R·I·T

Single Photon Counting Detectors for Low Light Level Imaging Applications

by

Kimberly Kolb

B.S. Rochester Institute of Technology (2008)

M.S. Rochester Institute of Technology (2011)

A thesis submitted in partial fulfillment of the
requirements for the degree of Doctorate of Philosophy in Imaging Science
in the Chester F. Carlson Center for Imaging Science
of the College of Science
Rochester Institute of Technology

April 15, 2015

Signature of the Author_____

Accepted by _____
Coordinator, Ph.D. Degree Program Date

Chester F. Carlson Center for imaging Science
College of Science
Rochester Institute of Technology
Rochester, New York

CERTIFICATE OF APPROVAL

Ph.D. Degree Dissertation

The Ph.D. degree dissertation of Kimberly Kolb
has been examined and approved by the
thesis committee as satisfactory for the
thesis requirement for the
Doctorate of Philosophy degree

Committee Approval:

Dr. Santosh Kurinec, Committee Chairperson Date

Dr. Donald F. Figer, Thesis Advisor Date

Dr. Zoran Ninkov, Committee Member Date

Dr. John Kerekes, Committee Member Date

Dr. Shouleh Nikzad, Committee Member Date

ABSTRACT

This dissertation presents the current state-of-the-art of semiconductor-based photon counting detector technologies. HgCdTe linear-mode avalanche photodiodes (LM-APDs), silicon Geiger-mode avalanche photodiodes (GM-APDs), and electron-multiplying CCDs (EMCCDs) are compared via their present and future performance in various astronomy applications. LM-APDs are studied in theory, based on work done at the University of Hawaii. EMCCDs are studied in theory and experimentally, with a device at NASA's Jet Propulsion Lab. The emphasis of the research is on GM-APD imaging arrays, developed at MIT Lincoln Laboratory and tested at the RIT Center for Detectors. The GM-APD research includes a theoretical analysis of SNR and various performance metrics, including dark count rate, afterpulsing, photon detection efficiency, and intrapixel sensitivity. The effects of radiation damage on the GM-APD were also characterized by introducing a cumulative dose of 50 krad(Si) via 60 MeV protons. Extensive development of Monte Carlo simulations and practical observation simulations was completed, including simulated astronomical imaging and adaptive optics wavefront sensing. Based on theoretical models and experimental testing, both the current state-of-the-art performance and projected future performance of each detector are compared for various applications. LM-APD performance is currently not competitive with other photon counting technologies, and are left out of the application-based comparisons. In the current state-of-the-art, EMCCDs in photon counting mode out-perform GM-APDs for long exposure scenarios, though GM-APDs are better for short exposure scenarios (fast readout) due to clock-induced-charge (CIC) in EMCCDs. In the long term, small improvements in GM-APD dark current will make them superior in both long and short exposure scenarios for extremely low flux. The efficiency of GM-APDs will likely always be less than EMCCDs, however, which is particularly disadvantageous for moderate to high flux rates where dark noise and CIC are insignificant noise sources. Research into decreasing the dark count rate of GM-APDs will lead to development of imaging arrays that are competitive for low light level imaging and spectroscopy applications in the near future.

ACKNOWLEDGEMENTS

At first, I endeavored to make this short - to say my piece and move on. But while the rest of this document is full of math and physics and science, I wanted this section to be a heartfelt thank you to all the players in my life that have helped me get so far, and who I am sure will help me go even farther. No one on Earth got to where they are today without the aid of others. Sometimes this aid comes in the form of encouragement, other times in the form of obstacles, other times in the form of indifference. Every interaction in life shapes who we are, who we will become. Positive and negative, our life is what we choose to make of it. Some of us get luckier than others, and I am grateful every day to be so lucky as to have found myself where I am now.

I'd like to thank my grandfather, Daniel Manser, for never doubting that I would be great. For constantly telling me stories about NASA, and how special and important it was to explore space – how amazing humans are. For telling me that I could go there one day, if I put my mind to it. I might not go in person, but I hope to keep exploring the infinite through its photon fingerprints. I'd like to thank my father, James Manser, for being the best man in my life. For always supporting me, and challenging me to be a better person. For showing me how to be kind to strangers and optimistic for tomorrow. For always believing in the best of others, even when they seem determined to have us see the worst. I miss you both every day.

I'd like to thank my husband, for being understanding and patient, for letting me drag him around the country, for going on this adventure with me. I do not exaggerate when I say I could not have done this without him. He makes me laugh, which has been invaluable to my sanity. Throughout more than a few existential crises during the course of my graduate work, he has helped to ground me and focus me. Here's to our next adventure.

I'd like to thank my family for their unconditional love and support. To my mother, who always tells me I can do anything, and to my sister who always understands me, no matter how strange I am. To my mother-in-law, who has always treated me like her own daughter, to my father-in-law who makes me feel loved, even from a distance. To my brother- and sisters-in-law, who accepted me and loved me, and made me feel like part of the family. To my grandmother Betty, who taught me the value of patience and calming work for my hands – I miss you. To my

grandmother Shirley, who is always proud of me, no matter how small the accomplishment. To my aunts and uncles and cousins, who always feel like home, no matter where or when I see you. Thank you.

To my friends, thank you for being so encouraging and warm. Thank you for the distractions, and for the help in focusing. To Christine Trombley, my favorite astrophysicist, thanks for commiserating and laughing and sewing and baking. Thanks for your help - personal and academic and professional.

To my colleagues at the CfD, who stimulate my mind and motivate me to do my best work. Brandon, Joong, Iain, and the student workers - thank you for your support, and for the many conversations I've had with you, both personal and professional. For the good times we've had on trips to conferences and to do work, thanks for making difficult work seem easier and even more rewarding.

To my advisor, Don Figer, who never made it easy. Behind all the short emails and never-ending revisions and criticism, I have always known that you only wanted the best. And I was always inspired by the idea that you thought I could be better. You can provide more motivation in three words or less than anyone else. Thank you for pushing me to be better, and thank you for providing me with world-class opportunities and connections.

To my committee members, thank you for helping me to reach the best version of this work. Zoran Ninkov, who never stops telling me about the wonderful things I could do. John Kerekes, who has given me great advice and guidance in all things PhD. Santosh Kurinec, who inspires me to believe in myself and who has always had encouraging words for me. Shouleh Nikzad, who inspires me professionally, encourages me personally, and who helped me craft the proposal that won my fellowship.

Thank you to all the women in science who came before me. You showed me that gender has nothing to do with aptitude for discovery or capacity for innovation. I never accepted that being a woman would hinder any of my ambitions, because of examples like you. I look forward to a day when "women in science" isn't a special case to be encouraged or promoted. In the meantime, I will do my best to raise expectations and inspire the women who will come after me.

This work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program – Grant NNX13AO54H.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

As imaging science pushes towards the limits of physics, scientific detectors have come closer and closer to being perfect, noiseless observers. New frontiers become available for discovery and study with every reduction in detector noise. And with every new discovery comes the desire to make another, to delve deeper into the mysteries of the universe. Whether those discoveries lie on the edge of a distant galaxy or inside the cells of the human body, the goal is always the same: better measurements.

For imaging detectors, this goal translates to counting every photon from the scene being imaged. While counting a few photons may seem unimportant, each photon counted reduces the uncertainty about a potentially critical measurement: Is there life on that planet? Will that asteroid collide with Earth? Are those cells cancerous? Low light level imaging applications have signals that are on the order of average detector noise for each image, which makes signal detection difficult and time-consuming, unless the noise can be reduced. Photon-counting detectors generally eliminate read noise, noise associated with the estimation of the collected signal in a pixel, in order to achieve single photon resolution.

This dissertation will investigate the current approaches to photon-counting and the state-of-the-art performance of current technologies. The detectors will be evaluated based on a specific set of criteria for various imaging and spectroscopy applications. Both the current performance and projected future performance of each detector will be compared. The goal of this research is to provide a comprehensive analysis of the future of photon-counting in semiconductor-based detectors.

1.1 THE HISTORY OF SINGLE PHOTON COUNTING

Since the inception of solid-state detectors, imaging scientists have looked for ways to count single photons, the smallest measurement of light. Each photon is a tiny bundle of a very specific amount of energy, depending on the origin of the photon. Counting individual photons means resolving the number of photons emitted from a source, and may even include other information

such as position, distance, temperature, energy, polarization, etc. And the goal of more information in a shorter amount of time is achieved with photon counting.

In order to acquire the near-noiseless operation required for photon-counting, integrated circuits (ICs) are required. ICs are used in all modern photon detectors in the readout electronics, which rely on the field effect transistors (FETs) to act as switches. While the theory of the FET was invented by Lilienfeld in 1926, the solid state transistor was not invented until 1947 at Bell Laboratories by William Shockley, John Bardeen, and Walter Brattain, for which they won the 1956 Nobel Prize in physics (Brinkman, et al., 1997).

Figure 1 – This figure shows the first FET, invented by Bardeen, Shockley, and Brattain (Brinkman, et al., 1997).

The solid-state transistor was pursued as a way to replace the vacuum tube, which could not operate at the high frequencies desired at the time. In 1957, Texas Instruments fabricated the first integrated circuit using discrete wire interconnections, though the modern IC using thin-film metal interconnections was first produced by Fairchild Semiconductor in 1959. However, it was not until the early 1970s that the impurity levels in the materials were controlled enough to yield reliable devices (Brinkman, et al., 1997).

Solid state transistors paved the way for array-based solid state photodiodes. The charge-coupled device (CCD) was invented in 1970 by Willard Boyle and George Smith at Bell Laboratories as an analogue to a magnetic bubble device (for storing information). Astronomers quickly saw the potential of the CCD for scientific imaging: in 1973, workers at NASA's Jet Propulsion Laboratory (JPL) initiated a program to develop large area CCD arrays for space-based imaging instruments. JPL built a traveling CCD camera system, the first of its kind, to promote the program by its use at major observatories around the world. The CCD had a much higher sensitivity than film, which was the detection method at the time, such that the same measurements could be made 100x faster with the CCD. This not only allowed more data to be produced in the same amount of time, but also allowed imaging of astronomical objects that were invisible to the detection methods available up to that point (Janesick & Elliot, 1992).

The most ubiquitous method of measuring a single photon was external amplification of the signal using photomultiplier tubes (PMTs). PMTs were first invented in 1930 by L. A. Kubetsky, a Soviet-Russian physicist and engineer, to amplify weak photocurrents (Lubsandorzhiev, 2006). Figure 2 shows a rendering of Kubetsky's tube, the first PMT.

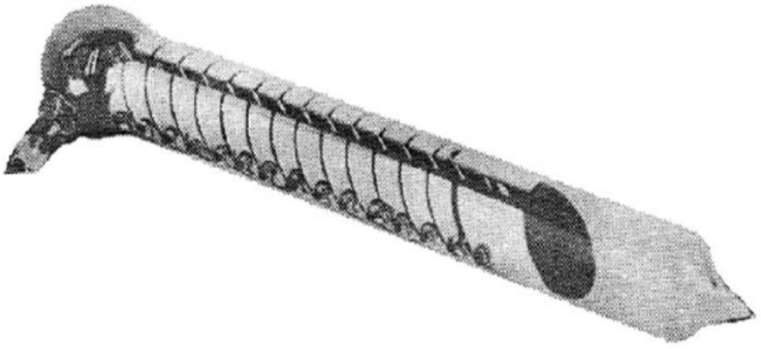

Figure 2 – This figure shows the first PMT, invented by Kubetsky (Lubsandorzhiev, 2006).

PMTs work by using Einstein's photo-electric effect: a photo-cathode (which converts photons into photo-electrons) is linked to a series of secondary electron emitters (called dynodes), which have successively higher voltages to create acceleration between units. A photo-electron from the photo-cathode reaches the first dynode through an electric field, accelerating the electron.

Upon impact, the electron ionizes more electrons, which are all accelerated to the next dynode through an electric field. This progression continues until the end of the PMT, where the total output gain is g^n, where g is the gain of the dynode stages (which is constant) and n is the number of dynodes. Later versions of the PMT included magnets for focusing the electrons. The total gain of these early devices was 10^3-10^4 (Lubsandorzhiev, 2006). The rest of the scientific community was not far behind the invention of the PMT, with other versions of the device being invented by RCA laboratories in 1936, among others. Later versions of the PMT incorporated simple detectors at the end of the tube to either measure average current or individual pulses of current from photons or other particles. This method resulted in calibration of the system to specific operation conditions for each new application (Darland, et al., 1979). This system was common well into the 1990s for both photon-counting and ranging applications (Donovan, et al., 1993), and continues to be used in niche applications even today.

Even though the CCD was invented in 1970, scientists were already in pursuit of measuring single photons with CCDs in the mid-1970s. Some first attempts included a mixture of PMT and CCD technology. The intensified CCD (ICCD) instrument used a photo-cathode to generate electrons from incoming photons, which were then focused via an electrostatic focus cone onto the CCD (Currie & Choisser, 1976). Figure 3 shows an example of an ICCD device.

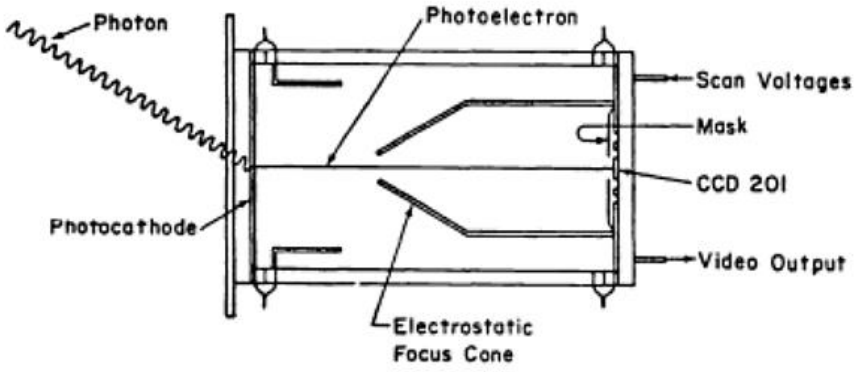

Figure 3 – This figure shows a schematic diagram of the ICCD device (Currie & Choisser, 1976).

The photo-electrons would hit the CCD and create a charge packet, which was then read out by a simple parallel-to-serial transfer system to an on-chip preamplifier. The system was read out during integration of the next frame (Currie & Choisser, 1976). PMTs were also used in conjunction with CCDs, though some difficulties associated with PMTs, such as fatigue and gain instability, became more significant as CCDs (or other detectors) became more sensitive (Coates, 1975).

This method of first converting incoming photons into photo-electrons via a photo-cathode, separate from the detector itself, continued as a common method well into the 1980s and was notably pursued by Hamamatsu Photonics (Tsuchiya, et al., 1985). They used a 3-stage microchannel plate (MCP) as the gain mechanism, which consisted of an array of narrow tubes that amplify electron signal through accelerated electron collision with the sidewalls. This implementation allowed for PMT-like amplification while maintaining spatial resolution. Instead of a CCD as the detector, the sensor was a position sensitive detector (PSD), which has no discrete spatial bins (like pixels), but instead uses surface resistance to measure continuous position data. The astronomy community also used MCPs as a means of external amplification in space-based missions (Siegmund, et al., 1986). Eventually, alternative approaches to signal enhancement were pursued, including the electron bombarded CCD (EBCCD) and electron bombarded CMOS (EBCMOS). Whereas ICCDs used phosphor screens, MCPs, and fiber optic components to retain the spatial information of the signal, EBCCDs and EBCMOS devices use a

photocathode to generate high-energy electrons from the incoming photon signal. The electrons that exit the photocathode are proximity-focused onto a thinned, back-side illuminated CCD or CMOS (compensated metal oxide semiconductor) array, and the electron is absorbed directly into the array. Electron-bombardment gain then amplifies the signal with nearly noiseless gain inside of the array and the signal is read out (Williams, Jr., et al., 1995).

But PMTs were not completely replaced by MCPs, as they still were used a variety of applications, including biological imaging of chemical reaction phot-emission (Seliger, 1980) and fluorescence modulation (Murray, et al., 1986).

Avalanche photodiodes also became a popular area of research in the 1960s and early 1970s, with both linear- and Geiger-mode functions being topics of interest. The motivation for this interest was the replacement of photo-multiplier tubes (PMTs), as modern experiments were likely to contain high electric fields and high magnetic fields (particularly in medical imaging), both of which degrade PMT performance (Jackson, et al., 2003). High PMT fabrication costs were also a factor; much of the internal architecture had to be handmade. Geiger-mode avalanche photodiodes (GM-APDs) could replace PMTs as both a signal amplifier for low flux imaging applications and as a means to detect single photons. The first silicon (solid state) single photon counting detectors were developed by Robert McIntyre at RCA Laboratories (a reach-through type of device) and Roland Haitz at Shockley research laboratory (a planar type of device) (Renker, 2006). Figure 4 shows diagrams of the two devices.

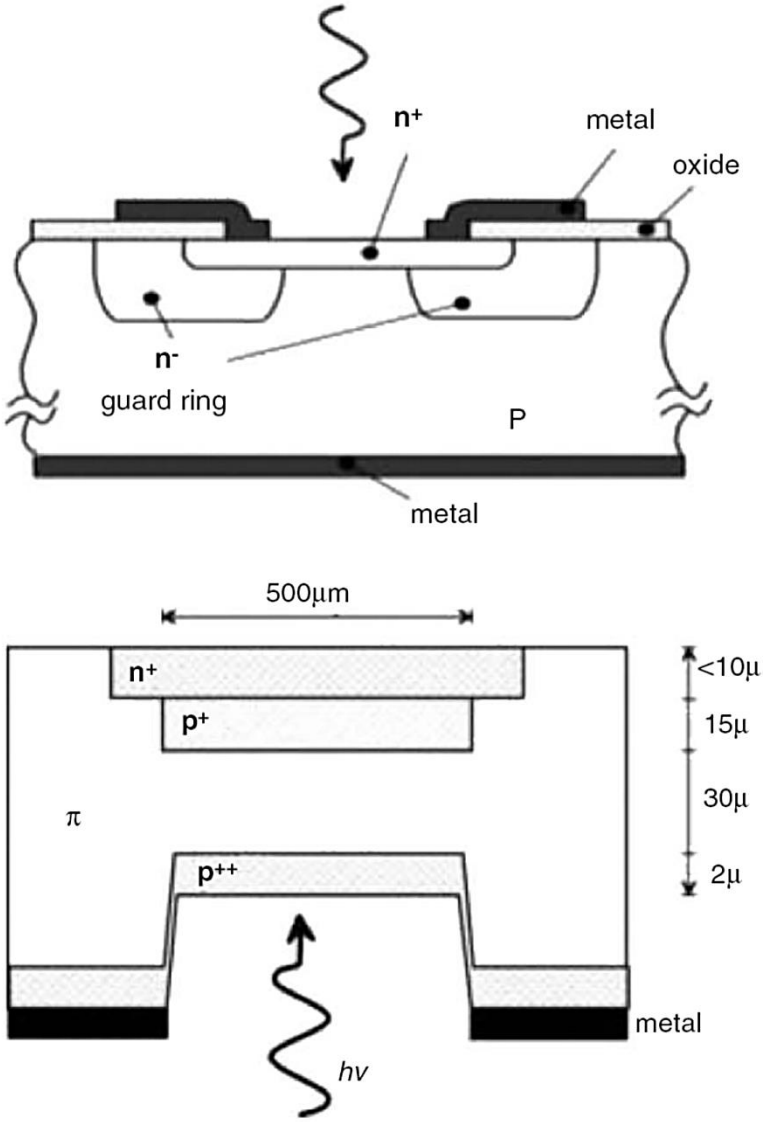

Figure 4 – This figure shows diagrams of single elements in the first two silicon single photon detectors. The top diagram is Haitz's planar device, and the bottom diagram is McIntyre's reach-through device (Renker, 2006).

The next generation of single photon-counting APDs came from Rockwell International Science Center in 1987 when they developed the solid state photomultiplier (SSPM). The device design utilized very high donor concentrations, which allowed the device to be sensitive even into the near IR (Petroff & Stapelbroek, 1989). The structure was later modified to have less sensitivity in the longer wavelengths and exists now as the visible light photon counter (VLPC). One of the

last radical improvements to the APD design occurred in Russia around 1990, called the metal-resistor-semiconductor (MRS) APD. A thin metal layer coats either SiC or Si_xO_y (the resistor layer), allowing for a local reduction of the electric field, which in turn limits the magnitude of the diode breakdown. A smaller breakdown magnitude leads to faster quenching, which in turn leads to faster cycle times (Renker, 2006).

By the mid-1990s, silicon APDs were the subject of promising research in the area of photon-counting, due in no small part to their commercial availability. By 1993, GM-APDs could be purchased with photon detection efficiencies (PDEs) near 70% near 600 nm, dark count rates (DCR) of 1 kHz, and relatively low operating voltages of 200-600 V (low compared to the PMTs, their main competitor). The notable Slik™ device by Perkin Elmer, available around 1990, was designed specifically for high PDE. Scientists also began to investigate the pros and cons of different quenching schemes (the mechanism by which the internal gain pulse is stopped) (Dautet, et al., 1993). These improvements in design and operation opened the door for more sensitive applications, including time-correlated single photon counting (TCSPC) for time of flight ranging, time-domain reflectometry, quantum cryptography (Hiskett, et al., 2000), and photon correlation experiments in quantum physics (Noh, et al., 1991). The 1990s also saw significant development in non-silicon APDs, including germanium (Owens, et al., 1994) and variations on InGaAs and InP such as InGaAs/InP (Ribordy, et al., 1998) and InGaAsP/InP (McIntosh, et al., 2002). Devices based on these semiconductor materials started to become available commercially for extended wavelength sensitivities into the infrared (IR) in the mid-1990s. Materials that are sensitive at longer wavelengths are useful for direct imaging of IR phenomena, but are also useful for applications in which the signal being detected must also be generated: communications, quantum computing, quantum key distribution, etc. Some materials even offer specific advantages, such as noiseless gain HgCdTe in linear-mode (Kinch, et al., 2004) or the fast reset times of InGaAs in Geiger-mode (Dixon, et al., 2009).

The 1990s also saw the introduction of superconducting materials for photon detection, though mainly in x-ray applications. Previously, superconducting detectors (most commonly, superconducting tunnel junction detectors) were used primarily for detection of high-energy

particles. By the early 2000s, superconducting materials and detectors were being used with great success in a variety of photon-counting applications, including quantum communications (Robinson, et al., 2006). Microwave kinetic inductance detectors (MKIDs) were a breakthrough in superconducting detectors, and have since been the subject of much research (see section 1.3.2). They are easy to fabricate, requiring only a few layers of material and a very simple readout system, and easy to mosaic into larger arrays. Initially, they were used mostly for astronomy applications, given their wide range of wavelength sensitivity (from x-ray to sub-millimeter photons), though other disciplines have recently begun investigating their potential. MKIDs are particularly promising for quantum optics and communications (Gao, et al., 2012).

Scientists also began experimenting with on-chip gain in CCDs in the early 2000s, leading to the electron-multiplying CCD (EMCCD). While reading out the CCD while taking many samples, as in the CCD Skipper readout (Janesick, et al., 1990), can reduce the read noise to sub-electron levels, this comes at the cost of a very slow frame rate. In EMCCDs, a gain register at the output of the device multiplied the standard CCD signal by a pre-determined gain, making the read noise of the device negligible and allowing for fast readouts. There were some drawbacks, however, as the gain was stochastic in nature. The more photo-electrons were multiplied at the gain stage, the more uncertain the measurement became. Therefore, it was important to balance the gain value and exposure settings with the needs of the experiment (Basden, et al., 2003). However, APDs still had a number of advantages, such as a fast reset for resolving higher flux rates for single photons Geiger-mode operation. And with the APDs in a linear-mode (limited gain), distinction between different numbers of simultaneous photons was also possible (Eraerds, et al., 2007). Distinguishing between different energies is also possible, as long as the photon has sufficient energy to ionize more than one electron, e.g. x-rays. This capability has a wide range of applications, including medical imaging and spectroscopy, synchrotron and x-ray tube spectra, radioactive materials analysis, and astronomy. Certain computational considerations need to be used in the analysis, however, since the relationship between photon energy and number of electrons is non-linear and depends on the detector material (Sievers, et al., 2012).

More recently, advancements in CMOS APS direct readout (DRO) devices have led to sCMOS (scientific CMOS) device architecture, which bypasses traditional trade-offs between readout rate and read noise (Andor, 2012). By using new CMOS design technology, sCMOS can achieve frame rates of 100 frames per second (10 ms readout times) with just 1.3 electrons read noise, and at 30 frames per second with just 1 electron read noise. The devices do sacrifice some quantum efficiency (roughly 57%) due to the monolithic device architecture. This could conceivably be increased with back-illumination processing and hybridization (as is done with current state-of-the-art CCD devices). These devices also have high dynamic range (30,000:1 at 30 frames per second) and can be fabricated in large array formats.

1.2 Applications of Photon Counting

Generally, photon counting is useful in low-light-level imaging scenarios. In these photon-starved applications, lower noise can make a significant impact on required exposure time and image quality. There are numerous reasons why an application is photon-starved, such as a faint object or the necessity of fast exposures, but the need for low noise and single-photon resolution is universal.

Astronomy and adaptive optics (AO) applications are discussed here for their relevance to semiconductor-based detectors that measure intensity. Other applications, such as laser ranging and quantum computing, rely on measurements of arrival time to calculate either time of flight (distance) or frequency of signal on very short scales. Often, these applications rely on syncing the detector with the source pulses. This represents a different detection paradigm, and those applications are not discussed here. For superconductors, medical imaging applications are relevant due to the need for precise energy resolution. Because the focus of this dissertation is semiconductor-based detectors, the details of superconducting applications left to Appendix B:.

1.2.1 *Astronomy*

There are many demanding applications in the field of astronomical spectroscopy and imaging. Two examples are spectroscopy and imaging of exoplanets (both direct and transit) and pulsars. Both applications require fast exposures that preserve signal-to-noise ratio (SNR).

Exoplanets (also referred to as extrasolar planets) are planets that orbit stars other than the Sun. There are three common approaches for the detection of exoplanets: the Doppler method, transit photometry, and direct imaging. In the decade that followed the detection of the first exoplanet, the Doppler method was the most common method of detection. This method measures small sinusoidal changes in velocity of a star as it orbits the center of the star-planet system. These measurements require very small changes. For example, a Jupiter-like planet requires measuring a long-term wavelength shift in a stellar absorption line of 0.1% of its width. The Doppler method of exoplanet detection requires a high dispersion spectrograph, high SNR of the spectral data, and extremely accurate wavelength calibration. The Doppler method has yielded the most confirmed exoplanets, but the Kepler mission's transit photometry method has recently yielded the most candidates (Impey, 2013).

The first exoplanet orbiting a sun-like star was discovered in 1995 using transit photometry (Mayor & Queloz, 1995), which works by observing a star while a planet is in transit in front of it. By taking spectroscopic data before, during, and after the transit, scientists can calculate the degree to which the planet blocks some wavelengths. The larger the planet, the more light is blocked. Figure 5 shows an example of transit photometry data (Sing, et al., 2009). The 3% decrease in signal is for a very large planet. A Jupiter-like planet will induce a photometric variation of just 1%, while Earth-like planets have an even smaller effect. Generally, transit times are on the order of hours for short-period planets (Santos, 2008).

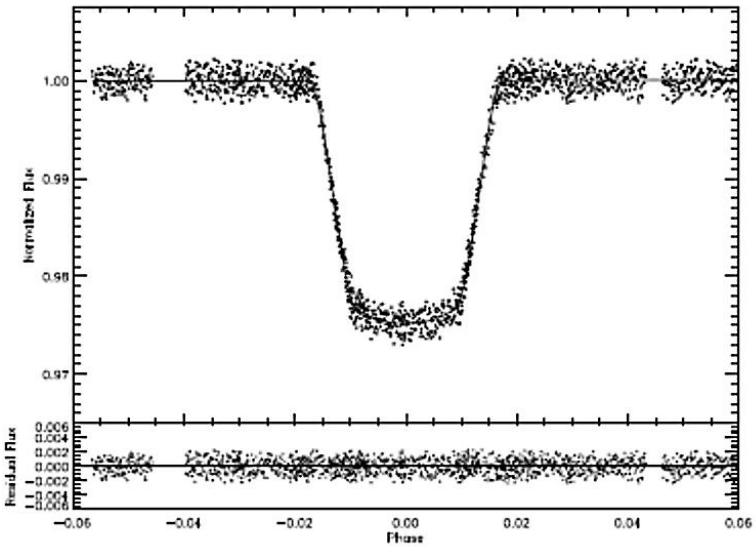

Figure 5 – An example of relative flux signal vs phase for a transit photometry measurement (Sing, et al., 2009). The 3% decrease in flux at zero phase corresponds to the transit of the exoplanet across the star.

NASA's Kepler planet detection mission alone has detected hundreds of candidate planets via transit photometry, making baseline measurements such as planet size, mass, and density, as well as average temperature and distance from the planets' stars (Borucki, et al., 2010). The NICMOS instrument on the Hubble Space Telescope (HST) has also been used for transit photometry of candidate exoplanets. Some observations focus on planet composition (e.g., metal enrichment) (Carter, et al., 2009), while others are in search of evidence for water in the atmosphere (Sing, et al., 2009). The limitation of transit photometry as a means of exoplanet detection is that the orbital axis of the planet must be closely perpendicular to the detector's line of sight. This is reasonably common (10% probability) for a Jupiter-like planet in a very short orbit (~3 days), but the probability of a compatible orbit for an Earth-like planet (both size and orbit) is only 0.5%. Planets with orbital paths that do not meet the transit requirement must be imaged directly (Santos, 2008).

Direct imaging observes the planet during its orbit, detecting the light reflected off of the planet from the star while not in transit. In this case, the star may be occulted, or blocked out, as on NASA's Terrestrial Planet Finder coronagraph, or TPF-C (Ford, et al., 2004). With the starlight

blocked, the exposure can be optimized for the exoplanet signal. Direct imaging is also done without a coronagraph, as on the MOST satellite (Rowe, et al., 2006), or the ground-based Keck and Gemini telescopes (Marois, et al., 2008). Characterization of exoplanets requires multiple measurements. The distance between the star and planet candidate is important for determining the likelihood of planet formation (Carson, et al., 2013), and albedo (average reflection coefficient) measurements can be taken at multiple wavelengths to construct a detailed model of the atmosphere. With this model, one can identify biomarkers, or atmospheric characteristics that suggest a biological origin, marking a planet for future long-term study. Large arrays will become more desirable as the need for increased spatial resolution grows (Schneider, et al., 2010).

X-ray detection instruments, such as the satellite Chandra, are often used to study pulsars (Ng, et al., 2005) and their nebulas (Gaensler, et al., 2002), as well as stellar flares (Mitra-Kraev, et al., 2013). Pulsars are rotating neutron stars that emit a beam of electromagnetic radiation, and are believed to be formed in supernova explosions. These stars are very dense and have short rotational periods – as the star rotates, the electromagnetic beam can be seen, causing a pulsing behavior in the signal (Radhakrishnan & Cooke, 1969). Figure 6 shows an image and relative spectral flux of the Crab pulsar, one of the first pulsars to be discovered and studied (Buhler & Blandford, 2014).

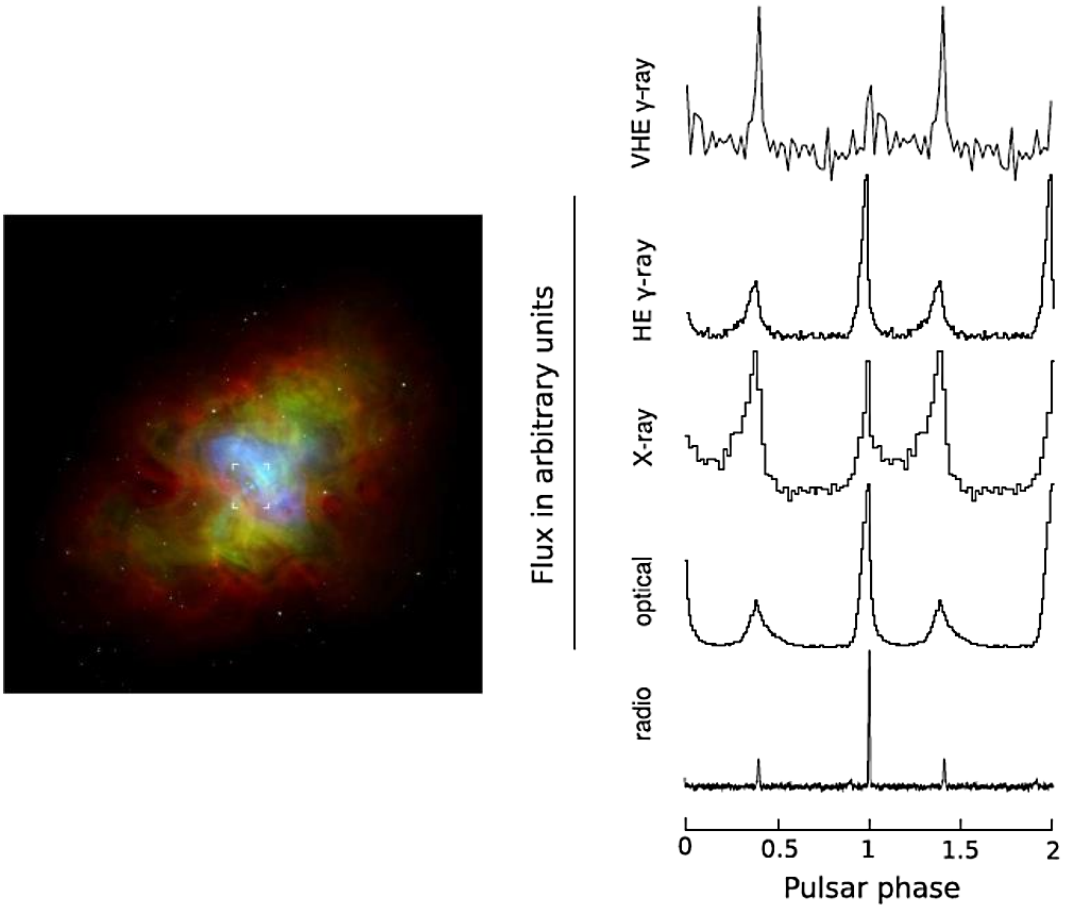

Figure 6 – A false-color image (left) and spectral signal phase diagram of the Crab Pulsar and nebula (right) are shown. In the false-color image, blue is x-rays, green is visible, and red is radio signal (Buhler & Blandford, 2014).

The longest period ever recorded for a pulsar is 8.5 s, requiring sampling periods on the order of milliseconds to reliably capture the fluctuation of pulsar signals. Figure 7 shows the distribution of known pulsars as of 1999, plotted by pulsar period P (in seconds) and the surface magnetic dipole field strength B_s (in Gauss) (Young, et al., 1999).

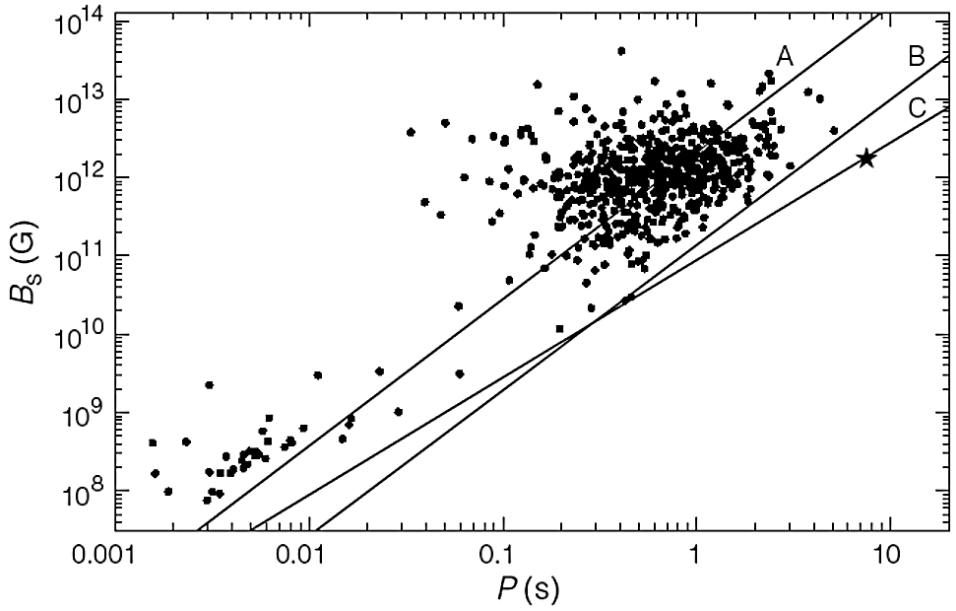

Figure 7 – The surface magnetic dipole field strength B_s vs the pulsar period P for a large population of known pulsars is shown. Most pulsars have a period less than 1 s. The lines labeled A, B, and C represent competing models defining the minimum magnetic field for a given pulsar period (Young, et al., 1999).

Pulsar observations are most often limited to x-rays and radio waves, but and some key features can also be measured in optical/IR signals (or the lack thereof) (Gotthelf, et al., 2004).

1.2.2 *ADAPTIVE OPTICS*

Adaptive optics uses wavefront sensing to correct for wavefront aberrations. Originally, AO was developed in the late 1960s to allow the US Air Force to image satellites from ground-based telescopes to correct for atmospheric aberrations (Platt & Shack, 2001). Over the course of an observation period, even while accounting for the rotation of the Earth, an object's signal will shift in the focal plane due to interaction with the Earth's atmosphere. AO is not limited to applications with only atmospheric aberrations, however, and has also been used to characterize aberrations of the human eye, the uniformity of laser beams, and optical system alignment (Neal, et al., 2002). Wavefront sensing systems can measure the tilt or curvature of a wavefront using optical components and a detector. In a standard setup, a beam splitter divides the signal into two portions: one part of the signal goes to the wavefront sensing detector and the other to the

imaging detector. In order to correct aberrant waveforms, the information collected by the wavefront-sensing detector feeds back into the system controlling the imaging optical components. These optical components manipulate the image signal so that the scene remains stationary on the focal plane of the imaging detector.

The most common approach to wavefront sensing is the Shack-Hartmann wavefront sensor. This sensor is based on the Hartmann screen test, which was invented by the distinguished astrophysicist Johannes Hartmann to evaluate the quality of an 80 cm refracting telescope at Potsdam. While the screen test is still used today, applications requiring wavefront characterization in low-light conditions necessitated the evolution of the concept. In the early 1970s, Roland Shack and his graduate student Ben Platt expanded on Hartmann's system, introducing a lenslet array to maximize the signal collection and focus the light at a predictable point behind the optics. After the initial development of the sensor, the military also used the technology to adaptively correct the distortions of mirrors used in high-energy laser systems, which would become deformed under the high heat load (Schwiegerling & Neal, 2005). Figure 8 shows an example of a Shack-Hartmann wavefront sensor (Platt & Shack, 2001). By measuring the displacement of the focal spots, the wavefront curvature could be calculated.

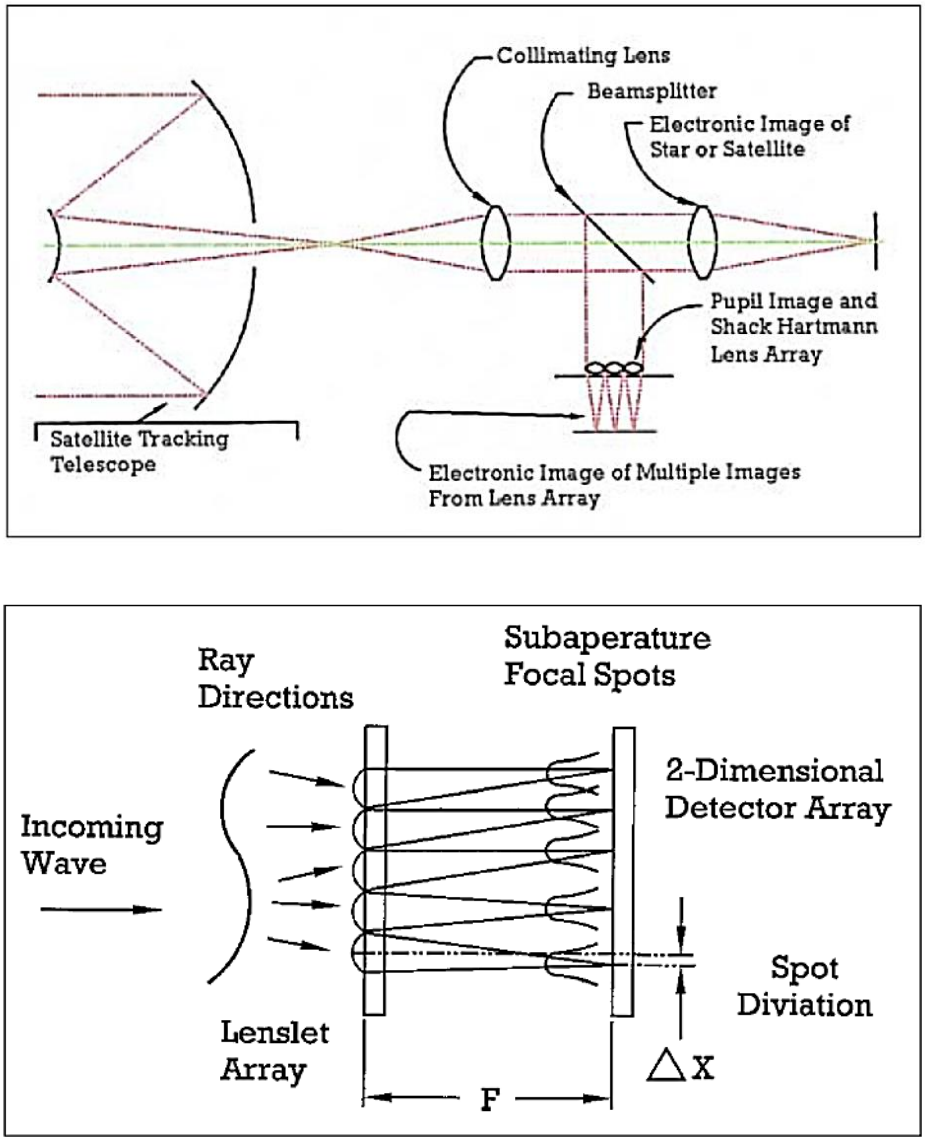

Figure 8 – An optical layout of the original Shack-Hartmann wavefront sensing system for AO is shown (Platt & Shack, 2001). The top figure shows the path of the incoming beam and the bottom figure shows a magnified view of the lens array.

In the mid-1980s, Josef Bille at the University of Heidelberg was the first to use the Shack-Hartmann technique on the human eye. It was first used to characterize the surface of the cornea, but later was applied to aberrations of the entire eye. These experiments led directly to the modern-day treatment of visual aberrations with ecximer laser sculpting (Schwiegerling & Neal,

2005). Zernike polynomials, coefficients used to characterize visual aberrations, can be precisely measured to high order (complicated aberrations, such as astigmatism and spherical aberration) to better inform lens prescriptions and vision correction surgeries (Liang, et al., 1994). In some cases, patients can have their vision improved beyond 20/20 (supernormal vision). Advanced AO systems have also allowed high-quality, in-vivo imaging of the retina (Liang, et al., 1997).

While the concept of AO was appealing for astronomy applications, the sensor technology was behind what was needed to correct the wavefront in real-time. It wasn't until the 1980s that AO was feasible, with the advent of the CCD camera, and development lagged behind that of the military due to scarce funding and classified research (Platt & Shack, 2001). The first actively-corrected astronomical telescope was the 3.6 m telescope at the European Southern Observatory (ESO), which saw first light in 1989 (Schwiegerling & Neal, 2005).

AO applications benefit from detectors with many pixels, high frame rates, and low noise. More pixels means finer spatial sampling; high frame rates mean finer temporal sampling; lower noise means more precise optical adjustments. Unfortunately for CCD imagers (the main alternative for AO applications), these three traits are generally mutually exclusive. A higher frame rate decreases the image latency but increases the read noise of a CCD. There are alternative methods, e.g. a multiple readout architecture (small sections of the larger array are read out individually and then combined), but this increases the complexity of a system as well as its power usage. In recent years, the Shack-Hartmann AO system has been applied to 2-D arrays of GM-APDs with promising success (Aull, et al., 2010; Aull, et al., 2015). Each lens is paired with a 4-element quad cell, capable of calculating the centroid of the focal spot based on the disparity in signal between each element. Multiple simulations of APD and CCD detectors in AO systems have shown that there is currently little difference in performance (Craven-Bartle, et al., 2000; Aull, et al., 2015).

Figure 9 shows the displacement estimation error for the centroid of a focal spot as a function of the number of photons detected per sample for a CCD and a GM-APD with various levels of crosstalk (Aull, et al., 2015). The detectors are assumed to be quad-cell devices.

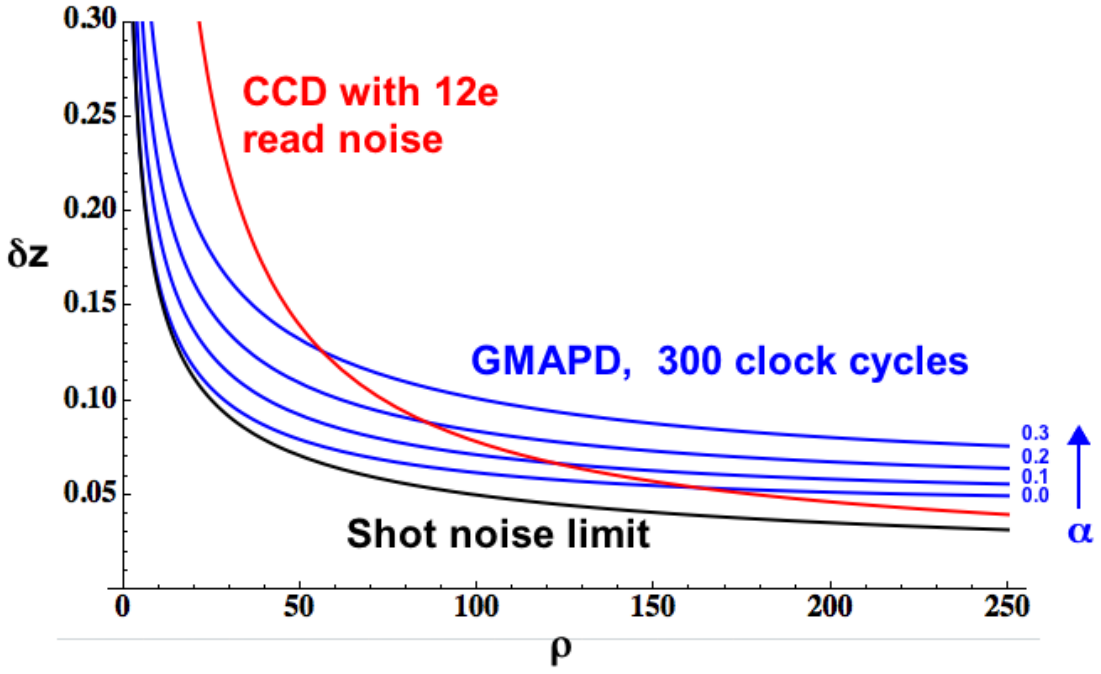

Figure 9 – This figure shows the displacement estimation error, δz, of the focal spot centroid as a function of the number of photons detected per integration time. α is the crosstalk probability for each blue curve. The limiting factor of the GM-APD curve with no crosstalk probability is the dead time of the device, and the limiting factor for the CCD is the read noise (12 e^-). Both detectors reach the Shot noise limit in the ideal case: no read noise in the CCD and no crosstalk or dead time in the GM-APD (Aull, et al., 2015).

If APD technology can reduce dark count noise (its current disadvantage), then APD could easily overtake CCD performance in AO systems.

1.3 Approaches to Photon Counting

Generally, photon counting is most relevant for low-light-level applications. In these photon-starved applications, lower noise can make a significant impact on required exposure time and SNR. There are numerous reasons why an application might be photon-starved, such as a faint object or the necessity of fast exposures, but the need for low noise and single-photon resolution is universal in these applications. Applications that require advances in photon counting (excluding time-correlated methods that require a free-running detection system) are described below.

There are two main approaches to photon counting detectors: semiconductors and superconductors. The main focus of this research is semiconductor-based detectors, given that semiconductor technology is more advanced for imaging applications. Superconductor-based detectors will be discussed and presented as a viable alternative for future applications, given some key improvements.

Semiconductor detectors use materials that have been studied for decades, such as Si and HgCdTe. Other materials, such as InGaAs, Ge, and InP, have unique advantages in the fabrication process and in material properties. Generally, semiconductor-based approaches to photon counting use the internal avalanche gain mechanism, which amplifies the signal of a single electron or hole many times over (gains in excess of 10^6 are not uncommon). Semiconductor detector arrays are fabricated on wafers, which are thin disks of semiconductor material sliced from an ingot. Some devices are built into the wafer material, while others are built on top of the wafer in a separate material grown epitaxially. Usually, more than one device can be fabricated on a single wafer, but large arrays are limited by wafer size. State-of-the-art silicon wafers are currently 300 mm, but the fabrication industry will transition to 450 mm in the near future (Watanabe & Kramer, 2006). Wafer size is smaller (~70 cm) for more expensive detector materials like HgCdTe (which generally require CdZnTe wafers on which to grow the HgCdTe), though efforts to grow HgCdTe on Si wafers have had promising success on wafers up to 150 mm (Reddy, et al., 2008). It is important to note that the detector price increases exponentially with increasing device size as fewer detector arrays can be fit on a single wafer. The cost of processing the wafer (which does not depend on the number of detectors per wafer) is split across fewer detectors. Also, if the number of defects per wafer remains constant, the number of defects per detector increases. To produce a detector with very few defects, more wafers must be processed, and the cost of those wafers is also added to the cost of the large detector. Pixel size is limited only by microlithography capabilities, though the required size is often set by system constraints. For example, if Nyquist sampling is required, then approximately two pixels must be sized to cover a resolution element. Common pixel sizes range from 10 μm to 30 μm.

Superconductor detectors have begun to emerge as promising alternatives for imaging and spectroscopy applications. Superconducting materials are usually chosen for the critical transition temperature (T_c), usually on the order of 1 K, though materials with lower T_c values are being investigated for future instruments. Some superconductor detectors use an LC circuit configuration and measure photon detection by way of a change in the circuit's resonance frequency, which is affected by a change in resistivity induced by absorbed energy. Others measure changes in current or voltage draw (indicating a change in resistivity). The current limitations on superconductor-based detector technologies are scalability, pixel size, and cooling requirements. While semiconductor-based detectors do not have inherent, noiseless energy resolution, they are very scalable, can have pixel sizes on the order of tens of microns, and require only modest cooling in most applications (on the order of -20˚ C in many instances).

1.3.1 *SEMICONDUCTOR-BASED DETECTORS*

There are many different implementations of semiconductor-based imaging detectors, and they can be divided into three categories: devices with no gain, devices with non-avalanche gain, and devices with avalanche gain.

The first category, devices with no gain, includes CCDs and CMOS active pixel sensor (APS) devices. Modern CCDs generally use buried channel MOS capacitors to store charge collected during an exposure. Reading out the device requires physically moving the charge through the pixels, which is called a destructive read. CMOS APS devices use a photodiode to store charge in each pixel, which can be read out by measuring the change in voltage during an exposure, which is a non-destructive read.

Detectors with non-avalanche gain, such as microchannel plates, ICCDs, and electron bombarded CCDs, are not included in this project. Generally, these devices use the photoelectric effect and gain due to electron bombardment, where a photon with high energy generates more than one electron-hole pair.

Avalanche gain can take place either in a multiplication register or in the pixel. EMCCDs can be operated in either linear mode or photon counting mode and have the same pixel architecture as a

standard CCD. The difference is in a special gain register that amplifies the signal from the pixels during the readout (Daigle, et al., 2004). Linear mode operation measures the analog signal to estimate the number of photons that were absorbed. In photon counting mode, on the other hand, they are read out very quickly and a threshold circuit is used to decide whether or not a photon was absorbed during the very short exposure. The short exposures are repeated many times to estimate the probability of the arrival of one or more photons in one short exposure. This probability is used to calculate the average number of photons per second per pixel. Linear-mode avalanche photodiodes (LM-APDs) and GM-APDs both have in-pixel avalanche multiplication, but use it differently. LM-APDs use low gain (<100) and measure the analog signal from each pixel to estimate the number of photons that were absorbed. GM-APDs use very high gain and a thresholding circuit to determine whether or not a photon was absorbed during a very short exposure. Much like the EMCCD, these short exposures are repeated many times to estimate the probability of the arrival of one or more photons in one short exposure. This probability is used to calculate the average number of photons per second per pixel. Although these approaches to semiconductor-based imaging detectors are fundamentally different, there are a few fundamental concepts that they share.

1.3.1.1 The PN Junction Diode

A diode is a device with a junction of two oppositely doped semiconductor materials. A material is considered doped when impurities are intentionally added to the crystal structure. For instance, doping silicon with boron will cause the material to become p-type, or have a majority of positive carriers (holes) in the material. If silicon were doped with phosphorus instead, the material would become n-type, or have a majority of negative carriers (electrons). These extra carriers enable higher conductivity than in pure silicon. Figure 10 shows a band diagram, which illustrates the available energy levels for a carrier in a crystal structure with no pn junction (Pierret, 1996).

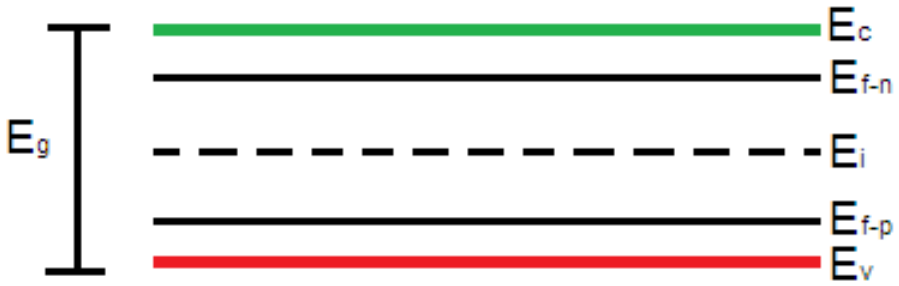

Figure 10 – A simple energy band diagram for a semiconductor is shown.

E_c refers to the energy level of the conduction band (where electrons can freely move within the lattice), and E_v represents the energy of the valence band where electrons reside in their bound state. E_g is the band gap energy, or the amount of energy needed for an electron to move to the conduction band from the valence band, and is a constant for a semiconducting material. For silicon, E_g is 1.12 eV at 300 K. E_i (intrinsic energy level) refers to the energy level that has a 50% probability of being filled in an intrinsic semiconductor, which is half the band gap (0.56 eV at room temperature for silicon). The probability that a carrier will occupy a certain energy state is a function of the Fermi energy level ($E_{f\text{-}n}$ for n-type dopants and $E_{f\text{-}p}$ for p-type dopants) and the temperature of the material. E_i is often shown in the band diagram as a reference for doped materials. For instance, if the Fermi level is above E_i, the material is negatively doped. The Fermi function describes this probability (see Eq. 1) and the expressions for the corresponding energy levels are shown in the Fermi level expressions (Eq. 2 and Eq. 3) (Pierret, 1996).

$$f(E) = \frac{1}{1 + e^{\frac{E - E_f}{kT}}} \qquad \text{Eq. 1}$$

f(E) denotes the probability (under equilibrium conditions) that an electron will occupy an available state of energy E. k is the Boltzmann constant and T is the temperature of the material in Kelvin. E_f is the Fermi level, which can be calculated as in Eq. 2 and Eq. 3 (Pierret, 1996).

$$E_{f-n} = E_i + kTln\left(\frac{N_D}{n_i}\right) \quad \text{Eq. 2}$$

$$E_{f-p} = E_i - kTln\left(\frac{N_A}{n_i}\right) \quad \text{Eq. 3}$$

Eq. 2 applies to n-type (donor) dopants, with N_D being the number of dopant atoms per unit volume. Eq. 3 applies to p-type (acceptor) dopants, with N_A having the same units as N_D. n_i is the intrinsic carrier concentration for the semiconductor (~1e10 cm^{-3} for silicon). For most applications, only one type of dopant is used in a particular volume of the semiconductor (either p-type or n-type). A mode of doping exists called compensated doping, which includes both n- and p-type dopants in the same region, but imaging detectors do not generally utilize these doping schemes (Pierret, 1996).

The carrier densities n and p (of electrons and holes, respectively), shown in Eq. 4 and Eq. 5, represent the number of carriers per unit volume. These values are dependent on the material's intrinsic characteristics, the Fermi level, and the temperature.

$$n = n_i e^{(E_f - E_i)/kT} \quad \text{Eq. 4}$$

$$p = n_i e^{(E_i - E_f)/kT} \quad \text{Eq. 5}$$

Eq. 6, called the law of mass action, directly relates n and p to the intrinsic carrier concentration of the material. Using Eq. 4 – Eq. 5, one can calculate n, p, or E_f from only one of the variables in the group.

$$\boldsymbol{np = n_i^2} \qquad \text{Eq. 6}$$

Current is defined as the net motion of carriers through the semiconductor. This does not refer to one carrier traveling the full length of a device, but to the average carrier motion that favors one direction. For an electron to be free to move about the lattice, it must gain enough energy to exist in the conduction band. When the electron makes this jump, it leaves behind a hole, which is a positive carrier. Eventually, the electron will encounter another hole and recombine. The mean distance that the carriers can travel before recombining is the diffusion length, and the mean time the carrier takes to travel that length is the carrier lifetime.

The semiconductor physics of a common structure called the pn junction include the basic relationships already discussed. As its name indicates, a pn junction is a p-type layer adjacent to an n-type layer. The majority carriers (electrons in an n-type material and holes in a p-type material) on one side of the junction diffuse to the other side, giving rise to space charge layers on each side of the junction; the resulting electric field creates a built-in potential barrier to the further diffusion of carriers. This layer is also called a depletion region because the carrier concentrations in this region are negligible compared to the majority carrier concentrations outside of the layer. Eq. 7 gives an expression for the width of the depletion region (Pierret, 1996).

$$\boldsymbol{W = \sqrt{\frac{2\epsilon_r\epsilon_0}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\left(V_{bi-}V_{app}\right)}} \qquad \text{Eq. 7}$$

V_{app} is the voltage applied to the junction, and V_{bi} is the built-in potential of the junction, described in Eq. 8 (and seen in Figure 11 as $q\emptyset_i$).

$$V_{bi} = \frac{kT}{q} ln\left(\frac{N_D N_A}{{n_i}^2}\right) \quad \text{Eq. 8}$$

One side of the junction is usually much more heavily doped than the other. In that case, the larger value will fall out of the dopant term in Eq. 7, revealing a dependence on the dopant concentration on the lightly doped side (Pierret, 1996).

The band structure across the depletion width changes for a pn junction as compared to a bulk material with no junction as in Figure 10. Figure 11 shows the altered band structure for a pn junction with no applied bias.

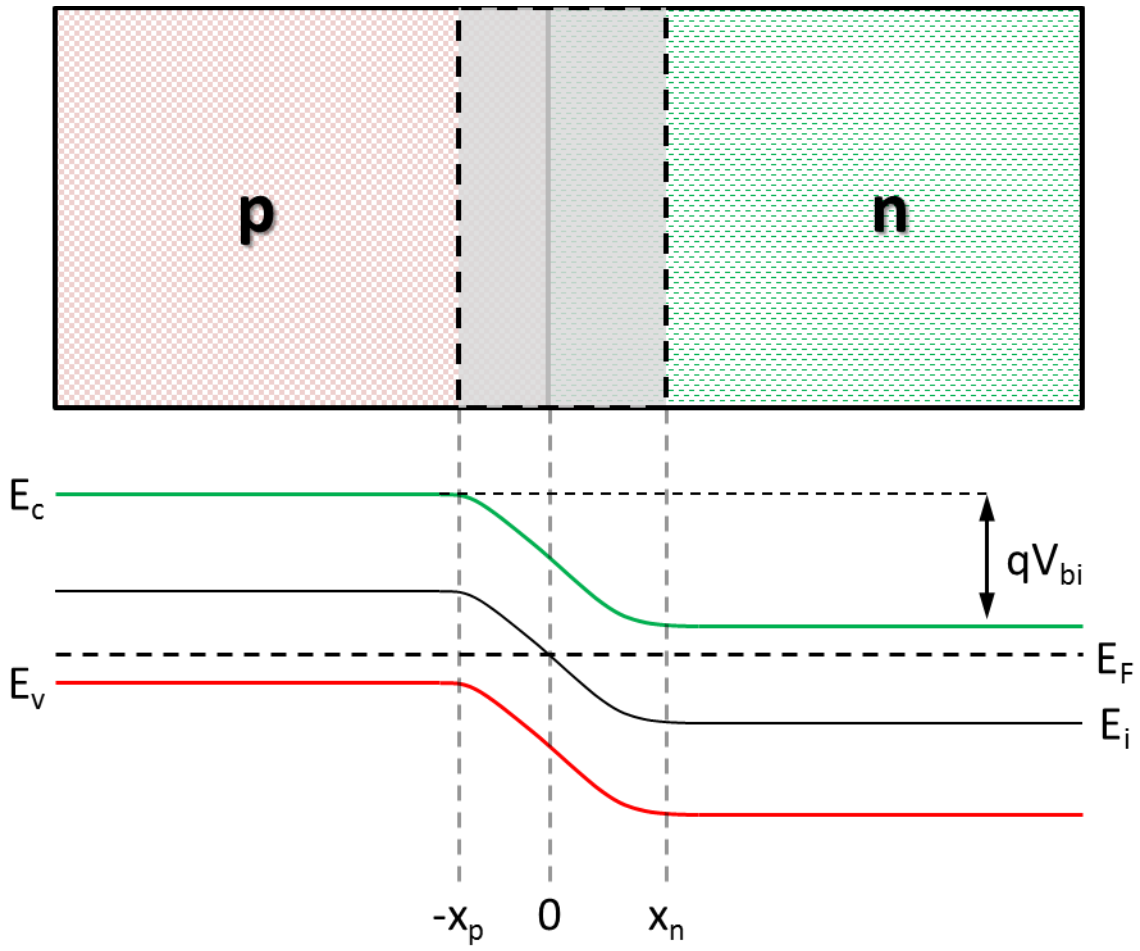

Figure 11 – A diagram of a standard PN junction is shown.

