UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Self-quenching single photon avalanche photodiodes for near-infrared detection

Permalink

https://escholarship.org/uc/item/2hq9h6nq

Author

Cheng, James C.

Publication Date

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Self-Quenching Single Photon Avalanche Photodiodes for Near-Infrared Detection

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Nanoscale Devices and Systems)

by

James Cheng

Committee in charge:

Professor Yu-Hwa Lo, Chair
Professor Peter Asbeck
Professor Prabhakar Bandaru
Professor Chung-Kuan Cheng
Professor Sadik Esener

2012

Copyright

James Cheng, 2012

All rights reserved.

The dissertation of James Cheng is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

DEDICATION

To my loving family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Prof. Yu-Hwa Lo, for his guidance throughout my entire Ph.D. career. His experience and knowledge was instrumental to my development as a successful researcher, and was always able to provide advice and creative solutions to any problems I faced. I would also like to thank the rest of my committee members: Prof. Peter Asbeck, Prof. Prabhakar Bandaru, Prof. Chung-Kuan Cheng, and Prof. Sadik Esener for the assistance and guidance they provided.

I would like to thank all my past and present lab members for the support they have provided over the many years. I thank Dr. Arthur Zhang, Dr. Kai Zhao, and Dr. Sifang You for their help and mentorship during the beginning of my research career, and Samia Rahman for the many insightful discussions we had. I also thank the rest of the group for their constant friendship and support: Dr. HongKwon Kim, Ash Arianpour, Dr. Randy Chen, Roger Chiu, Dr. Sung Hwan Cho, Dr. Jessica Godin, David Hall, Hosuk Lee, Yu-hsin Liu, Zhe Mei, Wen Qiao, Dr. Frank Tsai, TsungFeng Wu, and Yuchun Zhou.

I would also like to thank Dr. Wayne Chen, Dingbo Chen, Wei Lu, and Winnie Chen for the tricks and companionship they provided during the many hours I spent in the cleanroom. I also need to thank the Nano3 staff for the excellent support they provided which made my research possible: Larry Grissom, Dr. Xuekun Lu, Dr. Ahmet Erten, Sean Parks, Ryan Anderson, Dr. Maribel Montero, and Dr. Bernd Fruhberger. I also need to thank Dr. Mark Hsu and Dr. Michael Benchimol for the critical pieces of equipment required in my experiment, and the useful advice they gave regarding measurement techniques.

Most importantly I would like to thank my family. They have always been

supportive of me and encouraged me to follow my dreams. They're there for me through any ups or downs, and provided a solid foundation for my journey through life.

Portions of Chapter 2 is a reprint of materials as it appears in the following publications: James Cheng, Sifang You, Kai Zhao, and Yuhwa Lo, "Self-quenched InGaAs single-photon detector", Proc. SPIE 7320, 732010 (2009); Sifang You, James Cheng, and Yu-Hwa Lo, "Physics of Single Photon Avalanche Detectors With Built-In Self-Quenching and Self-Recovering Capabilities", IEEE J. Quantum Electronics, v 48, n 7, p 960 (2012). The dissertation author was the primary/secondary investigator and author of these materials.

Portions of Chapter 3 and 4 is a reprint of materials as it appears in the following publications: James Cheng, Sifang You, Samia Rahman, and Yu-Hwa Lo, "Self-quenching InGaAs/InP single photon avalanche detector utilizing zinc diffusion rings", Optics Express, v 19, p 15149 (2011); James Cheng, Yu-Hwa Lo, "Patterned zinc-diffused structures for improved avalanche probabilities in InGaAs/InP single photon detectors", IEEE Photonics Conference 2011. The dissertation author was the primary investigator and author of this material.

Portions of Chapter 5 is a reprint of materials as it appears in the following publications: James Cheng, Yu-Hwa Lo, "Modeling of Self-Quenching Single Photon Avalanche Photodiodes for System Applications", IEEE J. Sel. Top. Quantum Electronics, submitted. The dissertation author was the primary investigator and author of this material.

VITA

2007	Bachelor of Science, University of California, Berkeley Electrical Engineering and Computer Sciences Materials Science and Engineering
2009	Master of Science, University of California, San Diego Electrical Engineering (Nanoscale Devices and Engineering)
2012	Doctor of Philosophy, University of California, San Diego Electrical Engineering (Nanoscale Devices and Engineering)

PUBLICATIONS

James Cheng, Yu-Hwa Lo, "Modeling of Self-Quenching Single Photon Avalanche Photodiodes for System Applications", IEEE J. Sel. Top. Quantum Electronics, submitted.

Sifang You, James Cheng, and Yu-Hwa Lo, "Physics of Single Photon Avalanche Detectors With Built-In Self-Quenching and Self-Recovering Capabilities", IEEE J. Quantum Electronics, v 48, n 7, p 960 (2012).

James Cheng, Sifang You, Samia Rahman, and Yu-Hwa Lo, "Self-quenching InGaAs/InP single photon avalanche detector utilizing zinc diffusion rings", Optics Express, v 19, p 15149 (2011).

James Cheng, Yu-Hwa Lo, "Patterned zinc-diffused structures for improved avalanche probabilities in InGaAs/InP single photon detectors", IEEE Photonics Conference 2011.

Sifang You, James Cheng, and Yuhwa Lo,"Physics of self-recovering single photon avalanche detectors", Proc. SPIE 8155, 81551H (2011).

James Cheng and Yu-Hwa Lo, "Self-recovered InGaAs single photon avalanche detector with patterned Zn-diffused structure", Proc. SPIE 8154, 81540H (2011).

Arthur Zhang, Hongkwon Kim, James Cheng, Yu-hwa Lo, "Ultrahigh responsivity visible and infrared detection using silicon nanowire phototransistors", Nano Letters, v 10, n 6, p 2117 (2010).

Hongkwon Kim, Arthur Zhang, James Cheng, Yu-hwa Lo, "High-sensitivity visible and IR (1550nm) Si nanowire photodetectors", Proc. of SPIE 7608, 76082F (2010).

Arthur Zhang, James Cheng, Hongkwon Kim, Yisi Liu, Yu-hwa Lo, "Characterization and physics of top-down silicon nanowire phototransistors", Proc. of SPIE 7608, 76081D (2010).

James Cheng, Sifang You, Kai Zhao, and Yuhwa Lo, "Self-quenched InGaAs single-photon detector", Proc. SPIE 7320, 732010 (2009).

Kai Zhao, Sifang You, James Cheng, and Yu-Hwa Lo, "Self-quenching and self-recovering InGaAs/InAlAs single photon avalanche detector", Appl. Phys. Lett. 93, 153504 (2008).

ABSTRACT OF THE DISSERTATION

Self-Quenching Single Photon Avalanche Photodiodes for Near-Infrared Detection

by

James Cheng

Doctor of Philosophy in Electrical Engineering (Nanoscale Devices and Systems)

University of California, San Diego, 2012

Professor Yu-Hwa Lo, Chair

This thesis presents work on two significant design areas which improve the performance of single photon avalanche photodiodes (SPAD) for near-infrared detection. The first is a replacement for conventional zinc diffusion patterns, in which the field crowding effect of the junction is purposely utilized throughout the device. This design intends to address the practical non-uniformities that arises with device fabrication. Measurement results show >10x improvement in single photon detection efficiency (SPDE) while the dark count rate (DCR) is reduced

by >10x, when compared to single diffused wells.

The second is on a self-quenching epitaxial design, created through the use of a heterojunction energy barrier. Self-quenching removes the need for external quenching circuits, allowing for simpler manufacturing of large imaging arrays, faster timing performance, and higher efficiencies. Results from modeling, fabrication, and experimental measurements of these devices are discussed.

In particular, the dependence of the recovery time and afterpulsing rate on the device design and operating conditions are extensively examined experimentally. Implications of these effects, along with the SPDE, total DCR, and timing jitter of the device are predicted using a Monte Carlo model.

Chapter 1

Single Photon Detectors

The ultimate sensitivity limit for a photodetector is one single photon. To achieve this limit and obtain a good signal to noise ratio from the electronic readout circuit, the detector must utilize an internal gain mechanism. This chapter introduces single photon detectors (SPD) and their critical performance parameters for various single photon applications, with a focus on near-infrared detection [1, 2]. A review of quenching mechanisms and key figures of merit for avalanche photodiodes is also given.

1.1 Review of Single Photon Detectors

1.1.1 Photomultiplier Tubes

Photomultiplier tubes (PMT) were the first demonstration of a single photon detector. Based on two physical processes, the photoelectric effect and secondary emission, PMTs accelerate electrons in a vacuum tube with a high electric field (Figure 1.1, [3]). Secondary emission of electrons at each stage (dynode) produces gain, with each successive stage further multiplying the gain. The initial photon is absorbed by the photocathode and generates the first electron due to

the photoelectric effect. The output current from the PMT is measured at the anode.

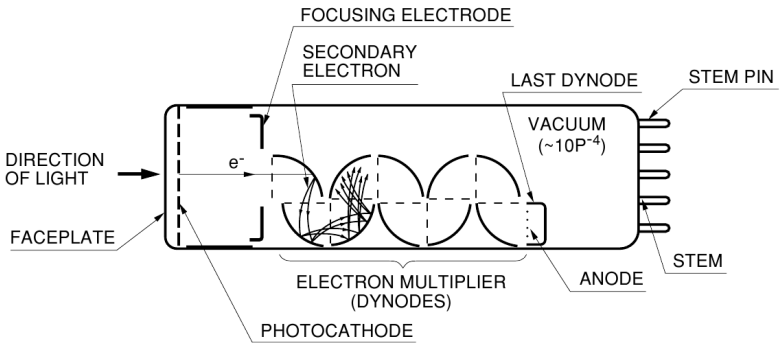

Figure 1.1: Operating principle of a photomultiplier tube.

PMTs have very low noise, with the main contribution of the dark current being thermionic emission from the photocathode and the dynodes. Even so, at typical operating conditions, the dark count rate is limited to less than 100 counts per second.

However, for near-infrared applications, PMTs are a poor choice due to the lack of a suitable photocathode material. State of the art PMTs for 1550 nm detection, utilizing InP/InGaAs as the photocathode, currently show a quantum efficiency of only 2%. PMTs also require very high voltages (thousands of volts), and are quite bulky (due to the vacuum tube) and too expensive for many applications.

1.1.2 Superconducting Nanowires

A relatively new development, superconducting nanowire photodetectors (SNPD) [4] are based on Cooper pairs breaking in the presence of light (Figure 1.2, [5]). The incident photon forms a localized hot spot which disrupts the su-

perconducting current. The current level around this localized hotspot builds up until it exceeds the critical current density and the nanowire is no longer superconducting, but rather has a measurable resistance. Monitoring of the wire resistance then provides information on whether the detector has sensed any radiation.

SNPD can show very good dark counts (<1 kHz [6]), detection efficiencies (57% at 1550 nm [7]), and low timing jitter (< 50 ps). However, current challenges with SNPD are the low temperature requirement for superconductivity (liquid helium cooling), small detector photosensitive area size due to the nanoscale size of the wires (electron beam lithography is required for fabrication), and nonuniformity of nanowires when attempting to fabricate large area devices (limiting how close the SNPD can be biased to its critical current level, thus reducing sensitivity).

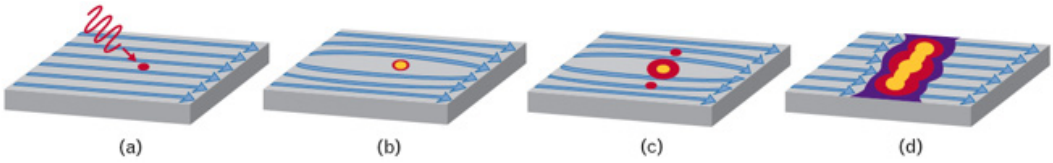

Figure 1.2: Operating principle of a superconducting nanowire photodetector.

1.1.3 Avalanche Photodiodes

Avalanche photodiodes (APD) are semiconductor diodes (p-n junctions) which uses a high internal electric field in order for electrons and holes to cause impact ionization. Impact ionization, the inverse of the Auger process, is when a carrier gains enough energy to create a secondary electron hole pair. The impact ionization process can be described as depending only on the local electric field [8] (contrast to the nonlocal theory presented in [9]), with the generation rate per

centimeter for electrons (α) and holes (β) fitting the form

$$\alpha(E) = \alpha_0 \exp(-[c_n/E]^{m_n}) \tag{1.1}$$

$$\beta(E) = \beta_0 \exp(-[c_p/E]^{m_p}) \tag{1.2}$$

The above rates can be derived theoretically from first principles based on the band structure [10–13], but in practice the coefficients are fitted to experimental data [14–16]. Note that the rates are also an implicit function of temperature; as the temperature is reduced, energy loss to phonon scattering is less, which increases the ionization rate [17].

Both electrons and holes can go through impact ionization and generate additional carriers. The newly generated carriers can also obtain enough energy as they travel through the high electric field to undergo additional impact ionizations. This results in an "avalanche." The multiplication gain of a p-i-n, with a multiplication width of W, is

$$M = \left\{1 - \int_0^W \alpha \exp\left[-\int_x^W (\alpha - \beta)dx'\right] dx\right\}^{-1} \tag{1.3}$$

where α and β are spatially dependent [18], due to non-uniformities in either the electric field or changes in the multiplication material itself (e.g. superlattice structures). For an APD with a uniform electric field and ionization rate across the multiplication region, the gain equation simplifies to

$$M = \frac{1-k}{\exp(-\alpha W[1-k]) - k} \tag{1.4}$$

where $k \equiv \beta/\alpha$. Note that when the denominator becomes zero, the gain of the APD becomes infinite. The voltage at which this occurs is called the "breakdown" voltage, and operating above this voltage is called "Geiger-mode." Single photon

sensitivities are reached in Geiger-mode operation. Note that a thicker multiplication region requires less electric field to reach the breakdown condition.

APDs for near-infrared detection typically utilizes a separate absorption and multiplication (SAM) structure (Figure 1.3). A low bandgap material, such as InGaAs, is used as the absorption layer for the near-infrared photons. A large bandgap material, such as InP, is placed next to the absorption layer and is used as the multiplication material. Implicit in this design is a proper grading and charge layer in between the absorption and multiplication regions.

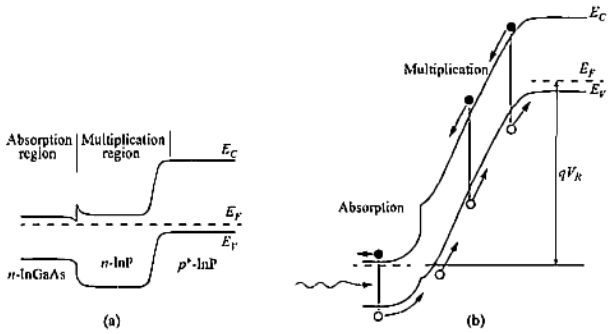

Figure 1.3: Typical separate absorption and multiplication structure for near-infrared APDs, under (A) equilibrium and (B) reverse bias conditions.

This SAM design has the benefit of decoupling the photo-generation and the avalanche processes from each other [17]. A high electrical field is required in the multiplication region to initiate the avalanche process. If this region uses the same material as the low bandgap absorption region, there would be significant tunneling current. A charge layer can be place between the multiplication p-i-n junction and the absorption region, with the correct dosage to balance the fields in the two regions. The electric field profile under different reverse bias voltages is shown in Figure 1.4(A).

When the initial bias is applied, the voltage is dropped across the p-i-n

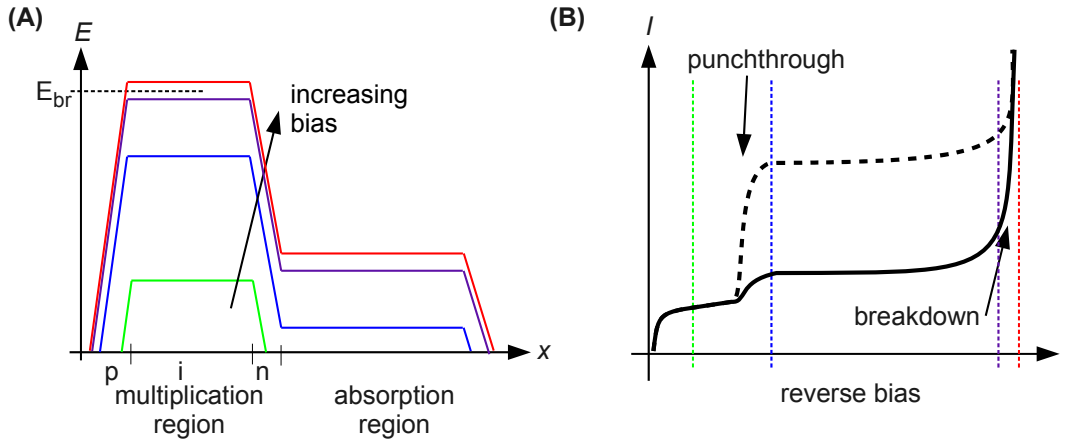

Figure 1.4: (A) Schematic electric field profile of SAM-APD structure at different reverse bias voltages. (B) Schematic semi-log I-V plot of SAM-APD, under dark (solid) and light (dashed) conditions, with lines depicting the corresponding voltages of the field profile.

junction. There is little field in the absorption region, and thus the current is low (Figure 1.4(B), green line). When the reverse bias is higher (same figure, blue lines), the charge control layer of the p-i-n becomes fully depleted, and there is some electric field in the absorption region. This corresponds to a large increase in the current, and is called the "punchthrough" voltage. Further increases in the reverse bias (purple lines) results in gain in the current, and when the electric field in the multiplication region becomes greater than the breakdown field (red lines), the current becomes infinite. The device is single photon sensitive in this mode.

1.2 Conventional SPAD Quenching Circuits

When the device is operating in Geiger-mode, a single carrier can trigger infinite current. The avalanche must be stopped in order for the device to be able to detect another photon, and must be done quickly to prevent any damage to the

device. There are several conventional methods of "quenching" the avalanche [19].

1.2.1 Passive Quenching

The simplest method of creating a quenching circuit is to place a large resistor (e.g. MΩ's) in series with the SPAD, so the applied operating bias is divided over the resistor and the SPAD. When there is little current flow, the voltage drop across the resistor is minimal. When the SPAD triggers an avalanche and generates a large current (e.g. 10-100 μA's), the voltage drop across the resistor becomes substantial enough to reduce the voltage applied on the diode to that below the breakdown voltage, stopping the avalanche process.

While the quenching process is relatively fast, the discharge of the current through the large resistance is quite slow. The dead time between photon sensing using passive quenching is related to the RC of the circuit. With an estimated 0.1 pF for a 20 μm radius SPAD and a 1 MΩ quenching resistor, the time constant is on the order of 100 nanoseconds. Additionally, it is typically difficult to reliably create uniform resistors on III-V materials.

1.2.2 Active Quenching

In contrast to passive quenching, an active quenching circuit monitors the current output from the SPAD and reduces the applied bias once it senses the avalanche current. These are typically rather complex, difficult to use for large array detectors, and are a very active research field in themselves [19–23].

After detection of the avalanche pulse, the applied voltage bias on the device is reduced for some hold-off time. It is important to design the hold-off time to completely quench the avalanche and minimize any afterpulsing effects. After the

hold-off time, the device is actively reset to the high voltage state, ready to detect another photon.

A promising external quenching circuit utilizes a passive quenching mechanism along with an active reset (Figure 1.5, [22]). This is achieved by using a normally off transistor as the quenching resistor, and turning the transistor on to allow for high current flow during the reset process. The benefits of this design is a much simpler quenching circuit, without the drawbacks of passive quenching.

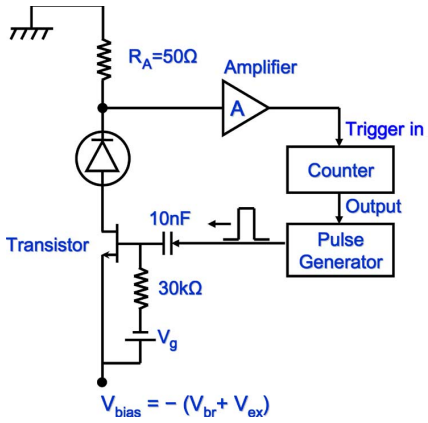

Figure 1.5: Passive quenching active reset circuit.

1.2.3 Gated Mode

The two previous methods operate the device in the free-running mode; that is, the detector is always ready to sense a photon until it fires an avalanche pulse. An alternate method of operating the SPAD is in gated mode. In this mode, the voltage applied on the device is usually kept below the breakdown voltage. Only for a short time interval is the voltage brought above the breakdown voltage, and it is during this gate that the detector is active.

Of course this mode is only suitable when the timing of the arriving photon

is known. It also introduces another class of problems: transients canceling. The voltage pulse used to gate the SPAD, when passed through the diode, usually results in a voltage signal much larger than the device's own avalanche response. There has been much active work on designing suitable canceling circuits to operate in gated mode [24]. A crucial issue is the accurate removal of these transient signals when operating the SPAD at very high speeds (e.g. GHz). A cursory overview of transients canceling will be briefly mentioned.

One method uses dual coaxial transmission lines of equal lengths before and after the diode, with one terminated by an open and the other line terminated by a short circuit [25]. The shorted line will invert the transient, which cancels out the non-inverted transient that traveled through the open line. This leaves just the avalanche signal intact at the output.

Another method utilizes a delay line corresponding to the gate frequency [26]. Shifting the transient by one clock cycle, the delayed signal is then fed into a differentiator along with the original non-shifted device output. The resulting differential signal should only contain the desired avalanche pulse. Alternatively, one can use a sinusoidal voltage gate rather than a square gate, and then using the appropriate filter to remove the gating frequency [27, 28].

1.3 Important Figures of Merit

1.3.1 Single Photon Detection Efficiency

The single photon detection efficiency (SPDE) is the probability in which the detector will fire an output signal when a single photon is incident on the detector. For a device with no false counts, this can be obtain by illuminating the

detector with exactly one photon for N trials, recording the number of counts C, and the efficiency is C/N.

However, true single photon sources are rather difficult to create [2, 29, 30]. Most single photon sources are simulated (i.e. "weak coherent pulses") using very fast lasers (e.g. picosecond pulse widths) and attenuating by the proper amount. Due to the Poisson nature of the laser source [31, 32], even when the mean number is 1 photon/pulse, there is a significant chance ($1 - 1/e \approx 63\%$) of not obtaining 1 photon/pulse.

We would like to calculate the true SPDE when provided with a photon source which follows the Poisson distribution. Given a mean λ photons/pulse, the probability of actually obtaining k photons is given as:

$$P(X = k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1.5}$$

The probability of a detection event is the probability of at least one of the multiple photons triggering an event, summed over the probability distribution for the number of photons:

$$\begin{aligned} P_{DE} &= \sum_{k=0}^{\infty} P(k, \lambda)[1 - (1 - SPDE)^k] \qquad (1.6) \\ &= \sum_{k=0}^{\infty} P(k, \lambda) - \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}(1 - SPDE)^k \\ &= 1 - \sum_{k=0}^{\infty} \frac{[\lambda(1 - SPDE)]^k}{k!} e^{-\lambda(1-SPDE)} e^{-\lambda \cdot SPDE} \\ &= 1 - \sum_{k=0}^{\infty} \frac{(\lambda')^k}{k!} e^{-\lambda'} e^{-\lambda \cdot SPDE} \\ &= 1 - e^{-\lambda \cdot SPDE} \qquad (1.7) \end{aligned}$$

In the case $\lambda \ll 1$, $P_{DE} \approx \lambda \cdot SPDE$. When there are C detection events in N

trials, we get the true SPDE as

$$SPDE = \frac{C}{N\lambda} \tag{1.8}$$

The SPDE of a SPAD depends on two physical processes. The first is the probability that the incident photon will generate an electron-hole pair which is able to drift out of the absorption region before recombining: the quantum efficiency QE. The second is the probability that the carrier injected into the multiplication region from the absorption region will initiate an avalanche: the avalanche probability η_{av}. The SPDE can then be express as

$$SPDE = QE * \eta_{av} \tag{1.9}$$

Operating the SPAD as a regular p-i-n photodiode, we can see that the QE of a well designed detector is quite high. Most of the work on improving the SPDE for SPADs goes into improving the avalanche probability, through either practical fabrication considerations (see Chapter 3) or other design parameters. Note that the increase in SPDE with bias voltage is due mostly to the increase in η_{av}, which is highly dependent on the electric field.

