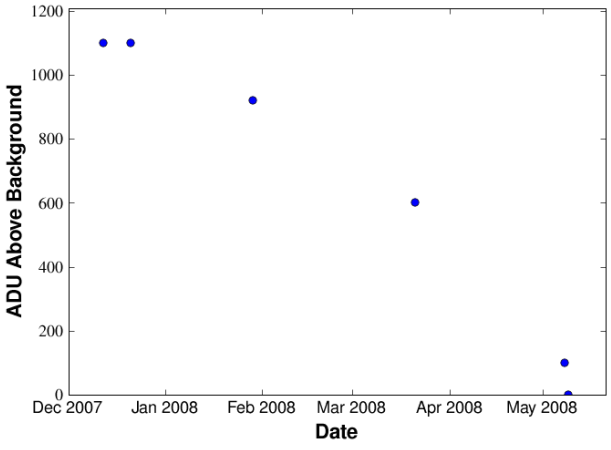


Figure 7.14: A plot showing the average LBI offset in H2RG-32-029 over a 6 month period. The spot was burned in by a monochromatic laser with a 650 nm wavelength. At best focus the beam produced about 3.54 kW/m^2. The detector was kept at room temperature throughout the six months. The plot shows that the spot eventually disappeared. The gain was approximately 6 e$^-$/ADU.

7.4 Effect of Forward Biasing

In his studies on persistence in infrared hybrid CMOS detectors with an InSb detector layer, Solomon noted that forward biasing the photodiodes effectively removed latent images [42]. He found that while the latent image could no longer be detected when the reverse bias was restored, the entire detector suffered from an instability in dark current. Similarly, Smith et al. demonstrated that the rise in dark current after incrementing the reverse bias across photodiodes in HgCdTe arrays is almost identical to the rise in dark current after exposure to a light source [116].

In HyViSI detectors we find a nearly analogous behavior. Forward biasing the photodiodes by bringing V_{RESET} higher than V_{SUB} and resetting the pixels clears any spatial signature of the latent image. In normal operation, this amounts to decreasing the output voltage on the power source that supplies V_{SUB} since V_{RESET} should not be brought above 1.7 volts. Unfortunately, when we reverse bias the diodes again and take another exposure we find the dark current has risen by several orders of magnitude and takes on the same exponential behavior as the persistence signal.[6] Effectively, **every pixel is afflicted with persistence** after the forward bias. It requires minutes to hours for the dark current to settle nonlinearly back to its equilibrium value.

In the following sections, we present an overview of the theory behind switching from forward to reverse bias in a PIN diode. We then describe experiments which show that the dark current signal behavior—the characteristic timescales and magnitude—are essentially the same in latent images and after forward biasing. Lastly, we show that varying the number of resets and the readout sequence after the reverse bias has been restored does not alter this behavior.

7.4.1 Theory of Bias-Direction Switch

The time required for the dark current to settle to equilibrium after going from forward to reverse bias is a major consideration in the design of PIN diodes for switching applications [54]. When a PIN diode is forward biased, majority carrier electrons are injected into the intrinsic or ν-type region from the n^+ region and majority carrier holes are injected into the ν-type region from the p^+ region (see Figure 2.1). The greater rate of recombination of carriers when holes from the p^+ side meet the electrons from the n^+ side over generation leads to an accumulation of charge in the ν region. Excess carriers also accumulate in the heavily doped regions. To reach equilibrium in the reverse bias state, these injected carriers must be swept out of the diode to form the depletion region.

Simulations and theoretical calculations show that the heavily doped contacts are depleted quickly (on the order of $1/20^{th}$ the lifetime of the minority carriers) relative to the intrinsic, or lightly doped ν region [120, 121]. The time it takes for the latter to occur is directly related to the width of the diode W, and is unaffected by the minority carrier lifetimes as long as the exterior circuitry permits large sufficiently large reverse currents [121]. Interestingly, in PIN diodes used for

[6] Unless a distinction is necessary, the process of forward biasing the diodes and returning them to their original reverse bias value will be referred to as a *bias switch* for the sake of brevity.

power switching applications, this time is on the order of milliseconds or microseconds. For the HyViSI PIN diodes, it is on the order of minutes, suggesting that another mechanism is responsible for the long time taken (or that the reverse current cannot reach some spots in the detector layer).

7.4.2 Forward Bias to Full Reverse Bias of $V_{SUB} = 15$V

After observing the removal of latent images with a forward bias, one is inclined to ask: *does this method offer a better solution for persistence mitigation than simply waiting for the latent images to subside*? Further, *does varying the time the pixels are held in reset versus the time they are allowed to integrate affect the time required for the detector to return to its equilibrium state*? An experiment was carried out to answer these questions and gain a better understanding of the mechanisms at play with latent images.

7.4.2.1 Experimental Description

The experiment is carried out as follows:

1) After the detector dark current has reached its equilibrium value, an exposure is taken consisting of 1 reset (line by line) and 10 reads at $V_{SUB} = 15$ V and $V_{RESET} = 300$ mV. Line by line reset is used.

2) The backside voltage is brought down to $V_{SUB} = 0$ V and another exposure with 10 (line by line) resets is taken at $V_{RESET} = 300$ mV. The PIN diodes are forward biased while the reset is held down and the pixel values are read out while the resets are performed.

3) The backside voltage is brought back to $V_{SUB} = 15$V. Exposures are then taken with a cadence of N_{Resets} and N_{Reads} while maintaining $V_{RESET} = 300$ mV until the dark current returns to thermal equilibrium.

4) Steps 1-3 are repeated with a different cadence.

All data was taken with the H1RG-022 at 170 K in 16 output full frame mode. The power supply that sourced V_{SUB} was adjusted through an RS-232 interface by the same IDL script that controlled the imaging sequence, which provided precise, uniform timing for each operation. Approximately 34 seconds elapsed between both the drop in V_{SUB} from 15 to 0 V and the forward bias reset to the return of $V_{SUB} = 15$V. The voltage was stepped in increments of $\Delta V_{SUB} = 5$ V. Based on a similar sequence used with H2RG-32-147 to erase latent images in telescope observations, the results appear to be applicable to all HyViSI detectors.

7.4.2.2 Results

Figure 7.15 shows the average dark current measured after step **3** of the experiment for several different values of N_{Resets} and N_{Reads}. The dark current at each time (read) was measured by

differencing the pixel values in successive reads of the detector, averaging a 150×150 box of pixels, and dividing by the time between reads. The main result highlighted by the figure is: **the return of the dark current to equilibrium after forward biasing the diodes and restoring reverse bias is not influenced by the number of resets or number of reads performed**. In addition, the sequence of $N_{Resets} = 1$ and $N_{Reads} = 250$ was performed with and without drop frames, indicating that leaving the detector idle yields the same result.

The average dark current as a function of time, $D_{avg}(t)$ (in e^-/s/pix), is well fit by an exponential decay of the form

$$D_{avg}(t) = D_1 e^{-t/\tau_1} + D_2 e^{-t/\tau_2} + D_{equil}, \tag{7.17}$$

where τ_1 and τ_2 are time constants associated with dark current amplitudes D_1 and D_2, respectively, and D_{equil} is the average dark current after the return to equilibrium. The average value of the coefficients from fits to the five readout sequences are:

$$\begin{aligned} \tau_1 &= 13.1 \text{ s} & \tau_2 &= 91.1 \text{ s} \\ D_1 &= 244.4 \text{ e}^-/\text{s/pix} & D_2 &= 121.17 \text{ e}^-/\text{s/pix} \\ D_{equil} &= 11.5 \text{ e}^-/\text{s/pix} \end{aligned}$$

The fit with these average values is shown in the black curves of Figure 7.15. While the fit describes the data well for $t = 0$ to $t = 400$ s, a very slow decay continues to occur for $t > 400$ s and brings D_{equil} to about 7.5 e^-/s/pix, suggesting a third time constant may be involved. A similar long time constant was observed for the latent images in InSb arrays [42].

The time constants associated with the decay are slightly longer than the time constants associated with the persistence signal. Possible reasons for this will be discussed in Section 7.5. If the source of the two phenomenona are the same, then the total amount of carriers released by a given pixel en route to equilibrium after the bias switch should still be the same as amount of carriers released by that pixel after it has reached saturation in persistence. The total amount of carriers can be calculated by integrating Equation 7.17 from $t = 0$ to $t = \infty$:

$$N_1 = D_1\tau_1 \cong 3200 \text{ e}^- \qquad N_2 = D_2\tau_2 \cong 11,000 \text{ e}^- \tag{7.18}$$

The sum of these two terms is about 14,000 e^-, which is very close to the maximum signal rise observed in a latent image with H1RG-022. The physical explanation of these these time constants and coefficients, as well as the relationship of the dark currents after bias switching and latent images will be discussed further. First, a discussion of effects seen after incrementing V_{SUB} in steps is given to shed light on where in the detector the dark carriers are originating.

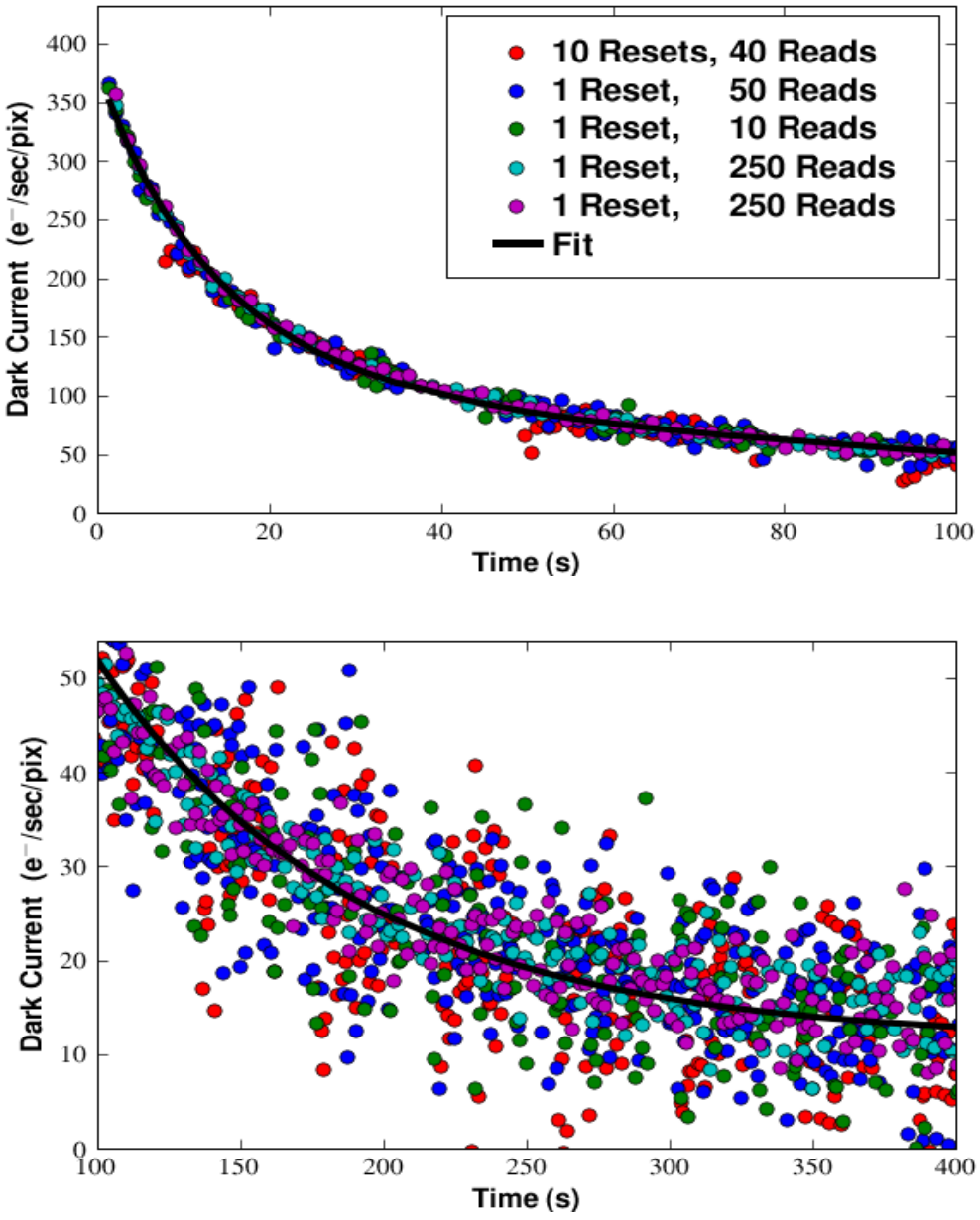


Figure 7.15: Dark current after forward biasing the PIN diodes and then restoring the reverse bias. The number of reads in an exposure and the number of resets between them is indicated in the legend. The outlying points for $N_{Resets} = 10$ and $N_{Reads} = 40$ is caused by a droop in signal that was not fully corrected by reference pixel subtraction. The fit from Equation 7.17 is shown with the black curve.

7.4.3 Incrementing Reverse Bias with $\Delta V_{SUB} = 2$V

A similar type of procedure to the one described in the previous section was carried out at the Independent Detector Testing Laboratory by Don Figer and colleagues with H1RG-018. Instead of switching from forward to reverse bias, the backside voltage was incremented in steps of $\Delta V_{SUB} = 2$ V, starting at $V_{SUB} = 0$ V and going up to 10 V. At each voltage, 20 or more dark exposures with 33 reads and an exposure time of 200 seconds were recorded. The detector was held at a temperature of 160 K and readout in full frame mode. No record of the number of resets between exposures or exact timing between them was available from the FITS headers for this data, so it will only be used to provide a qualitative look at the dark current after incrementing V_{SUB}.

Figure 7.16 shows the average dark current (measured by fitting slopes to the ramps of pixels and taking an average of a 150 × 150 box) as a function of time after each increase in V_{SUB}. It is immediately clear that the time constant associated with the decay to equilibrium for the steps of 0 to 2 V and 2 to 4 V is much greater than it was for the case when the bias was ramped from 0 to 15 V. Further, it is observed that nearly all the decay occurs between 0 and 4 V, with only a very small change in dark current for the increments in the 4 to 10 V range.

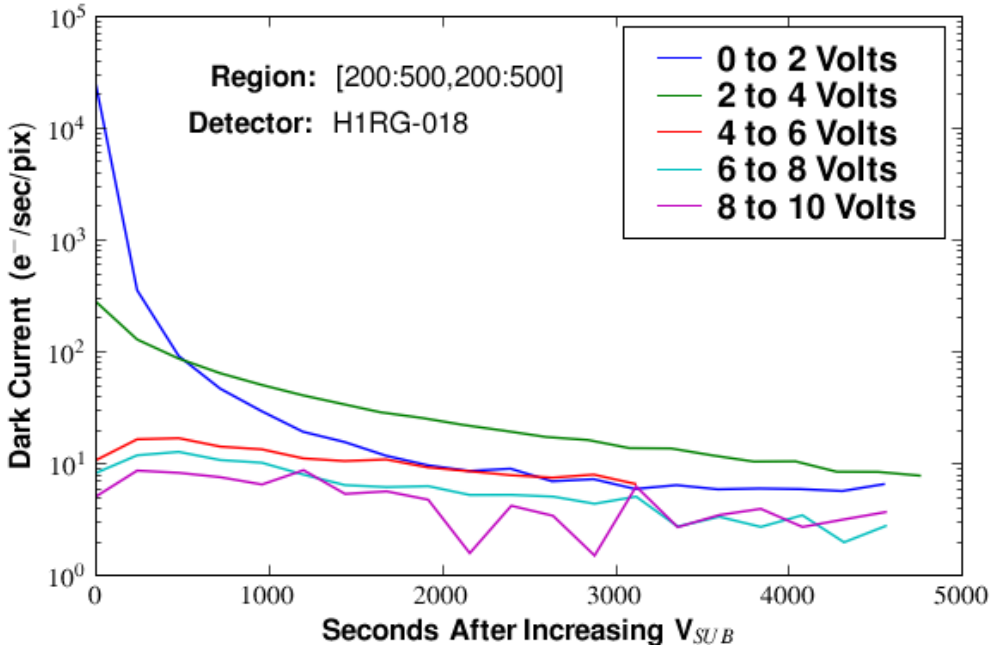


Figure 7.16: The mean dark current in 33 read/200 second integrations after increasing V_{SUB} in 2 volt increments. The dark current is estimated in each integration by fitting a slope to the pixels in a subregion of the array and taking the mean.