In Figure 11, x_n and $-x_p$ represent the bounds of the depletion width on the n- and p-sides of the junction, respectively, and the zero point is the physical location of the pn junction. qV_{bi} is the potential difference due to charge distribution across the depletion region.

Diffusion current is the movement of carriers due to a carrier density or thermal gradient, and is independent of applied bias. Drift current is the movement of carriers due to an applied bias. Electrons from the p-type region flow easily from left to right as diffusion current), but flowing from right to left requires more energy; the inverse is true for holes. A positive voltage applied to the p-side of the junction in excess of the built-in voltage V_{bi} eliminates the potential barrier for

drift current, turning the device on (Pierret, 1996). Eq. 9 describes the current flowing through a diode.

$$\boldsymbol{I = I_0\left(e^{\frac{qV_{app}}{kT}} - 1\right)} \qquad \text{Eq. 9}$$

I_0 is the leakage or dark current associated with the reverse drifting of carriers for any device, k is Boltzmann's constant, T is the temperature of the device, and V_{app} is the applied voltage. It can be obtained by I-V measurements at a large reverse bias (Pierret, 1996).

In a regular, forward-biased diode, a positive voltage across the junction removes the potential barrier and free carriers will drift (current due to an electric field) across the depletion width. This is not the case for a photodiode, which operates in reverse bias (increasing the potential barrier for drift current). The only significant carrier movement is dark current (diffusion current due to a density or thermal gradient, independent of applied bias), which is usually very small (~1 nA) because it is limited by the number of minority carriers available. When reverse-biased, the depletion area behaves like a capacitor storing charge with capacitance defined by Eq. 10.

$$\boldsymbol{C_j = \frac{A_j}{2}\sqrt{\frac{2q\epsilon_r\epsilon_0}{V_{bi} - V_{app}}N_B}} \qquad \text{Eq. 10}$$

N_B is the smaller of the two doping levels and A_j is the cross-sectional area of the junction. This equation has a built-in calculation of the depletion width. If the depletion width at the set voltage is known, however (or if the voltage indicates maximum depletion width), then the equation reverts to the standard parallel plate calculation for a capacitor with silicon as the dielectric layer (Pierret, 1996).

Photodiodes operate in reverse bias, increasing the potential barrier for drift current. The only significant carrier movement under reverse bias is diffusion current, which is usually very small

because it is limited by the number of minority carriers available. In photon detection applications, the diffusion current under reverse bias is referred to as dark current. While the diode is in this reverse-biased state, a photon absorbed in the device would excite a carrier to the conduction band, generating an electron-hole pair. Each carrier would then travel to opposite sides of the junction: the electron goes to the n-side and the hole goes to the p-side of the depletion region. This alters the total charge contained in the junction and the capacitance value, which modifies the voltage across the junction (Pierret, 1996).

1.3.1.2 Avalanche Gain

Avalanche gain is gain due to impact ionization in semiconductors. In a strong electric field, a carrier may have sufficient energy to dislodge another carrier, creating an electron-hole pair. The newly freed carriers immediately accelerate and impact other carriers, dislodging more electron-hole pairs. This causes an avalanche of charge to propagate through the lattice. (Pierret, 1996). The gain is the number of carriers generated during an avalanche by a single initiating carrier. Figure 12 shows the band diagram of a pn junction with avalanching carriers. Note the large change in the reverse bias as compared to the built-in voltage in Figure 11.

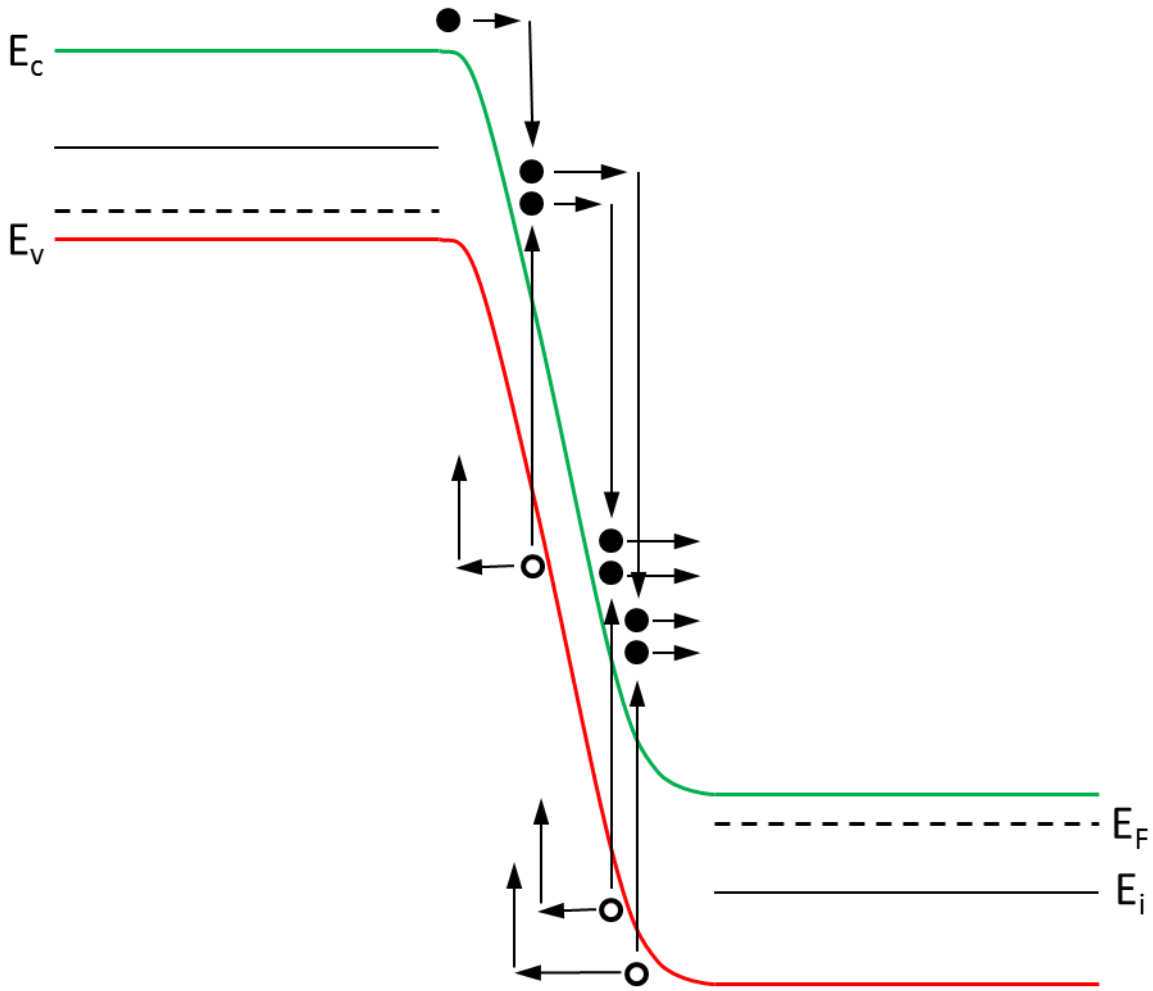

Figure 12 – A diagram of a pn junction with avalanche gain is shown.

Once an avalanche has begun, a competition develops between the rate at which electron-hole pairs are generated and the rate at which they are collected at the device terminals. At biases below some breakdown voltage (V_{BR}), collection dominates, causing the avalanche current to decay and ultimately stop. The gain (the number of carriers generated during an avalanche by a single initiating carrier) in this scenario is finite, and is determined by the statistics of the avalanche process. This type of operation is called linear-mode, since photocurrent is proportional to incident photon flux.

For linear-mode operation, Eq. 11 defines an empirical relationship between the multiplication factor (gain) M and the reverse bias V_{app}.

$$M = \frac{1}{1 - \left[\frac{|V_{app}|}{V_{BR}}\right]^m} \qquad \text{Eq. 11}$$

m takes on values between three and six, and is dependent on the material. M applies to the ideal diode equation when the reverse bias voltage approaches the breakdown voltage (Pierret, 1996).

In the case where the bias is above V_{BR}, multiplication outpaces collection. Initially, this causes exponential growth of the current. After some length of time, electrons and holes accumulate at the n- and p-edges of the depletion region, respectively. This creates an internal electric field that is in opposition to the applied bias and arrests the growth of the current. The device remains in an on state, however, until the circuit reduces the applied bias (quenches the device), allowing the APD to turn off. In this type of operation, known as Geiger-mode, the gain would be infinite if the bias were held above breakdown. By reducing the applied bias, called quenching, the gain is determined by the circuit. In the Geiger-mode case, an electrical event resulting from a single incident photon is indistinguishable from one initiated by a larger number of photons arriving simultaneously.

It is important to note that even in Geiger-mode, there is a probability that the avalanche may dwindle in its earliest stages and result in a non-detectable signal. While gain is an important metric for linear-mode operation, Geiger-mode operation is better characterized by the probability that the avalanche will become self-sustaining, referred to as the avalanche initiation probability. This probability can be calculated independently in theory (McIntyre, 1973). It is a function of carrier ionization rate (a material property), the applied electric field, and the position of the carrier in the high-field region. Carriers that are generated at the beginning of the high-field region are more likely to initiate an avalanche than carriers generated in the middle.

Applications that utilize APDs take advantage of the avalanche mechanism to boost signal from a single incident photon. For example, an observation of a target emitting very low photon flux may have a low SNR if the photocurrent is similar in magnitude to the dark current or to the

readout noise. When an avalanche multiplies the signal without multiplying the dark current, the SNR greatly improves and the target becomes discernable in the observation (Pierret, 1996).

1.3.1.3 MOS Capacitor

Metal-oxide-semiconductor (MOS) capacitors are based on the same principles as the PN junction, though they are utilized differently. The capacitor consists of a semiconductor, an insulating layer (SiO_2 in silicon devices, usually), and a metal contact on the insulating layer. Figure 13 shows an example of a biased MOS capacitor. A positive voltage applied to the capacitor's metal contact (the top layer) attracts free electrons in the semiconductor.

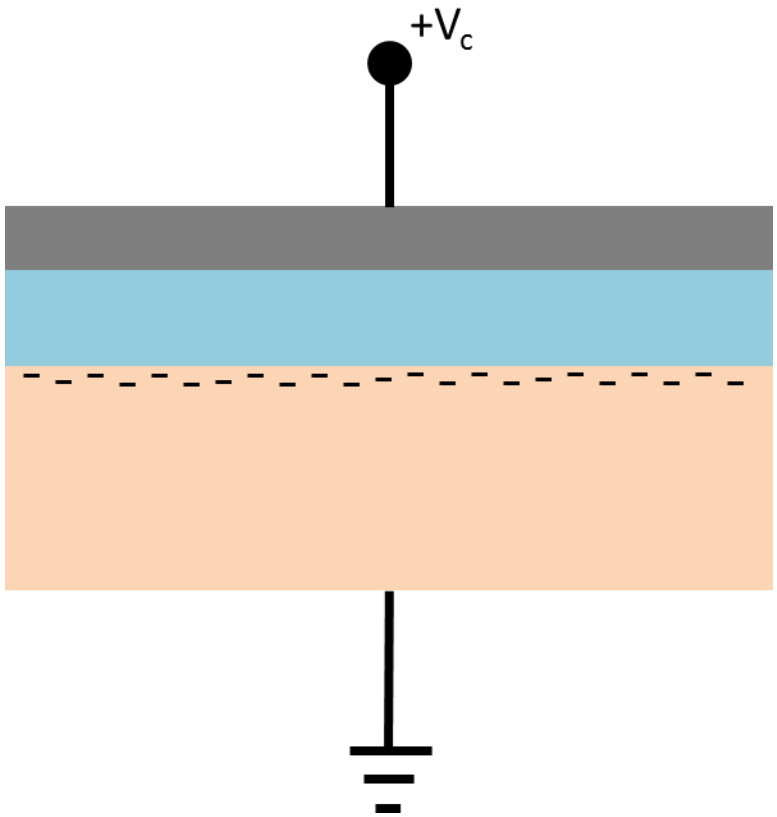

Figure 13 – A diagram is shown depicting the structure of a MOS capacitor. The "-" signs represent electrons that are drawn to the positive voltage (V_c) on the capacitor. The top layer is the metal contact, the middle layer is the insulating material, and the bottom layer is the semiconcductor.

MOS capacitors are used to move charge in a CCD by carefully changing the voltage across multiple MOS capacitors in a pixel and in neighboring pixels. As one voltage is increased, the

voltage next to it decreases, which causes the accumulated charge to move towards the capacitor with the higher voltage.

1.3.2 *Superconductor-Based Detectors*

Superconduction was first discovered by Kamerlingh Onnes in 1911, during the course of his experiments with liquefied helium. He discovered that the resistivity of many metals and alloys drops to zero below some material-specific critical temperature (Kittel, 2005). The transition between the superconducting state and a normal resistive state is leveraged in superconductor-based detectors to measure both the number and energies of individual photons.

Normally, nearly all incident photons are reflected off of metals because of the impedance mismatch at the metal-vacuum boundary. However, in sufficiently thin films (~20 Å), more photons can be absorbed in the superconducting state (Kittel, 2005).

A bulk superconductor in a weak magnetic field will have zero magnetic induction in the interior of the material. If the material is heated above the critical temperature, and then a magnetic field is introduced, the resulting magnetic field will no longer be sustained inside the material when it is again cooled below the critical temperature. This is called the Meissner effect, discovered by Meissner and Ochsenfeld in 1933 (Kittel, 2005). There are thirty elemental metals that exhibit superconductivity and the Meissner effect, called type 1 superconductors. Type 2 superconductors consist of alloys that also exhibit superconductivity, though they usually exist in a mixed state of normal and superconducting regions. This is sometimes referred to as a vortex state. Therefore, type 2 superconductors exhibit the Meissner effect, but not in every location in the material. They can tolerate higher critical magnetic fields, are mechanically harder than Type 1 superconductors, and also have higher critical temperatures. Table 1 shows a list of selected examples of superconductors and their critical temperatures (Rohlf, 1994; Blatt, 1992).

Table 1 – Examples of superconductors, both type 1 and type 2, are shown with their critical temperatures (Rohlf, 1994; Blatt, 1992).

Material (Chemical Symbol)	**Critical Temperature (K)**
Tungsten (W)	0.015
Iridium (Ir)	0.10
Titanium (Ti)	0.39
Cadmium (Cd)	0.56
Molybdenum (Mo)	0.92
Aluminum (Al)	1.20
Indium (In)	3.408
Mercury (Hg)	4.15
Niobium-Titanium Alloy (NbTi)	10
Vanadium-Gallium Alloy (V_3Ga)	14.8
Niobium-Nitrogen Alloy (NbN)	15.7

1.3.2.1 Cooper Pairs

Because electrons in a superconductor exist in a Fermi gas, the band gap as understood in semiconductors is not a valid model for superconductors. Instead of being dependent on the electron-lattice interaction, the "band gap" in superconductors is based on electron-phonon interactions. Below the critical temperature (T_c) of a superconductor, the conducting electrons in the material are loosely associated in pairs, called Cooper pairs. The binding energy of Cooper pairs (the effective "band gap" of the superconductor) is described in Eq. 12 (Day, et al., 2003).

$$E_{bind} \approx 3.5kT_c \quad \text{Eq. 12}$$

k is the Boltzmann constant and T_c is the critical transition temperature of the material. Since T_c is material property, different superconducting materials can be used for different wavelength ranges, much like semiconductor material selection. A few materials with a band gap acceptable for ultraviolet (UV), optical, and NIR have been characterized and used in MKID arrays, such as TiN and W (Mazin, et al., 2012).

Above the critical temperature, the electrons are disordered, creating resistivity. In the superconductive state, Cooper pairs that are broken and subsequently excited above the band gap energy cause resistance in AC circuits. In DC circuits, these electrons are shorted by the remaining superconducting Cooper pairs (Kittel, 2005). Non-paired (excited) electrons are called quasiparticles, and the number of quasiparticles generated by a photon is a function of the photon energy, as shown in Eq. 13 (Day, et al., 2003).

$$N_{qp} = 2\frac{\eta h\nu}{E_{bind}} \quad \text{Eq. 13}$$

η is the absorption efficiency of the material and hν is the energy of the photon. The factor of 2 results from the paired nature of the electrons: breaking the bond between Cooper pairs results in two quasiparticles. Because of the relationship in Eq. 13, superconducting detectors have the capability to collect both spatial and spectral information in a single pixel. This gives them an advantage over traditional refractive spectroscopy instruments that sacrifice spatial resolution for spectral information.

The Meissner effect generally applies to DC-operated superconductors. In AC-operated circuits, there is a short penetration of the magnetic field (~50 nm from the surface of the material) due to stored kinetic energy of the electrons. This interaction results in a surface inductance, which contributes to a surface impedance described in Eq. 14 (Day, et al., 2003).

$$Z_s = R_s + i\omega L_s \quad \text{Eq. 14}$$

R_s is the surface resistivity and L_s is the surface inductance of the superconducting material. When the superconductor is cooled to temperatures much less than the critical temperature, the resistance has a sharp threshold such that a single absorbed photon can cause the shift in state. A change between the superconducting and normal states can also be triggered by a change in temperature (Kittel, 2005). Generally, Cooper pairs will recombine on the order of milliseconds or microseconds. Regardless of the direct cause of the transition, the important effect for photon detection is the change in the resistivity of the material.

1.3.2.2 MKIDS and TKIDS

MKIDs and TKIDS (thermal kinetic inductance detectors) both detect photons by indirectly measuring the change in resistivity of the superconducting material. To achieve this, the superconductor is placed in an LC circuit, where any change in the quasiparticle density results in an increase in surface resistivity and therefore decreases the circuit's resonant frequency. As a result, a microwave probe signal transmitted through the circuit will show a change in both power and phase at the output, which is proportional to the number of quasiparticles generated in the superconductor (Day, et al., 2003). A figure describing the MKID detection process is shown in Figure 14 (Mazin, et al., 2012).

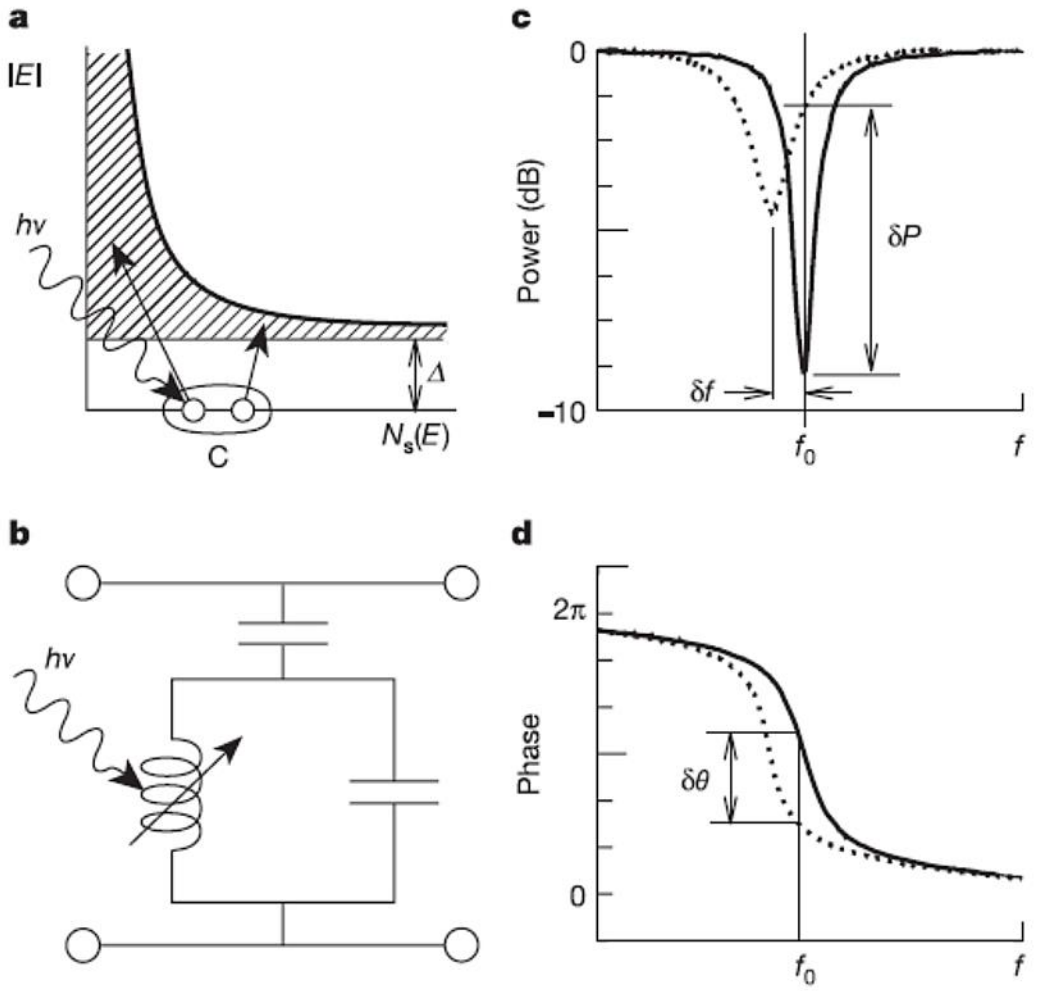

Figure 14 – A diagram of the detection of a photon in an MKID detector is shown (Mazin, et al., 2012). Figures (a) and (b) show the absorption of a photon, causing the generation of two quasiparticles from a Cooper Pair (C). Figures (c) and (d) show the shift in power and phase associated with the resonant frequency of the pixel after the generation of the quasiparticles.

As a practical method of reading out an array of KIDs, a frequency-domain multiplexer is used. Each pixel (resonator) in an array is designed to have a slightly different resonant frequency. If the operational temperature is kept constant, they will have zero false counts.

The drawback to this readout approach is that the number of pixels that can be multiplexed through a single read-out channel is limited. The number depends on the frequency separation required (separation must be greater than the width of typical resonance), the level of crosstalk between pixels, and the total power limit and bandwidth of the cryogenic microwave amplifier.

Anti-aliasing filters are often required to reject signals above the frequency domain of interest, as well. As an example, mosaicked array of 1k x 1k pixels would require 400 readout channels with the given current constraints (Mazin, 2005). This is far more than the 16-, 32-, or even 64-channel readouts in semiconductor-based detectors. However, this is not a fundamental limitation of the technology, and may be overcome in future development and research efforts. One permanent disadvantage is the super-cooled temperatures required for superconductor operation, usually down to a few tenths of a Kelvin. While semiconductor-based detectors are cooled to improve performance, they still operate at warmer temperatures. If the cooling system for a superconducting detector fails, the device stops working completely.

KID detectors also suffer from low fill factor, even with fiber optics or a microlens array, because pixels are patterned with inter-digitated fingers of the superconducting material (Mazin, 2005). And while KIDs may be immune to dark noise, they are affected by other noise sources, the most significant being two-level system (TLS) noise and the frequency resolution of the readout electronics. TLS noise is an excess frequency noise caused by defects that allow tunneling in amorphous dielectrics at very cold temperatures (~1 K). Physically, this noise manifests as a jitter in the resonance frequency measurement from the capacitors in each pixel (Kumar, et al., 2008).

One way to reduce TLS noise is to change the method of excitation of electrons, as is done in TKIDs. The response of a KID to excess quasiparticles generated by an absorbed photon is nearly identical the response to a change in temperature. The number of quasi-particles as a function of temperature is shown in Eq. 15 (Miceli, et al., 2014).

$$N_{qp}(T) = 2e^{\left(\frac{-E_{bind}}{kT}\right)} N_0 \sqrt{2\pi k_B T E_{bind}} \qquad \text{Eq. 15}$$

Once again, k is the Boltzmann constant, T is the temperature of the superconductor, E_{bind} is the Cooper pair binding energy, and N_0 is the single spin density of states at the Fermi level. As the temperature increases, more and more quasiparticles are generated. If the incident photons are

absorbed in a layer that is coupled to the superconductor material, the superconductor's temperature will increase in proportion to the energy of the photon absorbed (Miceli, et al., 2014). The absorber must be a material with low heat capacity, such as Ta or Sn, and must have a high absorption coefficient. The material must also have good thermalization properties to avoid non-uniformity in efficiency and detection time within a pixel (Quaranta, et al., 2014). Effectively, the KID can be used as an inductive thermometer. This leads to the familiar response of increase in resistivity and inductance, which results in the shift in resonant frequency. While the TKID is currently being investigated for use in thermalized x-ray detection, there is a possibility that such a detector could be used for UV/Optical/IR measurements.

1.3.2.3 Transition Edge Sensor Devices

Transition edge sensor (TES) devices use the same theory of superconductivity as KIDs, but instead of measuring the resonant frequency of an LC circuit, TES devices output a change in current. The superconducting material is biased and cooled to just below T_c. The bias is kept constant during an exposure. When a photon is absorbed, the resistivity of the material changes and the current flowing through the superconductor decreases. The change in current is proportional to the number of quasiparticles in the material, and therefore to the energy of the absorbed photon (Chervenak, et al., 1999). While no large-array TES devices have been used for imaging, single-element devices in small arrays have been used as spectrophotometers for astronomical observations to obtain time-resolved (up to 30 kHz) and energy-resolved (up to 0.05 eV) data (Romani, et al., 1999).

TES devices are generally read out using a superconducting quantum interference device (SQUID). When coupled with an input coil that converts current to magnetic flux, the output voltage of the SQUID varies proportionally to the change in current. The drawback to this device for imaging applications is the practical readout implementation. Each SQUID current amplifier has eight terminals, though some wires can be connected in series for multiple devices. For example, an array of N x M SQUIDs can be read out using as few as 2M+N+1 terminals instead of 8MN, though this is still prohibitive given the standard 2k x 2k arrays used in semiconductor-based detectors (Chervenak, et al., 1999).

There are two approaches to SQUID multiplexing: time-division multiplexing (TDM) and frequency-division multiplexing (FDM). In a TDM system, the number of detectors is limited by SQUID-noise aliasing (which increases as the square root of the number of detectors). FDM systems are limited by crosstalk in narrow-band filters, since they operate at MHz frequencies. Recently, alternate readout implementations have been devised to increase the scalability of TES devices, dubbed code-division multiplexing (CDM). Instead of modulating square waves (TDM) or the frequency of sinusoids (FDM) to address different pixels, CDM modulates the polarity of each pixel's coupling to a SQUID amplifier. This approach eliminates SQUID-noise aliasing and suppresses 1/f noise. CDM also reduces the number of addressing lines to N for an array of 2^N pixels, or 22 leads for a 2k x 2k array (Niemack, et al., 2010).

1.4 Comparison Criteria

In order to compare various detector technologies, each will be evaluated in terms of SNR for both short and long exposure scenarios. Specific applications will be used to benchmark each detector's performance.

For an exoplanet transit photometry experiment, it is reasonable to assume that the star is a sun-like star that has an apparent magnitude of 12 in the g-band, centered at 520 nm (Borucki, et al., 2010). Apparent magnitude is a measure of the brightness of a star as seen by an observer just outside of Earth's atmosphere (e.g., in low-Earth orbit). An apparent magnitude of 12 in the g-band corresponds to ~$3x10^5$ photons/s for a 2m telescope. An Earth-like planet orbiting such a star would introduce a dip in the signal of about 0.01% (Brown, et al., 2011). For a 1 m telescope like the one on Kepler, this translates to a signal of 98,000 photons/s from the star and a decrease of roughly 10 photons/s during transit. For short transit times, the temporal sampling must be sufficient to accurately characterize the transit signal, requiring short exposure times. For a direct imaging scenario, an Earth-like planet orbiting a sun-like star at a distance of 10 parsecs has a magnitude of 30, a signal of 0.1 photons/s. This imaging scenario requires long exposures.

Pulsars require short exposures to capture the oscillation of their signal. The Crab Pulsar is one of the first pulsars to be discovered, and has a magnitude in the V-band of 16.75 (Nasuti, et al., 1996), which is a signal of 1755 photons/s on a 2 m telescope.

For AO applications, it is reasonable to assume that the system is using a bright star to track wavefront aberration. If the star has a magnitude of 12 in the V-band, centered at 550 nm, then a 1 m telescope would collect 35,000 photons/s. Assuming that the signal is split across four detectors, each detector would receive a signal of 8700 photons/s. Given the short exposure times required for AO, this results in a signal of only 870 photons in a 0.1 s exposure (relatively long in AO) and only 87 photons in a 10 ms exposure.

A summary of the expected signal and exposure type for each application is shown in Table 2.

Table 2 – Signal and background flux levels for various missions and applications are shown.

Application	**Wavelength Region**	**Expected Signal (photons/s)**	**Exposure Type**
Exoplanet (direct imaging)	Optical / IR	0.1	Long
Exoplanet (transit photometry)	Optical	98000*	Short
Pulsar	X-Ray / UV / Optical	1800	Short
Adaptive Optics	Optical	8700	Short

*This signal is from the star, and is very large. The detector must be able to measure very small changes in the signal from the star during the transit.

2 THEORETICAL SNR OF SCIENTIFIC DETECTORS

The SNR of a measurement with a specific detector or system is the mean signal detected divided by the variance of the measurement. It is an important metric for comparison between detectors or between different operating conditions of the same detector. The ideal SNR for any detector or system is the SNR of the incoming signal. In the case of single photon counting detectors, this is the SNR of the photons arriving at the detector, which follow Poisson statistics. For Poisson statistics, the variance of the signal is equal to the mean of the signal. The standard deviation is the square root of the variance, and in this case it is called shot noise (Janesick, 2001). Therefore,

$$\boldsymbol{SNR} = \frac{\boldsymbol{S_p}}{\sqrt{\boldsymbol{S_p}}} = \sqrt{\boldsymbol{S_p}} \qquad \text{Eq. 16}$$

where S_p is the signal in photons.

Deriving the SNR expression for detectors operating in linear mode is straight-forward: add all the noise sources (standard deviation) in quadrature and apply any efficiency losses to the signal. This is true for detectors with and without gain, such as a CCD, a CMOS APS detector, an EMCCD operating in analog-mode, or an LM-APD.

The SNR expression for a photon counting mode detector becomes less intuitive, however, due to the digital nature of the measurements. For instance, the output of a photon counting GM-APD pixel is a string of ones and zeroes, representing whether or not individual gates registered an avalanche. The same is true of an EMCCD in photon counting mode, although the ones and zeroes are assigned by an thresholding algorithm, not read from the pixel. In this case, the probability of a one must be translated to the photon flux; this translation adds some noise that is not Poissonian.

The following sections detail a Monte Carlo simulation of the SNR for a variety of detectors, which was used to verify the derivation of the SNR expressions for CCDs, CMOS APS

detectors, EMCCDs in analog and photon counting mode, LM-APDs, and GM-APDs. In the following sections, a gate refers to a short exposure for a detector in photon counting mode. An exposure for these detectors is made up of many gates. Although the readout method for each gate is different for EMCCDs vs GM-APDs, the concept of many gates per exposure is constant.

2.1 SIMULATIONS

Monte Carlo simulations rely on random probability to generate the estimated output of complicated systems. The output of each process in a system is randomly generated using the constraining statistics of each process. The entire system is simulated in this way, multiple times, in order to estimate the mean output and the standard deviation of that output. Monte Carlo simulations are well-suited to detector physics, since all processes are governed by basic statistical processes. Photo-generation of charge is Poissonian, as is dark current generation, read noise is Gaussian, and low-probability processes like per-stage gain are described by the binomial distribution, as is the final output from a photon counting detector.

2.1.1 *PHOTON DETECTION AND SIGNAL RECONSTRUCTION*

A Monte Carlo simulation was created to calculate the observed SNR for any detector characteristics or operational settings. The simulation accepts a detailed list of inputs, including DCR / dark current, afterpulsing probability (p_{aft}), PDE / quantum efficiency (QE), gate time (t_{gate}), exposure time, signal flux, and the number of gates in the exposure (n_{gates}). The simulation repeats a single exposure (consisting of many gates) 10^7 times, and then the mean and standard deviation of the results are calculated, giving the SNR. For devices with linear (or non-existent) gain, carriers are generated based on Poisson statistics and the length of the exposure, given dark current and QE. The readout is also simulated with charge transfer efficiency (CTE), in the case of CCDs, and read noise. In the case of EMCCDs, the simulation is the same as CCDs with the addition of clock-induced charge (CIC). At the readout, each electron is assigned a random gain value using a statistical model. At the output of the gain register, read noise is applied (though in most cases this will be negligible). The analog-mode simulation divides the total output (quantized by the analog to digital converter, or ADC) by the average gain to determine the pixel

signal, while the photon counting mode simulation applies a thresholding routine to decide whether to return a 0 or a 1.

In the case of GM-APDs, carriers are generated from a random number Poisson distribution for each gate, based on the PDE, signal flux, and DCR. For gates where the number of generated carriers is greater than 0, the simulation records a 1. To add afterpulsing, a simple binomial random number generator is used to generate a carrier with probability p_{aft}, given that the simulation recorded a 1 in the previous gate. Examples of the results from the Monte Carlo simulation are shown below in section 2.

To add afterpulse carriers, a simple binomial random number generator is used to generate a carrier with probability p_{aft}, given that the simulation recorded a 1 in the previous gate. In this case, the avalanche probability is calculated with Eq. 52 (which simplifies to Eq. 29 if $p_{aft} = 0$), and SNR is calculated by dividing the mean by the standard deviation of the 10^7 trial results.

2.1.2 *Adaptive Optics and Centroid Estimates*

Using the detection simulation as a foundation, the AO simulation generates a calculated centroid position for a focal spot in a Shack-Hartmann (see section 1.2.2) quad-cell detector. In a quad-cell implementation, only four pixels are used to estimate the location of the focused beam from the lenslet array. Figure 15 shows a quad-cell group of pixels, labeled for reference. The lenslet focal spot is centered at (0,0), representing zero wavefront tilting.

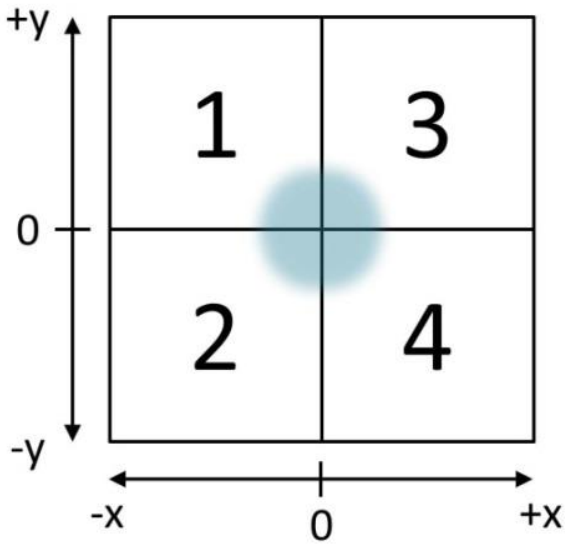

Figure 15 – A quad-cell group of pixels is shown, labeled for reference. The spot in the center shows the lenslet focal point with no wavefront tilt.

As discussed above, any shift in the center of the focal spot is linearly proportional to the wavefront tilt. With no wavefront tilting, the focal spot is centered on the intersection of all four pixels. When the focal spot moves, a simple center-of-mass centroid equation is used to estimate the center of the spot, shown in Eq. 17.

$$X = \frac{(S_3 + S_4) - (S_1 + S_2)}{S_1 + S_2 + S_3 + S_4}$$

$$Y = \frac{(S_1 + S_3) - (S_2 + S_4)}{S_1 + S_2 + S_3 + S_4} \qquad \text{Eq. 17}$$

X and Y are the coordinates of the position of the centroid of the focal spot relative to the center of the quad-cell group. S_n are the signals from each pixel, as shown in Figure 15.

It is important to note that the estimate is not perfect. The error is dependent on the focal spot size, as illustrated in Figure 16. When the spot no longer overlaps the cross-section of all four pixels, the estimate is saturated in one or both directions. If most of the spot overlaps the cross-section, there is a linear relationship between the actual location and the estimate of the location.

However, once the spot begins to move off of the cross-section, the relationship ceases to be linear until the location estimate saturates. For spots with sigmas greater than or equal to the size of 1 pixel, the estimate is linear throughout the entire quad cell. If less wavefront tilt is expected, e.g. such that the spot is only expected to travel within ± 0.25 pixels, then a smaller spot size is acceptable.

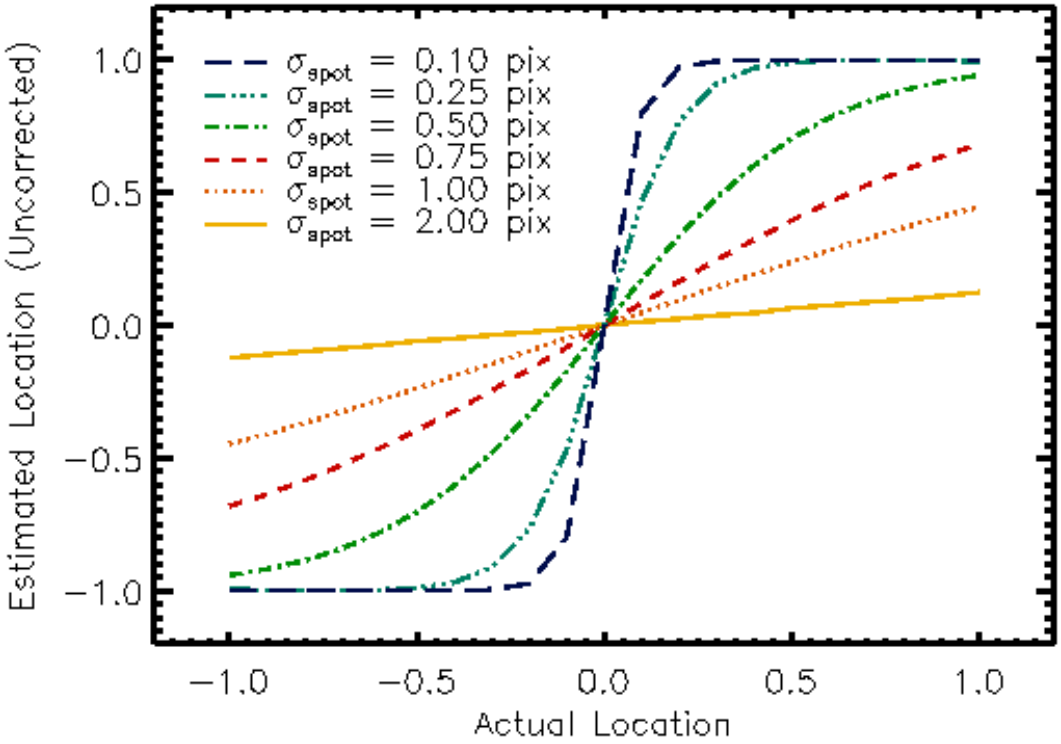

Figure 16 – This figure shows the estimated location of the centroid as a function of actual location without correcting for the spot size. The linear region of the estimate function decreases with decreasing spot size.

In order to correctly estimate the location of the centroid, the simulation must be limited to the linear region of the function (based on the spot size) and a linear correction factor must be used.

For ease of evaluation, the direct distance, D, to the focal spot is calculated based on the estimated values of X and Y, as in Eq. 18. The reported SNR is for the estimate of the direct distance.

$$D = \sqrt{X^2 + Y^2} \qquad \text{Eq. 18}$$

The simulation assumes that all four pixels share identical properties, including afterpulsing probability, dark current, detection efficiency, and read noise. Crosstalk is not considered in this simulation, but the method for correcting for crosstalk events has been studied previously (Aull, et al., 2015). The simulation can accept a non-ideal intrapixel sensitivity (IPS) function into account, but this introduces an error into Eq. 17 which is severe at large distances. The lenslet focal spot is approximated by a symmetrical 2D Gaussian function, and the signal on each pixel is determined mathematically using an oversampled version of the generated function. The Gaussian is scaled for the total fluence specified in the input to the simulation. The standard deviation in both x and y are assumed to be one pixel width. This assumption allows the actual pixel dimensions to be irrelevant, simplifying the simulation. Each pixel is simulated independently, based on the expected fluence determined by the relative shift in the center of the focal spot.

Figure 17 shows the results from the AO SNR simulation of a GM-APD detector. R is the magnitude in the R band of each focal spot (not the target, which is split across many lenslets). The detection efficiency is 0.8, the duty cycle (the percentage of time the detector is able to detect photons) is 0.89, the DCR is 0.04 e^-/s/pixel, the afterpulsing probability is zero, and the data for each point was simulated over ~9000 gates in 0.1 seconds. Each data point represents at least 10,000 iterations of the simulation and was evaluated at a unique set of simulated (x,y) coordinates. The magnitude reported is the effective magnitude of the light through a single lenslet; the total light from the guide star is spread across the lenslet array.

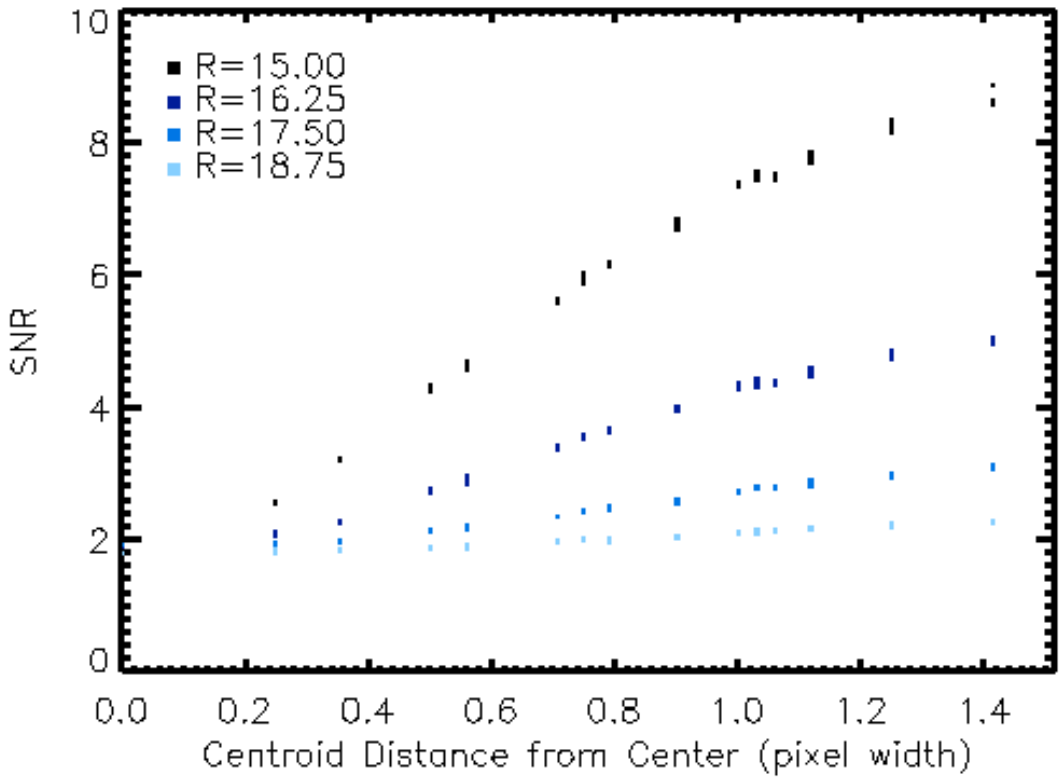

Figure 17 – This figure shows the Monte Carlo results for the AO simulation. SNR for the estimate of direct distance is shown for various fluence levels – R is the magnitude in the R band for the focal spot. The focal spot is assumed to be Gaussian with a symmetrical standard deviation of 1 pixel.

There is a pronounced SNR dependence on both the distance from the center of the quad-cell group and the signal from the lenslet array. The SNR of the estimate of centroid position increases as the distance from the quad cell center increases, and SNR also increases with increased fluence.

2.2 CCD

Modern CCDs generally use buried channel MOS capacitors to store charge collected during an exposure. Reading out the device requires physically moving the charge through the pixels, which is called a destructive read. This movement of electrons can lead to a loss of signal when moved over thousands of pixels (as in large-array devices), and the CTE becomes more important as the number of pixels increase. As CTE increases, it is more convenient to refer to the charge transfer inefficiency (CTI), which is 1-CTE. While state-of-the-art, buried-channel

CCDs may have CTIs on the order of $1x10^{-4}$ (a CTE of 0.9999), this can still represent a significant loss (Marcelot, et al., 2014). The pixel furthest from the readout in a 4k x 4k array, assuming a 32-channel readout, will lose 50 electrons with a CTI of $1x10^{-4}$ after transfer. CCDs and CMOS detectors also suffer from high read noise when the frame rate is high and reduced sampling is required (Janesick, 2001).

2.2.1 *THEORY OF OPERATION*

The basic theory of CCD operation is the collection of charge in a pixel and the transfer of that charge through other pixels in the array to the readout circuit. The most common method of charge transfer is the 3-stage transfer, which refers to voltages applied to three electrodes in each pixel in succession. Consider a CCD that collects electrons during an exposure. When the charge is being collected, only one electrode is set high, which attracts the electrons to the electrode. When the charge is read out, the adjacent electrode is also set high, spreading the charge between the two electrodes. Next, the first electrode is set low, so that the charge is only attracted to the second electrode. The process repeats with the third electrode, moving the charge again towards the readout. The cycle then begins again by setting the first electrode voltage high while the third electrode voltage remains high, spreading the charge between pixels. When the third electrode is set low, the charge has moved by one pixel. This gradual movement helps to reduce CTI. Figure 18 shows a schematic view of this three-phase clocking cycle (Janesick, 2001).

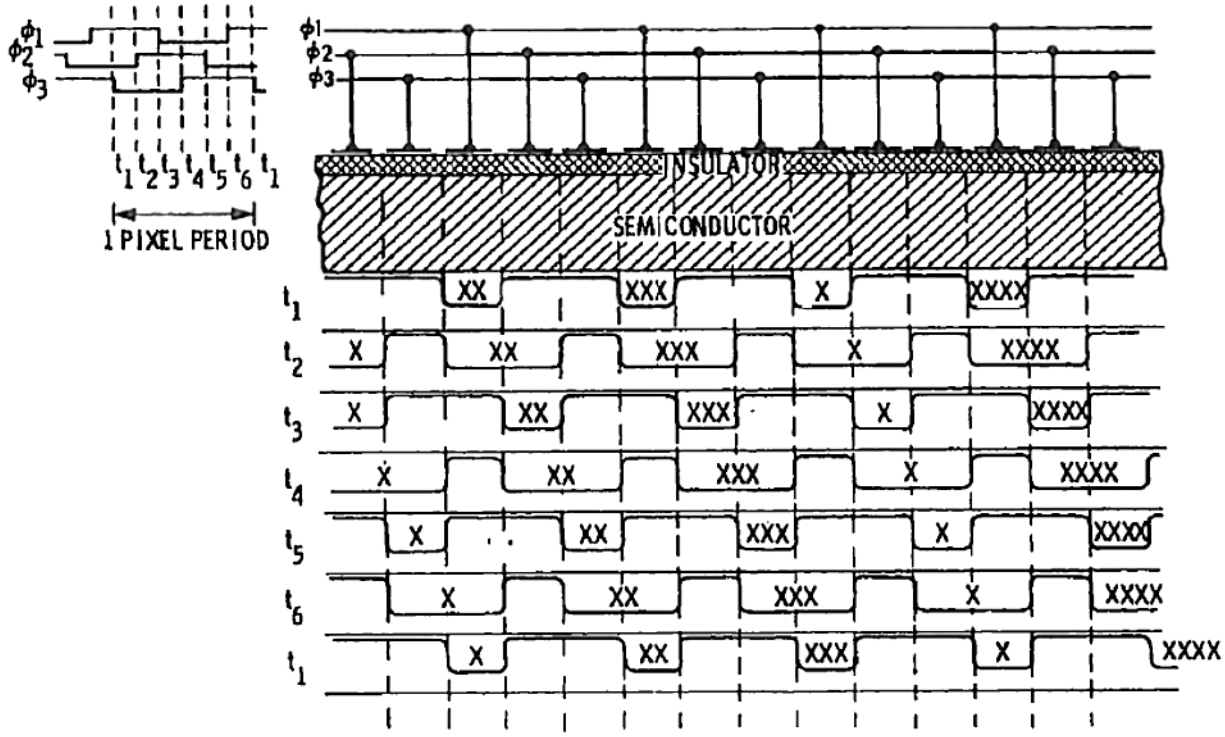

Figure 18 – A standard three-phase CCD clocking scheme is shown (Janesick, 2001). The Xs represent electrons that are collected during integration. Four pixels are shown, each with a different number of electrons to keep track of the charge packet as it is clocked through the device.

Frame transfer quickly transfers the signal from an imaging array to an on-chip storage array that is not illuminated. The storage array is read out while the integrating array begins another exposure.

Once the charge reaches the readout circuit, the analog signal is converted to a digital signal with an analog-to-digital converter (ADC) by quantizing the analog signal. Modern AD converters have many bits of quantization, and so the noise from quantization is considered negligible here.

2.2.2 *SNR*

For a non-ideal detector, the variance is not only the variance of the signal, but the sum of all the variances from each manipulation of the signal to the output of the detector (Janesick, 2001). For example, the SNR of a CCD with less than unity QE, exposure time t_{int}, dark current i_{dark}, and read noise σ_{read} is shown in Eq. 19 (ignoring any losses from CTI).

$$SNR = \frac{S_p \cdot QE}{\sqrt{S_p \cdot QE + i_{dark} \cdot t_{int} + (\sigma_{read})^2}} \qquad \text{Eq. 19}$$

Eq. 19 is valid in the linear operating region of the device, which assumes that the pixel output is linearly proportional to the signal. Figure 19 shows the theoretical SNR and Monte Carlo results according to Eq. 19 for the same CCD detector. The results are in good agreement.

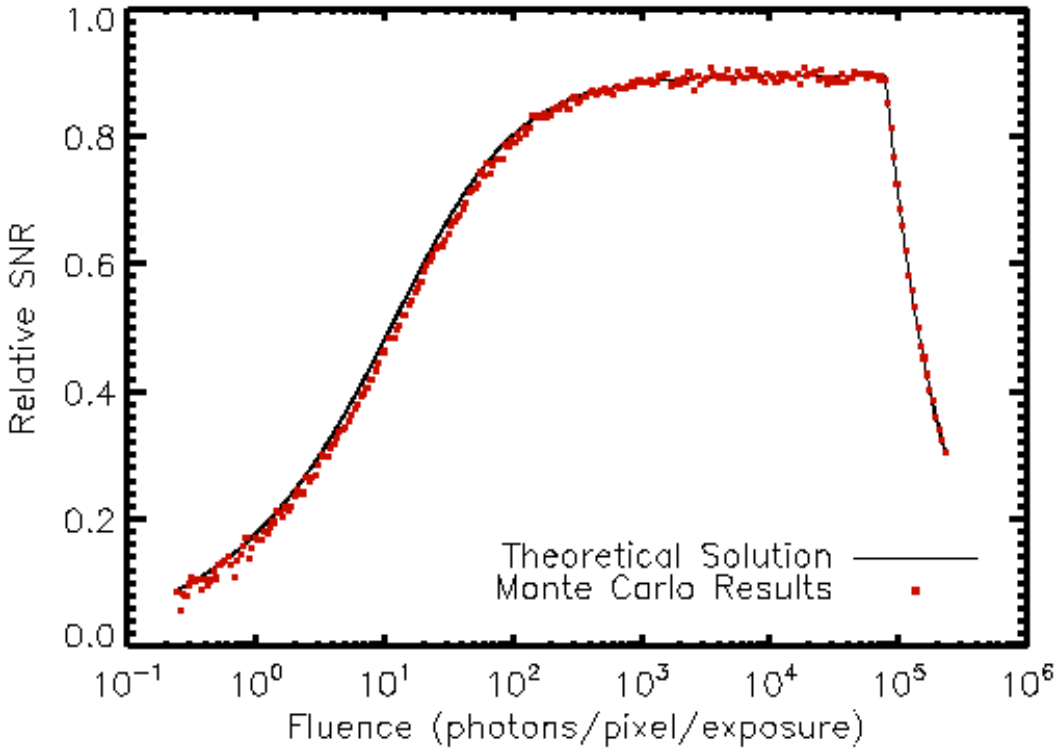

Figure 19 – This plot shows Monte Carlo results (individual points) and the theoretical solution (solid line) for the relative SNR of an LM-APD detector.

2.3 CMOS Detectors

CMOS direct readout (DRO) APS devices use a photodiode to store charge in each pixel, which can be read out by measuring the change in voltage during an exposure. This is called a non-destructive read. Read noise increases with increasing pixel rates, just as in a CCD. CTE is not

an issue, since charge is not being transferred, and CMOS detectors offer a distinct advantage in signal processing in their "up the ramp" sampling capability. In this type of sampling, a pixel's signal can be read many times in the same exposure. For the purposes of this document, CMOS will refer to direct readout active pixel sensor devices.

2.3.1 *THEORY OF OPERATION*

Instead of transferring the charge through the array, as in a CCD, a CMOS detector uses column- and row-select signals. Using this method, each pixel can be addressed in succession by connecting the output of the photodiode to the readout circuit and estimating the change in voltage. The most common circuit for accomplishing this is called a 3T circuit, which has three transistors (Janesick, 2001). One transistor controls the reset of the pixels. When the reset signal is high, the reset transistor is turned on and the reset voltage is applied to the photodiode, clearing the charge in the pixel. The second transistor is a source-follower buffer, which allows the pixel voltage to be measured without removing the charge in the pixel. The third transistor is the select transistor, which allows a pixel to be read by the readout electronics. The row signal is tied to the gate of the transistor, and the column signal is tied to the source / drain. When the row and column signals are high, the voltage is transferred from the buffer transistor to the readout circuit. Although this readout method is non-destructive, the readout rate is still limited by read noise, just as with CCDs. Since CTE is neglected in Eq. 19, the SNR for CCD and CMOS detectors is the same.

2.4 ANALOG-MODE EMCCD

The difference between a standard CCD and an EMCCD is a special gain register that amplifies the signal from the pixels during the readout (Daigle, et al., 2004). When operating in analog-mode, the number of photons that were absorbed is estimated based on the mean gain. As the number of electrons per pixel increases past 2 or 3, the estimate becomes so noisy that the number of electrons cannot be determined with any useful certainty. Therefore, short exposures that are read out quickly are necessary for good analog-mode EMCCD performance to keep the per-exposure flux near 1. Figure 20 shows an example of the degeneracy of measurements with many electrons (Robbins, 2003).

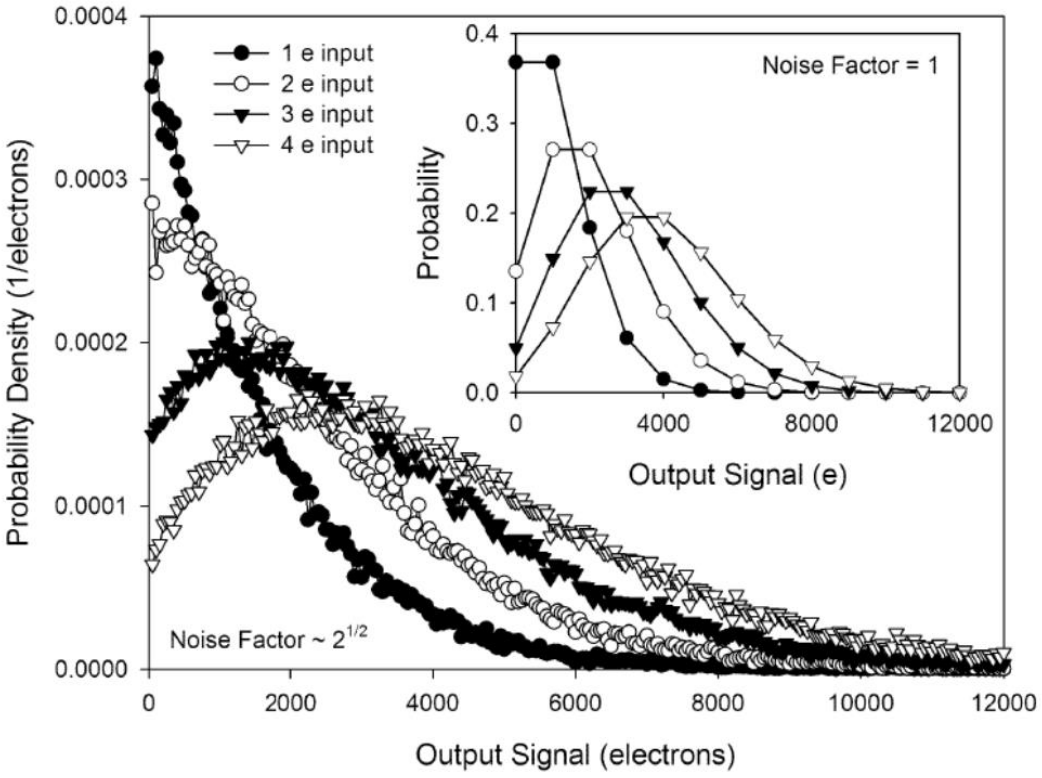

Figure 20 – The experimental (main) and theoretical (inset) distributions for the multiplication register output for 1, 2, 3, and 4 electrons are shown (Robbins, 2003). The multiplication register had 536 stages. For the theoretical case, the excess noise factor is 1, but for the experimental case the excess noise factor was measured to be ~1.41. This deviates from the theoretical value of 2.

2.4.1 *THEORY OF OPERATION*

The multiplication register is a 1D gain register that amplifies the signal as it passes through each stage. The gain in each stage is barely greater than 1. However, with an N-stage gain register, the total gain is G_{stage}^{N}, where N is the number of individual stages. For the case where G_{stage}=1.013, the total gain for a 512-stage gain register would be ~750. This gain, however, is stochastic, with a standard deviation equal to $\sqrt{2G_{total}}$. Figure 21 illustrates the gain mechanism in the multiplication register of an EMCCD (Robbins, 2003).

