UNIVERSITY OF CALIFORNIA SAN DIEGO

Integrated Pixel Technologies for Photon Counting and SWIR Imaging in CMOS Processes

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

in

Electrical Engineering (Photonics)

by

Yunrui Jiang

Committee in charge:

Professor Yu-Hwa Lo, Chair
Professor Kenji Nomura, Co-Chair
Professor Tse Nga (Tina) Ng
Professor Andrea Tao
Professor Paul Yu

2025

The Dissertation of Yunrui Jiang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

DEDICATION

To all my family and loved ones.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would first like to express my deepest gratitude to Prof. Yu-Hwa Lo for giving me the opportunity to join his group and for welcoming me into his research team at a time when doing so required a significant five-year funding commitment. I am truly grateful for the trust he placed in me and for his willingness to support my long-term development from the very beginning. Throughout my doctoral studies, his guidance, high standards, and scientific vision shaped the way I learned to approach research problems with clarity and rigor. He consistently encouraged me to think independently, explore new ideas, and push beyond what I initially believed I could achieve. His patience, generosity, and steady support carried me through both the exciting moments of discovery and the challenging periods of uncertainty. I am sincerely thankful for the mentorship and opportunities he provided, and for the confidence he showed in my potential during every stage of this journey.

I am equally grateful to Prof. Kenji Nomura, my co-advisor, whose expertise and mentorship have played a significant role in shaping this dissertation. His deep knowledge of thin-film transistors and device physics guided many of the ideas and experiments in this work. I truly appreciate his insightful feedback, his careful attention to technical details, and the inspiring scientific conversations that pushed me to refine my understanding. His support and encouragement throughout the course of my research were invaluable, and I am thankful for the opportunity to learn from him.

I am deeply thankful to my dissertation committee members: Prof. Tse Nga (Tina) Ng, Prof. Andrea Tao, and Prof. Paul Yu. Their time, expertise, and thoughtful feedback greatly strengthened this dissertation. I am grateful for the knowledge and experience they shared with me throughout this process, and for their guidance that helped keep me on the right track during the

pursuit of my Ph.D. degree. I want to express my special thanks to Prof. Longhui Zeng, Prof. Ayaskanta Sahu, Dr. Alan Davis, Dr. Shlok Paul, Yalun Tang for their insightful input into my Ph.D. research.

I am grateful to the former members of Professor Lo's research group from whom I learned a great deal. I would especially like to thank Dr. Jiayun Zhou, Dr. Alex Ce Zhang, Dr. Mohammad Abu Raihan Miah, Dr. Surya Arya, Dr. Yugang Yu, Dr. Zunming Zhang, and Brendan Schuster for their guidance and support when I first joined the group. I also extend my thanks to Hexuan Peng, Babusha Singh, and Jeemin Kang for their help and collaboration. I am deeply appreciative of the friends who supported me throughout this journey.

Finally, I am profoundly grateful to my parents, Haiying Wang and Jiang Xiangyang, for their unconditional love and constant support.

Chapter 1, in part, is a reprint of the material as it appears in Jiang, Y., Peng, H., Tang, Y., Zeng, L., Paul, S.J., Davis, A., Sahu, A., Nomura, K. and Lo, Y.H., 2025. Monolithically integrated pixel architecture with colloidal quantum dots and thin film transistors for SWIR imaging. APL Electronic Devices, 1(4). The dissertation author was the first author of this paper.

Chapter 1, in part, is a reprint of the material as it appears in Jiang, Y., Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS. IEEE Access, accepted. The dissertation author was the first author of this paper.

Chapter 1, in part, is a reprint of the material as it appears in Miah, M.A.R., Jiang, Y. and Lo, Y.H., 2021. A physics based unified circuit model for single photon and analog detector. IEEE Access, 9, pp.129571-129581. The dissertation author was the second author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Jiang, Y., Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS. IEEE Access, accepted. The dissertation author was the first author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Jiang, Y., Peng, H., Tang, Y., Zeng, L., Paul, S.J., Davis, A., Sahu, A., Nomura, K. and Lo, Y.H., 2025. Monolithically integrated pixel architecture with colloidal quantum dots and thin film transistors for SWIR imaging. APL Electronic Devices, 1(4). The dissertation author was the first author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Miah, M.A.R., Jiang, Y. and Lo, Y.H., 2021. A physics based unified circuit model for single photon and analog detector. IEEE Access, 9, pp.129571-129581. The dissertation author was the second author of this paper.

VITA

2019 Bachelor of Science in Opto-electronics, Tianjin University

2021 Master of Science in Electrical Engineering (Photonics), University of California San Diego

2025 Doctor of Philosophy in Electrical Engineering (Photonics), University of California San Diego

PUBLICATIONS

- **Jiang, Y.,** Peng, H., Tang, Y., Zeng, L., Paul, S.J., Davis, A., Sahu, A., Nomura, K. and Lo, Y.H., 2025. Monolithically integrated pixel architecture with colloidal quantum dots and thin film transistors for SWIR imaging. APL Electronic Devices, 1(4).

- **Jiang, Y.,** Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS. IEEE Access, accepted.

- **Jiang, Y.,** Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Colloidal Quantum Dot IR Imagers. In Proc. GOMACTech.

- Arya, S., **Jiang, Y.,** Jung, B.K., Tang, Y., Ng, T.N., Oh, S.J., Nomura, K. and Lo, Y.H., 2023. Understanding colloidal quantum dot device characteristics with a physical model. Nano Letters, 23(21), pp.9943-9952.

- Miah, M.A.R., **Jiang, Y.** and Lo, Y.H., 2021. A physics based unified circuit model for single photon and analog detector. IEEE Access, 9, pp.129571-129581.

- Schuster, B., Peng, H., Arya, S., **Jiang, Y.,** Miah, M.A.R., Zhou, J., Davis, A. and Lo, Y.H., 2023. CMOS-compatible self-quenched active recovery (SQUARE) single-photon detector. APL Photonics, 8(9).

- Zhou, Z., Liu, K., Wu, D., **Jiang, Y.,** Zhuo, R., Lin, P., Shi, Z., Tian, Y., Han, W., Zeng, L. and Li, X., 2024. On-chip integrated GeSe2/Si vdW heterojunction for ultraviolet-

enhanced broadband photodetection, imaging, and secure optical communication. Nano Research, 17(7), pp.6544-6549.

- Han, W., Wang, Z., Guan, S., Wei, J., **Jiang, Y.,** Zeng, L., Shen, L., Yang, D. and Wang, H., 2024. Recent advances of phase transition and ferroelectric device in two-dimensional In2Se3. Applied Physics Reviews, 11(2).

- Zhu, M., Liu, K., Wu, D., **Jiang, Y.,** Li, X., Lin, P., Shi, Z., Li, X., Ding, R., Tang, Y. and Yu, X., 2024. In-situ fabrication of on-chip 1T'-MoTe2/Ge Schottky junction photodetector for self-powered broadband infrared imaging and position sensing. Nano Research, 17(6), pp.5587-5594.

- Long, S., Zhu, Y., Hu, M., Qi, Y., **Jiang, Y.,** Liu, B. and Zhang, X., 2018. Thin-core fiber-optic biosensor for DNA hybridization detection. Optoelectronics Letters, 14(5), pp.346-349.

- Dai, S., Yang, C., Wang, Y., **Jiang, Y.** and Zeng, L., 2025. In Situ TEM Studies of Tunnel-Structured Materials for Alkali Metal-Ion Batteries. Advanced Science, 12(19), p.2500513.

- Wei, J., Xu, Y., Shen, Y., Shen, L., Wu, H., Gao, Z., Zhou, H., **Jiang, Y.,** Zeng, L., Wan, H. and Han, W., 2025. Indium-doped α-Ga 2 O 3 nanorod arrays for ultrasensitive solar-blind UV photodetector application. Journal of Materials Chemistry C, 13(24), pp.12254-12262.

ABSTRACT OF THE DISSERTATION

Integrated Pixel Technologies for Photon Counting and SWIR Imaging in CMOS Processes

by

Yunrui Jiang

Doctor of Philosophy in Electrical Engineering (Photonics)

University of California San Diego, 2025

Professor Yu-Hwa Lo, Chair
Professor Kenji Nomura, Co-Chair

Modern optical sensing increasingly relies on detector platforms that combine high sensitivity, broad spectral reach, and compatibility with large-scale CMOS fabrication. Applications in quantum information, time-resolved imaging, autonomous systems, and biomedical diagnostics require reliable sensing under extremely low photon flux as well as effective imaging in the short-wave-infrared region. This dissertation develops two separate

CMOS-compatible pixel technologies that address these needs across different operating regimes: a silicon-based avalanche device with controlled self-quenching and rapid recovery for high-throughput single-photon detection, and a monolithically integrated quantum-dot photodiode with thin-film-transistor amplification for scalable short-wave-infrared imaging. A complementary circuit-modeling framework captures the essential electrical behavior of these detectors, including avalanche dynamics, junction operation, noise processes, and voltage-mode signal transfer, and provides a predictive basis for pixel and system optimization. Together, these contributions expand the design space for integrated optical sensors and highlight how device engineering, materials processing, and circuit modeling can be combined to enable advanced photon-counting and short-wave-infrared imaging within CMOS processes.

.

Chapter 1 Introduction

Modern optical sensing is progressing rapidly in response to applications that demand high sensitivity, spectral specificity, and large-scale integration. Quantum information systems, depth sensing, biomedical imaging, and autonomous platforms increasingly require detectors capable of operating at extremely low photon flux while remaining compatible with high-volume semiconductor manufacturing. At the same time, the short-wave infrared spectral region has become important for imaging through scattering media, material identification, and eye-safe active illumination. These diverse requirements are difficult to satisfy within a single fabrication platform because no existing detector technology simultaneously provides single-photon sensitivity, broadband infrared response, and full CMOS compatibility. This dissertation addresses these challenges by developing pixel architectures for photon counting and short-wave infrared imaging that can be realized entirely within standard CMOS processes and low-temperature post-processing. The work spans device structures, pixel-level integration, and circuit-level analysis. In particular, accurate circuit modeling is essential for understanding detector behavior in an integrated environment, evaluating quenching and recovery strategies, and predicting noise and timing performance in practical pixel implementations. The following sections establish the scientific foundation for this work. They review the physics of photodetection, the principles of single-photon sensing, the operation of avalanche devices, the opportunities associated with the short-wave infrared band, the material characteristics of colloidal quantum dot photodiodes, and the role of circuit modeling in linking device physics with pixel-level design.

1.1 Photodetection

Photodetection is a fundamental process that enables the interaction between optical signals and electronic systems. At its core, photodetection converts incident photons into electrical carriers through the absorption processes governed by semiconductor band structure, density of states, carrier generation and recombination kinetics, and electric field driven charge separation. When a photon with energy greater than the bandgap is absorbed in a semiconductor, it excites an electron from the valence band to the conduction band, generating an electron hole pair. The probability that a photon is converted into mobile charge is determined by material absorption coefficients, the spatial overlap between the absorption depth and the depletion region, the presence of recombination centers, and the overall device geometry.

Silicon has historically served as the dominant platform for photodetection because it offers long minority carrier diffusion lengths, low defect densities, and excellent compatibility with planar CMOS fabrication. For visible wavelengths, the absorption coefficient of silicon is sufficiently high that photocarriers generated within a few micrometers of the surface can be collected efficiently by the depletion region of a p-n junction. This characteristic, together with the low cost of large wafer substrates and the highly refined control of doping and oxidation processes, established silicon photodiodes as the backbone of imaging systems for several decades. Silicon based imagers have benefitted significantly from innovations such as pinned photodiodes, correlated double sampling, deep trench isolation, and charge transfer techniques that reduce read noise and improve uniformity across large arrays.

Despite this progress, conventional silicon photodiodes face fundamental physical limitations that constrain their performance in advanced applications. As illumination intensity decreases to extremely low levels, the analog photo charge generated during an integration period

becomes comparable to or even smaller than noise sources that include thermal dark current, reset noise, and amplifier noise. In such regimes, analog detection becomes unreliable and the problem of sensing light reduces to the detection of discrete photon arrivals. For wavelengths beyond approximately 1000 nm, silicon becomes effectively transparent because the photon energy falls below the bandgap. This property makes it impossible for silicon photodiodes to detect light in the short-wave infrared spectral region, regardless of device geometry. These limitations motivate the development of new types of detectors that either amplify the signal generated by a single photon or employ materials with narrower bandgaps that are sensitive in the infrared.

The evolution of photodetector technology has therefore progressed in two complementary directions. The first direction involves the enhancement of sensitivity through internal gain mechanisms. Avalanche photodiodes exploit high electric fields to induce impact ionization, which multiplies photogenerated carriers and produces detectable current pulses even when only a single photon is absorbed. Although these devices provide very high sensitivity, they require careful control of breakdown phenomena, quenching dynamics, and excess noise associated with multiplication. The second direction involves the exploration of new material systems, including III-V semiconductors, two dimensional materials, perovskites, and colloidal quantum dots. These materials offer bandgaps that can be tuned through composition, thickness, or quantum confinement to enable detection beyond the visible spectrum. Many of these materials can be integrated with CMOS back-end processes, allowing the construction of hybrid or monolithic imaging systems that combine novel absorbers with advanced readout circuitry.

In parallel, modern imaging increasingly relies on integrated pixel architectures in which analog, mixed signal, and sometimes digital circuitry is co-located with each photodetector. This integration allows for low noise amplification, active reset, dynamic gain control, charge domain

and voltage domain readout, time stamping, and even primitive on pixel computation. The complexity of these architectures has increased over time as performance expectations have grown. Applications such as quantum imaging, time of flight depth sensing, fluorescence lifetime imaging, and SWIR imaging place stringent demands on sensitivity, timing accuracy, dark noise, and dynamic range. Meeting these demands requires solutions that originate not only from materials science but also from semiconductor device engineering and circuit design.

The work presented in this dissertation sits at the intersection of these developments. It focuses on the creation of advanced integrated pixel technologies that enable photon counting and SWIR imaging within standard CMOS processes. The first part of the thesis investigates a new type of silicon avalanche device that provides rapid self-quenching and active-recovery for high throughput single photon detection. The second part of the thesis investigates the integration of colloidal quantum dot photodiodes with IGZO thin film transistors and CMOS readout structures to create low-cost, high-performance SWIR pixels. These two threads share a common goal, which is the development of scalable, CMOS compatible photodetection solutions that meet the sensitivity and spectral requirements of next generation imaging systems.

1.2 Single Photon Detection

Single photon detection represents the fundamental sensitivity limit of photodetection and is governed by the discrete and stochastic nature of light. When optical signals reach extremely low intensity, averaging based on continuous photocurrent is no longer possible because only a few photon events occur during the relevant temporal window. In these conditions, the information in an optical signal is encoded not in an analog current but in a sequence of probabilistic photon

arrival times. Detecting these individual quanta requires devices that can convert a single absorbed photon into a measurable electrical transition that exceeds all noise sources present in the system.

The physics of single photon detection cannot be understood without considering the quantum statistics of light. In many applications, such as time correlated single photon counting or quantum communication, the arrival times of photons carry essential information. The timing resolution required for these applications often reaches tens of picoseconds, which demands extremely fast avalanche buildup and low timing jitter in the detector and readout circuitry. Furthermore, the Poissonian nature of photon arrivals imposes statistical limits on the achievable signal to noise ratio. A single photon detector must therefore not only be sensitive to individual quanta but must also preserve temporal information with minimal distortion from electronic noise, avalanche variability, or carrier trapping.

The need for single photon sensitivity spans a wide range of application domains. In time-of-flight LiDAR, the emission of short optical pulses and the detection of weak return signals require detectors that can register very sparse photon returns with high timing precision. In fluorescence lifetime imaging, the radiative decay of excited fluorophores is typically measured using time tagged photon arrivals, which makes timing jitter and afterpulsing critical device parameters. In astronomy and space communication, the ability to count individual photons allows the detection of extremely weak sources, such as distant stars or long-range optical communication channels. In quantum optics and quantum information systems, single photon detection is indispensable for qubit readout, entanglement verification, and secure communication protocols.

Achieving reliable single photon detection requires internal signal amplification on the order of several hundred thousand to several million carriers per absorbed photon. Traditional photodiodes cannot satisfy this requirement because their photocurrent is proportional to the

number of photo generated carriers and is therefore limited to a few tens of electrons for extremely low illumination levels. Superconducting nanowire detectors provide exceptionally high internal gain and low timing jitter. However, they require cryogenic temperatures and cannot be integrated monolithically with CMOS electronics. Photomultiplier tubes also provide high sensitivity but are bulky, fragile, and incompatible with semiconductor processes.

These limitations have made avalanche-based silicon detectors the most practical approach for single photon detection in CMOS processes. By operating a p-n or p-i-n junction above its breakdown voltage, the absorption of a single photon can trigger a self-sustaining avalanche through impact ionization. The resulting current pulse can be detected by on pixel circuitry, allowing direct time stamping of photon events with high temporal resolution. Although this approach solves the fundamental sensitivity problem, it introduces new challenges related to breakdown physics, quenching dynamics, correlated noise, and power consumption. The need to manage these characteristics has shaped the evolution of single photon avalanche diode arrays and has driven the development of new device architectures and circuit techniques.

This dissertation addresses these challenges by developing a single photon detection platform that incorporates a self-quenching mechanism and an actively controlled recovery path, achieved through a novel device structure that is compatible with standard CMOS design rules. The architecture reduces dead time, improves pulse shape consistency, and enables high throughput photon counting in dense arrays. The physical principles, device modeling, and experimental results associated with this platform are detailed in the first half of this thesis.

1.3 Single Photon Avalanche Diodes

Single Photon Avalanche Diodes, often abbreviated as SPADs, are the most widely adopted solid state detectors for photon counting in CMOS compatible processes. Their operation relies on the controlled initiation of an impact ionization cascade within a reverse biased p-n or p-i-n junction that is biased above its breakdown voltage. In this regime, referred to as Geiger mode operation, a single primary carrier generated by a photon is sufficient to trigger a self-sustaining avalanche. This avalanche produces a macroscopic current pulse that can be sensed easily by CMOS circuitry without analog amplification. The ability of the SPAD to convert a single absorbed photon into a distinctive, high amplitude pulse enables the direct detection of the arrival time of individual photons, which is essential for applications that require high timing precision.

The physical processes that govern SPAD operation are complex and involve multiple carrier transport, scattering, and multiplication mechanisms. The onset of avalanche multiplication occurs when the electric field inside the multiplication region exceeds a critical value, typically on the order of several hundred kilovolts per centimeter. At these high fields, carriers gain sufficient kinetic energy between collisions to create additional electron hole pairs through impact ionization. The rate of impact ionization is often described by ionization coefficients that follow an exponential dependence on the electric field. Because holes and electrons have different scattering behaviors and different ionization coefficients, the symmetry of the avalanche multiplication depends strongly on the doping profile and the orientation of the electric field. Device engineers therefore carefully design the depletion region to produce the desired balance between electron initiated and hole initiated breakdown, which in turn affects timing jitter, avalanche buildup time, and the probability that an absorbed photon triggers an avalanche.

Avalanche initiation is inherently stochastic. The time required for the avalanche current to reach a detectable level varies from event to event due to random fluctuations in carrier trajectories, scattering paths, and the spatial distribution of the first ionization event. These fluctuations contribute to timing jitter, which is one of the most important performance metrics for time resolved imaging. The avalanche buildup time is also influenced by the electric field distribution inside the device. A well confined high field region produces a rapid and uniform multiplication process that reduces jitter and improves the consistency of the output pulse. Silicon junctions created using standard CMOS processes often have relatively shallow depletion regions that restrict the geometries available for optimizing avalanche buildup. This places constraints on the achievable performance and motivates the exploration of modified device structures that introduce engineered multiplication regions or auxiliary field shaping layers.

A critical aspect of SPAD operation is the quenching of the avalanche. Once triggered, the avalanche is self-sustaining because the high electric field continues to accelerate carriers, creating more impact ionization events. To prevent permanent breakdown and thermal damage, the avalanche must be interrupted by reducing the current through the device below a sustaining level or by reducing the bias below the breakdown voltage. Passive quenching relies on a series resistor and the natural voltage drop that occurs as the avalanche current increases. Although this approach is simple, it results in slow recharge dynamics and limited control over dead time, which is the interval during which the detector is unable to detect another photon. Active quenching circuits can interrupt the avalanche rapidly using feedback controlled current reduction or voltage pulldown mechanisms. These circuits allow much shorter and more predictable dead time but require additional circuitry at the pixel level, which consumes area and power.

SPAD performance is strongly influenced by several forms of noise and correlated spurious events. The dark count rate arises from thermally generated carriers, trap assisted tunneling, and band to band tunneling, all of which can initiate avalanches in the absence of photons. In small geometry SPADs, the high electric fields required for Geiger mode operation cause tunneling to dominate and can produce extremely high dark count rates if the electric field profile is not carefully controlled. Another major source of noise is afterpulsing. During avalanche conduction, large numbers of carriers flow through the device and may become trapped in deep level defects. These trapped carriers can later be released with a time constant determined by trap energy levels and temperature, potentially initiating a new avalanche shortly after the original one. Afterpulsing introduces false counts that are correlated with prior events, which complicates applications that depend on precise inter photon timing. The prevalence of afterpulsing is influenced by the total avalanche charge that flows during each event, which is why many high performance SPAD architectures attempt to minimize avalanche charge by using rapid and controlled quenching.

Integration of SPADs within CMOS processes introduces additional challenges. Standard CMOS technologies were not originally intended to support high electric field avalanche regions, which means that designers must work within the constraints of limited junction depths, fixed doping profiles, and restricted isolation structures. Shallow junctions can lead to high tunneling induced dark counts, while lateral electric fields can cause premature edge breakdown if guard rings are not carefully designed. Optical crosstalk becomes problematic in dense arrays because avalanches emit secondary photons that can propagate into neighboring pixels and trigger correlated events. Electrical crosstalk through shared supply lines or substrate coupling can also create spurious signals. To mitigate these effects, SPAD arrays often employ deep trench isolation, guard ring structures, optical absorbers, or custom process modifications. However, not all

fabrication lines allow such modifications, which motivates the investigation of SPAD architectures that can be realized entirely with standard design rules.

The work introduced in this thesis is motivated by the need to overcome key limitations of traditional SPAD architectures in CMOS processes. In conventional designs, the avalanche pulse shape and recharge behavior are strongly coupled to the quenching element, resulting in variable dead time, long recharge intervals, and reduced temporal precision. To address these challenges, the dissertation presents a self-quenching and actively recovering SPAD structure that incorporates a high-speed recovery path built directly into the device. This architecture enables rapid restoration of the bias voltage after an avalanche and produces more uniform pulse amplitude across a wide range of event conditions. The improved temporal response and reduced dead time allow higher throughput photon counting and enhance the suitability of SPAD arrays for high dynamic range and high frame rate sensing applications. The detailed device analysis, circuit modeling, and experimental validation of this architecture form a major portion of the first half of the dissertation.

1.3 Short Wave Infrared Imaging

Short wave infrared imaging has emerged as an essential modality in many scientific, industrial, and consumer applications because it offers optical and material advantages that are not available in the visible spectrum or in the mid and long wave infrared ranges. The SWIR region is typically defined from approximately 1000 nanometers to 2500 nanometers. Photons in this spectral band possess energies that are sufficiently high to interact with electronic transitions in narrow bandgap semiconductors while still being low enough to penetrate through many scattering and absorptive media that strongly attenuate visible light. This combination creates a unique

balance between optical transparency, material compatibility, detector sensitivity, and safely accessible illumination sources.

One of the primary reasons why SWIR imaging is attractive lies in its interaction with the atmosphere and common optical materials. Atmospheric scattering decreases significantly in the SWIR band because the scattering cross section for most aerosols and water droplets follows an inverse fourth power dependence on wavelength. As a result, SWIR illumination experiences less attenuation in fog, haze, smoke, and other scattering environments. This property enables scene visibility and depth sensing in conditions where visible light based imaging fails. The SWIR spectrum also includes several atmospheric transmission windows, such as the region around 1550 nanometers, where absorption by water vapor is minimal. These windows are widely used in telecommunications because they allow low loss propagation through optical fibers and free space. Imaging systems that operate in these windows can exploit high power eye safe illumination sources that are already well developed for communication applications.

Biological tissue exhibits reduced scattering in the SWIR region as well. Water absorption dominates in certain parts of the spectrum, but several sub ranges provide deeper penetration than visible or near infrared light. This property has motivated the use of SWIR imaging in biomedical applications such as tissue inspection, vascular imaging, and fluorescence imaging with specialized contrast agents. Industrial inspection systems also benefit from SWIR imaging because many materials that appear identical in visible light exhibit distinct absorption or reflection characteristics in the SWIR range. For example, polymers, pharmaceuticals, agricultural products, and silicon wafers display spectral features in the SWIR domain that allow sorting, quality control, and nondestructive structural analysis.

Despite these advantages, SWIR imaging has not achieved the same level of commercial ubiquity as visible imaging. The primary barrier is the lack of a detector technology that is both high performing and compatible with low cost, high volume semiconductor manufacturing. State of the art SWIR detectors are based mainly on III V semiconductors such as indium gallium arsenide grown on indium phosphide substrates. These materials possess direct bandgaps that naturally cover the SWIR range and exhibit high quantum efficiency, low noise, and fast response. However, their fabrication requires epitaxial growth on expensive lattice matched substrates, which limits wafer size and increases cost. Because III V devices are not compatible with standard CMOS fabrication, they must be hybridized to a silicon readout integrated circuit using indium bump bonding or advanced hybrid bonding. This process introduces significant constraints on pixel pitch, alignment accuracy, and yield. It also increases thermal mismatch related stress, limits scalability, and results in per pixel costs that are orders of magnitude higher than those of visible CMOS image sensors.

These limitations are particularly important for applications that require megapixel resolution, small pixel pitch, or dense arrays integrated with computational electronics. As pixel pitch decreases, the number of bumps required for hybridization increases dramatically, and the probability of bonding defects rises. The capacitance introduced by the bump interconnect and the need for high current driving capabilities in the readout circuit restrict overall noise performance and dynamic range. Thermal and mechanical constraints also limit the possibility of large area sensors. These factors create a strong motivation to explore detector materials that can be integrated directly on top of CMOS wafers through processes that do not rely on lattice matching or high temperature epitaxy.

Several alternative approaches have been explored in recent years. Germanium and germanium tin alloys provide a narrower bandgap than silicon and can potentially extend absorption into the short-wave infrared. However, their integration remains challenging because the lattice mismatch between germanium-based materials and silicon introduces dislocations that increase dark current and degrade responsivity. Two dimensional materials, such as transition metal dichalcogenides, offer strong optical absorption per unit thickness and can be tuned through band structure engineering. However, their lateral transport characteristics and scalability pose practical challenges for large area imaging.

A particularly promising direction arises from solution processed semiconductors, especially colloidal quantum dots. These materials benefit from quantum confinement, which allows the bandgap to be tuned by adjusting the size of the nanocrystals. They can be synthesized in solution with precise control over surface ligands, doping, and electronic properties. More importantly, they can be deposited on large wafers at low temperature using techniques such as spin coating or inkjet printing. This allows monolithic integration with CMOS readout circuits without the need for specialized high temperature or lattice matched epitaxial growth. Quantum dot-based SWIR detectors have demonstrated substantial improvements in detectivity, responsivity, and stability over the past decade and are now sufficiently advanced to serve as viable alternatives to III V technologies for many applications.

However, the integration of such materials introduces new complexities that arise from their fundamentally different charge transport mechanisms. Carrier motion occurs through hopping between adjacent quantum dots, which leads to mobility values that are far lower than those of crystalline semiconductors. This property introduces tradeoffs between junction thickness, responsivity, transit time, and bandwidth. Surface ligands and trap states strongly influence dark

current and noise. Device characteristics depend critically on film uniformity, energetic disorder, and built-in potential. These constraints place significant demands on the design of the readout circuitry, particularly in relation to noise, pixel capacitance, charge integration dynamics, and voltage domain stability.

For these reasons, the development of a SWIR imaging solution that achieves high detectivity, low noise, and large-scale manufacturability must involve coordinated advances in material science, device engineering, and pixel level circuit design. The second half of this dissertation is devoted to addressing these challenges through the development of a fully integrated pixel architecture that combines a colloidal quantum dot photodiode with a thin film transistor based amplifier and a CMOS backplane. The combination allows analog voltage domain sensing with enhanced gain, reduced noise, and improved compatibility with the electrical characteristics of quantum dot photodiodes. The architecture supports high sensitivity, low dark noise, and scalable fabrication, which together form the foundation for cost effective SWIR imaging systems that can operate within standard CMOS environments.

1.5 Colloidal Quantum Dots

Colloidal quantum dots have emerged as one of the most promising material platforms for short wave infrared photodetection because they offer size tunability, solution based fabrication, and direct compatibility with CMOS back end processes. These nanocrystals, typically composed of lead chalcogenides such as lead sulfide or lead selenide, exhibit electronic and optical properties that are dominated by quantum confinement rather than by the bulk semiconductor band structure. When the radius of the nanocrystal approaches the exciton Bohr radius of the material, the energy levels become quantized, and the effective bandgap increases as the crystal size decreases. This

phenomenon allows fine control of the absorption edge by adjusting nanocrystal diameter during synthesis. As a result, quantum dots can be engineered to detect light from the visible range through the near infrared and into the short-wave infrared.

The density of states in a quantum dot film differs significantly from that of a crystalline semiconductor. Each dot contributes discrete energy levels that broaden into quasi continuum bands when many dots are coupled through ligand mediated electronic interactions. The presence of size distribution, surface defects, and variations in ligand coverage introduces energetic disorder that broadens these bands and creates localized tail states. Charge carriers move through the film predominantly by thermally assisted hopping between neighboring nanocrystals. The hopping rate is determined by wavefunction overlap, inter dot spacing, ligand length, and the energy difference between sites. Several models have been developed to describe transport in these films, such as nearest neighbor hopping and variable range hopping. Regardless of the model, the transport mechanism leads to mobility values that are typically in the range of 10^(-4) to 10^(-2) square centimeters per volt second. These values are several orders of magnitude lower than those of crystalline silicon or III-V materials.

Because carrier mobility is low, the absorber thickness must be carefully optimized. A thicker film increases optical absorption and responsivity but reduces charge collection efficiency because carriers must traverse a longer path through a hopping dominated transport network. A thinner film improves extraction efficiency and reduces transit time but sacrifices quantum efficiency at longer wavelengths. The optimum thickness is therefore a function of absorption length, built in potential, trap density, and readout noise. Engineering the photodiode structure often involves creating a p-i-n or p-n junction using selective ligand exchange to form regions with different doping characteristics. For example, surface treatments that expose lead rich facets create

n type behavior, while iodide or halide-based ligands can produce p type or intrinsic behavior. The heterojunction must establish a strong built-in electric field to assist in carrier extraction while minimizing diffusion driven leakage.

The formation of the depletion region and the resulting band profile are strongly influenced by surface chemistry. Ligands not only determine inter dot spacing but also passivate surface states that would otherwise act as recombination centers or generate mid gap trap states. Incomplete or unstable passivation can lead to high dark current because trap assisted tunneling or thermally activated hopping pathways become available. Dark current in quantum dot photodiodes arises from several mechanisms, including thermal generation in deep traps, band to band diffusion, and surface leakage along sidewalls or interfaces. The dependence of dark current on temperature often reveals the activation energies associated with these mechanisms. A well-designed device suppresses these leakage currents through careful choice of ligand systems, stoichiometric tuning of the quantum dots, and optimized deposition of buffer layers that provide conformal passivation.

Another major factor that affects device performance is film uniformity. Because quantum dots are solution processed, the formation of a continuous and uniform film requires precise control of solvent evaporation, ligand removal, and inter layer diffusion. Variations in nanocrystal concentration or ligand density can cause spatial fluctuations in local bandgap, trap density, or mobility. These fluctuations directly impact pixel to pixel uniformity in imaging arrays. Furthermore, the formation of cracks or non-uniform thickness during deposition can create high conductance pathways that significantly increase dark current. The optimization of coating methods, such as spin coating, blade coating, or spray deposition, plays a critical role in achieving reproducible performance.

Charge transport across the junction is also influenced by the alignment of energy levels at interfaces with transport layers or contact electrodes. For instance, electron extraction may be hindered by an unfavorable conduction band offset, which increases the likelihood of recombination before collection. Similarly, hole extraction can be limited by barriers at the valence band interface. Device engineers often employ metal oxide transport layers, such as zinc oxide or titanium oxide for electron transport, and materials such as nickel oxide for hole transport. The goal is to create selective contacts that enhance unidirectional carrier flow while minimizing leakage. However, these oxides must be deposited at low temperature to preserve the integrity of the quantum dot film and the underlying CMOS structures.

In addition to these material and device level considerations, the integration of quantum dot photodiodes with CMOS readout circuits introduces new constraints. Quantum dot photodiodes often exhibit higher dark current, larger junction capacitance, and lower dynamic resistance compared to silicon photodiodes. These characteristics affect the design of the readout interface. For example, the conventional three transistor pixel architecture used in CMOS imagers is not well suited for quantum dot detectors because its noise performance and gain characteristics were originally optimized for low dark current silicon diodes. Voltage domain readout architectures that employ thin film transistors as local amplifiers are often more compatible with quantum dot photodiodes because they provide additional gain, improved impedance matching, and more flexible control over reset and integration behavior.

Thin film transistors based on amorphous oxide semiconductors, such as indium gallium zinc oxide, offer high transconductance at low leakage levels and can be processed at temperatures that are compatible with quantum dot films. Incorporating such transistors as part of the pixel level architecture creates a hybrid system in which the quantum dot photodiode provides the optical

absorption while the thin film transistor provides analog amplification and buffering. This combination allows reduced noise, improved signal to voltage conversion, and greater flexibility in pixel design. The integration strategy must ensure that the deposition of the quantum dot layer does not degrade the performance of the thin film transistor, and vice versa.

Noise behavior in quantum dot photodiodes is a subject of considerable interest. Because transport is mediated by thermally activated hopping, the low frequency noise spectrum often exhibits significant 1/f components, which can degrade imaging performance at low signal levels. Shot noise associated with the dark current, thermal generation noise in trap states, and capacitive kTC noise at the readout interface must all be considered. Careful control of trap densities, contact resistances, and readout bandwidth is essential for achieving high detectivity.

These scientific and engineering challenges motivate the development of customized pixel architectures that are tailored specifically to the electrical behavior of quantum dot photodiodes. The second half of this dissertation presents a complete pixel design that integrates an optimized quantum dot absorber, a thin film transistor amplifier, and a CMOS backplane. The design supports voltage mode sensing with enhanced gain and reduced noise, enabling high sensitivity in the SWIR band while maintaining compatibility with low temperature fabrication and large area manufacturing. Detailed characterization of the optical response, dark current, noise spectrum, and pixel level transfer behavior demonstrates the feasibility of this approach for future large scale SWIR imaging systems.

1.6 Circuit Modeling for Advanced Photodetectors

Modern imaging systems increasingly rely on pixel architectures that combine complex device physics with mixed-signal readout circuitry. As detectors evolve to include avalanche

multiplication, voltage-mode sensing, heterogeneous material stacks, and nonstandard gain mechanisms, there is a growing need for circuit models that can translate these physical processes into forms that can be simulated and optimized at the system level. Device behavior such as avalanche initiation, quenching, charge transport, junction dynamics, and stochastic photon arrivals strongly influences timing accuracy, noise performance, and pixel-to-pixel uniformity. These effects must be represented in a way that is compatible with standard circuit design tools so that the interactions between the detector and the surrounding readout electronics can be understood and optimized.

Circuit-level modeling serves as the bridge between device physics and pixel design. It allows designers to explore quenching conditions, bias strategies, dynamic node behavior, and signal integrity without relying solely on experimental prototypes. Accurate models are especially important for integrated pixels because the detector and readout circuitry influence each other through shared capacitances, impedances, and transient charge flow. A physics-informed model makes it possible to evaluate sensitivity, noise, dynamic range, and timing characteristics under realistic operating conditions.

The final part of this dissertation introduces a circuit modeling framework that captures the essential electrical behavior of the detectors studied in this work. The goal is not to replicate every microscopic detail, but to provide a compact and flexible description that reflects the dominant physical mechanisms and allows full pixel and system simulations. This framework supports the analysis of emerging pixel architectures, enables comparison across detector types, and provides insight into how device properties translate into circuit-level performance.

