Characterization of avalanche photodiode arrays for temporally resolved photon counting

Jana Dee Strasburg

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2004

Program Authorized to Offer Degree: Physics

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

University of Washington

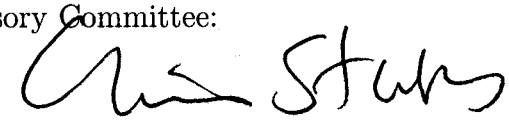
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Jana Dee Strasburg

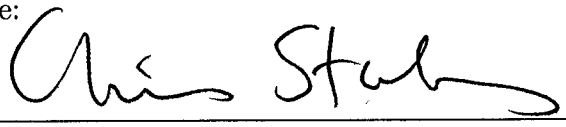
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

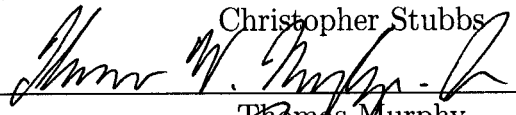


Christopher Stubbs

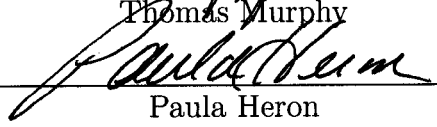
Reading Committee:



Christopher Stubbs



Thomas Murphy



Paula Heron

Date: May 21, 2004

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Bell and Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Jana Da Strasburg

Date May 21, 2004

University of Washington

Abstract

Characterization of avalanche photodiode arrays for temporally resolved photon counting

by Jana Dee Strasburg

Chair of Supervisory Committee:

Professor Christopher Stubbs
Physics

The Apache Point Observatory Lunar Laser-ranging Operation (APOLLO) is a next-generation campaign aimed at measuring the earth-moon separation with millimeter precision. Doing so requires precision measurements of the time-of-flight of photons between the earth and moon. APOLLO will utilize new technology in the form of avalanche photodiode (APD) arrays to detect and time-tag the lunar return photons. The APD arrays were provided by Massachusetts Institute of Technology, Lincoln Laboratory but with no accompanying electronic circuitry for biasing or reading out the detectors. I describe the design, construction and testing of the electronics to bias the arrays and detect photons. Once a final prototype version of the electronics was developed, I characterized the temporal and spatial response of the detectors at two different wavelengths: 786 nm and 668 nm. A simple model of APD performance was developed and compared reasonably well with the experimental data. This model was used to predict the detector performance at 532 nm–the wavelength used for APOLLO.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I'd like to begin by thanking the members of my committee–Christopher Stubbs, Thomas Murphy and Paula Heron–who have taken the time to read (and re-read and re-re-read) this dissertation and provide comments and suggestions. Chris has been my advisor and mentor since I was an undergraduate. During the last eight years he has been patient and supportive, always allowing me to find my own scientific path and "keeping a light on" for me to return to time and time again.

When Chris was unavailable or involved with other projects, Tom did a great job of picking up the ball. He stepped up to the plate and handled much of my day-to-day (and sometimes hour-to-hour) advising. He was infinitely patient with me and I am grateful for everything he has taught me. Many of his words and phrases are scattered throughout this dissertation, mostly in Chapter 2. During the course of APOLLO, Tom (and Chris and Eric Adelberger) wrote several funding proposals describing APOLLO's scientific goals and experimental overview. I used these proposals as jumping off points for the dissertation and thank Tom for allowing me to stand on his shoulders so to speak.

I'm also grateful to Paula for stepping in rather late in the game and bringing a unique perspective to the table. I had the opportunity to work with Paula for several years as part of the Physics Education Group, and found her to be a great role model. Her comments on my dissertation have helped me clarify points and "explain my reasoning" for the intelligent reader with no knowledge of the subject.

I'd also like to thank other faculty members from the University of Washington and elsewhere who played a role in getting me where I am: Gordon Watts, Ellen Covey,

Eric Adelberger, Stamatis Vokos, Bill Greenwood, Richard Louie, Tosh Kakar, Kieth Clay, Laura McGintie and K. T. Tang.

I am grateful to Peter Doherty, the astronomy department's resident electrical engineer, who I had the pleasure to work with for over seven years. His skill and experience were an excellent resource as he taught me the ropes of electronics design and layout.

None of this work would have been possible without the help of MIT Lincoln Laboratory. Special thanks go out to Brian Aull and his colleagues for sharing two prototype devices with us and providing technical support on their behavior and performance along the way.

In a rather bizarre twist, I would also like to thank Harvard and the University of California at San Diego for making offers to Chris and Tom that they couldn't refuse. Their imminent departure from the University of Washington sent me to look for other possible sources of funding for my last year. At the end of that search I arrived at Pacific Northwest National Laboratory in eastern Washington. A National Security Internship introduced me to Warren Harper, Richard Williams, Pam Aker and other members of the Infrared Technologies Team. In this environment, and with their support, I began to learn what it really means to be a scientist. Despite the fact that I was still finishing my Ph.D. (something they all did years ago), they welcomed me as a colleague and an equal. Had I stayed at the University of Washington or under the close watch of Chris or Tom, I'm not sure I would be the burgeoning young scientist I am today.

I am also incredibly grateful to my friends and family for helping me make it this far. I can credit Michael Bush for single-handedly helping me through my first year and encouraging me to stick with the program. I also had frequent friends and study mates who provided support as we struggled through the hard times: Chris,

Andrew, Lisa, Gajus, Armin, Ricardo, Antony, Graeme, Seth, Kareem and others. I wouldn't be here without the love and support of my parents who encouraged me and always told me the sky's the limit. I'm thankful to my non-physics friends who always supported me and reminded me that there was such a thing as life outside the physics building.

Last, but not least, I'd like to thank my husband Michael. He's been there for the good times and the bad, the late nights and the breakdowns. From the beginning he's stood by my side and helped me to remember where my priorities are. Without him and his patience, support, encouragement and love I never would have made it.

Thank you everyone!

DEDICATION

To Michael

Chapter 1

INTRODUCTION

This dissertation represents over three and a half years of work on the Apache Point Observatory Lunar Laser-ranging Operation (APOLLO). During that time, I held the primary responsibility for the detector development and implementation. This required designing electronics to bias and readout the avalanche photodiodes (APDs) we chose to use as the detectors for this time-of-flight measurement. Once a suitable prototype was arrived at, I collected several sets of data aimed at characterizing the temporal and spatial performance of the APDs. We developed a simple model of APD behavior and achieved reasonable agreement between theory and experiment. This allowed for the prediction of the detector performance at 532 nm.

Chapter 2 will be an introduction to the Apache Point Observatory Lunar Laser-ranging Operation (APOLLO). The chapter begins with a brief narration of the history of lunar laser-ranging (LLR). Section 2.2 provides motivation for trying to achieve 1 mm resolution in the distance between the earth and moon. At first glance the usefulness of such a measurement may appear dubious, but LLR can potentially contribute to the areas of cosmology, scalar field and string theory. Section 2.3 lays out some of APOLLO's specific scientific goals. Section 2.4 is a general overview of APOLLO. Section 2.5 discusses APOLLO's estimated error budget and the importance of temporally resolved photon counting. The requirements set forth in this error budget drive the work that is described in the remainder of this dissertation.

Chapter 3 is a more detailed discussion of the APD array detectors and the elec-

tronics necessary to use them. Section 3.1 provides an introduction to the geometry and general behavior of the APDs we are using. Section 3.2 discusses some of the considerations that go into the design of electronics for the biasing, readout and quenching of these devices. This is followed in Section 3.3 by a more detailed description of the various quenching techniques. Section 3.4 details the final circuit design that was used to collect the data presented in Chapter 5.

Chapter 4 contains a more detailed discussion of the theory of APD behavior and provides the background for the numerical simulations. Section 4.1 discusses photon penetration in silicon and its dependence on the wavelength of incident photons. Section 4.2 outlines the theory behind electron diffusion and drift in silicon. These two phenomena play an important role in the temporal and spatial performance of the APDs. Section 4.3 provides a brief discussion on the lateral propagation of an avalanche and its consequences for the temporal resolution of the devices. Section 4.4 describes numerical simulations of the APD that indicate this is a promising avenue for further research.

Chapter 5 holds the bulk of the dissertation and characterizes the temporal and spatial performance of the APDs. The chapter begins with a discussion of the biases inherent in detecting only the first photon. Section 5.2 describes the basic experimental setup. The remaining four sections are particularly data intensive, each dealing with a different aspect of the APD characterization. In an effort to help the reader navigate this portion of the dissertation, each section begins with a bulleted summary of the major results for that section. Section 5.3 focuses on the purely temporal response. Section 5.4 discusses the spatial response to the incident flux or the spatial detection efficiency. Section 5.5 explores the spatial variation of the temporal parameters. Section 5.6 considers the distribution of detected photons in the temporal response and uses comparisons with numerical simulations to estimate the depth of the drift/diffusion region interface.

Finally, Chapter 6 brings us back to the big picture. This chapter begins with

a brief re-cap of Chapter 5. All of the background, modeling and characterization necessary to predict and optimize the detector performance at 532 nm, which is the wavelength that will be used in APOLLO, are laid out in Section 6.2. Using the detector performance predictions and the specifications of the 532 nm laser, Section 6.3 goes on to predict the temporal profiles of the full lunar response. The final section of the dissertation, 6.4, discusses avenues of future work, both in terms of optimizing and analyzing APOLLO data and also in understanding the performance of the APDs.

Chapter 2

LUNAR LASER-RANGING AND APOLLO

Lunar Laser-Ranging (LLR) efforts have been under way since the early 1970s following the placement of the first retro-reflector arrays on the lunar surface by the Apollo astronauts. The Apache Point Observatory Lunar Laser-ranging Operation (APOLLO) will be a next generation LLR campaign aimed at measuring the earth-moon separation with millimeter precision. This increase in range precision will provide an order-of-magnitude improvement in fundamental tests of gravity.

2.1 History of LLR

Lunar Laser-Ranging is a continuing legacy of NASA's Apollo program. A total of five arrays were left on the moon between July 1969 and January 1973. After more than 30 years, the corner cube arrays left on the lunar surface by American astronauts and Soviet unmanned missions are still providing relevant scientific data. Bender et al. (1973); Dickey et al. (1994); Nordtvedt (1996) and Samain et al. (1998) were frequent resources of mine which discuss various aspects of the history and scientific goals of LLR.

The retro-reflector arrays placed on the surface by the Apollo missions 11, 14 and 15 are still functioning without any measurable degradation in their performance. In addition to the three Apollo arrays, two French-built arrays were placed on the surface by Soviet missions 17 and 21. Figure 2.1 shows the different arrays and their locations on the lunar surface. The two French arrays were mounted on mobile rovers Lunokhod 1 and Lunokhod 2 respectively. Lunar return photons from Lunokhod 1

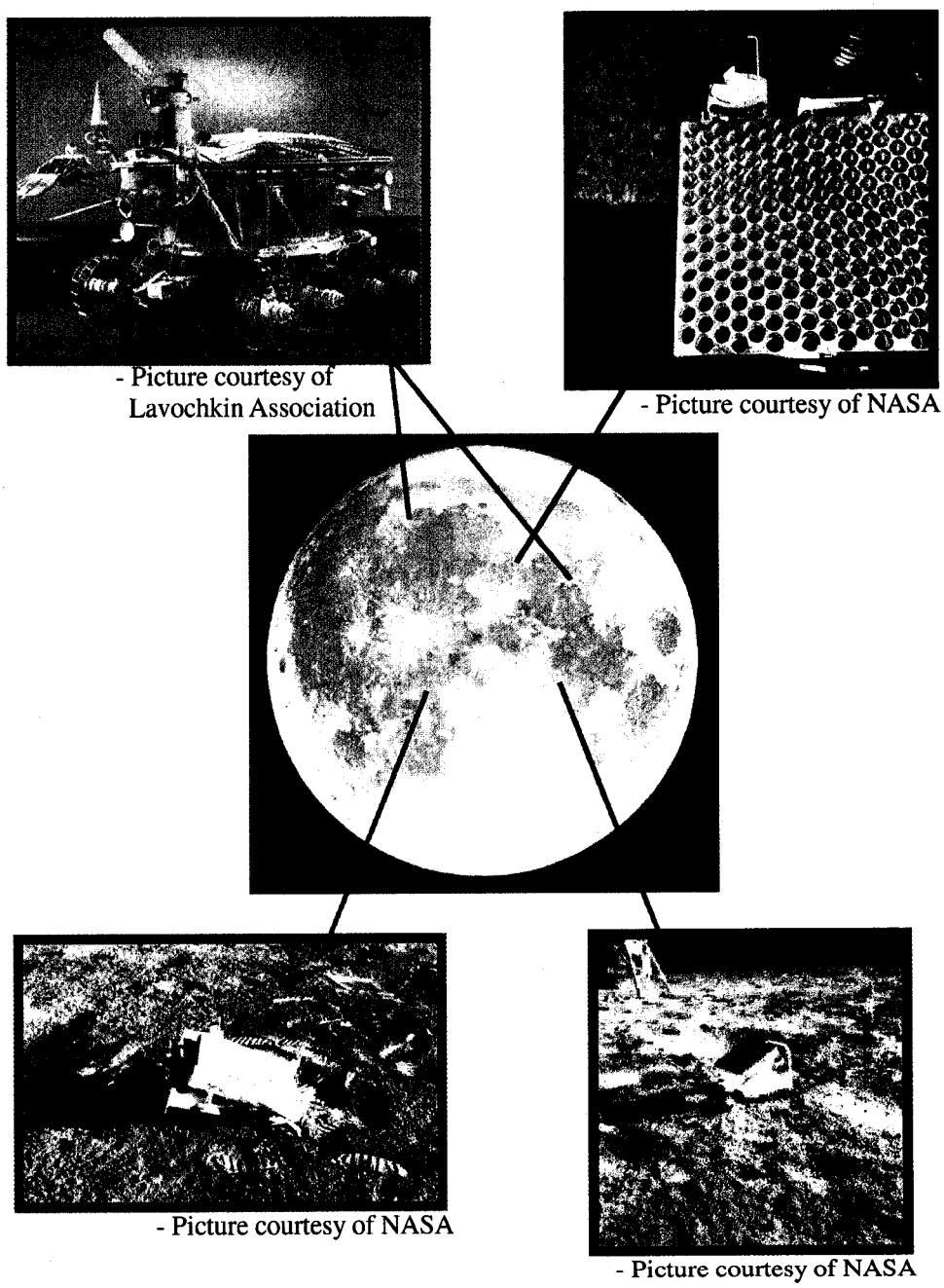


Figure 2.1: Location and images of the five different retro-reflector arrays on the lunar surface. The upper left image is the mobile rover deployed on Soviet missions 17 and 21. The upper right image is the 300-element, Apollo 15 array. The lower two images show the deployed Apollo 11 and 14 arrays.

were detected within a few days of landing, but further attempts to re-acquire a signal from the array have been unsuccessful. It is possible that the array was damaged or covered with dust during surface explorations.

The Apollo 11 and 14 arrays are made up of 100 individual corner cubes mounted in a square panel, 46 cm on a side. The Apollo 15 array consists of 300 corner cubes hexagonally mounted with overall dimensions of 104 cm by 61 cm. Lunokhod 1 and 2 are both arrays similar in size to those from the Apollo 11 and 14 missions and consist of 14 corner cubes, each of which is nearly three times larger than its Apollo counterpart. Unlike the Apollo arrays, Lunokhod 1 and 2 are not designed for operation during lunar daylight. When exposed to direct sunlight, thermal effects cause the divergence of the corner cubes to increase and reduces the lunar detection rates by about a factor of 30 (Kokurin, 1979). Typically Apollo 15 is ranged to most frequently, due to its larger size and thus higher photon return rates.

Dedicated LLR began in 1970 using the 2.7 m telescope at the McDonald Observatory near Fort Davis, Texas. The uncertainty in the earth-moon separation determined from the early data was 250 mm. This increased resolution drastically improved our knowledge of the lunar orbit, previously known only to a few hundred meters. In the mid-1980's the 2.7 m site was decommissioned and the McDonald Laser-Ranging System (MLRS) was initiated at the same site using a dedicated 0.76 m telescope. Even though the light collecting area was much smaller, technological improvements led to an increase in the resolution to the 20–30 mm level. At the same time, two other LLR stations came online: the Observatoire de la Côte d'Azur (OCA) in Grasse, France using a dedicated 1.5 m telescope, and the Haleakala station located in Maui, Hawaii. MLRS and OCA are the only currently operational LLR sites.

2.2 *Scientific motivation for further research*

In the last thirty years, LLR has increased our knowledge of the earth, moon and solar system, and has played an important role in our understanding of gravitational physics. A next-generation LLR endeavor, such as APOLLO, has the ability to further our understanding of gravity and even has the potential to discover fundamentally new physics.

The Equivalence Principle (EP), or the universality of free-fall, is one of the most fundamental principles in all of physics. Essentially it says that all bodies experience the same acceleration in a gravitational field, regardless of their mass or composition. For example, without air resistance, a feather and a bowling ball both fall toward the ground with the same acceleration. The EP is at the very heart of Einstein's Theory of General Relativity, and thus at the heart of our understanding of cosmology and the universe in which we live. With so much of our cosmological framework resting on this principle, it is imperative that we continue to test it whenever possible.

There are many questions that arise when trying to merge classical gravitational theory, such as General Relativity, with quantum theory. What is the quantum analog of gravity? Why is gravity so weak compared to the other fundamental forces of nature? Many different theories are being explored in an attempt to answer these questions (including string theory and brane theory). Many of these incorporate scalar fields as a natural part of their formalism. As we will see in a moment, this has important consequences for the EP.

In addition, recent astrophysical measurements, including the anisotropy scale of the cosmic microwave background (Hanany et al., 2000), the distance measurements of Type Ia supernovae (Riess et al., 1998), and the gravitational behaviors of galactic superclusters (Peacock et al., 2001), point to the conclusion that the universe is expanding at an accelerating rate, thus implying some form of fundamentally new gravitational phenomenon. Scalar fields are also a possible solution to the problem of

an accelerating universe.

Nearly all models that incorporate scalar fields to solve quantum or cosmological problems predict a violation of the equivalence principle at some level; however, the magnitude of the violation is highly model dependent. Therefore high-precision tests of the equivalence principle have a discovery potential with a very broad reach, and in fact provide the most sensitive low-energy window available for probing new physics.

Recent results indicating a possible time variation of the fine structure constant, α, (Murphy et al., 2001) further motivate tests of the equivalence principle. In some models (Damour and Polyakov, 1994b; Damour and Vokrouhlický, 1996) equivalence principle violation and time-varying fundamental constants, such as the fine structure constant and Newton's gravitational constant, G, go hand in hand. With this in mind, LLR provides high-precision, simultaneous tests of the equivalence principle and $\dot{G}$—either or both of which could uncover fundamentally new physics.

2.3 APOLLO's scientific goals

LLR currently provides the best available way to test a variation of the EP known as the strong equivalence principle. LLR also provides measurements of the weak equivalence principle to a precision competitive with current laboratory experiments and furthermore delivers the most precise determination of the time variation of Newton's gravitational constant, G. It also tests the geodetic precession, a relativistic effect predicted by de Sitter in 1916. In addition, the interior, tidal response and physical librations (rocking) of the moon are all probed by LLR.

Each of these will be described briefly in the following sections. I frequently referred to two books by Clifford Will while writing these sections and encourage the interested reader to check them out as well. Will (1986) is written at a popular science level and provides an excellent introduction to these topics. Will (1993) is a more mathematical intensive discussion which ties in the post-Newtonian formalism

and is aimed at a scientific audience.

2.3.1 Strong equivalence principle

The EP can be thought of in several different ways. As mentioned earlier, the EP is also known as the universality of free-fall and says that the behavior of all objects in an external gravitational field is the same regardless of the mass, composition, color, flavor, etc. Another way of thinking of this is that the gravitational mass found in $F = Gm_1m_2/r^2$ is equal to the inertial mass found in $F = ma$. Alternatively, experiments performed in an accelerating reference frame, such as a spaceship speeding up at the rate of 9.8 m/s^2, would have the same results as those performed in a gravitational field at the surface of the earth.

The strong equivalence principle (SEP) is the application of the EP to *all* forms of mass and energy, including gravitational self-energy. Gravitational self-energy can be thought of as the gravitational binding energy of a body. The earth is made up of chunks of material, each of which is gravitationally attracted to every other chunk. The binding energy of the earth is the energy that would be needed to break the gravitational bonds between each chunk of material and move each infinitely far away. Quantitatively the mass due to the gravitational self-energy, M_{SE}, is defined as

$$M_{SE} = \frac{G}{2c^2} \iint \frac{\rho(\vec{x})\,\rho(\vec{y})}{|\vec{x} - \vec{y}|}\, \mathrm{d}^3x\, \mathrm{d}^3y \approx \frac{GM^2}{2c^2R}, \tag{2.1}$$

where G is Newton's gravitational constant, c is the speed of light, ρ is the mass density, M is the total mass and R is the radius of the object in question.

The contribution of mass due to the gravitational self-energy of a body only becomes non-negligible when dealing with objects of astronomical size. For example, gravitational self-energy contributes less than half a part per billion to the total mass of the Earth. The contribution of gravitational self-energy to the total mass of various objects is shown in Table 2.1.

Table 2.1: Contribution of gravitational self-energy to total mass for various objects

Object	Contribution of gravitational self-energy to total mass
One kg of lead	8.1 parts in 10^{26}
Human being	8.5 parts in 10^{26}
Moon	1.9 parts in 10^{11}
Earth	4.6 parts in 10^{10}
Sun	1.1 parts in 10^{6}
Neutron star	1 part in 10

The SEP asks the question, "Do massive bodies such as the earth and the moon behave the same in a gravitational field? Do they both fall towards the sun at the same rate?" Because the SEP applies to massive bodies with non-negligible gravitational self-energy, laboratory tests are not practical.

Since the gravitational self-energy is essentially gravity in a gravitational field, the question arises, is there some kind of gravity-gravity interaction that is different than the interaction between gravity and the other forms of mass and energy? Does the gravitational self-energy contribute equally to both the gravitational, M_g, and inertial mass, M_i? Quantitatively this can be represented by the following:

$$\frac{M_g}{M_i} = 1 - \eta \frac{M_{SE}}{M}. \tag{2.2}$$

If η, also known as the Nordtvedt parameter, is non-zero, then there is a violation of the EP (Nordtvedt, 1968a,b). While η is exactly zero in General Relativity, this is not the case in alternate theories of gravity. By testing the EP, it may be possible to provide constraints on these alternate theories.

At this point I'll digress slightly into a brief discussion of parameterized post-Newtonian (PPN) formalism. PPN formalism is a description of extensions to Newtonian gravity that encompasses most metric theories of gravity. The mathematical expression for the metric is identical for all theories, but the numerical values of the constants change from theory to theory. Mathematical expressions for the subsequent fields, potentials, tensors and equations of motion can be worked out independent of a specific gravitational theory. Only when the appropriate numerical values of the various parameters are introduced do the results change. Two of the ten PPN parameters are β, which is a measure of the linearity of the superposition of gravity and related to the gravitational self-energy, and γ, which is a measure of the amount of space-time curvature produced by a unit mass. Under General Relativity, $\beta = 1$ and $\gamma = 1$. In Brans-Dicke theory, a scalar-tensor theory of gravity, $\beta = 1$, but $\gamma = \frac{1+\omega}{2+\omega}$ where ω is a free parameter of the theory. For very large ω, Brans-Dicke theory reduces to General Relativity. The Nordtvedt parameter, η is related to the PPN parameters by $\eta = 4\beta - \gamma - 3$. γ can also be measured by the time-delay of light as it passes through a gravitational field. Radar ranging of the Viking spacecraft is commonly used to measure γ. The perihelion shift of Mercury also measures a linear combination of γ and β.

Returning now to the discussion of EP violation, any EP signal resulting from the differential acceleration of the earth and moon toward the sun would manifest itself as an apparent shift of the moon's orbit toward or away from the sun relative to the earth. In other words, the moon would be closer to or farther from the earth when it was new than two weeks later when it was full. This apparent shift is illustrated in Figure 2.2. The magnitude of the shift is given by

$$\delta_r = 12.9\,\eta\,\cos(D) \ \text{ meters} \tag{2.3}$$

where D is the elongation angle between the sun, the moon and the earth. By measuring the earth-moon separation with a high degree of precision throughout the

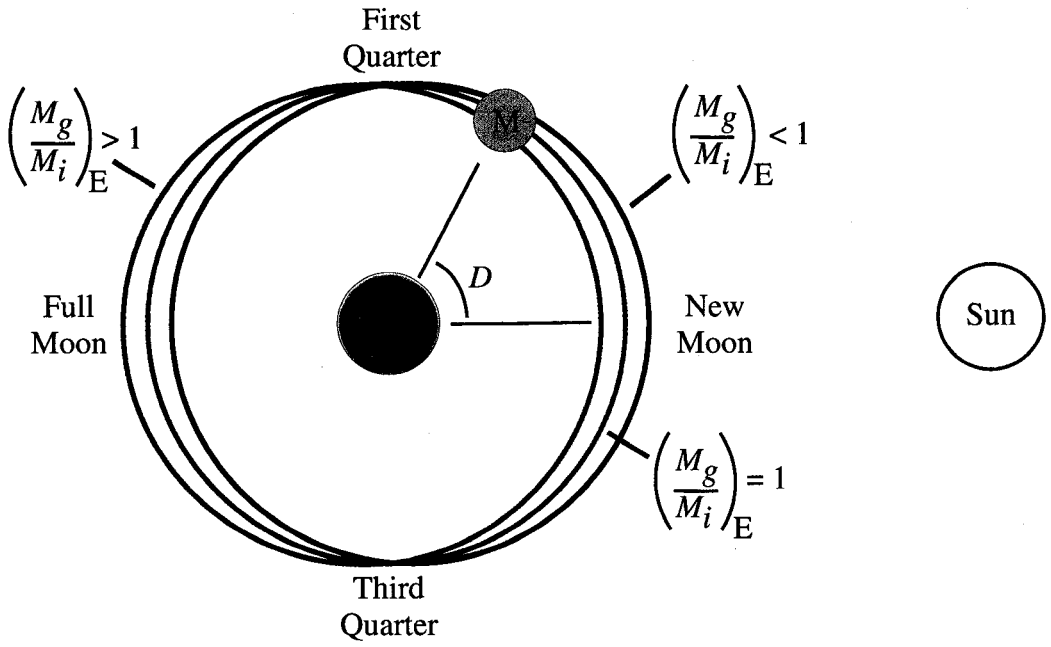


Figure 2.2: Equivalence principle violation causes orbital shifts. Depending on the ratio of the gravitational and inertial mass of the earth, the moon's orbit would appear to shift towards or away from the sun, relative to the earth.

lunar month, it is possible to detect such a signal as a range variation to the moon with a period equal to the synodic period of 29.53 days.

The SEP is currently tested to about one part in 10^3 with no apparent violation (Anderson and Williams, 2001). APOLLO has the potential to test the SEP to a part in 10^4.

2.3.2 Weak equivalence principle

Since the earth and moon have different compositions, LLR is also sensitive to any violation of the weak equivalence principle (WEP). The Earth possesses a massive iron and nickel core, while the moon is thought to contain mostly silicates. If the electromagnetic and nuclear energies of these materials were to interact differently with gravity, then they would accelerate toward the sun differently, resulting in a

signal with the same frequency as any SEP violation.

Because LLR is sensitive to both WEP and SEP violations, we must rely on laboratory-based constraints of the WEP to disentangle the two effects. One way of testing the WEP is the Eötvös experiment where the differential acceleration of two test masses is measured, often using a torsion balance. Landmark experiments at Princeton (Roll et al., 1964) and Moscow (Braginsky and Pavov, 1972) used torsion balances to measure the differential acceleration of various substances. More recently, experiments conducted by the Eöt-Wash group at the University of Washington have constrained any WEP violation to one part in 10^{13} (Su et al., 1994; Baeßler et al., 1999). Laboratory tests of earth-like and moon-like bodies can test the compositional dependence of the EP. These measurements can then be combined with LLR tests of the EP to determine the constraints on the SEP and WEP separately.

Current LLR tests of the WEP are at a level competitive with laboratory-based experiments, namely one part in 10^{13} (Anderson and Williams, 2001). APOLLO will be able to push this limit by a factor of ten. If an EP violation is observed with APOLLO's improved precision it will not be possible to determine if it is due to a weak or strong EP violation until the precision of laboratory tests of the WEP are improved.

2.3.3 Time variation of G

As mentioned earlier, many of the theories that predict a violation of the EP also predict that Newton's gravitational constant, G, will vary in time at a rate proportional to the expansion of the universe. One of the first arguments for a time-varying G grew out of Dirac's large number hypothesis, where fundamental constants can be combined into dimensionless quantities. For example, the ratio of the gravitational and electrostatic force between an electron and proton is a dimensionless quantity on the order of 10^{40}. Another example is the age of the universe expressed in terms of the time light takes to cross the classical electron radius, also about 10^{40}. Dirac asked if

the fact that these two unit-less numbers are the same was a coincidence, or the result of a deeper and as yet unknown law of nature. The latter would suggest that (since the age of the universe has been increasing with time) the ratio of the gravitational and electrostatic force between an electron and proton must also be changing with time. As a result, one or more of the "constants" — the mass of the electron and proton, the electric charge, the speed of light or Newton's gravitational constant — must also be changing with time.

While the large number hypothesis may be viewed as scientific hocus-pocus, other theories tie the expansion of the universe in with a time-varying gravitational constant. Newton's law determines the gravitational force between two bodies regardless of all the other bodies in the universe. But what if the force did depend (ever so slightly) on the distribution of the surrounding matter? In our small corner of the universe G has the value of 6.673×10^{-11} Nm^2/kg^2, but maybe in some other far off corner the value is 6.5×10^{-11} Nm^2/kg^2. This being the case, G could vary at every point in space-time based on the distribution of matter around that space-time point. An expanding universe will certainly change the distribution of matter and could therefore cause G to change with time.

If G were decreasing with time, the gravitational self-energy of massive bodies would decrease. This would cause their radii to increase and their rotation rates to slow in order to conserve angular momentum. The orbital radii of objects would also increase. This means that the moon's orbital radius and period would increase at a rate proportional to the expansion of the universe. Tidal friction also changes the mean radius and period of the lunar orbit, but not in the same ratio as a time-varying G (also known as $\dot{G}$). The moon is receding from the earth at the rate of about 2.6 cm per year due to tidal friction. By studying the recession over long periods of time (several decades), it has been possible to separate out the signals due to $\dot{G}$ and the diurnal and semi-diurnal tides.

$\dot{G}/G$ is currently constrained at the level of 3×10^{-12} per year (Williams et al.,

Table 2.2: Current precision of scientific objectives. APOLLO should achieve an order-of-magnitude improvement on these measurements.

Measurement	Current Precision
Strong Equivalence Principle	1 part in 10^3
Weak Equivalence Principle	1 part in 10^{13}
$\dot{G}/G$	3 parts in 10^{12}
Geodetic Precession	3.5 parts in 10^3

2002). APOLLO should improve this constraint by an order of magnitude.

2.3.4 Geodetic de Sitter precession

In General Relativity, a gyroscope in a free (and therefore geodetic) orbit about a massive body will precess due to the gravitational curvature caused by the body. This phenomenon was first predicted by Willem de Sitter in 1916 (de Sitter, 1916). He calculated a precession of 19.2 milli-arc-seconds per year for the motion of the earth-moon "gyroscope" in its orbit around the sun.

The geodetic precession is currently measured to a precision of 3.5 parts in 10^3 (Williams et al., 2002) and agrees with de Sitter's prediction. APOLLO should increase the precision of this measurement by an order of magnitude.

2.4 APOLLO overview

APOLLO, like other laser-ranging systems, measures the time-of-flight of a short-pulse laser beam reflected off a distant target—in this case the retro-reflector arrays on the lunar surface. One millimeter of range precision corresponds to 7 ps of round-trip travel time.

We will use the 3.5 meter telescope at the Apache Point Observatory (APO),

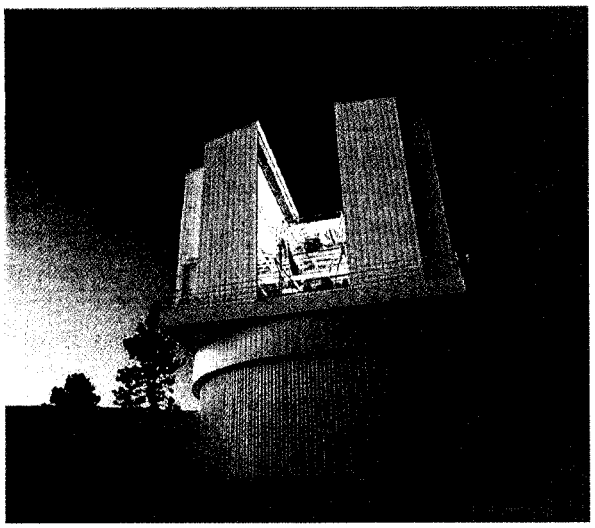


Figure 2.3: Apache Point Observatory 3.5 m telescope enclosure. Photograph courtesy of Dan Long.

pictured in Figure 2.3, as the platform for this operation. Located in southern New Mexico at an altitude of 2788 meters, this site regularly experiences one arc-second seeing or better. The 3.5 m telescope is run by a consortium of universities with the University of Washington owning about 30% of the telescope time. The APO telescope is flexibly scheduled so that coordinating half-hour observing sessions throughout the lunar month poses no serious problems for other users.

The fraction of return photons detected by APOLLO (or any other LLR operation), ϵ, can be found using the following equation:

$$\epsilon = \eta^2 f Q \left(\frac{nd^2}{r^2\Phi^2}\right)\left(\frac{D^2}{r^2\phi^2}\right) \tag{2.4}$$

where η is the one-way telescope and atmospheric efficiency, f is the receiver throughput (dominated by the narrow-band filter), Q is the detector efficiency, n is the number of retro-reflector elements in an array, d is the diameter of the retro-reflectors, r is the distance between the earth and the moon, Φ is the atmospheric divergence, D is the telescope diameter and ϕ is the retro-reflector divergence. The first parenthetical

term represents the loss on the out bound trip due to the relative areas of the retro-reflector arrays and the laser spot on the lunar surface. The second parenthetical term represents the loss on the return trip due to the relative areas of the telescope and the reflected laser spot. APOLLO should typically experience one arc-second atmospheric divergence, 40% telescope, 25% receiver and 30% detector efficiencies. For the Apollo 15 array, the link efficiency is 1.7×10^{-17}. A 115 mJ laser pulse at 532 nm contains roughly 3×10^{17} photons so APOLLO should have a signal rate of roughly 2 photons per pulse.

The size of the telescope and the excellent seeing conditions combine to improve the signal rate over that of current LLR operations by nearly three orders-of-magnitude. The diameter of APOLLO's telescope is larger by a factor of 4.5 and 2.3 than that of the MLRS and OCA operations respectively. This corresponds to detecting 20 and 5 times as many lunar return photons.

The size of the laser spot on the lunar surface is proportional to the atmospheric seeing of the telescope and site. A system experiencing one arc-second seeing produces a laser spot approximately two kilometers in diameter on the lunar surface. MLRS and OCA typically experience 2–4 arc-second image quality, meaning that the number of photons that arrive at the retro-reflector is reduced by a factor of 4–16 as compared to APOLLO. Combining this with the collecting area, APOLLO should realize a factor of 20 gain over OCA and more than 300 over MLRS.

The increased signal strength will improve the statistical error and allow us to investigate possible systematic effects that could not be addressed previously. Under optimum conditions, MLRS and OCA detect at most 1 photon for every 100 launched laser pulses. This translates into detecting an average of one photon every ten seconds. In comparison, APOLLO should be able to detect on the order of 1–10 photons per pulse or several hundred photons every ten seconds.

Unfortunately, optimum conditions can be difficult to achieve. For example, if the laser beam is not well collimated upon exiting the telescope aperture, the beam

diameter on the lunar surface will be further degraded. One additional arc-second of divergence at the telescope would increase the diameter of the beam footprint on the moon by 2 km. If, for example, the lunar tracking algorithm is not working properly and drifts at the rate of one arc-second per minute, it would only take a few minutes to lose the lunar return signal completely. If only one photon is detected every 10, 20, or 30 seconds, it is very difficult to adjust any system parameters such as the pointing, focus or transceiver co-alignment in an effort to optimize the system. On the other hand, if APOLLO's performance is degraded by a factor of 50, it will still detect a photon every few seconds. At this rate, it is much easier to change different system parameters and achieve optimum performance.

The high signal rate will enable APOLLO to operate in high-background conditions such as the full moon and twilight. Because of the difficulty in ranging during high-background conditions, the majority of previous lunar ranges have been centered around the quarter moons—where the equivalence principle signal passes through a null.

Figure 2.4 is a schematic of APOLLO's optical layout. APOLLO will use a commercially available, Q-switched, mode-locked, Nd:YAG laser, operating at a wavelength of 532 nm (Continuum Laser). This laser has a pulse energy of 115 mJ, a repetition rate of approximately 20 Hz, and a pulse width of 40 ps root-mean-square (RMS). The laser will be expanded to fill the entire aperture of the 3.5 m telescope. This will prevent harm to the optics from the very high peak power (in the GW range) as well as reduce the eye-safety threat to pilots.

Filling the the entire aperture of the telescope with the outgoing laser beam requires switching between the transmit and receive modes. This will be done using a rotating optic. This optic is mostly transmissive and anti-reflection coated, but has a small high-reflectivity patch designed to reflect the outgoing laser beam. The laser will be slaved to the motor of the rotating optic, ensuring that the reflective patch always intercepts the laser beam. Figure 2.5 shows the laser mounted on the back

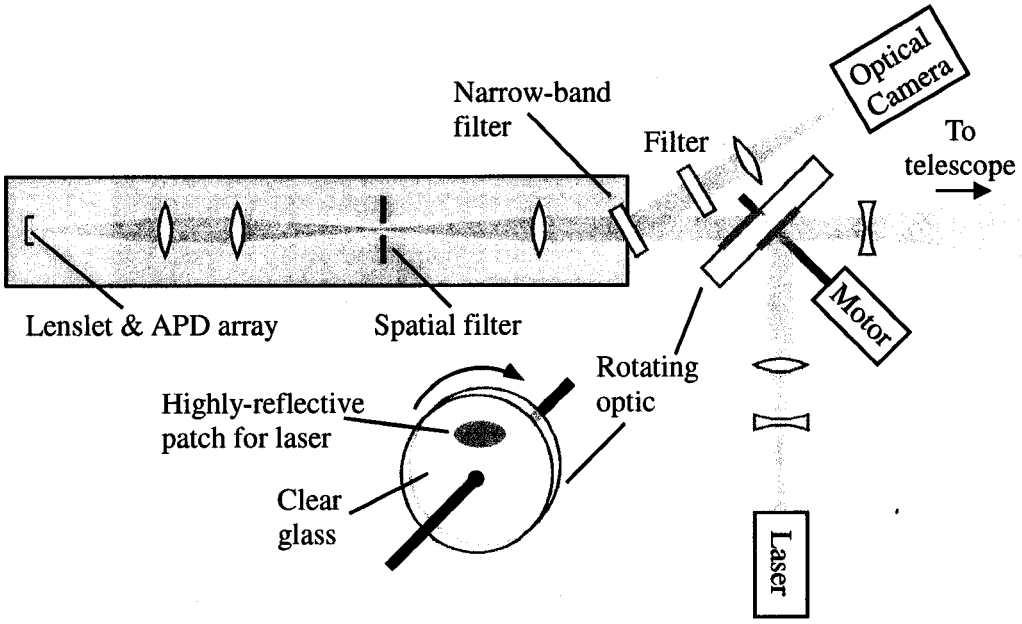


Figure 2.4: Schematic layout of the APOLLO optical path

panel of the telescope.

A small corner-cube prism will be mounted at the telescope exit aperture and will reflect a tiny portion of the outgoing pulse back toward the rotating optic. This reflected pulse will be significantly attenuated so that the number of photons arriving at the APD is comparable to that from the lunar return (on the order of 1–5 photons per pulse). These reflected photons will determine the departure time of the outgoing laser pulse. The location of the corner cube will be mechanically referenced to the intersection of the telescope axes and will serve as the fixed position reference for the differential LLR timing measurement. Experimentally APOLLO will measure the separation between the intersection of the telescope axes and the retro-reflector arrays on the lunar surface.

In addition, the photons from the corner cube enable characterization of the convolution of time responses of the detector, timing electronics and laser pulse shape. Detailed knowledge of this effective profile will be invaluable in interpreting the tem-

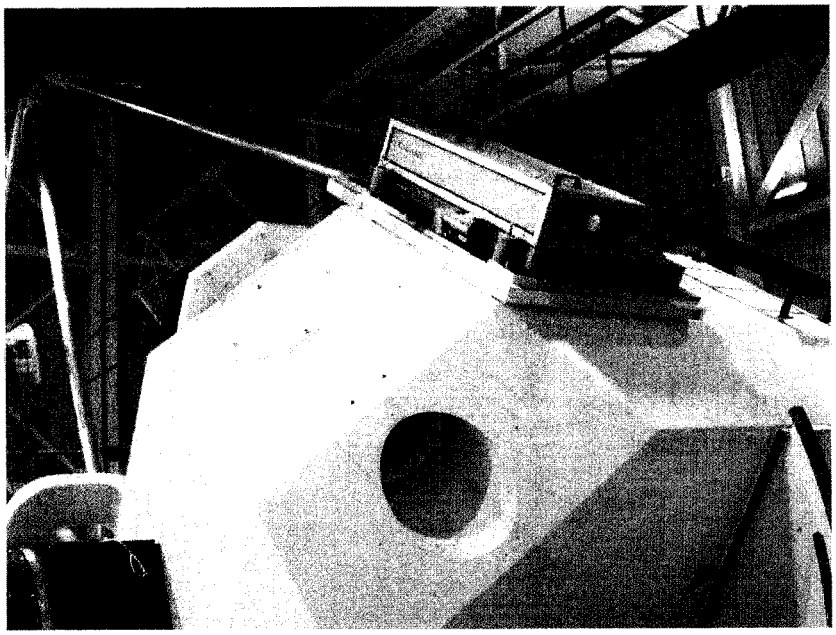


Figure 2.5: Laser mounted on the back port of the 3.5 m telescope. Photograph courtesy of Tom Murphy.

poral profile returning from the moon.

Under optimum conditions the 3.5 m telescope will typically collect several photons per pulse. Ideally, we would like to tag each arriving photon with a time of arrival. Conventional single photon detectors would only be able to detect the first arrival photon, which would significantly bias the results to earlier measured times (and thus shorter ranges). To avoid this, we need to oversample the return pulse so that less than one photon is expected in each detector element. We are introducing new technology in the form of avalanche photodiode arrays to individually detect and time-tag each returning photon. The output signals from these detectors will trigger a sophisticated timing system built around a GPS-slaved 50 MHz clock and a bank of 12-bit time-to-digital-converters.

APOLLO measures the time-of-flight of photons making the $\sim$ 2.5 second round-trip from the telescope to the lunar surface and back. We want to ultimately convert

this to a distance from the center-of-mass of the earth to the center-of-mass of the moon. In order to do so, it is necessary to understand many other phenomena including the body rotation and positional displacements of the surface sites with respect to the center-of-masses of both bodies. These displacements can result from solid earth tides as well as ocean, atmospheric and ground water loading. There is also an additional effect that comes from the path delay through the atmosphere due to its non-unity refractive index.

These effects have traditionally been combated through a combination of meteorology and modeling. Typically, ranging stations measure local temperature, pressure and relative humidity at the surface in order to model the atmospheric path delay and loading. In addition to these measurements, APOLLO will incorporate a precision gravimeter at the site to measure vertical site displacements as small as a few tenths of a millimeter in magnitude. These measurements, along with atmospheric readings, will be used to constrain and improve the model that will ultimately produce the final result.

In addition to the site displacements, modeling is also needed to account for other known effects on the range and center-to-center distances. For example, the center-to-center distance is perturbed by solar and planetary interactions, relativistic gravitational interactions, higher-order gravitational harmonics and tidal friction and dissipation, just to name a few. The range distance is perturbed by an atmospheric path delay, and a time delay of light propagating in the gravitational fields of the Sun and Earth.

The model incorporates a simultaneous least squares fit to over a hundred different parameters in order to produce numerical results for EP violation, $\dot{G}$, geodetic precession and other scientific goals.

2.5 *APOLLO's estimated error budget*

In order to achieve millimeter resolution, it is necessary to measure the time-of-flight of the lunar return photons with an uncertainty of 7 ps; however, the uncertainty of each individual return photon is typically a few hundred picoseconds. By averaging the time-of-flight of several hundred individual photons, it should be possible to arrive at our goal of 7 ps. The uncertainty of a given photon is affected by the retro-reflector array orientation, the laser pulse shape, the detector and timing electronics, each of which will be described in the following sections.

2.5.1 Retro-reflector array orientation

The major contribution to the range uncertainty is the retro-reflector array orientation. The orientation changes as a result of the lunar libration or rocking. The moon's orbit is elliptical causing its speed to vary slightly as it goes around the earth. This in turn causes the lunar face to appear to rock back and forth in the horizontal direction. The inclination of the lunar orbit also causes an apparent rocking in the vertical direction. In addition to these optical librations, there are small physical librations due to the moon's non-uniform rotation causing small discrepancies between the rotation and the revolution. The combined physical and geometrical librations can be as great as 10 degrees. This means that the retro-reflector array orientation can also change by as much as 10 degrees. The arrays are about a half meter on a side so a 10 degree change in orientation can result in a 85 mm difference between the nearest and furthest element of the array. Since a return photon can be reflected from any element, this serves to temporally smear the return laser pulse. At the maximum libration angle, the Apollo 15 retro-reflector array orientation contributes roughly 300 ps RMS of uncertainty to the error budget.

The discrete nature of the retro-reflector arrays actually results in a time response described by a series of delta functions. These delta functions are then convolved

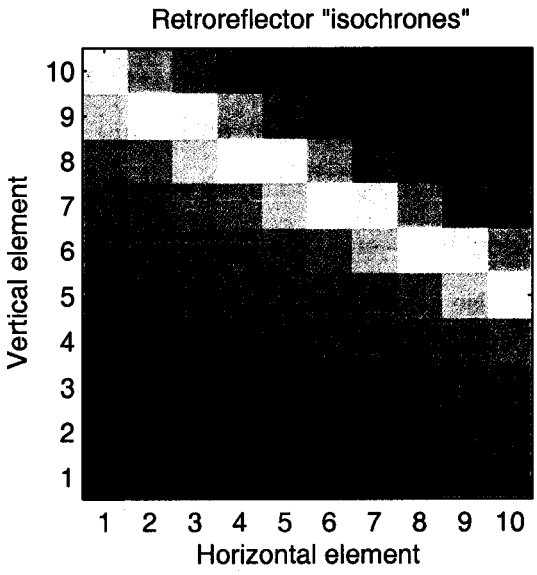


Figure 2.6: Retro-reflector "isochrones" for $(5°, 2.5°)$ libration angles. Elements with the same color will have the same range. The corner-to-corner spread is about 110 mm or 770 ps.

with the laser pulse, detector and timing response and other noise sources to give the observed temporal profile. Obviously the smaller the contribution from the laser pulse, detector and timing electronics, the better the resolution of the retro-reflector array.

Figure 2.6 shows a 100 element retro-reflector array with a libration orientation of 5° latitude and 2.5° longitude. Elements having the same time of flight are labeled with different colors resulting in a series of "isochrones" across the array. For this particular orientation, the resulting probability distribution is trapezoidal and consists of evenly spaced delta functions like those shown in the top plots of Figures 2.7 and 2.8. If the two libration angles are nearly equal in magnitude, the distribution will be triangular and if either one of the angles is near zero, the distribution will be rectangular.