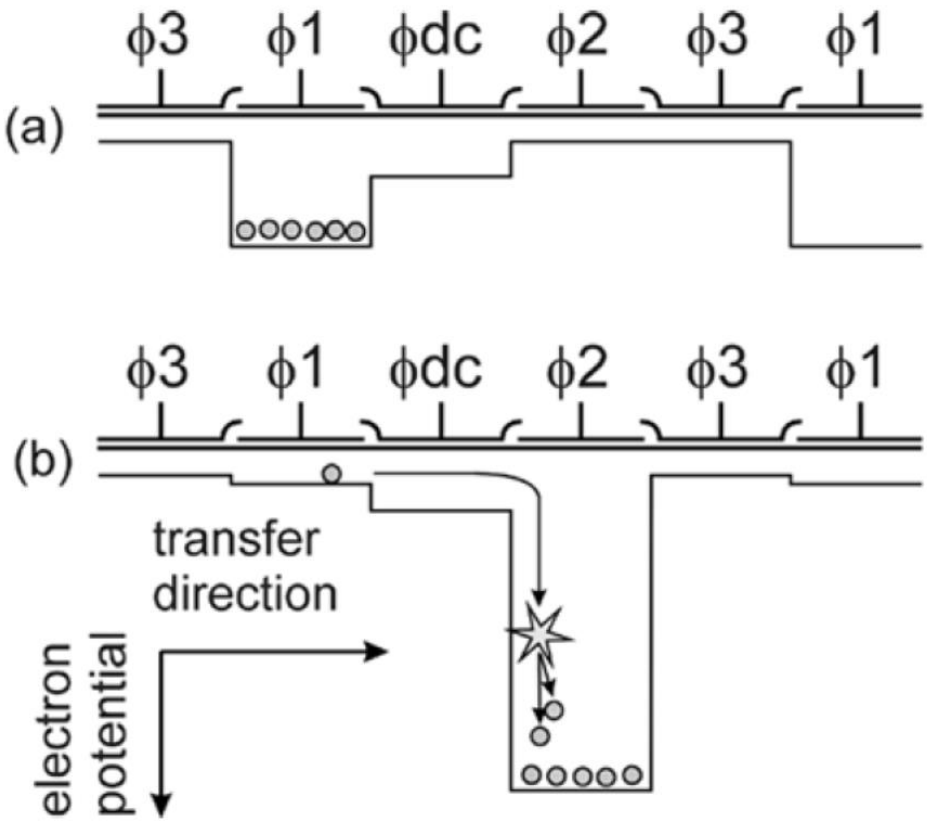

Figure 21 – This figure shows the electron movement for a low-voltage three-phase clocking strategy (a) and a high-voltage three-phase clocking strategy (b). The high-voltage case is used in the multiplication register of an EMCCD, resulting in gain (Robbins, 2003).

CIC in the array is not unique to EMCCDs, and is significant in any CCD that is read out very quickly. CIC is charge added to the original signal during the transfer from one pixel to another due to the high electric fields required for fast clocking during transitions.

2.4.2 *SNR*

CIC is best expressed as a probability of an injected carrier. As long as this probability is low, the probability distribution is assumed to only include 0 and 1 as outcomes, and therefore has the statistical characteristics of a Bernoulli trial (where p is the probability of a 1 and $q = 1 - p$ is the probability of a 0). The variance of a Bernoulli trial (one transfer) is

$$\boldsymbol{\sigma^2 = p(1-p)} \qquad \text{Eq. 20}$$

The total variance is the single-trial variance of Eq. 20 multiplied by the number of transfers. The number of transfers is dependent on the number of pixels in the detector and the location of the signal in the focal plane, but for simplicity the average number of transfers is N/2 for an NxN array, assuming the CIC is negligible in one direction. Therefore, the SNR of a the CCD in Eq. 19, adding CIC, is

$$\boldsymbol{SNR = \frac{S_p \cdot QE}{\sqrt{S_p \cdot QE + i_{dark} \cdot t_{int} + \left(\frac{N}{2}\right) \cdot p_{cic}(1-p_{cic}) + (\sigma_{read})^2}}} \qquad \text{Eq. 21}$$

From Eq. 21, the SNR of a CCD with gain G becomes

$$\begin{aligned} \boldsymbol{SNR} &= \boldsymbol{\frac{G \cdot S_p \cdot QE}{\sqrt{G^2\left(S_p \cdot QE + i_{dark} \cdot t_{int} + \left(\frac{N}{2}\right) \cdot p_{cic}(1-p_{cic})\right) + (\sigma_{read})^2}}} \\ &= \boldsymbol{\frac{S_p \cdot QE}{\sqrt{S_p \cdot QE + i_{dark} \cdot t_{int} + \left(\frac{N}{2}\right) \cdot p_{cic}(1-p_{cic}) + \left(\frac{\sigma_{read}}{G}\right)^2}}} \end{aligned} \qquad \text{Eq. 22}$$

In the first half of Eq. 22, the read noise is not multiplied by G because read noise is applied after the gain stage. However, G does apply to the standard deviation (and so G^2 to the variance) of the charge carriers entering the gain register. Suppose now that the gain for an EMCCD is not constant, or that the gain value for any carrier has a probability distribution with mean G. This requires the addition of an excess noise factor (ENF), which will be denoted here as F, defined in Eq. 23 (Robbins, 2003).

$$\boldsymbol{F^2 = \frac{{\sigma_{out}}^2}{G^2 {\sigma_{in}}^2}}$$ Eq. 23

The denominator is the expected output variance for noiseless gain, so F is the degree to which the output standard deviation exceeds that for a noiseless output. Due to the quantum nature of electrons (i.e., a non-integer value of electrons cannot deterministically exist), the variance of the output from an EMCCD gain register (as long as the number of multiplication stages is large and the gain is high) approaches $2(G\sigma_{in})^2$, or twice the expected variance. Therefore, we assume that $F = \sqrt{2}$ in Eq. 7.

$$\boldsymbol{SNR = \frac{G \cdot S_p \cdot QE}{\sqrt{F^2 G^2 \left(S_p \cdot QE + i_{dark} \cdot t_{int} + \left(\frac{N}{2}\right) \cdot p_{cic}(1 - p_{cic}) \right) + (\sigma_{read})^2}}}$$

$$\boldsymbol{= \frac{S_p \cdot QE}{\sqrt{2 \left(S_p \cdot QE + i_{dark} \cdot t_{int} + \left(\frac{N}{2}\right) \cdot p_{cic}(1 - p_{cic}) \right) + \left(\frac{\sigma_{read}}{G}\right)^2}}}$$ Eq. 24

Figure 22 shows the theoretical and Monte Carlo results based on Eq. 24 for the same EMCCD detector in analog-mode. The results are in good agreement.

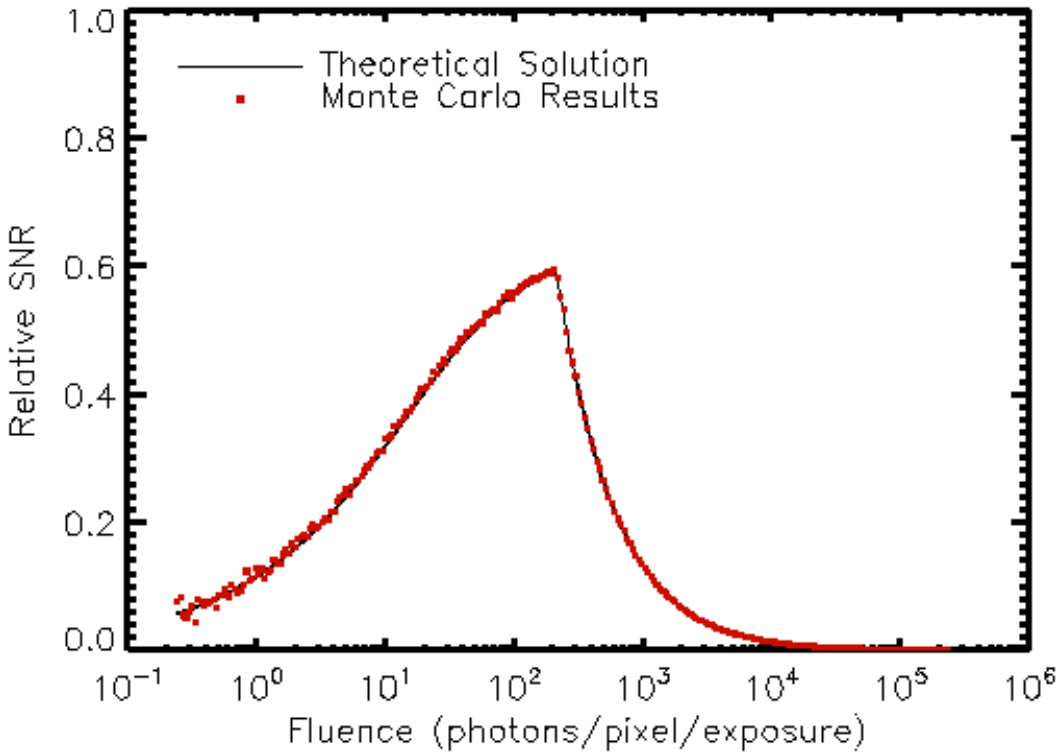

Figure 22 – This plot shows Monte Carlo results (individual points) and the theoretical solution (solid line) for the relative SNR of an EMCCD detector in analog-mode.

2.5 LM-APD

The LM-APD device considered in this document is a HgCdTe detector, which has effectively zero noise in the avalanche gain – e.g., there is no ENF with the proper junction grading and doping (Kinch, et al., 2004; Vojetta, et al., 2012). HgCdTe LM-APD devices are also referred to as e-APDs, since only electrons are avalanching and producing the gain. Other material types are used for LM-APD devices – silicon LM-APDs are widely available in single-element implementation. But the noiseless gain of the HgCdTe LM-APDs make them good candidates for photon-counting applications.

2.5.1 *THEORY OF OPERATION*

The output from an LM-APD is the voltage level of a pixel, which is estimated with a CMOS circuit or transistor-transistor logic (TTL) circuit, etc. – much like CMOS APS detectors. Because the gain is effectively noiseless, discrete voltage levels each represent a photon number

in an ideal detector. Dark current can also be multiplied, however, and so must be subtracted from the final signal. A more problematic source of noise is un-multiplied leakage current. This leakage current increases the signal voltage at a constant rate, and could saturate the AD converter for low photon flux levels and long exposures. Figure 23 shows an example of the signal from a pixel on a LM-APD detector (Cottingame, 2010). The red lines indicate the constant slope due to the un-multiplied leakage current, as well as the discrete signal levels associated with photon detection. The uncertainty in the trends is the result of read noise, which is not eliminated completely in these detectors, but is effectively reduced by the gain.

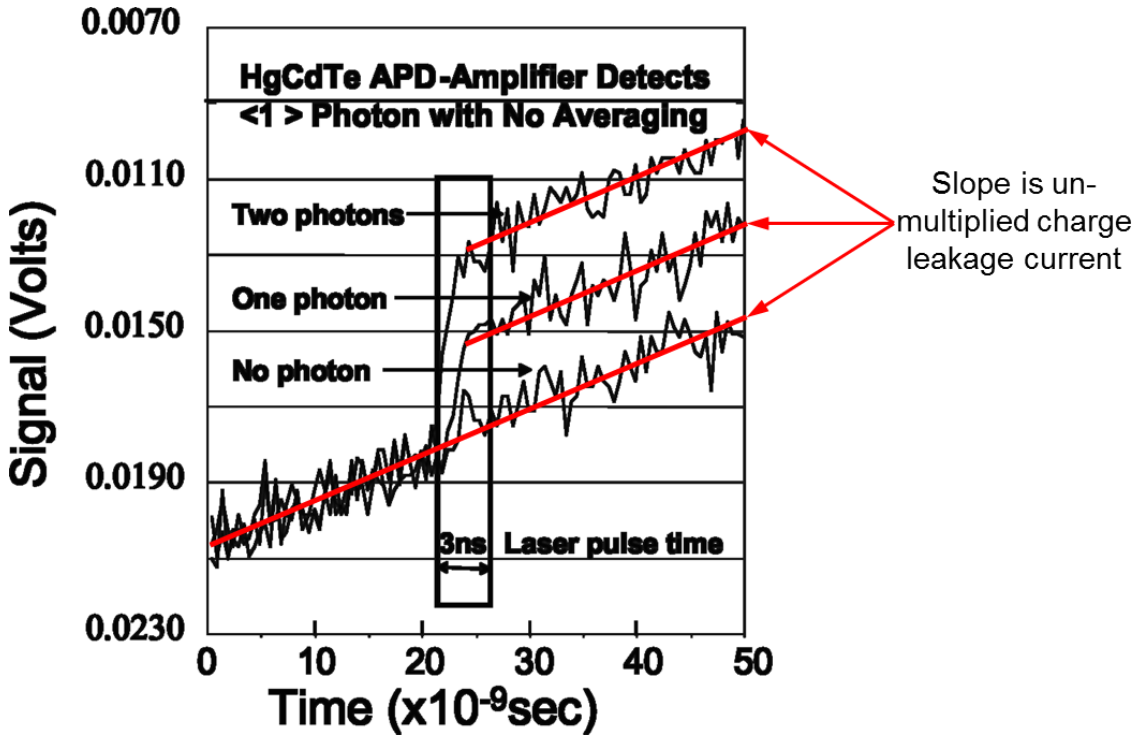

Figure 23 – This plot shows a sample of the output of a Raytheon eAPD, a HgCdTe LM-APD. Because the gain is effectively noiseless at low levels, it is easy to distinguish between one and two photons. The slope (noted by the red lines) in the data is due to the un-multiplied leakage current (Cottingame, 2010).

2.5.2 *SNR*

The SNR expression for an LM-APD is straight-forward, since the device uses linear gain to estimate the signal. There is un-multiplied leakage current that must be subtracted periodically to

avoid saturation during an exposure. The resulting noise affects the estimation of the signal, since the gain is relatively low in these devices (10-100). Eq. 25 shows the full SNR equation.

$$SNR = \frac{S_p \cdot QE}{\sqrt{S_p \cdot QE + i_{dark} \cdot t_{int} + \frac{i_{leakage} \cdot t_{int}}{G^2} + \left(\frac{\sigma_{read}}{G}\right)^2}} \qquad \text{Eq. 25}$$

Note that the leakage current noise contribution and the read noise are divided by G^2. This is because those noise contributions occur after the signal is multiplied, which diminishes their effect on the overall noise. Figure 24 shows the theoretical SNR and Monte Carlo results according to Eq. 25 for the same LM-APD detector. The results are in good agreement.

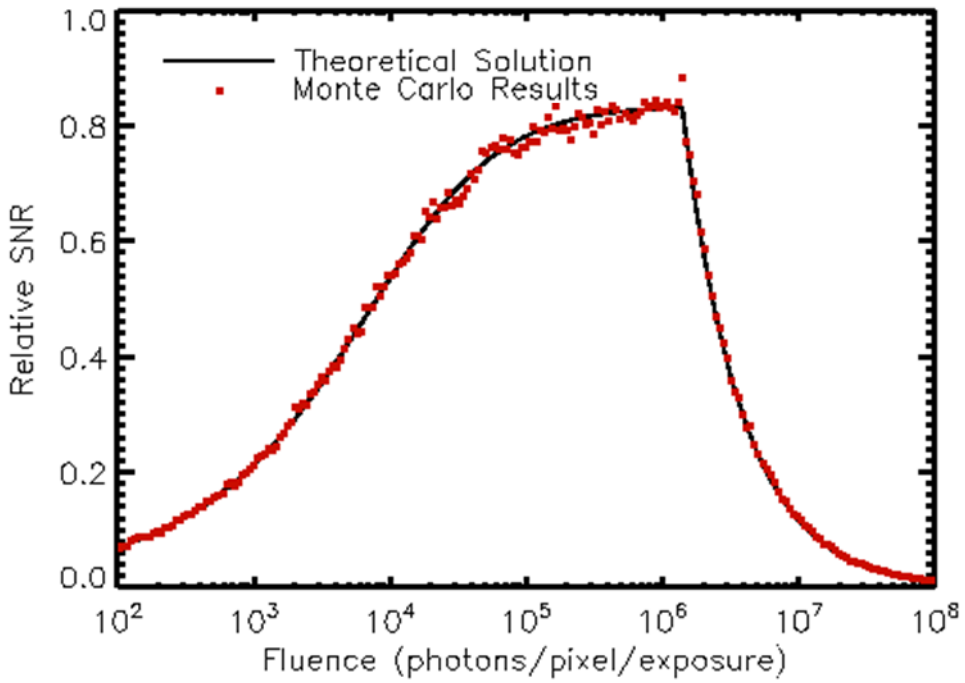

Figure 24 – This plot shows Monte Carlo results (individual points) and the theoretical solution (solid line) for the relative SNR of an LM-APD detector.

2.6 GM-APD

The array-based GM-APDs discussed in this thesis have been developed by Massachusetts Institute of Technology Lincoln Laboratory and characterized by the Center for Detectors to determine their suitability for space-based imaging applications, specifically for exoplanet missions (Figer & Kolb, 2014). A GM-APD can count single photons, which is very useful for low-light imaging applications. In Geiger-mode (digital) operation, the APD pixel counters simply record whether or not an avalanche occurred in a given exposure window (a gate), and repeats the process over many samples. From these samples, an avalanche probability can be calculated, and from that probability the estimated flux is derived (Kolb, 2014). Since this operation is fundamentally different from CCD or CMOS detectors, the form of the SNR expression is fundamentally different as well. The final form of the expression depends on DCR, afterpulsing probability (p_{aft}), PDE, gate length, number of gates (n_{gates}), and signal flux. The inclusion of afterpulsing probability also results in a means to estimate the signal when significant afterpulsing is present.

2.6.1 *THEORY OF OPERATION*

It is important to note that the majority of GM-APD applications use temporal measurements or avalanche totals to count photons. Unlike most applications, the detector described here is used to measure intensity by measuring avalanche probability during a set exposure window (usually on the order of microseconds). SNR has been presented for temporal measurements previously for similar devices (Gatt, et al., 2009), but it is important to emphasize that the operation is fundamentally different for the device presented here. For the GM-APD detector presented here, the measurement is actually of the probability of an avalanche during a short exposure, and nearly all temporal information is lost. Timing jitter is irrelevant, and the output from each pixel is digital. Afterpulsing effects on noise have been investigated by others (Vinogradov, et al., 2009), though no integration of afterpulsing statistics has been presented in the context of intensity imaging. Given the fundamental difference in operation between most GM-APDs and the device presented in this paper, a new expression for SNR must be derived from fundamentally different first principles.

Most currently-accepted NEP expressions for GM-APDs are irrelevant, given their time-measurement paradigm (Hadfield, 2009). Other accepted expressions for NEP are derived with the assumption that the intensity output from the pixel is analogue, not digital (i.e., number of avalanches is given rather than avalanche probability), which results in a completely different relationship between signal and noise (Lacaita, et al., 1996).

For this device, the detection cycle is clocked externally and reset at regular intervals. Each exposure is comprised of five distinct stages, repeated many times over. The first stage is the arming of the device, when the bias on the pixel is increased above the breakdown voltage. A set delay (the second stage), which constitutes the exposure gate, occurs before the next clock pulse. After the gate, a recording pulse is asserted (the third stage) that transfers the state of the pixel (1 or 0) to the readout circuit. Immediately after the recording pulse is complete, the pixel is forcefully disarmed (the fourth stage), meaning that the voltage is set below the breakdown voltage. A final delay (the fifth period) is observed after the disarm signal, called the hold-off time. This delay is usually used to mitigate afterpulsing events. At the end of the hold-off time the pixel is armed again and the cycle repeats. In an ideal device, the forced disarm would be unnecessary because avalanches between gates (quenched normally by the thresholding circuit) would not affect the occurrence of avalanches during the gates. In practical use, forced disarm is required because of the afterpulsing mechanism, which can induce an avalanche in a subsequent gate with a characteristic exponential decay probability. Figure 25 shows an example of the clocking signals described above.

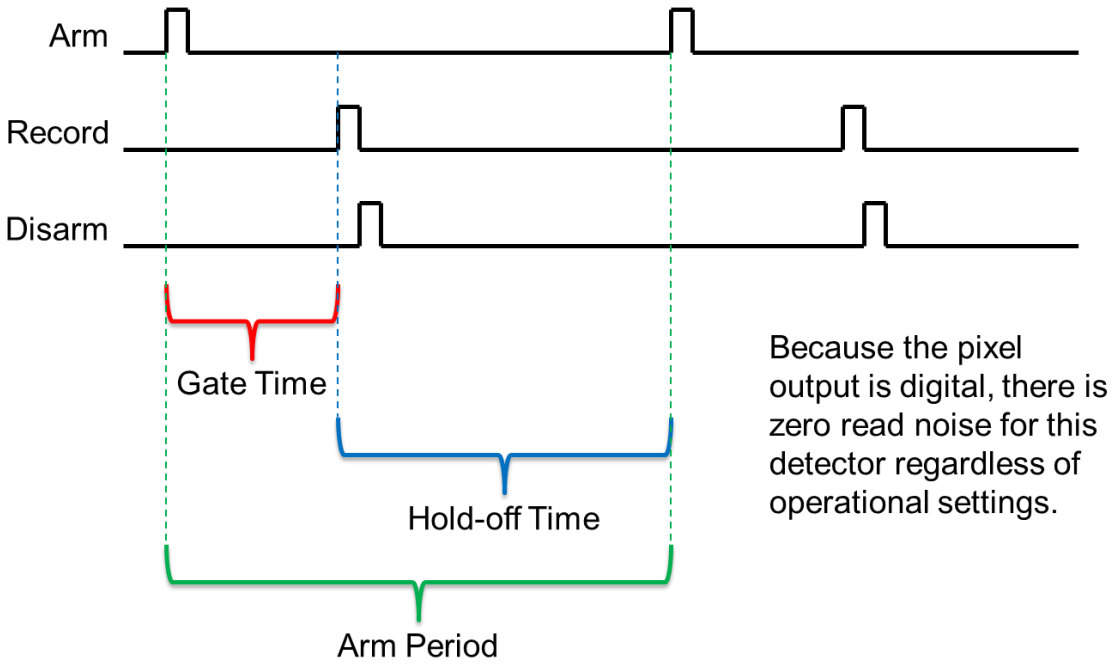

Figure 25 – This plot shows a sample of the clocking signals required for the GM-APD array operation.

Afterpulsing occurs when a trap (an intermediate energy state that exists in the band gap of the material) releases a carrier that initiates an avalanche in the absence of a photo-generated electron or dark count. Carriers are released from traps at random times, with the average de-trapping time defined as the trap lifetime. The length of the trap lifetime depends on the type of trap and its energy level. Avalanching carriers can remain in these states longer than the hold-off time of the device, becoming free again during the next detection cycle. The electric field immediately accelerates the newly mobile electron or hole, which may initiate an avalanche of carriers in the absence of photon signal (Kindt & de Langen, 1998).

2.6.2 *SNR NEGLECTING AFTERPULSING*

To simplify the relationships, the initial assumptions are that QE is unity and there are no other sources of noise (these non-idealities will be added later). Instead of counting each 1 as a single photon, the ratio of 1s to total number of gates is measured. That ratio is an estimate of the probability of one or more photons arriving within a single gate. The probability of n_p photons in a specific interval for a an average number of photons λ_p is

$$\boldsymbol{P(n_p) = \frac{e^{-\lambda_p} {\lambda_p}^{n_p}}{n_p!}} \quad \text{Eq. 26}$$

For a Poisson distribution, since each trial is memoryless (independent),

$$P(n_p = a \; or \; n_p = b) = P(n_p = a) + P(n_p = b) = e^{-\lambda_p}\left(\frac{{\lambda_p}^a}{a!} + \frac{{\lambda_p}^b}{b!}\right) \quad \text{Eq. 27}$$

and the total probability than n_p equals any positive integer is 1, or

$$\sum_{n_p=0}^{\infty} \frac{e^{-\lambda_p} {\lambda_p}^{n_p}}{n_p!} = 1 \quad \text{Eq. 28}$$

Since dark carrier generation is also a Poisson process, the average number of dark carriers per gate λ_d is additive in the exponential function. Changing n_p to n to represent the total number of carriers per gate (assuming dark current is the only source of noise and that there is no photon loss) and λ_p to λ to reflect the average number of total carriers per gate, the ratio of 1s to the total number of gates can therefore be defined as:

$$\frac{n_{ones}}{n_{gates}} = P(n \geq 1) = 1 - P(n = 0) = 1 - \frac{e^{-\lambda}\lambda^0}{0!} = 1 - e^{-\lambda} \quad \text{Eq. 29}$$

where $\lambda = \lambda_p + \lambda_d$

From this simplified solution, the estimate of the average number of photons per gate is

$$\lambda_p = -\ln\left(1 - \frac{n_{ones}}{n_{gates}}\right) - \lambda_d \qquad \text{Eq. 30}$$

In practice, λ_d may be estimated by applying Eq. 29 to measurements made under dark conditions for an isolated pixel or by directly measuring the current under reverse bias conditions where the electric field is not sufficient for any multiplication. The estimate of the total number of photons in a pixel over an exposure (the fluence for the exposure) with PDE less than unity is

$$PDE \cdot \lambda_{tot} = \left(-\ln\left(1 - \frac{n_{ones}}{n_{gates}}\right) - \lambda_d\right) \cdot n_{gates} \qquad \text{Eq. 31}$$

The variance of the estimate is slightly more complicated. To simplify the equations, let $\frac{n_{ones}}{n_{gates}}$ = x and λ_{tot} = y. The variance of x is the variance of a binomial distribution based on the number of trials, n_{gates}, and the probability of an avalanche during a gate, P(gate=1) = p:

$$\sigma_x{}^2 = \frac{n_{gates} \cdot p(1-p)}{n_{gates}{}^2} = \frac{p(1-p)}{n_{gates}} \qquad \text{Eq. 32}$$

The variance of the total number of ones is the Bernoulli trial variance, p(1-p), multiplied by the number of trials, n_{gates}. However, x is the ratio of total number of ones to the total number of gates, and so the variance must be divided by the square of the scale variable, resulting in Eq. 32.

Now, the variance of the estimate (which is a function of x) must be defined. To begin, the expected value of y(x) is defined in terms of the probability density function of x, or f(x), as follows.

$$E[y(x)] = \int_{-\infty}^{\infty} y(x) f(x) dx \qquad \text{Eq. 33}$$

If f(x) is concentrated about the mean (a valid assumption since x is the outcome of a series of Bernoulli trials, a binomial distribution), then f(x) is assumed to be negligible outside the range (μ – ε , μ + ε), where μ is the mean of f(x), and y(x) becomes y(μ):

$$E[y(x)] = y(\mu) \int_{\mu-\varepsilon}^{\mu+\varepsilon} f(x) dx = y(\mu) \qquad \text{Eq. 34}$$

This estimate may be improved by a polynomial expansion (Papoulis & Pillai, 2002).

$$y(x) = y(\mu) + y'(\mu)(x - \mu) + \cdots + y^n(\mu) \frac{(x - \mu)^n}{n!} \qquad \text{Eq. 35}$$

Inserting Eq. 35 into Eq. 34 and neglecting higher-order terms for a parabolic approximation:

$$E[y(x)] = y(\mu) + y''(\mu) \frac{(x - \mu)^2}{2} = y(\mu) + y''(\mu) \frac{{\sigma_x}^2}{2} \qquad \text{Eq. 36}$$

To find the variance of y:

$${\sigma_y}^2 + {\mu_y}^2 = E[y^2(x)] \qquad \text{Eq. 37}$$

$$\mu_y{}^2 = E[y(x)]^2 = \left(y(\mu) + y''(\mu)\frac{\sigma_x{}^2}{2}\right)^2$$
$$= y^2(\mu_x) + y(\mu_x)y''(\mu_x)\sigma_x{}^2 + [y''(\mu_x)]^2\frac{\sigma_x{}^4}{4} \quad \text{Eq. 38}$$

$$E[y^2(x)] = y^2(\mu_x) + \left[\frac{d^2}{dx^2}\, y^2(\mu_x)\right]\frac{\sigma_x{}^2}{2}$$
$$= y^2(\mu_x) + \sigma_x{}^2[|y'(\mu_x)|^2 + y(\mu_x)y''(\mu_x)] \quad \text{Eq. 39}$$

Since σ_x is always less than 1, the $\sigma_x{}^4$ term in Eq. 38 is negligible. Substituting the simplified Eq. 38 and Eq. 39 into Eq. 37 and solving for $\sigma_y{}^2$ (Papoulis & Pillai, 2002):

$$\sigma_y{}^2 = |y'(\mu_x)|^2\sigma_x{}^2 \quad \text{Eq. 40}$$

Eq. 40 is the variance of the estimate of the number of photons. The derivative of y(x) (Eq. 30) is

$$y'(\mu_x) = \frac{1}{1-\mu_x} \quad \text{Eq. 41}$$

Substituting Eq. 32 and Eq. 41, and letting μ_x equal P(gate=1) = p, the total variance for the exposure is

$$\sigma_{tot}{}^2 = |y'(\mu_x)|^2 \sigma_x{}^2 n_{gates}{}^2 = \frac{p}{(1-p)} n_{gates} \qquad \text{Eq. 42}$$

Since the output of each gate is either a 1 or a 0 in the case of a photon counting mode detector, each gate may be modeled as a Bernoulli trial with the probability of a 1 equal to $p = \mu_x$ and the probability of a 0 equal to $q = 1 - p$. Assuming that only one electron is necessary to register a 1, p is defined by Eq. 29. Now $\sigma_y{}^2$ may be evaluated as a function of the average number of photon-generated carriers (λ_p) and the average number of dark-current-generated carriers (λ_d), substituting Eq. 29 and taking PDE into account, as

$$\sigma_{tot}{}^2 = \left(\frac{E\left[\frac{n_{ones}}{n_{gates}}\right]}{1 - E\left[\frac{n_{ones}}{n_{gates}}\right]} \right) n_{gates} = \left(e^{(PDE \cdot \lambda_p + \lambda_d)} - 1\right) n_{gates} \qquad \text{Eq. 43}$$

Therefore, the SNR for a GM-APD with no afterpulsing is

$$SNR = \frac{PDE \cdot \lambda_p \cdot n_{gates}}{\sqrt{\frac{p}{(1-p)} \cdot n_{gates}}} \qquad \text{Eq. 44}$$

where $p = 1 - e^{-(PDE \cdot \lambda_p + \lambda_d)}$

λ_p is the number of photons absorbed per gate, and λ_d the number of dark current carriers generated per gate. Figure 26 shows an overlay of the Monte Carlo results and the analytical solution in Eq. 44, normalized to the ideal shot noise limit of SNR.

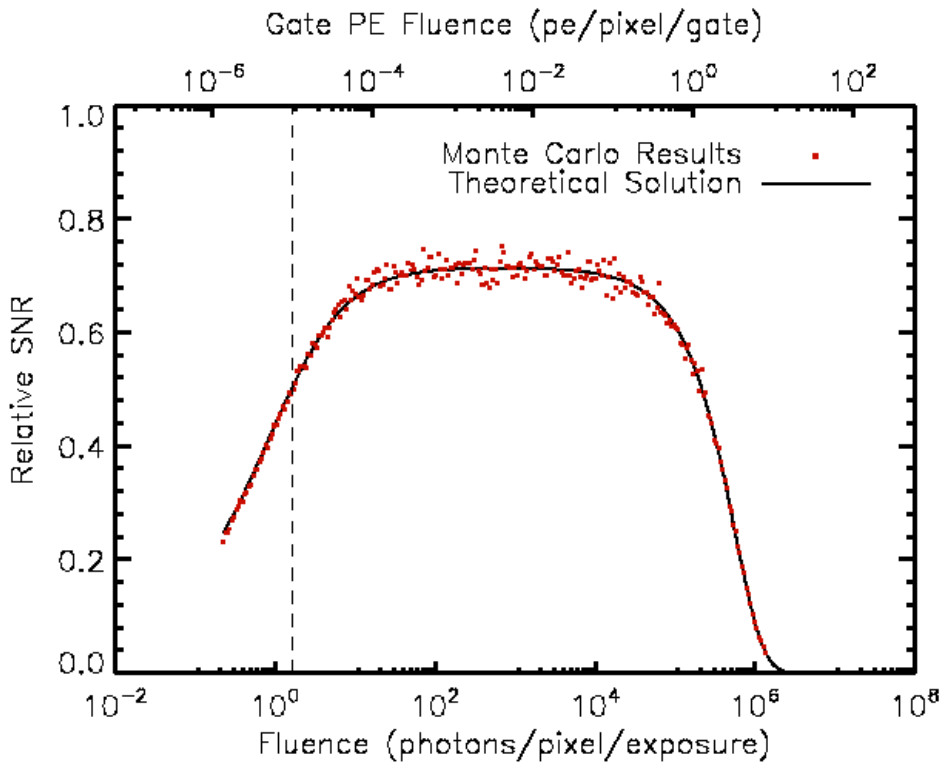

Figure 26 – This plot shows Monte Carlo results vs. analytical solution for the relative SNR of a GM-APD in photon counting mode over a range of fluence values. DCR is 1 Hz. The dashed vertical line notes the fluence at which photo-generated signal and noise contributions are equal. Gate length is 10 μs, exposure time is 1 s, PDE is 60%, and duty cycle is ~85%. Relative SNR is normalized to the ideal SNR, the shot-noise limited case where SNR = $\sqrt{Fluence}$.

Since there is no read noise and no CIC in a GM-APD, Eq. 43 correctly models the variance of the estimate neglecting afterpulsing. Any dead time losses are built into the signal and the noise because the estimates are always based on multiples of gate fluence. For example, if the estimate of the mean number of photons per gate is 2 (assume that QE is 1), the number of gates is 100, but the duty cycle is 50%, then the estimate of the total number of photons in the exposure is $2 \cdot 100 = 200$ (instead of the total incident fluence of 400).

2.6.3 *SNR WITH AFTERPULSING*

To account for the variance added due to afterpulsing, the probability of a gate equal to 1 must be amended to include the probability of an afterpulse carrier. To begin, the new definition of p (the probability of a gate equal to 1) is

$$P(gate = 1) = P(afterpulse \cup 1|\lambda)$$
$$= P(afterpulse) + P(1|\lambda) - P(afterpulse)P(1|\lambda) \quad \text{Eq. 45}$$

Here, P(afterpulse) is the probability that one or more afterpulse carriers are present during the gate, and P(1|λ) is the probability of one or more photon-generated or dark carriers are present during the gate (where $\lambda = QE \cdot \lambda_p + \lambda_d$). The probability of an afterpulse carrier being present during the gate, while dependent on the previous gate avalanche probability, is statistically independent of the current gate. Therefore, they are not mutually exclusive and the probability of either happening is not a simple sum. This leads to the subtraction of the cross term in Eq. 45. P(1|λ) is equal to Eq. 29, but the derivation of P(afterpulse), which will be referred to as p_{aft} going forward, is more involved. The derivation of avalanche probability given a certain afterpulse probability requires a few assumptions. For this derivation to be valid, the following must be true.

1. The pixels are disarmed at the end of the gate and no avalanche events occur between gates (see section 2.6.1).
2. There is no (or insignificant) dependence on gates previous to the gate immediately preceding to the gate of interest (i.e., the probability of an avalanche in the current gate is only a function of the state of the gate immediately before it and the photo- and dark carrier generation process).
3. There is no significant delayed crosstalk from neighboring pixels (i.e., afterpulsing occurs only as a result of the same pixel's previous state, not a neighboring pixel's previous state).

Physically, p_{aft} is the integral of the exponential decay function of afterpulse arrival time from the beginning of the gate to the end of the gate (i.e., from $t_{hold\text{-}off}$ to $t_{hold\text{-}off} + t_{gate}$ if t = 0 is the time of the previous avalanche). With passive quenching and clocking relative to the avalanche, this estimate has zero error. However, with gated clocking of the circuit, paft theoretically changes from gate to gate because the effective quench time is dependent on when the previous

gate's avalanche occurred. In gated operation, the precise avalanche arrival time is unknown, though an average value of p_{aft} may be sufficient since the measurement involves many gates over the course of the exposure.

The first assumption listed above is easily confirmed by determining the method of operation of the device in question, while the third can be confirmed with avalanche correlation tests between neighboring pixels. The second assumption does imply some constraints in the amount of afterpulsing in order for this analysis to be relevant. The assumption hinges on the behavior and amount of traps in the pixel. If the exponential decay function of de-trapped carriers extends past two arm periods with significant probability, then there is a certain probability that an avalanche will occur due to an avalanche in a gate more than one arm period before. If one assumes that 1% is a negligible probability for an afterpulse-induced avalanche two arm periods later, then it is easy to calculate the maximum decay lifetime allowed for assumption 2 to be valid. For example, in an exposure with t_{gate} = 10 μs and an 85% duty cycle, the total arm period is ~11.8 μs. If we assume that dark- or photo-induced avalanches occur near the beginning of the gate on average (Poisson arrival statistics), then the effective time between populating a trap and the beginning of the next gate is roughly equal to the arm period. Therefore, the decay lifetime must be less than or equal to $\frac{2}{-\ln(0.01)} t_{arm\ period}$, or in this case the decay lifetime must be less than or equal to 5.1 μs. For valid results, the maximum p_{aft} measured using the expression derived in this paper is the integral of the exponential decay over the following gate, or p_{aft} = 0.09. However, as gate times decrease and duty cycle increases, the maximum valid p_{aft} also increases.

In practice, the afterpulse lifetime can be estimated by varying the gate length and accounting for dark count probability (Nakata, et al., 2008). P_{aft} is the integral of the exponential decay function of afterpulse arrival time from the beginning of the gate to the end of the gate (i.e., from $t_{hold\text{-}off}$ to $t_{hold\text{-}off} + t_{gate}$ if t = 0 is the time of the previous avalanche). With passive quenching and clocking relative to the avalanche, this estimate has zero error. However, in gated operation, p_{aft} changes from gate to gate because the effective quench time is dependent on when the previous gate's avalanche occurred. In gated operation, the precise avalanche arrival time is unknown, though an average value of paft may be sufficient since the measurement involves many gates over the

course of the exposure. Some researchers choose to "blank" or ignore the gates following a recorded event to reduce the afterpulsing noise (Ben-Michael, et al., 2006). This may be useful for instances of high afterpulse probabilities, but the drawback is a significant decrease in the number of gates (the number of samples) and the duty cycle of the device. Depending on the afterpulsing probability, blanking may not increase the SNR of the measurement.

To continue the derivation, the probability of an afterpulse in first gate (n=0) is zero, since there were no previous gates. Therefore, the probability of an avalanche is

$$P_0(gate = 1) = 0 + \left(1 - e^{-\lambda}\right) - 0 = 1 - e^{-\lambda} \qquad \text{Eq. 46}$$

For the second gate (n=1), the probability becomes more complicated.

$$\begin{aligned} P_1(gate = 1) &= p_{aft}\left(1 - e^{-\lambda}\right) + \left(1 - e^{-\lambda}\right) - p_{aft}\left(1 - e^{-\lambda}\right)\left(1 - e^{-\lambda}\right) \\ &= \left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda}\right) \end{aligned} \qquad \text{Eq. 47}$$

The first term, $p_{aft}\left(1 - e^{-\lambda}\right)$, is the probability that an afterpulse carrier is present in the second gate, and the second term, $\left(1 - e^{-\lambda}\right)$, is the probability that a photo-generated electron or dark carrier is present in the second gate (recall that $\lambda = PDE \cdot \lambda_p + \lambda_d$). The third term is the cross term that must be subtracted, since the first two terms (probabilities) are independent of one another (see Eq. 45).

Moving on to the third gate (n=2), a pattern begins to emerge.

$$P_2(gate = 1) = p_{aft}\left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda}\right)$$
$$+\left(1 - e^{-\lambda}\right) - \left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda}\right)\left(1 - e^{-\lambda}\right) \quad \text{Eq. 48}$$
$$= \left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda} + \left(p_{aft}e^{-\lambda}\right)^2\right)$$

That pattern continues for the fourth gate (n=3):

$$P_3(gate = 1) = p_{aft}\left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda} + \left(p_{aft}e^{-\lambda}\right)^2\right)$$
$$+\left(1 - e^{-\lambda}\right) - \left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda} + \left(p_{aft}e^{-\lambda}\right)^2\right)\left(1 - e^{-\lambda}\right) \quad \text{Eq. 49}$$
$$= \left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda} + \left(p_{aft}e^{-\lambda}\right)^2 + \left(p_{aft}e^{-\lambda}\right)^3\right)$$

The expression can now be simplified into sum notation.

$$P(gate = 1) = \left(1 - e^{-\lambda}\right)\sum_{n=0}^{N}\left(p_{aft}e^{-\lambda}\right)^n \quad \text{Eq. 50}$$

where P(gate = 1) is the avalanche probability. N is the total number of gates in the exposure (n_{gates}), and n is the number of an individual gate. P_{aft} is the probability that an afterpulsing carrier is present during a given gate, and $\lambda = PDE \cdot \lambda_p + \lambda_d$. For large values of N, the upper limit of the sum can be assumed infinite since $p_{aft}\left(1 - e^{-\lambda}\right)$ is always less than one, and higher order terms will be very small. Making these assumptions, the sum in Eq. 50 becomes a Maclaurin series that converges.

$$\sum_{n=0}^{\infty}\left(p_{aft}e^{-\lambda}\right)^{n} = \frac{1}{1-p_{aft}e^{-\lambda}} \qquad \text{Eq. 51}$$

And therefore

$$P(gate = 1) = \frac{1-e^{-\lambda}}{1-p_{aft}e^{-\lambda}} \qquad \text{Eq. 52}$$

for large values of n_{gates}. This probability (the probability of one or more electrons being present in a gate given a certain afterpulsing probability) is based on a compound Poisson distribution that skews from the standard distribution given the same mean (Vinogradov, et al., 2009).

Eq. 52 behaves as expected. When $p_{aft} = 0$, the probability of a gate equal to 1 is simply Eq. 29. Similarly, when ($\lambda = PDE \cdot \lambda_p + \lambda_d$) >> 1, the probability of a gate equal to one is Eq. 29 again; the additional number of gates triggered due to afterpulse carriers approaches zero because the other carrier generation rates are very high. Figure 27 shows the relationship between avalanche probability, gate fluence, and afterpulse probability.

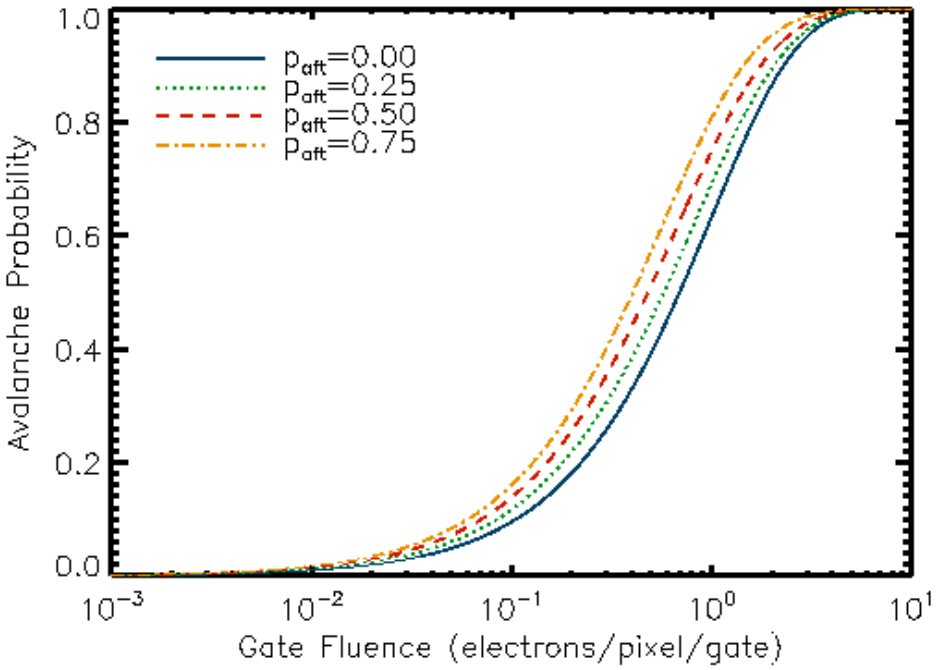

Figure 27 –Analytical solution for the probability of a triggered gate for various afterpulse probabilities. Higher afterpulse probabilities cause higher avalanche probabilities for the same gate fluence. Monte Carlo simulations match this behavior.

Since $p_{aft} < 1$ and $e^{-\lambda} \leq 1$, Eq. 50 quickly converges for large values of N. As long as the total number of gates is greater than 10^3, the error due to assuming an average avalanche probability will be negligible even for very high values of p_{aft}.

It is interesting to note that the probability of a gate equal to 1 due solely to afterpulse carriers has a peak value. At a certain point, the probability of a coincidence of afterpulse and any other carrier is significant, and so most of the gates with afterpulse carriers in them would have been triggered anyway. The probability that an afterpulse carrier alone is responsible for a triggered gate is

$$P(afterpulse\ only) = P(afterpulse\ \cap 0|\lambda) = P(afterpulse)P(0|\lambda) \qquad \text{Eq. 53}$$

Again, because the afterpulse condition is independent of the current gate, the two probabilities can be multiplied to form the overall probability. For the first gate (n=0):

$$P(afterpulse\ only) = 0 \cdot e^{-\lambda} = 0 \qquad \text{Eq. 54}$$

This makes sense, since there was no previous gate to create an afterpulse carrier. Moving on to the second gate (n=1),

$$P(afterpulse\ only) = \left(1 - e^{-\lambda}\right)p_{aft}e^{-\lambda} \qquad \text{Eq. 55}$$

where Eq. 46 provides the probability of an avalanche in the previous (first) gate to define P(afterpulse). Using Eq. 47 in the same fashion, the probability in the third gate is

$$\begin{aligned} P(afterpulse\ only) &= p_{aft}\left(1 - e^{-\lambda}\right)\left(1 + p_{aft}e^{-\lambda}\right)e^{-\lambda} \\ &= \left(1 - e^{-\lambda}\right)\left(p_{aft}e^{-\lambda} + \left(p_{aft}e^{-\lambda}\right)^2\right) \end{aligned} \qquad \text{Eq. 56}$$

Here, a familiar pattern emerges, except that the first term in the sum is now 0 instead of 1. Using the same assumptions as for Eq. 51, the expression can be simplified to

$$\begin{aligned} P(\text{afterpulse only}) &= \left(1 - \mathrm{e}^{-\lambda}\right)\sum_{\mathrm{n}=1}^{\infty}\left(\mathrm{p_{aft}e}^{-\lambda}\right)^{\mathrm{n}} = \frac{1}{1 - \mathrm{p_{aft}e}^{-\lambda}} - 1 \\ &= \left(\frac{1 - \mathrm{e}^{-\lambda}}{1 - \mathrm{p_{aft}e}^{-\lambda}}\right)\mathrm{p_{aft}e}^{-\lambda} \end{aligned} \qquad \text{Eq. 57}$$

As shown in Figure 28, P(afterpulse only) is low for low fluence and increases for higher fluences until the probability that an avalanche would occur without afterpulsing (afterpulse carriers coincide with a dark count or photon-generated carriers) becomes significant.

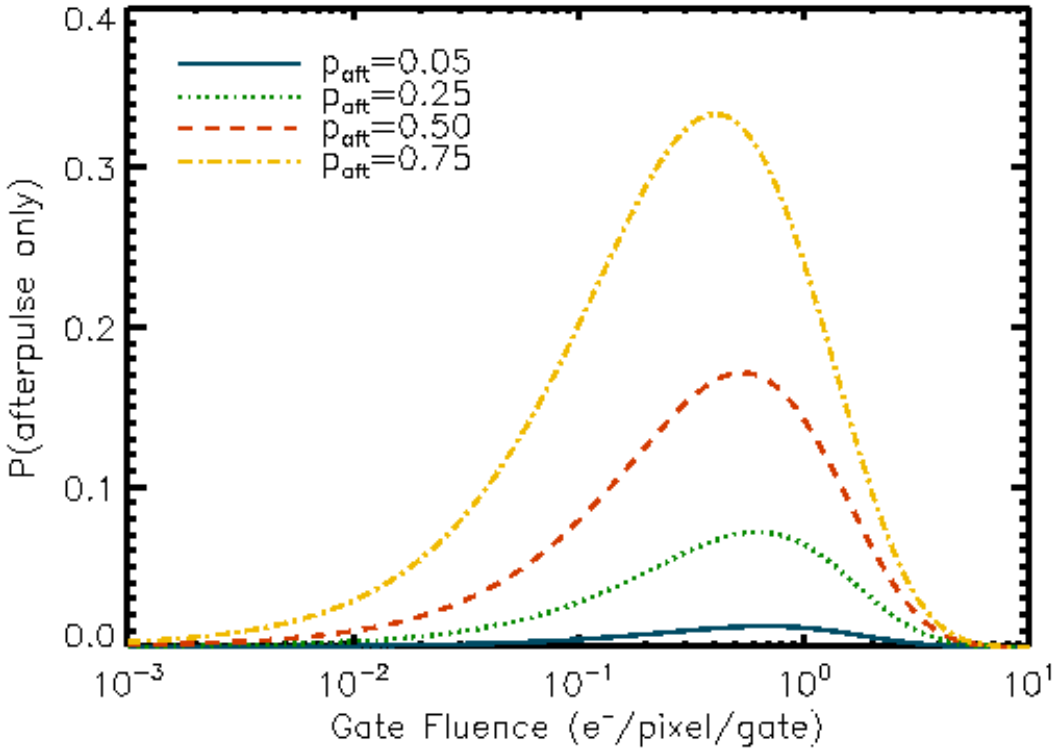

Figure 28 – Theoretical solutions for the probability of a triggered gate due solely to the afterpulse contribution as a function of gate electron fluence (photon-generated and dark carriers) for various afterpulse probabilities.

The peak value of the probability function (the gate fluence with the most noise due to afterpulsing) is where the derivative of Eq. 57 with respect to λ equals 0, or

$$\lambda_{peak} = ln\left(1 + \sqrt{1 - p_{aft}}\right) \qquad \text{Eq. 58}$$

To find a new estimate for the average number of photons per gate, Eq. 29 must be solved again, but with the new expression for P(n≥1). $\frac{n_{ones}}{n_{gates}}$ (avalanche probability) is now equal to Eq. 52, and the estimate of the mean number of photo-generated carriers per gate (λ_p) is now

$$\hat{\lambda}_p = -ln\left[\frac{1-\frac{n_{ones}}{n_{gates}}}{1-p_{aft}\frac{n_{ones}}{n_{gates}}}\right] - \lambda_d \qquad \text{Eq. 59}$$

Going back to Eq. 40, the variance is a function of the first derivative of the estimate and the variance of $\frac{n_{ones}}{n_{gates}} = x$. The latter value is not the same as Eq. 42, however, and Figure 29 illustrates this point. For a standard binomial distribution the comprising Bernoulli trials are independent tests. In the case of afterpulsing, however, each trial has a correlation to the trial before. This means that the distribution will have the same mean (the probability derived in Eq. 52), but the variance will be increased. Some trials will have many interactions between Bernoulli trials and the total number of triggered gates will be large, while other trials may result in very few triggered gates. The variance will increase as a function of the afterpulse probability.

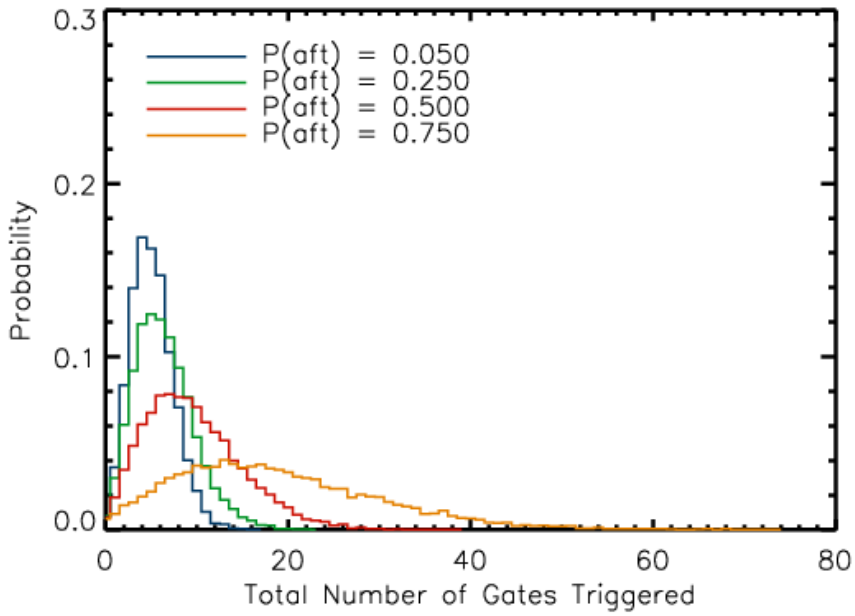

Figure 29 – Monte Carlo results for the probability density function of the total number of triggered gates in an exposure in 1000 gates for various afterpulse probabilities. Each curve is for a gate electron fluence of 0.0048 electrons.

As a point of comparison, Figure 30 shows the expected binomial distribution for an exposure with the avalanche probability equal to Eq. 52, but with no correlation between gates (Bernoulli trials). The real distribution with the self-correlated binomial distribution is overlaid.

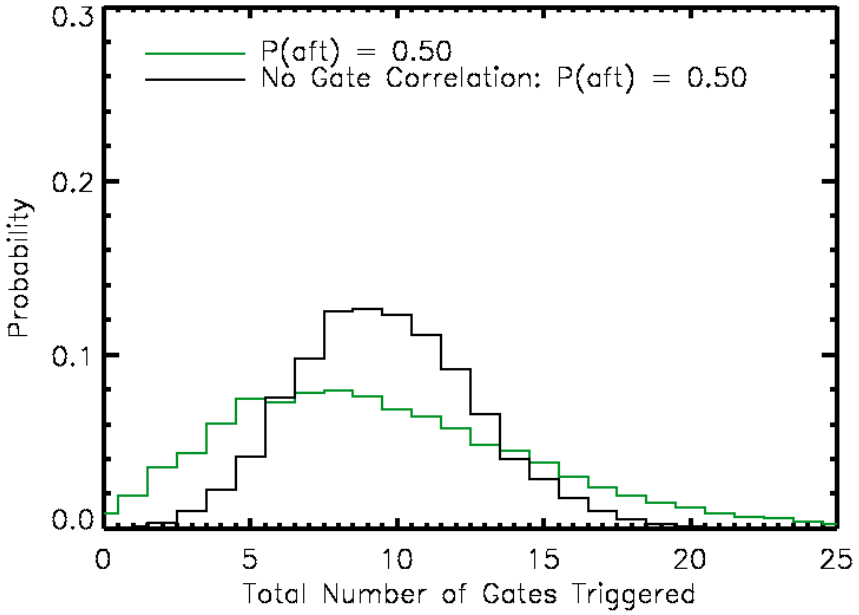

Figure 30 – Monte Carlo results for the probability density function of the total number of triggered gates in an exposure of 1000 gates, with a constant electron flux per gate of 0.0048 electrons. The black curve shows the un-correlated Bernoulli trial experiment, while the green curve shows the real distribution with inter-gate dependencies due to afterpulsing.

The correlated data set has the same mean (~9.5), but larger variance (wider distribution) than the standard binomial distribution, which indicates that the variance of x must be dependent on the afterpulsing probability as well. Unlike the previous derivation of the variance for the number of counts, however, this variance is derived via the Markov Chain method.

The Markov Chain method ascertains the distribution of events that are dependent on the present state and nothing else (Papoulis & Pillai, 2002). In this case, the "present state" is the n^{th} gate value, while the predicted value is for the $(n+1)^{th}$ gate (dependent on the present state). This particular process is discrete, which simplifies the derivation somewhat.

To start the derivation, the probabilities of every state transition must be defined. Figure 31 shows a state diagram for the case of afterpulsing and gate values.

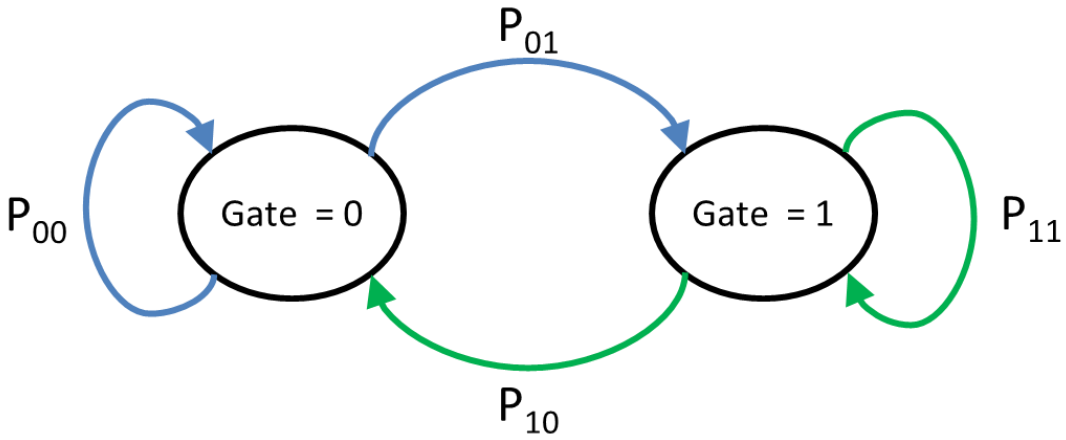

Figure 31 – State diagram for gate values from the n^{th} to $(n+1)^{th}$ gates

P_{01} is the probability of a 1 in the $(n+1)^{th}$ state given a 0 in the n^{th} state, and so on for the other probabilities. The sum of the probabilities leaving a state must equal 1. Eq. 60 – Eq. 63 show the expressions for each probability (λ is the total electron fluence from photon and dark current processes during a single gate).

$$P_{01} = 1 - e^{-\lambda} \quad \text{Eq. 60}$$

$$P_{00} = 1 - P_{01} = e^{-\lambda} \quad \text{Eq. 61}$$

$$P_{11} = p_{aft} + P_{01} - p_{aft}P_{01} = 1 - e^{-\lambda}(1 - p_{aft}) \quad \text{Eq. 62}$$

$$P_{10} = 1 - P_{11} = e^{-\lambda}(1 - p_{aft}) \quad \text{Eq. 63}$$

Since afterpulsing has no effect on gates after a zero (per the assumptions in this derivation), the probability for P_{00} and P_{01} are straightforward Poisson probabilities. For P_{11}, the probability of an afterpulse or a λ-generated carrier is the sum of both probabilities minus the cross term. The equation for P_{10} follows. These probabilities comprise the matrix M that calculates the probability of a one and zero based on the probabilities for the previous gate (see Eq. 64). The columns of M are the probabilities leaving each state.

$$\begin{bmatrix} p_0(n+1) \\ p_1(n+1) \end{bmatrix} = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \begin{bmatrix} p_0(n) \\ p_1(n) \end{bmatrix} \quad \text{Eq. 64}$$

In a steady state approximation, where the probability of 1 or 0 is independent of the outcome of a gate many gates before, Eq. 64 reduces to Eq. 65 and further to Eq. 66.

$$\begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} \quad \text{Eq. 65}$$

The determinant of M is zero as a result of the defined state relationships, and $P_1 + P_0 = 1$. Using these relationships, Eq. 65 can be solved to find P_1 and P_0 (the steady state probabilities for 1 and 0, respectively). The expression for P_1 in Eq. 66 should match the expression previously derived in Eq. 52, as a check.

$$\begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} e^{-\lambda} & e^{-\lambda}(1-p_{aft}) \\ 1-e^{-\lambda} & 1-e^{-\lambda}(1-p_{aft}) \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} \rightarrow P_1 e^{-\lambda}(1-p_{aft}) = P_0(1-e^{-\lambda})$$

$$P_1 = \frac{1-e^{-\lambda}}{1-p_{aft}e^{-\lambda}} \qquad \text{Eq. 66}$$

$$P_0 = 1 - P_1 = \frac{e^{-\lambda}(1-p_{aft})}{1-p_{aft}e^{-\lambda}}$$

To calculate the variance of the number of counts, the standard definition of variance will be used and interpreted in terms of M and the state probabilities, as shown in Eq. 67.

$$\begin{aligned} {\sigma_c}^2 = E[c^2] - (E[c])^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} \overline{p_1(i)p_1(j)} - \overline{p_1(i)}\ \overline{p_1(j)} \\ &= \sum_{k=-(N-1)}^{N-1} (N-|k|)\left(P_1 p_1(k) - {P_1}^2\right) \\ &= (N-0)P_1(1-P_1) \quad + 2P_1 \sum_{k=1}^{N-1}(N-k)(p_1(k)-P_1) \\ &= NP_1P_0 + 2p_1 \sum_{k=1}^{N-1}(N-k)(p_1(k)-P_1) \end{aligned} \qquad \text{Eq. 67}$$

where N is the number of gates (n_{gates}), k = € – j, and $p_1(k)$ is the probability of a 1 given that a 1 was recorded k gates before. $P_1(k)$ is also, by definition, the bottom right corner term of M^k, or

$$p_1(k) = [0 \quad 1]M^k \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{Eq. 68}$$

To find a closed-form solution for $p_1(k)$, M must be diagonalized, or a matrix A must be found such that

$$A^{-1}MA = D$$
$$M^k = A^{-1}D^kA \qquad \text{Eq. 69}$$

where the columns of A are the eigenvectors of M and D is a diagonal matrix of the eigenvalues of M (Anthony & Harvey, 2012). The eigenvalues of M and the matrices A and D are defined in Eq. 70-Eq. 72.

$$eigenvalues = \begin{bmatrix} 1 \\ p_{aft}e^{-\lambda} \end{bmatrix} \qquad \text{Eq. 70}$$

$$A = \begin{bmatrix} \frac{e^{-\lambda}(1 - p_{aft})}{1 - e^{-\lambda}} & -1 \\ 1 & 1 \end{bmatrix} \qquad \text{Eq. 71}$$

$$D = \begin{bmatrix} 1 & 0 \\ 0 & p_{aft}e^{-\lambda} \end{bmatrix} \qquad \text{Eq. 72}$$

Since D is a diagonal matrix, D^k (Eq. 69) is calculated simply as a matrix of the k^{th} power of the individual terms. Referring back to Eq. 68 and Eq. 69, $p_1(k)$ can be calculated as in Eq. 73.

$$p_1(k) = [0 \quad 1]A^{-1}D^kA\begin{bmatrix}0\\1\end{bmatrix} = \frac{\left(1-e^{-\lambda}\right)+e^{-\lambda}\left(1-p_{aft}\right)\left(p_{aft}e^{-\lambda}\right)^k}{1-p_{aft}e^{-\lambda}}$$

Eq. 73

$$= \frac{p_{01}+p_{10}\left(p_{aft}e^{-\lambda}\right)^k}{1-p_{aft}e^{-\lambda}}$$

Figure 32 shows both the Monte Carlo results and analytical solution for $p_1(k)$. Note that as k approaches infinity (many gates), $p_1(k)$ approaches P_1, which is the steady state value of avalanche probability.