ACKNOWLEDGEMENTS

Chapter 1, in part, is a reprint of the material as it appears in Jiang, Y., Peng, H., Tang, Y., Zeng, L., Paul, S.J., Davis, A., Sahu, A., Nomura, K. and Lo, Y.H., 2025. Monolithically integrated pixel architecture with colloidal quantum dots and thin film transistors for SWIR imaging. APL Electronic Devices, 1(4). The dissertation author was the first author of this paper.

Chapter 1, in part, is a reprint of the material as it appears in Jiang, Y., Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS. IEEE Access, accepted. The dissertation author was the first author of this paper.

Chapter 1, in part, is a reprint of the material as it appears in Miah, M.A.R., Jiang, Y. and Lo, Y.H., 2021. A physics based unified circuit model for single photon and analog detector. IEEE Access, 9, pp.129571-129581. The dissertation author was the second author of this paper.

Chapter 2 Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS

Single-photon avalanche diodes (SPADs) are used in applications such as quantum communication, LiDAR, low-light imaging, and biosensing. However, conventional CMOS SPADs face a fundamental trade-off: passive quenching results in long RC recovery times, while active quenching leads to a reduced fill factor and increased complexity. We presented a self-quenched SPAD with an active self-recovery mechanism that can substantially accelerate the self-recovery process following self-quenching, making it a self-quenched, actively recovered (SQUARE) SPAD. Fabricated in a standard CMOS process, the SQUARE detector integrates a carbon-incorporated amorphous silicon (a-Si) layer above a crystalline silicon p/n junction, attaining Geiger-mode gains exceeding 10^5 with self-recovery below 30 ns. In comparison, the self-quenched, passively recovered SPAD of a similar structure but using a defected a-Si shows the same Geiger-mode gain but a self-recovery time of >1us. The result demonstrates 50 times faster active recovery due to the gain of a-Si layer during the recovery cycle. To facilitate adoption of the device in applications using commercial CMOS process, we developed a theoretical framework and a Verilog-A circuit model that accurately capture avalanche dynamics and practical foundry constraints. Single-photon imaging demonstrations under low-light conditions confirm the advantage of enhanced speed and sensitivity of the SQUARE detector, positioning it as a scalable, cost-effective solution for single-photon imaging and sensing applications.

2.1 Introduction

Single-photon detection is crucial for a broad range of emerging technologies [1], spanning quantum communication [2], [3], LiDAR [4], [5], imaging [6], and spectroscopy [7], [8]. By enabling the detection of extremely low light levels with high fidelity, single-photon detectors

have radically expanded our capacity to probe and manipulate light-matter interactions. [9] While photomultiplier tubes (PMTs) [10], [11] and superconducting nanowire single-photon detectors (SNSPDs) [12], [13] offer excellent sensitivity and timing characteristics, their reliance on high voltage, vacuum environments, or cryogenic cooling restricts their scalability and cost-effectiveness, particularly in size and power limited settings.

In contrast, complementary metal–oxide–semiconductor (CMOS) single-photon avalanche diodes (SPADs) represent a compelling alternative [14]–[16]. Fabricated using mature and widely available semiconductor processes, CMOS SPADs can be integrated into high-density arrays at a relatively low cost and operate at room temperature [17]–[19]. These features are especially advantageous for applications demanding compact form factors and real-time photon counting under low-light conditions. However, conventional CMOS SPAD implementations typically require external passive [20], [21] or active [22]–[26] quenching circuits to terminate the avalanche triggered by a single photon—an approach that increases system complexity and diminishes the effective fill factor in densely packed array designs.

To address these limitations, a lot of research has focused on developing self-quenching SPAD architectures that integrate quenching functionality directly within the detector [27]–[29]. Self-quenched SPADs use either an integrated (non-linear) resistor or a transient carrier barrier (TCB) via bandgap engineering to quench the photo response triggered by single photons. While the quenching speed is fast, resulting in a narrow (100 – 200 ps) single-photon current pulse, but the self-recovery time has been long, limited by either the RC response or carrier transport across an energy barrier in a bandgap engineered design [30]. To overcome the limitations of self-quenched SPADs, we introduced an active recovery mechanism in a self-quenched SPAD to significantly reduces the self-recovering time [31]. Our design includes a secondary gain region,

formed by depositing carbon incorporated (a-Si:H:C) amorphous silicon (a-Si) via in-situ PECVD on top of the top metal layer of a foundry-fabricated CMOS wafer. Together with a conventional silicon p/n junction, this secondary gain region produces carrier multiplication during the recovering cycle, shortening the charging time to restore the bias voltage across the primary gain region. Building on our previously introduced SQUARE concept (self-quenched active-recovery) [31], this work demonstrates a 20-30 ns circuit recovery time, roughly 50 times faster than passive self-quenched SPADs without recovery gain, while maintaining stable Geiger operation.

The central elements of this work are experimental demonstration of the gain produced by a-Si to accelerate self-recovery process and development of a compact device circuit model of the SQUARE detector implemented in Cadence Virtuoso (Verilog-A) [32]–[34]. The model captures avalanche onset, self-quench, and active-recovery dynamics under foundry-specific constraints, explicitly representing the gain- assisted recharge pathway. It enables system co-design by predicting operating points and transients, thereby guiding parameter choices within the CMOS node. Using this framework, we designed a transimpedance amplifier (TIA) readout circuit tailored to SQUARE [35], [36], achieving voltage mode detection while minimizing parasitics to support accurate single-photon sensing at high count rates.

To demonstrate the viability of the device, we implement a single-pixel imaging system to use the output of SQUARE detector to reconstruct an image under low-light conditions. This demonstration effectively highlights the key advantages of fast recovery, providing a strong foundation for future multi-pixel or array-based implementations. In summary, the development of a CMOS SPAD featuring a self-quenched active recovery mechanism, supported by a circuit model and integrated with a TIA readout circuit, marks a significant step forward in single-photon detection technology.

2.2 Methods and Experimental Results

The device structure is illustrated in Fig. 2.1(a). The device integrates a dual gain architecture to optimize the performance of a single-photon avalanche diode by combining a conventional silicon p/n junction with a secondary gain medium. The conventional silicon (Si) p/n junction, fabricated with Tower Semiconductor's 180 nm SBC18H3 CMOS process, serves as the photon detection and carrier multiplication element, leveraging impact ionization to initiate an avalanche effect upon photon absorption. The secondary gain region, consisting of a 100 nm thick film of 5% carbon-incorporated amorphous silicon (a-Si), is deposited directly on the top metal layer of the CMOS wafer using plasma-enhanced chemical vapor deposition (PECVD) at 270°C. This a-Si layer not only provides additional gain during the recovery phase but also creates an electronic transport barrier through the band offsets between aluminum and a-Si. The barrier and the secondary gain during device recovering establish a negative feedback mechanism that stabilizes the Geiger gain and significantly accelerates the self-recovery process. The device structure is inherently simple and easy to fabricate, eliminating the need for external active quenching circuits that typically require additional area, power supply lines, and design complexity. Fig. 2.1(b) shows the top view layout of the SQUARE detector, showing the Si p–n junction region and the a-Si region. Detailed back-end-of-line integration process and deposition parameters are provided in Supplementary Section II. The fabricated device was evaluated under DC conditions using a continuous-wave 635 nm pigtailed laser diode. Fig. 2.1(c) shows the current-voltage (I-V) characteristics of the integrated device measured under dark and illuminated conditions, with the photocurrent acquired at an incident laser power of 10.6pW.

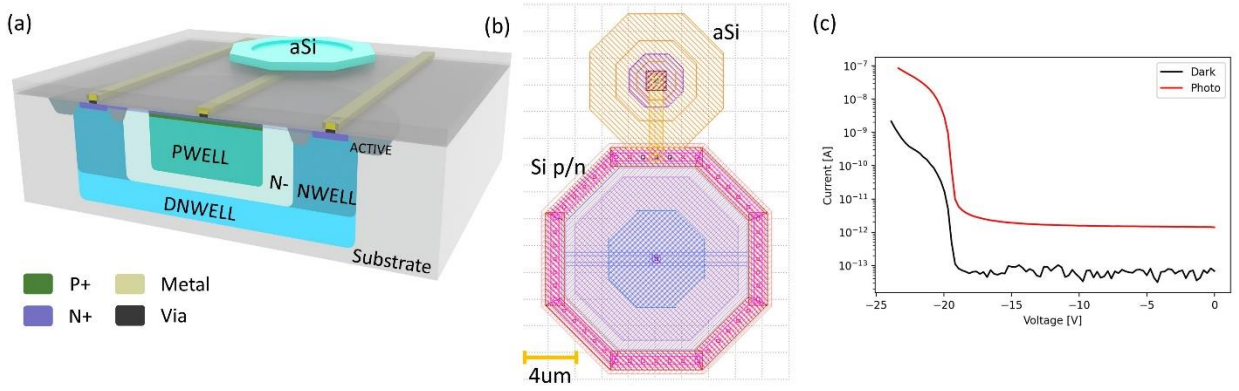

Figure 2.1 Device structure and characterization. (a) Artistic illustration of the device cross-sectional structure, showing the device stack and the contact of the a-Si layer to the underlying Si p–n junction. For illustration purpose, the figure is not drawn to scale. The carbon-incorporated a-Si mesa is deposited directly on the top metal (M6) and contacted through the M1-M6 via ladder. (b) Top-view layout of the device, showing the Si p–n junction region and the a-Si region. (c) Integrated device current–voltage (IV) characteristics under dark and illuminated conditions.

The SQUARE detector operation principle is summarized in Fig. 2.2. A photon absorbed in the Si depletion region creates a primary electron–hole pair and triggers impact-ionization gain in crystalline Si, launching carrier multiplication and producing a current surge, as shown in Fig. 2.2(a). These carriers exit the p/n junction and charge the a-Si/metal capacitance, increasing the voltage drop across the a-Si branch while reducing the over-bias of Si p/n junction. When the bias across the silicon p-n junction falls below the breakdown voltage, the impact ionization process ceases, resulting in instantaneous current quenching, as shown in Fig. 2.2(b). As the accumulated charge across the a-Si layer begins to inject into the a-Si bulk, it seeds a secondary carrier multiplication process under the high electric field (Fig. 2.2(c)). The resulting electron-hole pairs rapidly recharge the silicon p-n junction, restoring the device to its initial state and preparing it to detect subsequent photons (Fig. 2.2(d)) [31].

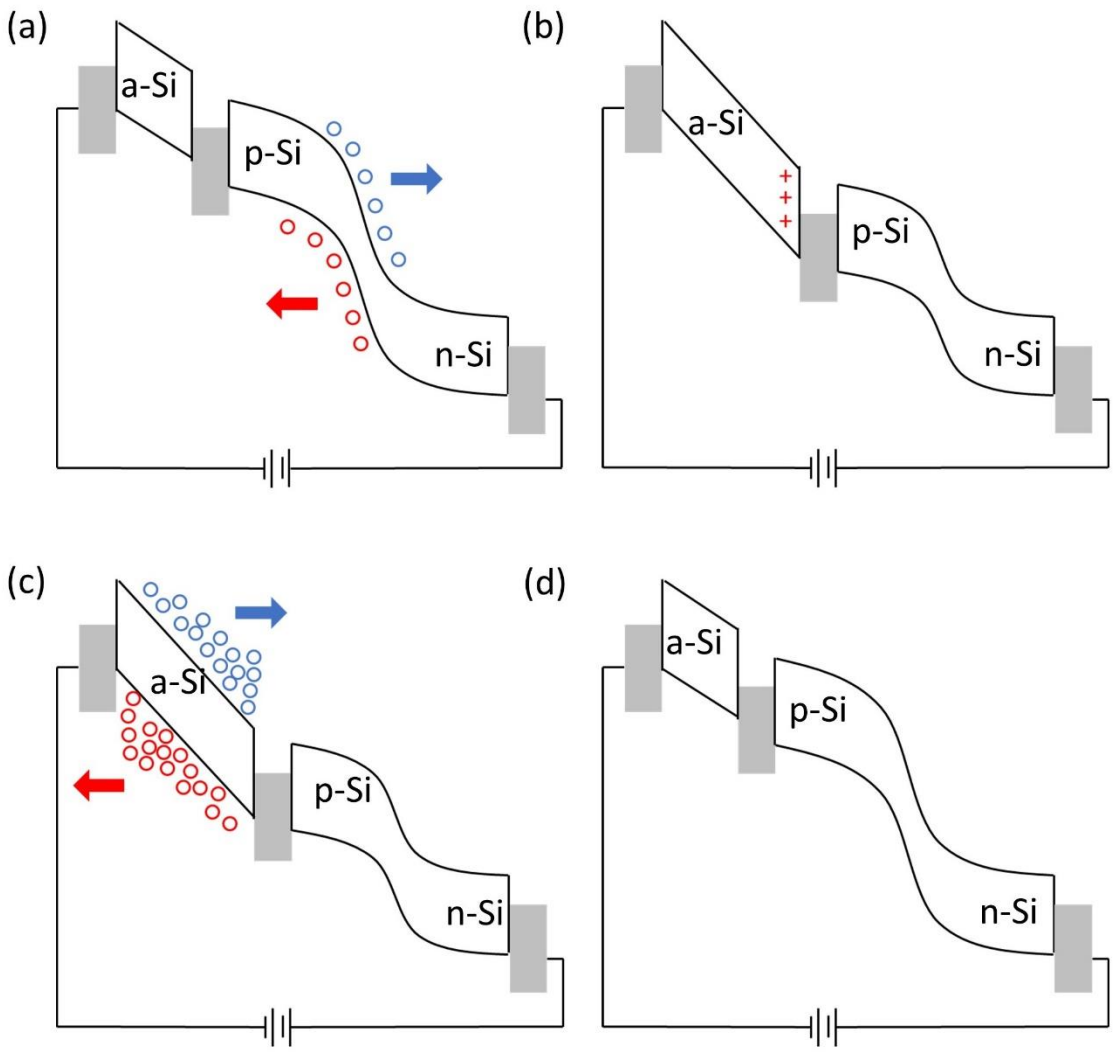

Figure 2.2 Device operation principle: (a) A single photon absorbed in the reverse-biased Si p/n junction triggers impact ionization. (b) The resulting current charges the a-Si/metal capacitance, raising the drop across a-Si while pulling the Si junction below breakdown, which quenches the avalanche. (c) The heightened a-Si field drives carrier transport with band-tail–mediated carrier multiplication, yielding an enhanced recharge current. (d) This current rapidly charges up the Si junction capacitance and re-establishes over-bias of the Si p/n junction, and the device is ready to respond to the next incoming photon.

The unique attribute of the SQUARE detector is its active self-recovery mechanism, which leverages carrier multiplication within the a-Si layer to significantly accelerate the self-recovery

process. The detailed process of how carrier multiplication can occur in disordered material such as a-Si under high electric field has been discussed in the references [37],[38].

The device was characterized in the AC domain using a 40 GHz ground–signal–ground (GSG) probe, with a bias tee employed to decouple the AC and DC components. The AC signal is routed to an oscilloscope operating at a 20 Gsamples/s sampling rate, while the DC bias is supplied by a Keysight precision source meter via the inductor branch of the bias tee. Waveform data are acquired over 1 ms at a reverse bias voltage of 26V. A threshold of 60 μA was applied during SPAD pulse extraction, which is well below the extracted peak heights to ensure reliable detection of all valid photon events. Fig. 2.3(a) shows a sample of the recorded dark pulses at 26 V reverse bias. The recorded raw data and the extracted dark pulses are shown in Fig. 2.3(a)(i) and (ii), respectively. The bias dependent dark count rate of the device is 1 to 10 kHz. Fig. 2.3(b) shows the device output when the device is illuminated with a continuous-wave (CW) laser source at 2.95e6 photons/s with a detection efficiency of 17%. The uniform pulse peak heights and temporal shapes suggests that the SPAD maintains fast quenching and recovery during continuous photoexcitation. This consistency in pulse characteristics indicates that the device can recover rapidly after each detection, allowing it to reliably respond to successive photon arrivals without distortion. In comparison, Fig. 2.3(c) shows data from a similar SPAD structure of the same thickness of defect-laden a-Si unable to produce gain for active recovery due to high interfacial states. The device in Fig. 2.3(c) shows markedly different behavior from the previous device. Under the same photon rate of 2.95e6 photons/s, the device without a-Si gain produces lower pulse amplitudes and large variations in the output, pointing to a slower and less efficient recovery mechanism such that many output responses are generated before the device is fully recovered.

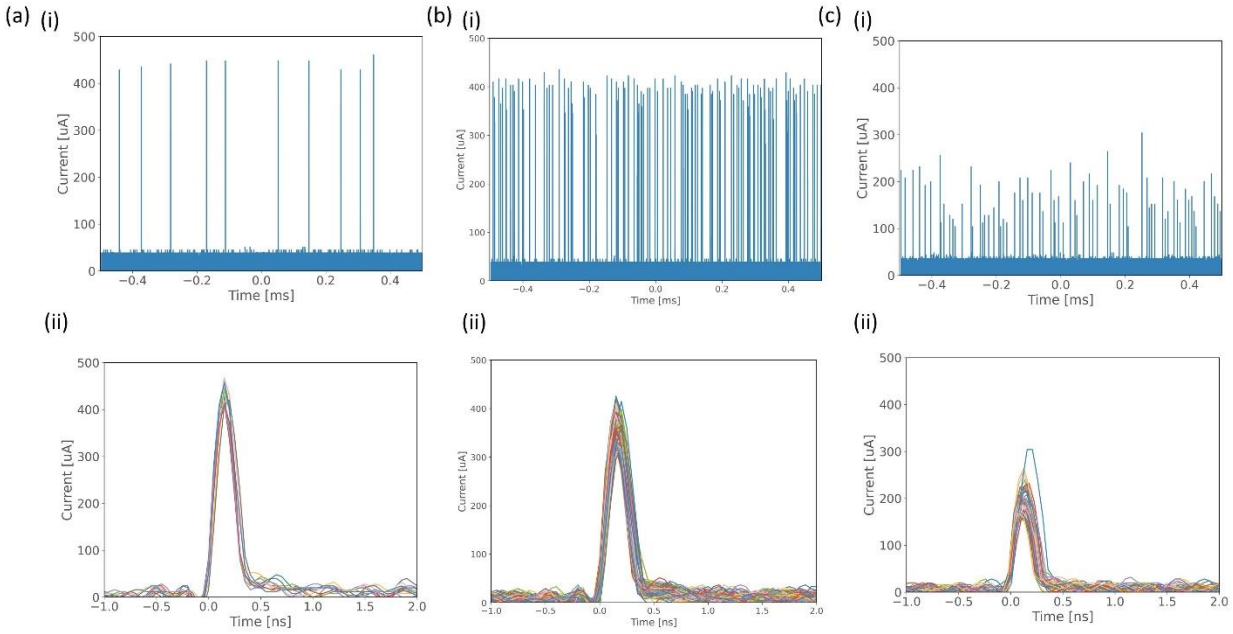

Figure 2.3 (a)(i) Waveform measured under dark conditions; (ii) overlay of noise traces without photon input. (b)(i) Waveform under illumination for the device with an a-Si gain layer at 2.95e6 photons/s; (ii) overlay of photon-induced pulses. (c)(i) Waveform under illumination for the device without an a-Si gain layer at 2.95e6 photons/s; (ii) overlay of corresponding detection pulses. Illumination was provided by a continuous-wave 635nm fiber-pigtailed laser diode.

To quantify the a-Si gain mechanism, we compare the results of standard SQUARE detectors and devices that shares the same silicon p–n junction but has the a-Si layer with a high density of interfacial states. Fig. 2.4(a) illustrates the energy band diagrams of both device structures. In the standard SQUARE detector configuration, an amorphous silicon layer contains low density of deep traps and interfacial states. Under high electric field conditions, carriers injected into the a-Si layer undergo multiplication process to produce many carriers to charge up the silicon junction during self-recovery. In contrast, the device with the same thickness but low-quality a-Si layer exhibits significant interfacial trapping, where trap states at the electrode/a-Si interface pin the potential, resulting in a flattened band profile under bias. This suppression of

field-induced gain prevents the injection of amplified carriers into the Si junction, leading to slower recharge dynamics.

Fig. 2.4(b) presents a quantitative comparison of the two devices by plotting the average current peak height as a function of Δt, defined as the time interval between consecutive detection events. Each data point represents the mean peak height for events occurring within a specific Δt bin. Comparing the two curves reveals that the device with a-Si gain achieves significantly faster response times, around 50 times faster (Fig. 2.4(b)), than the device without gain, when evaluated at equivalent current peak levels.

Figs. 2.4(c) and (d) present the measured dependence of Geiger-mode gain on photon arrival rate for SQUARE detector and the detector in the same architecture with and without a-Si gain. The Geiger-mode gain is defined as the total number of electrons generated per detection event, calculated by integrating the area under the current pulse waveform. The measurements were performed under increasing photon arrival rates, spanning from 10 kHz to 10 GHz, at two different bias voltages: 26 V and 30 V. When the SPAD is fully recovered between events, the Geiger-mode gain remains constant regardless of photon flux. However, as the arrival rate increases and events occur in closer succession, the device has less time to recharge, leading to a reduction in output signal amplitude. We define the recovery time as the inverse of the photon arrival rate at which the peak-to-valley amplitude of the detection signal drops to 50\% of its maximum value. The SQUARE detector with a-Si gain, shown in Fig. 2.4(c), exhibits fast recovery times ranging from 20 to 30 ns. In contrast, the device without the a-Si gain layer, shown in Fig. 2.4(d), demonstrates a much earlier roll-off in gain as the photon rate increases, indicating significantly slower recovery time at around 1.6 us. These results highlight the effectiveness of the integrated a-Si gain mechanism in enabling rapid self-recovery without external quenching

circuitry, representing an orders of magnitude improvement over conventional uncooled self-quenched SPADs, which typically recover on the microsecond scale. Table 2.1 shows a comparison of the dead time / recovery time between this work and the passively-quenched SPADs and SiPMs in the literature.

Table 2.1 Comparison of the dead time / recovery time for passively quenched Si SPADs, SiPMs and SQUARE detector.

Quenching Mechanism	Device Type	Temperature (K)	Dead Time / Recovery Time (ns)	Work cited
Passively quenched	Si SPAD	RT[a]	900	[39]
Passively quenched	Si SPAD	RT	970	[40]
Passively quenched	Si SPAD	RT	300 or higher	[41]
Passively quenched	SiPM	RT	31.8 ± 1.8	[42]
Passively quenched	SiPM	163	35 ± 4	[43]
Self quenched	SQUARE	RT	< 30	This work

[a] Room Temperature.

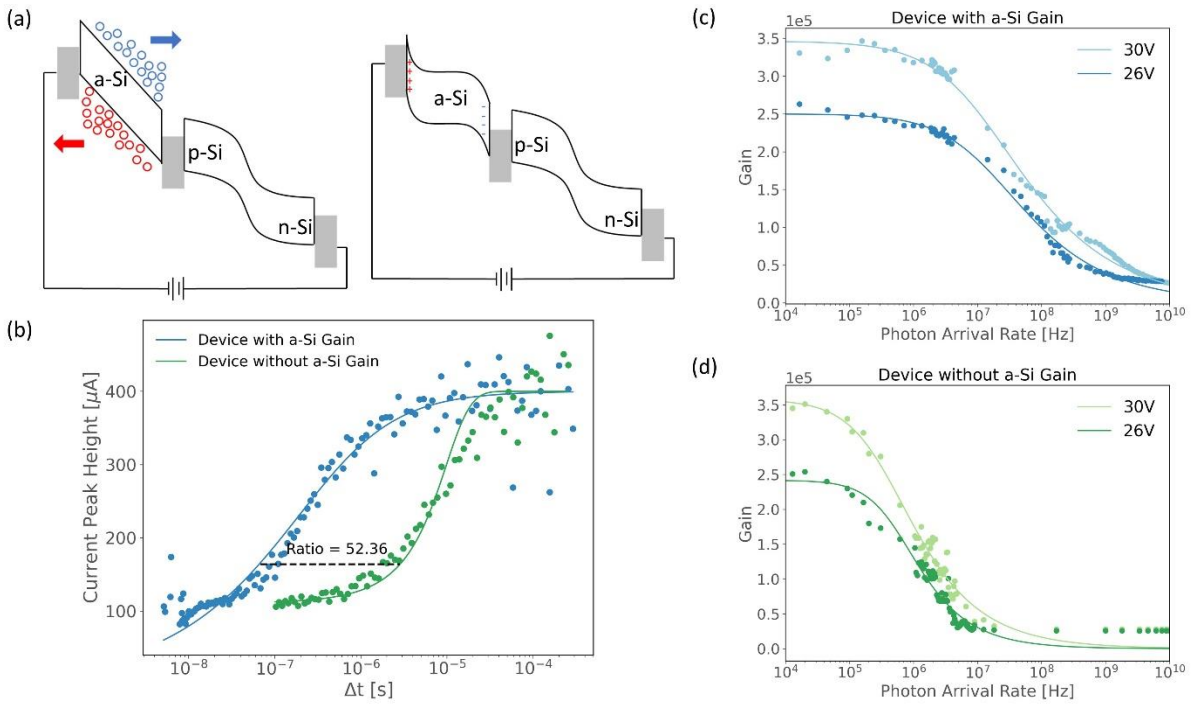

Figure 2.4 (a) Comparison of SPAD with and without a-Si gain. (a) Energy band diagrams of devices with (left) and without (right) a-Si gain layer. (b) Average pulse peak height versus the time interval since the previous detection event. (c) Geiger-mode gain versus photon arrival rate for device with a-Si gain at 26 V and 30 V. (d) Geiger-mode gain versus photon arrival rate for device without a-Si gain at 26 V and 30 V.

2.3 Device Fabrication Process and Standalone a-Si Characterization

In Tower Semiconductor's 180 nm SBC18H3 CMOS process, the fabrication of a SPAD device typically begins by defining a Deep N-Well (DNWELL) to create the buried, high-voltage region and provide electrical isolation from the substrate. Next, a P-Well (PWELL) is introduced to form the main avalanche junction, followed by an N-Well (NWELL) implant to support peripheral structures and shape the electric field. A guard ring, typically implemented using N-doping, is placed around the diode to reduce edge breakdown and improve reliability. Following these well implantations, N+ and P+ diffusions are introduced within their respective wells to create ohmic contacts for biasing and signal extraction. Contacts are then patterned and formed to

electrically connect the diode to external circuitry. Finally, multiple layers of metal interconnects are formed to route the signals and create the contact pads, and oxide passivation layers are applied to protect the device.

Following the foundry process, a second gain medium composed of carbon-incorporated hydrogenated amorphous silicon (H:a-Si) is deposited on top of the wafer. The top aluminum layer (metal 6) is first dipped in buffered oxide etch (BOE) for five seconds to remove any native oxide. Then, in a PECVD reactor at 270°C, a 100 nm-thick layer of amorphous silicon—doped with 5% carbon by introducing 550 sccm of silane (SiH4) and 28 sccm of methane (CH4)—is deposited. Patterning of the a-Si layer is carried out via reactive-ion etching (RIE) using negative photoresist as an etch mask, after which a 200 nm sputtered SiO_2 layer is applied to protect the sidewalls of the a-Si mesa and mitigate leakage. Finally, a third lithographic step with negative photoresist defines a 110 nm-thick ITO film that forms the electrical connections between the probing pads and the devices.

To visualize lateral registration between the amorphous-silicon (a-Si) stack and the photosensitive Si p–n junction, we acquired micrographs from two vertically adjacent fabricated devices that expose complementary regions of the layout. Fig. 2.5(a) highlights the a-Si region, including its connection to the contact pad and the interface aligned to the underlying p–n junction. Fig. s.5(b) focuses on the photosensitive Si p–n junction region and its contact geometry. The composite in Fig. 2.5(c) overlays (a) and (b), demonstrating spatial correspondence between the a-Si features and the p–n junction.

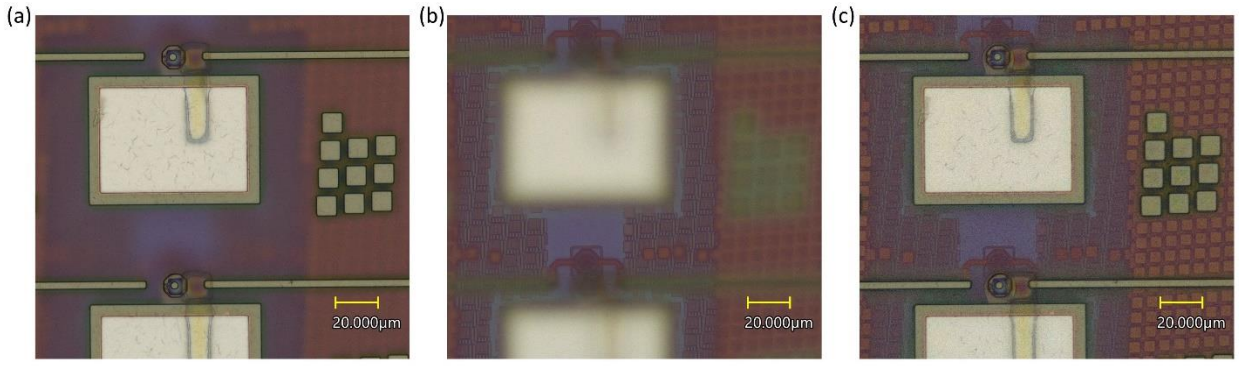

Figure 2.5 Microscopic graphs of two vertically adjacent fabricated devices. (a) Image focused on the a-Si region connected to the contact pad and the p/n junction underneath (out of focus). (b) Image focused on the photosensitive Si p/n junction region while the a-Si on top is out of focus. (c) Overlay of the graphs captured in (a) and (b), showing both the a-Si layer and the Si p/n junction.

To quantify the intrinsic gain behavior of the amorphous-silicon layer independent of the Si SPAD, a thin a-Si layer was deposited on top of the n-Si substrate. The device cross section and layout are shown in Fig. 2.6(a) and (b), respectively. We then measured the standalone a-Si device (60 nm film) under 50 MHz operation across a range of temperatures and DC bias conditions through the setup in Fig. 2.6(c). Fig. 2.7(a) shows the gain–bias characteristics at several temperatures, demonstrating a monotonic increase in gain with applied bias and a modest temperature sensitivity at lower fields; the curves converge at higher bias, consistent with field-enhanced transport via tail states/tunneling dominating over phonon-assisted processes. Fig. 2.7(b) presents the gain–temperature dependence at fixed biases, highlighting that higher bias suppresses the temperature coefficient of gain. In the integrated SQUARE stack, the a-Si layer is connected in series with the Si p–n junction; device dimensions were chosen such that, within the operating regime, the a-Si drop is ~5–7 V, corresponding to an average field of ~0.8–1.2 MV cm^{-1} for a 60 nm film, sufficient to realize the gain levels observed here while preserving stable SPAD

operation. These standalone results underpin the voltage-division, gain-assisted quench/recovery mechanism discussed in the main text.

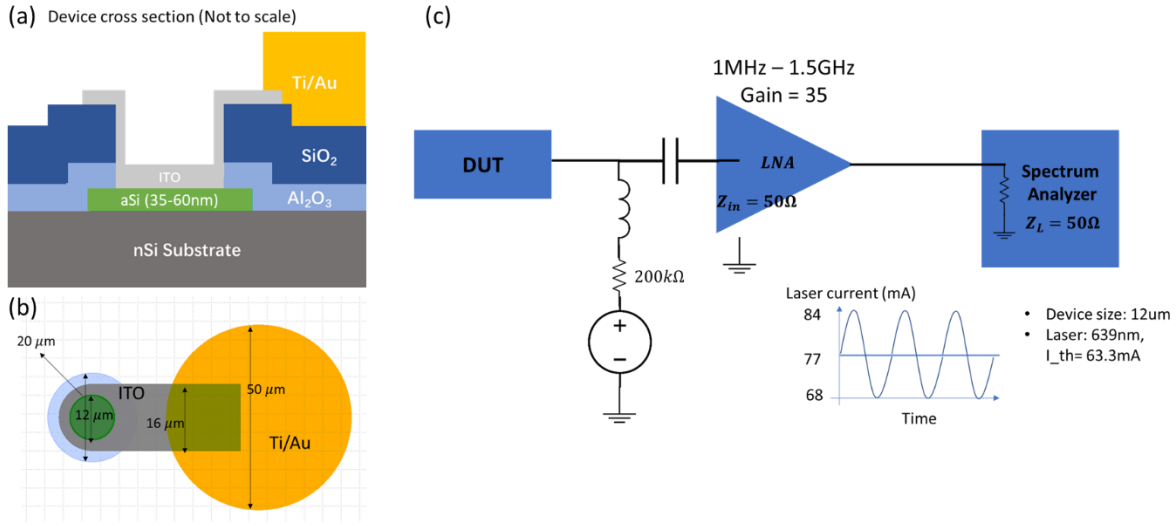

Figure 2.6 Standalone a-Si device (a) cross section and (b) layout. (c) Characterization setup.

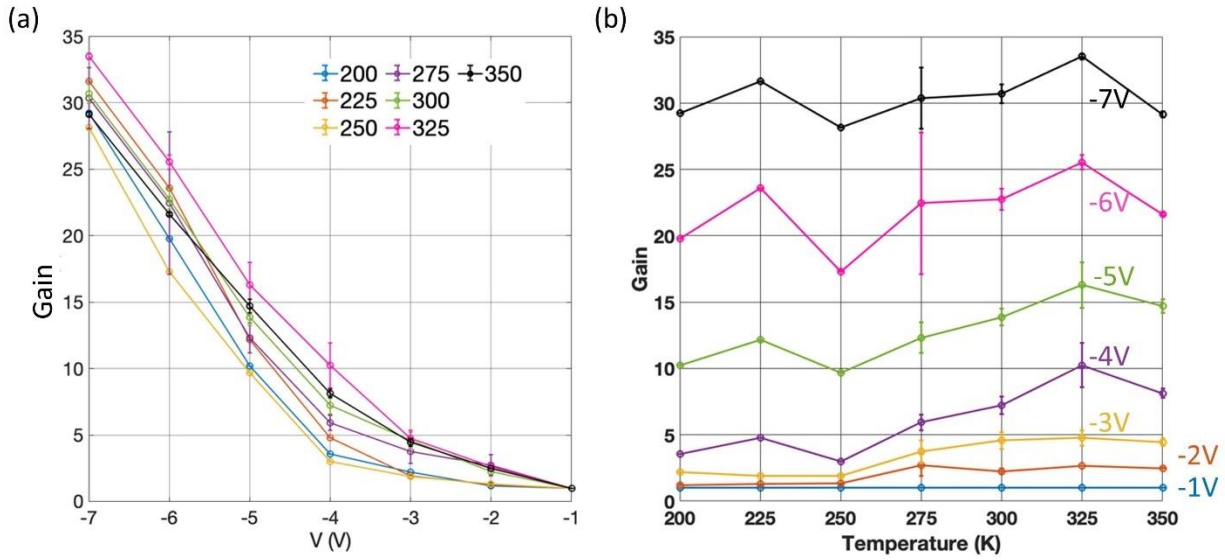

Figure 2.7 Standalone a-Si device gain-dependence characterization results at 50MHz. (a) Gain-bias dependence characterization results under various temperature. (b) Gain-temperature dependence characterization results under various bias conditions.

2.3 Device Circuit Model and Imager Output

2.3.1 Theoretical Calculations and Device Circuit Model

To model the time-dependent behavior of the SQUARE detector, we decompose its output current into two primary components: one originating in the crystalline silicon p/n junction and one associated with the carbon-incorporated amorphous silicon (a-Si) layer [31]. At any time t, the total applied bias satisfies

$$V_{bias} = V_{Si}(t) + V_{a-Si}(t)$$

Where $V_{Si}(t)$ and $V_{a-Si}(t)$ denotes the instantaneous voltage drop across the p/n junction and the a-Si layer, respectively.

The current at the p/n junction side can be written as the sum of the displacement current through the junction capacitance C_{Si} and the particle current generated by avalanche multiplication of carriers that originate at earlier times u:

$$I(t) = \int_0^t [en_{Si}(u) + h(V_{Si})] \frac{dG_{Si}^u(t-u)}{dt} du + C_{Si} \frac{d}{dt} V_{Si}(t) \quad (1)$$

where $n_{Si}(u)$ is the rate of electron–hole pairs generated by photon absorption in the p–n region at time u, $h(V_{Si})$ is the reverse-bias current of the Si p/n junction, which can be neglected for our operating bias due to its small magnitude. $\frac{dG_{Si}^u(t-u)}{dt}$ describes the time-dependent avalanche multiplication for carriers created at time u.