1.3.2 Dark Counts

Dark counts are false detection events from the detector, in the absence of any photo input. Since the detector is single photon/electron sensitive, any electrons generated in the dark will produce the same current response as the desired photons. They are caused by the random generation processes intrinsic to any semiconductor device, and are of great importance when designing a good APD. Here we will qualitatively discuss the dominant sources of dark counts:

thermal generation in the absorption region, tunneling in the multiplication region, and afterpulsing. Numerical analysis of these rates are discussed in Section 5.3.

The first source of dark counts are from thermally generated carriers. For a SAM-APD, the dominant source of thermally generated carriers are from the low bandgap absorption region material, rather than the high bandgap multiplication material. As thermally generated carriers have a $\exp(-E/kT)$ dependence [33], it can be greatly reduced by decreasing the temperature. For cases where the temperature cannot be reduced too low, minimizing the volume of the absorption material and improving the material defects are alternative options.

Secondly, thin multiplication regions suffer from high tunneling current due to the high field necessary for impact ionization [34]; whereas the charge layer in the SAM-APD structure can be designed to control the field in the absorption region so tunneling in this region is minimal. To prevent significant tunneling in the multiplication region, it should be sufficiently thick. A thick multiplication region has two effects: less electric field required for breakdown, and smaller tunneling probability given the same electric field. Unfortunately a thick multiplication region also degrades the timing jitter (see Section 1.3.4) and excess noise (Section 1.3.5).

During the avalanche process, carriers can become trapped in defect states within the multiplication material. When they are later released from the trap states, they can trigger their own avalanche. This "afterpulsing" effect is dependent on the size of the avalanche pulse (i.e. total current flow through the material), the hold off time after triggering the primary avalanche (dependent on the employed quenching mechanism), and the intrinsic material property. As improvements in the material growth progresses slowly, most work in SPADs try to limit afterpulsing

by controlling the first two effects. The work in this thesis, using an internal self-quenching mechanism, also attempts to limit the total avalanche gain and engineer the correct hold off time to minimize afterpulsing.

1.3.3 Counting Rate / Hold-off Time

The saturation rate is the maximum photon illumination rate that the detector is capable of observing before the efficiency drops, while the hold-off time, or dead time, is the period after the initial detection event in which the SPDE is zero. Both are highly dependent on the particular quenching circuit used to operate the SPAD.

For active reset circuits, the rate is proportional to the inverse of the hold-off time; whereas in passive reset circuits, there is no hard threshold as the SPDE recovers monotonically with time. For gated mode operation, this is simply the gating rate.

The maximum counting of the detector is the rate that the detector is capable of producing an output signal. For devices with a passive reset circuit or the self-recovery devices presented in this thesis, there is no hard cutoff in the SPDE recovery mechanism, and the maximum counting rate can be higher than the saturation rate. For example, a device may saturate at 1 MHz and have 20% SPDE, but at 10 MHz the SPDE drops to 15%. However, the latter case yields a faster overall counting rate.

1.3.4 Timing Jitter

Timing jitter is the variation in when the detector produces a detectable signal. In a time correlate counting histogram, this can be represented by the pulse

width of the correlation peak, where, depending on the application, either the full width half maximum, or full width at 1/10th or 1/100th maximum is important (Figure 1.6, [35]). The timing jitter of the detector is usually the limiting factor of the temporal resolution of the measurement system.

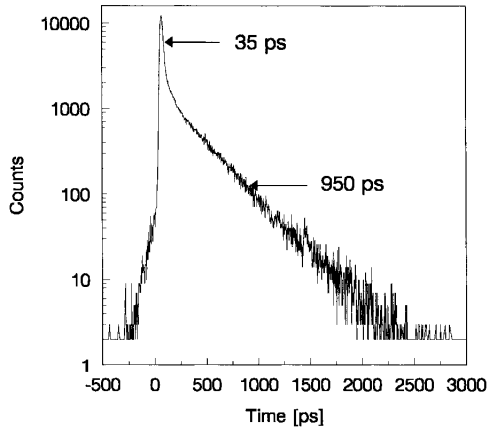

Figure 1.6: Example timing characteristic of a shallow junction SPAD, with arrows indicating full widths at half and 1/100th maximum.

The jitter in a SAM-APD is comprised of two components: jitter in the avalanche buildup process, and variations in the depth of photogenerated carrier. For the latter, the actual absorption location of the photon within the material follows an exponential dependence, and thus there would be variations in the time it takes for the carrier to drift out of the absorption region. Assuming a carrier velocity near the saturation velocity of 10^7 cm/s, and a 1 μm thick absorption region, this corresponds to approximately 10 ps timing jitter from this component.

The former component, jitter in the avalanche buildup, is usually the more significant component, on the order of hundreds of picoseconds. A high electric field, thin multiplication region, and lower ionization coefficient ratio ($k = \beta/\alpha$, Section 1.3.5) will all reduce the timing jitter by reducing the noise of the avalanche process.

1.3.5 Excess Noise

Ideally the avalanche response to a single photon should remain the same under constant operating conditions. The excess noise factor defines the fluctuation in the signal and is expressed as

$$F = \frac{\langle M^2 \rangle}{\langle M \rangle^2} = 1 + \frac{\langle \Delta M^2 \rangle}{\langle M \rangle^2} \tag{1.10}$$

where M is the multiplication gain. Good excess noise is necessary for linear mode APDs (as opposed to Geiger mode), where the gain is finite and the desired amplification response should be linear. It is also useful for SPADs which are capable of photon number counting (see Section 1.3.6).

Following McIntyre's model for multiplication noise [36], the excess noise is shown to be related to the ratio of the impact ionization coefficients $k = \beta/\alpha$. An intuitive explanation of the relation is obtained by observing the two extreme cases: $k \to 0$, and $k = 1$. In the first case, only a single carrier can create impact ionization events. Thus the multiplication gain only occurs during a single pass of the initial carrier, and is easily predicted. On the other hand, if both carriers contribute equally to impact ionization (i.e. $k = 1$), there is a positive feedback mechanism between hole impact ionizations and electron impact ionizations, and thus the total gain fluctuation is more. Figure 1.7 plots the noise factor as a function of the multiplication gain and the ionization ratio [18].

Previous SPAD designs have tried to minimize the excess noise by using materials which are intrinsically single carrier ionizing, or by engineering superlattices to have the same effect (e.g. regularly create band discontinuities so only one carrier type obtains enough energy for impact ionization [37, 38]). Thinner multiplication regions at higher electric fields also demonstrate better noise fac-

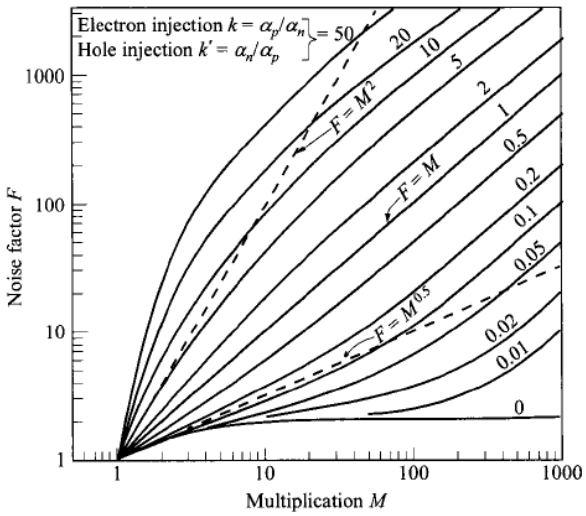

Figure 1.7: Theoretical noise factor F vs multiplication gain M, at different ionization coefficient ratios.

tors [14, 39]. The work is this thesis limits excess noise due to the self-quenching negative feedback mechanism, which stops the avalanche after a certain gain is reached [40].

1.3.6 Photon Number Resolution

Photon number resolution is the ability to resolve the number of incident photons, when the number is larger than one. Devices with infinite gain such as SPADs cannot differentiate between one and multiple photons. The ideal device will scale a large but finite output directly in proportion to the number of incident photons.

Note that any regular single photon detector can obtain photon number resolution via multiplexing of multiple devices in parallel. A large matrix of individual devices collectively form a single readout line. When multiple photons hit different pixels, if the excess noise of the device is good, the pulse amplitude will

directly depend on the number of pixels firing. The resolution of course depends on the SPDE of the individual pixel, and the photon flux low enough so the same pixel will not receive more than one photon at a time.

1.4 Dissertation Outline

The rest of the thesis will describe the work on self-quenching designs and patterned zinc diffusion designs. While both aspects can be independently applied to any APD to improve its performance, the fabricated devices in this work will utilize both. This chapter already introduced the important figures of merit to consider when designing SPADs, which will need to be constantly taken into consideration throughout the work.

Chapter 2 describes the operating principle of the self-quenching and self-recovering design, and creates a mathematical model to describe the process as a function of the device structure. Chapter 3 describes a new zinc patterned design to improve avalanche efficiencies, and fabrication and measurement of the device. The performance is shown to be limited by the self-recovery process. Chapter 4 then presents how to reduce the self-recovery time, experimental verification of the change, and analysis of the implications of the smaller barrier. Afterpulsing is shown to be a key limitation of the device, and a complete device performance model to take this effect into account is described in Chapter 5.

Chapter 2

Self Quenching and Self Recovering Design

The previous chapter introduced the complexity and inefficiencies of conventional quenching circuits for operating a SPAD in Geiger-mode. In this chapter we use bandgap engineering to create a design which removes the need for external quenching circuits. The intentional bandgap discontinuity creates a negative feedback mechanism after an avalanche event in order to self-quench it.

The benefits and qualitative understanding of this design and the negative feedback mechanism is first described. We then present a deterministic model of the self-quenching, self-recovering SPAD which incorporates a TCB heterojunction barrier. We can quantify the transient characteristics of the self-quenching process, and its dependence on the device design. This model can be used to determine the timing performance of the device. In particular, the recovery time of the device will be examined in detail, as this parameter directly correlates with the maximum speed the device can operate at.

The first section of the model details a derivation in which a time constant is assumed, which relates the relative ease of the carriers to escape from the barrier. The latter section will remove this constant by deriving the current equations based off the physical escape mechanisms of the carrier, namely tunneling and thermionic

emission.

2.1 Design Motivation

Chapter 1 introduced the need of quenching the avalanche after a SPAD detects a photon. The motivation for a self-quenching design are many. The first is by removing the need for external circuits, the fabrication process is greatly simplified. This is especially beneficial for III-V materials where it is difficult to fabricate complex circuits, reduces the device cost, increases the fill factor of photosensitive area, and makes large area imaging arrays possible. This last point is of particular interest, as most of the high performance external circuits can reported in literature can operate on only a single device. On the other hand the self-quenching design needs only a very simple readout circuit associated with each pixel.

The second is that by incorporating the quenching mechanism within the device, instead of being external to the device, the quenching and recovering processes become much faster. This means a device with faster response, less timing jitter, and less excess noise. The low noise avalanche process in turn means a device capable of resolving the number of photons, and not saturate at just one photon.

2.2 Self-Quenching Design Concept

The core of the self-quenching mechanism is a layer called "Transient Carrier Buffer," or TCB. In practice, this is formed by creating a heterojunction barrier next to the multiplication region. Any avalanche generated carriers will temporar-

ily stop at this barrier (Figure 2.1(A)). The collected charges screen the electric field across the multiplication region, reducing it to below the breakdown field and stopping the avalanche process, hence self-quenching. As the carriers escape from the barrier, the field in the multiplication region returns to the original field, and the device is self-recovered.

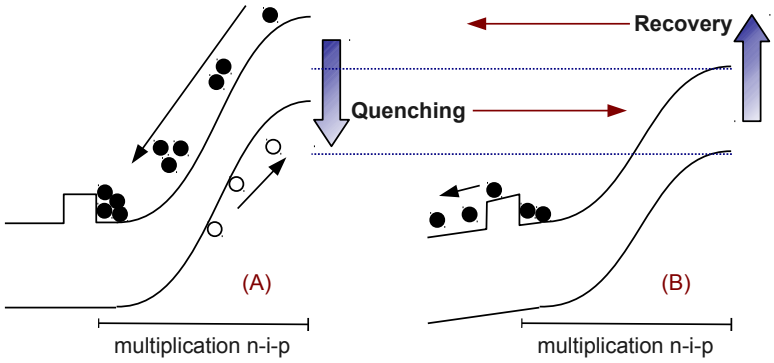

Figure 2.1: Schematic band diagram of the SPAD illustrating the (A) self-quenching and (B) self-recovery operating principles. Note the change in the voltage in the multiplication region during this process, due to electron collection and release at the conduction band offset.

Note that during the entire process, the applied bias on the terminals of the device is constant. Only the voltage distribution within the device changes, which quenches or recovers the device. The constant DC bias allows for convenient and easy usage of the SPAD. Note that both a conduction band barrier next to the n layer of the multiplication p-i-n or a valence band barrier next to the p layer can produce a self-quenching SPAD, by stopping either electrons or holes, respectively.

2.3 Model Setup

A 1-D model of the SPAD is created with the two device regions of interest: the multiplication region from $x = 0$ to $x = W$, and the TCB region from $x = W$ to $x = W + L$ (Figure 2.2). Holes travel through the multiplication region in the $+x$ direction, and are injected from the absorption region into the multiplication region at $x = 0$. A single hole (e.g. generated by a single photon or thermally) entering the multiplication region is then expressed as:

$$J_p(x = 0, t) = e\delta(t)/S \tag{2.1}$$

where S is the effective area of the diode.

Figure 2.2: Regions of interest in the self-quenching SPAD model, with a valence band barrier for holes at $x = W$.

The ionization rates of electrons and holes are given by the local field approximation model, and the electric field throughout the multiplication region is assumed to be uniform. Then the ionization rates are given by Equation (1.1) and the generation rate due to impact ionization is:

$$G = G_n = G_p = \beta J_p + \alpha J_n \tag{2.2}$$

We will neglect thermal generation in the high bandgap multiplication region for simplicity. Using the above generation rates, the continuity equations

within the multiplication region are given as

$$e\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}J_p(x,t) + G \tag{2.3}$$

$$-e\frac{\partial}{\partial t}n(x,t) = -\frac{\partial}{\partial x}J_n(x,t) - G \tag{2.4}$$

Using $J_p = epv_p$ and $J_n = env_n$, the previous equations can be expressed as

$$\frac{1}{v_p}\frac{\partial}{\partial t}J_p(x,t) = -\frac{\partial}{\partial x}J_p(x,t) + (\beta J_p + \alpha J_n) \tag{2.5}$$

$$-\frac{1}{v_n}\frac{\partial}{\partial t}J_n(x,t) = -\frac{\partial}{\partial x}J_n(x,t) - (\beta J_p + \alpha J_n) \tag{2.6}$$

where it is assumed the carriers travel at their saturation velocities due to the high electric field. The total current is constant with respect to position, and is the sum of the hole, electron, and displacement currents:

$$J(t) = J_p(x,t) + J_n(x,t) + \frac{\epsilon}{W}\frac{dV_m}{dt} \tag{2.7}$$

As there are only two regions of interest in this model, the voltage across the TCB region is $V_{TCB} = V - V_m$, where V is the total applied bias on the device (i.e. constant with respect to time) and V_m is the voltage across the multiplication region. We can also model the TCB as a leaky capacitor, e.g. a resistor and capacitor in series. A time constant, denoted τ, will refer to the escape time of the carrier from the barrier interface, and have an associated RC.

$$\begin{aligned} J(t) &= \frac{V}{R} + C\frac{dV}{dt} \\ &= \frac{\epsilon}{L\tau}V_{TCB} + \frac{\epsilon}{L}\frac{d}{dt}V_{TCB} \end{aligned} \tag{2.8}$$

2.4 Transients Derivation

We would like to solve the equations given in the previous section for $V_m(t)$. Once V_m is obtained, the transient avalanche pulse can be examined. Re-arranging

(2.7) and (2.5),

$$J(t) - \frac{\epsilon}{W}\frac{dV_m}{dt} - J_p = J_n$$
$$\alpha(J(t) - \frac{\epsilon}{W}\frac{dV_m}{dt} - J_p) = \frac{1}{v_p}\frac{\partial}{\partial t}J_p(x,t) + \frac{\partial}{\partial x}J_p(x,t) - \beta J_p \tag{2.9}$$

Likewise with (2.7) and (2.6),

$$\alpha(J(t) - \frac{\epsilon}{W}\frac{dV_m}{dt} - J_p) = \frac{1}{v_n}\frac{\partial}{\partial t}J_n(x,t) - \frac{\partial}{\partial x}J_n(x,t) - \beta J_p \tag{2.10}$$

For ease lets calculate the relations

$$\frac{dJ_p}{dx} = -\frac{dJ_n}{dx} \tag{2.11}$$
$$\frac{\partial J_p}{\partial t} + \frac{\partial J_n}{\partial t} = \frac{\partial J}{\partial t} - \frac{\epsilon}{W}\frac{d^2V_m}{dt^2} \tag{2.12}$$

Equation (2.11) is obtained by taking the spatial derivative of (2.7), which is independent of position, while (2.12) is simply the time derivative of (2.7).

Using these relations, we find $v_n\times$ (2.10) + $v_p\times$ (2.9) as

$$\alpha(J - \frac{\epsilon}{W}\frac{dV_m}{dt} - J_p) + \beta J_p = \frac{1}{v_p + v_n}(\frac{dJ}{dt} - \frac{\epsilon}{W}\frac{d^2V_m}{dt^2}) + \frac{\partial J_p}{dx}$$

Or arranging so the J_p terms are on the LHS:

$$\frac{\partial J_p}{dx} - (\beta - \alpha)J_p = \alpha(J - \frac{\epsilon}{W}\frac{dV_m}{dt}) - \frac{1}{v_p + v_n}(\frac{dJ}{dt} - \frac{\epsilon}{W}\frac{d^2V_m}{dt^2}) \tag{2.13}$$

Let $A \equiv \beta - \alpha$. Integrating J_p in (2.13) through the multiplication region (i.e. x from 0 to W) results in

$$e^{-AW}J_p(W) - J_p(0) = (1 - e^{-AW})\frac{1}{A}\times$$
$$\left[\alpha(J - \frac{\epsilon}{W}\frac{dV_m}{dt}) - \frac{1}{v_p + v_n}(\frac{dJ}{dt} - \frac{\epsilon}{W}\frac{d^2V_m}{dt^2})\right] \tag{2.14}$$

At the anode of the multiplication p-i-n region (i.e. $x = W$), there is zero electron current, or $J_n(x = W) = 0$. From (2.8) and (2.7)

$$J_p(x = W) = \frac{\epsilon}{L\tau}(V - V_m) + \epsilon\left(\frac{1}{L} + \frac{1}{W}\right)\frac{d(V - V_m)}{dt} \tag{2.15}$$

Solving (2.8), (2.14) and (2.15) for V_m, using the initial condition (2.1), we obtain

$$\begin{aligned}\left[-\frac{1 - e^{-AW}}{A}\frac{1}{(v_p + v_n)}\left(\frac{1}{W} + \frac{1}{L}\right)\right]\frac{d^2V_m}{dt^2} \\ + \left[\left(\frac{1}{W} + \frac{1}{L}\right)e^{-AW} + (1 - e^{-AW})\frac{\alpha}{A}\left(\frac{1}{W} + \frac{1}{L}\right) - \frac{1 - e^{-AW}}{A(v_p + v_n)L\tau}\right]\frac{dV_m}{dt} \\ + \left[\frac{1}{L\tau}e^{-AW} - (1 - e^{-AW})\frac{\alpha}{AL\tau}\right](V - V_m) - \frac{e}{S\epsilon}\delta = 0\end{aligned} \tag{2.16}$$

Equation (2.16) is an ordinary differential equation that can then be numerically computed for V_m as a function of time. Keep in mind that the ionization coefficients α and β have an exponential dependence on the multiplication voltage ($E = V_m/W$). We can vary the device design parameters multiplication width W, TCB width L, applied bias V, and escape time τ (e.g. barrier height), and study how these effect the device performance in metrics such as recovery time and gain.

2.5 Simulation Results

The device is simulated to operate in Geiger-mode, so the breakdown voltage needs to be calculated first. The breakdown voltage is when the denominator of the gain equation in (1.4) becomes zero. After determining this voltage, the bias voltage on the device is set to a few percent above the breakdown voltage. Figure 2.3 shows the simulation response to a single photon arriving at t=100 under a small percentage overbias. Part (A) demonstrates the fast multiplication voltage drop during the avalanche process, which in turn quenches the avalanche

current seen in part (B) of the figure. After the avalanche event, the multiplication voltage slowly recovers, and this corresponds to a small discharge current. This simulation demonstrates that the key effects of our design works in producing a self-quenching, self-recovering SPAD.

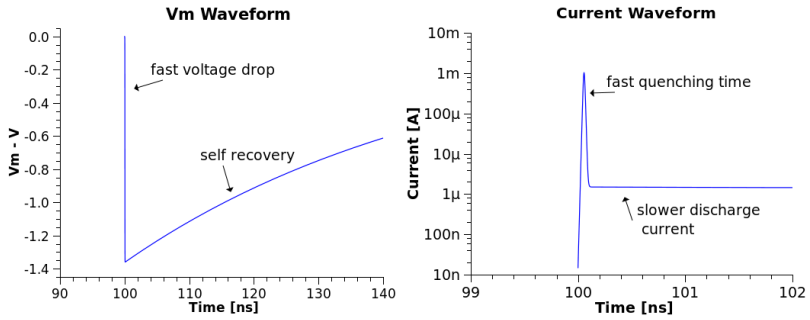

Figure 2.3: Simulated response to a single photon: (A) multiplication voltage waveform, (B) current waveform.

We will next want to examine the response dependence on carrier escape time τ. Voltage waveforms at different τ are shown in Figure 2.4(A). We see that as the escape time decreases, the device recovery is faster. On the other hand, the device gain is not a strong function of the escape constant, as the quenching speed is already much faster than recovery time. The gain of this device can be calculated by integrating the area of the current pulse, and its dependence on τ is seen in (B). Only when the recovery time is on the order of the quenching speed (e.g. sub-nanosecond) does the gain increase. As the barrier height approaches zero (e.g. $\tau \to 0$), the gain approaches infinity, like a regular SPAD operating in Geiger mode.

Lastly, we demonstrate the device response when it is not fully recovered, and show how the detection efficiency and photon count rate is limited by the recovery time of the device. In Figure 2.5 we input a second photon after some

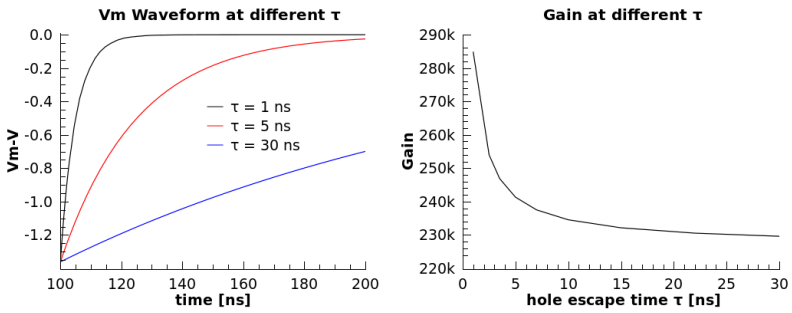

Figure 2.4: Dependence on escape time: (A) multiplication voltage waveform, (B) total gain.

time delay from the first input, and observe the device's current response to the second photon. As seen in the 180 ns case, the device has not recovered from the first photon input and produces no detectable current response. At 200 ns between photon arrival, the device recovered enough to produce a weak avalanche response. At 260 ns, there is enough time between photons for the second input to produce a full size avalanche with the same gain as the first input response.