The dependence of the dark current on the change in V_{SUB} is evidence that the dark carrier generation sites are located near the front surface. If they were located uniformly throughout the bulk along the z-axis (from the frontside to backside), then when the diodes are not fully depleted, increasing the reverse bias should sweep out a new set of carriers from the layer of material that was previously field free, which would result in a large dark signal. The results of the numerical simulations shown in Figure 2.6 indicate the change of the width of this field free layer goes from 24 μm for $0 \rightarrow 2$ volts to 20 μm for $2 \rightarrow 4$ volts and continues to get smaller as V_{SUB} is incremented higher, reaching about 10 μm for $8 \rightarrow 10$ volts. If the dark carrier generation sites were indeed located uniformly throughout the diode, one would expect to see a difference in the dark rates when the bias voltage is switched from $4 \rightarrow 6$, $6 \rightarrow 8$, and $8 \rightarrow 10$ volts. One can see from Figure 7.16 that this is not so. Thus, we conclude that the dominant dark generation sites after the switch from forward to reverse bias, which are also believed to be the hole storage sites that participate in persistence, are concentrated within 20-40 μm of the front surface.

Another important consideration is that the metallurgical junction is a two dimensional surface. When V_{SUB} is raised from $0 \rightarrow 2$ V, the volume of bulk material in between the p^+ implants, above the front surface is also being swept of carriers. Stored holes that cross the junction into the p^+ implant will cause a large dark current. The one-dimensional numerical simulation does not treat this region, and the two dimensional geometry of the junction is not known. It is very likely that a large contribution to the dark rate arises in this region due to trapped minority carriers being swept out when the reverse bias is applied. When V_{SUB} is increased further, this region is already near or at full depletion, depending on the potential above the surface passivation. If it is fully depleted, it will not contribute to the dark rate. This may explain why the change in dark rate is so large when V_{SUB} is increased in the 0-4 V range, and near zero when increased in the 4-10 V range.

7.5 Model, Simulations, and Summary

Even with the many tests performed on the 100 μm thick HyViSI detectors and the abundance of data exhibiting the core-halo persistence, a clear cut model for the persistence remains elusive. To truly pin down the physics behind this phenomenon, we require a more detailed knowledge of the detector layout and the ability to "turn more knobs". For instance, changing the thickness of the detector and the dimensions of the p^+ implant and gradient in the doping density at its metallurgical junction (neither of which are known because of proprietary reasons), using a metal grid under the front surface to adjust the surface potential above the SiO_2 layer, etc., and observing the effect on the latent images would reveal a great deal about how and where in the detector they arise.

Despite the lack of information and knobs, based upon what is known of the detector layout and the empirical evidence presented in the preceding sections, we can present a sound, qualitative hypothesis that explains the key features. Some aspects of the model have been touched on in the previous sections. What follows is a comprehensive description that aims to address all of the components of the core-halo persistence.

7.5.1 Phenomenological Description

As alluded to previously, the similarity of the persistent signal induced by illumination and that brought about by forward biasing the photodiodes suggests that excess minority carriers stored in the depletion region are responsible for the latent images and persistence. In p^+-n junctions used for switching applications, it is assumed that the stored minority carrier charge is dominated by holes in the bulk n-region closest to the junction when the bias direction is switched [122] from forward to reverse. We will assume the same thing here, and our assumption is supported by the fact that the persistence in HyViSIs occurs at low values of V_{SUB} when the depletion region in the PIN diodes is confined to the area near the metallurgical p^+-n junction. On the slightly doped n-side of the junction (the I region), minority carrier holes are trapped near this junction when the diode is forward biased or when it is illuminated. In the former case, the holes diffuse to to the region from the p^+ implant and have a density profile that exponentially decays with distance into the n type bulk. In the latter, a fraction of the photo-generated holes are trapped there either before or after they drift under the influence of the electric field with a similar profile.

The role that the stored holes play in a subsequent, reverse-biased exposure of the detector is somewhat indirect. Because a large reverse current cannot flow across all depleted portions of the PIN diode, they are not immediately swept out of the diode during the reset. Instead, they add positive charge to the charge density profile, which in turn requires an additional amount of electrons to be injected into the p^+ side of the junction to restore the voltage in the implant to V_{RESET}. The situation is illustrated in Figure 7.17 and explained in more detail in the figure caption. Using Gauss's Law and integrating the electric field for a given charge distribution shows that the

number of electrons needed to compensate the trapped holes far exceeds the number of trapped holes themselves. However, the exact number will depend on how the trapped holes are distributed near the junction and how the injected electrons fill the p^+ implant during reset. Since the doping profiles of the PIN diodes is not known, neither of these two distributions can be calculated. As seen in Figure 7.17, it is assumed that the trapped holes and injected electrons are exponentially distributed around the metallurgical junction.

After the reset switch is closed, as illustrated in Figure 7.18, the effect of the injected electrons is two-fold. First, electrons that diffuse to the sites of the trapped holes and recombine with them will cause a rise in signal. This occurs because of the fact that the positive charge from the trapped hole is effectively removed from the charge density after the electron recombines with it, which lowers the

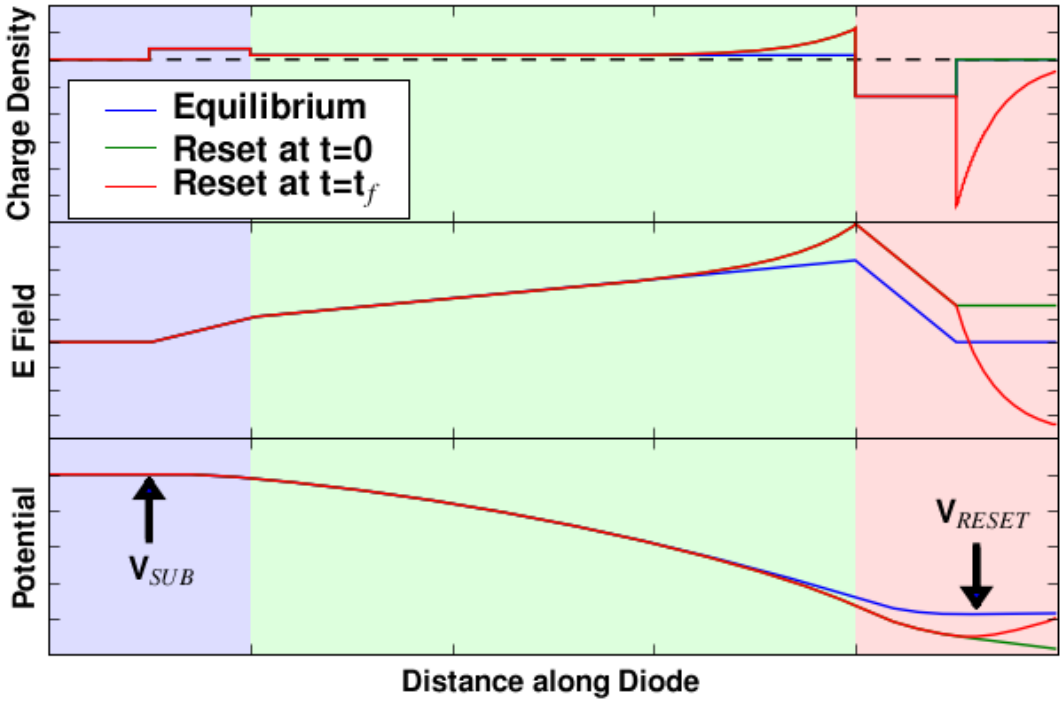


Figure 7.17: A toy model showing the charge density, electric field, and electric potential in a PIN diode in various circumstances. The electric field is obtained using Poisson's equation and integrating the field yields electric potential. The p^+ region is shaded in light red, the n bulk in green, and the n^+ region in blue. The diode in equilibrium, i.e. after a long series of resets in the dark, is shown by the blue curves. After illumination or forward bias, there is an excess of holes stored near the p^+-n junction. At reset, if they were not compensated by injected electrons, the green curves would show the three distributions (the green curves are hidden in some regions because they lie beneath the red curves). With the injected electrons from the reset well, the potential at the p^+ implant is brought to V_{RESET} and the distributions look like the ones shown by the red curves.

magnitude of electric field and decreases the potential across the PIN diode. This process, indicated by $U_n = U_p$ in Figure 7.18, increases V_{node} and hence the signal measured, i.e. $\Delta S > 0$. Second, the drift and diffusion of the electrons itself can cause both an increase or decrease in signal, depending on the relationship between the direction of their motion and the direction of the electric field. This follows from the fact that the work done on an electron moving through an electric field (by the field) is the line integral of the electric field over the path taken by the electron. If the field does positive work on the electron the voltage across the diode will decrease and if it does negative work then the voltage across the diode will increase.[7] The field around the metallurgical junction, an approximation of which can be viewed in Figures 2.6 and 2.7, possesses a large amplitude and is directed to do positive work by pulling holes toward the p^+ implant and electrons away from it. If this field is preserved and an electron diffuses or drifts from the p^+ implant to the n type bulk, then $\Delta S > 0$. For an electron diffusing in the opposing direction, $\Delta S < 0$.

This bi-modality of signal change is believed to be the source of the core-halo structure. Free electrons—sourced by the reset and accumulated near the front surface in between the p^+ implants—are transported between pixels via the currents $J^{n,x}_{DRFT}$ and $J^{n,x}_{DIFF}$. Pixels that see a decrease in the number of electrons over time in the collecting node will have $\Delta S > 0$ and those that see an increase will have $\Delta S < 0$. The exact rates of signal change will be determined by the drift, diffusion, and recombination as well as the intricate interactions of the electrons near the Si-SiO_2 interface and at the p^+-n junctions. However, it will now be shown that one can produce many of the key features of the core-halo persistence with the first three terms alone.

7.5.2 Persistence Simulations

According to the model just described the dominant interactions in the core-halo persistence, which are illustrated in Figure 7.18, occur at or near the front surface of the detector and are governed by the following equations

$$\frac{\partial n}{\partial t} = \nabla \cdot (\mathbf{J}_{Diff} + \mathbf{J}_{Drft}) - U_n \tag{7.19}$$

$$\frac{\partial p}{\partial t} = -U_p. \tag{7.20}$$

No transport currents exist for the holes because they are confined to deep level trapping sites. If the doping densities of the silicon and geometries of the pixel were known, these equations could be solved on a two or three dimensional grid by choosing an appropriate set of initial conditions

[7] An easy way to see this without working through the math is to consider two relevant cases. In one, the electric field in the depletion region of the reverse biased PIN diode does *positive* work on an electron-hole pair generated by a photon, and this leads to a *decrease* in the voltage across the diode. In the other, the built in voltage across a PN junction is established by the electrons and holes diffusing against the electric field arising from the ionized donors left behind. The electric field does *negative* work on the diffusing carriers and this corresponds to an *increase* in the voltage across the junction.

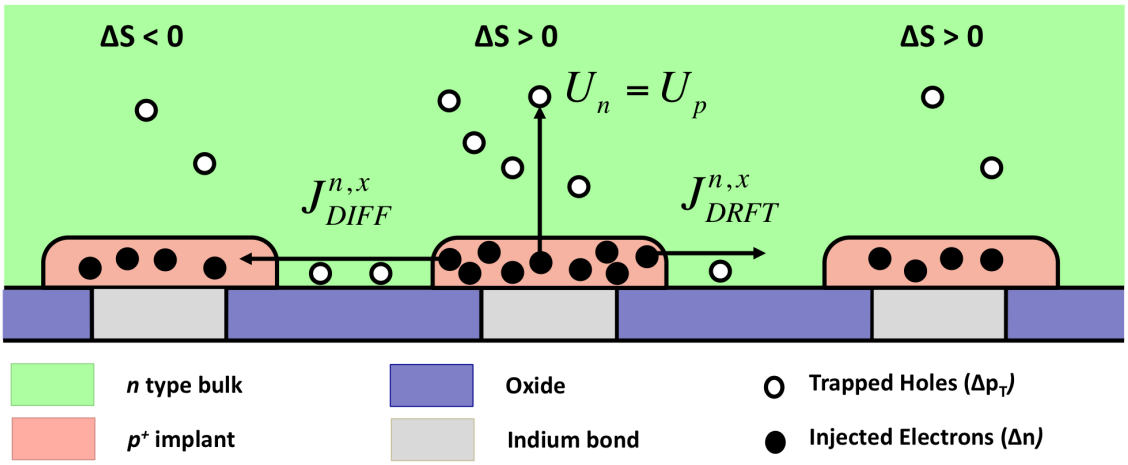


Figure 7.18: A diagram showing the diffusion and drift currents and recombination believed to be responsible for the Core-Halo persistence. The trapped holes require a surplus of injected electrons at reset to restore the pixel potential to V_{RESET}. Some of the injected electrons drift or diffuse to the trapping sites and recombine with the holes at the rate $U_n = U_p$, which causes an increase in the pixel signal, $\Delta S > 0$. Other electrons drift or diffuse away from the pixel with a current of J^x_{DRFT} or J^x_{DIFF}, respectively. Pixels that lose the electrons see a signal gain, $\Delta S > 0$, and those that receive them see a decrease, $\Delta S < 0$. Note that the plane 1/2 of a pixel length behind the plane of the drawing would have only bulk n type material, making it an avenue for the electrons to travel through if the surface is accumulated.

and using a finite difference scheme like the one outlined in Appendix C.3. Since these specifics are unknown, a more simplified approach shall be taken to model the system.

To start with, the two-dimensional doping pattern of the detector created by the p^+ implants is ignored and the detector is treated as being uniformly doped. This treatment is justified if the electron transport occurs primarily in the n type bulk in the "alleys" between implants. Then, assuming the core-halo persistence occurs in a sufficiently thin layer near the front surface of the detector ($z = 0$) and utilizing rotational symmetry around the core center, the effect can be simulated using a quasi-one-dimensional computational scheme. In this scheme, the computational details of which are outlined in Appendix C.5, an initial distribution of excess holes, $p(r, t = 0)$, and electrons, $n(r, t = 0)$—both of which are assumed to be present after the detector is illuminated and the pixels are reset—are evolved in time in cylindrical coordinates. For simplicity, the ionized donors and equilibrium carrier distributions present in the silicon are ignored and attention is paid only to the trapped holes and injected electrons. The vertical component of electric field is also neglected so that only the radial component, which is assumed to arise solely from $p(r, t)$ and $n(r, t)$, affects the evolution of the system. While the explicit currents and field are confined to the $z = 0$ plane, the simulation effectively allows for the loss of carriers in the transverse direction through the terms U_n

and U_p, making it quasi-1d (this is explained in more detail in Appendix C.5).