Nothing can be done to improve the uncertainty due to the retro-reflector array

orientation; however, careful experimental design can assure that this is the dominant term in the error budget. Figures 2.7 and 2.8 show the distribution of the retro-reflector array with $(5°, 2.5°)$ libration angle and the total distribution when the convolution of the laser, detector and timing electronics contributes 14 ps and 71 ps respectively. Comparison between the two bottom plots shows the advantages of using a very short laser pulse with very precise detectors and timing electronics. In this case, the peaks due to the various isochrones are still resolvable. However, if the laser and electronics have a more realistic total contribution of 71 ps, the individual peaks are smeared out.

2.5.2 Laser pulse width

The next significant contributor to the error budget is the laser pulse itself. The size of this contribution depends on the choice of lasers. When LLR first began at MacDonald Observatory, they used a ruby laser with pulse widths on the order of tens of nanoseconds (or pulse thicknesses of several meters). Under these circumstances, the uncertainty of the laser pulse dominated over the retro-reflector response. APOLLO will use a Nd:YAG laser operating at 532 nm with a FWHM of about 100 ps or 30 mm. The uncertainty contribution of this laser is about 40 ps RMS, which is significantly smaller than the maximum retro-reflector orientation contribution of 300 ps RMS. However, inspection of the bottom plot of Figure 2.8 indicates that this will not be enough to resolve the individual peaks of the retro-reflector response within the total distribution.

At this point you may be asking, "Why not use an even tighter laser pulse? After all, femtosecond lasers are commercially available." The advantage of a shorter laser pulse must be weighed against other factors, including the output power, the cost and availability, operational and maintenance requirements. In addition, if the detectors and timing electronics are only capable of 50 ps of uncertainty, the advantages of the shorter laser pulse are lost due to the coarseness of the detection capabilities.

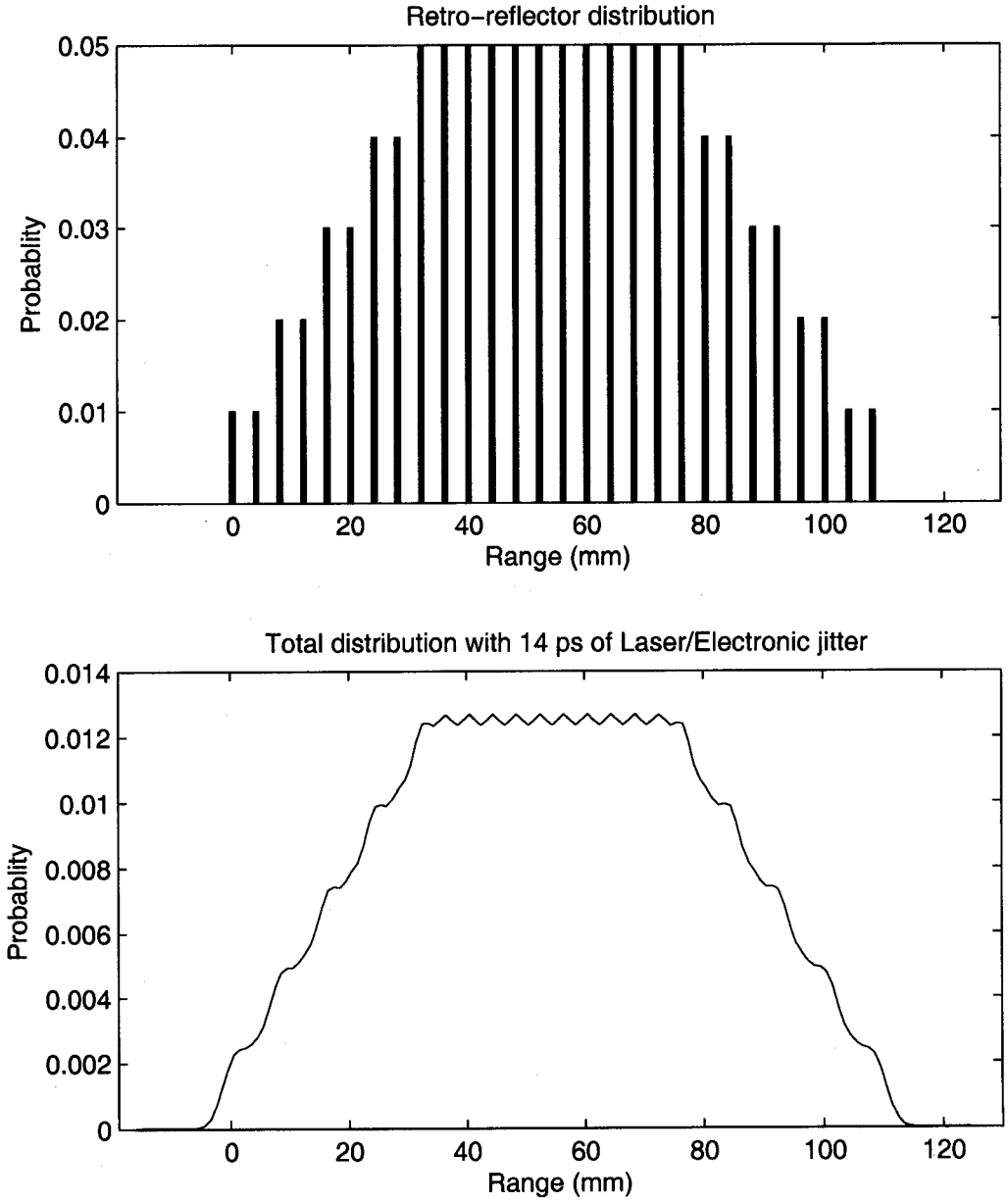


Figure 2.7: *Top:* Retro-reflector array response for $(5°, 2.5°)$ libration angles. *Bottom:* Total distribution if array response is convolved with 14 ps of laser/detector jitter (10 ps each). The individual peaks of the array response are clearly resolvable in the total distribution.

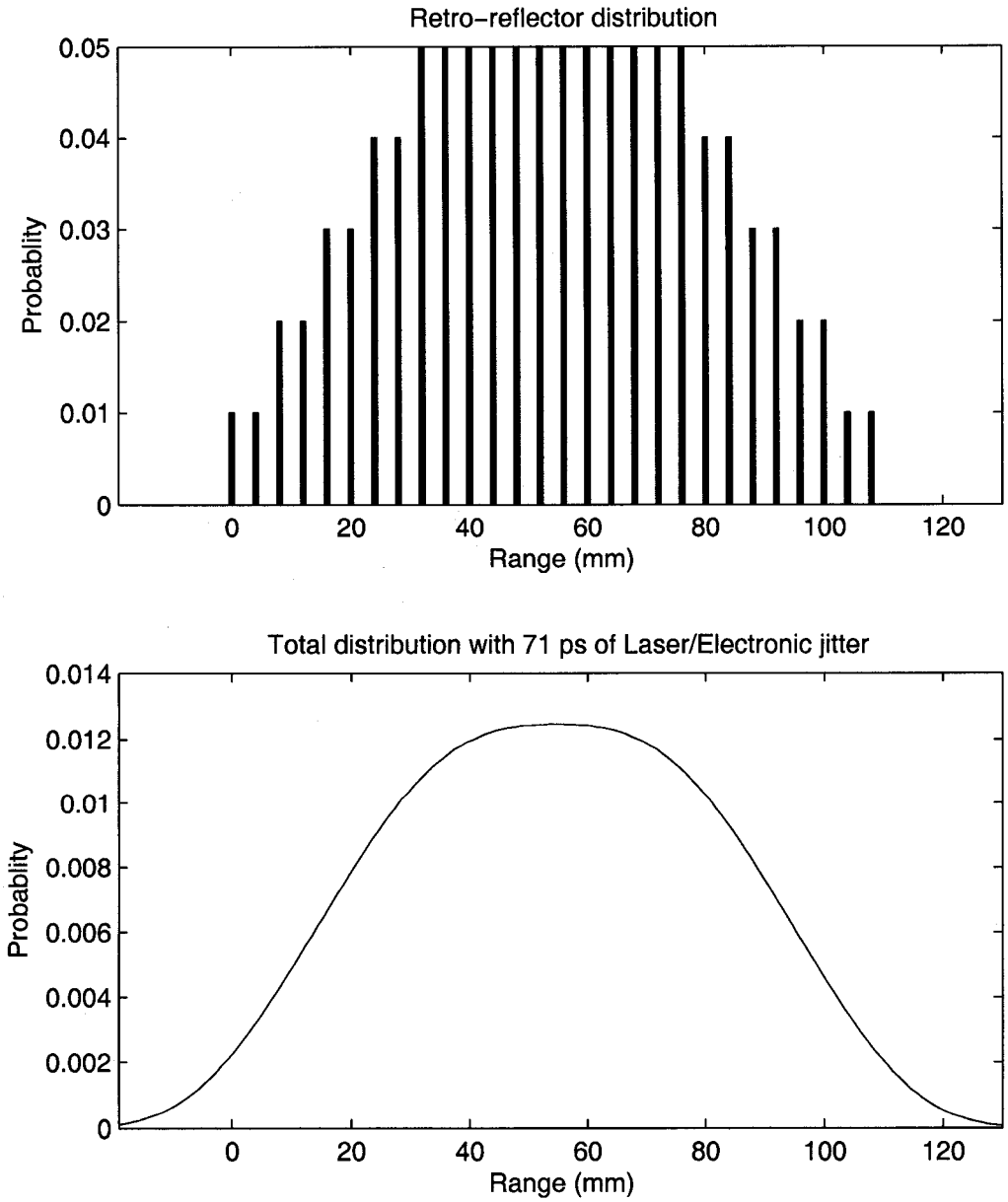


Figure 2.8: *Top:* Retro-reflector array response for $(5°, 2.5°)$ libration angles. *Bottom:* Total distribution if array response is convolved with 71 ps of laser/detector jitter (50 ps each). The individual peaks of the array response are no longer resolvable in the total distribution.

2.5.3 Detector and timing electronics

Ideally one would like the contribution due to the laser pulse and the retro-reflector array orientation to dominate over all the other sources of random error. This requires that the uncertainty due to the detector and timing electronics be less than 40 ps RMS. In the same way that shorter laser pulses better characterize the retro-reflector array response, the more precise the contribution from the detectors and timing electronics, the better the characterization of the laser pulse.

This is illustrated by Figures 2.9–2.11. The top panel of all three figures shows the very distinct laser pulse profile of the Hamamatsu Picosecond Light Pulser (PLP) test laser at 786 nm. These data were streak camera results of the pulse profile provided by Hamamatsu. If this laser pulse is convolved with a very low jitter detector and electronics response (e.g., $\sigma = 5$ ps for Figure 2.9) the result is still very much like the laser pulse. As the uncertainty of the detector and timing electronics increases to $\sigma = 25$ ps (Figure 2.10) and $\sigma = 75$ ps (Figure 2.11), the convolved response looks less like the laser pulse and begins to be dominated by the electronics or detector timing response. This being the case, our goal was to choose detectors and design electronics that would measure the arrival time of an incident photon with a precision of less than 40 ps in order to best characterize the laser's temporal response.

2.5.4 Complete error budget

Table 2.3 shows APOLLO's estimated random error budget. As discussed earlier, the retro-reflector array orientation typically dominates the error budget followed by the laser pulse width and the detector and timing electronics. Millimeter resolution is achieved by a $\sqrt{N}$ improvement over single photon timing, where N is the number of photons collected. For example, if the uncertainty in the time-of-flight of each individual photon is 140 ps (or 20 mm), averaging 400 photons will result in an uncertainty of $140/\sqrt{400} = 7$ ps. If APOLLO receives 1 photon per pulse, it would

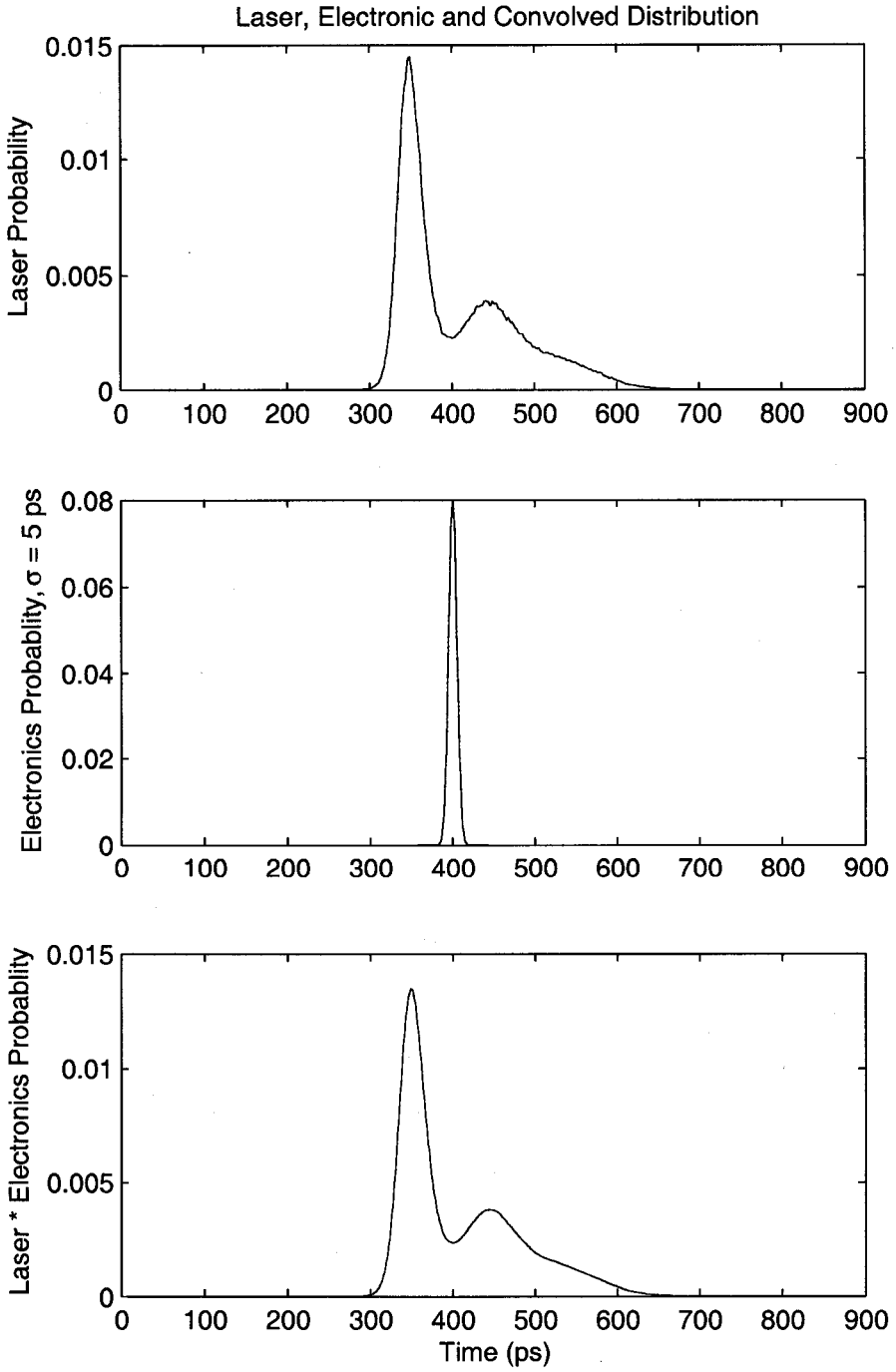


Figure 2.9: *Top:* Laser pulse probability of Picosecond Light Pulser (PLP) at 786 nm; *Middle:* Probability distribution of electronic jitter with $\sigma = 5$ ps; *Bottom:* Convolved distribution of laser and electronics. There is very little difference between the top and bottom frames.

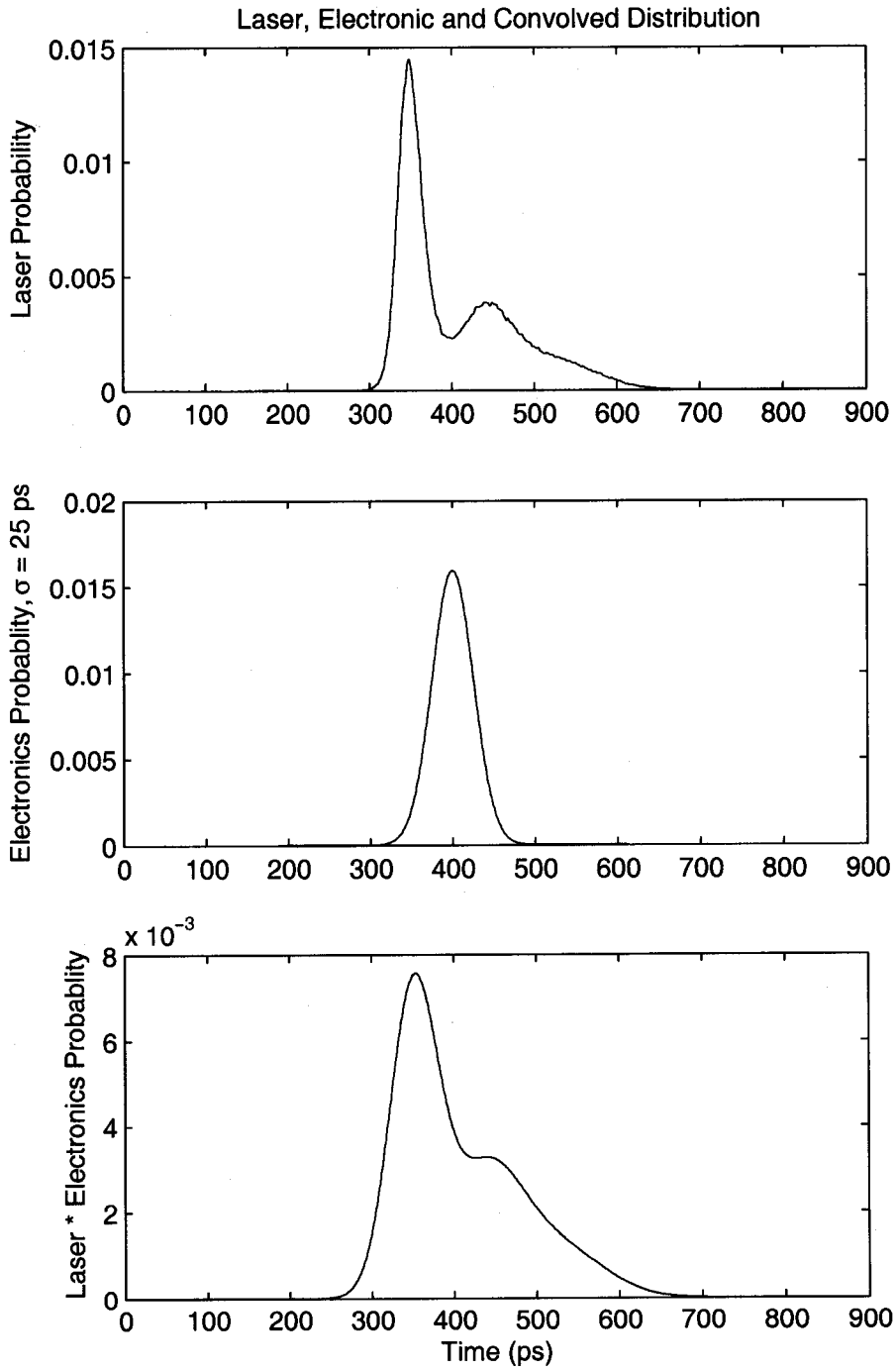


Figure 2.10: *Top:* Laser pulse probability of PLP at 786 nm; *Middle:* Probability distribution of electronic jitter with $\sigma = 25$ ps; *Bottom:* Convolved distribution of laser and electronics. Distinguishing features of the laser pulse become less distinct, though the secondary hump is still prominent.

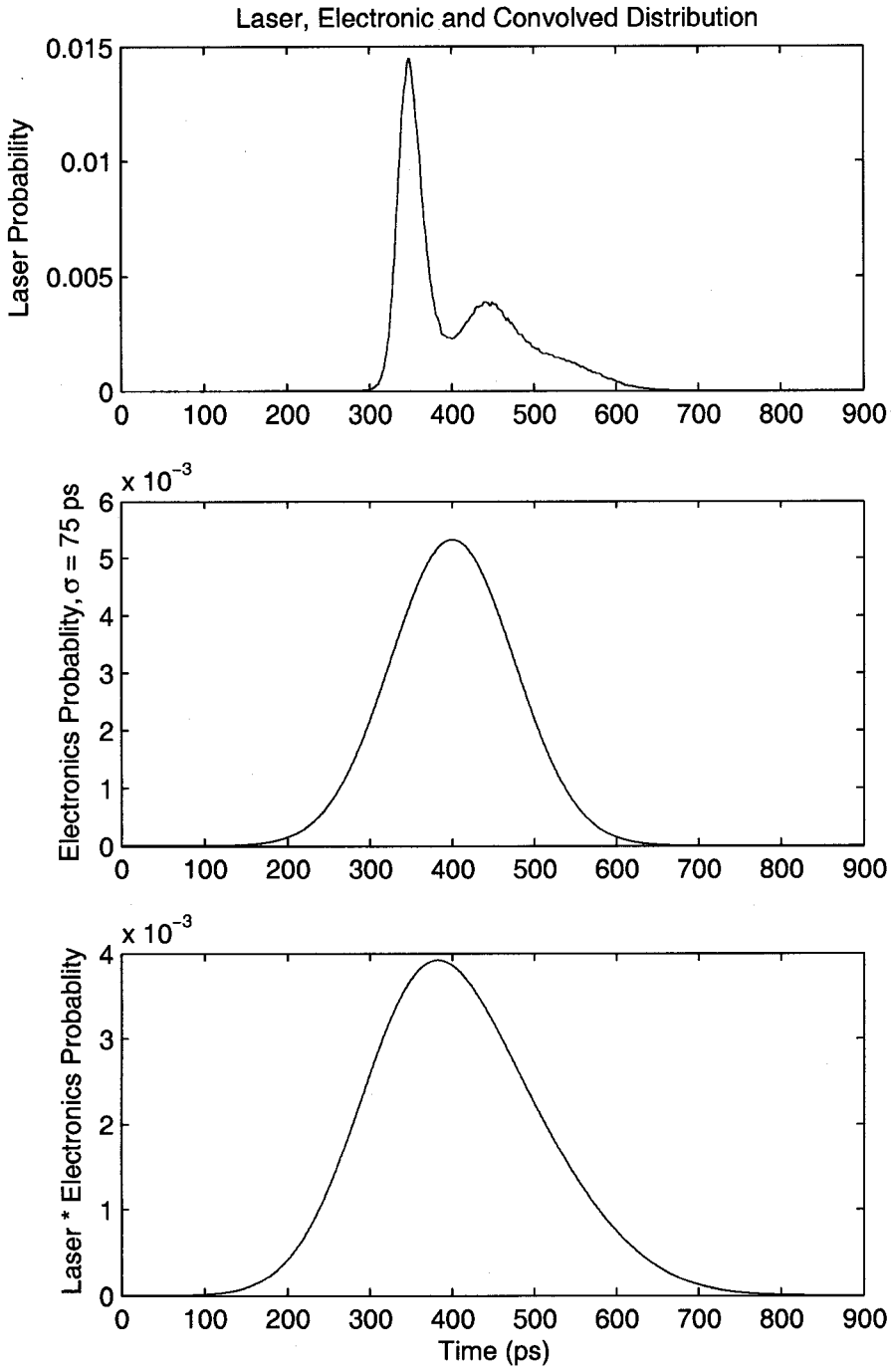


Figure 2.11: *Top:* Laser pulse probability of PLP at 786 nm; *Middle:* Probability distribution of electronic jitter with $\sigma = 75$ ps; *Bottom:* Convolved distribution of laser and electronics. There is little difference in the profile between the middle and bottom frames. Only the asymmetry of the laser pulse is still apparent.

Table 2.3: APOLLO's estimated random error budget

Source	Estimated RMS error
Laser pulse width	40 ps
Detectors and timing electronics	50 ps
Time to digital converter	15 ps
Retro-reflector array orientation	0–300 ps
Time stability of clock over 2.5 seconds	< 10 ps
Noise from background photons and dark counts	< 1 ps
Start photodiode timing and calibration	20–30 ps
Total quadrature combination	70–310 ps
Events required for millimeter resolution	100–1960 photons

take less than 30 seconds to collect enough data to reach our goal of 1 mm precision.

Chapter 3

INTRODUCTION TO AVALANCHE PHOTODIODES

As discussed in the previous chapter, APOLLO will determine the separation between the earth and moon by high-precision measurements of the time-of-flight of photons from the Apache Point Observatory to the moon and back. The location and size of the telescope should allow APOLLO to receive between one and ten return photons per pulse, an improvement of two to three orders-of-magnitude over current operations. The increased signal rate also presents its own challenges, since each photon must be detected and time-tagged. Several different methods could achieve this end, such as using beam-splitters or fiber-optic bundles to send the light to individual detectors, or using an array of single-photon detectors. We have chosen to do the latter.

Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) has been developing arrays of avalanche photodiodes for use in laser-radar applications. In the winter of 2001 we received two 16-element prototype arrays from MIT LL. The devices were mounted in 40-pin DIP chips and have element diameters of 20 and 30 microns with a 100 micron pitch. Figure 3.1 is an image of the 30 micron device. If used as is, the large spacing between elements would mean that the array would have a fill-factor of less than 10%; however, if a lenslet array with the proper spacing is utilized, a nearly 100% fill-factor can be recovered.

MIT LL had developed their own CMOS timing circuitry, but the temporal jitter was on the order of a few hundred picoseconds (Albota et al., 2002b). This is understandable since laser radar applications do not typically require millimeter resolution. As such, when the APDs arrived there were no accompanying electronics to

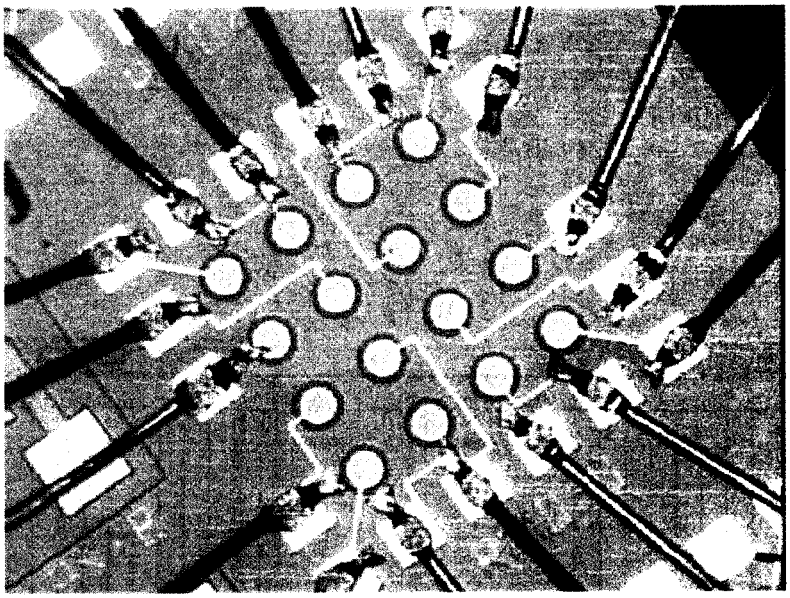


Figure 3.1: Image of 16-element APD array from Lincoln Laboratory. The diameter of each element is 30 microns and the spacing between elements is 100 microns. A lenslet array will be placed in front of the array to recover a nearly 100% fill factor.

bias or readout the devices, and I spent the next two years developing hardware to accomplish this task.

This chapter will start with an introduction to the geometry and general behavior of the APD elements followed by a description of the electronics necessary to bias and readout the array. This chapter is aimed at those who may be interested in using APDs as high temporal-resolution, photon detectors. Readers interested in a more theoretical description of APD behavior will be more interested in Chapter 4.

3.1 Geometry and general behavior

A schematic of the cross section of a single APD element is shown in Figure 3.2. The basic structure of the diode is a central $n-p-\pi-p$ region. The p region is silicon heavily doped with receptor-sites (boron); the π region is lightly doped boron-doped silicon and n is heavily donor-doped silicon using arsenic and phosphorus (Aull et al.,

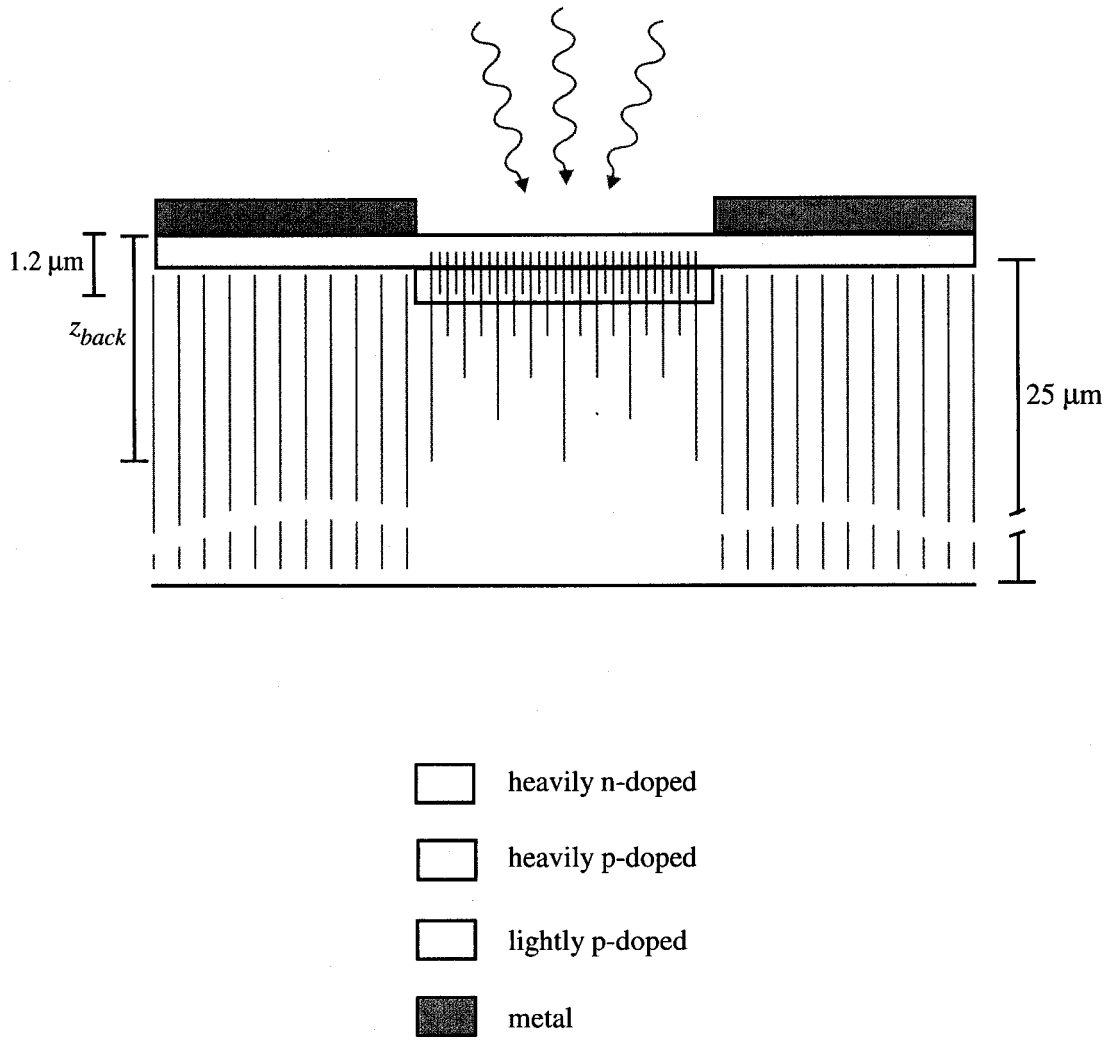


Figure 3.2: Schematic of APD cross section. The APD has a $n - p - \pi - p$ structure where π means mean lightly p-doped. The gray lines represent electric field lines whose density reflects the strength of the electric field. The electric field is strongest in the region of the p implant and decreases to zero deeper in the silicon. The central, active region of the APD is surrounded by a $n - \pi - p$ region that serves as a guard ring that limits the volume for potential avalanche-initiating carriers.

2002).

This structure sets up a very high electric field at the interface of the p implant. Photons arriving in this region can impact ionize an atom in the lattice and produce a photoelectron-hole pair. The high electric field causes the pair to separate. The photoelectron in turn has sufficient energy to impact ionize another atom and so on. Thus the arrival of a photon in this region can initiate an avalanche. The region is therefore referred to as "the avalanche region." Depending on the doping properties of the silicon, impact ionization can be dominated by either the electrons or the holes. The avalanche region is on the order of half a micron.

Deeper into the silicon, the magnitude of the electric field decreases. I will call this region "the drift region." Here the electric field is not strong enough to cause impact ionization, but serves to move electrons toward the avalanche region. The electric field is strong enough to cause the electrons to move at their saturation velocity of 1×10^5 m/s. Because the avalanche region is so thin and close to the surface, the vast majority of incident photons penetrate well into the silicon before photo-converting into an electron-hole pair. For example, at 786 nm more than 30% of the incident photons penetrate deeper than 10 microns. The electric field in the drift region sweeps photoelectrons towards the avalanche region where they can be detected as a steady-state current as mentioned above. The drift region has a thickness of a few microns and depends on the bias voltage applied to the diode and the diode's breakdown voltage.

Below z_{back}, the back of the drift region, there is no electric field and electrons diffuse isotropically. This process will be discussed in more detail in Section 4.2. I will call this region "the diffusion region." At longer wavelengths more than 50% of the incident photons can penetrate into the diffusion region before photo-converting, in which case they are only able to reach the drift region (and thus any hope of detection) by a random-walk process.

Depending on the application, the APDs can also be "back illuminated" where-

upon the substrate is thinned to a few tens of microns. When illuminated from the back, longer wavelength photons will penetrate through the substrate and photo-convert near the drift region. This allows for better detection probabilities and timing results because fewer photons get converted in the substrate and are only detected after a random-walk process.

Surrounding the central $n-p-\pi-p$ region is a $n-\pi-p$ region with a moderate electric field that acts as a guard ring, so that only electrons from the central portion of the APD can diffuse or drift into the avalanche region. This minimizes the effects of background or spurious hits from dark-counts (electron-hole pairs that are not created by photons) and crosstalk between elements.

Based on the description above, it is clear that there are two important components for detecting a photon: the likelihood that a photon converts into an electron-hole pair, and the likelihood that the photo-electron initiates a detectable avalanche. The first quantity is also known as the *Quantum Efficiency* and is familiar to anyone who uses solid-state devices (such as CCD arrays) to detect photons. The second, known as *Avalanche Probability* is more particular to APDs and has to do with the likelihood that a photo-electron initiates a self-sustaining avalanche. Statistically, there is a certain probability that the charge in the avalanche region could fluctuate to zero and result in a non-event. As a result, not every photo-generated electron which passes through the avalanche region will initiate a self-sustaining avalanche. Together these can be combined into the *Photon Detection Efficiency*, typically 30–50% for non-anti-reflection-coated devices.

Since the electrons and holes are constantly being "swept" into either the p or n regions by the electric field, there is a relationship between how fast the charge is being created, either by incident photons or impact ionization, and how fast the charge is being removed. This is controlled by V_d, the voltage drop across the diode, and $V_{breakdown}$, the breakdown voltage of the diode. If $V_d < V_{breakdown}$, charge is created slower than it is swept away and the avalanche eventually tapers off. This

results in a well-defined relationship between the number of incident photons and the current through the diode with a typical gain of tens to hundreds. On the other hand, if $V_d > V_{breakdown}$, charge is created faster than it can be swept away and the avalanche will continue indefinitely. This is also known as "Geiger-Mode." In this case the steady state current, I_d, will be related to the electrical properties of the diode by

$$I_d = \frac{V_d - V_{breakdown}}{R_d} = \frac{V_{excess}}{R_d}. \tag{3.1}$$

where R_d is the internal resistance of the diode, on the order of 20 kΩ, and V_{excess} is the excess voltage beyond the breakdown voltage.

When operating in Geiger-Mode only the first photon is detected, meaning that subsequent photons that arrive while an avalanche is already in progress will go unnoticed. In addition, there is no detector noise to speak of. Instead, the noise source is from spurious events. If an avalanche is initiated by a thermally generated electron (resulting in a dark count), or a background photon before a lunar return photon arrives, then the lunar return photon will go undetected. Thus background hits limit the amount of available "real-estate" where we can detect lunar return photons.

Figure 3.3 shows the first-order equivalent circuit for the APD. While waiting for a photon to arrive, the diode is mainly a capacitive device with a capacitance of a few picofarads. When a photon initiates an avalanche, it is akin to closing the switch and allowing the current to flow through R_d.

3.2 Bias, readout and quenching electronics

Having discussed how APDs work, it is important to note that simply providing a voltage drop across the APD will not allow for the detection of the steady-state current caused by an incident photon, nor a way to quench or shut-off the current so that the APD can detect another photon. This must be accomplished by supplementary power, readout and quenching electronics. APDs can be quenched using one of three

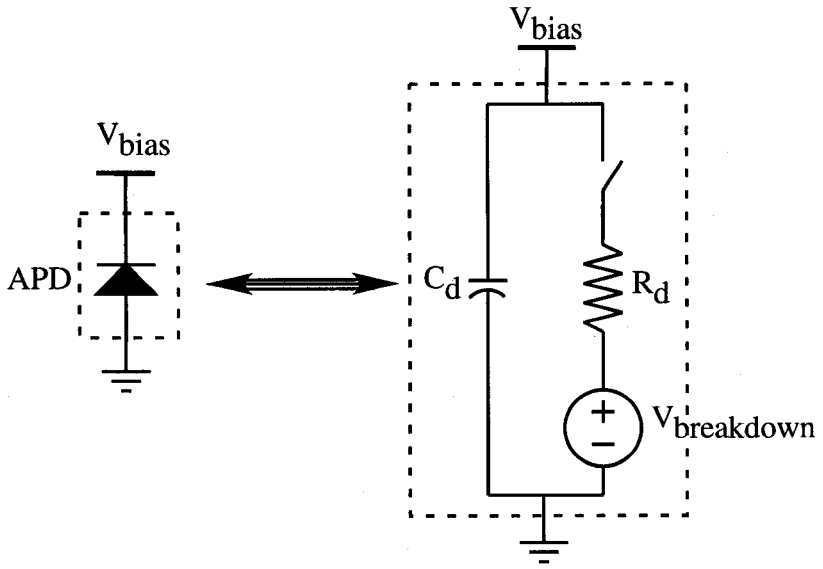


Figure 3.3: First-order equivalent circuit for APD. For our devices, R_d is on the order of 20 kΩ and C_d is a few picofarads, $V_{breakdown}$ is about 25 V and V_{bias} is about 28 V, resulting in an excess voltage of approximately 3 V.

different methods: passive quenching, active quenching and gated quenching. During the last decade, researchers from the Department of Electronics and Information at Milan Polytechnic have published several papers on APDs and associated electronics. Specifically, I found Ghioni et al. (1996); Cova et al. (1996, 1989); Lacaita et al. (1995) extremely useful. The doctoral dissertation of Willem Kindt was also very helpful (Kindt, 1999).

As mentioned earlier, APDs can work as single photon detectors when they are run in Geiger-Mode. Under these circumstances the APDs are biased a few volts above the breakdown voltage, producing a strong electric field within the diode. When a charge carrier is released in this field, it is rapidly accelerated, colliding with and ionizing other atoms along the way. These secondary charges are in turn accelerated and continue to impact ionize other atoms, thus creating an avalanche of charge that can be detected from a single photoelectron.

Several features of the APD behavior are related to the excess voltage or the voltage beyond breakdown. The photon detection efficiency and inherent timing resolution are increased with the excess voltage. Unfortunately the dark counts also increase, both in the form of thermal excitations and trapped charge (also known as after-pulsing). As a result, the excess voltage must be carefully chosen to balance the benefits of improved detection efficiency and timing resolution with the higher dark count rate.

The excess voltage can vary as a result of several things. To begin with, the breakdown voltage varies with temperature. As the temperature increases, the breakdown voltage can increase by 0.3% per degree K (Cova et al., 1996). If a constant bias is applied, the excess voltage can vary with temperature, changing the performance of the detector significantly. For example, consider a device with a 25 V breakdown voltage that is biased at a constant 30 V. If the breakdown voltage increases by 5% (corresponding to a $\sim$ 17 degree temperature change), the excess voltage goes from 5 V to 3.75 V, a 25% reduction. This translates into a 1.5% per degree K change in the excess voltage. Secondly, the stability of the bias voltage, including overshoots and ringing, can play a role in the excess voltage. When gating the APDs, the resistor-capacitor combination of the ballast (or quenching) resistor together with the APD capacitance and any stray capacitance in the circuit, means the excess voltage will be exponentially rising and decaying to a steady-state value after the gate pulse and could result in a greater variation in pulse heights.

3.3 Quenching options

The behavior of the APD is similar regardless of the quenching method used. When an avalanche is initiated, the current through the diode increases with a rise time of a few nanoseconds or less. The current reaches a maximum value based on the excess voltage and the internal resistance of the APD.

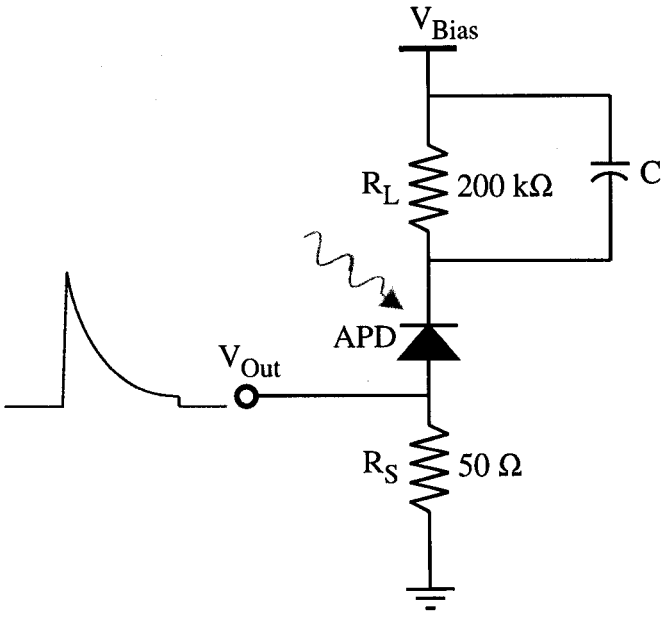


Figure 3.4: APD Passive quenching circuit.

3.3.1 Passive quenching

Figure 3.4 shows a typical passive quenching circuit consisting of the series combination of a large ballast resistor, R_L, of a few 100 kΩ, the APD and a small sensing resistor, R_S on the order of 50 Ω. Before the APD fires there is no current flowing through the 50 Ω sensing or 200 kΩ ballast resistors, so the voltage drop across the diode is equal to V_{bias}. When an avalanche is initiated the current through the diode rapidly increases to its steady state value of I_d, which is related to the excess voltage and internal resistance of the diode by Equation 3.1. When the current is initiated by the avalanche, the voltage drop across the sensing resistor increases rapidly while the voltage drop across the resistor-capacitor (RC) network of the 200 kΩ in parallel with C (which is on the order of a few picofarads) remains nearly unchanged because the voltage drop across a capacitor (in this case initially zero volts) cannot change rapidly. As the voltage across the RC network increases, the voltage drop across the APD and sensing resistor begin to decrease. When the voltage drop across the APD

is reduced below the breakdown voltage, the avalanche is quenched and the current through the APD drops to zero. At the time of quenching the voltage drop across the RC network is nearly equal to the excess voltage of $V_{bias} - V_{breakdown}$. The RC network begins to discharge, reducing the voltage drop across it and therefore increasing the voltage drop across the APD so that the APD is once again biased beyond its breakdown voltage and capable of detecting another photon. As soon as the APD is biased beyond breakdown, it is capable of detecting another photon, regardless of the fact that the RC network has not yet fully discharged. This means that the excess voltage could be considerably reduced, which would result in a lower peak current I_d for a subsequent photon. The time constant for the RC network is on the order of microseconds, so while an avalanche can be quenched in a matter of nanoseconds, the recovery time is several orders-of-magnitude longer.

The shortcomings of this particular design for photon detection are readily apparent. The long recharge rates make the detectors highly sensitive to background and dark counts. If an avalanche has recently been initiated, the APD may not be fully recovered by the time a lunar return photon arrives. This would lower the efficiency of detecting the photon we care about as well as varying the amplitude of the voltage spike across the sensing resistor. This would require a constant fraction discriminator to sense the rise as opposed to a simple threshold comparator.

3.3.2 Active quenching

Another method of APD quenching is known as active quenching, where the onset of an APD avalanche is sensed using a comparator. The signal from the comparator then actively shuts down the APD by lowering the applied bias voltage below that of breakdown. After some designated hold-off time (as short as tens to hundreds of nanoseconds) the APD bias voltage can be raised back above breakdown and the APD is ready to detect another photon. Gating is easily incorporated into the active quenching scheme using TTL pulses and logic gates.

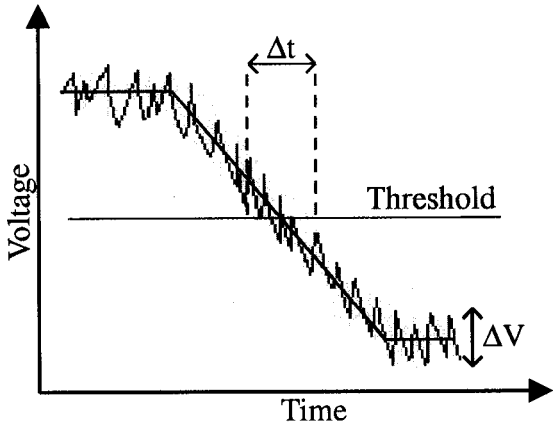


Figure 3.5: Impact of signal noise on timing jitter of a comparator

There are several things to consider with active quenching. First, the use of a comparator must be carefully designed so that it responds to the several-millivolt pulse of the APD but is unaffected by the several-volt pulse that will quench the APD. Various methods have been employed to deal with this problem. For example, a comparator with a differential input can be used: the quenching pulse is applied to both terminals (common mode signal) while the APD pulse is only applied to one side (differential signal). In this way, only the small APD pulse triggers the comparator. One advantage of the comparator is that the output is a logic pulse that can be very fast and very large compared to the APD output. As a result, amplifiers and/or discriminators are not needed between the quenching circuit and the TDC.

At this point it is important to consider the effect a simple threshold comparator can have on the temporal jitter. The comparators trigger at a constant threshold level; therefore the timing jitter of the logic pulse is directly related to the uncertainty in the voltage at the threshold crossing. This is illustrated by Figure 3.5. When there is noise of magnitude ΔV on a trigger signal this noise is translated into a temporal jitter Δt if a threshold crossing method is used. The timing jitter can be reduced by reducing the noise on the triggering signal or by increasing the slope of the trigger

signal. To minimize the timing jitter due to a threshold crossing, a quiet signal with a sharp or steep transient is required. It should also be noted that there could also be noise on the threshold signal. This would further degrade the temporal jitter of the threshold crossing.

While active quenching is easily applied to single APD elements, it would not necessarily be effective for an array. All 16 elements of the APD have the same common anode and would therefore experience the same bias voltages. If an active quenching method were used, the first photon would effectively turn off the remaining 15 elements. If the first photon were a lunar or calibration photon, there is a chance that other photons from the same laser pulse could squeeze in before the bias voltage was reduced; but if a background photon or dark count was detected first, the entire array could be shut down before the calibration or lunar photons arrive. It could be possible to have separate gates for each element, but this would increase the complexity of the circuitry and could cause interference between channels. For example, rapidly turning off the gate signal on one channel could produce a transient or extra noise on another.