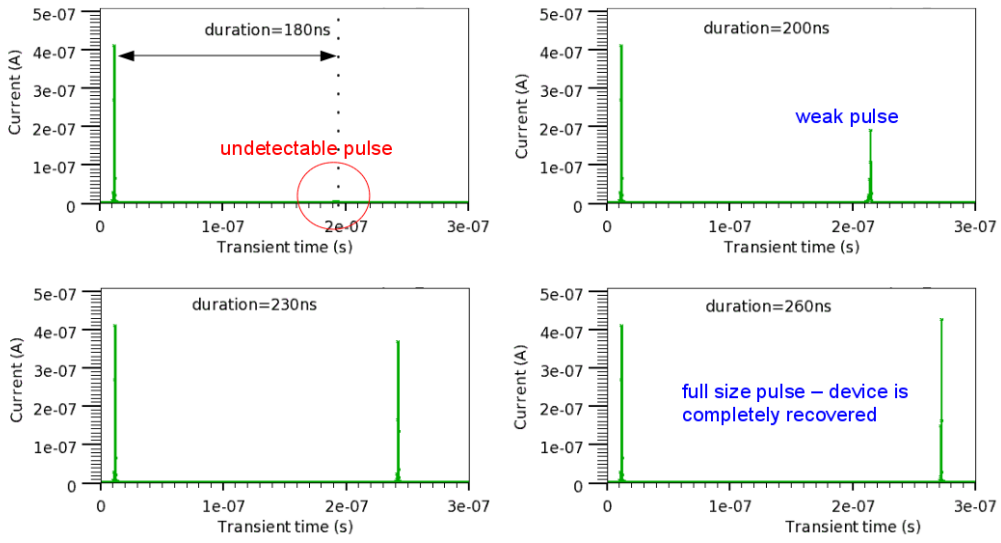

Figure 2.5: Simulated current response to multiple input photons with different arrival times between inputs: 180 ns, 200 ns, 230 ns, and 260 ns.

We conclude that the hole escape time is an important parameter of interest for any real SPAD. In contrast to a conventional SPAD where the maximum detection rate depends on the external quenching circuit, for the self-quenching SPAD it depends on the carrier escape time from the barrier. As this physically depends on the barrier design, both the self-quenching process and the self-recovery process are interlinked to each other.

2.6 Physical Interface Escape Model

While the time constant τ is useful to intuitively understand the device response, we would like to show an explicit relationship between the device recovery time and the physical heterostructure barrier design. We do this by calculating the escape currents from the interface due to thermionic emission and intraband tunneling. Once we know these escape current, we can then replace any τ dependencies in the model with the equivalent escape current. To match the notation from [41], in this section we will examine electron escape from a conduction band TCB barrier. By relating the escape current to its physical mechanism, with this modification the derivation can therefore account for bias and temperature dependent performance of the recovery time, without the need of any experimental fitting parameters.

2.6.1 Thermionic Emission

Following [18], given the coordinate system where the conduction band at the side with lower energy $E_{c-} \equiv 0$, for a barrier height defined as $\phi_B = \Delta E_c - E_{fn}$,

the thermionic emission current J_{th} is calculated as

$$J_{th} = \int_{\Delta E_c}^{\infty} q v_x dn \tag{2.17}$$

$$dn = N(E)F(E)dE \tag{2.18}$$

$$\approx \frac{4\pi(2m^*)^{3/2}}{h^3}\sqrt{E-E_c}\exp\left(-\frac{E-E_c+q\phi_n}{kT}\right)dE \tag{2.19}$$

$$\approx 2\left(\frac{m^*}{h}\right)^3 \exp\left(-\frac{q\phi_n}{kT}\right)\exp\left(-\frac{m^*v^2}{2kT}\right)(4\pi v^2 dv) \tag{2.20}$$

$$J_{th} = A^*T^2\exp\left(-\frac{q\phi_n}{kT}\right)\exp\left(-\frac{m^*v_{0x}^2}{2kT}\right) \tag{2.21}$$

$$= A^*T^2\exp\left(-\frac{q\phi_B}{kT}\right) \tag{2.22}$$

where $A^* \equiv 4\pi q m^* k^2/h^3$ is the effective Richardson constant.

2.6.2 Intraband Tunneling

From the WKB approximation, the tunneling probability through a triangular barrier is:

$$T_t(E_x) = \exp\left(-\frac{4\sqrt{2m^*}}{3qF\hbar}(\Delta E_c - E_x)^{3/2}\right) \tag{2.23}$$

where the electric field F across the barrier is, in the case of a self-quenching SPAD, $F_{TCB} = (V - V_m)/L$. The tunneling current is:

$$J_{tu} = \frac{4\pi q m^*}{h^3}\int_0^{\Delta E_c} T_t(E_x)dE_x \int_0^{\infty}(f_1(E) - f_2(E))dE_\perp \tag{2.24}$$

$$\approx \frac{4\pi q m^*}{h^3}\left(\exp(\frac{E_{fn1}}{kT}) - \exp(\frac{E_{fn2}}{kT})\right) \times \int_0^{\Delta E_c} T_t \exp(\frac{E_x}{kT})dE_x \int_0^{\infty} T_t \exp(\frac{E_\perp}{kT})dE_\perp \tag{2.25}$$

$$\approx \frac{4\pi q m^* kT}{h^3}\exp(\frac{E_{fn1}}{kT})\int_0^{\Delta E_c} T_t(E_x)\exp(-\frac{E_x}{kT})dE_x \tag{2.26}$$

2.6.3 Model Modifications

With the expression for the total escape current $J_t = J_{th} + J_{tu}$, the total current in the TCB region (compare to Equation 2.8) becomes

$$J(t, x > W) = J_t + \frac{\epsilon}{L}\frac{d}{dt}V_{TCB} \tag{2.27}$$

The E_{fn} terms in the thermionic emission and tunneling currents can be related to the electron concentration at the TCB barrier as a function of $V_m(t)$ by

$$\sigma = \frac{kTL}{q(V - V_m))}N_c \exp\left(\frac{E_{fn}}{kT}\right) \tag{2.28}$$

where the electric fields in the two regions are related by the sheet charge σ accumulated at the TCB interface:

$$E_m = E_{TCB} - \frac{q\sigma}{\epsilon} \tag{2.29}$$

Following Section 2.4, the rest of the equations can be calculated with the new escape current, modifying the boundary condition to $J_n(x = W) = 0$, resulting in

$$\begin{aligned}\left[-\frac{1 - e^{-AW}}{A}\frac{1}{(v_p + v_n)}\left(\frac{1}{W} + \frac{1}{L}\right)\right]\frac{d^2V_m}{dt^2} \\ + \left[\left(\frac{1}{W} + \frac{1}{L}\right)e^{-AW} + (1 - e^{-AW})\frac{\beta}{A}\left(\frac{1}{W} + \frac{1}{L}\right)\right]\frac{d(V - V_m)}{dt} \\ + \frac{1 - e^{-AW}}{A(v_p + v_n)\epsilon}\frac{dJ_t}{dt} - \left[e^{-AW} - (1 - e^{-AW})\frac{\beta}{A}\right]\frac{J_t}{\epsilon} - \frac{e}{S\epsilon}\delta = 0\end{aligned} \tag{2.30}$$

This equation can be numerically solved for $V_m(t)$, and the thermionic and tunneling components examined as a function of the temperature and applied bias. A more detailed analysis using these equations and their implications in device design are presented in Chapter 5.

2.7 Conclusion

This chapter introduced the self-quenching SPAD design by utilizing a TCB energy barrier. A deterministic model is formulated to explain the self quenching and self recovering process, and the design is extended by using thermionic and tunneling models to physically explain the escape mechanism of the carriers from the interface. The self-quenching speed is found to be much faster than the recovery process, and a preliminary examination of the recovery time as a function of barrier strength is examined.

This chapter, in part, is a reprint of material as it appears in the following publications:

- James Cheng, Sifang You, Kai Zhao, and Yuhwa Lo, "Self-quenched InGaAs single-photon detector", Proc. SPIE 7320, 732010 (2009). The dissertation author was the primary investigator and author of this material.

- Sifang You, James Cheng, and Yu-Hwa Lo, "Physics of Single Photon Avalanche Detectors With Built-In Self-Quenching and Self-Recovering Capabilities", IEEE J. Quantum Electronics, v 48, n 7, p 960 (2012). The dissertation author was the secondary author of this material.

Chapter 3

Patterned Zinc Diffusion Design for Improved Efficiencies

This chapter will describe a novel zinc diffusion design which allows for higher avalanche probabilities. This design is independent of the epitaxial design, and can be incorporated into any APD device. Simulation, fabrication, and measurement results of self-quenching SPADs utilizing this zinc diffusion design will be presented. Finally, the limitations of the this design will be discussed.

3.1 Conventional Zinc Diffusion Design

The simplest, fastest, and cheapest method of fabricating a photodiode and defining its photosensitive area is to grow the complete epitaxial structure, including the p-i-n junction, and etch everything but the device area. However, these "mesa-etched" devices have two significant shortcomings: surface states and non-uniform electric field [42–46]. Surface states greatly increase the dark count rate and afterpulsing probability, and the increased number of avalanches further decreases the detection efficiency. Additionally, the increased afterpulsing effects increases the device's recovery time, limiting the device bandwidth. Non-uniform

electric field creates its own problem: the highest avalanche probability is at the surface while most of the photo-generation is at the center, greatly reducing the device's sensitivity.

Thus most advance devices have utilized zinc diffusion to form the p-junction in InP and isolate the high field region from any mesa surfaces. Again, the simplest method (a single diffusion well) introduces another problem: a higher electric field at the junction edges due to the field crowding effect [47]. The conventional answer to this problem is to engineer the diffusion profile to remove the field crowding effect, either through double diffusion wells, guard rings, trench isolations, or any combination thereof [48–51]. An example diagram of such a design is shown in Figure 3.1.

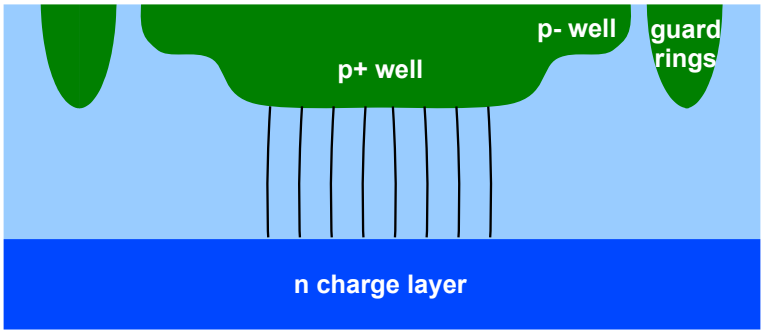

Figure 3.1: Ideal zinc diffusion profile demonstrating a uniform electric field within the center of the photoactive area.

While such a design works in theory, the practical repeatability of the design is questionable. The heavily doped junction must be precisely aligned with the second lightly doped junction, and the spacing to the guard rings very well controlled. Additionally, this design is critically dependent on the actual diffusion profile. If the profile changes a little bit due to process variations, the dimensions and spacings of the lightly doped second junction and the guard rings must be

adjusted.

As a baseline reference for the pattern zinc ring design to be discussed momentarily, we fabricated a device utilizing a guard ring design and measured its efficiency as a focused laser spot position is swept across the device (see Section 3.5 for experimental details). The results in Figure 3.2 demonstrate that it is very difficult to fabricate the conventional design with uniform efficiencies across the photosensitive area. Even with the guard rings, there is significant field enhancement at the edges of the primary diffusion well, resulting in an increase in detection efficiency only in that region.

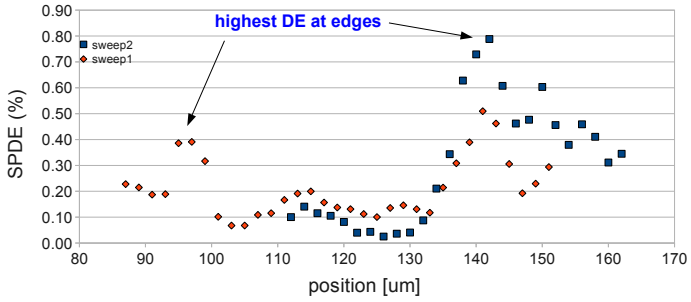

Figure 3.2: Measured SPDE over spatial position of a diffused guard ring device.

3.2 Patterned Zinc Diffusion

Due to the factors aforementioned, this section proposes a design in which, instead of fighting the field crowding effect and trying to minimize it, the effect is purposely utilized to our advantage. To maximize the area with high field crowding effects, multiple thin diffusion lines are created and connected together to form the device anode. Figure 3.3(B) shows the cross section of the simulated electric field contour plot of this design. In the simulation, the device is radially symmetric

about x=0. When compared to the conventional single diffusion well design in (A), which has a very high field crowding effect only at the edge, the Zn-ring device has a higher percentage of high field regions within the photosensitive area of the device.

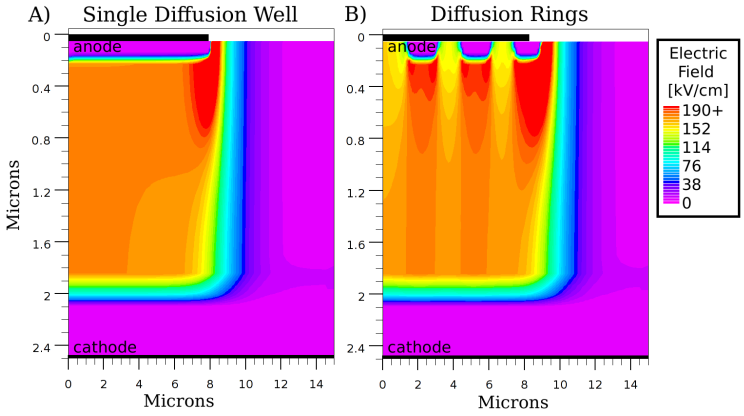

Figure 3.3: Simulated electric field contour plots for a single diffusion well and the diffusion ring design. The black lines indicate the locations of the electrodes.

The advantage of this design is that it is independent of the resulting diffusion profile, since there is only a single diffusion step to the process instead of multiple tightly controlled diffusion steps. All the lines will experience the same edge crowding effect, and thus have the same breakdown voltage. To realize the device in practice, concentric rings of thin zinc diffusion wells are created to avoid any issues with the termination of the zinc diffusion line.

3.3 Epitaxial Design

The next few sections will demonstrate the practical performance of this patterned diffusion design, through fabrication and testing of a self-quenching SPAD structure. The first version of the epitaxial wafer design is shown in Ta-

ble 3.1, where all the quaternary and ternary compounds are latticed matched to InP (see Figure 3.4 for the band alignment of this system). In contrast to the self-quenching designs from [52], where a valence band offset is used to stop avalanche generated holes, a conduction band offset is created here to stop avalanche generated electrons. This moves the TCB region to below the multiplication region, exposing the multiplication layer to the surface and allowing us to perform Zn diffusion. However, as an initial confirmation of whether the conduction band barrier created by this TCB layer is strong enough to quench the avalanche, a conventional p-i-n multiplication layer was grown to quickly fabricate a mesa etched device and test its properties. The band diagram of this design is simulated using Silvaco Atlas, and the results at equilibrium is shown in 3.5.

Table 3.1: Epitaxial design (T) of first generation electron TCB.

Layer	Material	Thickness [μm]	Doping [cm^{-3}]	Comment
0	InP		n-type	Substrate
1	InP	0.5	n=1×10^{17}	Buffer
2	InGaAs	1.5	-	Absorption
3	InAlGaAs	0.3	-	Grading
4	InAlAs	0.8	-	TCB
5	InGaAsP(1.1μm) to InP	0.10	-	Grading
6	InP	0.25	n=1×10^{17}	Field-control
7	InP	0.8	-	Multiplication
8	InP	0.2	p=2×10^{17}	
9	InGaAsP(1.3μm)	0.1	p=1×10^{18}	Contact

After initial tests confirmed this epi structure was suitable for self-quenching and single photon sensitivities, an epi structure suitable for zinc diffusion was

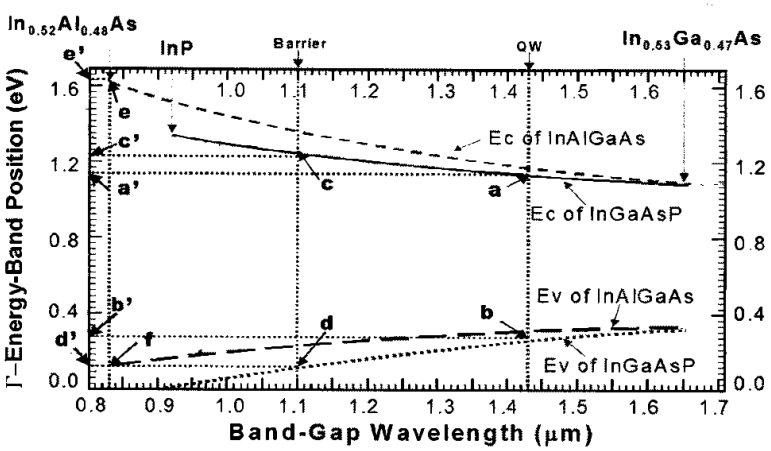

Figure 3.4: Band position of III-V materials latticed matched to InP.

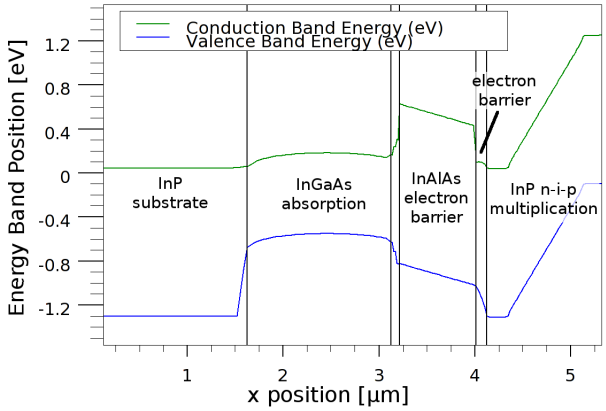

Figure 3.5: Simulated band diagram of the epi design under zero bias.

created and is shown in Table 3.2. Also, from the experimental results of the first epi wafer the quantum efficiency of the device was found to be relatively low. The second wafer improves this in two ways: a step-like transition layer between the absorption and TCB layers instead of a smooth quaternary grading, and a slightly reduced dosage in the field-control n layer.

For ease of creating the patterned p-junction, a heavily doped layer was placed on top of the multiplication region to act as a Zn-source. The idea is

Table 3.2: Epitaxial design (W) for Zn-diffusion drive-in, with modified grading layers. This epi can also be used for conventional Zn diffusion processes by removing layers 10-13.

Layer	Material	Thickness [μm]	Doping [cm^{-3}]	Comment
0	InP		n-type	Substrate
1	InP	0.5	n=1×10^{17}	Buffer
2	InGaAs	1.5	-	Absorption
3	InGaAsP(1.55μm)	0.03	-	Transition
4	InGaAsP(1.3μm)	0.03	-	Transition
5	InGaAsP(1.1μm)	0.03	-	Transition
6	InAlAs	0.8	-	TCB
7	InGaAsP(1.1μm) to InP	0.1	-	Grading
8	InP	0.24	n=1×10^{17}	Field-control
9	InP	1.5	-	Multiplication
10	InGaAsP(1.3μm)	0.03	-	Etch stop
11	InP	0.5	-	Spacer
12	InP	0.6	p=2×10^{18}	Zn-source
13	InGaAsP(1.3μm)	0.05	p=2×10^{18}	Cap

to pattern and etch the source and spacer layers, and then simply drive in the zinc from the source layer to the multiplication layer through a high temperature process. This is to avoid the need for "proper" diffusion through sealed ampoules [53–56], usage of a MOCVD chamber [57–62], or spin-on-dopants [63–67]. Further references on zinc diffusion processing, characterization, and problems can be found in [68–73]. After the zinc drive in, the source layer can be etched away and a metal electrode is deposited.

On the other hand, by removing layers 10-13 on this epi wafer, we ex-

pose the intrinsic multiplication layer and can perform conventional zinc diffusion for reference. The sealed ampoule technique produces junctions which are too deep for SPADs, and spin-on-dopants is an ineffective choice for doping InP with zinc. While more complicated, the diffusion process using a MOCVD chamber was deemed better, and device fabrication using this process will be described in further detail in the next section.

It should also be noted that this design is compatible with electroluminescence measurements. Unlike the design used in [52], the multiplication region is brought close to the surface. Any hot carriers from the avalanche process which recombine will emit light with a wavelength higher than that of the multiplication material (e.g. InP). In this way, the device can be used to up-convert light from near-infrared wavelengths to visible light [74]. If the multiplication region was buried deep within the device, most of the light emission would be reabsorbed before escaping from the device.

3.4 Device Fabrication

The process flow for the patterned zinc-diffused design is outlined in Figure 3.6. As previously mentioned, the epi design shown in Table 3.2 first has a blanket removal of the top Zn-diffusion drive in layers. This is done with selective etching of InGaAsP from InP using 1:1:10 H_2SO_4:H_2O_2:H_2O, selective etching of InP from InGaAsP using 3:1 HCl:H_2O, and H_2SO_4:H_2O_2:H_2O again to remove layer 10 (see [75] for III-V etching). During the etching of InP with HCl, gas bubbles should be observable on the reacting surface. The good etching selectivity between the materials allows us to over etch and ensure that the layer is completely removed.

The final thickness etched was measured using a DekTak 150 Stylus Profilometer on the corner of the sample which was protected by resist. Prior to the etch process, the sample is thoroughly cleaned in solvents (acetone, IPA, then DI water rinse). Any particles remaining before the etch step will become a etch mask and severely affect the diffusion results.

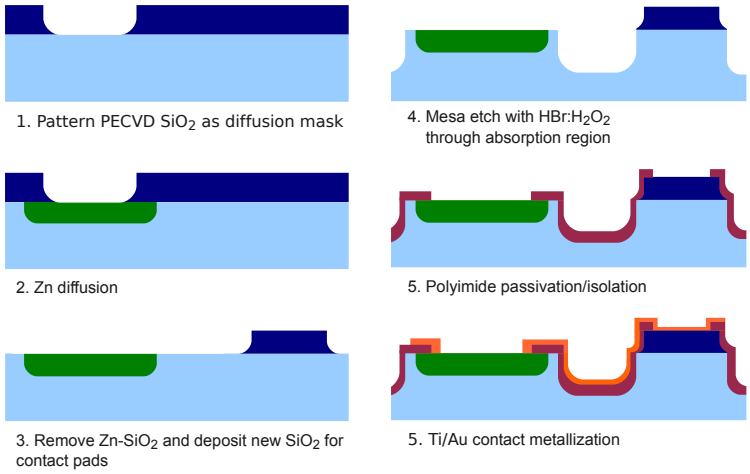

Figure 3.6: Process flow for a patterned Zn-diffused SPAD.

After the wafer is prepped, the first step is to create the zinc diffusion mask. SiO_2 is deposited using a Oxford PECVD at 350 °C using a silane gas mixture. The total thickness deposited was optically measured to be 400 nm with a refractive index of 1.45 at 635 nm. The sample is patterned through conventional photolithography using Shipley S1818 positive photoresist (nominal thickness 1.8 μm), exposed using a Karl Suss MA6 mask aligner, and developed in Microposit MF-319 for 40 seconds. After development, the sample undergoes a short O_2 descum to remove any residual photoresist, and is ready for the oxide etching step. The samples are placed in an Oxford Plasmalab 80 RIE chamber to etch the SiO_2, using 25 sccm Ar and 25 sccm CHF_3 at 35 mTorr, 200 W. Figure 3.7 shows a

photograph of the patterned oxide mask for zinc diffusion rings.

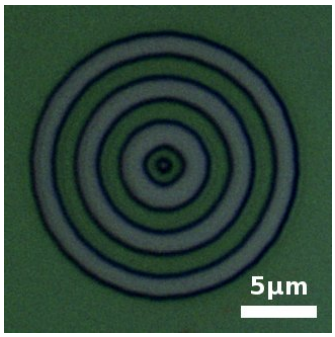

Figure 3.7: Micro-photograph of patterned SiO_2 mask used for Zn diffusion.

Once the diffusion mask is patterned, the sample is placed in a MOCVD chamber and dimethylzinc is diffused to create a p concentration of 2×10^{17} cm^{-3} and a junction depth of approximately 0.7 µm. The electrochemical C-V profile of the doping is shown in Figure 3.8.