After establishing the mesh for the computation and the current transport equations, the only things that remain are to determine the recombination rate and the initial conditions. In this model, the trapped holes are allowed only to recombine with the injected electrons so that the recombination rates are necessarily the same $U_n = U_p$. A very simplified form of the Shockley-Read-Hall Equation [112] is used to describe the rate of recombination:

$$U_n = U_p = -\frac{pn}{\tau_p(n+p)}, \tag{7.21}$$

where $\tau_p = \tau_n$ is meant to represent the lifetime of the free electrons when present amongst the trapped holes. The initial distribution of trapped holes $p(r, t = 0)$ is based directly on the distribution of collected charge generated by the illumination source before reset, i.e. the signal measured in the previous exposure. In the case where the detector is undepleted, a piecewise function like the one in Equation 2.14 is appropriate. For the overdepleted case, the following form is used:

$$p(r, t=0) = \begin{cases} Q_{FW} + Q_{SAT} * \exp^{-r^2/(2\sigma_{sat}^2)} & 0 \quad < r < r_{sat} \quad : \quad \text{Saturated by Illumination} \\ Q_{FW} * \exp^{-(r-r_{sat})^2/(2\sigma_{psf}^2)} & r_{sat} < r < r_{max} : \quad \text{Not Saturated} \end{cases} \tag{7.22}$$

Based upon the argument presented in the previous section, the electron distribution will have the same functional form, but the number of electrons in a given pixel will exceed the holes by some amount: $n(r, t = 0) = \gamma p(r, t = 0)$. γ is in general a nonlinear function that depends on the number of holes as well as their true arrangement along the z axis. Solving for it would involve not only a knowledge of this arrangement, but an iterative numerical procedure to solve for the electron distribution that sets $V_{node} = V_{RESET}$. Here it will be treated as a constant: $\gamma \sim 4$.

The simulation itself only provides the number of electrons and holes at a given location and time. The key assumption made in comparing the simulation to real data is that the change in $n(r, t)$ for the pixels is directly related to the change in signal, ΔS_{ADU}:

$$\Delta S_{ADU}(r, t) = G_{NET} \Delta n(r, t) = G_{NET} \cdot [n(r, t) - n(r, t = 0)] \,. \tag{7.23}$$

G_{NET} is the conversion gain between e^- and ADU.

7.5.2.1 Simulation Results

The simulation that will be described here was intended to mimic core-halo persistence induced by the star HD53791 during a 30 second g band exposure, in which the star saturated out to a radius of $r_{sat} = 6$ pixels. Following the exposure, the detector was blanked off and a 200 second, 500 read dark exposure was recorded in window mode with a frame time of $t_{frame} = 0.46$ s. The illuminated exposure was used to deduce the initial conditions. The parameter Q_{FW} was computed using the

Table 7.1: Parameters for the core-halo simulation shown in Figure 7.19.

Q_{FW} (e^-)	5560	μ_n ($cm^2/s/V$)	$3.24 \cdot 10^{-7}$
Q_{SAT} (e^-)	585	D_n (cm^2/s)	$1.62 \cdot 10^{-10}$
σ_{sat} (pix)	2.15	τ_p (sec)	50
σ_{psf} (pix)	3.70	r_{sat} (pix)	6

approximate saturation of the center pixel and the relationship of 0.027 latent e^-/e^- noted in Section 7.2.2.2, and σ_{psf} was based on the FWHM of the stars in the saturating exposure. τ_p was set at 50 seconds based upon the longest time constant, τ_2, observed in the experiment described in Section 7.2.2.3 under the assumption that this is the time constant associated with the recombination of trapped holes and injected electrons. The other values are based largely on trial and error until a suitable match between the simulation and the real data was found.

Figure 7.19 shows the results for a simulation of the latent image in the dark exposure following the illumination; the initial conditions and other relevant variables are listed in Table 7.1. The radial profile taken at $t = 200$s for the simulated data shows good quantitative agreement with an

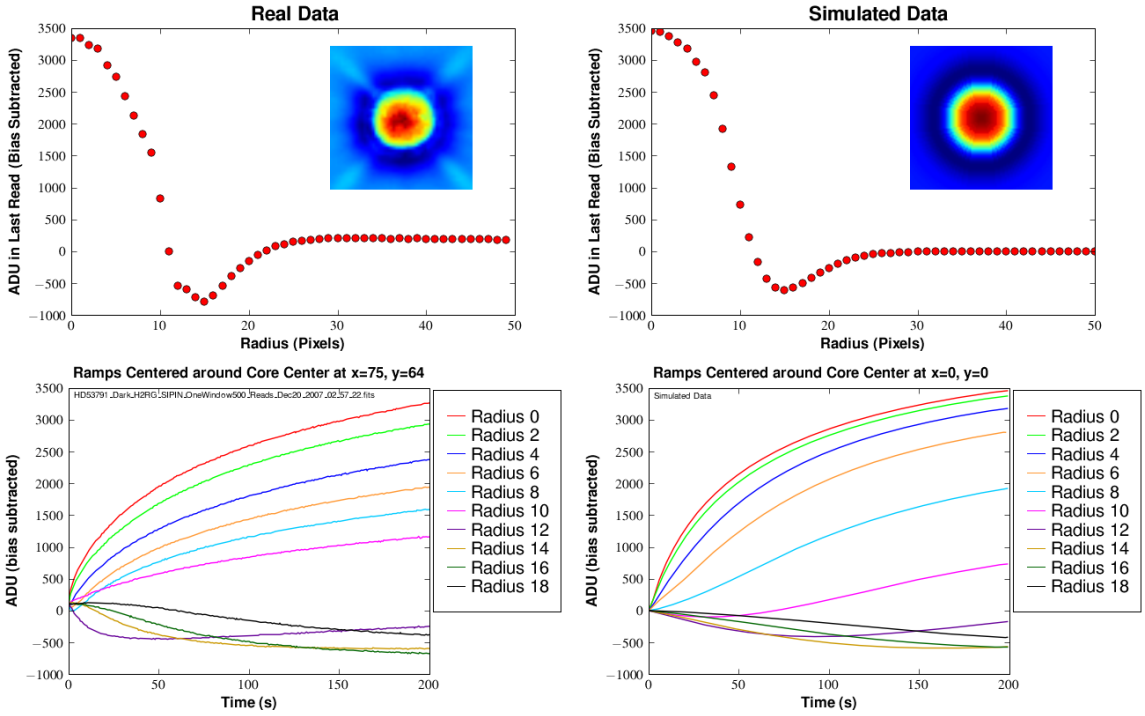


Figure 7.19: A comparison of persistence observed after a long exposure and simulated persistence.

averaged profile from the real data. The average for the real data was taken across several rows in order to avoid the non-radially symmetric diffraction spikes that show up in the latent image. One can still see some oddity in the real data near the halo minimum—it is not smoothly varying—that has to do with the diffraction pattern and the telescope tracking error. The latter can be observed in the sub-images included with the plots. The pixel ramps for the simulated data show good qualitative agreement with the real data, especially for the pixels near the core center and at the outer edge of the halo, but there are some large discrepancies for the pixels between $R_{core} = 9$ and $R_{min} = 15$. In general, the real data shows that the halo pixels closest to the latent image center see a very rapid drop in counts. In fact, Figure 7.8 shows that the time constants associated with the decay of the halo are very close to those of the rise of the core. However, this is not reproduced in the simulation, where the halo pixels have a much lower rate for the decay (this can be seen most easily by comparing the purple and pink curves). The junction fields or the two-dimensional structure of p^+-n implants—the diffusion and drift is being modeled as occurring only in the n type bulk, but neighboring pixels can actually be thought of as a p channel transistor where rapid carrier transfer may take place—may be responsible for the rapid decay of the halo minimum. Despite these shortcomings, the overall similarities are remarkable considering this is a simple quasi-1d simulation.

Another noteworthy point is that the diffusion coefficient and mobility obtained from the simulation are extremely low for electrons in silicon. Typical values for the mobility of electrons in silicon are $\mu_n = 10^2 = 10^4$ cm^2/V/s, which is more than 10 orders of magnitude larger than the value obtained in the simulation. Small diffusion coefficients on the order of 10^{-10} cm^2/V/s are more typical for ions in SiO_2 [123], but the presence of ions in the HyViSI SiO_2 layer is extremely unlikely (this can be said for any device fabricated with modern semiconductor processing techniques in a clean environment). Further, the value of D_n from the simulation is about 10^3 times as large as μ_n, which conflicts with the Einstein relation $D_n/\mu_n = kT/q$. Since the data was taken at 160 K, $kT/q \sim 0.014V$. One might therefore guess that the electric field is negligible in the simulation, but as Figure 7.20 shows, this is not the case. The final electron and hole distributions are very different depending on whether or not drift is included. The primary difference is that with drift included, the mutual repulsion of the electrons drives them away from the core at a much faster rate than with diffusion alone. This steepens the profile between R_{core} and R_{min} and creates a closer match to the data. It also aids in decreasing the decay time of the halo pixels, bringing it closer to the rise time for the halo pixels. And for long simulation times, the field actually causes some of the electrons that left the core to return there. This behavior is essential to account for the decrease in core signal for long integration times, as exemplified by Figure 7.4.

Another important point made by Figure 7.20 is that after 200 seconds, a large number of trapped holes still remain. If the detector is again reset, this distribution of holes will cause another population of electrons to be injected into the core pixels. The holes would thus serve to determine the initial conditions for a simulation of an exposure following the reset. In simulations where t_{final}

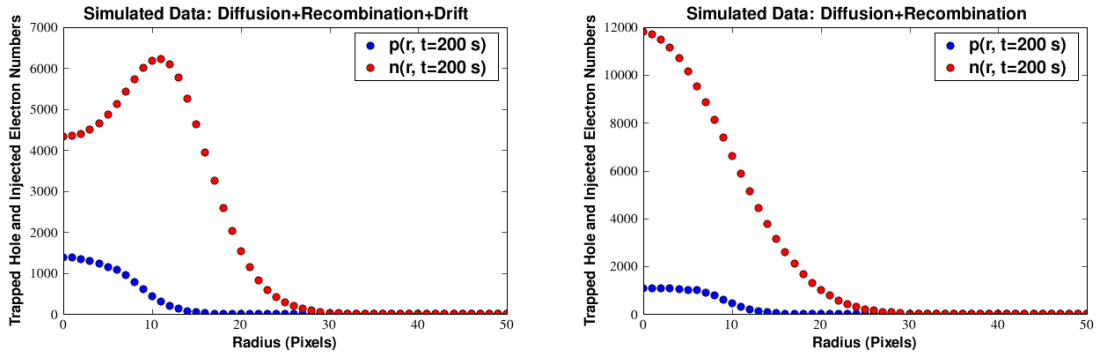


Figure 7.20: A comparison of the simulated hole and electron distributions with and without drifting the electrons. The field that creates the drift is due to the injected electrons and trapped holes.

is on the order of 10 minutes, most of the electrons migrate to the outer radii, creating a scenario where the number of holes in the core pixels exceeds the number of electrons. Since the number of electrons must be at least as great as the number of trapped holes to remove the trapped holes through recombination, this scenario explains how the persistence can last for hours.

Although not shown here, with the appropriate initial conditions, the simulations can also account for the double core-halo structure when the detector is not fully depleted. In both the over-depleted and under-depleted case, the simulations are successful in reproducing the transition of a pixel far away from the stimulated region from integrating dark current to suddenly seeing a signal change in the opposite direction. It has been noted that the dark current volcanoes also show a halo of pixels that decrease in signal, so one might hypothesize that this model could also be applied to them. However, the transitional behavior just described does not occur for the pixels far away from the center of the volcanoes, suggesting a different mechanism behind them.

7.5.3 Summary

To summarize, a model has been proposed to explain latent images in the HyViSI detectors, particularly the core-halo structure. The model states that a fraction of holes generated from illumination, electrical stimulus, or cosmic ray events are caught in deep level traps somewhere near the front surface of the detector. At reset, the presence of these trapped holes forces a large number of electrons to be injected by the reset transistor in order to bring the voltage at the integrating node to the reset voltage. These injected electrons either recombine with the trapped holes or migrate away from the collecting node. Pixels that see a net increase in the number of electrons at the collecting node decrease in voltage and vice versa. The persistence subsides after the fraction of traps that are populated by holes is restored to its equilibrium value.

This model accounts for many of the traits exhibited by the persistence and latent images. The

most important are:

1) **The core-halo structure itself.** Emission of holes from traps in the near vicinity of the p^+ implant alone cannot account for the negative signal in the halo. Charge must be transferred between pixels. The drift and diffusion of the injected electrons explains both the positive signal in the core and the negative signal in the halo.

2) **Persistence at low V_{SUB}.** The model states that the stored charge after a stimulus is located near the metallurgical p^+-n junction at the front side of the detector. Hence, the layer of undepleted bulk for values of $V_{SUB} <$ 7-8 V does not prevent the persistence from occurring. It only changes the behavior by allowing a wide diffusion of holes to occur during the stimulus and create stored charge at larger radii.

3) **The double core-halo structure at low V_{SUB}.** When the detector is not fully depleted and the pixels are saturated by light, diffusion in the undepleted regions spreads the holes out to far radii and they fill trapping sites along the way. The existence of two strong gradients in the trapped hole concentration—one at the photocurrent source and one near the edge of the diffusion envelope—results in a double core-halo structure in the latent image.

4) **The temperature dependence of the halo width.** To produce a halo, the model requires that electrons move along the front surface. The diffusion and drift of electrons along this surface requires that it is in an accumulated state. At a fixed surface potential, the induced charge density depends very strongly on temperature through the product of the Debye length, $L_D \propto T^{1/2}$, and the intrinsic carrier concentration, $n_i \propto T^{3/2} \exp(-E_g/2kT)$, where E_g is the energy gap [124]. For low temperatures ($T \sim 110K$) the persistence is dominated by recombination of the injected electrons with the trapped holes and very little transport along the front surface takes place, so the halo has a very small extent. As the temperature increases, the accumulation along the surface allows for easier transport of the injected electrons and the halo width grows accordingly.

5) **Persistence without saturation of pixels.** According to the model, stored charge will be present after stimulus regardless of whether or not the pixels are saturated. For signal in the range $0 - FW$, the amount of stored charge should be proportional to the amount of holes reaching the front side of the detector. This supports the linear relationship between the integrated photocurrent in the stimulus exposure and the latent image maximum in the one following.

6) **Saturation of persistence signal.** The nonlinearity and eventual saturation of the latent image maximum as a function of integrated photocurrent in a previous exposure occurs because the number of available trapping sites decreases as the photoholes accumulate. While a more sophisticated model would be required to explain the manner in which these sites

are occupied—and where exactly they are located for that matter—this model requires the persistence to saturate at some level since the number of traps is finite.

7) **Sensitivity of Core-Halo Pixels.** The name "desensitization" has been used to describe this phenomenon under the assumption that the pixels suffered a loss in sensitivity. However, the empirical evidence shows that both the core and halo pixels in a latent image still integrate photocurrent (if starlight is shining on them, for instance). The charge transport of injected electrons in the model does not preclude the pixels in the latent image halo from integrating signal charge, so they are not desensitized.