Similarly, at the a-Si side, the measured current contains the displacement current of the a-Si capacitance C_{a-Si} and the particle current associated with photogenerated or injected carriers in the a-Si:

$$I(t) = \int_0^t [en_{a-Si}(u) + f(V_{a-Si})] \frac{dG^u_{a-Si}(t-u)}{dt} du - C_{a-Si} \frac{d}{dt} V_{Si}(t) \quad (2)$$

where $n_{a-Si}(u)$ is the rate of carriers photo- or thermally generated in the a-Si, $f(V_{a-Si})$ represents carriers injected into the a-Si under the applied voltage, and $\frac{dG^u_{a-Si}(t-u)}{dt}$ captures the local carrier multiplication dynamics in the a-Si layer. In our device, the a−Si layer is not directly illuminated, and its intrinsic dark current is negligible within the quenching and recovery time window; therefore, $n_{a-Si}(t) \approx 0$.

A key simplifying assumption in this analysis is that carrier multiplication within each gain region develops almost instantaneously upon carrier injection. In other words, the avalanche buildup and decay times are considered negligible compared to the detector's overall self-recovery time. Mathematically, we represent this by approximating $G^u_{Si}(t-u)$ and $G^u_{a-Si}(t-u)$ using step functions:

$$\boldsymbol{G^u_{Si}(t-u) \cong G_{Si}(u)\theta(t-u)} \quad (3\text{-A})$$

$$\boldsymbol{G^u_{a-Si}(t-u) \cong G_{a-Si}(u)\theta(t-u)} \quad (3\text{-B})$$

where $\theta(\cdot)$ is the Heaviside step function, $G_{Si}(u)$ and $G_{a-Si}(u)$, represent the instantaneous multiplication gains in the silicon and a-Si layers, respectively. While this approximation does not capture the fine details of the current pulse shape produced by a single photon, it is valid for determining the longer self-recovery timescale, which typically extends beyond the sub-nanosecond regime where gain builds up or decays.

By substituting these step-function approximations back into Eqs. (1) and (2), we obtain simplified forms for the detector current, here denoted for the Si p/n region [Eq. (4)] and for the a-Si layer [Eq. (5)], respectively:

$$I(t) = e\sum_{i=0}\delta(t - i\Delta)\,G_{Si}(t) + C_{Si}\frac{d}{dt}V_{Si}(t) \tag{4}$$

$$I(t) = f(V_{a-Si})G_{a-Si}(t) - C_{a-Si}\frac{d}{dt}V_{Si}(t) \tag{5}$$

By combining these expressions and recognizing that both regions share the same node voltage $V_{Si}(t)$, we arrive at:

$$(C_{Si} + C_{a-Si})\frac{d}{dt}V_{Si}(t) = -e\sum_{i=0}\delta(t - i\Delta)\,G_{Si}(t) + f(V_{a-Si})G_{a-Si}(t) \tag{6}$$

To finalize the solution of Eq. (6), we require an expression for the injection current $f(V_{a-Si})$ at the metal/a-Si interface. Consistent with our prior investigations, direct tunneling is considered the dominant conduction mechanism in this material system under relevant bias conditions, and so we employ a Fowler–Nordheim formulation for $f(V_{a-Si})$. This approach allows us to capture the interplay between the avalanche multiplication in the a-Si layer and the quenching behavior that drives the rapid recovery of the SQUARE device. Although the instantaneous multiplication approximation omits details of sub-nanosecond transients, it reliably predicts the longer timescale evolution of the detector voltage, which is critical for optimizing self-quenching and recovery characteristics. Fig. 2.8(a) presents a model simulated example of the output signal under low photon arrival rates, where full recovery between consecutive pulses is observed. As the photon arrival rate increases, the degree of recovery diminishes, leading to reduced signal amplitude. This trend is quantified in Fig. 2.8(b), which shows the dependence of the peak-to-valley amplitude on the photon arrival rate, illustrating the impact of pulse overlap on signal magnitude.

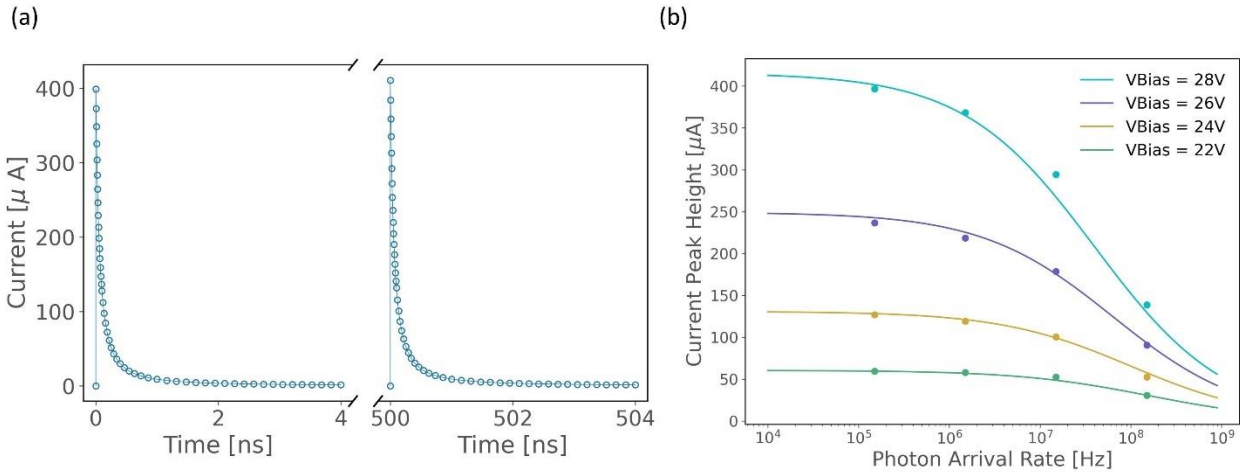

Figure 2.8 (a) Simulated output signal illustrating full recovery between pulses at low photon arrival rates. (b) Fitting result of current peak height to the applied bias and photon arrival rate using Eq. (7).

A circuit model is essential for enabling users to design electronic components, such as amplifiers, that interface with the detector [32]. In the present work, we develop an empirical circuit model in Cadence Virtuoso (Verilog-A) that captures the detector's transient response by directly relating the applied bias and photon arrival rate to the current peak amplitude given by Eq. (7). Concretely, the model generates a current pulse of empirically determined amplitude, which charges a capacitor at the photon arrival instant and subsequently discharges through a defined RC time constant of 100ps which corresponds to our experimentally observed value. This approach reproduces the critical quenching and recovery behavior of the SQUARE without resorting to detailed device-level simulations.

$$Current\ Peak\ Height\ (uA) = \frac{6V^2 - 240.8V + 2454}{1 + (Photon\ Arrival\ Rate)^{0.6} \times e^{(0.138V - 14.368)}} \tag{7}$$

here V denotes the magnitude of the externally applied bias across the series combination of the Si p–n junction and the a-Si layer.

To more accurately represent real-world SPAD operation, we enhanced our circuit-level representation to account for additional stochastic factors that arise under practical operating conditions. First, photon arrival times are treated as a random variable, recognizing that the detector's response depends on how fully it has recovered since the preceding event. Second, a Poisson distribution is used to model the occurrence of multiple photons arriving nearly simultaneously, where the resultant peak current scales approximately with the number of coincident photons. Finally, for largerarea devices, the spatial variation of absorption sites is taken into consideration: subsequent photons striking different regions may not exhibit the reduced peak amplitude typically observed in smaller SPADs. By incorporating these temporal, multiplicity, and spatial considerations, our refined model provides a more faithful simulation of the detector response in practical photon-counting scenarios.

Fig. 2.9(a) shows the Verilog-A implementation framework of the device circuit model. In this Verilog-A framework, a noise-driven pseudo-random process simulates the stochastic timing of photon arrivals, reflecting their inherently Poissonlike distribution in practical systems. Each arriving event is then evaluated to determine whether multiple photons arrive nearly simultaneously; if so, these events merge into a single, higher-amplitude pulse. Such dynamic amplitude adjustments capture realistic scenarios where closely spaced photon arrivals are indistinguishable from a single, more intense event. Moreover, the code accounts for large-area behavior by tracking whether subsequent photons land in spatially distinct regions of the detector. In larger SPADs, photons striking different locales do not diminish each other's amplitude to the extent seen in smaller devices, thus preserving signal strength and enhancing modeling fidelity. By combining these elements into a single circuit-level description, the model effectively

reproduces the key stochastic influences, including temporal dispersion, photon multiplicity, and spatial variation, governing real SPAD operation.

Fig. 2.9(b) shows the simulated output signals from the device circuit model under low photon arrival rates with various bias voltages. Fig. 2.9(c) shows the simulation result of our device circuit model. The top panel and the bottom panel illustrate how the model simulates the number of photons, and the photon arrival events, respectively. The detailed implemented code can be found in the supplementary material.

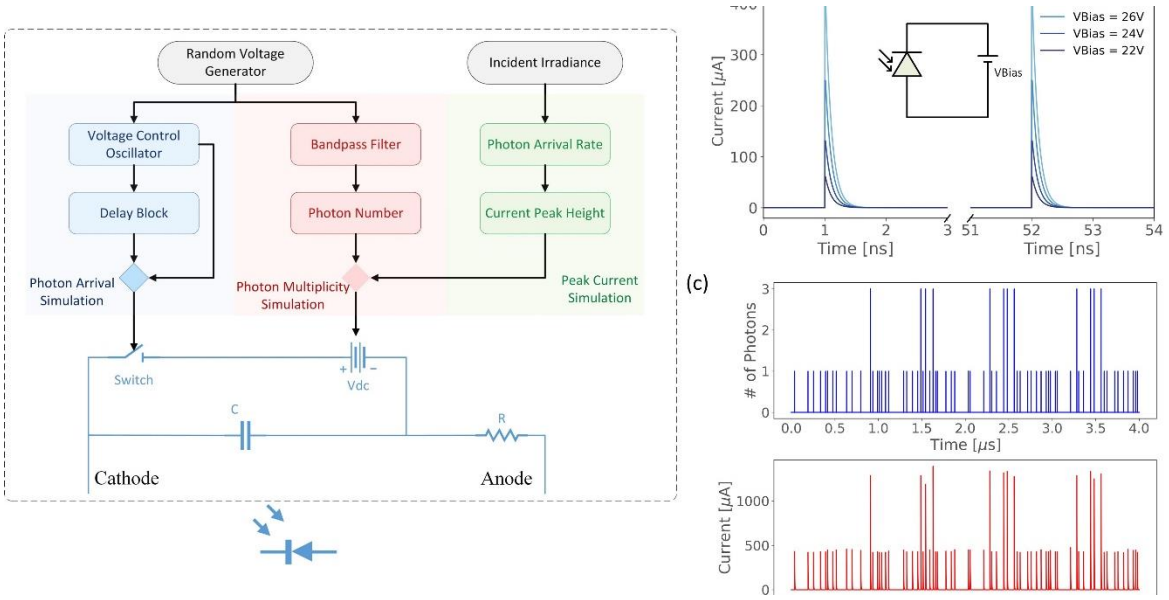

Figure 2.9 Implementation framework and behaviors of device circuit model. (a) Verilog-A implementation framework of the device circuit model. The device circuit model includes the simulation of photon arrival, peak current, and photon multiplicity. (b) Photon arrival events simulated results under various bias voltages for a given light power of 100pW. (c) Stochastic simulated results of the number of photons and the corresponding photon arrival events.

2.3.2 Verilog-A Implememtation of Device Circuit Model

Figures 2.10-2.12 illustrates how various aspects of SQUARE device operation, such as random photon arrivals, avalanche triggering, quenching, and recovery, are captured and emulated under realistic operating conditions at the circuit level. The code integrates multiple modeling

elements to achieve a powerful yet flexible simulation framework. First, a parameterized noise generation routine periodically updates a noise variable using the built-in $random function; this noise is then scaled by a user-defined factor to control the frequency and amplitude of simulated photon arrivals. A square-wave signal, produced by integrating the instantaneous frequency over time to obtain a phase variable, emulates the fundamental "arrival or no arrival" behavior, while a small delay introduces a realistic timing offset.

To refine the spectral content and rate of these arrivals, the code applies a bandpass filter implemented via the $laplace\_nd$ function, effectively shaping the noise signal within prescribed low and high cutoff frequencies. The filter's parameters, including center frequency, quality factor, and gain, are exposed as user-tunable values, allowing fine control over the appearance rate and intensity of photon-like events. The simulation then derives the number of coincident photons from thresholding this filtered output at distinct levels, enabling the modeling of multiple photon arrivals in close temporal proximity.

A switching mechanism, governed by a user-defined voltage threshold (V_{th}), transitions between high (R_{off}) and low (R_{on}) resistances to emulate the rapid quenching process upon photon detection. Internal device parameters such as R_{device} and C_{device} approximate the parasitic resistance and junction capacitance of the SQUARE device, dictating how it charges and discharges once an avalanche is triggered. Meanwhile, an empirical expression calculates the avalanche peak voltage (V_{peak}) based on the applied bias (V_{app}), the number of photons arriving simultaneously, and the device area. This expression additionally considers whether subsequent photons strike the same region of a large-area device, capturing spatial variability and preventing the reduction in peak height typically seen in smaller detectors.

Through this combination of parameterized noise, frequency modulation, bandpass filtering, switching, and empirical peak-height calculations, the Verilog-A module faithfully reproduces the essential dynamics of SPAD avalanche and quenching behavior within a foundry-specific CMOS environment. The result is a versatile modeling approach that accommodates rapid exploration of design trade-offs—such as bias levels, detector geometry, and readout circuitry requirements—without resorting to computationally intensive device-level simulations. Consequently, the framework expedites the validation and refinement of both the SQUARE device itself and its accompanying electronics, allowing researchers to pinpoint optimal configurations for high-performance single-photon detection in a variety of applications.

```
// VerilogA for layout2024, SQUARE_YJ, veriloga

`include "constants.vams"
`include "disciplines.vams"

module SQUARE_YJ(Anode, Cathode);
  inout Anode, Cathode;

  electrical Anode, Cathode;
  electrical arrival, photons;
  electrical wn, filter_out, filter_out1;
  electrical nodeRC, nodeES;
  electrical Vapp, Peak;
  electrical same_place;

  // Parameters
  parameter real area            = 4.34e-7;
  parameter real hc              = 1.986e-25;
  parameter real wavelength      = 518e-9;
  parameter real irradiance      = 2.3e-8;
  parameter real step_time       = 1e-11;
  parameter real char_area       = 4.34e-7;
  parameter real rate            = irradiance * area * wavelength / hc;

  parameter real Rdevice         = 1e3 from [0:inf);
  parameter real Cdevice         = 100e-15 from (0:inf);

  parameter real sample_interval = 1e-9;
  parameter real MAX_INT = 2147483648.0;
  parameter real center_freq   = rate;
  parameter real sensitivity   = 2*rate;
  parameter real v_high        = 1.0;
  parameter real v_low         = 0.0;

  parameter real fp_low = 2e6;
  parameter real fp_high = 4e6;
  parameter real fc = sqrt(fp_low * fp_high);
  parameter real BW = fp_high - fp_low;
  parameter real Q = fc / BW;

  parameter real filter_gain = 1.0;
  parameter real epsilon = 1e-6;

  parameter real Ron           = 1e-3 from (0:inf);
  parameter real Roff          = 1e10 from (0:inf);
  parameter real Vth           = 0.5;
```

Figure 2.10 Verilog-A Implementation of Device Circuit Model. Part A.

```
47
48  real last_update_time;
49  real phase;
50  real noise_value;
51  real vco_out;
52  real delayed_vco;
53  real filter_output;
54  real noise_amplitude = 0.5;
55  real R_switch;
56
57
58  analog begin
59    if ($abstime == 0) begin
60      last_update_time = 0;
61      noise_value = 0;
62    end
63    if (($abstime - last_update_time) >= sample_interval) begin
64      last_update_time = $abstime;
65      noise_value = $random;
66      noise_value = noise_amplitude * (noise_value / MAX_INT);
67    end
68    V(wn) <+ noise_value;
69
70    if ( V(wn) > (0.5-char_area / area) ) begin
71      V(same_place) <+ 1;
72    end else begin
73      V(same_place) <+ 0;
74    end
75
76    phase = idt((center_freq + sensitivity * noise_value) * 2.0 * `M_PI, 0);
77    vco_out = ((phase % (2.0 * `M_PI)) < `M_PI ? v_high : v_low);
78    delayed_vco = absdelay(vco_out, step_time);
79    V(arrival) <+ abs(delayed_vco - vco_out);
80
81    if (V(arrival)*V(photons) > Vth) begin
82      R_switch = Ron;
83    end else begin
84      R_switch = Roff;
85    end
86    I(nodeES, nodeRC) <+ V(nodeES, nodeRC) / R_switch;
87
88    I(nodeRC, Anode) <+ V(nodeRC, Anode) / Rdevice;
89
90    I(Cathode, nodeRC) <+ Cdevice * ddt(V(Cathode, nodeRC));
91
92    V(Vapp) <+ V(Cathode) - V(Anode);
```

Figure 2.11 Verilog-A Implementation of Device Circuit Model. Part B.

```
93
94      filter_output = laplace_nd(V(wn),  { 0, filter_gain * (2.0 * `M_PI * fc /
        Q), 0},  { (2.0 * `M_PI * fc) * (2.0 * `M_PI * fc), 2.0 * `M_PI * fc / Q, 1
        });
95      V(filter_out) <+ filter_output;
96      V(photons) <+ ( V(filter_out) < -2e-3 ? 0 : ( V(filter_out) < 8e-3 ? 1 :
        ( V(filter_out) < 18e-3 ? 2 : 3 ) ) );
97
98      V(Peak) <+ 1e-3 * (V(Vapp)<20 ? V(Vapp)-1 : (3*pow(V(Vapp), 2) - 120.4*V(
        Vapp) + 1227)) / (1 + V(same_place) * pow(rate/1000, 0.6)*exp(0.138*V(Vapp)-
        14.368));
99      V(Cathode, nodeES) <+ V(Vapp) - V(Peak) * V(photons);
100
101   end
102
103 endmodule
```

Figure 2.12 Verilog-A Implementation of Device Circuit Model. Part C.

2.3.3 Readout Circuit Design

In a single-photon imaging system, the readout circuit converts the avalanche signals generated by each photon detection into a usable electrical output for subsequent processing. We employ the aforementioned SQUARE detector circuit model in our design of a trans-impedance amplifier (TIA)-based readout. By integrating this empirical model into Cadence Virtuoso, we obtain realistic simulation waveforms that capture the stochastic avalanche dynamics, quenching behavior, and photon arrival statistics tailored to our device. These waveforms inform key design decisions, such as bias volt ages, transistor sizing, and feedback resistor selection, for the TIA, ensuring that the amplifier is well-matched to the detector's transient current pulses.

The TIA itself employs a common-source N-MOSFET with a feedback resistor that sets the trans-impedance gain. A diode-connected N-MOSFET provides the active load, establishing the DC operating point while supplying an effective resistance in a compact form. This configuration achieves efficient current-to-voltage conversion while maintaining good noise performance. The output node is subsequently buffered by a source follower stage, which delivers

a low-impedance, robust output signal suitable for driving downstream circuitry. By integrating this TIA with the SQUARE detector within the CMOS process, parasitic capacitances are minimized, preserving the integrity of the short, delicate current pulses generated by avalanche events. In tandem, the co-simulation framework and empirical device model guide iterative refinements that optimize performance metrics such as dynamic range, timing accuracy, and overall signal-to-noise ratio. Fig. 2.13(a) illustrates the proposed TIA architecture, implemented in Tower Semiconductor's 180 nm SBC18H3 CMOS process, along with its connection to our SQUARE detector model. Pre-layout simulations indicate that the designed TIA achieves a transimpedance gain of 54.23 $dB\Omega$ with a -3dB bandwidth of 34.67 GHz. The reported bandwidth represents the intrinsic circuit performance prior to layout; post-layout parasitic capacitances, routing, and packaging are expected to reduce the achievable bandwidth. Based on our DC simulation with a 1.8 V supply and a quiescent current of approximately 147 uA, the total static power consumption of the amplifier is about 265 uW. Fig. 2.13(b) shows the simulated performance of the combined system. The top panel depicts the transient current pulses at the SQUARE device node, while the lower panel displays the corresponding TIA output.

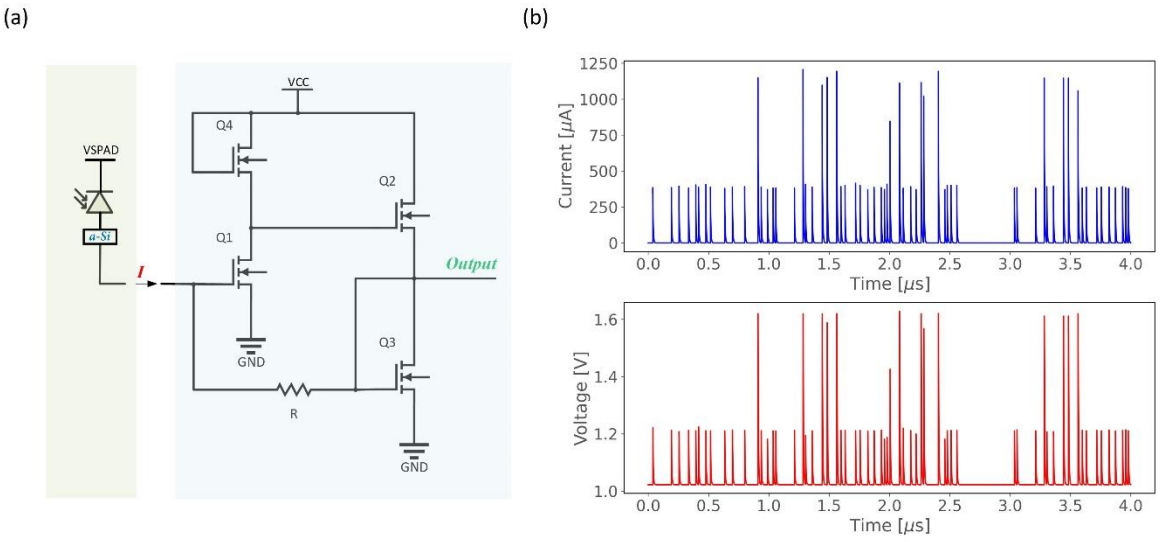

Figure 2.13 (a) TIA design with the SQUARE detector in an equivalent circuit model. (b) The photon arrival events simulation results from the SQUARE device circuit model (top panel) and the corresponding outputs from the TIA (bottom panel).

The gain bandwidth simulation result of the proposed TIA design is shown in Fig. 2.14 under Tower Semiconductor's 180 nm SBC18H3 CMOS process. From the simulation result, we can see that a gain of 54.23dB and a pre-layout -3dB bandwidth of 34.67GHz are achieved. Note that this presents the intrinsic circuit performance prior to layout; post-layout parasitic capacitances, routing, and packaging are expected to reduce the achievable bandwidth. A zoom-in view of the input current and output voltage is shown in Fig. 2.15, confirming the fast response of the TIA circuit.

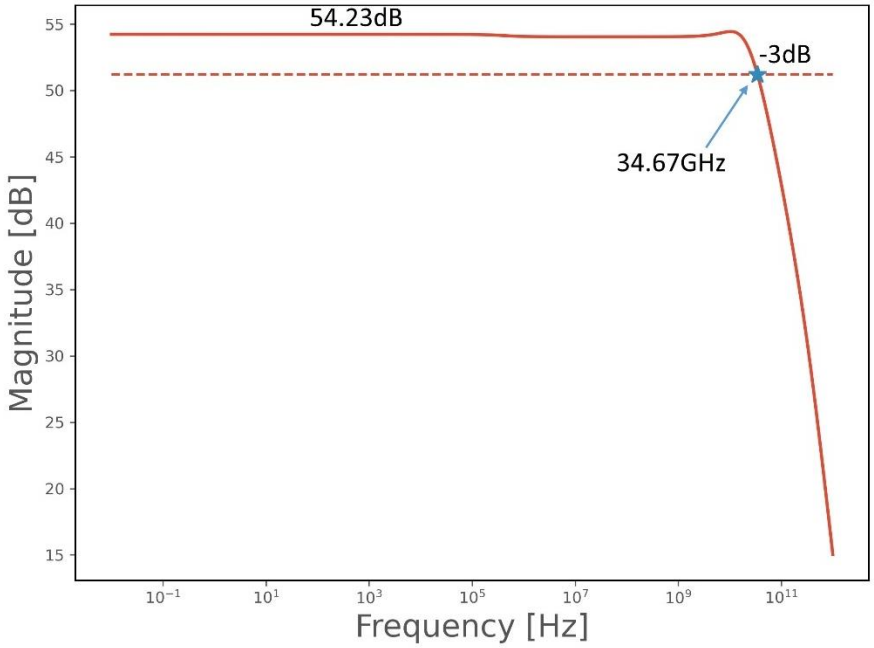

Figure 2.14 Gain bandwidth simulation of the TIA design.

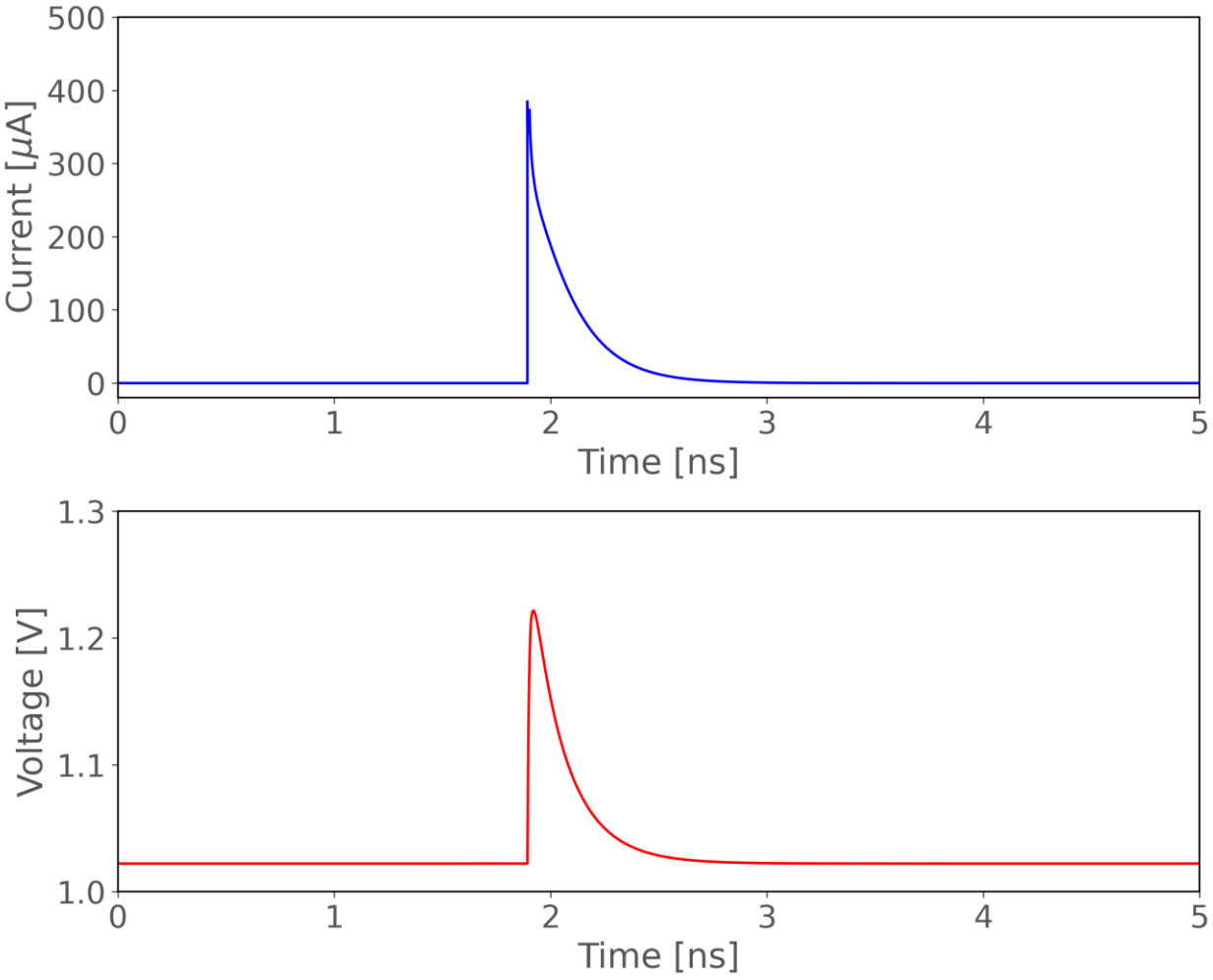

Figure 2.15 Zoom-in view of the single photon response from the SQUARE detector and the corresponding output waveform from the readout TIA circuit.

2.3.4 CMOS SQUARE Detector Imaging Experiment

An SPAD imaging setup developed in this work leverages the enhanced response time provided by the self-quenched active recovery (SQUARE) mechanism to detect and resolve individual photon events with superior timing precision compared to conventional designs. This architecture enables each pixel to operate in photon-counting mode, where the number of photon detection events within a defined integration window is recorded to construct an image. Such a

method is inherently suited for high temporal resolution imaging due to the rapid recovery and low dead time of the SQUARE-enhanced SPAD pixels.

The experimental setup is shown in Figure 2.16. A pulse-modulated fiber-pigtailed 635nm laser is used in this setup. The two beam splitters are employed to direct both the laser beam and the LED illumination onto the device under test (DUT), enabling simultaneous imaging of the laser spot and the DUT. A 10X objective lens focuses the combined light onto the DUT. A 40 GHz ground-signal-ground (GSG) probe establishes contact with the device. A bias tee separates the AC and DC signals: the DC bias is routed through the inductor branch, while the AC response is coupled via the capacitor and directed to the oscilloscope for data capture.

A photo mask is introduced into the optical path and mounted on a motorized two-axis stage. The mask modulates the transmitted laser intensity according to its local transparency pattern, effectively encoding spatial information into the optical signal that reaches the DUT. During operation, the stage translates the mask in a raster-scan fashion across the entire field of view (FOV). The travel step size along the x- and y-axes defines the effective pixel size, and the timing of the stage motion is synchronized with the acquisition system to ensure precise spatial registration of the measured signals.

As the mask is scanned, regions of higher transparency permit stronger laser transmission, while opaque or transparent regions attenuate the beam accordingly. The DUT response is therefore modulated in direct correspondence with the mask features. By correlating the acquired signal intensity with the known stage position at each step, a two-dimensional image can be reconstructed. This process yields a spatially resolved map of the mask pattern projected through the optical system, with resolution determined by the step size, stage stability, and optical point spread function.

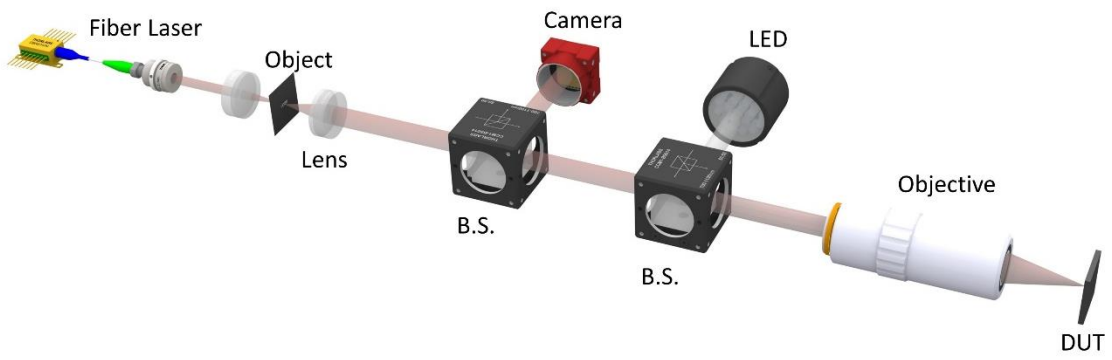

Figure 2.16 Artistic Illustration of the experimental setup. B.S.: Beam Splitter; DUT: Device Under Testing.

Imaging experiments were performed using a set of structured test patterns to assess the imager's practical capabilities. The first target, the UCSD logo, was imaged using a per-pixel integration time of 100 us (Fig. 2.17(a)). This relatively long exposure enabled high photon accumulation, resulting in a high signal-to-noise ratio and a well-resolved image. To examine the system's behavior under high-speed conditions, the integration time was progressively reduced. At a reduced exposure of 500 ns per pixel, the UCSD logo remained clearly discernible (Fig. 2.17(b)). The ability to maintain recognizable structure at sub-microsecond exposure underscores the efficient avalanche quenching recovery offered by the SQUARE architecture.

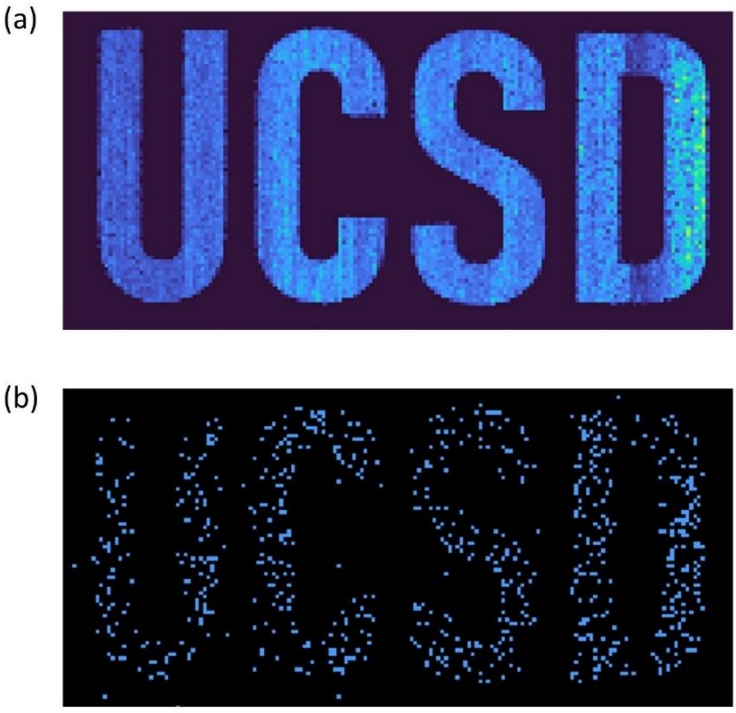

Figure 2.17 Imaging results obtained using the CMOS SQUARE imaging system. (a) UCSD pattern captured with a pixel acquisition time of 100 μs, demonstrating high-fidelity image reconstruction. (b) UCSD pattern captured with a reduced acquisition time of 500 ns, illustrating the device ability to maintain pattern visibility while operating at high speed.

2.4 Summary

In summary, we have developed a CMOS-compatible SQUARE SPAD architecture that achieves self-quenching and accelerated self-recovery. We have demonstrated 20-30ns self-recovery time in free-running mode operation, which is over 50 times faster than passive self-quenched SPADs without gain during recovery.

To facilitate adoption of the device in applications using commercial CMOS process, we have developed a Verilog-A circuit model that incorporates the stochastic nature of single-photon detection. The model considers random photon arrival times and can be extended across an array by assigning position-dependent photon flux. The SQUARE detector architecture offers an

attractive solution for high-speed, high-resolution photon counting and imaging without external quenching circuits.

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in Jiang, Y., Peng, H., Tang, Y., Zeng, L., Paul, S.J., Davis, A., Sahu, A., Nomura, K. and Lo, Y.H., 2025. Monolithically integrated pixel architecture with colloidal quantum dots and thin film transistors for SWIR imaging. APL Electronic Devices, 1(4). The dissertation author was the first author of this paper.

REFERENCE

[1] C. J. Chunnilall, I. P. Degiovanni, S. Kück, I. Müller, and A. G. Sinclair, ‘‘Metrology of single-photon sources and detectors: a review,’’ Optical Engineering, vol. 53, no. 8, pp. 081 910–081 910, 2014.

[2] R. H. Hadfield, ‘‘Single-photon detectors for optical quantum information applications,’’ Nature photonics, vol. 3, no. 12, pp. 696–705, 2009.

[3] J. Zhang, M. A. Itzler, H. Zbinden, and J.-W. Pan, ‘‘Advances in InGaAs/InP single-photon detector systems for quantum communication,’’ Light: Science & Applications, vol. 4, no. 5, pp. e286–e286, 2015.

[4] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, ‘‘Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers,’’ Nature communications, vol. 10, no. 1, p. 4984, 2019.

[5] Y. Guan, H. Li, L. Xue, R. Yin, L. Zhang, H. Wang, G. Zhu, L. Kang, J. Chen, and P. Wu, ‘‘Lidar with superconducting nanowire single-photon detectors: Recent advances and developments,’’ Optics and Lasers in Engineering, vol. 156, p. 107102, 2022.