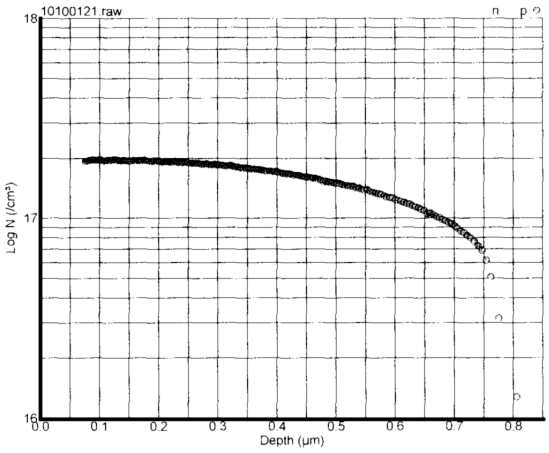

Figure 3.8: Electrochemical C-V doping profile of Zn diffused process.

Post diffusion, the oxide is stripped using BOE (buffered oxide etch, 6:1 NH_4F:HF). This oxide cannot be used as a contact pad insulator since the zinc diffusion process degrades its electrical properties. The surfaces exposed to the

zinc diffusion process undergoes a significant change, and after oxide removal the zinc diffused areas can still be seen (Figure 3.9). This provides enough contrast for aligning the further steps without additional alignment markers. Alternatively, instead of a blanket oxide strip, the Zn-doped oxide can be patterned as alignment markers for later steps.

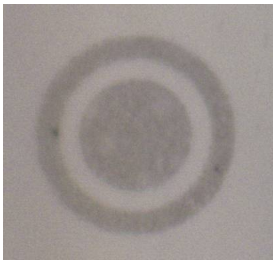

Figure 3.9: Bright field image of the device post Zn-diffusion, with the mask stripped. The device shown here is for a conventional single well with guard ring structure.

To form an insulator for the contact pads, another layer of SiO_2 is deposited, with a total thickness of 300 nm. This thickness should be able to withstand >100 V of applied bias, making it suitable as a contact pad for testing SPADs. The oxide is patterned with a standard S1818 photolithography process, and the oxide etched with BOE. S1818 has poor adhesion to oxide, and will sometimes lift off during the wet etch process. This can be prevented by using a post development bake, or through the use of an adhesion promotion layer such as HMDS. The resist is stripped and the device is patterned for the mesa etch step with S1818.

The etching solution is prepared by mixing 4:1:25 of $HBr:H_2O_2:H_2O$ vigorously for several minutes. The solution is then allowed to sit for another 10-15 minutes, as the etch rate changes rapidly in the first few minutes of etching, and then agitated again right before mixing. Care should also be taken to not let the solution sit for too long (e.g. > 2 hours) during etching of many samples. This so-

lution will isotropically etch InP, InGaAs, InGaAsP, InAlAs, and their compounds, with a small difference in etching rate (see Figure 3.10). During etching, the sample needs to be agitated very vigorously to produce uniform and clean etching. For easier handling during this step, the samples are attached to a carrier glass slide with photoresist.

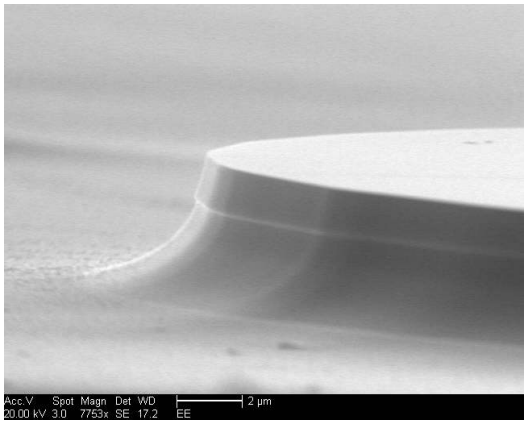

Figure 3.10: SEM image of test structure after HBr mesa etching, illustrating different etch rates between the different epi layers.

The mesa etch height can be confirmed by measuring its step height. For best results, all the layers through the absorption region should be removed to reduce the dark current. After etching, the surface states of the mesa changes rapidly. Polyimide (HD-4104 or PI-2610 from HD Microsystems) is spun on immediately as a passivation layer. Following the standard process, the PI-2610 is cured at 350 °C for around 30 minutes, with a slow temperature ramp up (80 minutes) and ramp down (30 minutes) to reduce the thermal shock on the material. The exact process depends on the type of polyimide used. Note that the PI-2610 used here was first thinned to obtain a desirable thickness, using the ratio 2:1 PI-2610:T9039. After mixing, this diluted PI was left to settle overnight to remove any bubbles.

PI-2610 is not photo-definable, and must be etched away after curing in

order to contact the device area. O_2 RIE etching is recommended to remove cured PI, as there are very few wet processes capable of etching the material. Fluorine based etching can also be used for faster etching, but at a risk of etching the mask and other layers on the device. In the case of a pure O_2 RIE, an SiO_2 hard mask must be used first, as no resist would completely survive the PI O_2 etch step. If resist is used as the only mask, the mesa edges are usually exposed prematurely due to the non-uniformity of the resist spin-coating onto the complex surface topology. Before SiO_2 deposition, the cured PI surface goes through a quick O_2 plasma etch to roughen the surface and promote adhesion of the deposited thin film. A thin film of 100-120 nm of SiO_2 is sputtered onto the PI using an AJA Orion RF sputterer (300 W, 10 sccm Ar, 10 mTorr, DC 200 V). Sputtering was preferred over PECVD due to deposition at room temperature. After SiO_2 deposition, the PI opening pattern is created in NR9-3000PY (negative resist, nominal thickness 3 μm), SiO_2 etched with a Ar/CHF_3 plasma, and PI etched with a O_2 plasma (40 sccm O_2, 100 mTorr, 100 W RIE). As the etch process is long, and it is easy to burn the resist, the actual etch was done at 50% duty cycle with the RIE power off, in increments of 1 minute per half-cycle. The sample is then ready for metalization.

For device structures consisting of only large mesas, photo-definable polyimide HD-4104 can be used instead for a simpler process. The resolution limit of this negative tone PI is >10μm; therefore smaller openings must be created with PI-2610 using the process mentioned above. After spin-coating and exposure, the PI is developed in PA-401D and rinsed in IPA. It is important to not rinse the sample in water immediately after developing or residual scum will form on the sample. After developing, the samples are then cured (30 minutes hold at an intermediate 200 °C, 60 minutes at 375 °C, with 10 °C/min ramp up and down)

and ready for metalization.

Metalization of the samples involve patterning with NR9-1500PY (negative resist, nominal thickness 1.5 μm), developing in RD6 for 6 seconds, and an O_2 descum. The samples are placed into a Temescal BJD 1800 electron beam evaporator, and 20 nm/200 nm of Ti/Au deposited at a rate of approximately 1 Å/s. Titanium serves as the adhesion layer for gold, which is needed for the for the gold wirebonds. After deposition, the metal can be lifted off by stripping the photoresist in acetone, and rinsed in IPA and DI water. The backside contact to the device is formed using indium solder, taking care to scratch the backside to remove any native oxide and wet the surface nicely with indium. A photograph of the final die with multiple devices is seen in Figure 3.11.

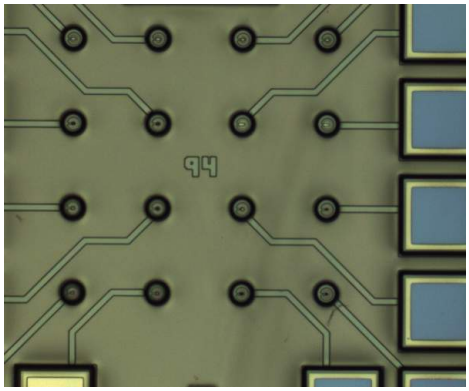

Figure 3.11: Micro-photograph of device after metalization step.

Devices are tested initially through probe measurements for I-V characteristics. Suitable devices with low dark current and good photoresponse are then chosen for wirebonding. The chip is cleaved into smaller pieces (<.220in x .375in) to fit into ceramic dual-in-line packages. The die backside is attached to the package using silver paste (Pelco conductive silver 187) and baked briefly to drive out any of the solvents. Wirebonding is done with a West Bond ball bonder to attach

gold wires from the on-die contact pads to the DIP.

3.5 Measurement Setup

Time correlated single photon counting (TCSPC) is used to measure the SPDE and DCR of the device at various operating bias and temperatures. First, we will need to build the optical and electrical setup to measure these quantities. A brief overview of the system is shown in Figure 3.12.

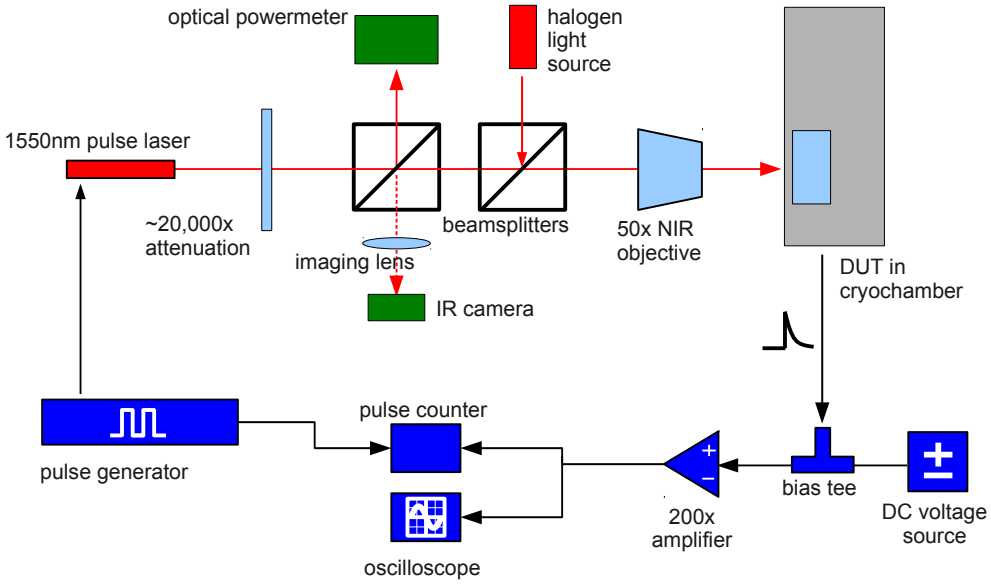

Figure 3.12: Overview schematic of the measurement setup.

The sample is placed in a cryochamber (Janis VPF-100) to provide cooling from liquid nitrogen (77 K) or heating up to 400 K. A LakeShore 325 temperature controller is used to control the heating resistor and read the two silicon thermal diodes (DT-470). One diode is placed near the heating element to help the PID feedback controller, and another closer to the package to gauge the sample

temperature more accurately.

3.5.1 Optical Path

In order to easily control the position of a tightly focused spot onto the device, a compact imaging system was built which can be mounted on top of a three-axis micrometer stage. To create this system, a cage optics framework was used to hold the light sources, beamsplitters, imaging lenses, and camera together. A 50X Mitutoyo M Plan Apo NIR Infinity-Corrected Objective (NA=0.42, working distance=17.0 mm) is used to focus the collimated light source to the device residing in the cryochamber. The long working distance objective was necessary to enable alignment of the spot through the chamber window while maintaining a small spot size. The 1550nm laser source is a Id Quantique id300 Fabry-Perot laser with 300 ps wide pulses. The output of the laser is fiber coupled and connected to a fiber mounted graded-index (GRIN) lens, which provides a collimated output.

As the objective is infinity corrected, the imaging optics becomes greatly simplified. A single plano-convex lens was placed 11 cm away from the objective, and a NIR camera (Spiricon phosphor coated CCD focal plane array) placed 13 cm away from the lens. A 50/50 NIR AR-coated non-polarizing cube beamsplitter was placed in between the objective, laser source, and imaging lens. To illuminate the sample, a 10/90 beamsplitter was placed in between, with a fiber bundled halogen light source as one of its input. After visually aligning the laser to the device, neutral density filters are placed after the GRIN lens output to reduce the number of photons per pulse. The ND filters are mounted to a separate stage as to not disturb the optical alignment of the laser. An optical power meter can also read the power from one of the beamsplitters and confirm the quantum efficiency of the

SPAD.

3.5.2 Electrical Path

The electrical signal from the device is connected to a bias-tee (Picosecond 5530A), where the bias-tee's inductor port (i.e. DC bias) is connected to a voltage supply (Keithley 2400 sourcemeter) and the capacitor port (i.e. AC signal) is connected to a cascaded amplifier chain (Mini-circuits ZFL-1000LN and Miteq AM-1300). The total amplification provided, after a 3 dB attenuator at the end to reduce reflections, is 200x. The output can be observed either through an oscilloscope or photon counter/multiscaler (Becker&Hickl MSA-1000). An Agilent 81110A pattern generator is used to trigger both the laser and the counter for the TCSPC measurements.

3.5.3 Time Correlated Single Photon Counting

By using the same periodic signal to trigger both the pulsed laser and the counter acquisition start time, any signals arriving at the counter due to the photo-response of the SPAD are correlated to each other and the trigger. Dark counts, which are random in nature, are uncorrelated to each and produces a constant background level. By repeating this measurement many times, a counting histogram like Figure 3.13 can be produced.

From this data, the SPDE and DCR can be extracted. The DCR is the average count per collection bin, divided by the length of the bin (typically 1 ns) and the number of measurement sweeps N. The detection efficiency is simply the total number of counts in the correlated peak, minus the background DCR level, and divided by the number sweeps N. The SPDE is the detection efficiency divided

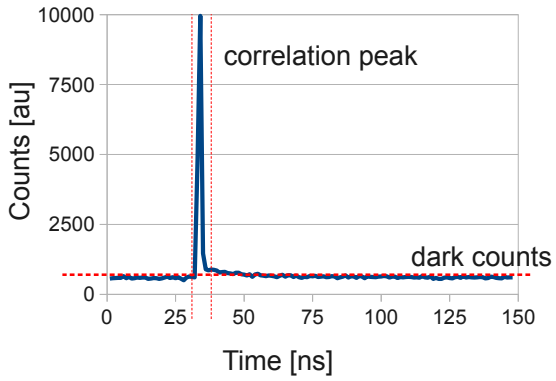

Figure 3.13: Counting histogram for time correlated single photon counting.

by the number of photons per pulse, given a Poisson distribution of photons and $\ll 1$ photon/pulse (see Section 1.3.1).

3.6 Experimental Results and Analysis

The I-V curves for a self-quenching SPAD is measured at various temperatures (Figure 3.14) and demonstrates the typical SPAD behaviors (e.g. a suitable punchthrough voltage compared to the breakdown voltage, and decreasing temperature reducing the breakdown voltage). Atypical to normal SPADs, the breakdown current is not as sharp due to the self-quenching effect. The slope of the breakdown current also flattens out as the temperature is decreased, indicating stronger self-quenching at lower temperatures.

The device is connected according to Section 3.5.2 to observe the avalanche pulses through the oscilloscope, operating the device in free-running Geiger mode without any external quenching circuits. Shown in Figure 3.15 is the image of a dark pulse from the device at 160 K and a DC bias of 42.2 V. The peak height is strong at hundreds of millivolts (after 200x amplification), yet is self-quenching

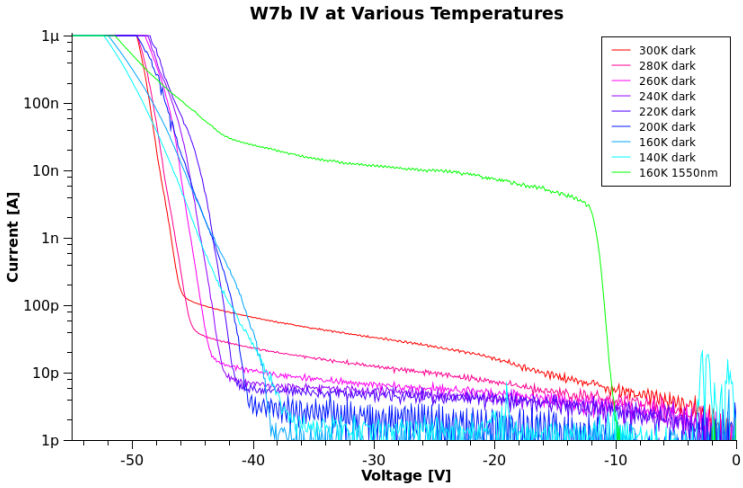

Figure 3.14: I-V plots of a Zn-ring device at low temperatures.

since the pulse terminates by itself. The pulse width is also quite narrow, on the order of several nanoseconds (note the recording instrument has a bandwidth of only 500 MHz).

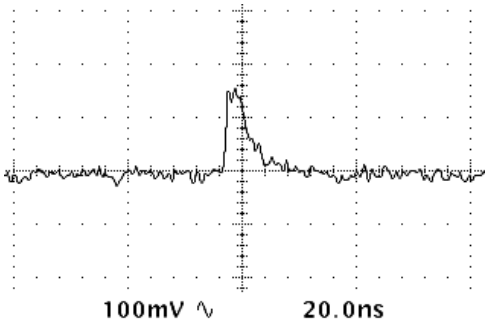

Figure 3.15: Oscilloscope trace of the self-quenched avalanche pulse, at 160 K and 42.2 V reverse bias.

Devices with various spacings and widths of the zinc-rings were fabricated to test their effect on the device performance, with the I-Vs compared in Figure 3.16. The breakdown voltages depend on the different spacings and widths of the zinc-rings, and are actually a good indicator of how well the SPAD can detect

single photons, as will be seen shortly.

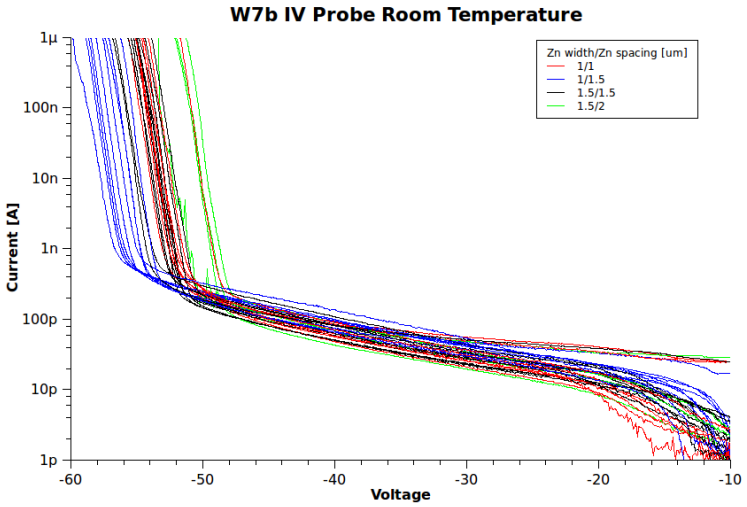

Figure 3.16: I-V plots of Zn-ring devices of different geometries.

SPDE is measured using the TCSPC method described previously, with the counter detection threshold typically around 40-50 mV. This corresponds to roughly 5 µA of current from the SPAD. Lower thresholds are possible if the electrical system is well isolated from any noise, while higher thresholds would reduce the number of detected events.

The SPDE vs DCR of the Zn-ring SPADs of various geometries and temperatures is shown in Figure 3.17. A regular zinc diffused single junction well is also included for comparison. Not surprisingly, the SPDE is very low due to the high field crowding effects appearing only at the edge of the well. The 1.5 µm wide devices show better performance, but the small improvement means that the width of the thin diffusion line is still too thick, and thus the fill factor of high field regions is reduced. The 1.0 µm spaced devices also show better results, but do not have the best performance due to the rings being too close to each other for the given diffusion depth, and producing a guarding effect and reducing the electric

field. For this given process window, the 1.0 µm wide/1.5 µm spaced devices show the best performance. At 160 K, these devices demonstrate a 12x improvement in SPDE while decreasing the DCR by more than 10x.

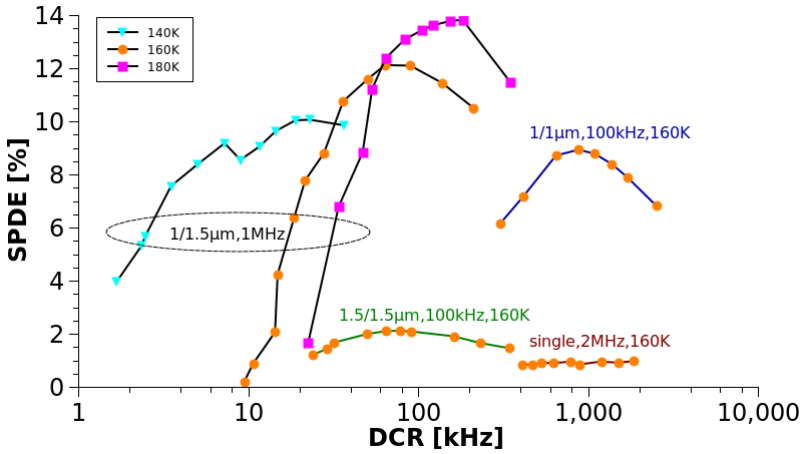

Figure 3.17: SPDE vs DCR of devices with various Zn-ring width/spacings.

It's reasonable that the geometry that has the highest breakdown voltage from Figure 3.16 also has the best SPDE. The highest breakdown voltage indicates that the field uniformity throughout the diode is the best, as there are no particular regions which would break down first and produce the high current. The uniformity of the field improves SPDE as the photosensitive area coincides with the high field regions, giving each photogenerated hole the maximum chance of avalanching.

Also of particular interest is the decrease in SPDE when the DCR is large. This saturation in detection efficiency is limited by the self-recovery nature of the device, and will be discussed in great detail in the next chapter.

3.7 Limitations of Patterned Diffusion Design

In the previous section, due to limitations in the photolithography process, only devices with widths down to 1.0 μm were tested. The general trend at larger widths is that the smaller the diffusion line, the better the performance. To test the limitations of this design, electron beam lithography was used to create smaller features.

PECVD SiO_2 is used as the diffusion mask as before, but the thickness is limited to 200 nm to ensure complete pattern transfer from the resist. Devices with a nominally 100 nm wide opening are patterned using PMMA A4 as the ebeam resist (nominal thickness 200 nm) and a Raith 50 electron beam writer (30 keV, 100 pA, areal dosage: 600 $\mu C/cm^2$, line dosage: 1.8 nC/cm). The exposed resist is developed with 1:3 MIBK:IPA and rinsed in IPA. The SEM images of the resulting patterns are shown in Figure 3.18.

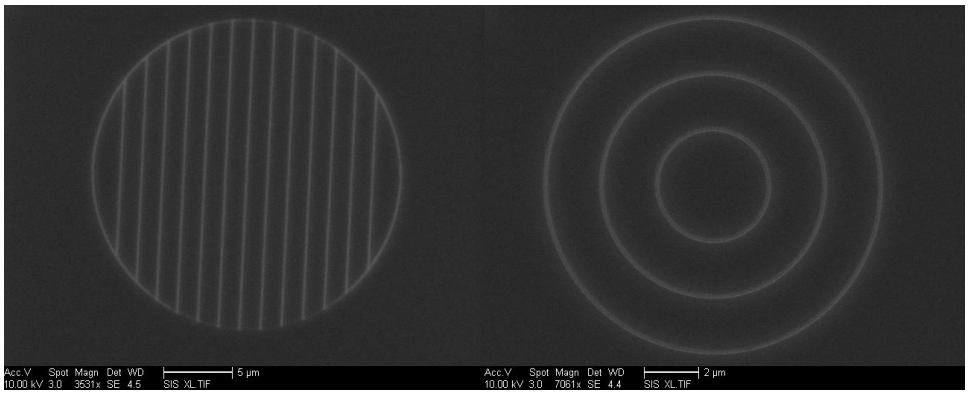

Figure 3.18: SEM of ebeam lithography patterned ring structures, top view.

The SiO_2 is dry etched using a Ar/CHF_3 RIE plasma, which is mostly anisotropic and does not etch the InP substrate. The remaining PMMA is removed

using O_2 plasma. Figure 3.19 shows a cross sectional view of the etched oxide pattern to be used as the zinc diffusion mask. The thinnest portion of the oxide mask opening is 55 nm. The rest of the fabrication process is the same as before.

