Modeling and Development of Superconducting Nanowire Single-Photon Detectors

Thesis by
Jason Paul Allmaras

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Applied Physics

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 19, 2020

ACKNOWLEDGEMENTS

First, I'd like to thank my advisor, Dr. Matthew Shaw for granting me the opportunity to do my Ph.D. studies with the SNSPD group at JPL. Your willingness to support my diverse efforts has encouraged and allowed me to broaden my academic horizons. I especially appreciate your guidance in the early years of my journey. I'm also grateful to Professor Keith Schwab for enabling the arrangement with the JPL team, and the other members of my committee, Professor Andrei Faraon and Professor Austin Minnich for their time.

I'd like to express my sincerest thanks to my theory mentor Dr. Alex Kozorezov. The work presented here has only been possible due to your patience and guidance.

I've had the privilege of learning from a number of talented experimentalists at JPL, and I'd particularly particularly like to thank Boris Korzh, Emma Wollman, Francesco Marsili, and Angel Velasco for teaching me proper laboratory techniques, and for answering my late-night and weekend calls when things went wrong.

My work in the cleanroom was only made possible through the incredible patience and support of Andrew Beyer who taught me everything I know about nanofabrication. Thank you for humoring my questions and debugging the endless woes of the stepper. I would also like to thank Ryan Briggs for teaching me to use the EBPG, and Ekkehart Schmidt, Bruce Bumble, Rick LeDuc, Cliff Frez, and Mike Fitzsimmons for their advice, their troubleshooting expertise, and for broadening my understanding of cleanroom techniques.

I've had the pleasure of working with many postdocs, fellow graduate students, and interns while at JPL and more recently at Caltech. While there are too many to list, I would particularly like to thank Simone, Eddy, Andrew, and Sahil for the learning opportunities, camaraderie, and interesting conversations over the years. I am also grateful to Professor Maria Spiropulu for making the JPL team feel at home at Caltech. I would additionally like to thank my friends Joe, Jeremy, Aaron, Kohtaro, Tomas, Reid, Ian, and Kenny for making my graduate experience more enjoyable.

Finally, I'd like to thank my family for their support throughout my entire education, and I'd like to thank my partner Sam for her encouragement and constant reminder that there is more to life than research.

ABSTRACT

Superconducting nanowire single-photon detectors (SNSPDs) have demonstrated remarkable efficiency, timing resolution, and intrinsic dark count rate properties, but the SNSPD community currently lacks a comprehensive model of the single-photon detection process. In this work, we conduct a detailed examination of the current detection mechanism models and compare their predictions to new experimental measurements of the intrinsic timing properties and polarization dependence of specialized NbN test devices. First, we consider the energy downconversion cascade using the kinetic equations to describe the non-equilibrium electron and phonon systems immediately following photon absorption. These calculations provide estimates for the energy loss and fluctuations during this process, and provide qualitative information about the way energy is partitioned between the electron and phonon systems. To study the suppression of superconductivity following down-conversion, we apply the most advanced existing model, that of Vodolazov [6], but find it inadequate to quantitatively describe the timing properties of these detectors. By extending the model to use the generalized time-dependent Ginzburg–Landau equations, we achieve better quantitative agreement with experiment. However, the generalized model still provides only a qualitative picture of the detection process.

We also conduct an experimental examination of the heat transfer process in WSi nanowires by examining the nanowire reset dynamics, steady-state dissipation, and crosstalk between elements of an array. The results are compared to existing electrothermal models, but these models fail to adequately describe the dynamics of the system. A generalized form of the electrothermal model provides better fitting to experiment, but incorporation of non-equilibrium effects is likely needed to provide a fully quantitative description of the system. These results are directly connected to some of the thermal challenges of SNSPD array development. Informed by the crosstalk results, we demonstrate a new multiplexing technique based on thermal coupling between two active nanowire layers, known as the thermal row-column. This method promises to enable kilopixel to megapixel scale imaging arrays for low photon-flux applications. Finally, we discuss the design and characterization of the ground detector for the Deep Space Optical Communication (DSOC) demonstration mission.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] J. P. Allmaras, E. E. Wollman, A. D. Beyer, R. M. Briggs, B. A. Korzh, B. Bumble, and M. D. Shaw, "Demonstration of a thermally coupled row-column SNSPD imaging array", Nano Letters **20**, 2163–2168 (2020) `10.1021/acs.nanolett.0c00246`,
J. P. Allmaras designed and fabricated the devices, performed the experiments, analyzed the data, and wrote the paper with input from the other authors.