8) **Eventual decay of core pixels in long time exposures.** A large fraction of stored holes may be left behind in the core pixels after the injected electrons have migrated to the halo. The electric field created by these holes eventually pulls the electrons back into the core. Those that diffuse back into the p^+ implants in the core cause a decrease in signal.

9) **Reappearance of latent images after long time exposures.** The left-behind holes can survive for times much greater than τ_p if there are no electrons with which to recombine. If stored holes still exist at the core when the pixels are reset, a new swarm of electrons will be injected, giving rise to another latent image. The model can thus account for the disappearance of persistence in 200-300 seconds when regular resets are applied and the reappearance of persistence after 1 hour or more when only one or two resets are applied.

10) **Similarity between illumination and electrical stimulus.** Both illumination and a forward bias on the diodes creates an increase in the number of minority carrier holes in the n type bulk right around the p^+-n junction. Thus, according to the model, both result in occupancy of the deep level traps. With illumination, gradients in the charge distribution result because of non-uniformity of the photon flux, and these gradients create a core-halo structure. With a forward bias all pixels are uniformly affected (ignoring pixel to pixel non-uniformity) and so no core-halo structure is formed. Instead, the persistence is dominated by the recombination of the injected electrons with the stored holes so that all pixels see an increase in signal. And since the electrons do not migrate horizontally, they stay with the trapped holes and recombine. This means that given sufficient time, the persistence will disappear independent of the number of resets performed. With the same line of reasoning, the model predicts that the time constants for the rise of the core pixel signals after illumination will be shorter than those for the rise of the pixel signals after electric stimulus.

11) **Persistence is not detected in reset frames.** With the reset transistor actively delivering electrons to compensate for the stored charge, the measured voltage at the integrating node stays at V_{RESET}.

7.6 Dealing with Persistence

Smith et al. present a number of possible ways to lower persistence in HgCdTe arrays through alterations in design or by using physical techniques such as waiting for the fast time constant portion of the decay to take place before beginning another exposure [116]. They also present methods to calibrate the persistence so that it can be accounted for and removed if present in science data [125]. Some of these strategies are applicable to SiPIN diodes, but others are not due to the fact that HyViSI detectors are operated in full depletion while the HgCdTe are per-pixel depleted detectors. The disparity in the depletion state of the two types of detector creates some fundamental differences, most notably with the time constants involved–the decay time constants in HgCdTe detectors depend on stimulus while they appear to be independent of stimulus at a given temperature in HyViSI detectors–and the lack of a core-halo structure in the HgCdTe detectors.

As discussed, the core-halo latent images in HyViSI detectors are only well fit with Equation 7.13 for a very narrow range of circumstances. And even with the simulations described in Section 7.5.2, there are significant residuals after subtracting the real data from the simulated data. Calibrating and removing the persistence from science data is thus not likely to be fruitful for these detectors. Instead, the course should be to taken to prevent the persistence from occurring in the detector through design improvements or minimizing it with operational strategy.

7.6.1 Persistence Reduction by Design

It is very difficult to suggest design improvements without knowing the design itself. Such information is considered proprietary to Teledyne Imaging Sensors. But with or without an intimate knowledge of the architecture, what follows are some possible approaches to take.

7.6.1.1 Frontside Passivation Treatments

The theory that the persistent charge is due to stored minority holes in the vicinity of the p^+-n junction would imply that attention should be paid to this region. It is likely that the majority of the trapping sites are located at the interface between the SiO_2 passivation layer and the Si bulk since these interfaces are notorious for having large trap densities. Some manufacturers aim to reduce the number of these trapping sites by treating the surface with special techniques during fabrication. In fact, at least three of the devices tested during the course of this thesis work were made with dissimilar surface treatments (Bai, private communication). Since the persistence was equally strong in all of them, alleviating the problem in this way may not be a viable option.

7.6.1.2 Controllable Potential Metal Grid

The use of a metal grid below the surface passivation (see Figure 1.7) that can be held at a controllable potential is one option to consider. Raytheon Vision Systems uses this approach [38], which

was successful in drastically reducing persistence in InSb arrays [42]. Building a HyViSI with the metal grid would, at the very least, allow one to confirm whether or not the stored charge is located near the front surface passivation. In the best case scenario, placing the surface at a higher potential than the p^+ implants could be used to effectively create true potential wells and prevent holes from reaching the trapping sites altogether. The downside of this grid is that it increases the capacitance of the pixels, thereby reducing the sensitivity, and may result in yield issues.

7.6.1.3 Anti-blooming Diode

The first generation H4RG had an anti-blooming diode in the multiplexer pixel. As evidenced by the data collected with H4RG-10-007, this diode prevented column bleeding, output crosstalk, and blooming. It did not eliminate persistence, but set a threshold on the maximum observed persistence signal that was well below the other HyViSI detectors. This is in agreement with the model presented since the number of holes available for trapping does not increase after the blooming diode starts to bleed them off. The problem with the anti-blooming diode implemented is that it creates unacceptable leakage currents, which led TIS to remove it from their next generation H4RG [88]. Because of the benefits it presents, though, future consideration should be given to reimplementing the anti-blooming diode.

7.6.1.4 Capacitive TransImpedance Amplifier (CTIA)

In the SFD architecture, photo-charge accumulates in the collecting p^+ implant, which allows charge to be trapped in the vicinity of the p^+-n junction as it builds up. In the CTIA architecture, the p^+ side of the diode is connected to the integrating node of a feedback circuit so that the dominant accumulation takes place on a feedback capacitor in the ROIC. In theory, this should all but eliminate persistence since the ability of the minority carriers to accumulate around the front side of the detector is significantly hindered. However, the CTIA has many disadvantages for astronomy when compared to the SFD. It consumes more power, has a higher readout noise, and requires a large pixel footprint in the multiplexer, which limits the possibility of reducing pixel size [116, 23].

7.6.2 Reduction After Design

So far, no operational techniques have been successful at eliminating persistence in the sensors where it has been observed. With the current design of the detector it appears that the best one can do is minimize its strength and duration using the following strategies.

7.6.2.1 Short Exposure Times

Because the latent image maximum increases with increasing fluence, the time taken for the persistence signal to reach the read noise level does as well. Short exposure times will limit the fluence

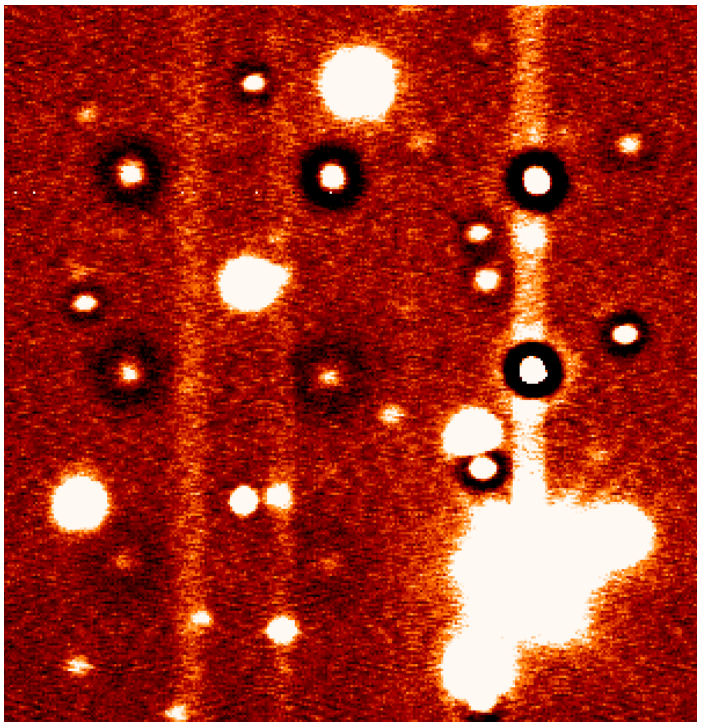


Figure 7.21: An example of a poorly chosen 3×3 dither sequence that shows latent images in all of the previous 8 dither locations. The exposure time was 21s and the time between exposures was 16s so the dither period was $T_{Dither} = 333$s. This is shorter than the time it takes for the persistence for the brightest star in the image to fall below the read noise level, so the latent image from the first dither is still detectable at about the 1σ level. When the dithers are combined, the core-halo images will show up in several locations (see the mosaic in Figure 6.20). An interesting overlap between the light from a star and the halo of a latent image is also shown.

seen by the pixels so that the latent images fall below the read noise more quickly. The latent images will still decay with the same time constant, of course, so care must be taken when planning a dither sequence. If the ratio of the time period for the dither sequence, T_{Dither}, to the decay time of the persistence is too short, the result will look like Figure 7.21, in which latent images show up at all of the previous dither locations. In this instance, when combining the dithered exposures, the mean or median of the nine pixel values used to estimate the flux at a given location on the sky will be thrown off by up to 4 values tainted by persistence. One solution is to simply do more dither locations with the same exposure time, thereby increasing the number of samples used to estimate the flux. Another solution is to use a slightly longer exposure time. For instance, if the exposure time in the 3×3 sequence in Figure 7.21 was increased to 120 s, at least four out of the nine locations would not have latent images above the read noise level. It is always possible to mask the core-halo structures in software when combining the dithers, but this will result in increased noise for certain spatial locations.

7.6.2.2 Frequent Resets

It may seem that frequent resets and short exposures imply the same thing. This is true if the detector is operated in full frame mode since two separate exposures means a reset of all pixels between the two. However, if the detector is operated in window mode, a reset sequence only resets

the pixels within the window; the other pixels on the detector continue to integrate charge.

For window mode applications such as telescope guiding, whenever window data is not required, the detector should be switched to full frame mode so that all the pixels on the array can be reset. The reason for this is that if intermittent full frame resets are not performed when the detector is operated in window mode, bright stars may saturate pixels outside the window and generate persistence. This persistence will cause havoc when the detector is later read in full frame mode or a new guide window is selected that happens to fall on a previously saturated set of pixels. The full array may be reset line by line to ensure stability in signal after the reset, but this comes at the expense of a large reset time for the full frame. The reset time can be reduced dramatically if the detector is switched to full frame, global reset mode and then a global reset is performed (the detector manual states this reset should be greater than 10 μs in order to fully reset the pixels and empirical evidence shows it should be less than $\sim$ 100 ms to prevent strong nonlinearities). However, after the global reset, when the detector is switched back to window mode, a line by line reset of the window should be employed to prevent nonlinearities in the early reads of the window.

The empirical data shows that a reset should be performed at least every 120-200 seconds at T=160 K to clear the trapped charge in the shortest possible time. With the right initial conditions, simulations show that after this time period, trapped holes still exist near the core with no electrons available for recombination, so another reset must be applied to provide them. When resets are only performed every 20-30 minutes, the persistence can endure for 1-2 hours. So if long exposures are a must, 300-400 second long reset periods should be issued between them to prevent the latent images from showing up in each of the long exposures.

7.6.2.3 Lower Temperature and Higher V_{SUB}

Persistence is still observed at the lowest temperature, T=100K, tested for this thesis work and the highest backside voltage, $V_{SUB} = 40$V. But because the halo width decreases when the temperature is lowered and the PSF is decreased when V_{SUB} is raised, both of these changes result in a smaller number of pixels affected by the latent images.

7.6.2.4 Reset of Saturated Regions with Guide Windows

One possible way to prevent the buildup of stored charge via large fluences is to reset saturated regions during long exposures using the guide mode of the HxRG multiplexers. This requires either prior knowledge of the relevant pixel locations or "smart firmware" that automatically detects saturated regions and programs the HxRG serial register with the appropriate window coordinates and then resets these regions.[8] Complex clocking patterns and instruction sets would be thus be involved. The signal change of the rows and the columns of the window, which were discussed in

[8]Detection of the saturated regions could also potentially be done on the DAQ, but a dedicated set of instructions in the control electronics microcontroller or FPGA would not impose on the flow of science data between the two systems.

Section 6.1.4.3 and are shown most clearly in Figure 5.11, would create an additional complication when the data is reduced.

Appendix A

Data Reduction

A.1 IRAF Parameters

The parameters we used in **DAOFIND** to reduce our H4RG-10-007 photometric data are chosen according to the recipe laid out in Davis [93]. In particular, for a given sky value, s (in ADU), number of photons per ADU, p, and read noise, r (in e^-), the expected 1σ variance in the sky will be

$$(\sqrt{s \times p + r^2})/p \tag{A.1}$$

For our images, with $s = 2.5$, $p = 1$, and $r = 0.3$ for the combinations of dithers, we have $1\sigma = \sqrt{2.5 \times 1 + 0.3^2} = 1.609$.

Most of the parameters were kept at default. We adjusted **fwhmpsf** according to the seeing for each night. It was typically between 11 pixels and 14 pixels (larger in g than in i and y) , corresponding to the 1.375-1.75 arcsecond seeing at the site. Following Davis [93], we set **psfrad** = 4.5∗**fwhmpsf** and **fitrad** = **fwhmpsf**. We also adjusted **sigma** according to the number of dithers used to form the final image and the gain of the preamplifiers.

The parameter to which the finding algorithm was most sensitive was **threshold**. Several "eyeball" tests for each image were performed to determine a reasonable value for **threhsold**. Fortunately, doing a few iterations of detection, psf fitting, and subtraction eliminated the need to find a perfect value for this parameter.

Appendix B

Signal to Noise Ratio

When observing a celestial source with an apparent magnitude M_s and a background sky flux B in $e^-/s/pix$, we can calculate the Signal to Noise Ratio (S/N) for a given exposure as

$$S/N = \frac{I_o A \eta T t * 10^{-(M_s - M_o)/2.5}}{\sqrt{\pi r_{source}^2 (Bt + Dt + R_n/\sqrt{N}) + I_o A \eta T t * 10^{-(M_s - M_o)/2.5}}}, \tag{B.1}$$

where

- M_o is a magnitude of zero that corresponds to a flux of $I_o = 10^6$ photons/s/cm^2/band
- A is the area of the telescope in cm^2
- D is the dark current in $e^-/s/pix$
- R_n is the CDS read noise of a given pixel. Note that this can be decreased by the factor $1/\sqrt{N}$ by sampling the pixel N times during the integration. The floor of $R_n/\sqrt{N}$ will likely be limited by $1/f$ noise and not zero.
- r_{source} is the approximate radius subtended by the source on the detector
- T is the transmitted fraction of light through the atmosphere and optical system
- η is the quantum efficiency of the detector

To obtain a signal to noise of S, then, we should expose for a time t given by

$$t = \frac{(S/N)^2 (I_o A \eta T * 10^{-(M_s - M_o)/2.5} + \pi r_{source}^2 (B + D))}{I_o^2 A^2 \eta^2 T^2 * 10^{-2(M_s - M_o)/2.5}}, \tag{B.2}$$

where we have neglected the contribution of R_n since it becomes negligible in comparison to B and D in the limit of large t.

Appendix C

Numerical Simulations

C.1 PN Junctions

In order to verify the validity of the numerical solutions and simulations used in the analysis of the detector PIN photodiodes, a numerical simulation of a simple one-dimensional PN junction has been conducted. The methodology follows closely that of Kurata [57] and MacCormack [126]. The results of these simulations match well with the analytical solution based upon the full depletion approximation as well as other simulations. A brief review of the techniques used will be given in this section.