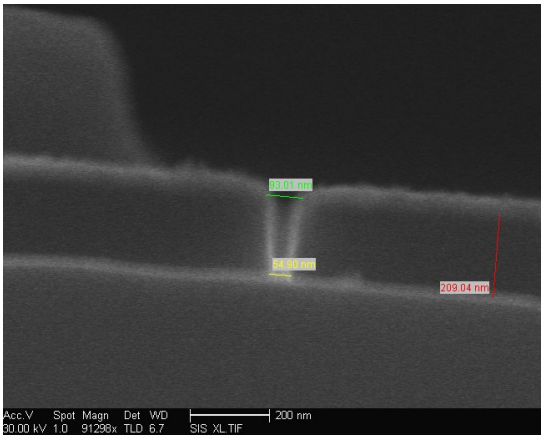

Figure 3.19: Cross-section SEM of etched ebeam patterned SiO_2 .

Testing the finished devices shows that many (>75%) of the devices have very high current. Of the remaining devices, the measured SPDE is typically very low. For example, at 160 K, the SPDE is 0.7% at a DCR of 40 kHz in free running mode. The performance limitation is attributed to the insufficient flux of the zinc source entering the small openings. Thus while a planar diffusion sample has 2×10^{17} cm^{-3} doping to a depth of 0.7 μm, the actual dosage entering the small opening is much less. Lack of doping means many devices do not even form a p-i-n junction (thus the high current). Obtaining the correct dosage for the smaller feature sizes will require extensive additional testing of the MOCVD diffusion process.

3.8 Conclusion

This chapter introduced a new single step design pattern for zinc diffusion, which can greatly improve the efficiency of any APD device. Devices were fabricated and measured to test the performance enhancement due to the patterned design. The improvement due to this design is summarized in Table 3.3, with the mesa-etched devices measured in gated-mode of this epi wafer characterized in [76], and better Zn-ring performance described in Figure 4.1.

Table 3.3: Zn-Ring performance comparison.

Device	SPDE	DCR	Operational Mode	Reference
Mesa	1%	5 MHz	Gated Mode	[76]
Zn-ring	20%	8 kHz	Free Running	Fig. 4.1

Compared to conventional methods of producing a uniform electric field, this method is simpler and more robust. The design performance is limited by the current status of the diffusion process technology, and even better performing devices should be possible given a better diffusion technique.

Initial characterization of the electron TCB self-quenching design was also provided. The self-quenching effect can be observed through the avalanche pulse waveforms, the DC I-V measurements, and SPDE saturation measurements. This data suggests the TCB is too strong in this preliminary design.

This chapter, in part, is a reprint of material as it appears in the following publications:

- James Cheng, Sifang You, Samia Rahman, and Yu-Hwa Lo, “Self-quenching

InGaAs/InP single photon avalanche detector utilizing zinc diffusion rings", Optics Express, v 19, p 15149 (2011).

- James Cheng, Yu-Hwa Lo, "Patterned zinc-diffused structures for improved avalanche probabilities in InGaAs/InP single photon detectors", IEEE Photonics Conference 2011.

The dissertation author was the primary investigator and author of this material.

Chapter 4

Recovery Time and Afterpulsing

As the total count rate increases for a SPAD operating in free-running mode, the total dead time the device experiences increases. Thus the likelihood of a photon arriving in the dead time increases, decreasing the overall detection efficiency of the device - the device is saturated. In this chapter we examine the effects of saturation on a self-quenching SPAD, its dependence on the device recovery time, reducing the recovery time, and the implications of high dark count rates and afterpulsing.

4.1 SQ-SPAD Device Saturation

At the end of the last chapter, in Figure 3.17, we see that there is a maximum SPDE, which will start decreasing as the DCR increases further. The total avalanche rate is faster than the recovery time in this case, limiting the SPDE. This effect can be easily understood from the simulation results previously shown in Figure 2.5, where the device response is shown to be limited if the photon arrival time is too close to one another. The voltage in the multiplication region must be given enough time after the first avalanche event to recover to a high enough value in order to produce single photon response for the second photon.

We can deduce that the device's recovery time is inversely related to the total count rate at which the device starts to saturate. For example, peaking at 100 kHz DCR corresponds to a recovery time of approximately 10 μs. Then, if the laser illumination rate on these devices was 1 MHz (e.g. faster than the recovery time), the SPDE would not be optimal. Thus, to measure the true SPDE of these devices, both the DCR and the laser illumination rate must be minimized to ensure the response is not saturated. In Figure 4.1 we vary the laser illumination rate to observe the saturation effect, and see the SPDE decreasing when the illumination rate is faster than 100 kHz. At slower illumination rates, when the avalanche rate does not saturate the device, the SPDE of this device (1.0 μm width / 1.5 μm spacing Zn-ring) reaches 20%.

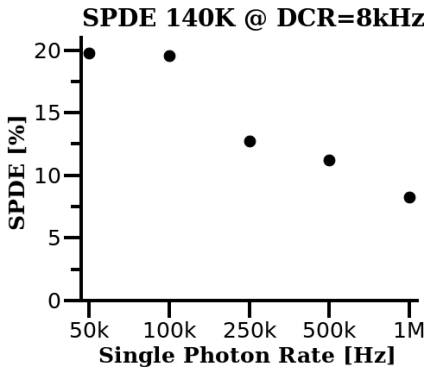

Figure 4.1: Measured SPDE at various laser rates, demonstrating saturation at higher photon rates.

4.2 Recovery Time Measurement

To get a better understanding of the recovery time, we can measure it experimentally. By also triggering the counter with avalanche events (e.g. dark counts), there should be some dead time in the counting histogram after the initial

avalanche event in which the probability of a second avalanche is reduced due to the device not being fully recovered. To create this histogram, the SPAD signal is connected to both the trigger and signal port of the counter, instead of the usual time-correlated setup. The time constant associated with the exponential recovery in the avalanche probability is denoted as the recovery time (Figure 4.2). The counting histogram is recorded at different applied bias voltages, with the inset of this figure showing the time constant as a function of bias voltage.

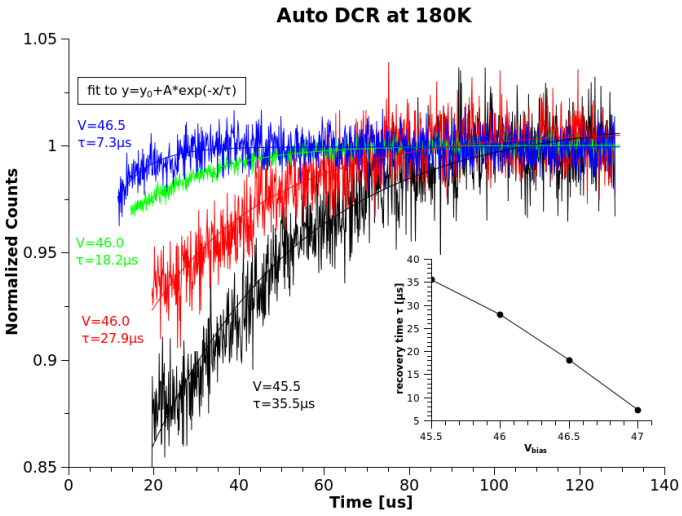

Figure 4.2: Auto-correlated counting histogram to extract the recovery time.

The complete dependence of the recovery time on bias level and temperature is shown in Figure 4.3. As expected, the recovery time reduces when increasing the bias voltage (escape via tunneling), and also reduces with increasing temperature (escape via thermionic emission). While the recovery time is adjustable over three orders of magnitude depending on operating conditions, under typical experimental conditions (e.g. 200 K, 1-2 V overbias) the recovery time is 10's of microseconds. This verifies the earlier SPDE measurement of saturation at greater than 100 kHz,

and the next section describes how to decrease the recovery time.

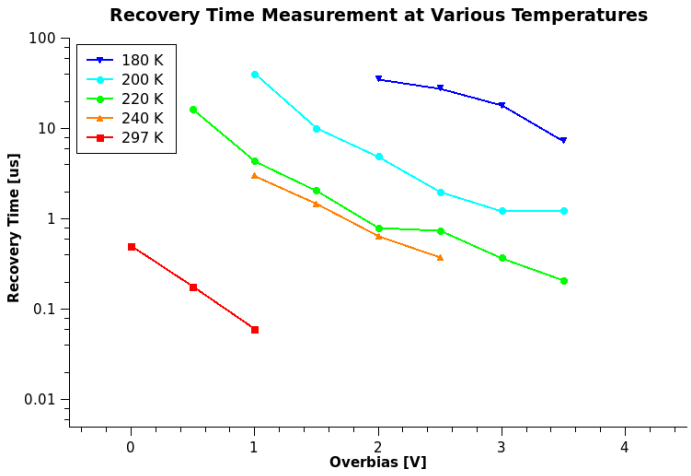

Figure 4.3: Measured recovery time at different bias voltages and temperatures.

4.3 Reducing the Recovery Time

For a more useful device, we would like to reduce the recovery time in order to obtain a higher operating count rate. We can do so by reducing the barrier height of the TCB heterojunction, using Figure 3.4 for reference. The new epitaxial design is reported in Table 4.1, and a visualization of the new design compared to the old design is shown in Figure 4.4. The barrier in this case is between InP and 1.4 µm InGaAsP, yielding an estimated conduction band offset ΔE_c of 150 meV, compared to the 350 meV of the original design.

Note that theoretically, the design in Table 3.2 could have been used as the baseline, and replacing Layer 6 InAlAs with the correct InAlGaAs composition to obtain the desired ΔE_c to InP. In practice, however, it is difficult to grow a clean interface between InAlGaAs and InGaAsP without growing several layers of

Table 4.1: Epitaxial design (X) for reduced electron barrier.

Layer	Material	Thickness [μm]	Doping [cm^{-3}]	Comment
0	InP		n-type	Substrate
1	InP	0.5	n=1×10^{17}	Buffer
2	InGaAs	1.5	-	Absorption
3	InGaAsP(1.55μm)	0.03	-	Transition
4	InGaAsP(1.3μm)	0.03	-	
5	InGaAsP(1.1μm)	0.03	-	
6	InGaAsP(1.0μm)	0.03	-	
7	InP	0.8	-	TCB
8	InGaAsP(1.4μm)	0.04	-	Transition
9	InGaAsP(1.3μm)	0.04	-	
10	InGaAsP(1.2μm)	0.04	-	
11	InGaAsP(1.1μm)	0.04	-	
12	InGaAsP(1.0μm)	0.04	-	
13	InP	0.24	n=1×10^{17}	Field Control
14	InP	1.5	-	Multiplication
15	InGaAsP(1.3μm)	0.03	-	Cap

high bandgap material, greatly disrupting the ΔE_c we would like to form. Thus we chose to use only InP and InGaAsP. The transition layers 8-12 of this design also need to be thick enough to cool down the hot electrons from the high field multiplication region before they see the TCB barrier.

The devices were fabricated as previously described in Section 3.4, and the recovery time measured according to Section 4.2. The recovery time as a function of temperature and bias are shown in Figure 4.5, overlayed on top of the recovery

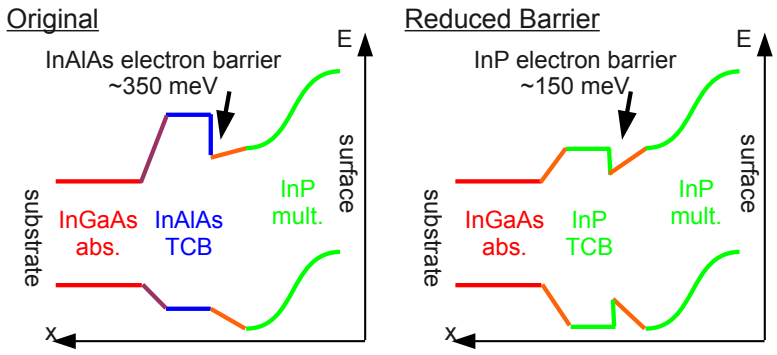

Figure 4.4: Band alignment of the reduced ΔE_c epi compared to the original epitaxial design. The TCB barrier has been reduced from 350 meV to 150 meV.

time measurements of the previous epitaxial design. The design shows a two orders of magnitude improvement in recovery time, with recovery time ranging from 10 ns to 1 µs, depending on the bias voltage.

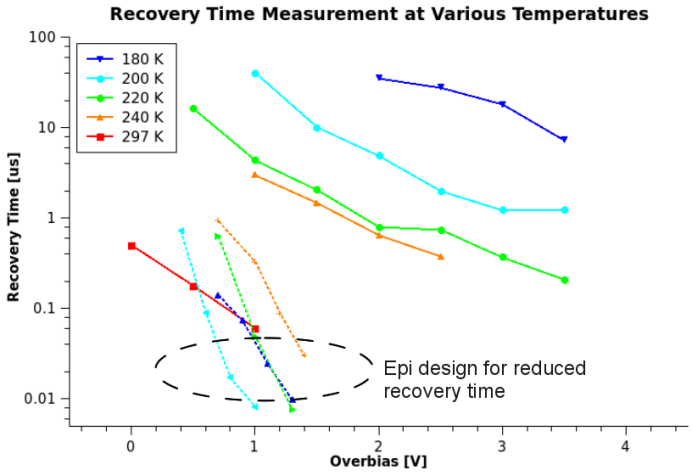

Figure 4.5: Comparison of measured recovery time for reduced recovery time design vs the original epitaxial design.

From this data, we also see that the bias dependence of the recovery time is much stronger in the case of the small barrier height design compared to that of the high barrier height. Additionally, there is no notable correlation between the temperature and the recovery time. We can conclude that at such a small ΔE_c,

the tunneling escape mechanism is dominating over thermionic emission.

4.4 Reduced Recovery Time and Afterpulsing

There are two practical limits to how small of a heterojunction barrier, and thus recovery time, can be used in a self-quenching SPAD. The first is due to the physical link between the quenching mechanism and the recovery mechanism; in order for the recovery time to be as fast as possible, the barrier must be weak, and thus the self-quenching effect is reduced. The second limit is due to afterpulsing; if the device is immediately ready after the first avalanche detection, the probability of a trapped carrier in the multiplication region triggering an afterpulse is exponentially increased.

Both of these effects increase the total dark count rate, and thus the total counting rate of the SPAD. As mentioned previously, a high avalanche rate can saturate the device and decrease the SPDE. Using a state of the art double zinc diffusion process, courtesy of Princeton Lightwave Inc., SPADs with the reduced barrier epitaxial design were fabricated and tested. The I-V plot of this device at room and low temperature is shown in Figure 4.6. Due to the smaller barrier height of this design, the self-quenching effect is not as noticeable in the I-V plot as it is for the original epitaxial design (refer to Figure 3.14).

From oscilloscope observations and the counting histogram of this device, we can see a very high afterpulsing rate, due to the small barrier height. The normalized counting histograms at various laser illumination rates is shown in Figure 4.7. On this scale, the primary dark count level (e.g. measured with no light input) is less than 10^{-5}. As the frequency of the laser input increases, the total

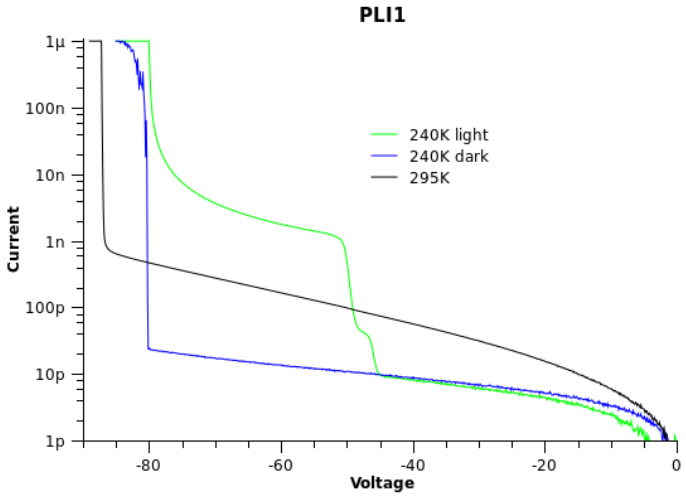

Figure 4.6: Current-Voltage plot of epitaxial wafer with low barrier height design.

number of uncorrelated dark counts increases due to afterpulsing [77]. There is less time between measurements for the afterpulsing rate to decay, thus increasing the background level. This effect is pronounced when changing from 10 µs to 1 µs time intervals, but the fact that there is still a measurable background level at 1 ms time intervals signifies the dominating effect afterpulsing has on the device. Ideally the counts before the main correlation peak should be zero, given enough time between measurements for the afterpulsing probability to become zero.

The SPDE, DCR, and afterpulsing rate of the device is shown in Figure 4.8. As the afterpulsing probability is a exponential function of time, here we characterize the afterpulsing rate as the count rate 250 ns after the initial photon input. To compensate for afterpulsing from the photo responses, the DCR is measured with zero illumination. From the figure we see that while the primary DCR is relatively low, the high afterpulsing rate limits the bias voltage we can apply on the device. This in turn limits the SPDE achievable with the device to a low value of less than 1%.

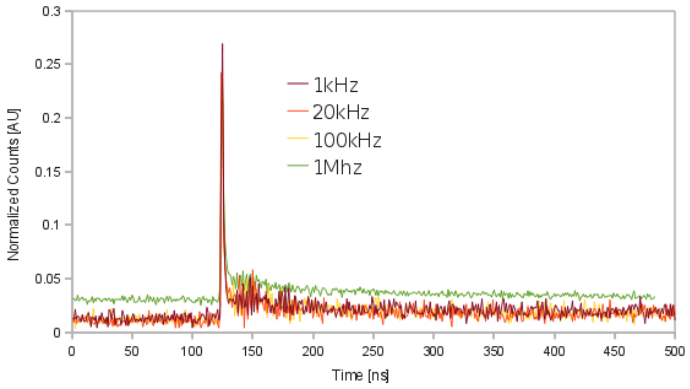

Figure 4.7: Counting histogram showing relative afterpulsing rate. The DCR is $< 10^{-5}$ on this relative scale. Ideally the counts before the main peak should be zero.

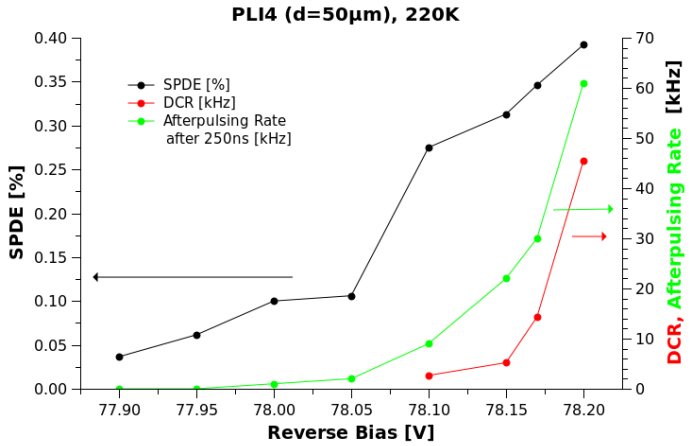

Figure 4.8: SPDE (left axis), DCR (right) and afterpulsing rate (right) as a function of applied bias. The afterpulsing rate is measured 250 ns after the primary avalanche event.

It becomes clear that afterpulsing is the dominating limitation to the performance of these devices. Unlike conventional SPADs which utilize an external quenching circuit to set the device recovery time, the recovery time of these free-running SPADs are characterized by their epitaxial design and cannot be independently adjusted during operation. The next chapter will include this important effect to create a model that is useful for designing the optimal self-quenching

SPAD.

4.5 Conclusion

This chapter examined the self-recovery nature of the SPAD in greater detail, providing experimental evidence of the dependence of the recovery time on the epitaxial design, the operating temperature, and the applied bias. It is shown that the recovery time can be easily adjustable from tens of nanoseconds to tens of microseconds. For long recovery times, the operating count rate limits the usefulness of the device. For short recovery times, afterpulsing becomes dominating and also decreases the device performance.

This chapter, in part, is a reprint of material as it appears in the following publications:

- James Cheng, Sifang You, Samia Rahman, and Yu-Hwa Lo, "Self-quenching InGaAs/InP single photon avalanche detector utilizing zinc diffusion rings", Optics Express, v 19, p 15149 (2011).

The dissertation author was the primary investigator and author of this material.

Chapter 5

A SQ-SPAD Monte Carlo Model for Device Optimization

In this chapter we formulate a Monte Carlo model of the self-quenching SPAD. By including variability in the avalanche buildup between multiple detection events, the timing jitter, excess noise, gain variations, and detection efficiency can be determined. Knowing these additional characteristics allows us to design the self-quenching structure to optimize the device for various single photon applications. In particular, the model will calculate the effective bitrate and noise equivalent power (NEP) of the device.

5.1 Motivation

The previous model we have developed demonstrated the self-quenching and self-recovering nature of our SPAD. By using deterministic rate equations, the current response to a single carrier was shown. The model included experimental input parameters such as applied bias and operating temperature, device design parameters such as barrier height and epi-layer thicknesses, along with material parameters such as impact ionization rates and effective masses. Except for the

imprecise nature intrinsic to impact ionization modeling, no special fitting parameters have been made in this model.

Using this model we extracted the recovery time of the device by creating two input events at a fixed time interval. We determined the recovery time by varying the time interval until the device response to a second event reaches a certain value. This model provides valuable insight on how the recovery time is effected by the device design and operating conditions.

From a timing perspective, a faster recovery time means a device operating at a faster data rate. One can easily and effectively reduce the barrier size to reduce the recovery time, achieving the best recovery time with no barrier. On the other hand, without a barrier, then the device also loses its self-quenching abilities. Practically, afterpulsing will also provide the lower limit of the self-recovery time. Afterpulsing is primarily determined by two factors: material defects, and avalanche gain. The former depends on the quality of the traps during material growth, and cannot be easily reduced. The latter is directly linked to the number of carriers that are trapped during avalanche and can later be released creating an afterpulsing event.

Thus we would like to limit the total gain of the device, while providing minimal noise in terms of pulse height and timing jitter, along with optimal self-recovery time. However, for a self-quenching SPAD, there are often trade-offs when trying to optimize for one value.

Table 5.1: General trends in a self-quenching SPAD.

	Gain	Pulse Width	Pulse Height	Excess Noise	Jitter	DE	Recovery Time
↑ Applied Bias	↑	↑↓	↑	↓	↓	↑	↓
↑ Temperature	↑↓	↑↓	↑↓	↑	↑↓	↑↓	↓
↑ Mult. Width	↑	↑	↑	↑	↑	↑	↑↓
↑ TCB Strength	↓	↓	↓	↓	↓	↑↓	↑

5.2 Device Optimization for Bitrate

Previous works have clearly derived the statistics of timing jitter of a conventional SPAD [78] along with the recovery mechanism of self-quenching SPAD [41]. The goal of this work is to build an explicit relation between these parameters and the maximum bitrate the self-quenching SPAD is capable of operating at.

First, let us qualitatively examine the dependence of gain, pulse width, pulse height, excess noise, timing jitter, detection efficiency, and recovery time on various operating and design parameters. The general trends are summarized in Table 5.1. Not directly mentioned is afterpulsing, which is strongly correlated with the gain, pulse width, and recovery time. The key dependencies are listed in the following:

- Increasing bias increases the gain, which increases the afterpulsing effect.
- Increasing bias reduces timing jitter.
- Increasing bias increases the TCB field, decreasing the recovery time.
- Increasing temperature increases thermally generated dark counts.
- Increasing temperature reduces the recovery time.
- Increasing multiplication width increases timing jitter, while reducing dark carriers from trap assisted tunneling.

– Increasing TCB strength reduces gain, while increasing the recovery time.

As the timing jitter of the SPAD is usually much shorter than the recovery time, we can increase the device bitrate using a simple pulse position modulation (PPM) scheme, where we encode M bits in 2^M time slots during a period T. The self recovery time will limit the period T, while the timing jitter σ will limit how fine the time slots can be ($2^M \approx T/\sigma$). Then the maximum bitrate, in the absence of any false counts, will be M/T bits/second, or $\log_2(T/\sigma)/T$.