[2] B. Korzh, Q.-Y. Zhao, J. P. Allmaras, S. Frasca, T. M. Autry, E. A. Bersin, A. D. Beyer, R. M. Briggs, B. Bumble, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, A. E. Lita, F. Marsili, G. Moody, C. Peña, E. Ramirez, J. D. Rezac, N. Sinclair, M. J. Stevens, A. E. Velasco, V. B. Verma, E. E. Wollman, S. Xie, D. Zhu, P. D. Hale, M. Spiropulu, K. L. Silverman, R. P. Mirin, S. W. Nam, A. G. Kozorezov, M. D. Shaw, and K. K. Berggren, "Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector", Nature Photonics, 1–6 (2020) `10.1038/s41566-020-0589-x`,
J. P. Allmaras contributed to the experimental measurements and data analysis, contributed to the theory development, performed the numerical modeling, and contributed to writing the paper.

[3] J. P. Allmaras, A. G. Kozorezov, B. A. Korzh, K. K. Berggren, and M. D. Shaw, "Intrinsic timing jitter and latency in superconducting nanowire single-photon detectors", Physical Review Applied **11**, 034062 (2019) `10.1103/PhysRevApplied.11.034062`,
J. P. Allmaras contributed to the theory development, performed the numerical modeling, and contributed to writing the paper.

[4] A. N. McCaughan, V. B. Verma, S. M. Buckley, J. P. Allmaras, A. G. Kozorezov, A. N. Tait, S. W. Nam, and J. M. Shainline, "A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors", Nature Electronics **2**, 451–456 (2019) `10.1038/s41928-019-0300-8`,
J. P. Allmaras contributed to the theory development and performed the numerical modeling.

[5] J. P. Allmaras, A. G. Kozorezov, A. D. Beyer, F. Marsili, R. M. Briggs, and M. D. Shaw, "Thin-film thermal conductivity measurements using superconducting nanowires", Journal of Low Temperature Physics **193**, 380–386 (2018) `10.1007/s10909-018-2022-0`,
J. P. Allmaras designed the devices, performed the experiments, analyzed the data, performed the modeling, and wrote the paper with input from the other authors.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Ever since the superconducting nanowire single-photon detector (SNSPD or SSPD) was first demonstrated by Goltzmann et al. in 2001 [7], there has been rapid development of this technology [8–10]. Single pixels have shown photon-counting detection efficiency exceeding 90% in the near-infrared [11–13], intrinsic dark count rates below 0.5 counts per hour [14, 15], and timing jitter below 5 ps at telecommunication wavelengths [2]. Detectors have been optimized for performance over a wide spectral range, from the UV [14, 16] through the mid-infrared [17, 18], fabricated out of a variety of material systems [7, 19–21], and directly integrated on waveguides [22–24] for cryogenic integrated photonics applications. The maximum count rate (MCR) for a single pixel is typically a few tens of megacounts per second, leading to a remarkable dynamic range for a photon-counting detector. Most of these laboratory demonstrations have only achieved a single high performance metric with a given device, so recent attention has focused on demonstrating high-efficiency, low dark count rates, and high timing resolution simultaneously [25, 26].

The problem of scaling the performance of single pixels to large-format arrays has also been given much attention. Direct readout arrays have been demonstrated up to 64 pixels [27], but scaling to the kilopixel range is challenging due to the heat load placed on the cryogenic stage by each high bandwidth RF cable. To overcome this challenge, several multiplexing strategies have been proposed and demonstrated. One approach uses a row-column biasing scheme [28, 29] and correlations between detections on row and column readout channels to determine which pixel registered a detection. Multiplexing based on single-flux-quantum (SFQ) readout has been demonstrated in both standard arrays [30] and row-column arrays [31]. A more recent approach used the finite signal propagation speed of a long nanowire and differential readout to infer the photon absorption location along the length of the nanowire [32, 33]. Frequency multiplexing has also been demonstrated in both DC [34] and AC [35] nanowire bias configurations.

Owing to their high performance in the near-infrared, SNSPDs have become an integral component in quantum key distribution [36–38], laser communication from deep space [27, 39], and quantum science experiments [40]. If large-format arrays

can be further developed, emerging applications such as detectors for dark matter search efforts, telecom imaging LIDAR, and quantum imaging would benefit from the high-efficiency and low dark count rate performance offered by SNSPDs. By extending the wavelength sensitivity further into the mid-infrared, SNSPD arrays could be used as the focal plane arrays for astronomical spectroscopy [41]. Nearly two decades after their first demonstration, SNSPDs continue to be the focus of research and development, and their presence in commercial and research applications is only expected to grow as cryocooler technology continues to miniaturize and become more economically viable.

While this technological progress is remarkable, the SNSPD community still lacks a quantitatively accurate model of the photon-detection process in SNSPDs. Many theoretical works [3, 6, 7, 42–55] have improved the qualitative understanding of these devices, but they fail to make a direct connection between experiment and first principles calculations based on a microscopic model of the device using measurable material properties. Such a quantitative model would be extremely useful for device design and understanding the fundamental limits of detector performance. Furthermore, there are open questions about the thermal coupling of thin films to various dielectric substrates, and how heat propagates through these dielectrics. This has implications for the reset dynamics of SNSPDs [56, 57], coupling between pixels in arrays [1, 27], and the design of multilayer superconducting switches [4, 58].