3.3.3 Gated quenching

Gated quenching is a slight modification of the active quenching method. In this case the APD is biased above the breakdown voltage for short periods of time (subnanosecond to microsecond) and, once the APD fires, current is allowed to flow through it until it is "quenched" by the termination of the gate pulse. The gate pulse can be either DC or AC coupled with the bias voltage. With DC coupling, the limit of the gate duration, repetition and duty cycle is only determined by the thermal and trapping effects of the APD elements. In other words, we have to consider the amount of power dissipated by the APD if it fires early in the gate and remains "on" for the entire duration of the gate which could be over 100 ns. Like the passive gating and active quenching, care must be taken in reading out the APD

since the current through the sensing resistor responds both to the APD pulse and the transients associated with the gate opening and closing.

The main advantage of this method is that it works well with arrays. Each element will fire independently on only the first photon it detects. In addition, using smaller resistors means that the excess voltage should be steady for the large majority of the 100 ns window. Since the APD will be fully charged for the arrival of the first photon, the pulse height will be constant, meaning a simple threshold discriminator can be used.

With this method, we could have a rather simple setup containing a few passive components and a comparator for each channel. The external gating would be the same for all the elements and would be applied at the input of the array, making only one pulse producing/conditioning circuit necessary. This circuit could be driven using logic pulses from the control computer that would tell the gate when to open. Consider the scheme of a single lunar range. Once the laser fires, the control computer sends a signal turning on the APD, since it will be receiving photons from the telescope's internal corner-cube in a few tens of nanoseconds. Two and a half seconds later, when the computer expects a return photon from the moon, it again tells the gate to open.

3.4 Final circuit design

From the spring of 2001 thru the spring of 2003 I designed, built and tested six generations of electronic circuitry. The first stage began with a simple passive circuit similar to that shown in Figure 3.4. This was followed by more complicated passive quenching and gated quenching circuits. The final design is very similar to one published by Cova et al. (1989); Lacaita et al. (1995). Figure 3.6 shows the schematic of the Mark VI circuit which was used to bias and gate-quench a single APD element. All the data presented in the remainder of this dissertation were collected with this design.

The details of the circuit design will not be discussed, but there are a few things worth noting. The circuit is highly symmetric with the APD being matched to a variable capacitor, C_{var}, and the 12 kΩ resistor being matched to a 12 kΩ resistor in series with a 10 kΩ potentiometer. The lower portion of the circuit is the detector channel while the upper portion is the transient or dummy channel.

V_{gate} is a logic pulse that turns the APDs on and off. The excess voltage is determined by V_1. All the data discussed in this dissertation were collected with V_1 equal to 5 V or 5.5 V. This resulted in an excess voltage of approximately 3 V. V_2 is chosen so that a specified amount of current flows through the 2 kΩ resistor when V_{gate} is in the "on" state. We set V_2 equal to $V_1 + 2$ V (7 or 7.5 V) for our experiments. V_{Bias} is a constant value that represents the bias of the APD in the "off" state. For the work presented later, V_{Bias} was set equal to 22.5 V.

C_{var} is on the order of a few picofarads and is used to mirror the APD's response to the gate signal. This is because the APD is primarily a capacitive device prior to breakdown. The use of the other diodes and transistors mean that the voltage at the nodes to the left of the the transistors are the same for both the detector and dummy channels. Adjusting the 10 kΩ potentiometer changes the current that flows through the 500 Ω resistor located below V_A. This produces an offset ΔV in the voltage between V_A and V_B, as illustrated by the two traces. If no avalanche is initiated, the voltages at V_A and V_B track each other, regardless of V_{gate}, and there is no threshold crossing to trigger an output pulse for the comparator. On the other hand, if an avalanche is initiated in the APD, the voltage at V_B drops below V_A resulting in a comparator flip. Once an avalanche is initiated, current will continue to flow through the APD until the gate is turned off.

The signals at the bottom of Figure 3.6 also illustrate some of the challenges of this particular scheme. Ideally V_A and V_B would be square pulses with no noise or ringing, but in practice this is not the case. The response of the APD is super-imposed on this pulse so the slope of the threshold crossing could change significantly depending on

where in the gate the photon arrives. Adjusting the potentiometer essentially adjusts the threshold level of the comparator and can also impact the temporal jitter.

For the final electronic design, it will be important to optimize the location of the calibration and lunar return photons in the temporal gate as well as the appropriate values of C_{var} and the 10 kΩ potentiometer in order to find the "sweet spot" of minimum temporal jitter. Our experience is that this circuit design, in combination with the TDC, is capable of better than 25 ps of resolution when fully optimized. However, without optimization, the jitter can be on the order of 100 ps. This optimization was not done on the Mark VI circuit because of the prototype nature of the electronics and it was decided such an effort was not a good use of time and resources. In addition, all of this circuitry was fabricated in the electronics shop at the University of Washington on two-layer boards. The restriction to two layers makes the use of ground planes and short traces difficult if not impossible. As a result the signals will have significantly more noise and variation than the final design, which will be implemented on a multi-layer board and should perform much better.

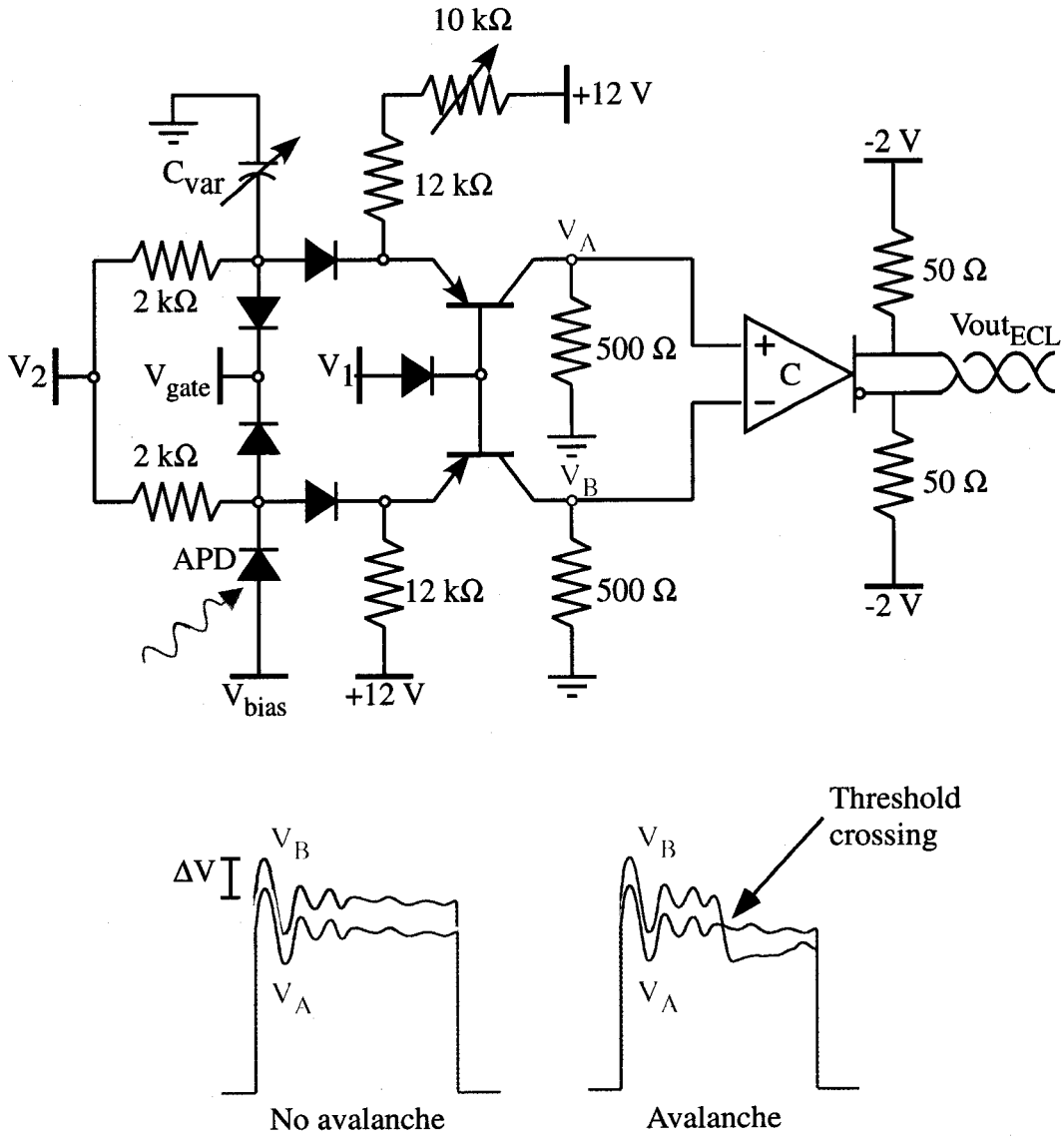


Figure 3.6: Schematic of the final Mark VI circuit which utilizes gated quenching. All the data discussed in this thesis were collected using this design. For our experiments, V_{gate} was a 0–5 V pulse, 100 ns in duration; V_{bias} was 22.5 V; V_1 was 5 V and V_2 was 7 V. This resulted in a 3.5–4 V excess voltage.

Chapter 4

THEORY OF APDS

Having given a brief introduction to the basics of APDs and the electronics necessary to use them, I'll now spend a bit more time discussing our model for their performance and the theory behind it. The first few sections serve as background for the reader and are the basis for the numerical simulations described in the last section. While studying the theory of electron movement in silicon, I frequently used Rieke (2003) as a resource. The mathematical development of Sections 4.1 and 4.2 is drawn from this text. Kindt (1999) was also a frequent resource that addressed the photon penetration and drift and diffusion properties of electrons in silicon as well as the lateral avalanche growth in APDs.

As discussed earlier, the APD can be thought of as having four distinct regions: an avalanche region where electrons have enough energy for impact ionization to initiate an avalanche; a drift region where a weaker electric field moves electrons toward the avalanche region—by-and-large at a terminal velocity on the order of 10^5 m/s (0.1 μm/ps) related to the strength of the electric field; a diffusion region where there is no electric field and electrons move isotropically in a random-walk fashion; and a guard ring region that limits the volume for potential avalanche initiating carriers. The avalanche region begins less than 0.5 microns below the surface and has a thickness of about 1 micron, the drift region is a few microns thick (depending on V_{excess}) and the diffusion region extends for several tens of microns.

As shown in Figure 3.2 an APD consists of essentially four layers. There is a 650 micron thick p substrate, a 25 micron π epitaxial layer, a roughly 0.5 micron p implant and a roughly 0.5 micron n layer. Recall that p denotes heavily receptor-site-doped

silicon, π is lightly receptor-site-doped silicon and n is heavily donor-doped silicon.

In this chapter we will only concern ourselves with the electron carriers. The holes are also able to move in the silicon, but they have a much lower mobility so all the "action" is dominated by the electron behavior.

When the APD is reverse-biased, there is a separation of charge that creates an electric field. In the outer guard ring region, there are many fewer receptor sites in the π "epilayer," so the electric field must penetrate deeper into the silicon in order to deplete the available receptors. In the active area where the implant exists there are a great many receptor sites so the depletion is concentrated—giving rise to a much stronger, localized electric field. As the bias voltage is increased, the magnitude of the electric field increases and continues to penetrate deeper into the silicon. Once the bias voltage reaches the breakdown voltage of the diode, the APD is capable of detecting photons.

The relative doping levels in the n layer and p implant determine the depth of the electric field when the diode reaches breakdown. According to Brian Aull, one of the main designers of the APDs, the diodes are designed so that the electric field just penetrates to the bottom of the p implant at the breakdown voltage. If the bias voltage continues to increase beyond breakdown, the electric field begins to extend into the π epilayer. Because there are fewer receptor sites in the π epilayer, the electric field rapidly penetrates deeper into the silicon, thus extending the drift region of the diode at the approximate rate of 1 μm per applied volt of bias. The magnitude of the electric field in the avalanche region also continues to increase as the bias voltage is increased. This has ramifications on the lateral avalanche propagation and temporal resolution of the APDs, which will be discussed in Section 4.3.

Table 4.1: Inverse absorption coefficient in silicon

Wavelength	z_o
532 nm	1.0 μm
668 nm	3.5 μm
786 nm	9.0 μm

4.1 Photon penetration in silicon

The depth that photons penetrate into the silicon before photo-converting falls off exponentially at different rates, depending on the wavelength of light. Longer wavelengths penetrate deeper into the silicon before photo-converting. The normalized photon flux at a depth, z is given by

$$P(z) = \frac{1}{z_o} \exp\left(-\frac{z}{z_o}\right). \tag{4.1}$$

The APDs were tested with 786 nm and 668 nm light and will eventually be used with 532 nm for APOLLO. Table 4.1 gives the inverse absorption coefficient, z_o, at the different wavelengths and Figure 4.1 shows how the photon absorption depends on depth for the three different wavelengths.

We have constructed a simple model for the performance of the APD. Photons that convert in the avalanche region may promptly initiate an avalanche. Photons that convert in the drift region will move toward the avalanche region at a speed on the order of 0.1 μm/ps. Since the drift region penetrates on the order of 1–5 microns below the avalanche region, there will be a delay of roughly 10–50 ps before an avalanche is initiated. Photons that convert deeper than the extent of the electric field can only reach the drift region and eventually initiate an avalanche by diffusion. This is a much slower process and contributes a long tail on the temporal distribution.

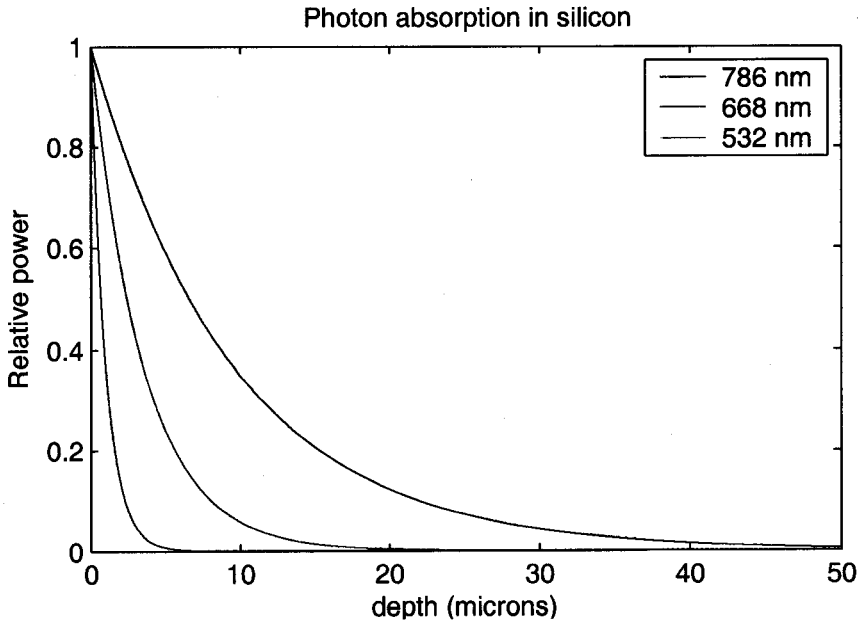


Figure 4.1: Photon absorption in silicon

4.2 Electron diffusion and drift

This section will begin with a derivation of the Einstein relation that can be used to determine the diffusion constant of electrons in silicon. This will be followed by descriptions of the effective thermal velocity and the mean free path.

Diffusion

Assume that there is an electron density of $n(z)$ at depth, z, and an electric field, $\mathcal{E}_z$, positive in the direction of increasing z. Electrons in the presence of the electric field experience a force per unit volume of

$$F_e(z) = +q\mathcal{E}_z n(z) \tag{4.2}$$

which for negatively charged electrons is directed toward the direction of decreasing z.

If there were no electric field present, the electrons would diffuse isotropically via a random walk process as a result of their thermal energy. This diffusion is driven by the osmotic pressure

$$P_o(z) = n(z)kT \tag{4.3}$$

where k is Boltzmann's constant and T is the temperature. This results in an osmotic force per unit volume of

$$F_o(z) = -\frac{\mathrm{d}P_o(z)}{\mathrm{d}z} = -kT\frac{\mathrm{d}n(z)}{\mathrm{d}z}. \tag{4.4}$$

Under equilibrium conditions, there is no net force on the electrons so

$$F_e(z) + F_o(z) = 0 \tag{4.5}$$

resulting in the differential equation

$$\frac{\mathrm{d}n(z)}{\mathrm{d}z} + \frac{q\mathcal{E}_z n(z)}{kT} = 0 \tag{4.6}$$

with a solution of

$$n(z) = n(0)\exp\left(-\frac{q\mathcal{E}_z z}{kT}\right). \tag{4.7}$$

Consider now the flux of particles as a result of these two processes. The flux due to the applied electric field is

$$S_e(z) = - < v_z > n(z) = -\mu_e \mathcal{E}_z n(z) \tag{4.8}$$

where $< v_z >$ is the terminal velocity of electrons and μ_e is the electron mobility. The osmotic flux is

$$S_o(z) = -D_e\frac{\mathrm{d}n(z)}{\mathrm{d}z} \tag{4.9}$$

where D_e is the diffusion constant of electrons in silicon. Again under equilibrium conditions the net flux is zero so that

$$S_e(z) + S_o(z) = 0 \tag{4.10}$$

resulting in the differential equation

$$\frac{\mathrm{d}n(z)}{\mathrm{d}z} + \frac{\mu_e \mathcal{E}_z n(z)}{D_e} = 0 \tag{4.11}$$

with a solution of

$$n(z) = n(0) \exp\left(-\frac{\mu_e \mathcal{E}_z z}{D_e}\right). \tag{4.12}$$

Comparing Equations 4.7 and 4.12 one arrives at the Einstein relation

$$\frac{\mu_e}{D_e} = \frac{q}{kT}. \tag{4.13}$$

The electron mobility, μ_e, in silicon is 1.35 $\times 10^3$ cm^2/V·s (Rieke, 2003) resulting in a diffusion constant, D_e, of 0.0033 m^2/s at room temperature. This diffusion constant can be thought of as the rate of increase of the surface area of a sphere likely containing the diffusing electron. The diffusion constant plays an important role in the determination of the mean free path and the numerical simulations of diffusion in the APDs.

Thermal velocity

The thermal velocity of the electrons also plays a role in the performance of the APD. In the absence of an electric field electrons have thermal energy

$$E = \frac{3}{2}kT = \frac{1}{2}m^* v_{eff}^2 \tag{4.14}$$

where $m^* = 0.26m_e$ is the effective mass of conduction electrons in silicon. For electrons at room temperature this results in an effective thermal velocity of 2.3×10^5 m/s in silicon.

Mean free path

The diffusion constant and thermal velocity can now be combined to determine the mean free path (MFP) of an electron in silicon. Consider the case of a one-dimensional

random walk. The distance traveled by a particle can be determined in terms of the diffusion constant, D_e, or the MFP, l_o, of the particle

$$Distance = \sqrt{D_e \tau} = l_o \sqrt{N} \tag{4.15}$$

where τ is the time, N is the number of steps and $\tau_o = \tau/N$ is the mean free time. The (thermal) velocity of this particle is

$$v_{th} = \frac{N l_o}{\tau} = \frac{l_o}{\tau_o}. \tag{4.16}$$

By combining Equations 4.15 and 4.16 we can arrive at a relationship between the MFP and the diffusion constant

$$l_o = \frac{D_e}{v_{th}}. \tag{4.17}$$

The result is a one-dimensional MFP of 14.3 nm for electrons in silicon at room temperature. For the case of the APDs, the diffusion is isotropic in three dimensions so a factor of $\sqrt{3}$ must be included resulting in a three-dimensional MFP of 24.8 nm. The electron travels at a speed v_{th} over this distance which gives an associated τ_o of 0.12 ps.

4.3 Lateral growth of avalanche

Once an avalanche is initiated, the current through the diode will only reach its maximum current when the avalanche fills the entire area of the element. This lateral growth can happen by one of two methods: photon emission and lateral diffusion.

When electrons have enough energy to impact ionize, excess energy could go into isotropic photon emission; however, photon emission in silicon is quite low. Lacaita et al. (1988, 1990, 1993) discuss the impact of photon emission in the lateral growth of an avalanche across an element and found that one photon was emitted for every 10^6 charge carriers crossing the junction. For larger APDs, on the order of 100s of microns in diameter, photon emission plays a role in the lateral growth of the avalanche. For

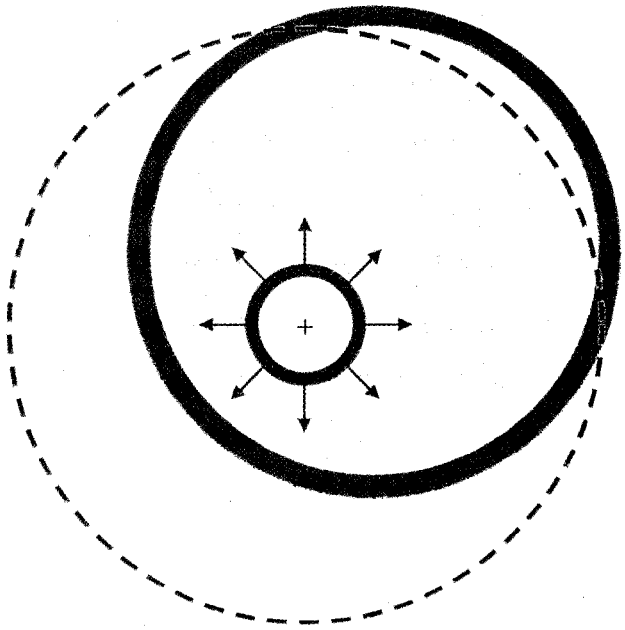


Figure 4.2: Lateral growth of an avalanche. The circles of increasing size represent the temporal growth of the extent of the avalanching region in the APD.

elements on the order of 10s of microns in diameter, however, the lateral growth of the avalanche is dominated by the lateral diffusion of charge carriers.

Lacaita et al. (1990) used a 140 μm $\times$ 14 μm rectangular element illuminated with a 15 μm laser spot to study how fast the avalanche propagated across the element. They found that the one-dimensional propagation speed, v_p, was directly proportional to the steady state current $I_d = V_{excess}/R_d$.

Consider the case where a photon has initiated an avalanche at the position indicated by a "+" symbol in Figure 4.2. After a short period of time (several picoseconds) the avalanche will have spread over a small region of the element. In the central portion (shaded light green) subsequent impact ionizations have no impact on the lateral growth of the avalanche; however, this is not the case for the edges highlighted in bright green. I will call these two regions the avalanching region and the edge plasma. When impact ionization occurs, the initial momentum vector of the freed

carriers in the plasma is isotropic perpendicular to the applied electric field. In other words, newly freed charge carriers in the edge plasma may have an equal likelihood of heading in towards the avalanching region or out towards the inactive area of the element. Due to the high electric field, the lateral diffusion constant and horizontal MFP of these electrons is reduced compared to those calculated in Section 4.2.

If an electron were to travel directly across the diameter of the 20 micron element it would take about 800 low-electric-field MFP steps (or 100 ps). It would take about 1200 steps (or 145 ps) to cross the 30 micron device. This suggests a temporal difference in achieving full avalanche of at least 50 ps between the center and edge for the 20 micron device and about 75 ps for the 30 micron device. This is the lower limit for the avalanche propagation across each element. As mentioned above, the mean free path of electrons will be shorter in the avalanche region as a result of the strong electric field. In addition, this argument assumed that at each impact ionization event, the freed charge carrier moved exactly perpendicular to the edge plasma. It is safe to assume that there would be an additional efficiency term that would represent what fraction of the MFP the charge traveled into the inactive area of the element before impact ionizing. As the number of charge carriers participating in the avalanche increases, the efficiency factor approaches unity.

As the excess voltage is increased, the electric field within the avalanche region is also increased. This improves the efficiency of the impact ionization and increases the number of charge carriers in the junction. It is this mechanism that accounts for the improved temporal response of the APDs when the excess voltage is increased.

4.4 Numerical simulations

A numerical simulation was developed to explore the effect of the diffusion region on the APD detection efficiency and temporal response. An electron was placed at a depth z below the drift region, along the central axis of the detector. It was then

allowed to move a distance of l_o in any direction. This was achieved using two random numbers, c_1 (cosine of the polar angle) which is uniformly distributed between -1 and 1 and c_2 (azimuthal angle) which is distributed between 0 and 2π. For each step

$$x = x_o + l_o\sqrt{1 - c_1^2}\cos(c_2), \tag{4.18}$$

$$y = y_o + l_o\sqrt{1 - c_1^2}\sin(c_2), \tag{4.19}$$

and

$$z = z_o + l_o c_1, \tag{4.20}$$

thus achieving random, isotropic displacements of length l_o.

If $z < 0$, the electron entered the drift region and was considered detected. All of the simulations I conducted used a cylindrical drift region with a flat bottom. If $\sqrt{x^2 + y^2}$ was greater than the radius of the detector, then the electron was considered lost to the guard ring. I only considered a purely cylindrical guard ring. The electron was allowed to make up to one million steps until it was detected or lost to the guard ring, corresponding to roughly 120 ns.

4.4.1 Spatial results

Figure 4.3 shows the results of the simulation for the 20 micron diameter device. The simulation was repeated 10,000 times at each depth from 1 to 25 microns below the drift region. The results indicate that very few electrons (<5%) continue diffusing after 120 ns, but are instead detected or intercepted by the guard ring. Electrons created in the first four microns are more likely to be detected, while those created deeper than four microns are more likely to encounter the guard ring.

Figure 4.4 again shows the detected photons as a function of depth but with error bars (based on Poisson statistics from the simulation) included. The data are well represented by a cubic function until a depth of 24 microns. Since the number of detected electrons from that depth represent less than 0.25% of those incident, the

detection probability was set equal to zero for depths greater than 24 microns. This functional form was used to determine the number of detections in the "core" versus the "tail" of the temporal response that will be discussed in Section 5.6.

The diffusion simulations were repeated for the 30 micron device and the results are shown in Figure 4.5. With a larger diameter device, a significant number ($\sim 30\%$) of the electrons are still diffusing after 120 ns. This is a six-fold increase over the 20 micron device. The 50% cutoff for detected photons also extends deeper into the silicon — nearly twice as deep as the 20 micron device. This behavior is likely a strong function of any taper to the guard ring field. If the diffusion region were to be conical instead of cylindrical, we would expect to see more guard ring captures at larger depths at the expense of those still diffusing.

Again the number of detected electrons as a function of depth is well-described by a cubic function as shown in Figure 4.6. This function, like its counterpart for the 20 micron device, will be used to estimate the back of the depletion layer based on the experimental data that will be discussed in Section 5.6.

4.4.2 Temporal results

The nature of the simulations also allows us to explore the effect of the diffusion region on the temporal response of the APDs. When an electron was detected (crossed $z = 0$), the elapsed time since its creation was also recorded. From Section 4.1 we know that the number of electrons created at a particular depth depends on the wavelength of incident light. We can use this as a weighting factor in determining the overall temporal response of the diffusion tail.

Figure 4.7 shows the temporal results of the diffusion simulations for the 20 micron device illuminated with 668 nm light. The conversion likelihood versus depth at this wavelength was used to scale the contributions from the different depths. Each different color represents the temporal response from a different depth from 1–5 microns. As the electrons start out at deeper and deeper levels the peak of the distribution

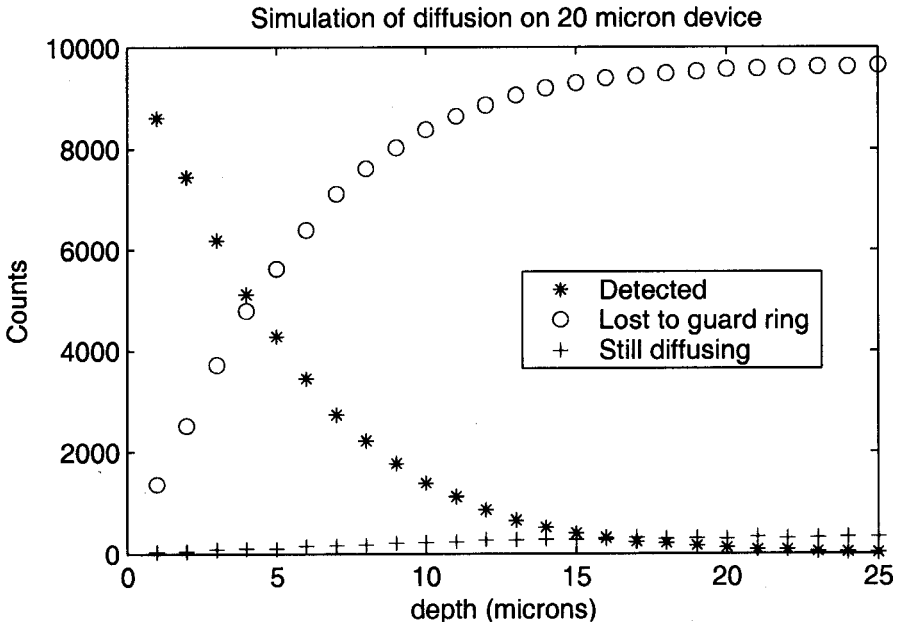


Figure 4.3: Results of diffusion simulation in the 20 micron device. The blue $*$ represent electrons that reached the drift region, green $\circ$ represent electrons which intercepted the guard ring and the red $+$ represent electrons which are still diffusing after 120 ns.

becomes much less prominent, and the location of the peak begins to move toward longer times. Figure 4.8 is a similar plot that is weighted based on the 786 nm penetration depth. Figure 4.9 shows the cumulative temporal response integrated over depth for both wavelengths. As is expected, the 668 nm data fall off faster than the 786 nm data since the 668 nm photons do not penetrate as far into the silicon.

The simulations were also used to determine the temporal response of the 30 micron device. Figures 4.10 and 4.11 show the temporal response at different depths within the silicon where the data have been weighted to reflect the penetration depth at the two different wavelengths. Figure 4.12 shows the depth integrated temporal response at each wavelength.

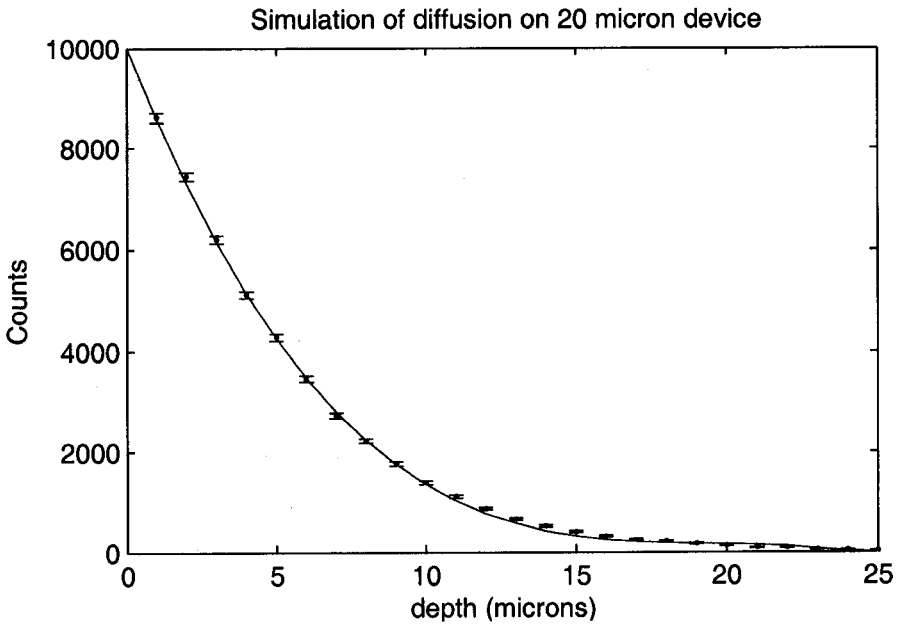


Figure 4.4: Functional fit to detected electrons in the 20 micron device. The blue points represent the simulated data and are well represented by a cubic fitting function (red line).

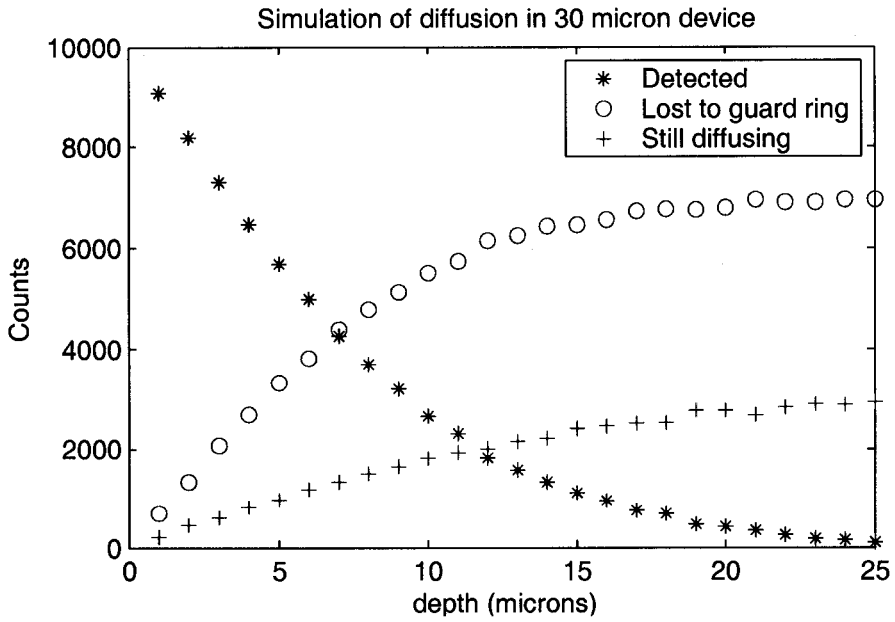


Figure 4.5: Results of diffusion simulation in the 30 micron device. The blue $*$ represent electrons that reached the drift region, green $\circ$ represent electrons which intercepted the guard ring and the red $+$ represent electrons which are still diffusing after 120 ns.

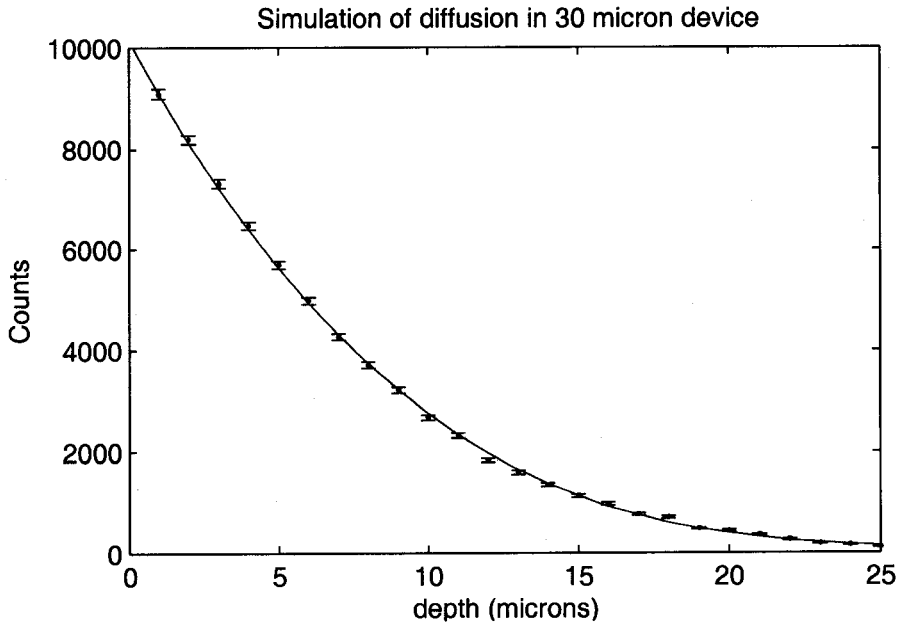


Figure 4.6: Functional fit to detected electrons in the 30 micron device. The blue points represent the simulated data and are well represented by a cubic fitting function (red line).

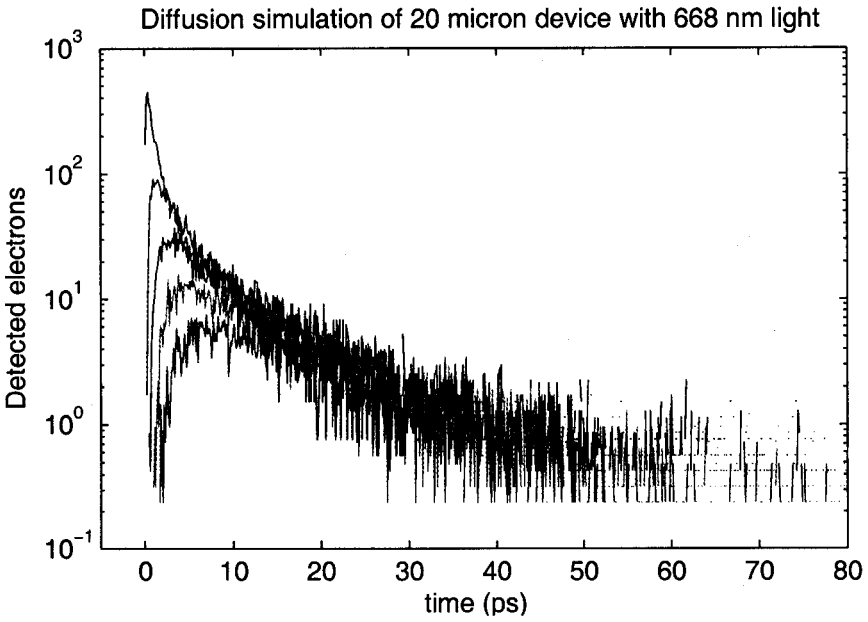


Figure 4.7: Temporal results of the diffusion simulation on 20 micron device. Each color represents the temporal response at a different depth. The magnitude of the response has been weighted by the number of electrons produced at each depth as determined by Equation 4.1 for a wavelength of 668 nm.

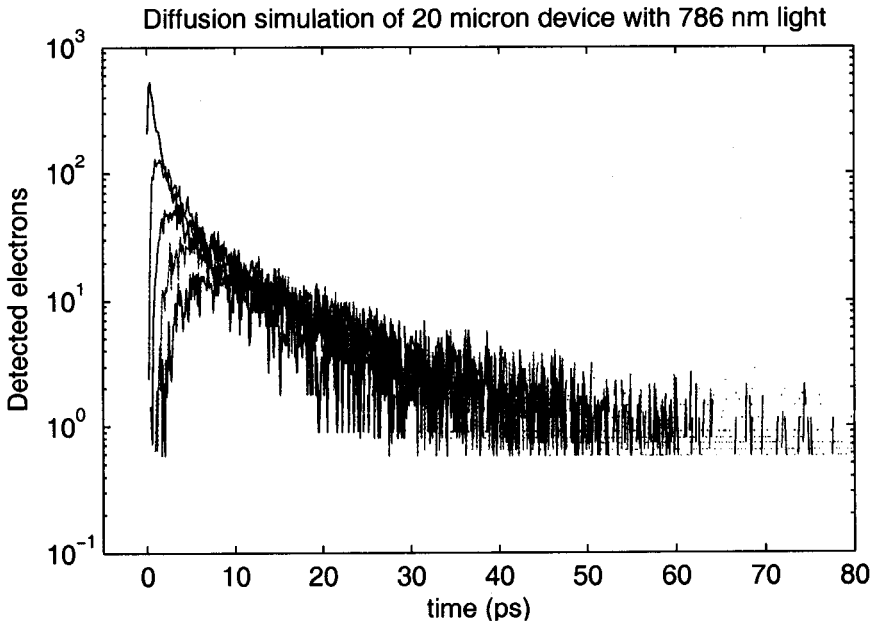


Figure 4.8: Temporal results of the diffusion simulation on 20 micron device. Each color represents the temporal response at a different depth. The magnitude of the response has been weighted by the number of electrons produced at each depth as determined by Equation 4.1 for a wavelength of 786 nm.

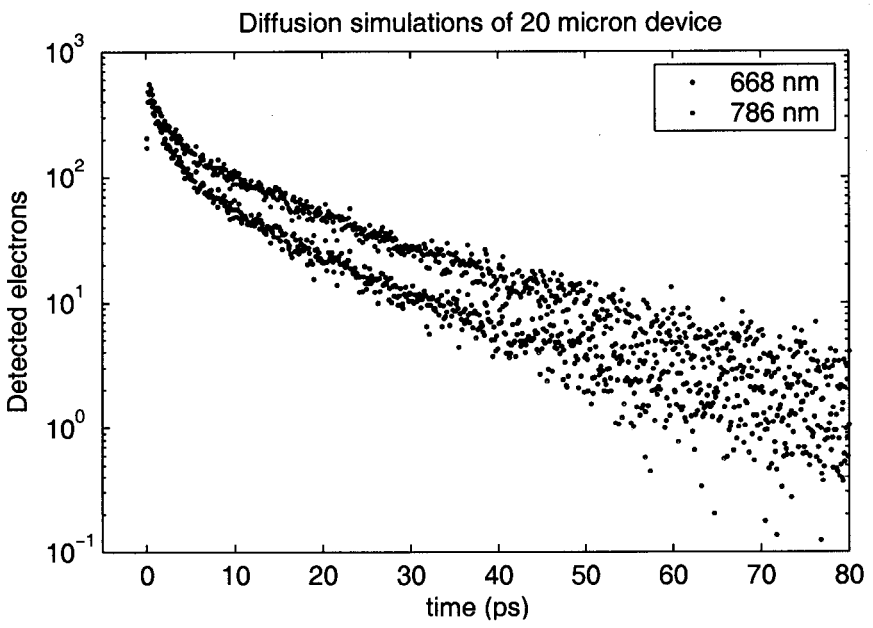


Figure 4.9: Depth-integrated temporal results of the diffusion simulation; red indicates the 786 nm response and blue the 668 nm response.

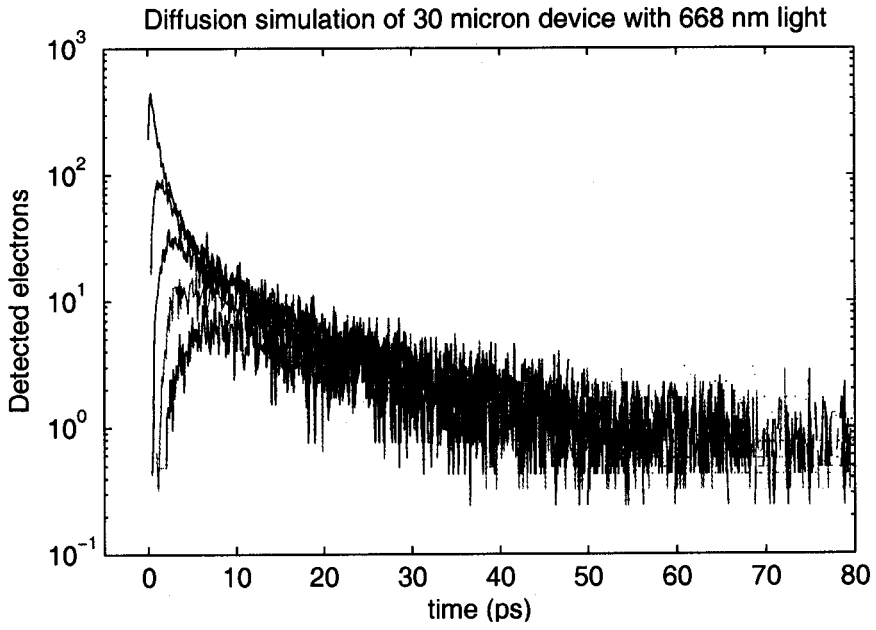


Figure 4.10: Temporal results of the diffusion simulation on the 30 micron device. Each color represents the temporal response at a different depth. The magnitude of the response has been weighted by the number of electrons produced at each depth as determined by Equation 4.1 for a wavelength of 668 nm.

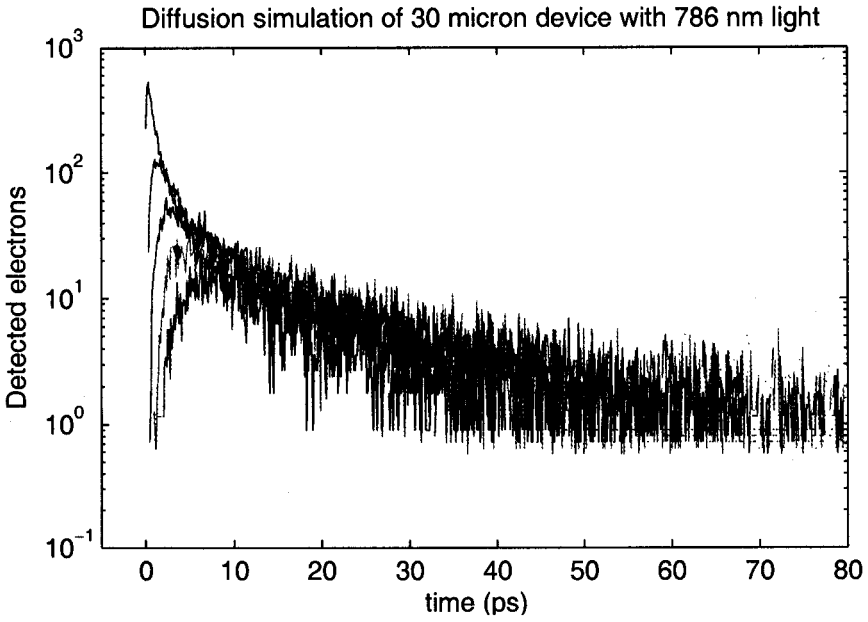


Figure 4.11: Temporal results of the diffusion simulation on the 30 micron device. Each color represents the temporal response at a different depth. The magnitude of the response has been weighted by the number of electrons produced at each depth as determined by Equation 4.1 for a wavelength of 786 nm.

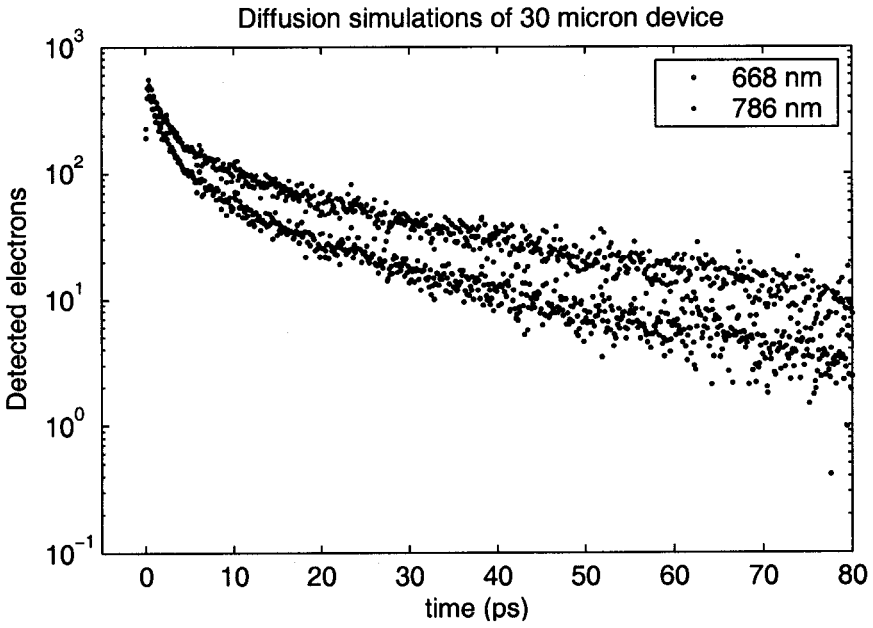


Figure 4.12: Depth-integrated temporal results of the diffusion simulation for 30 micron device; red indicates the 786 nm response and blue the 668 nm response.