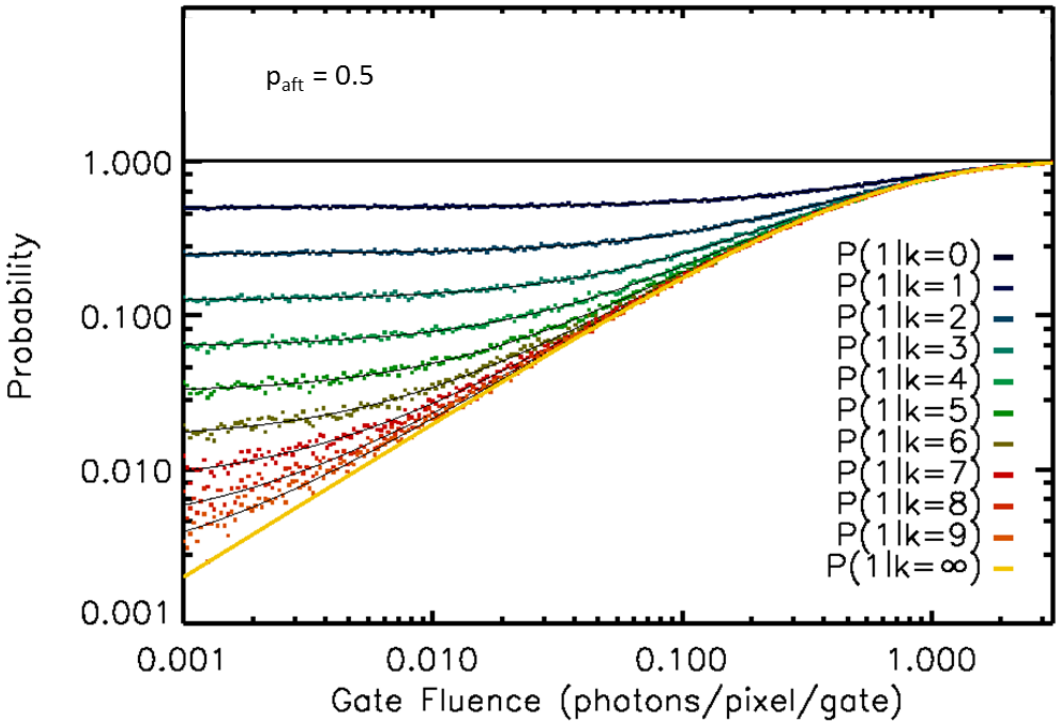

Figure 32 – $p_1(k)$ as a function of fluence and afterpulse probability. Colored points represent Monte Carlo results and the solid lines represent the analytical solution for each value of k.

Substituting Eq. 73 into Eq. 67, and using some geometric series identities, the variance of the total number of counts is

$$\sigma_c{}^2 = P_1 P_0 \left(N + 2 \frac{p_{aft} e^{-\lambda} \left(\left(p_{aft} e^{-\lambda} \right)^N + N\left(1 - p_{aft} e^{-\lambda}\right) - 1 \right)}{\left(1 - p_{aft} e^{-\lambda}\right)^2} \right) \quad \text{Eq. 74}$$

If N is assumed to be very large, the expression simplifies to

$$\sigma_c{}^2 = N P_1 P_0 \left(1 + 2 \frac{p_{aft} e^{-\lambda}}{1 - p_{aft} e^{-\lambda}} \right) \quad \text{Eq. 75}$$

It is interesting to note that the variance is a scaled version of the standard binomial variance. Eq. 75 is the variance of the total number of counts, but the variance of x is a scaled version, or

$$\sigma_x{}^2 = \frac{\sigma_c{}^2}{N^2} = \frac{P_1 P_0}{N} \left(1 + 2 \frac{p_{aft} e^{-\lambda}}{1 - p_{aft} e^{-\lambda}} \right) \quad \text{Eq. 76}$$

Now, going forward with the derivation of the SNR, Eq. 40 still provides the correct expression for the variance of the estimate.

$$\hat{\lambda}_p = y(\mu_x) = -ln\left(\frac{1 - \mu_x}{1 - p_{aft} \cdot \mu_x} \right) \quad \text{Eq. 77}$$

Solving Eq. 59 in terms of λ and letting $y = \lambda$ and μ_x = P(gate=1) from Eq. 52, the total variance for the entire exposure is

$$\sigma_y{}^2 = |y'(\mu_x)|^2 \sigma_x{}^2 n_{gates}{}^2$$
$$= NP_1P_0\left(\frac{1-p_{aft}}{P_0\left(1-p_{aft}\cdot P_1\right)}\right)^2\left(1+2\frac{p_{aft}e^{-\lambda}}{1-p_{aft}e^{-\lambda}}\right) \quad \text{Eq. 78}$$

When p_{aft} is zero, the variance of the estimate reduces to Eq. 40. Combining Eq. 78 with the first half of Eq. 43, the SNR of a GM-APD in photon counting mode is

$$SNR = \frac{QE\cdot\lambda_p\cdot n_{gates}}{\sqrt{P_1P_0\left(1+2\frac{p_{aft}e^{-\lambda}}{1-p_{aft}e^{-\lambda}}\right)\left(\frac{1-p_{aft}}{P_0\left(1-p_{aft}\cdot P_1\right)}\right)^2 n_{gates}}} \quad \text{Eq. 79}$$

where

$$P_1 = \frac{1-e^{-\lambda}}{1-p_{aft}e^{-\lambda}}$$

$$P_0 = \frac{e^{-\lambda}\left(1-p_{aft}\right)}{1-p_{aft}e^{-\lambda}}$$

$$\lambda = PDE\cdot\lambda_p + \lambda_d$$

Note that Eq. 79 simplifies to Eq. 44 (SNR neglecting afterpulsing) when $p_{aft} = 0$. Figure 33 shows an overlay of Monte-Carlo simulation results and the theoretical solution according to Eq. 79 for the same detector. The simulation agreed with the theoretical data in both mean and standard deviation.

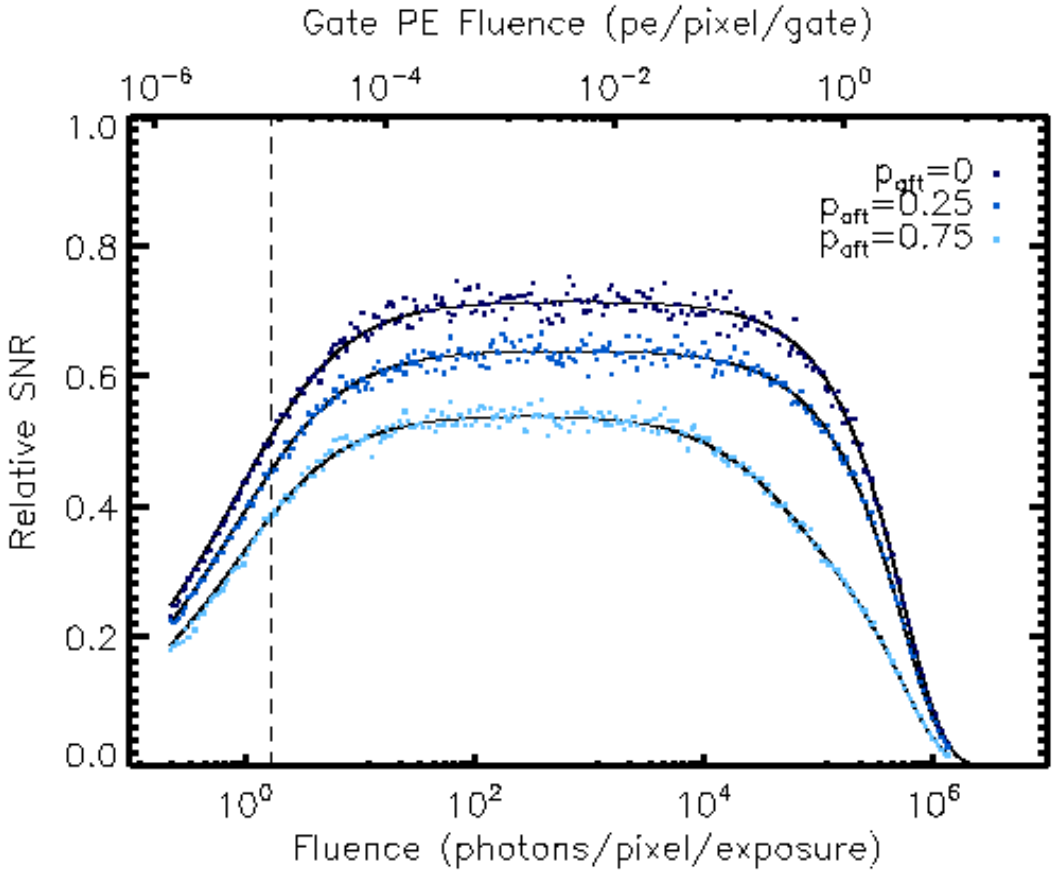

Figure 33 – This plot shows Monte Carlo results (individual points) and analytical solutions (corresponding solid lines) for the relative SNR of a GM-APD in photon counting mode vs gate fluence for multiple afterpulse probabilities. The dashed vertical line notes the fluence at which photo-generated signal and DCR contributions are equal. Relative SNR is normalized to the ideal SNR, the shot-noise limited case where SNR = $\sqrt{Fluence}$.

It is important to note that, as explained at the beginning of section 2.6.3, an actual value of p_{aft} = 0.75 or even 0.25 is unlikely for most operating conditions if the assumptions stated at the beginning of section 2.6.3 hold. However, since the simulation is based on the same assumptions as the derived expression for SNR, and the inputs are given without regard to feasibility, the comparison of simulated to calculated results in Figure 33 is valid. The exaggerated values of p_{aft} more easily illustrate the overall trends in SNR behavior across a range of fluence values as the afterpulsing probability changes.

The earlier onset of roll-off at high fluence for larger values of p_{aft} is due to an effective decrease in saturation level. Given the same fluence, the avalanche probability will increase with

increasing afterpulse probability. The roll-off at low fluence is still due to background noise (DCR). While the relative SNR still has a maximum of $\sqrt{PDE}$ for the case of $p_{aft} = 0$, the maximum for cases where $p_{aft} > 0$ is lower.

2.7 PHOTON-COUNTING MODE EMCCD

In photon-counting mode, EMCCDs are read out very quickly and a threshold circuit is used to decide whether or not a photon was absorbed during the very short exposure. The short exposures are repeated many times to estimate the probability of the arrival of one or more photons in one short exposure. This probability is used to calculate the average number of photons per second per pixel.

2.7.1 *THEORY OF OPERATION*

The operation of an EMCCD in photon-counting mode is very similar to operation in analog-mode. The only difference is in the timing and the data analysis. In terms of timing, the short exposures should be set such that the average electron fluence (e^-/pixel/gate) is near 1.

2.7.2 *SNR*

The derivation of an SNR expression for EMCCDs in PC mode builds on concepts encountered in the GM-APD SNR derivation. To start, consider the physical meaning of the string of 1s and 0s that are the output of the detector. The system will output a 1 when the magnitude of carriers at the output is greater than or equal to a threshold value. Therefore, the probability of a gate equal to 1 is the probability density function of the carriers at the output of the gain register, integrated from the threshold value to infinity. Figure 34 shows an example of a gain register output probability density function and the probability of a gate equal to 1.

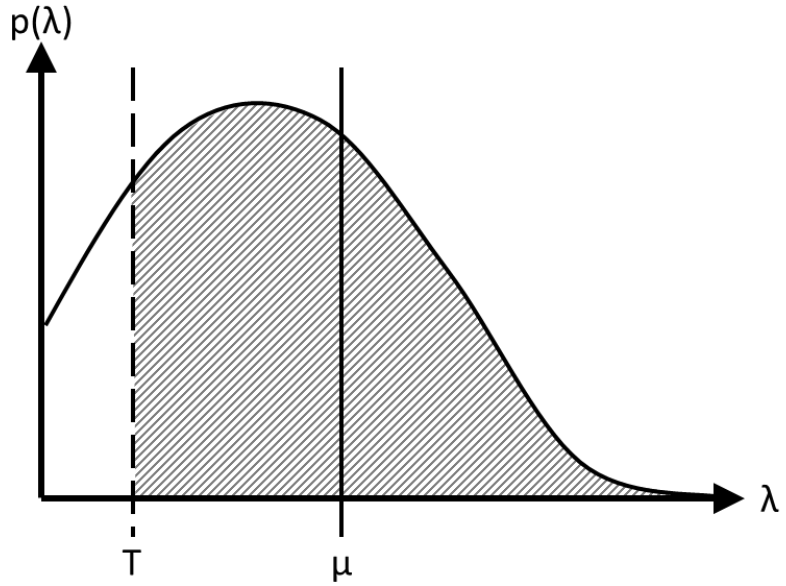

Figure 34 – This figure shows a sample output probability density function and the key values to derive the probability of a 1 for an EMCCD in PC mode. T is the threshold, μ is the distribution mean, and the shaded area is the integral representing the probability. Λ is the number of carriers at the output of the gain register.

The probability distribution of the number of output carriers is simplified when one considers the output to be the sum of the outputs for each individual input carrier. For example, if there are 3 electrons at the input of the gain register, the number of output electrons is the sum of 3 random variables chosen from the correct gain probability density function. Since, for a single carrier, the gain is equal to the number of output carriers, this simplifies the derivation.

The probability distribution for the number of output carriers for one input carrier (and thus the distribution of the gain value) is an exponential distribution with mean G (Daigle, 2009). The probability distribution of the sum of multiple random variables is the convolution of the probability distributions of each random variable. Conveniently, the convolution of n exponentially-distributed random variables (with the same rate constant) has a closed-form solution, shown in Eq. 80 (Akkouchi, 2008).

$$f_n(t) = \frac{\beta^n t^{n-1}}{(n-1)!} e^{-t\beta} \qquad \text{Eq. 80}$$

$f_n(t)$ is the probability density function for n exponentially-distributed random variables, t is the random variable assigned to the output, and β is the rate constant of the input distributions. For the case of the EMCCD gain register, n is the number of input carriers (the number of input distributions to convolve), t is the actual gain value for any given trial, and β is the reciprocal of the mean gain. Therefore,

$$f_n(g) = \frac{(G)^{-n} g^{n-1}}{(n-1)!} e^{-\frac{g}{G}} \qquad \text{Eq. 81}$$

where g is the actual gain (random variable), G is the mean gain, and n is the number of input carriers to the gain register. The integral of Eq. 81 from the threshold value to infinity is the shaded region of Figure 34, or the probability of a gate equal to 1. Even though the random variable for the gain must be discrete, since the gain values are large (>100), it is reasonable to treat the distribution as continuous.

$$\begin{aligned} P(gate = 1 \mid n) &= \int_T^\infty \frac{(G)^{-n} g^{n-1}}{(n-1)!} e^{-\frac{g}{G}} \, dg \\ &= \frac{(G)^{-n}}{(n-1)!} \int_T^\infty g^{n-1} e^{-\frac{g}{G}} \, dg \end{aligned} \qquad \text{Eq. 82}$$

Here it is helpful to note the definition of the incomplete gamma function,

$$\Gamma(n,x) \equiv \int_x^{\infty} \frac{u^{n-1}}{e^u} du \qquad \text{Eq. 83}$$

With a simple substitution of u = g/G, Eq. 82 becomes

$$\int_T^{\infty} \frac{(G)^{-n} g^{n-1}}{(n-1)!} e^{-\frac{g}{G}}\ dg = \frac{G \cdot (G)^{-n} G^{n-1}}{(n-1)!} \int_T^{\infty} \frac{u^{n-1}}{e^u}\ dg$$

$$= \frac{\Gamma\left(n, \frac{T}{G}\right)}{(n-1)!} \qquad \text{Eq. 84}$$

where x → T/G because the probability can be thought of as the probability that the input is greater than or equal to the threshold scaled by the mean gain (as opposed to the probability that the input, when multiplied, is greater than or equal to the threshold). This simplifies the expression.

At this point in the derivation, the probability of a gate equal to 1 is known for a given number of input carriers, n. The probability of a gate equal to 1 for a random number of input carriers is the more useful probability, however. This probability is the sum of the product of the probability of n carriers at the input and the probability that n carriers will be greater than the threshold at the output, over all possible values of n.

$$P(gate = 1) = \sum_{k=1}^{\infty} \left(\frac{e^{-\lambda}\lambda^k}{k!}\right)\left(\frac{\Gamma\left(k, \frac{T}{G}\right)}{(k-1)!}\right) \qquad \text{Eq. 85}$$

The sum in Eq. 85 begins at 1 because the probability that zero input carriers will be greater than the threshold at the output is zero. The first term is the Poisson probability of k carriers, where λ

is the number of absorbed photons per gate on average, while the second term is taken from Eq. 84. Eq. 85 does not take into account the dark noise, the parallel CIC in the array, the serial CIC in the gain register or the read noise.

Since the probability of a parallel CIC carrier is small in general, it can be described as a Bernoulli trial variable (see Eq. 20). However, when the Bernoulli trial results for each transfer are added together, the distribution can be described as Poissonian, since each carrier represents a rare event. Once this assumption is made, the parallel CIC rate can be added to the photo-generated carrier and dark carrier rates to create a Poisson random variable that defines the distribution for the number of carriers at the input of the gain register.

$$\frac{n_{ones}}{n_{gates}} = P(gate = 1)$$

$$= \sum_{k=1}^{\infty} \left(\frac{e^{-(\lambda_p + \lambda_d + \lambda_{pCIC})} \left(\lambda_p + \lambda_d + \lambda_{CIC}\right)^k}{k!} \right) \left(\frac{\Gamma\left(k, \frac{T}{G}\right)}{(k-1)!} \right) \quad \text{Eq. 86}$$

Rearranging Eq. 86 and solving for $\hat{\lambda}_p$,

$$\hat{\lambda}_p = -ln\left(\frac{x}{\sum_{k=1}^{\infty} \frac{\left(\hat{\lambda}_p + \lambda_d + \lambda_{CIC}\right)^k \Gamma(k, T/G)}{k!\,(k-1)!}} \right) - (\lambda_d + \lambda_{CIC}) \quad \text{Eq. 87}$$

where x is the ratio of the number of 1s recorded to the total number of gates recorded, $P(gate = 1)$. The estimate is a function of itself because of the exponential term, but the solution may be found arithmetically.

The variance of the estimate in Eq. 87 is now more complicated to derive, since the estimate is now a function of itself. However, by rearranging Eq. 40, the variance can be found when the

equation is inverted (solved for the variable as a function of the estimate). The function x(y) has no intuitive meaning, but the mathematics still hold. Letting $\frac{n_{ones}}{n_{gates}} = x$ and $\hat{\lambda}_p = y$, the variance per gate is

$$\sigma_y{}^2 = \frac{\sigma_x{}^2}{|x'(y)|^2} \qquad \text{Eq. 88}$$

The derivative of the function x(y) is

$$x'\left(\hat{\lambda}_p\right) = \frac{d}{d\hat{\lambda}_p}\left(e^{-\left(\hat{\lambda}_p+\lambda_d+\lambda_{CIC}\right)}\sum_{k=1}^{\infty}\frac{\left(\hat{\lambda}_p+\lambda_d+\lambda_{CIC}\right)^k\Gamma\left(k,\frac{T}{G}\right)}{k!\,(k-1)!}\right) \qquad \text{Eq. 89}$$

$$= e^{-\left(\hat{\lambda}_p+\lambda_d+\lambda_{pCIC}\right)}\sum_{k=1}^{\infty}\frac{\left(\hat{\lambda}_p+\lambda_d+\lambda_{CIC}\right)^{k-1}\Gamma\left(k,\frac{T}{G}\right)\left(k-\left(\hat{\lambda}_p+\lambda_d+\lambda_{CIC}\right)\right)}{k!\,(k-1)!}$$

The variance of x is simply Combining Eq. 88 and Eq. 89 yields Eq. 90.

$$SNR = \frac{QE \cdot \lambda_p \cdot n_{gates}}{\sqrt{x(1-x)n_{gates}}} \cdot \left(e^{-(\lambda)} \sum_{k=1}^{\infty} \frac{(\lambda)^{k-1} \Gamma\left(k, \frac{T}{G}\right)(k-\lambda)}{k!\,(k-1)!} \right)$$

Where

$$\lambda = QE \cdot \lambda_p + \lambda_{\mathrm{d}} + \lambda_{\mathrm{CIC}} \qquad \text{Eq. 90}$$

$$x = \frac{n_{ones}}{n_{gates}} = \sum_{k=1}^{\infty} \left(\frac{e^{-(\lambda)}(\lambda)^k}{k!} \right) \left(\frac{\Gamma\left(k, \frac{T}{G}\right)}{(k-1)!} \right)$$

This equation assumes that read noise and serial CIC from the multiplication register are negligible. The thresholding may be set so that it is above the sCIC level for much of the time, though the probability distribution of the output of the gain register for sCIC is not derived here. In order for the read noise to be negligible, e.g. an error less than 0.1%, Eq. 91 must be true.

$$\frac{1}{2}\left(1 - erf\left(\frac{T}{\sigma_r\sqrt{2}}\right)\right) \leq 0.001$$

$$\sigma_r \leq \sim\frac{T}{3} \;\; or \;\; T \geq 3\sigma_r \qquad \text{Eq. 91}$$

The first expression is the integral from T to infinity of a Gaussian distribution with mean zero and standard deviation σ_r. The contribution from the read noise causing a change in gate status from 0 to 1 is assumed to cancel itself out (an equal number of trials will results in the change from 1 to 0). Eq. 91 represents the cases where, absent any input to the gain register, the read noise alone will cause the gate to be recorded as a 1. It is also assumed that any secondary interaction between the read noise and the sCIC noise is negligible.

Figure 35 shows the SNR of an EMCCD in photon-counting mode, both Monte Carlo results and the analytical solution from Eq. 90. Monte Carlo results below 0.02 e^-/s/pixel have a problem with the estimation of very small numbers in the incomplete gamma function.

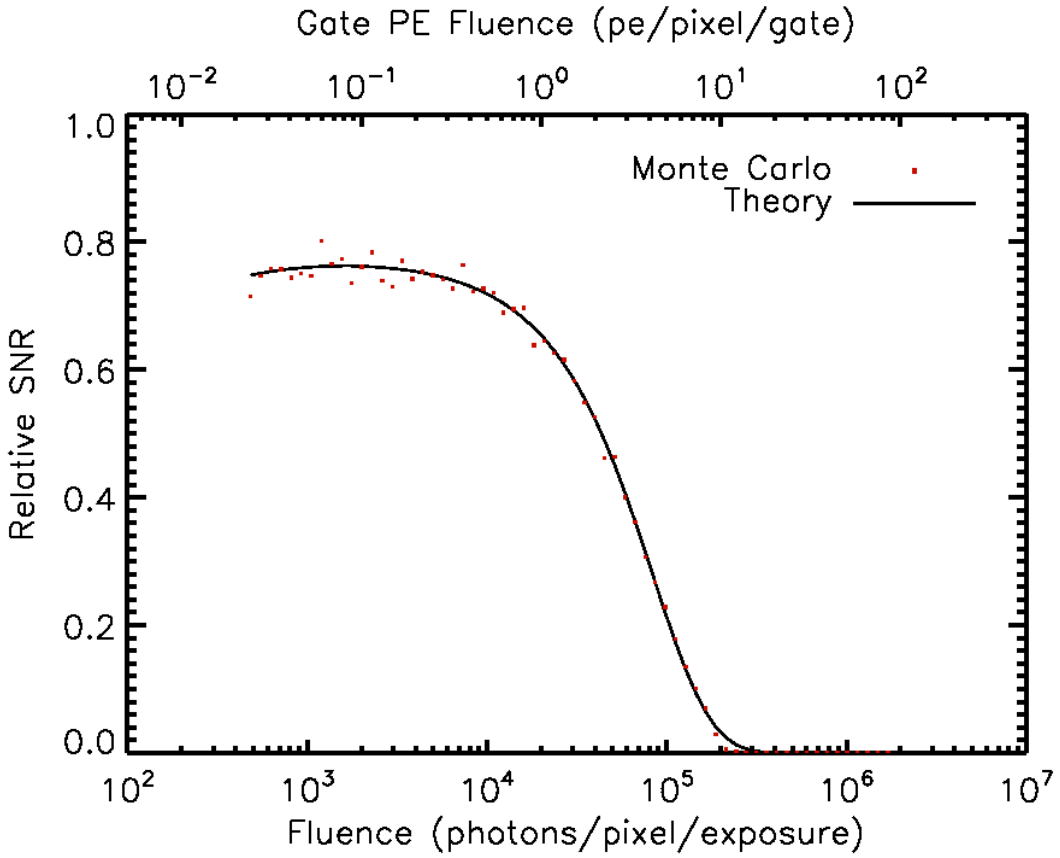

Figure 35 – Monte Carlo results vs. analytical solution for the SNR of an EMCCD in photon-counting mode over a range of fluences. The gate length is set to 0.05s, the QE is 80%, and the exposure time is 800s (15686 gates with a frame read time of 1 ms). The dark noise (both dark current and pCIC) is 0.002 electrons per gate. The gain is set to 100σ_r (read noise is negligible). Relative SNR is the SNR of the device normalized to the ideal SNR, the shot-noise limited case.

Figure 36 shows the derived SNR for this research and the experimentally-calculated SNR for identical detector characteristics (Daigle, 2009). The gate time is 50 ms, the dark current and CIC contribution per gate is 0.0023 e^-/pix, the QE is 95%, and the threshold is set to 5x the read noise with a gain of 3000.

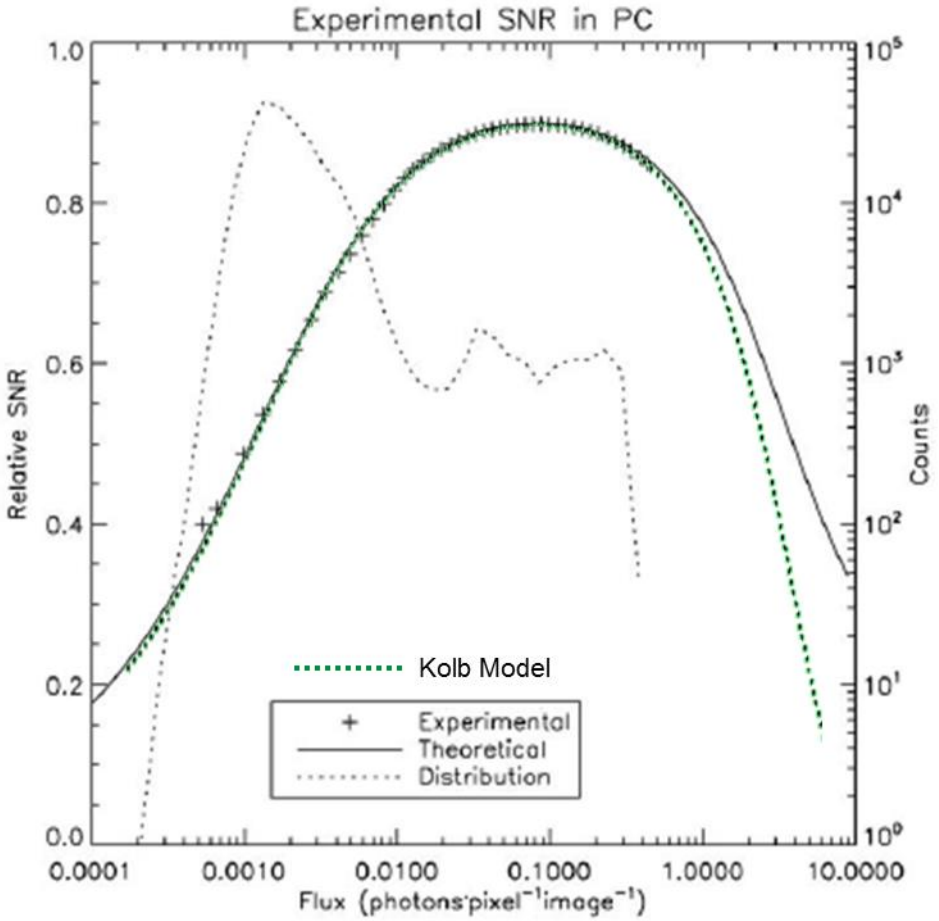

Figure 36 – This plot shows derived results (green dotted line) and published experimental results (individual + signs) for an EMCCD in photon-counting mode with the same settings and noise values. The thin solid line is the SNR expression used in the reference. The only discrepancy is in the shape of curve during saturation, but there is no experimental data to compare the theoretical curves. The derived results have a slightly better fit to the data points at the highest flux values. The derived results are overlaid on the published data figure with the same axis scaling (Daigle, 2009).

The derived results are in good agreement with the published experimental results, though there is a discrepancy in the saturation behavior of the model used by the published source and the derived model presented here. The model used in the paper is very simplistic and does not take into account the noise due to thresholding, which adversely affects the SNR when the device begins to saturate. The derived model in this proposal matches the measured data at the beginning of saturation better than the model presented in the published source, though the difference in fit is small.

Eq. 90 can be simplified if a few assumptions are made. Instead of simplifying the SNR equation as a whole, a more convenient simplification is available at the beginning of the derivation. If the read noise contribution is assumed to be negligible and the thresholding loss is also considered to be negligible, then the probability of a 1 (see Eq. 85) can be expressed exactly like that of a GM-APD, as in Eq. 29.

Comparing Eq. 85 to the Poisson probability expression in Eq. 28, the similarities are clear. As $\left(\frac{\Gamma\left(k,\frac{T}{G}\right)}{(k-1)!}\right)$ Approaches 1, Eq. 85 approaches P(k≥1). For an error of less than 1%, the ratio of the threshold to the mean gain must be less than or equal to 0.01, while maintain the threshold at $\geq 3\sigma_r$ as in Eq. 91. In this case, the term may be dropped from the expression. Figure 37 shows the incomplete gamma function term subtracted from 1 as a function of k and T/G. At T/G ratios less than 0.01, the error in assuming that the function is always 1 becomes very small.

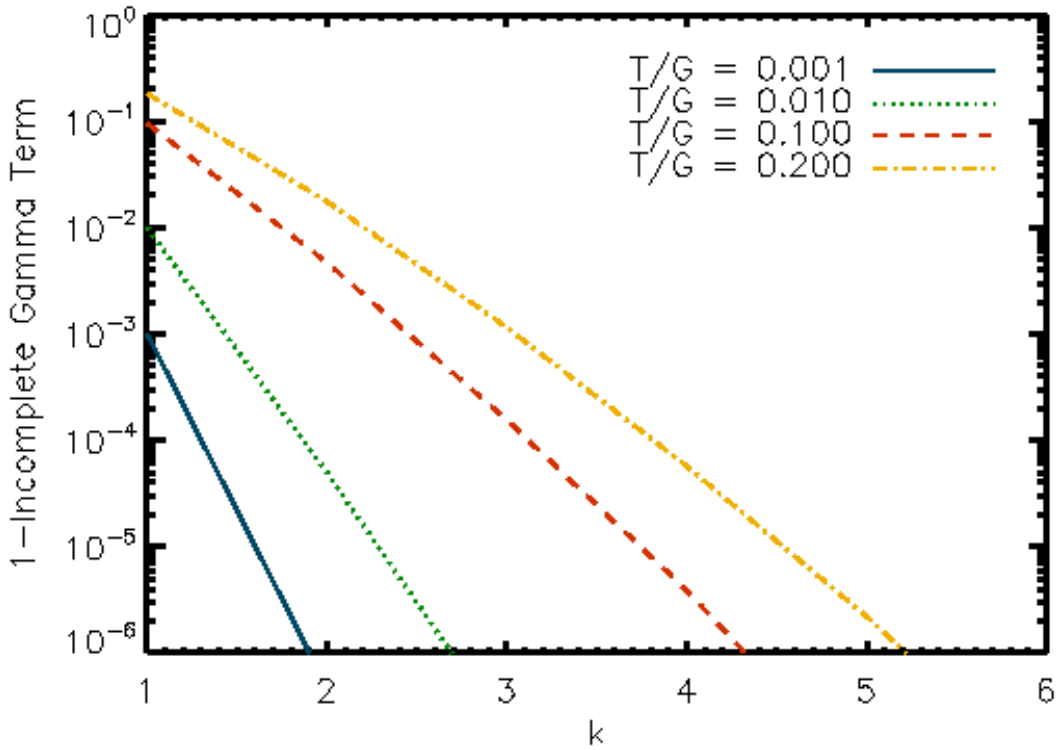

Figure 37 – Incomplete gamma function term for various k and fractional threshold (T/G) values

In this approximation, any carrier that is multiplied is counted (no thresholding loss) and no read noise is counted (zero read noise). Therefore, the derivation becomes identical to the GM-APD case with no afterpulsing probability. The only difference is that the CIC must be counted as a contributor to the dark current and scaled appropriately, since it is applied per read as opposed to per unit time.

3 GM-APD TESTING

The GM-APD devices testing in this project were the product of MIT Lincoln Laboratory (MIT LL). The CfD and MIT LL worked together on a project funded by the Gordon and Betty Moore Foundation, entitled "Zero Noise Detector for the Thirty Meter Telescope." The detector arrays were fabricated at MIT LL facilities, and the readout integrated circuits (ROICs) were fabricated by IBM and bump-bonded to the detector arrays to make the complete device. Early on in the project, testing of on-wafer test devices characterized multiple internal architectures for the detector arrays (Kolb, 2011). These architectures included low-fill-factor (LFF), medium-fill-factor (MFF), and high-fill-factor (HFF) architectures. Figure 38 shows the internal structure of a single pixel in an LFF device, which was originally designed for light detection and ranging (LIDAR) systems. The MFF and HFF designs had larger multiplication and absorption regions. The HFF devices collect more photons and have higher efficiencies than the LFF devices. However, due to extraordinarily high crosstalk (see section 3.11), only the LFF devices are tested fully for this project.

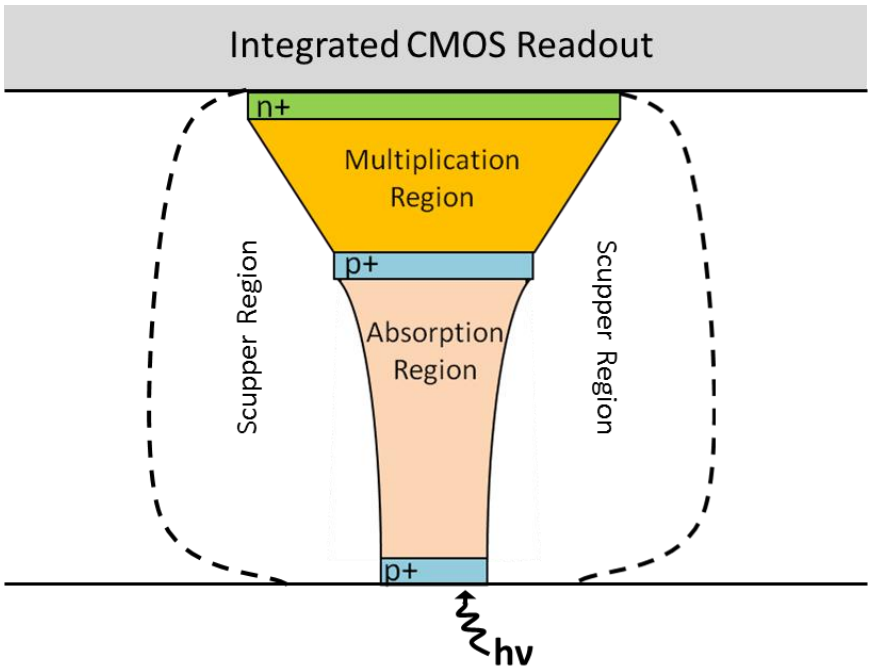

Figure 38 – This figure shows the GM-APD design for one pixel (not to scale).

Each pixel in the LFF devices is made up of three distinct regions, each with a specific function. The absorber depletion region has a medium-strength electric field that moves carriers to the multiplier region, which has a strong electric field (above breakdown voltage) to facilitate avalanches. A weak electric field, called a "scupper," surrounds the absorption and multiplication regions of each pixel to direct carriers generated outside of these regions to the cathode without initiating an avalanche, reducing the DCR (Aull, et al., 2010).

Except for the implanted regions, the silicon material is nearly intrinsic. It is very lightly p-doped with boron with a carrier concentration of 8.5×10^{13} cm^{-3}. The boron-doped silicon is epitaxially grown and is 15 μm thick. The carrier lifetime is 10-100 μs. The n+ implant is circular and doped with arsenic, with a peak concentration of 4×10^{18} cm^{-3} at the surface. The spread of this implant is narrow, with a spatial distribution decay constant of 0.1 μm. The n+ implant has a diameter of 13 μm. The p+ implant between the absorption and multiplication regions is also circular, and has a diameter of 9 μm. The peak concentration of the implant is 8.5×10^{16} cm^{-3}, with a narrow distribution ($\pm$ 0.17 μm) centered at 1 μm from the silicon surface. There is a phosphorus guard ring implant that encircles the n+ implant to spread out the field lines at the edge of the multiplication region. The inner diameter of the ring is 12 μm and the outer diameter is 14 μm. The implanted ring creates a low-doped halo around the periphery of the diode. Between pixels, there is a channel-stop implant, which is designed to electrically isolate each pixel. It is blocked in each pixel over a circular region, which encompasses the implants shown in Figure 38, with a diameter of 19 μm. The illuminated surface of the device is underneath 725 μm of quartz, which does not significantly affect the reflectance of light at visible wavelengths. There is an anti-reflection coating at the illuminated surface of the device: a 65.8 nm thick layer of Ti_3O_5. Figure 39 shows the reflectance at the silicon/quartz interface with the anti-reflective coating for various wavelengths.

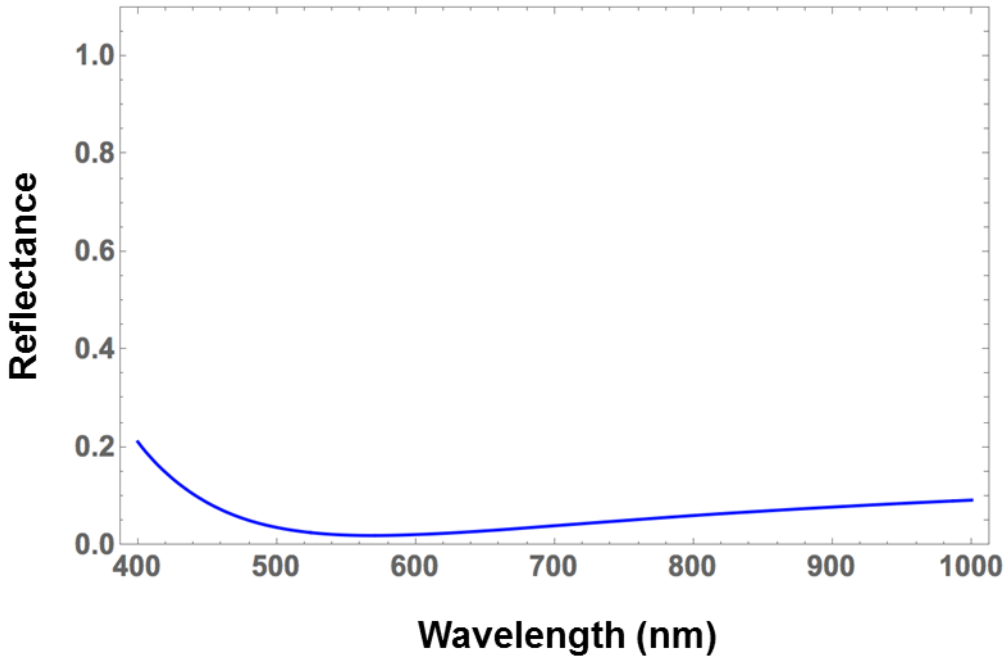

Figure 39 – This figure shows reflectance vs wavelength at the quart-silicon interface with a 68.5 nm layer of Ti_3O_5. The incident light is assumed to be perpendicular to the surface of the detector.

A GM-APD exposure is comprised of many individual detection cycles, each of which contains three main signals. At the start of the cycle, an arm pulse is asserted to set the reverse bias across the diodes above the breakdown voltage. The voltage is kept high for a short period of time (generally on the order of 0.1 μs), after which an individual pixel may or may not avalanche. In the event of an avalanche, the quenching circuit detects an increase in current and actively sets the voltage below the breakdown voltage, or disarms the pixel, to stop the avalanche. At the end of the gate (the time that the voltage is allowed to be high), the pixel state is recorded, a "one" or "zero," corresponding to whether the pixel experienced an avalanche. After the state is recorded, the pixels are actively disarmed, regardless of state, and after a specified delay (the hold-off time) the pixel is armed again.

The output of a pixel from a given exposure is the total number of avalanches recorded. The ratio of avalanches to number of gates is the avalanche probability, from which the flux can be estimated. Since this estimation method is fundamentally different from that used for CCD or CMOS detectors, the form of the expression for the SNR is fundamentally different, as well. For

the detector discussed here, the dominant noise characteristics for SNR are determined by DCR, afterpulsing probability, and PDE. DCR and PDE are analogous to dark current and QE, but DCR and PDE values include the avalanche initiation probability, which degrades the device's ability to detect a carrier when it is less than 1 (McIntyre, 1973).

The GM-APD array devices were mounted in the dewar with a custom mount. The detectors were positioned so that they could be exposed to light when the dewar port cover was removed (see Figure 58). Figure 40 shows three GM-APD detectors mounted in the dewar before installing the heat shield and the dewar lid.

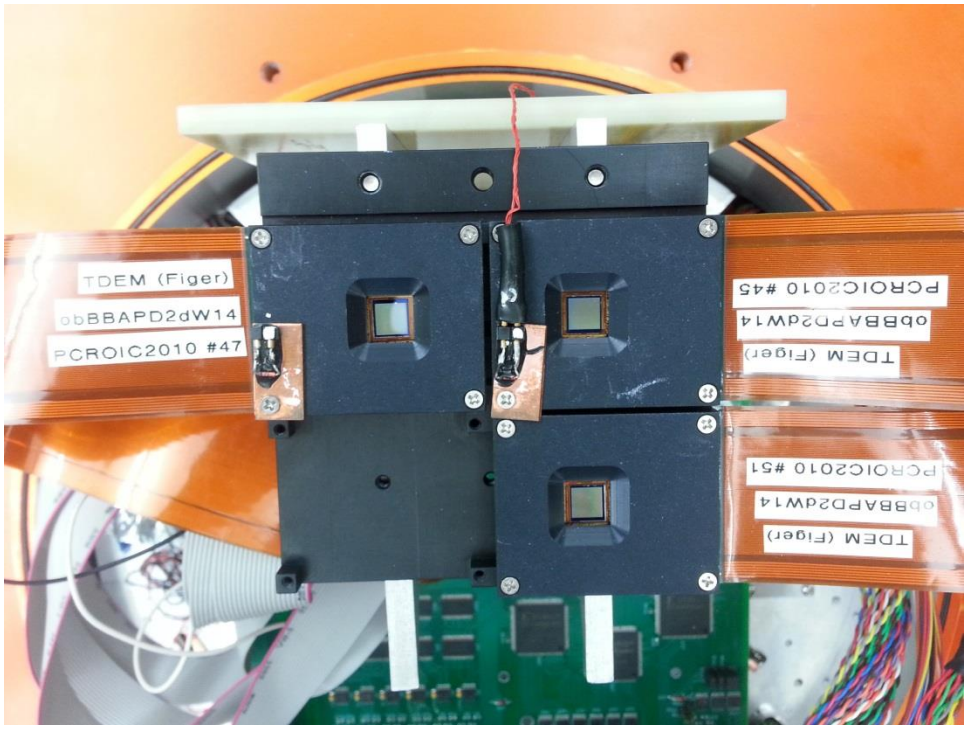

Figure 40 – Three GM-APD detectors are shown mounted in the dewar. The empty quadrant was used to house a temperature sensor mounted to a flex package for detector temperature measurements.

GM-APD device characterization included evaluation of DCR, afterpulsing probability (P_{aft}), intra-pixel sensitivity (IPS), PDE, and the effects of radiation damage on each. Crosstalk for these devices is negligible, roughly 0.5%, and was not affected by radiation damage. Post-radiation data was taken at the same applied bias as the pre-radiation data, though this represents

a change in applied overbias due to a shift in breakdown voltage after irradiation. Although three devices were tested, the devices were very consistent with each other and only one set of data is presented here for brevity.

3.1 ELECTRONICS

A number of circuit boards and external power supplies are required to operate the GM-APD detectors. The system of control is referred to as the Photon Counting ReadOut Integrated Circuit (PCROIC) system, after the ROIC implemented in the GM-APD imaging devices. Four FPGAs are required to do the necessary signal processing. The FPGAs in this system have very basic programming that consists of simple in/out and enable commands, and they are programmed via JTAG protocols.

If only one device were being operated, a single circuit board would be required (called the Warm Electronics Board, or WEB). The WEB has an Opal Kelly clocking board and some signal conditioning components. Two power supplies are required to operate the WEB, one at 10 V and another at 3.3 V. The connector for supplying those voltages is shown on the bottom wall of the warm electronics housing in Figure 41.

In order to test up to four devices at once, however, two more circuit boards had to be used. One is used to split the signals to each detector into differential signals for better signal integrity while being fed through the dewar wall (called the Warm Daughter Board, or WDB), and the other is used to make the signals single-ended again before being sent on to the detectors (called the Cold Fanout Board, or CFB). The high voltage applied to the GM-APD devices is provided by dedicated power supply connections on the WDB, via ports on the outside of the warm electronics housing (top wall in Figure 41). A separate power supply is required for each detector that is installed in the dewar for testing. The WEB (left) and WDB (right) are shown in Figure 41 inside the warm electronics housing. The Opal Kelly control (USB connection) and the CameraLink connector are accessible through the left wall of the housing, and the differentiated signals exit through cables on the right wall of the housing.

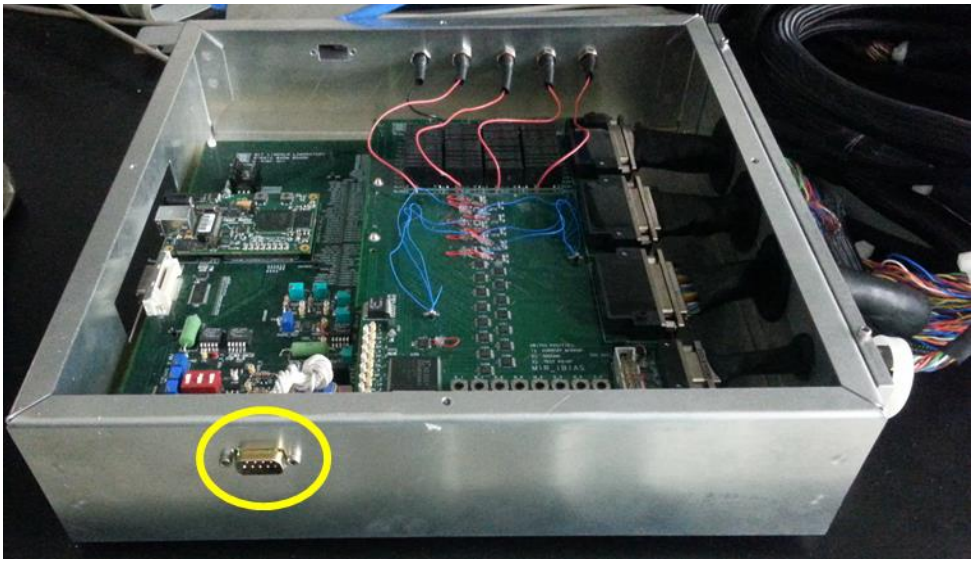

Figure 41 – The warm electronics housing with power supply connector (circled in yellow) is shown.

Along with the signals from the WDB, temperature sensor signals were connected via ports through the dewar wall. Figure 42 shows a top-down view of the electrical connections for the control and monitoring of the GM-APD devices and CFB. PCROIC PWR carries the supply voltage to the detectors as well as enabling voltages for a variety of internally-driven clock signals. PCROIC DIN carries data acquisition commands to the detectors, and PCROIC DOUT carries data acquisition responses from the detectors. PCROIC test carries signals that control the gating and timing of the detection cycle. The Lakeshore C/D cable relays temperature sensor data for monitoring the temperature of the circuit board, and Lakeshore A/B carries temperature sensor data for the control heater and the detectors. Also visible in Figure 42 are the detector raft mounts on the cold plate, as well as the thermal strap that connects the detector raft to the final stage cold finger of the closed-system cryogenic cooler.

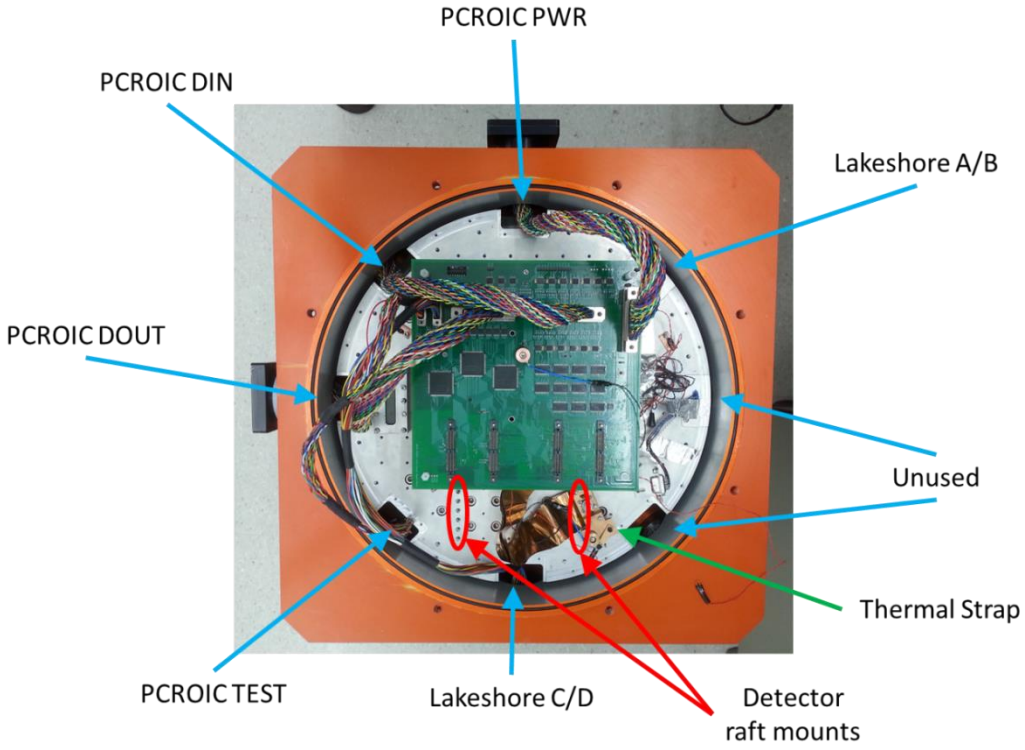

Figure 42 – An internal view of the wiring for electronic control and monitoring of the GM-APD devices is shown. The detectors are not mounted in this view.

Thermal heating straps were attached to each dewar port flange to prevent condensation and shorting of signals going into and out of the dewar while cold.

3.2 Breakdown Voltage

All GM-APD bias settings are relative to the breakdown voltage. The applied bias is defined in terms of the overbias, or the number of volts applied above the breakdown voltage. Even though the breakdown voltage may vary between devices, tests done at the same overbias setting are valid for comparison.

To estimate the breakdown voltage, the current draw from the array is measured as the voltage is varied. Ideally, every pixel in the array would have the same effective breakdown voltage, and the current-voltage curve would have a steep rise in current at the breakdown voltage. In reality, voltage sagging across the array leads to a slightly lower applied voltage in the center of the array, causing a higher effective breakdown voltage in the center. Voltage sagging occurs when

the voltage applied at the edges of the array is higher than the voltage supplied at the center of the arrays due to parasitic resistance. Process variations across the array also lead to small differences in the breakdown voltage of each device. The result of these variations is a slow increase in current as more and more pixels are biased above their breakdown voltage. Figure 43 shows an example of the measured current vs applied voltage.

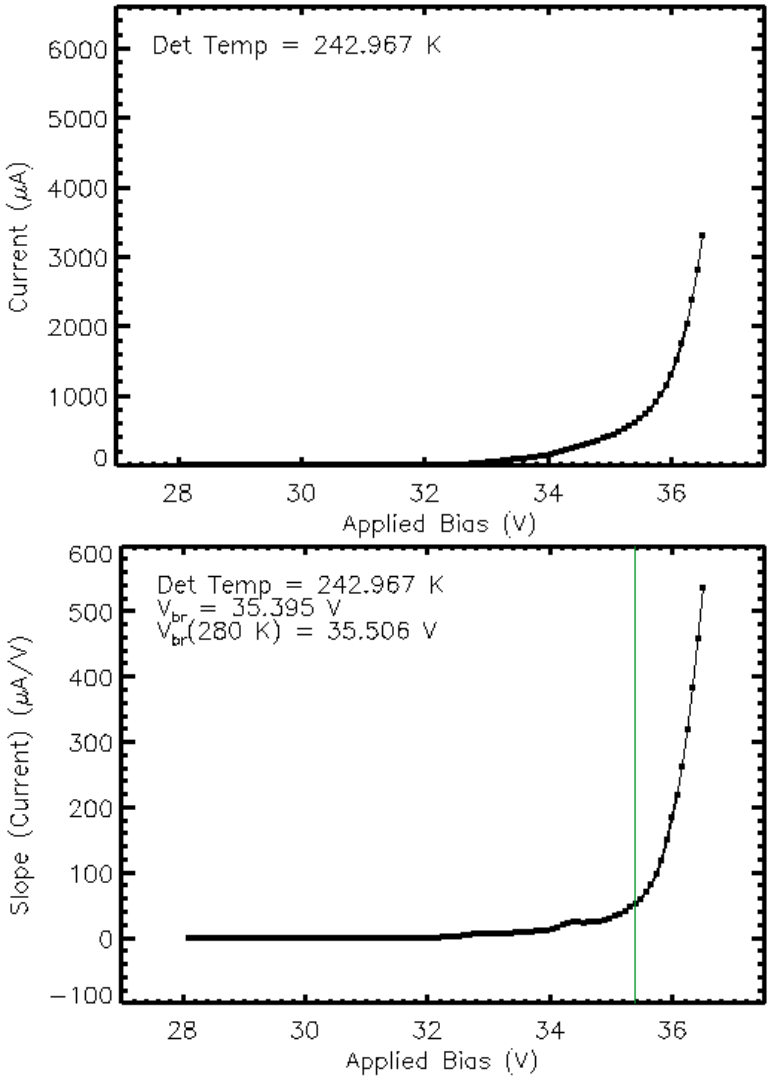

Figure 43 – This figure shows a plot of the array current vs applied voltage for a GM-APD array device (top) and the derivative of the current vs voltage curve (bottom).

Since the slope of the current is shallow until most of the pixels are in breakdown, it is easier to find the point at which most pixels are avalanching by taking the derivative of the current vs

voltage curve. The inflection point of the exponential increase is defined as the effective breakdown voltage for the array. A fitting routine attempts to fit the derivative curve to an exponential function, but the curve itself is not a perfect exponential. In the case of an exponential fit, the characteristic voltage value would be the breakdown voltage. However, failing an exponential fit, the first voltage value where the current exceeds some threshold is chosen as the breakdown voltage. As a final resort, if neither the fit nor the threshold yielded a value that represents the breakdown voltage, the value is chosen by eye. It is important to note that the applied bias in Figure 43 is an externally-applied voltage to the system. The control electronics reduce the voltage by 5 V by default (the disarm state of the device), and so the actual breakdown voltage is 5 V less than reported. In normal operation, the full bias is applied to the detector when armed.

The effective breakdown voltage decreases linearly with temperature, but the slope must be confirmed to accurately adjust the applied bias based on the detector temperature. If the applied bias is not adjusted, the overbias will vary across different temperature settings. Figure 44 shows the breakdown voltage of a GM-APD array vs temperature.

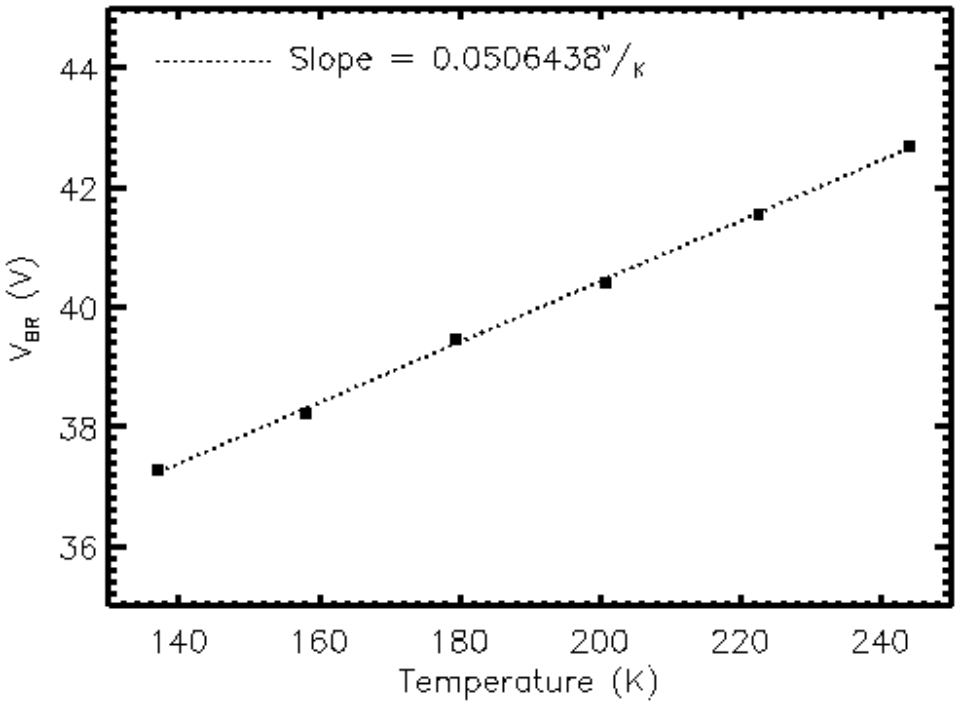

Figure 44 – This plot shows breakdown voltage vs temperature (individual points) and a linear fit to the data (dashed line).

The breakdown voltage at 280 K and the change in breakdown voltage per Kelvin were stored in a configuration file for each detector. During each experiment, the settled temperature of the detector was used to calculate the appropriate applied bias to achieve the specified overbias.

3.3 Radiation Damage Simulations

The radiation testing plan was done by CfD engineer Dr. Joong Lee (Figer, 2010). A summary of his work is provided here for reference.

In order to simulate detector performance degradation in space, due to radiation damage, equivalent radiation damage must be introduced to the detectors in a controlled environment while characterizing detector performance. Radiation damage is categorized into two types: ionizing radiation and nonionizing radiation. Generally, ionizing radiation effects scale with total ionizing dose (TID), while nonionizing radiation effects scale with displacement damage dose (DDD) in the absorber layer(s).

Because the variety of energies and particles in space will normally exceed what is practically available on the ground, some equivalence must be used to simulate the same amount and type of damage. It is often the case that monoenergetic high-energy protons are the only radiation source used in testing. A dose of 60 MeV protons that provides the same TID and DDD as the spectrum of high-energy radiation particles on orbit can be used to simulate the damage received as well as the performance degradation over time. This approach has worked well in space-qualifying several missions for imagers for *NASA*, including Chandra, ASCA, SUZAKU, and HETE (Figer, 2010). The energy of 60 MeV has been used widely for testing CCDs since it is one of the most common energies for several low-earth orbits (Edmonds, et al., 2000).