C.2 Basic Semiconductor Physics

The crucial quantities in semiconductor analysis are the charge density, ρ, electric field, $\mathbf{E}$, electric potential, ϕ, the number of electrons in the conduction band, n_c (cm^{-3}), and the number of holes in the valence band p_v (cm^{-3}). As noted in Ashcroft and Mermin [127], conduction is entirely due to electrons in conduction band levels or holes in valence band levels. In order to describe the behavior of these quantities in a semiconductor with a dielectric constant of ϵ, we must apply the basic governing equations: Gauss's law,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon}, \tag{C.1}$$

Poisson's equation,

$$\nabla^2 \phi = -\frac{\rho}{\epsilon}, \tag{C.2}$$

and the continuity equations,

$$\frac{\partial n}{\partial t} = G_n - U_n + \frac{1}{q} \nabla \cdot \mathbf{J}_n, \tag{C.3}$$

$$\frac{\partial p}{\partial t} = G_p - U_p + \frac{1}{q}\nabla \cdot \mathbf{J}_p. \tag{C.4}$$

G_n and G_p are the electron and hole generation rate ($\mathrm{cm}^{-3}/\mathrm{s}$), respectively, and U_n and U_p are the corresponding recombination rates. $\mathbf{J}_n$ and $\mathbf{J}_p$ are the electron and hole current densities, respectively,

$$\mathbf{J}_p = q\mu_p p\mathbf{E} - qD_p\nabla p \tag{C.5}$$

$$\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n\nabla n. \tag{C.6}$$

μ_p and μ_n are the mobilities and D_p and D_n are the diffusion constants for holes and electrons, respectively. Each can be determined from the other via the Einstein relations $D_{p,n} = (kT/q)\mu_{p,n}$. In principle, the mobilities might be dependent on the electric field or position in the material. Here we will neglect any such dependencies. The charge density in the material has contributions from n_c and p_v along with the ionized donor impurity atoms, N_d^+, and ionized acceptor impurity atoms, N_a^-:

$$\rho = q(p_v - n_c + N_d^+ - N_a^-) \tag{C.7}$$

With these equations and proper expressions for G and R, we can appropriately describe the dynamics of the system. We shall neglect magnetic fields and external electric fields.

C.3 Numerical Methods: Finite Volume Scheme

To calculate the fields and concentrations numerically, we use a finite-difference scheme and a simple one-dimensional model represented by the diagram in Figure C.1 The concentrations and potentials are defined at grid points i and the electric field and current densities at grid points $i \pm 1/2$. The material extends from $x_0 = 0$ μm to $x_I = 2$ μm. In this simple case, a uniform grid is used, so $\Delta x_i = \Delta x_{i\pm1/2}$ is a constant. However, for other simulations where I is sufficiently large (i.e. for the 100 μm thick PIN diodes) a non-uniform grid is used. In these cases, the grid spacing is made very fine in regions where the physical quantities are expected to change rapidly and coarse in areas where they are expected to vary slowly. The separations between adjacent points are then given by the relation

$$\Delta x_{i\pm1/2} = (\Delta x_i + \Delta x_{i\pm1})/2 \tag{C.8}$$

The extension to 2 dimensions is straightforward. It simply involves placing another grid in the y direction and overlaying these two grids to form a mesh. In a similar fashion, we discretize time at points t_k, with $k = 0, 1, 2, ...$, and use a constant separation

$$\Delta t = t_{k+1} - t_k. \tag{C.9}$$

This allows approximations of time-derivatives.

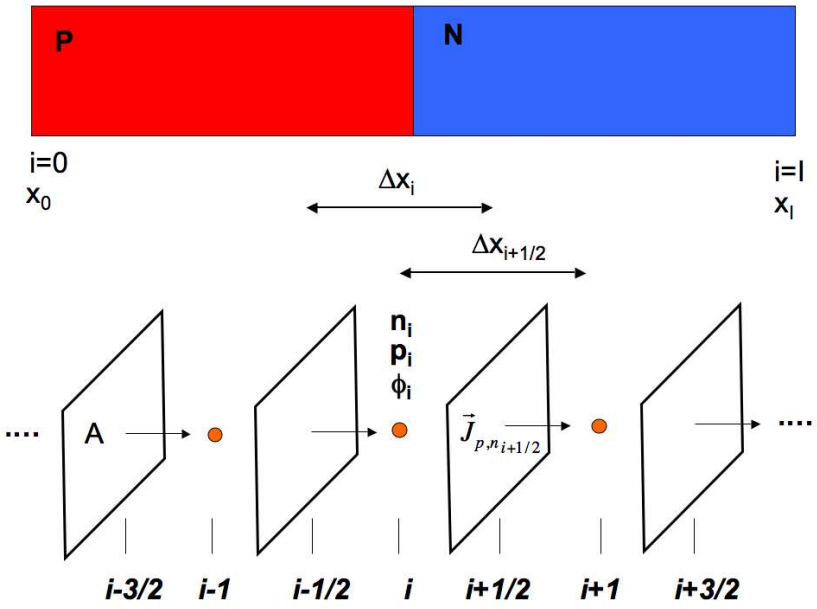


Figure C.1: Diagram showing configuration for the PN junction. The grid is staggered so that $\mathbf{E}$ and $\mathbf{J_{p/n}}$ are defined at half-integer values of $i \pm 1/2$ while the charge density and potential are defined at integer values i. The distance between x_i and x_{i+1} is $\Delta x_{i+1/2}$ and the distance between $x_{i-1/2}$ and $x_{i+1/2}$ is Δx_i. The surface enclosing the charge has area A on each side.

In the discrete approximation over the 1-d grid, Equations C.2-C.4 take the form,

$$\frac{p_i^{k+1} - p_i^k}{\Delta t} = -(1-\alpha)\left(\frac{J_{p,i+1/2}^k - J_{p,i-1/2}^k}{q\Delta x_i} - G_{p,i}^k + U_{p,i}^k\right) - \alpha\left(\frac{J_{p,i+1/2}^{k+1} - J_{p,i-1/2}^{k+1}}{q\Delta x_i} - G_{p,i}^{k+1} + U_{p,i}^{k+1}\right) \tag{C.10}$$

$$\frac{n_i^{k+1} - n_i^k}{\Delta t} = (1-\alpha)\left(\frac{J_{n,i+1/2}^k - J_{n,i-1/2}^k}{q\Delta x_i} - G_{n,i}^k - U_{n,i}^k\right) + \alpha\left(\frac{J_{n,i+1/2}^{k+1} - J_{n,i-1/2}^{k+1}}{q\Delta x_i} - G_{n,i}^{k+1} - U_{n,i}^{k+1}\right) \tag{C.11}$$

$$\frac{\phi_{i+1}^k - 2\phi_i^k + \phi_{i-1}^k}{\Delta x_i^2} = \frac{-q(p_i^k - n_i^k + N_{d,i}^+ - N_{a,i}^-)}{\epsilon} = \frac{-\rho_i}{\epsilon} \tag{C.12}$$

If we set $\alpha = 0$, we will have an *explicit* set of equations that can be solved with a very easy to code

algorithm for a steady state solution. At each time step k, the potential ϕ^k is solved for based upon the charge density at k. This potential is then used to calculate the new charge density at $k+1$. However, this method converges extremely slowly, taking anywhere from hours to days depending on the initial conditions used. This is because the method is unstable and requires a very small value for Δt to advance towards the solution.

If, on the other hand, we set $\alpha = 1/2$ or $\alpha = 1$, we will be using a *semi-implicit* or *implicit* method, respectively. The case of $\alpha = 1/2$ is often refereed to as the Crank-Nicolson method [126]. These methods converge much faster than the explicit method due to their increased stability. However, the algorithms are much more difficult to implement numerically due to the fact that we must solve the three coupled equations simultaneously because we do not know what values the quantities have at $k+1$. Not to mention, two of them are nonlinear.

To overcome the difficulty of nonlinearity, we must first linearize the equations. The current densities as well as the generation and recombination rates all require linearization. As an example, we first write the hole current density at time t_{k+1} as

$$J_{p,i+1/2}^{k+1} = J_{p,i+1/2}^{k} + \delta J_{p,i+1/2}. \tag{C.13}$$

If we neglect second-order terms and higher, the Taylor expansion of the change in $J_{p,i}$ from time-step k to $k+1$ can be expressed as

$$\delta J_{p,i+1/2} = \frac{\partial J_{p,i+1/2}}{\partial p_i}\delta p_i + \frac{\partial J_{p,i+1/2}}{\partial p_{i+1}}\delta p_{i+1} + \frac{\partial J_{p,i+1/2}}{\partial \phi_i}\delta \phi_i + \frac{\partial J_{p,i+1/2}}{\partial \phi_{i+1}}\delta \phi_{i+1}. \tag{C.14}$$

Similar expressions can be written for G_p, U_p, J_n, G_n, and U_n. The usefulness of the Taylor expansion becomes apparent after a few more steps and substitutions. We first write Equations C.10- C.12 as:

$$\frac{1}{\Delta t}\delta p + \alpha\frac{D_i}{q}\delta J_{p,i} + \alpha\delta G_{p,i} - \alpha\delta U_{p,i} = -\frac{D_i}{q}J_{p,i}^k + G_{p,i}^k - U_{p,i}^k \tag{C.15}$$

$$\frac{1}{\Delta t}\delta n - \alpha\frac{D_i}{q}\delta J_{n,i} + \alpha\delta G_{n,i} - \alpha\delta U_{n,i} = \frac{D_i}{q}J_{n,i}^k + G_{n,i}^k - U_{n,i}^k \tag{C.16}$$

$$D_i^2\delta\phi_i = -\frac{\rho_i^k}{\epsilon} + D_i^2\phi_i^k \tag{C.17}$$

where D_i and D_i^2 are second-order centered difference operators defined such that

$$D_i(f) = \frac{f_{i+1/2} - f_{i-1/2}}{\Delta x_i} \tag{C.18}$$

$$D_i^2(f) = \frac{f_{i+1}}{\Delta x_{i+1}\Delta x_{i+1/2}} - \frac{f_i}{\Delta x_i\Delta x_{i+1/2}} - \frac{f_i}{\Delta x_i\Delta x_{i-1/2}} + \frac{f_{i-1}}{\Delta x_{i-1}\Delta x_{i-1/2}}. \tag{C.19}$$

Then we introduce the vectors

$$\Theta = \begin{pmatrix} p \\ n \\ \phi \end{pmatrix} \qquad \delta\Theta = \begin{pmatrix} \delta p \\ \delta n \\ \delta\phi \end{pmatrix},$$

With these vectors and the introduction of several new matrices, the three coupled equations can be written as one matrix equation:

$$\begin{aligned} \alpha\left[\frac{\tilde{J}_{i-1/2,i-1}}{q} + \tilde{G}_{i,i-1} + \tilde{U}_{i,i-1}\right]\delta\Theta_{i-1} + \\ \left[T_i + \alpha\left(\frac{\tilde{J}_{i-1/2,i} - \tilde{J}_{i+1/2,i}}{q} + \tilde{G}_{i,i} + \tilde{U}_{i,i}\right)\right]\delta\Theta_i + \\ \alpha\left[\frac{\tilde{J}_{i+1/2,i+1}}{q} + \tilde{G}_{i,i+1} + \tilde{U}_{i,i+1}\right]\delta\Theta_{i+1} = F_i \end{aligned} \tag{C.20}$$

The notation is quite cumbersome and care must be taken with the double indices to ensure the derivatives are being approximated correctly. $\tilde{J_{i+1/2,i}}$, $\tilde{G_{i,i}}$, $\tilde{U_{i,i}}$, etc., are the 3×3 Jacobian of their respective variables and coordinates that handle the linearization. As an example,

$$\tilde{G}_{i,i+1} = \begin{pmatrix} \frac{\partial G_{p,i}}{\partial p_{i+1}} & 0 & \frac{\partial G_{p,i}}{\partial \phi_{i+1}} \\ 0 & \frac{\partial G_{n,i}}{\partial n_{i+1}} & \frac{\partial G_{n,i}}{\partial \phi_{i+1}} \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{U}_{i,i+1} = \begin{pmatrix} \frac{\partial U_{p,i}}{\partial p_{i+1}} & 0 & \frac{\partial U_{p,i}}{\partial \phi_{i+1}} \\ 0 & \frac{\partial U_{n,i}}{\partial n_{i+1}} & \frac{\partial U_{n,i}}{\partial \phi_{i+1}} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\tilde{J}_{i+1/2,i} = \begin{pmatrix} \frac{\partial J_{p,i+1/2}}{\partial p_i} & 0 & \frac{\partial J_{p,i+1/2}}{\partial \phi_i} \\ 0 & \frac{\partial J_{n,i+1/2}}{\partial n_i} & \frac{\partial J_{n,i+1/2}}{\partial \phi_i} \\ 0 & 0 & 0 \end{pmatrix}$$

and T_i and F_i are given by:

$$T_i = \begin{pmatrix} \frac{1}{\Delta t} & 0 & 0 \\ 0 & \frac{1}{\Delta t} & 0 \\ 0 & 0 & D_i^2 \end{pmatrix}, \quad F_i = \begin{pmatrix} -\frac{D_i}{q}J_{p,i}^k + G_{p,i}^k - U_{p,i}^k \\ \frac{D_i}{q}J_{n,i}^k + G_{n,i}^k - U_{n,i}^k \\ \frac{\rho_i^k}{\epsilon} + D_i^2\phi_i^k \end{pmatrix}$$

All of the quantities on the right-hand side of Equation C.20 are known at time-step t_k (at t_0 we supply a suitable initial guess) and the left-hand side represents a block tri-diagonal matrix multiplying the array of unknown column vectors $\delta\Theta$. To solve for $\delta\Theta$ we must invert this matrix with techniques such as the ones in [126] and [57]. In the case where a steady-state solution is sought, we solve for $\delta\Theta$ at each time t_k until $\delta\Theta \sim 0$.

The elements with values at $i = 0$ and $i = I$ must be handled separately, first, because they do

not have surrounding points at $i = -1$ and $i = I + 1$, and second, because they physically represent the supplied boundary conditions. For an ideal PN junction, the requirement is that the space charge density and electric field at the boundaries vanish and that the built-in electric potential takes its thermal equilibrium values. The first of these conditions, along with the law of mass-action that relates the electron and hole concentrations to the intrinsic concentration, n_{int}, yields

$$\begin{aligned} p_0 &= -\frac{N_{d,0} - N_{a,0}}{2}\left\{1 + \left[1 + \left(\frac{2n_{int}}{N_{d,0} - N_{a,0}}\right)^2\right]^{1/2}\right\} \quad , \quad n_0 = \frac{n_{int}^2}{p_0} \\ n_I &= \frac{N_{d,I} - N_{a,I}}{2}\left\{1 + \left[1 + \left(\frac{2n_{int}}{N_{d,I} - N_{a,I}}\right)^2\right]^{1/2}\right\} \quad , \quad p_I = \frac{n_{int}^2}{n_I} \end{aligned} \tag{C.21}$$

while the second results in

$$\phi_0 = -\frac{kT}{q}\ln\left[\frac{p_0}{n_{int}}\right] \qquad , \qquad \phi_I = V_{bias} + \frac{kT}{q}\ln\left[\frac{n_I}{n_{int}}\right] \tag{C.22}$$

where V_{bias} is the applied bias voltage. In the case of the PIN diodes, this will take the value of the substrate voltage, V_{SUB}, and will be enforced on the n^+ side of the junction.