Chapter 5

APD CHARACTERIZATION

5.1 Discussion of first-photon arrival distribution

When characterizing the performance of single-photon detectors and timing electronics, it is necessary to understand the importance of signal rate. Some aspects of the overall timing response will change significantly with increased photon rate while others will not. For instance, the readout and timing electronics do not care if they are triggered by one photon or a thousand, a photon at the nose of a distribution or one at the tail end. On the other hand, the apparent distribution of the laser pulse and detector will change with signal strength.

Because APDs are only capable of detecting the *first* photon, strong signals mean that instead of sampling the complete distribution, one ends up effectively sampling only the leading edge. If the signal is strong, the likelihood of a photon in the tail of the laser distribution being the first photon detected is very low. The detector response follows along the same line. With a strong signal, the first photon detected is the earliest arrival but all information about the total distribution is lost.

Figure 5.1 illustrates this phenomena. The blue line shows the probability distribution of the PLP-02 lasing at 786 nm. If less than one photon is detected per pulse, the data would follow this distribution. If the incident power is increased, only the first photon will be detected and the distribution changes. If five photons are expected per pulse, all trace of the second after-pulse is gone and the distribution is a Gaussian with significantly less temporal jitter than the single photon distribution. As the signal continues to increase, the leading edge of the distribution becomes more

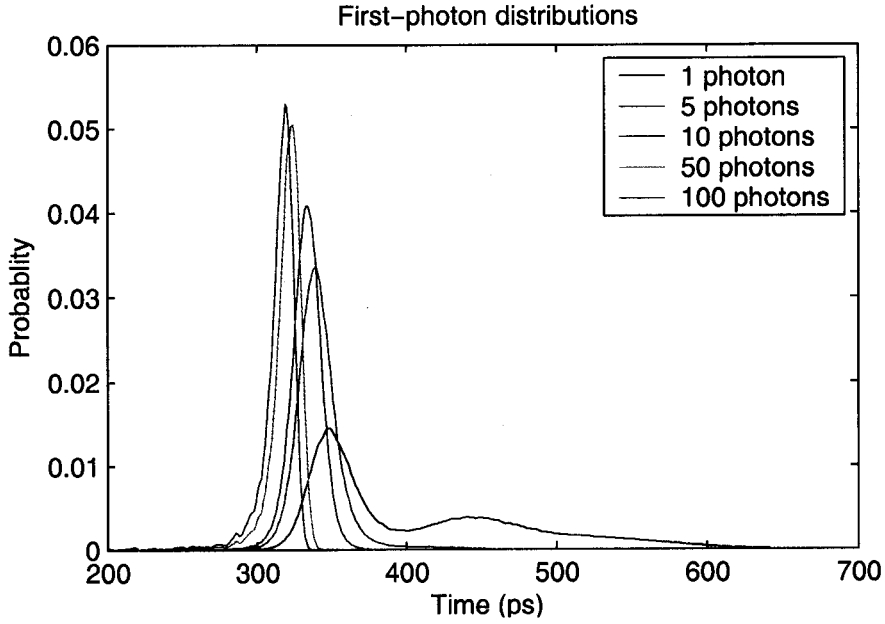


Figure 5.1: First-photon biasing of the PLP-01 laser at 786 nm

ragged as a direct result of noise on the leading edge of the laser pulse. This causes the temporal jitter to increase, meaning that there is an optimum signal strength, resulting in a minimum amount of temporal jitter.

Detected photons will obey Poisson statistics because they are discrete, uncorrelated events. This information can be used to correct the data for any first-photon biases. According to Poisson statistics,

$$P_n = \frac{\mu^n}{n!} \exp(-\mu), \tag{5.1}$$

where P_n is the probability of observing n events, in this case the detection of a photon, and μ is the expected average number of events.

Consider the experiment where at least one photon is detected one out of every four gates or 25% of the time. In other words, no photons were detected 75% of the time. In that case

$$\sum_{n=1}^{\infty} P_n = 0.25 \tag{5.2}$$

and

$$P_0 = 1 - 0.25 = \exp(-\mu) \quad \Longrightarrow \quad \mu = 0.28768. \tag{5.3}$$

It then follows that

$$P_1 = \mu \exp(-\mu)\Big|_{\mu=0.28768} = 0.21576, \tag{5.4}$$

$$P_2 = \frac{\mu^2}{2!} \exp(-\mu)\Big|_{\mu=0.28768} = 0.03103, \tag{5.5}$$

and

$$P_3 = \frac{\mu^3}{3!} \exp(-\mu)\Big|_{\mu=0.28768} = 0.00298. \tag{5.6}$$

Even though only one event was detected for every four gates, the statistics show that nearly 14% $(1.0 - 0.21576/0.25)$ of those detections were from two or more photons. For many of our results, we are interested in the *total* photon flux or the number of expected photons and not just the *detected* flux. Looking at the example given above, consider that the data were collected over 40,000 gates. In that example, photons were detected 25% of the time (or in 10,000 gates) but because of the statistics we know that the actual number of incident photons is closer to $\mu \times 40,000$ or 11,500 because some positive detections resulted from multi-photon events.

First-photon biasing also gives us a way to test the timing response of just the electronics. If the APDs are bombarded with many photons, then only the earliest arriving photon, with the smallest penetration depth from the leading edge of the laser, will be detected. On the other hand, the timing electronics contribution remains the same, so the temporal response of the bombarded APDs is dominated by the electronics jitter. As a result, flooding the APDs with pulsed light by removing the neutral density filters gives us a way to examine the contribution of just the readout and timing electronics. Figure 5.2 shows the best strong signal returns we obtained with the electronics. In this case the RMS jitter was 24 ps.

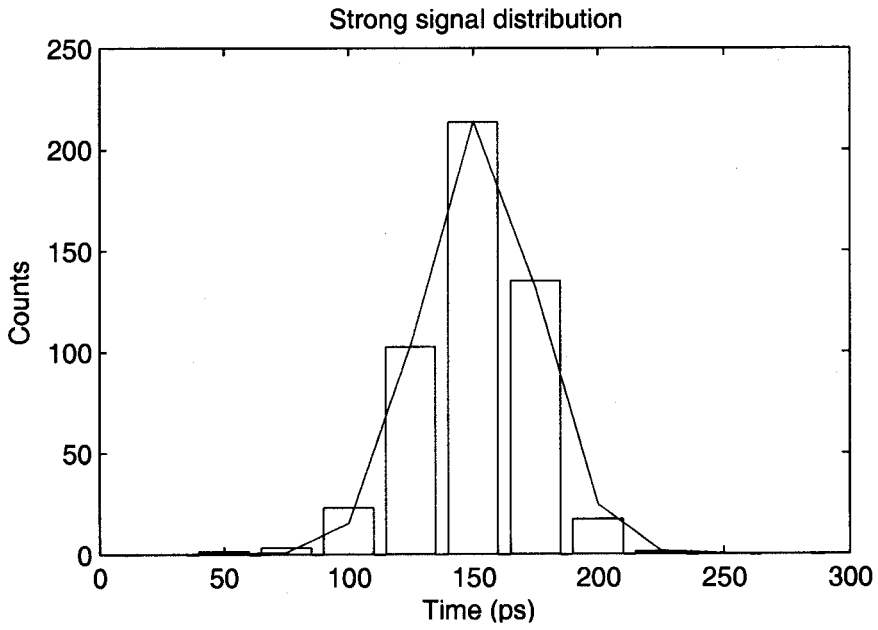


Figure 5.2: Strong signal response of the TDC and timing electronics

5.2 Experimental setup

Time resolution data were taken using a Hamamatsu PLP-02 short-pulsed laser at 786 nm or a Hamamatsu PLP-01 lasing at 668 nm, a single APD element with associated readout electronics and a Phillips Scientific 7186H Time to Digital Converter (TDC). A trigger signal from the laser biased the APD above breakdown for 140 ns. There was a user specified delay of 20–100 ns between the trigger signal and photon emission.

Figure 5.3 shows the layout of the experiment. The APD was placed approximately 40 cm from the laser and a microscope objective was used to focus the laser light onto a small spot with a FWHM of less than 5 microns. Filters with an effective neutral density of 4.0 (10^4 reduction) at 786 nm, and 3.0 (10^3 reduction) at 668 nm, were placed in the diverging laser beam so that fewer than one photon per pulse arrived at the detector. The microscope objective was mounted on an XYZ stage. The depth of focus was adjusted by a manual translation stage and the position of the beam on

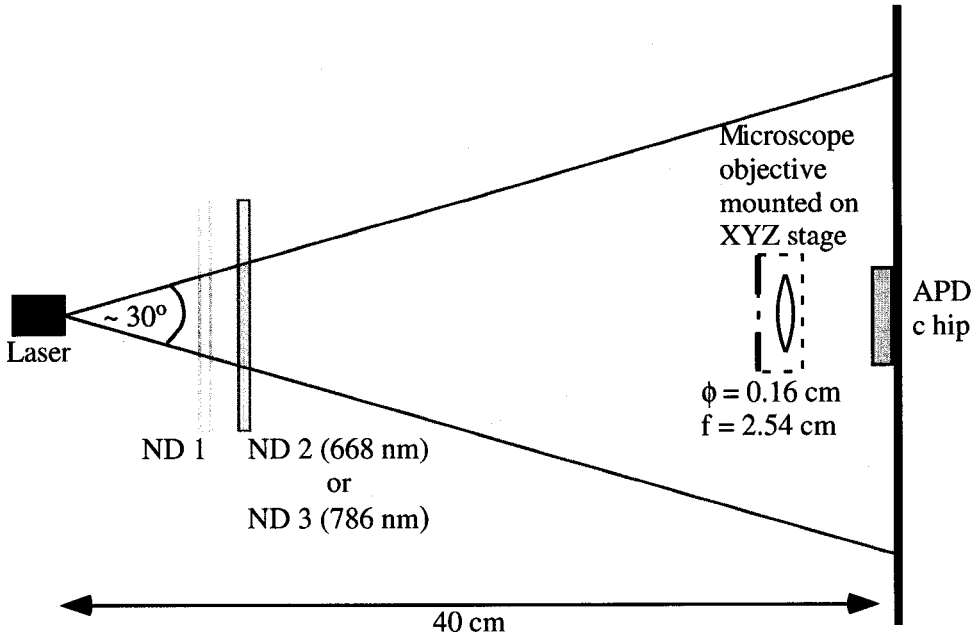


Figure 5.3: Experimental setup

the APD was controlled by two New Focus Picomotors.

The 75 mV signals from the APD readout electronics were compared with a mirror channel having the same transients and resistive-capacitive characteristics. If the APD avalanched then the comparator output acted as the *start* signal for the TDC. The end of the gate served as the *stop* signal. The TDC measured the elapsed time between *start* and *stop* (comparator trip relative to the end of the gate) in 25 ps bins. After many pulses, or alternatively many gates, it was possible to build up a histogram of the temporal response of the APD. Figure 5.4 shows the histogram built up over 43.2 million pulses. The results in Figure 5.4 reflect the convolution of the APD response, the pulse shape of the PLP-02 lasing at 786 nm and the temporal jitter of the electronics and TDC. The full temporal response can be characterized by a Gaussian core with a FWHM of 250 ps and a modified exponential tail that decays over tens of nanoseconds.

Two different types of data will be discussed in this chapter: element-centered

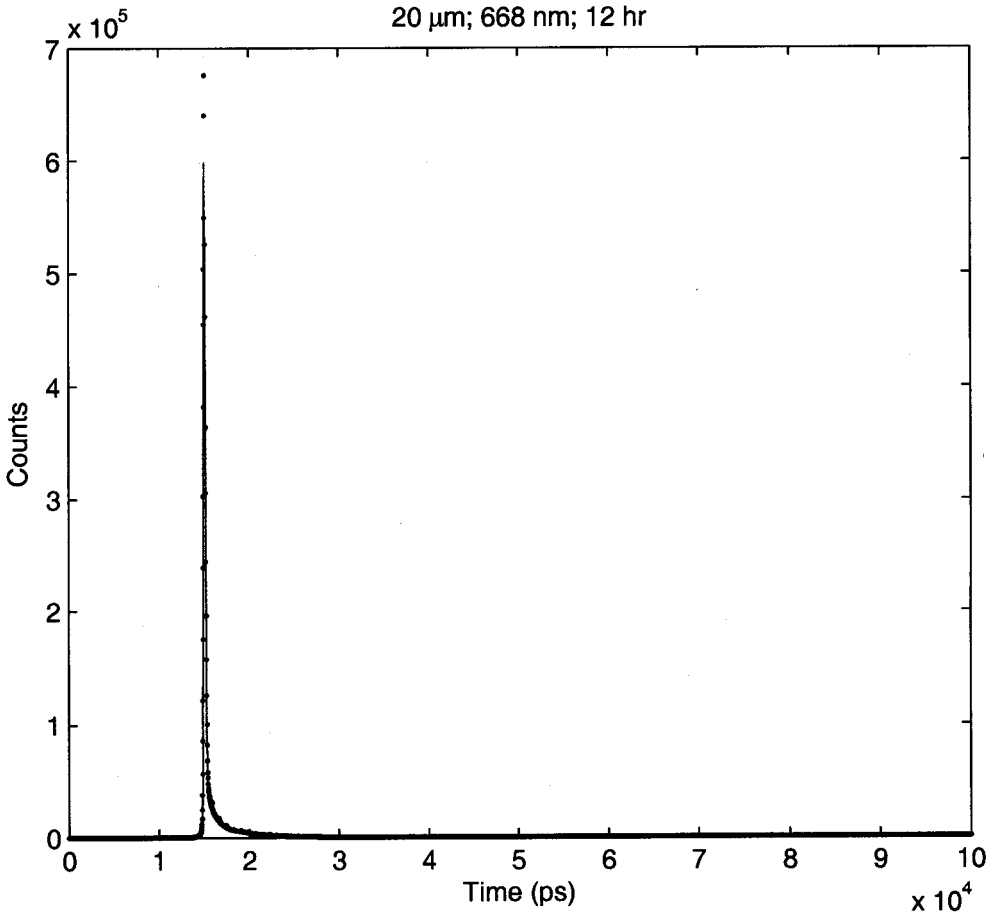


Figure 5.4: Temporal response of 20 μm device at 668 nm during a 12 hour exposure

long exposures and spatially-scanned short exposures. The element-centered, long exposures began by centering the laser beam within the element by determining the location with the highest count rate. Using the variation of the Picomotors and spatial dependence of the detection efficiency, it was estimated that the element centered data were within 2.5 microns of the center of the element. Once the laser beam was centered, the detector was exposed to 7.2 million (or 43.2 million) pulses over a 2 hour (or 12 hour) period. A photon was detected roughly 25% of the time. The spatially-scanned, short exposures involved scanning the laser beam across the element in one micron steps using the Picomotors. At each position, the APD was exposed to 40,000 pulses. Near the center of the element, a photon was detected roughly 25% of the time. The scans were conducted in only one dimension, but were within 2.5 microns of center in the orthogonal direction. As a result, the length of the chord scanned by the laser pulse was greater than 96% of the diameter of the element.

The experimental results will be separated out into different sections: one for the purely temporal results, one for the spatial dependence of the flux, one for the spatial dependence of the temporal parameters and one for the distribution of power.

5.3 Temporal response

- Temporal profiles are reasonably well described by a Gaussian core and a pseudo-exponential, diffusion-driven tail.

- The Gaussian core has an RMS width on the order of 100–125 ps. This is a convolution of the laser pulse, detector response, electronics and TDC. This response is independent of wavelength.

- Once the other contributions have been estimated, an upper limit of 75 ps is placed on the detector response width.

- The diffusion tail falls off faster for smaller devices and shorter wavelengths.

Table 5.1: Description of the canonical datasets used to examine the temporal response of the APDs

Device	Wavelength	Time	Gates	Observed counts	Corrected counts
20 micron	668 nm	12 hr	43.2e6	9.098e6	10.218e6
20 micron	668 nm	2 hr	7.2e6	1.511e6	1.696e6
20 micron	786 nm	12 hr	43.2e6	10.059e6	11.450e6
30 micron	668 nm	12 hr	43.2e6	10.249e6	11.699e6
30 micron	668 nm	2 hr	7.2e6	2.294e6	2.774e6
30 micron	786 nm	2 hr	7.2e6	2.767e6	3.491e6

The 2 and 12 hour-long exposures provided high signal-to-noise datasets to examine the temporal performance of the APDs. Six different datasets, consisting of different device elements illuminated with different wavelength light for different exposure periods, were used to examine the temporal response of the APDs. Each dataset has been corrected for first-photon bias to determine the incident flux. Table 5.1 summarizes the six datasets that will be discussed in this section.

5.3.1 First-photon correction of temporal response

In section 5.1 the biases introduced by detecting only the first photon were discussed. If all we want to do is determine the actual number of incident photons, we only need to multiply μ, the expected number of photons per gate, by the total number of gates. However, this method does not give any way to determine *when* the additional photons landed within the temporal distribution. To find this, we must use a slightly different technique. We know that a second photon is more likely to occur later in the gate, so the first few bins up to and including the core should have very little correction, while photons from the tail are much more likely to be missed since so

many photons have been detected ahead of them. In other words, we need a method for boosting the power in the tail while leaving the core relatively untouched. This can be achieved by using the cumulative probability distribution of the data.

Figure 5.5 shows the normalized cumulative probability distribution for the data shown in Figure 5.4. From this we can determine the likelihood that a photon has been detected. For instance, at 1×10^4 ps there is very little likelihood of already detecting a photon, while at 1.5×10^4 ps there is a 20% chance of already detecting a photon and by 5×10^4 ps there is a 99% chance that a photon has already been detected.

The corrected number of counts in each bin can be found by

$$N_{corrected} = (1 + \alpha P(\tau))N \tag{5.7}$$

where N is the number of counts in each bin,

$$P(\tau) = \int_{-\infty}^{\tau} p(t)\,\mathrm{d}t \tag{5.8}$$

is the cumulative probability distribution, $p(t)$ is the probability density function and α is a constant that is chosen so that $N_{corrected}/N_{gates}$ is equal to μ, the expected photon arrival rate.

Figure 5.6 shows the raw and corrected data for the 20 micron device illuminated with 668 nm light for 12 hours. There is no difference over the first 1×10^4 ps while the number of photons in the tail is increased slightly.

5.3.2 20 micron device

Figure 5.4 shows the complete temporal response of the 20 micron device illuminated with 668 nm light for 12 hours. It can be characterized by a Gaussian core with a FWHM of 250 ps and a modified exponential tail that lasts for several tens of nanoseconds. It was found that the temporal response could be fit with the following

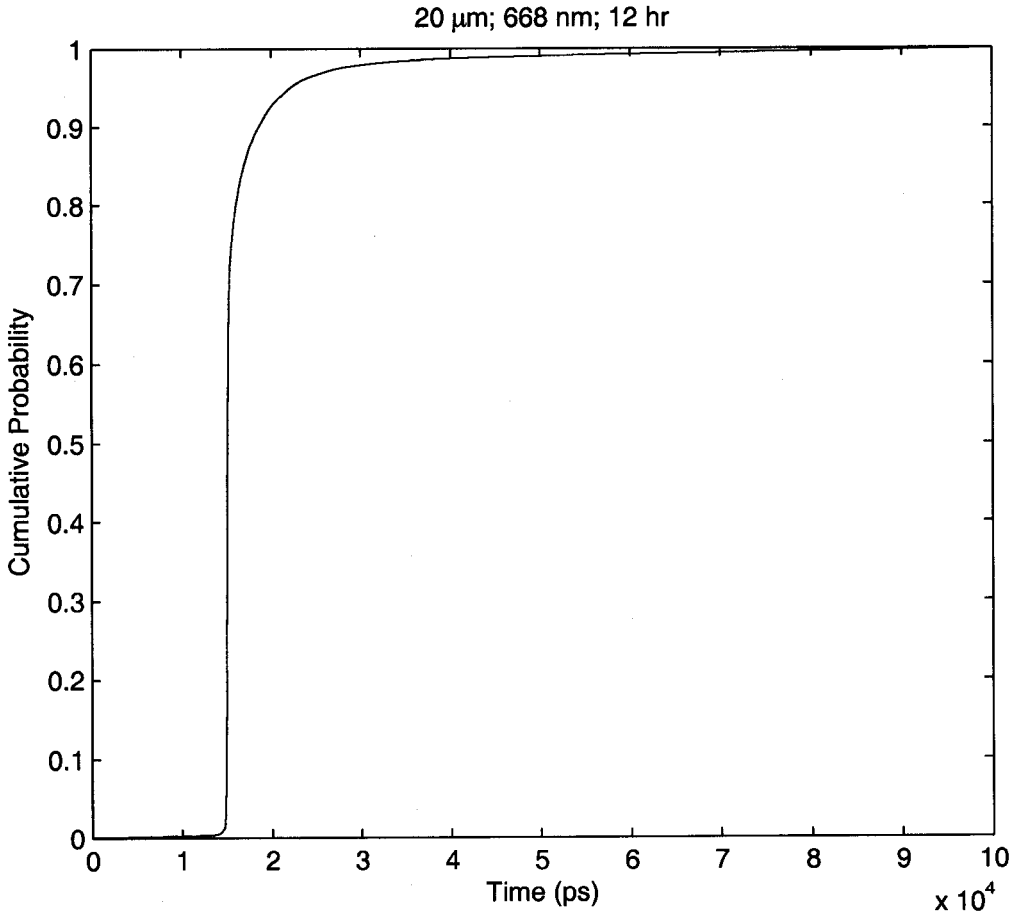


Figure 5.5: Normalized cumulative probability distribution of 20 μm device at 668 nm during a 12 hour exposure

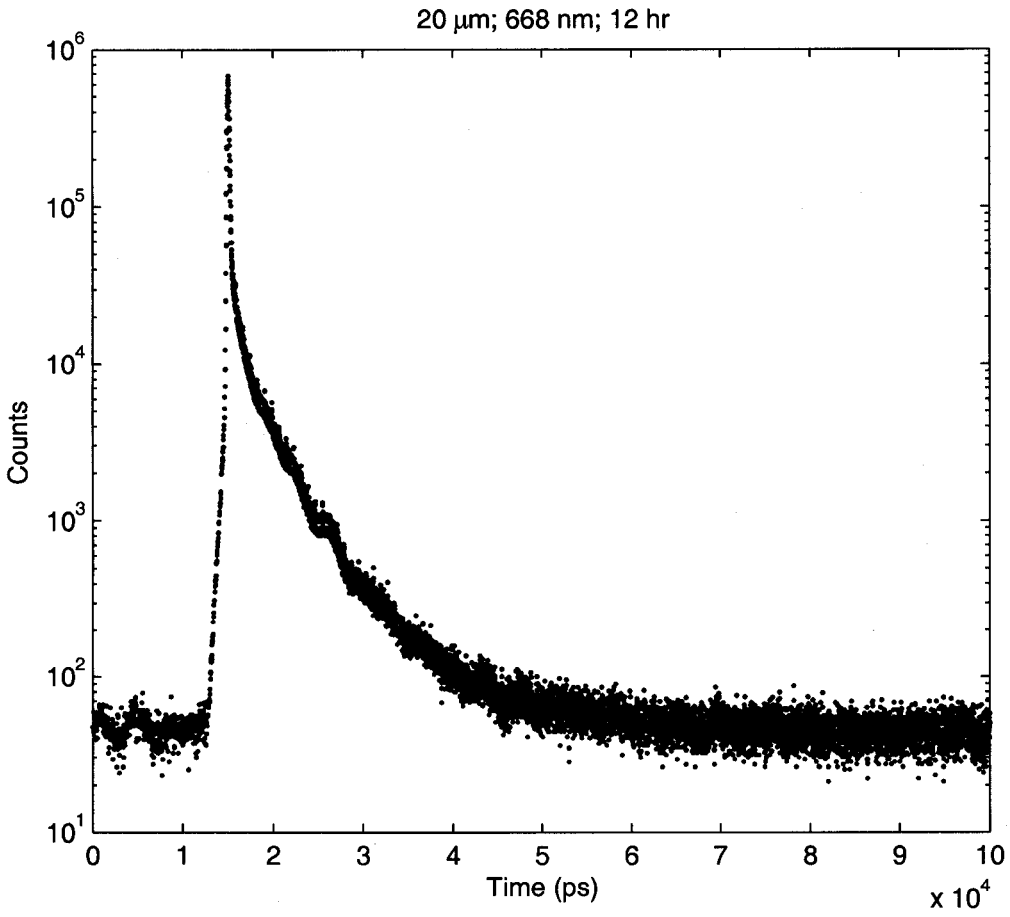


Figure 5.6: Raw (red) and corrected (blue) temporal distributions of 20 μm device at 668 nm during a 12 hour exposure

function

$$\begin{aligned} F(t) &= A\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] \\ &\quad + B\left(1-\exp\left[-\frac{t-(\mu+d)}{t_o}\right]\right)\exp\left[-\left(\frac{t-(\mu+d)}{\tau}\right)^{0.5}\right] \\ &\quad + C \end{aligned} \tag{5.9}$$

where A is the Gaussian amplitude in counts, t is the time in picoseconds, μ is the Gaussian mean in picoseconds, σ is the Gaussian spread in picoseconds, B is the amplitude of the exponential in counts, d is the offset between the core and the tail (fixed to 50 ps), t_o is the smoothing time constant for the transition between Gaussian and tail (fixed to 25 ps), τ is the exponential decay factor in bins, and C is the background in counts. The three terms represent the core, tail and background respectively. The $(1 - \exp[-t/t_o])$ factor smoothes the "turn on" of the pseudo-exponential tail term so that it is not so abrupt. As a result of the smoothing function the tail term takes approximately 125 ps to reach its peak height before it starts to decay.

Figures 5.7 and 5.8 illustrate that the fitting function, depicted by the blue curve, does a reasonable job of describing both the core and the tail of the distribution of arrival times.

Figures 5.9 and 5.10 show the first-photon corrected data and the best fit to the data for the 20 micron device illuminated with 668 nm light for 2 hours.

Figures 5.11 and 5.12 illustrate that the fitting function, depicted by the blue curve, does a reasonable job of describing both the core and the tail of the distribution of arrival times.

Table 5.2 gives the temporal fit parameters for the 20 micron device. The reported errors are the formal errors returned by the least squares fitting routine. I will not report χ^2 values for these fits for reasons discussed in section 5.3.4, but the fitting routine uses the square of the residuals to determine the uncertainties in each

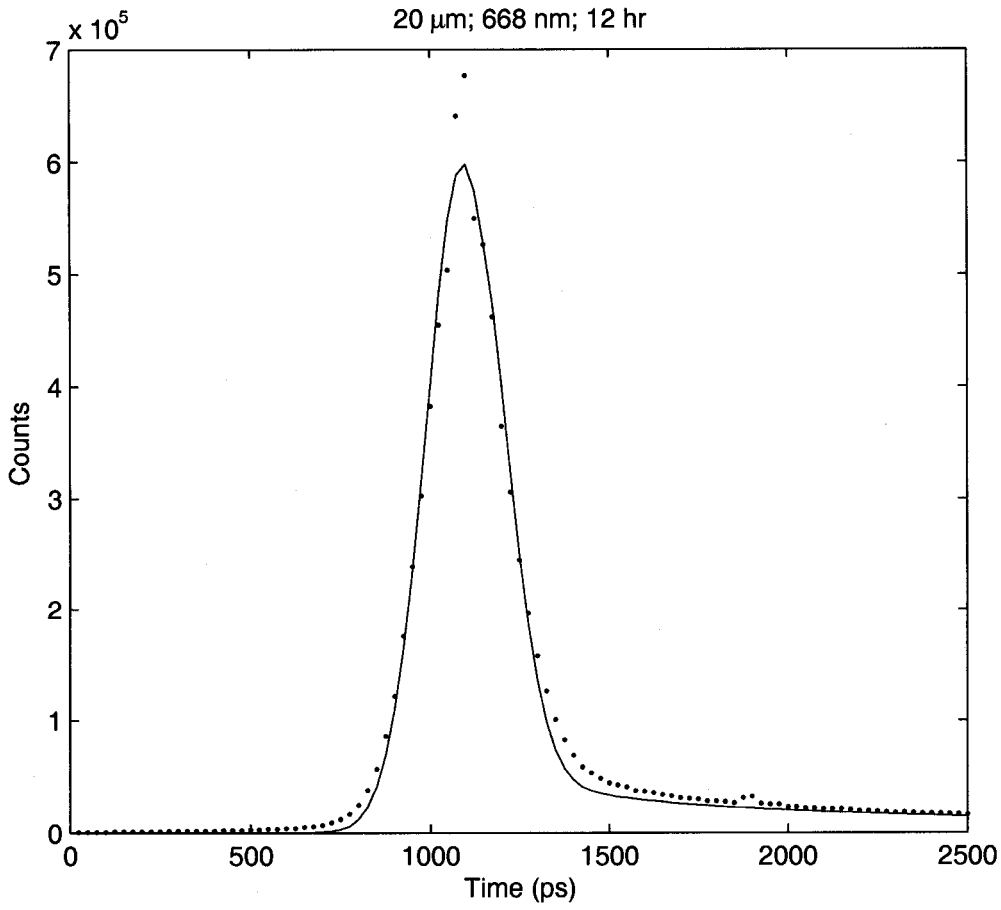


Figure 5.7: Temporal response of 20 μm device at 668 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

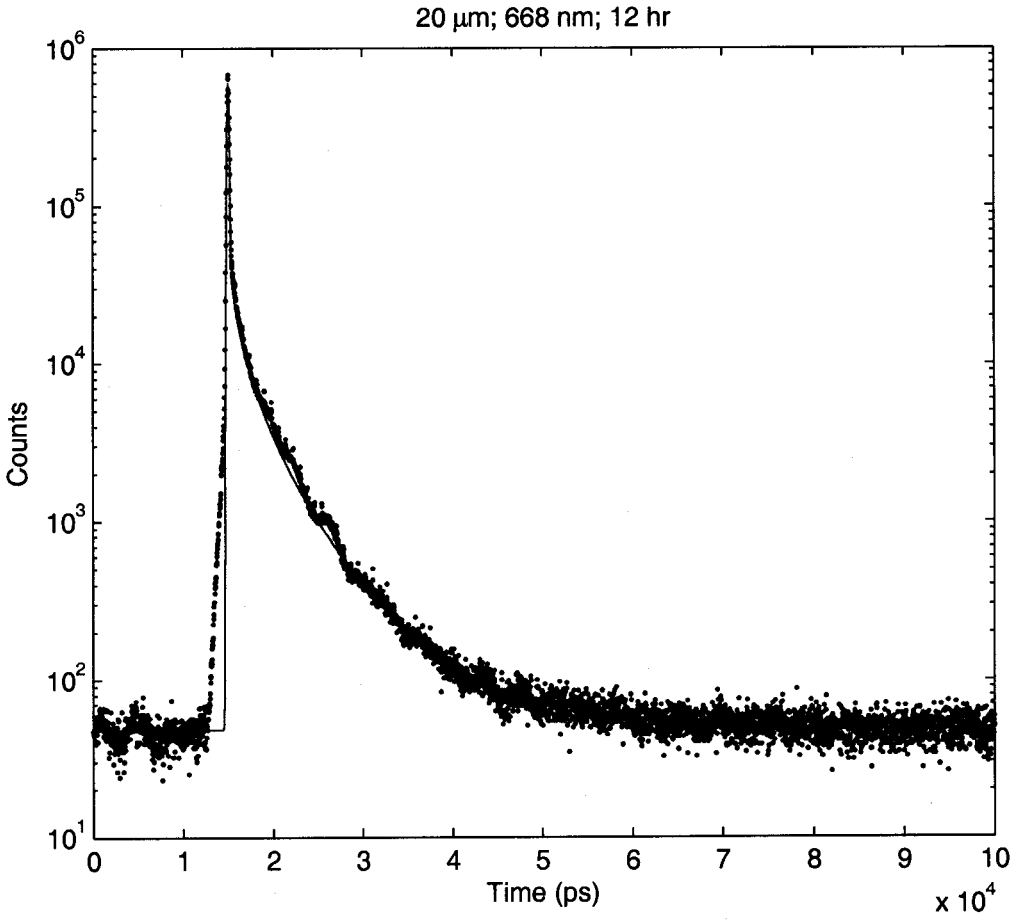


Figure 5.8: Temporal response of 20 μm device at 668 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

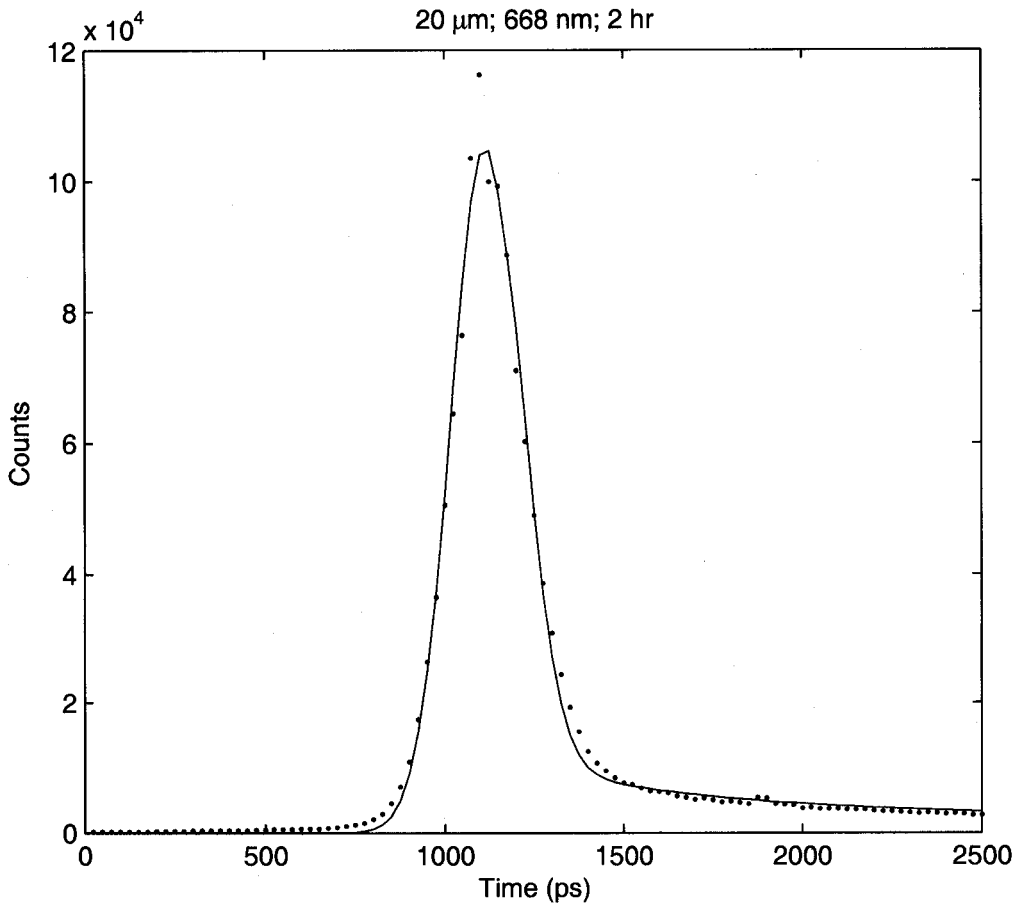


Figure 5.9: Temporal response of 20 μm device at 668 nm during a 2 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

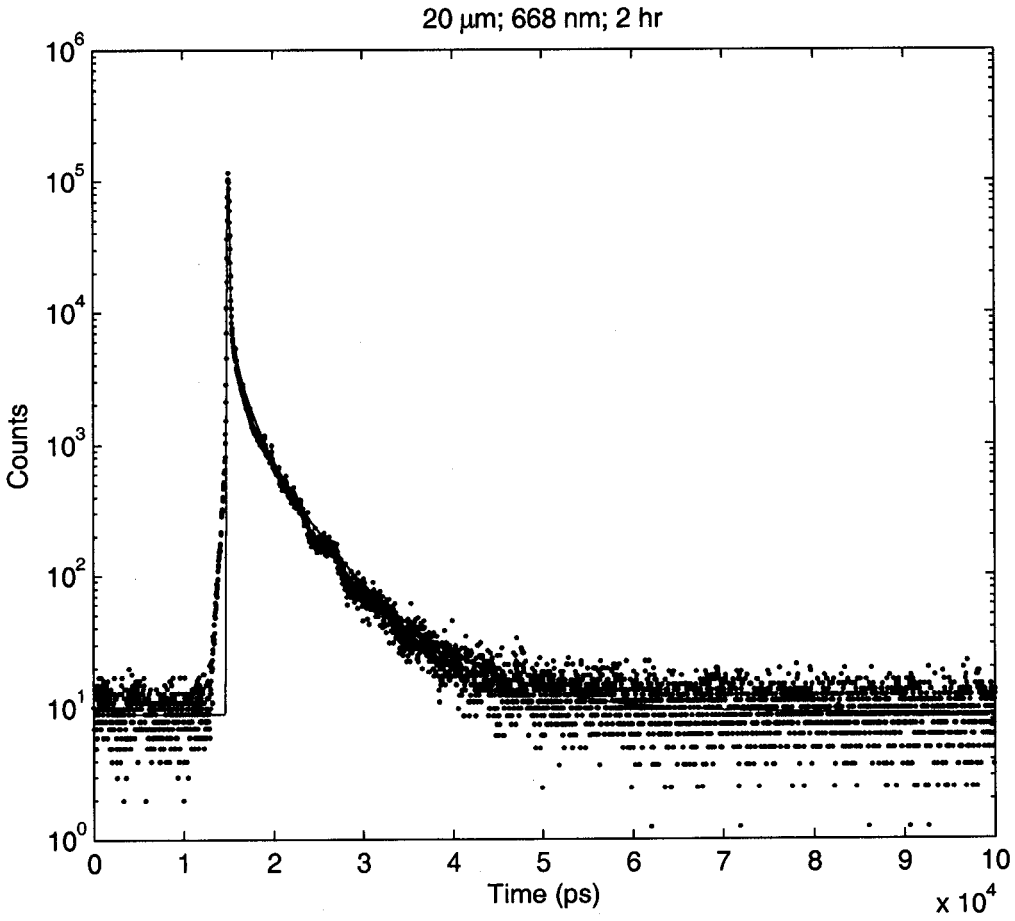


Figure 5.10: Temporal response of 20 μm device at 668 nm during a 2 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

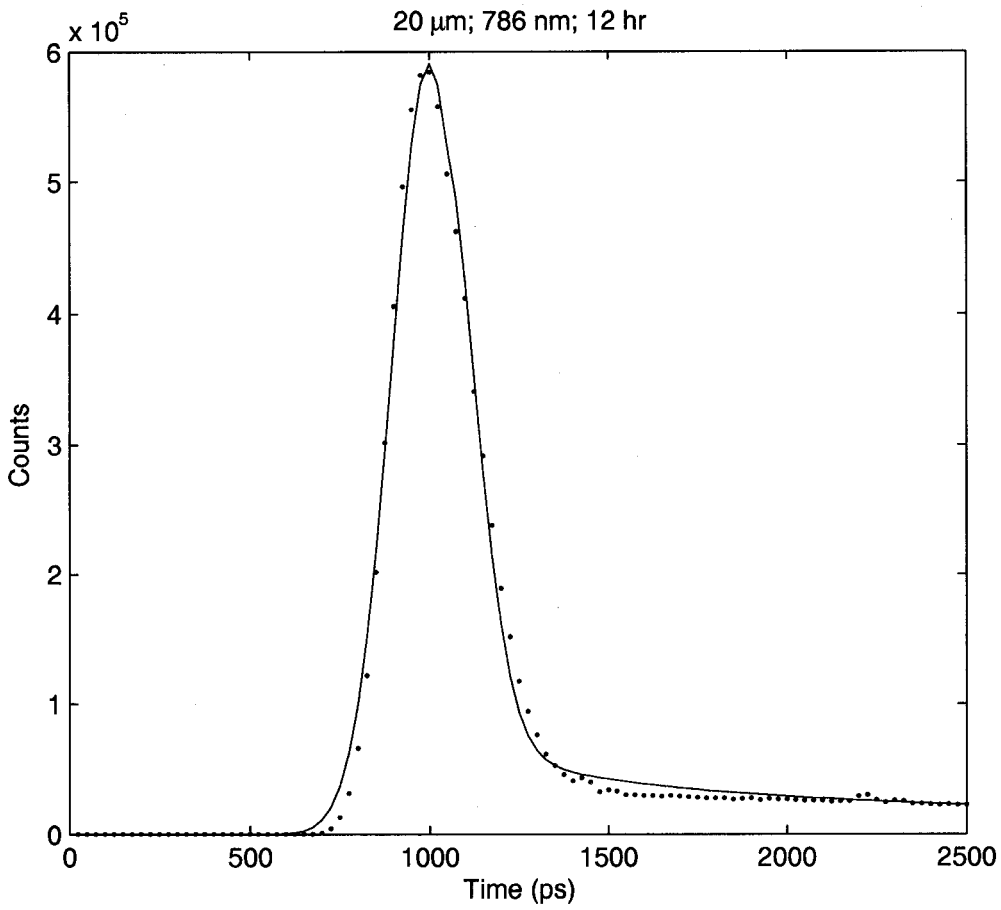


Figure 5.11: Temporal response of 20 μm device at 786 nm during a 12 hour exposure The red points represent the raw data and the blue line is the least-squares best fit.

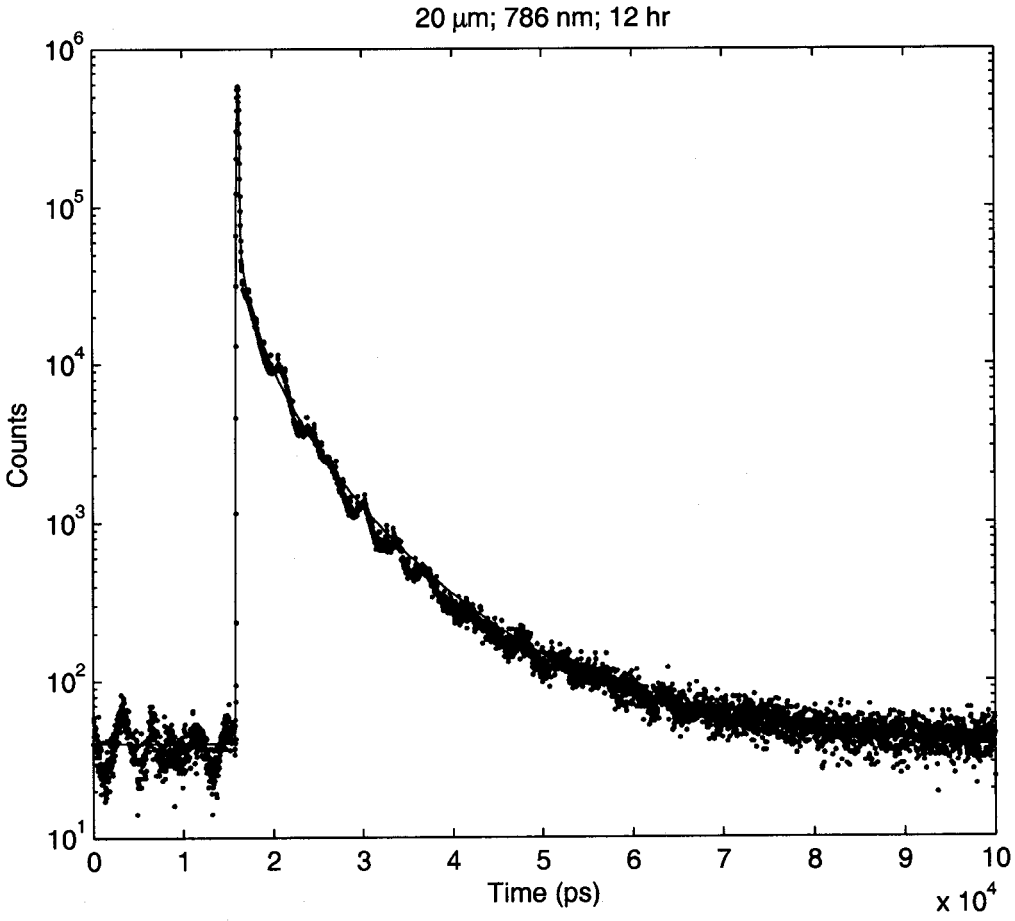


Figure 5.12: Temporal response of 20 μm device at 786 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

Table 5.2: Temporal fit parameters for the 20 micron device. Reported errors are the formal errors of the fit parameters as discussed in the text.

Fit parameters for 20 μm device						
	668 nm, 12 hr		668 nm, 2 hr		786 nm, 12 hr	
Parameter	Fit	Error	Fit	Error	Fit	Error
GssAmp, A (cnts)	5.985e5	1030	1.053e5	162.5	5.902e5	825.5
GssMn, μ (ps)	1.510e4	0.2713	1.512e4	0.2174	1.625e4	0.2157
GssSig, σ (ps)	105.6	0.2580	96.88	0.2078	106.1	0.2012
ExpPeak, B (cnts)	7.171e4	1970	1.587e4	284.3	8.997e4	1221
ExpTau, τ (ps)	532.3	20.43	527.5	13.46	744.5	14.92
Bkgrnd, C (cnts)	48.02	38.76	8.892	5.839	39.59	32.08

parameter.

5.3.3 30 micron device

Figures 5.13 and 5.14 show the temporal response of the 30 micron device illuminated with 668 nm light for 12 hours. The data found in Figures 5.15 and 5.16 were also collected with 668 nm light, but for a period of 2 hours. Figures 5.17 and 5.18 show the distribution of photon arrival times for the 30 micron device illuminated with 786 nm light for 2 hours. The photon detection rates for both two hour exposures were significantly higher than the other long exposures. The first-photon corrections were at the 20-25% level, as opposed to the 10% corrections of the other four datasets.

Table 5.3 gives the results of the least-squares fits of the temporal profiles for the 30 micron device.