[6] Y. Wang, K. Huang, J. Fang, M. Yan, E. Wu, and H. Zeng, ‘‘Mid-infrared single-pixel imaging at the single-photon level,’’ Nature communications, vol. 14, no. 1, p. 1073, 2023.

[7] S. Komiyama, O. Astafiev, V. Antonov, T. Kutsuwa, and H. Hirai, ‘‘A single-photon detector in the far-infrared range,’’ Nature, vol. 403, no. 6768, pp. 405–407, 2000.

[8] E. Reiger, S. Dorenbos,V. Zwiller, A.Korneev, G. Chulkova, I. Milostnaya, O. Minaeva, G. Gol’tsman, J. Kitaygorsky, D. Pan et al., ‘‘Spectroscopy with nanostructured superconducting single photon detectors,’’ IEEE Journal of Selected Topics in Quantum Electronics, vol. 13, no. 4, pp. 934–943, 2007.

[9] H. De Riedmatten, M. Afzelius, M. U. Staudt, C. Simon, and N. Gisin, ‘‘A solid-state light–matter interface at the single-photon level,’’ Nature, vol. 456, no. 7223, pp. 773–777, 2008.

[10] K. Hamamatsu Photonics, ‘‘Photomultiplier tubes: Basics and applications,’’ Edition 3a, vol. 310, 2007.

[11] J. Rolfe and S. Moore, ‘‘The efficient use of photomultiplier tubes for recording spectra,’’ Applied Optics, vol. 9, no. 1, pp. 63–71, 1970.

[12] B. G. Oripov, D. S. Rampini, J. Allmaras, M. D. Shaw, S. W. Nam, B. Korzh, and A. N. McCaughan, ‘‘A superconducting nanowire single photon camera with 400,000 pixels,’’ Nature, vol. 622, no. 7984, pp. 730–734, 2023.

[13] E. E. Wollman, V. B. Verma, A. B. Walter, J. Chiles, B. Korzh, J. P. Allmaras, Y. Zhai, A. E. Lita, A. N. McCaughan, E. Schmidt et al., ‘‘Recent advances in superconducting nanowire

single-photon detector technology for exoplanet transit spectroscopy in the mid-infrared,'' Journal of Astronomical Telescopes, Instruments, and Systems, vol. 7, no. 1, pp. 011 004–011 004, 2021.

[14] D. P. Palubiak and M. J. Deen, ''Cmos spads: Design issues and research challenges for detectors, circuits, and arrays,'' IEEE Journal of Selected Topics in Quantum Electronics, vol. 20, no. 6, pp. 409–426, 2014.

[15] F. Gramuglia, M.-L. Wu, C. Bruschini, M.-J. Lee, and E. Charbon, ''A low-noise CMOS SPAD pixel with 12.1 ps sptr and 3 ns dead time,'' IEEE Journal of Selected Topics in Quantum Electronics, vol. 28, no. 2: Optical Detectors, pp. 1–9, 2021.

[16] D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, ''Spad figures of merit for photon-counting, photon-timing, and imaging applications: a review,'' IEEE Sensors Journal, vol. 16, no. 1, pp. 3–12, 2015.

[17] R. K. Henderson, N. Johnston, H. Chen, D. D.-U. Li, G. Hungerford, R. Hirsch, D. McLoskey, P. Yip, and D. J. Birch, ''A 192× 128 time correlated single photon counting imager in 40nm cmos technology,'' in ESSCIRC 2018-IEEE 44th European Solid State Circuits Conference (ESSCIRC). IEEE, 2018, pp. 54–57.

[18] I. Cusini, D. Berretta, E. Conca, A. Incoronato, F. Madonini, A. A. Maurina, C. Nonne, S. Riccardo, and F. Villa, ''Historical perspectives, state of art and research trends of spad arrays and their applications (part ii: Spad arrays),'' Frontiers in Physics, vol. 10, p. 906671, 2022.

[19] N. Na, Y.-C. Lu, Y.-H. Liu, P.-W. Chen, Y.-C. Lai, Y.-R. Lin, C.-C. Lin, T. Shia, C.-H. Cheng, and S.-L. Chen, ''Room temperature operation of germanium–silicon single-photon avalanche diode,'' Nature, vol. 627, no. 8003, pp. 295–300, 2024.

[20] J. Zheng, X. Xue, C. Ji, Y. Yuan, K. Sun, D. Rosenmann, L. Wang, J. Wu, J. C. Campbell, and S. Guha, ‘‘Dynamic-quenching of a single photon avalanche photodetector using an adaptive resistive switch,’’ Nature Communications, vol. 13, no. 1, p. 1517, 2022.

[21] M. M. Hayat, M. A. Itzler, D. A. Ramirez, and G. J. Rees, ‘‘Model for passive quenching of spads,’’ in Quantum Sensing and Nanophotonic Devices VII, vol. 7608. SPIE, 2010, pp. 709–716.

[22] F. Nolet, S. Parent, N. Roy, M.-O. Mercier, S. A. Charlebois, R. Fontaine, and J.-F. Pratte, ‘‘Quenching circuit and spad integrated in cmos 65 nm with 7.8 ps fwhm single photon timing resolution,’’ Instruments, vol. 2, no. 4, p. 19, 2018.

[23] W. Jiang, R. Scott, and M. J. Deen, ‘‘High-speed active quench and reset circuit for spad in a standard 65 nm cmos technology,’’ IEEE Photonics Technology Letters, vol. 33, no. 24, pp. 1431–1434, 2021.

[24] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, ‘‘Avalanche photodiodes and quenching circuits for single-photon detection,’’ Applied optics, vol. 35, no. 12, pp. 1956–1976, 1996.

[25] A. Gallivanoni, I. Rech, D. Resnati, M. Ghioni, and S. Cova, ‘‘Monolithic active quenching and picosecond timing circuit suitable for large-area single-photon avalanche diodes,’’ Optics Express, vol. 14, no. 12, pp. 5021–5030, 2006.

[26] Y. Xu, J. Lu, and Z. Wu, ‘‘A compact high-speed active quenching and recharging circuit for spad detectors,’’ IEEE Photonics Journal, vol. 12, no. 5, pp. 1–8, 2020.

[27] S. You, J. Cheng, and Y.-H. Lo, ‘‘Physics of single photon avalanche detectors with built-in self-quenching and self-recovering capabilities,’’ IEEE Journal of Quantum Electronics, vol. 48, no. 7, pp. 960–967, 2012.

[28] K. Zhao, S. You, J. Cheng, and Y.-h. Lo, ‘‘Self-quenching and selfrecovering ingaas/ inalas single photon avalanche detector,’’ Applied Physics Letters, vol. 93, no. 15, 2008.

[29] D. Hall, Y.-H. Liu, and Y.-H. Lo, ‘‘Single photon avalanche detectors: prospects of new quenching and gain mechanisms,’’ Nanophotonics, vol. 4, no. 4, pp. 397–412, 2015.

[30] J. Cheng, S. You, S. Rahman, and Y.-H. Lo, ‘‘Self-quenching InGaAs/InP single photon avalanche detector utilizing zinc diffusion rings,’’ Optics express, vol. 19, no. 16, pp. 15 149–15 154, 2011.

[31] B. Schuster, H. Peng, S. Arya, Y. Jiang, M. Miah, J. Zhou, A. Davis, and Y.-H. Lo, ‘‘Cmos-compatible self-quenched active recovery (square) single photon detector,’’ APL Photonics, vol. 8, no. 9, 2023.

[32] M. A. R. Miah, Y. Jiang, and Y.-H. Lo, ‘‘A physics based unified circuit model for single photon and analog detector,’’ IEEE Access, vol. 9, pp.129 571–129 581, 2021.

[33] S. Arya, Y. Jiang, B. K. Jung, Y. Tang, T. N. Ng, S. J. Oh, K. Nomura, and Y.-H. Lo, ‘‘Understanding colloidal quantum dot device characteristics with a physical model,’’ Nano Letters, vol. 23, no. 21, pp. 9943–9952, 2023.

[34] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen, and H. Peng, ‘‘A comprehensive and accurate analytical spad model for circuit simulation,’’ IEEE Transactions on Electron Devices, vol. 63, no. 5, pp. 1940–1948, 2016.

[35] X. Wang, Y. Liu, J. Hu, D. Li, R. Ma, and Z. Zhu, ‘‘An analog sipm based receiver with on-chip wideband amplifier module for direct tof lidar applications,’’ IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 70, no. 1, pp. 88–100, 2022.

[36] Y. Liu, L.Wang, L. Gao, R. Fan, X. Su, L. Shen, S. Pu, L.Wang, and Z. Zhu, ‘‘Frontiers and challenges in silicon-based single-photon avalanche diodes and key readout circuits,’’ Microelectronics Journal, vol. 147, p. 106165, 2024.

[37] Y.-H. Liu, L. Yan, A. C. Zhang, D. Hall, I. A. Niaz, Y. Zhou, L. Sham, and Y.-H. Lo, ‘‘Cycling excitation process: An ultra efficient and quiet signal amplification mechanism in semiconductor,’’ Applied Physics Letters, vol. 107, no. 5, 2015.

[38] L. Yan, Y. Yu, A. C. Zhang, D. Hall, I. A. Niaz, M. A. Raihan Miah, Y.-H. Liu, and Y.-H. Lo, ‘‘An amorphous silicon photodiode with 2 thz gain bandwidth product based on cycling excitation process,’’ Applied Physics Letters, vol. 111, no. 10, 2017.

[39] M. Stipčević, H. Skenderović, and D. Gracin, ‘‘Characterization of a novel avalanche photodiode for single photon detection in vis-nir range,’’ Optics express, vol. 18, no. 16, pp. 17 448–17 459, 2010.

[40] L. Neri, S. Tudisco, F. Musumeci, A. Scordino, G. Fallica, M. Mazzillo, and M. Zimbone, ‘‘Dead time of single photon avalanche diodes,’’ Nuclear Physics B-Proceedings Supplements, vol. 215, no. 1, pp. 291–293, 2011.

[41] J. Zheng, X. Xue, C. Ji, Y. Yuan, K. Sun, D. Rosenmann, L. Wang, J. Wu, J. C. Campbell, and S. Guha, ‘‘Dynamic-quenching of a singlephoton avalanche photodetector using an adaptive resistive switch,’’ Nature Communications, vol. 13, no. 1, p. 1517, 2022.

[42] J. Jiang, J. Jia, T. Zhao, K. Liang, R. Yang, and D. Han, ‘‘Recovery time of silicon photomultiplier with epitaxial quenching resistors,’’ Instruments, vol. 1, no. 1, p. 5, 2017.

[43] G. Gallina, P. Giampa, F. Retière, J. Kroeger, G. Zhang, M. Ward, P. Margetak, G. Li, T. Tsang, L. Doria et al., ‘‘Characterization of the hamamatsu vuv4 mppcs for nexo,’’ Nuclear

Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 940, pp. 371–379, 2019.

Chapter 3 Monolithically Integrated Pixel Architecture with Colloidal Quantum Dots and Thin Film Transistors for SWIR Imaging

Infrared (IR) imaging, particularly in the short-wave infrared (SWIR) range, is important for applications such as remote sensing, biomedical diagnostics, and machine vision. Yet widespread deployment remains constrained because today's IR imaging primarily use epitaxial III–V photodetectors, which deliver high responsivity and low dark current but require expensive, (nearly) lattice-matched substrates, high-temperature growth, and complicated hybrid integration of focal plane array and ROIC. Alternatively, colloidal quantum dot (CQD) photodiodes are attractive for their solution processability with tunable bandgap on large-area, low-cost crystalline or non-crystalline substrates. Here we present a monolithic pixel architecture that integrates a PbS CQD photodiode in photo-voltaic mode and an IGZO TFT in subthreshold mode. This configuration achieves high optical-to-current conversion gain of the order of 10,000 at 1060nm wavelength with low dark current and over 1MHz bandwidth to support high frame rate. The demonstrated approach offers a scalable and manufacturable path toward low-cost SWIR imaging over large, panel-sized platforms.

3.1 Introduction

State-of-the-art short-wave infrared (SWIR) imagers are predominantly based on epitaxial III–V materials such as InGaAs, which deliver high responsivity and low dark current but require lattice-matched substrates and complex, high-temperature fabrication processes.[1], [2] These constraints drive up cost and limit wafer size and IR spectral range, making large-area, low-cost SWIR imaging arrays difficult to realize. Colloidal quantum dot (CQD) photodiodes offer a promising alternative by leveraging solution processing and bandgap tunability.[3]

Colloidal quantum dots (CQDs) are an attractive material system for infrared (IR) detection and imaging spanning the near-IR (NIR), short-wave IR (SWIR), and mid-IR (MIR) regimes.[3]-[6] Material systems such as PbS, HgTe, and Ag_2Te provide bandgap tunability through size control and surface chemistry while enabling low-temperature, solution-processed deposition over large areas on crystalline or non-crystalline substrates without the constraints for lattice matching.[7]-[10] In CQD thin films, ligand exchange and transport-layer selection govern interdot coupling, carrier polarity, and Fermi-level position, allowing electrical and spectral properties to be engineered at the film level.[11], [12] PbS-based devices, in particular, have achieved high room-temperature quantum efficiencies in the SWIR, underscoring the potential of solution-processed absorbers for scalable imagers.[13]-[15]

Despite their attractive properties, CQD imagers face some material challenges.[16]-[19] In particular, surface and interfacial trap states lead to significant dark current and noise, and the soft, solution-processed films are sensitive to subsequent processing steps, complicating monolithic integration with silicon readout electronics. These limitations underscore the need for device architectures and integration strategies that can simultaneously preserve SWIR sensitivity, suppress dark current, and maintain uniform performance over large areas.[20], [21]

Amorphous-oxide thin-film transistors (TFTs), notably In-Ga-Zn-O (IGZO) TFTs, offer uniform electrical characteristics, low leakage, and favorable subthreshold behavior on large substrates, suitable for back-end-of-line integration.[22]-[24] With appropriate passivation, oxide TFTs exhibit stable thresholds and minimal drift during operation and have been widely deployed as switches and buffers on display-class backplanes.[25]-[27] Their large-area manufacturability and intrinsically high input impedance make them well matched to solution-processed photodetectors.

Here, we employ a voltage-mode readout strategy in which the CQD photodiode operates under open-circuit-voltage (V_{oc}) conditions. In this configuration, the photogenerated V_{oc} is directly applied to the high-impedance gate of the IGZO TFT, which in turn converts the voltage signal into a current through its transconductance response.[28] Our design exhibits several key advantages: (i) It effectively suppresses the CQD photodiode's dark current by operating in an open-circuit configuration. (ii) It eliminates carrier transport issues, as the photo-generated signal (an open-circuit voltage, or V_{oc}) does not require carriers to traverse transport layers. (iii) Its output is proportional to the photon flux ($^{W}/_{cm^2}$) instead of the absolute power on each pixel, making it highly suitable for pixel scaling. (iv) It achieves high, controllable transducer gain, the ratio of TFT output current to optical power with a low background (dark) current.

To ensure integration process compatibility, we adopt an inverted PbS CQD photodiode structure where the n-type PbS CQD is atop the p-type CQD. The IGZO TFT is fabricated and passivated before CQD deposition. The CQD p/n junction absorber is subsequently deposited on top, having the anode of CQD junction in direct contact with the gate of the prefabricated TFT. This architecture keeps the absorber isolated from a continuous dark-current path, allowing gain and operating condition set by the TFT bias in the subthreshold regime. This design also ensures process compatibility and avoids any post-deposition and annealing steps that could degrade the CQD junction, while the passivated layer protects the TFT under the solution process of QD deposition.[29], [30] Using this architecture, we have demonstrated an optical power to current output transition gain of around 10,000 with low background current.

Temporal response is characterized by first measuring the TFT and CQD components separately, and then the integrated pixels. Small-signal frequency responses for the TFT (electrical gate excitation) and the CQD device (intensity-modulated optical excitation) are both flat to 1

MHz, the highest frequency our test setup supports. The integrated pixel, illuminated with the same modulated optical input and measured at the TFT output, retains a flat response up to the highest measured frequency of 1 MHz. These results confirm that the TFT and CQD dynamics is well suited for high frame rate imaging systems.

We evaluated the pixel's functionality by performing single-pixel imaging of a grayscale target. The reconstructed image reproduced the target with high fidelity, manifesting the wide dynamic range and low noise. Circuit-level modeling further confirmed array compatibility: a 4T-1C unit cell (functionally a five-transistor pixel with the TFT transducer) reliably sampled the CQD photovoltage and buffered it to the column for readout.[31], [32] Simulations captured the expected timing of reset, transfer, and readout, with node voltages settling within the sampling window, demonstrating that the pixel design is broadly applicable to TFT and CMOS ROIC architectures.[33]-[35]

Taken together, these results show that integrating PbS CQD junction operating in V_{oc} mode with the TFT provides a manufacturable path to low-cost, high-sensitivity SWIR imaging on large, panel-sized platforms. The approach supports MHz-class temporal response, enables straightforward pixel-level gain tuning, and is compatible with monolithic integration on scalable backplanes.

3.2 Results

3.2.1 Stack Selection for Gate-Coupled Integration Process

We use a monolithic integration process in which the transistor front end is fabricated and passivated prior to CQD deposition. This process sequence necessitates the use of an inverted CQD photodiode architecture, where the anode is deposited directly on the pre-patterned IGZO TFT

gate contact. In contrast, the standard (non-inverted) CQD diode structure has its anode on top of the cathode. These two configurations exhibit distinct characteristics, as depicted next.

Figure 3.1(a)(i) shows the schematic of a standard (non-inverted) CQD photodiode, where a ZnO electron transport layer is deposited on the ITO electrode, followed by a PbS-TBAI absorption layer and a PbS-EDT hole transport layer. The corresponding illumination-dependent current-density-voltage characteristics in Fig. 3.1(a)(ii) serve as the baseline for comparison. Figure 3.1(b)(i) illustrates the inverted device structure, in which the layer order is reversed and a 15 nm C_{60} interlayer is inserted between the sputtered ZnO and the PbS-TBAI layer to mitigate sputter-induced damage to the CQDs.[36]-[38] As shown in Fig. 3.1(b)(ii), the inverted device exhibits a higher dark current but similar photo response in comparison with the non-inverted device. In our architecture, the most relevant characteristics are photovoltage dependence on the photon flux, as shown in Fig. 3.1(c)(i). The standard (non-inverted) CQD photodiode achieves a higher V_{oc} across most of the measured range but saturates at a lower illumination level. In contrast, the inverted device produces a slightly lower maximum V_{oc} but exhibits a larger dynamic range. We adopt the inverted CQD photodiode structure primarily because the inverted structure is compatible with the monolithic integration flow with n-channel IGZO TFTs.

The open-circuit voltage (V_{oc}) versus incident power density is plotted for different device diameters (15 μm, 50 μm, and 100 μm) in Fig. 3.1(c)(ii). The overlapping V_{oc} curves confirm that the device performance is maintained during scaling and that perimeter effects are negligible.

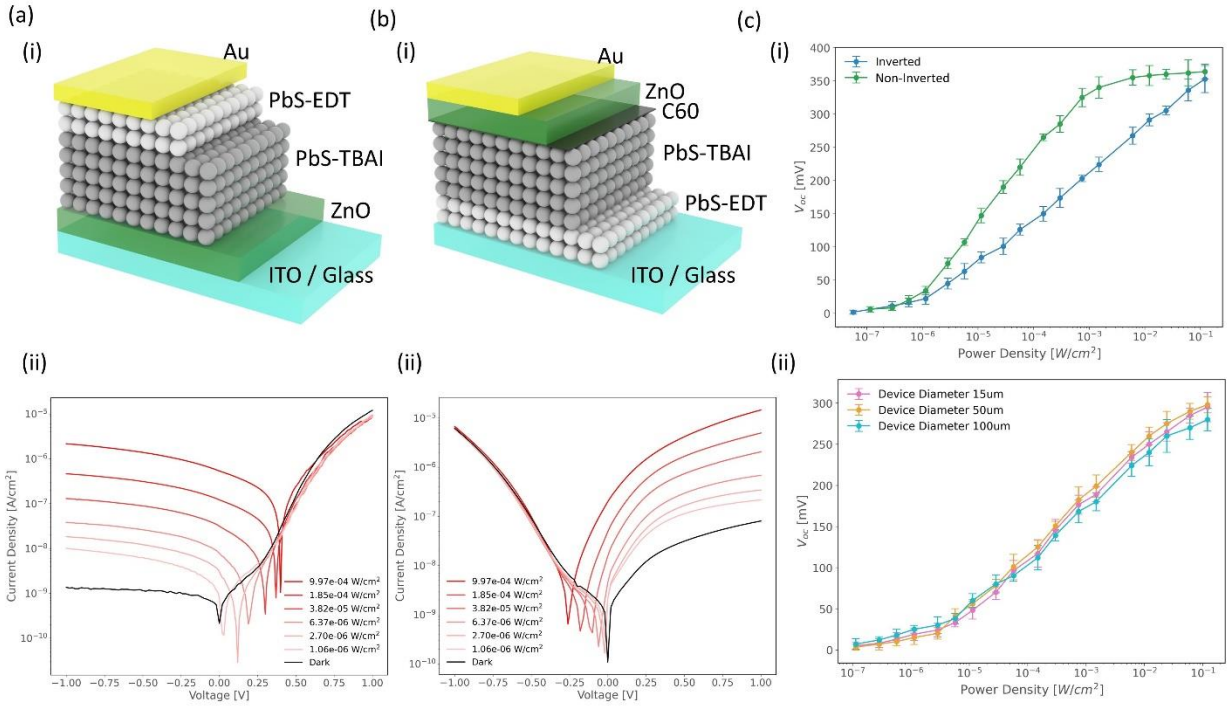

Figure 3.1 Comparison of non-inverted and inverted stack of Colloidal Quantum Dot (CQD) Photodiode. (a) Non-inverted stack of CQD Photodiode: (i) Artistic visualization of the non-inverted stack; (ii) I-V characteristics of the non-inverted CQD Photodiode under different illumination levels. (b) Inverted stack of CQD Photodiode: (i) Artistic visualization of the inverted stack; (ii) I-V characteristics of the inverted CQD Photodiode under different illumination levels. (c) Open circuit voltage (V_{oc}) measurement results: (i) Comparison of V_{oc} measurement results v.s. incident power density of the non-inverted stack and the inverted stack; (ii) Comparison of V_{oc} measurement results v.s. incident power density across different device diameters.

3.2.2 TFT Electrical Characteristics and Uniformity

Figure 3.2(a) shows the structure of the passivated IGZO TFT with 2um gate length, serving as the front-end transistor in our design. The stack consists of a patterned ITO gate electrode, a gate dielectric, an amorphous-IGZO oxide semiconductor channel with source/drain contacts, and a passivation layer that protects the TFT during subsequent CQD processing. The transfer characteristics are shown in Fig. 3.2(b), where I_{DS} - V_{GS} is plotted at stepped V_{DS}, exhibiting the typical turn-on and subthreshold behaviors. The I_{DS} - V_{DS} characteristics are plotted at stepped V_{GS} in Fig. 3.2(c). Uniformity across the 2-inch wafer, evaluated from 468 devices, is

summarized in Figs. 3.2(d)-(f). The overlaid transfer curves in Fig. 3.2(d) indicate decent process device uniformity and yield. The threshold-voltage histogram in Fig. 3.2(e) is centered at approximately -0.55 V with a spread of ~0.20 V across all measured sites. The wafer-scale map of extracted V_{TH}, shown in Fig. 3.2(f) reveals no systematic spatial trends, such as radial gradients or edge-related deviations. Figure 3.3 summarizes wafer-scale subthreshold-slope (SS) and mobility statistics for the TFTs. The histogram in Fig. 3.3(a) is centered at around 0.10 V/dec and exhibits a spread of ~0.02 V/dec across all measured sites. The mobility distribution is centered around $\sim 11 cm^2/(V \cdot s)$ with a spread of $\sim 7 cm^2/(V \cdot s)$ across the wafer. The wafer-level map in Fig. 3.3(b) and Fig. 3.3(d) shows no systematic spatial variation, indicating uniform process control across the substrate.

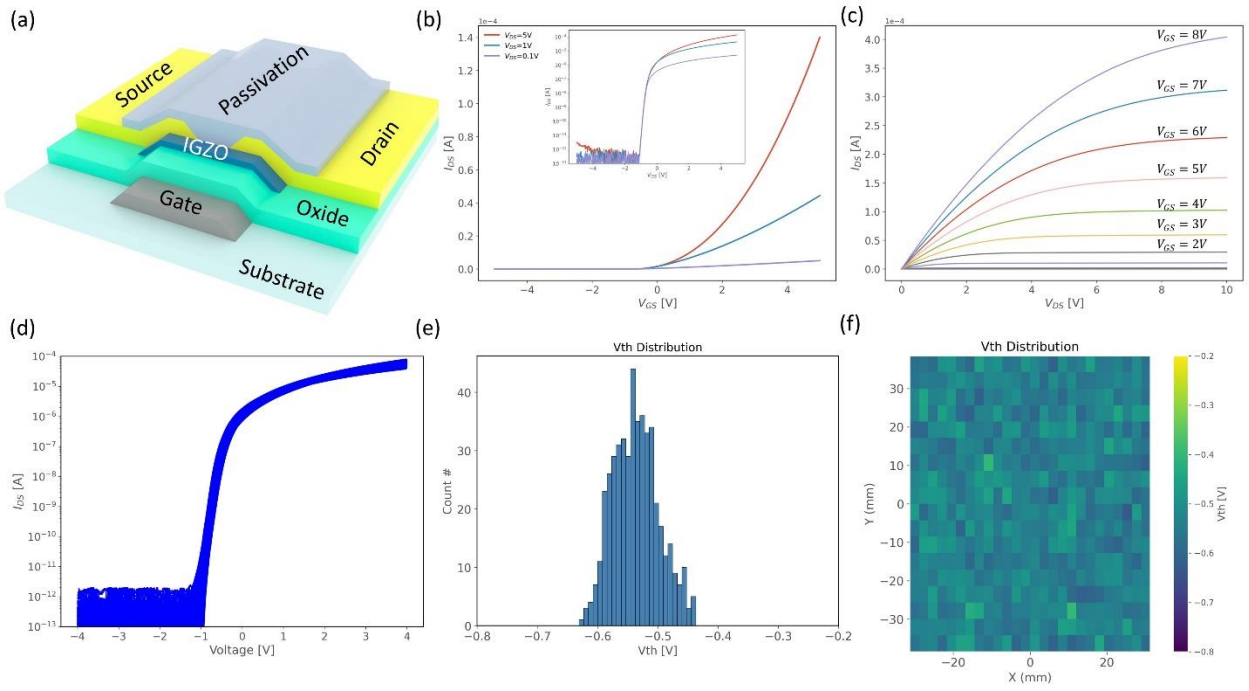

Figure 3.2 TFT devices statistics. (a) Artistic visualization of the TFT device structure. (b) Transfer curves of the TFT device under different V_{DS} bias levels (logarithmic scale and linear scale). (c) Characteristic curves of the TFT device under different V_{GS} bias levels. (d) Overlaid wafer-scale transfer curves of the spatially distributed TFT devices. (e) Histogram distribution of the threshold voltage. (f) Wafer-scale map of the threshold voltage.

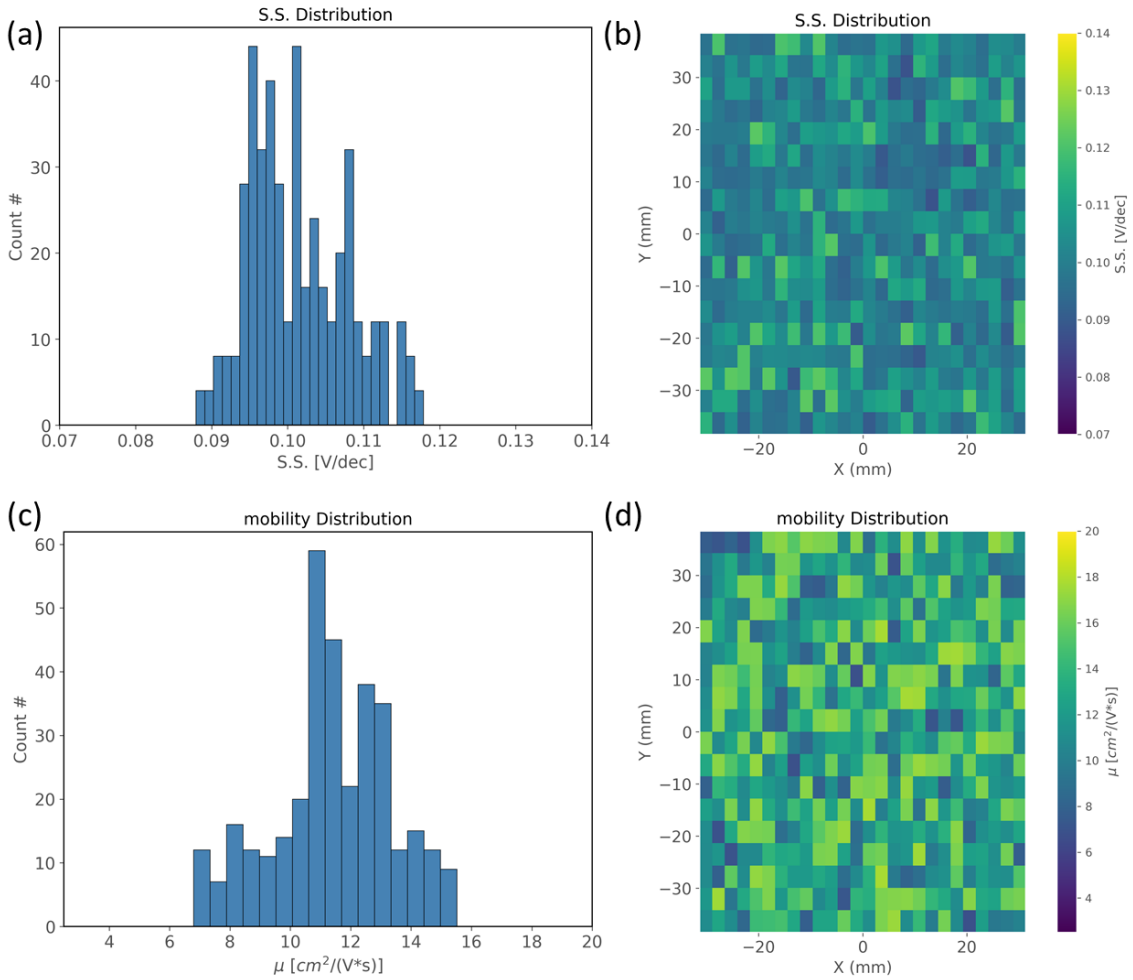

Figure 3.3 TFT wafer-scale subthreshold slope (S.S.) and mobility statistics. (a) Histogram of the distribution of S.S.. (b) Wafer-level map of S.S.. (c) Histogram of the mobility distribution. (d) Wafer-level map of mobility.

3.2.3 Integrated QD-TFT Pixel and Electrical Response

Figure 3.4(a) shows a schematic of the monolithically integrated CQD photodiode and IGZO TFT. The TFT features a lithographically defined gate beneath the IGZO channel. The exposed channel area is passivated by an Al_2O_3 layer, on which the spin-coated PbS quantum dot layer is formed. The p-type quantum dots are in direct contact with the TFT's molybdenum gate. Subsequently, the n-type quantum dots, a C_{60} layer, a ZnO electron transport layer, and the top

ITO contact are deposited using shadow masks. This stack forms an inverted CQD photodiode connected to the TFT gate, creating the circuit illustrated in Fig. 3.4(b). The quantum dot (QD) photodiode operates in open-circuit voltage (Voc) mode. Its output voltage is applied directly to the gate of the TFT, modulating the drain current. This combination of the QD photodiode and TFT converts the optical flux into a current signal. An external gate bias sets the TFT's operating point, enabling control of the conversion efficiency and gain.

Figure 3.4© shows the modulation of the integrated pixel's transfer curves by optical flux at 1060nm wavelength. This leftward shift indicates that under a fixed gate bias, the CQD-induced photovoltage changes the TFT's drain current by several orders of magnitude when the TFT operates in the subthreshold regime.

The characteristics in Fig. 3.4(d) were measured at three gate biases (-0.5 V, -0.7 V, and -0.8 V). At a more positive gate bias to move the TFT from the subthreshold regimes towards the saturation regime, the output current at a given photon flux increases whereas the baseline current of the TFT also increases. This bias-controlled gain allows weak optical inputs to be transduced into detectable currents with a decent contrast, underscoring the pixel's ability to resolve very weak infrared signals. When the gate is biased more negatively, the device maintains this sensitivity to low-light levels while simultaneously offering a broader dynamic range at the target irradiance level. To quantify this, we extract the effective responsivity (R) as a function of incident optical power for three bias conditions in Fig. 3.4(e), each corresponding to a different dark-current baseline. The pixel responsivity reaches a maximum of $10^4 A/W$ under low optical power (< pW) illumination. This occurs when the gate bias is set for a higher transconductance while the TFT is still in the subthreshold regime. The responsivity gradually decreases with increasing optical power. This behavior follows an approximation relationship of $I_d \propto P^{\gamma}$ where $\gamma \approx 0.3, 0.5, 0.8$

for $V_G = -0.5V, -0.7V, and\ -0.8V$, respectively. These results indicate that a more positive gate bias enhances amplification of weak optical signals for high gain, while a more negative gate bias provides a higher dynamic range.

The specific detectivity was calculated using $D^* = R\sqrt{Area}\Big/\sqrt{2qI_{dark}}$ and the extracted values are shown in Fig. 3.4(f) with a peak detectivity of 3.3×10^{13} Jones. For comparison, Zhou et al. reported a PbS CQD-gated carbon-nanotube TFT SWIR photodetector with a peak detectivity of 5.6×10^{13} Jones and a response time of 0.57 ms.[39] While the reported detectivity is slightly higher, their millisecond-level response time is significantly slower than the sub-microsecond operation demonstrated here. This difference arises from the longer channel transit time and increased trapping effects typically associated with carbon-nanotube conduction. In contrast, the high electron mobility and low trap density of IGZO TFTs enable our architecture to operate in the sub-microsecond regime, supporting MHz-class readout suitable for fast-frame SWIR imaging.

Next, we derive a simple model for the drain current of the integrated CQD-TFT pixel as a function of gate bias and optical power, which yields an analytic expression for the pixel responsivity.

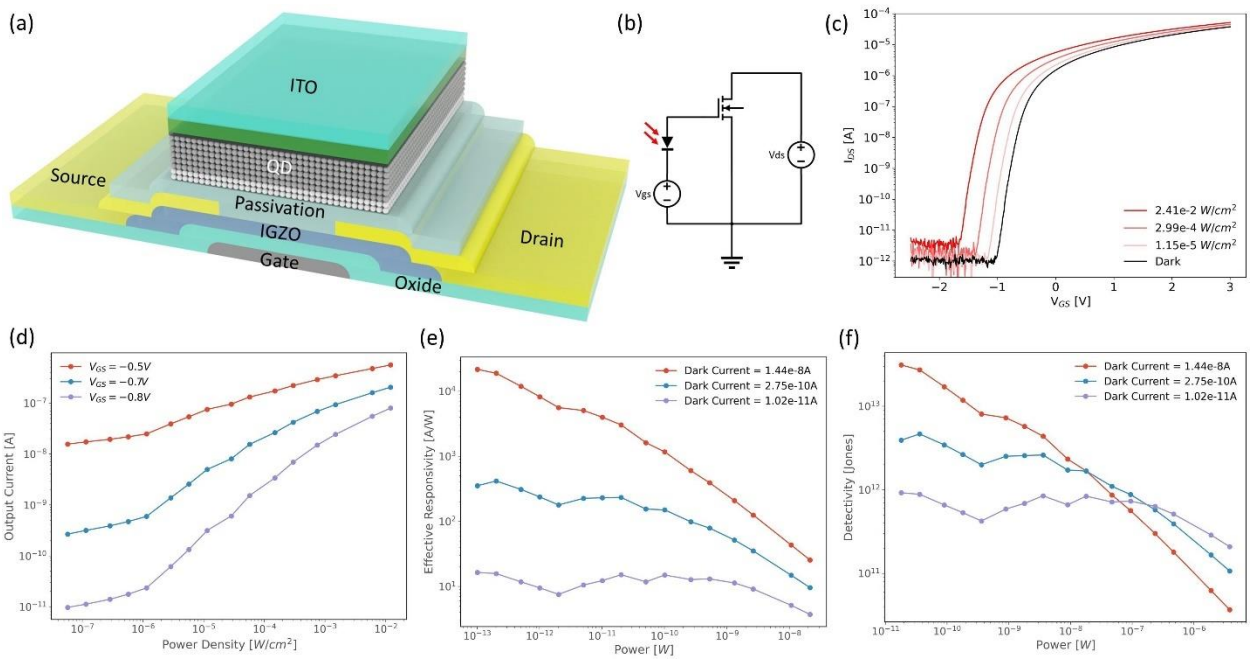

Figure 3.4 Integrated device characterizations. (a) Artistic visualization of the integrated device. (b) Representative operational circuitry for the integrated device. (c) Transfer curves of the integrated device under different incident illumination power densities. (d) The output current v.s. incident illumination power density under various gate bias levels. (e) The calculated effective responsivity v.s. incident illumination power density under various gate bias levels. (f) Specific detectivity v.s. incident illumination power density under various gate bias levels.