Eq. 92 relates the change in dark current for an exposure to a proton spectrum ($\Delta I_{D,spectrum}$) to the change in dark current for samples exposed to a monoenergetic beam of 60 MeV radiation ($\Delta I_{D,60\ MeV}$)..

$$\frac{\Delta I_{D,spectrum}}{\Delta I_{D,60\ MeV}} = \frac{\int NIEL(E)\phi(E)dE}{NIEL_{60\ MeV}\phi_{60\ MeV}FWHM_{60\ MeV}} \qquad \text{Eq. 92}$$

$NIEL_E$ and ϕ€ are the energy-dependent non-ionizing energy loss (NIEL) and flux of the proton spectrum (transported through any shielding) on orbit, respectively. $NIEL_{60\ MeV}$ and $\phi_{60\ MeV}$ are the NIEL and flux used in the proton irradiation experiment at the radiation testing facility, and $FWHM_{60\ MeV}$ is the full-width-at-half-maximum energy spread of the proton beam.

Any ground-based radiation testing program begins with simulation of the radiation environment and transport simulation of the radiation spectrum in space. Together, these simulations are used to calculate the radiation dose on the detector on orbit. For this purpose, SPENVIS[1] (and other simulation packages, such as SPACERAD[2]) were used. The radiation testing program considered both a low Earth orbit (LEO) and an L2 orbit location[3]. Calculations of the radiation environment were done using SPENVIS (see Figure 45). The calculations assume a 2015 launch date and five

[1] http://www.spenvis.oma.be/
[2] http://www.spacerad.com/
[3] http://earthobservatory.nasa.gov/Features/OrbitsCatalog/page2.php

year mission duration. The relevant mission parameters for TESS and WMAP, representative of the missions at LEO and L2, respectively, were used in the calculations. Because the actual dose over the mission lifetime is dependent on the relative phasing between the mission launch date and the solar cycle, Figure 45 also shows the expected radiation fluence for the duration of an entire solar cycle (Figer, 2010). A solar cycle refers to the 11-year cycling of solar radiation and ejection of solar material, which influences space weather and the amount of radiation from the sun for the Earth and objects in orbit. The mission in the L2 orbit presents a harsher radiation environment than the LEO orbit.

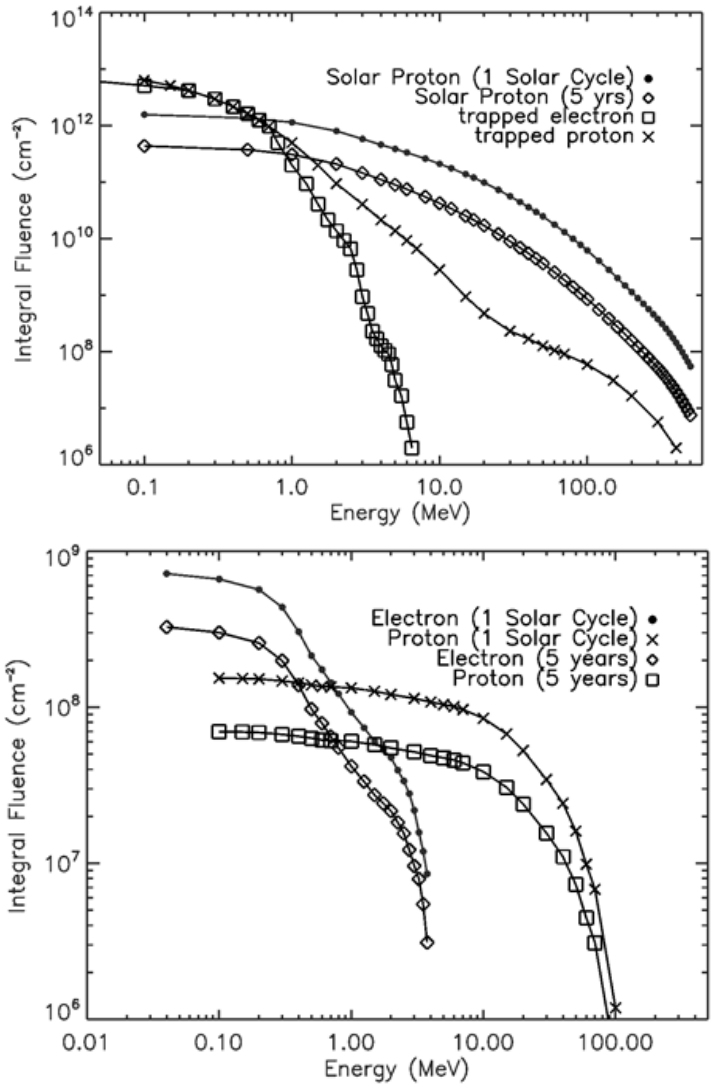

Figure 45 – This figure shows the integrated fluence, from indicated energy to infinity, vs particle energy at L2 (top) and LEO (bottom), assuming mission durations of 5 years and one solar cycle (Figer, 2010).

The radiation spectrum seen by the detector in the radiation environment is used in a radiation transport simulation to calculate the ionizing and non-ionizing radiation dose on the detector. There are a number of radiation transport packages available, in either SPENVIS or as a standalone application, such as Mulassis[4]. The results of the simulation, using Mulassis to calculate the radiation dose at the L2 orbit, are shown in Figure 46 (Figer, 2010). The L2 orbit

[4] http://reat.space.qinetiq.com/mulassis

presents a relatively benign radiation environment as the expected dose for a 5 year mission is 1 krad(Si), and for an 11 year mission is 5 krad(Si). To mitigate the uncertainties in radiation simulation, a mission duration of 11 years was assumed for the radiation testing, the length of an entire solar cycle. In addition to a single solar cycle, a margin of 10×, or 50 krad(Si), was assumed for the simulated radiation dose on orbit for the case of extreme events or extended mission lifetimes.

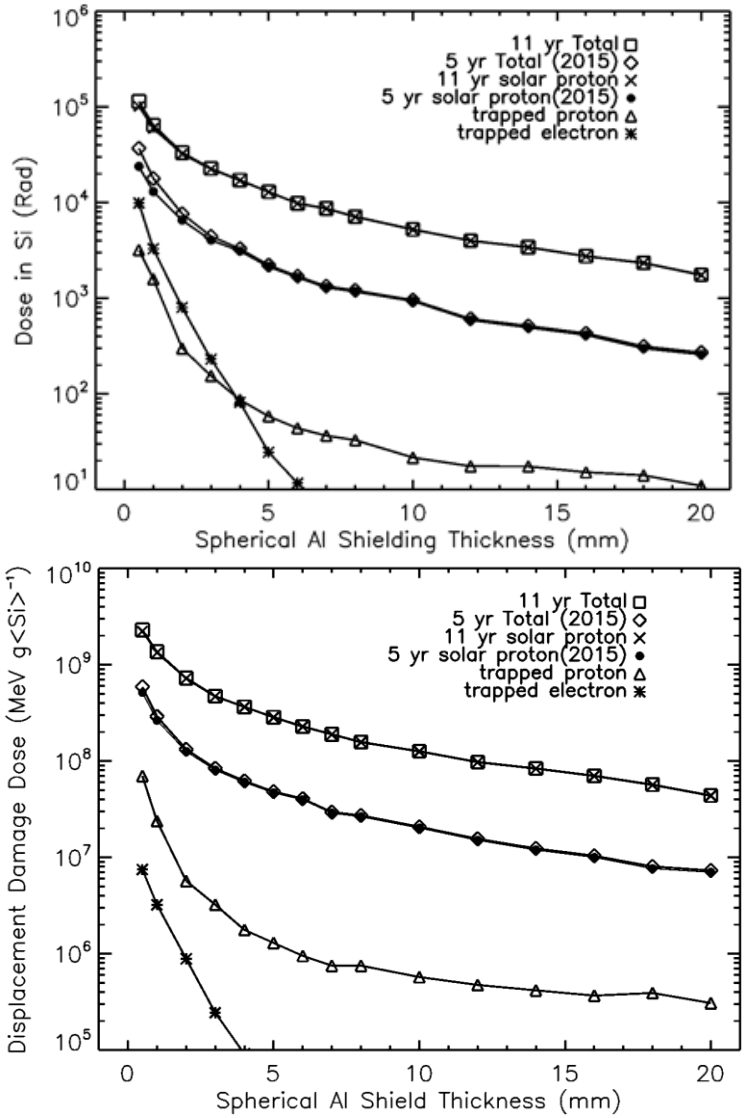

Figure 46 – This figure shows ionizing radiation dose for a mission at L2 (top) and displacement damage dose expected for a mission at L2 (bottom) (Figer, 2010).

3.4 RADIATION TESTING

Three GM-APD devices were irradiated at Massachusetts General Hospital's (MGH) Francis H. Burr Proton Therapy Laboratory with monoenergetic 60 MeV protons. They were exposed to a cumulative dose of 50 krad(Si) in geometrically spaced doses, simulating 10 solar cycles at an L2 orbit (assuming a 1 cm Al shield). The dose was calibrated behind a 0.08 in aluminum plate to simulate the dewar cover between the detectors and the radiation beam. The energy spread of the protons was 6.6 MeV at full-width half-maximum (FWHM), or 2.8 MeV standard deviation (assuming a Gaussian fit). This distribution is insignificant to the total dose, given that there is no significant difference between 55 MeV protons and 60 MeV protons in terms of dose in silicon.

The entire testing system was transported and set up at the proton beam facility so that the detectors could be tested between radiation doses in a vacuum- and temperature-controlled environment. The system was set up so that the detectors were in the beam path inside the dewar, with the radiation passing through a thin metal cover, which kept the dewar completely dark. Figure 47 shows the setup of the dewar, control electronics, and cooling system used at the MGH facilities. Lead bricks were used to shield the control electronics behind the dewar, as well as the electronics inside the dewar with strategic placement (see Figure 49).

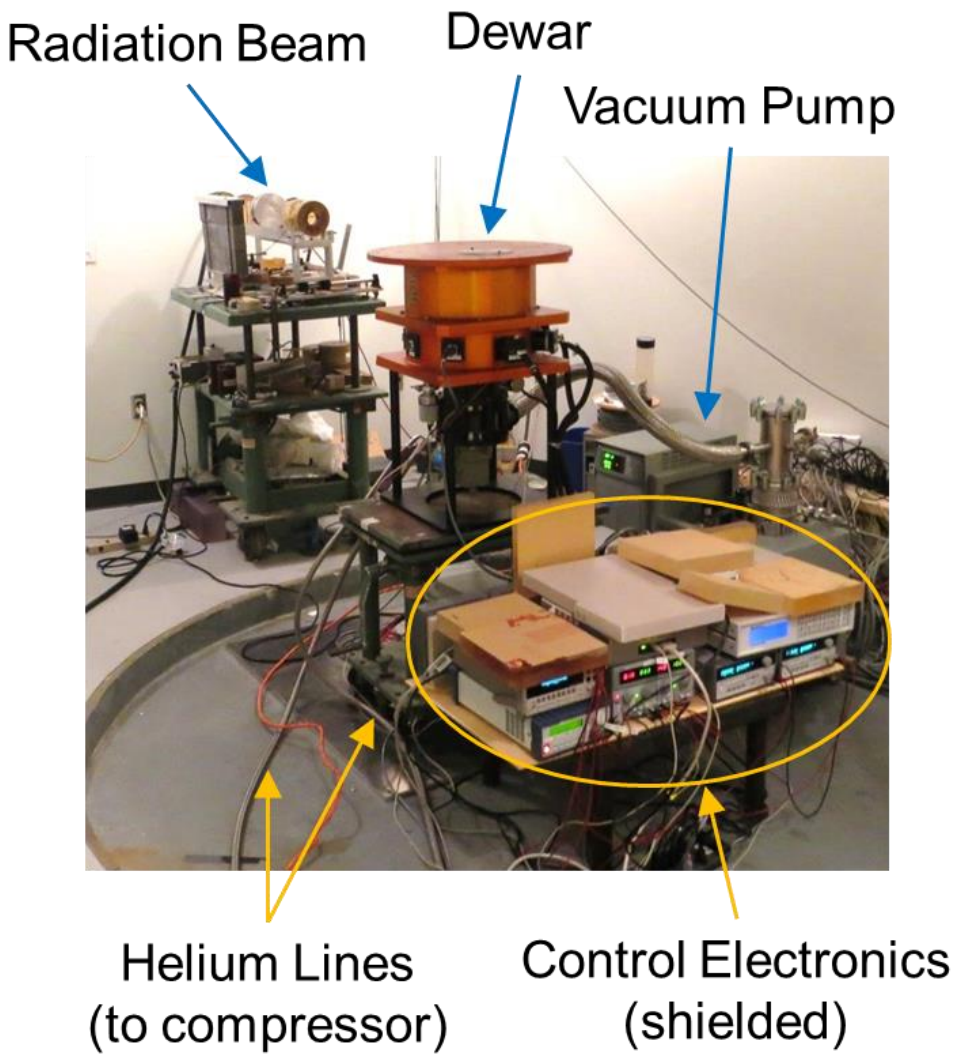

Figure 47 – The experimental setup at MGH is shown. The vacuum pump and dewar were shipped from the CfD, the compressor used for the cryogenic pump was borrowed from the MGH facility, and the control electronics were shielded behind a wall of lead bricks.

Figure 48 shows the remote control room for the radiation experiments. Due to the danger of radiation exposure, the detectors had to be controlled from a remote location. The computer shown, called the remote computer, was logged into the computer near the dewar, called the local computer. The local computer controlled the detector electronics as well as the periphery electronics (power supplies, temperature control, etc.) and had to be physically connected to

them. The remote computer was connected to the local computer via a 500 ft. ethernet cable and was used to monitor the detector output data in real time during the experiment and note the precise start and end times of the radiation beam for each radiation dose.

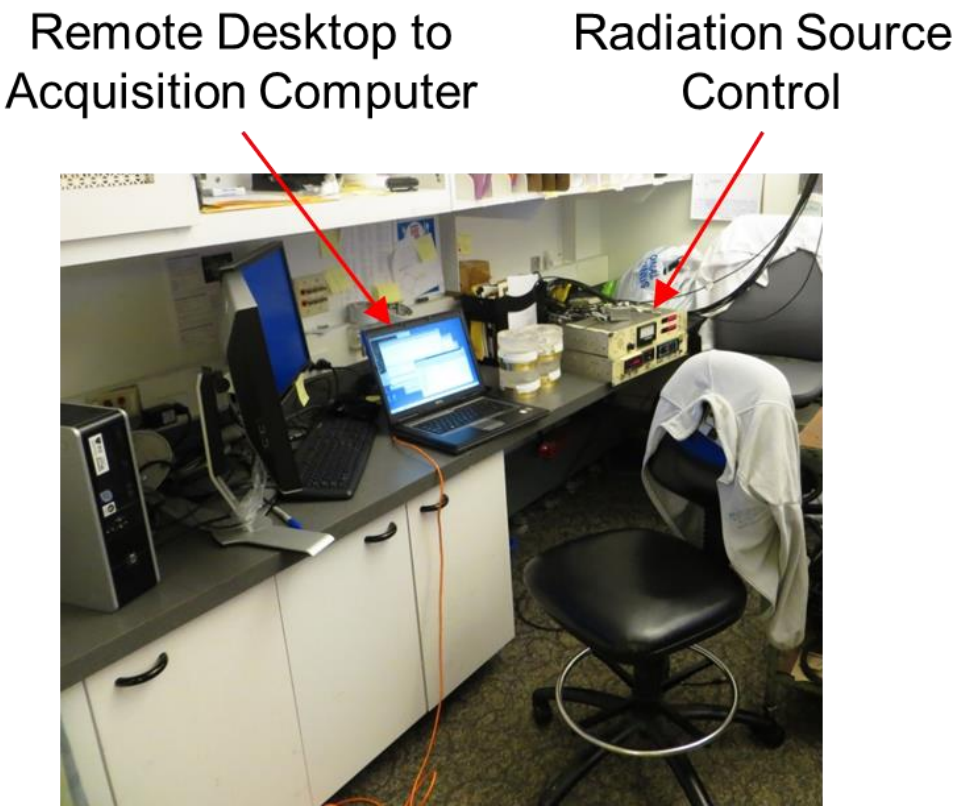

Figure 48 – The experimental setup at MGH is shown. The experiments were controlled remotely from a radiation-safe location.

In order to measure the spatial distribution of the radiation dose, radiochromic film was placed over the dewar port cover during the experiments. Figure 49 shows the sheet as well as the lead bricks used to shield the electronics inside the dewar.

Figure 49 – This figures shows a close-up of the radiation testing setup around the dewar port. The top image shows a side view, and the bottom image shows the view as seen from the end of the radiation beam hardware (see Figure 47).

The radiochromic film was not left on for the duration of the experiment. It was left on during the very first radiation dose of 0.1 krad(Si) and then removed. The radiation profile for the first dose is a valid measurement of the total radiation profile. Figure 50 shows a greyscale image of the radiochromic film and the pixel values along the cut line shown in white. The spatial distribution of the radiation beam is within 1.7% across each device and 1.4% between devices.

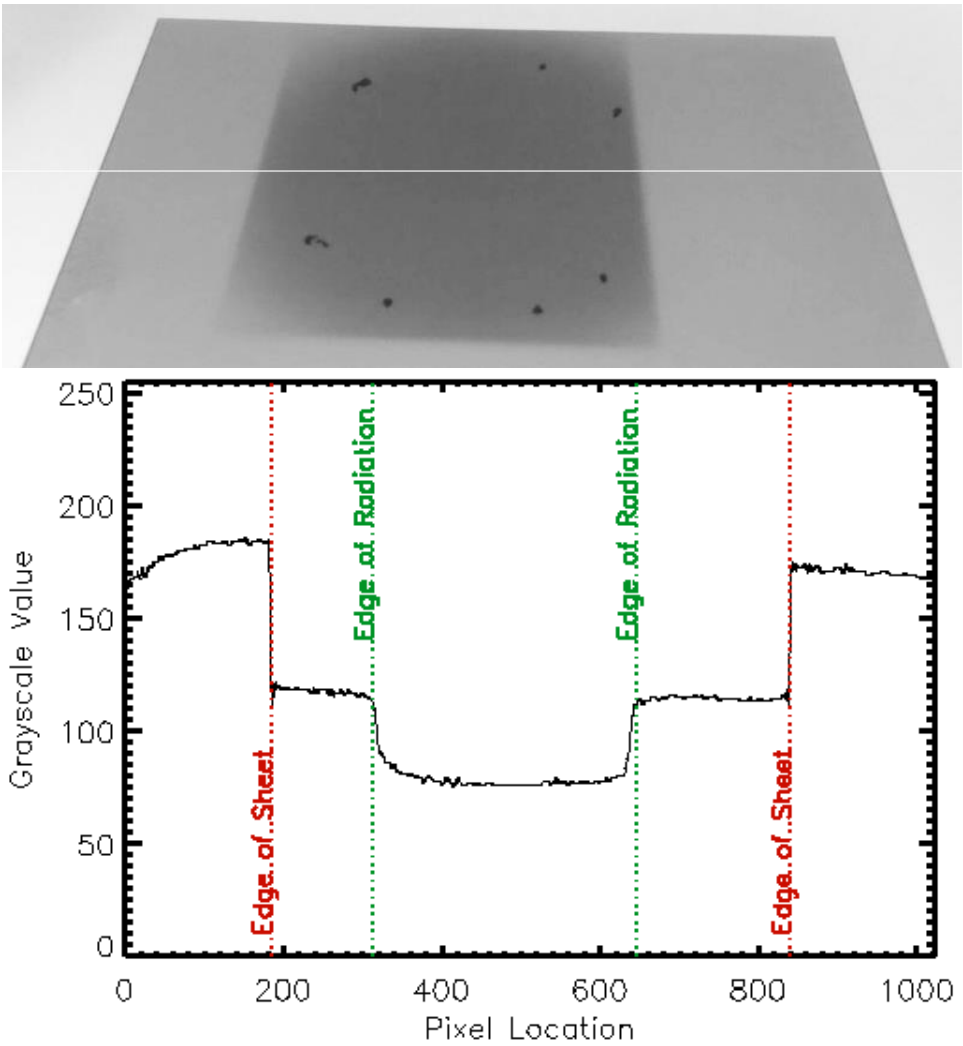

Figure 50 – The exposed radiochromic film is shown after a 0.1 krad(Si) dose (top) with a cut line of the value of each pixel marked by the white line (bottom) .The dark dots in the upper image note the placement of the screws on the dewar port cover for reference. The sharp edges of the radiation exposure are due to the shielding effect of the lead bricks in Figure 49.

Figure 51 shows the in-situ results for the radiation testing of the devices. Data was taken for 20 minutes (wall time) per detector between radiation doses and overnight after the final radiation dose. The detectors were kept cold (~220 K) for the duration of the experiment. After the final dose, the electronics suffered some failures, which were likely single-event upsets from the secondary neutron scattering. The electronics were reset and reprogrammed before starting the final data set, leaving a small gap in the data.

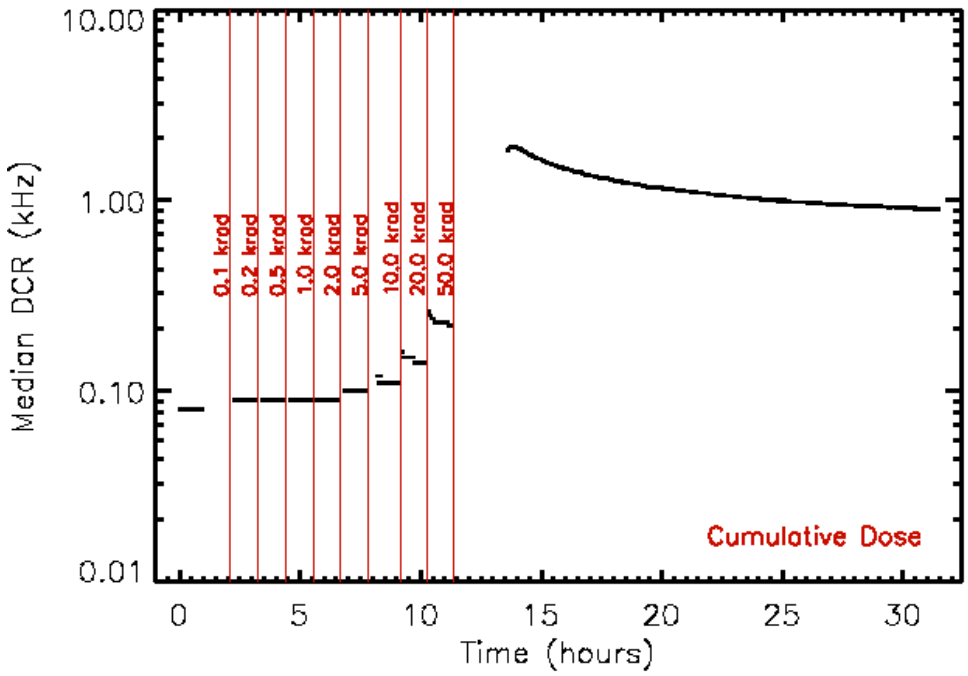

Figure 51 – Median DCR vs time over incremental radiation doses is shown for a GM-APD device. The radiation dose is marked at the time when the radiation beam stopped for that dose.

In Figure 51**,** some exponential decay occurs in the DCR following each radiation dose, and the $^{1}/_{e}$ lifetime and steady-state DCR appear to increase with increasing dose. After the final radiation dose, the data for each detector showed an exponential decay. The measured 1/e lifetime for the final data set was 5.28 hours. The temporary rise in DCR immediately following a radiation dose is likely due to secondary radiation from the irradiated dewar and system components. At lower radiation doses, it is likely that there was a small increase, but that the increase was not large enough to detect. The increase in steady-state DCR is likely due to lattice damage caused by the proton radiation. Atoms that are dislodged from the lattice structure create intermediate energy states and become carrier generation sites.

There were no detector failures during in-situ radiation testing. However, despite being behind a shield of lead bricks, the electronics suffered a single-event upset that resulted in the failure of one readout channel on all the detectors. After resetting the electronics, they operated normally. The four FPGAs in the control electronics were re-programmed, as a precaution, before testing continued.

Following the cold in-situ radiation testing, the detectors were warmed to 300 K and brought back to the CfD. Cold testing resumed when the DCR reached 99% of the settling point, calculated with an exponential decay function. Figure 52 shows warm data, taken with the same settings as when the detectors were cold.

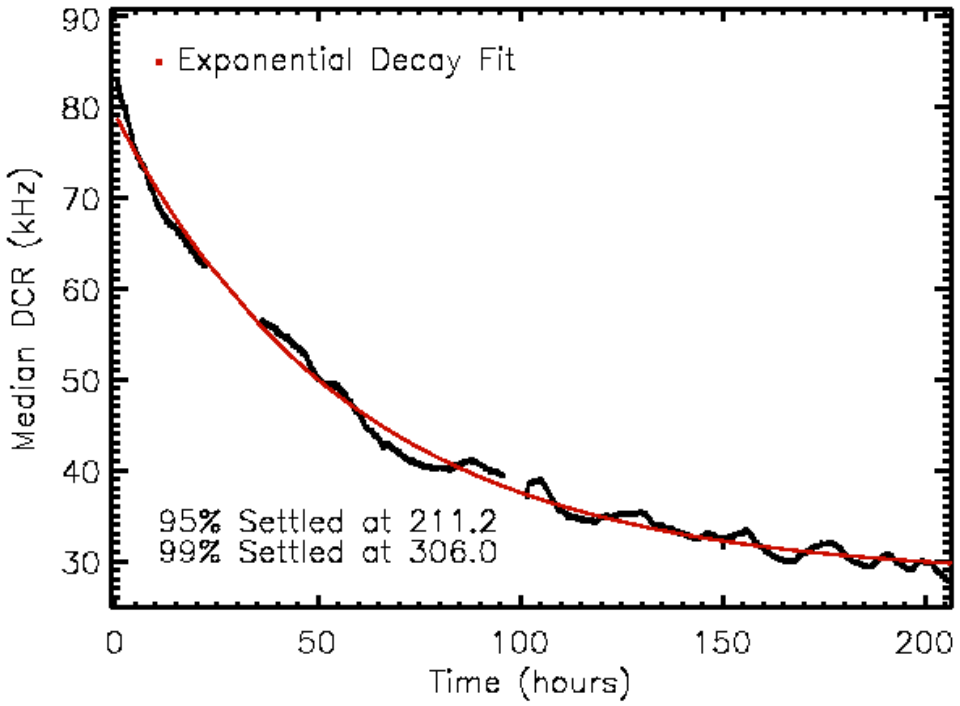

Figure 52 – This figure shows median DCR vs time at room temperature for one of the GM-APD devices.

The $^{1}/_{e}$ lifetime of the DCR decay at room temperature in Figure 52 is ~66.5 hours. Unlike the decay in in-situ radiation data, the decay at room temperature was likely due to annealing, or self-healing, of the lattice using energy from increased temperature. Because the crystal lattice structure of a material represents the least amount of potential energy stored in the material, atoms that are dislocated from the lattice will move back into position to rest in the lowest potential energy state. Over time, the DCR will approach a new steady-state value, which will be higher than pre-radiation levels because some lattice damage will not anneal at room temperature. Once the DCR settled, the detectors were cooled. Over the course of three weeks, DCR data was taken at multiple temperatures in four separate experiments. Figure 53 shows the results from these four runs overlaid on the same plot. The mean percent standard deviation

between the DCR data points at each temperature was ~1%. This consistency verified that the DCR had reached its new post-radiation steady-state level.

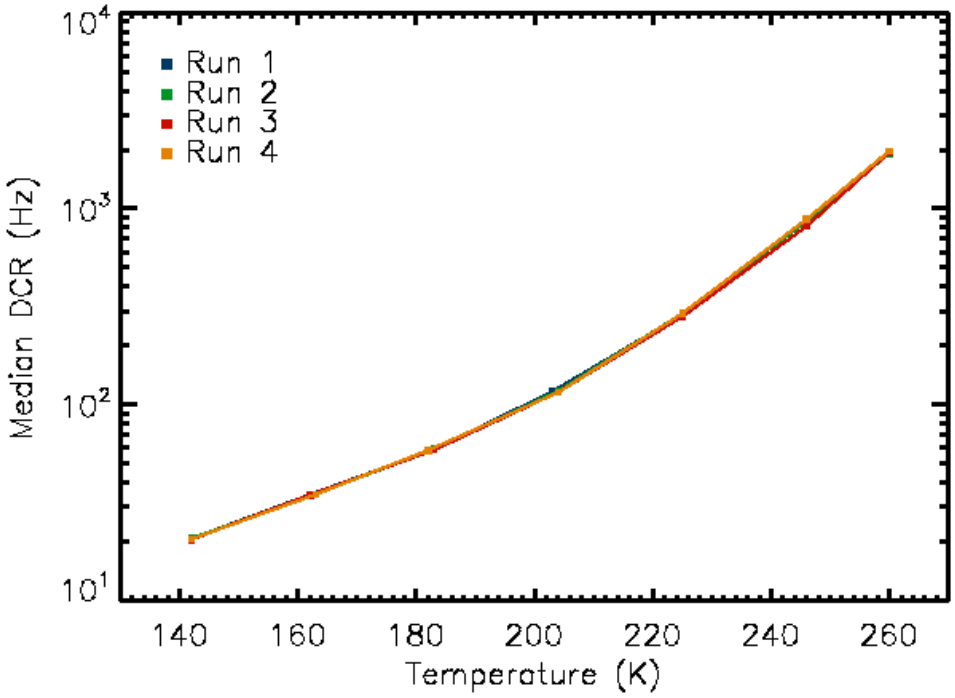

Figure 53 – This figure shows median DCR vs temperature for a GM-APD detector. Each run is a separate experiment in a three week period.

3.5 Post-Radiation Breakdown Voltage

The breakdown voltage of all the devices increased by ~1.5 V after irradiation. This increase corresponds to a ~4.5% increase in absorber resistivity due to radiation damage, calculated based on the dimensions and doping profiles of the pixels as presented in the introduction to section 3. As atoms in the silicon lattice become displaced by radiation damage, the resistivity of the bulk material increases (Li, 2002). The applied voltage in a GM-APD is across the entire diode structure, so the voltage is divided between the multiplication and absorption regions (see Figure 38). The absorber region sustains more damage because of its larger volume. If the resistance of the absorber region significantly increases after irradiation, less of the applied voltage falls across the multiplication region, and more applied voltage is necessary to achieve the critical electric field necessary for breakdown.

To match the testing conditions of the pre-radiation characterization, post-radiation testing at 2.0 V was done. The post-radiation DCR of the devices were measured at 2.0 V overbias and 0.5 V overbias and the results are shown in Figure 54. At 2.0 V overbias, the DCR was very high (nearly 200x higher at 150 K) and dominated by afterpulsing, especially at temperatures below 200 K.

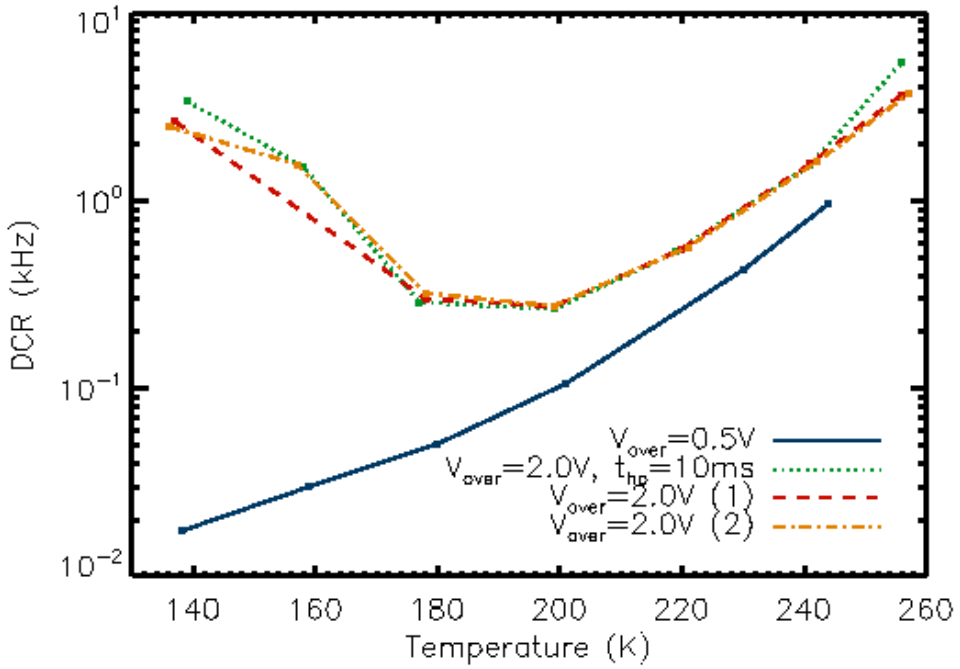

Figure 54 – Post-radiation DCR results for S47 are shown with an overbias of 2.0 V and 0.5 V. The hold-off time for each run was 5 ms unless otherwise noted. The DCR for the higher overbias results at temperatures below 200 K are significantly affected by afterpulsing. Afterpulse testing results for the 0.5 V overbias confirm that afterpulsing does not affect the DCR at a hold-off time of 5 ms.

The 2.0 V overbias results are significantly affected by afterpulsing at temperatures below 200 K, even at hold-off times of 10 ms. The increase in DCR at colder temperatures implies that there are many traps with long lifetimes. There was no measured improvement between hold-off times of 5 ms and hold-off times of 10 ms. This result implies that the trap lifetimes of the device are much longer than 10 ms.

In order to measure the afterpulse probability, data must be taken at a hold-off time long enough to assume there is no afterpulsing. In the post-radiation case at 2.0 V overbias, the longest hold-

off time required may be on the order of 1 s. Measuring the afterpulsing with hold-off times longer than 1 s is not practical using the standard method (described in section 3.8); a single data set at one temperature would take over 11 days.

Given the length of the trap lifetimes expected at 2.0 V overbias, it was reasonable to assume that the effects of afterpulsing could be measured in a persistent charge experiment. In this experiment, the detectors are illuminated so that the avalanche probability is 1, and then the light source is shut off while data collection continues. Figure 55 shows the persistent charge results for an overbias of 2.0 V.

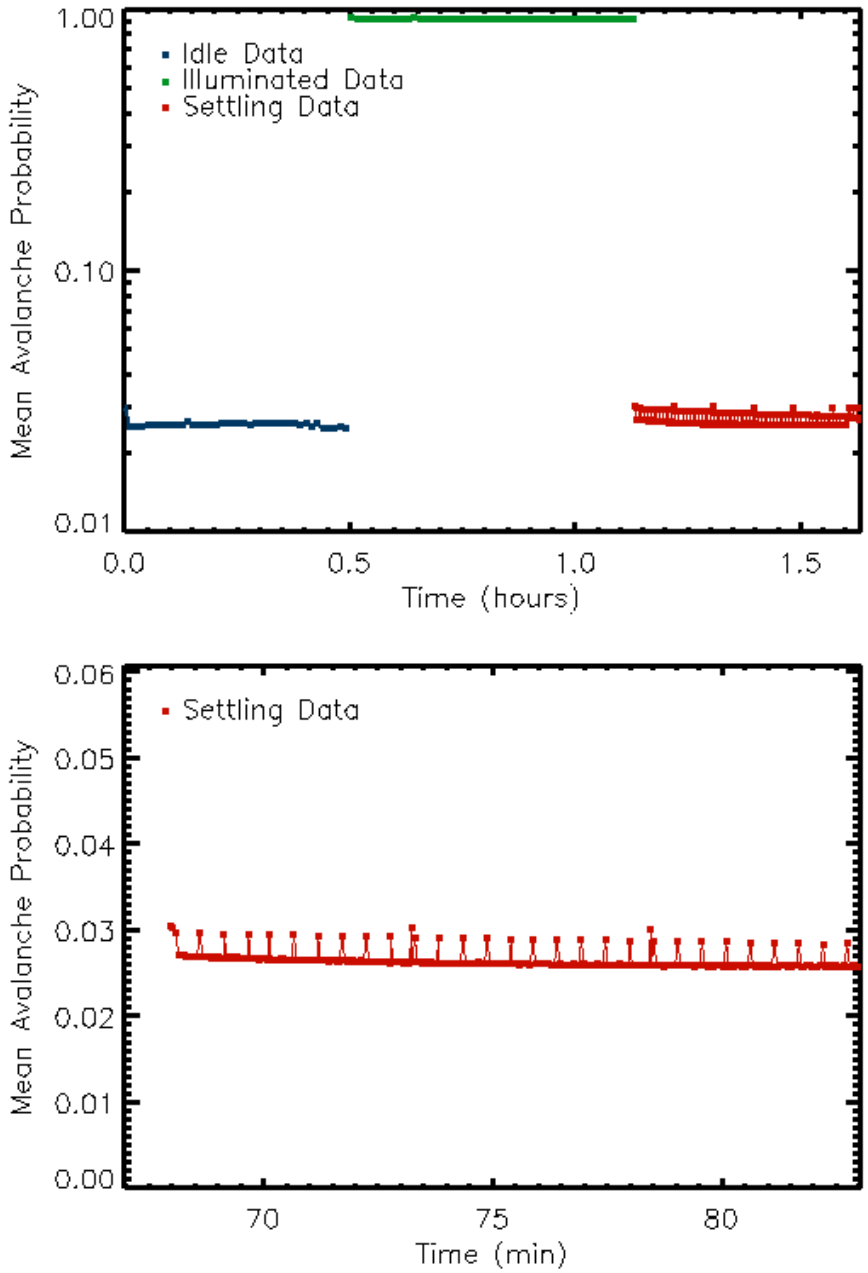

Figure 55 – Post-radiation persistent charge results for S47 are shown with an overbias of 2.0 V. The $^{1}/_{e}$ lifetime of the post-illuminated decay in the settling data is ~6 minutes.

The post-radiation persistence at 2.0 V overbias is significant with a $^{1}/_{e}$ lifetime of 5.8 minutes at 116 K. This level of persistence makes operation of the device at 2.0 V overbias impractical. Therefore, after 50 krad(Si) of radiation, the device was operated at a lower overbias setting of 0.5 V. Because dark exposures could not be taken with zero afterpulsing probability, post-

radiation results at 2.0 V overbias are not possible for the DCR, afterpulsing, IPS, PDE, or crosstalk experiments. Instead, the post-radiation results below are presented for an overbias of 0.5 V. While the comparison is not ideal, an overbias of 2.0 V would not be viable for testing or for general operation. A 0.5 V overbias in the post-radiation data represents the realistic operating conditions for a device with a 50 krad(Si) radiation dose.

3.6 READ NOISE

The read noise for the GM-APD array devices is expected to be zero due to the way in which signal is stored and read out. The pixel signal is stored in an in-pixel counter that increments when an avalanche occurs (see Figure 25). A pixel can only be in one of two states: avalanched or not avalanched. The state of counter is read out by a digital circuit and no estimation is required. Therefore, there is no estimation noise, and there should be no noise in the transfer of the signal.

To verify that there was no read noise, the output of each pixel was artificially set to either a 1 or a 0 and read out. In pre-radiation testing, there were zero instances of a false output out of 10,000 trials for each of the 256 x 256 pixels in an array. This demonstrates the read noise is less than one count per 6.55×10^8 reads before radiation. No post-radiation testing was done, but the nature of the threshold circuit and robustness of the pre-radiation testing suggests that there would be no affect due to radiation damage.

3.7 DCR

DCR is the rate of counts generated in the absence of light. In this report, DCR is corrected for dead time and the experiments to measure it are designed to minimize counts from afterpulsing. As such, the DCR is expressed as electrons/s/pixel (Hz). The necessary exposure time was set by calculating the number of gates required for the desired SNR. The SNR of a measurement is proportional to the square root of the total number of gates. With this in mind, the number of gates was set to 1 million, equivalent to an exposure time of 10 s, and the experiment was done in a light-tight dewar. The minimum SNR for the entire DCR experiment was 14 using these settings for the DCR levels measured.

The probability of an avalanche in a single gate (P) is the number of avalanches divided by the number of gates. Since the rate at which electrons enter the multiplication region is Poissonian, the Poisson probability model can be used to solve for the mean electron flux, given P and t_{gate}. Eq. 93 shows the relationship between P, t_{gate}, and the DCR (λ_e). Note that Eq. 93 cannot be used if the afterpulsing probability is not zero. Afterpulse counts are not Poissonian, and so they cannot be included in λ_e. The statistics regarding the characterization of afterpulsing are discussed in the next section.

$$P = 1 - e^{-\lambda_e t_{gate}}$$

$$\lambda_e = \frac{-\ln(1-P)}{t_{gate}} \qquad \text{Eq. 93}$$

The expected trend for a silicon device is the doubling of dark current (which is proportional to DCR) roughly every 8 K (Fry, 1975), but these devices show doubling roughly every 17 K in pre-radiation testing. This shallow slope may be caused by biasing effects, or it may be cause by dark carrier generation sites with low activation energies (as might be caused by backside defects introduced in the thinning process – see section 6.5). Although the bias is adjusted with temperature to account for the change in breakdown voltage (see section 3.1), there is no way to know if the adjustment compensates perfectly for the change. Additionally, the breakdown voltage for each pixel in a detector array is not uniform. This is likely due to processing non-uniformity across the wafer, specifically implant depths and doses, as well as local defects in the material. This non-uniformity in breakdown voltage, combined with changes in the shape of the effective multiplication region due at different biases, leads to uncertainty in the overbias for any given pixel.

Figure 56 shows the pre- and post-radiation DCR results for a GM-APD device operated at the same applied bias setting. The pre-radiation testing is at an overbias of 2.0 V and the post-radiation testing is at an overbias of 0.5 V (see section 3.5).

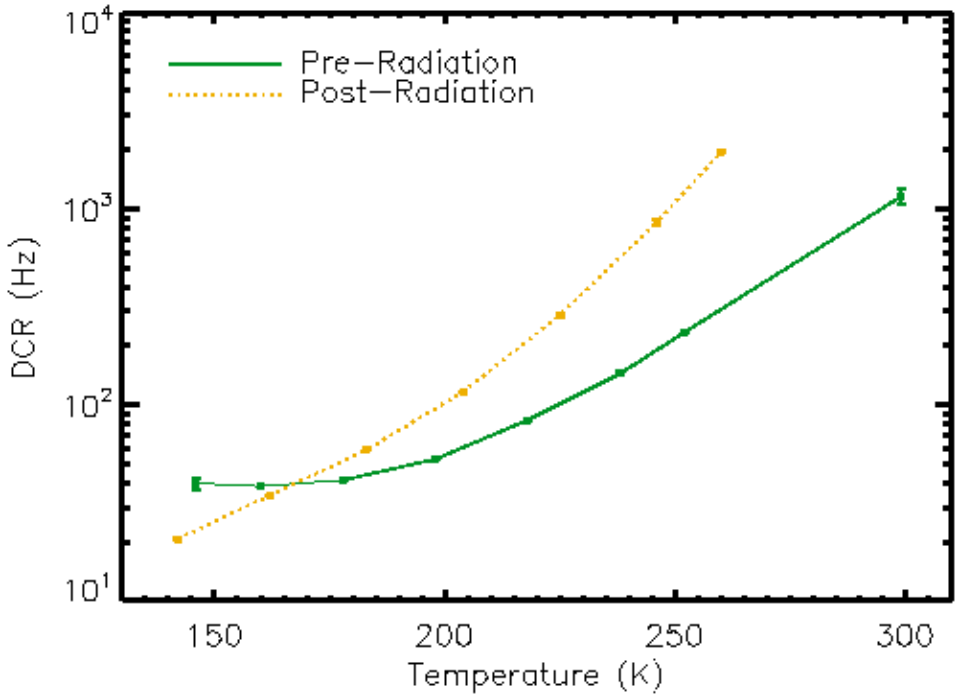

Figure 56 – This plot shows median DCR vs temperature for a GM-APD device, pre- and post-irradiation. Longer arm periods were used to ensure that afterpulsing was insignificant at lower temperatures.

The apparent decrease in DCR at lower temperatures after irradiation is due to the lower overbias for the post-radiation data. Because the overbias is lower, the avalanche initiation probability is also lower, which reduces the DCR as well as the efficiency. The post-radiation DCR curve is shifted down by a factor proportional to the decrease in avalanche initiation probability at all temperatures. It is important to note that, even at a reduced overbias, the post-radiation DCR is still greater than the pre-radiation DCR at temperatures above 165 K.

3.8 AFTERPULSING

Afterpulsing is an increase in count rate following an avalanche in the same pixel. Afterpulsing is often caused by traps, which are energy levels that exist only around material defects in the detector material. Instead of being collected, charge becomes "trapped" in these energy states and released a random amount of time later. Afterpulsing increases the measured DCR and decreases the SNR. Afterpulsing probability (p_{aft}) can be measured experimentally by observing

individual gates (Itzler, et al., 2012; Jensen, et al., 2006), but the operation of these devices is not conducive to that method.

Here, P_{aft} is calculated by comparing measured DCR at various hold-off times. At hold-off times that are much longer than the trap lifetime of traps in the detector, afterpulsing probability is near zero. By taking data at a very long hold-off time, the measured DCR (λ_e) at that hold-off time can be assumed to be unaffected by afterpulsing. The afterpulsing probability at shorter hold-off time settings was determined using a theoretical model for avalanche probability and λ_e. In this case, afterpulsing probability (p_{aft}) is the probability of an afterpulse carrier initiating an avalanche during a gate, given an avalanche in the previous gate. This model assumes that the state of a gate is only dependent on the state of the gate immediately before it. This assumption allows for the representation of avalanche probability (P) for each gate in an infinite series (n=0,1,2,…). With the use of the Maclaurin series identity, that infinite series simplifies to the closed-form solution in Eq. 94. It is important to note that Eq. 94 assumes that there are no incident photons, though it would be possible to calculate p_{aft} if the photon signal were constant by substituting ($PDE \cdot \lambda_p + \lambda_e$) for λ_e.

$$P = \frac{1 - e^{-\lambda_e t_{gate}}}{1 - p_{aft} e^{-\lambda_e t_{gate}}}$$

$$p_{aft} = \frac{1 - e^{\lambda_e t_{gate}}(1 - P)}{P} \qquad \text{Eq. 94}$$

Recall Figure 27, which shows that the change in avalanche probability is not linearly proportional to afterpulsing probability. The figure shows avalanche probability (P) vs. gate fluence for various values of p_{aft}. Gate fluence is the number of electrons per pixel per gate that could initiate an avalanche. The avalanche probability values are calculated using Eq. 94. Note that the avalanche probability converges at small and large gate fluence values. When gate fluence is low, any additional avalanches from afterpulsing are unlikely because the avalanche probability is very low to begin with. At high gate fluence, afterpulse events are often coincident with avalanches that would have occurred because of the gate flux.

In the GM-APD array devices, the traps that cause afterpulsing are most likely located in the multiplication region. It is the only region that is subjected to high current density during an avalanche. There is a smaller probability that traps in the absorption region will also be populated, but the trapping of carriers in the absorption region occurs at random times, as opposed to during an avalanche, and the carriers are less likely to reach the multiplication region and initiate an avalanche than carriers released from traps in the multiplication region. While the number of traps in the material remains the same regardless of the overbias, the avalanche initiation probability at various positions inside the multiplication region increases with increased electric field. Traps positioned at the boundary of the avalanche region, near the implant that separates the multiplication and absorption regions (see Figure 38), will release carriers with a higher avalanche initiation probability than traps near the middle of the multiplication region. Although traps will likely be populated during an avalanche regardless of location, the released carriers will have limited impact if the location decreases the avalanche initiation probability. However, with increased overbias, the electric field in the multiplication region increases so that carriers released further towards the middle of the region have a significant avalanche initiation probability. This leads to a large increase in afterpulsing probability with a higher applied overbias.

Since afterpulsing was too high to measure at 2.0 V overbias post-radiation, the afterpulsing probability results after irradiation are shown for an overbias of 0.5 V. Afterpulsing probability decreases with decreasing avalanche initiation probability and therefore with decreasing overbias. Figure 57 shows P_{aft} vs t_{ho}, pre- and post-irradiation.

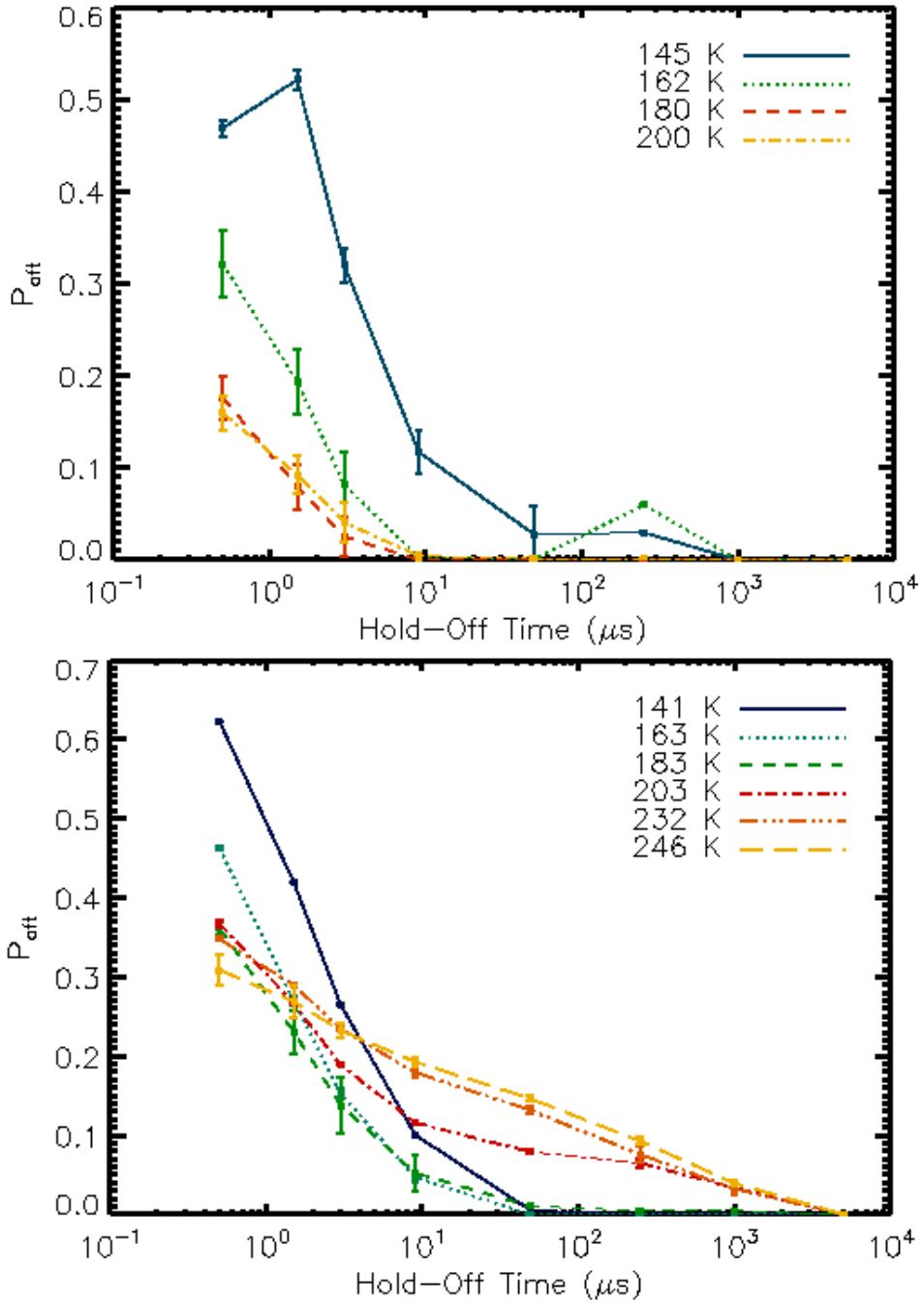

Figure 57 – Median afterpulsing probability vs hold-off time is shown for various temperatures. The plot on the left shows pre-radiation results, and the plot on the right shows post-radiation results.

It is important to note that measurements of the afterpulsing probability in excess of ~0.1 are not physically possible with the model used (see section 2.6.3). This means that measurements in

excess of $p_{aft} = 0.1$ imply that the effects of traps extend beyond a single arm period. In these cases, the afterpulsing cannot be corrected in the estimate of the signal, and the detector should not be operated at the hold-off times that correspond to $p_{aft} > 0.1$. For $p_{aft} \leq 0.1$, the measurements are valid for a single arm period. An increase in the measured afterpulsing probability indicates an increase in the actual afterpulsing probability, but the exact afterpulsing probability is indiscernible with the current model. When the state of a gate depends on avalanches more than one gate prior, the Markov chain model is no longer applicable.

The decrease in afterpulse probability at the shortest hold-off time in the pre-radiation results is not likely due to any physical phenomenon, but is more likely a result of an experimental inconsistency. The afterpulsing probability is very sensitive to small changes in avalanche probability when the measured probability is so high.

P_{aft} increased at all temperatures after irradiation, even at the lower overbias. After irradiation, a new trend emerged in which p_{aft} increased at temperatures above 183 K with a nearly exponential slope. The minimum t_{ho} required to avoid afterpulsing was 1 ms before irradiation. After irradiation the minimum t_{ho} was 1 ms below 183 K and 5 ms for warmer temperatures. For t_{ho} up to 10 ms, there was no discernible decrease in post-radiation afterpulsing probability at 0.5 V overbias.

The increase in afterpulsing probability at warmer temperatures may be due to self-emission re-triggering: photons emitted during the avalanche are reabsorbed in the same pixel and trigger an avalanche during the next gate. It stands to reason that crosstalk may also be higher at shorter higher temperature, though an analysis of crosstalk at multiple temperatures was not done. Future iterations of device design would benefit from such a study.

3.9 IPS

IPS is the PDE as a function of 2D location inside a pixel. Ideally, the IPS function would be 1 across the entire pixel, but for the GM-APDs tested, the active area is concentrated at the center of each pixel and can be modeled as a 2D Gaussian function. IPS was measured by projecting a small pinhole image, with a FWHM of 4.1 μm, onto the detectors. The focal spot size was

measured outside of the dewar using a knife-edge test, and the focal spot size inside the dewar was simulated using a program called OpticalRayTracer.[5]

A homing routine was run before each scan to center the laser on the pixel of interest. The spot was moved through a 2x2 pixel grid of 2.5 μm steps, with exposures at each grid location. It was important to keep the exposures short to avoid spot drift due to the heavy load on the linear stage motors (see Figure 58). The measured IPS function was de-convolved with the expected laser spot size to calculate the actual IPS function of the pixel. Reported values for fill factor and FWHM are calculated from IPS functions corrected in this manner.

The pre-radiation IPS data was taken at an overbias of 2.0 V, and the post-radiation IPS data was taken at an overbias of 0.5 V. No change in IPS due to radiation damage was expected. The shape of the implants and the electric field inside the pixels would not change with the uniform damage introduced by radiation. IPS is normalized to the peak sensitivity, and reflects only the efficiency as a function of location.

[5] http://arachnoid.com/OpticalRayTracer/

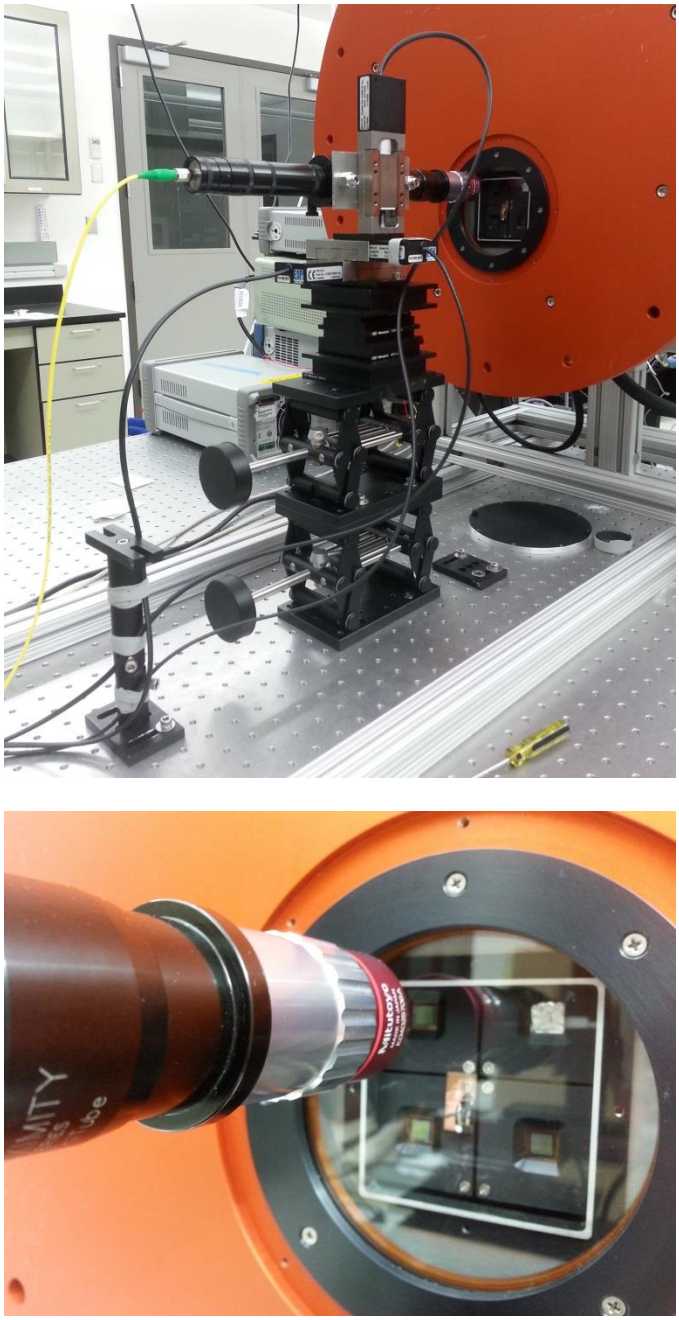

Figure 58 – The testing setup at the CfD for measuring IPS is shown. During experiments, a light-tight box is placed over the entire assembly. On the top, the full setup is shown: three linear stages, mounted orthogonally, are secured to a support structure that allows the laser spot to scan across a detector inside the dewar. On the bottom, a close-up of the objective lens is shown scanning across a detector behind the dewar window.

Figure 59 shows the measured IPS of two pixels on one detector. Note that the scan area is 2x2 pixels, but each plot only shows the signal from one pixel (outlined).

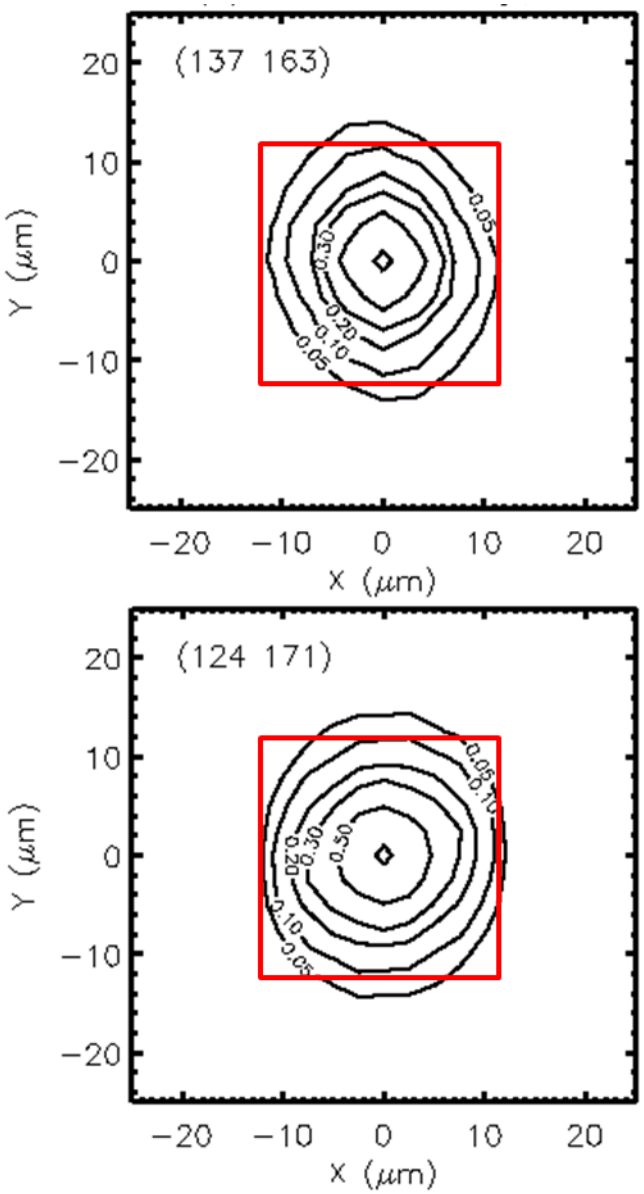

Figure 59 – Sample IPS results from pixels on a GM-APD detector are shown. The centers of the pixel are inside the central contours. The upper plot shows pre-radiation results, and the lower plot shows post-radiation results. There is no statistically significant difference between the FWHM values. The boxed outlines show the pixel boundaries.