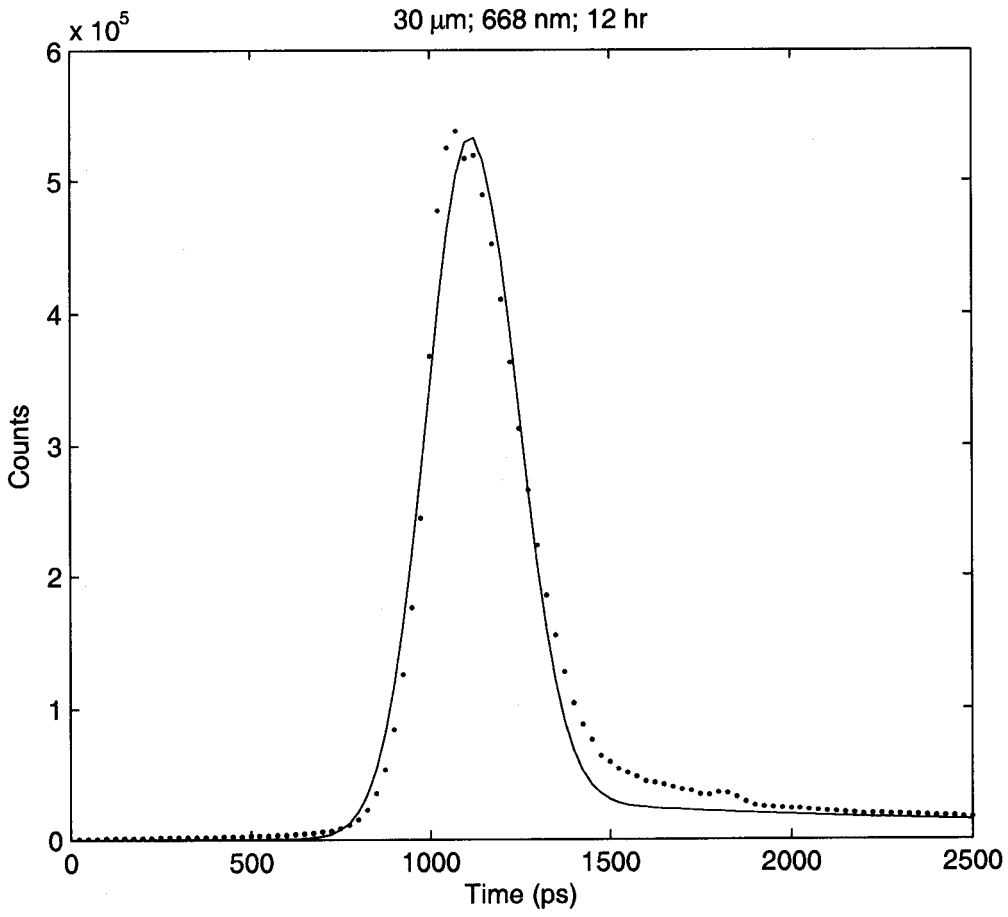


Figure 5.13: Temporal response of 30 μm device at 668 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

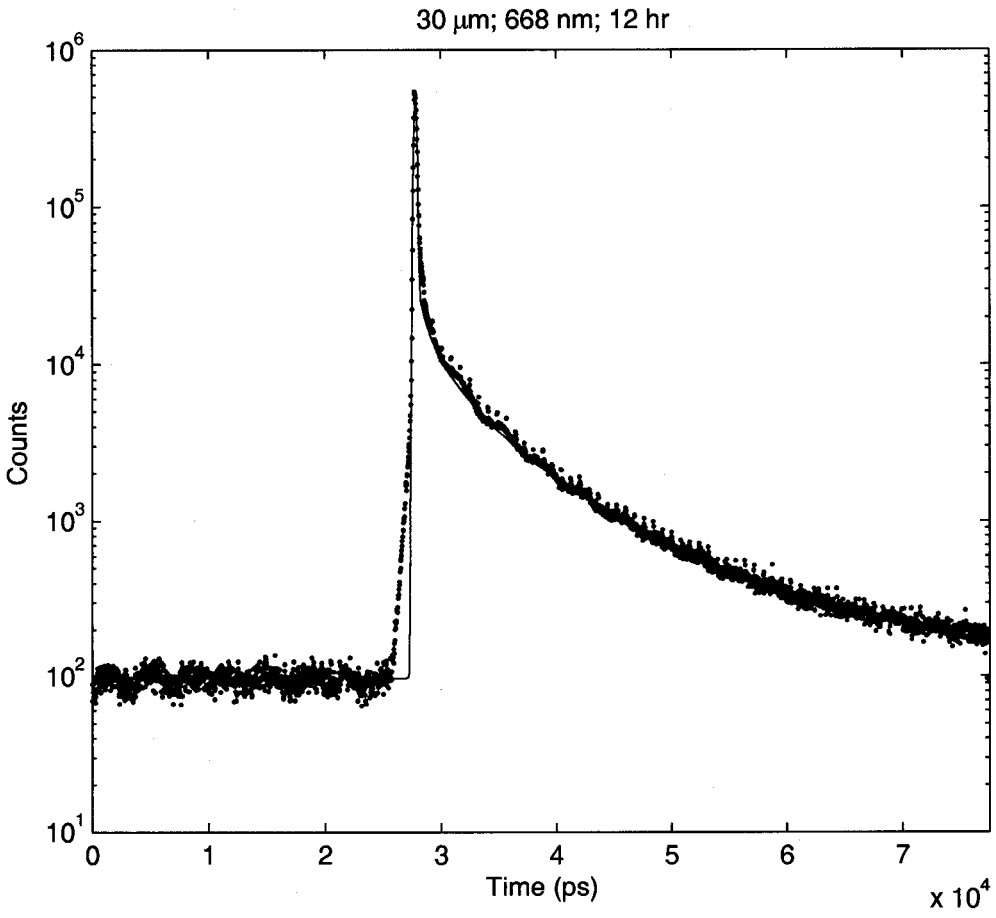


Figure 5.14: Temporal response of 30 μm device at 668 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

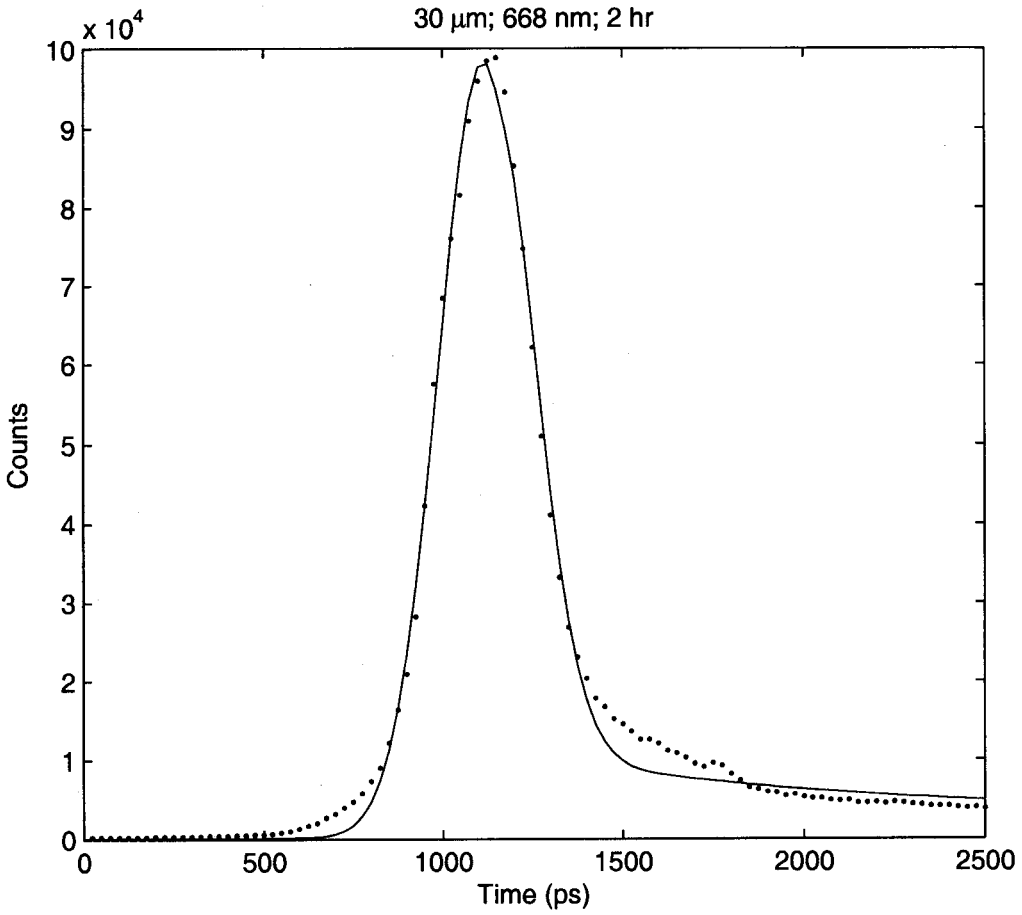


Figure 5.15: Temporal response of 30 μm device at 668 nm during a 12 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

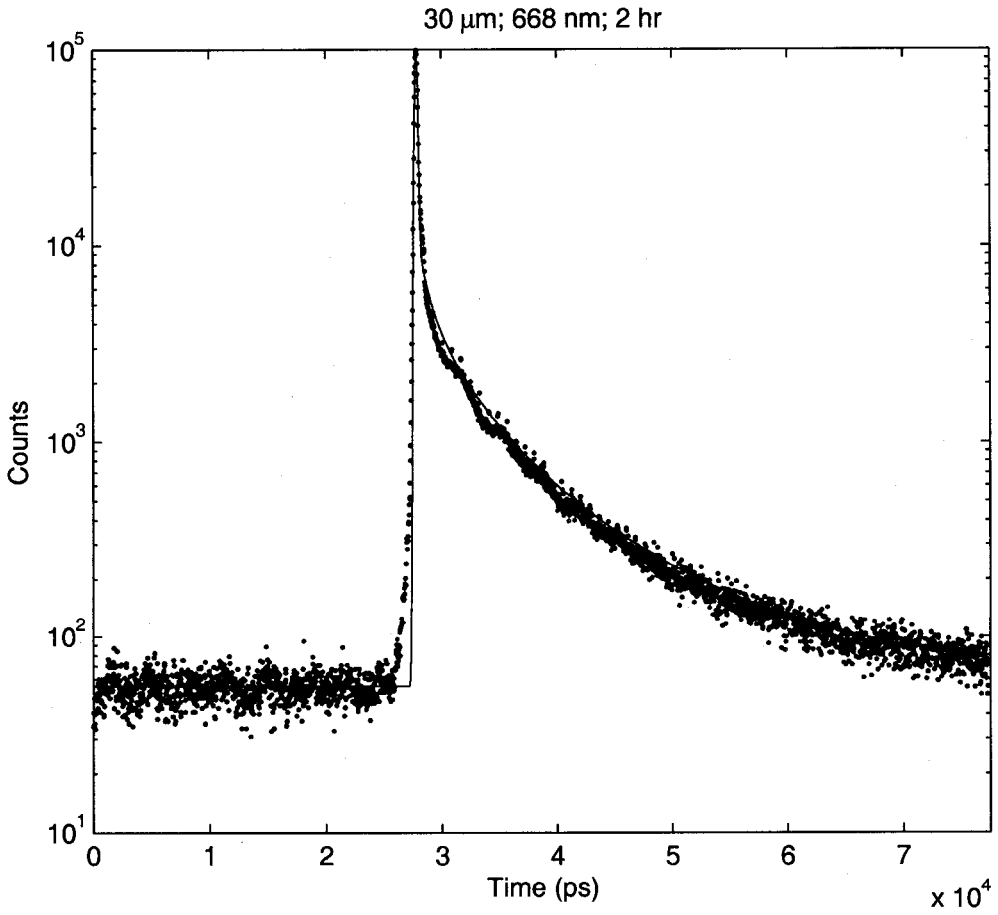


Figure 5.16: Temporal response of 30 μm device at 668 nm during a 2 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

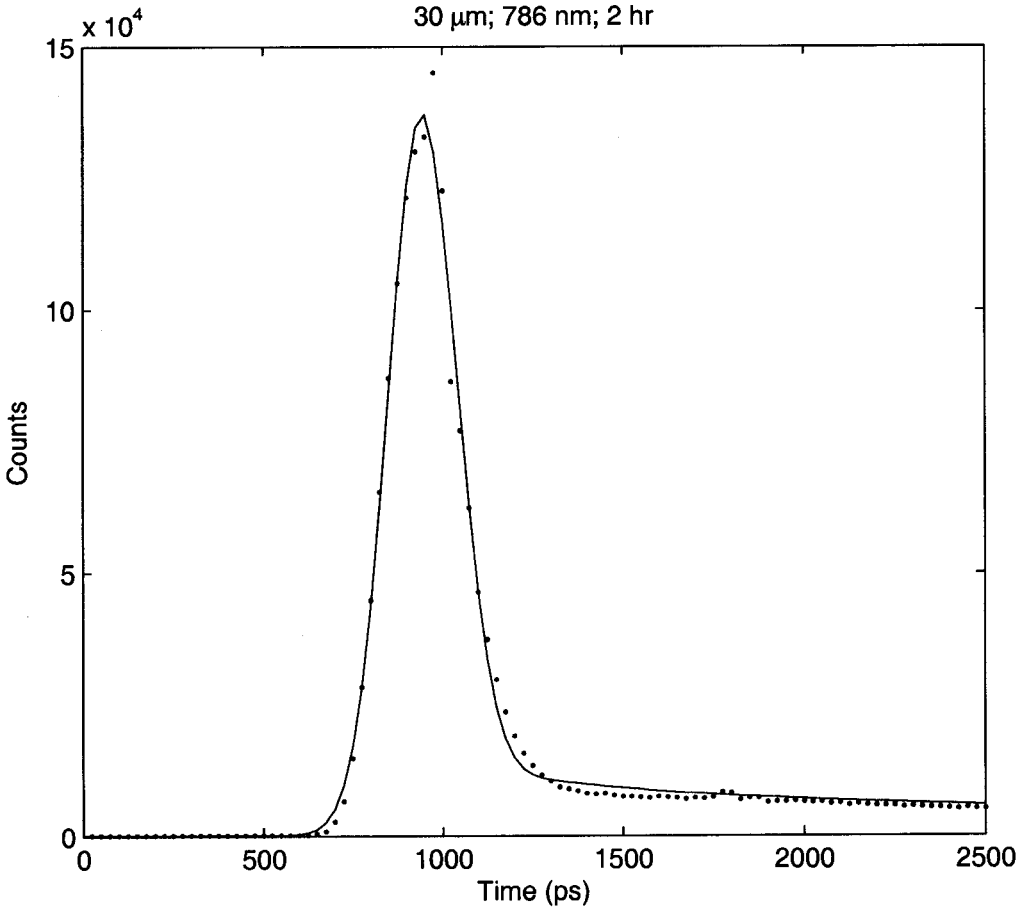


Figure 5.17: Temporal response of 30 μm device at 786 nm during a 2 hour exposure. The red points represent the raw data and the blue line is the least-squares best fit.

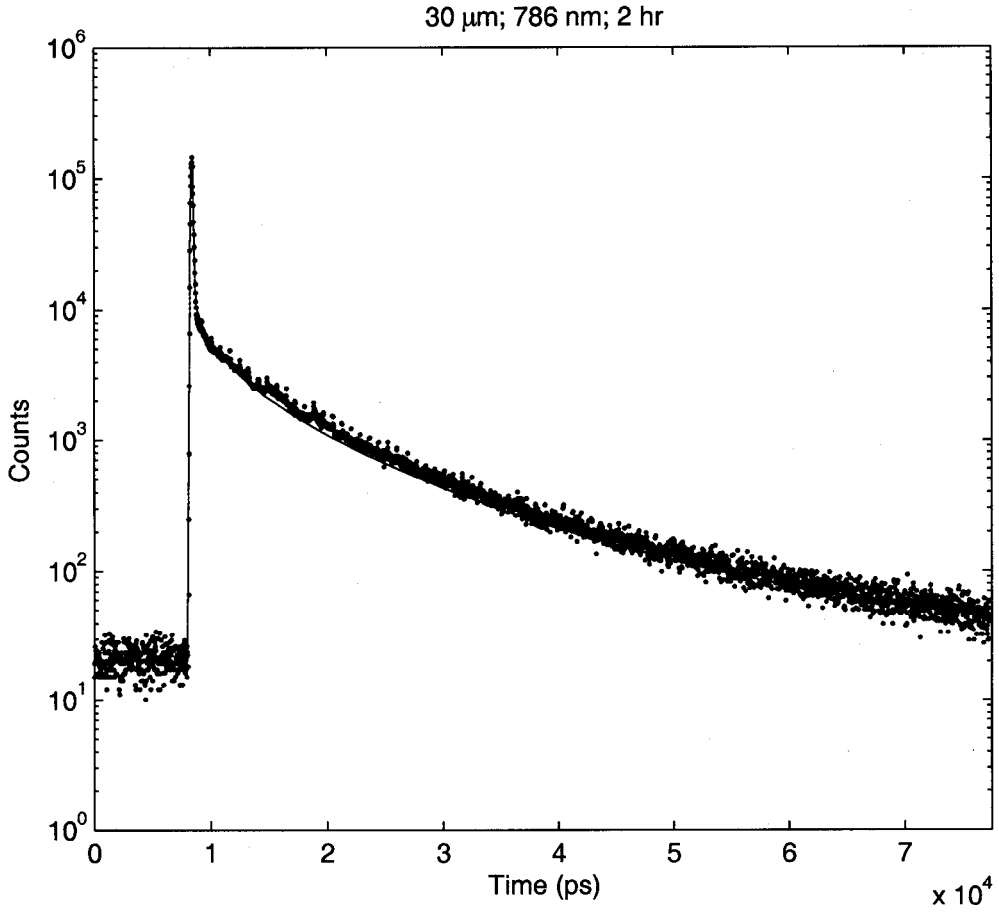


Figure 5.18: Temporal response of 30 μm device at 786 nm during a 2 hour exposure.The red points represent the raw data and the blue line is the least-squares best fit.

Table 5.3: Temporal fit parameters for the 30 micron device. Reported errors are the formal errors as discussed in the text.

Fit parameters for 30 μm device						
	668 nm, 12 hr		668 nm, 2 hr		786 nm, 2 hr	
Parameter	Fit	Error	Fit	Error	Fit	Error
GssAmp, A (cnts)	5.337e5	1386	9.832e4	159.3	1.371e5	204.32
GssMn, μ (ps)	5.283e4	0.4273	5.288e4	0.2843	3.345e4	0.1816
GssSig, σ (ps)	124.2	0.4165	127.4	2.685	95.03	0.1747
ExpPeak, B (cnts)	3.876e4	1420	1.311e4	173.5	1.491e4	150.6
ExpTau, τ (ps)	1270	79.90	1203	26.90	1669	29.58
Bkgrnd, C (cnts)	96.83	75.71	55.59	8.726	20.15	8.452

5.3.4 Discussion

Several conclusions can be drawn from these datasets and I will deal with each in turn.

Overall pulse shape and goodness of fit

To first order, the temporal profiles are well described by a Gaussian core and a modified exponential tail, but this is not a complete model of the device physics. The fitting function used to characterize the data does not do a good job of fitting the "shoulder" of the profile, located about 500 ps after the peak. This is not surprising since we have not included any mechanism for modeling the drift region of the APD in this fitting function. The 668 nm data also deviates significantly in the rising edge of the distribution. More will be said about this later. Both of these un-modeled phenomena contribute to a reduced χ^2 on the order of 10–100 for these six datasets.

Figures 5.19 and 5.20 show the data, fits and normalized, squared residuals for

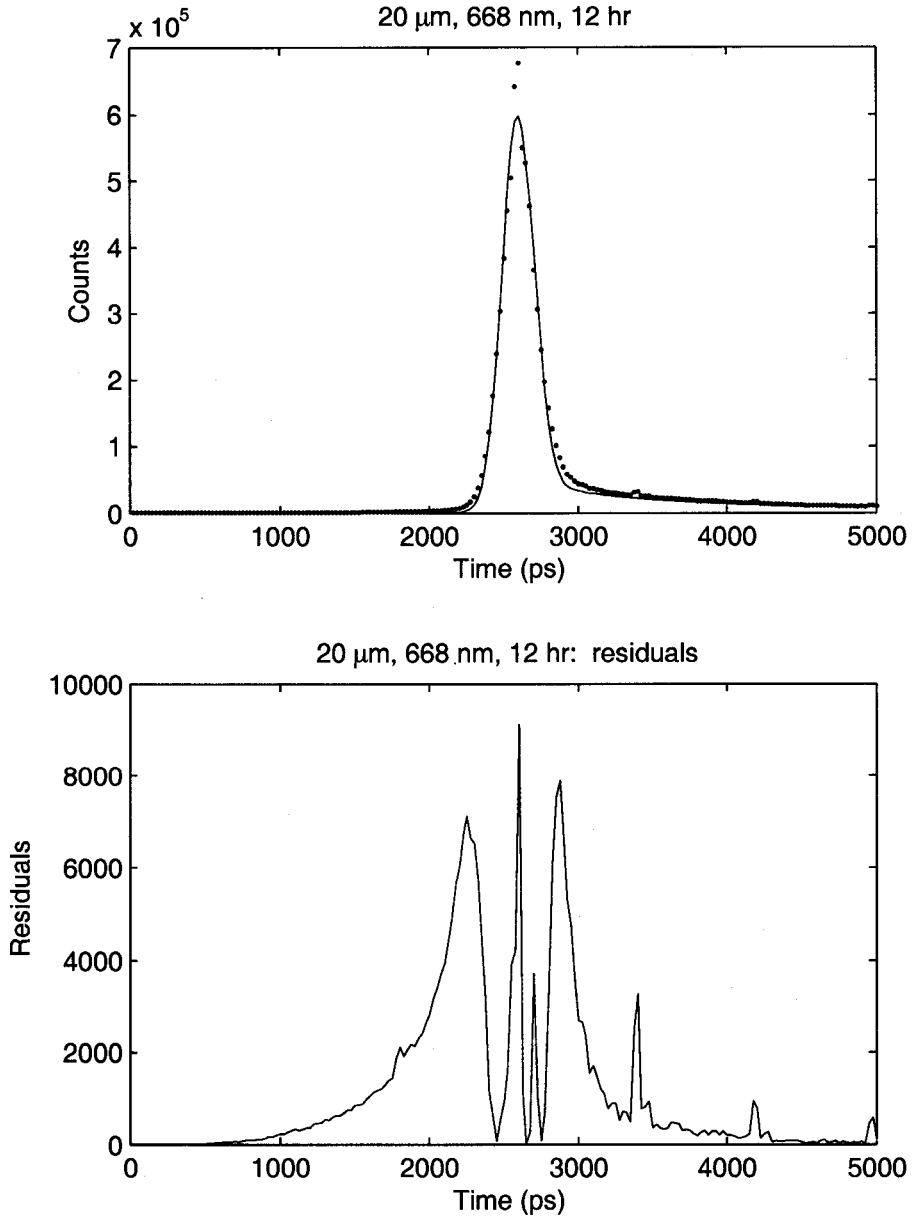


Figure 5.19: Data (red), fit (blue) and residuals (bottom graph) for the core response of the 20 micron device illuminated with 668 nm light for 12 hours.

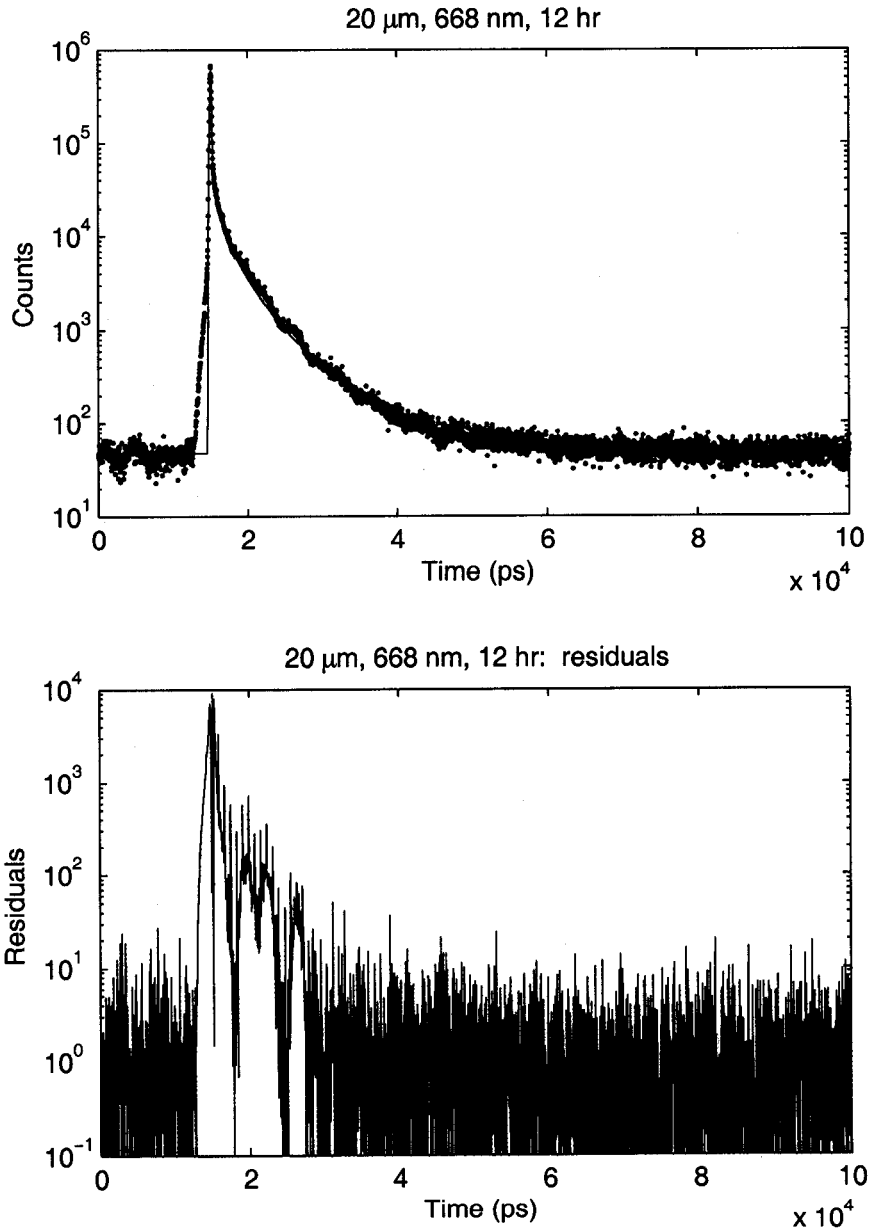


Figure 5.20: Data (red), fit (blue) and residuals (bottom graph) for the full response of the 20 micron device illuminated with 668 nm light for 12 hours.

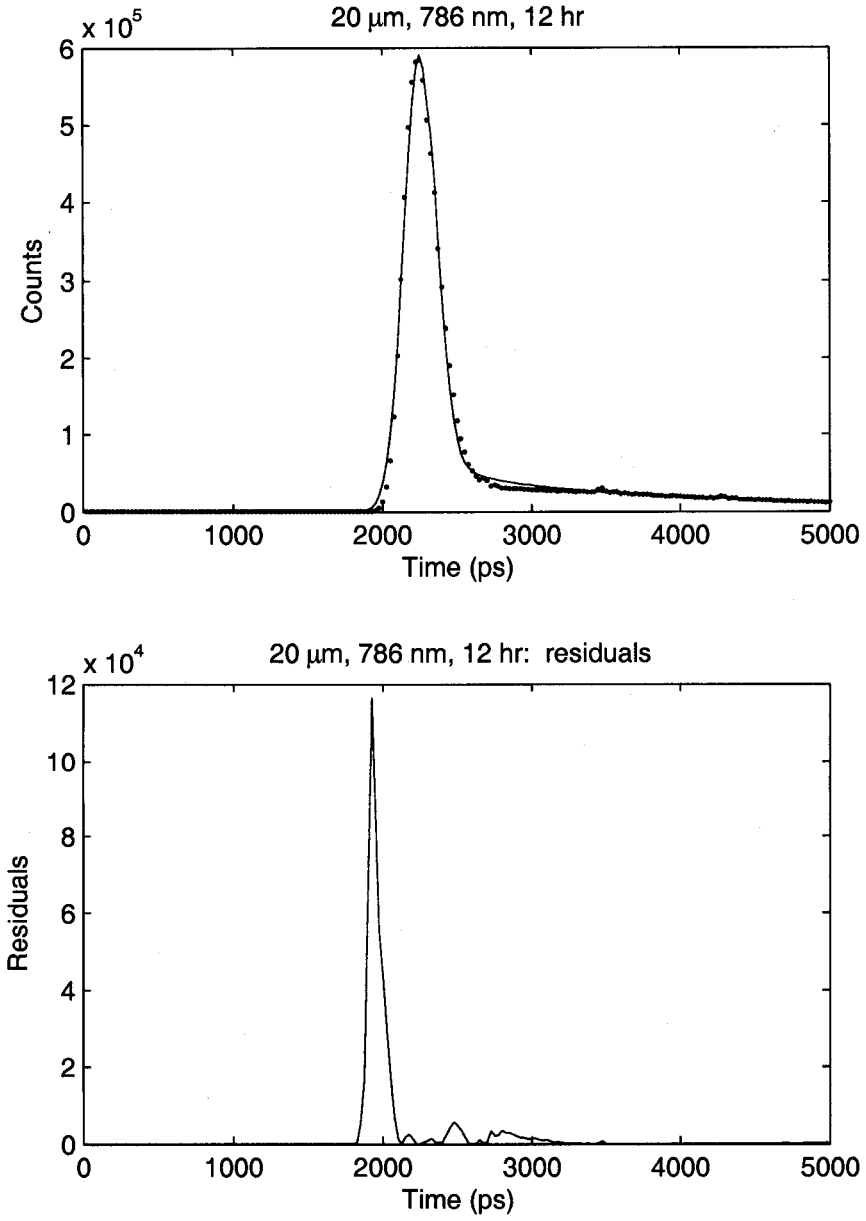


Figure 5.21: Data (red), fit (blue) and residuals (bottom graph) for the core response of the 20 micron device illuminated with 786 nm light for 12 hours.

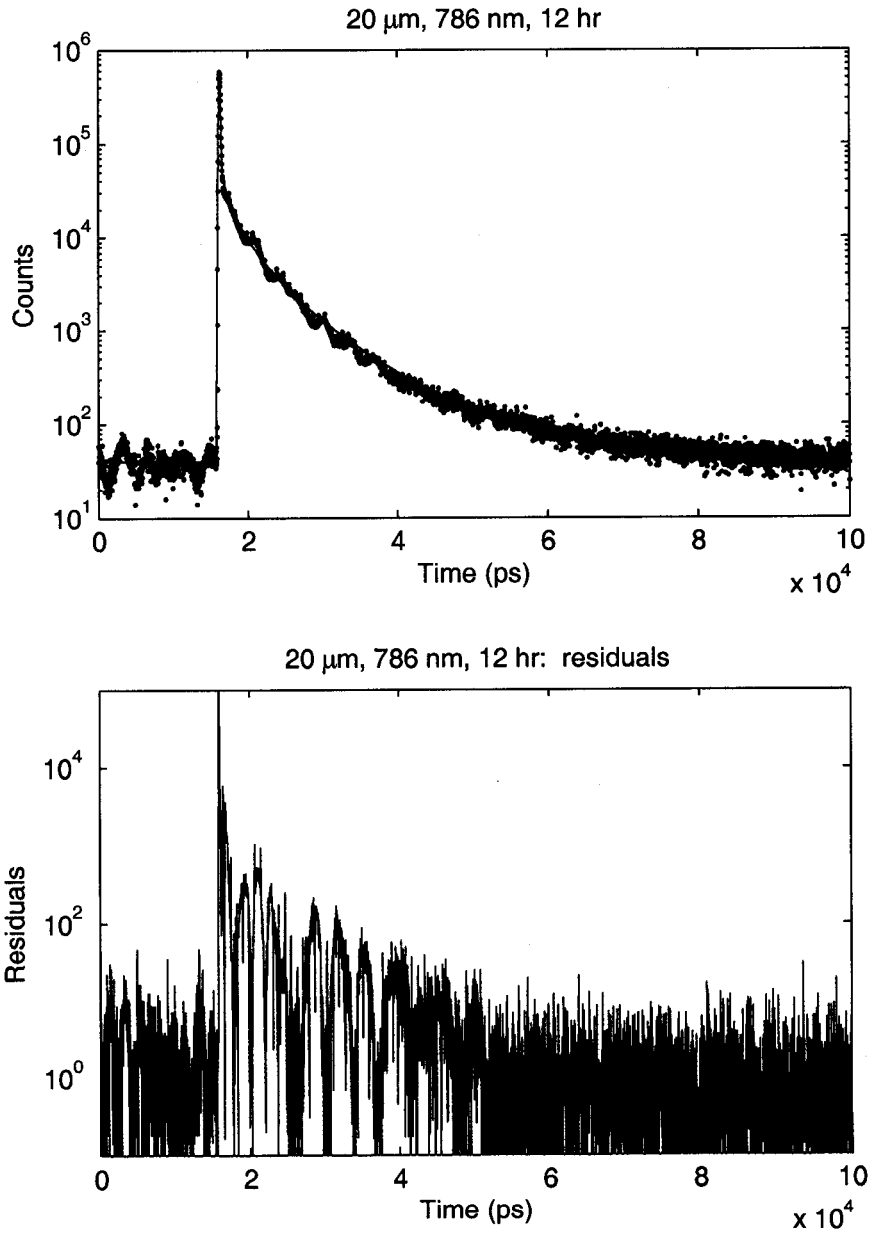


Figure 5.22: Data (red), fit (blue) and residuals (bottom graph) for the full response of the 20 micron device illuminated with 786 nm light for 12 hours.

the 20 micron device illuminated with 668 nm light for 12 hours. Summing these residuals and dividing by the number of degrees of freedom (in this case 4090) results in a reduced χ^2 of 65. The residuals of the core clearly show the contributions from the un-modeled physics. The full response verifies that the residuals are dominated by the discrepancies in the core.

Figures 5.21 and 5.22 show the data, fits and normalized, squared residuals for the 20 micron device illuminated with 786 nm light for 12 hours. These data have a χ^2 of 143. The χ^2 value is dominated by the handful of points on the leading edge of the laser pulse.

The least-squares routine used to fit the data uses χ^2 in its determination of the uncertainties in the temporal parameters. Because χ^2 is dominated by unmodeled physics, I have chosen not to report the values for specific datasets.

Width of Gaussian core

For all six datasets the Gaussian spread, σ is very similar, ranging from 100–125 ps. In our simple model for the APD, the Gaussian core is the result of photons that photo-convert in the avalanche or drift regions of the diode. As a result, the width of this Gaussian core should be wavelength independent. At shorter wavelengths more photons would photo-convert within these regions than at longer wavelengths, but the temporal response, or the width of the Gaussian, should remain unchanged.

Estimate of detector contribution to temporal width

Several factors should be considered when examining the width of the Gaussian core. The temporal profiles are a convolution of the laser pulse, the response from the APD, the readout electronics and the TDC. We have observed that the readout electronics and TDC are capable of a combined timing response with a spread of less than 25 ps RMS as shown in Figure 5.2. However, this is highly dependent on the threshold level of the comparator. Figure 3.5 illustrates how temporal jitter is dependent on

the slope of and noise on the crossing signal. The best temporal resolution is achieved with steep slopes and low noise. As the threshold is adjusted, the first crossing (which flips the comparator) is moved across different slopes and the temporal jitter of the readout electronics will change.

The location of the photon arrival within the gate also affects the timing jitter. We observed transients and ringing associated with the turn-on and turn-off of the gate. Any APD signals will be superimposed on this gate. As a result, any noise on the APD gate itself could translate into temporal jitter of the electronics. If the delay between the turn-on of the gate is too short, the gate voltage itself may still be fluctuating.

The two traces at the bottom of Figure 3.6 illustrate how the temporal jitter can depend on the threshold level and the location within the gate. If the photon signal is superimposed on a down-swing of the gate, then the slope at the threshold crossing will be much steeper than if the gate is swinging up. Changing ΔV will effectively change the threshold and effect the temporal jitter. For the sake of argument, let's assume that the leading edge of the current spike triggered by a photon arrival increases slowly to a constant slope and then slowly rolls off to a steady state value. If the threshold is set too small, then the avalanche will just be beginning and the slope at the threshold crossing will be small. On the other hand, if the threshold is set too high the current may have started to roll off, resulting in a shallow slope at the threshold crossing.

It should also be noted that the TDC was beginning to malfunction during the collection of much of this data. Calibrations of the TDC frequently showed that it had an uncertainty of over 50 ps RMS, more than triple its specifications. This uncertainty would be convolved with the threshold or slope dependent uncertainty described above.

The laser pulses also play a role in the temporal jitter of the core. Figures 5.23 and 5.24 show the streak camera results for the two laser pulses used for these experiments.

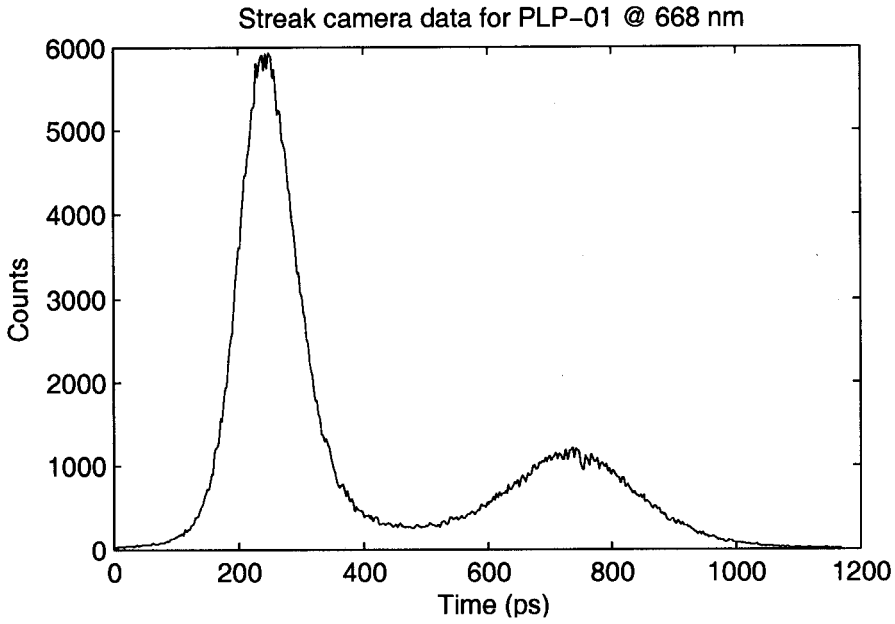


Figure 5.23: Streak camera data for the PLP-01, lasing at 668 nm

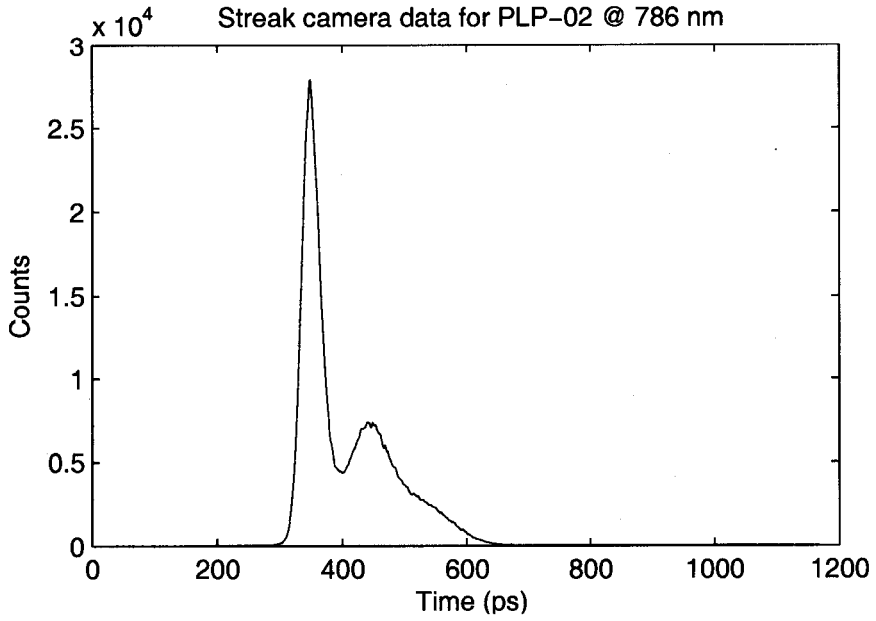


Figure 5.24: Streak camera data for the PLP-02, lasing at 786 nm

These data were provided by Hamamatsu. The 668 nm laser has a centroid of 398 ps and an RMS of 233.5 ps. The 786 nm laser has a centroid of 406 ps and an RMS of 74.1 ps. The entirety of the 786 nm pulse fits within the Gaussian core of the convolved temporal profiles and it is not surprising that its presence is not manifest in the data. The profile of the 668 nm laser provided by Hamamatsu is not consistent with the data we collected. The second after-pulse should be clearly distinguishable, but it is not present in any of the 668 nm profiles.

Inspection of Figures 5.23 and 5.24 show that the 668 nm laser pulse also has a much slower leading edge compared to the 786 nm data. We see indications of this in the convolved temporal profiles. The residuals in Figure 5.19 show that the 668 nm convolved temporal profiles deviate significantly from a Gaussian shape at low count numbers. This discrepancy is also easily seen on the logarithmic plots of each of the 668 nm datasets. In contrast, with the exception of a handful of points, the leading edge of the 786 nm temporal profile is well described by a Gaussian. This is evident by the good fit to the leading edge on the logarithmic plots and the residuals shown in Figure 5.21.

Hamamatsu can create a faster leading edge but at the expense of an after-pulse. The convolved temporal profiles indicate that the leading edge of the 668 nm laser is much slower than the 786 nm. This being the case, the possibility exists that the 668 nm temporal profile provided by Hamamatsu does not go with the slow-rising, non-after-pulsing laser they sent. Perhaps somewhere else in the world some other graduate student is pondering over the strange after-pulse present in her data. The laser or its driving circuitry could also have been damaged during shipping. Disregarding the conspiracy theories, if the after-pulse is ignored and we only consider the first 500 ps of data, the centroid is 258 ps with an RMS of 67.0 ps.

An effort was made to reasonably adjust the threshold level in all of the discussed datasets so that the timing jitter of the readout electronics and TDC were minimized. While we have observed that the combined electronic and TDC contribution could be

less than 25 ps, it was typically on the order of 50–75 ps. Consider for the moment that both the 786 nm and 668 nm laser pulses contribute on the order of 70 ps to the overall temporal jitter. The quadrature combination of these two sources alone is between 85–100 ps RMS. The total temporal profiles are between 100–125 ps RMS. As a result, the detectors could contribute anywhere from a few picoseconds (using 100 ps and 100 ps) to 90 ps (using 125 ps and 85 ps) of RMS jitter to the error budget.

I have chosen to estimate an *upper limit* on the detector contribution of 75 ps RMS. If this contribution is convolved with a 50 ps electronic jitter and a 70 ps jitter from the laser, one would expect a total temporal width of 114 ps. This is certainly in the ball park of what we observe. It is important to remember that this is only an upper limit. The electronics *could* be contributing much more than 50 ps to the temporal jitter, in which case the detector's contribution could decrease to the point of being less than the expected 40 ps laser pulse at 532 nm. Based on the observed variations in the contribution from the timing electronics and TDC, I would speculate that the detector will contribute 45–50 ps to the random error budget.

The work presented here was aimed at determining an appropriate readout electronics scheme for the APD arrays. All the fabrication was done on two-layer printed circuit boards. The final Mark VI design that was used to collect all of these data is something of a proof-of-concept. In order to achieve quiet signals, multi-layer circuit boards with good ground planes must be designed and attention paid to high frequency noise pickup. This work is being undertaken by the electronics team at the University California at San Diego. It is expected that the final APD electronics will be much quieter and contribute significantly less to the overall temporal jitter. Unfortunately, we will have no *a priori* information about the pulse shape of APOLLO's Nd:YAG laser other than the specifications. This will make any deconvolution of the detector and laser pulse responses more challenging.

Diffusion tail dependence

A significant pseudo-exponential tail was observed in both devices at both wavelengths, however the decay constant of the tail is about 25% shorter at 668 nm than it is at 786 nm. Under our simple model for APD performance, the diffusion tail comes from photons that photo-convert deeper in the silicon where there is no electric field and are only detected after diffusing to the drift region. A greater portion of the 786 nm photons will be found in the tail because of the longer penetration depth in silicon. The details of the process will be discussed further in Section 5.6.

The decay constant of the tail was also shorter for the 20 micron devices. This is consistent with the results of the simulations discussed in Section 4.4. The 20 micron devices have a smaller diffusion volume, so electrons which penetrate deeper are more likely to be swept into the guard ring or go undetected.

Summary

In this section we have discussed the temporal profiles of six canonical datasets. The data are well described by a Gaussian core and a pseudo-exponential, diffusion-driven tail. Un-modeled physics, in the form of the leading edge of the 668 nm laser pulse and the effect of the drift region (at the interface between the core and tail), account for the largest residuals. The Gaussian core has a wavelength-independent width of 100–125 ps RMS. I have estimated that the laser pulse and readout/timing electronics contribute 75 and 50 ps, respectively, to this convolved temporal response. This sets an upper limit on the detector contribution of 75 ps, though I would speculate that the true value is closer to 50 ps. The pseudo-exponential tail decays faster for smaller devices and shorter wavelengths, which is consistent with a diffusion process.

5.4 Spatial profile of response

The main points learned from the spatial profiles are

- The spatial profile of the total detected flux is well described by a "Bowler Hat" function.
- The spatial widths of the profiles are consistent with the sizes of the detector elements.
- While peak heights were similar, the flux detected at longer wavelengths falls off more rapidly toward the edges of the array.

In order to characterize the spatial response of the individual detector elements, we scanned a small laser spot across the element and illuminated it at several locations. For the one-dimensional, high-resolution scans, the laser spot was scanned across either the 20 or 30 micron element in 1.5 micron steps. The scans are approximately twice as wide as the diameter of each element and the element is roughly centered in each case. At each point in the scan the detector was gated 40,000 times. Neutral density filters were inserted (as shown in Figure 5.3) until the number of positive detections was less than one per gate, typically receiving two or three positive hits out of every ten gates.

To determine the optimum separation between the lens and detector, we arranged them so that their separation was greater than the estimated best focus distance for the lens. The spot was then scanned through the center of the detector element and a simple fitting routine was used to estimate the spot size. The distance between the lens and detector was then decreased and the scans and fitting were repeated. In this way, we stepped through the waist of the focused beam until the spot size began to increase. This method allowed us to create the smallest possible spot on the detector. This procedure was repeated for each wavelength and each detector studied. Because

of the short focal lengths of the optics used, we were not able to quantitatively verify the smallest attainable spot size by imaging the spot on a CCD array or some other method. It should also be mentioned that for all the analysis, we assumed that the laser had a Gaussian spatial profile.

The one-dimensional response that we measure is the result of a convolution of the APD spatial response with the laser's spatial profile. If the laser spot could be made infinitely sharp and narrow, the measured response would be essentially that of the APD by itself. On the other hand, if the laser spot size was much larger than the detector, the measured response would be essentially that of the laser spot. In our case, the laser spot has a FWHM on the order of 4–5.5 microns, smaller than the detector diameter—but not greatly so.

When analysis of the spatial response was first begun, we had not yet developed a robust fitting function for the temporal profiles. As a result, we began analysis of the spatial response by solely examining the total number of detected photons at each point within the scan.

As a first approximation of the detector response, we assumed that it was uniform over the element in a sort of "Top Hat" function. It was this "Top Hat" function that was convolved with a Gaussian laser spot to determine the optimum separation between the lens and the detector mentioned above. However, when the optimum separation was reached, there were significant differences between the best fit function and the raw data. This prompted us to modify our "Top Hat" function into a "Bowler Hat" function by adding a term with negative parabolic curvature. When the "Bowler Hat" was convolved with the laser pulse, the fits were much better.

The function for the laser was

$$F_{las} = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_l}\right)^2\right]. \tag{5.10}$$

where σ_l was the standard deviation of the laser's spatial profile in microns and m was the mean in microns.

The final "Bowler Hat" function was

$$F_{APD} = \begin{cases} (A-B)(1-C_1x^2) & : \quad m-R \leq x \leq m+R \\ 0 & : \quad m-R > x \quad \text{or} \quad m+R < x \end{cases} \tag{5.11}$$

where R is the detector radius in microns, A is the maximum amplitude in counts, B is the constant background in counts and C_1 is the curvature of the APD response in inverse microns squared. A sample detector response is shown in Figure 5.25 with m = 30 μm and R = 10 μm.

The final fitting function was a convolution of the laser response and the detector response plus a constant background and a second order baseline,

$$F_{fit} = (F_{las} * F_{APD}) + B - C_2x^2. \tag{5.12}$$

The last term was included to account for the slight downward bowing of the baseline. This is particularly apparent for the 30 micron device illuminated with 786 nm light shown in Figure 5.29. This bowing could be the result of Fraunhofer diffraction from the small aperture in front of the lens. In that case, the 3–5 micron FWHM would represent the central core up to the first minima.

All the data in Figures 5.26–5.29 have been corrected for first-photon biases in the following manner. The raw data were fit using Equation 5.12. The background, B, was then removed to determine N_2, the number of positive detections associated with the laser. Following a procedure similar to that shown in Equations 5.3–5.5 the number of incident photons, $N_{corrected}$, can be found by

$$N_{corrected} = -N_{gates} \ln\left(1 - \frac{N_2}{N_{gates}}\right), \tag{5.13}$$

where N_{gates} is the number of gates or total number of possible detections, in this case 40,000.