This thesis aims to refine the current understanding of SNSPD physics through modeling and experimental characterization. By making a detailed comparison between new experimental data and model predictions, we identify the key weaknesses of the existing models and make refinements to address these shortcomings. The result is a model which comes much closer to quantitative accuracy, but further development is needed to produce a predictive quantitative model for use in designing new SNSPD architectures. This work also addresses the problem of thermal coupling in the thin dielectric layers typically used to enable high-efficiency SNSPDs. This first chapter provides relevant background about SNSPD operation which is needed to understand the more detailed modeling of the remaining work. Chapter 2 discusses the energy downconversion process when a photon is first absorbed in a nanowire. This downconversion process determines the partitioning of energy between the electron and phonon systems, and fluctuations during this process play a key role in understanding the internal efficiency and timing properties of these

detectors. Chapter 3 describes the recent experimental efforts which have enabled new methods of validating the predictions of microscopic detection models. This is followed by a comparison of model predictions based on the time-dependent Ginzburg–Landau (TDGL) equations to this new experimental work. To account for the failure of the existing model, several refinements are introduced which lead to improved agreement between experiment and model. Chapter 4 considers the challenges of understanding and modeling electrothermal evolution in the WSi material system. This addresses the difficulties in understanding reset dynamics and thermal coupling in WSi nanowire arrays. Chapter 4 also applies this electrothermal modeling to multilayer heater switches which are of technological interest for SNSPD readout and interfacing low impedance superconducting electronics with high impedance loads. Chapter 5 discusses the design and demonstration of a new SNSPD multiplexing technique which relies upon thermal coupling in a multilayer SNSPD array to achieve imaging capability in a row-column architecture, and Chapter 6 discusses the design and characterization of a prototype detector for the Deep Space Optical Communication (DSOC) demonstration mission. The final chapter provides a summary and an outlook for the field.

1.1 Basic SNSPD Operation

At a fundamental level, an SNSPD is simply a current-carrying superconducting nanowire. The wires are typically a few nm thick, have widths ranging from a few tens of nm [59] to a few µm [60–62], and have lengths on the order of hundreds of µm to several mm. In typical devices, this nanowire is meandered over an active area (often on the order of several hundred $µm^2$ for single pixels) in order to increase the optical absorption of the device. SNSPDs are usually current-biased and operated at temperatures of 1–4 K depending on the material system used for the device. Current-biasing is achieved through the use of a bias-tee, but the readout portion of the tee can be AC, DC, or transformer coupled to an amplifier based on the system requirements. The basic detection process of an SNSPD is shown in Figure 1.1a. The detection process begins with the nanowire in the current-carrying state until a photon is absorbed, triggering a cascade of events which lead to the registered detection of a single photon.

In order for an incoming photon to be detected by the nanowire, it must first be absorbed by the active region of the detector. This absorption process is well understood. Classical electromagnetic theory is suitable for describing the absorption of photons in SNSPDs, but this typically requires the use of modeling software in order
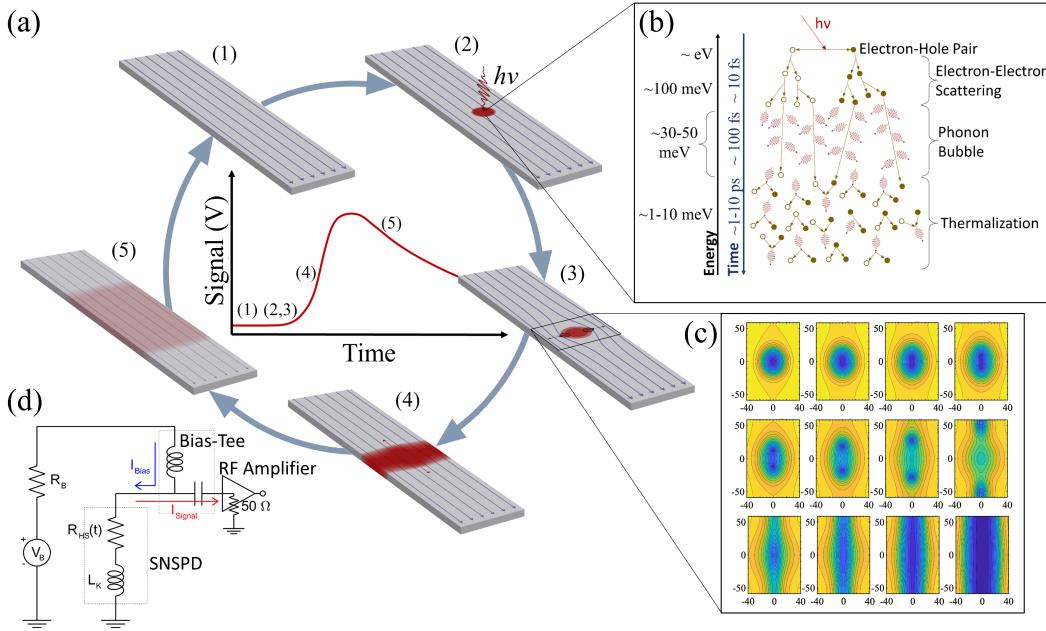
Figure 1.1: (a) Schematic of the SNSPD detection process. The device begins in the current-carrying superconducting state (1) until a photon is absorbed (2) and generates a hotspot of excited quasiparticles and high-energy phonons through a process of downconversion shown in (b). The evolution of these quasiparticles leads to the instability of the superconducting state, resulting in the formation of vortices (vortex-antivortex pair generation if the hotspot is in the center of the nanowire or vortex entry from the edge if the absorption is near the edge) as shown in (c). As the vortices move due to the current flow (3), they dissipate energy, leading to the formation of a normal domain across the entire cross-section of the nanowire. Once formed, the normal domain grows along the length of the wire due to Joule heating (4), increasing the impedance of the device. This change in impedance diverts current from the nanowire to the readout circuitry, leading to a voltage transient (center). Once current is diverted to the readout, the nanowire cools, recovers to the superconducting state, and current returns to the device. (d) A typical SNSPD is operated using a bias-tee with the DC port carrying the bias current and the RF port coupled to a low-noise amplifier.