The raw measured data at each scan location was dark-subtracted and normalized to the peak signal. At this point in the data reduction, the measured IPS is spatially convolved with the laser

spot. In order to correct for the convolution, the laser spot profile (assumed to be Gaussian with a FWHM of 3.8 μm) and the measured IPS function were converted to Fourier space. In Fourier space, a spatial convolution becomes a frequency-domain multiplication. A simple division of the Fourier transforms of the measured data by the laser spot frequency data results in the Fourier transform of the de-convolved IPS data. A reverse Fourier transform is applied to the de-convolved frequency data to yield the de-convolved spatial IPS function. The equation used for the de-convolution is shown in Eq. 96, $IPS_{de\text{-}convolved}(x,y)$ is the de-convolved IPS function, $IPS_{convolved}(x,y)$ is the as-measured IPS function (normalized and centered), and Laser(x,y) is the laser spot profile. $\Im\{\}$ is the Fourier transform operator, and $\Im^{-1}\{\}$ is the inverse Fourier transform operator.

$$IPS_{de-convolved}(x,y) = \Im^{-1}\left\{\frac{\Im\{IPS_{convolved}(x,y)\}}{\Im\{Laser(x,y)\}}\right\} \qquad \text{Eq. 95}$$

The pre-radiation FWHM was 10.2 μm $\pm$ 5 μm, and the post-radiation FWHM was 11.7 μm $\pm$ 1.3 μm. The pre- and post-radiation FWHM values are not statistically different. The size of the laser spot is highly dependent on the focus, so it is likely that the variance in FWHM is a function of the variance of the spot size when moving the spot projector and re-focusing. The mean pre-radiation fill-factor was 18.7% and the mean post-radiation fill-factor was 24.2%.

3.10 PDE

PDE is the inferred photo-generation rate divided by the incident photon rate. The GM-APD data presented here is for a LFF architecture, which limits PDE. This performance is not representative of what is expected for detectors optimized for imaging applications, which have a peak PDE of ~15% (Figer & Kolb, 2014). Additionally, the PDE was measured at a moderate overbias (2 V) to limit DCR – a higher applied overbias would result in higher PDE. The PDE of a device is very dependent on overbias because overbias increases the avalanche initiation probability. Figure 60 shows PDE at 520 nm as a function of overbias. DCR increases with

overbias due to increased avalanche initiation probability as well. In order to avoid high DCR, PDE was measured with a moderate overbias of 2 V.

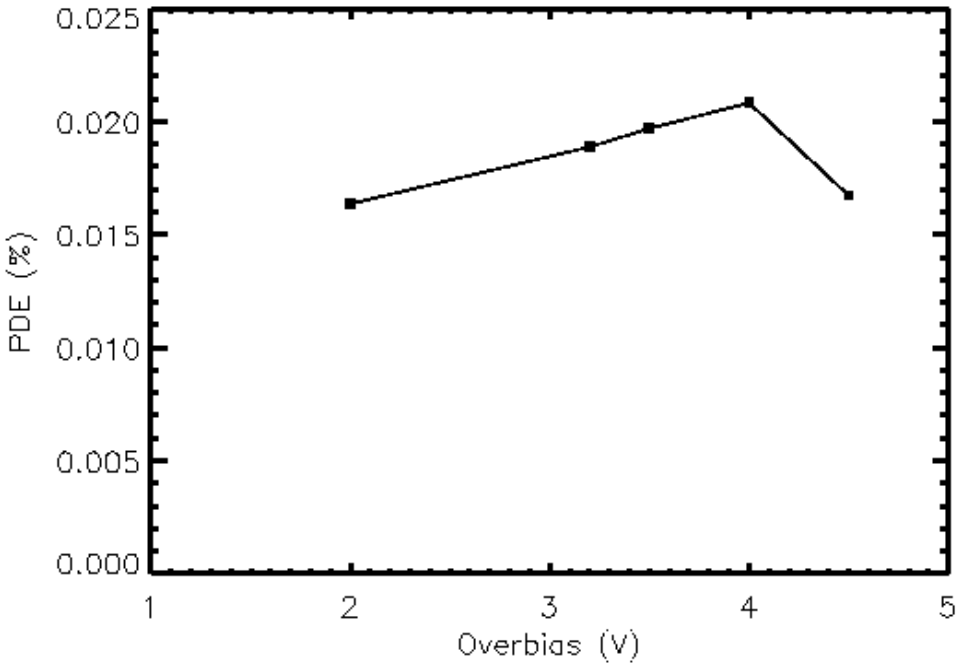

Figure 60 – PDE (%) vs. overbias for a single pixel is shown for a wavelength of 520 nm. PDE increases with overbias until a certain point at which the multiplication region begins to be pinched off by the high electric field in the absorber region.

Eq. 96 gives the expression for the number of photo-generated electrons entering the multiplication region during a gate that can initiate an avalanche. The expression is a function of avalanche probability, DCR, and t_{gate}. The equation assumes that afterpulsing is negligible.

$$P = 1 - e^{-t_{gate}(\lambda_p + \lambda_e)}$$

$$\lambda_p = \frac{-\ln(1-P)}{t_{gate}} - \lambda_e \qquad \text{Eq. 96}$$

A calibrated photodiode (UV-enhanced Si) was used to determine the number of photons incident on the detectors. This was done by measuring the flux inside of an integrating sphere illuminated by a monochromator. In order to account for changes in flux at the detectors vs. the

calibrated photodiode, an experiment was run with two diodes simultaneously: one in the integrating sphere position and one inside the dewar at the same position as the detectors (see Figure 61). The ratio of photon flux between the two photodiodes was calculated and used to estimate the flux on the detectors based on the diode readings inside the integrating sphere.

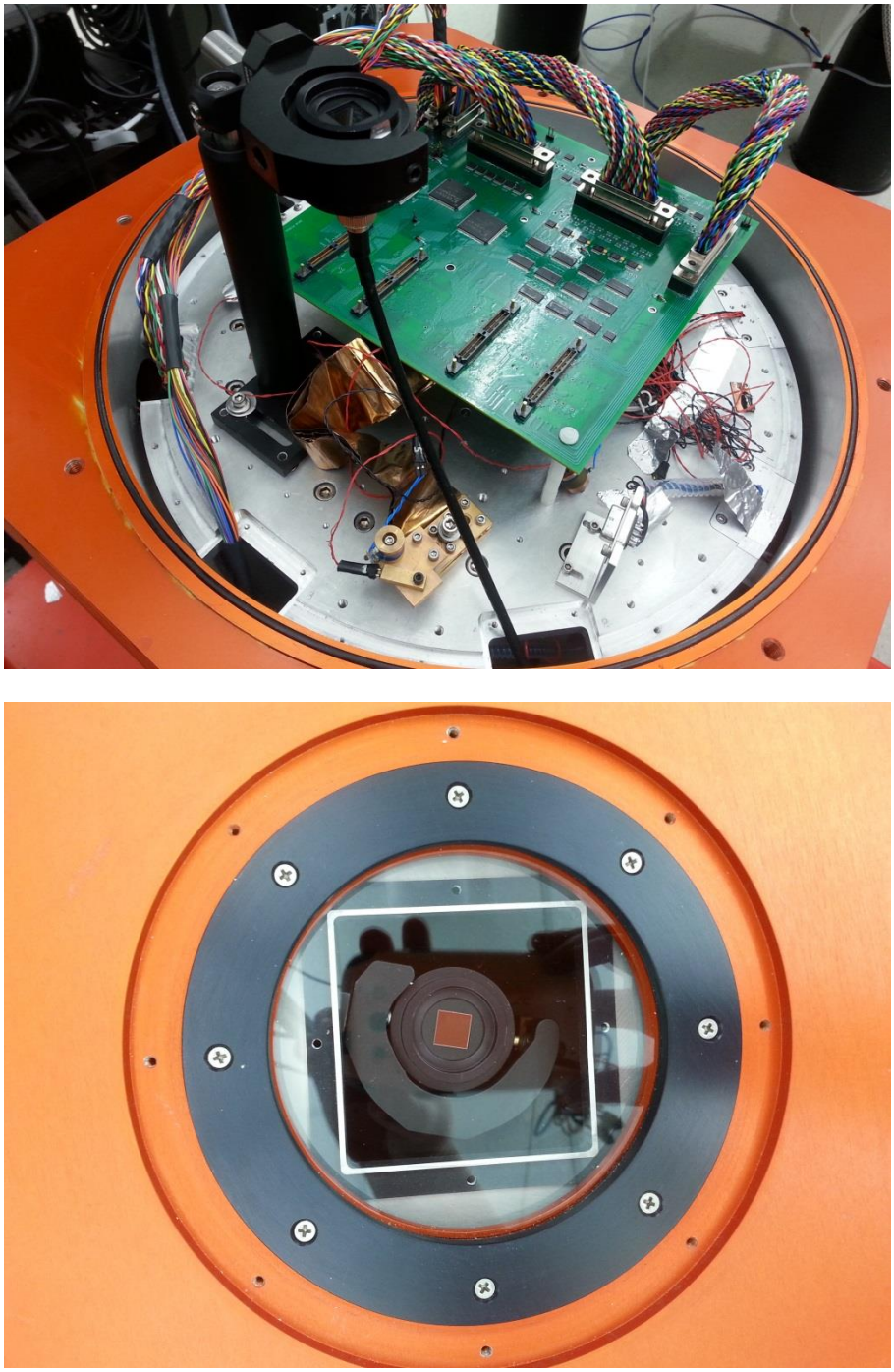

Figure 61 – The detector replacement calibration setup is shown. On the top, the calibrated diode can be seen mounted inside the dewar at the same height and location as the detectors when mounted. On the bottom, the diode as seen from the outside of the closed dewar is shown.

Eq. 97 shows the expression for PDE in terms of λ_p (photoelectrons/gate/pixel), S_D (diode signal in photons/s/mm^2), K (the ratio of photons incident on the detectors to photons incident on the calibrated diode in the integrating sphere), A_{pix} (the area of a pixel in mm^2), and t_{gate}.

$$PDE = \frac{\frac{\lambda_p}{t_{gate} \cdot A_{pix}}}{S_D \cdot K} \qquad \text{Eq. 97}$$

For each new wavelength setting, the mechanical parts of the monochromator move, which causes the background light level to change at each wavelength. Therefore, dark exposures were taken at each wavelength setting to account for the changes in background light level. Λ_e was calculated from each dark exposure and subtracted from the total number of electrons detected to calculate λ_p as in Eq. 96. Eq. 97 was then used to calculate the PDE for each pixel.

The maximum SNR occurs at P=0.797, which can be arranged by setting the gate time, assuming that the photon flux is constant for all gates. In order to ensure statistically relevant results, the gate time was optimized for four separate wavelength intervals based on the signal level expected at the detectors. Choosing a minimum SNR of 3, minimum and maximum gate times were calculated using extrapolated values for photon flux, PDE, and DCR from a preliminary experiment. Figure 62 shows the ideal gate time vs. wavelength for the pre-radiation PDE experiment, which was repeated for the post-radiation experiment, but is not shown. The minimum and maximum gate times at each wavelength are the error bar bounds. The green line shows the gate times that were chosen for the experiment.

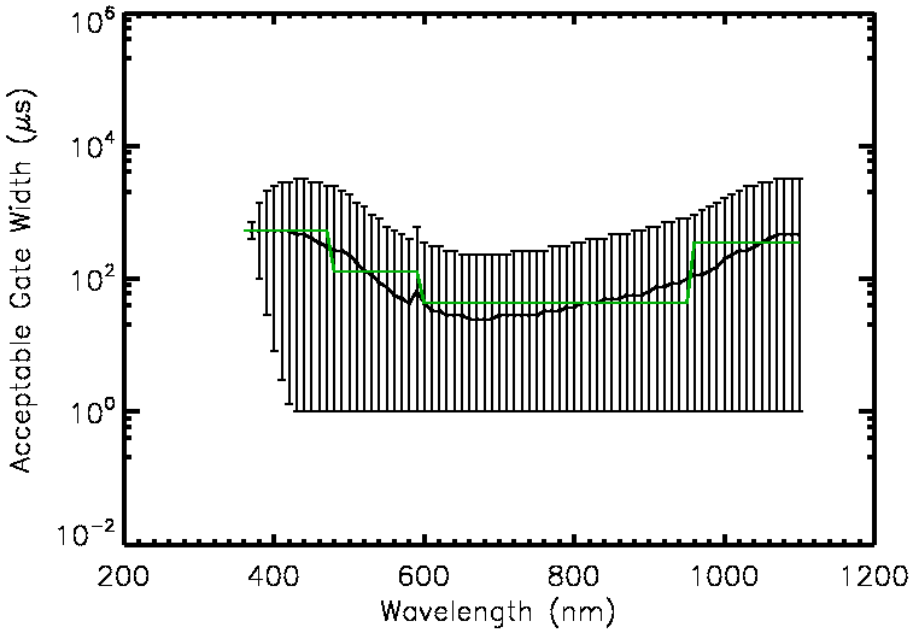

Figure 62 – This plot shows the ideal gate time (solid black line) vs. wavelength for the pre-radiation PDE experiment. The ideal gate time gives an avalanche probability of 0.797 per gate. The minimum and maximum gate times (error bars) are calculated by assuming a minimum SNR of 3, with a minimum allowable gate time of 1 μs. The stepped line shows the chosen gate times at each wavelength for the experiment.

Figure 63 shows the expected SNR vs. wavelength for the PDE experiment. The expected SNR is calculated using the chosen gate time values from Figure 62 and extrapolated values for photon flux, PDE, and DCR from a preliminary experiment. Some data points for S47 fall below an expected SNR of 3 at wavelengths less than 400 nm because S47 has higher DCR than the other two detectors. The same gate time setting is applied to all three detectors, even though they each have different DCR levels, and the gate times chosen yield greater than an SNR of 3 on average.

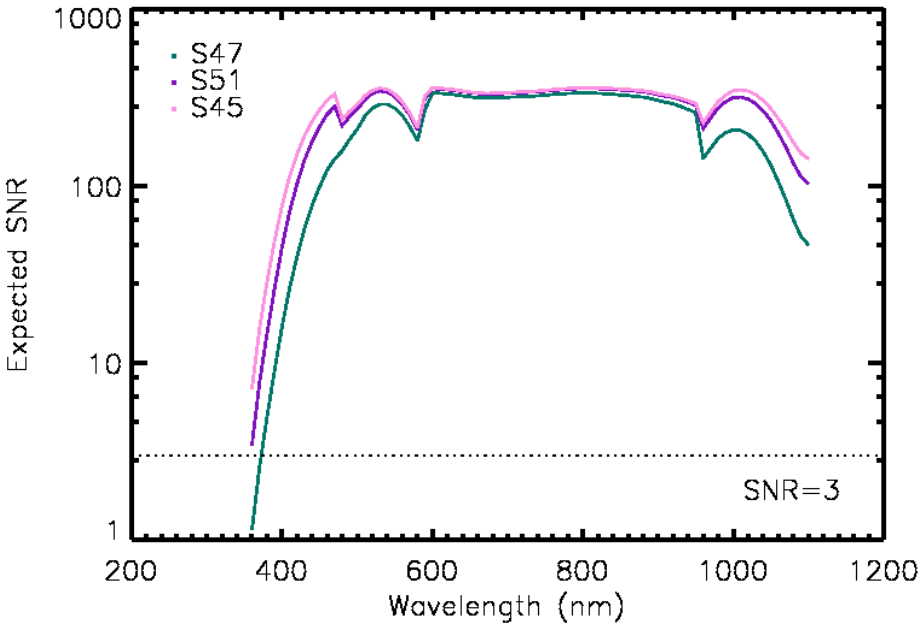

Figure 63 – This plot shows the expected SNR vs. wavelength for each of three GM-APD detectors (pre-radiation). Expected SNR is calculated based on chosen gate times and extrapolated values for photon flux, PDE, and DCR from a preliminary experiment.

Figure 64 shows the median pre- and post-radiation PDE of a GM-APD device for a range of wavelengths at the same applied bias and temperature (130 K). The results are corrected for transmission loss at the detector surface (see Figure 39). The short-wavelength cutoff was ~480 nm in pre- and post-radiation results. The peak PDE occurred at 730 nm, and was 0.27% ± 0.03% before irradiation, and 0.19% ± 0.002% after. In pre-radiation testing, 46% of pixels were within 10% of the median and 7% were dead or hot pixels. In post-radiation testing, 55% of pixels were within 10% of the median and 8% were dead or hot pixels. The decreased sensitivity after irradiation is likely due to the shift in breakdown voltage discussed in section 3.1. The PDE experiment was repeated after adjusting the applied bias to reflect the new breakdown voltage. The PDE at 730 nm was 0.25% ± 0.0002%, which is not statistically different from pre-radiation results.

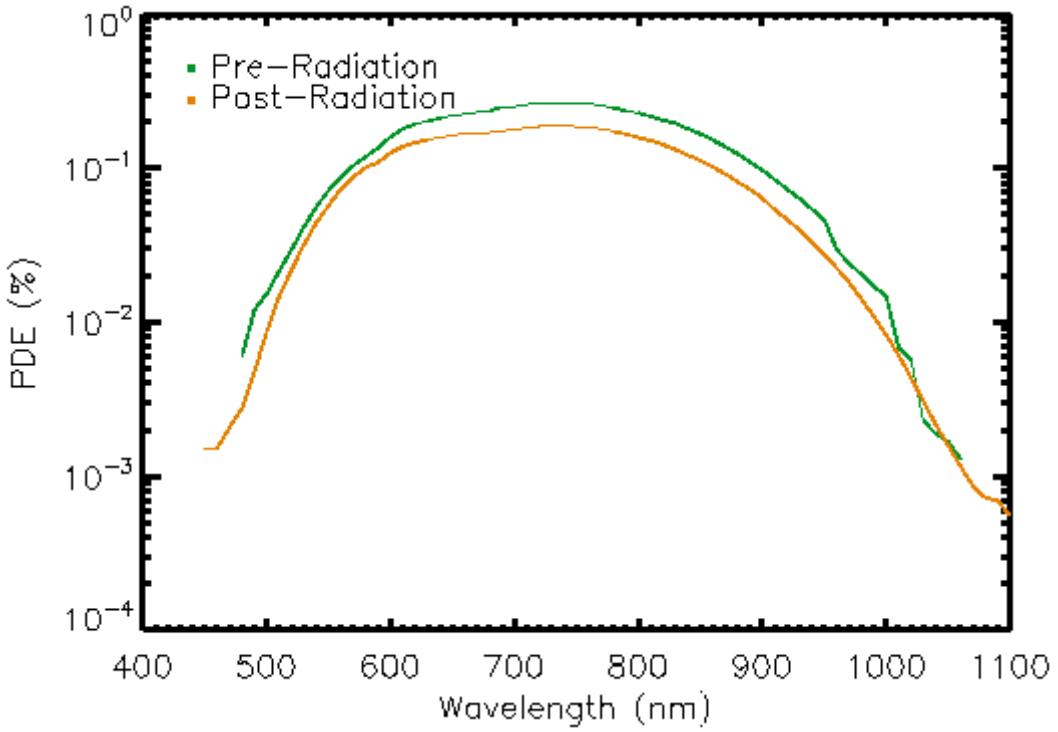

Figure 64 – This figure shows pre- and post-radiation PDE (%) vs wavelength for a GM-APD detector.

Post-radiation PDE was measured at 2 V overbias at 0.25% $\pm$ 0.001%. It is important to note that afterpulsing probability is not well-characterized at 2 V overbias, though it is high. Therefore, the measured PDE is likely an underestimate, given that the probability of additional counts due to afterpulsing is higher for lower count rates (dark measurements) than for high count rates (illuminated measurements). If corrected for afterpulsing, the PDE might increase to the 0.27% of pre-radiation levels, though it is not statistically different from the pre-radiation PDE without correction.

The PDE presented in Figure 64 is not corrected for fill factor, and is equivalent to the mean of the sensitivity over the entire pixel. The measured PDE would be much higher if the signal were concentrated in the center of the pixel (e.g., by a microlens (Itzler, et al., 2010)). The IPS function and the measured PDE were used to calculate the sensitivity at the center of the pixel. The pre- and post-radiation peak sensitivity at 730 nm were 1.3% and 0.7% (1.2% for the bias-corrected results), based on the IPS functions in section 3.9.

3.11 CROSSTALK

Crosstalk is the correlation of events in neighboring pixels. To measure crosstalk, the IPS experiment was updated to minimize the impact of the laser spot size on the result. Since the laser spot size is smaller than the pixel size (even when slightly out of focus) crosstalk was measured as the nearest neighbor trigger probability when the laser spot was focused on the center of the pixel. The laser had a wavelength of 632 nm and was pulsed at 40 MHz, averaging 400 pulses per gate.

Crosstalk probability was not expected to change due to radiation damage. Like IPS and PDE, crosstalk is a function of the internal structure of the GM-APD pixels. Since the structure itself is not affected by radiation damage, it is unlikely that radiation damage would induce any change in crosstalk probability.

The devices showed very low crosstalk probabilities before irradiation, with a mean nearest neighbor crosstalk probability $0.44\% \pm 0.23\%$. The post-radiation crosstalk probability was $0.49\% \pm 0.17\%$, which falls within the range of pre-radiation crosstalk probability. Pre- and post-radiation crosstalk probabilities are not statistically significant.

For crosstalk generated by avalanche emission, wavelength does not influence crosstalk probability. As long as the photo-generated carrier initiates an avalanche, the photon emission from the avalanching carriers will follow the same distribution. For crosstalk due to insufficient electrical isolation between pixels (e.g., a weak electric field), shorter wavelengths absorbed closer to the back side may be more likely to find their way to an adjacent pixel. This type of crosstalk is the most likely cause of crosstalk in the LFF devices because of the small volume of the multiplication region. The low level of crosstalk measured is likely due to insufficient electrical isolation, and is mitigated by the scupper regions that surround each pixel's active area (see Figure 38). These regions prevent carriers that are generated outside of the active region from initiating an avalanche.

Poor optical and electrical isolation can result in very high crosstalk, as was seen in HFF devices when tested at the CfD. Nearest neighbor firing probabilities of ~2% were measured (after

background subtraction), most likely due to optical crosstalk as a result of large avalanche volumes. Figure 65 shows crosstalk probability as a function of pixel location from a central pixel that fires (Figer & Kolb, 2014). The point spread function extends with significant afterpulsing probability up to 3 pixels away from the firing pixel at low overbiases (shown). At higher voltage settings, group firing occurs, where large blocks of pixels fire with very high probability, as the optical crosstalk propagates during a gate.

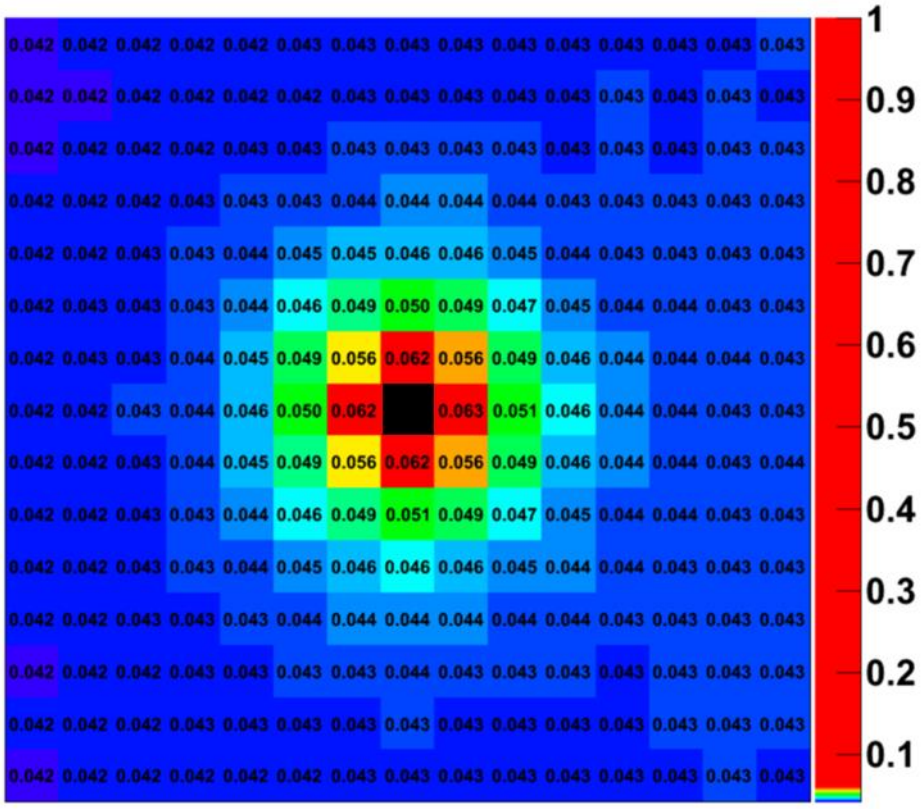

Figure 65 – This figure shows crosstalk probability as a function of distance from the central pixel (in black). Crosstalk probability was calculated using a correlation matrix method from individual gate data (Figer & Kolb, 2014).

3.12 GM-APD PERFORMANCE SUMMARY

For the 1 solar cycle post-radiation interpolated values, the effects of radiation on PDE are considered to be linear and the relative trend in DCR is assumed to be the same as in pre-radiation measurements. The post-radiation results presented in previous sections represent a

cumulative radiation dose of 50 krad(Si), or 10 solar cycles at an L2 orbit. Based on in-situ radiation testing, extrapolated performance after 1 solar cycle, 5 krad(Si), are also included in simulations. For all simulations, t_{gate} = 300 μs and λ = 730 nm. The duty cycle is calculated by assuming that t_{ho} is the minimum required to avoid afterpulsing at the optimum operating temperature.

Table 3 – SNR characteristics for a GM-APD device at various radiation levels are shown.

Parameter	**Pre-Radiation Value**	**Post-Radiation Value (1 solar cycle)**	**Post-Radiation Value (10 solar cycles)**
DCR (Hz)	38.2	50.7	17.4
PDE (%)	0.3	0.3	0.2
Duty Cycle (%)	96.8	96.8	85.7
Optimum Operating Temperature (K)	160	160	140

The effective overbias for the post-radiation data at 10 solar cycles is 0.5 V, which is why the DCR is lower than the other two radiation levels. For reference, at the same temperature and overbias, the pre-radiation DCR was 5.9 Hz. This means that the DCR at those settings increased by 11.5 Hz after 10 solar cycles of simulated radiation, or 2.3 e^-/s/pix/krad(Si).

Figure 66 shows the expected SNR for pre- and post-radiation performance characteristics. The projected SNR of a detector at a specific fluence helps determine if the detector could be used for a given application.

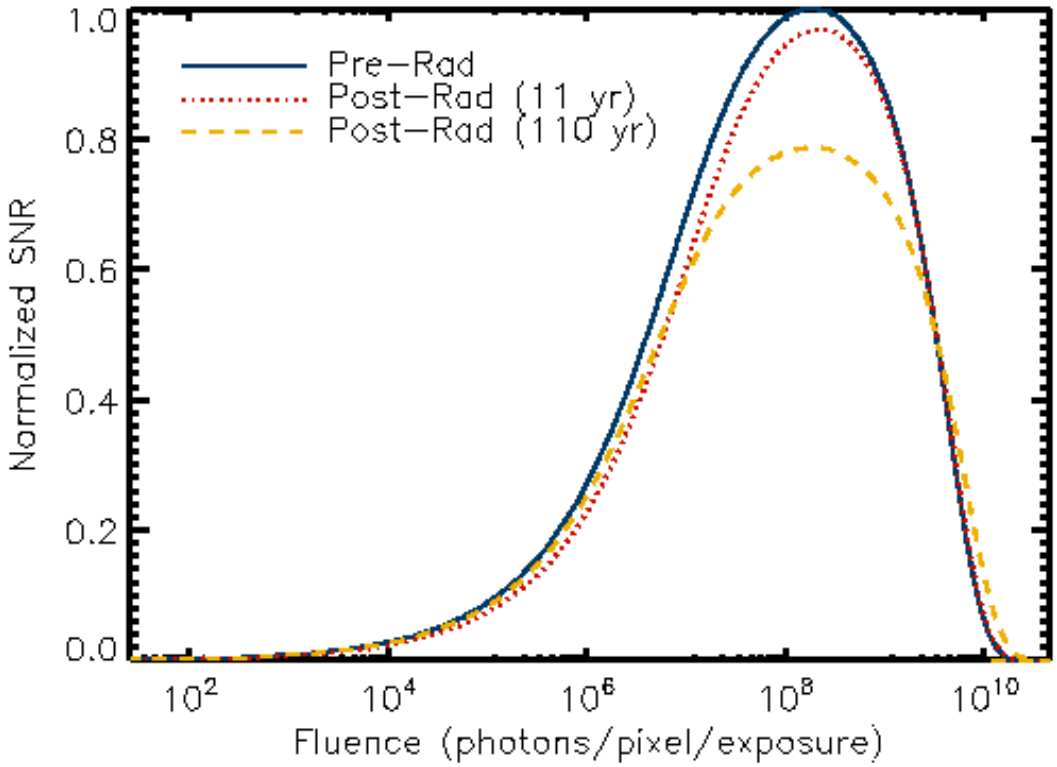

Figure 66 – This figure shows pre- and post-radiation SNR of a GM-APD for a 1000 s (wall time) exposure. The SNR is normalized to the maximum relative SNR before irradiation.

4 EMCCD TESTING

The EMCCD testing was done at NASA JPL under the direction of Dr. Shouleh Nikzad. The device was the e2v CCD201, which is being considered as a possible candidate for the WFIRST mission (NASA, 2014). Figure 67 shows a schematic of the device, including an overview of the active area and readout elements (e2v technologies, Nov. 2011).

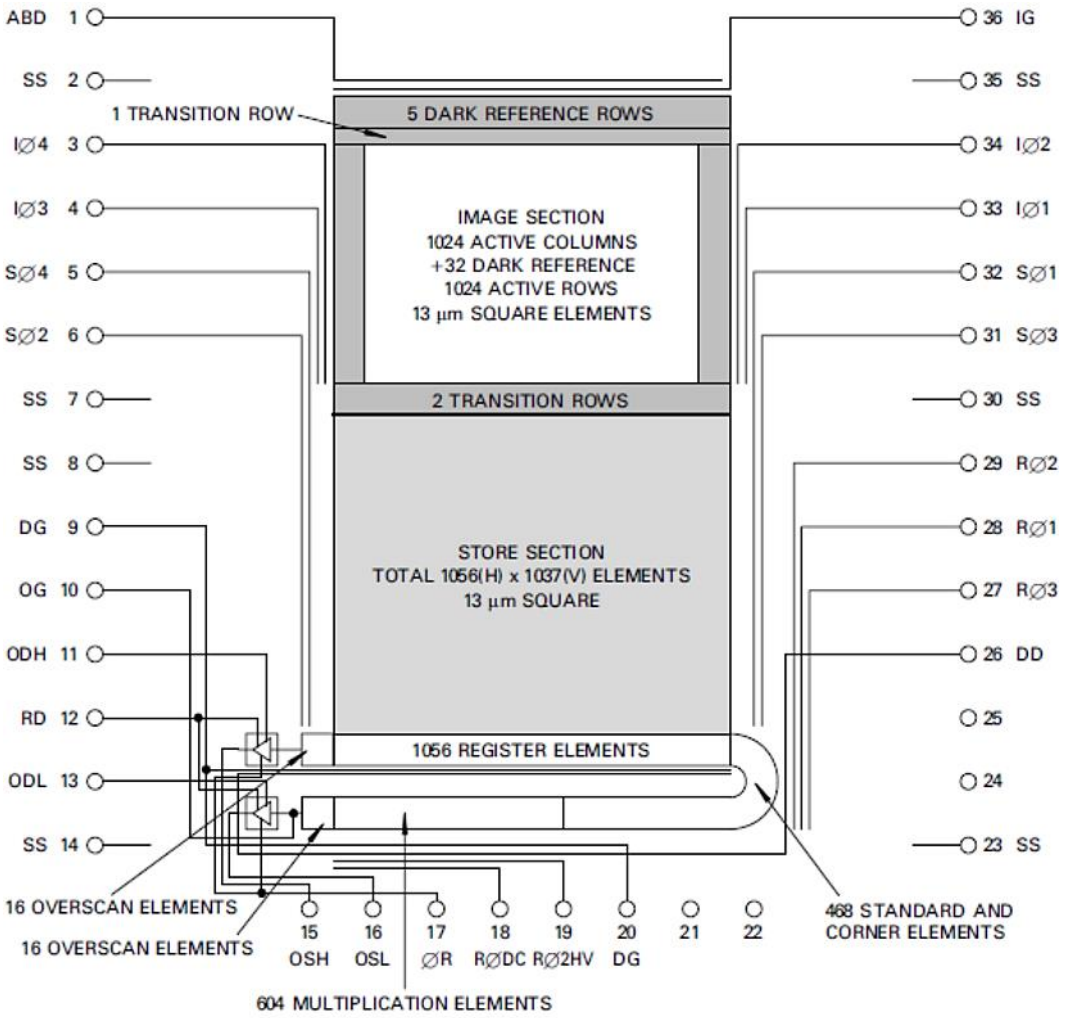

Figure 67 – This schematic shows the layout of the e2v CCD201 EMCCD chip (e2v technologies, Nov. 2011).

The device can be read out in two different ways. The first is through the standard output amp after the transfer register at the bottom of the store section. The second is through the

multiplication register, which is shown beneath the transfer register and has 604 multiplication elements or stages. Effectively, the signal is either read out through a standard output amplifier or it is routed through a multiplication register and then to another output amplifier. The voltage across each stage in the multiplication register is set by RØ2HV, a high voltage clocked input. Figure 68 shows the clocking pattern for RØ2HV in relation to the standard register clocking signals RØ1, RØ2, and RØ3 (e2v technologies, Nov. 2011). This is the pattern that was used for the testing presented, with the maximum amplitude of RØ2HV equal to 42 V.

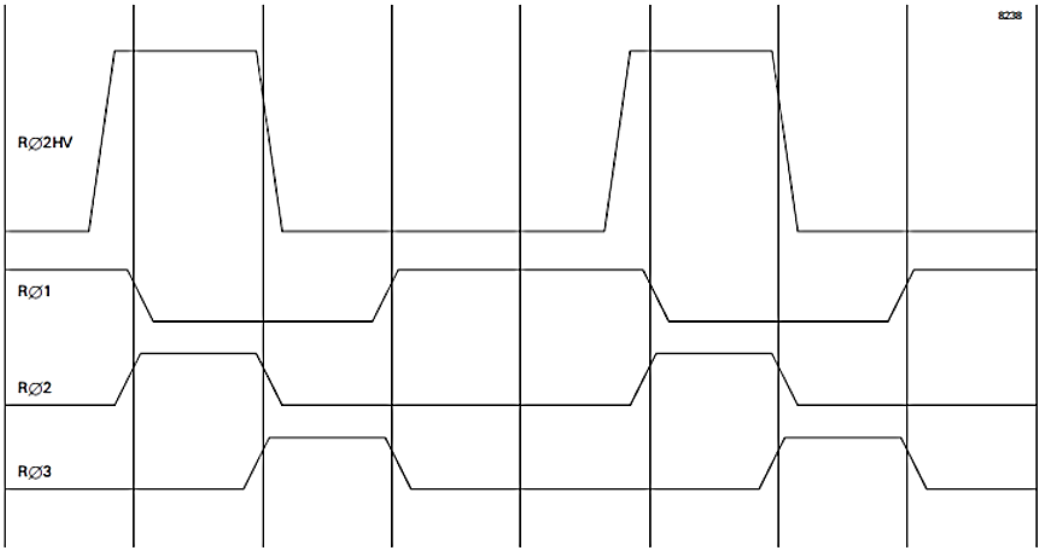

Figure 68 – This figure shows the clocking pattern required to operate the multiplication register in high gain mode (e2v technologies, Nov. 2011).

To reduce CIC in the readout stemming from high fields during the rise and fall of the RØ2HV signal, a sine wave input can also be used, as shown in Figure 69 (e2v technologies, Nov. 2011).

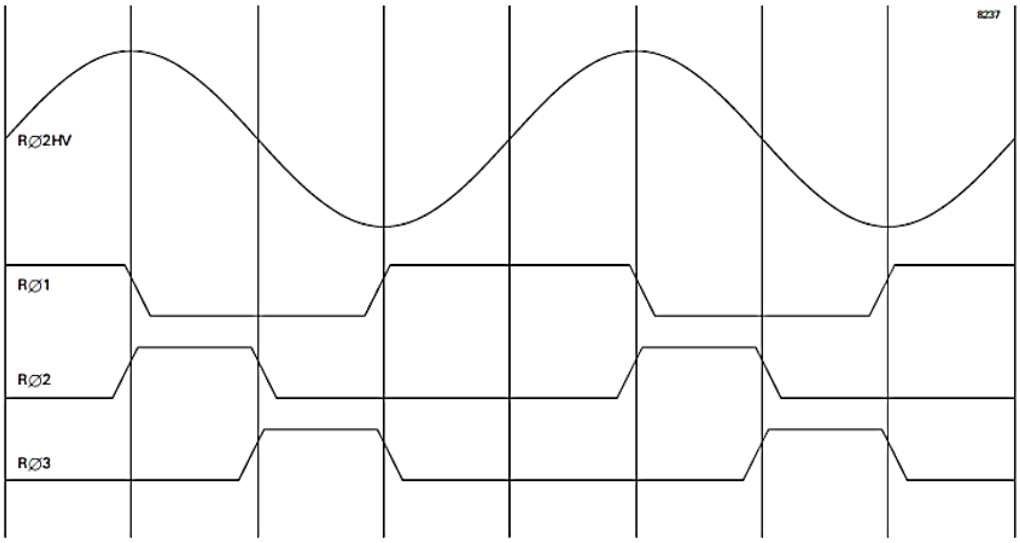

Figure 69 – This figure shows an alternate clocking pattern for reducing CIC in high gain mode (e2v technologies, Nov. 2011).

Figure 70 shows multiplication gain as a function of RØ2HV voltage and temperature (e2v technologies, Nov. 2011). As the voltage across the multiplication register elements increases, the gain increases. Similarly, as the temperature decreases and thermal resistance decreases, the gain associated with the same voltage increases. This is similar to the shift in breakdown voltage seen in GM-APDs when the temperature changes (see Figure 44). The testing presented here was done at 168 K, which is not shown in Figure 70.

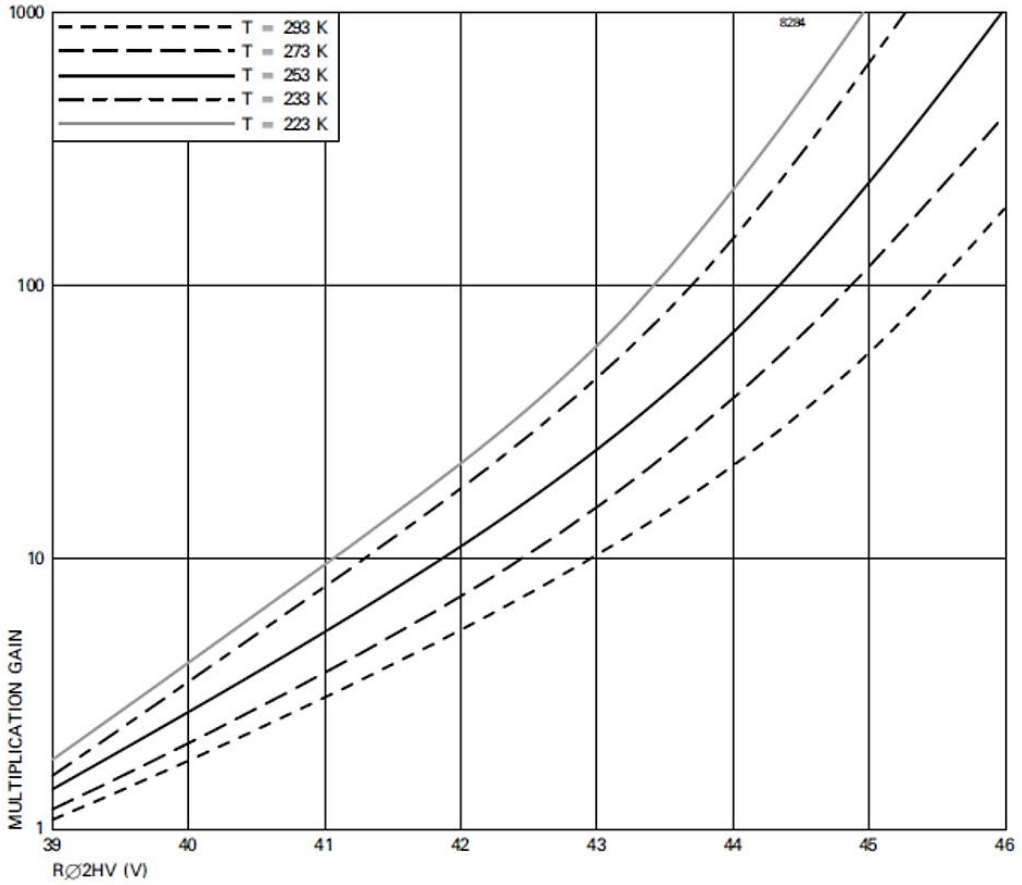

Figure 70 – The relationship between RØ2HV amplitude, temperature, and multiplication gain through the multiplication register is shown (e2v technologies, Nov. 2011).

The EMCCD was read out using LEACH[6] electronics, developed by Dr. Bob Leach at Astro-Cam, and coding and clocking developed by Dr. Timothy Goodsall of JPL.

4.1 EMCCD GAIN CHARACTERIZATION

In order to measure the gain, the RØ2HV voltage was changed from 12.5 V to 42 V in 9 steps. Then, ten frames were taken with varying exposure times from 0.1 s to the shorter of 250 s or the 90% saturation time. The slope was measured in digital number (DN) per second with a linear fit to the mean measured data in a central region of the detector. Figure 71 shows an example of the data at a single RØ2HV setting of 37.8 V.

[6] http://www.astro-cam.com/

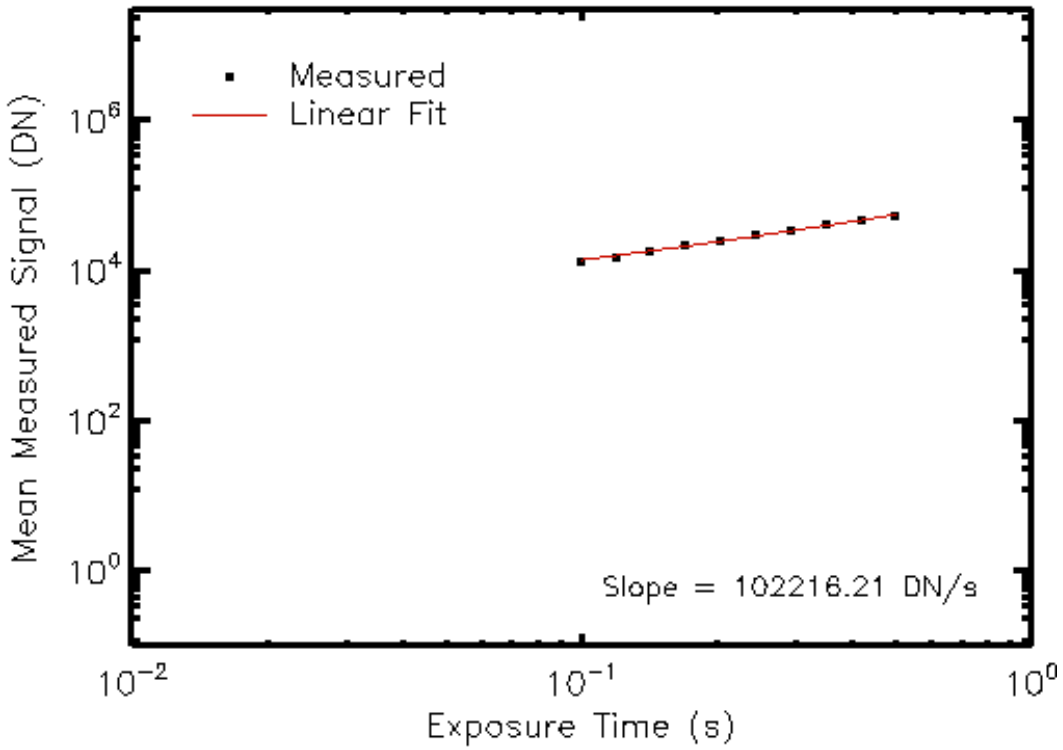

Figure 71 – This figure shows the mean measured signal from a 300 x 1000 region of pixels in the center of the detector as a function of exposure time for RØ2HV = 37.8 V.

Assuming that the multiplication register gain is equal to 1 e^-/e^- at the RØ2HV = 12.5 V setting, the mean measured "dark current" in DN/s at each voltage setting can be divided by the dark current at the 12.5 V setting to calculate the gain. Figure 72 shows the results from the experiment for various values of RØ2HV, labeled as V_{reg}, assuming that the output from the supply voltage scales linearly. V_{stage}, in mV, is the voltage across each multiplication element in the register. The gain does not start to increase significantly until V_{reg} is greater than roughly 35 V, and V_{stage} approaches 60 mV.

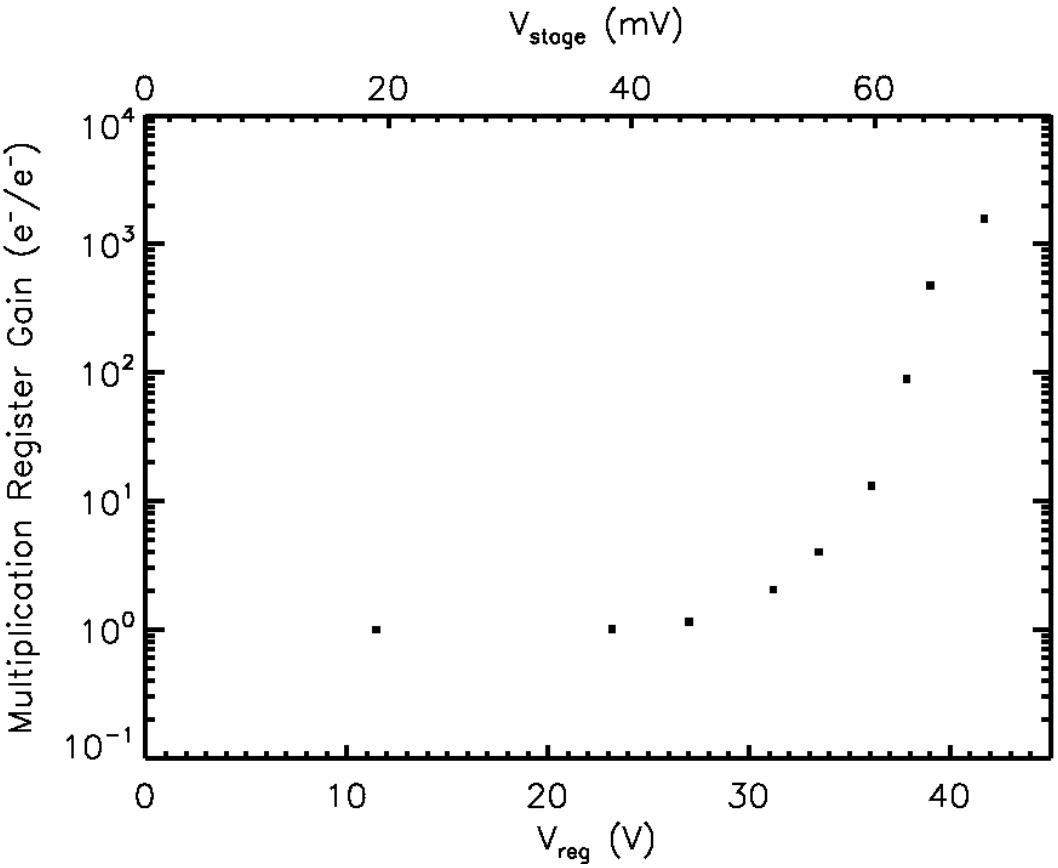

Figure 72 – This figure shows the measured multiplication register gain as a function of the voltage across the register (RØ2HV) and across each stage.

4.2 SERIAL CIC VS GAIN

As the voltage across the gain register increases, the serial CIC increases as well. Serial CIC is caused by high electric fields in the gain register between each stage, and results in tunneling and charge injection, even when there is no signal. To measure serial CIC, the values in the overscan rows and columns are measured for different gain voltage settings (see Figure 67). The overscan regions are extra reads taken when reading out the array. It contains no pixel information, and is an estimate of read noise for standard CCDs, and read noise and serial CIC for EMCCDs in high gain mode. Figure 73 shows the measured serial CIC at each voltage setting.

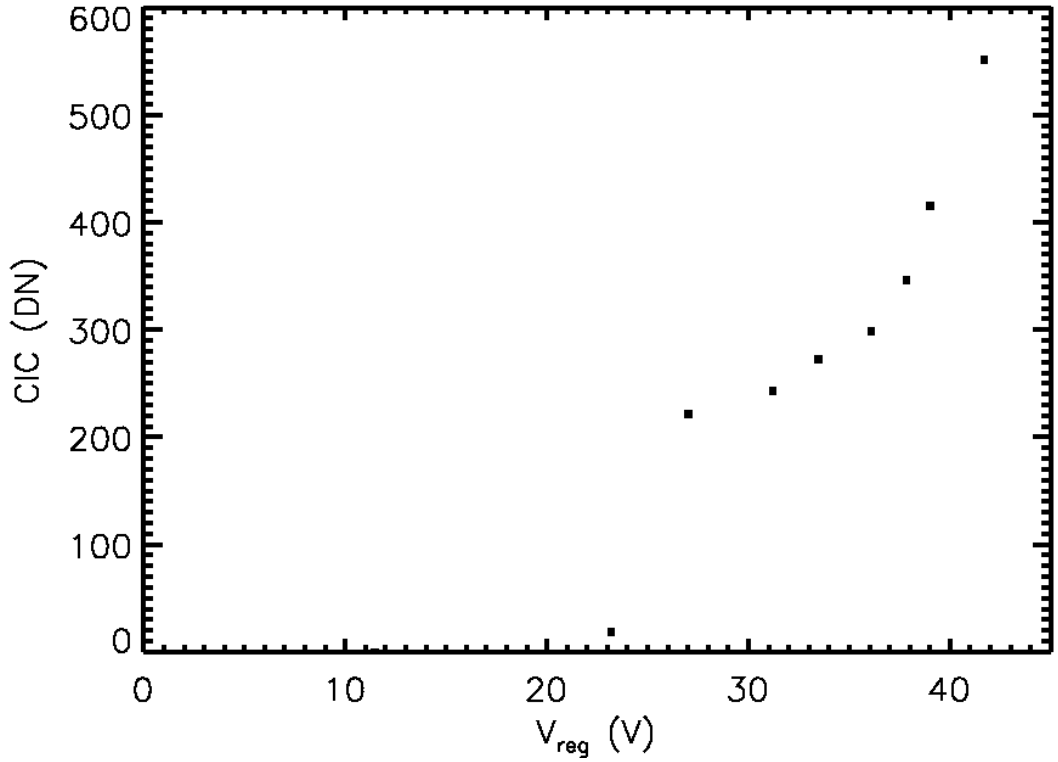

Figure 73 – This figure shows the measured serial CIC for each gain register voltage setting.

Serial CIC may be generated in any one of the multiplication register stages, and so the carriers may be multiplied by the gain in the register. This leads to a wide distribution, characterized by a non-Poissonian increase in the variance of the output at high gain settings. Figure 74 shows the measured variance of the overscan rows and columns as a function of the multiplication register voltage.

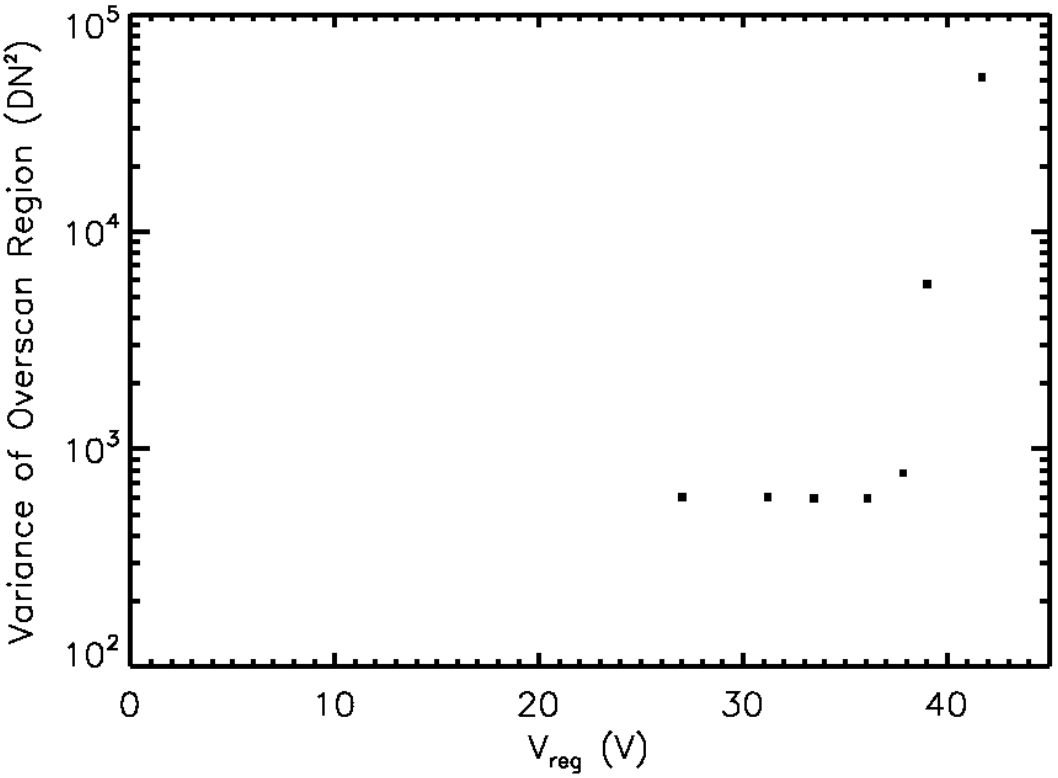

Figure 74 – This figure shows the measured variance of the overscan region for each gain register voltage setting.

In order to avoid counting serial CIC in photon-counting mode, the threshold must be set sufficiently high so that a non-detection does not artificially trigger a photon count with any significant probability. This condition may become prohibitive at the higher gain values when balanced in the overall SNR (see section 2.7.2).

5 LM-APD PERFORMANCE

As discussed in section 2.5, only HgCdTe LM-APD devices are discussed in this paper due to their noiseless gain properties. Other detector materials, such as silicon or InGaAs, have an ENF due to uncertainty in the gain.

The sensitivity range of HgCdTe devices can be tuned by engineering the band gap of the material. When the ratio of HgTe and CdTe is altered, the band gap decreases with the increased presence of mercury (a metal). Given the temperature, T, and a material composition of $Hg_{1-x}Cd_xTe$ the band gap of the device can be described by Eq. 98 (Norton, 2002).

$$E_g = -0.302 + 1.93x - 0.81x^2 + 5.35(1 - 2x)T^2 \qquad \text{Eq. 98}$$

Therefore, the cut-off wavelength varies with both temperature and composition, as shown in Figure 79.

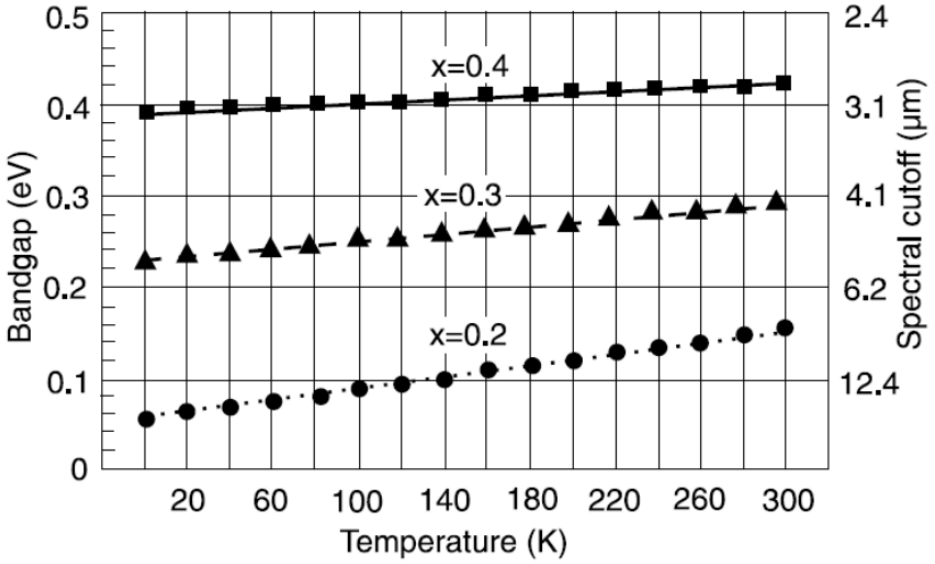

Figure 75 – The band gap and cut-off wavelength for a HgCdTe device are shown for various temperatures and stoichiometric ratios (Norton, 2002).

By tuning the band gap of the absorption and multiplication regions, the device can have sensitivity ranging from the visible to the IR.

5.1 Gain Uncertainty

In the ideal case, a deterministic number of electrons will be generated by the avalanche process, but there are exceptions that must be taken into consideration when designing the device. Electrons travelling the full width of the multiplication region of the device will have a noiseless gain as anticipated. Electrons that are generated inside the multiplication region, however, will have decreased gain. The gain is deterministic for any given location, since the gain will simply scale with the width of the multiplication region that is travelled. The uncertainty in aggregate gain comes from uncertainty in the position of the generated electron.

The photon absorption distribution function is exponentially distributed from the surface of the device. For shorter wavelengths, almost all of the photons are absorbed within a short distance from the surface because the mean absorption depth is small. For longer wavelengths, the mean absorption depth is large, resulting in a flatter distribution and a significant number of photons

being absorbed deep in the material. When the distribution extends into the multiplication region, the randomly-distributed absorption depth causes uncertainty in the gain for a photon of a given wavelength. This noise is exacerbated when using band-pass filters, as photons can have a range of wavelengths within the band. The problem is worse at longer wavelengths, as shown in Figure 76 (Finger, et al., 2013). While the figure shows two different growth styles, the important distinction is the difference in the ENF for the K (2.2 μm) and H (1.65 μm) bands in the same growth style (MOVPE, or metal-organic vapor phase epitaxy). The ENF for the longer wavelength in the K band is higher than for the H band.

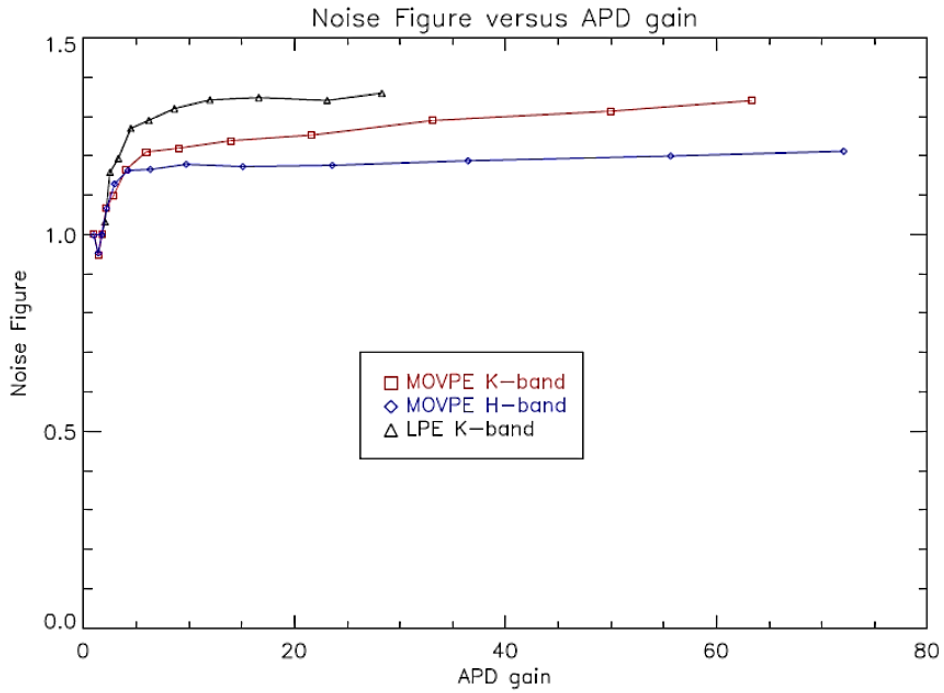

Figure 76 – This figure shows the excess noise factor, F, as a function of the gain of a HgCdTe LM-APD. The highest values of F are for the LPE (liquid phase epitaxy material growth style) device in the K-band (2.2 μm). The MOVPE (metal-organic vapor phase epitaxy growth style) devices show a difference in F for K and H (1.65 μm) bands, where the longer wavelengths in the K band have a higher excess noise factor (Finger, et al., 2013).