5.4.1 20 micron device

The two scans of the 20 micron device include 41 data points. The fitting function has seven parameters so each dataset has $\nu = 41 - 7 = 34$ degrees of freedom. This

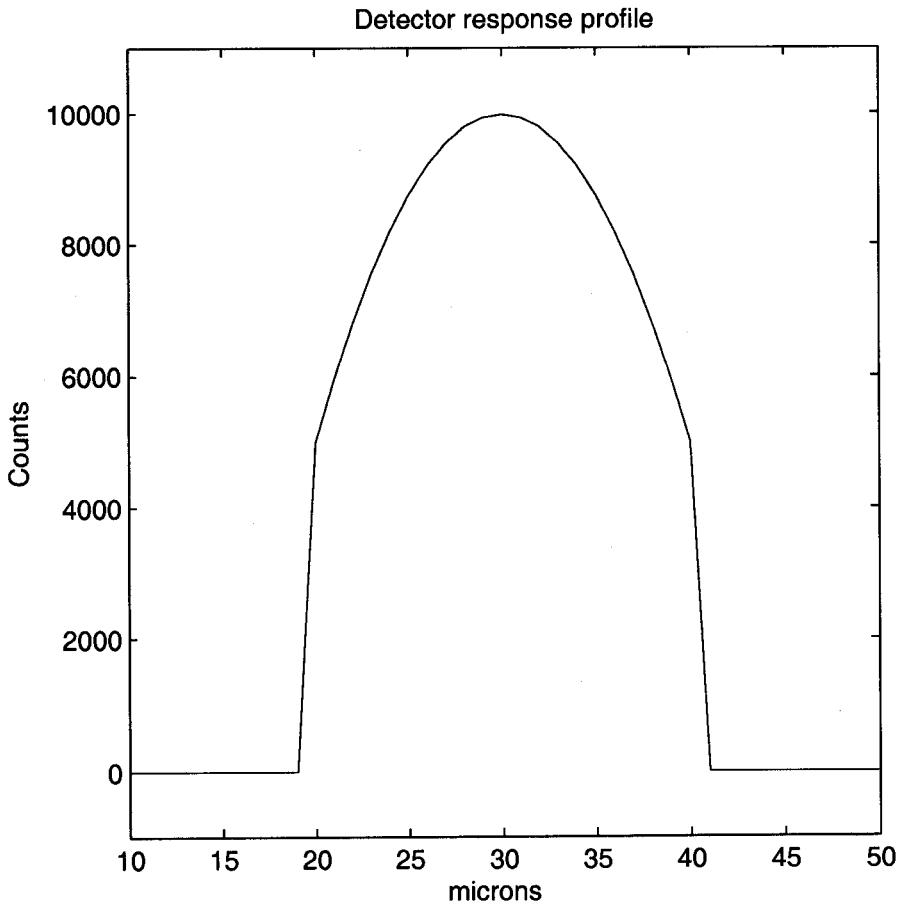


Figure 5.25: Sample detector response model, F_{APD}, with $m = 30$ μm and $R = 10$ μm.

Table 5.4: Spatial fit parameters for 20 μm device at 668 nm and 786 nm. The fit parameters correspond to those found in Equations 5.10–5.11. The errors are the 1σ errors for each parameter.

Fit parameters for 20 μm device				
	668 nm response		786 nm response	
Parameter	Fit	Error	Fit	Error
Amplitude, A (cnts)	11,460	46.93	12,690	61.73
Mean, m (μm)	25.49	0.01356	32.64	0.01437
Radius, R (μm)	10.58	0.2016	9.918	0.1347
Curvature, C_1 ($1/\mu m^2$)	1.018e-2	3.769e-4	1.221e-2	4.280e-4
Background, B (cnts)	6.399	52.45	0.8130	65.07
Laser RMS, σ_l (μm)	1.838	0.1517	1.629	0.1086
2^{nd} order baseline, C_2 (cnts/μm^2)	0.5014	0.1828	0.5620	0.2482

is important for calculating the reduced χ^2.

Figure 5.26 shows the corrected data collected using the 20 μm device illuminated with 668 nm photons. The error bars depict the $\pm 1\sigma$ errors determined by Poisson statistics (i.e. $\sigma = \sqrt{N}$). The solid red line is the least-squares fit of F_{fit}. The parameters of the fit can be found in Table 5.4. For the most part, the fit function is within one or two standard deviations of the raw data. The reduced χ^2 for this fit is 171, but is dominated by one outlying point (located at 3 microns). If this point is removed, the reduced χ^2 becomes 1.21.

Figure 5.27 shows the corrected data collected using the 20 μm device illuminated with 786 nm photons. Again, the error bars come from the Poisson statistics and the solid red line is the least-squares fit of the data. The parameters for the fit function can also be found in Table 5.4. The reduced χ^2 is 1.21 for this dataset and includes all 41 points.

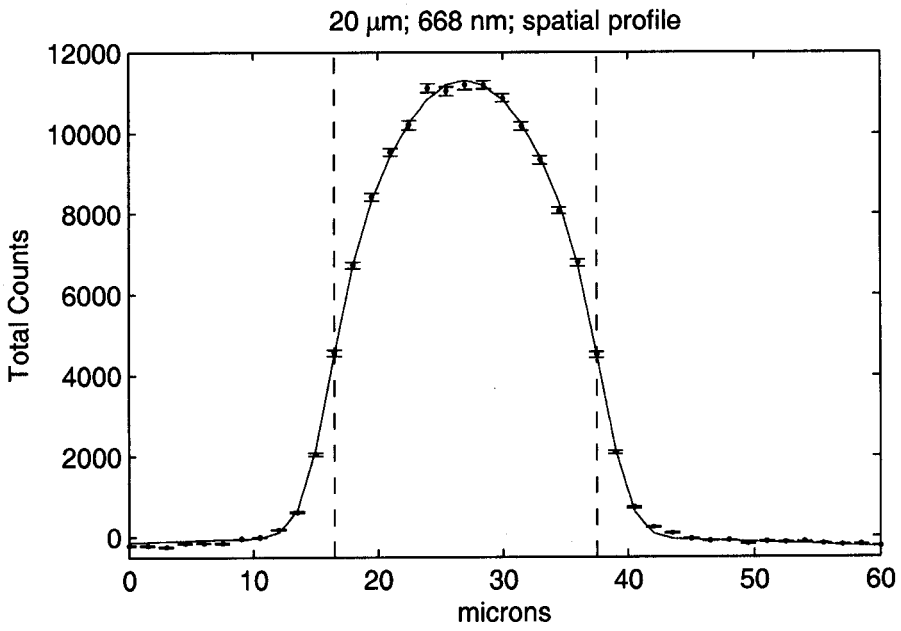


Figure 5.26: Profile of 20 μm device at 668 nm. The data have been corrected for first-photon biasing and the background has been removed. The blue points represent the raw data and the red fitting function is a "Bowler Hat" convolved with a Gaussian laser spot.

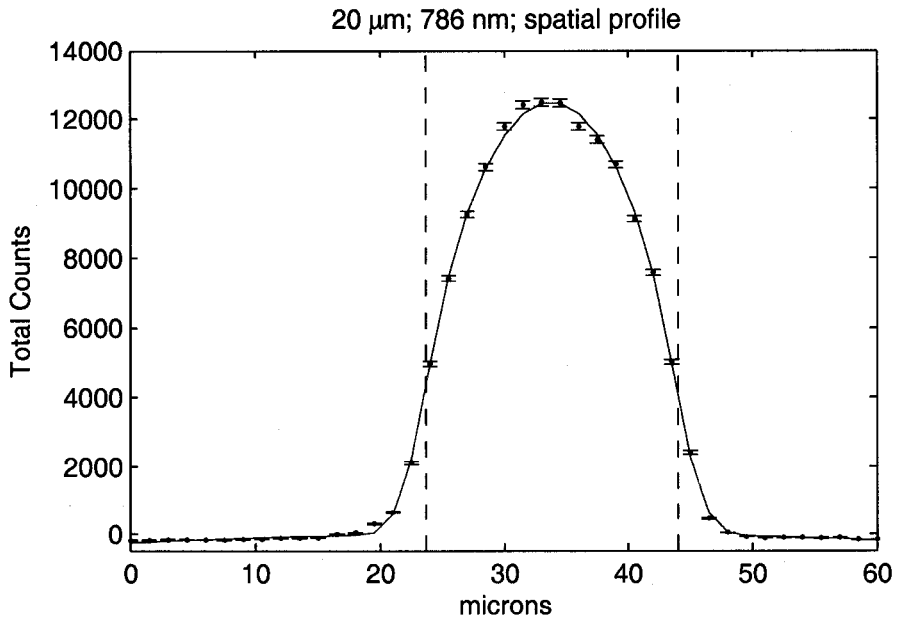


Figure 5.27: Profile of 20 μm device at 786 nm. The data have been corrected for first-photon biasing and the background has been removed. The blue points represent the raw data and the red fitting function is a "Bowler Hat" convolved with a Gaussian laser spot.

5.4.2 30 micron device

The 668 nm scan of the 30 micron device used 61 data points while the 786 nm scan only had 51. Combining this with the number of fit parameters results in 54 and 44 degrees of freedom for the two scans.

Figure 5.28 shows the corrected data collected using the 30 μm device illuminated with 668 nm photons. As before, the error bars come from the Poisson statistics and the solid red line is the least-squares fit of the data. The parameters for the fit function can be found in Table 5.5. The reduced χ^2 for this dataset is 153, but again this is dominated by one point (located at 4.5 microns). If this point is removed, the χ^2 value becomes 1.13.

Figure 5.29 shows the corrected data collected using the 30 μm device illuminated with 786 nm photons. Again, the error bars come from the Poisson statistics and the solid red line is the least-squares fit of the data. The parameters for the fit function can also be found in Table 5.5. The reduced χ^2 value for this dataset is 42.5, and dominated by the first three data points (0–3 microns). Without these three outliers, χ^2 is 1.17.

5.4.3 Discussion

Spatial profile and goodness of fit

The spatial profiles are well described by a "Bowler Hat" function. If a few outliers are removed from each dataset, all four datasets have a reduced χ^2 between 1.1 and 1.2. Completely random errors around the fit would result in an expected χ^2 of 1. A "χ by eye" examination of Figures 5.26–5.29 also indicates that the "Bowler Hat" function does a reasonable job of fitting the spatial profiles.

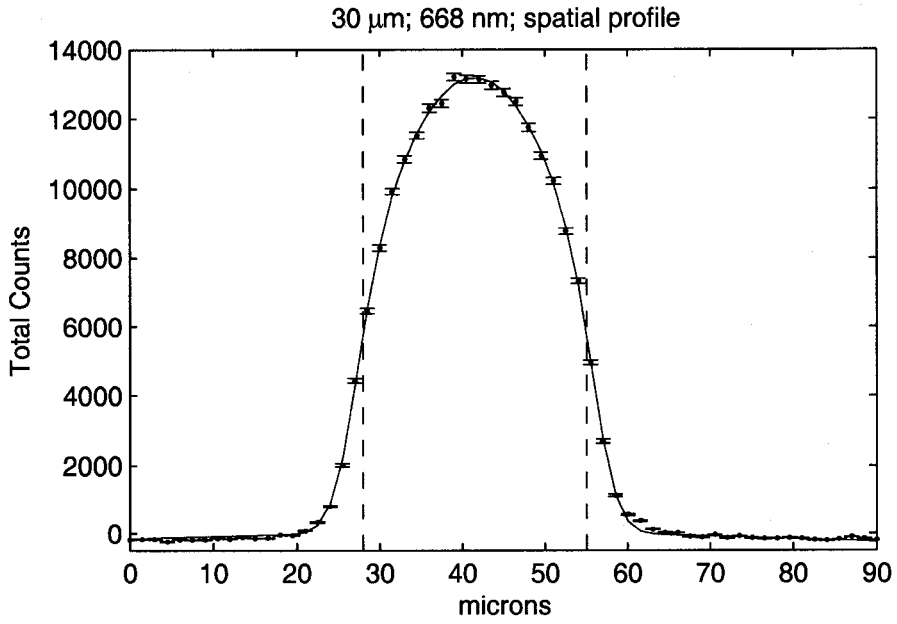


Figure 5.28: Profile of 30 μm device at 668 nm. The data have been corrected for first-photon biasing and the background has been removed. The blue points represent the raw data and the red fitting function is a "Bowler Hat" convolved with a Gaussian laser spot.

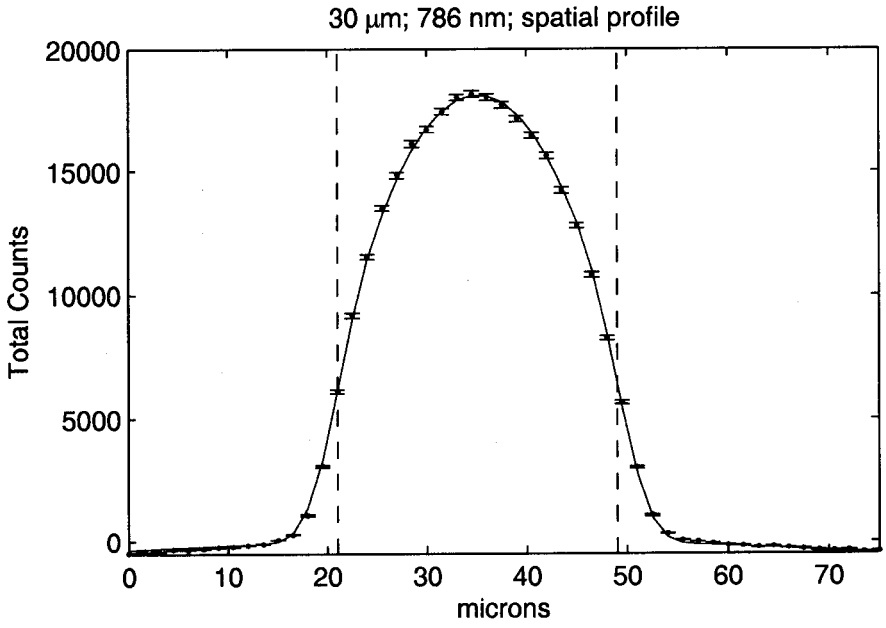


Figure 5.29: Profile of 30 μm device at 786 nm. The data have been corrected for first-photon biasing and the background has been removed. The blue points represent the raw data and the red fitting function is a "Bowler Hat" convolved with a Gaussian laser spot.

Table 5.5: Fit parameters for 30 μm device at 668 nm and 786 nm. The fit parameters correspond to those found in Equations 5.11–5.10. The errors are the 1σ errors for each parameter.

Fit parameters for 30 μm device				
	668 nm response		786 nm response	
Parameter	Fit	Error	Fit	Error
Amplitude, A (cnts)	13,370	44.06	18,370	54.50
Mean, m (μm)	40.24	0.01317	34.04	0.01285
Radius, R (μm)	13.96	0.1221	14.13	0.1336
Curvature, C_1 ($1/\mu m^2$)	5.481e-3	1.642e-4	6.377e-3	1.346e-4
Background, B (cnts)	5.355	41.22	7.053	71.895
Laser RMS, σ_l (μm)	2.331	0.1060	2.229	0.1082
2^{nd} order baseline, C_2 (cnts/μm^2)	0.2328	0.07158	0.7079	0.1690

Radius of "Bowler Hat"

The radius of the "Bowler Hat" is a quick sanity check for the fits. For both the detectors, the radius of the "Bowler Hat" is roughly equal to the radius of the element. Discrepancies between the two could be explained in two ways: a non-uniform step size for the Picomotors which move the laser spot across the element, or edge effects caused by a non-uniform electric field. The Picomotors are not stepper motors but use piezoelectric transducers to turn a screw. This design is essentially a friction-driven system and thus the step size can vary depending on the load the motor is moving, direction of motion, temperature and overall mechanical wear. This is likely the dominant source of error. The electric field will not be uniform over the entire element but will have curvature near the edges. Though there is a parabolic term applied over the whole element, it isn't designed to account for these edge effects.

(The source of the parabolic curvature will be discussed later on in this chapter).

Parabolic curvature

For both the 20 and the 30 micron devices, the 786 nm data have a greater parabolic curvature, C_1, than the 668 nm data. The difference between them is roughly four times the uncertainty and could be be considered statistically significant, depending on how stringent your requirements are. This was our first indication that the spatial detector performance was dependent on wavelength.

Summary

The spatial profiles are well described by a "Bowler Hat" function. The spatial widths of the profiles are consistent with the element radii, especially when the uncertainties due to the non-uniform step size of the Picomotors are considered. The different values of the parabolic curvature, C_1, indicates a difference in the shape and extent of the electric field in the two devices.

5.5 Spatial variation of temporal parameters

- The mean value of the Gaussian core shifts by more than 150 ps depending on whether the laser spot (and thus avalanche initiation location) is at the center or edge of the element, and is well described by a parabola.
- The RMS width of the Gaussian core is essentially uniform over the element.
- The decay rate of the pseudo-exponential tail is much longer at longer wavelengths.

Once the temporal fitting function had been developed, I re-analyzed the scans discussed in the previous section to determine how the various temporal fit parameters varied spatially over the array.

Recall the temporal fitting function that was introduced in Section 5.3. Equation 5.9 consisted of three terms for describing the Gaussian core, the pseudo-exponential tail and the background. For each point in the scan, the temporal profiles were corrected for first-photon biases following the procedure outlined in Section 5.3 and fit with Equation 5.9.

5.5.1 20 micron device

The two 20 micron scans consisted of 41 spatial points spaced approximately 1.5 microns apart. At each position, the detector was gated on and off 40,000 times.

668 nm response

Figure 5.30 shows the Gaussian amplitude, A, for the 20 micron device illuminated with 668 nm light. The amplitude has been fit with a "Bowler Hat" function convolved with a Gaussian laser spot. The black dashed lines are placed at the edges of the active area and are located in the same spot for Figures 5.30–5.34.

The error bars in Figure 5.30 and throughout the rest of the section represent the uncertainty in the temporal fit parameter, in this case A, determined by the least-squares routine. This is only one source of error. The 20 micron device was only scanned once with the 668 nm laser. If the scans were repeated, I would expect subsequent scans to scatter around the red best-fit line. This would effectively increase the size of the error bars and improve the goodness of fit. Because I have no way of evaluating the scan-to-scan error contribution, I will not report the χ^2 values since they will be severely biased by the under-estimated uncertainties.

Figure 5.31 shows the spatial dependence of the Gaussian mean, μ, for the 20 micron device illuminated with 668 nm light. The best-fit line is a parabola convolved with a Gaussian laser spot. If the scans were repeated, I would again expect the data to scatter around the best-fit line, effectively increasing the error bars and improving the goodness of fit. The center-to-edge difference in the mean is roughly 160 ps.

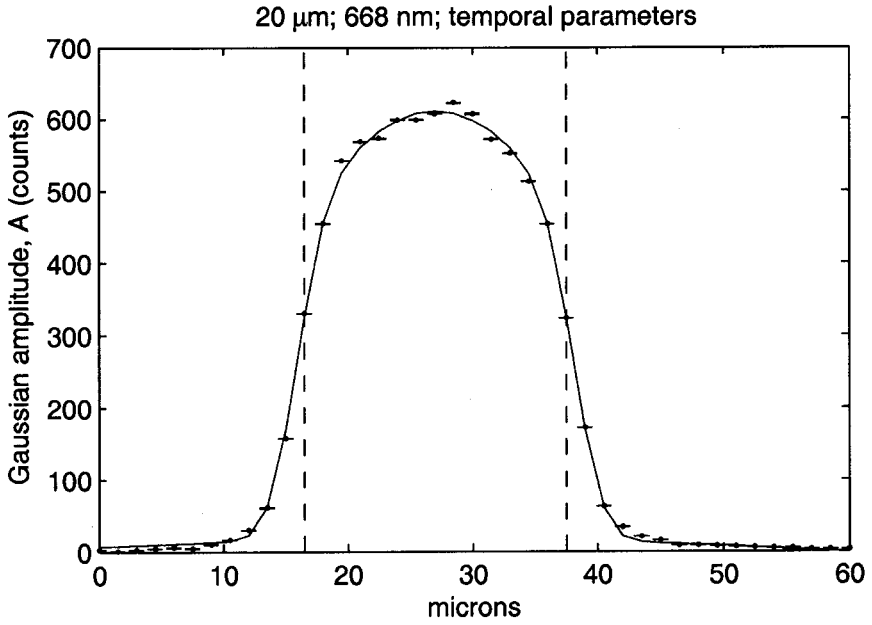


Figure 5.30: Spatial dependence of the Gaussian amplitude, A, for the 20 micron device at 668 nm (blue points). The fit function (red line) is a "Bowler Hat" convolved with a Gaussian laser spot. The black dotted lines represent the edge of the active area.

Figure 5.32 shows how the Gaussian spread, σ, varies over the 20 micron device. This spread varies between 100–120 ps and is consistent with the values determined in Section 5.3. Figure 5.33 shows the fit of the exponential amplitude, B. Near the center of the element, the exponential amplitude is on the order of 20% of the Gaussian amplitude. The decay constant, τ, of the exponential is shown in Figure 5.34. At 668 nm, the decay constant is on the order of several hundred picoseconds.

786 nm response

Figure 5.35 shows the Gaussian amplitude, A, for the 20 micron device illuminated with 786 nm light. As before the amplitude has been fit with a "Bowler Hat" function convolved with a Gaussian laser spot. The black dashed lines are placed at the edges of the active area and are located in the same spot for Figures 5.35–5.39.

Figure 5.36 shows the spatial dependence of the Gaussian mean, μ, for the 20 micron device illuminated with 786 nm light. The best-fit line is a parabola convolved with a Gaussian laser spot. The difference between the center and the edge is on the order of 150 ps.

Figure 5.37 shows how the Gaussian spread, σ, varies over the 20 micron device. This spread is roughly uniform over the entire element and varies between 120–125 ps. Figure 5.38 shows the fit of the exponential amplitude, B. Near the center of the element, the exponential amplitude is roughly 15% of the Gaussian amplitude. The decay constant, τ, of the exponential is shown in Figure 5.39. At 786 nm, the decay constant is on the order of a thousand picoseconds.

5.5.2 30 micron device

668 nm response

Figure 5.40 shows the Gaussian amplitude, A, for the 30 micron device illuminated with 668 nm light. As before, the amplitude has been fit with a "Bowler Hat" function convolved with a Gaussian laser spot. The black dashed lines are placed at the edges of the active area and are located in the same spot for Figures 5.40–5.44.

Figure 5.41 shows the spatial dependence of the Gaussian mean, μ, for the 30 micron device illuminated with 668 nm light. The best-fit line is a parabola convolved with a Gaussian laser spot. The difference between the center and the edge is roughly 200 ps.

Figure 5.42 shows how the Gaussian spread, σ, varies over the 30 micron device. This spread varies between 90–110 ps. Figure 5.43 shows the fit of the exponential amplitude, B. Near the center of the element, the exponential amplitude is on the order of 15% of the Gaussian amplitude. The decay constant, τ, of the exponential is shown in Figure 5.44. At 668 nm, the decay constant is on the order of several hundred picoseconds.

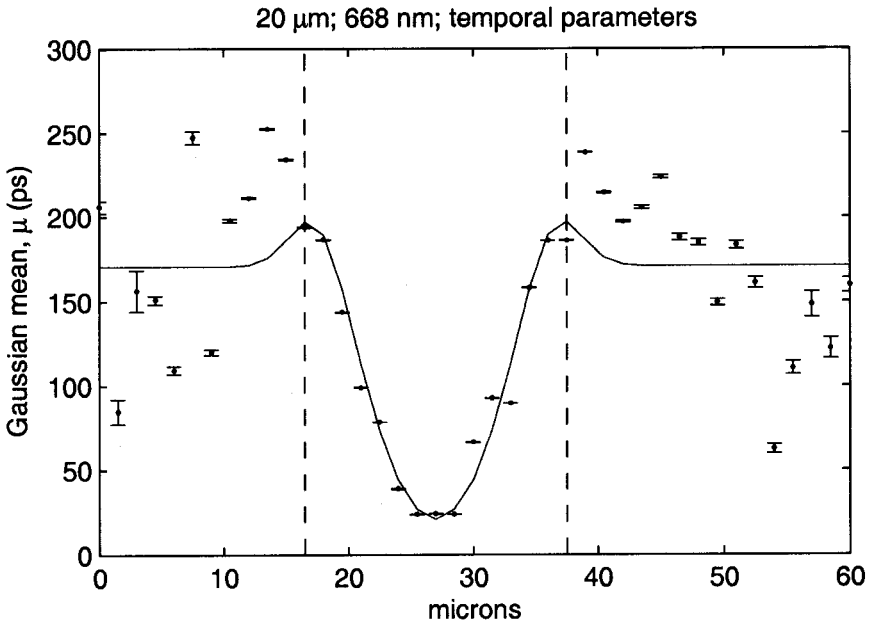


Figure 5.31: Spatial dependence of the Gaussian mean, μ, for the 20 micron device at 668 nm (blue points). The fit function (red line) is a parabola convolved with a laser spot. The black dotted lines represent the edge of the active area.

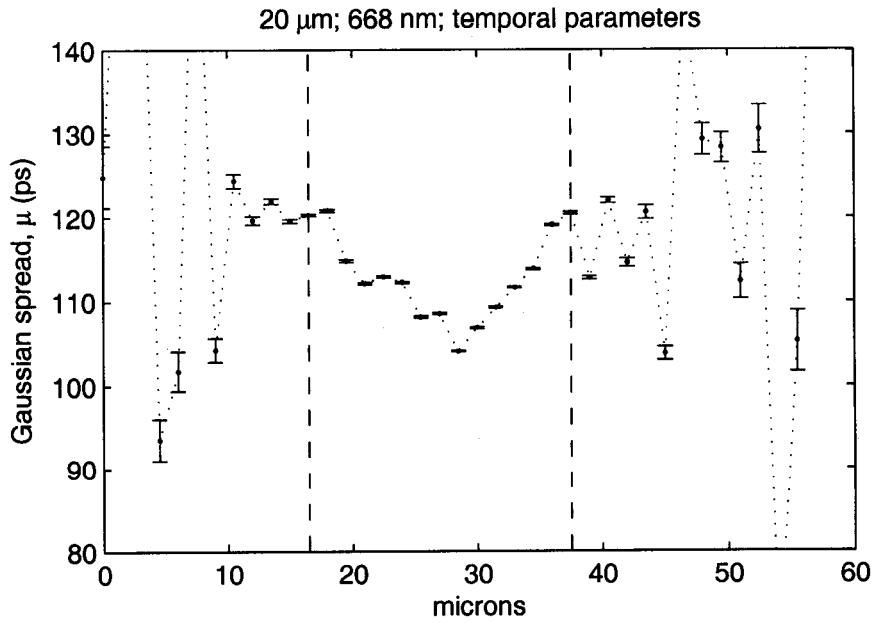


Figure 5.32: Spatial dependence of the Gaussian spread, σ, for the 20 micron device at 668 nm. The black dotted lines represent the edge of the active area.

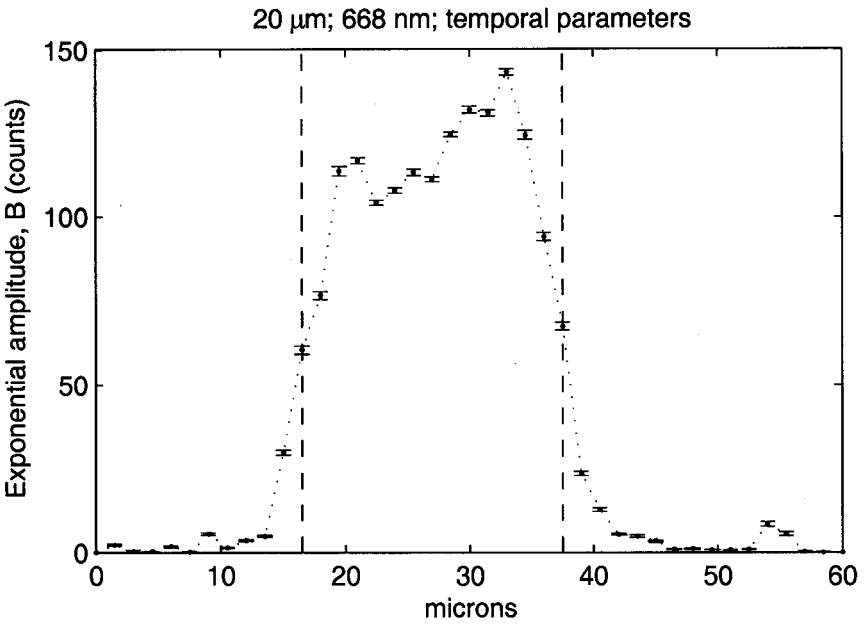


Figure 5.33: Spatial dependence of the exponential peak, B, for the 20 micron device at 668 nm. The black dotted lines represent the edge of the active area.

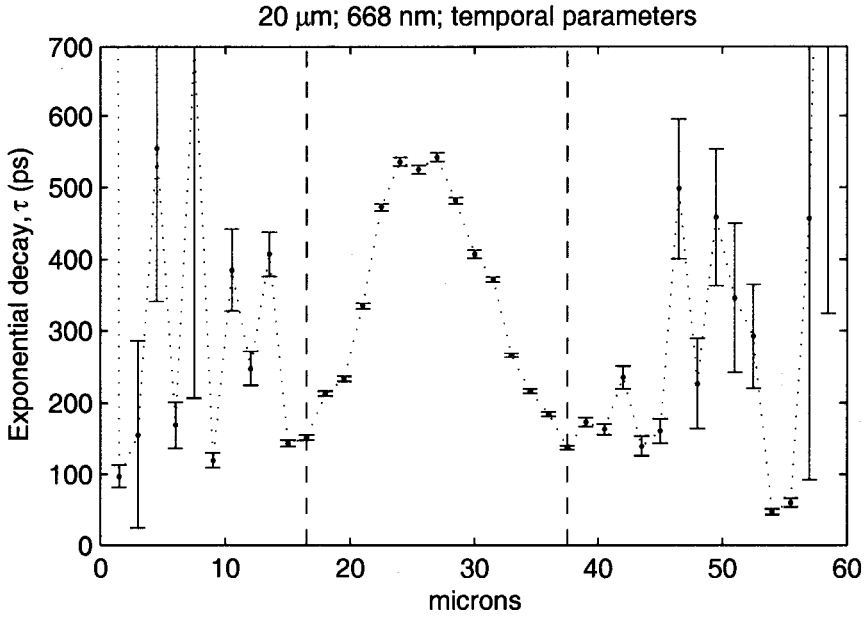


Figure 5.34: Spatial dependence of the exponential decay constant, τ, for the 20 micron device at 668 nm. The black dotted lines represent the edge of the active area.

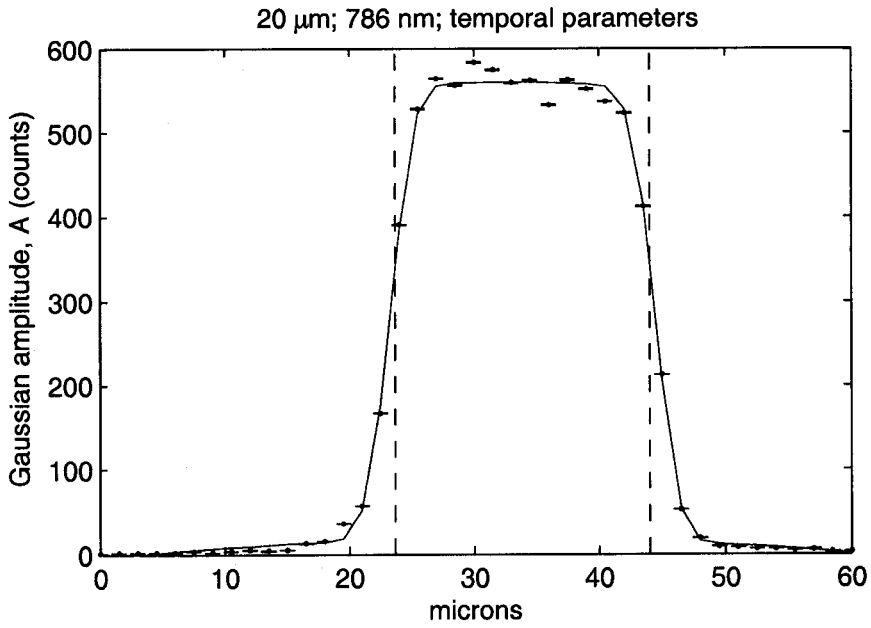


Figure 5.35: Spatial dependence of the Gaussian amplitude, A, for the 20 micron device at 786 nm (blue points). The fit function (red line) is a "Bowler Hat" convolved with a Gaussian laser spot. The black dotted lines represent the edge of the active area.

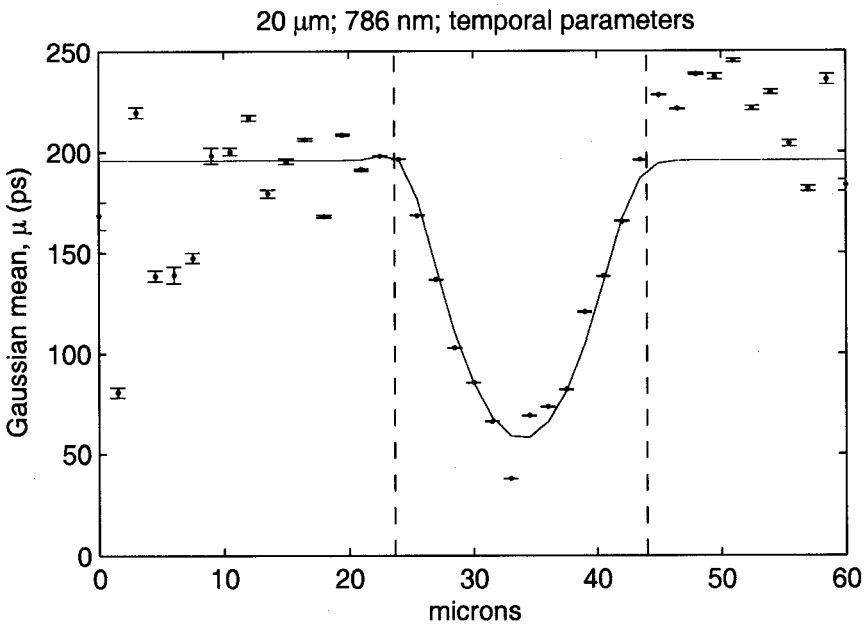


Figure 5.36: Spatial dependence of the Gaussian mean, μ, for the 20 micron device at 786 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dotted lines represent the edge of the active area.

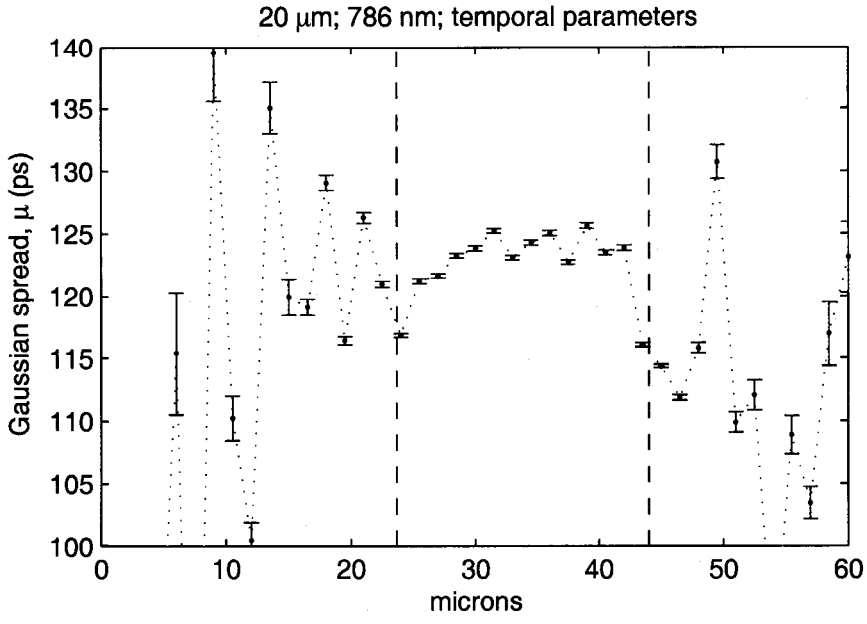


Figure 5.37: Spatial dependence of the Gaussian spread, σ, for the 20 micron device at 786 nm. The black dotted lines represent the edge of the active area.

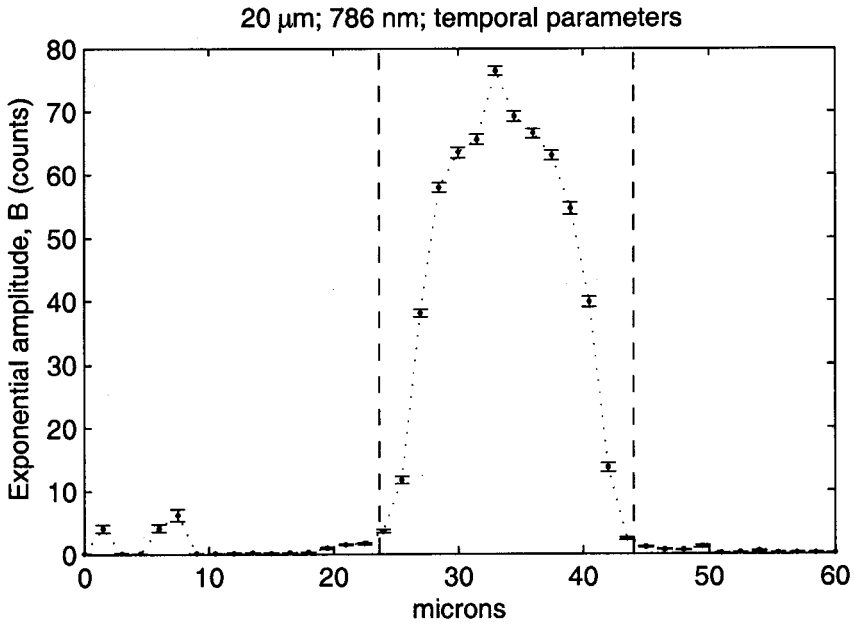


Figure 5.38: Spatial dependence of the exponential peak, B, for the 20 micron device at 786 nm. The black dotted lines represent the edge of the active area.

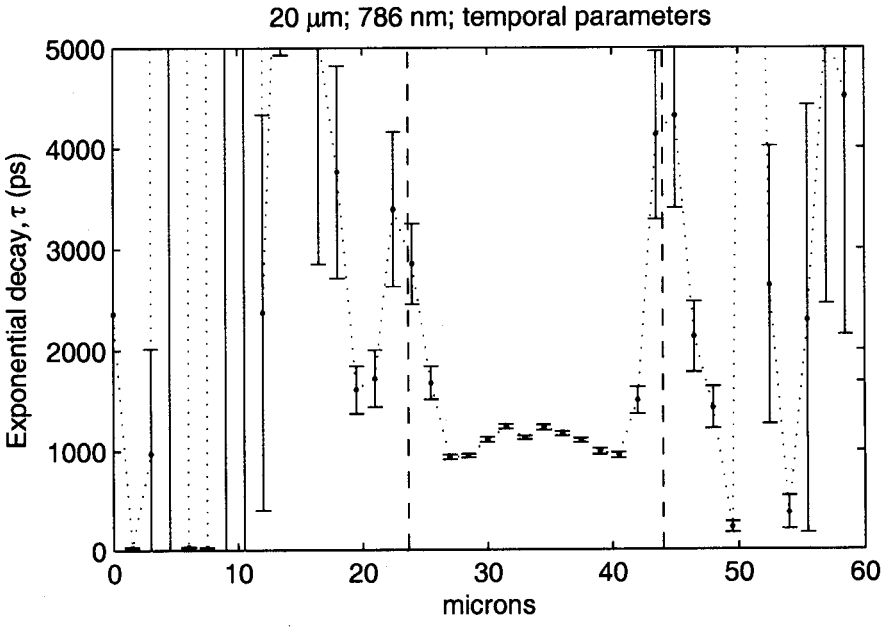


Figure 5.39: Spatial dependence of the exponential decay constant, τ, for the 20 micron device at 786 nm. The black dotted lines represent the edge of the active area.

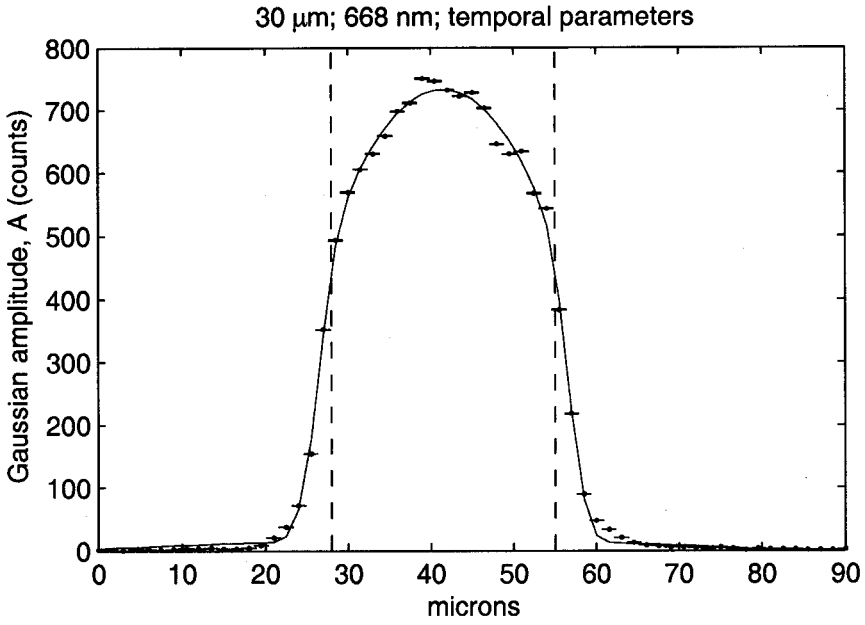


Figure 5.40: Spatial dependence of the Gaussian amplitude, A, for the 30 micron device at 668 nm (blue points). The fit function (red line) is a "Bowler Hat" convolved with a Gaussian laser spot. The black dashed lines represent the edges of the active area.

786 nm response

Figure 5.45 shows the Gaussian amplitude, A, for the 20 micron device illuminated with 668 nm light. As before the amplitude has been fit with a "Bowler Hat" function convolved with a Gaussian laser spot. The black dashed lines are placed at the edges of the active area and are located in the same spot for Figures 5.45–5.49.

Figure 5.46 shows the spatial dependence of the Gaussian mean, μ, for the 30 micron device illuminated with 786 nm light. The best-fit line is a parabola convolved with a Gaussian laser spot. The difference between the center and the edge is roughly 200 ps.

Figure 5.47 shows how the Gaussian spread, σ, varies over the 30 micron device. This spread varies between 90–120 ps. Figure 5.48 shows the fit of the exponential amplitude, B. Near the center of the element, the exponential amplitude is on the

order of 10% of the Gaussian amplitude. The decay constant, τ, of the exponential is shown in Figure 5.49. At 786 nm, the decay constant is on the order of a few thousand picoseconds.

5.5.3 Discussion

Errors and goodness of fit

As I briefly mentioned at the beginning of Section 5.5.1 the error bars shown in Figures 5.30–5.49 are the uncertainties in each parameter that were determined by the least-squares fitting routine. If the scans were repeated ten times, I would expect the uncertainties of the fit to remain about the same, while the values of the parameters would change. This scan-to-scan noise would effectively increase the error bars and increase the goodness of fits. Because we only have one scan for each device and wavelength combination, there is no way to estimate the scan-to-scan error contribution.

If there were more hits at each point in the scan, it would be possible to get a rough estimate of the scan-to-scan noise by breaking the data up into chunks and analyzing each separately. The scatter in the subsets of data could then be used to estimate the scan-to-scan uncertainty. Unfortunately, we only have 40,000 gates at each point in the scan and detected at most about 10,000 positive hits. If this is broken up into five equal subsets, each one would only have about 2000 hits spread out over 4096 temporal bins. Even when all 10,000 hits are considered there is only roughly 100 hits in any bin of the core. If this is decreased by a factor of five, then at most we'd only get 20 hits per temporal bin and the Poisson noise is on the order of 25%.

Because of the underestimation of the errors, χ^2 values for the fits are extremely high (on the order of several hundred to a thousand). However, the fits of the Gaussian amplitude, A, and the Gaussian mean, μ, do show the overall trends of the data and

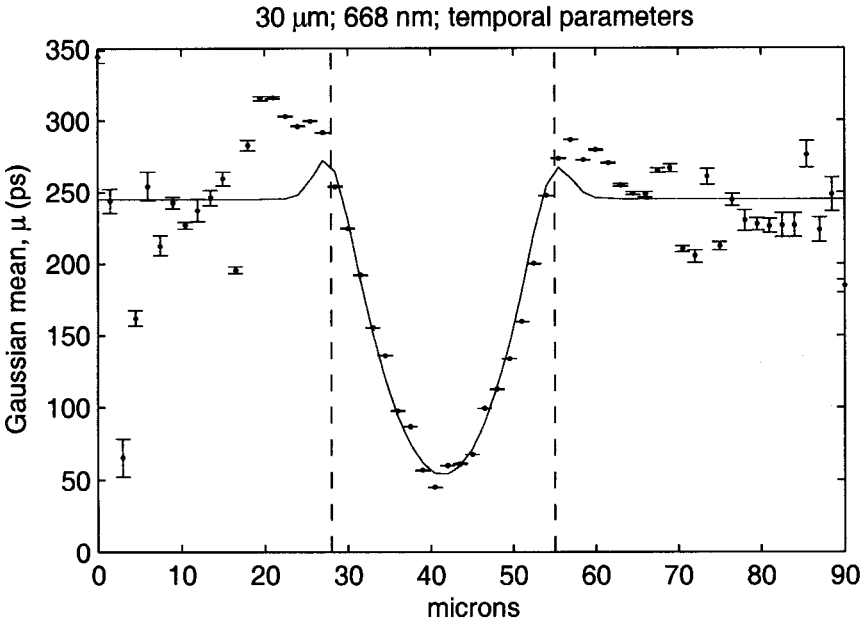


Figure 5.41: Spatial dependence of the Gaussian mean, μ, for the 30 micron device at 668 nm. The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines represent the edges of the active area.

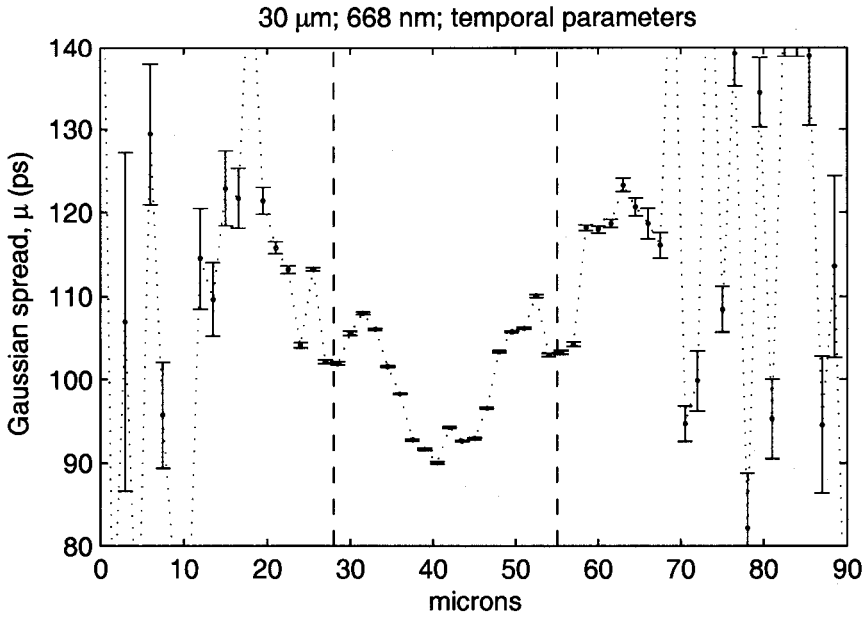


Figure 5.42: Spatial dependence of the Gaussian spread, σ, for the 30 micron device at 668 nm. The black dashed lines represent the edges of the active area.