The last point to consider before the equations are solved is a very subtle one; one that often causes headaches when implementing numerical methods. As shown by Scharfetter and Gummel [128], instability of the solution occurs when $|\phi_{i+1} - \phi_i| > 2kT/q$. One way to get around this is to make the space between grid points sufficiently small. The price one pays for this is increased computation time. An alternative way is to assume the current densities and electric field are constant in between grid points and instead to solve a differential equation to approximate $J_{p,n}$ at these location. The technique is referred to as Scharfetter-Gummel discretization and it results in the following expressions for the current densities:

$$\begin{aligned} J_{p,i+1/2} &= -\frac{q\mu_p}{\Delta x_{i+1/2}}\left[\left(\frac{\phi_i - \phi_{i+1}}{1 - e^{-q(\phi_i - \phi_{i+1})/kT}}\right)p_i + \left(\frac{\phi_i - \phi_{i+1}}{1 - e^{q(\phi_i - \phi_{i+1})/kT}}\right)p_{i+1}\right] \\ J_{n,i+1/2} &= -\frac{q\mu_n}{\Delta x_{i+1/2}}\left[\left(\frac{\phi_i - \phi_{i+1}}{1 - e^{q(\phi_i - \phi_{i+1})/kT}}\right)n_i + \left(\frac{\phi_i - \phi_{i+1}}{1 - e^{-q(\phi_i - \phi_{i+1})/kT}}\right)n_{i+1}\right] \end{aligned} \tag{C.23}$$

It is from these expressions that the partial derivatives in the Jacobians $\tilde{J}$ are calculated. For the exact matrix elements, the reader is referred to Kurata [57].

C.4 Results for Abrupt PN Junction

The actual implementation of this numerical method is carried out via a python script that makes use of the numpy and scipy libraries. As a test, we consider the abrupt junction shown in Figure

C.1 in which the P and N regions are both 1 μm in length. The doping densities are set such that $N_d = 0$ cm^{-3} and $N_a = 10^{18}$ cm^{-3} on the P side and $N_d = 10^{19}$ cm^{-3} and $N_a = 0$ cm^{-3} on the N side. The value of the intrinsic carrier density is $n_{int} = 10^{10}$ cm^{-3}. An implicit method ($\alpha = 1$) is used and 70 time steps of $\Delta t = 10^{-12}$ s are taken to reach the steady state solution. Larger values of Δt result in instabilities that cause the solution to diverge.

The results of a simulation with V_{bias} =0 V are shown in Figure C.2. The dashed lines on the plots indicate the depletion region boundaries obtained from the full depletion approximation. $x_p = 34.6$ nm is the distance the depletion region extends into the p material and $x_n = 3.46$ nm is the distance it extends into the n material. In the full depletion approximation, the changes in space charge densities occur as step discontinuities at the boundaries of the depletion region. The numerical solution does not have this simplification built in and shows that these transitions are indeed smooth as one would expect. They do not begin or end at the calculated boundaries, but rather are centered around them. The potential and number densities also vary in agreement with the analytical calculation, but show some variation outside the depletion region boundaries. The case of $V_{bias} = 1.2$ V is shown in Figure C.3. As expected, the size of the depletion region is increased. It reaches further into the less heavily doped p region then it does into the n region.

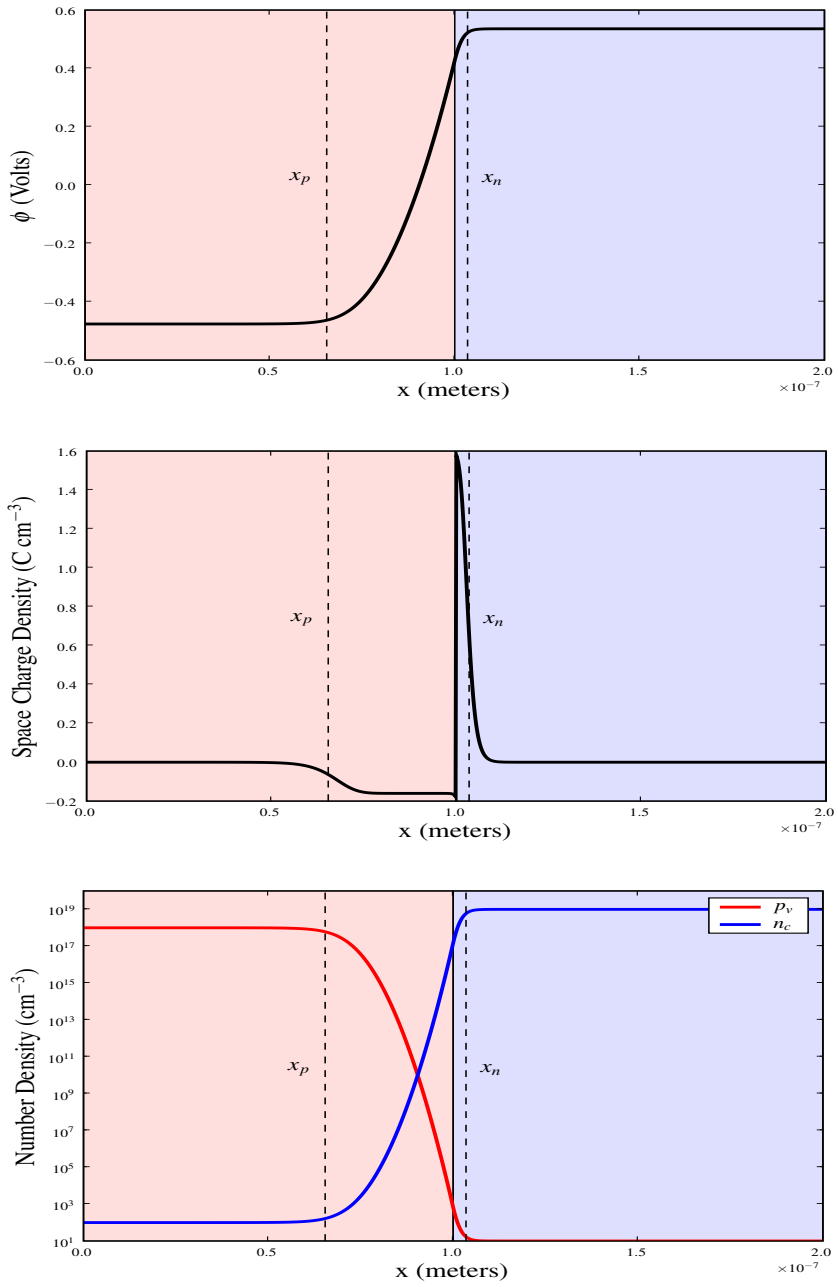


Figure C.2: Potential (ϕ), charge density, and number density of holes and electrons for the numerical simulation of an abrupt PN junction with no bias voltage applied. The depletion region boundaries x_p and x_n obtained from the analytical calculation are shown on the plots.

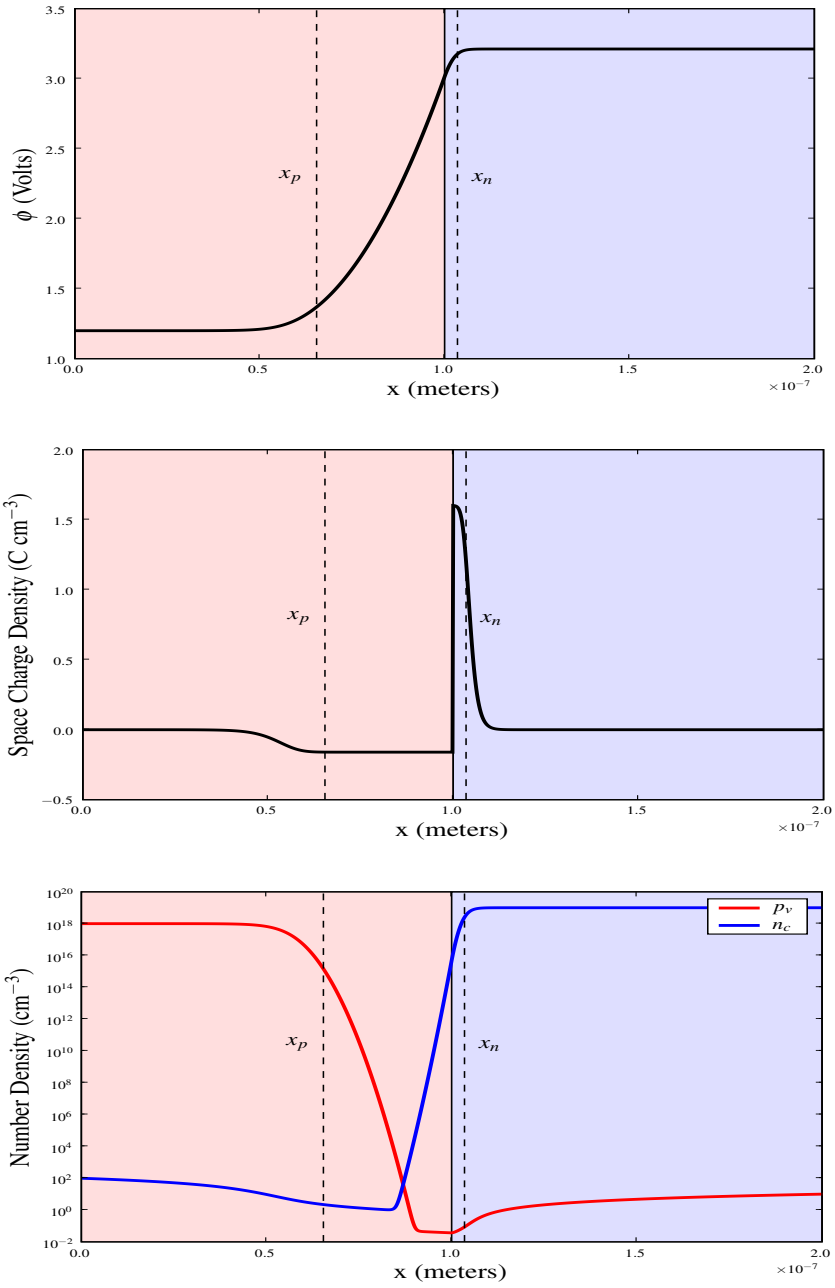


Figure C.3: Potential (ϕ), charge density, and number density of holes and electrons for the numerical simulation of an abrupt PN junction under reverse bias with $V_{bias} = 1.2$ V. The depletion region boundaries x_p and x_n shown are the ones calculated for the case where $V_{bias} = 0$ V to illustrate the increase in the width of the depletion region.

The same methodology used here will be applied to the PIN diodes in the HyViSI detectors. With the proper grid spacing and time step-size, we should expect that valid results will be obtained for structures more complex than the simple PN junction.

C.5 Cylindrically Symmetric Persistence Simulations

Two or three dimensional finite difference numerical simulations are difficult to carry out. This is especially true when an implicit method is used, as the two or three dimensions must be handled with different techniques. For radially symmetric phenomena, such as persistence in hybrid detectors, using polar coordinates reduces the complexity of the difference scheme, saves computation time and memory, and may potentially increase numerical stability.

We begin by considering cylindrically symmetric distributions of electrons, n, and holes, p, whose dynamics are governed by the semiconductor Equations C.1-C.7. In the context of the actual detector layer, these distributions would obviously have some extent in the z direction, but we ignore this and collapse them into a plane at $z = 0$, which corresponds to the front side of the detector. We will assume that excess carriers can still be lost to diffusion and subsequent recombination in the z direction, but after these particles have left the plane they will no longer be tracked. We will further make the simplifying assumption that the equilibrium carrier concentrations, p_v^o and n_v^o, balance the ionized donors, N_d^+ and N_a^-, so that these four species make no net contribution to the charge density, and thus no contribution to the radial electric field. The only contribution to the charge density and radial electric field then arises purely from the excess of carriers n and p. We will denote the number of electrons at radius r and time t as $n(r,t)$ and the number of holes as $p(r,t)$.

We will assume that the electrons are free to move in the radial direction and that generation is negligible, i.e. $G_n = 0$. The continuity equation governing the electron distribution can then be written as:

$$\frac{\partial n}{\partial t} = \nabla \cdot (\mathbf{J}_{Diff} + \mathbf{J}_{Drft}) - U_n = \frac{2J_{Diff}}{r} + \frac{\partial J_{Diff}}{\partial r} + \frac{2J_{Drft}}{r} + \frac{\partial J_{Drft}}{\partial r} - U_n, \tag{C.24}$$

where J_{Diff} is the electron diffusion current given by

$$J_{Diff} = D_n \nabla n(r,t) = D_n \frac{\partial n}{\partial r}\hat{r}, \tag{C.25}$$

J_{Drft} is the electron drift current given by

$$J_{Drft} = \mu_n n E_r \hat{r}, \tag{C.26}$$

and U_n is the rate of recombination, which will be considered shortly. D_n is the diffusion coefficient with units of $[length^2][time^{-1}]$, μ_n is the mobility with units of $[length^2][V^{-1}][time^{-1}]$, and E_r is

the radial electric field. Note that the currents are in terms of particles per unit time and not charge per unit time.

For reasons discussed in the text, the hole distribution is assumed to consist of immobile holes and contain no source of generation when the detector is not illuminated, i.e. $G_p = 0$.[1] It only changes through recombination with electrons,

$$\frac{\partial p}{\partial t} = -U_p = -\frac{pn}{\tau_p(n+p)}. \tag{C.27}$$

Under the assumption that the traps responsible for recombination are deep level traps, this recombination requires that an electron be removed from the conduction band as well, which means $U_n = U_p$. The trapped holes may very well exist outside of the plane $z = 0$, meaning the electrons lost through U_n need to drift or diffuse in the z direction to reach the recombination sites, and a true 2d simulation would need to account for this. Since transport of the electrons to the recombination sites is neglected, the simulation is only quasi-2d.

Substituting Equations C.25 and C.27 into C.24 yields the form that will be used to be solved for the electron distribution evolution. Along with the hole distribution equation and Gauss's Law for the electric field, the governing equations then have the form:

$$\frac{\partial n}{\partial t} = D_n\left(\frac{2}{r}\frac{\partial n}{\partial r} + \frac{\partial^2 n}{\partial r^2}\right) + \frac{2J_{Drft}}{r} + \frac{\partial J_{Drft}}{\partial r} - \frac{pn}{\tau_p(p+n)} \tag{C.28}$$

$$\frac{\partial p}{\partial t} = -\frac{pn}{\tau_p(n+p)} \tag{C.29}$$

$$\frac{\partial E_r}{\partial r} = \frac{p-n}{\epsilon} - \frac{2}{r}E_r \tag{C.30}$$

These are the equations that will be used to evolve the particle distributions in time.