To quantify the effects of false counts, we will use Reed Solomon error correction codes and calculate the resulting effective bitrate. With Reed Solomon codes, R data symbols along with $2t$ parity symbols are transmitted ($n = R + 2t$), and the system is able to correct for t symbols of error. The t errors in this case is the total dark count rate. The symbol size m will determine the maximum code word length n with the relationship $n = 2^m - 1$. For example, let's take the symbol size $m = M$, although there may be more efficient schemes. In terms of single photon counting, the total count rate n will consist of t false counts, t overhead to correct for errors, and R signal, where $t + R$ is the photon count rate DE/T. That is to say, the maximum effective bitrate R is

$$R = \left(\frac{DE}{T} - DCR\right) \log_2 \left(\frac{T}{\sigma}\right) \tag{5.1}$$

It is clear from this equation that we need to minimize the recovery time T, the timing jitter σ, and the DCR. But first, we need to examine the dependence of the DCR on the applied bias, temperature, and device design.

5.3 Dark Count Rate

The total dark count rate of a well-design SPAD is mainly due to three parts: thermally generated carriers from the absorption region, generation by trap-assisted tunneling in the multiplication region, and afterpulsing. Thermally generated carriers in the high bandgap multiplication region is negligible compared to that from the low bandgap absorption region, and tunneling in the absorption region can be limited by proper design of the field charge layer in between the absorption and multiplication layers.

Thermally generated carriers can be found by calculating from the primary dark current of the measured devices. For example, 30 pA of primary current at 300 K corresponds to roughly 180 MHz DCR. At lower temperatures where the primary dark current is below the measurement equipment sensitivity, equation (5.2) can be used by fitting the measured results at different temperatures [33, 79]. For InGaAs, the activation energy of the generation-recombination center is around 0.39 eV, which is close to half the bandgap [79]. For example, based off the 30 pA measured at 300 K, the estimated primary dark current at 200 K for this device is 7 fA, corresponding to a DCR of 46 kHz.

$$I_{G-R} \propto T^2 \exp(-E_a/kT) \tag{5.2}$$

Tunneling in the high field multiplication region is also of concern. Trap assisted tunneling is found to be dominant over direct tunneling [34]. The trap assisted tunneling (TAT) rate is given by

$$J_{TAT} = \frac{AF(x)^2 N_T \exp\left(\frac{-(B_1 E_{B1}^{3/2} + B_2 E_{B2}^{3/2})}{F(x)}\right)}{N_v \exp\left(\frac{-B_1 E_{B1}^{3/2}}{F(x)}\right) + N_c \exp\left(\frac{-B_2 E_{B2}^{3/2}}{F(x)}\right)} \tag{5.3}$$

$$A = q^3(2m_r/(qE_g))^{1/2}/(4\pi^3\hbar^2) \tag{5.4}$$

$$B_1 = \pi(m_{lh}/2)^{1/2}/(2q\hbar) \tag{5.5}$$

$$B_2 = \pi(m_c/2)^{1/2}/(2q\hbar) \tag{5.6}$$

where $F(x)$ is the electric field, m_c is the conduction band effective mass, m_{lh} is the light hole effective mass, $m_r = 2(m_c m_{lh})/(m_c + m_{lh})$ is the reduced effective mass, $E_{B1} = aE_g$ is the tunneling barrier height from valence band to trap state, and $E_{B2} = (1-a)E_g$ is the barrier height from trap to conduction band. N_T is the number of defects per volume, and along with a, the position of the trap state, are two fitting parameters. They are found to be about $a = 0.75$ and $N_T \approx 2 \times 10^{14}$ cm^{-3} for InP. For example, a 100 μm^2 area device with a multiplication thickness of 1.0 μm, operating at 4 V overbias will produced 100 kHz DCR due to TAT.

Lastly, the afterpulsing rate can be directly fitted to a counting histogram of a free-running self-quenching SPAD. The afterpulsing probability at a given time is represented by a characteristic de-trapping time τ_d [80],

$$AP(t) = C_0 \exp(-t/\tau_d) \tag{5.7}$$

where τ_d is a strong function of temperature, and the coefficient C_0 depends on material property and represents the defect density. For a single trap model, τ_d is $\propto \exp(E_a/kT)$, and is found to be around around 65 ns [80]. E_a is around 0.11 eV for InP at 290 K. For a self-quenching SPAD, there is some dead time T during which the avalanche probability is reduced. The expression for the afterpulsing probability in this case can be expressed as:

$$AP(t) = C_1 \exp(-t/\tau_d)(1 - \exp(-t/T)) \tag{5.8}$$

C_1 relates the avalanche probability time constant to the multiplication voltage

recovery time constant, and also normalizes C_0 to the Geiger-mode gain (i.e. total current flow) per avalanche event (more current means more trapped carriers and increased afterpulsing). The afterpulsing probability can be integrated with respect to time in order to find the total afterpulsing probability per avalanche event,

$$P_{AP} = \int_0^\infty AP(t)dt \tag{5.9}$$

With the above probability per event, the total afterpulsing rate is related to the primary count rate (photo plus dark) R_{p+d} as

$$\langle R_{AP} \rangle = R_{p+d} P_{AP} \tag{5.10}$$

5.4 Model Formulation

In order to examine the stochastic properties of the avalanche process, we turn to a simplified Monte Carlo model. In this model, we need to simulate the locations of each electron and hole in the multiplication region. A 1D model is assumed, with holes traveling in the $-x$ direction and electrons in the $+x$ direction. Holes are injected into the multiplication region at $x = W$ to simulate photo-generated holes drifting from the absorption region, and are removed from the simulation when they reach $x = 0$, contributing to the measured current. Impact ionization generated electrons are removed from the simulation when they reach $x = W$ (i.e. TCB region), adding to the trapped interface charge, which in turn reduces the multiplication voltage.

The ionization rates are given by the local field approximation model (see Eqn. (1.1)). Then the probability distribution function of an electron ionizing is

$$\text{PDF}(x) = \alpha \exp(-\alpha(x - x_0 - d_e)) \tag{5.11}$$

where $x - x_0$ is the travel distance of the carrier since being created. The PDF represents the probability of an electron ionizing at exactly x. The cumulative probability of ionization of the electron is then

$$\mathrm{CDF}(x) = 1 - \exp(-\alpha(x - x_0 - d_e)) \tag{5.12}$$

and represents the total ionization probability of the electron which has traveled x distance. The "dead space" of the carrier, d_e, is given by the hard threshold ionization energy divided by the local electric field [14, 35]. From this reference, the threshold energy is given as 2.05 eV and 2.20 eV for electrons and holes, respectively. In the hard threshold model, the ionization probability is identically zero within the dead space, and rises at distances greater than the dead space. Example ionization probabilities at different electric fields are shown in Figure 5.1, using the same width-independent ionization coefficients for InP from [14] (reproduced in Table 5.2).

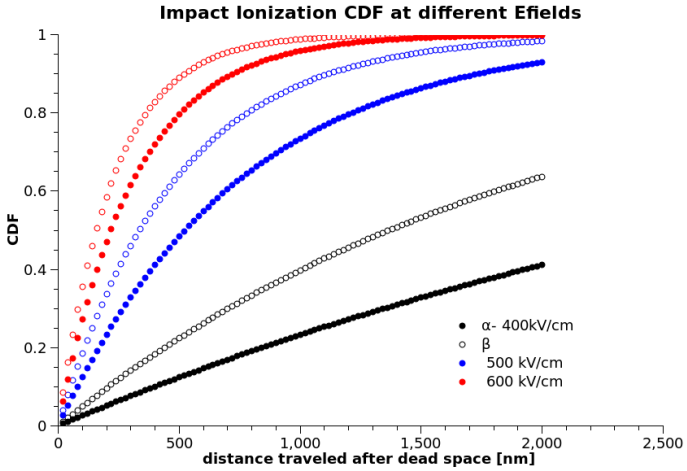

Figure 5.1: Impact ionization probability over distance traveled, at different electric fields.

To simplify matters, all carriers are assumed to travel at their saturation velocity throughout the entire multiplication region. The positions of each carrier

Table 5.2: Impact ionization coefficients for InP used in this model.

α_0 [/cm]	c_n [cm/V]	m_n	β_0 [/cm]	c_p [cm/V]	m_p
3.01×10^6	2.45×10^6	1.08	4.29×10^6	2.08×10^6	1.12

are updated at discrete time step by $v_{sat} \times dt$. At the birth of each carrier, it is given a "luck" between 0 and 1 (uniformly distributed). As the carrier travels, when the CDF of the ionization probability becomes larger than the carrier's luck, it will create another electron hole pair at that location, and the parent carrier's birth location x_0 and luck resets. This is the same random path length (RPL) model described in [78].

To clarify how the probability of creating an impact ionization event is determined, consider a hole with a luck of 0.3 while referring to Figure 5.1. At an electric field of 500 kV/cm, it will need to travel roughly 200 nm after the dead space before impact ionization; whereas at 400 kV/cm, the hole will need to travel around 700 nm after the dead space. Figure 5.2 plots an example time evolution of the location of holes, electrons, and impact ionization events over the first few picoseconds of one trial, for a multiplication width W of 1.0 μm.

Assume all electrons leaving the the multiplication region and reaching the conduction band barrier at the interface will instantly lose all excess energy. To model the electron escape from the barrier, we first get the electron sheet concentration σ at the interface is related to the quasi Fermi level E_{fn} by the following equation:

$$\sigma = \frac{kT}{qE_{tcb}} N_c \exp(\frac{E_{fn}}{kT}) \tag{5.13}$$

Knowing the Fermi level, the thermionic and intra-band tunneling rates of electrons crossing an energy barrier are given by the standard equations (see Sections 2.6.1

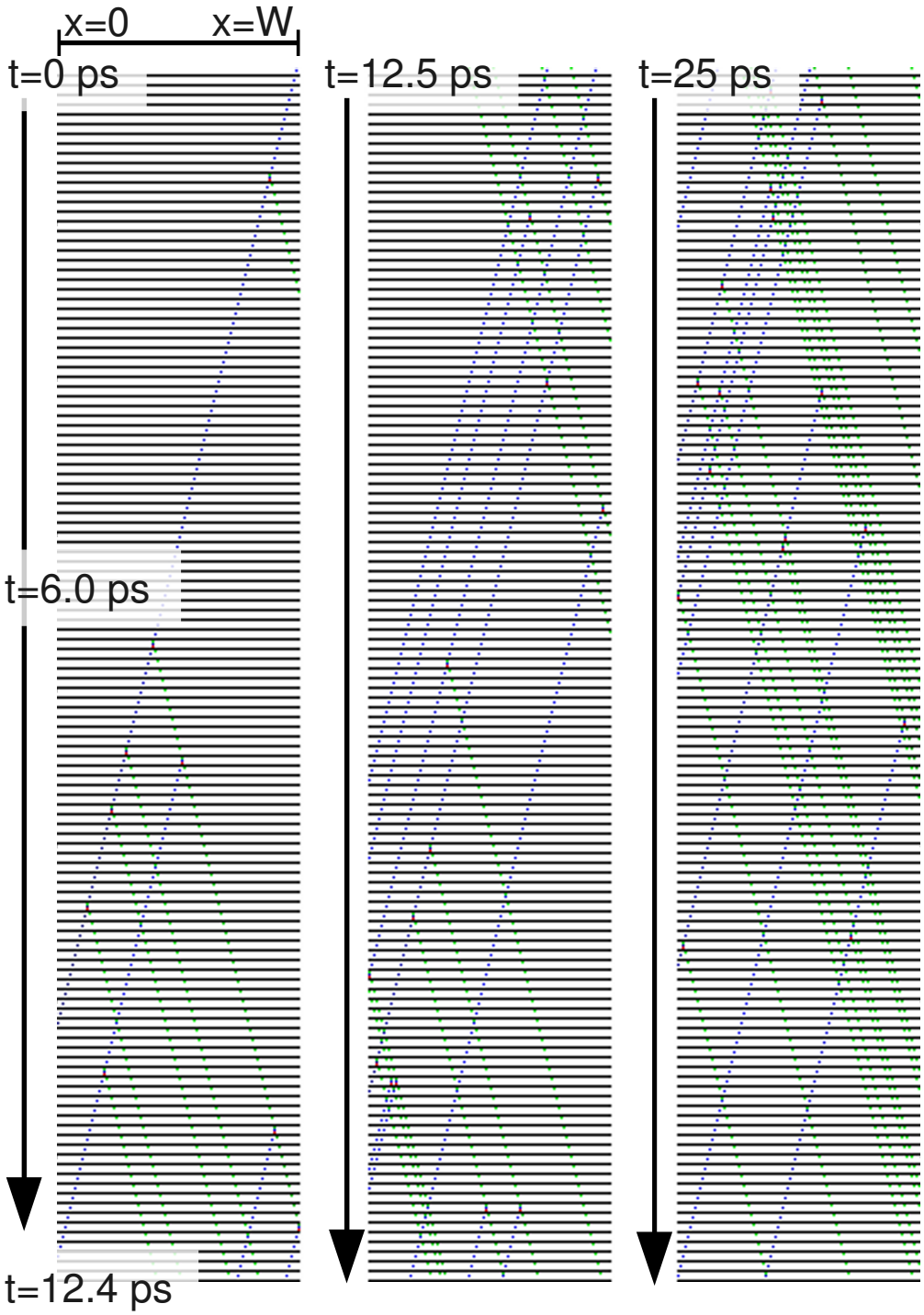

Figure 5.2: Location of holes (blue dots), electrons (green), and impact ionization events (red) in the multiplication region as a function of time (vertical axis). Each horizontal line denotes one time step (0.1 ps) in the simulation.

and 2.6.2):

$$J_{th} = \frac{4\pi m^* qkT}{h^3} \exp(\frac{E_{fn}}{kT}) kT \exp(-\frac{\Delta E_c}{kT}) \tag{5.14}$$

$$J_{tu} = \frac{4\pi m^* qkT}{h^3} \exp(\frac{E_{fn}}{kT}) \times \int_0^{\Delta E_c} \exp\left(-\frac{8\pi\sqrt{2m^*}}{3qE_{tcb}h}(\Delta E_c - E_x)^{3/2}\right) \exp(-\frac{E_x}{kT}) dE_x \tag{5.15}$$

Finally, the change of the electron concentration at the interface over a discrete time interval is given by:

$$\Delta n_w = (J_{th} + J_{tu}) dt \times S/q \tag{5.16}$$

where S is the effective area of the device (taken to be 10 μm^2 in this chapter).

After determining Δn_w, we can recalculate V_m for the next time step, and repeat the process. The simulation ends when the number of carriers in the multiplication region becomes zero. We measure the device current as

$$I(t) = I_h(t, x = 0) + \frac{\epsilon S}{W} \frac{dV_m}{dt} \tag{5.17}$$

The "measurement" of the current waveform can be repeated many times to obtain a standard counting histogram. Detection efficiency of the device is given by a fixed current threshold (e.g. 10 μA), and timing jitter by the timing information of when the measured current crosses this threshold. Recovery time can be given by measuring the detection efficiency as a function of photon inter-arrival times (see Section 4.2).

5.5 Simulation Results

To ease the computation complexity of this very large parameter space, we will fix the operating temperature at 240 K (easily obtainable by thermoelectric

cooling), and the multiplication region at 1.0 μm to balance tunneling with timing jitter. We will then focus on operating bias versus TCB strength. As the TCB strength depends on both the conduction band offset ΔE_c and the TCB length length L, here we will only examine the contributions of ΔE_c and fix L at 0.8 μm.

Let's first compare some simulated waveforms obtained by this model with those obtained by the previous deterministic model. Figure 5.3 shows several examples of the current output from the Monte Carlo simulation and compares them to the deterministic model presented earlier in Chapter 2, clearly demonstrating the timing jitter and pulse height variations due to the stochastic impact ionization process. In this simulation the device is operated at a small overbias above the breakdown voltage.

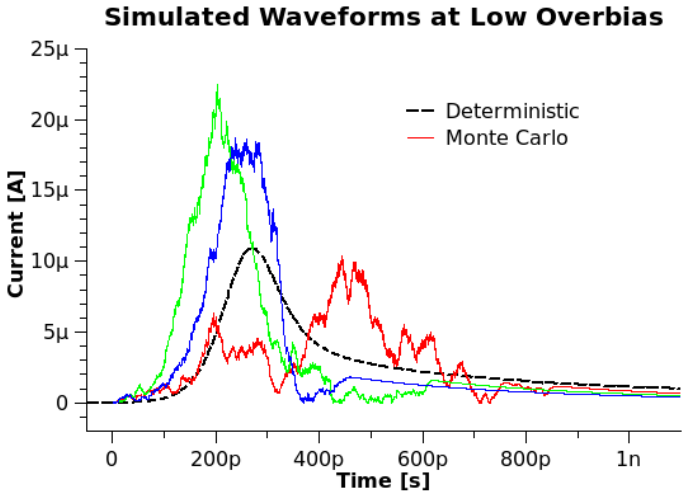

Figure 5.3: Simulated current waveforms from the Monte Carlo model (solid color lines) and the deterministic model (dashed black line).

By running the simulation many times, a counting histogram can be generated (Figure 5.4). Here the threshold for detection is set at 10 μA, and 1000 trials are run at an overbias voltage of 7% and a ΔE_c of 0.25 eV. Since this is an ideal model which simulates the response of the device to a single photo input, all the

collected counts contribute to the detection efficiency of the device. The timing variation of the detection threshold can be quantified as the width of the counting histogram peak.

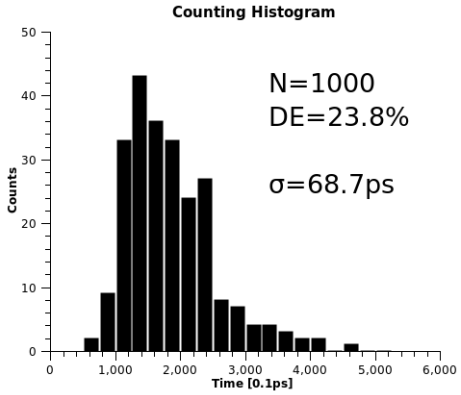

Figure 5.4: Counting histogram resulting from the Monte Carlo simulation.

We can also collect data regarding the variations in pulse shapes. Histograms of the avalanche gain, pulse peak, and pulse width are displayed in Figure 5.5. Gain is defined by the area of the avalanche pulse, and is of particular interest since it directly affects the afterpulsing rate. Pulse peak variations is useful since it characterizes the device's excess noise, but applications requiring good excess noise will not be discussed here.

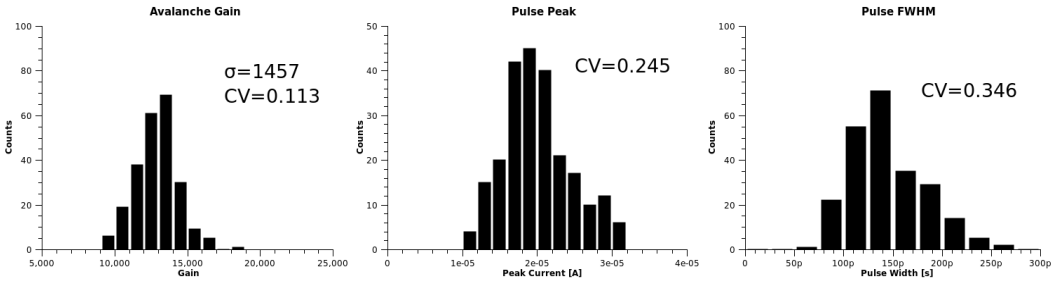

Figure 5.5: Histogram plots of the gain, pulse peak, and pulse width.

Most interesting are the dependencies of the SPDE (Figure 5.6(A)), timing jitter (Figure 5.6(B)), gain (Figure 5.7(A)), and recovery time (Figure 5.7(B)) as a function of the applied bias level and the TCB barrier height. The SPDE and timing jitter show no dependencies on the barrier height, and only depend on the applied voltage like a conventional SPAD. This is because the avalanche buildup time is much faster than the self-quenching speed, so the current crosses the detection threshold before it is able to self-quench. Note that these simulations represent the response to only a single photon isolated in time - in a real device the SPDE would depend on the TCB strength if it is operating close to the saturation limit mentioned in Chapter 4.

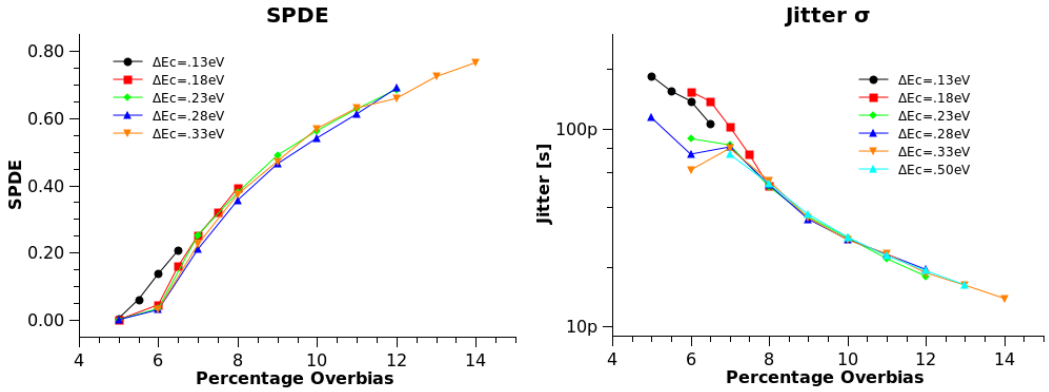

Figure 5.6: (A) SPDE as a function of percent overbias, at various barrier heights. (B) Jitter as a function of percent overbias, at various barrier heights.

The gain does depend on the barrier height, although only to a limit. From the figure, we see that at barrier heights larger than 0.23 eV, there is no further change in gain as the barrier height increases. This is due to the fact that the recovery time of the device is already much longer than the self-quenching speed of the device. At lower barrier heights and faster recovery times, the carriers can escape before the device is completely self-quenched, allowing for additional

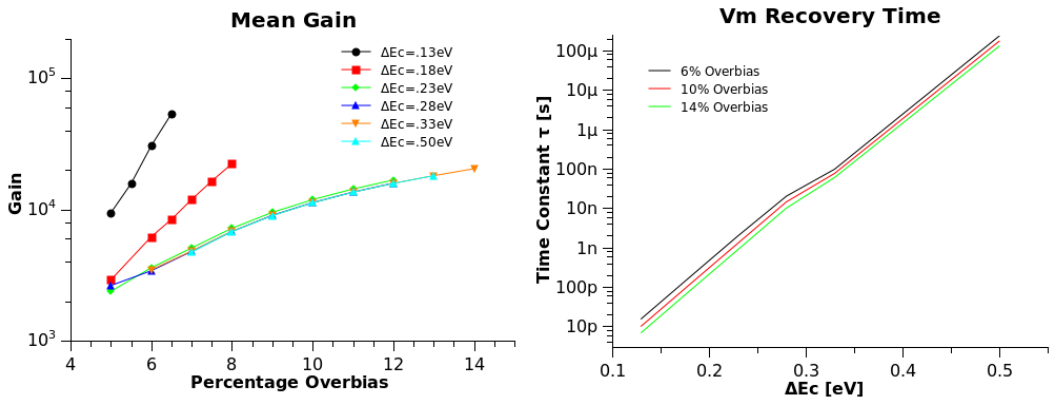

Figure 5.7: (A) Gain as a function of percent overbias, at various barrier heights. (B) Recovery time as a function of barrier height, at various applied biases.

carriers to be generated in the avalanche. In the ultimate limit of no barrier, the gain becomes infinite, just like a conventional SPAD operating in Geiger-mode (refer to Figure 2.4(B) for comparison).

The recovery time of the device can be calculated deterministically from the multiplication voltage recovery. Since the escape equations (due to tunneling and thermionic emission) mentioned previously have a time constant associated with the process, the recovery time could be fitted to this time constant. It needs to be mentioned that a more precise way of obtaining this value (and also realize the saturation effect of the SPDE), is to run the Monte Carlo simulation with multiple input carriers per run, and vary the time interval between the carriers for each trial. However, this calculation is too resource intensive to be done at the moment.