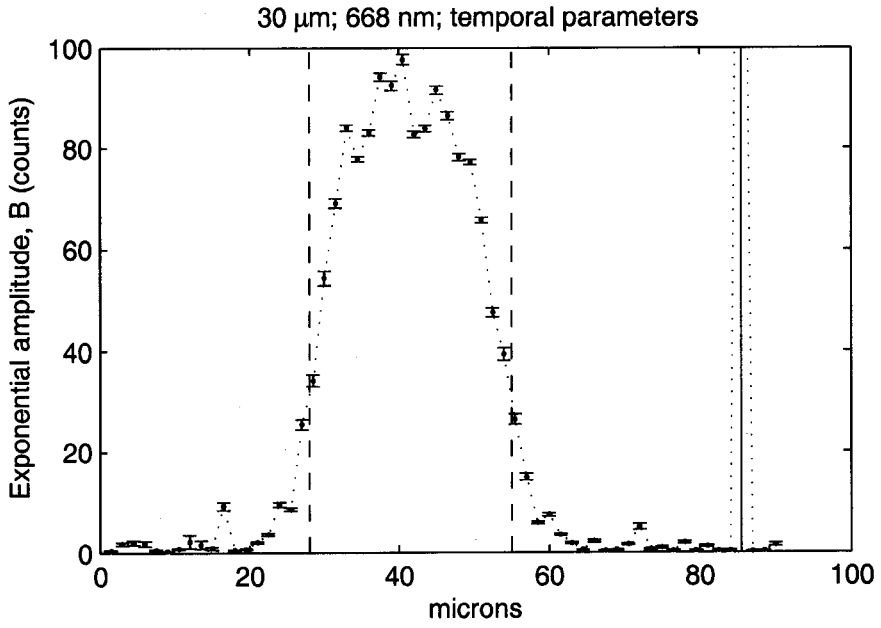


Figure 5.43: Spatial dependence of the exponential peak, B, for the 30 micron device at 668 nm. The black dashed lines represent the edges of the active area.

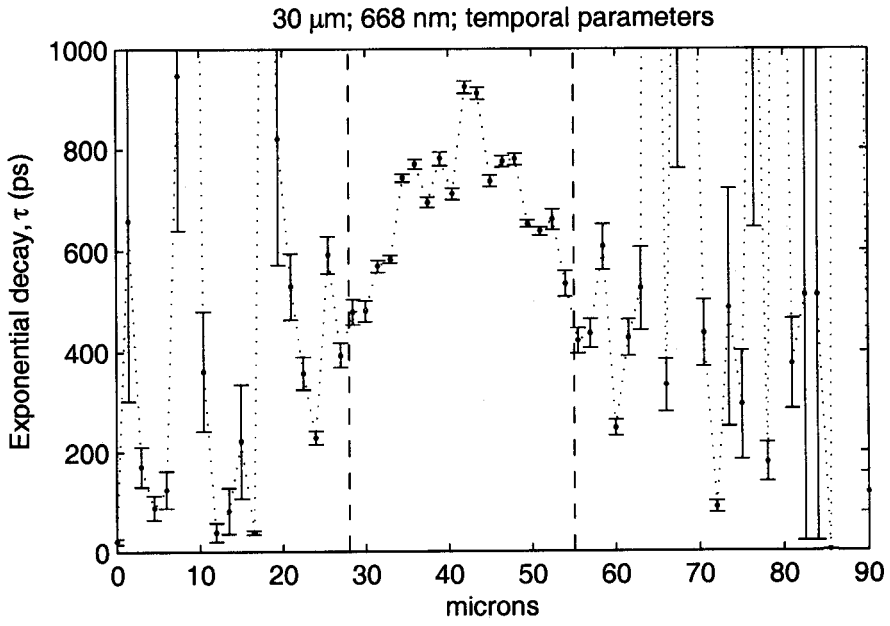


Figure 5.44: Spatial dependence of the exponential decay constant, τ, for the 30 micron device at 668 nm. The black dashed lines represent the edges of the active area.

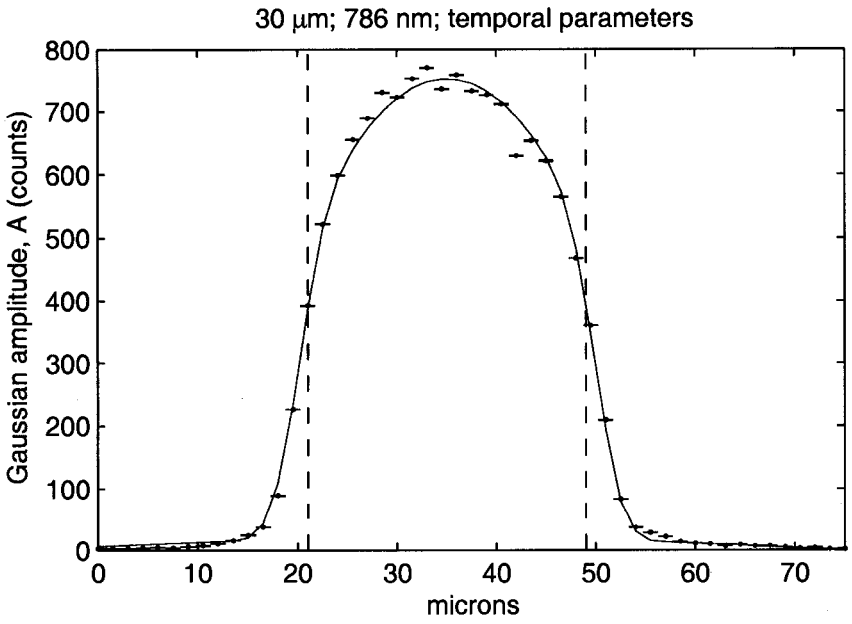


Figure 5.45: Spatial dependence of the Gaussian amplitude, A, for the 30 micron device at 786 nm (blue points). The fit function (red line) is a "Bowler Hat" convolved with a Gaussian laser spot. The black dashed lines represent the edges of the active area.

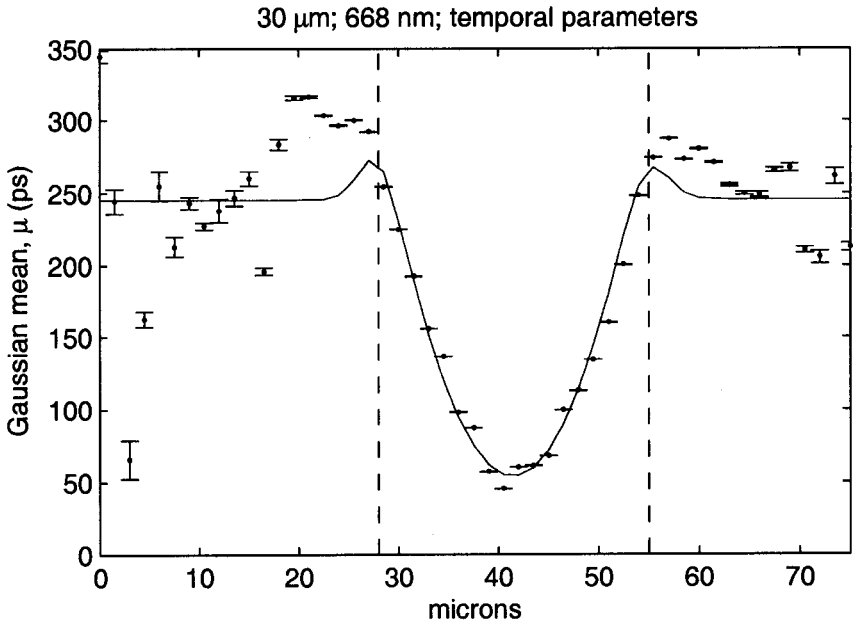


Figure 5.46: Spatial dependence of the Gaussian mean, μ, for the 30 micron device at 786 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines represent the edges of the active area.

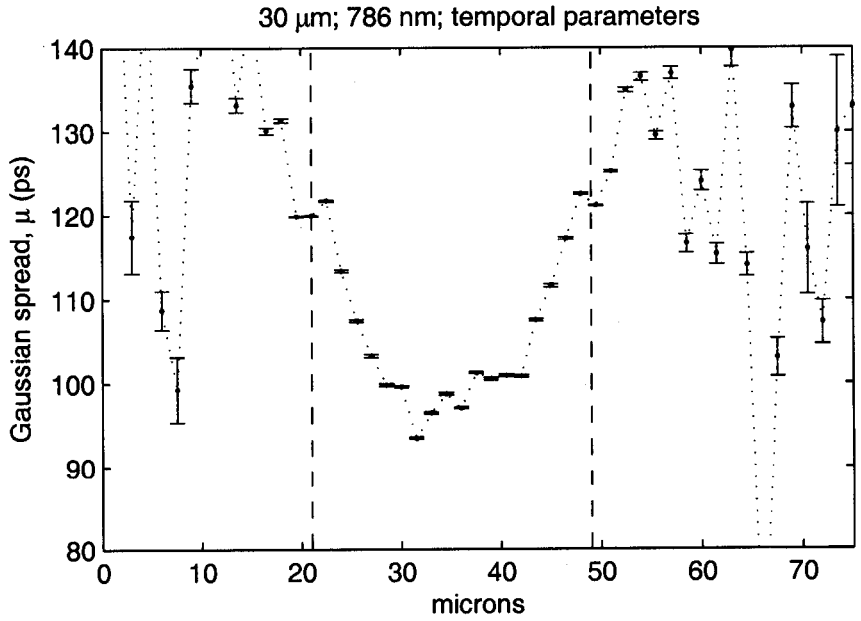


Figure 5.47: Spatial dependence of the Gaussian spread, σ, for the 30 micron device at 786 nm. The black dashed lines represent the edges of the active area.

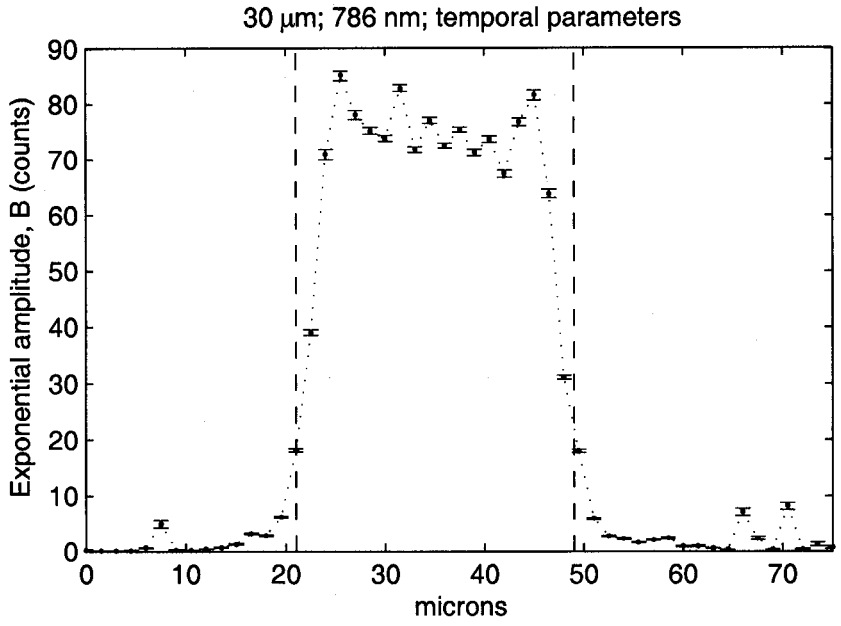


Figure 5.48: Spatial dependence of the exponential peak, B, for the 30 micron device at 786 nm. The black dashed lines represent the edges of the active area.

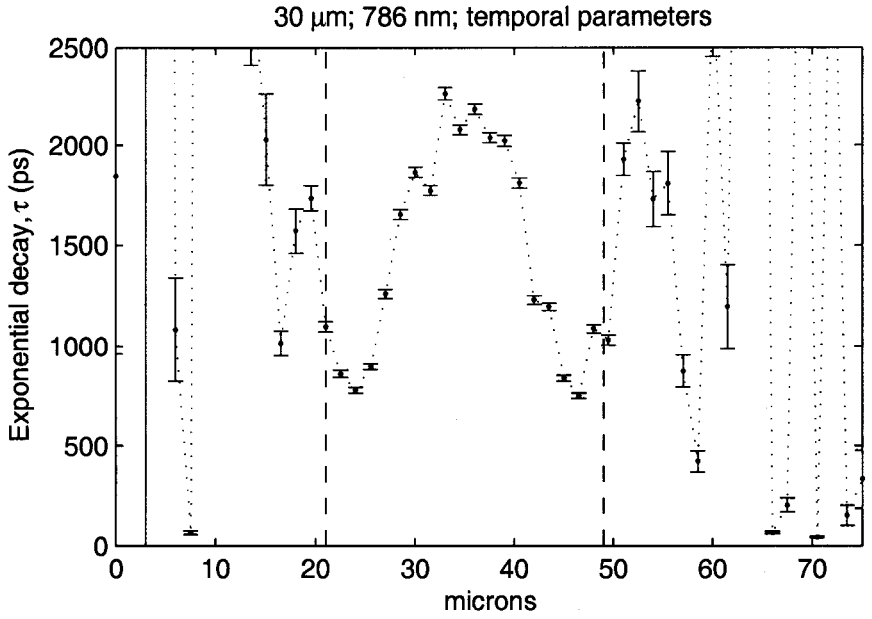


Figure 5.49: Spatial dependence of the exponential decay constant, τ, for the 30 micron device at 786 nm. The black dashed lines represent the edges of the active area.

are consistent between all the datasets. If more scans were available, it would be possible to gain a better understanding of the trends of the other temporal parameters. For instance, we would be able to see if the Gaussian spread, σ, tends to increase toward the edges of the element or if the exponential decay constant was significantly higher at the center of the element.

Temporal Walk of Gaussian core

One important conclusion from this data is that the mean of the Gaussian core changes significantly depending on where the avalanche is initiated. For the 20 micron device, the center-to-edge difference was on the order of 150 ps. For the 30 micron device, the temporal walk is roughly 200 ps between the center and the edge.

This temporal walk is not unexpected. Section 4.3 discussed the lateral growth of the avalanche and argued that the minimum center-to-edge difference would be 50 ps for the 20 micron device and 75 ps for the 30 micron device. This was an underestimate because it assumed that the lateral mean free path (MFP) of the electrons would remain 24.8 nm. In fact, the MFP is reduced in regions of high-electric field requiring more time for the avalanche to grow across the entire element. In addition, this estimate assumed that at each impact ionization event the charge carrier always moved deeper into the non-avalanching region of the detector. It should be noted there is also a mechanism for decreasing the edge-to-center difference. In Section 4.3 I considered the time for the avalanche to grow across the entire element or reach the peak current. However, the threshold of the comparator is likely set somewhere in the 30–70% level (i.e. the comparator will flip when the current reaches 30–70% of its peak value). A lower threshold could decrease the center-to-edge temporal walk, but could also decrease the slope at the threshold crossing and increase the error contribution of the timing electronics. Given these arguments, the observed temporal walk of the Gaussian core is reasonable and consistent with the simple model of APD performance we've been using thus far.

If the entire element is illuminated, the avalanche initiation location will cause an additional source of temporal jitter. An avalanche that is initiated near the edge of the element will take significantly longer to reach the peak current than an avalanche that begins in the center. The consequences of the avalanche initiation location will be discussed further in Section 6.2.

Width of Gaussian core

For all four device and wavelength combinations, the spread of the Gaussian core, σ, was 90–125 ps. This is consistent with the results of Section 5.3. Since we only have one scan for each combination, it is difficult to determine how σ changes spatially over the element. Figure 5.37 indicates that the spread of the Gaussian core is uniform over the element, while the other three datasets suggest that the temporal jitter is slightly lower near the center of the element. During the strong-signal calibrations I observed that the contribution of the timing electronics and TDC could vary by several tens of picoseconds from run to run (each run consisting of 1000-5000 gates). This contributes to the point-to-point differences in the spatial variation of the spread of the Gaussian core and makes it more difficult to determine any overall trends in the data.

There is also a slight correlation between the amplitude and spread of the Gaussian core. When the Gaussian amplitude exhibits negative curvature, the Gaussian spread exhibits positive curvature. In the case of Figures 5.35 and 5.37, neither one exhibits much curvature. In Section 5.6 we will see what this means for how the detections are distributed between the core and the tail.

Given these factors, it is reasonable to consider the RMS spread of the Gaussian core to be essentially uniform over the element. The values observed are also consistent with a 75 ps RMS error from the detector, the upper limit for the detector contribution determined in Section 5.3.

Decay rate of exponential tail

The decay rate of the pseudo-exponential tail, τ, is more than a factor of two larger for the 786 nm data than the 668 nm data. In Section 5.3 it was also determined that the tail decayed faster for shorter wavelengths. The difference in the magnitude of τ could be because of the much lower signal-to-noise in the scan data. When the laser spot was centered, there would be at most roughly 100 points in any given bin of the core. This means that the bins in the tail would contain less than 10 positive hits. It would become difficult, in that case, to separate out the pseudo-exponential tail from the background, especially after several tens of nanoseconds.

Summary

Fitting the temporal profiles obtained at different spatial locations within the element allows us to examine the trends in the device behavior. If several scans were available, a more detailed determination of the device behavior could be possible. The data indicate that the mean location of the Gaussian core can change by more than 100 ps for an avalanche initiated at the center vs. the edge of the element. This dependence on avalanche initiation location introduces another source of temporal jitter when the entire element is illuminated. This will be discussed further in Section 6.2. The spread of the Gaussian core is on the order of 90–125 ps RMS and is essentially uniform over the element. These observations are consistent with the 75 ps upper limit for the detector's contribution to the error budget determined in Section 5.3. The pseudo-exponential tail decays faster for shorter wavelengths and is consistent with a diffusion process.

5.6 Distribution of detected photons

- The integrated flux in the Gaussian core is uniform over the detector and is well characterized by a "Top Hat" convolved with a Gaussian laser pulse.

- The integrated flux in the tail falls off dramatically toward the edges of the element and is well characterized by a parabola.

- Comparison between the flux in the core relative to the tail, and the numerical simulations, result in a $\sim$ 3 micron depth for the drift/diffusion region interface when the detectors are biased $\sim$ 3.5–4 V beyond their breakdown voltage.

At this point it is important to recall the big picture. Temporal information from the return photons will be used to measure the distance between the earth and moon with millimeter precision. The retro-reflector arrays and the laser pulse each contribute up to 300 ps and 40 ps of RMS jitter respectively. Ideally, we would like the detectors to contribute less than 40 ps of jitter in order to have some hope of characterizing the laser pulse.

By looking at the temporal profiles, we see that the temporal jitter of the detectors is probably greater than 40 ps. The Gaussian cores have an RMS spread of 100–125 ps and suggest an upper limit on the detector contribution of 75 ps RMS and an estimated value of around 50 ps. There are also long pseudo-exponential tails that can last for several tens of nanoseconds. The behavior of these tails is consistent with a diffusion-driven process. As the wavelength of incident light is decreased, the penetration depth of the photons should also decrease and the number of detections, or the power in the tail, should also decrease.

In order to study this phenomenon, we began by examining the number of detections, or the flux, in the Gaussian core relative to the pseudo-exponential tail. The

number of detections in the core can be determined by

$$P_{core} = \int_{-\infty}^{\infty} A \exp\left[-\frac{t^2}{2\sigma^2}\right] = A\sigma\sqrt{2\pi} \tag{5.14}$$

where A is the Gaussian amplitude and σ is the Gaussian RMS spread. The number of detections in the tail can be determined by

$$P_{tail} = \int_{0}^{\infty} B(1 - \exp[-t]) \exp\left[\left(-\frac{t}{\tau}\right)^{0.5}\right] \tag{5.15}$$

where B is the exponential peak and τ is the decay constant. There is no simple analytical solution to Equation 5.15, so it was determined using numerical quadrature integration.

5.6.1 Spatial distribution of power

The high-resolution scans of the 20 and 30 micron devices discussed in Sections 5.4–5.5 were used to examine how the number of detections in the core versus the tail varied spatially over the elements.

20 micron device

Figure 5.50 shows the spatial dependence of P_{core} for the 20 micron device illuminated with 668 nm light. The fitting function is a uniform "Top Hat" response convolved with a Gaussian laser spot. Figure 5.51 shows the spatial dependence of P_{tail} for the same dataset. This time the fitting function is a parabola convolved with a Gaussian laser spot. Figure 5.52 shows the ratio of P_{core} and P_{tail}.

The $\pm 1\sigma$ error bars in all three of these plots (as well as those shown in Figures 5.53 - 5.61) were determined using Equations 5.14 and 5.15 and standard error-propagation techniques such as those discussed by Bevington and Robinson (2003), namely

$$\sigma_f^2 = \sigma_x^2 \left(\frac{\partial f}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial f}{\partial y}\right)^2 + \cdots \tag{5.16}$$

where $f(x, y, \ldots)$ and σ_x and σ_y are the standard deviation of x and y respectively.

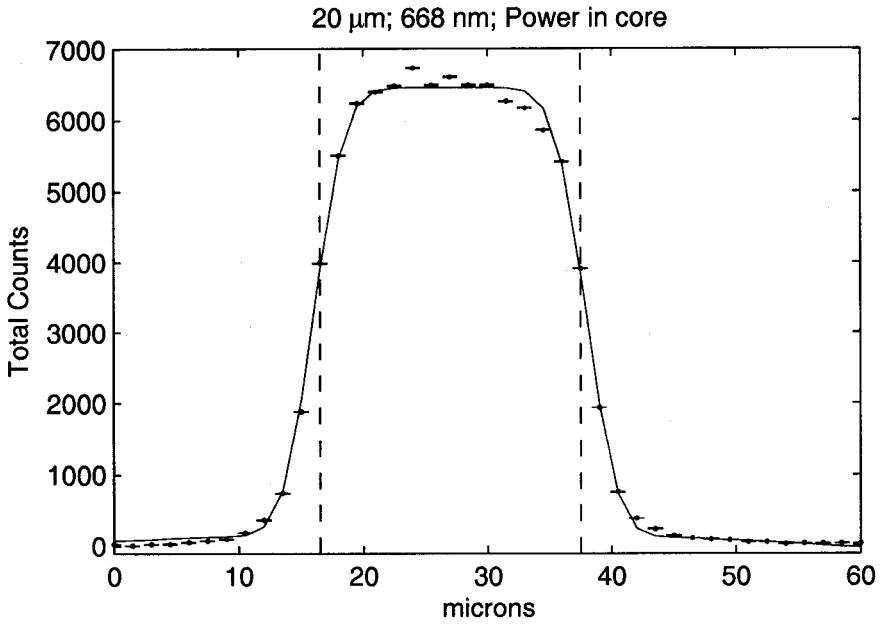


Figure 5.50: Spatial dependence of P_{core} for the 20 micron device at 668 nm (blue points). The fit function (red line) is a "Top Hat" convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

Because the errors of each of the fitted parameters are underestimates, the propagated errors in P_{core}, P_{tail} and the ratio of the two are also underestimates. The use of these estimates results in very large determinations of χ^2. If multiple scans were available, I would expect them to scatter about the best-fit lines.

Figure 5.53 shows the spatial dependence of P_{core} for the 30 micron device illuminated with 786 nm light. The fitting function is a uniform "Top Hat" response convolved with a Gaussian laser spot. Figure 5.54 shows the spatial dependence of P_{tail} for the same dataset. This time the fitting function is a parabola convolved with a Gaussian laser spot. Figure 5.55 shows the ratio of P_{core} and P_{tail}.

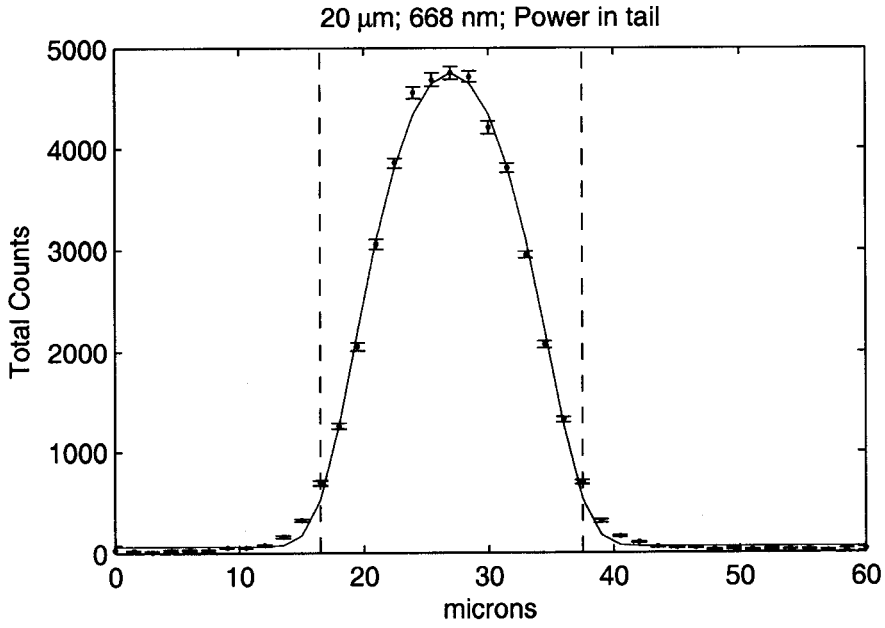


Figure 5.51: Spatial dependence of P_{tail} for the 20 micron device at 668 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

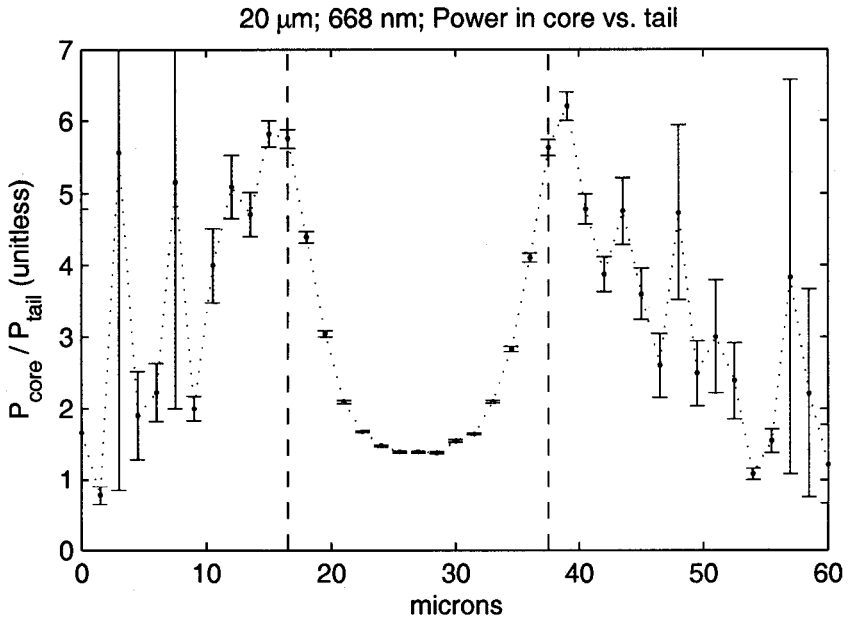


Figure 5.52: Spatial dependence of the power in the core relative to the power in the tail for the 20 micron device at 668 nm. The black dashed lines indicate the edges of the active area.

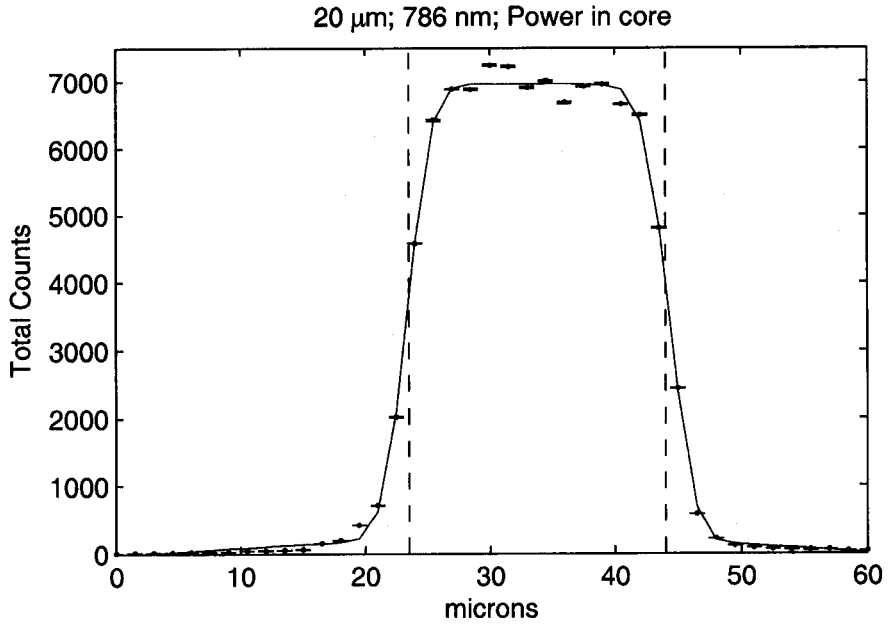


Figure 5.53: Spatial dependence of P_{core} for the 20 micron device at 786 nm (blue points). The fit function (red line) is a "Top Hat" convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

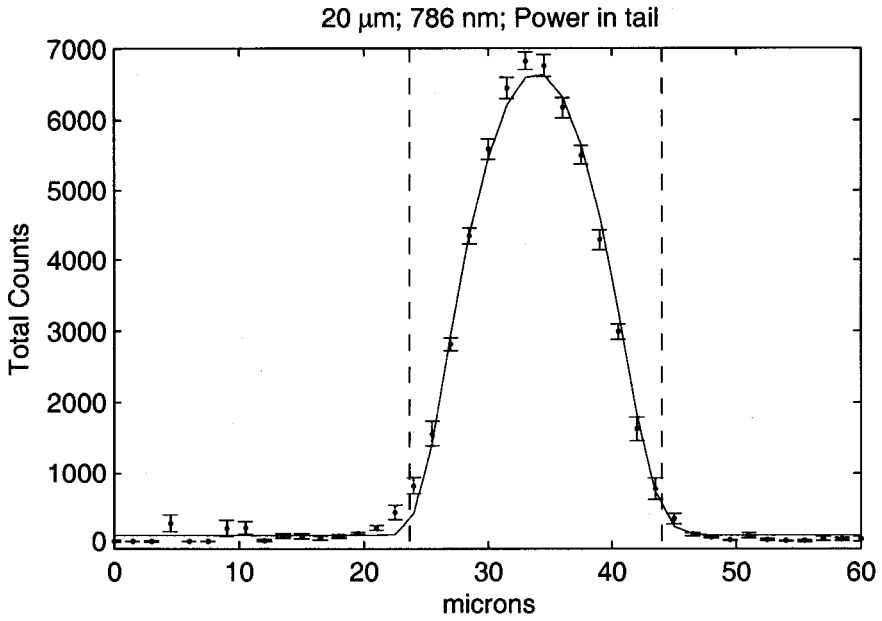


Figure 5.54: Spatial dependence of P_{tail} for the 20 micron device at 786 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

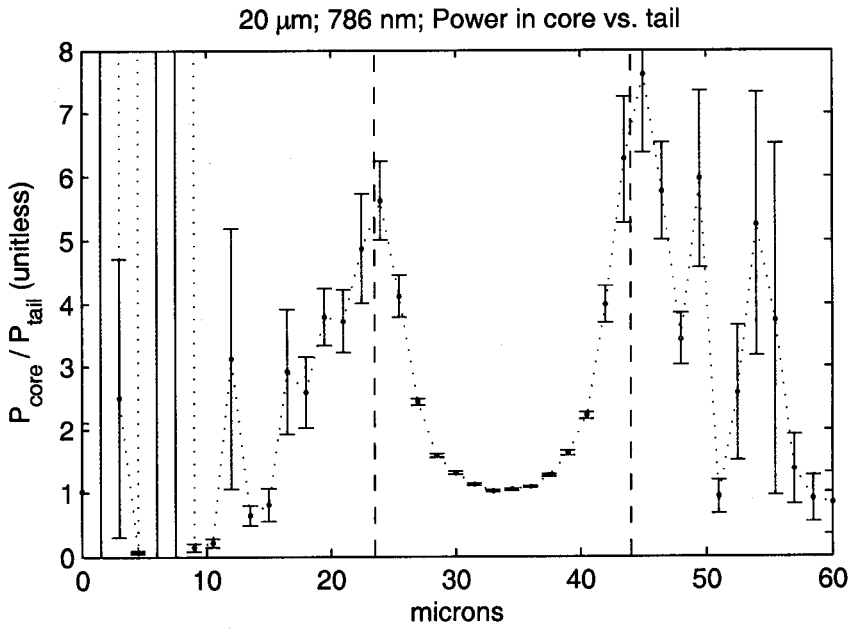


Figure 5.55: Spatial dependence of the power in the core relative to the power in the tail for the 20 micron device at 786 nm. The black dashed lines indicate the edges of the active area.

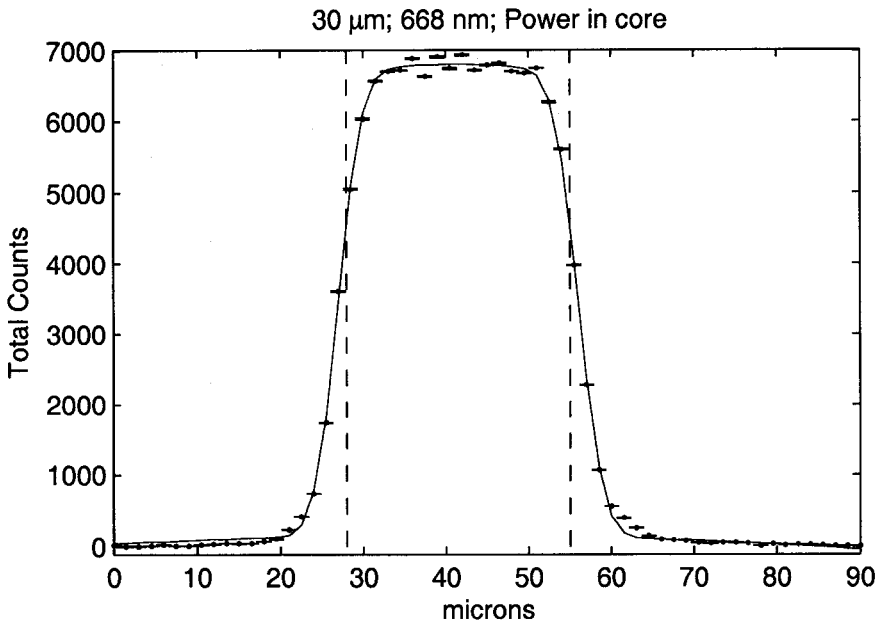


Figure 5.56: Spatial dependence of P_{core} for the 30 micron device at 668 nm (blue points). The fit function (red line) is a "Top Hat" convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

30 micron device

Figure 5.56 shows the spatial dependence of P_{core} for the 30 micron device at 668 nm and is fit by a uniform "Top Hat" response convolved with a Gaussian laser spot. Figure 5.57 shows the spatial dependence of P_{tail} for the same dataset. The fitting function is a parabola convolved with a Gaussian laser spot. Figure 5.58 shows the ratio of P_{core} and P_{tail}.

Figure 5.59 shows the spatial dependence of P_{core} for the 30 micron device illuminated with 786 nm light. The fitting function is a uniform "Top Hat" response convolved with a Gaussian laser spot. Figure 5.60 shows the spatial dependence of P_{tail} for the same dataset. This time the fitting function is a parabola convolved with a Gaussian laser spot. Figure 5.61 shows the ratio of P_{core} and P_{tail}.

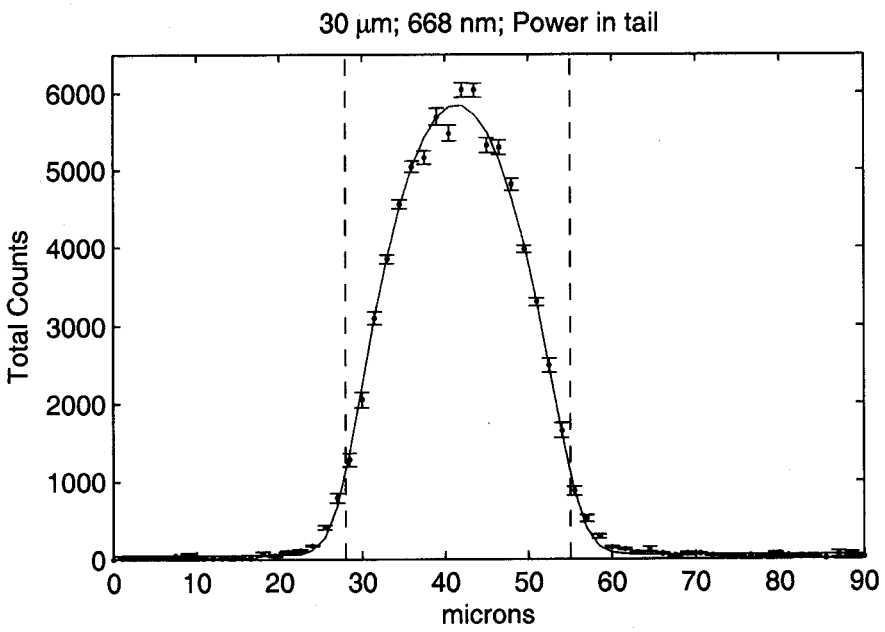


Figure 5.57: Spatial dependence of P_{tail} for the 30 micron device at 668 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

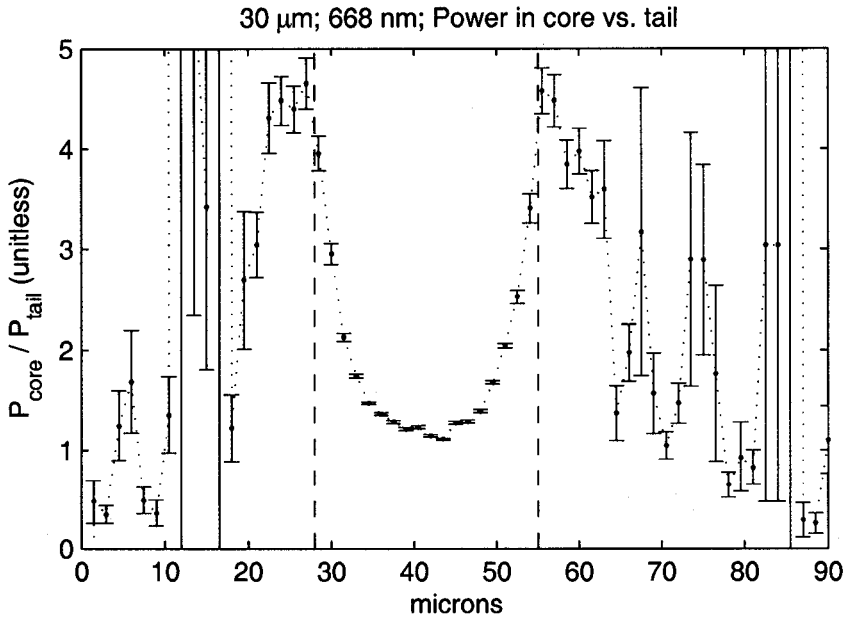


Figure 5.58: Spatial dependence of the power in the core relative to the power in the tail for the 30 micron device at 668 nm. The black dashed lines indicate the edges of the active area.

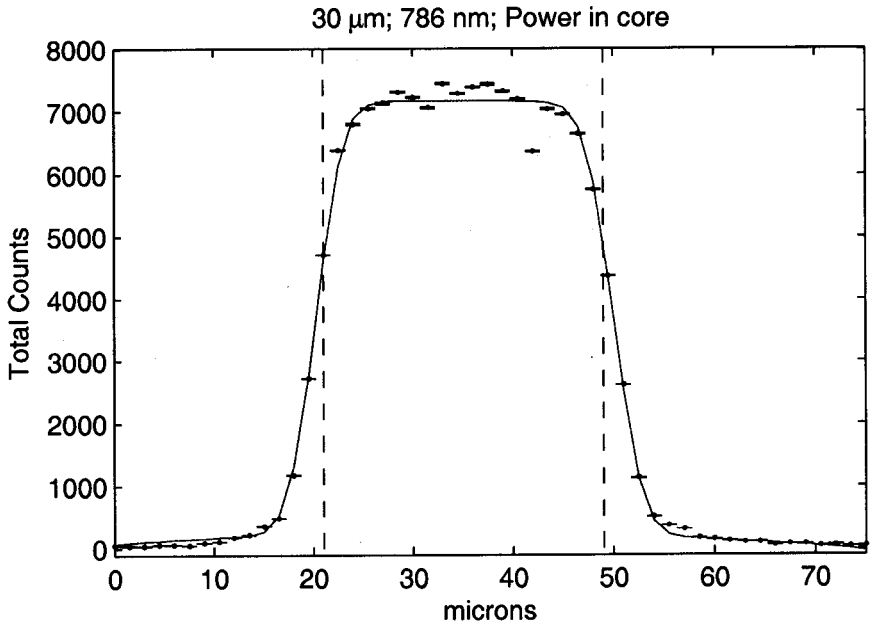


Figure 5.59: Spatial dependence of P_{core} for the 30 micron device at 786 nm (blue points). The fit function (red line) is a "Top Hat" convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

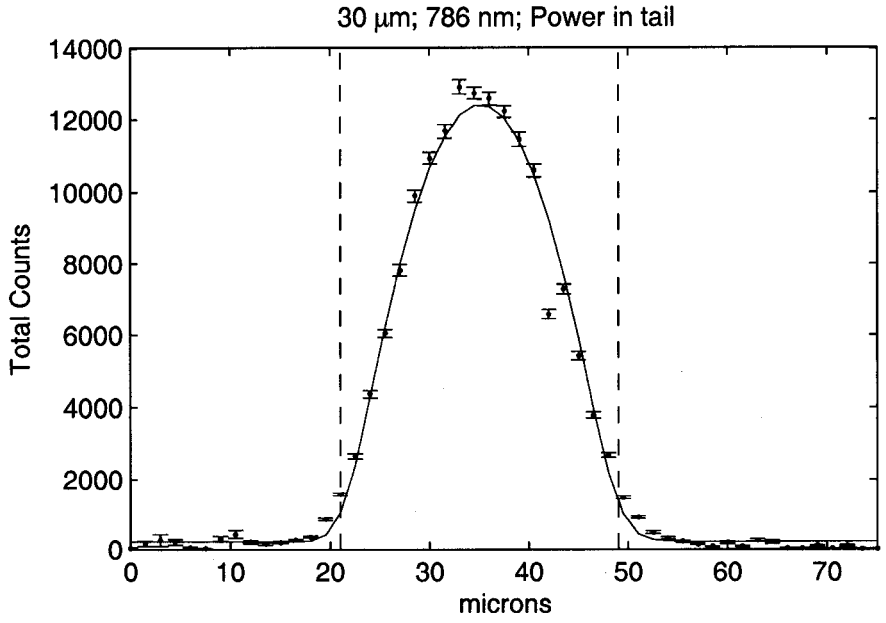


Figure 5.60: Spatial dependence of P_{tail} for the 30 micron device at 786 nm (blue points). The fit function (red line) is a parabola convolved with a Gaussian laser spot. The black dashed lines indicate the edges of the active area.

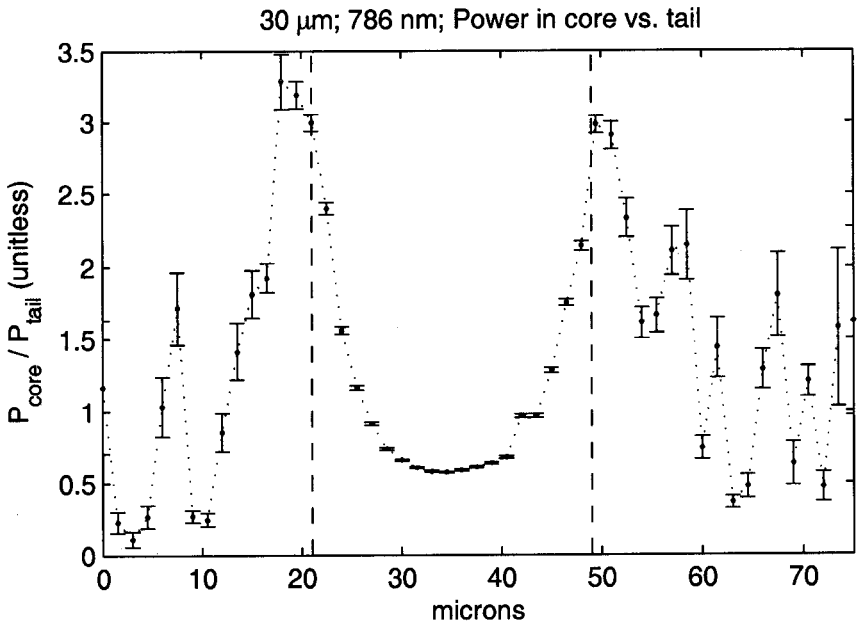


Figure 5.61: Spatial dependence of the power in the core relative to the power in the tail for the 30 micron device at 786 nm. The black dashed lines indicate the edges of the active area.

5.6.2 Model comparison

We can now use the simple model of the APD discussed in Section 3.1 and Chapter 4 to estimate z_{back}, the back of the drift region. Photo-generated electrons above this depth will be swept into the avalanche region and detected. Below this depth there is no electric field and electrons can only reach the avalanche region by diffusion. For the time being, I will assume that the avalanche probability is one. In that case, every electron that makes it to the drift region will eventually avalanche. The numerical simulations discussed in Section 4.4 can be used to determine the detection probability for photons that photo-convert at depths greater than z_{back}. Combining the results of the numerical simulations shown in Figures 4.4 and 4.6, and the uniform avalanche probability discussed above, we arrive at Figure 5.62 which shows how the avalanche probability depends on the depth at which photons photo-convert. For this figure, z_{back} is 3 microns. Photo-generated electrons within 0.5 microns of the surface will not be able to cross the avalanche region and initiate an avalanche. This causes the avalanche probability very near the surface to drop to zero.

The avalanche probability shown in Figure 5.62 can be combined with the penetration depth of photons as a function of wavelength to determine how many photons arrive in the core (above z_{back}) and in the tail (below z_{back}). The particular value of z_{back} can be adjusted until the ratio of the P_{core} to P_{tail} is equal to what we observe.

Because the simulations were only done on the axis of the detector, we will only consider the behavior at the center of the element. We have ten different datasets that we can analyze: six centered long-exposures, and the center of four short-exposure scans. Table 5.6 shows the results of this analysis on each of these ten datasets.