One solution to this problem is to increase the optical thickness of the absorption region of the LM-APD so that nearly all photons travel the full width of the multiplication region. A thicker absorption region will decrease the QE at shorter wavelengths, however, because the electrons will have farther to travel to reach the multiplication region before recombining.

Another solution is to redesign the structure of the pixel itself. Instead of using the vertical GM-APD design (see Figure 38), a ring structure can be used to eliminate the effects of wavelength on absorption depth and the ENF. This ring structure is common for APD structures in all material types, though they are usually used in large-scale, single-element devices. In an imaging device, each pixel is built as a set of concentric rings, with the absorption region in the outermost ring encircling the multiplication region. Figure 77 shows an example of this structure (Beck, et al., 2006). The ring structure minimizes the amount of time required for carriers to reach the central depletion region and limits the effects of wavelength-dependent ENF. However, there is still an ENF associated with all wavelengths, since photons may still be absorbed in the multiplication region and have reduced travel time (and therefore reduced gain).

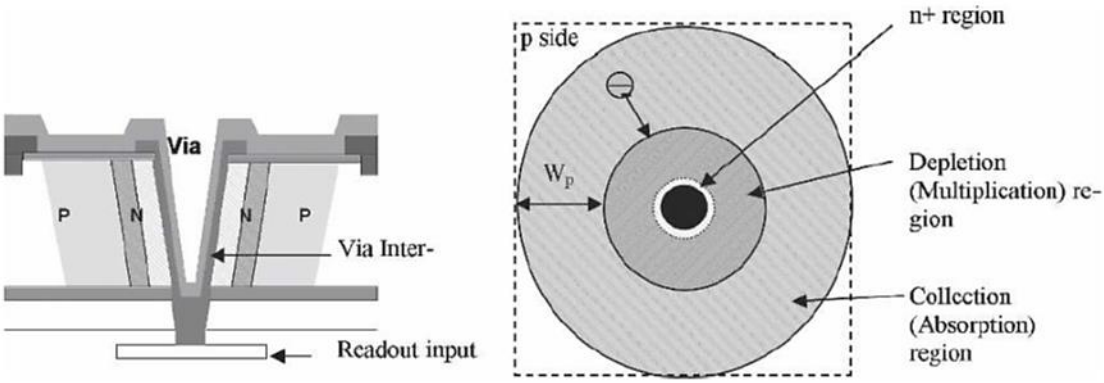

Figure 77 – Top-down and cross-sectional views of a HgCdTe ring APD are shown (Beck, et al., 2006).

5.2 TUNNELING CURRENT

In addition to keeping the ENF low, the leakage current and DCR must be kept low. For detectors with long wavelength cut-offs (small band gaps), the thermal dark current is too high to use for astronomy applications unless the detector is cooled to cryogenic temperatures. This is a fundamental limitation of small band gap devices and is common to all IR detectors. In a LM-APD, there is another concern due to the high electric field in the multiplication region: tunneling. With high electric field and significant trap density, electrons can tunnel through the band gap, assisted by a trap energy state in the middle of the band gap (Beck, et al., 2006).

Figure 78 shows the relationship between electric field in the depletion (multiplication) region and tunneling current (Finger, et al., 2013).

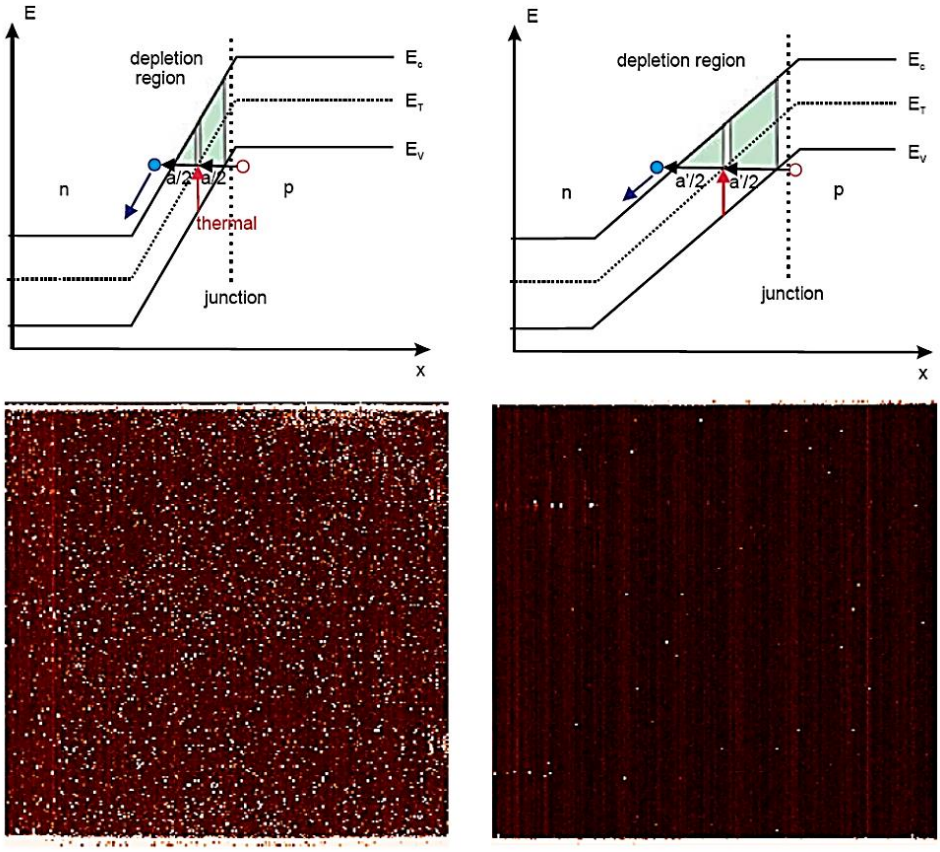

Figure 78 – This figure illustrates the relationship between electric field and tunneling current. Top left: band structure diagram for a narrow depletion width with high electric field that is conducive to tunneling. Bottom left: dark image taken with the device design shown in top left. Top right: band structure diagram for a wider depletion region and lower electric field to reduce tunneling. Bottom right: dark image taken with the device design shown in top right (Finger, et al., 2013).

While some users choose to operate HgCdTe LM-APDs at very high gains (on the order of 10,000), the benefits of the noiseless gain mechanism are better realized as smaller gains ≤ 100 due to tunneling current. Tunneling probability decreases exponentially with decreasing electric field, which means that decreasing the gain also helps to reduce tunneling noise. There is a balance to be struck, however, since the gain must be high enough to make the read noise

negligible, or photon counting is not possible. Depending on the readout rate, the read noise might be on the order of tens of electrons, which necessitates gain on the order of hundreds.

6 CURRENT AND FUTURE STATE-OF-THE-ART

In order to objectively compare detectors, the state-of-the-art performance for each detector type must be defined. CCDs have had a long history of improvement, and as such are a very mature technology with very little noise under optimum conditions. CMOS APS detectors are also a mature technology with low noise and certain advantages in imaging applications, including a non-destructive read out for up-the-ramp sampling. EMCCDs, both analog and photon counting, are based on CCD detectors, the main difference being an added multiplication register in the readout circuitry. Because EMCCDs are based on an already-mature technology, the CCD, they also have low noise in optimum conditions. They are limited in their performance, however, by the ENF associated with the gain and the maximum frame rate for a large format device.

Geiger-mode and linear-mode APDs both make use of avalanche gain, but avalanche noise is not a concern for either. GM-APDs make us of a digital readout that detects any self-sustaining avalanche. The LM-APDs presented here avoid excess noise due to the avalanche gain by using electron injection in HgCdTe, which has noiseless avalanche gain.

It is important to understand not only the noise source specifications, but the way in which each noise source affects the SNR of the detector. Each detector's state-of-the-art performance is outlined below, as well as predictions for future state-of-the-art.

6.1 CCDS

Current state-of-the-art CCDs have very low dark current and very low read noise at slower frame rates and heavily-sampled signals. Wide Field Camera 3 (WFC3) on HST has dark current on the order of 0.0002 e-/s/pixel at standard operating temperature and a minimum read noise of just 2 e^-/frame (Gilliland, et al., 2010; Basden, et al., 2003). Other state-of-the-art CCDs show the same low noise (Basden, et al., 2003). The full-well saturation limit for WFC3 is 72,000 e^-, which is typical. Duty cycles of CCDs for long exposures are nearly 100% because the only dead time the detector sees during operation is the final readout. For short exposures, if the readout time is greater than the exposure time, the detector cannot be read out during the subsequent

exposure. This leads to smaller duty cycles. Additionally, faster readouts lead to higher read noise that can be 10x greater than the minimum read noise of the detector.

As array sizes grow, faster frame rates become more difficult to achieve. Pixel readout rates rise geometrically with increasing format, but even maintaining pixel readout rates leads to geometrically longer readout times. As pixel rates increase, CIC becomes a concern due to the high electric fields during the faster clocking, and CTE starts to suffer as well. The solution is more parallel channels to read out all the pixels, but data rates eventually become a limiting factor as well. What this means for future state-of-the-art is that large-format CCDs will be difficult to use for high frame rate applications, even more than they are now. And for noise levels, there is not much room left for improvement. Faster reads will exacerbate read noise because of reduced sampling capability, and dark current is already at a minimum.

6.2 EMCCDs

Analog-mode EMCCDs make use of a multiplication register in the detector read out circuit that applies a gain to the signal in each pixel. Due to the nature of the multiplication, an ENF is introduced that effectively halves the QE in SNR calculations. However, the read noise associated with fast readouts in CCDs and CMOS APS detectors is reduced by a factor equal to the gain, making the read noise negligible. Because EMCCDs are based on CCD arrays, they have the same low dark current (~0.0002 e^-/s/pixel) without the higher read noise associated with fast readouts.

Unfortunately, CIC becomes significant at high frame rates. CIC is charge created when signal is quickly transferred between pixels, and has noise characteristics identical to dark current. Therefore, the dark signal associate with EMCCDs must include both dark current and CIC (which add linearly). Typical CIC at high frame rates is 0.003 e^-/pixel/frame in the array, a 10x increase in the total dark signal from CCDs (Basden, et al., 2003). Saturation levels in the pixel are not as important as saturation levels in the readout circuitry, which is typically on the order of 1,000,000 e^- after multiplication. With gains ~1000, the saturation of the readout is on the order of 1,000 e^- before multiplication. Without major innovations in clocking patterns, CIC will

continue to be a constraining factor for EMCCDs at high frame rates, eclipsing the read noise and the dark current noise.

Recall that photon counting EMCCDs also use multiplication registers to apply a gain to the signal in each pixel of the array. However, instead of using the analog output signal to estimate the number of photons absorbed, photon counting EMCCDs use a thresholding circuit to determine if there were any carriers in each pixel. This approach is similar to the GM-APD thresholding and digital output. If the threshold is set correctly, any read noise and CIC in the multiplication register can be ignored and the ENF introduced by the multiplication register gain is negligible as well. Therefore, the only remaining noise sources are the dark current and CIC in the array, which are the same as the state-of-the-art analog-mode EMCCD detectors (Daigle, et al., 2012). Slower frame rates can lead to reduced or negligible CIC, though the frame rate must be matched to the signal so the number of photons per pixel per frame is on the order of 1-3. Signals in excess of a few photons per frame lead to saturation. Generally, the frame rates required for this small signal condition are associated with significant CIC. Therefore, photon counting EMCCDs also require a step forward in signal clocking to reduce CIC at high frame rates.

6.3 CMOS APS DETECTORS

CMOS APS detectors have a non-destructive read out system, where the pixel signals can be sampled during an exposure without resetting the pixel. CMOS APS devices are newer than CCD devices, and as such the state-of-the-art noise levels are slightly higher than in CCDs, but still competitive, especially given the advantages of the non-destructive read out. Dark current on the order of 0.02 e^-/s/pixel are common in infrared detectors for IR imaging and spectroscopy (one of the main applications), and minimum read noise levels are on the order of 3 e^-/read (Bai, et al., 2004). Like in CCDs, the read noise increases with increasing frame rates, to 10x the minimum read noise and higher.

With improvements in processing and fabrication, the dark current may decrease, though it is near a minimum now in terms of the physical processes that govern carrier generation. Read noise still has room for improvement, which might be done through advanced clocking

innovations or increased data rates and more parallel processing. However, as with CCDs, read noise of less than an electron or two is incredibly difficult to achieve during ideal conditions, and certainly not at high frame rates. This is a fundamental function of the sampling required for low-noise measurements, and so there is little room for improvement.

6.4 LM-APD

LM-APDs, specifically the HgCdTe detectors, have two types of dominant noise. The first is multiplied dark current, which is analogous to dark counts in GM-APDs. Dark carriers generated in the avalanche structure can be multiplied and counted along with photo-generated charge. The second dominant source of noise is un-multiplied dark current, which does not enter the avalanche region of the device. The dark current is still counted by the readout system, and adds uncertainty to the final signal. A long integration with significant un-multiplied dark current and few photo-generated carriers results in significant uncertainty in the number of photons absorbed. Although read noise is present in LM-APDs due to the traditional readout structure and analog output, it is effectively negligible due to the avalanche gain of the detector.

There is significant improvement to be made in the leakage current of LM-APD devices, hinging mostly on device material quality and fabrication capabilities. When the material quality increases, the leakage current should decrease significantly. The DCR will also decrease, and the noiseless gain can have more of an impact.

6.5 GM-APD

Current state-of-the-art silicon detectors have dark current on the order of 8 e-/s at room temperature (Jackson, et al., 2002), achieved through various processing and design improvements. Efficiency in other GM-APD devices is near 80% (Renker & Lorenz, 2009).

Although the first-generation devices had significant noise and very low efficiency, a few simple, targeted improvements would greatly improve their performance. As alluded to in section 3.10, a microlens array would greatly improve PDE by focusing incident photons in the center of the pixel. However, re-designing the internal device structure would lead to the most significant gains in performance.

As designed, the scupper region in Figure 38 mitigates DCR at the expense of efficiency. Carriers generated by photons in the absorption region have a significant probability of moving to the scupper region. However, the scupper would no longer be necessary if the dark current were not so high. The high dark current in these devices is due to a number of factors, including damage introduced by thinning the detector. Improving the dark current would eliminate the need for the scupper region and allow the efficiency to increase to the levels seen in other GM-APD devices, near 80% (Renker & Lorenz, 2009). Increasing the quality of the material will also lead to a decrease in afterpulsing. Ideally, there should be no traps, and therefore no afterpulsing, in a majority of the pixels.

The DCR in the first-generation devices was also high, even when cooled. This suggests that the dominant noise source is not thermally-generated dark current due to the shallow decrease in DCR with decreasing temperature. Other noise sources, such as trap-enhanced tunneling current and surface defect-generated current, are not temperature-dependent and would keep the noise floor high. For the devices presented in section 3, the noise floor seems to be caused by backside defects introduced during the thinning process. This conclusion is based on testing of an un-thinned device. As shown in Figure 79, the noise floor for the un-thinned device is much lower than for the thinned device. The DCR at 150 K is three orders of magnitude lower for the un-thinned device.

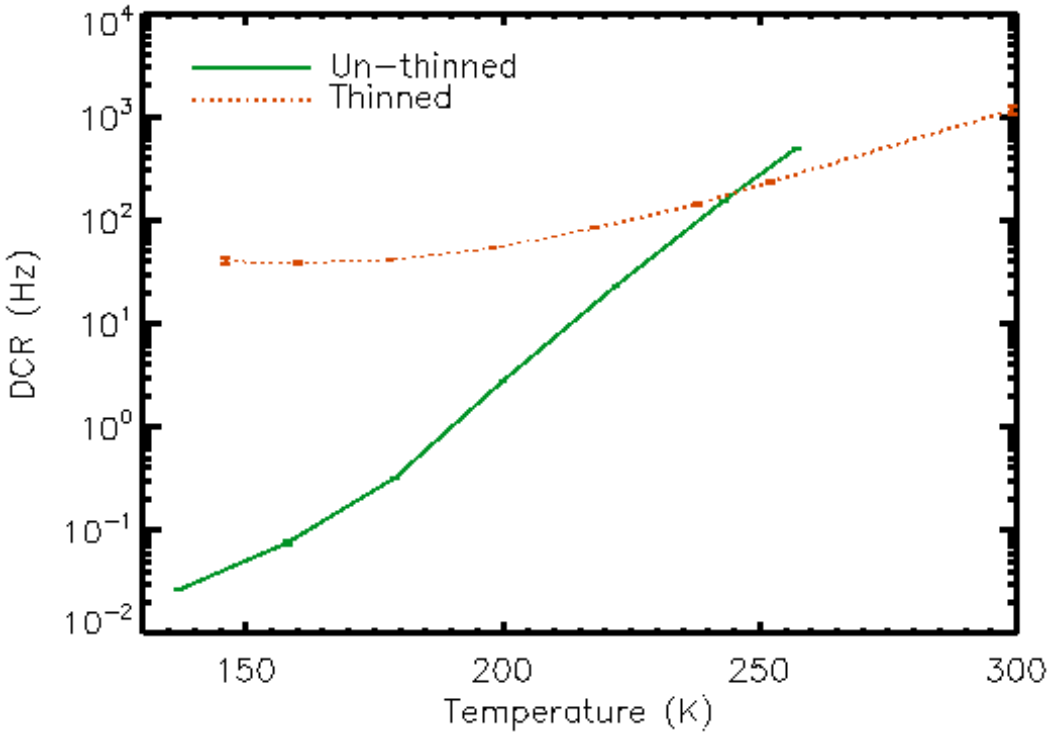

Figure 79 – This plot shows the median DCR for an un-thinned GM-APD array device and a thinned device presented in section 3. The DCR at warmer temperatures (above 175 K) for the un-thinned device doubles every 8 K, which is expected for thermally-generated DCR.

Silicon on insulator (SOI) fabrication of devices would improve GM-APD performance in three ways: decreased damage at the back surface, increased radiation tolerance, and decreased power dissipation. Rather than using a mechanical thinning process to thin the wafer before dicing (and introducing damage to the back surface), SOI processing allows for a more gentle removal of the substrate. This leads to a large decrease in current generation at the back surface. The use of SOI processing also has major advantages with respect to radiation effects, as compared to processes that use bulk silicon wafers. This is because the volume of the APD absorber layer can be made much smaller than that in bulk CMOS, thereby reducing the cross section to radiation and the resulting radiation damage.

SOI-based APD structures would have another advantage for the hybridized detector as a whole, as well. Thinner absorber and multiplication regions require less applied voltage to provide the electric field required for avalanche breakdown. This would make the devices more compatible with low-voltage SOI CMOS circuits, since commercial SOI CMOS readout circuit technologies

generally have more limited voltage swings than bulk processes. A low-voltage-optimized APD paves the way for scaling to low-voltage CMOS, making many of the SOI technology nodes viable for this application, and, therefore, enabling very radiation-hard focal planes.

Lower volume also has the added advantage of lower DCR, since thermal carrier generation is a function of volume, and lower crosstalk probability. When designing GM-APDs, care should be taken to keep the volume of the multiplication region as small as possible. Increased volume leads to more carriers participating in each avalanche. This increases optical crosstalk between pixels, which has been measured in devices that have large multiplication regions. Figure 80 shows the effect that high optical crosstalk can have on a pixel's signal (Aull, et al., 2015). A smaller multiplier volume means smaller avalanches, which results in less optical crosstalk.

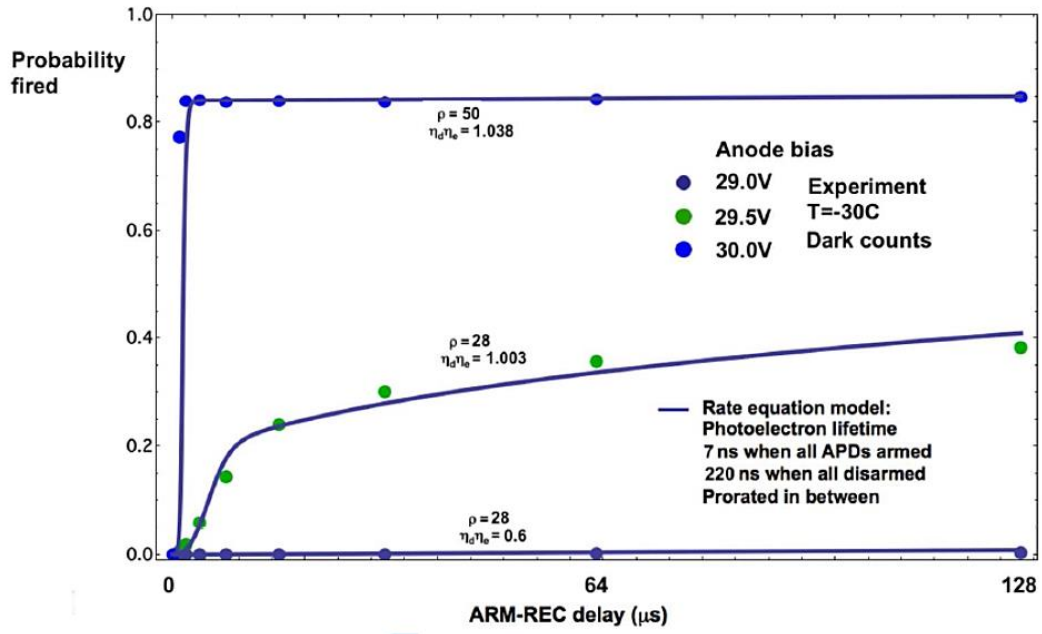

Figure 80 – This plot shows avalanche probability vs hold-off time for a GM-APD with high optical crosstalk probability. The data is fit with a rate model equation that takes the geometry of the devices and the avalanche photon emission spectrum into account (Aull, et al., 2015).

Significant optical crosstalk leads to large groups of pixels firing during a single gate, which makes signal estimation nearly impossible. In order to mitigate the effects of a larger multiplication area, which must occur if the active area is expanded, optical isolation trenches should be added between pixels.

7 Detector Comparison and Conclusions

Which is the best photon counting detector? To answer that question, all the strengths and weaknesses of each detector type should be considered as a whole for various applications. The best detector for AO will likely not be the best detector for photon-starved astronomy – not only because of different strengths and weaknesses, but because the strengths and weakness of each detector changes based on how it is operated. The following sections aim to compare CCDs / CMOS APS, EMCCDs, and GM-APDs both generally speaking and under specific conditions. LM-APDs are not included in this discussion due to their current limitations. It is unclear whether or not the dark current or leakage current can be reduced to make the LM-APD competitive at this time. Further, the development of the readout electronics has been stalled as of this writing due to the noise levels and ADC saturation triggered by noise alone, even for short exposures. At this time, the HgCdTe LM-APD requires further development to make realistic prognostications about its future use in photon counting applications.

Broad comparisons can be made about each detector to establish their unique characteristics, including saturation behavior, frame rate limitations, and radiation tolerance. Relative SNR vs fluence is a comparison tool that combines a variety of performance characteristics into a convenient, single metric. And as discussed previously in section 3.12, different pixel architectures are affected differently by radiation damage.

7.1 Radiation Tolerance

CCD- and CMOS-based detectors experience increases in dark current after radiation damage from two main sources: bulk damage and ionization effects. (Janesick, et al., 1989) The latter source is caused by damage at the surface of the devices at the silicon/insulator interface. In GM-APDs, this type of damage does not affect the DCR because the avalanche initiation probability for carriers generated at the surface of the device is effectively zero. The bulk damage is mostly comprised of deep-level defects (lattice displacement), which act as generation / recombination centers in the material. Carrier generation at deep-level defect sites requires thermal energy, and

is very sensitive to changes in temperature – the generated current increases exponentially with temperature.

It is important to note that EMCCDs are more susceptible to radiation damage than traditional CCDs (Hadwen, et al., 2004; Smith, et al., 2006; Pool, et al., 2005). This is due to radiation-induced energy states between the valence and conduction bands near the high field regions of the gain register elements. After exposure to 5 krad(Si) in high energy protons, the sum of the induced dark current in the output and gain registers is 0.4 e^-/gain element/pixel for a nominal read rate of 11 MHz at −20 C (Robbins, 2009).

At 160 K, the increase in DCR after one solar cycle (11 years) was 12.5 e-/s/pix. In contrast, an x-ray detection CCD device on board the ASCA satellite (with similar shielding) experienced an increase of 0.8 e-/s/pix/yr, which would extrapolate to an increase of 8.8 e-/s/pix after one solar cycle (a rate of increase of 1.8 e-/s/pix/krad(Si), assuming that the measurement window was representative of the flux distribution for the entire solar cycle). (Yamashita, et al., 1997) While the CCD experiences less radiation damage, the GM-APD is not susceptible to surface-generated dark current. This gives it the potential to surpass the CCD's radiation tolerance with some design improvements. In addition to an increase in dark current, however, the CTE for CCD-based devices decreases significantly after irradiation (Mori, et al., 2013; Janesick, et al., 1991; Noeske, et al., 2012). These radiation effects are common to standard CCDs, analog-mode EMCCDs, and photon counting EMCCDs since they all utilize the same basic structure and operation. It should be noted that increased shielding, such as on the STIS instrument on HST, can significantly decrease the radiation dose per year and the radiation-induced dark current (Kimble, et al., 2000).

7.2 Current Best Candidates

For an overall comparison of each detector type, a plot of SNR vs fluence is useful. Depending on the scenario, the required exposure(s) must be short or long. Transit photometry, adaptive optics, and pulsar imaging all require short exposure times, while direct imaging of exoplanets requires long integration times.

Table 4 shows the characteristic noise values for a 0.1 s exposure for each detector type: CCD, CMOS APS, analog-mode EMCCD, photon counting EMCCD, and GM-APD imaging arrays. Due to the high frame rate required for short exposures, the read noise for the CCD and CMOS-APS detectors is high and the readout time is a significant portion of the total 0.1 s wall time, reducing the duty cycle. The values shown for the photon counting EMCCD and the GM-APD assume that the gate times have been optimized for the fluence range simulated. The probability of one or more photon-generated electrons per gate in each detector is 0.797 at a fluence of 800 photons/pixel/exposure. For the purposes of the following simulations, all detectors are assumed to be sensitive in the visible spectrum.

Table 4 – This table shows state-of-the-art performance characteristics for short exposures of 0.1 s (requiring pixel rates in the tens of MHz range for moderately-sized CCD- and CMOS-based imagers.

Parameter	**CCD**	**CMOS APS**	**Analog-Mode EMCCD**	**Photon Counting EMCCD**	**GM-APD**
Dark Current (e-/s/pix)	0.00021	0.015	0.00021	0.00021	0.027
CIC (e-/pix/frame)	0	0	0.0025	0.0025	0
Read Noise (e- rms)	10	10	<<1	<<1	0
QE	90%	90%	90%	90%	70%
Duty Cycle	90%	90%	90%	100%	100%
ADC Saturation (1000s e-)	72	100	1,000	1,000	N/A

Figure 81 shows the simulated SNR for the detectors using the values shown in Table 4. Common fluxes for the short exposure applications are marked and labeled (see Table 2 for reference).

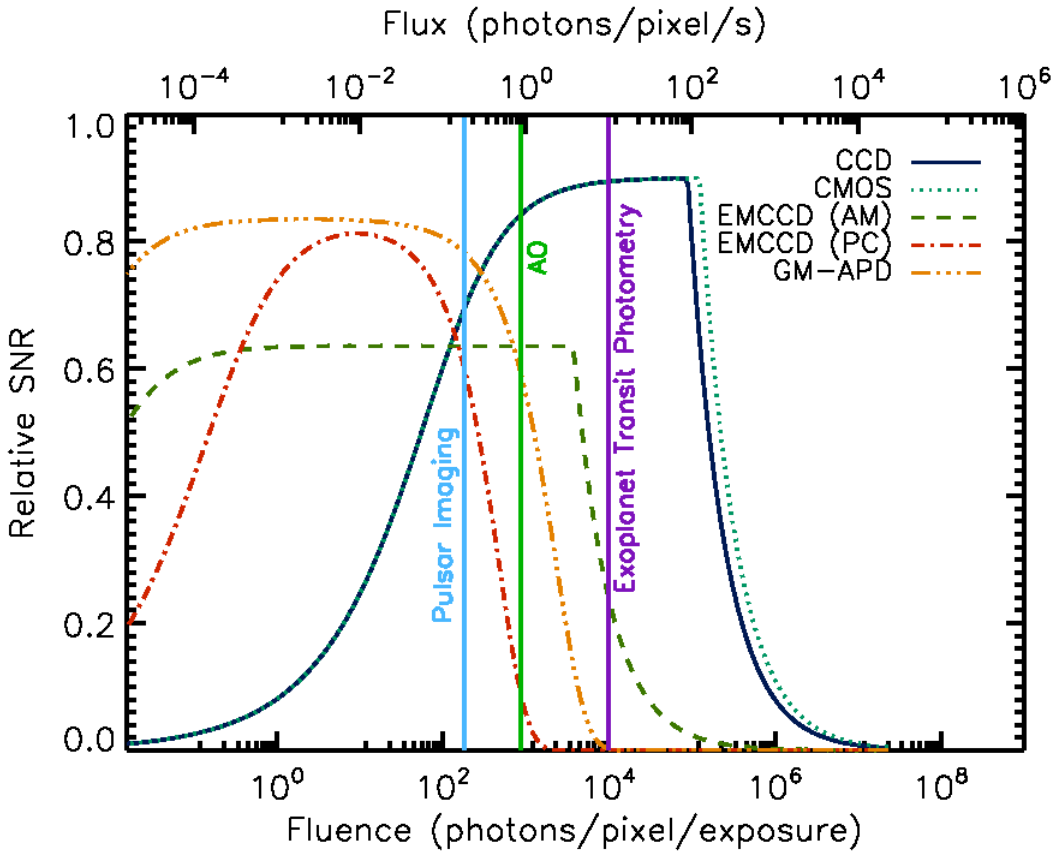

Figure 81 – This plot shows the SNR of various detectors vs fluence. The settings for each detector are optimized for the short exposure scenario of 0.1 s wall time per image. The labeled vertical lines show reasonable fluence levels for the applications noted.

The long exposure scenario is also an interesting case, and requires different settings for the detectors. Table 5 shows the characteristic noise values for a 1000 s exposure for each detector. The main distinction between the values in Table 5 and the values in Table 4 is the read noise. In a long exposure scenario, the readout time can be much slower while still maintaining high duty cycle, which reduces read noise for the CCD and CMOS APS detectors. The GM-APD and EMCCD gate times are set to 1 ms, which has an avalanche probability of 0.797 at 1500 photons/pixel/s.

Table 5 – This table shows state-of-the-art performance characteristics for long exposures of 1000 s.

Parameter	**CCD**	**CMOS APS**	**Analog-Mode EMCCD**	**Photon Counting EMCCD**	**GM-APD**
Dark Current (e-/s/pix)	0.00021	0.015	0.00021	0.00021	0.027
CIC (e-/pix/frame)	0	0	0	0	0
Read Noise (e- rms)	2	3	<<1	<<1	0
QE	90%	90%	90%	90%	70%
Duty Cycle	100%	100%	100%	100%	100%
ADC Saturation (1000s e-)	72	100	1,000	1,000	N/A

Figure 90 shows the simulated SNR for a range of fluence values for each detector, based on the values in Table 5. The expected flux for direct imaging of an exoplanet (see Table 2) is shown.

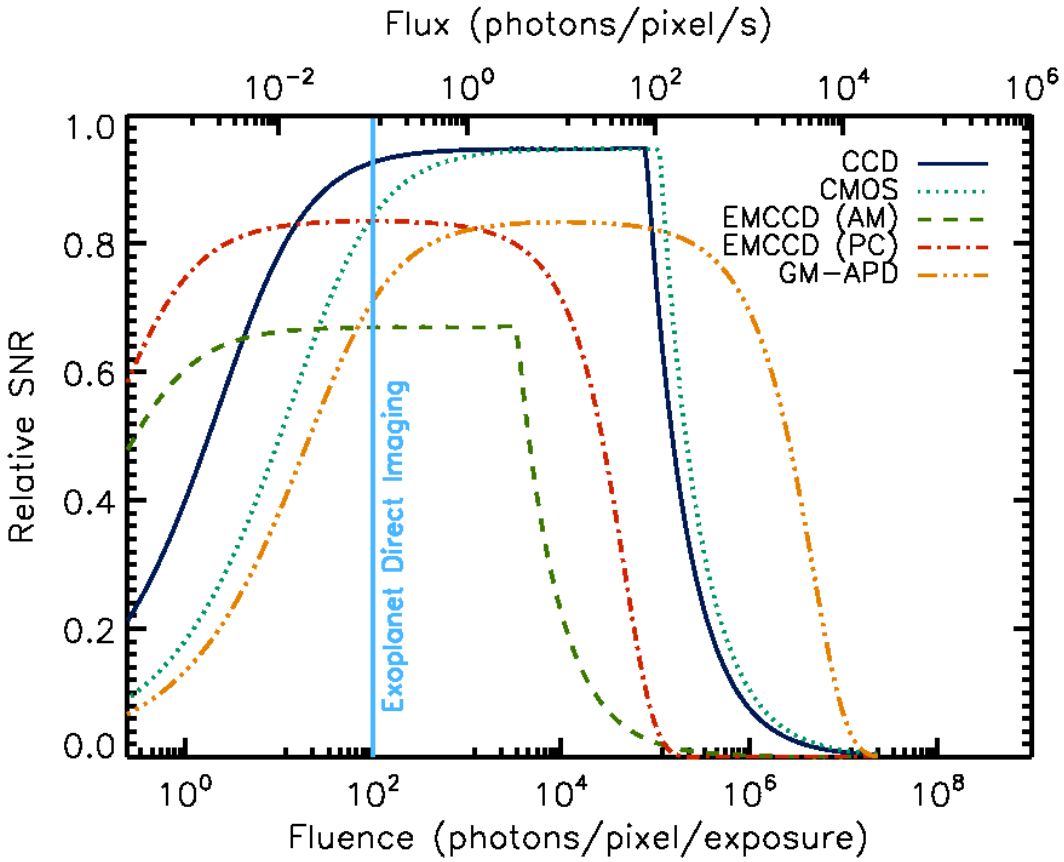

Figure 82 – This plot shows the SNR of various detectors vs fluence. The settings for each detector are optimized for the long exposure scenario of 1000 s wall time per image. The labeled vertical line shows a reasonable fluence level for the applications noted.

The CCD and CMOS detectors have nearly the same performance except for a small change in ADC saturation. Their poor performance at low fluence levels in the fast exposure scenario is due to the high read noise necessary to read out the array quickly. This disadvantage does not apply for long exposures where the read noise decreases significantly. The analog-mode EMCCD is limited to 70% of the shot-noise-limited SNR because of the ENF caused by uncertainty in the gain. It also saturates more quickly, even though its ADC has a higher saturation level than the CCD or CMOS devices, because of the gain. The EMCCD suffers from very low duty cycle in the fast exposure scenario due to the maximum pixel readout rate (tens of MHz). Additionally, the CIC, which introduces as much noise as an equivalent amount of dark current, is high when the pixel readout rate is as high as required here. In the fast exposure scenario, the GM-APD clearly dominates at fluence levels between 1 and 100 photons

(corresponding to 10-1000 photons/s in this simulation). If t_{gate} were shorter (e.g., if the readout electronics were optimized for fast readout), the detector would saturate later and offer significant competition to the CCD and CMOS detectors at higher fluence levels. The absence of read noise for the GM-APD, even though the dark current is higher, makes it the best candidate for fast imaging if the projected performance levels can be met. For long exposures, the higher dark current of the GM-APD results poor performance below a total fluence of ~100 photons (0.1 photons/s), and it is out-performed by the CCD and CMOS devices between 1,000 and 100,000 photons (1-100 photons/s). However, if DCR were to improve to the levels of CCD and CMOS devices, the GM-APD would out-perform the photon counting EMCCD and the analog-mode EMCCD at all fluence levels, and the CCD and CMOS devices below 10 photons (0.01 photons/s).

7.2.1 *ASTRONOMY APPLICATIONS*

The proposed WFIRST-AFTA mission currently includes a coronagraph instrument. The details of this proposed instrument provide a convenient way to compare the performance of two detector types for direct imaging applications. For short exposures, CCD and CMOS APS sensors are ill-suited due to the high read noise associated with fast readout times. For the short exposure scenarios (transit photometry and pulsar imaging), only photon counting EMCCDs and GM-APDs are considered. Long exposure scenarios (direct imaging) require sensitive detectors with very low noise. The best detector candidates are detectors with zero read noise (photon counting), high efficiency, and low dark current. Although analog-mode EMCCDs have low dark current and read noise, the ENF in the multiplication register effectively halves the QE, making them inferior to photon counting EMCCDs in this application. GM-APD imaging arrays are also considered for the long exposure application.

WFIRST-AFTA uses a coronagraph instrument to block light from the star and directly image the orbiting exoplanet. The mission will give the first reflected (visible) light images of the planetary systems of nearby stars (NASA, 2014). The 2.4 m aperture provides a total field of 0.4˚. The coronagraph has an inner working angle of 100 mas, equivalent to 1 AU at 10 pc, and an outer working angle of 750 mas. The instrument is designed to image planets that have a

planet:star contrast of $5x10^{-10}$ to $5x10^{-8}$, focusing on bright stars. The wavelength range of the instrument is 400 nm – 1000 nm (NASA, 2014).

One of the current detector candidates for the WFIRST-AFTA coronagraph instrument is an e2v CCD201-20, a photon counting EMCCD with 1k x 1k pixels. The detector has a dark current of 0.0003 e-/pix/s and CIC of 0.001 e-/pix/frame, and the QE in the V-band (550 nm) is 93%. The read noise when operated at modest frame rates is 8 e^- rms, which is effectively 0.04 e^- rms at a gain of 200 (suggested gain value). The gain is provided by 604 multiplication elements in the readout register (e2v technologies, Nov. 2011). To evaluate the performance of GM-APDs for a mission such as WFIRST-AFTA, the EMCCD described above will be used as a basis of comparison.

One way to evaluate the performance of the EMCCD and GM-APD for exoplanet imaging is to compare the SNR of each device across a range of relevant signal levels. The theoretical SNR equations for the EMCCD in photon counting mode and the GM-APD are actually the same in certain circumstances. The EMCCD must be operated with a detection threshold at least three times that of the read noise (in order to avoid significant counts from read noise) and at least 10 times smaller than the total gain (to avoid lost signal due to high thresholding). The GM-APD must be operated with a hold-off time such that the afterpulsing probability is zero. If all of these assumptions are true, then Eq. 99 gives the theoretical SNR of both an EMCCD in photon counting mode and a GM-APD (see section 2.6.2).

$$SNR = \frac{\eta \cdot \lambda_p \cdot n_{gates}}{\sqrt{\frac{p}{(1-p)} \cdot n_{gates}}} \qquad \text{Eq. 99}$$

where $$p = 1 - e^{-(\eta \cdot \lambda_p + \lambda_d)}$$

η is the efficiency of the detector (QE for EMCCDs and PDE for GM-APDs), λ_p is the number of incident photons per gate, λ_d is the number of dark current carriers per gate (for an EMCCD,

this includes CIC), n_{gates} is the number of gates in the exposure, and p is the avalanche probability.

Table 2 gives the relevant performance parameters for both detectors. The values for the EMCCD are taken from the manufacturer (e2v technologies, Nov. 2011) and from projected use in the WFIRST-AFTA mission (NASA, 2014). The GM-APD parameter values are based on the state-of-the-art performance metrics detailed in Section 3 and the same use as the EMCCD.

Table 6 – Performance parameters for two devices are shown: an e2v CCD201-20 EMCCD in photon-counting mode and a GM-APD array.

Parameter (units)	**EMCCD (PC)**	**GM-APD**
QE / PDE (%)	93	70
Dark Current (e^-/pix/s)	0.0003	0.03
CIC (e^-/pix/frame)	0.001	0
Duty Cycle (%)	100	99.98
Gate Time (ms)	55	55
Read Noise (e^- rms/frame)	0.04 (effective)	0

Figure 83 shows the SNR of the EMCCD and the GM-APD over a range of signal levels. The simulated exposure time is 10 hours in the V-band (550 nm), and the signal level is per pixel, not per object. Each exposure is made up of 55 ms, and the GM-APD has a hold-off time of 10 μs.

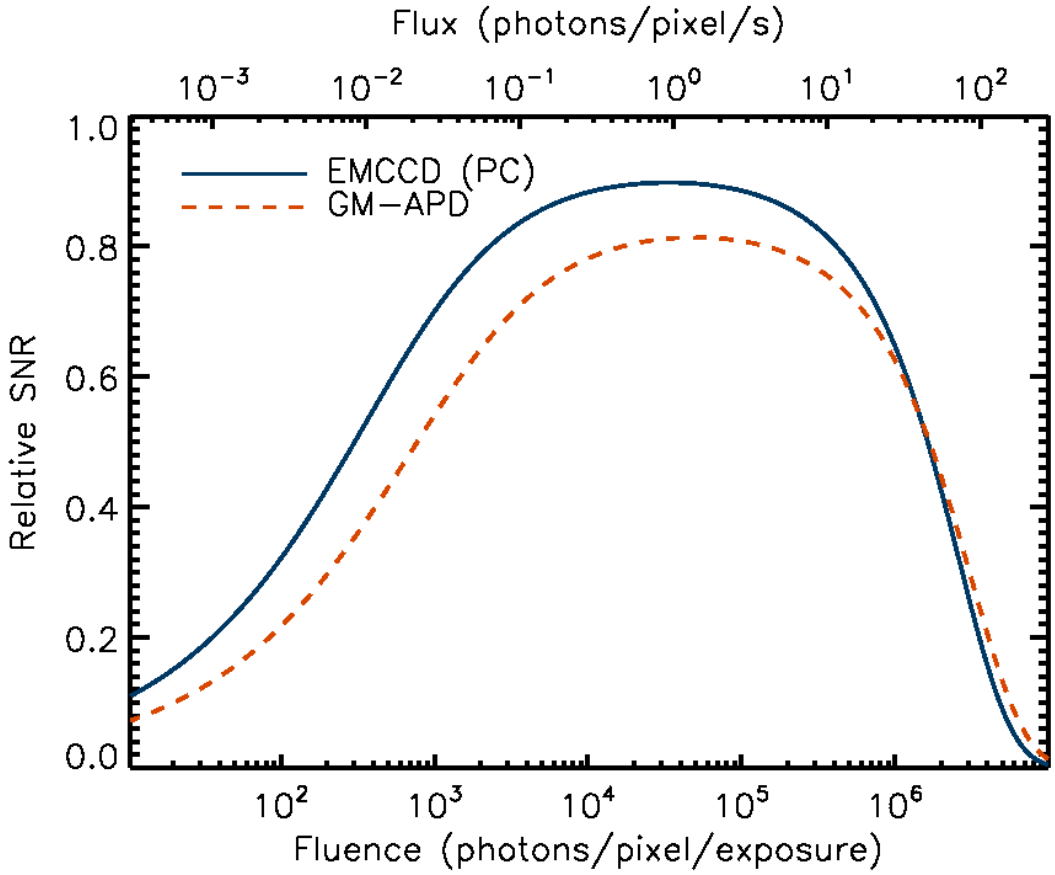

Figure 83 – This plot shows the simulated results for an e2v CCD201-20 operated in photon-counting mode and a GM-APD array. Relative SNR is the SNR normalized to the shot noise limit. Both detectors have the same gate time and exposure settings, but unique efficiency and noise values. The total exposure time is 10 hours.

The EMCCD has a wider range of high SNR than the GM-APD does due to its lower dark noise, and it also has a higher peak due to its higher efficiency. To illustrate the implications of the difference in SNR, Figure 84 shows the wall time required to reach an SNR of 10 for both detectors given the settings described above.

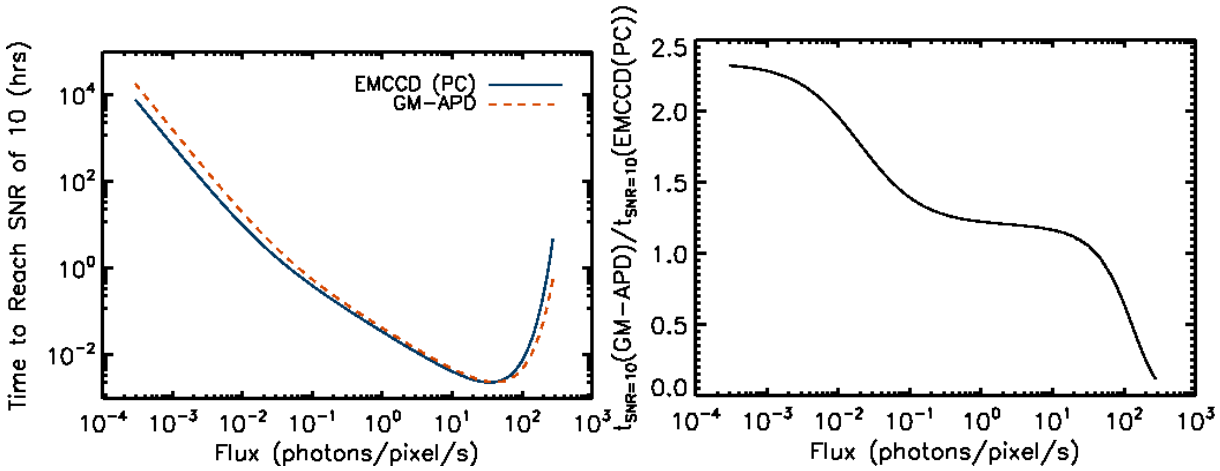

Figure 84 – This plot shows the wall time required to reach SNR= 10 (top) and the ratio of time required for the two detectors (bottom).

While the results in Figure 83 and Figure 84 are useful for comparison, they do not consider the imaging or spectroscopy cases, where the light from the object is spread across multiple pixels. In an imaging scenario, assuming that the focal spot size is diffraction-limited, Eq. 100 gives the angular width (in radians) of the central lobe of the airy disk pattern (between the first dark minima).

$$\theta = 2.44\frac{\lambda}{D} \qquad \text{Eq. 100}$$

λ is the wavelength of the light and D is the diameter of the aperture. The WFIRST-AFTA coronagraph aperture is 2.4 m and the plate scale is 17 mas/pix (NASA, 2014), so the central lobe of the diffraction pattern at 550 nm is 115 mas or 6.78 pixels. Alternatively, the full width at half maximum (FWHM) can be calculated using Eq. 101.

$$\theta = 1.03\frac{\lambda}{D} \qquad \text{Eq. 101}$$

The FWHM is 48.7 mas, or 2.86 pixels, at 550 nm. Figure 85 shows a 1D cross-section of the diffraction-limited, airy disk point spread function (PSF). The solid line shows the theoretical

function and the dashed line shows the function as sampled by the pixels in the simulated sub-array.

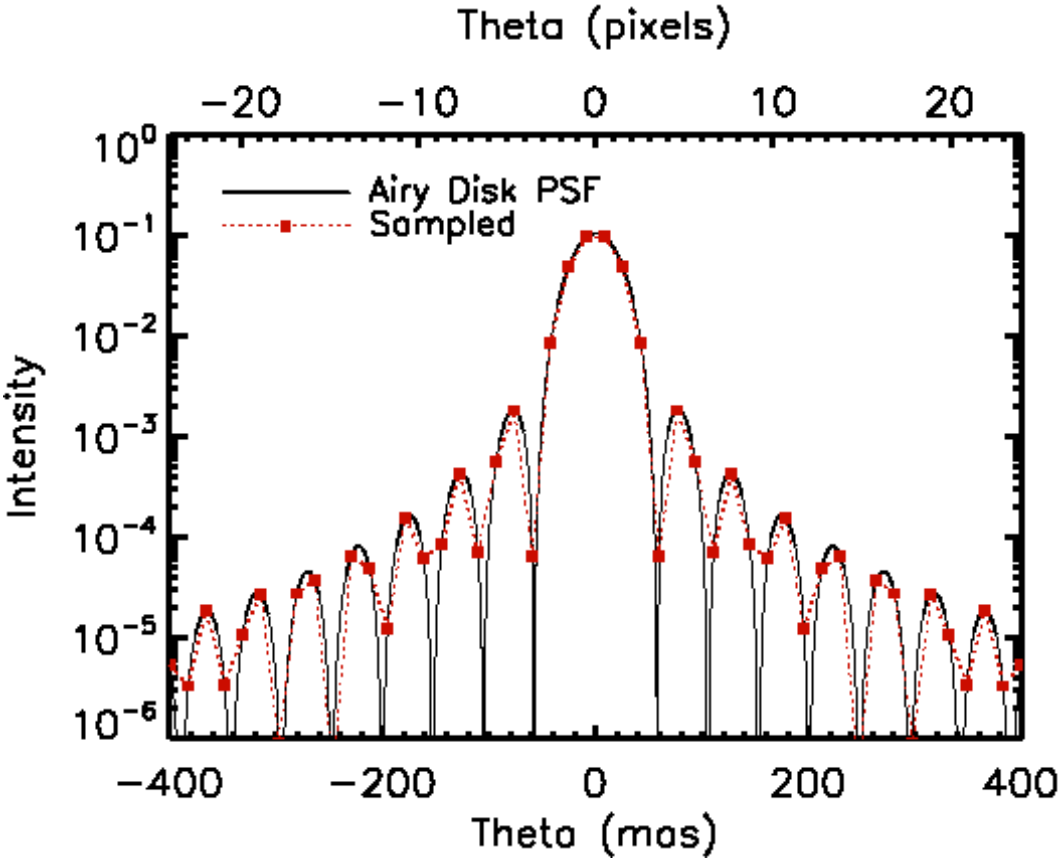

Figure 85 – This plot shows a 1D cut of the simulated PSF for the WFIRST coronagraph instrument. The solid line is the theoretical PSF and the dashed line is the PSF as sampled by the detectors (17 mas/pix).

Figure 86 shows simulated images of exoplanets with various magnitudes for both the EMCCD and the GM-APD. The simulated images assume that the planet is not in a debris field and that any remaining diffracted light from the star is not significant. Zodiacal light is assumed to be negligible as well.

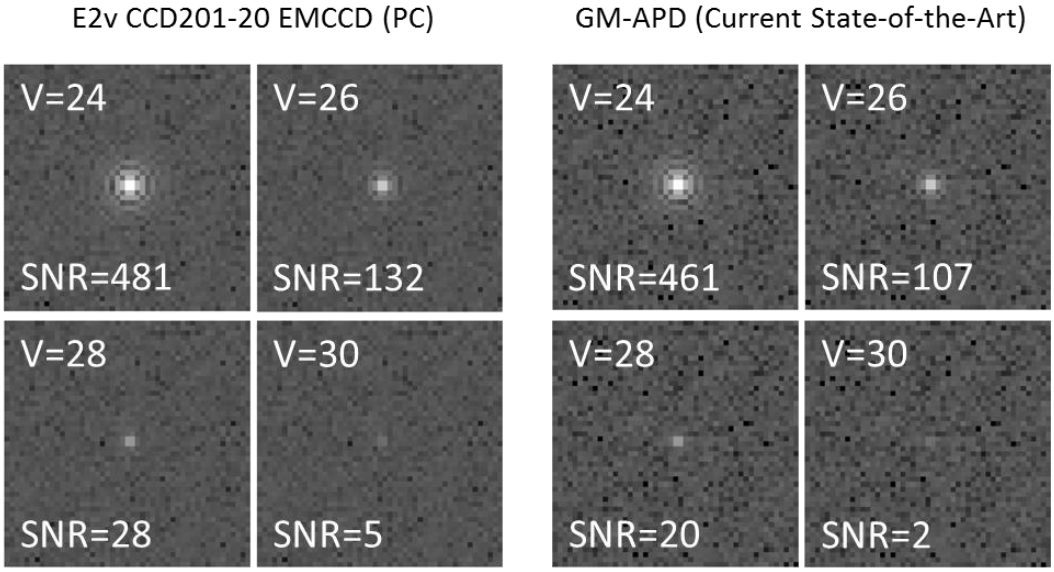

Figure 86 – This plot shows simulated images of exoplanets of varying magnitudes for an EMCCD (left) in photon-counting mode and a current state-of-the-art GM-APD (right). The corresponding apparent magnitude in the V band for each simulation is noted in the top left corner of each image. SNR is calculated for the signal that falls inside of the first dark ring, which includes 37 pixels. The value range is constant for each image.

The background noise is notably greater for the GM-APD, which has a total dark noise contribution of 0.0017 e^-/gate compared to the EMCCD's combined dark current and CIC of 0.0011 e^-/gate. While both detectors resolve the first few maxima of the signal for the brightest planet (V = 24), the EMCCD does notably better for the faintest planet (V = 30) due to a combination of lower dark noise and higher efficiency.

7.2.2 *Adaptive Optics*

As discussed in section 2.1.2, AO performance will be evaluated based on a Monte Carlo simulation that calculates the SNR of estimate of the centroid location. Because of the nature of AO, only short exposure conditions are considered. The CCD, CMOS APS, and analog-mode EMCCD are excluded based on read noise levels in the first two, and the ENF in the latter. Only the GM-APD and the photon counting EMCCD are simulated.

Figure 87 shows the results for the GM-APD described in Table 4. The SNR increases with increasing signal strength (lower magnitude values) and distance from the center of the 4-pixel quad cell. Higher contrast in pixels signals leads to higher SNR. With an SNR threshold of 3, the

brightest signal (R=15) is only distinguishable once the centroid is at least 0.3 pixels from the center. At R=16.25, it is only distinguishable at 0.6 pixels from the center, and the faintest two focal spots do not reach the threshold at all.

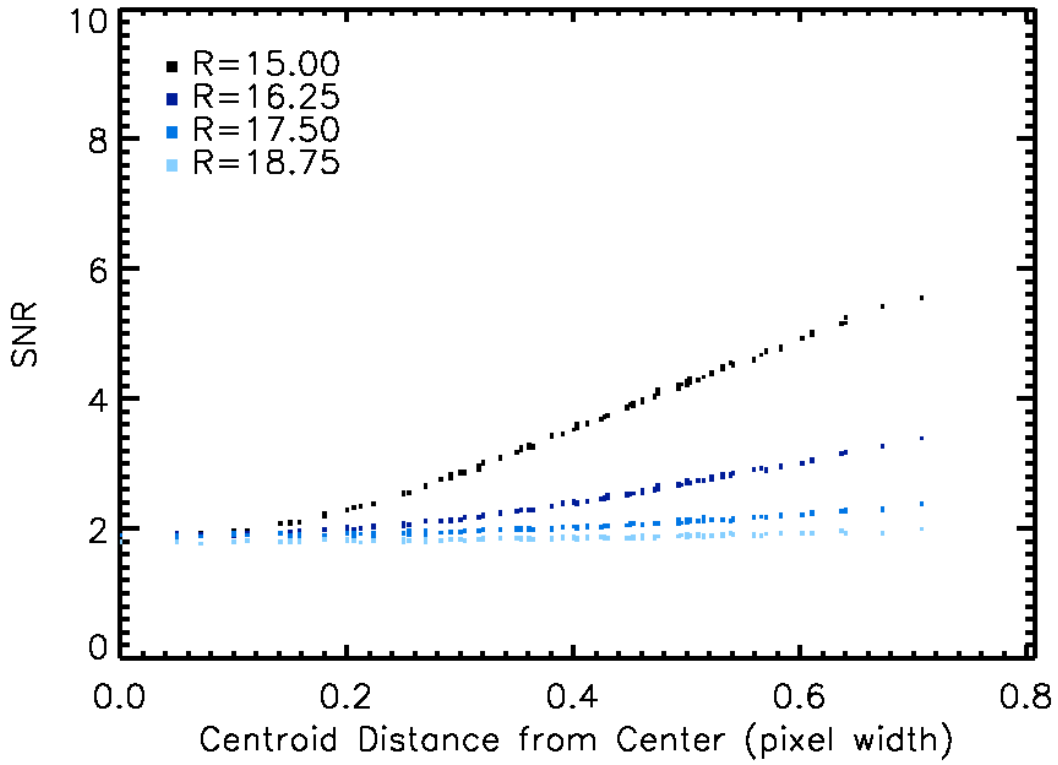

Figure 87 – This plot shows simulated SNR of the centroid location estimate as a function of distance and focal spot magnitude for a current state-of-the-art GM-APD. Exposure time for each measurement is 0.1 s.

Figure 88 shows the photon counting EMCCD performance (see Table 4) under the same conditions as the GM-APD. For the brightest focal spot, it does not reach the SNR threshold of 3 until the centroid is 0.38 pixels from the center of the quad cell, and achieves lower SNR across all values of magnitude and distance.

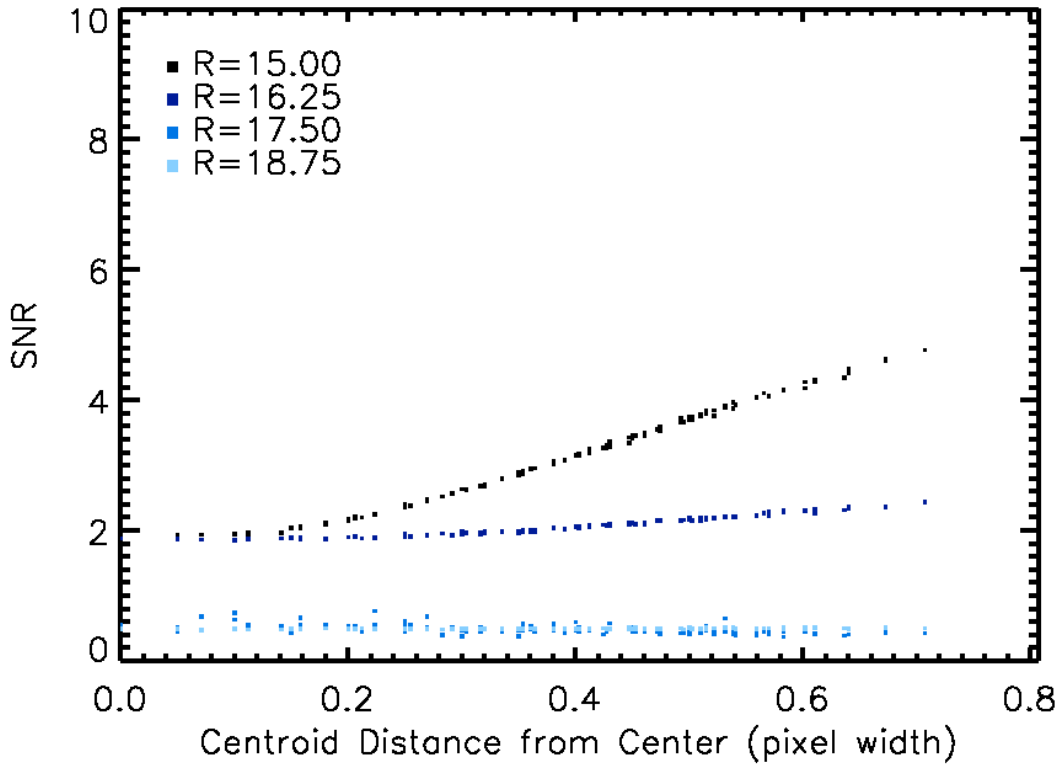

Figure 88 – This plot shows simulated SNR of the centroid location estimate as a function of distance and focal spot magnitude for a current state-of-the-art EMCCD in photon-counting mode. Exposure time for each measurement is 0.1 s.

7.3 Long-Term Best Candidates

The most significant change to detector performance in the long term will be the noise reduction achieved in GM-APD detectors. CCD and CMOS APS detectors for UV/Optical applications have little room for improvement. Dark current is nearly negligible, even for faint objects, and read noise is fundamentally limited by sampling and readout speed. There is some improvement to be made in IR detectors and applications, but the dark current noise will always be a function of material band gap, regardless of detector implementation.

Analog-mode EMCCDs are similarly limited in a fundamental way by the ENF in the multiplication register. There is no way around the excess noise introduced by the way in which gain is applied. And while photon counting EMCCDs use the gain in a way that makes the excess noise insignificant to the detection process, the main constraint is currently CIC. CIC has

been a problem for a number of years now, and it will only get worse as array sizes increase, especially for short exposure scenarios. Unlike dark current, which scales with exposure time, CIC is noise applied to each read, effectively increasing the noise per second as the exposure time decreases. Without significant strides in CIC mitigation, this will be the limitation of EMCCDs. Although CIC may only be on the order of 0.001 e^-/pixel/read, in gated operation this is very large. Consider moderately-long gates of 200 µs, which begin to saturate at ~800 photons/pix/s. CIC of 0.001 e^-/pixel/read translates to an equivalent noise source of 5 e^-/pix/s. Even for long gates of 1 ms, the CIC is still equivalent to 1 e^-/pix/s.