To interpret the trends in Fig. 3.4(d),(e), we model the gate-coupled pixel in the regime where the TFT operates in subthreshold and the PbS QD junction operates in V_{oc} mode. In the gradual channel approximation for a TFT in subthreshold, the sheet density of electrons in the channel varies exponentially with the local channel potential V(x):

$$n_s(x) = \frac{C_I}{q}\eta V_T \exp\left(\frac{V_{GS}-V_{TH}-V(x)}{\eta V_T}\right) \quad (1)$$

Where C_I is the gate-insulator capacitance per unit area, η is the subthreshold-slop factor, V_T is the thermal voltage kT/q, V_{TH} is the threshold voltage. With one-dimensional drift-diffusion transport:

$$I = qW\mu n_s(x)\frac{dV}{dx} + qWD\frac{dn_s}{dx},\ D = \mu V_T \quad (2)$$

Integration from source ($x = 0, V = 0$) to drain ($x = L, V = V_{DS}$) gives the long-channel subthreshold current:

$$I_{ds}^{\text{sub}} = \frac{W}{L}\mu C_I(\eta V_T)^2 \exp\left(\frac{V_{GS}-V_{TH}}{\eta V_T}\right)\left(1 - e^{-V_{DS}/V_T}\right) \tag{3}$$

Defining the bias-independent pre-factor $I_0 = \frac{W}{L}\mu C_I(\eta V_T)^2$, the small signal transconductance in subthreshold is

$$g_m = \frac{\partial I_{DS}}{\partial V_G}\Big|_{V_{DS}} = \frac{1}{\eta V_T} I_{ds}^{\text{sub}} = \frac{1}{\eta V_T} I_0 \exp\left(\frac{V_{GS}-V_{TH}}{\eta V_T}\right)\left(1 - e^{-V_{DS}/V_T}\right) \tag{4}$$

The QD p-n junction operated at open circuit obeys

$$I = I_s\left(e^{\frac{V_g}{\eta' V_T}} - 1\right) - I_{ph} \tag{5}$$

$$I_{ph} = \gamma\left(\frac{e}{h\nu_i}\right) P_i \tag{6}$$

where I_s and η' are the dark-saturation current and ideality factor, P_i is the incident optical power at the device, γ lumps optical coupling and quantum efficiency, and $h\nu_i$ is the photon energy. Setting $I = 0$ gives the photovoltage

$$V_g' = \eta' V_T ln\left(1 + \frac{I_{ph}}{I_s}\right) = \eta' V_T ln\left(1 + \frac{\gamma e}{h\nu_i I_s} P_i\right) \tag{7}$$

One can identify the baseline drain current, which can be treated as dark current of the pixel, at the chosen gate bias as

$$I_{ds,dark} = I_{ds}^{sub}\big|_{V_{GS}=V_{bias}} \tag{8}$$

The photo current with incident optical power of P_i at a gate bias point of V_{bias} can be expressed as

$$I_{ds,photo} = \frac{W}{L}\mu C_I(\eta V_T)^2 \exp\left(\frac{V_{bias}+\eta' V_T ln\left(1+\frac{\gamma e}{h\nu_i I_s}P_i\right)-V_{TH}}{\eta V_T}\right)\left(1 - e^{-V_{DS}/V_T}\right) \tag{9}$$

The effective responsivity can be calculated as:

$$R(P_i) = \frac{\Delta I_{ds}}{P_i} = \frac{\eta'}{\eta} I_{ds,dark} \frac{\ln(1+\alpha P_i)}{P_i} \quad (10)$$

where $\alpha \equiv \frac{\gamma e}{h\nu_i I_s}$

3.2.4 Small-Signal Frequency Response of Constituent and Integrated Devices

Figure 3.5 shows the frequency responses of the TFT, the QD photodiode, and the integrated QD-TFT device using small-signal measurements. For each case, a sinusoidal voltage or optical modulation of fixed amplitude within the operating regime was applied, and the resulting amplitude response was measured across the frequency range of interest. In these experiments, a transimpedance amplifier (TIA) both supplied the bias to the device and amplified the output current, which was then monitored on a spectrum analyzer to capture the frequency response. Panels (i) illustrate the measurement configurations, and panels (ii) present the corresponding frequency characteristics up to 1 MHz.

The frequency response of the IGZO TFT device was characterized by applying a 100 mVpp sinusoidal modulation to the gate while monitoring the drain current as shown in Fig. 3.5(a)(i). The resulting magnitude responses in Fig. 3.5(a)(ii) remain flat up to the MHz range for gate biases of 2 V, -0.3 V, -0.6 V, and -0.8 V, indicating no detectable input- or transit- limited pole over the band of interest. For the QD photodiode, as illustrated in Fig. 3.5(b)(i), a 1060 nm laser diode was intensity-modulated with a sinusoidal waveform at a fixed-offset optical power density, and the corresponding frequency response, as shown in Fig. 3.5(b)(ii), shows no noticeable roll-off within the measured band. Finally, the integrated CQD-TFT pixel was tested under the same modulated optical excitation as shown in Fig. 3.5(c)(i), where the photodiode-generated voltage drives the TFT gate to produce a drain current output. The measured responses

in Fig. 3.5(c)(ii) remain flat through the low-MHz range across different gate biases, demonstrating that the integrated device maintains a fast response over the target bandwidth.

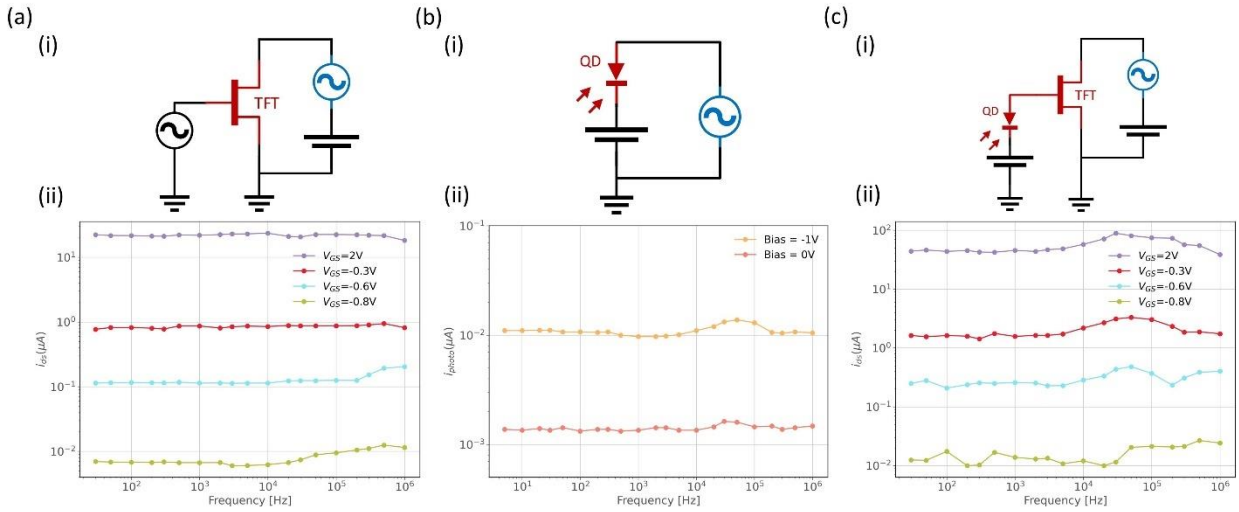

Figure 3.5 AC Measurement results. (a) Measurement results of the TFT IGZO device: (i) Measurement setup; (ii) AC response under various gate bias levels. (b) Measurement results of the PbS CQD device: (i) Measurement setup; (ii) AC response under different bias levels. (c) Measurement results of the integrated device: (i) Measurement setup; (ii) AC response under various gate bias levels.

Figure 3.6 outlines the AC measurement setup, which comprises two subsystems: an optical excitation path and an electrical acquisition path.

In the optical path, a fiber-coupled laser diode driven by a current controller provides the illumination. A function generator imposes a sinusoidal modulation on the driver to set the test frequency. The fiber output is collimated and focused onto the device through a movable convex lens, allowing adjustment of spot size and focus on the sample plane. Two beam splitters direct the beam to the device under test (DUT) while simultaneously relaying the field of view to a CCD camera for alignment. An objective positioned above the sample provides magnification and final focusing so that the beam is confined to the active area.

In the electrical path, the TFT gate is biased with a programmable DC source. The drain–source terminals are connected to a low-noise transimpedance amplifier; the drain is biased using the amplifier's internal bias, the AC photocurrent is converted to a voltage, and the output is routed to an RF spectrum analyzer to record magnitude response versus modulation frequency.

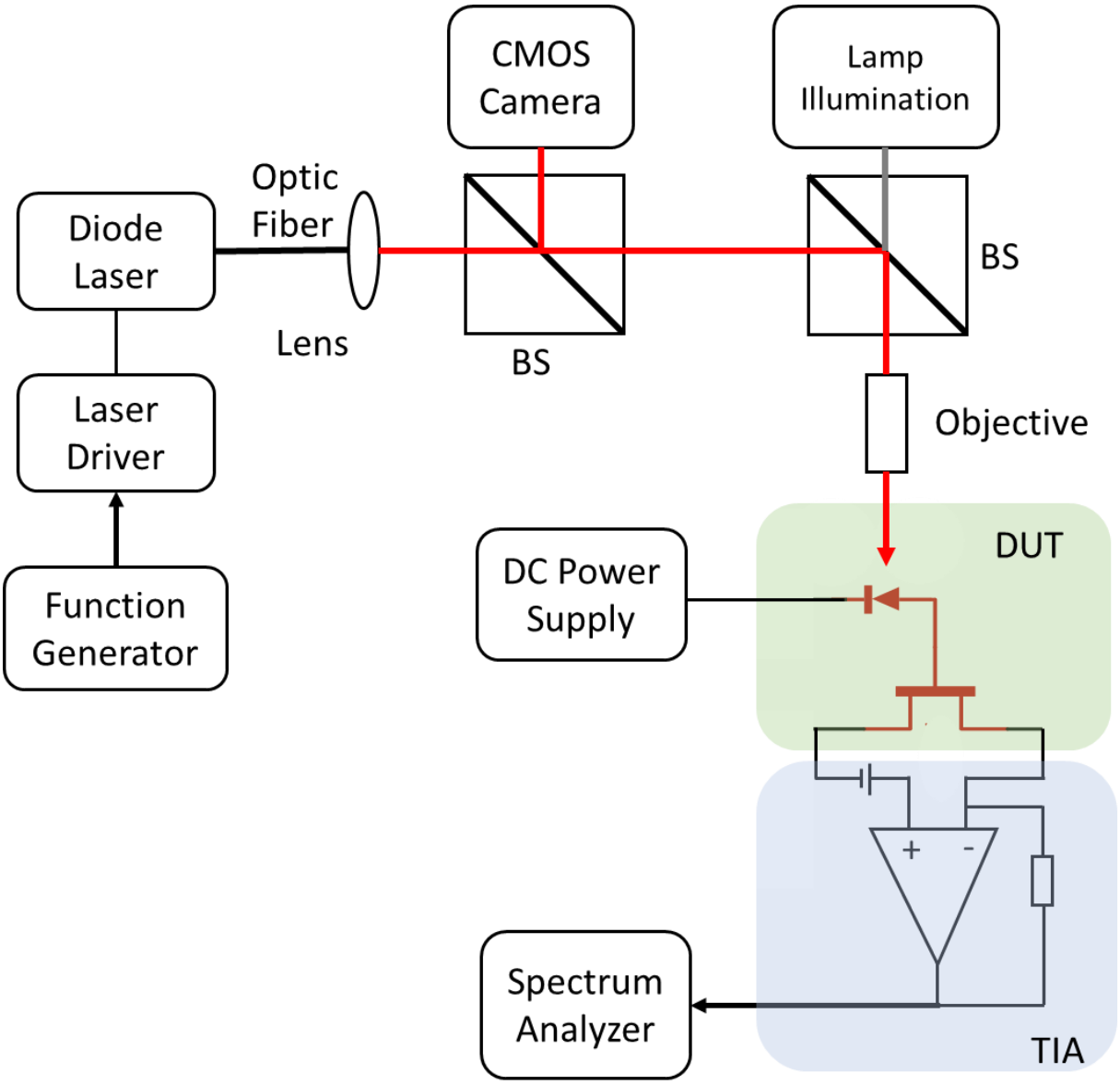

Figure 3.6 AC measurement setup for the integrated device. BS: Beam Splitter, DUT: Device Under Test, TIA: Trans-Impedance Amplifier.

3.2.5 Imgaing Setup and Result

The experimental imaging setup is shown in Figure 3.7. A pulse-modulated fiber-pigtailed 1060nm laser is used in this setup. Two beam splitters combine the laser path with LED illumination along a common optical axis, enabling simultaneous visualization of the laser spot and the device under test. A 10× objective focuses the co-propagating beams onto the sample. The integrated pixel is biased with a DC source at the gate, while the drain and source are connected to the Agilent B1500A parameter analyzer; the resulting currents are recorded by the B1500A during acquisition.

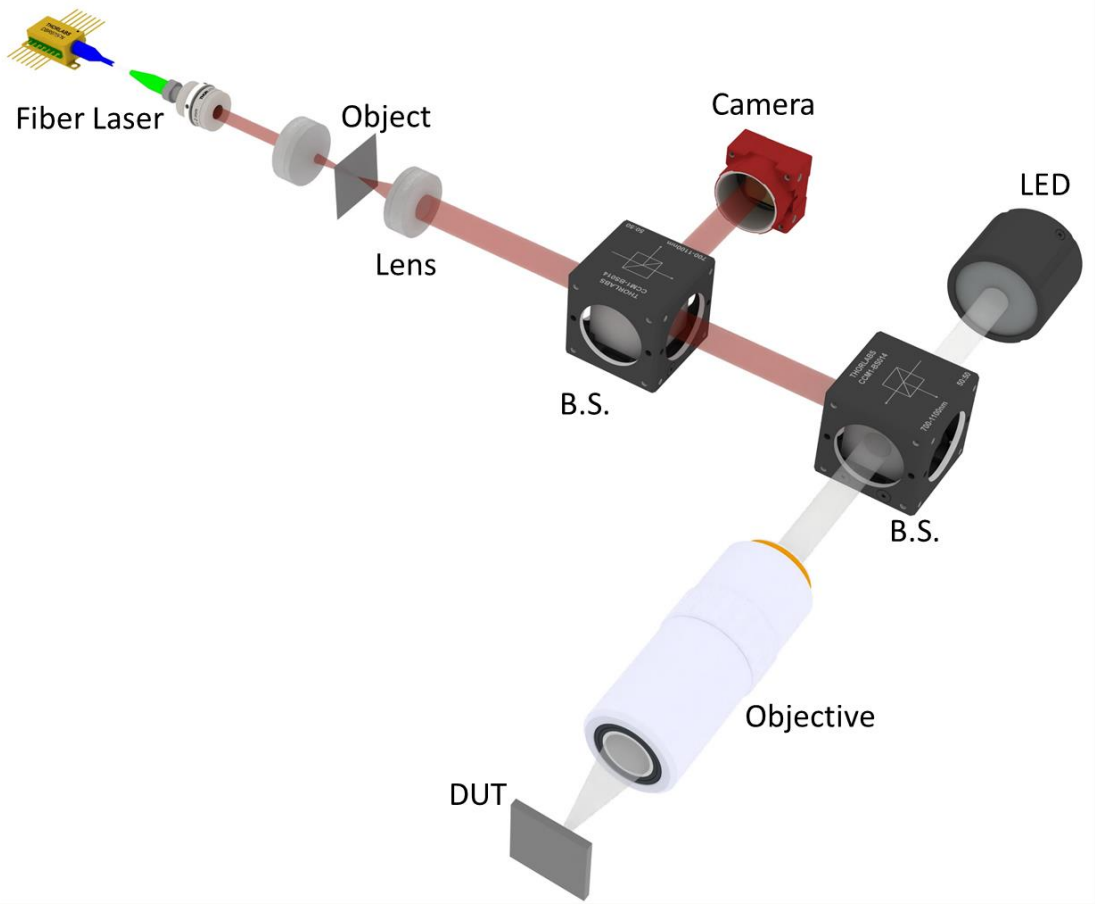

Figure 3.7 Artistic Illustration of the experimental imaging setup. B.S.: Beam Splitter; DUT: Device Under Testing.

A grayscale “Mickey Mouse” target was used to investigate the imaging capability of the integrated pixel. The target was scanned across the detector in a uniform spatial grid, and at each position the drain current was measured twice: once with the incident light on and once with the light off. Subtracting the background signal from the illuminated response and normalizing the result produced the final image. As shown in the reconstructed image in Fig. 3.8(a), the pixel faithfully reproduces both the shape and intensity variations of the target illuminated by an infrared (1060nm) light source.

A representative line profile along the $A - A'$ trajectory in Fig. 3.8(a) is shown in Fig. 3.8(b), where the per-sample drain current on a logarithmic scale was presented. The continuous drain-current response is shown in blue, while discrete markers identify the sampled light-off (green) and light-on (red) points at each pixel. The profile tracks the expected intensity variations across features, maintaining a clear separation between light-on and light-off currents over the entire scan. The result suggests that the integrated QD-TFT pixel is able to deliver a stable, high-speed response across the MHz frequency range.

The current prototype uses a TFT channel with W/L = 10um / 2um, and the photodiode active region (CQD absorber) has a diameter of 15um. Accounting for device footprint and routing, the integrated pixel can be implemented within a 25um X 25um area. This footprint represents the achievable spatial resolution of our present proof-of-concept design, and further scaling is expected to follow standard display-grade TFT design rules.

This demonstration validates the fundamental signal transduction mechanism of the integrated pixel. The present experiment uses a single-pixel scanning approach, and scaling to full-array imaging will require additional integration of interconnects, pixel addressing, and optical

fill-factor optimization. The successful proof-of-concept imaging confirms the readiness of the device architecture for these next development steps.

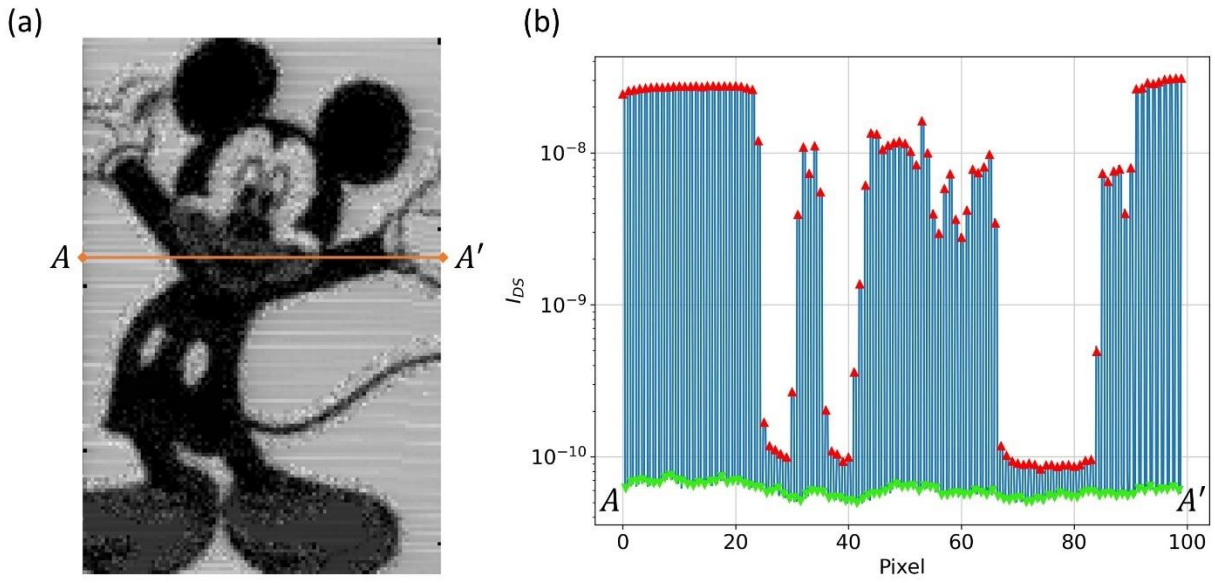

Figure 3.8 Imaging results. (a) Reconstructed "Micky Mouse" image from the captured data. (b) a representative line profile from along the $A - A'$ trajectory.

3.2.6 Circuit-Level Readout Modeling and Time-Domain Simulation

3.2.6.1 Generic Power-Law TFT expression

We consider a long-channel TFT described by the gradual-channel approximation (GCA). Let $V(x)$ denote the channel potential measured from source to drain with $V(0) = V_S$ and $V(L) = V_D$, width W, length L, and gate-insulator capacitance per unit area C_{ox}. The local gate overdrive is:

$$V_{ov}(x) = V_G - V_{TH} - V(x) \tag{11}$$

And the channel charge per unit area follows charge control, $Q(x) = -C_{ox}V_{ov}(x)$. Transport is assumed drift-dominated above threshold so that the local current density is $j(x) =$

$\mu_{eff}(V_{ov})Q(x)E(x)$ with $E(x) = -{}^{dV}/_{dx}$. In disordered semiconductor, trap filling and percolation lead to a geate dependent mobility. We adopt the generic power low

$$\mu_{eff}(V_{ov}) = \mu_0 V_{ov}^{\gamma} \tag{12}$$

where $\gamma > -2$ and μ_0 is a constant pre-factor (absorbing any characteristic voltage so that $[\mu_0] = cm^2V^{-(1+\gamma)}s^{-1}$, ensuring μ_{eff} has mobility units). The total current $I_D = Wj(x)$ is independent of x; substituting Q and E gives

$$I_D = \frac{W}{L}\int_0^L \mu_{eff}(V_{ov})C_{ox}V_{ov}(x)\frac{dV}{dx}dx = \frac{WC_{ox}\mu_0}{L}\int_{V_S}^{V_D}(V_G - V_{TH} - V)^{1+\gamma}dV \tag{13}$$

Introducing $u = V_G - V_{TH} - V$, have $du = -dV$ and transform the integral limits to $u_S = V_G - V_{TH} - V_S$ and $u_D = V_G - V_{TH} - V_D$, yielding the closed form

$$I_D = \frac{WC_{ox}\mu_0}{L}\frac{u^{2+\gamma}}{2+\gamma}|_{u_D}^{u_S} = \frac{WC_{ox}\mu_0}{L}\frac{(V_G - V_{TH} - V_S)^{\gamma+2} - (V_G - V_{TH} - V_D)^{\gamma+2}}{\gamma+2} \tag{14}$$

To ensure numerical smoothness and physical consistency near the threshold region, the hard cutoff function $max(0,\ V)$ traditionally used to enforce positive overdrive is replaced by a differentiable softplus formulation. Specially, we define

$$\Phi(V) \equiv V_{SS}\ln\left(1 + e^{V/V_{SS}}\right), \qquad V_{SS} > 0 \tag{15}$$

Which is infinitely differentiable, asymptotically equivalent to $max(0,\ V)$ for $|V| \gg V_{SS}$, and reduces to the rectified linear function V^+ in the limit $V_{SS} \to 0$. This substitution provides a smooth transition around threshold while preserving the correct asymptotic behavior in strong accumulation or inversion.

Accordingly, the source- and drain-end overdrive are regularized as

$$\tilde{V}_{ov,S} = \Phi(V_G - V_{TH} - V_S), \tilde{V}_{ov,D} = \Phi(V_G - V_{TH} - V_D) \tag{16}$$

Substituting $\tilde{V}_{ov,S/D}$ for the hard overdrive terms yields

$$I_{TFT} = \frac{W}{L}\frac{C_{ox}\mu_0}{V_\gamma^\gamma}\frac{(\tilde{V}_{ov,S})^{\gamma+2} - (\tilde{V}_{ov,D})^{\gamma+2}}{\gamma+2} \tag{17}$$

Applying the definition of $\Phi(V)$, the regularized overdrives can be expressed explicitly as

$$\tilde{V}_{ov,S}{}^{\gamma+2} = V_{SS}^{\gamma+2}\left[\ln\left(1+e^{(V_G-V_{TH}-V_S)/V_{SS}}\right)\right]^{\gamma+2},$$

$$\tilde{V}_{ov,D}{}^{\gamma+2} = V_{SS}^{\gamma+2}\left[\ln\left(1+e^{(V_G-V_{TH}-V_D)/V_{SS}}\right)\right]^{\gamma+2} \quad (18)$$

Substitute (S8) into (S7) and factoring out $V_{SS}^{\gamma+2}$ gives the compact soft-regularized current expression

$$I_{TFT} = \frac{W}{L}\frac{C_{ox}\mu_0}{V_\gamma^\gamma}\frac{V_{SS}^{\gamma+2}}{\gamma+2}\left\{\left[\ln\left(1+e^{(V_G-V_{TH}-V_S)/V_{SS}}\right)\right]^{\gamma+2} - \left[\ln\left(1+e^{(V_G-V_{TH}-V_D)/V_{SS}}\right)\right]^{\gamma+2}\right\} \quad (19)$$

3.2.6.2 Current-dependent series contacts

Non-ohmic contacts in oxide/organic TFTs are well captured by a current-softening series-resistance at source and drain:

$$R_c(I) = R_c + R_{c,max}\left(\frac{I_c}{I_c+|I|}\right)^{n_c} \quad (20)$$

where R_c is the minimum contact resistance (Ω) at one terminal in the high-current (ohmic) limit. This represents the baseline ohmic series resistance of the source/drain contact to the channel. $R_{c,max}$ is the contact resistance elevation at low currents (Ω). At $|I| \ll I_c$, the effective resistance tends toward $R_c + R_{c,max}$. This parameter quantifies the degree of current crowding or injection limitation. For typical oxide TFTs, a rule of thumb is $R_{c,max} = 3R_c$. Setting $R_{c,max} = 0$ recovers the simple constant-resistance case. I_c denotes the characteristic current scale (A) governing the transition from injection-limited to ohmic behavior. When $|I| \approx I_c$, the effective resistance starts to fall significantly from $R_c + R_{c,max}$ towards R_c. Physically, I_c represents the current density at which the contact barrier is sufficiently lowered by carrier injection; in practice, values are typically in the nA range for large-area oxide TFTs. n_c is the exponent controlling the sharpness

of the transition (dimensionless). Smaller values give a gradual roll-off of resistance with current; larger values make the drop sharper. Specific cases have physical interpretations:

So that the terminal voltage drops obey $V_S - V_{S,int} = R_c(I_S) * I_S$ and $V_D - V_{D,int} = R_c(I_D) * I_D$. at low $|I|$, $R_c \approx R_c + R_{c,max}$ (injection-limited); at high $|I|$, $R_c \to R_c$ (ohmic).

3.2.6.3 Extrinsic overlap and junction capacitances

Extrinsic parasitics are added as linear capacitors:

$$I_{GS} = C_{\text{ov}} \frac{d}{dt}(V_G - V_S),\ I_{GD} = C_{\text{ov}} \frac{d}{dt}(V_G - V_D),\ I_{DS} = C_{DS,\text{ov}} \frac{d}{dt}(V_D - V_S) \qquad (21)$$

Figure 3.9 shows the high-level implementation workflow of the TFT model. The model was implemented in Cadence Virtuoso using Verilog-A.

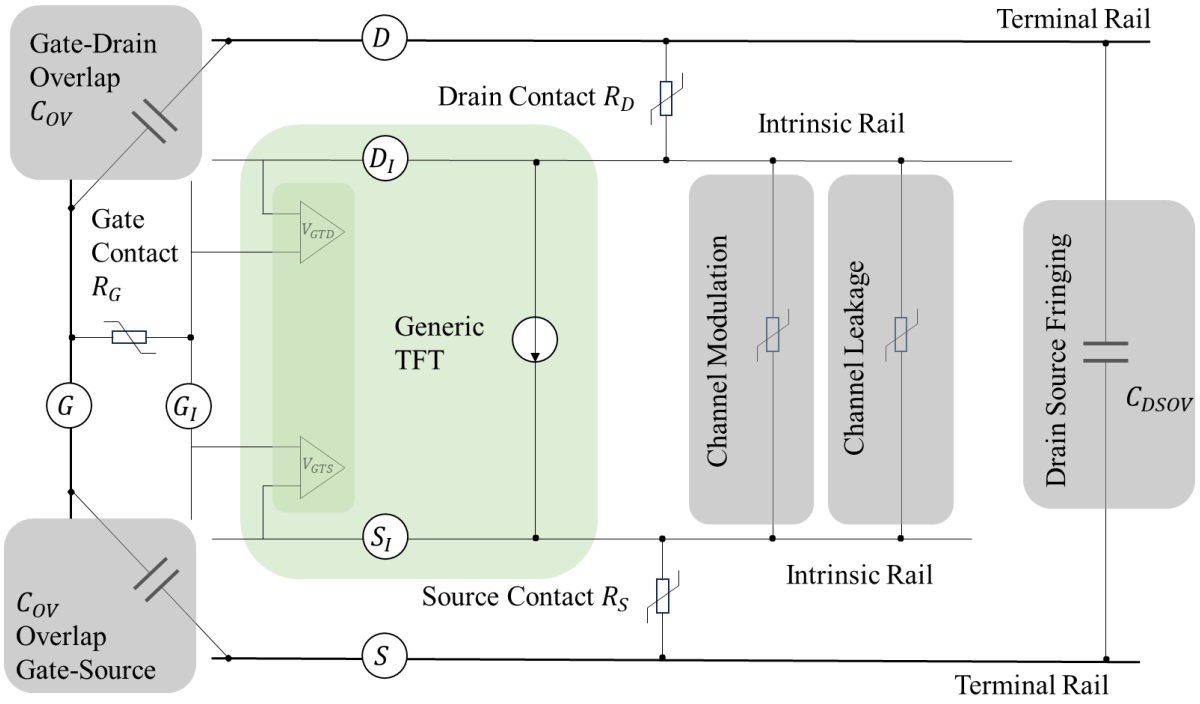

Figure 3.9 TFT model implementation high level workflow.

To extract the key model parameters, we begin by defining a function $H(V_G)$ that can be directly computed from the measured saturation current characteristics$I_D^{sat}(V_G)$. The function is defined as:

$$H(V_{\mathrm{G}}) = \frac{\int_0^{V_{\mathrm{G}}} I_{\mathrm{D\ sat}}(V_{\mathrm{G}})\mathrm{d}V_{\mathrm{G}}}{I_{\mathrm{D\ sat}}} \tag{22}$$

where $V_{G,0}$ is any gate voltage chosen above the threshold voltage so that the integral is well defined over the region of interest.

To develop an analytical approximation for $H(V_G)$, the integral in Eq. (22) can be separated into two components,

$$H(V_{\mathrm{G}}) = \frac{\int_0^{V_{\mathrm{T}}} I_{\mathrm{D\ sat}}(V_{\mathrm{G}})\mathrm{d}V_{\mathrm{G}} + \int_{V_{\mathrm{T}}}^{V_{\mathrm{G}}} I_{\mathrm{D\ sat}}(V_{\mathrm{G}})\mathrm{d}V_{\mathrm{G}}}{I_{\mathrm{D\ sat}}} \tag{23}$$

If V_G is selected such that the device operates in strong inversion, the first integral in Eq. (23) contributes negligibly because the drain current below threshold is several orders of magnitude smaller. As a result, $H(V_G)$ in strong inversion can be approximated solely by the second term,

$$H(V_{\mathrm{G}}) \approx \frac{\int_{V_{\mathrm{T}}}^{V_{\mathrm{G}}} I_{\mathrm{D\ sat}}(V_{\mathrm{G}})\mathrm{d}V_{\mathrm{G}}}{I_{\mathrm{D\ sat}}} \tag{24}$$

For the saturation current, one can express as:

$$I_{D,sat} = K(V_G - V_T)^m \tag{25}$$

Substituting the saturation current expression of Eq. (25) into Eq. (24), and assuming that the dependence of the transconductance parameter K on V_G is weak enough to be treated as approximately constant, we obtain

$$H(V_{\mathrm{G}}) = \frac{(V_{\mathrm{G}} - V_{\mathrm{T}})}{m+1} \tag{26}$$

This expression indicates that $H(V_G)$ is linear when plotted as a function of $(V_G - V_T)^{m+1}$. In the strong inversion region, this relation becomes effectively linear in V_G, where the slope determines the value of mmm, and the intercept with the V_G axis yields the threshold voltage V_T.

The complete extraction procedure therefore consists of three steps:

1. **Computation of $H(V_G)$:** Numerically evaluate Eq. (22) using the measured $I_{D,sat}$ data.
2. **Linear fitting:** Plot $H(V_G)$ in the strong inversion region and perform a linear fit. The slope of this fit provides the value of m, and the intersection with the V_G axis gives the extracted threshold voltage V_T according to Eq. (26).
3. **Extraction of K:** Once m and V_T are known, the transconductance parameter KKK can be obtained directly by rearranging Eq. (25) and fitting the measured saturation current.

Figure 3.10 shows the fitting result of $H(V_G)$. From the fitting, one can extract the corresponding m and V_T. For this experiment, fitted results show that $m = 2.29$ and $V_{\mathrm{T}} = -0.44V$. The full parameter list for the model simulation is shown in Table 3.1. Figure 3.11 compares the simulated waveforms with the experimental measurements.

The full Verilog-A TFT model was implemented as:

```
// VerilogA for tft, tft, veriloga
`include "constants.vams"
`include "disciplines.vams"
module tft_dg(G_T, G_B, D, S, testSIGNAL);
        parameter real version_TFT = 1.01; // TFT model version = 1.01
  electrical G_T, G_B, D, S; // terminals of the TFT for connection in circuits
  electrical GI,DI,SI; // internal terminals of the TFT channel, for inspection only
```

electrical virtualNode; // virtual node for summing capacitive currents of the quasi-static model. The node is grounded, since the sum of currents to this node is zero +/- numerical errors

electrical testGND, testSIGNAL; // test nodes for monitoring, isolated from the TFT model. See and modify end of the code to select and route quantity to the test terminals

parameter integer np=+1 from [-1:1] exclude 0; //TFT polarity, np=+1 for n-type TFT, np=-1 for p-type TFT

parameter real W=60.0e-6 from (0:inf); // channel width in [m] (default W=60um)

parameter real L=10.00e-6 from (0:inf); // channel length in [m] (default L=10um)

parameter real CI=0.0021 from [0:inf); // unit-area capacitance of the gate dielectric in [F/m^2]; 0.0021F/m^2 = 210nF/cm^2 (default,40nm Al2O3)

parameter real VTH=-0.4479 from (-inf:inf); // threshold voltage in [V] with its polarity (default VT=0 V)

parameter real uo=9e-4 from (0:inf); // (low-field) mobility in [m^2/Vs] at gate overdrive Vgamma=1V=np(VGgamma-VT). Default uo=9e-4 m^2/Vs=9 cm^2/Vs.*

parameter real Vgamma=1.0 from (0:inf); // Gate overdrive voltage in [V] for uo. Default, Vgamma = 1.0 V.

*parameter real gamma=0.1174 from [-1:inf); // Mobility enhancement factor [numerical exponent]. Default, gamma=max(0,2*To/T-2)=0.1174, for To~400K and T~300K.*

parameter real VSS=0.1 from (0:inf); // sub-threshold slope voltage in [V], VSS[V/Np]=dV/d(ln(ID))= 0.43(gamma+2)V/dec = V/dec for gamma=0.33.

parameter real epsT = 3.9 from (0:inf);

parameter real epsB = 10 from (0:inf);

parameter real toxT = 20.0e-9 from (0:inf);

parameter real toxB = 20.0e-9 from (0:inf);

parameter real VTTG = -0.5;

parameter real RC=100.0e3 from (0:inf); // minimum contact resistance at one contact (to the channel)

*parameter real RCmax=300.0e3 from [0:inf); // elevation of contact resistance at low currents. Typically, RCmax=3*RC. RCmax=0 for constant resistance=RC.*

parameter real ICmax=1e-9 from (0:inf); // max current for max contact resistance of value=(RC+RCmax). Typically, ICmax is in nA-range.

parameter real nIC=0.75 from [0.125:4]; // reduction exponent for RCmax. Meaningful values: nIC=0.5=SCLC; nIC=1=constant voltage drop.

parameter real RGmin=1.0 from (0:inf); // negligible gate contact resistance.