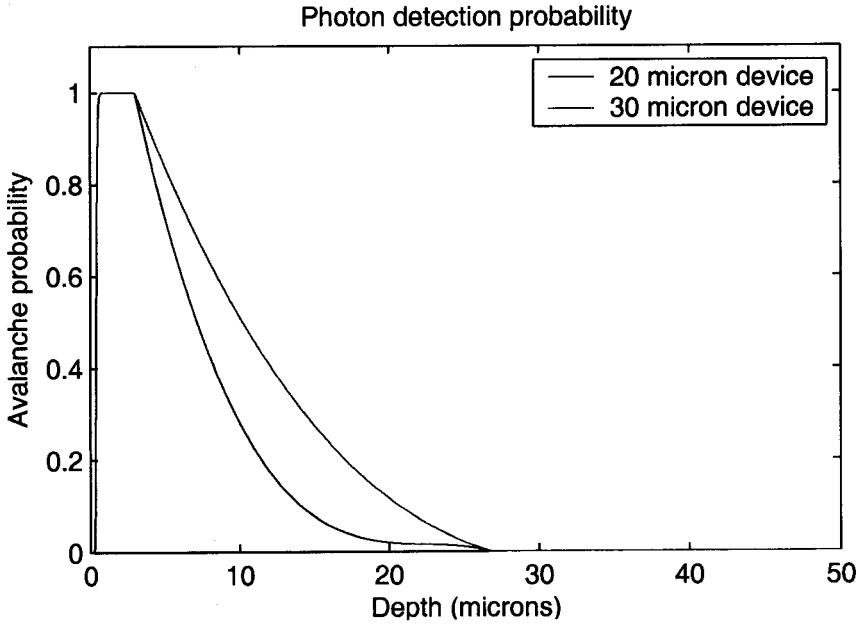


Figure 5.62: Probability of detecting a photon as a function of the depth at which it photo-converts

5.6.3 Discussion

Flux in Gaussian core

The number of detections, or power, in the Gaussian core is well described by a uniform "Top Hat" distribution convolved with a Gaussian laser pulse. The slight correlation between the Gaussian amplitude, A, and the Gaussian spread, σ, that was discussed in Section 5.5.3 combines in such a way that the total number of counts in the core is uniform over the element.

Flux in pseudo-exponential tail

The number of counts in the pseudo-exponential tail falls off dramatically toward the edges of the element and is well described by a parabola. It is the parabolic response of the tail that contributed to the "Bowler Hat" fits of the spatial response discussed

Table 5.6: Flux distributed in core versus tail

Diameter (μm)	λ (nm)	Gates	P_{core} (counts)	P_{tail} (counts)	$\frac{P_{core}}{P_{tail}}$	z_{back} (μm)
20	668	43.2e6	6.340e6	2.995e6	2.117	3.39
20	668	7.2e6	1.023e6	6.566e5	1.558	2.82
20	668	4.0e4	6494	4679	1.388	2.62
20	786	43.2e6	6.281e6	5.282e6	1.189	3.89
20	786	4.0e4	7218	6446	1.120	3.73
30	668	43.2e6	6.648e6	3.903e6	1.703	3.32
30	668	7.2e6	1.256e6	1.250e6	1.004	2.40
30	668	4.0e4	6928	6030	1.149	2.61
30	786	7.2e6	1.307e6	1.977e6	0.6610	3.17
30	786	4.0e4	7440	1.292e4	0.5759	2.85

in Section 5.4. If the depth of the drift region was increased, or the wavelength of the light decreased such that very few of the photons penetrated into the drift region, then the response of the detectors should be essentially uniform over the element.

Estimate of z_{back}

Using the simple model developed and discussed in earlier chapters, I was able to estimate the depth of z_{back}, the back of the drift region and the cutoff between the Gaussian core and pseudo-exponential tail. The value of z_back ranged from 2.5–4 microns over the ten datasets considered. For the 20 micron device, $z_{back} = 3.29 \pm 0.56$. For the 30 micron device, $z_{back} = 2.87 \pm 0.38$. This has important consequences for device performance at 532 nm and will be discussed in more detail in Section 6.2.

Summary

The number of electrons contributing to the Gaussian core is essentially uniform over the element. The number of electrons that contribute to the diffusion tail falls off quadratically toward the edges of the element. Using the simple model developed in earlier chapters, it is estimated that z_{back} was $\sim$ 3 microns for these devices.

Chapter 6

CONCLUSIONS

6.1 Summary of APD characterization

The temporal response of the APDs are reasonably well described by a Gaussian core and a pseudo-exponential, diffusion-driven tail. The Gaussian core has an RMS width of 100–125 ps and is essentially uniform over the element. This a convolution of the laser pulse, detector response, electronics and TDC and is independent of wavelength (668 nm or 786 nm) and device used (the 20 or 30 micron diameter). Estimating the error contributions of the laser pulse and electronics, an upper limit of 75 ps RMS is placed on the detector response width, though the true value is likely to be closer to 50 ps. The diffusion tail falls off faster for shorter wavelengths. Spatial scans of the elements revealed that the Gaussian core shifts by more than 150 ps depending on where the avalanche is initiated within the element (i.e. the center or near the edge).

An analysis of the distribution of detected photons revealed that the integrated power or number of counts in the Gaussian core is uniform over the element. On the other hand, the power in the tail falls off towards the edges of the element and is well characterized by a parabola. If the excess voltage is increased, or the wavelength of light is decreased, fewer photons will convert in the diffusion region—reducing the parabolic contribution of the tail. In that case, the response of the devices should become more and more uniform. Comparison between the amount of power in the core vs. the tails and the numerical simulations suggest a 3 micron depth for the drift/diffusion region interface when the detectors are biased roughly 3.5–4 V beyond their breakdown voltage.

6.2 Predictions of 532 nm temporal response

6.2.1 Temporal response

Based on the results described in Section 5.6 the drift region extends roughly 3 microns beneath the surface. Given that the penetration depth of 532 nm photons is on the order of one micron, fewer than 5% will penetrate into the diffusion region. If the excess voltage is increased, the depth of the drift region will also increase and further reduce the effect of any diffusion tail. The data discussed in this simulation were collected with an excess voltage of roughly 3.5–4 V. This was limited by the electronic design. When the new circuit boards are fabricated, they should be capable of producing an excess voltage of 5–7 V. Experimental results indicate that the APDs require 1.6 V of excess voltage before the drift region starts to penetrate into the π region, located roughly 1.3 microns below the surface. After that, the depth of the electric field increases by one micron for every volt of excess voltage. Given 5–7 V of excess voltage, the drift region should extend roughly 5–7 μm below the surface and less than 1% of the photons should penetrate into the diffusion region.

It should also be noted that a significant fraction (roughly 30 %) of the 532 nm photons will convert early in the n layer—within 0.4 microns of the surface. There is no electric field in this region so the photo-converted electron-hole pairs diffuse isotropically. Because of the direction of the electric field, only the holes which diffuse into regions of non-zero electric field will traverse the avalanche region and have the possibility of detection. Since the avalanche process in silicon is dominated by electrons, both the quantum efficiency and the avalanche probability are decreased, resulting in a two-fold reduction in the detection efficiency at 532 nm.

Assuming that the bias voltage is increased so that V_{excess} is on the order of 5–7 V, the diffusion tail can be neglected, resulting in a Gaussian temporal response. The results discussed in Sections 5.3 and 5.5 place an upper limit of 75 ps on the APD's contribution to the temporal jitter. This appears to be essentially uniform over the

element. Recall that this is a worst case scenario, but because of the variation in the electronics and the total jitter, it is difficult to determine tighter constraints on the detector performance.

The 532 nm laser that will be installed in the APOLLO apparatus is thought to have a Gaussian pulse profile with an RMS of 40 ps. Indications are that the contribution of the detectors and timing electronics will likely be greater than this, but by less than a factor of two. As a result, it will probably not be possible to characterize the 532 nm laser pulse, but any strong asymmetries or after-pulsing should be apparent in the convolved response.

6.2.2 Spatial response

The spatial response of temporal parameters also has consequences for the APOLLO performance. As discussed earlier, the diffusion tail should be negligible at 532 nm, especially if the excess voltage is increased to 5–7 V. All the power of the response will be from the Gaussian core which is observed to be uniform over the element.

The profile of the Gaussian response of the detectors has an RMS width of under 75 ps; however, the mean value of this Gaussian changes spatially over the element. Avalanches that are initiated near the center of the element reach the peak current faster than those that are initiated near the edges of the detector. If the entire area of the detector is illuminated, this will add an additional component to the per-photon jitter that could be greater than 75 ps.

In Section 5.5.2 we saw that the Gaussian mean can change by as much as 250 ps between the edges and the center of the 30 micron element. The parabolic fitting function shown in red on Figure 5.41 can be used to determine analytically the temporal jitter associated with this wandering mean. Figure 6.1 shows this analytical result. If the entire 30 micron array is illuminated, the avalanche initiation location contributes 90 ps to the total temporal uncertainty. This particular component could be reduced by changing the spot size at the detector. For instance, if all of the light

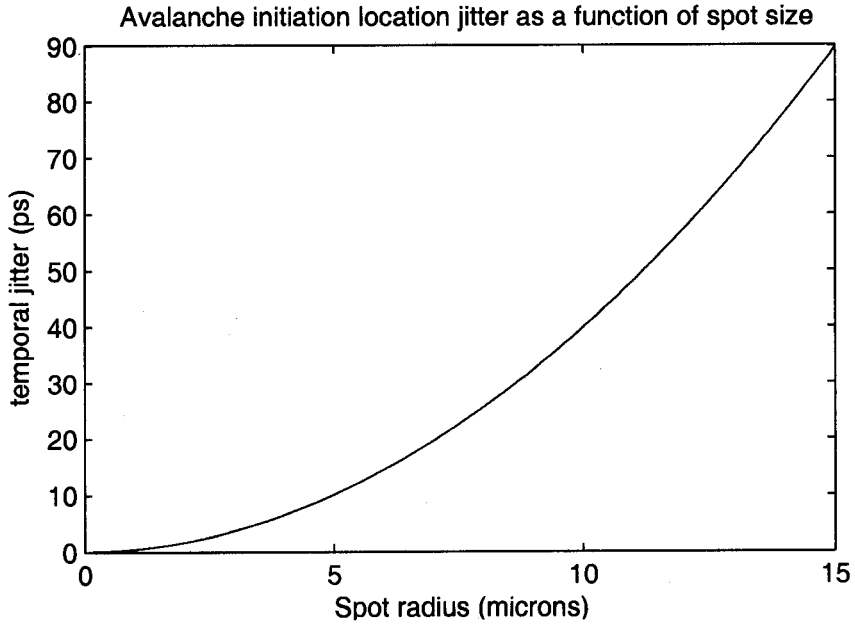


Figure 6.1: Temporal jitter introduced by the avalanche initiation location for different spot sizes on the 30 micron APD element.

was focused into the central 5 micron diameter, then the temporal jitter associated with the avalanche initiation location would be about 10 ps and the detector response would be on the order of 75 ps or better.

Unfortunately, decreasing the spot size on the detector doesn't come without disadvantages. Figure 6.2 shows the details of the optical path designated "Lenslet and APD array" on Figure 2.4. The lenslet array is placed at the focal plane of the telescope and the detector array is placed at the focus of the lenslet array. Since a lens maps parallel light rays to points, each focus point on the lenslet corresponds to light from a particular angle on the sky. The same phenomenon forms a pupil at the focus of the lenslet array. All incident light on the lens passes through this pupil, regardless of its incident direction.

Combining this with the geometry of the remaining optical layout results in the

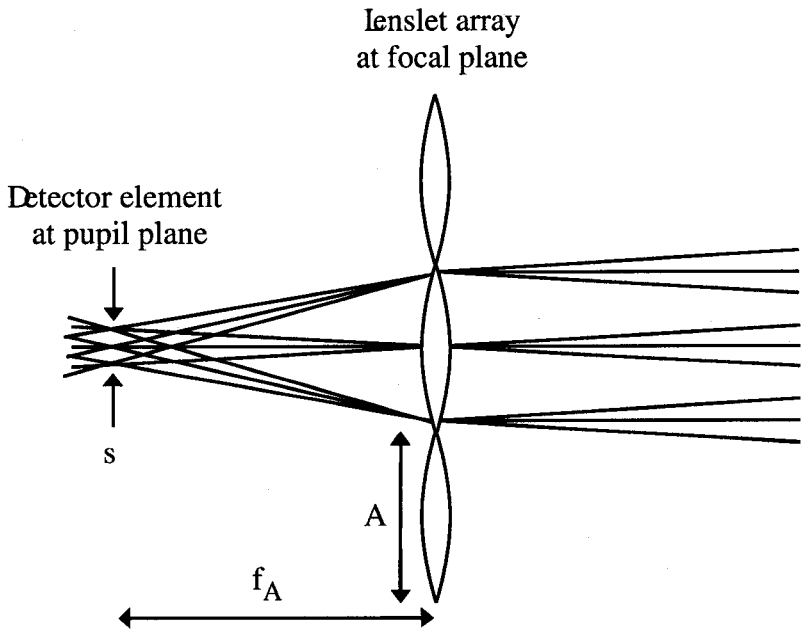


Figure 6.2: Ray trace diagram for the lenslet and detector arrays. Placing the lenslet array at the focal plane of the telescope and the detector array at the focus of the lenslets creates a pupil image on the detector. Since the lenslet array is at the focal plane of the telescope, the light incident on the array from the left comes from different angles on the sky.

relationship

$$\theta \approx \frac{As}{Df_A} \tag{6.1}$$

between θ, the angle on the sky subtended by a single element of the lenslet array, A, the size of one lenslet element, s, the pupil diameter at the APD, D, the diameter of the telescope and f_A, the focal length of the lenslet. This approximation holds for small θ such that $\tan\theta \approx \theta$. From this we can see that the field of view (FOV) of each APD element is directly proportional to the pupil or spot diameter created on the element. Lenslet arrays with a pitch of 100 microns and a focal length of 500 microns are commercially available. Combining this with the 3.5 m telescope diameter we find that $\theta = s \times 0.0118''$. In other words, a 10 micron spot subtends 0.118 arc-seconds on the sky and the whole 4×4 APD array would have a FOV

of less than 0.5 arc-seconds, about half the typical seeing conditions at APO. This would make acquiring and tracking the lunar signal much more difficult. In order to effectively use the spatial information of the array, the FOV of the array should be larger than the seeing conditions of one arc-second. A small FOV effectively turns the array into a single detector for the purposes of tracking and acquisition. The 20 micron device would allow a maximum FOV of just under an arc-second and the 30 micron device would allow a maximum FOV on the order of 1.5 arc-seconds. For this reason we intend to use the 30 micron device for APOLLO.

In Section 5.5 we observed that while the behavior of the APD elements was roughly uniform, edge effects were present. Imaging a 25 micron spot on the 30 micron avoids these edge effects without significantly degrading the light-collecting area. A 25 micron spot would have a field of view of 0.3 arc-seconds per element, giving the 4×4 array a FOV of 1.2×1.2 arc-seconds.

The FOV effectively sets the minimum outer diameter of the spot, but it is still possible to reduce the temporal jitter by masking off the central portion of the beam at the expense of losing photons. For example, a thin ring would decrease the admitted intensity, but also greatly decrease the temporal jitter due to avalanche initiation location. On the other hand, the entire element could be illuminated but at the cost of increasing the temporal jitter of the APD. The question then arises "is there an ideal size for a central obscuration?" For instance, let's say the use of a 5 micron central obscuration in a 25 micron spot reduced the temporal jitter such that we can achieve the same statistical error with 10% fewer photons. From Section 5.6 we know that the power in the core is uniform over the element so a 5 micron central obscuration only reduces the number of collected photons by 4%. In that case, the 10% gain is greater than the 4% loss so the system efficiency can be improved by including a 5 micron central obscuration.

There is, in fact, already a central obscuration caused by the APO secondary mirror. Ray tracing can show that every point in the pupil plane where the detector

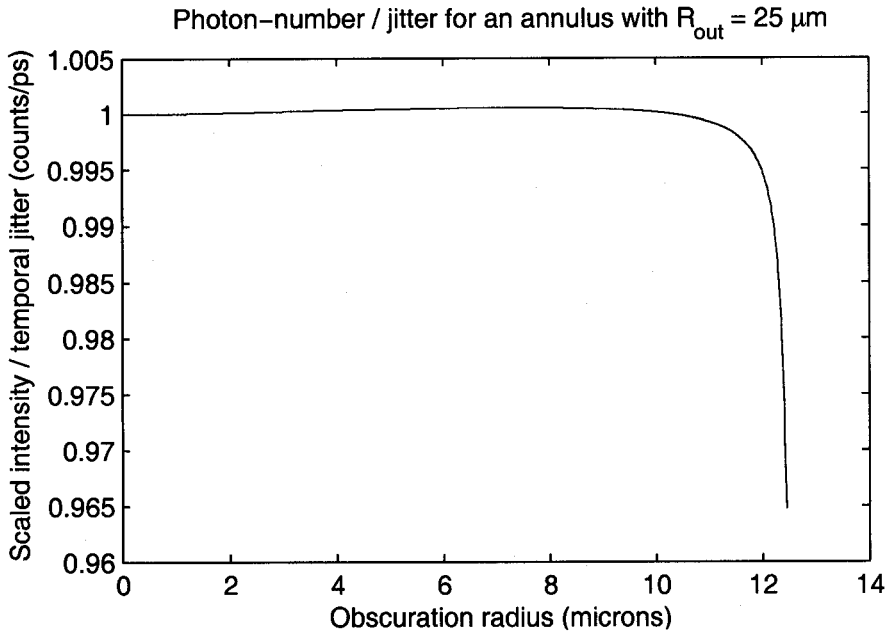


Figure 6.3: Effect of central obscuration on the ratio between the photon number and the temporal jitter for a spot with an outer diameter of 25 microns.

sits can be mapped to a point on the primary mirror. This is because the primary mirror is itself a pupil. As a result, the central obscuration caused by the 0.8 m diameter secondary mirror is mapped to the center of the APD element. If we consider a 25 micron spot, then the central 5.7 microns of the APD will not be illuminated.

An interesting side note is that a central obscuration will further reduce the impact of any remaining diffusion tail. As mentioned earlier, the tail is significantly reduced at 532 nm, but up to 1% of the flux could still come from the diffusion region and degrade the temporal response. The number of detected photons in the tail falls off rapidly toward the edges of the element while the power in the core is essentially uniform. If the center of the element is obscured, any flux remaining in the tail should be further reduced.

Figure 6.3 shows how the ratio of the number of collected photons (an effective intensity) to the temporal jitter due to avalanche initiation location changes for different

central obscurations. The plot is essentially flat (only changing at a few parts in 10^4) until the central 20 microns of the 25 micron spot is obscured. In other words, the benefit of decreased temporal jitter is almost completely canceled by the detriment of the decreased collecting area and therefore the number of photons collected. This being the case, changing the size of the central obscuration has little impact on the performance or efficiency of the APDs. If we consider an annulus with a 25 micron outer diameter and a 5.7 micron inner diameter (due to the secondary mirror) then the avalanche initiation location will contribute roughly 60 ps RMS to the temporal uncertainty. Combining this with the better than 75 ps Gaussian detector performance, we find that the APDs will contribute at most 96 ps to the total temporal error budget. As discussed earlier, this is an upper limit on the detector performance.

It could turn out that the detector response is dominated by the avalanche initiation location. In that case one way to reduce the temporal jitter of the APDs to something comparable with the laser pulse would be to invest in custom lenslet arrays with shorter focal lengths. Any reduction in the focal length directly translates into a smaller spot size with the same FOV. A 20% reduction in the focal length would produce a 20 micron spot which has an avalanche initiation location jitter of 40 ps which is comparable with the laser.

6.3 Predictions of full lunar response

Given the information presented in this dissertation, it is now possible to make a prediction of the full temporal profiles of the lunar photons.

I will assume that the 532 nm laser will produce a Gaussian beam with a pulse width of 40 ps RMS as specified by the manufacturer. I will also assume a worst-case scenario for the uniform Gaussian temporal response of the APD, namely 75 ps, though this could be significantly lower. Based on the analysis discussed in Section 6.2.2 the avalanche initiation location will contribute 60 ps. At one point we achieved

better than 25 ps resolution with the TDC and circuitry that I designed. Considering the crudeness of the boards and my limited experience with low-noise electronics, I am confident that the electronics designed by the electrical engineers at UC San Diego will perform at least as well.

The quadrature combination of these contributions results in a combined laser, electronics, and detector temporal jitter of 107 ps. If we ignore the effect of the retro-reflector for a moment, we would need to collect 235 photons in order to reach our goal of 7 ps resolution. At a collection rate of one photon per pulse, this would take less than 15 seconds. On those few instances where the libration angles are near zero, this would be the best performance we could get.

This is also the temporal profile of the calibration pulse that is used to determine the start of the time-of-flight. Keeping track of the temporal response of the calibration photons will give us a way to track the long-term performance of the laser pulse, detectors, electronics and TDC. If a distortion in the calibrated temporal response is observed, strong-signal tests could determine if the problem is associated with the electronics or TDC. Unfortunately, decoupling the detectors and lasers is more difficult. A fast photodiode will also be used to monitor the output power of the laser. The photodiode results used in conjunction with the calibration temporal response history could tease out any discrepancies with the laser.

When the retro-reflector array is included in the temporal profile, it becomes the dominant source of uncertainty. Figure 6.4 shows the response of the retro-reflector with an orientation of $(5°, 2.5°)$ by itself, and convolved with the predicted temporal response at 532 nm. The temporal jitter of this convolved response is 210 ps. Collecting photons for 45 seconds at the rate of one photon per pulse (and 20 pulses per second) would give the desired 7 ps resolution.

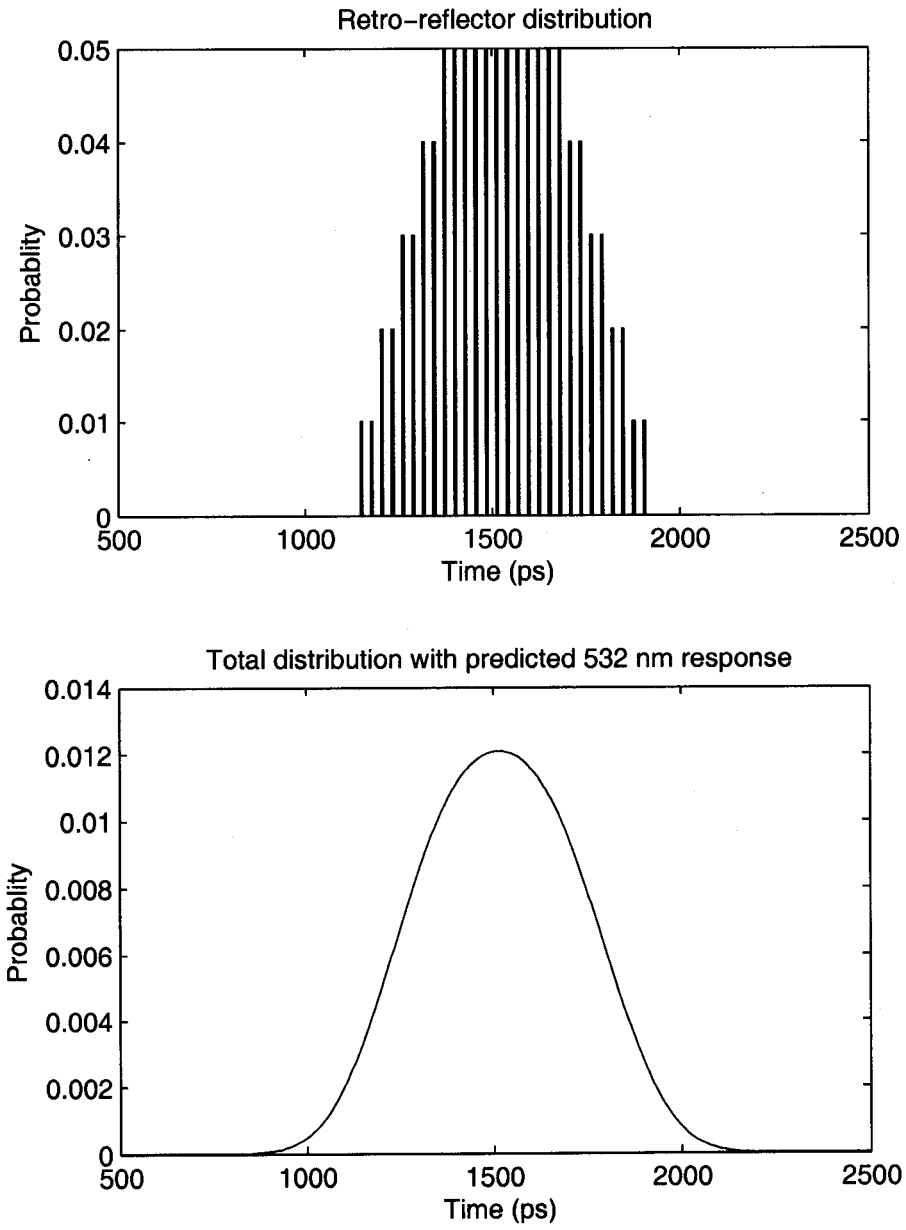


Figure 6.4: *Top:* Retro-reflector array response for $(5°, 2.5°)$ libration angles. *Bottom:* Total distribution if array response is convolved with 107 ps jitter that is predicted at 532 nm. The total temporal response has a jitter of 210 ps.

6.4 *Future work*

A wise person once told me that the dissertation is never finished, it's only submitted. As a result there is always work to do in the future.

6.4.1 Electronic characterization

New versions of the APD bias and readout electronics are currently being designed by engineers at UC San Diego. To my knowledge the basic design is similar to that shown in Figure 3.6. The use of multi-layer boards with continuous ground planes should greatly reduce the noise of the electronics and allow for better temporal performance. Once the boards are finished, it will be necessary to optimize the electronics in order to achieve the best temporal resolution possible. All 16 channels will respond to the same gate pulse so the photons will return at the same point in the gate. The appropriate separation between the gate-on signal and the expected photon arrival will have to be determined for all 16 channels simultaneously.

In order to avoid ringing or variations in the excess voltage, the photon arrival should likely be positioned later in the gate; however, this also increases the chances that a background photon or thermal event initiates an avalanche in an element before the lunar return or calibration photons can arrive. The dark count rates for the devices are on the order of 20 kHz, or 1 every 50 μs. The APDs will be gated on for roughly 100 ns 40 times a second (20 calibration returns and 20 lunar returns), so the APDs will be active for about 4 μs every second. Comparing these two numbers, a dark count is only expected in about one out of every 10–12 gates and is only detrimental if it arrives before a calibration or lunar return photon. The background rates will also degrade the performance in a similar fashion. The full moon has a surface brightness of -13 magnitudes per square arc-second. This translates into 0.056 photons per nanosecond per nanometer per square arc-second arriving at the telescope. This is reduced by the telescope (40%), receiver (25%) and detector (30%) efficiencies

discussed in Section 2.4, resulting in 0.002 photons per nanosecond per nanometer per square arc-second *detected* by the APDs. A narrow-band wavelength filter and a small field-of-view both mean that only 532 nm photons, from a specific region of the sky that arrive during the appropriate time window, have the possibility of being detected. Assuming we use a 100 ns temporal gate, a 2 nm wavelength filter and a $1.2'' \times 1.2''$ FOV, APOLLO should detect ~ 0.5 photons per gate, spread over all 16 elements. If one element detects a background photon, the other 15 are still available to detect a calibration or lunar return photon. Again, if a background photon arrives after the arrival of a calibration or lunar return it has no effect on our system. The final design of the optical train will have to balance the pros and cons involved with the choice of the filters, fields of view, background contamination, location of photon arrival within gate and temporal jitter.

In addition to the location of the photon arrival within the gate, each channel will have to be optimized by adjusting the variable capacitor and potentiometer. The values of these components should be similar for all the channels, but each will have to be tweaked to find the "sweet spot" of minimum temporal jitter. The TDC also has a slightly different temporal performance depending on the delay between the photon arrival (start pulse) and the stop pulse. The use of a delay line between the detectors and the TDC could be used to minimize the temporal jitter. Each channel could have its own delay line of slightly different lengths, but this is probably a bit over the top, at least to begin with.

6.4.2 Numerical simulations

Another area of future work would be numerical simulations of the APD performance. I began exploring the numerical simulations in the last month or two of my thesis work. There is a wealth of temporal and spatial information that can be explored by modeling the diffusion process. I had the opportunity to run only a handful of simulations, and only went as far as to say, "Gee, that temporal response looks

familiar" but didn't make any quantitative comparisons between the results of the simulation and the experimental results. The numerical simulations hold a wealth of information and are open to many avenues of exploration, but I'll point out a few that I would find interesting.

As of now I have explored the behavior only within the diffusion region. Once an electron region passed the boundary into the drift region I considered it detected and stopped the simulation. This first stage could be followed by a second stage that incorporates the drift of the electron from the diffusion region to the avalanche region. Dr. Murphy has developed a model for the electric field as a function of depth which could be combined with analytical equations available in the literature in a straight forward manner. I could also conceive of a relatively simple model of the behavior in the avalanche region, perhaps using ballistic motion and conservation of momentum techniques to model the lateral propagation of the avalanche across the element. There may also be work available in the literature, where others have developed a model for avalanche behavior, that could be incorporated into a numerical simulation. I haven't looked for such papers myself, but would not be surprised to find them out there.

In the course of my work, I simulated only the diffusion of electrons created on-axis. Exploring the off-axis behavior is as simple as changing the initial position of the electron. After that, it is only a matter of time to run simulations at different radii and depths. Based on the distribution of power in the diffusion tail that we observe, these simulations should result in a roughly parabolic dependence on the radius.

A model that incorporates all three regions (the diffusion, drift and avalanche region) offers nearly endless paths of exploration. Such a model would allow its user to explore both the spatial and temporal aspects of the APD performance. The results of the simulation could be compared with the experimental data collected in this dissertation or other sources. For example, the simple model of a Gaussian core and diffusion tail ignores any effect of the drift region. This is likely the cause of the

poor fits on the shoulder of the temporal distributions, about 500 ps after the peak. Incorporation of a drift region in the numerical simulations could lead to a better temporal response profile.

The shape and extent of the electric field in the avalanche, drift region and guard ring could also be explored. In all of my simulations, I assumed that the guard ring was completely cylindrical. In fact, the electric field could look more like a champagne-glass. If the guard ring is not cylindrical, but rather tapers inward at deeper depths, more electrons would encounter the guard ring, further reducing the flux in the tail, especially at longer times. In comparing the diffusion tails of the simulations with those observed, we saw that the experimental tails did in fact fall off much faster than the simulations. Comparisons between the simulation, the temporal profiles and spatial profiles could yield a more detailed model for APD performance.

The work discussed in this thesis has been used to construct and verify a simple model of the APD performance. At present, it is too simplistic to describe such details as the shape of the electric field and the consequent values and locations of the doping levels in the various layers. On the other hand, this work has opened the door and laid a framework that could be used to provide much more detailed characterization of the APDs. This framework could be used to verify and inform the future designs of the APDs.

6.4.3 Optimized photon analysis techniques

Finally, we return to the big picture and the point of all this work—measuring the distance between the earth and the moon with millimeter precision. As I've discussed already, assuming the photon arrival rates can be optimized to one photon per pulse or better, APOLLO should be able to achieve millimeter resolution in a matter of minutes. The next crucial step is deciding how to best implement the data analysis techniques. A couple of possibilities exist.

During the design of the apparatus, we took the approach of spreading the returned

light over many elements and in so doing preserving some spatial resolution to aid in the acquisition and tracking process. Another design option would be to focus all the light down on to one element and only detect the first, or earliest, photon that arrives. As we've seen (and exploited), detecting only the first photon biases the temporal profile towards shorter times and produces a distorted temporal distribution.

Atmospheric turbulence causes the outgoing laser beam to break up into bright and dark spots. These intensity variations are known as scintillation. Scintillation also explains why stars appear to twinkle. As a result of the beam break-up, sometimes the retro-reflector array will reflect many photons, while other times it will reflect very few. The signal variation on the return trip is dominated by speckle interference. The retro-reflector array elements aren't phased so the signals from individual elements can interfere, producing a speckle pattern of bright and dark regions. The combination of atmospheric turbulence and speckle interference patterns from the incoherent retro-reflectors cause large fluctuations in the intensity of light reaching the retro-reflectors and returning to the telescope. If only one detector is used to collect all the light then there will be no way of knowing if a signal resulted from the first photon out of ten or one photon all by itself. If it is the first of ten, the time of flight will likely be much earlier in the distribution than its lone counterpart. In order to develop an unbiased distribution we would need to weight the two signals differently, but have no way of telling them apart.

The advantage of the array detector, beyond the potential spatial information, is that it allows us to determine a weighting factor. If all 16 elements detect a photon, we can be certain that those result from a very strong return and are going to be strongly biased towards shorter times. If only one element detects a photon, then only a small first-photon correction is needed since only one out of sixteen channels detected a photon. The details of the first-photon bias corrections have to be worked out. For example, if there is a point where signals are too strong (such as all sixteen, ten, or six elements firing) to correct the data.

In addition, there could also be ways to improve the temporal resolution by more than the $\sqrt{N}$ improvement discussed so far. The lunar libration angles are known for any given night thereby giving us *a priori* knowledge of the retro-reflector array orientation and its contribution to the temporal profile. Bayesian analysis techniques could make use of this *a priori* information to better determine the centroid of the return pulse. For example, let's assume that the retro-reflector array is oriented in such a way that it provides a purely rectangular pulse that has a width of 150 ps. Let's also assume for the moment that the temporal contribution of the laser, detector and electronics is a delta-function. If two photons are detected with a 75 ps separation between them, the centroid of the rectangular pulse can only be shifted in time so far until it cannot keep both the detected photons within the distribution. If the two photons are 100 ps apart, the uncertainty in the position of the centroid is reduced, even though we still only have two photons. From this it is clear that the first and last photons provide the greatest constraint on the location of the centroid. Under these conditions, Bayesian analysis techniques could reduce the uncertainty in the centroid of the pulse much faster than $1/\sqrt{N}$, approaching $1/N$ for a small uncertainty due to the laser, detector and timing electronics.

For Bayesian analysis to be the most effective, we would need sharp edges to the temporal distribution, hence the rectangular pulse and delta-function contribution from the laser, detector and timing electronics. If the total distribution is like that shown in Figure 6.4, then Bayesian analysis is not nearly as useful because of the wide wings of the distribution. Nevertheless, we should continue to explore other techniques in an effort to exploit our knowledge of the various contributions to the temporal distribution.

BIBLIOGRAPHY

R. I. Abbot, P. J. Shelus, J. D. Mulholland, and E. C. Silverberg. Laser observations of the Moon: Identification and construction of normal points for 1969-1971. *Astronomical Journal*, 78:784, October 1973.

M. A. Albota, R. M. Heinrichs, D. G. Kocher, D. G. Fouche, B. E. Player, M. E. O'Brien, B. F. Aull, J. J. Zayhowski, J. Mooney, and B. C. Willard. Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser. *Applied Optics*, 41:7671–7678, December 2002a.

Marius A. Albota, Brian F. Aull, Daniel G. Fouche, Richard M. Heinrichs, David G. Kocher, Richard M. Marino, James G. Mooney, Nathan R. Newbury, Michael E. O'Brien, Brian E. Player, Berton C. Willard, and John J. Zayhowski. Three-Dimensional Imaging Laser Radars with Geiger-Mode Avalanche Photodiode Arrays. *Lincoln Laboratory Journal*, 13(2):351–367, 2002b.

J. D. Anderson and J. G. Williams. Long-range tests of the equivalence principle. *Classical and Quantum Gravity*, 18:2447–2456, July 2001.

Brian F. Aull, Andrew H. Loomis, Douglas J. Young, Richard M. Heinrichs, Bradley J. Felton, Peter J. Daniels, and Deborah J. Landers. Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging. *Lincoln Laboratory Journal*, 13(2):335–348, 2002.

S. Baeßler, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, U. Schmidt, and H. E. Swanson. Improved Test of the Equivalence Principle for Gravitational Self-Energy. *Physical Review Letters*, 83:3585–3588, November 1999.

P. L. Bender, D. G. Currie, R. H. Dicke, D. H. eckhardt, J. E. Faller, W. M. Kaula, J. D. Mulholland, H. H. Plotkin, S. K. Poultney, E. C. Silverberg, D. T. Wilkinson, J. G. Williams, and C. O. Alley. The Lunar Laser Ranging Experiment. *Science*, 182:229, 1973.

B. Bertotti, I. Ciufolini, and P. L. Bender. New test of general relativity - Measurement of de Sitter geodetic precession rate for lunar perigee. *Physical Review Letters*, 58:1062–1065, March 1987.

Philip R. Bevington and D. Keith Robinson. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw Hill, 3rd edition, 2003.

J.-P. Boy, J. Hinderer, and P. Gegout. Global atmospheric loading and gravity. *Physics of the Earth and Planetary Interiors*, 109:161–177, December 1998.

V. B. Braginsky and V. I. Pavov. Verification of the equivalence of inertial and gravitational mass. . *Sov. Phys. JETP.*, 34:463–466, 1972.

Continuum Laser. `http:www.continuumlasers.com/resmainswf.html`.

S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied Optics*, 35:1956, April 1996.

S. Cova, A. Lacaita, M. Ghioni, G. Ripamonti, and T. A. Louis. 20-ps timing resolution with single-photon avalanche diodes. *Review of Scientific Instruments*, 60: 1104–1110, June 1989.

T. Damour. Testing the equivalence principle: why and how? . *Classical and Quantum Gravity*, 13:A33–A41, November 1996.

T. Damour and K. Nordtvedt. Tensor-scalar cosmological models and their relaxation toward general relativity. *Physical Review D*, 48:3436–3450, October 1993.

T. Damour, F. Piazza, and G. Veneziano. Runaway dilaton and equivalence principle violations. *ArXiv General Relativity and Quantum Cosmology e-prints*, April 2002.

T. Damour and A. M. Polyakov. The string dilaton and a least coupling principle. *Nuclear Physics B*, 423:532, July 1994a.

T. Damour and A.M. Polyakov. String Theory and Gravity. *General Relativity and Gravitation*, 26:1171–1176, 1994b.

T. Damour and D. Vokrouhlický. Equivalence principle and the Moon. *Physical Review D*, 53:4177–4201, April 1996.

H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb. Photon counting techniques with silicon avalanche photodiodes. *Applied Optics*, 32:3894–3900, July 1993.

P. de Bernardis, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, G. De Gasperis, P. C. Farese, P. G. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, A. E. Lange, L. Martinis, S. Masi, P. V. Mason, P. D. Mauskopf, A. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, D. Sforna, and N. Vittorio. A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404:955–959, April 2000.

W. de Sitter. On Einstein's thory of gravitation and its astronomical consequences. *Mon. Not. Royal Astron. Soc.*, 77:155–184, 1916.

J. O. Dickey, P. L. Bender, J. E. Faller, X. X. Newhall, R. L. Ricklefs, J. G. Ries, P. J. Shelus, C. Veillet, A. L. Whipple, J. R. Wiant, J. G. Williams, and C. F. Yoder. Lunar Laser Ranging - a Continuing Legacy of the Apollo Program. *Science*, 265: 482, July 1994.

M. Ghioni, S. Cova, F. Zappa, and C. Samori. Compact active quenching circuit for fast photon counting with avalanche photodiodes. *Review of Scientific Instruments*, 67:3440–3448, October 1996.

I. Goldman. PSR 0655 + 64 - an astrophysical laboratory for testing relativistic gravity theories. *Astrophysical Journal*, 390:494–498, May 1992.

Hamamatsu Picosecond Light Pulser. `http://usa.hamamatsu.com`.

S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, V. V. Hristov, A. H. Jaffe, A. E. Lange, A. T. Lee, P. D. Mauskopf, C. B. Netterfield, S. Oh, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu. MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of 10'-5deg. *Astrophysical Journal Letters*, 545:L5–L9, December 2000.

B. Heckel, E. Adelberger, S. Baessler, J. Gundlach, M. Harris, C. Hoyle, S. Merkowitz, U. Schmidt, A. Sharp, G. Smith, and E. Swanson. Results on the Strong Equivalence Principle, Dark Matter, and New Forces. In *Fundamental Physics in Space*, page 1225, 2000.

W. J. Kindt. *Geiger Mode Avalanche Photodiode Arrays for spatially resolved single photon counting.* Delft University Press, 1999.

Yu L. Kokurin. Laser Ranging of the Moon. In N. G. Basov, editor, *Lasers and Their Applications in Physical Research*, volume 91, pages 161–226. Plenum Pr, 1979.

A. Lacaita, S. Cova, and M. Ghioni. Four-hundred-picosecond single-photon timing with commercially available avalanche photodiodes. *Review of Scientific Instruments*, 59:1115–1121, July 1988.

A. Lacaita, S. Cova, C. Samori, and M. Ghioni. Performance optimization of active quenching circuits for picosecond timing with single photon avalanche diodes. *Review of Scientific Instruments*, 66:4289–4295, August 1995.

A. Lacaita, S. Cova, A. Spinelli, and F. Zappa. Photon-assisted avalanche spreading in reach-through photodiodes. *Applied Physics Letters*, 62:606–608, February 1993.

A. Lacaita, M. Mastrapasqua, M. Ghioni, and S. Vanoli. Observation of avalanche propagation by multiplication assisted diffusion in p-n junctions. *Applied Physics Letters*, 57:489–491, July 1990.

D. E. Lebach, B. E. Corey, I. I. Shapiro, M. I. Ratner, J. C. Webber, A. E. E. Rogers, J. L. Davis, and T. A. Herring. Measurement of the Solar Gravitational Deflection of Radio Waves Using Very-Long-Baseline Interferometry. *Physical Review Letters*, 75:1439–1442, August 1995.

J. Müller and K. Nordtvedt. Lunar laser ranging and the equivalence principle signal. *Physical Review D*, 58:62001, September 1998.

M. T. Murphy, J. K. Webb, V. V. Flambaum, V. A. Dzuba, C. W. Churchill, J. X. Prochaska, J. D. Barrow, and A. M. Wolfe. Possible evidence for a variable fine-structure constant from QSO absorption lines: motivations, analysis and results. *Mon. Not. Royal Astron. Soc.*, 327:1208–1222, November 2001.

T. W. Murphy, E. G. Adelberger, J. D. Strasburg, and C. W. Stubbs. APOLLO: Multiplexed Lunar Laser Ranging. In *13th International Workshop on Laser Ranging; Washington D.C.*, 2002. `http://cddisa.gsfc.nasa.gov/lw13/docs/papers/llr_murphy_1m.pdf`.

T. W. Murphy, J. D. Strasburg, C. W. Stubbs, E. G. Adelberger, J. Angle, K. Nordtvedt, J. G. Williams, J. O. Dickey, and B. Gillespie. The Apache Point Observatory Lunar Laser Ranging Operation (APOLLO).

In *12th International Workshop on Laser Ranging; Matera, Italy*, 2000. http://geodaf.mt.asi.it/GDHTL/news/iwlr/Murphy_et_al_apachepoint.pdf.

K. Nordtvedt. Equivalence principle for massive bodies. I. Phenomenology. *Physical Review*, 169:1014–1016, 1968a.

K. Nordtvedt. Equivalence Principle for Massive Bodies. II. Theory. *Physical Review*, 169:1017–1025, May 1968b.

K. Nordtvedt. Testing Relativity with Laser Ranging to the Moon. *Physical Review*, 170:1186–1187, June 1968c.

K. Nordtvedt. Lunar laser ranging reexamined: The non-null relativistic contribution. *Physical Review D*, 43:3131–3135, May 1991.

K. Nordtvedt. From Newton's moon to Einstein's moon. *Physics Today*, 49:26–33, May 1996.

K. Nordtvedt. LETTER TO THE EDITOR: Testing Newton's third law using lunar laser ranging. *Classical and Quantum Gravity*, 18:L133–L137, October 2001.

E. Overbeck, C. Sinn, I. Flammer, and J. Ricka. Silicon avalanche photodiodes as detectors for photon correlation experiments. *Review of Scientific Instruments*, 69: 3515–3523, October 1998.

J. A. Peacock, S. Cole, P. Norberg, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. D. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, W. J. Percival, B. A. Peterson, I. Price, W. Sutherland, and K. Taylor. A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410:169–173, March 2001.

Phillips Scientific Time to Digital Converter.
`http://www.phillipsscientific.com/pdf/7186ds.pdf`.

R. D. Reasenberg, I. I. Shapiro, P. E. MacNeil, R. B. Goldstein, J. C. Breidenthal, J. P. Brenkle, D. L. Cain, T. M. Kaufman, T. A. Komarek, and A. I. Zygielbaum. Viking relativity experiment - Verification of signal retardation by solar gravity. *Astrophysical Journal Letters*, 234:L219–L221, December 1979.

George Rieke. *Detection of Light: From the Ultraviolet to the Submillimeter, 2^{nd} edition.* Cambridge University Press, 2003.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astronomical Journal*, 116: 1009–1038, September 1998.

P. G. Roll, R. Krotkov, and R. H. Dicke. The equivalence of inertial and passive gravitational mass. *Ann. Phys. (N.Y.)*, 26:442–517, 1964.

E. Samain, J. F. Mangin, C. Veillet, J. M. Torre, P. Fridelance, J. E. Chabaudie, D. Feraudy, M. Glentzlin, J. Pham van, M. Furia, A. Journet, and G. Vigouroux. Millimetric Lunar Laser Ranging at OCA (Observatoire de la Cote d'Azur). *Astronomy and Astrophysics Supplement Series*, 130:235–244, June 1998.

G. L. Smith, C. D. Hoyle, J. H. Gundlach, E. G. Adelberger, B. R. Heckel, and H. E. Swanson. Short-range tests of the equivalence principle. *Physical Review D*, 61: 22001, January 2000.

J. D. Strasburg, T. W. Murphy, E. G. Adelberger, C. W. Stubbs, D. W. Miller, and J. I. Angle. The advantages of avalanche Pho-

todiode (APD) arrays in laser-ranging applications. In *13th International workshop on Laser Ranging; Washington D.C.*, 2002a. http://cddisa.gsfc.nasa.gov/lw13/docs/papers/detect_strasburg_1m.pdf.

J. D. Strasburg, T. W. Murphy, C. W. Stubbs, E. G. Adelberger, D. W. Miller, and J. I. Angle. Lunar laser ranging using avalanche photodiode (APD) arrays. In *Survey and Other Telescope Technologies and Discoveries. Edited by Tyson, J. Anthony; Wolff, Sidney. Proceedings of the SPIE, Volume 4836, pp. 387-394 (2002).*, pages 387–394, December 2002b.

Y. Su, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, M. Harris, G. L. Smith, and H. E. Swanson. New tests of the universality of free fall. *Physics Review D*, 50: 3614–3636, September 1994.

D. Vokrouhlicky. A Note on the Solar Radiation Perturbations of Lunar Motion. *Icarus*, 126:293–300, April 1997.

Clifford M. Will. *Was Einstein Right.* Basic Books Inc., 1986.

Clifford M. Will. *Theory and Experiment in Gravitational Physics.* Cambridge University Press, 1993.

J. G. Williams, D. H. Boggs, J. O. Dickey, and W. M. Folkner. Lunar Laser Tests of Gravitational Physics. In Vahe G. Gurzadyan, Robert T. Jantzen, and Remo Ruffini, editors, *The Ninth Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories*, page 1797, 2002.

J. G. Williams, D. H. Boggs, C. F. Yoder, J. T. Ratcliff, and J. O. Dickey. Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research*, 106:27933–27968, November 2001.

J. G. Williams, X. X. Newhall, and J. O. Dickey. Relativity parameters determined from lunar laser ranging. *Physical Review D*, 53:6730–6739, June 1996.

F. Zappa, M Ghioni, S. Cova, C. Samori, and A. Giudice. An Integrated Active-Quenching circuit for Single-Photon Avalanche Diodes. *IEEE Transactions on Instrumentation and Measurement*, 49:1167–1175, 2000.

VITA

Jana Dee Strasburg was born May 29, 1976 in Helena, Montana. She spent the next 18 years of her life in Montana before attending Pacific Lutheran University in Tacoma, Washington in 1994. During her time at PLU she earned a double major in physics and electrical engineering and graduated Magna Cum Laude. In 1998 she began graduate school at the University of Washington in Seattle and earned her Master's degree three years later. In 2002 she accepted a National Security Internship at Pacific Northwest National Laboratory, where she studied laser beam propagation through atmospheric turbulence as part of the Electronics, Optics and Radar group while completing her dissertation.