Since the governing equation involves only one spatial dimension, a regularly, finely spaced grid of radial points r_i (from $r = r_o$ to $r = r_{max}$) can be used without creating memory allocation problems or significantly slowing the evolution of the system. This means all $\Delta r = r_{i+1} - r_i$ are equal in the grid. With second order centered difference schemes for the first and second spatial derivatives and a first order explicit time derivative, Equations C.29 and C.28 become

$$\frac{n_i^{k+1} - n_i^k}{\Delta t} = D_n\left(\frac{2}{r_i}\frac{n_{i+1}^k - n_{i-1}^k}{2\Delta r} + \frac{n_{i+1}^k - 2n_i^k + n_{i-1}^k}{\Delta r^2}\right) + \frac{2J_i^k}{r_i} + \frac{J_{i+1}^k - J_{i-1}^k}{2\Delta r} - \frac{p_i n_i}{\tau_p(n_i + p_i)} \tag{C.31}$$

$$\frac{p_i^{k+1} - p_i^k}{\Delta t} = -\frac{p_i n_i}{\tau_p(p_i + n_i)} \tag{C.32}$$

[1] A constant term G_p can be included to extend to the case where the detector is illuminated with a weak background flux. The results do not change significantly, so G_p is kept at zero here.

where k is the time index that relates the time elapsed t to the time increment Δt via $t = k\Delta t$. $J_i^k = J_{Drft}$ is the electron drift current at $r = r_i$ and time-step k given by

$$J_i^k = \mu_n n_i^k E_i^k. \tag{C.33}$$

Since Equation C.31 requires the electric field at each grid point be known, at each time-step, the electric field is first solved for using a backward difference scheme:

$$E_{i-1} = \Delta r \left(-\frac{2}{r_i} E_i + \frac{1}{\Delta r} E_i + \frac{p_i - n_i}{\epsilon} \right). \tag{C.34}$$

Starting at the maximum radius in the grid, $r = r_{max}$, far away from the electron and hole distributions so that $p = n = 0$ and $E = 0$, we integrate this toward $r = 0$ to obtain the electric field. Because the field would diverge at the origin, E_o is not included in the calculation, and instead set to zero. At each time-step k, Equation C.34 is used to solve for the field and then C.31 and C.32 are used to advance the solution forward in time.

The boundary condition for the electron distribution at the origin r_o is easily handled by recognizing that the first derivative is necessarily zero and then using the radial symmetry to create an imaginary point $r_{-1} = r_1$ that can be used for the second derivative:

$$n_0^{k+1} = n_0^k + D_n \Delta t \left(\frac{2n_1^k - 2n_0^k}{\Delta r^2} \right) - \frac{p_o}{\tau_p}. \tag{C.35}$$

And to reiterate, if r_{max} is made sufficiently large then the particles and field will not reach the other boundary, so $n_{i_{max}}^{k+1} = n_{i_{max}}^k = 0$, and so on for the other variables. With the computational method in place, the only thing that remains is to specify the initial distributions $n(r, 0)$ and $p(r, 0)$.

In choosing a set of units for the simulation, it turns out to be easiest to treat r in terms of pixels. This is because small numbers like the pixel pitch of 18×10^{-6} m necessitate very small time steps and contribute significantly to roundoff error in the divisions and multiplications used in the calculation, and this in turn decreases numerical stability. After the calculation is finished, the results can be easily converted to physically meaningful quantities via the following equations:

$$D_n^{phys} = D_n^{sim} * p^2 \tag{C.36}$$

$$\mu_n^{phys} = \mu_n^{sim} * p^2, \tag{C.37}$$

where p is the pixel pitch in cm.

Appendix D

Conversion Gain Reference Sheet

The equation relating the number of electrons (e^-) in a pixel to the recorded data number (DN or ADU) goes as:

$$G_{NET} = G_{PIXEL} * G_{UC} * G_{OUT} * G_{AMP} * G_{A/D} \tag{D.1}$$

Below is a short reference sheet of the experiments and what each yields.

Fe^{55} Calibration:

Method: Collect set of exposures that record Fe^{55} hits in the detector.
Histogram the hit values in ADU. The peak corresponds to 1660 e^-.
Notes: Value will depend on the gain of the control of A/D converter, G_{AMP}.
Provides: G_{net} (e^-/ADU)

Electronic Gain with V_{RESET}:

Method: Program set of voltages for V_{RESET}. Read detector output while reset switch is closed. Plot DN vs. V_{RESET} and obtain slope.
Provides: $G_{UC}(V/V) = \Delta V_{OUT\ NOSF}/(\Delta V_{RESET} * G_{ELEC})$
$G_{SF}(V/V) = \Delta V_{OUT\ SF}/(\Delta V_{RESET} * G_{UC} * G_{ELEC})$

A/D or Control Electronics Calibration:

Method: Use a set of known voltages as input to the A/D converter in control electronics.

Notes: G_{AMP} used here should correspond to G_{AMP} used in the Fe^{55} calibration.

Provides: $G_{ELEC}(V/ADU) = G_{AMP} * G_{A/D}$ – If amplification stages are included
$G_{A/D}$ (V/ADU) – If amplification stages are bypassed
G_{AMP} (V/V)

Well Depth from Saturated Images

Method: Use an exposure or set of saturated exposures to find the full range of the pixels in ADU.

Notes: The full range is the average taken over all pixels of the quantity $FR = I_{max} - I_{min}$, where I_{min} is the pixel value immediately after reset and I_{max} is the pixel value before the output becomes nonlinear and saturates.

Provides: $WellDepth$ (ADU)
$WellDepth$ $(e^-) = WellDepth$ (ADU) $* G_{net}$

Bibliography

[1] S. E. Holland, "Fully depleted charge-coupled devices," SNIC Symposium, (Stanford, CA), 2006.

[2] B. E. Burke, J. L. Tonry, M. J. Cooper, D. J. Young, A. H. Loomis, P. M. Onaka, and G. A. Luppino, "Development of the orthogonal-transfer array," *SPIE Proc.* **6068**, 2006.

[3] J. R. Janesick, *Scientific Charge-Coupled Devices*, SPIE-The International Society for Optical Engineering, Bellingham, Washington, 2001.

[4] I. S. McLean, *Electronic Imaging In Astronomy: Detectors and Instrumentation*, Praxis Publishing, Chichester, UK, 1997.

[5] P. Martinez and A. Klotz, *A Practical Guide to CCD Astronomy*, Cambridge University Press, Cambridge, UK, 1998.

[6] I. Deyhimy, R. C. Eden, R. J. Anderson, and J. S. Harris Jr., "A 500-MHz GaAs charge-coupled device," *Applied Physics Letters* **36**, January 1980.

[7] S. Lauxtermann, G. Izrael, P. Seitz, H. Bloss, J. Ernst, H. Firla, and S. Gick, "A mega-pixel high speed CMOS imager with sustainable gigapixel/sec readout rate," *2001 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors* , June 2001.

[8] T. Brandt, "High-speed CMOS imagers are flexible," *Laser Focus World* , February 2004.

[9] J. L. Lowrance, V. J. Mastrocola, G. F. Renda, P. K. Swain, R. Kabra, M. Bhaskaran, J. R. Tower, and P. A. Levine, "Ultrahigh-frame CCD imagers," *SPIE Proc.* **5210**(67), 2004.

[10] M. Downing, G. Finger, D. Baade, N. Hubin, O. Iwert, and J. Kolb, "Detectors for AO wavefront sensing," *SPIE Proc.* **7015**(62), 2008.

[11] J. Janesick, T. Elliot, and J. Tower, "CMOS detectors: Scientific monolithic CMOS imagers come of age," *Laser Focus World* , July 2008.

[12] M. Sirianni and M. Mutchler, "Radiation damage in HST detectors," in *Scientific Detectors For Astronomy 2005*, pp. 171–178, Springer Netherlands, 2006.

[13] C. J. Bebek, J. H. Bercovitz, D. E. Groom, S. E. Holland, R. W. Kadel, A. Karcher, W. F. Kolbe, H. M. Oluseyi, N. P. Palaio, V. Prasad, B. T. Turko, and G. Wang, "Fully depleted back-illuminated p-channel CCD development," *SPIE Proc.* **5167**, pp. 50–62, 2004.

[14] J. Janesick, J. T. Andrews, and T. Elliot, "Fundamental performance differences between CMOS and CCD imagers: Part 2," *SPIE Proc.* **6276**, 2006.

[15] S. V. W. Beckwith et al., "The Hubble ultra deep field," *The Astronomical Journal* **132**, pp. 1729–1755, November 2006.

[16] J. Mack, J. Biretta, S. Bagget, and C. Proffitt, "WFPC2 dark current vs. time," *Instrument Science Report WFPC2 2001-2005* , 2001.

[17] P. Magnan, "Detection of visible photons in CCD and CMOS: A comparative view," *Nuc. Inst. and Meth.* **504**, pp. 199–212, 2003.

[18] J. W. Beletic, "Optical and infrared detectors for astronomy: Basic principles to state-of-the-art," in *Optics in Astrophysics*, pp. 123–154, Springer Netherlands, 2005.

[19] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of CMOS image sensors," *Microelectronics Journal* **37**, pp. 443–451, 2006.

[20] E. Goujou, P. Gorria, E. Fauvet, M. Robert, and G. Cathebras, "Study and characterization of photosensitive cells in ASICs," *Industrial Electronics, Control and Instrumentation* **2**, pp. 911–914, 1994.

[21] S. K. Mendis, S. E. Kemeny, R. C. Gee, B. Pain, C. O. Staller, Q. Kim, and E. R. Fossum, "CMOS active pixel image sensors for highly integrated imaging systems," *IEEE Journal of Solid-State Circuits* **32**(2), 1997.

[22] A. Hoffman, M. Loose, and V. Suntharalingam, "CMOS detector technology," in *Scientific Detectors For Astronomy 2005*, pp. 377–402, Springer Netherlands, 2006.

[23] J. W. Beletic, R. Blank, D. Gulbransen, D. Lee, M. Loose, E. C. Piquette, T. Sprafke, W. E. Tennant, M. Zandian, and J. Zino, "Teledyne Imaging Sensors: Infrared imaging technologies for astronomy and civil space," *SPIE Proc.* **7021**(20), 2008.

[24] A. Moore, Z. Ninkov, and W. J. Forrest, "Quantum efficiency overestimation and deterministic cross talk resulting from interpixel capacitance," *Opt. Eng.* **45**(7), p. 076402, 2006.

[25] Y. Bai, J. Bajaj, J. W. Beletic, M. C. Farris, A. Joshi, S. Lauxtermann, A. Petersen, and G. Williams, "Teledyne Imaging Sensors: Silicon CMOS imaging technologies for x-ray, uv, visible and near infrared," *SPIE Proc.* **7021**(01), 2008.

[26] K. D. Munck, D. S. Tezcan, T. Borgers, W. Ruythooren, P. D. Moor, S. Sedky, C. Toccafondi, J. Bogaerts, and C. V. Hoof, "High performance hybrid and monolithic backside thinned CMOS imagers realized using a new integration process," *Electron Devices Meeting* , pp. 1–4, Dec. 2006.

[27] B. Burke, P. Jorden, and P. Vu, "CCD technology," *Experimental Astronomy* **19**, pp. 69–102, 2005.

[28] J. Ohta, *Smart CMOS Image Sensors and Applications*, Taylor and Francis Group, Boca Raton, FL, 2008.

[29] C. Hong and R. I. Hornsey, "CMOS active pixel sensor with in-pixel contrast stretch," *2001 IEEE Workshops on CCDs and Advanced Image Sensors* , June 2001.

[30] J. P. Crooks, S. E. Bohndiek, C. D. Arvanitis, R. Speller, H. XingLiang, E. G. Villani, M. Towrie, and R. Turchetta, "A CMOS image sensor with in-pixel ADC, timestamp, and sparse readout," *IEEE Sensors Journal* **9**, pp. 20–28, 2009.

[31] D. X. D. Yang, B. Fowler, and A. E. Gamal, "A nyquist-rate pixel-level ADC for CMOS image sensors," *IEEE Journal of Solid-State Circuits* **34**(3), 1999.

[32] J. Vaillant and F. Hirigoyen, "Optical simulation for CMOS imager microlens optimization," *SPIE Proc.* **5459**, 2004.

[33] J. Yoshida, "OmniVision adopts backside illumination technology for CMOS imager," *EE Times* , May 2008.

[34] J. W. Beletic, "IR focal-plane arrays enable imaging that is out of this world," *Laser Focus World* , October 2007.

[35] B. Pain and E. R. Fossum, "A review of infrared readout electronics for space science sensors," tech. rep., Jet Propulsion Laboratory, July 1993.

[36] G. H. Rieke, "Infrared detector arrays for astronomy," *Annual Review of Astronomy And Astrophysics* **45**, pp. 77–115, 2007.

[37] Y. Bai, S. G. Bernd, J. R. Hosack, M. C. Farris, J. T. Montroy, and J. Bajaj, "Hybrid CMOS focal plane array with extended UV and NIR response for space applications," *SPIE Proc.* **48**, 2003.

[38] S. Kilcoyne, N. Malone, M. Harris, J. Vampola, and D. Lindsay, "Silicon p-i-n focal plane arrays at raytheon," *SPIE Proc.* **7082**, 2008.

[39] L. M. Simms, D. F. Figer, B. J. Hanold, D. J. Kerr, D. K. Gilmore, S. M. Kahn, and J. A. Tyson, "First use of a HyViSI H4RG for astronomical observations," *SPIE Proc.* **6690**, August 2007.

[40] B. N. Dorland, G. S. Hennessy, N. Zacharias, D. G. Monet, H. Harris, C. Rollins, P. Shu, L. Miko, B. Mott, A. Waczynski, and E. Kan, "Laboratory and sky testing results for the TIS H4RG-10 4kx4k, 10 micron visible CMOS-hybrid detector," *SPIE Proc.* **6690**, 2007.

[41] M. J. McCaughrean, *The Astronomical Application of Infrared Array Detectors*. PhD thesis, University of Edinburgh, 1988.

[42] S. L. Solomon, *Near Infrared Spectroscopic Imaging of NGC 7023 and Charge Trapping in InSb Photodiode Arrays*. PhD thesis, University of Rochester, 1998.

[43] A. C. Moore, *Operating, Testing, and Evaluating Hybridized Silicon P-I-N Arrays*. PhD thesis, Rochester Institute of Technology, 2005.

[44] H. Bushouse, C. Skinner, and J. MacKenty, "The STSCI NICMOS pipeline: CALNICA, single image reduction," June 10, 1996.

[45] Z. Ivezic et al., "LSST: From science drivers to reference design and anticipated data," *arXiv:astro-ph/0805.2366v1* , 15 May 2008.

[46] B. R. Hancock, T. J. Cunningham, K. McCarty, G. Yang, C. Wrigley, P. G. Ringold, R. C. Stirbl, and B. Pain, "Multi-megarad (Si) radiation tolerant integrated CMOS imager," *SPIE Proc.* **4306**, 2001.

[47] M. Bruzzi, "Radiation damage in silicon detectors for high-energy physics experiments," *IEEE Transactions on Nuclear Science* **48**(4), 2001.

[48] A. D. Falcone, D. N. Burrows, Y. Bai, M. Farris, R. Cook, and S. Bongiorno, "Hybrid CMOS x-ray detectors: The next generation for focused x-ray telescopes," *SPIE Proc.* **6686**, 2007.

[49] R. J. Dorn, S. Eschbaumer, G. Finger, L. Mehrgan, M. Meyer, and J. Stegmeier, "A CMOS visible silicon imager hybridized to a Rockwell 2RG multiplexer as a new detector for ground based astronomy," *SPIE Proc.* **6276**, 2006.

[50] S. W. Han and E. Yoon, "Low dark current CMOS image sensor pixel with photodiode structure enclosed by p-well," *Electronics Letters* **42**(20), 2006.

[51] C. Y. Wu, Y. C. Shih, J. F. Lan, C. C. Hsieh, C. C. Huang, and J. H. Lu, "Design, optimization, and performance analysis of new photodiode structures for CMOS active-pixel-sensor (APS) imager applications," *IEEE Sensors Journal* **4**(1), 2004.