*parameter integer nSCLC=4 from [4:5] exclude 5; // Selector for channel modulation model. nSCLC=4 is only implemented: gamma model with CL=eB*eo/pi/L*

parameter real eB=2 from [0:inf); //relative permittivity of the medium in the back of the semiconducting film, eB~2, for channel modulation.

`define CL  eB * `P_EPS0 / `M_PI / L

parameter real eBleak=0.5 from [0:inf);

`define CLL  eBleak * `P_EPS0 / `M_PI / L

parameter real VTL=0 from (-inf:inf); // threshold voltage in [V] with its polarity for leakage, ideally VTL=VT (default VTL=0 V)

parameter real RBS=10.0e12 from (0:inf); // [Ohm/square] Sheet resistance of the bulk of the semiconducting film, for Ohmic leakage

`define Vmin 10.0e-6

*`define CG0 W*L*CI*

parameter integer selectQS=1 from [0:1]; // selectQS=0 suppresses the quasi-static capacitances. The quasi-static charges are still calculated.

parameter real LOV=30.0e-6 from [0:inf); // Geometrical overlap of gate conductor with drain/source contact pad. LOV is approximately the length of the pads perpendicular to channel width W

*`define COV CI*W*LOV*

parameter real eBov=1 from [0:inf); //relative permittivity of the medium in the back of the semiconducting film for geometrical capacitance between drain and source terminals, eBov~1 ~ (eB~2)/2.

`define CDSOV W * 2 * eBov * `P_EPS0 / `M_PI

// variables

real VGTS, VGTD; // for generic TFT charge drift DC model, etc. VGTS~np(VG-VT-VS) and VGTD~np*(VG-VT-VD) are overdrive voltages at source and drain ends of the intrinsic channel of the TFT*

real VSSS, VSSD; // sub-threshold slope voltage for VGTS and VGTD.

real IDSgen; // Channel current per the generic TFT charge drift DC model

real numerator, denominator, fDeltaID; // for channel modulation

real VSSSLF, VGTSLF, VSSDLF, VGTDLF; // VSS_ and VGT_for forward leakage diode. (See above VGTS, VGTD, VSSS and VSSD)

real VSSSLR, VGTSLR, VSSDLR, VGTDLR; // VSS_ and VGT_for reverse leakage diode. (See above VGTS, VGTD, VSSS and VSSD)

real contactIC, contactVC; // current and voltage across one contact of drain/source terminals

real QG,QS,QD; // quasi-static charges of gate, source and drain terminals

real ksiS, ksiD, ksiG; // normalized overdrive voltages for quasi-static charges

real den, numS, numD; // denominator and numerators of the normalized functions of QS and QD

real VT;

real CoxT, CoxB;

analog begin

/ ============ Generic TFT charge drift DC model ============ */*

*CoxT = epsT * `P_EPS0 / toxT;*

*CoxB = epsB * `P_EPS0 / toxB;*

*VT = VTH - (CoxT / CoxB) * (V(G_T) - VTTG);*

VSSS = sqrt(pow(VSS,2) + pow((np(V(GI) - VT - V(SI)) + abs(V(GI) - VT - V(SI)))/64,2));*

*VGTS = VSSS * ln(1+exp(np*(V(GI) - VT - V(SI))/VSSS));*

VSSD = sqrt(pow(VSS,2) + pow((np(V(GI) - VT - V(DI)) + abs(V(GI) - VT - V(DI)))/64,2));*

*VGTD = VSSD * ln(1+exp(np*(V(GI) - VT - V(DI))/VSSD)); // Drain overdrive voltage*

// Generic TFT charge drift DC model

*I(DI,SI) <+ np*W/L*uo/pow(Vgamma,gamma)*CI*(pow(VGTS,gamma+2)-pow(VGTD,gamma+2))/(gamma+2); // Generic TFT charge drift DC model*

/ ============ Contact model ============ */C*

*V(G_B,GI) <+ RGmin*I(G_B,GI); // negligible gate contact resistance*

*V(G_B,GI) <+ 0 *I(G_B,GI) * pow(ICmax/(ICmax + abs(I(G_B,GI))),nIC); // non-linear gate contact resistance is set to zero (ignored)*

*V(S,SI) <+ RC * I(S,SI); // constant (minimum) source contact resistance*

*V(S,SI) <+ RCmax*I(S,SI) * pow(ICmax/(ICmax + abs(I(S,SI))),nIC); // non-linear source contact resistance*

*V(D,DI) <+ RC * I(D,DI); // constant (minimum) drain contact resistance*

*V(D,DI) <+ RCmax*I(D,DI) * pow(ICmax/(ICmax + abs(I(D,DI))),nIC); // non-linear drain contact resistance*

/ ============ Channel modulation ============ */*

*I(DI,SI) <+ np*W/L*uo/pow(Vgamma,gamma)*6*`CL*(pow(VGTS,gamma+2)-pow(VGTD,gamma+2))/(gamma+2); // first component of channel modulation current*

*I(DI,SI) <+ W/L*uo/pow(Vgamma,gamma)*6*`CL*(V(DI)/2+V(SI)/2-V(GI)+VT)*(pow(VGTS,gamma+1)-pow(VGTD,gamma+1))/(gamma+1); // second component of channel modulation current*

*numerator = abs(pow(VGTS,3+2*gamma)-(3+2*gamma)*pow(VGTS,2+gamma)*pow(VGTD,1+gamma)+(3+2*gamma)*pow(VGTS,1+gamma)*pow(VGTD,2+gamma)-pow(VGTD,3+2*gamma));*

denominator = abs(pow(VGTS,2+gamma)-pow(VGTD,2+gamma)) + pow((1+3)`Vmin,2+gamma); // small voltage (1+3)*`Vmin~40uV added to avoid division on zero*

fDeltaID = numerator/denominator;

*I(DI,SI) <+ W/L*uo/pow(Vgamma,gamma)*6*`CL*(V(DI)-V(SI))*fDeltaID/2/(gamma+1)/(3+2*gamma);*

/ ============ Channel leakage ============ */*

*I(D,S) <+ V(D,S) * W / L / RBS;*

// Forward leakage diode

VSSSLF = sqrt(pow(VSS,2) + pow((np(V(D) - VTL - V(S)) + abs(V(D) - VTL - V(S)))/64,2));*

*VGTSLF = VSSSLF * ln(1+exp(np*(V(D) - VTL - V(S))/VSSSLF));*

VSSDLF = sqrt(pow(VSS,2) + pow((np(V(D) - VTL - V(D)) + abs(V(D) - VTL - V(D)))/64,2));*

*VGTDLF = VSSDLF * ln(1+exp(np*(V(D) - VTL - V(D))/VSSDLF));*

*I(D,S) <+ np*W/L*uo/pow(Vgamma,gamma)*6*`CLL*(pow(VGTSLF,gamma+2)-pow(VGTDLF,gamma+2))/(gamma+2);*

// Reverse leakage diode

VSSSLR = sqrt(pow(VSS,2) + pow((np(V(S) - VTL - V(S)) + abs(V(S) - VTL - V(S)))/64,2));*

*VGTSLR = VSSSLR * ln(1+exp(np*(V(S) - VTL - V(S))/VSSSLR));*

VSSDLR = sqrt(pow(VSS,2) + pow((np(V(S) - VTL - V(D)) + abs(V(S) - VTL - V(D)))/64,2));*

*VGTDLR = VSSDLR * ln(1+exp(np*(V(S) - VTL - V(D))/VSSDLR));*

*I(D,S) <+ np*W/L*uo/pow(Vgamma,gamma)*6*`CLL*(pow(VGTSLR,gamma+2)-pow(VGTDLR,gamma+2))/(gamma+2);*

/ ========== Quasi-static charge model ========== */*

ksiS = sqrt(pow(VGTS,2)+pow(`Vmin,2)) / (sqrt(pow(VGTS,2)+pow(`Vmin,2))+sqrt(pow(VGTD,2)+pow(`Vmin,2)));

ksiD = sqrt(pow(VGTD,2)+pow(`Vmin,2)) / (sqrt(pow(VGTS,2)+pow(`Vmin,2))+sqrt(pow(VGTD,2)+pow(`Vmin,2)));

ksiG = ksiS + ksiD;

den = pow((pow(ksiS,(2+gamma))/(2+gamma)-pow(ksiD,(2+gamma))/(2+gamma)),2);

*numS = (pow(ksiS,5+2*gamma)/(5+2*gamma))/(2+gamma)-(pow(ksiS,3+gamma)/(3+gamma))*(pow(ksiD,2+gamma)/(2+gamma))+(pow(ksiD,5+2*gamma)/(5+2*gamma))/(3+gamma);*

*numD = (pow(ksiS,5+2*gamma)/(5+2*gamma))/(3+gamma)-(pow(ksiS,2+gamma)/(2+gamma))*(pow(ksiD,3+gamma)/(3+gamma))+(pow(ksiD,5+2*gamma)/(5+2*gamma))/(2+gamma);*

QS = np(-`CG0)*(VGTS+VGTD)*sqrt((pow(numS,2)+pow(((1+ksiS)*`Vmin/6),2))/(pow(den,2)+pow(`Vmin,2)));*

QD = np(-`CG0)*(VGTS+VGTD)*sqrt((pow(numD,2)+pow(((1+ksiD)*`Vmin/6),2))/(pow(den,2)+pow(`Vmin,2)));*

QG = -(QS+QD);

*I(G_B,virtualNode) <+ ddt(QG)*min(1,selectQS);*

```
    I(S,virtualNode) <+ ddt(QS)*min(1,selectQS);
    I(D,virtualNode) <+ ddt(QD)*min(1,selectQS);
    V(virtualNode) <+ 0;

    /* ============ Overlap capacitances ============ */
    I(G_B,S) <+ `COV * ddt(V(G_B)-V(S));
    I(G_B,D) <+ `COV * ddt(V(G_B)-V(D));
    I(D,S) <+ `CDSOV * ddt(V(D)-V(S));

    /* ============ Monitor of quantities ============ */
        V(testSIGNAL) <+ 0;

  end // of "analog begin"
endmodule
```

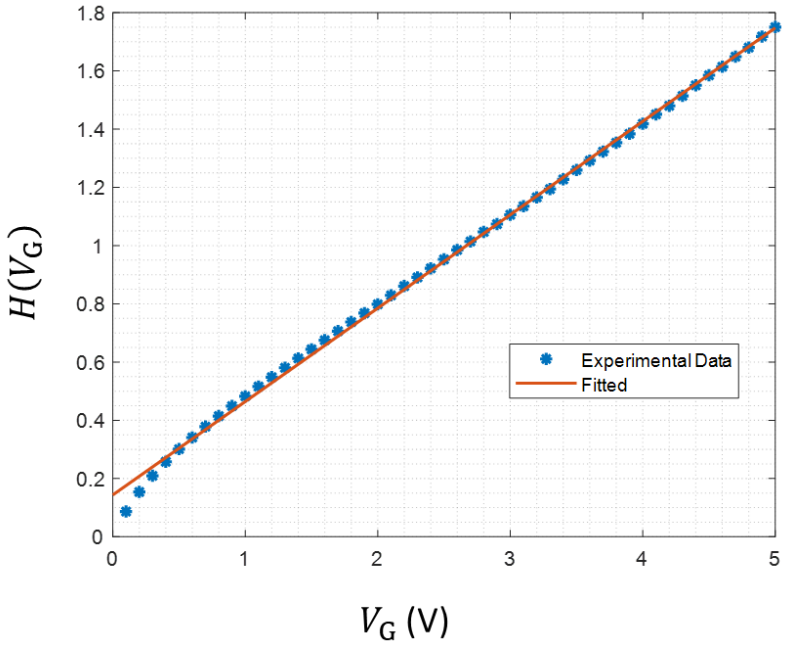

Figure 3.10 Fitting of $H(V_G)$ v.s. V_G using the experimental data.

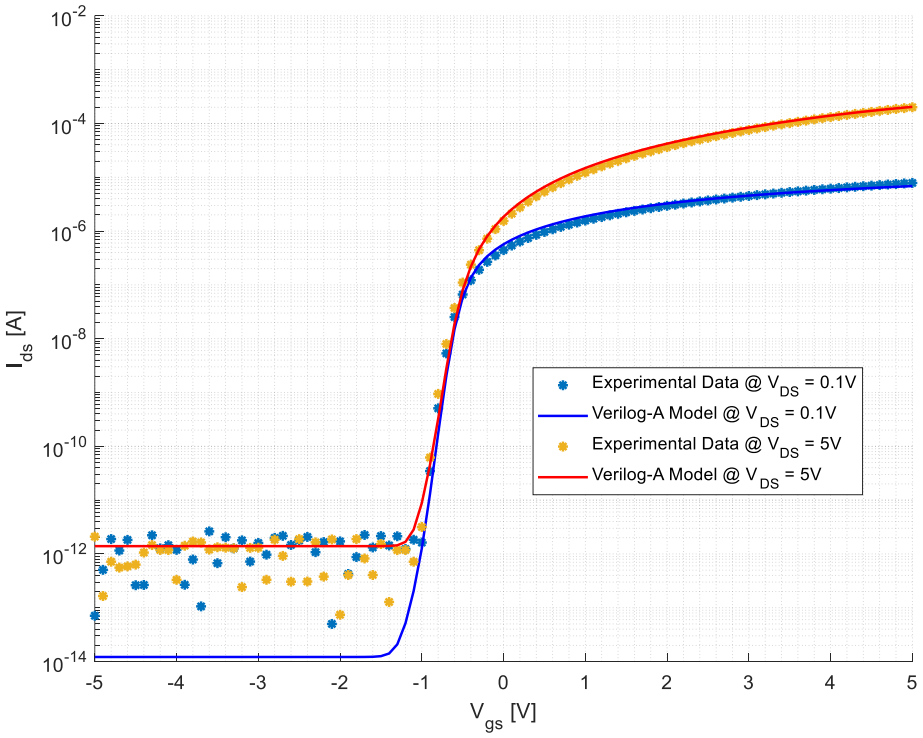

Figure 3.11 Comparison of the simulated and experimental transfer curves under different drain/source bias conditions.

Table 3.1 Extracted Parameters for Model Simulation

Model Parameter	Notation	Unit or Value	Default Value
Type of TFT	np	[-1, +1]	+1
Channel Width	W	m	60um
Channel Length	L	m	10um
Gate Dielectric Capacitance per Area	C_I	F/cm^2	0.0021
Threshold Voltage	V_T	V	-0.44
Low-Field Mobility	μ_0	m^2/Vs	1e-3
Mobility Enhancement Factor	γ	numeric	0.29
Subthreshold Slope Voltage	V_{SS}	V	0.08

Building on the device-level results, a natural next step is to integrate the sensing element with a column-addressable readout to enable array-scale operation. To this end, we evaluate our integrated device coupled to a representative ROIC topology and verify its behavior through circuit-level simulation. The unit cell follows a 4T-1C architecture, functionally a 5-transistor pixel when the CQD coupled TFT is included to enable open-circuit-voltage (V_{oc}) operation of the CQD photodiode. As shown in Fig. 3.12(a), a reset defines the baseline on the integration node V_{int}, a transfer switch connects the integrated device to V_{int}, and a source-follower/row-select pair buffers the level to the column for signal readout.

Figure 3.12(b) overlays the simulated control signals ($\overline{rst}$, tx, sel) with the principal node voltages V_{DS}, V_{int}, and the sampled output ($smpSig$). A reset pulse initializes V_{int} to the reference and potentially reset any readout capacitance; the subsequent transfer pulse produces the expected step on V_{int} that settles within the sampling window. The select phase forwards the buffered level to the column readout. Figure 3.12(c) maps incident power density to the sampled output voltage. The transfer is monotonic and approximately logarithmic over the mid-range, with the anticipated flattening at the lower intensities. Collectively, the simulations indicate that the V_{oc}-mode 4T-1C front end stably acquires the QD-TFT signal and delivers a well-conditioned irradiance-to-voltage conversion suitable for integration into array-level readout circuits.

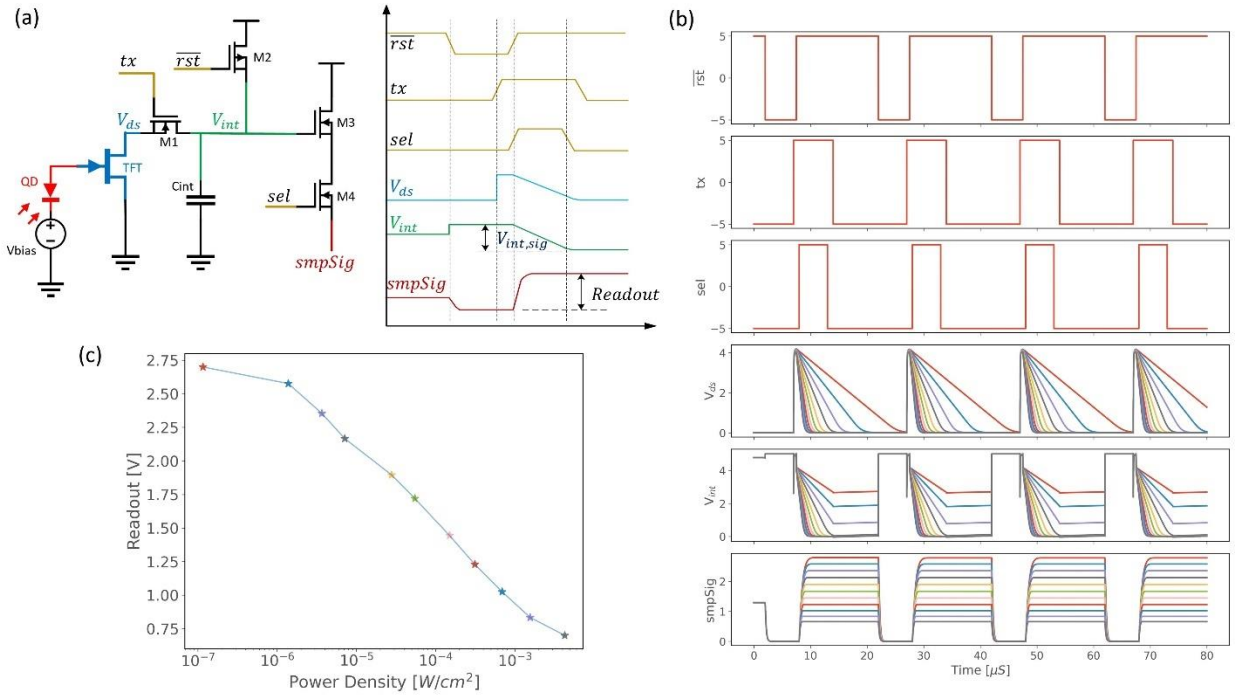

Figure 3.12 Circuit level readout modeling and simulation results. (a) Readout circuit architecture of the integrated device and the corresponding control signals and node voltages. (b) Simulation results of the node voltages under different incident illumination power. (c) Readout v.s. incident illumination power density.

3.3 Experimental Section

3.3.1 Thin-Film Transistor Fabrication

Bottom-gate, passivated a-IGZO TFTs were fabricated on glass using a standard photolithographic flow. Molybdenum (Mo) gate electrodes were deposited by DC sputtering at room temperature and patterned by photolithography followed by wet etching in a commercial Mo etchant. After resist strip and DI rinse, substrates were dehydrated at 120 °C. An Al_2O_3 gate dielectric was grown by thermal ALD using trimethylaluminum (TMA) and H_2O at 200 °C (growth rate ≈0.9 Å cycle^{-1}) to a thickness of 20 nm. The film serves as the only gate insulator directly above the Mo gate and establishes the gate-oxide capacitance used in device modeling.

The a-IGZO semiconductor was deposited by RF magnetron sputtering at room temperature under Ar/O_2 ambient (oxygen fraction a few percent), yielding a channel thickness in the 30 nm range. Channel was defined by photolithography and a dilute HCl/H_2O wet etch. The IGZO channel was post annealed on a hot plate in air at 300-350 °C for 1 h to activate the oxide and reduce defects prior to contact formation.

Source/drain contacts were realized by sputter deposition of indium tin oxide (ITO, ~70 nm) and lift-off. Probe pads (Ti/Au, 5/200 nm) were deposited by DC sputtering and patterned by lift-off. A single passivation layer of Al_2O_3 was deposited by plasma-enhanced ALD (PEALD) at 300 °C using TMA with an O_2 plasma co-reactant. Contact vias to gate, source, and drain were opened by wet etching using buffered oxide etchant (BOE) and a photoresist mask. Probe pads (Ti/Au, 5/200 nm) were deposited by DC sputtering and patterned by lift-off. A final air annealing at 300 °C for 1 h completed the TFT process and stabilized the threshold voltage and subthreshold characteristics.[40]

3.3.2 PbS Quantum Dot Synthesis and Spin Coating

PbS quantum dots with a first-exciton peak near 1060 nm were prepared by a modified hot-injection route using standard Schlenk techniques. Oven-dried glassware was assembled under inert gas, and a mixture of PbO (0.46 g, 2.06 mmol), oleic acid (2.4 g, 8.5 mmol), and 1-octadecene (10 mL) was charged to a 50 mL three-neck flask. The mixture was degassed at room temperature for 1 h, then heated to 120 °C under vacuum for an additional hour until a clear lead-oleate solution formed and residual moisture/oxygen were removed. The system was then placed under nitrogen while maintaining 120 °C.

Immediately before injection, a sulfur precursor was prepared by mixing bis(trimethylsilyl)sulfide, (TMS)2S (210 μL, ≈1.0 mmol), with 1-octadecene (8 mL) in a dry vial. The precursor was swiftly injected (≤2 s) into the hot lead-oleate solution under nitrogen. Heating and stirring were stopped at once to arrest further growth, and the reaction was allowed to cool naturally to room temperature.

Purification was carried out by three precipitation–redispersion cycles. PbS-OA solution was precipitated with acetone, centrifuged, and the supernatant removed. The precipitate was then redispersed in toluene, and the process was repeated. After the final cycle, the material was vacuum-dried and redispersed in octane to form the spin-coating solution used for device fabrication.

Small portions of the purified solution were analyzed by UV-Vis-NIR absorption spectroscopy to confirm the first-exciton peak at ≈1060 nm (Figure 3.13).

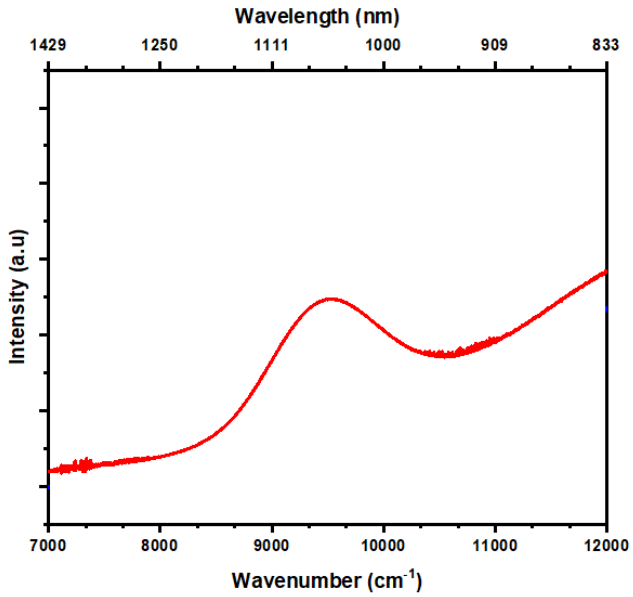

Figure 3.13 Absorption profile of the synthesized PbS Quantum Dot. An Absorption peak around 1060nm was observed.

Substrates were transferred from the cleanroom immediately after final passivation and gate-contact opening, rinsed with anhydrous isopropanol. After a 5-minute bake at 120 °C on a hotplate to remove adsorbed moisture, samples were allowed to cool to room temperature.

A 10 mg mL^{-1} PbS QD dispersion in octane (100-150 µL) was dispensed onto the substrate, spun at 2000 rpm for 15 s to form a uniform layer. A solid-state ligand exchange was performed directly on the spinning film by dispensing the ligand solution and maintaining for 30-45s before spin dry, and the sample was washed with mother solution 3 times to remove displaced oleate and excess ligand. For halide exchange, tetrabutylammonium iodide (TBAI, 10 mg mL^{-1} in anhydrous acetonitrile) was used; for short-chain crosslinking, 1,2-ethanedithiol (EDT, 0.02 vol% in acetonitrile) was used. Films were built up by repeating the cycle to reach the target thickness, with typical stacks formed from 6-12 cycles. A representative microscopic picture of the integrated device is shown in Figure 3.14.

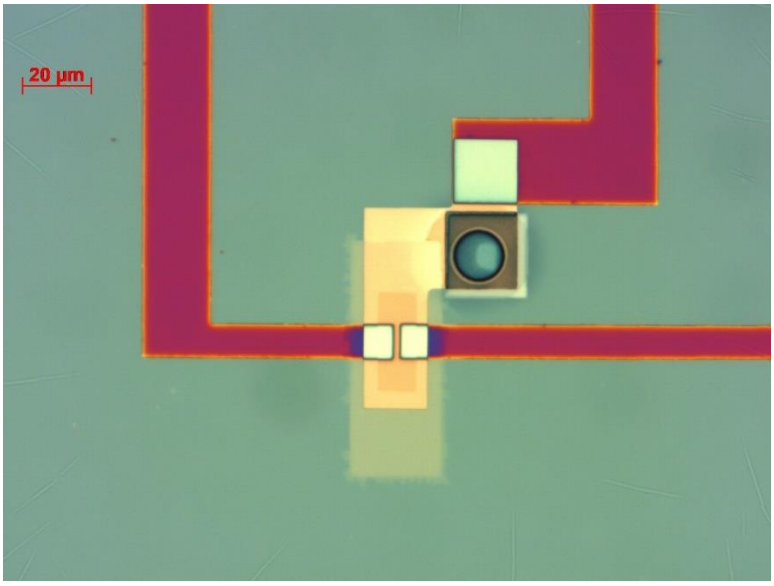

Figure 3.14 Microscopic image of the integrated pixel.

3.3.3 QD Integration Process Flow

After completing fabrication of the IGZO TFT, we deposited the CQD layers by spin coating. PbS CQDs with an absorption peak centered at 1060 nm and capped with oleic acid ligands were first dispersed in octane at a concentration of 10 mg mL^{-1}. For each spin-coating cycle, the PbS-OA solution was drop-cast onto the substrate, spun to form a uniform film, and subjected to a solid-state ligand exchange. In the case of the p-type transport layer, ligand exchange was performed using 0.02% 1,2-ethanedithiol (EDT) in acetonitrile: the film was soaked for 45 s, spun dry, and then rinsed three times with the mother solution to remove excess ligands. This cycle was repeated five times to yield an EDT-exchanged PbS layer with a total thickness of ~40-50 nm. The absorption layer was then deposited using the same procedure, with ligand exchange carried out in a 10 mg mL^{-1} tetrabutylammonium iodide (TBAI) solution in methanol, also repeated for five cycles.[41], [42]

A 15 nm C_{60} interlayer was deposited by thermal evaporation through a shadow mask, providing both efficient electron transport and protection of the CQDs from sputter damage. Using the same shadow mask, a ZnO electron transport layer was deposited by RF magnetron sputtering at room temperature in an Ar/O_2 atmosphere with 7% O_2. Subsequently, a smaller shadow mask was applied for the deposition of the ITO electrode by RF sputtering, ensuring direct contact to the exposed TFT gate pad while avoiding electrical shorts to neighboring electrodes.

3.3.4 Measurement Procedures

Electrical and optoelectronic characterizations of the individual QD devices, IGZO TFT device, and integrated pixels were performed using an Agilent B1500A Precision Semiconductor Parameter Analyzer. The high-resolution (HR) measurement unit was employed to ensure low-noise and precise current–voltage acquisition. For non-inverted IV measurements, either voltage sweeps or constant-voltage sampling modes were used to record the drain current. Open-circuit photovoltage (VOC) measurements were carried out in current-source mode, by sourcing 0 A and recording the resulting voltage.

For optical excitation, a 1060 nm single-mode fiber-coupled laser diode (LP1060-SF15A, Thorlabs) driven by a CLD1010LP Laser Diode Controller was used. The output beam was directed through a free-space optical path (details provided in the Supplementary Information) and focused onto the device active area with a spot diameter of approximately 150 μm.

Small-signal modulation experiments were performed by superimposing a sinusoidal drive from a Keysight 33612A Function/Arbitrary Waveform Generator. The modulation was applied either to the laser intensity for photodiode measurements or directly to the TFT gate for transistor AC characterization. The resulting photocurrent was amplified and converted to a voltage output

using an SR560 Low-Noise Voltage Preamplifier (Stanford Research Systems), which simultaneously provided the DC bias to the TFT drain and photodiode terminals. Frequency-domain responses were then recorded using a Keysight N9020A MXA Signal Analyzer.

3.4 Conclusion

We have demonstrated a monolithically integrated PbS QD-IGZO TFT device operated under photovoltaic mode. The IGZO TFT is fabricated and passivated before introducing the CQD layers, a sequence that safeguards the transistor's electrical performance while allowing the CQD photodetector to be deposited without exposure to processing steps that could degrade its properties. Under illumination, the CQD photodetector generates a light-dependent gate potential that directly drives the TFT, which then provides transconductance-based voltage-to-current amplification. The design can produce a large and tunable responsivity of greater than 10^4 A/W for an input optical power of sub-picowatt. It also obtains a large dynamic range and a flat frequency response of beyond 1 MHz. These results highlight the promise of integrated CQD-TFT pixels for low-cost, large-area IR imaging.

ACKNOWLEDGEMENTS

Chapter 3, in full, is a reprint of the material as it appears in Jiang, Y., Arya, S., Peng, H., Davis, A., and Lo, Y.H., 2025. Integration of Self-Quenched Active-Recovery SPADs in Standard CMOS. IEEE Access, accepted. The dissertation author was the first author of this paper.

REFERENCE

[1] E. Thimsen, B. Sadtler, and M. Y. Berezin, “Shortwave-infrared (swir) emitters for biological imaging: a review of challenges and opportunities,” Nanophotonics 6, 1043–1054 (2017).

[2] B. Zhu and H. Jonathan, “A review of image sensors used in near-infrared and shortwave infrared fluorescence imaging,” Sensors 24, 3539 (2024).

[3] J. Liu, P. Liu, D. Chen, T. Shi, X. Qu, L. Chen, T. Wu, J. Ke, K. Xiong, M. Li, et al., “A near-infrared colloidal quantum dot imager with monolithically integrated readout circuitry,” Nature Electronics 5, 443–451 (2022).

[4] D. Shin, Y. Park, H. Jeong, H.-C. V. Tran, E. Jang, and S. Jeong, “Exploring the potential of colloidal quantum dots for near-infrared to short-wavelength infrared applications,” Advanced Energy Materials 15, 2304550 (2025).

[5] T. Nakotte, S. G. Munyan, J. W. Murphy, S. A. Hawks, S. Kang, J. Han, and A. M. Hiszpanski, “Colloidal quantum dot based infrared detectors: extending to the mid-infrared and moving from the lab to the field,” Journal of Materials Chemistry C 10, 790–804 (2022).

[6] E. Lhuillier, S. Keuleyan, and P. Guyot-Sionnest, “Colloidal quantum dots for mid-ir applications,” Infrared Physics & Technology 59, 133–136 (2013).

[7] J. Ouyang, N. Graddage, J. Lu, Y. Zhong, T.-Y. Chu, Y. Zhang, X. Wu, O. Kodra, Z. Li, Y. Tao, et al., “Ag2te colloidal quantum dots for nearinfrared- ii photodetectors,” ACS Applied Nano Materials 4, 13587–13601 (2021).

[8] S. Keuleyan, E. Lhuillier, V. Brajuskovic, and P. Guyot-Sionnest, “Midinfrared hgte colloidal quantum dot photodetectors,” Nature Photonics 5, 489–493 (2011).

[9] K. A. Sergeeva, H. Zhang, A. S. Portniagin, E. Bossavit, G. Mu, S. V. Kershaw, S. Ithurria, P. Guyot-Sionnest, S. Keuleyan, C. Delerue, et al., “The rise of hgte colloidal quantum dots for infrared optoelectronics,” Advanced Functional Materials 34, 2405307 (2024).

[10] A. De Iacovo, C. Venettacci, L. Colace, L. Scopa, and S. Foglia, “Pbs colloidal quantum dot photodetectors operating in the near infrared,” Scientific reports 6, 37913 (2016).

[11] S. Coe-Sullivan, J. S. Steckel,W.-K.Woo, M. G. Bawendi, and V. Bulovi´c, “Large-area ordered quantum-dot monolayers via phase separation during spin-casting,” Advanced Functional Materials 15, 1117–1124 (2005).

[12] J. Y. Kim, V. Adinolfi, B. R. Sutherland, O. Voznyy, S. J. Kwon, T. W. Kim, J. Kim, H. Ihee, K. Kemp, M. Adachi, et al., “Single-step fabrication of quantum funnels via centrifugal colloidal casting of nanoparticle films,” Nature communications 6, 7772 (2015).

[13] S. Zhou, Y. Wang, C. Deng, P. Liu, J. Zhang, N. Wei, and Z. Zhang, “Highly sensitive swir photodetector using carbon nanotube thin film transistor gated by quantum dots heterojunction,” Applied Physics Letters 120 (2022).

[14] X. Yin, C. Zhang, Y. Guo, Y. Yang, Y. Xing, and W. Que, “Pbs qd-based photodetectors: future-oriented near-infrared detection technology,” Journal of Materials Chemistry C 9, 417–438 (2021).

[15] A. Stavrinadis, D. So, and G. Konstantatos, “Low-temperature, solutionbased sulfurization and necking of pbs cqd films,” The Journal of Physical Chemistry C 120, 20315–20322 (2016).

[16] S. Kumar and S. Pradhan, “Colloidal quantum dot-based near and shortwave infrared light emitters: Recent developments and application prospects,” Advanced Optical Materials 12, 2400993 (2024).

[17] J. Kim, J. Roh, M. Park, and C. Lee, “Recent advances and challenges of colloidal quantum dot light-emitting diodes for display applications,” Advanced Materials 36, 2212220 (2024).

[18] F. P. García de Arquer, D. V. Talapin, V. I. Klimov, Y. Arakawa, M. Bayer, and E. H. Sargent, “Semiconductor quantum dots: Technological progress and future challenges,” Science 373, eaaz8541 (2021).

[19] H. Jung, N. Ahn, and V. I. Klimov, “Prospects and challenges of colloidal quantum dot laser diodes,” Nature Photonics 15, 643–655 (2021).

[20] Y. Qin, T. Guo, J. Liu, T. Lin, J. Wang, and J. Chu, “Colloidal quantum dots in very-long-wave infrared detection: Progress, challenges, and opportunities,” ACS omega 8, 19137–19144 (2023).

[21] L. Yu, P. Tian, and K. Liang, “Advancements and challenges in colloidal quantum dot infrared photodetectors: Strategies for short-wave infrared, mid-wave infrared, and long-wave infrared applications,” Quantum Beam Science 9, 9 (2025).

[22] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, “Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors,” nature 432, 488–492 (2004).

[23] Y. Zhu, Y. He, S. Jiang, L. Zhu, C. Chen, and Q. Wan, “Indium–gallium–zinc–oxide thin-film transistors: Materials, devices, and applications,” Journal of Semiconductors 42, 031101 (2021).

[24] T. Kamiya, K. Nomura, and H. Hosono, “Present status of amorphous in–ga–zn–o thin-film transistors,” Science and Technology of Advanced Materials (2010).

[25] K. Nomura, “Recent progress of oxide-tft-based inverter technology,” Journal of Information Display 22, 211–229 (2021).

[26] J. Sheng, H.-J. Jeong, K.-L. Han, T. Hong, and J.-S. Park, “Review of recent advances in flexible oxide semiconductor thin-film transistors,” Journal of Information Display 18, 159–172 (2017).

[27] Z. Pan, Y. Hu, J. Chen, F. Wang, Y. Jeong, D. P. Pham, and J. Yi, “Approaches to improve mobility and stability of igzo tfts: a brief review,” Transactions on Electrical and Electronic Materials 25, 371–379 (2024).