to design and optimize practical devices. For periodic structures, as is typically the case in meandered SNSPDs, the rigorous coupled-wave analysis (RCWA) technique can be used to efficiently solve Maxwell's equations for incident plane waves on the system [63–65]. The RCWA method converts the electromagnetic problem to the Fourier basis where it can be solved numerically to determine the fraction of the incident light absorbed in the nanowire. This technique can also extract the electric field profile within the nanowire. RCWA is numerically simple to solve, making it ideal for the optimization of optical cavities which are used to enhance the absorption of light in the active area of the nanowire. This is the technique used to achieve high efficiency in many devices [11, 13, 27]. For non-periodic structures, finite-difference time-domain (FDTD) solvers are employed to model the evolution of the electric field interacting with a device, but the computational complexity of this method is much higher than that of the RCWA method. As part of my graduate work, significant effort was devoted to improving the RCWA capabilities of the JPL group and optimizing the optical cavity design for the DSOC demonstration mission ground detector, the results of which are summarized in Chapter 6.

For the visible and near-infrared wavelengths typically measured with SNSPDs, the absorption of a photon generates an electron-hole pair. This excitation rapidly interacts with the electronic and lattice systems to generate a thermalized hotspot, as shown in Figure 1.1b. The details of this downconversion process are discussed and modeled in Chapter 2. First, the electron and hole interact with the electronic system through high momentum transfer collisions, leading to a small number of high-energy electron and hole excitations. Once the energy of these excitations approaches a few times the Debye energy, interactions with phonons dominate, and nearly all of the energy is transferred to the phonon system, in what is known as the phonon bubble. While the phonons of the phonon bubble interact with lower energy electrons, some of these phonons can escape to the substrate, contributing to fluctuations in the detection process [53]. After a certain time, which is on the order of a few picoseconds [6], the electron and phonon systems thermalize, meaning they approach their equilibrium distributions at an elevated effective temperature. This region of excitations is typically referred to as a hotspot [7, 66].

As the number of quasiparticles increases during the downconversion process, the superconducting condensate begins to feel the effect of these excitations. The presence of high-energy excitations breaks Cooper-pairs and suppresses superconductivity. If the photon has sufficient energy, the superconducting state becomes

unstable. The exact details of this process have been discussed in the literature for years [3, 6, 7, 42–55], but the current understanding [6] is that the suppression of superconductivity leads to the formation of vortices or phase-slip lines in the nanowire. The nature of this process depends on the bias current in the nanowire, the operating temperature, the width of the nanowire, and the location where the hotspot is formed. For typical geometries, if the hotspot is located in the center of the nanowire, a vortex-antivortex pair forms, with each moving toward the edge of the nanowire. If the photon is absorbed near the edge of the nanowire, a vortex enters from the edge and traverses the width of the wire. In either case, the motion of vortices heats the superconductor and contributes to suppression of the super-conducting state. In the limit of narrow nanowires, 2D vortices are not supported within the film, and instead, superconductivity is suppressed through 1D phase-slip lines. Chapter 3 is dedicated to modeling the details of this process of breaking superconductivity following the formation of a hotspot. As superconductivity is suppressed, this local region of the nanowire transitions to the normal state. This region of normal metal is known as a normal domain.