[52] G. Finger, R. Dorn, M. Meyer, L. Mehrgan, A. Moorwood, and J.Stegmeier, "Interpixel capacitance in large format CMOS hybrid arrays," *SPIE Proc.* **6267**, 2006.

[53] S. M. Sze, *Physics of Semiconductor Devices: Second Edition*, John Wiley & Sons, 1981.

[54] S. Kayali, G. Ponchack, and R. Shaw, "GaAs MMIC reliability assurance guideline for space applications," Tech. Rep. 96-25, Jet Propulsion Laboratory, December 1996.

[55] R. F. Pierret, *Volume I: Semiconductor Fundamentals*, Addison-Wesley, second ed., 1988.

[56] Y. Bai, M. C. Farris, A. Joshi, J. R. Hosack, J. Bajaj, and J. T. Montroy, "Recent progress of hybrid CMOS visible focal plane array technology," *SPIE Proc.* **5902**, 2005.

[57] M. Kurata, *Numerical Analysis for Semiconductor Devices*, Lexington Books, 1982.

[58] W. W. Gärtner, "Depletion-layer photoeffects in semiconductors," *Physical Review* **116**(1), 1959.

[59] B. J. Rauscher, O. Fox, P. Ferruit, R. J. Hill, A. Waczynski, Y. Wen, W. Xia-Serafino, B. Mott, D. Alexander, C. K. Brambora, R. Derro, C. Engler, M. B. Garrison, T. Johnson, S. S. Manthripragada, J. M. Marsh, C. Marshall, R. J. Martineau, K. B. Shakoorzadeh, D. Wilson, W. D. Roher, M. Smith, C. Cabelli, J. Garnett, M. Loose, S. Wong-Anglin, M. Zandian, E. Cheng, T. Ellis, B. Howe, M. Jurado, G. Lee, J. Nieznanski, P. Wallis, J. York, M. W. Regan, D. N. B. Hall, K. W. Hodapp, T. Böoker, G. D. Marchi, P. Jakobsen, and P. Strada, "Detectors for the james webb space telescope near-infrared spectrograph I: Readout mode, noise model, and calibration considerations," *arXiv:0706.2344v1* , 15 Jun 2007.

[60] M. Loose, J. Beletic, J. Garnett, and M. Xiu, "High-performance focal plane arrays based on the HAWAII-2RG/4RG and the SIDECAR ASIC," *SPIE Proc.* **6690**, 2007.

[61] M. Vadipour, "Capacitive feedback technique for wide-band amplifiers," *IEEE Journal of Solid-State Circuits* **28**, January 1993.

[62] D. F. Figer, M. Agronin, J. Balleza, R. Barkhouser, L. Bergeron, G. R. Greene, S. R. McCandliss, B. J. Rauscher, T. Reeves, M. W. Regan, U. Sharma, and H. S. Stockman, "The Independent Detector Testing Laboratory and the NGST detector program," *SPIE Proc.* **4850**(981), 2003.

[63] I. S. Glass, *Handbook of Infrared Astronomy*, Cambridge University Press, Cambridge, UK, 1999.

[64] J. R. Janesick, G. B. Soli, T. S. Elliot, and S. A. Collins, "Effects of proton damage on charge coupled devices," *SPIE Proc.* **1447**(87), 1991.

[65] D. F. Figer, B. J. Rauscher, M. W. Regan, E. Morse, J. Balleza, L. Bergeron, and H. S. Stockman, "Independent testing of JWST detector prototypes," *SPIE Proc.* **5167**(270), 2004.

[66] B. Pain and T. J. Cunningham, "Fundamentals of CMOS imagers, operation, performance, trends." Lecture Notes, 2009.

[67] O. Yadid-Pecht, B. Mansoorian, E. R. Fossum, and B. Pain, "Optimization of noise and responsivity in cmos active pixel sensors for detection of ultra low light levels," *SPIE Proc.* **2019**, pp. 125–136, 1997.

[68] H. Tian and A. E. Gamal, "Analysis of 1/f noise in CMOS APS," *SPIE Proc.* **3965**, pp. 168–176, 2000.

[69] H. Tian, B. Fowler, and A. E. Gamal, "Analysis of temporal noise in CMOS photodiode active pixel sensor," *IEEE Journal of Solid-State Circuits* **36**(1), 2001.

[70] C. Leyri, S. Pilorget, M. Marin, M. Minondo, and H. Jaouen, "Random telegraph signal noise SPICE modeling for circuit simulators," *Solid State Devices Research Conference* **37**, 2007.

[71] A. M. Fowler and I. Gatley, "Demonstration of an algorithm for read-noise reduction in infrared arrays," *Astrophysical Journal* **353**(33), 1990.

[72] J. D. Garnett and W. J. Forrest, "Multiply sample read limited and background limited noise performance," *SPIE Proc.* **1946**(395), 1993.

[73] J. D. Offenberg, D. J. Fixsen, M. A. Nieto-Santisteban, R. Sengupta, J. C. Mather, and H. S. Stockman, "Uniform data sampling: Noise reduction & cosmic rays," *ASP Conference Series: Astronomical Data Analysis Software and Systems X* **238**, 2001.

[74] A. M. Fowler and I. Gatley, "Noise reduction strategy for hybrid IR focal plane arrays," *SPIE Proc.* **1541**, 1991.

[75] R. Chapman, S. Beard, M. Mountain, D. Pettie, and A. Pickup, "Implementation of a charge integration system in a low background application," *SPIE Proc.* **1235**, 1990.

[76] N. Bezawad and D. Ives, "High-speed multiple window readout of HAWAII-1RG detector for a radial velocity measurement," *SPIE Proc.* **6276**, 2006.

[77] I. M. Peters, E. W. Bogaart, W. Hoekstra, A. C. M. Kleimann, and J. T. Bosiers, "Very-low dark current in FF-CCDs," *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, (Karuizawa, Japan), 2005.

[78] R. J. Dorn, S. Eschbaumer, D. N. B. Hall, G. Finger, L. Mehrgan, M. Meyer, and J. Stegmeier, "Evaluation of the Teledyne SIDECAR ASIC at cryogenic temperature using a visible hybrid H2RG focal plane array in 32 channel readout mode," *SPIE Proc.* **7021-29**, 2008.

[79] J. R. Janesick, *Photon Transfer*, SPIE-The International Society for Optical Engineering, Washington, 2007.

[80] D. F. Figer, M. Regan, and E. Morse, "Independent testing of silicon PIN detector arrays for LSST," *American Astronomical Society Meeting 2005* **36**, p. 1528, 2004.

[81] P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge University Press, Cambridge, UK, 1989.

[82] W. H. Press, W. T. Vetterling, S. A. Teukolsky, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, NY, 1992.

[83] D. Bacon, A. Refregier, and R. Ellis, "Detection of weak gravitational lensing by large-scale structure," *Mon. Not. Roy. Astron. Soc.* **318**, p. 625, 2000.

[84] C. Sterken, *Astronomical Photometry*, Springer Verlag, New York, LLC, 2002.

[85] G. Verdoes, R. Vermeij, E. Valentijn, and K. Kuijken, "The secondary standards program for OmegaCAM at the VST," *arXiv:astro-ph/0612469v1* , 18 Dec 2006.

[86] P. Stetson, "Some factors affecting the accuracy of stellar photometry with CCDs," *DAO Preprint* , September 1988.

[87] S. Jester et al., "The Sloan Digital Sky Survey view of the Palomar-Green bright quasar survey," *The Astronomical Journal* **130**(3), pp. 873–895, 2005.

[88] B. N. Dorland, R. Dudik, G. S. Hennessy, A. Waczynski, B. J. Rauscher, P. K. Shu, and J. E. Hubbs, "Laboratory and sky testing of the second generation Teledyne Imaging Sensor (TIS) 16.7 megapixel visible/near-IR CMOS-hybrid array H4RG-10 A2," *SPIE Proc.* **7439**, 2009.

[89] A. U. Landolt, "UBV photoelectric sequences in the celestial equatorial selected areas 92-115," *The Astronomical Journal* **78**(959), 1973.

[90] A. U. Landolt, "UBVRI photometric standard stars in the magnitude range $11.5 < V < 16.0$ around the celestial equator," *Astrophysical Journal* **104**(1), 1992.

[91] C. A. Gullixson, P. C. Boeshaar, J. A. Tyson, and P. Seitzer, "The BjRI photometric system," *The Astrophysical Journal Supplement* **99**, pp. 281–293, 1995.

[92] L. A. Hillenbrand, J. B. Foster, S. Persson, and K. Matthews, "The Y band at 1.035 microns: Photometric calibration and the dwarf stellar/substellar color sequence," *Publications of the Astronomical Society of the Pacific* **114**(797), pp. 708–720, 2002.

[93] P. Massey and L. E. Davis, "A user's guide to stellar CCD photometry with IRAF." 1992.

[94] T. Oja, "The variable star HD 79889," *Astronomy and Astrophysics* **184**, pp. 215–218, 1987.

[95] L. Yanying, J. Shiyang, and C. Ming, "Period changes in HD 79889," *Information Bulletin on Variable Stars* **3607**, 1992.

[96] T. Qingquan, Y. Dawei, and J. Shiyang, "Period changes in HD 79889," *Information Bulletin on Variable Stars* **3771**, 1992.

[97] L. Zhongli and J. Shiyang, "Period change of Delta Scuti star HD 79889," *Information Bulletin on Variable Stars* **4077**, 1994.

[98] A. Derekas, L. L. Kiss, P. Szekely, E. J. Alfaro, B. Csak, S. Meszaros, E. Rodriguez, A. Rolland, K. Sarneczky, G. M. Szabo, K. Szatmary, M. Varadi, and C. Kiss, "A photometric monitoring of bright high-amplitude delta scuti stars. ii. period updates for seven stars," *Astronomy and Astrophysics* **402**, p. 733, 2003.

[99] L. L. Kiss and K. Szatmary, "Has the Delta Scuti star BE Lyn a companion?," *Information Bulletin on Variable Stars* **4116**, 1995.

[100] G. Finger, R. J. Dorn, S. Eschbaumer, D. N. B. Hall, L. Mehrgan, M. Meyer, and J. Stegmeier, "Performance evaluation, readout modes, and calibration techniques of HgCdTe HAWAII-2RG mosaic arrays," *SPIE Proc.* **7021-29**, 2008.

[101] E. W. Greisen and M. R. Calabretta, "Representations of world coordinates in FITS," *Astronomy and Astrophysics* **395**, pp. 1061–1075, 2002.

[102] D. G. Monet, "Introduction to CCD astrometry," *Astronomical CCD Observing and Reduction Techniques, ASP Conference Series* **23**, 1992.

[103] N. Zacharias, "Measuring the atmospheric influence on differential astrometry: A simple method applied to wide-field CCD frames," *Publications of the Astronomical Society of the Pacific* **108**, pp. 1135–1138, 1996.

[104] N. Zacharias et al., "The USNO CCD Astrograph Catalog (UCAC) project and beyond," *Astrophysics and Space Science Library* **288**, p. 67, 1996.

[105] J. Koornneef, "Near-infrared photometry: II. intrinsic colours and the absolute calibration from one to five microns," *Astronomy and Astrophysics* **128**, pp. 84–93, 1983.

[106] G. G. Pavlov and J. A. Nousek, "Charge diffusion in CCD x-ray detectors," *Nuclear Instrumentation and Methods A* **428**(348), 1999.

[107] P. O'Connor, V. Radeka, D. Figer, J. G. Geary, D. K. Gilmore, J. Oliver, C. W. Stubbs, P. Z. Takacs, and J. A. Tyson, "Study of silicon thickness optimization for LSST," *SPIE Proc.* **6276**, 2006.

[108] D. Groom, "Cosmic rays and other nonsense in astronomical CCD imagers," *Exp. Astronomy* **14**(1), 2002.

[109] C. Amsler et al. (Particle Data Group) *Phys. Lett. B* **667**(1), 2008.

[110] A. R. Smith, R. J. McDonald, D. L. Hurley, S. E. Holland, D. E. Groom, W. E. Brown, D. K. Gilmore, R. J. Stover, and M. Wei, "Radiation events in astronomical CCD images," *IS&T/SPIE's Electronic Imaging 2002* , 2002.

[111] M. Brown, M. Schubnell, and G. Tarlé, "Correlated noise and gain in unfilled and epoxy-underfilled hybridized HgCdTe detectors," *The Publications of the Astronomical Society of the Pacific* **118**(848), pp. 1443–1447, 2006.

[112] W. Shockley and W. T. Read Jr., "Statistics of the recombination of holes and electrons," *Physical Review* **87**, pp. 835–842, September 1952.

[113] M. J. McNutt and W. E. Meyer, "Bulk impurity charge trapping in buried channel charge coupled devices," *Journal of the Electrochemical Society* **128**, pp. 892–896, April 1981.

[114] G. L. Miller, D. V. Lang, and L. C. Kimerling, "Capacitance transient spectroscopy," *Annual Review of Material Science* **7**, p. 377, 1977.

[115] L. Trombetta, F. Feigi, and R. Zeto, "Positive charge generation in metal-oxide-semiconductor capacitors," *Journal of Applied Physics* **69**, February 1991.

[116] R. M. Smith, M. Zavodny, G. Rahmer, and M. Bonati, "A theory for image persistence in HgCdTe photodiodes," *SPIE Proc.* **7021**, 2008.

[117] G. H. Rieke, *Detection of Light: From the Ultraviolet to the Submillimeter*, Cambridge University Press, Cambridge, UK, 2nd ed., 2003.

[118] D. M. Fleetwood, P. S. Winokur, R. A. Reber Jr., T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Riewe, "Effects of oxide traps, interface traps, and border traps on metal-oxide-semiconductor devices," *Journal of Applied Physics* **73**(10), 1993.

[119] J. J. Makwana and D. K. Schroder, "A non-volatile memory overview." http://aplawrence.com/Makwana/nonvolmem.html, 2004.

[120] D. A. Aaronov and R. Mamatkulov, "On the theory of transient process after reversal of a P-I-N diode current from forward to reverse direction (i)," *Physica Status Solidi* **14**(2), pp. 695–704, 1972.

[121] A. Hoffman and E. Spenke, "Switching processes in alloyed PIN rectifiers," *Solid State Electronics* **8**(8), pp. 693–697, 1965.

[122] G. W. Neudeck, *Volume II: the PN Junction Diode*, Addison-Wesley, second ed., 1989.

[123] S. R. Hofstein, "An investigation of instability and charge motion in metal-silicon oxide-silicon structures," *IEEE Transactions on Electron Devices* **13**(12), 1966.

[124] A. S. Grove and D. J. Fitzgerald, "Surface effects on p-n junctions: Characteristics of surface space-charge regions under non-equilibrium conditions," *Solid State Physics* **9**(8), p. 783, 1966.

[125] R. M. Smith, M. Zavodny, G. Rahmer, and M. Bonati, "Calibration of image persistence in HgCdTe photodiodes," *SPIE Proc.* **7021**, 2008.

[126] B. MacCormack, "Numerical computation of compressible viscous flow." Course Reader for Stanford Computational Fluid Dynamics, 2006.

[127] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Saunders College Publishing, Philadelphia, 1976.

[128] D. Scharfetter and H. Gummel, "Large-signal analysis of a silicon read diode oscillator," *IEEE Transactions on Electron Devices* **16**, pp. 64–77, January 1969.