Since the GM-APD has no read noise and no CIC, small improvements in the DCR and afterpulsing probability offer significant improvements in overall SNR.

7.3.1 *ASTRONOMY APPLICATIONS*

The first-generation GM-APD array-based detectors demonstrated zero read noise and modest afterpulsing at temperatures above 140 K. Radiation-induced dark current was 1.4x higher than that of a CCD with the same radiation dose and shielding. However, the devices had high dark current and low efficiency. The dark current contribution was 38 e^-/s/pix, and the PDE was only 0.3% at its peak. However, the causes of these shortcomings are known, and a second generation of devices is currently being tested that have addressed the problems (Figer & Kolb, 2014).

The state-of-the-art GM-APD performance is roughly equivalent to the EMCCD for fluxes greater than 1 photon/s (see Figure 84) when considering the time required to reach an SNR of 10. While GM-APD performance lags behind EMCCDs for lower signal levels, targeted research to reduce the dark current would significantly improve SNR for faint objects. Given the limitation of CIC in EMCCDs, GM-APDs with dark current comparable to state-of-the-art CCD levels would offer an advantage for low-light-level imaging and spectroscopy. Figure 89 shows the theoretical SNR of a GM-APD with various dark current values along with the SNR of a state-of-the-art EMCCD. The lowest simulated value is equal to that of an EMCCD (0.0003 e^-/s/pix). Because the EMCCD has CIC noise due to high pixel readout rates, the GM-APD performance at the lower light levels exceeds that of the EMCCD even with only a modest decrease in the dark current to 0.01 e^-/s/pix.

The improvement in SNR shown in Figure 89 is also evident when comparing the time required to reach an SNR of 10 for the GM-APD device and the e2v EMCCD (see Figure 90).

A shot-noise-limited detector would reach an SNR of 10 in 10 hours for a fluence of 100 photons (0.003 photons/s). The e2v EMCCD in photon-counting mode would require 94.7 hours and a GM-APD with the same dark current would require only 16.3 hours. With current state-of-the-art dark current, the GM-APD would require 207.8 hours (2.2x the exposure time required for the EMCCD), though a GM-APD with a 3x decrease in dark current to 0.01 e^-/s/pix would require 85.8 hours – less than the EMCCD.

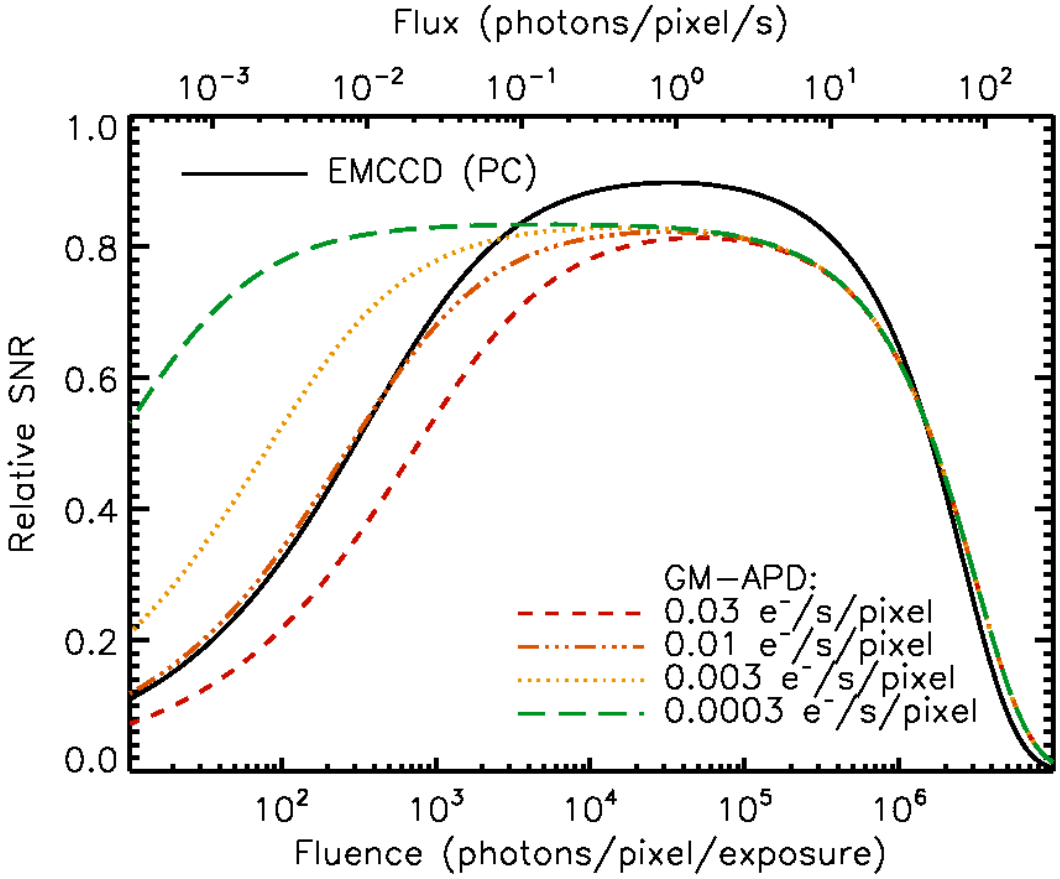

Figure 89 – This plot shows the relative SNR of an e2v EMCCD in photon-counting mode and a theoretical GM-APD device with dark current equal to that of the EMCCD. Both detectors have the same gate time and exposure time settings. The total exposure time is 10 hours. The 0.03 e^-/s/pixel curve is the current state-of-the-art.

A shot-noise-limited detector would reach an SNR of 10 in 10 hours for a fluence of 100 photons (0.003 photons/s). The e2v EMCCD in photon-counting mode would require 94.7 hours and a GM-APD with the same dark current would require only 16.3 hours. With current state-of-the-art dark current, the GM-APD would require 207.8 hours (2.2x the exposure time required for the EMCCD), though a GM-APD with a 3x decrease in dark current to 0.01 e^-/s/pix would require 85.8 hours – less than the EMCCD.

Improvements in PDE could bring GM-APD efficiency closer to that of EMCCDs, though avalanche initiation probability limits the total efficiency.

The GM-APD devices also have similar radiation tolerance to existing CCD devices. While the ASCA satellite CCD experienced less radiation damage when compared to an equivalent simulated environment for the GM-APD, the latter are not susceptible to surface-generated dark current. This gives them the potential to surpass the CCD's radiation tolerance with targeted design improvements. Increased shielding, such as on the STIS instrument on HST, can also significantly decrease the radiation dose per year and the radiation-induced dark current (Kimble, et al., 2000). Another effect of radiation damage on CCD-based devices (including EMCCDs) is a decrease in the CTE (Mori, et al., 2013; Janesick, et al., 1991). GM-APDs are not affected by CTE, and so are immune to CTE-related decreases in performance.

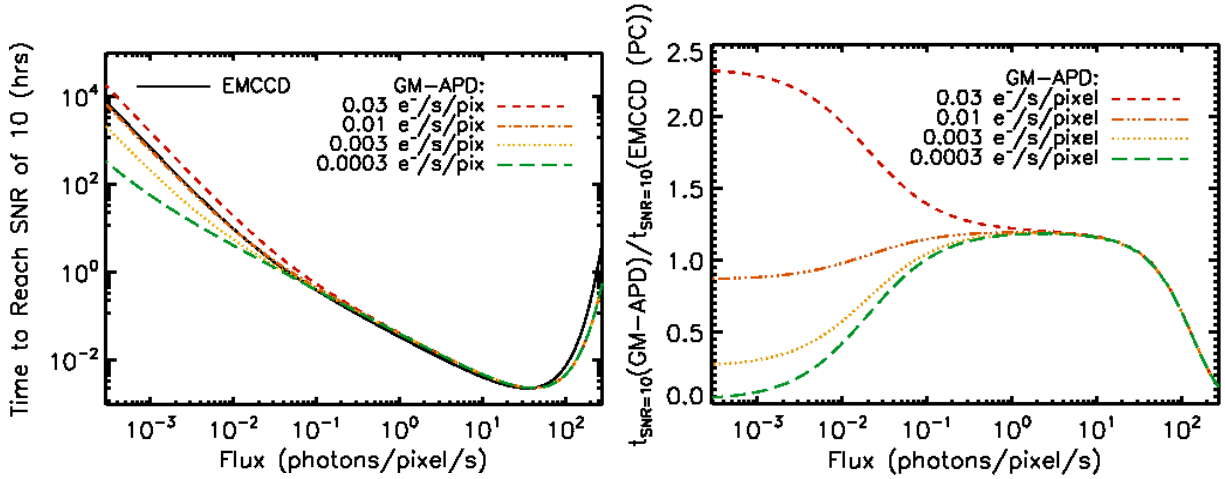

Figure 90 – This plot shows the wall time required to reach SNR= 10 (top) and the ratio of time required (bottom) for the e2v EMCCD and the theoretical GM-APD with dark current equal to that of the EMCCD. The 0.03 e⁻/s/pixel curve is the current state-of-the-art.

State-of-the-art performances for GM-APDs lag significantly behind that of EMCCDs for low fluxes (< 1 photon/s), but are roughly equivalent at higher fluxes, as shown in Figure 83 and Figure 84. The only difference in performance at fluxes greater than 1 photon/s is the lower efficiency associated with GM-APDs due to non-ideal avalanche initiation probability. However, improvements in dark current could increase the performance of the GM-APD at low signal levels. If the dark current were improved by an order of magnitude, the performance of the GM-APD would be higher for fluxes less than 0.1 photons/s, as shown in Figure 90. EMCCDs are fundamentally limited by CIC noise, which only increases as the number of pixels increases for large-format arrays due to high pixel readout rates. Theoretically, GM-APD devices are less limited by fundamental noise sources for extremely low signal levels and may ultimately be the better solution with more advanced research.

In order to grasp how much better the GM-APD would be at faint fluxes with dark current equal to that of EMCCDs, the exoplanet imaging scenario is revisited. Assuming the same exposure settings and the same objects as Figure 86, Figure 91 shows simulated images with the current state-of-the-art EMCCD and a GM-APD with the same dark current (0.0003 e⁻/s/pix).

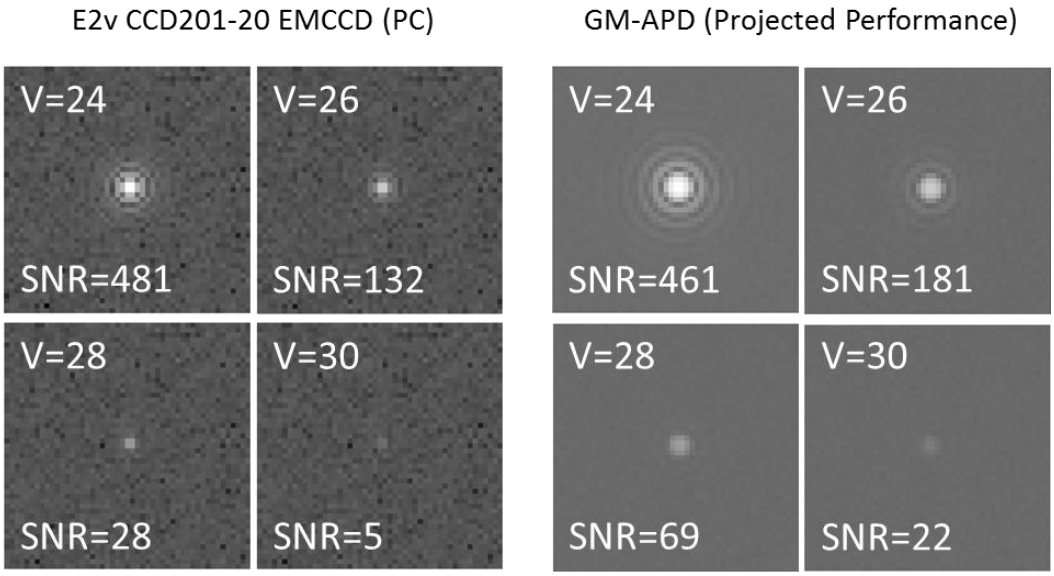

Figure 91 – This plot shows simulated images of exoplanets of varying magnitudes for an EMCCD (left) in photon-counting mode and a GM-APD with future projected performance (right). The corresponding apparent magnitude in the V band for each simulation is noted in the top left corner of each image. SNR is calculated for the signal that falls inside of the first dark ring, which includes 37 pixels. The value range is constant for each image.

In this scenario, the GM-APD out-performs the EMCCD for every object except the brightest, which is dominated by shot noise and determined mainly by efficiency. The faintest object (V=30), while not resolvable with an SNR cutoff of 10 on the EMCCD, is resolved on the GM-APD.

7.3.2 *Adaptive Optics*

As with the astronomy imaging applications, the GM-APD will surpass the photon counting EMCCD with modest decreases in DCR. While the current state-of-the-art GM-APD out-performs the photon counting EMCCD by a small margin, it may not be significant enough to sway users to change their current systems. With a 10x decrease in the DCR to 0.003 e^-/s/pixel, the GM-APD outpaces the EMCCD, though the improvement in centroid estimation is limited. Most of the noise is due to shot noise and the estimation function, and so a decrease in dark current does not affect the SNR very much. Figure 92 shows the simulated results for measurements taken with an improved GM-APD. The results are nearly indistinguishable from the current state-of-the-art case.

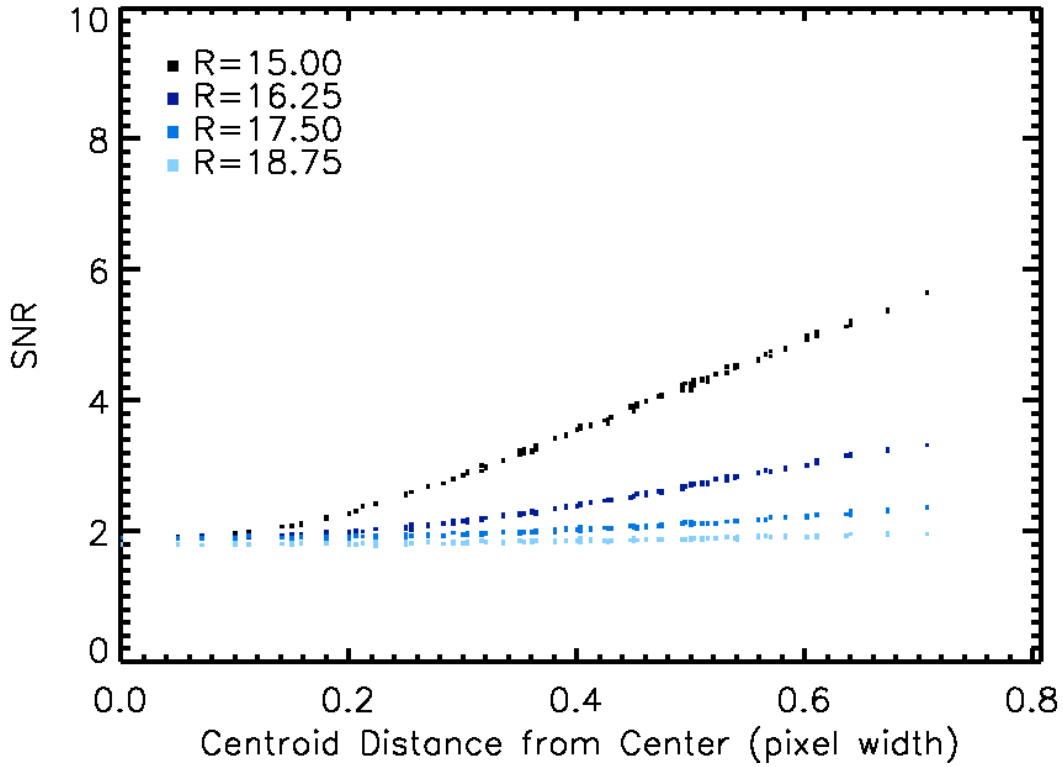

Figure 92 – This plot shows simulated SNR of the centroid location estimate as a function of distance and focal spot magnitude for a future state-of-the-art GM-APD. Exposure time for each measurement is 0.1 s.

7.4 Future Work

As of March 2015, a second-generation GM-APD device is being delivered to the CfD for testing. The device has been fabricated and joined to a ROIC via the Ziptronix[7] process, which aligns and bonds dielectric layers together, matching the metal vias to form vertical connections. This reduces parasitic resistance and capacitance associated with bump bonds and back-filling, and the detector was fabricated using an SOI process, which should result in a more pristine back surface. Fewer defects on the backside of the detector should result in DCR more similar to the un-thinned detector shown in Figure 79.

The new devices are fabricated using the HFF design, which should result in much higher PDE. In order to counteract the high crosstalk probability present in the first HFF iteration, a scupper

[7] http://www.ziptronix.com/

circuit was introduced to the device design. A small circular implant is located at the intersection of each row and column and biased just below the anode supply voltage when turned on. The scupper implant forms a small diode with the backside implant and helps to collect carriers that are generated between pixels by the avalanche emission. This comes with a reduction in fill factor and therefore efficiency, and so the scupper voltage setting must be balanced to achieve the most crosstalk suppression without compromising PDE.

The devices will undergo the same testing described in section 3, using the same electronics and dewar system. Radiation testing is not planned at this time, though it may be included in future proposals for space qualification and TRL advancement. The new devices should be more radiation hard than the first generation due to the SOI fabrication.

Appendix A: List of Acronyms and Definitions

Acronym	Definition
ADC	Analog-to-Digital Converter
AO	Adaptive Optics
CCD	Charge-Coupled Device
CDM	Code-Division Multiplexing
CFB	Cold Fanout Board
CfD	Center for Detectors
CIC	Clock-Induced Charge
CMOS APS	Compensated Metal-Oxide-Semiconductor Active Pixel Sensor
CT-scan	Computational Tomography scan
CTE	Charge Transfer Efficiency
CTI	Charge Transfer Inefficiency
DCR	Dark Count Rate
DDD	Displacement Damage Dose
DN	Digital Number
DRPE	Double-Random Phase Encoding
EBCCD	Electron Bombarded CCD
EBCMOS	Electron Bombarded CMOS detector
EMCCD	Electron-Multiplying CCD
ENF	Excess Noise Factor
ESO	European Southern Observatory
FDM	Frequency-Division Multiplexing
FET	Field Effect Transistor
FWHM	Full-Width at Half-Maximum
GM-APD	Geiger-Mode Avalanche PhotoDiode

HFF	High-Fill-Factor
HST	Hubble Space Telescope
ICCD	Intensified CCD
IPS	IntraPixel Sensitivity
IR	InfraRed
L2 orbit	2nd Lagrangian orbit point in the Earth-Sun system
LEO	Low Earth Orbit
LFF	Low-Fill-Factor
LIDAR	LIght Detection And Ranging
LM-APD	Linear-Mode Avalanche PhotoDiode
LPE	Liquid Phase Epitaxy
MCP	MicroChannel Plate
MFF	Medium-Fill-Factor
MGH	Massachusetts General Hospital
MIT LL	Massachusetts Institute of Technology Lincoln Laboratory
MKID	Microwave Kinetic Inductance Detector
MOS capacitor	Metal-Oxide-Semiconductor capacitor
MOVPE	Metal-Organic Vapor Phase Epitaxy
MRI	Magnetic Resonance Imaging
MRS APD	Metal-Resistor-Semiconductor Avalanche PhotoDiode
NASA JPL	National Aeronautics and Space Administration Jet Propulsion Laboratory
NIEL	Non-Ionizing Energy Loss
NIR	Near InfraRed
PCROIC	Photon Counting ROIC
PCSCT	Photon Counting Spectral Computational Tomography
PDE	Photon Detection Efficiency

PMT	PhotoMultiplier Tube
PSD	Position Sensitive Detector
PSF	Point Spread Function
QE	Quantum Efficiency
ROIC	ReadOut Integrated Circuit
SNR	Signal to Noise Ratio
SOI	Silicon-On-Insulator
SQUID	Superconducting QUantum Interference Device
SSPM	Solid State PhotoMultiplier
TCSPC	Time-Correlated Single Photon Counting
TDM	Time-Division Multiplexing
TES	Transition Edge Sensor
TID	Total Ionizing Dose
TKID	Thermal Kinetic Inductance Detector
TLS noise	Two-Level System noise
UV	UltraViolet
VLPC	Visible Light Photon Counter
WDB	Warm Daughter Board
WEB	Warm Electronics Board
WFC3	Wide Field Camera 3 (HST)

Appendix B: Superconductor-Based Detector Applications

The energy resolution available in superconductor-based detectors lends itself to many fields, particularly spectroscopy in astronomical applications (see section 1.2.1) and in the medical field. Biological imaging of tumors (Weissleder, et al., 1999) and the brain (Hillman, 2007) has become a popular alternative to radiation-intensive imaging like CT-scans (computational tomography scans) and x-ray images. Magnetic resonance images (MRIs) introduce less radiation to the patient, but are computationally intensive and the equipment required is large and stationary. Near infrared (NIR) tissue imaging can image human tissue up to a few centimeters thick, which is sufficient to test for surface tissue cancers affecting the breast or skin. Even brain imaging, using some interferometry, shows promising results, especially as a non-invasive and low-radiation alternative children and the elderly, or people who are otherwise at risk of increased harm from standard tests.

The medical community also began to see the potential in newer photon counting detectors. CT-scans use x-rays to distinguish various densities in the human body to build a 3D model (via a series of 2D slices) of internal structures. Traditionally, CT detectors consist of silicon devices optically coupled to a scintillator (a material that emits photons of lower energy when it absorbs a photon of high energy). The silicon devices simply integrate the signal over the length of the exposure, which results in the integration of the x-ray energy and no information on the number or energies of the original x-rays. There are currently two limitations to the technology: patient radiation doses and object contrast. Implementing a photon counting spectral CT (PCSCT) system increases the contrast to noise ratio (the figure of merit for CT images) while using the same or a fewer number of photons. Higher sensitivity and faster reset times allow lower patient doses. And although density-only resolution may be sufficient in certain applications like a spectral mammography (Ding & Molloi, 2012), increased energy resolution increases contrast when density alone is not enough (Barber, et al., 2009). The addition of energy resolution would help doctors distinguish between similar (but crucially different) substances, such as the difference between arterial plaque and a stent in a coronary artery. Because x-rays of different energies produce a distinct number of electrons in a semiconductor, photon counting detectors

with fast resets can be used to measure both the number (by resolving pulses) and energy (by resolving pulse height) of the photons collected (Feuerlein, et al., 2008).

For semiconductor devices, it is important to note that this type of application would require a free-running detection mode, where each incident photon generates a pulse (as opposed to gated operation that bins photons inside of a gate). Generally, free-running mode requires control electronics with very fast quench and rest times that are triggered by an avalanche, not an external clock. However, given the relatively small array sizes currently used (on the order of 2 x 256), superconductors would also be an option for this application due to their built-in energy resolution. Recent advances have decreased the superfluous x-ray photons in a CT exposure by inserting a K-edge energy filter between the patient and the x-ray source, making each exposure more efficient and less noisy (Shikhaliev, 2012).

A free-running system would also be necessary for time-correlated single photon counting (TCSPC) of fluorescence lifetime imaging. In this imaging scenario, the time of each photon arrival is measured for the duration of the fluorescence event, which is generally on the order of nanoseconds. The arrival times follow Poisson statistics, where the number of photons at any time decreases exponentially from the start of the experiment. At the end of the experiment, the total number of photons can be calculated from the arrival time distribution and used to make an intensity image, while the arrival time distribution itself is used to calculate the $^{1}/_{e}$ lifetime of the fluorescence decay. TCSPC of fluorescence lifetime imaging is used to characterize fluorescent materials, whether they occur naturally or are used for medical imaging as fluorescent markers. Figure 93 shows an example of the results from a TCSPC fluorescence lifetime imaging experiment (Chessel, et al., 2013).

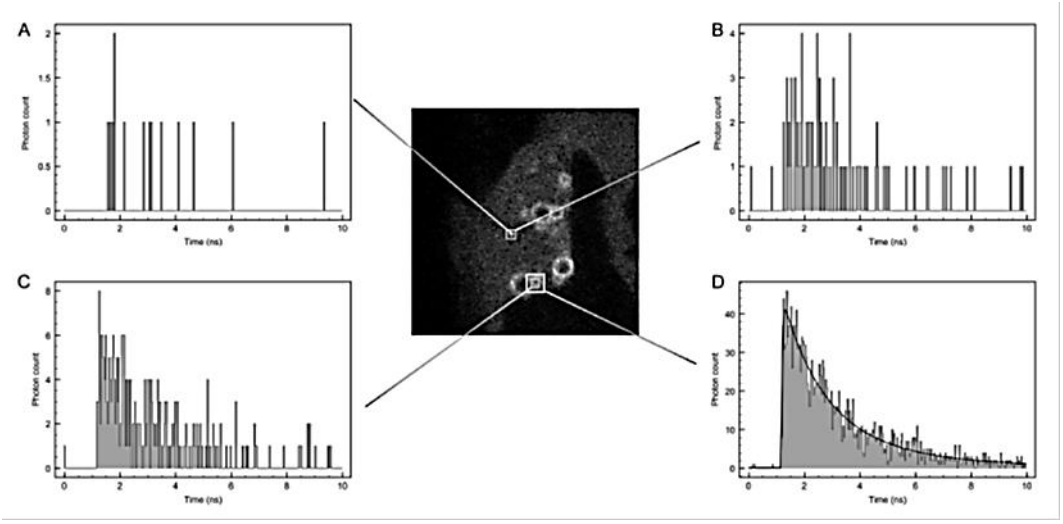

Figure 93 – The results of a TCSPC fluorescence lifetime imaging experiment are shown. The central image is an intensity map of the fluorescence, while the outer figures show the photon arrival time distribution of various regions (Chessel, et al., 2013).

Appendix C: TIME-BASED GM-APD APPLICATIONS

It is important to note that these applications are not compatible with the GM-APD imaging arrays presented in this dissertation. Due to the myriad users of GM-APDs that use them for time-based measurements, this appendix is included for reference. Brief discussions of the applications are provided.

Time-based applications require circuits that record the time of photon arrival. These detectors can be gated or free-running, but they always record the arrival time. Optical communications setups require precise timing to distinguish between digital bursts of photons (representing 1s in a digital word) and the spaces between them, representing 0s. Another time-based application is laser ranging, which syncs the detection cycle of the detector to a pulsed laser.

Optical Communication

Communication at telecom wavelengths for deep space and near-Earth applications is essential for extra-terrestrial exploration. Difficulties with the technology include the challenge of sending a signal over long distances in space (loss of information and flux from diffraction) as well as the tradeoff between lower energy signals (long wavelengths) and the detection efficiency of those signals. Lower energy photons (IR) travel well through space, but they are generally more difficult to detect and the materials required to do so are more expensive. These obstacles are compounded by the fact that detectors are often the limiting factor in a communication system's performance, making low noise and high efficiency desirable. Wavelength manipulation via waveguides offers a solution by allowing for the use of low-energy signal photons desired for space-based telecom with silicon-based detectors. An example of such a system is depicted in Figure 94 (Grein, et al., 2010).

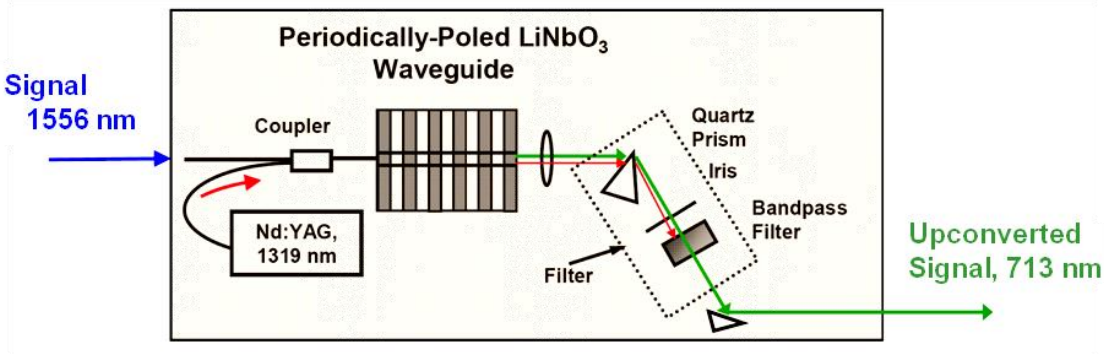

Figure 94 – A wavelength up-converter based on periodically-poled lithium-niobate (PPLN) is illustrated (Grein, et al., 2010).

The waveguide (with Periodically-Poled $LiNbO_3$ – PPLN), which is only 48 mm long and temperature-stabilized, facilitates the conversion of 1556 nm photons to 713 nm photons using a 1319 nm pump (the original 1319 nm pump signal is filtered out via a bandpass filter before the signal reaches the detector). A silicon GM APD detects the up-converted signal at 713 nm, where silicon has high detection efficiency. The system has the potential to reach greater than 90 % efficiency, and the noise produced from the pump (evidenced in an increase in dark counts) was not the limiting noise factor of the system. Detector parameters to consider in such a system include the DCR, PDE, and timing jitter (which in this application constrain the data rate) (Grein, et al., 2010).

Another use for photon counting in communications is encryption. Specifically, double-random-phase encoding (DRPE) for image verification and retrieval is a popular application. By controlling and reading the phase value, it can be used as an encryption tool in addition to the arrival time spacing (essentially making the encrypted data have a complex value). State of the art simulations and theory have shown that photon-limited encrypted distributions have sufficient information for decryption, authentication, and signal retrieval. According to the DRPE algorithm, an image can be turned into a noisy, complex-valued distribution that does not reveal its content when two random phase filters are used in the spatial and Fourier domains. On the user end, the image can be decrypted using the decryption key provided (per the random phase masks) and taking the inverse Fourier transform. Figure 95 shows two methods for using photon counting with encryption (Perez-Cabre, et al., 2012).

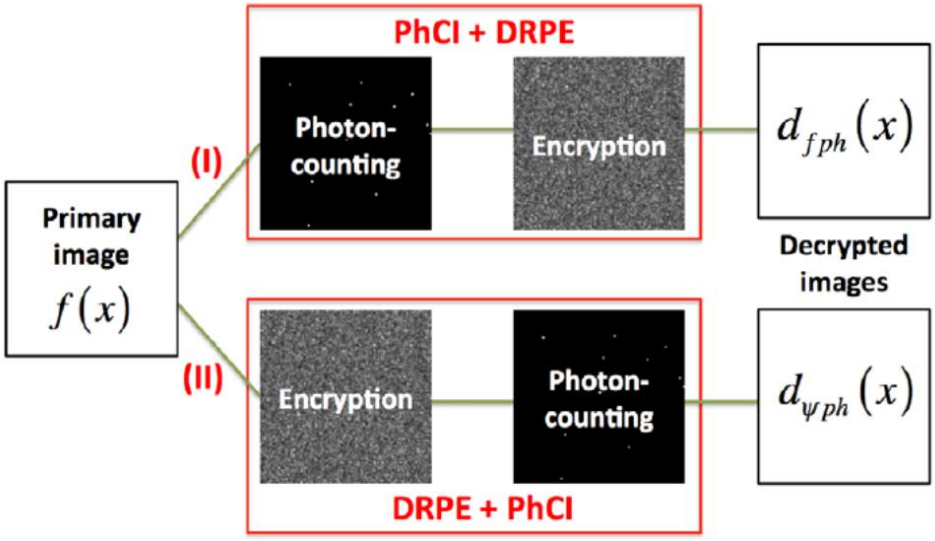

Figure 95 – Two procedures for combining photon counting imaging with encryption. PhCI stands for photon counting imaging and DRPE stands for double-random-phase encryption (Perez-Cabre, et al., 2012).

Others have recently used the DRPE for encryption of 3D images, as well. While 3D images are more complicated to encrypt in general, due to the extra information that must be transferred, the problem is tackled as a series of 2D images that must be reconstructed (Cho & Javidi, 2013).

LIDAR

Another free-running application is LIDAR, which uses photon time of flight information to estimate elevation (distance from an airborne imaging system or satellite). A laser pulse is aimed at an object or area, and the detector is synced with the laser pulses. The lasers generally output short wave (near) IR signals (e.g. 1.06 μm) with a very narrow filter over the detector, ensuring that the only photons collected are those originally from the laser (Yuan, et al., 2010). The time-correlated signals are collected, and the distance the light traveled can be calculated based on the speed of light. In some cases, the distance resolution is on the order of centimeters in ideal conditions. Figure 96 shows an example of 3D LIDAR imaging (Van Ardt, 2008). In this particular image, warmer colors represent objects with more height (close to the aerial detector). Scenes imaged using LIDAR generally take advantage of multiple capture perspectives to form a point cloud and render a 3D representation of the scene.

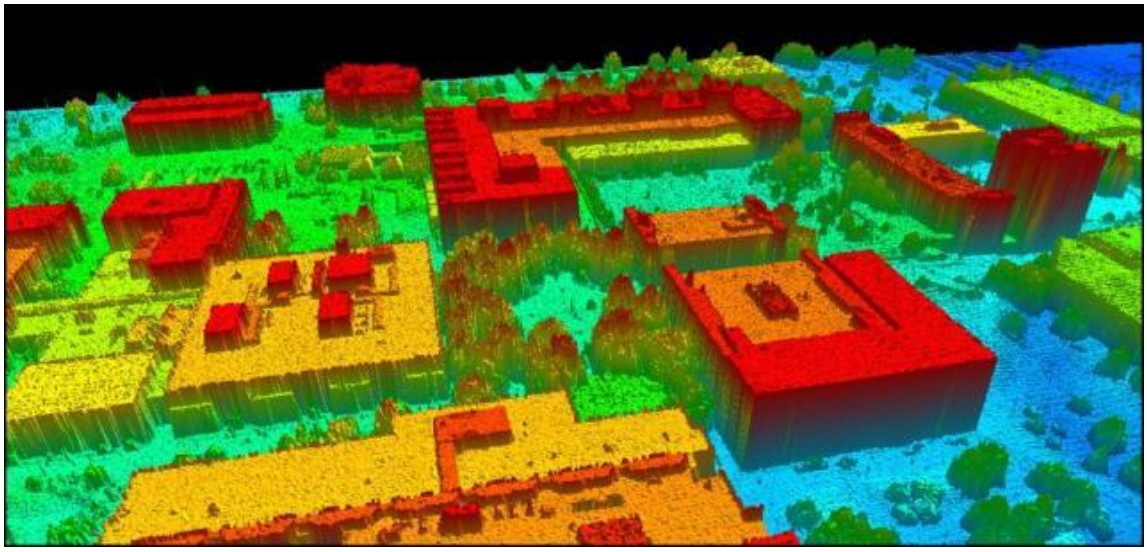

Figure 96 – A 3D LIDAR image of a portion of the RIT campus is shown (Van Ardt, 2008).

REFERENCES

Akkouchi, M., 2008. On the Convolution of Exponential Distributions. *Journal of the Chungcheong Mathematical Society*, 21(4).

Andor, 2012. *Andor Learning: Comparing sCMOS with other detectors.* [Online] Available at: http://www.andor.com/learning-academy/comparing-scmos-compare-scmos-with-other-detectors
[Accessed 13 April 2015].

Anthony, M. & Harvey, M., 2012. *Linear Algebra: Concepts and Methods.* New York: Cambridge University Press.

Aull, B. F. et al., 2015. Detection Statistics in Geiger-Mode Avalanche Photodiode Quad-Cell Arrays With Crosstalk and Dead Time. *IEEE Sensors Journal,* 15(4), pp. 2133-2143.

Aull, B. F., Schuette, D. R., Reich, R. K. & Johnson, R. L., 2010. *Adaptive optics wavefront sensors based on photon-counting detector arrays.* San Diego, USA, s.n., pp. 1-11.

Aull, B. F. et al., 2015. A Study of Crosstalk in a 256 x 256 Photon Counting Imager Based on Silicon Geiger-Mode Avalanche Photodiodes. *IEEE Sensors Journal,* 15(4), pp. 2123-2132.

Bai, Y. et al., 2004. *Hybrid CMOS Focal Plane Array with Extended UV and NIR Response for Space Applications.* San Diego, USA, SPIE, pp. 83-93.

Barber, W. C. et al., 2009. Characterization of a novel photon counting detector for clinical CT: count rate, energy resolution, and noise performance. *Proc. of SPIE, Medical Imaging 2009: Physics of Medical Imaging,* 7258(725824), pp. 1-9.

Basden, A. G., Haniff, C. A. & Mackay, C. D., 2003. Photon counting strategies with low-light-level CCDs. *Monthly Noticies of the Royal Astronomical Society,* 345(3), pp. 985-991.

Beck, J. et al., 2006. The HgCdTe Electron Avalanche Photodiode. *Journal of Electronic Materials,* 35(5), pp. 1166-1173.

Ben-Michael, R., Itzler, M. A. & Nyman, B., 2006. Afterpulsing Effects in 1.5um Single Photon Avalanche Photodetectors. *19th Annual Meeting of the IEEE Lasers and Electro-Optics Society,* pp. 783-784.

Blatt, F. J., 1992. *Modern Physics.* 1st ed. USA: McGraw-Hill.

Borucki, W. J. et al., 2010. Kepler Planet-Detection Mission: Introduction and First Results. *Science,* 19 February, 327(5968), pp. 977-980.

Brinkman, W. F., Haggan, D. E. & Troutman, W. W., 1997. A History of the Invention of the Transistor and Where It Will Lead Us. *IEEE Journal of Solid-State Circuits,* 32(12), pp. 1858-1865.

Brown, T. M., Latham, D. W., Everett, M. E. & Esquerdo, G. A., 2011. Kepler Input Catalog: Photometric Calibration and Stellar Classification. *The Astronomical Journal,* 142(4), pp. 1-76.

Buhler, R. & Blandford, R., 2014. The surprising Crab pulsar and its nebula: a review. *Reports on Progress in Physics,* 77(6), pp. 1-15.

Carson, J. et al., 2013. Direct Imaging Discovery of a "Super-Jupiter" Around the Late B-Type Star k And. *Astrophysical Journal Letters,* 763(2), p. L32.

Carter, J. A., Winn, J. N., Gilliland, R. & Holman, M. J., 2009. Near-Infrared Transit Photometry of the Exoplanet HD 149026b. *Astrophysical Journal,* 696(1), pp. 241-253.

Chervenak, J. A. et al., 1999. Superconducting multiplexer for arrays of trasition edge sensors. *Applied Physics Letters,* 74(26), pp. 4043-4045.

Chessel, A., Waharte, F., Salamero, J. & Kervrann, C., 2013. A Maximum Likelihood Method for Lifetime Estimation in Photon Counting-Based Fluorescence Lifetime Imaging Microscopy. *Proc. of the 21st European Signal Processing Conference (EUSIPCO),* pp. 1-5.

Cho, M. & Javidi, B., 2013. Three-dimensional photon counting double-random-phase encryption. *Optics Letters,* 38(17), pp. 3198-3201.

Coates, P. B., 1975. Fatigue and its correction in photon counting experiments. *Journal of Physics E: Scientific Instruments,* 8(3), pp. 189-193.

Cottingame, W., 2010. *Requirements and Candidates for Ladar Single-Photon Detector Arrays.* Pasadena, KECK Institute for Space Studies Single-Photon Counting Detectors Large Scale Study.

Craven-Bartle, T. V., Dorn, R. J. & Beletic, J. W., 2000. Computer simulation comparison of CCDs and APDs for curvature wavefront sensing. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series,* July, Volume 4007, pp. 444-451.

Currie, D. C. & Choisser, J. P., 1976. A Photon Counting Array Photometer. *SPIE/SPSE Technical Symposium East: Low Light Level Devices,* Volume 78, pp. 83-94.

Daigle, O., 2009. Extreme Faint Flux Imaging with an EMCCD. *Publications of the Astronomical Society of the Pacific,* pp. 866-884.

Daigle, O. et al., 2012. *Characterization results of EMCCDs for extreme low light imaging.* Amsterdam, Netherlands, SPIE, pp. 1-9.

Daigle, O. et al., 2004. L3CCD results in pure photon counting mode. *Proceedings of SPIE: Astronomical Telescopes and Instrumentation,* Volume 5499, pp. 219-227.

Darland, E. J., Leroi, G. E. & Enke, C. G., 1979. Pulse (Photon) Counting: Determination of Optimum Measurement System Parameters. *Analytical Chemistry,* 51(2), pp. 240-245.

Dautet, H. et al., 1993. Photon counting techniques with silicon avalanche photodiodes. *Applied Optics,* 32(21), pp. 3894-3900.

Day, P. K. et al., 2003. A broadband superconducting detector suitable for use in large arrays. *Nature,* Volume 425, pp. 817-821.

Ding, H. & Molloi, S., 2012. Quantification of breast density with spectral mammography based on a scanned multi-slit photon-counting detector: a feasibility study. *Physics in Medicine and Biology,* 57(15), pp. 4719-4738.

Dixon, A. R. et al., 2009. Ultrashort dead time of photon-counting InGaAs avalanche photodiodes. *Applied Physics Letters,* 94(231113), pp. 1-3.

Donovan, D. P., Whiteway, J. A. & Carswell, A. I., 1993. Correction for nonlinear photon-counting effects in lidar systems. *Applied Optics,* 32(33), pp. 6742-6753.

e2v technologies, Nov. 2011. *CCD201-20 Back Illuminated 2-Phase IMO Series Electron Multiplying CCD Sensor,* Chelmsford, Essex, UK: e2v technologies (uk) limited.

Edmonds, L. D., Barnes, C. E. & Scheik, L. Z., 2000. An Introduction to Space Radiation Effects on Microelectronics. *JPL Publication 00-06.*

Eraerds, P. et al., 2007. SiPM for fast Photon-Counting and Multiphoton Detection. *Optics Express,* 15(22), pp. 14539-14549.

Feuerlein, S. et al., 2008. Multienergy Photon-counting K-edge Imaging: Potential for Improved Luminal Depiction in Vascular Imaging. *Radiology,* 249(3), pp. 1010-1016.

Figer, D., 2010. A Photon-Counting Detector for Exoplanet Missions. *JPL Document D-66493*.

Figer, D. & Kolb, K., 2014. A Photon-Counting Detector for Exoplanet Missions. *Technology Development for Exoplanet Missions: Technology Milestone Report,* JPL Document(D-93392).

Finger, G. et al., 2013. NIR HgCdTe Avalanche Photodiode Arrays for Wavefront Sensing and Fringe Tracking. *Proceedings of the Scientific Detector Workshop.*

Ford, V. G. et al., 2004. The Terrestrial Planet Finder coronagraph: technology and mission design studies. *Astronomical Telescopes and Instrumentation,* pp. 1274-1283.

Fry, P. W., 1975. Silicon Photodiode Arrays. *Journal of Physics E: Scientific Instrumentation,* Volume 8, pp. 337-349.

Gaensler, B. M. et al., 2002. Chandra Imaging of the X-Ray Nebula Powered by Pulsar B1509-58. *Astrophysical Journal,* 569(2), pp. 878-893.

Gao, J. et al., 2012. A titanium-nitride near-infrared kinetic inductance photon-counting detector and its anomalous electrodynamics. *Applied Physics Letters,* 101(142602), pp. 1-4.

Gatt, P., Johnson, S. & Nichols, T., 2009. Geiger-mode avalanche photodiode ladar receiver performance characteristics and detection statistics. *Applied Optics,* 48(17), pp. 3261-3276.

Gilliland, R. L., Rajan, A. & Deustua, S., 2010. WFC3 UVIS Full Well Depths, and Linearity Near and Beyond Saturation. *Space Telescope WFC3 Instrument Science Report,* October, Volume 10, pp. 1-21.

Gotthelf, E. V., Halpern, J. P., Buxton, M. & Bailyn, C., 2004. Imaging X-Ray, Optical, and Infrared Observations of the Transient Anomalous X-Ray Pulsar XTE J1810-197. *Astrophysical Journal,* 605(1), pp. 368-377.

Grein, M. E. et al., 2010. Efficient communication at telecom wavelengths using wavelength conversion and silicon photon-counting detectors. *Proceedings of SPIE,* Volume v6709, 670910-1.

Hadfield, R., 2009. Single photon detectors for optical quantum information applications. *Nature Photonics,* Volume 3, pp. 696-705.

Hadwen, B. J., Camas, M. A. & Robbins, M. S., 2004. The Effects of Co60 Gamma Radiation on Electron Multiplying Charge-Coupled Devices. *IEEE Transactions on Nuclear Science,* 51(5), pp. 2747-2752.

Hillman, E. M. C., 2007. Optical brain imaging in vivo: techniques and applications from animal to man. *Journal of Biomedical Optics,* September/October, 12(5), p. 051402.

Hiskett, P. A. et al., 2000. Performance and design of InGaAs/InP photodiodes for single-photon counting at 1.55um. *Applied Optics,* 39(36), pp. 6818-6829.

Impey, C., 2013. The First Thousand Exoplanets: Twenty Years of Excitement and Discovery. In: D. A. Vakoch, ed. *Astrobiology, History, and Society: Life Beyond Earth and the Impact of Discovery.* Berlin Heidelberg: Springer, pp. 201-212.

Itzler, M. A. et al., 2010. *Design and performance of single photon APD focal plane arrays for 3-D LADAR imaging.* San Diego, USA, s.n., pp. 77801M1-5.

Itzler, M. A., Jiang, X. & Entwistle, M., 2012. Power law temporal dependence of InGaAs/InP SPAD afterpulsing. *Journal of Modern Optics,* 59(17), pp. 1472-1480.

Jackson, J. C., Morrison, A. P., Phelan, D. & Mathewson, A., 2002. *A Novel Silicon Geiger-Mode Avalanche Photodiode.* San Francisco, USA, IEEE, pp. 797-800.

Jackson, J. C. et al., 2003. Toward integrated single-photon-counting microarrays. *Optical Engineering,* 42(1), pp. 112-118.

Janesick, J. & Elliot, T., 1992. History and Advancements of Large Area Array Scientific CCD Imagers. *ASP Conference Series: Astronomical CCD Observing and Reduction Techniques,* Volume 23, pp. 1-67.

Janesick, J. et al., 1990. New advancements in charge-coupled device technology - sub-electron noise and 4096x4096 pixel CCDs. *SPIE - Charge-Coupled Devices and Solid State Optical Sensors,* Volume 1242, pp. 223-237.

Janesick, J., Elliott, T. & Pool, F., 1989. Radiation Damage in Scientific Charge-Coupled Devices. *IEEE Transactions on Nuclear Science,* 36(1), pp. 572-578.

Janesick, J. R., 2001. *Scientific Charge-Coupled Devices.* Bellingham, USA: SPIE Press.

Janesick, J., Soli, G., Elliott, T. & Collins, S., 1991. *The Effects of Proton Damage on Charge-Coupled Devices.* San Jose, USA, SPIE, pp. 87-108.

Jensen, K. E. et al., 2006. Afterpulsing in Geiger-mode avalanche photodiodes for 1.06um wavelength. *Applied Physics Letters,* 88(133503), pp. 1-3.

Kimble, R. A., Goudfrooij, P. & Gilliland, R. L., 2000. *Radiation damage effects on the CCD detector of the Space Telescope Imaging Spectrograph.* Munich, Germany, SPIE, pp. 532-544.

Kinch, M. A. et al., 2004. HgCdTe Electron Avalanche Photodiodes. *Journal of Electronic Materials,* 33(6), pp. 630-639.

Kindt, W. & de Langen, K., 1998. Integrated readout electronics for Geiger mode avalanche photodiodes. *Solid-State Circuits Conference,* pp. 216-219.

Kittel, C., 2005. *Introduction to Solid State Physics.* 8th ed. New York: John Wiley & Sons, Inc.

Kolb, K., 2014. Signal-to-noise ratio of Geiger-mode avalanche photodiode single-photon counting imaging detectors. *Optical Engineering,* 53(8), p. 081904.

Kolb, K. E., 2011. *Characterization of silicon Geiger-mode avalanche photodiodes with novel device architecture.* Rochester Institute of Technology: Thesis.

Kolb, K., Hanold, B., Lee, J. & Figer, D. F., 2013. *Test Results for an Array-Based GM-APD Detector Before and After Irradiation.* Florence, s.n.

Kolb, K., Hanold, B., Lee, J. & Figer, D. F., 2014. *Evaluation of GM-APD array devices for low-light-level imaging.* Baltimore, MD, s.n.

Kumar, S. et al., 2008. Temperature dependence of the frequency and noise of superconducting coplanar waveguide resonators. *Applied Physics Letters,* 92(123503), pp. 1-3.

Lacaita, A., Zappa, F., Cova, S. & Lovati, P., 1996. Single-photon detection beyond 1um: performance of commerically available InGaAs/InP detectors. *Applied Optics,* 35(16), pp. 2986-2996.

Liang, J., Grimm, B., Goelz, S. & Bille, J. F., 1994. Objective measurement of wave aberrations of the human eye with the use of a Hartmann-Shack wave-front sensor. *J. Opt. Soc. Am. A,* pp. 1949-1957.

Liang, J., Williams, D. R. & Miller, D. T., 1997. Supernormal vision and high-resolution retinal imaging through adaptive optics. *J. Opt. Soc. Am. A,* pp. 1884-2892.

Li, Z., 2002. *Radiation Hardness/Tolerance of Si Sensors/Detectors for Nuclear and High Energy Physics Experiments.* Carmel, USA, s.n.

Lubsandorzhiev, B. K., 2006. On the history of photomultiplier tube invention. *Nuclear Instruments and Methods in Physics Research A,* 567(1), pp. 236-238.

Marcelot, O. et al., 2014. Study of CCD Transport on CMOS Imaging Technolog: Comparison Between SCCD and BCCD, and Ramp Effects on the CTI. *IEEE Transactions on Electron Devices,* 61(3), pp. 844-849.

Marois, C. et al., 2008. Direct Imaging of Multiple Planets Orbiting the Star HR 8799. *Science,* 322(5906), pp. 1348-1352.

Mayor, M. & Queloz, D., 1995. A Jupiter-mass companion to a solar-type star. *Nature,* 378(23), pp. 355-359.

Mazin, B. A., 2005. Microwave Kinetic Inductance Detectors. *Doctoral dissertation, California Institute of Technology.*

Mazin, B. A. et al., 2012. A superconducting focal plane array for ultraviolet, optical, and near-infrared astrophysics. *Optics Express,* 20(2), pp. 1503-1511.

McIntosh, K. A. et al., 2002. InGaAsP/InP photodiodes for photon counting at 1.06um. *Applied Physics Letters,* 81(14), pp. 2505-2507.

McIntyre, R. J., 1973. On the Avalanche Initiation Probability of Avalanche Diodes Above the Breakdown Voltage. *IEEE Transactions on Electron Devices,* July, Vol. ED-20(No. 7), pp. 637-641.

Miceli, A., Cecil, T. W., Gades, L. & Quaranta, O., 2014. Towards X-ray Thermal Kinetic Inductance Detectors. *Journal of Low Temperature Physics,* pp. 1-7.

Mitra-Kraev, U., Harra, L. K., Williams, D. R. & Kraev, E., 2013. The first observed stellar X-ray flare oscillation: Constraints on the flare loop length and the magnetic field. *Astronomy and Astophysics,* Volume 431, p. 679–686.

Mori, K. et al., 2013. Proton radiation damage experiment on P-Channel CCD for an X-ray CCD camera onboard the ASTRO-H satellite. *Nuclear Instruments and Methods in Physics Research A,* Volume 731, pp. 160-165.

Murray, J. G. et al., 1986. A single-photon-counting Fourier transform microfluorometer. *Journal of Physics E: Scientific Instruments,* 19(5), pp. 349-355.

Nakata, T. et al., 2008. Gated Geiger mode operation and after pulse probability measurement of the InAlAs APD. *21st Annual Meeting of the IEEE Lasers and Electro-Optics Society,* pp. 228-229.

NASA, 2014. *WFIRST-AFTA Science Definition Team Interim Report.* [Online] Available at: http://wfirst.gsfc.nasa.gov/science/sdt_public/WFIRST-AFTA_SDT_Interim_Report_April_2014.pdf
[Accessed 9 October 2014].

Nasuti, F. P., Mignani, R., Caraveo, P. A. & Bignami, G. F., 1996. Spectrophotometry of the Crab pulsar. *Astronomy and Astrophysics,* Volume 314, pp. 849-852.

Neal, D. R., Copland, J. & Neal, D., 2002. *Shack-Hartmann wavefront sensor precision and accuracy.* Seattle, WA, SPIE, pp. 148-160.

Ng, C.-Y., Roberts, M. S. E. & Romani, R. W., 2005. Two Pulsar Wind Nebulae: Chandra/XMM-Newton Imaging of GeV J1417-6100. *Astrophysical Journal,* 627(2), pp. 904-909.

Niemack, M. D. et al., 2010. Code-division SQUID multiplexing. *Applied Physics Letters,* 96(163509), pp. 1-3.

Noeske, K. et al., 2012. WFC3 UVIS Charge Transfer Efficiency October2009 to October 2011. *Instrument Science Report WFC3,* 2012(09), pp. 1-15.

Noh, J. W., Fougeres, A. & Mandel, L., 1991. Measurement of the Quantum Phase by Photon Counting. *Physical Review Letters,* 67(11), pp. 1426-1429.

Norton, P., 2002. HgCdTe Infrared Detectors. *Opto-Electronics Review,* pp. 159-174.

Owens, P. C. M. et al., 1994. Photon counting with passively quenched germanium avalanche. *Applied Optics,* 33(30), pp. 6895-6901.

Papoulis, A. & Pillai, S. U., 2002. *Probability, Random Variables and Stochastic Processes.* New York: McGraw-Hill.

Perez-Cabre, E., Abril, H. C., Millan, M. S. & Javidi, B., 2012. Photon-counting double-random-phase encoding for secure image verification and retrieval. *Journal of Optics,* 14(094001), pp. 1-11.

Petroff, M. D. & Stapelbroek, M. G., 1989. Photon-Counting Solid-State Photomultiplier. *IEEE Transactions on Nuclear Science,* 36(1), pp. 158-162.

Pierret, R. F., 1996. *Semiconductor Device Fundamentals.* Reading, Mass: Addison-Wesley Publishing Company.

Platt, B. C. & Shack, R., 2001. History and Principles of Shack-Hartmann Wavefront Sensing. *Journal of Refractive Surgery,* 17(5), pp. 573-577.

Pool, P. J. et al., 2005. Applications of electron multiplying CCD technology in space instrumentation. *Proceedings of SPIE,* 5902(67).

Quaranta, O., Cecil, T., Gades, L. & Miceli, A., 2014. Mitigation of position dependence in x-rays absorbers for Thermal Kinetic Inductance Detectors. *Journal of Physics: Conference Series,* 507(042033), pp. 1-4.

Radhakrishnan, V. & Cooke, D. J., 1969. Magnetic poles and the polarization structure of pulsar radiation. *Astrophysical Letters,* Volume 3, pp. 225-229.

Reddy, M. et al., 2008. MBE Growth of HgCdTe on Large-Area Si and CdZnTe Wafers for SWIR, MWIR, and LWIR Detection. *Journal of Electronic Materials,* 37(9), pp. 1274-1282.

Renker, D., 2006. Geiger-mode avalanche photodiodes, history, properties, and problems. *Nuclear Instruments & Methods in Physics Research,* Section A(567), pp. 48-56.

Renker, D. & Lorenz, E., 2009. Advances in solid state photon detectors. *Journal of Instrumentation,* April, 4(4), p. P04004.

Ribordy, G., Gautier, J.-D., Zbinden, H. & Gisin, N., 1998. Performance of InGaAs/InP avalanche photodiodes as gated-mode photon counters. *Applied Optics,* 37(12), pp. 2272-2277.

Robbins, M., 2009. *Private Communication with D. Figer* [Interview] 2009.

Robbins, M. S., 2003. The Noise Performance of Electron Multiplying Charge-Coupled Devices. *IEEE Transactions on Electron Devices,* 50(5), pp. 1227-1232.

Robinson, B. S. et al., 2006. 781 Mbit/s photon-counting optical communications using a superconducting nanowire detector. *Optics Letters,* 31(4), pp. 444-446.

Rohlf, J. W., 1994. *Modern Physics from a to Z0.* 1st ed. USA: John Wiley & Sons, Inc.

Romani, R. W. et al., 1999. First Astronomical Application of a Cryogenic Transition Edge Sensor Spectrophotometer. *The Astrophysical Journal,* 521(2), pp. L153-L156.

Rowe, J. F. et al., 2006. An Upper Limit on the Albedo of HD 209458B: Direct Imaging Photometry with the MOST Satellite. *The Astrophysical Journal,* 646(2), pp. 1241-1251.

Santos, N. C., 2008. Extra-solar planets: Detection methods and results. *New Astronomy Reviews,* 52(2), pp. 154-166.

Schneider, J. et al., 2010. The Far Future of Exoplanet Direct Characterization. *Astrobiology,* 10(1), pp. 121-126.

Schwiegerling, J. & Neal, D. R., 2005. Historical development of the Shack-Hartmann wavefront sensor. In: J. H. a. R. Hooker, ed. *Robert Shannon and Roland Shack: Legends in Applied Optics.* Bellingham, WA: SPIE, pp. 132-139.

Seliger, H. H., 1980. Single Photon Counting and Spectroscopy of Low Intensity Chemiluminescent Reactions. In: C. Peng, D. L. Horrocks & E. L. Alpen, eds. *Liquid Scintillation Counting: Recent Applications and Development II.* New York: Academic Press, Inc, pp. 281-319.

Shikhaliev, P. M., 2012. Photon counting spectral CT: improved material decomposition with K-edge-filtered x-rays. *Physics in Medicine and Biology,* 57(6), pp. 1595-1615.

Siegmund, O. H. W., Vallerga, J. & Jelinsky, P., 1986. Calibration of photon counting imaging microchannel plate detectors for EUV astronomy. *Proc. of SPIE, X-Ray Calibration: Techniques, Sources, and Detectors,* Volume 689, pp. 40-48.

Sievers, P. et al., 2012. Bayesian deconvolution as a method for the spectrscopy of X-rays with highly pixelated photon counting detectors. *Journal of Instrumentation,* 7(3), pp. 1-19.

Sing, D. K. et al., 2009. Transit spectrophotometry of the exoplanet HD189733b. I. Searching for water but finding haze with HST NICMOS. *Astronomy and Astrophysics,* Volume 505, pp. 891-899.

Smith, D. R., Ingley, R. & Holland, A. D., 2006. Proton Irradiation of EMCCDs. *IEEE Transactions on Electron Devices,* 53(2), pp. 205-210.

Tsuchiya, Y., Inuzuka, E., Kurono, T. & Hosoda, M., 1985. Photon-Counting Imaging and Its Applications. *Advances in Electronics and Electron Physics,* Volume 64A, pp. 21-31.

Van Ardt, J., 2008. *Lidar over RIT's campus.* [Online] Available at: http://ipler.cis.rit.edu/node/24 [Accessed 5 July 2011].

Vinogradov, S. et al., 2009. Probability Distribution and Noise Factor of Solid State Photomultiplier Signals with Cross-Talk and Afterpulsing. *IEEE Nuclear Science Symposium Conference Record,* pp. 1496-1500.

Vojetta, G. et al., 2012. Linear photon-counting with HgCdTe APDs. *Proceedings of SPIE,* 8375(83750Y), pp. 1-18.

Watanabe, M. & Kramer, S., 2006. 450 mm Silicon: Opportunity and Wafer Scaling. *Interface - Electrochemical Society,* 15(4), pp. 28-31.

Weissleder, R., Tung, C.-H., Mahmood, U. & Bogdanov Jr., A., 1999. In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nature Biotechnology,* April, 17(4), pp. 375-378.

Williams, Jr., G. M., Reinheimer, A. L., Aebi, V. W. & Costello, K. A., 1995. Electron bombarded back-illuminated CCD sensors for low light level imaging applications. *Proceedings of SPIE: IS&T/SPIE's Symposium on Electronic Imaging - Science & Technology,* pp. 211-235.

Yamashita, A. et al., 1997. Radiation Damage to Charge Coupled Devices in the Space Environment. *IEEE Transactions on Nucelar Science,* June, 44(3), pp. 847-853.

Young, M. D., Manchester, R. N. & Johnston, S., 1999. A radio pulsar with an 8.5-second period that challenges emission models. *Nature*, 400(6747), pp. 848-849.

Yuan, P. et al., 2010. 32 x 32 Geiger-mode LADAR Cameras. *Proceedings of SPIE*, Volume v7648, 76840C-1.