Once a normal domain reaches across the width of the nanowire, it expands along its length due to Joule heating of the normal metal. Joule heating continues as long as current flows through the normal domain, but the normal domain has a large impedance. The typical sheet resistance for SNSPD materials ranges from a few hundred to a thousand ohms per square and the normal domain typically expands to a length of a few to a few tens of squares depending on the device geometry. The large impedance caused by the normal domain diverts the bias current from the nanowire to the readout amplifier, which typically has a $50\,\Omega$ input impedance. Once the current has been diverted to the amplifier, heating stops, and the normal domain cools and returns to the superconducting state. Once superconducting, the bias current returns to the nanowire.

The reset characteristics of SNSPDs are strongly influenced by the properties of the superconductor as well as the readout and biasing circuitry used to operate the device. In typical devices, the electrical performance can be understood from a simple lumped element model as shown in Figure 1.1c. The SNSPD is modeled as a variable resistor in series with an inductor. Thin-film superconducting nanowires have a large kinetic inductance L_k, which is proportional to the number of squares (length/width) of the device. This kinetic inductance typically far exceeds any geometric contribution to the inductance, so the series inductor in the lumped

element model represents the kinetic inductance. The variable resistor encodes the time-dependent resistance of the normal domain following the absorption of a photon. SNSPDs are typically current biased using either an inductive or resistive bias-tee, and a low-noise amplifier is used to increase the signal-to-noise ratio of the voltage transient. Based on this circuit model, it is clear that the electrical recovery of current from the amplifier to the nanowire after photon detection is governed by the $\tau_R = L_k/R_L$ time constant, where R_L is the load resistance (50 Ω in this example). This recovery time limits the MCR achieved by the detector.

Decreasing L_k is a common technique for increasing the MCR, but this only works up to a certain point. If the reset time is too fast, the nanowire can enter a stable state with a self-heating hotspot [56, 67, 68] in a process known as latching. When the kinetic inductance is small or the load resistor is large, current is shunted from and returns to the nanowire more quickly. If current returns to the nanowire before the superconductor has fully cooled to the superconducting state, the normal domain persists and the device does not self-reset. In a similar process, if the normal domain fully relaxes but current returns faster than the nanowire returns to the bath temperature, the current can exceed the temperature-dependent switching current, leading to a secondary detection pulse. This is known as afterpulsing. In practice, there are additional features which can influence the latching behavior of devices. Electrical reflections due to impedance mismatches at the bias-tee or amplifier input can return to the device, leading to an increase in bias current which can exceed the switching current. At high count rates, traditional AC coupling using a bias-tee can lead to additional current flow in the detector due to reflected low frequency components of the detector signal. This additional bias current can lead to switching of the device. Various strategies such as using DC coupled amplifiers[69], adding shunt resistors or inductors, and careful management of impedance mismatch have been employed to reduce these electrical contributions to latching.

Understanding the details of latching and afterpulsing requires a model of the coupled electrical and thermal evolution of the device. This electrothermal modeling has increased in sophistication since its first use in SNSPDs [56, 68], but its quantitative accuracy has not been verified in all of the material systems commonly used in these detectors. Chapter 4 considers the electrothermal behavior and modeling of the WSi material system, which is not well described by the existing electrothermal models. The current modeling only considers the lumped element form of Figure 1.1. For an accurate picture of the electrical dynamics of SNSPDs, it is sometimes necessary

to consider the microwave transmission properties of the nanowire [70]. Design of nanowires as transmission lines can enable imaging capabilities [32], reduction of timing jitter [26, 32], and improve signal amplification through impedance-matching tapers [71]. Fully understanding the implications of microwave signal propagation in SNSPDs is still an area of active research, but for most applications, the lumped element model is sufficient for semi-quantitative analysis.

Both the detection process and latching are strongly dependent on the material properties of the superconductor used in the SNSPD and on the device geometry. While there is general consensus on the types of materials which make good SNSPDs, the community does not have an accurate means of predicting the detector performance for an arbitrary set of material properties. Given that characteristics such as the long-wavelength sensitivity, MCR, and timing jitter all impact the applications which use SNSPDs, understanding the tradeoffs which occur due to design decisions is essential for providing fully optimized systems. Improving the semi-quantitative understanding of these detectors to enable better optimized designs is the underlying motivation for the modeling efforts of this thesis.

Chapter 2

ENERGY DOWNCONVERSION CASCADE

2.1 Introduction

In the initial moments after a photon is absorbed in a current-carrying nanowire, the behavior of the system is governed by the dynamics of a low number of high-energy excitations in the electronic and lattice systems. Despite their short lifetimes, the interactions of these excitations leave their imprints on the eventual suppression of superconductivity in the system. Understanding these dynamics is a prerequisite for generating a predictive model of the SNSPD detection mechanism.