[28] J. Tang and E. H. Sargent, “Infrared colloidal quantum dots for photovoltaics: fundamentals and recent progress,” Advanced materials 23, 12–29 (2011).

[29] S.-H. Kim, Y.-S. Kim, T. Hwang, T. H. Kim, H. Koo, J. S. Park, and J.-S. Park, “Reliability improvement of high mobility oxide tfts based on hydrogen-resistant peald al2o3 gate insulators grown with n2o plasma,” ACS Applied Materials & Interfaces 17, 14168–14178 (2025).

[30] D. A. Mourey, M. S. Burberry, D. A. Zhao, Y. V. Li, S. F. Nelson, L. Tutt, T. D. Pawlik, D. H. Levy, and T. N. Jackson, “Passivation of zno tfts,” Journal of the Society for Information Display 18, 753–761 (2010).

[31] Y. Zhang, X. Lu, G. Wang, Y. Hu, and J. Xu, “Modeling random telegraph signal noise in cmos image sensor under low light based on binomial distribution,” Chinese Physics B 25, 070503 (2016).

[32] S. B. Mahato, J. De Ridder, G. Meynants, G. Raskin, and H. Van Winckel, “Measuring intra-pixel sensitivity variations of a cmos image sensor,” IEEE Sensors Journal 18, 2722–2728 (2018).

[33] P. Martyniuk and A. Rogalski, “Quantum-dot infrared photodetectors: Status and outlook,” Progress in quantum Electronics 32, 89–120 (2008).

[34] A. D. Stiff-Roberts, “Quantum-dot infrared photodetectors: a review,” Journal of Nanophotonics 3, 031607 (2009).

[35] V. Pejovi´c, E. Georgitzikis, J. Lee, I. Lieberman, D. Cheyns, P. Heremans, and P. E. Malinowski, “Infrared colloidal quantum dot image sensors,” IEEE Transactions on Electron Devices 69, 2840–2850 (2021).

[36] E. Klem, C. Gregory, G. Cunningham, S. Hall, D. Temple, and J. Lewis, “Planar pbs quantum dot/c60 heterojunction photovoltaic devices with 5.2% power conversion efficiency,” Applied Physics Letters 100 (2012).

[37] D. M. N. M. Dissanayake, R. A. Hatton, T. Lutz, R. J. Curry, and S. R. P. Silva, “The fabrication and analysis of a pbs nanocrystal:c60 bilayer hybrid photovoltaic system,” Nanotechnology 20, 245202 (2009).

[38] J. W. Ryan, J. M. Marin-Beloqui, J. Albero, and E. Palomares, “Nongeminate recombination dynamics–device voltage relationship in hybrid pbs quantum dot/c60 solar cells,” The Journal of Physical Chemistry C 117, 17470–17476 (2013), https://doi.org/10.1021/jp4059824.

[39] S. Zhou, Y. Wang, C. Deng, P. Liu, J. Zhang, N. Wei, and Z. Zhang, “Highly sensitive swir photodetector using carbon nanotube thin film transistor gated by quantum dots heterojunction,” Applied Physics Letters 120 (2022).

[40] Y. Tang, Y. Song, and K. Nomura, “Long-term sustainability of highmobility ultrathin a-igzo tft with double-layered passivation,” ACS Applied Materials & Interfaces 17, 47170–47179 (2025), pMID: 40763264, https://doi.org/10.1021/acsami.5c08848.

[41] B. K. Jung, H. K. Woo, C. Shin, T. Park, N. Li, K. J. Lee, W. Kim, J. H. Bae, J.-P. Ahn, T. N. Ng, and S. J. Oh, “Suppressing the dark current in quantum dot infrared photodetectors by controlling carrier statistics,” Advanced Optical Materials 10, 2101611 (2022), https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/adom.202101611.

[42] B. K. Jung, T. Park, Y. K. Choi, Y. M. Lee, T. H. Kim, B. Seo, S. Oh, J. W. Shim, Y.-h. Lo, T. N. Ng, and S. J. Oh, “An ultra-sensitive colloidal quantum dot infrared photodiode exceeding 100,000% external quantum efficiency via photomultiplication,” Nanoscale Horiz. 9, 487–494 (2024).

Chapter 4 A Physics Based Unified Circuit Model for Single Photon and Analog Detector

Photodetectors having an internal carrier multiplication mechanism such as avalanche detectors and recently discovered detectors with cycling excitation process (CEP) have attracted tremendous interests for their high sensitivity and flexibility of being operable in analog, sub-Geiger and Geiger mode. Detection of single or few photons in Geiger mode or sub-Geiger mode is particularly interesting as LiDAR systems are finding a wide range of applications in autonomous driving, augmented and virtual reality, robotics, imaging, sensing, and communications. In this paper, we present a universal detector equivalent circuit model, applicable to all modes of operation, for photodetectors with carrier multiplication gain. The bias-dependent gain, bandwidth, as well as gain buildup dynamics are rooted in device physics and translated into an equivalent circuit model that can be readily incorporated into integrated circuit design. The model simulates the characteristics of devices biased below and above breakdown voltage, making seamless transitions between analog, sub-Geiger, and Geiger mode with the same set of parameters directly obtained from the material properties. As a result, circuit designers can choose detectors using different gain medium (e.g. Si, InP, InAlAs) to simulate device effects on system performance. The circuit model is implemented in Orcad PSpice circuit simulator. Its wide applicability is demonstrated by simulating the device in linear, sub-Geiger and Geiger mode with external amplifiers and quenching circuits, as well as the gain-bandwidth product of detectors with a Si and InAlAs gain medium.

4.1 Introduction

Semiconductor photodetectors with inherent carrier multiplication gain such as avalanche detectors and more recently, detectors with cycling excitation process (CEP) produce high sensitivity

in both conventional analog mode and in sub-Geiger mode or Geiger mode [1]–[3]. The latter gives rise to single photon sensitivity and has played an important role in quantum communications, deep space communication, bio-photonic and medical imaging, and LiDAR systems [4]–[8]. The LiDAR systems, in particular, have attracted tremendous interests because of their applications in autonomous driving, robotics, and augmented and virtual reality [9], [10]. While avalanche photodetectors in analog mode can be modeled in photoreceiver circuits for optical communications with a relatively simple equivalent circuit model, the physical operation of avalanche photodetectors in sub-Geiger mode and Geiger mode is more complicated and is difficult to have an adequate equivalent circuit that relate the circuit parameters to the physical parameters of the devices and materials.

In most prevalent equivalent circuit models for Geiger-mode detectors, a single-photon avalanche detector (SPAD) is modelled as an on-off switch with additional resistors, voltage sources and capacitors [11]–[14]. All parameters in the equivalent circuit are assigned from measured I-V response without relating them to the physical characteristics of the devices. For detectors made of different gain media (e.g. Si and InP) or having different designs, one cannot choose the equivalent circuit parameters to reflect such differences. On the other hand, there are a number of device models for SPADs that describe the device behaviors [15]–[18], but these device physics models, analytical or semi-empirical, cannot be directly converted into an equivalent circuit model and used for circuit simulations. In some cases, statistical behaviors including dark count rate and afterpulsing probability of SPADs were analytically modelled and added to the equivalent circuit model in hardware description language [19], [11], [12]. However, the parameters of the model are inconvenient to use for circuit designers as they are not attainable from the basic material properties or device geometries. In addition, almost all circuit models for SPADs are limited to Geiger mode

operation or for some specific quenching circuit design [20]. On the other hand, when designers need to make tradeoffs between sensitivity and dynamic range for different system requirements, the design can involve detectors that operate between Geiger mode, sub-Geiger mode, and analog mode.

Therefore, it is most desirable to create a unified circuit model for detectors with avalanche or CEP gain that work for all modes of operations, i.e., analog mode, sub-Geiger mode, and Geiger mode. This is equivalent to a transistor circuit model that works for the transistor to be in linear, saturation, or subthreshold conditions. Additionally, we wish to have all the parameters of the equivalent circuit related to and obtained from the device geometries and material properties. In this manner, as the designer chooses detectors made of different materials or having different geometries, an equivalent circuit model for the chosen detector types becomes readily available to show how different detectors can affect the system performance.

In this paper, we present a unified equivalent circuit model for photodetectors with built-in carrier multiplication gain such as avalanche photodetectors (APDs) and cycling excitation process (CEP) detectors. The equivalent circuit model has two salient features: (i) all the parameters in the equivalent circuit model can be obtained from the material parameters and rooted in device physics, and (ii) the circuit model works for all modes of operation (analog, sub-Geiger-mode, and Geiger-mode) and the transition between different modes is smooth without artificial separation of the circuit models between different modes of operation. When the device bias is switched between above-breakdown and below-breakdown voltage, the changes in device behaviors occur naturally, according to device physics. For example, below breakdown, the detector characteristics are subject to the gain-bandwidth limit while above breakdown (i.e. in Geiger mode), the device properties are no longer related to the gain-bandwidth product. Another unique feature is that the model treats dark current and photon current as individual electrical pulses, each of which contains a single photon or

carrier, thus enables simulations of quenching and recovery times, bias-dependence and frequency dependence of single-photon detection efficiency, and many statistic characteristics of the device. For instance, the model can show how increased dark current reduces the single-photon detection efficiency and shows circuit designers the limit of device dark current above which the detector can no longer detect single photons under Geiger-mode.

Generally, semiconductor single-photon detectors contain a carrier multiplication layer as gain medium. For conventional SPADs, the carrier multiplication layer has a p-n or p-i-n structure, and for CEP detectors, the gain medium is a thin disordered medium such as undoped amorphous-silicon (a-Si). During operation, a voltage is applied to reverse bias the p/n junction or the disordered layer, allowing photo generated electrons (holes) to gain sufficient energy from the applied bias, and subsequently generate secondary electron-hole pairs via the effect of impact ionization or cycling excitation process (CEP), the latter of which also involves localized states and phonons [4], [21], [22]. The device dynamics such as gain build-up time and device response time depends on the multiplication process. Depending on the magnitude of applied bias relative to the DC breakdown voltage, a detector with a carrier multiplication gain can operate in analog mode where the bias voltage is substantially below the breakdown voltage, or in sub-Geiger mode where the bias voltage is very close to the breakdown voltage, or in Geiger mode where the bias voltage is greater than the breakdown voltage. To achieve single-photon sensitivity, the detector usually needs to operate in Geiger mode, but SPADs in Geiger mode suffer from dark count, afterpulsing, optical crosstalk, and low dynamic range [23], [24]. On the other hand, detectors in sub-Geiger mode are able to detect few or even single photons with less serious negative effects associated with Geiger-mode detection. However, detectors in sub-Geiger mode are limited by the gain-bandwidth product and excess noise and require very precise bias and temperature control. In Geiger-mode operation, detectors require

passive or active quenching when operating continuously. The dynamic switching of the detectors between above breakdown and below breakdown bias prefer a unified circuit model applicable to all bias voltage and showing the dynamics of carrier multiplication. All parameters in our circuit model are related to fundamental material and design parameters: impact ionization coefficients for electrons and holes based on the local-field model [25], [26], and the carrier transit time across the multiplication region. The electron/hole impact ionization coefficients for commonly used semiconductors can be found from many published sources (Table I), and the carrier transit time is related to the specific device structure and can be estimated from the thickness of the gain region. In section II, we describe in detail how an equivalent circuit model from the physical picture is derived. In section III, we discuss how to apply the model under different conditions.

4.2 Model Derivation and PSPICE Implementation

The current of a photodetector is composed of particle current and displacement current. Particle current is due to the flow of electrons and holes generated by photon absorption, dark current, and carrier multiplication such as impact ionization or CEP effect [1], [21]. Displacement current is due to charging/discharging of the device capacitance (C_j). We describe the model for the particle current next.

For illustration purpose, here we assume that the photoexcited carriers are produced in a p-doped layer next to the carrier multiplication region. As a result, being minority carriers in p type semiconductors, photoexcited electrons are injected into the gain medium to initiate the multiplication process. We also assume that only one photo- or thermally excited carrier is injected into the gain region within the carrier transit time (t_r) across the high-field gain region, typically 1 to 10ps depending on the device structure. For example, if the device has a primary dark current

of 1.6pA, there is one thermal electron in the gain region every 100 ns, which is much longer than the transit time. This satisfies the condition that at most one electron is injected into the gain region to initiate carrier multiplication within a transit time, which is regarded as the smallest time unit in our model. In case the incident light is strong, and the above condition is not satisfied, we can divide a single detector into units of small detectors in parallel such that each detector unit meets the above criterion. In most cases of single photon detector, however, the light intensity is low enough that it is not necessary to divide the detector into multiple units for circuit simulations.

Figure 4.1(a) depicts a series of carrier generation events triggered by the injection of a single electron produced by photon absorption or thermal generation. It is assumed that the first electron is injected into the gain region at t=0 and it generates M_n second generation e-h pairs within a duration of t_r. Subsequently, these generated holes at t_r excite the 3rd generation e-h pairs with a multiplication factor M_p when time reaches $2t_r$, while the original electrons produced in the first t_r period leave the cathode as particle (electron) current. At $3t_r$, the 3rd generation electrons excite new e-h pairs with a multiplication factor M_n while the holes produced in the previous cycle leave the anode as particle (hole) current. To summarize the above process, we have electrons experience gain during the time intervals from $2Kt_r$ to $(2K+1)t_r$ and holes experience gain from $(2K+1)t_r$ to $(2K+2)t_r$, where $K = 0, 1, 2, 3$ .... Notably, as the above cycling process, initiated by a single electron injection at t=0, goes on, new photo- or thermally excited carriers can be injected into the gain region to start their own carrier multiplication cycles at different times, and the net particle current is the summation of all these events. Here we have simplified the carrier multiplication process by setting the smallest time increment to be carrier transit time and assuming that under high E-field, electrons and holes have the same transit time or saturation velocity. The first approximation can be improved by choosing a smaller time interval (e.g. a

fraction of transit time), and the second approximation can be easily changed by replacing t_r with $t_{r,n}$ and t_{rp} (i.e. transit time for electrons and holes). The mathematical expressions will be more complicated, but the SPICE implementation is essentially the same.

The above process can be represented by a block diagram in Fig. 4.1(b), in which M_n and M_p are single-path carrier multiplication factor for electrons and holes, and their value depends on the voltage drop in the gain region. Next, we will derive the expressions for gain, particle current, and gain-bandwidth product.

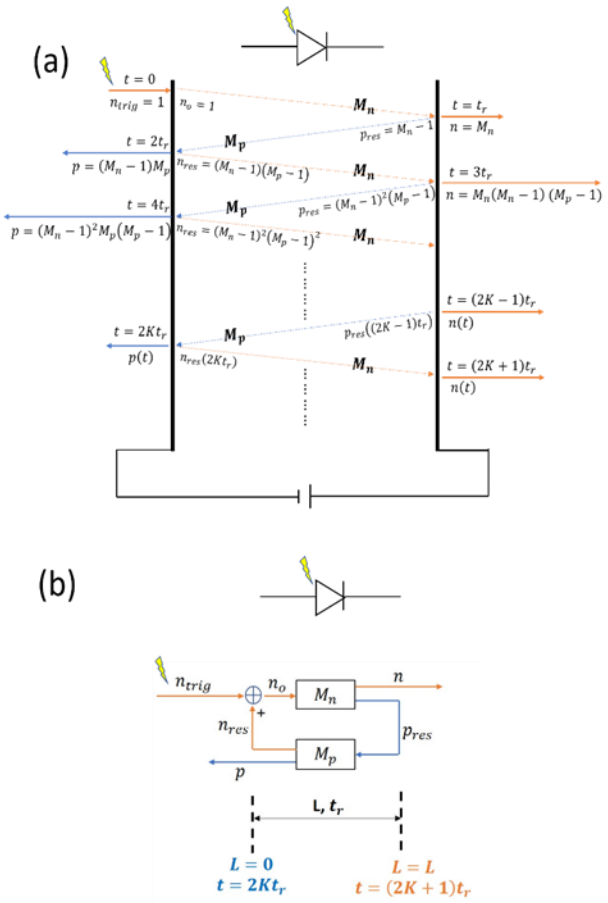

Figure 4.1 Carrier multiplication processes in the multiplication region of the device. (a) time evolution of a single triggering event and (b) block diagram. Here, t_r is the carrier transit time, L is the length of multiplication (gain) region, n_{trig} represents the triggering event (i.e. primary photoexcited or dark carrier injected into the gain region only at the even multiples of t_r (n_{trig} is 0 or 1 for a single electron), n_o is the total number of electrons experiencing the gain, $M_{n(p)}$ is multiplication factor for electrons(holes), $n(p)_{res}$ is the number of electrons(holes) remaining in the gain region, and n(p) is the number of electrons(holes) leaving the device within $2t_r$ to produce the particle current, K (K=0, 1, 2…) is a measure of time in the unit of $2t_r$.

Prior to the first triggering event at t=0, we set the initial state of the device as follows:

$$n_{res}(0) = 0, p_{res}(0) = 0, n(0) = 0, p(0) = 0 \quad (1)$$

Here, n_{res} (p_{res}) is the number of electrons (holes) remaining in the gain region, and n(p) is the number of electrons(holes) leaving the device to produce electron (hole) particle current within a duration of $2t_r$. The results in Figure 1(a) can be summarized by the following equations:

$$n_o(t = 2Kt_r) = n_{trig}(t = 2Kt_r) + n_{res}(t = 2Kt_r) \quad (2)$$

$$n(t = (2K + 1)t_r) = n_o\,(t = 2Kt_r)M_n \quad (3)$$

$$p_{res}(t = (2K + 1)t_r) = n_0(t = 2Kt_r)(M_n - 1) \quad (4)$$

$$p(t = 2(K + 1)t_r) = n_o(t = 2Kt_r)(M_n - 1)M_p \quad (5)$$

$$n_{res}(t = 2(K + 1)t_r) = n_o(t = 2Kt_r)(M_n - 1)\big(M_p - 1\big) \quad (6)$$

Here we divide time into discrete intervals by the unit of carrier transit time t_r. For an electron injected into the gain medium at t=0, the electron current is produced at each $(2K + 1)t_r$ (K =0, 1,2,3..) when electrons leave the cathode (Eq. 3) and the hole current is produced at each $2Kt_r$ (K =1,2,3..) when holes leave the anode (Eq.5). Equation (2) describes the number of electrons inside the gain region, being equal to the sum of the electrons excited by the on-going amplification process, n_{res} and new injected electron, n_{trig}, that will initiate new cycles of amplification ($n_{trig} = 0\ or\ 1$). Equations (4,6) describe the number of holes and electrons inside the gain medium. Using Fig. 1, the values of n, p, n_o, n_{res} and p_{res} at any given time can be calculated.

In (2)- (6), we choose the smallest time unit to be t_r. Within this time interval, the number of carriers inside the device should be at least 1 under Geiger mode (i.e. do not allow fractional photoexcited electrons). We enforce this condition by assuming that if at any moment the number of electrons entering into the gain region, n(t), falls below 1, we consider that no electrons are left inside the gain region and the carrier multiplication process stops, i.e.

$$n_o(t) = \begin{cases} 0;\ n_o(t) < 1 \\ n_0(t) \text{ from (2)};\ n_o(t) \geq 1 \end{cases} \quad (7)$$

We enforce the condition in Eq. (7) only in Geiger mode. In other modes where we are dealing with ensemble average, we accept the condition that the electron(hole) number can be lower than one as the carrier multiplication process continues.

We calculate the particle current (I_{part}) by multiplying a factor of $e/2t_r$ to the number of electrons and holes leaving the device within each duration of $2t_r$. Then,

$$I_{part}(2(K+1)t_r) = e/2t_r\left[n\left((2K+1)t_r\right) + p(2(K+1)t_r)\right] \quad (8)$$

From Eqs. (2-6) and the schematic in Fig. 1, we can find the bias (U) dependent DC gain since M_n and M_p are both bias dependent.

$$Gain_{DC}(U) = \frac{\sum_{K=0}^{\infty}\left[n\left((2K+1)t_r\right)+p(2(K+1)t_r)\right]}{\sum_{K=0}^{\infty} n_{trig}(2Kt_r)} = \frac{M_n+M_p(M_n-1)}{1-(M_n-1)(M_p-1)} \quad (9)$$

From (9), we can write,

$$Gain_{DC}(U) = \frac{\sum_{K=0}^{\infty}\left[n\left((2K+1)t_r\right) + p(2(K+1)t_r)\right]}{\sum_{K=0}^{\infty} n_{trig}(2Kt_r)}$$

For a single triggering event,

$$Gain_{DC}(U) = M_n + M_p(M_n-1) + M_n(M_n-1)\left(M_p-1\right) + M_p(M_n-1)^2\left(M_p-1\right)$$

$$+ M_n(M_n-1)^2\left(M_p-1\right)^2 + M_p(M_n-1)^3\left(M_p-1\right)^2 + \cdots$$

$$\Rightarrow Gain_{DC}(U) = \left[M_n + \frac{M_p}{M_p-1}\right]\left[1 + (M_n-1)\left(M_p-1\right) + (M_n-1)^2\left(M_p-1\right)^2 + \cdots\right]$$

$$- \frac{M_p}{M_p-1}$$

$$\Rightarrow Gain_{DC}(U) = \left[M_n + \frac{M_p}{M_p-1}\right]\left[\frac{1}{1-(M_n-1)\left(M_p-1\right)}\right] - \frac{M_p}{M_p-1}$$

$$\Rightarrow Gain_{DC}(U) = \frac{M_n + M_p(M_n-1)}{1-(M_n-1)\left(M_p-1\right)} \quad (9.A)$$

From (9), gain approaches infinity when $(M_n - 1)(M_p - 1) = 1$, which determines the breakdown voltage, (V_{BD}). We can also infer that, above the breakdown voltage, $(M_n - 1)(M_p - 1) > 1$.

Next, we calculate the gain decay time (τ_{decay}) in analog or sub-Geiger mode when the voltage is below V_{BD} and gain build-up time (τ_b) when the voltage is greater than the breakdown voltage and the device is in Geiger mode. From Fig. 1(a), we find that under a single triggering event, the total number of electrons and holes at the end of K^{th} round trip (i.e. $t=2Kt_r$) is

$$N(2Kt_r) = n_{res}(2Kt_r) + p(2Kt_r) = \frac{2M_p - 1}{M_p - 1} e^{\frac{t}{2t_r}\ln[(M_n-1)(M_p-1)]} \tag{10}$$

The detailed derivation of Eq. (10) from Eqs. (2-6) is given as following:

From (10), we can write,

$$N(t = 2Kt_r) = n_{res}(t = 2Kt_r) + p(t = 2Kt_r)$$

For a single triggering event at t=0 and using (2)-(6) we can write,

$$n_{res}(2t_r) = (M_n - 1)(M_p - 1);$$

$$p(2t_r) = M_p(M_n - 1)$$

$$n_{res}(4t_r) = (M_n - 1)^2(M_p - 1)^2;$$

$$p(2t_r) = M_p(M_n - 1)^2(M_p - 1);$$

In general,

$$n_{res}(2Kt_r) = (M_n - 1)^K(M_p - 1)^K \tag{10.A}$$

$$p(2Kt_r) = M_p(M_n - 1)^K(M_p - 1)^{K-1} \tag{10.B}$$

Using (10-A) and (10-B), N can be written as,

$$N(t = 2Kt_r) = (M_n - 1)^K(M_p - 1)^K\left[1 + \frac{M_p}{M_p - 1}\right]$$

$$\Rightarrow N(t = 2Kt_r) = \frac{2M_p - 1}{M_p - 1} e^{K \ln[(M_n-1)(M_p-1)]}$$

Since $t = 2Kt_r$, we can find,

$$N(2Kt_r) = \frac{2M_p-1}{M_p-1} e^{\frac{t}{2t_r}\ln[(M_n-1)(M_p-1)]} \quad (10.C)$$

Decay time and gain build up time are found by Writing (10-C) in an exponential decay function.

In Eq. (10), we define τ_{decay} and τ_b as

$$\tau_{decay} = -\frac{2t_r}{\ln[(M_n-1)(M_p-1)]}; U < V_{BD} \quad (11.A)$$

$$\tau_b = \frac{2t_r}{\ln[(M_n-1)(M_p-1)]}; U \geq V_{BD} \quad (11.B)$$

From (9) and (11-A), the gain-bandwidth product of the detector under the condition $U < V_{BD}$ (i.e. in analog mode or sub-Geiger mode) can be approximated as

$$GB \cong -\left[\frac{M_n+M_p(M_n-1)}{1-(M_n-1)(M_p-1)}\right]\frac{\ln[(M_n-1)(M_p-1)]}{4\pi t_r} \approx \frac{M_n+M_p(M_n-1)}{4\pi t_r} \quad (12)$$

If the detectors are triggered by photoexcited holes, a similar expression for Eqs. (9,11,12) can be obtained by switching the subscript n and p in all terms in the equations.

The dependence of $M_{n(p)}$ on the voltage drop in the device can be represented as

$$M_n = e^{\alpha(U)L}; M_p = e^{\beta(U)L} \quad (13)$$

where α and β are the multiplication coefficients for electrons and holes, and L is the thickness of the gain medium. We use empirical formulas for the electron and hole multiplication coefficients based on the local-field model that is widely used by device simulation programs such as Silvaco [27].

$$\alpha = A_n e^{-\left(\frac{B_n L}{U}\right)^{cn}}; \beta = A_p e^{-\left(\frac{B_p L}{U}\right)^{cp}} \quad (14)$$

A_n, A_p, B_n, B_p, c_n and cp are material specific parameters that can be found in numerous publications (Table 4.1). Transit time (t_r) is the time required for the carrier to cross the gain region and can be approximated from the length of the gain region and carrier velocity. For example, in Si, the saturation velocity of both electron and hole is $\sim 10^7$ cm/sec, producing a transit time of 1ps for a gain region of 100nm. Impact ionization coefficient parameters of Si, L = 100nm and t_r = 1ps are used as default parameters for the circuit model.

Table 4.1 Empirical Impact Ionization Coefficients of Different Materials

Material	Field (kV/cm)	A_n ($x10^5$ cm^{-1})	Bn ($x10^5$ cm^{-1})	c_n	A_p ($x10^5$ cm^{-1})	B_p ($x10^5$ cm^{-1})	c_p
Si [26]	400 – 800	7.03	12.3	1	6.71	16.9	1
InP [25], [28]	560 - 1250	2.32	8.46	2	2.48	7.89	2
$In_{0.52}Al_{0.48}$ As [29]	220 – 980	2.2	8.9	1.71	2.95	11.5	1.71
GaAs [25]	1100 - 1400	6.39	16	0.9	5.92	15.5	0.95
4H-SiC [30]	900 – 4000	27.8	105	1.37	35.1	103	1.09

Finally, to calculate the total detector current, we need to include displacement current through the device capacitance (Cj) in parallel with the particle current. It is to be noted that under Geiger mode, when the device is quenched by an external quenching circuit, particle current decreases rapidly and the device current is primarily produced by the displacement current through C_j. 1pF is used as the default value of the device capacitance C_j in our simulations.

Fig. 4.2(a) shows the device gain using Eqs. (9,13,14) with the default values of the model. The breakdown voltage is found to be 6.3V. The gain decay time below breakdown and gain buildup time above breakdown are plotted in Fig. 2(b) using the relations in Eq. 11. As expected, the gain decay time increases when the device approaches breakdown, limited by the gain-

bandwidth product. This can limit device sensitivity for photon counting in sub-Geiger mode. Also note that from Fig. 4.2(b), the gain buildup time decreases as the excess voltage (i.e. overbias above the breakdown voltage) increases. The model suggests that in Geiger-mode operation, detectors are not limited by the gain-bandwidth product and faster response can be achieved by increasing the overbias, which reduces the gain buildup time according to Eq. 11(b).

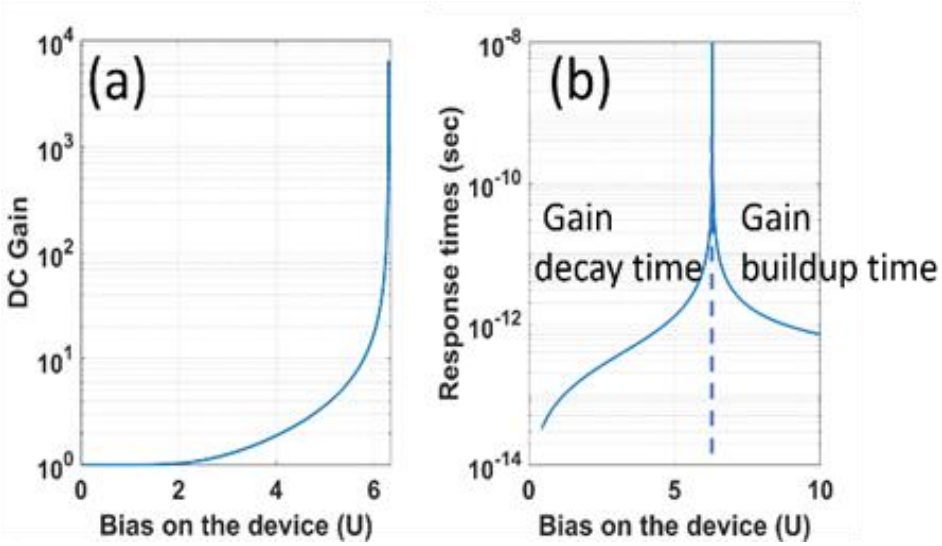

Figure 4.2 (a) DC gain dependence on device bias according to (9) for a Si APD with a 100nm multiplication layer. (b) Gain decay time and gain build up time dependence on device bias according to equation (11). Below the breakdown voltage of 6.3V, the plot represents decay time (τ_decay) and above breakdown voltage the plot represents gain build up time (τ_b). Parameters of Si in Table 1 are used to compute the gain and response time.

We implement the device model using behavioral and circuit components in Orcad PSpice [31] as an example. This model can also be implemented in Cadence Spectre [32] by importing this PSpice model. Figure 4.3(a) illustrates the implemented model using different behavioral PSpice components. V_trigger denotes the incoming triggering pulse and V_Anode and V_cathode represent the device anode and cathode voltage, respectively. At every instant, the model calculates bias (V_bias_on_dev) dependent ionization coefficients, i.e. α ,and β from Eq. (14) and multiplication factors, M_n, M_p from Eq. (13). The particle current I_part is set to be 0

before the first triggering event. Once a single triggering event is registered by ntrig =1 at $t=t_1$, the multiplication process starts and values of n, p and I_part at $t=t_1+2t_r$ are calculated. Note that, V_bias_on_dev($t=t_1+2t_r$) is also computed from the anode voltage at $t=t_1$ for numerical stability. Laplace blocks are used to realize the delay functions, and expressions from Eqs. (2)-(8), (13), (14) are implemented by 1 input and general purpose ABM blocks. For a note, ABM delay block can also be used in lieu of Laplace blocks for recent versions of PSpice. Even though PSpice simulates circuits under arbitrarily chosen time steps, the output current is only valid at the multiple of $2t_r$. Therefore, it is recommended to use 3-5 times of $2t_r$ as the "Maximum Step Size" in the "transient" options of PSpice such that PSpice arbitrary time steps are smaller than $2t_r$ and Laplace block does not show any non-causal warning. Finally, device capacitance is placed in parallel with the calculated particle current, I_part, to get the total current of the device. Fig. 4.3(b) shows the derived sub-circuit symbol from the model components. Device is triggered by the input pulse applied at terminal 1 (Trigger) of the symbol. Terminal 2 and terminal 3 represent the cathode and anode connection of the device. Since the particle current is calculated every $2t_r$, the input pulses have a width of $2t_r$ and subsequent inputs are separated by a time interval greater than 2tr. The amplitude of the pulse represents the number of photon/dark carrier per pulse. Moreover, if the DC primary dark/photo current, Ip, is known, then triggering pulses by primary dark/photo current need to be separated by a time interval of e/Ip. In cases that e/Ip becomes smaller than $2t_r$ due to a very high dark/photo current, the detector can be divided into identical subunits connected in parallel. The contents of the .LIB file of the model is included in Appendix C and the .OLB file of the model is attached in the supplementary material. We used the model shown in Fig. 4.3 along with the default parameters for simulations in this paper.

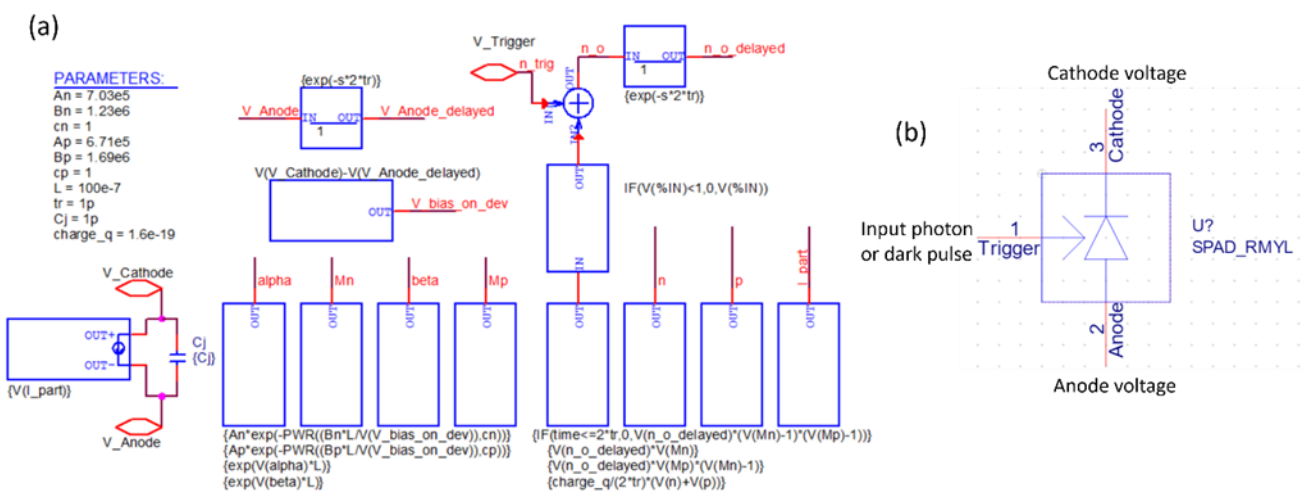

Figure 4.3 (a) Orcad PSPICE implementation of the model using Eqs. (2) – (14). (b) Modelled sub-circuit symbol used for different applications.

Following shows the source code for the LIB file of the model:

** source DEVICE_MODEL*

**$*

.SUBCKT SPAD_RMYL Trigger V_anode V_cathode +PARAMS: charge_q=1.6e-19 an=7.03e5 bn=1.23e6 cn=1 +ap=6.71e5 bp=1.69e6 cp=1 l=100e-7 cj=1p tr=1p

*E_LAPLACE1 V_ANODE_DELAYED 0 LAPLACE +{V(V_ANODE)} {({exp(-s*2*tr)})/(1)}*

E_ABM21 V_BIAS_ON_DEV 0 VALUE

+{ V(V_CATHODE)-V(V_ANODE_DELAYED) }

*E_ABM24 MN 0 VALUE { {exp(V(ALPHA)*L)} }*

E_ABM23 BETA 0 VALUE

*+{{Ap*exp(-PWR((Bp*L/V(V_BIAS_ON_DEV)),cp))}}*

*E_ABM25 MP 0 VALUE { {exp(V(BETA)*L)}}*

E_ABM22 ALPHA 0 VALUE

*+{{An*exp(-PWR((Bn*L/V(V_BIAS_ON_DEV)),cn))}}*

E_ABM11 I_PART 0 VALUE

*+{{charge_q/(2*tr)*(V(N)+V(P))}}*

E_ABM31 P 0 VALUE

*+{{V(N_O_DELAYED)*V(MP)*(V(MN)-1)}}*

*E_ABM30 N 0 VALUE { {V(N_O_DELAYED)*V(MN)}}*

E_SUM1 N_O 0 VALUE {V(TRIGGER)+V(N83646)}

E_ABM26 N83642 0 VALUE

*+{{IF(time<=2*tr,0,V(N_O_DELAYED)*(V(MN)1)*(V(M+P)-1))}}*

E_LAPLACE6 N_O_DELAYED 0

*+LAPLACE {V(N_O)} {({exp(-s*2*tr)})/(1)}*

E_ABM33 N83646 0 VALUE +{IF(V(N83642)<1,0,V(N83642))}

G_ABMI1 V_CATHODE V_ANODE VALUE +{{V(I_PART)}}

C_Cj V_ANODE V_CATHODE {Cj} TC=0,0

.ENDS

**$*

4.3 Model Application

In this section, we demonstrate the universal applicability of our model by simulating devices operating in different modes (sub-Geiger mode and Geiger-mode) with different circuits. We also show how material properties of the gain medium can affect the device performance.