The last value we will need to calculate is the afterpulsing component of the total dark counts. From the simulation results, afterpulsing depends on the total current flow (i.e. gain) and recovery time of the device. Section 5.3 covers the calculation of the afterpulsing probability per avalanche event, given by Equation

5.9 and plotted in Figure 5.8. Note that a shift in the afterpulsing defect density would move the entire series of curves uniformly.

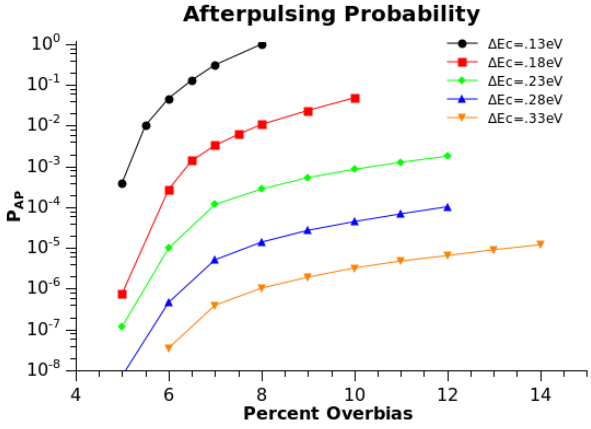

Figure 5.8: Afterpulsing probability per avalanche event as a function of percent overbias, at various barrier heights.

5.6 Application Design

With these simulation results we can then look into the design of the SPAD for various applications. For use in a communications link, the formulation derived in Section 5.2 is used, and the results for Equation 5.1 is plotted in Figure 5.9.

Note that in this figure, at higher overbiases than those depicted, the total DCR increases exponentially such that the effective bitrate is rapidly reduced to zero. We see from these results that there is an optimal barrier height in designing the SPAD for a maximum bitrate. At smaller barrier heights, the large gain and afterpulsing effects reduces the bitrate, while at larger barrier heights the long recovery time hinders device performance.

Single photon detectors are also used in many imaging applications, such as fluorescence lifetime imaging (FLIM) [81, 82] and time of flight laser imaging

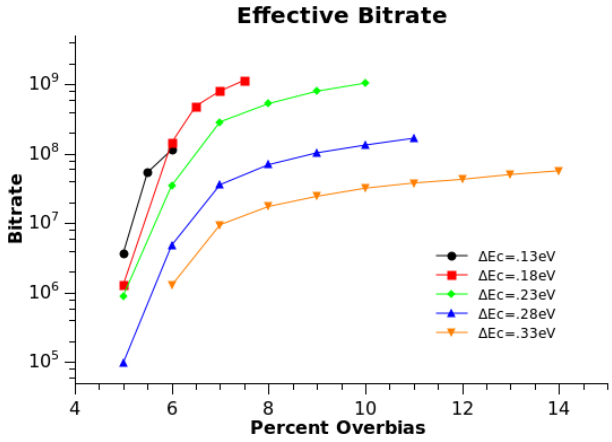

Figure 5.9: Effective bitrate of self-quenching SPAD using PPM and error correction codes.

(LIDAR). For these imaging applications, the noise equivalent power (NEP) is a useful figure of merit to examine. For single photon detectors, the NEP is given by [77, 83, 84]:

$$NEP = h\nu \frac{\sqrt{2DCR}}{SPDE} \tag{5.18}$$

and shown in Figure 5.10, assuming 1550 nm light is used. As expected, there is a minimum obtainable NEP with respect to voltage since the DCR usually rises faster than the SPDE at very high bias levels. Higher barrier heights also produce a lower NEP due to less afterpulsing.

In these correlated imaging applications, the timing of the jitter is often the limiting factor in determining the timing resolution of the measurement system. Thus it is important to examine the trade-offs between NEP and timing jitter when designing SPADs for these applications. Figure 5.11 shows these two values as an implicit function of applied bias, given various barrier heights. From the figure, when system counting rate is not important, a larger barrier height is optimal in terms of both NEP and timing jitter.

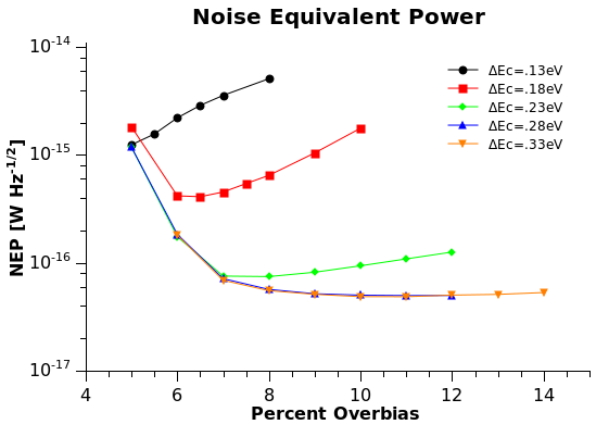

Figure 5.10: NEP as a function of bias, given various barrier heights.

On the other hand, for passive imaging applications where the photons are not correlated, of more importance is the total photon flux the detector is capable of sensing before saturating. The relationship between these two parameters are shown in Figure 5.12. This result demonstrates the clear trade-off between lower NEP and higher fluxes.

5.7 Conclusion

In this chapter we created a complete model of the self-quenching SPAD in order to determine the detection efficiency, photon count rate, dark count rate, timing jitter, and afterpulsing of the SPAD. We see a clear trade-off between afterpulsing and the device performance (both timing jitter and SPDE). By knowing these characteristics, we can optimize the SPAD for various applications depending on different criteria. We demonstrate the results of varying the barrier height of the SPAD in three such cases: communications using PPM, time correlated imaging, and passive imaging.

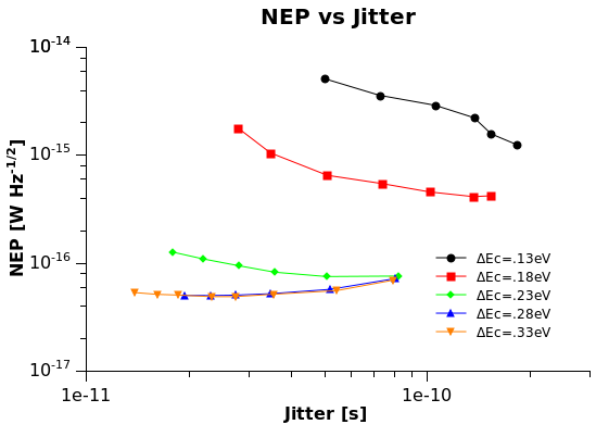

Figure 5.11: NEP vs timing jitter, at different barrier heights. This figure is useful in the design of SPADs of time-correlated imaging applications.

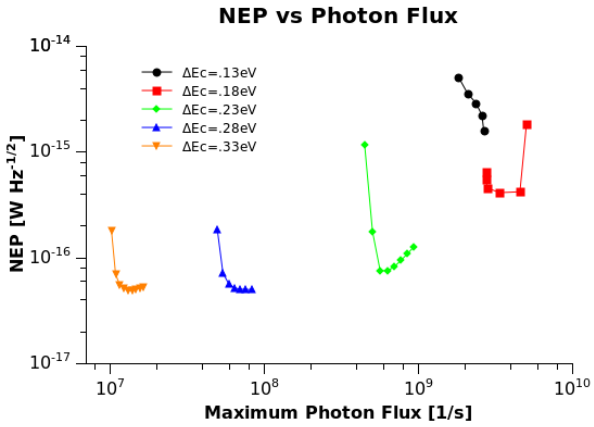

Figure 5.12: NEP vs maximum photon flux, at different barrier heights. This figure is useful in the design of SPADs for passive imaging applications.

This chapter, in part, is a reprint of material as it appears in the following publications:

- James Cheng, Yu-Hwa Lo, "Modeling of Self-Quenching Single Photon Avalanche Photodiodes for System Applications", IEEE J. Sel. Top. Quantum Electronics, submitted.

The dissertation author was the primary investigator and author of this material.

Chapter 6

Conclusion

6.1 Thesis Summary

This thesis presented several advancements in the development of single photon avalanche photodiodes for near-infrared detection. Design, fabrication, and test results for a self-quenching SPAD utilizing a patterned zinc ring diffusion design demonstrate several key contributions, summarized in the following:

- Design and fabrication of a self-quenching SPAD compatible with zinc diffusion and electroluminescence, through usage of an electron TCB.
- Development and demonstration of a practical zinc diffusion pattern which greatly improves the SPDE while reducing the fabrication complexity.
- Creating a compact optical setup capable of measuring the spatial dependence of SPDE.
- Measurement and characterization of the self-quenching SPAD's recovery time, as a function of operating bias and temperature, and epitaxial design.
- Simulation and formulation of a complete model describing the performance of the self-quenching SPAD.

By utilizing an internal self-quenching mechanism, the fabrication complexity of the SPAD is greatly reduced, allowing for the possibility of easily creating large array detectors. The self-quenching mechanism is achieved through the formation of a heterostructure barrier next to the multiplication region; in the case of this thesis work, a conduction band offset is created to temporarily stop electrons. A qualitative model of this self-quenching and self-recovery process is presented, verifying the physical processes involved in this design is enough to cause self-quenching.

Previously fabricated devices have shown that mesa-etched devices have very low SPDE due to the mesa etching process. The next step is to isolate the multiplication junction from the surface, through the process of zinc diffusion. However, this yields its own problems due to the non-uniform field crowding effects near the junction edge, and this can be seen by sweeping the laser illumination spot and measuring the SPDE as a function of position. A novel zinc diffusion pattern is created and demonstrated a significant increase in the device performance. Instead of trying to minimize the edge breakdown effects, these effects are purposely utilized throughout the device to maximize its performance. Another advantage of this design is its independence from the exact process step used to create the zinc diffusion.

The devices were tested and self-quenching characteristics quantified. It was shown that the self-recovery time of the device limits the SPDE performance, and the epitaxial design was modified to improve this parameter. However, by reducing the self-recovery time, the afterpulsing effects of the device also increased greatly. The high dark count rate due to the afterpulsing effect is shown to limit the SPDE. A model is presented to include this effect, demonstrating the best

device design in view of a particular application's design criterion.

6.2 Outlook

Much research remains in improving avalanche photodiodes for use as the ideal single photon detector. As discussed extensively in the thesis, the single photon detection efficiency, photon counting rate, dark count rate, afterpulsing, timing jitter, and excess noise can all be further improved. Ultimately, the particular SPAD design will depend on its intended application, as simultaneously improving all of the figures of merit proves difficult.

By far the largest factor limiting the SPDE is the DCR. Contributions to the DCR can be attributed to three main components: intrinsic material property, non-idealities introduced through processing, and operating conditions. Work from this and other groups to address each of these issues will be briefly discussed.

6.2.1 Intrinsic Material Property

The intrinsic material property of the semiconductor will always contribute to dark counts, with the most significant source being thermal generation in the low bandgap material of the SAM-APD.

One possibility to alleviate the usage of low-bandgap semiconductor materials is to use a hybrid APD structure, with another optical component to upcovert the near-infrared light to visible. This high energy photon can then be directly detected by an APD with a high bandgap and low noise, such as a silicon SPAD. Work in this area have demonstrated 1550 nm photons can be converted to 630 nm with a pump laser of 1064 nm using periodically poled lithium niobate (PPLN) [85].

Although this work showed a high frequency upconversion efficiency of 90%, fluorescence of the nonlinear optical crystal leads to a high background count rate.

6.2.2 Processing Non-idealities

While simple, mesa-etching the devices introduces a significant amount of defects which can increase the DCR. There is significant work in the area of mesa passivation [86–90], but ultimately zinc diffused devices still have better performance. The zinc ring design presented in Chapter 3 was created to address the non-idealities when creating the p-i-n junction, and its limitations were briefly discussed at the end of the chapter. Further work can be done to improve this, but in practice correctly engineering the electric field profile remains tedious and unrepeatable. The ideal detector should be able to remove this complex processing step.

There are several promising ideas on removing both the mesa surface state effect while not requiring zinc diffusion. Although currently only demonstrated in a conventional photodiode for mid-infrared wavelengths, a nBn type device can significantly reduce the dark current (Figure 6.1, [91, 92]). These minority carrier devices have an effective photo-active area defined not by mesa etching or junction doping, but by the diffusion length of the minority carriers. With the large conduction band barrier, any thermally generated carriers originating from outside the active area will recombine before reaching the electrode and not contribute to the current. The nBn design also removes any possible surface leakage current since no mesa etching is necessary to define the active area. For the InAs detector, the noise enhancement due to this design is similar to reducing the temperature by 100 K for a conventional p-n photodetector.

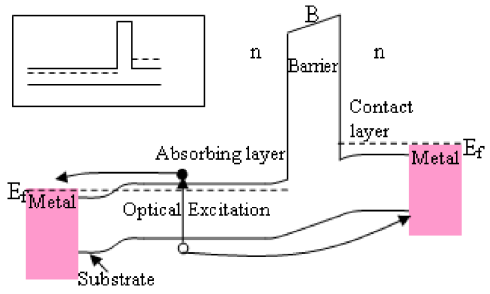

Figure 6.1: Schematic band diagram of a nBn detector.

Another method currently being investigated is the usage of a three terminal device (Figure 6.2) to reduce the surface leakage current from a mesa-etched device. Instead of a single anode contact, dual anodes are created: one at the center of the device, and another towards the outer edge. The outer electrode "guards" the center electrode from any of the dark counts arising from the surface states, while only the true photosensitive signals from the center is collected. The difficulties arising from this design are the fabrication challenges with adding the second top contact, a larger device area due to the three terminals, and leakage current between the center and outer electrodes.

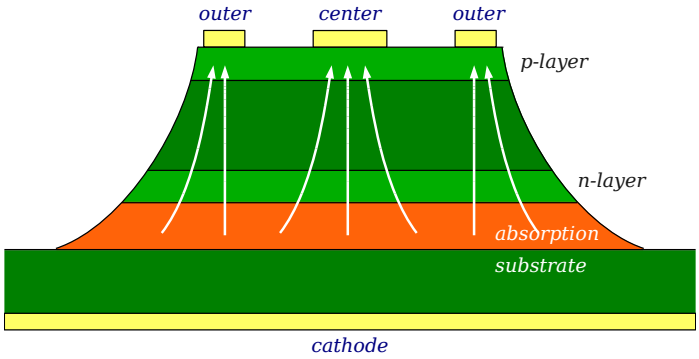

Figure 6.2: Schematic operating principle of a three terminal SPAD. Any surface induced dark currents are collected by the outer electrode and do not contribute to the primary signal read form the center electrode.

6.2.3 Operating Conditions

During operation of the SPAD, one can control the DCR through either the temperature or the bias. The effects of temperature are obvious, but reducing this value too much is not practical in many real world applications. The effects of bias mainly include afterpulsing, and to a lesser extent, dark current due to tunneling.

Afterpulsing remains the critical issue to overcome, as the bias is limited by the afterpulsing and dark count of the device. Conventionally, afterpulsing can be reduced by reducing the total current flow to the minimal detectable level [93], and properly adjusting the arming voltage to minimize the probability of a freed carrier from generating an afterpulsing event [21]. The former was also one of the goals of the self-quenching design discussed in this thesis.

An alternative method of achieving single photon sensitivities without the downside of afterpulsing is by combining multiple gain mechanisms within the device instead of the single high avalanche gain mechanism used in a SPAD. By keeping the avalanche gain low, and operating the APD portion in sub-Geiger mode, the afterpulsing effect becomes essentially eliminated. A structure that utilizes both avalanche gain and field-effect transistor gain mechanisms can reach the same overall total gain at a relatively low avalanche gain (Figure 6.3). Epitaxial structures to create this device are listed in Tables 6.1 and 6.2, while the band diagrams at equilibrium corresponding to these structures are shown in Figure 6.4.

The "APDFET" structure works by using an APD as the gate of a standard FET structure. After the photon is absorbed, the generated hole drifts into the multiplication region and undergoes moderate gain (e.g. ~100). By designing the valence band barrier which stops the holes versus the conduction band barrier used

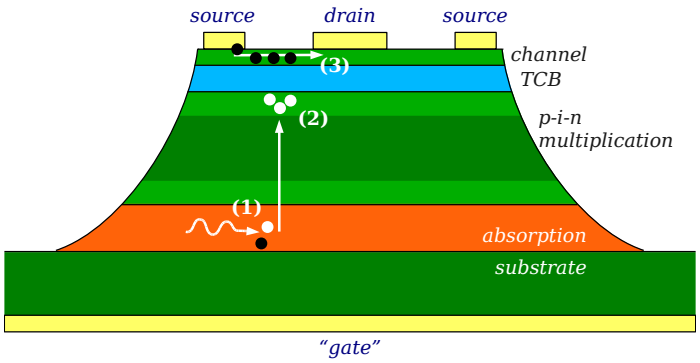

Figure 6.3: Schematic operating principle of an APDFET structure. (1) Photogenerated holes drift into the multiplication region, undergoing avalanche gain and collecting at the TCB (2). This turns on the FET electron channel (3).

Table 6.1: Epitaxial design (M) for APDFET structure.

Layer	Material	Thickness [μm]	Doping [cm^{-3}]	Comment
0	InP		n-type	Substrate
1	InP	0.5	n=1×10^{18}	Buffer
2	InGaAs	1.5	-	Absorption
3	InGaAsP(1.55μm)	0.03	-	Transition
4	InGaAsP(1.3μm)	0.03	-	
5	InGaAsP(1.1μm)	0.03	-	
6	InP	0.2	n=1.15×10^{17}	Field Control
7	InP	0.8	-	Multiplication
8	InP	0.2	p=2.5×10^{17}	Field Control
9	5x InAlAs	0.05	-	TCB superlattice
10	4x InP	0.05	-	TCB superlattice
18	InP	0.03	p=5×10^{16}	FET Channel
19	InGaAsP(1.3μm)	0.05	n=5×10^{17}	FET source/drain
20	InP	0.15	n=2×10^{18}	Contact

Table 6.2: Epitaxial design (L) for APDFET structure, with larger ΔE_c.

Layer	Material	Thickness [μm]	Doping [cm^{-3}]	Comment
0	InP		n-type	Substrate
1	InP	0.5	n=1×10^{18}	Buffer
2	InGaAs	1.5	-	Absorption
3	InGaAsP(1.55μm)	0.03	-	Transition
4	InGaAsP(1.3μm)	0.03	-	
5	InGaAsP(1.1μm)	0.03	-	
6	InP	0.2	n=1.15×10^{17}	Field Control
7	InP	0.8	-	Multiplication
8	InP	0.2	p=2.5×10^{17}	Field Control
9	InGaAlAs(1.1μm)	0.07	-	TCB
10	InAlAs	0.4	-	TCB
11	InP	0.03	p=1×10^{17}	FET Channel
12	InGaAs	0.03	n=5×10^{17}	FET source/drain
13	InGaAsP(1.3μm)	0.03	n=5×10^{17}	FET source/drain
14	InP	0.14	n=2×10^{18}	Contact

to hold the electrons within the channel, a high transistor gain (e.g. 100~1000) can be achieved. The output signal is read from the center drain electrode, and the overall total gain from the device produces a detectable signal even at the moderate avalanche gain.

Preliminary results from Epi (M) show a promising DC gain at a moderate bias, but the electron TCB designed to stop electrons leaking from the channel proved to be too weak. This results in a positive feedback mechanism between the avalanche gain and the transistor gain. The positive feedback created very

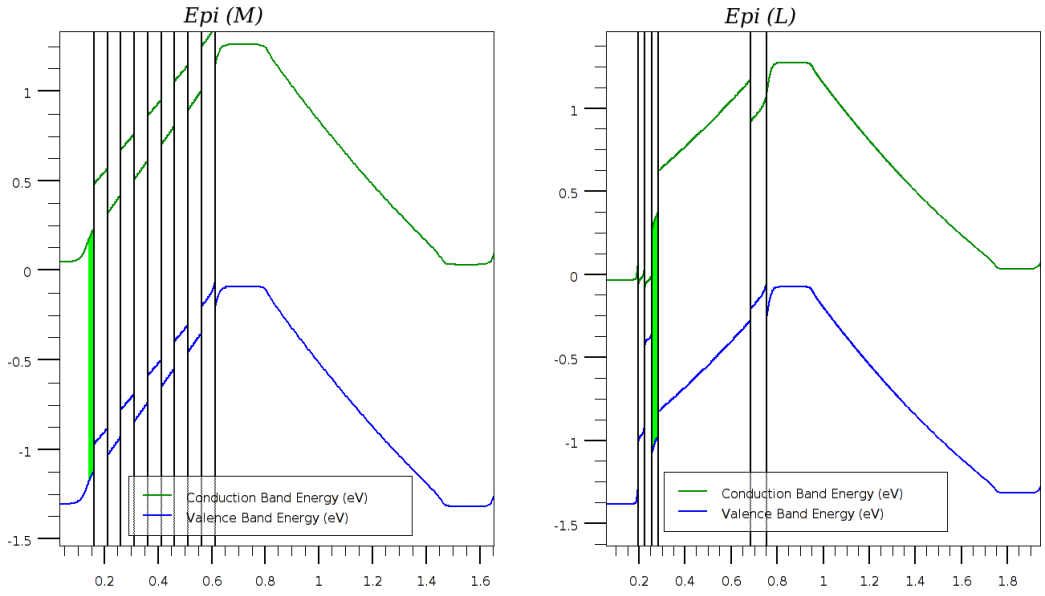

Figure 6.4: Simulated equilibrium band diagrams of APDFET structures (M) and (L). The green layer denotes the FET channel layer.

large current pulses which cannot be quenched, nor allow the device a chance to recover and detect another photon. Epi (L) was designed to increase the electron barrier and reduce the positive feedback mechanism. By cascading multiple gain mechanisms together, this structure is expected to reach single photon sensitivities without any afterpulsing effects.

Bibliography

[1] R. H. Hadfield. "Single-photon detectors for optical quantum information applications." *Nature Photonics*, volume 3, no. 12 pp. 696–705, 2009. doi:10.1038/nphoton.2009.230.

[2] G. S. Buller and R. J. Collins. "Single-photon generation and detection." *Measurement Science and Technology*, volume 21, no. 1 p. 012002, 2010. doi:10.1088/0957-0233/21/1/012002.

[3] Hamamatsu Photonics K.K. "Photomultiplier Tubes: Basics and Applications, 3rd E.", 2006.

[4] G. N. Gol'tsman, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, K. Smirnov, B. Voronov, A. Dzardanov, C. Williams, and R. Sobolewski. "Picosecond superconducting single-photon optical detector." *Applied Physics Letters*, volume 79, no. 6 pp. 705–707, 2001. doi:10.1063/1.1388868.

[5] E. A. Dauler. "Superconducting nanowire photodetector arrays." Technical report, MIT Lincoln Laboratory, 2010.

[6] J. Kitaygorsky, J. Zhang, A. Verevkin, A. Sergeev, A. Korneev, V. Matvienko, P. Kouminov, K. Smirnov, B. Voronov, G. Gol'tsman, and R. Sobolewski. "Origin of dark counts in nanostructured nbn single-photon detectors." *Applied Superconductivity, IEEE Transactions on*, volume 15, no. 2 pp. 545–548, june 2005. doi:10.1109/TASC.2005.849914.

[7] K. M. Rosfjord, J. K. W. Yang, E. A. Dauler, A. J. Kerman, V. Anant, B. M. Voronov, G. N. Gol'tsman, and K. K. Berggren. "Nanowire single-photon detector with an integrated optical cavity and anti-reflection coating." *Opt. Express*, volume 14, no. 2 pp. 527–534, Jan 2006. doi:10.1364/OPEX.14.000527.

[8] C. Canali, P. Pavan, A. Carlo, P. Lugli, R. Malik, M. Manfredi, A. Neviani, L. Vendrame, E. Zanoni, and G. Zandler. "Experimental and Monte Carlo analysis of impact-ionization in AlGaAs/GaAs HBT's." *Electron Devices, IEEE Transactions on*, volume 43, no. 11 pp. 1769–1777, nov 1996. doi:10.1109/16.542420.