Within the SNSPD community, there have been only a few publications aimed at explaining the details of the rapidly evolving downconversion process. The primary work on the subject was produced by Vodolazov in 2017 [6] and studies the downconversion cascade based on application of the kinetic equations. This work examines the non-equilibrium evolution of the electron and phonon systems when subject to a single high-energy excitation, as occurs during single-photon detection. There are several conclusions relevant to understanding the SNSPD detection mechanism. First, the work finds that for the energies involved for most SNSPDs, the process is not sensitive to the initial conditions. When the energy is concentrated in high-energy electron excitations (electron bubble), high-energy phonon excitations (phonon bubble), or an equilibrated electron system (equilibrated electrons), the final energy distribution between electrons and phonons is approximately equal after some thermalization time. The second result is that the downconversion process takes an amount of time which is comparable to or shorter than the characteristic timescale of order parameter evolution $\tau_{|\Delta|} = \hbar/k_B T_c$.[1] Fully modeling the detection process in SNSPDs requires simulating the downconversion process as well as the subsequent suppression of superconductivity, but it is not practical to use the kinetic equations of downconversion to describe the full evolution of the superconductor due to the numerical complexity. Defining the thermalization time as the time required for quasi-equilibration of the electron system, when the thermalization time is shorter than the order parameter evolution timescale, the solution of the evolution

---

[1]For simplicity, we will use the definition of $\tau_{|\Delta|} = \hbar/k_B T_c$ throughout this work, but this is only an approximation. The true timescale of the evolution of the order parameter depends on the magnitude of the perturbation to the system.

of the order parameter can be decoupled from the initial downconversion calculations. This is done by using an appropriate set of initial conditions, based on the downconversion calculations, when modeling the suppression of superconductivity [6].

A second work takes a higher level approach by using a simple model to understand fluctuations during the downconversion process [53]. This latter work builds on a number of earlier works describing the same process in other superconducting detectors [72, 73]. This model treats the energy downconversion as a lossy process where the initial generations of high-energy phonons can escape to the substrate before thermalizing. Fluctuations in the number of escaping phonons lead to fluctuations in the total energy deposited in the superconductor once the system has reached its quasi-equilibrated state, considered to be the initial conditions for the subsequent evolution of the system. The main parameters of this model are the mean fraction of energy deposited $\bar{\chi}$ by the photon and a Gaussian distribution of the fluctuations, parameterized by $\sigma_F \sim \sqrt{E_\lambda}$. The square root dependence of the standard deviation $\sigma_F \sim \sqrt{E_\lambda}$ on the photon energy E_λ originates from the number of Debye phonons ($N_{ph,D}$) being proportional to E_λ and the shot-noise fluctuations being proportional to $\sqrt{N_{ph,D}}$. These two parameters are used as free parameters to fit experimental data in WSi [53] and NbN [2, 3] with reasonable success.

In this chapter, we reexamined the downconversion process with the goals of estimating the energy loss due to phonon escape and characterizing the fluctuations in this energy escape. This seeks to justify the assumptions used in the fitting procedure of the work on downconversion fluctuations [53] through a more rigorous microscopic approach as employed in the work by Vodolazov [6]. Understanding both energy loss and fluctuations of this loss is crucial to developing a predictive model of the SNSPD detection process. Furthermore, we seek to understand the appropriate initial conditions for describing the nanowire system when modeling the subsequent suppression of superconductivity following downconversion. This modeling, based on the time-dependent Ginzburg–Landau (TDGL) equations, will be discussed in detail in Chapter 3, but the initial conditions used to seed the evolution of the system play a crucial role in the system dynamics. These equations also assume that the electron and phonon systems are well described by their equilibrium distribution functions, a criterion which is not met during the initial downconversion. Calculating the thermalization time required to reach near equilibrated electrons and phonons is another goal of this chapter. Understanding how to mesh the downconversion pro-

cess with the TDGL model is of significant importance to developing a model of photon detection. Therefore, $\tau_{|\Delta|}$ is a useful timescale to consider for its implications in how to transition from calculations of downconversion to calculations within the TDGL model.

We begin by examining Vodolazov's model [6] based on the kinetic equations to determine the accuracy of various simplifying assumptions. We consider the effects of the initial conditions, diffusion, and electron-electron interactions on the subsequent downconversion process. Next, we employ a more advanced kinetic equation to describe the non-equilibrium phonon distribution, which includes the angle and position dependence of the phonon modes, under the assumption of equilibrated electrons. This one-temperature approach greatly simplifies the computational complexity of the problem while enabling a more detailed calculation of the phonon trapping effects in the superconductor. By employing the one-temperature model, we estimate the energy loss and fluctuations during downconversion, which provides information about the initial state of the system when modeling the suppression of superconductivity during single-photon detection. Finally, we calculate the response of the system using the various models (kinetic, one-temperature, and two-temperature) to compare their accuracy, and consider the corrections which may be needed to improve the consistency between the simplified models based on equilibrated electrons and phonons and the full non-equilibrium formulation.