4.3.1 Impulse Response of the Device

We simulated the impulse response of the device in different modes of operation. Figure 4.4(a) shows the device response under low DC gain of around 17 (at 6V bias) when the device is triggered by a single electron. As the bias voltage is close to 6.3V to achieve a DC gain of 1895, the device response in sub-Geiger mode is slowed down significantly from 25ps to 2.5ns, as shown in Fig. 4.4(b). Due to the slow response time, even though the DC gain is high, the peak amplitude of the device response does not increase substantially compared to the peak amplitude in linear mode. Figure 4.4(c) shows the Geiger mode response at 0.7V excess bias. The output current rises exponentially within a very short time, thus requiring a quenching circuit to bring the bias below breakdown to prevent device damage.

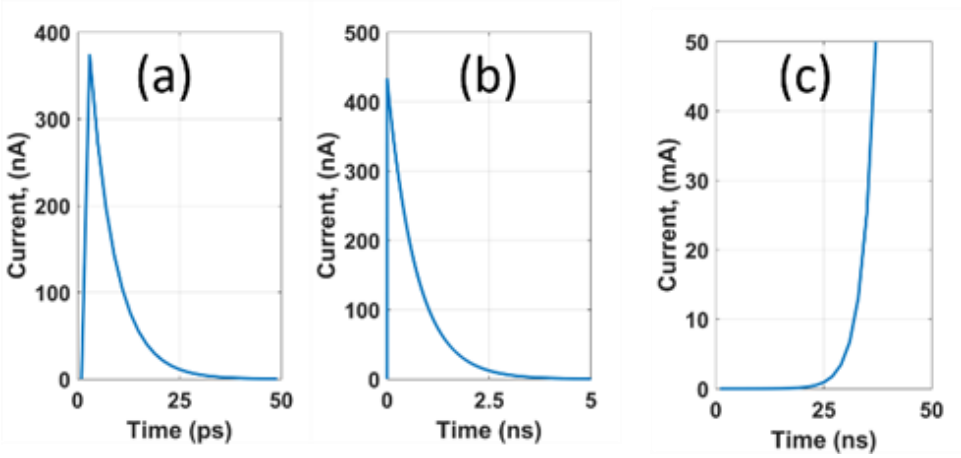

Figure 4.4 Impulse response of the device in (a) Linear mode (6V) with a gain of ~17, (b) Sub-Geiger mode with a DC gain of ~1895, and (c) Geiger mode (0.7V excess bias). The Geiger mode current response rises exponentially with time, thus requiring quenching.

4.3.2 Sub-Geiger Mode

Sub-Geiger mode operation can be used for detection of a few photons although with very low noise amplifiers, the device can achieve single photon sensitivity and even photon number

resolving capabilities [33]. By integrating a charge sensitive amplifier (CSA) with the detector in sub-Geiger mode and using the equivalent circuit in Fig. 4.3, Fig. 4.5(a) and parameter values of Si in Table 4.1, we show the circuit response in Fig. 4.5(b). For a DC gain of 1895, the peak amplitude of the output is around 142uV when triggered by a pulse containing 3 photons. As shown in Fig. 4.5(c), the output amplitude rises exponentially with the bias voltage in sub-Geiger mode, which stresses the importance of bias stability. Figure 4.5(d) shows the simulated circuit response to different number of photons. The device output increases with the photon number, thus providing a wider dynamic range than Geiger-mode operation. Using this example, we show that our detector circuit model can be used with an external circuit to simulate sub-Geiger mode characteristics.

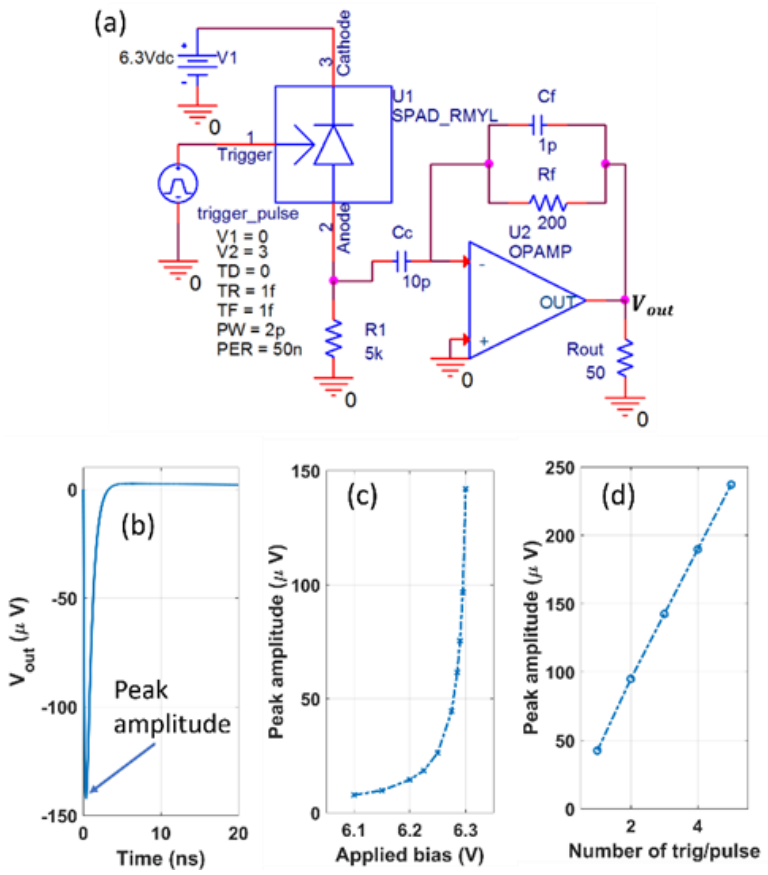

Figure 4.5 Simulation of sub-Geiger mode operation with the detector biased for a DC gain of 1895. (a) Simulated circuit diagram in PSpice. (b) Response at the output of a transimpedance amplifier (TIA). The TIA integration time is around 1.25ns. (c) Peak amplitude of the output voltage under different applied bias. (d) Dependence of the peak output amplitude on the number of photoexcited carriers per pulse.

4.3.3 Passive Quenching Circuit

For single photon detection in Geiger mode, a quenching circuit is needed to avoid device damage and to reset the device after detection of each triggering event. Next, we apply our detector model in a passive quenching circuit (PQC).

The passive quenching circuit in Fig. 4.6(a) uses a 100 kΩ resistor. The device is originally biased at 7V (0.7V excess bias) and the output signal is obtained from the 50Ω sensing resistor. We vary the rate of incoming photons to show the effect of recovery time and the performance

limits for passive quenching circuits. Figure 4.6(b) shows the device response to a series of single photon events separated by 40ns. The red arrows indicate the photon arrival times. The device responds strongly to the first triggering event, but the output responses to the subsequent events become much less pronounced because an interval of 40ns is too short for the detector to be fully recovered after quenching. The simulation shows that after 200ns or from the 6^{th} triggering event on, the voltage swing across the device, $V_{Cathode}$-V_{Anode}, is only between 6.2V and 6.4V, about $\pm 0.1V$ around the breakdown voltage, and never reaches the original value of 7V across the detector. The recovery time of a passive quenching circuit is determined by the time constant $(R_{passive}+R_{sense})Cj$ (≈500ns for the simulated circuit), which explains why the circuit cannot support single photon triggering events separated by 40ns. If we increase the time interval between single photon pulses to 1.2μs, the detector has time to be fully recovered and the output signal can reach the full level, as shown in Fig. 4.6(c).

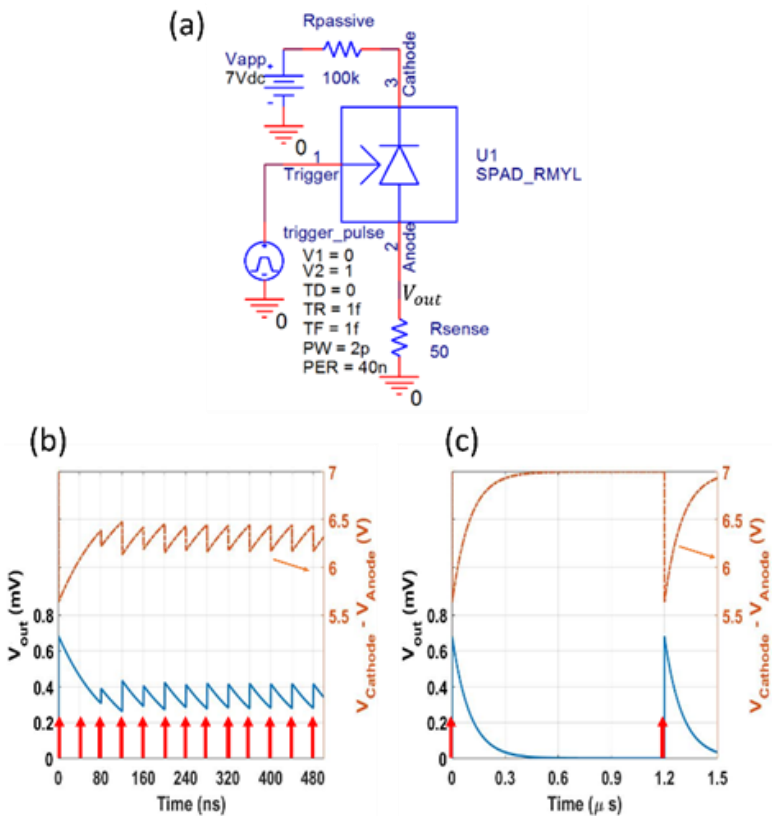

Figure 4.6 Simulation of a Geiger-mode detector with passive quenching circuit (PQC) using our model. (a) Simulated circuit in PSpice at 0.7V excess bias. (b) V_{out} and actual voltage drop ($V_{cathode}$ - V_{anode}) on the device with 40ns periodic triggering events. Device quenches after the first triggering event and cannot fully recover to due to the slow recovery time. (c) Device response to triggering events separated by 1.2µs. The detector can fully recover in this case. Red arrows mark the incidence of the triggering events for (b) and (c).

4.3.4 Mixed Active-Passive Quenching Circuit

To overcome the problem of slow recovery time for passive quenching circuits, active quenching circuits (AQC) with a fast recovery time can be integrated with Geiger-mode detectors [34], [35]. A high performance AQC requires elaborated design and optimization of circuit configuration and components, which is outside the scope of this paper. Instead we simulate an ideal mixed active-passive quenching circuit [12] with our detector model to demonstrate the viability of the detector model in Geiger mode operation with an active quenching circuit.

Figure 4.7(a) shows the circuit configuration in PSpice. When a photon or dark pulse triggers the detector, the carrier multiplication process is initiated and the detector is at first passively quenched by the resistors R_L and R_s. When the voltage V_{out} across R_s rises above the comparator threshold (1.5mV), the control logic block is activated and in 3ns, the delay block in the control logic closes the S_quench switch, reducing the voltage drop in the detector to 6V, which is about 0.3V lower than the breakdown voltage. After holding the detector at the quenched state for 3ns, the detector is reset to its initial state via the S_reset switch and the delay blocks in the control logic. In this design, in 6.5ns after detecting a single photon, the detector is fully recovered and able to respond to the next triggering event. In our simulation, we set four triggering events at 0ns, 10ns, 20ns, and 40ns, respectively. Figure 4.7(b) and 4.7(c) show the actual voltage drop in the device ($V_{Cathode}$-V_{Anode}) and output voltage across the sensing resistor R_s. We can clearly observe three actions upon each triggering pulse: (i) fast response upon photon arrival due to the short gain buildup time, (ii) passive quenching through R_L followed by active quenching at 3ns, and (iii) bias reset at 6.5 ns. In this design, the detector is fully recovered after 6.5ns. Moreover, the device produces no output in the absence of photon or dark pulses, as shown in the output between 28ns and 40ns in Figs. 4.7(b) and 4.7(c).

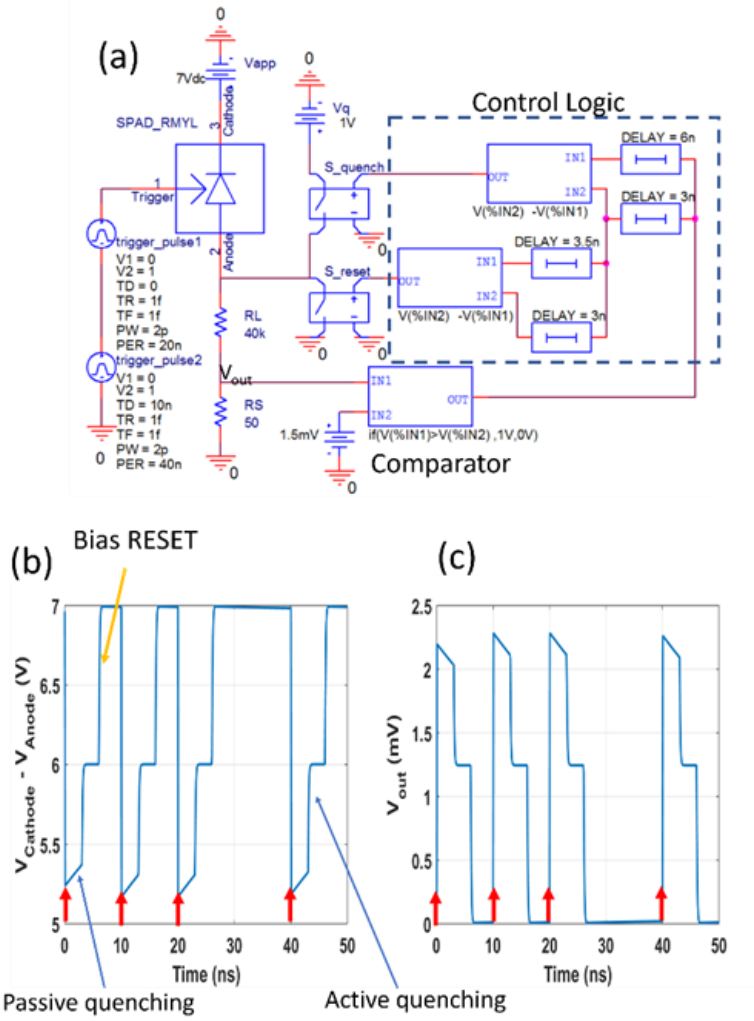

Figure 4.7 (a) Implementation of the model in PSpice for a mixed active-passive quenching circuit with a 7V applied bias on the cathode. (b) Voltage drop in the device with time and (c) Output signal, Vout, with time. Red arrows in (b) and (c) indicate the incoming triggering pulses. Device is actively quenched 3ns after being triggered. After 6.5ns since the detection of photon, the device is reset to have 7V drop across it and is able to detect the next triggering pulse.

4.3.5 Material Dependence of Device Response

Different materials can be used for gain medium in single photon detectors to achieve the desired operational wavelength. Si is used for visible photons [36]–[38], 4H-SiC for UV photons [39], and InP, InAlAs and GaAs for telecommunication wavelengths for near infrared photons [40]–[43]. All parameters in our circuit model are obtained directly from material parameters

related to the carrier multiplication process. Therefore, our circuit model can predict the performance of detectors made of different materials by virtue of material specific impact ionization coefficients. We list the published impact ionization coefficients under high electric field for different materials in Table 4.1. Moreover, for new materials, these coefficients can be obtained from well-established experimental techniques [29], [44]–[46].

We use our circuit model to simulate the intrinsic gain bandwidth (GBW) product for detectors with Si and InAlAs gain medium. Gain is calculated from the integrated impulse response, and bandwidth is calculated from the fall time of the device when it is triggered by a single carrier. The above procedure is equivalent to taking a Fourier transform of the impulse response. APDs with Si gain medium has a breakdown voltage of 6.302V and APDs with InAlAs gain medium has a breakdown voltage of 8.782V. Figure 4.8 shows the variation in GBW product from 6V to 6.3V for Si and from 8V to 8.78V for InAlAs devices. The intrinsic GBW product of APDs using Si and InAlAs gain medium is ~433GHz and ~355GHz, respectively. These results are consistent with the reported GBW product for Si and InAlAs APDs [47]–[49].

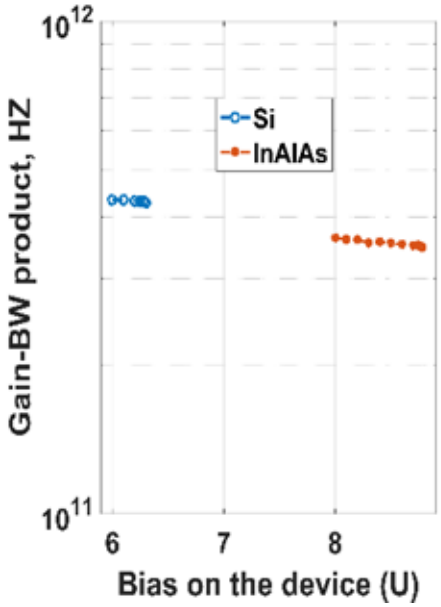

Figure 4.8 Simulated bias-dependent gain-bandwidth (GBW) product for detectors with 100nm Si and InAlAs gain medium. The intrinsic GBW of APDs is ~433GHz for Si gain medium and ~355GHz for InAlAs gain medium

4.3.6 Application and Simulation of the Model to Published Si APD

We simulated a Si APD reported in [50] with a passive quenching circuit. The default Si parameters for impact ionization coefficients, L = 1.5μm, tr = 15ps and Cj = 37.44fF [50] were used as the model parameter. Cj was extracted from the single photon gain curve of reported in [50]. The computed breakdown voltage was found to be 42.9V which was almost same as 43V, reported in [50]. Fig. 4.9(a) shows the amplified output voltage of the passive quenching circuit in Fig. 4.9(b) with Rq = 250k and VEX = 1.6V. The simulation result matches with the experiment very well.

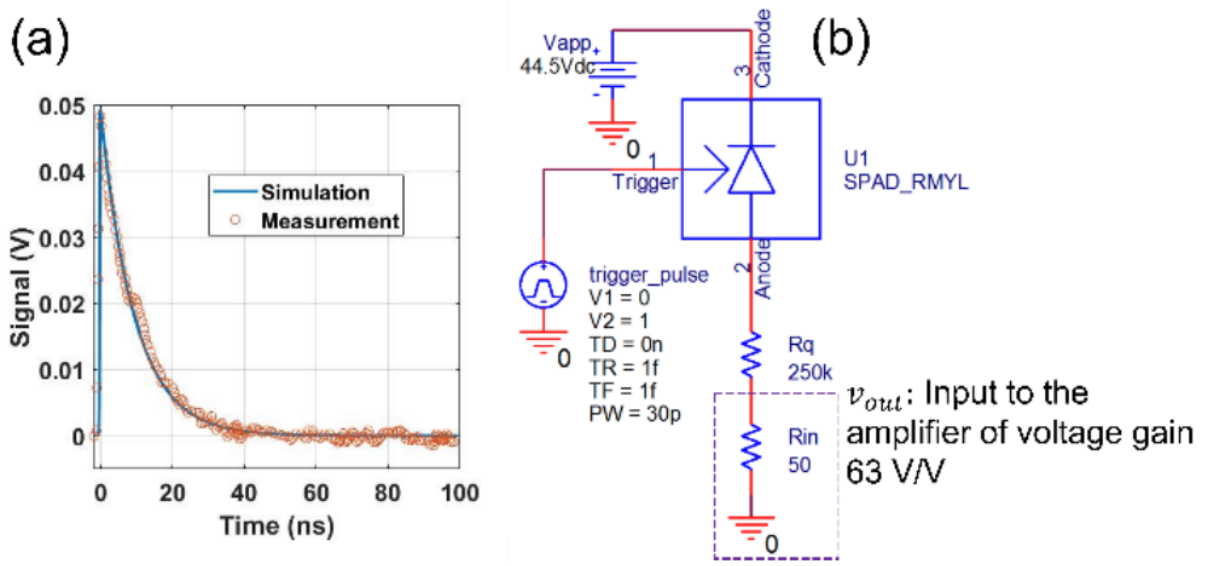

Figure 4.9 (a) Measured and simulated output signal of the circuit. (b) Schematic of the simulated circuit.

4.4 Conclusion

We have developed a universal circuit model in SPICE for photodetectors with carrier multiplication gain. Our model can simulate and predict both static and transient characteristics of the devices in analog, sub-Geiger and Geiger mode with charge sensitive amplifiers and passive/active quenching circuits. Moreover, using material dependent ionization coefficients as input parameters, designers can use our model to simulate circuit performance for detectors made of different materials. It offers a versatile and effective design and simulation tool for circuits including APDs or CEP detectors to serve the rapidly increasing applications that require high sensitivity photon detection.

ACKNOWLEDGEMENTS

Chapter 4, in full, is a reprint of the material as it appears in Miah, M.A.R., Jiang, Y. and Lo, Y.H., 2021. A physics based unified circuit model for single photon and analog detector. IEEE Access, 9, pp.129571-129581. The dissertation author was the second author of this paper.

REFERENCE

[1] L. Yan, Y. Yu, A. C. Zhang, D. Hall, I. A. Niaz, M. A. R. Miah, Y.-H. Liu, and Y.-H. Lo, ‘‘An amorphous silicon photodiode with 2 THz gainbandwidth product based on cycling excitation process,’’ Appl. Phys. Lett., vol. 111, no. 10, Sep. 2017, Art. no. 101104, doi: 10.1063/1.5001170.

[2] S. M. Sze, ‘‘Photodetectors,’’ in Physics of Semiconductor Device, 2nd ed. Hoboken, NJ, USA: Wiley, 1981.

[3] Y.-H. Liu, A. Zhang, M. Abu, R. Miah, D. Hall, I. A. Niaz, L. Yan, Y. Yu, M. S. Kavrik, and Y.-H. Lo, ‘‘Cycling excitation process for light detection and signal amplification in semiconductors,’’ Proc. SPIE, vol. 9933, Sep. 2016, Art. no. 99330C, doi: 10.1117/12.2238585.

[4] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, ‘‘Invited review article: Single-photon sources and detectors,’’ Rev. Sci. Instrum., vol. 82, no. 7, Jul. 2011, Art. no. 071101, doi: 10.1063/1.3610677.

[5] T. Baba, Y. Suzuki, K. Makino, T. Fujita, T. Hashi, S. Adachi, S. Nakamura, and K. Yamamoto, ‘‘Development of an InGaAs SPAD 2D array for flash LiDAR,’’ Proc. SPIE, vol. 10540, Jan. 2018, Art. no. 105400L, doi: 10.1117/12.2289270.

[6] M. Benetti, M. Popleteeva, G.-F.-D. Betta, L. Pancheri, C. Collini, E. Morganti, L. Lorenzelli, L. Lunelli, L. Pasquardini, C. Pederzolli, and D. Stoppa, ‘‘CMOS single-photon detector for

advanced fluorescence sensing applications,'' in Proc. Int. Workshop Biophoton., Jun. 2011, pp. 1–3, doi: 10.1109/IWBP.2011.5954798.

[7] R. H. Hadfield, ''Single-photon detectors for optical quantum information applications,'' Nature Photon., vol. 3, no. 12, pp. 696–705, Dec. 2009, doi: 10.1038/nphoton.2009.230.

[8] H. Hemmati, A. Biswas, and I. B. Djordjevic, ''Deep-space optical communications: Future perspectives and applications,'' Proc. IEEE, vol. 99, no. 11, pp. 2020–2039, Nov. 2011, doi: 10.1109/JPROC.2011.2160609.

[9] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, ''Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 8445–8453. Accessed: Mar. 31, 2021. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Wang_PseudoLiDAR_From_Visual_Depth_Estimation_Bridging_the_Gap_in_3D_ CVPR_2019_paper.html

[10] R. Wang, ''3D building modeling using images and LiDAR: A review,'' Int. J. Image Data Fusion, vol. 4, no. 4, pp. 273–292, Dec. 2013, doi: 10.1080/19479832.2013.811124.

[11] A. D. Mora, A. Tosi, S. Tisa, and F. Zappa, ''Single-photon avalanche diode model for circuit simulations,'' IEEE Photon. Technol. Lett., vol. 19, no. 23, pp. 1922–1924, Dec. 1, 2007, doi: 10.1109/LPT.2007.908768.

[12] F. Zappa, A. Tosi, A. D. Mora, and S. Tisa, ''SPICE modeling of single photon avalanche diodes,'' Sens. Actuator A, Phys., vol. 153, no. 2, pp. 197–204, Aug. 2009, doi: 10.1016/j.sna.2009.05.007.

[13] Y. Tian, J. Tu, and Y. Zhao, ''A Pspice circuit model for single-photon avalanche diodes,'' Opt. Photon. J., vol. 7, no. 8, pp. 1–6, Aug. 2017, doi: 10.4236/opj.2017.78B001.

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

[15] A. Gulinatti, I. Rech, M. Assanelli, M. Ghioni, and S. Cova, ‘‘A physically based model for evaluating the photon detection efficiency and the temporal response of SPAD detectors,’’ J. Mod. Opt., vol. 58, nos. 3–4, pp. 210–224, Feb. 2011, doi: 10.1080/09500340.2010.536590.

[16] A. Inoue, T. Okino, S. Koyama, and Y. Hirose, ‘‘Modeling and analysis of capacitive relaxation quenching in a single photon avalanche diode (SPAD) applied to a CMOS image sensor,’’ Sensors, vol. 20, no. 10, p. 3007, May 2020, doi: 10.3390/s20103007.

[17] S. You, J. Cheng, and Y. H. Lo, ‘‘Physics of single photon avalanche detectors with built-in self-quenching and self-recovering capabilities,’’ IEEE J. Quantum Electron., vol. 48, no. 7, pp. 960–967, Jul. 2012, doi: 10.1109/JQE.2012.2196679.

[18] S. You, J. Cheng, and Y. Lo, ‘‘Physics of self-recovering single photon avalanche detectors,’’ Proc. SPIE, vol. 8155, Sep. 2011, Art. no. 81551H, doi: 10.1117/12.895971.

[19] Z. Cheng, X. Zheng, D. Palubiak, M. J. Deen, and H. Peng, ‘‘A comprehensive and accurate analytical SPAD model for circuit simulation,’’ IEEE Trans. Electron Devices, vol. 63, no. 5, pp. 1940–1948, May 2016, doi: 10.1109/TED.2016.2537879.

[20] Y. Oussaiti, D. Rideau, J. R. Manouvrier, V. Quenette, B. Mamdy, C. Buj, J. Grebot, H. Wehbe-Alause, A. Lopez, G. Mugny, M. Agnew, E. Lacombe, M. Pala, and P. Dollfus, ‘‘Verilog—A model for avalanche dynamics and quenching in single-photon avalanche diodes,’’ in Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD), Sep. 2020, pp. 145–148, doi: 10.23919/SISPAD49475.2020.9241648.

[21] M. A. R. Miah, I. A. Niaz, and Y.-H. Lo, ‘‘Defect assisted carrier multiplication in amorphous silicon,’’ IEEE J. Quantum Electron., vol. 56, no. 3, pp. 1–11, Jun. 2020, doi: 10.1109/JQE.2020.2988263.

[22] M. Isler, ‘‘Phonon-assisted impact ionization of electrons in In0.53Ga0.47As,’’ Phys. Rev. B, Condens. Matter, vol. 63, no. 11, Mar. 2001, Art. no. 115209, doi: 10.1103/PhysRevB.63.115209.

[23] K. H. Kim and Y. S. Kim, ‘‘Design and analysis of CMOS single photon counting avalanche photodiodes integrated with active quenching circuits,’’ J. Nucl. Sci. Technol., vol. 45, no. 5, pp. 511–514, Jun. 2008, doi: 10.1080/00223131.2008.10875903.

[24] F. Guerrieri, S. Tisa, A. Tosi, and F. Zappa, ‘‘Two-dimensional SPAD imaging camera for photon counting,’’ IEEE Photon. J., vol. 2, no. 5, pp. 759–774, Oct. 2010, doi: 10.1109/JPHOT.2010.2066554.

[25] J. S. Cheong, M. M. Hayat, X. Zhou, and J. P. R. David, ‘‘Relating the experimental ionization coefficients in semiconductors to the nonlocal ionization coefficients,’’ IEEE Trans. Electron Devices, vol. 62, no. 6, pp. 1946–1952, Jun. 2015, doi: 10.1109/TED.2015.2422789.

[26] R. van Overstraeten and H. de Man, ‘‘Measurement of the ionization rates in diffused silicon p-n junctions,’’ Solid-State Electron., vol. 13, no. 1, pp. 583–608, May 1970, doi: 10.1016/0038-1101(70)90139-5.

[27] Silvaco. Atlas User Manual. Accessed: Mar. 31, 2021. [Online]. Available: https://silvaco.com/dynamicweb/silen/

[28] L. W. Cook, G. E. Bulman, and G. E. Stillman, ‘‘Electron and hole impact ionization coefficients in InP determined by photomultiplication measurements,’’ Appl. Phys. Lett., vol. 40, no. 7, pp. 589–591, Apr. 1982, doi: 10.1063/1.93190.

[29] Y. L. Goh, D. J. Massey, A. R. J. Marshall, J. S. Ng, C. H. Tan, W. K. Ng, G. J. Rees, M. Hopkinson, J. P. R. David, and S. K. Jones, ‘‘Avalanche multiplication in InAlAs,’’ IEEE Trans. Electron Devices, vol. 54, no. 1, pp. 11–16, Jan. 2007, doi: 10.1109/TED.2006.887229.

[30] W. S. Loh, B. K. Ng, J. S. Ng, S. I. Soloviev, H.-Y. Cha, P. M. Sandvik, C. M. Johnson, and J. P. R. David, ‘‘Impact ionization coefficients in 4H-SiC,’’ IEEE Trans. Electron Devices, vol. 55, no. 8, pp. 1984–1990, Aug. 2008, doi: 10.1109/TED.2008.926679.

[31] (Mar. 4, 2014). Overview Page—OrCAD PSpice Designer. Accessed: Mar. 1, 2021. [Online]. Available: https://www.orcad.com/products/orcadpspice-designer/overview

[32] Spectre Simulation Platform. Accessed: Mar. 6, 2021. [Online]. Available: https://www.cadence.com/en_US/home/tools/custom-ic-analog-rfdesign/circuit-simulation/spectre-simulation-platform.html

[33] Y. Miyamoto, K. Tsujino, J. Kataoka, and A. Tomita, ‘‘Sub-geiger mode single-photon detector using a low-dark-current InGaAs avalanche photodiode,’’ in Proc. Eur. Conf. Exhib. Opt. Commun., Sep. 2012, pp. 1–3, doi: 10.1364/ECEOC.2012.P7.04.

[34] B. Bérubé, V.-P. Rheaume, A. C. Therrien, S. Parent, L. Maurais, A. Boisvert, G. Carini, S. A. Charlebois, R. Fontaine, and J.-F. Pratte, ‘‘Development of a single photon avalanche diode (SPAD) array in high voltage CMOS 0.8 μm dedicated to a 3D integrated circuit (3DIC),’’ in Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec. (NSS/MIC), Oct. 2012, pp. 1835–1839, doi: 10.1109/NSSMIC.2012.6551428.

[35] S. Tisa, F. Guerrieri, and F. Zappa, ‘‘Variable-load quenching circuit for single-photon avalanche diodes,’’ Opt. Exp., vol. 16, no. 3, pp. 2232–2244, Feb. 2008, doi: 10.1364/OE.16.002232.

[36] C. Veerappan and E. Charbon, ‘‘A low dark count p-i-n diode based SPAD in CMOS technology,’’ IEEE Trans. Electron Devices, vol. 63, no. 1, pp. 65–71, Jan. 2016, doi: 10.1109/TED.2015.2475355.

[37] L. Yan, M. A. R. Miah, Y.-H. Liu, and Y.-H. Lo, ‘‘Single photon detector with a mesoscopic cycling excitation design of dual gain sections and a transport barrier,’’ Opt. Lett., vol. 44, no. 7, pp. 1746–1749, Apr. 2019, doi: 10.1364/OL.44.001746.

[38] S. Mandai, M. W. Fishburn, Y. Maruyama, and E. Charbon, ‘‘A wide spectral range single-photon avalanche diode fabricated in an advanced 180 nm CMOS technology,’’ Opt. Exp., vol. 20, no. 6, pp. 5849–5857, Mar. 2012, doi: 10.1364/OE.20.005849.

[39] X. Bai, D. Mcintosh, H. Liu, and J. C. Campbell, ‘‘Ultraviolet single photon detection with geiger-mode 4H-SiC avalanche photodiodes,’’ IEEE Photon. Technol. Lett., vol. 19, no. 22, pp. 1822–1824, Nov. 15, 2007, doi: 10.1109/LPT.2007.906830.

[40] J. Zhang, M. A. Itzler, H. Zbinden, and J.-W. Pan, ‘‘Advances in InGaAs/InP single-photon detector systems for quantum communication,’’ Light, Sci. Appl., vol. 4, no. 5, p. e286, May 2015, doi: 10.1038/lsa.2015.59.

[41] X. Meng, S. Xie, X. Zhou, N. Calandri, M. Sanzaro, A. Tosi, C. H. Tan, and J. S. Ng, ‘‘InGaAs/InAlAs single photon avalanche diode for 1550 nm photons,’’ Roy. Soc. Open Sci., vol. 3, no. 3, Mar. 2016, Art. no. 150584, doi: 10.1098/rsos.150584.

[42] K. Zhao, S. You, J. Cheng, and Y.-H. Lo, ‘‘Self-quenching and selfrecovering InGaAs/InAlAs single photon avalanche detector,’’ Appl. Phys. Lett., vol. 93, no. 15, Oct. 2008, Art. no. 153504, doi: 10.1063/1.3000610.

[43] A. C. Farrell, X. Meng, D. Ren, H. Kim, P. Senanayake, N. Y. Hsieh, Z. Rong, T.-Y. Chang, K. M. Azizur-Rahman, and D. L. Huffaker, ‘‘InGaAs-GaAs nanowire avalanche photodiodes

toward single-photon detection in free-running mode,'' Nano Lett., vol. 19, no. 1, pp. 582–590, Jan. 2019, doi: 10.1021/acs.nanolett.8b04643.

[44] S. L. Chuyang, ''Photodetectors,'' in Physics of Optoelectronic Devices, 1st ed. Hoboken, NJ, USA: Wiley, 1995.

[45] M. H. Woods, W. C. Johnson, and M. A. Lampert, ''Use of a Schottky barrier to measure impact ionization coefficients in semiconductors,'' SolidState Electron., vol. 16, no. 3, pp. 381–394, Mar. 1973, doi: 10.1016/ 0038-1101(73)90013-0.

[46] L. J. J. Tan, J. S. Ng, C. H. Tan, and J. P. R. David, ''Avalanche noise characteristics in submicron InP diodes,'' IEEE J. Quantum Electron., vol. 44, no. 4, pp. 378–382, Apr. 2008, doi: 10.1109/JQE.2007.914771.

[47] N. Duan, T.-Yang. Liow, A. E.-J. Lim, L. Ding, and G. Q. Lo, ''310 GHz gain-bandwidth product Ge/Si avalanche photodetector for 1550 nm light detection,'' Opt. Exp., vol. 20, no. 10, pp. 11031–11036, May 2012, doi: 10.1364/OE.20.011031.

[48] Y. Kang, H.-D. Liu, M. Morse, M. J. Paniccia, M. Zadka, S. Litski, G. Sarid, A. Pauchard, Y.-H. Kuo, H.-W. Chen, W. S. Zaoui, J. E. Bowers, A. Beling, D. C. McIntosh, X. Zheng, and J. C. Campbell, ''Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain–bandwidth product,'' Nature Photon., vol. 3, no. 1, pp. 59–63, Jan. 2009, doi: 10.1038/nphoton.2008.247.

[49] W. S. Zaoui, H.-W. Chen, J. E. Bowers, Y. Kang, M. Morse, M. J. Paniccia, A. Pauchard, and J. C. Campbell, ''Frequency response and bandwidth enhancement in Ge/Si avalanche photodiodes with over 840 GHz gain-bandwidth-product,'' Opt. Exp., vol. 17, no. 15, pp. 12641–12649, Jul. 2009, doi: 10.1364/OE.17.012641.

[50] G. Kawata, J. Yoshida, K. Sasaki, and R. Hasegawa, ‘‘Probability distribution of after pulsing in passive-quenched single-photon avalanche diodes,’’ IEEE Trans. Nucl. Sci., vol. 64, no. 8, pp. 2386–2394, Aug. 2017, doi: 10.1109/TNS.2017.2717463.