[9] Y. Okuto and C. R. Crowell. "Ionization coefficients in semiconductors: A nonlocalized property." *Phys. Rev. B*, volume 10 pp. 4284–4296, Nov 1974. doi:10.1103/PhysRevB.10.4284.

[10] F. Osaka, T. Mikawa, and O. Wada. "Analysis of Impact Ionization Phenomena in InP by Monte Carlo Simulation." *Japanese Journal of Applied Physics*, volume 25, no. Part 1, No. 3 pp. 394–401, 1986. doi:10.1143/JJAP.25.394.

[11] D. Harrison, R. A. Abram, and S. Brand. "Impact ionization rate calculations in wide band gap semiconductors." *Journal of Applied Physics*, volume 85, no. 12 pp. 8178–8185, 1999. doi:10.1063/1.370657.

[12] N. Sano, M. Tomizawa, and A. Yoshii. "Monte carlo analysis of hot electron transport and impact ionization in silicon." *Japanese Journal of Applied Physics*, volume 30, no. Part 1, No. 12B pp. 3662–36665, 1991. doi:10.1143/JJAP.30.3662.

[13] D. Harrison, R. A. Abram, and S. Brand. "Characteristics of impact ionization rates in direct and indirect gap semiconductors." *Journal of Applied Physics*, volume 85, no. 12 pp. 8186–8192, 1999. doi:10.1063/1.370658.

[14] M. Saleh, M. Hayat, P. Sotirelis, A. Holmes, J. Campbell, B. Saleh, and M. Teich. "Impact-ionization and noise characteristics of thin III-V avalanche photodiodes." *Electron Devices, IEEE Transactions on*, volume 48, no. 12 pp. 2722–2731, dec 2001. doi:10.1109/16.974696.

[15] C. A. Armiento, S. H. Groves, and C. E. Hurwitz. "Ionization coefficients of electrons and holes in InP." *Applied Physics Letters*, volume 35, no. 4 pp. 333–335, 1979. doi:10.1063/1.91111.

[16] L. W. Cook, G. E. Bulman, and G. E. Stillman. "Electron and hole impact ionization coefficients in InP determined by photomultiplication measurements." *Applied Physics Letters*, volume 40, no. 7 pp. 589–591, 1982. doi:10.1063/1.93190.

[17] L. Tarof, J. Yu, T. Baird, R. Bruce, and D. Knight. "Temperature measurements of separate absorption, grading, charge, and multiplication (SAGCM) InP/InGaAs avalanche photodiodes (APD's)." *Photonics Technology Letters, IEEE*, volume 5, no. 9 pp. 1044–1046, sept. 1993. doi:10.1109/68.257186.

[18] S. Sze and K. K. Ng. *Physics of semiconductor devices*. Wiley-Interscience, Hoboken, N.J, 2007.

[19] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. "Avalanche photodiodes and quenching circuits for single-photon detection." *Appl. Opt.*, volume 35, no. 12 pp. 1956–1976, Apr 1996. doi:10.1364/AO.35.001956.

[20] F. Zappa, A. Lotito, A. Giudice, S. Cova, and M. Ghioni. "Monolithic active-quenching and active-reset circuit for single-photon avalanche detectors." *Solid-State Circuits, IEEE Journal of*, volume 38, no. 7 pp. 1298–1301, july 2003. doi:10.1109/JSSC.2003.813291.

[21] M. Liu, C. Hu, J. Campbell, Z. Pan, and M. Tashima. "Reduce afterpulsing of single photon avalanche diodes using passive quenching with active reset." *Quantum Electronics, IEEE Journal of*, volume 44, no. 5 pp. 430–434, may 2008. doi:10.1109/JQE.2007.916688.

[22] C. Hu, M. Liu, X. Zheng, and J. Campbell. "Dynamic range of passive quenching active reset circuit for single photon avalanche diodes." *Quantum Electronics, IEEE Journal of*, volume 46, no. 1 pp. 35–39, jan. 2010. doi:10.1109/JQE.2009.2024086.

[23] M. Liu. *Infrared Single Photon Avalanche Diodes*. Ph.D. thesis, University of Virginia, may 2008.

[24] J. Zhang, R. Thew, C. Barreiro, and H. Zbinden. "Practical fast gate rate InGaAs/InP single-photon avalanche photodiodes." *Applied Physics Letters*, volume 95, no. 9 091103, 2009. doi:10.1063/1.3223576.

[25] D. Bethune and W. Risk. "An autocompensating fiber-optic quantum cryptography system based on polarization splitting of light." *Quantum Electronics, IEEE Journal of*, volume 36, no. 3 pp. 340–347, march 2000. doi:10.1109/3.825881.

[26] Z. L. Yuan, B. E. Kardynal, A. W. Sharpe, and A. J. Shields. "High speed single photon detection in the near infrared." *Applied Physics Letters*, volume 91, no. 4 041114, 2007. doi:10.1063/1.2760135.

[27] N. Namekata, S. Adachi, and S. Inoue. "1.5 GHz single-photon detection at telecommunication wavelengths using sinusoidally gated InGaAs/InP avalanche photodiode." *Opt. Express*, volume 17, no. 8 pp. 6275–6282, Apr 2009. doi:10.1364/OE.17.006275.

[28] N. Namekata, S. Sasamori, and S. Inoue. "800 MHz single-photon detection at 1550-nm using an InGaAs/InP avalanche photodiode operated with a sine wave gating." *Opt. Express*, volume 14, no. 21 pp. 10043–10049, Oct 2006. doi:10.1364/OE.14.010043.

[29] P. Michler, A. Imamoglu, M. D. Mason, P. J. Carson, G. F. Strouse, and S. K. Buratto. "Quantum correlation among photons from a single quantum dot at room temperature." *Nature*, volume 406, no. 6799 pp. 968–970, 2000. doi:http://dx.doi.org/10.1038/35023100.

[30] B. Lounis and W. E. Moerner. "Single photons on demand from a single molecule at room temperature." *Nature*, volume 407, no. 6803 pp. 491–493, 2000. doi:http://dx.doi.org/10.1038/35035032.

[31] F. T. Arecchi. "Measurement of the statistical distribution of gaussian and laser sources." *Phys. Rev. Lett.*, volume 15 pp. 912–916, Dec 1965. doi:10.1103/PhysRevLett.15.912.

[32] J. A. Fleck. "Quantum Theory of Laser Radiation. I. Many-Atom Effects." *Phys. Rev.*, volume 149 pp. 309–321, Sep 1966. doi:10.1103/PhysRev.149.309.

[33] M. A. Itzler, r. Ben-Michael, C. F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, and R. Ispasoiu. "Single photon avalanche diodes (spads) for 1.5 m photon counting applications." *Journal of Modern Optics*, volume 54, no. 2-3 pp. 283–304, 2007. doi:10.1080/09500340600792291.

[34] J. Donnelly, E. Duerr, K. McIntosh, E. Dauler, D. Oakley, S. Groves, C. Vineis, L. Mahoney, K. Molvar, P. Hopman, K. Jensen, G. Smith, S. Verghese, and D. Shaver. "Design Considerations for 1.06 um InGaAsP/InP Geiger-Mode Avalanche Photodiodes." *Quantum Electronics, IEEE Journal of*, volume 42, no. 8 pp. 797–809, aug. 2006. doi:10.1109/JQE.2006.877300.

[35] A. Spinelli and A. Lacaita. "Physics and numerical simulation of single photon avalanche diodes." *Electron Devices, IEEE Transactions on*, volume 44, no. 11 pp. 1931–1943, nov 1997. doi:10.1109/16.641363.

[36] R. McIntyre. "Multiplication noise in uniform avalanche diodes." *Electron Devices, IEEE Transactions on*, volume 13, no. 1 pp. 164–168, jan 1966. doi:10.1109/T-ED.1966.15651.

[37] F. Capasso, W. T. Tsang, A. L. Hutchinson, and G. F. Williams. "Enhancement of electron impact ionization in a superlattice: A new avalanche photodiode with a large ionization rate ratio." *Applied Physics Letters*, volume 40, no. 1 pp. 38–40, 1982. doi:10.1063/1.92910.

[38] T. Kagawa, Y. Kawamura, and H. Iwamura. "InGaAsP/InAlAs superlattice avalanche photodiode." *Quantum Electronics, IEEE Journal of*, volume 28, no. 6 pp. 1419–1423, jun 1992. doi:10.1109/3.135291.

[39] P. Yuan, C. Hansing, K. Anselm, C. Lenox, H. Nie, A. J. Holmes, B. Streetman, and J. Campbell. "Impact ionization characteristics of III-V semiconductors for a wide range of multiplication region thicknesses." *Quantum Electronics, IEEE Journal of*, volume 36, no. 2 pp. 198–204, feb. 2000. doi:10.1109/3.823466.

[40] K. Zhao, A. Zhang, Y. hwa Lo, and W. Farr. "InGaAs single photon avalanche detector with ultralow excess noise." *Applied Physics Letters*, volume 91, no. 8 081107, 2007. doi:10.1063/1.2772231.

[41] S. You, J. Cheng, and Y.-H. Lo. "Physics of Single Photon Avalanche Detectors With Built-In Self-Quenching and Self-Recovering Capabilities." *Quantum Electronics, IEEE Journal of*, volume 48, no. 7 pp. 960–967, July 2012. doi:10.1109/JQE.2012.2196679.

[42] H. Sudo and M. Suzuki. "Surface degradation mechanism of InP/InGaAs APDs." *Lightwave Technology, Journal of*, volume 6, no. 10 pp. 1496–1501, oct 1988. doi:10.1109/50.7907.

[43] B. Liu, J.-P. Landesman, J.-L. Leclercq, A. Rhallabi, C. Cardinaud, S. Guilet, F. Pommereau, M. Avella, M. González, and J. Jiménez. "InP surface properties under ICP plasma etching using mixtures of chlorides and hydrides." *Materials Science in Semiconductor Processing*, volume 9, no. 1–3 pp. 225–229, 2006. doi:10.1016/j.mssp.2006.01.064.

[44] H. Wieder. "Surface fermi level of III–V compound semiconductor-dielectric interfaces." *Surface Science*, volume 132, no. 1–3 pp. 390–405, 1983. doi:10.1016/0039-6028(83)90549-6.

[45] T. Sawada, K. ich Numata, S. Tohdoh, T. Saitoh, and H. Hasegawa. "*In-Situ* characterization of compound semiconductor surfaces by novel photoluminescence surface state spectroscopy." *Japanese Journal of Applied Physics*, volume 32, no. Part 1, No. 1B pp. 511–517, 1993. doi:10.1143/JJAP.32.511.

[46] H. Iber, S. Mo, E. Peiner, G. Vollrath, A. Schlachetzki, and F. Fiedler. "Characterization of surface damage in dry-etched InP." *Semiconductor Science and Technology*, volume 12, no. 6 p. 755, 1997. doi:10.1088/0268-1242/12/6/019.

[47] S. Sze and G. Gibbons. "Effect of junction curvature on breakdown voltage in semiconductors." *Solid-State Electronics*, volume 9, no. 9 pp. 831–845, 1966. doi:10.1016/0038-1101(66)90033-5.

[48] Y. Liu, S. Forrest, J. Hladky, M. Lange, G. Olsen, and D. Ackley. "A planar InP/InGaAs avalanche photodiode with floating guard ring and double diffused junction." *Lightwave Technology, Journal of*, volume 10, no. 2 pp. 182–193, feb 1992. doi:10.1109/50.120573.

[49] S. Cho, S. Yang, J. Ma, S. Lee, J. Yu, A. Choo, T. Kim, and J. Burm. "Suppression of avalanche multiplication at the periphery of diffused junction by floating guard rings in a planar InGaAs-InP avalanche photodiode." *Photonics Technology Letters, IEEE*, volume 12, no. 5 pp. 534–536, may 2000. doi:10.1109/68.841277.

[50] M. Kim, J. Baek, T. Kim, S. Kim, and K. Chung. "Characterization of double floating guard ring type InP-InGaAs avalanche photodiodes with Au/Zn low resistance ohmic contacts." *Thin Solid Films*, volume 514, no. 1–2 pp. 250–253, 2006. doi:10.1016/j.tsf.2006.02.019.

[51] G. Hasnain, W. Bi, S. Song, J. Anderson, N. Moll, C.-Y. Su, J. Hollenhorst, N. Baynes, I. Athroll, S. Amos, and R. Ash. "Buried-mesa avalanche photodiodes." *Quantum Electronics, IEEE Journal of*, volume 34, no. 12 pp. 2321–2326, dec 1998. doi:10.1109/3.736100.

[52] K. Zhao. *III-V Single Photon Avalanche Detector with built-in negative feedback for NIR photon detection.* Ph.D. thesis, University of California, San Diego, 2008.

[53] M. Yamada, P. K. Tien, R. J. Martin, R. E. Nahory, and A. A. Ballman. "Double zinc diffusion fronts in InP—Theory and experiment." *Applied Physics Letters*, volume 43, no. 6 pp. 594–596, 1983. doi:10.1063/1.94436.

[54] G. J. van Gurp, P. R. Boudewijn, M. N. C. Kempeners, and D. L. A. Tjaden. "Zinc diffusion in n-type indium phosphide." *Journal of Applied Physics*, volume 61, no. 5 pp. 1846–1855, 1987. doi:10.1063/1.338028.

[55] B. Tuck and A. Hooper. "Diffusion profiles of zinc in indium phosphide." *Journal of Physics D: Applied Physics*, volume 8, no. 15 p. 1806, 1975.

[56] S. He and Y. Zhao. "An experimental investigation of Zn diffusion into InP and InGaAs." *Semiconductor Science and Technology*, volume 20, no. 2 p. 149, 2005.

[57] Y. Huang, J.-H. Ryou, and R. D. Dupuis. "Control of Zn diffusion in InP/InAlGaAs-based heterojunction bipolar transistors and light emitting transistors." *Journal of Crystal Growth*, volume 310, no. 19 pp. 4345–4350, 2008. doi:10.1016/j.jcrysgro.2008.07.034.

[58] M. Glade, J. Hergeth, D. Grützmacher, K. Masseli, and P. Balk. "Diffusion of Zn acceptors during MOVPE of InP." *Journal of Crystal Growth*, volume 108, no. 3–4 pp. 449–454, 1991. doi:10.1016/0022-0248(91)90221-P.

[59] M. Wada, K. Sakakibara, M. Higuchi, and Y. Sekiguchi. "Evaluation of Surface Zn Concentration in Zn Diffusion into InP." *Japanese Journal of Applied Physics*, volume 31, no. Part 2, No. 5B pp. L597–L599, 1992. doi:10.1143/JJAP.31.L597.

[60] M. Wada, K. Sakakibara, M. Higuchi, and Y. Sekiguchi. "Investigation of Zn diffusion in InP using dimethylzinc as Zn source." *Journal of Crystal Growth*, volume 114, no. 3 pp. 321–326, 1991. doi:10.1016/0022-0248(91)90048-A.

[61] M. Wada, M. Seko, K. Sakakibara, and Y. Sekiguchi. "GaInAs/InP PIN Photodiodes Fabricated by MOVPE and a New Zn Diffusion Technique." *Japanese Journal of Applied Physics*, volume 29, no. Part 2, No. 3 pp. L401–L404, 1990. doi:10.1143/JJAP.29.L401.

[62] R. A. Logan, S. N. G. Chu, M. Geva, N. T. Ha, and C. D. Thurmond. "Zinc incorporation into InP grown by atmospheric pressure metalorganic vapor phase epitaxy." *Journal of Applied Physics*, volume 79, no. 3 pp. 1371–1377, 1996. doi:10.1063/1.361035.

[63] N. Arnold, R. Schmitt, and K. Heime. "Diffusion in III-V semiconductors from spin-on-film sources." *Journal of Physics D: Applied Physics*, volume 17, no. 3 p. 443, 1984.

[64] S. K. Rai. *Zinc diffusion in GaAsSb from spin–on glass dopant sources*. Master's thesis, University of Notre Dame, 2004.

[65] U. König, H. Haspeklo, P. Marschall, and M. Kuisl. "Spin-on and boat diffusion of Zn into InP and InGaAs grown by metalorganic vapor-phase epitaxy and liquid-phase epitaxy." *Journal of Applied Physics*, volume 65, no. 2 pp. 548–552, 1989. doi:10.1063/1.343138.

[66] S. Aytaç and A. Schlachetzki. "Shallow and selective diffusion of zinc in indium phosphide." *Solid-State Electronics*, volume 24, no. 1 pp. 57–61, 1981. doi:10.1016/0038-1101(81)90212-4.

[67] K. Yoon, Y. Lee, D. Yeo, and S. Kim. "The characteristics of Zn-doped InP using spin-on dopant as a diffusion source." *Journal of Electronic Materials*, volume 31 pp. 244–247, 2002.

[68] M. Ettenberg, M. Lange, A. Sugg, M. Cohen, and G. Olsen. "Zinc diffusion in InAsP/InGaAs heterostructures." *Journal of Electronic Materials*, volume 28 pp. 1433–1439, 1999.

[69] J. Zhao, Y. Li, M. Lange, M. Cohen, and G. Olsen. "Utilising Zn segregation at InP/InGaAs interface for as-deposited ohmic contact formation for photonic and electronic device applications." *Electronics Letters*, volume 37, no. 16 pp. 1048–1049, aug 2001. doi:10.1049/el:20010680.

[70] P. Kohl and F. Ostermayer. "Photoelectrochemical methods for III-V compound semiconductor device processing." *Annual Review of Materials Science*, volume 19 pp. 379–399, 1989. doi:10.1146/annurev.ms.19.080189.002115.

[71] T. Kallstenius, U. Smith, and B. Stoltz. "Studies of Internal Structure in InGaAsP/InP-Based Lasers Using Atomic Force Microscopy in Combination

with Selective Etching." *Journal of The Electrochemical Society*, volume 146, no. 2 pp. 749–754, 1999. doi:10.1149/1.1391675.

[72] J. B. Williamson and K. W. Carey. "Dopant–Type Selective Electroless Photoetching of Zn–Diffused InP and InGaAs/InP Heterostructures." *Journal of The Electrochemical Society*, volume 140, no. 7 pp. 2125–2128, 1993. doi:10.1149/1.2220777.

[73] M. N. Ruberto, X. Zhang, R. Scarmozzino, A. E. Willner, D. V. Podlesnik, and R. M. Osgood. "The Laser–Controlled Micrometer–Scale Photoelectrochemical Etching of III–V Semiconductors." *Journal of The Electrochemical Society*, volume 138, no. 4 pp. 1174–1185, 1991. doi:10.1149/1.2085737.

[74] H. Finkelstein, M. Gross, Y.-H. Lo, and S. Esener. "Analysis of hot-carrier luminescence for infrared single-photon upconversion and readout." *Selected Topics in Quantum Electronics, IEEE Journal of*, volume 13, no. 4 pp. 959–966, july-aug. 2007. doi:10.1109/JSTQE.2007.901884.

[75] A. Clawson. "Guide to references on III–V semiconductor chemical etching." *Materials Science and Engineering: R: Reports*, volume 31, no. 1–6 pp. 1–438, 2001. doi:10.1016/S0927-796X(00)00027-9.

[76] S. You. *Physics of self-quenching and self-recovering Single Photon Avalanche Detectors*. Ph.D. thesis, University of California, San Diego, 2012.

[77] R. Warburton, M. Itzler, and G. Buller. "Improved free-running InGaAs/InP single-photon avalanche diode detectors operating at room temperature." *Electronics Letters*, volume 45, no. 19 pp. 996–997, 10 2009. doi:10.1049/el.2009.1508.

[78] C. Tan, J. Ng, G. Rees, and J. David. "Statistics of Avalanche Current Buildup Time in Single-Photon Avalanche Diodes." *Selected Topics in Quantum Electronics, IEEE Journal of*, volume 13, no. 4 pp. 906–910, july-aug. 2007. doi:10.1109/JSTQE.2007.903843.

[79] S. Paul, J. B. Roy, and P. K. Basu. "Empirical expressions for the alloy composition and temperature dependence of the band gap and intrinsic carrier density in GaInAs." *Journal of Applied Physics*, volume 69, no. 2 pp. 827–829, 1991. doi:10.1063/1.348919.

[80] X. Jiang, M. Itzler, R. Ben-Michael, K. Slomkowski, M. Krainak, S. Wu, and X. Sun. "Afterpulsing Effects in Free-Running InGaAsP Single-Photon Avalanche Diodes." *Quantum Electronics, IEEE Journal of*, volume 44, no. 1 pp. 3–11, jan. 2008. doi:10.1109/JQE.2007.906996.

[81] H. C. Gerritsen, M. A. H. Asselbergs, A. V. Agronskaia, and W. G. J. H. M. Van Sark. "Fluorescence lifetime imaging in scanning microscopes: acquisition speed, photon economy and lifetime resolution." *Journal of Microscopy*, volume 206, no. 3 pp. 218–224, 2002. doi:10.1046/j.1365-2818.2002.01031.x.

[82] J. R. Lakowicz, editor. *Topics in Fluorescence Spectroscopy, Volume 1: Techniques*. Plenum Press, New York, 1991.

[83] F. Zappa, A. L. Lacaita, S. D. Cova, and P. G. Lovati. "Solid-state single-photon detectors." *Optical Engineering*, volume 35 pp. 938–945, April 1996. doi:10.1117/1.600702.

[84] A. Lacaita, P. A. Francese, F. Zappa, and S. Cova. "Single-photon detection beyond 1 um: performance of commercially available germanium photodiodes." *Appl. Opt.*, volume 33, no. 30 pp. 6902–6918, Oct 1994. doi:10.1364/AO.33.006902.

[85] M. A. Albota and F. C. Wong. "Efficient single-photon counting at 1.55 μm by means of frequency upconversion." *Opt. Lett.*, volume 29, no. 13 pp. 1449–1451, Jul 2004. doi:10.1364/OL.29.001449.

[86] V. Bessolov and M. Lebedev. "Chalcogenide passivation of III–V semiconductor surfaces." *Semiconductors*, volume 32 pp. 1141–1156, 1998. doi:10.1134/1.1187580.

[87] D. Schmidt and D. Trommer. "Conservation of low dark current of InGaAs photodiodes after NH3/HF etch with a BCB passivation layer." In *Indium Phosphide and Related Materials, 2000. Conference Proceedings.*, pp. 302–305, 2000. doi:10.1109/ICIPRM.2000.850292.

[88] H. Kim, J. Choi, H. Bang, Y. Jee, S. Yun, J. Burm, M. Kim, and A. Choo. "Dark current reduction in APD with BCB passivation." *Electronics Letters*, volume 37, no. 7 pp. 455–457, mar 2001. doi:10.1049/el:20010318.

[89] R. Yeats and K. V. Dessonneck. "Polyimide passivation of InGaAs, InP, and InGaAsP/InP p-n junction structures." *Applied Physics Letters*, volume 44, no. 1 pp. 145–147, 1984. doi:10.1063/1.94583.

[90] H. Tang, X. Wu, Q. Xu, H. Liu, K. Zhang, Y. Wang, X. He, X. Li, and H. M. Gong. "Effect of sulfur passivation on the InP surface prior to plasma-enhanced chemical vapor deposition of SiNx." *Semiconductor Science and Technology*, volume 23, no. 3 p. 035031, 2008. doi:10.1088/0268-1242/23/3/035031.

[91] S. Maimon and G. W. Wicks. "nBn detector, an infrared detector with reduced dark current and higher operating temperature." *Applied Physics Letters*, volume 89, no. 15 151109, 2006. doi:10.1063/1.2360235.

[92] H. S. Kim, E. Plis, J. B. Rodriguez, G. D. Bishop, Y. D. Sharma, L. R. Dawson, S. Krishna, J. Bundas, R. Cook, D. Burrows, R. Dennis, K. Patnaude, A. Reisinger, and M. Sundaram. "Mid-IR focal plane array based on type-II InAs/GaSb strain layer superlattice detector with nBn design." *Applied Physics Letters*, volume 92, no. 18 183502, 2008. doi:10.1063/1.2920764.

[93] R. E. Warburton, M. Itzler, and G. S. Buller. "Free-running, room temperature operation of an ingaas/inp single-photon avalanche diode." *Applied Physics Letters*, volume 94, no. 7 071116, 2009. doi:10.1063/1.3079668.