2.2   Vodolazov Model

We begin by summarizing the existing work by Vodolazov [6] which uses the numerical solution of a set of kinetic equations to describe the evolution of the electron and phonon distribution functions in response to a high-energy excitation. In that work, the full kinetic equations are given by

$$N_1 \frac{\partial n}{\partial t} = D\nabla \left[\left(N_1{}^2 - R_2{}^2 \right) \nabla n \right] - R_2 \frac{\partial n}{\partial \epsilon} \frac{\partial |\Delta|}{\partial t} + I_{e-ph}\left(n, N\right) + I_{e-e}(n) \qquad (2.1)$$

for the electron system and

$$\frac{\partial N}{\partial t} = -\frac{N - N_{eq}(T_{sub})}{\tau_{esc}} + I_{ph-e}(N, n) \qquad (2.2)$$

for the phonon system, where $n(\epsilon)$ is the electron distribution function, ϵ is the energy, $N(\epsilon)$ is the phonon distribution function, N_1 and R_2 are spectral functions, D is the diffusion coefficient, $|\Delta|$ is the magnitude of the order parameter, τ_{esc} is the phonon escape time, and $N_{eq}(T_{sub})$ is the equilibrium phonon distribution at

the substrate temperature T_{sub}. For the electron distribution function, the energy is relative to the Fermi energy. A set of collision integrals describes the interactions between the electron and phonon systems, as well as interactions within the electron system. These functions are I_{e-e} for electron-electron collision integral, I_{e-ph} for the electron-phonon collision integral, and I_{ph-e} for the phonon-electron collision integral.

During the initial downconversion, the distribution functions evolve on a timescale faster than $\tau_{|\Delta|}$, so it is convenient to assume that $|\Delta|$ is constant with time. Under this assumption Vodolazov finds that the influence of the spectral functions is minimal. It is not surprising that the superconducting state has little effect on the downconversion process because the spectral functions modify the density of states and coherence factors of the electron-electron collision integral most significantly near the energy gap, which is at an energy scale much smaller than the energies dominating the downconversion process. Using Vodolazov's conclusion [6], we treat the superconductor as a normal metal by neglecting the spectral functions, setting $N_1 \rightarrow 1$ and $R_2 \rightarrow 0$. We also normalize all energy scales by $k_B T_c$. In Vodolazov's calculations, diffusion is also neglected, but it is not yet clear how significantly this simplification modifies the results of these calculations. Under these simplifications, the kinetic equation for the electronic system is reduced to

$$\frac{\partial n}{\partial t} = I_{e-ph}(n, N) + I_{e-e}(n) \tag{2.3}$$

and the electron-electron collision integral is written as

$$
\begin{aligned}
I_{e-e}(n) = -\frac{\alpha_{e-e}}{\tau_0} \int_0^\infty \int_0^\infty d\epsilon_1 d\epsilon_2 \\
a \Bigg[\frac{1}{|\epsilon - \epsilon_1|} \Big[n_\epsilon (1 - n_{\epsilon_1})(1 - n_{\epsilon_2})(1 - n_{\epsilon - \epsilon_1 - \epsilon_2}) \\
- (1 - n_\epsilon) n_{\epsilon_1} n_{\epsilon_2} n_{\epsilon - \epsilon_1 - \epsilon_2} \Big] \Theta(\epsilon - \epsilon_1 - \epsilon_2) \\
+ \left(\frac{1}{|\epsilon + \epsilon_1|} + \frac{2}{|\epsilon - \epsilon_2|} \right) \Big[n_\epsilon n_{\epsilon_1} (1 - n_{\epsilon_2})(1 - n_{\epsilon + \epsilon_1 - \epsilon_2}) \\
- (1 - n_\epsilon)(1 - n_{\epsilon_1}) n_{\epsilon_2} n_{\epsilon + \epsilon_1 - \epsilon_2} \Big] \Theta(\epsilon + \epsilon_1 - \epsilon_2) \\
+ \left(\frac{1}{|\epsilon - \epsilon_1|} + \frac{2}{|\epsilon + \epsilon_2|} \right) \Big[n_\epsilon (1 - n_{\epsilon_1}) n_{\epsilon_2} n_{-\epsilon + \epsilon_1 - \epsilon_2} \\
- (1 - n_\epsilon) n_{\epsilon_1} (1 - n_{\epsilon_2})(1 - n_{-\epsilon + \epsilon_1 - \epsilon_2}) \Big] \Theta(-\epsilon + \epsilon_1 - \epsilon_2) \Bigg],
\end{aligned}
\tag{2.4}
$$

where a is a coefficient of order unity and Θ is the Heaviside function.[2] For all of our calculations, we take $a = 1$. The distribution function with a subscript indicates the value of the distribution function at the energy of the subscript. The strength of the electron-electron interaction is dominated by the term

$$\alpha_{e-e} = \tau_0 \frac{k_B T_c}{4\hbar} \frac{\rho_\square}{R_Q} \tag{2.5}$$

where τ_0 describes the strength of electron-phonon interactions, $\rho_\square$ is the film sheet resistance, and $R_Q = \frac{\hbar}{e^2}$ is the resistance quantum. A cursory inspection of (2.4) raises concerns about the presence of singularities at particular combinations of ϵ, ϵ_1, and ϵ_2. However, analysis in the region of the potential singularities reveals that terms of the integral cancel to prevent the integral from becoming unbounded, as shown in Appendix A.1. The electron-phonon collision integral is given by

$$I_{e-ph}(n, N) =$$
$$-\frac{1}{\tau_0}\left[\int_{\max(0,\epsilon-\Omega_D)}^{\epsilon} d\epsilon_1 (\epsilon - \epsilon_1)^2 \left[(1 + 2N_{\epsilon-\epsilon_1})(n_\epsilon - n_{\epsilon_1}) + n_\epsilon(1 - 2n_{\epsilon_1}) + n_{\epsilon_1} \right] \right.$$
$$+ \int_{\epsilon}^{\epsilon+\Omega_D} d\epsilon_1 (\epsilon - \epsilon_1)^2 \left[(1 + 2N_{\epsilon_1-\epsilon})(n_\epsilon - n_{\epsilon_1}) - n_\epsilon(1 - 2n_{\epsilon_1}) - n_{\epsilon_1} \right]$$
$$+ \left. \int_0^{-\epsilon+\Omega_D} d\epsilon_1 (\epsilon+\epsilon_1)^2 \left[(1 + 2N_{\epsilon_1+\epsilon})(n_\epsilon + n_{\epsilon_1} - 1) - n_\epsilon(1 - 2n_{\epsilon_1}) - n_{\epsilon_1} + 1 \right] \right] \tag{2.6}$$

where Ω_D denotes the Debye energy normalized by $k_B T_c$.[3] Finally, the phonon-electron collision integral is given by

$$I_{ph-e}(N, n) = \frac{\gamma}{\tau_0}\left[\int_0^{\epsilon} d\epsilon_1 \left(n_{\epsilon_1} n_{\epsilon-\epsilon_1} + N_\epsilon \left(n_{\epsilon-\epsilon_1} + n_{\epsilon_1} - 1 \right) \right) \right.$$
$$+ \left. \int_0^{\infty} d\epsilon_1 \left((1 - n_{\epsilon_1}) n_{\epsilon+\epsilon_1} + N_\epsilon \left(n_{\epsilon+\epsilon_1} - n_{\epsilon_1} \right) \right) \right], \tag{2.7}$$

where the parameter γ describes the ratio of heat capacities of the electron and phonon systems at T_c and has the definition

$$\gamma = \frac{8\pi^2}{5} \left.\frac{C_e}{C_{ph}}\right|_{T=T_c} = \frac{4\pi^4}{15} \frac{E_0}{B_{ph}T_c^4} \tag{2.8}$$

where $E_0 = 4N(0)k_B^2 T_c^2$ is a characteristic energy density, $N(0)$ is the electron single-spin density of states at the Fermi level, $B_{ph} = \frac{2\pi^2 k_B^4}{5\hbar^3 c_{avg}^3}$ describes the phonon

---

[2]A minus sign has been added to the equation compared to [6].

[3]The lower limit of integration on the first term of the electron-phonon has been modified from 0 of [6] to $\max(0, \epsilon - \Omega_D)$.

heat capacity according to the Debye model with $C_{ph}(T) = B_{ph}T^3$, and c_{avg} is the mode-averaged sound velocity.

Using this notation, the energy density of the electron system is given by

$$E_e = E_0 \int_0^\infty d\epsilon \, \epsilon \, n \tag{2.9}$$

and the energy density of the phonon system is

$$E_{ph} = \frac{E_0}{\gamma} \int_0^{\Omega_D} d\epsilon \, \epsilon^3 \, N. \tag{2.10}$$

The total escaped energy after a photon is absorbed at time $t = 0$ is given by

$$E_{esc}(t) = \frac{E_0}{\gamma} \int_0^t dt' \epsilon^3 \left(\frac{N(t') - N_{eq}(T_{sub})}{\tau_{esc}} \right). \tag{2.11}$$

It should be noted that this formulation is not entirely self-consistent. The electron-electron collision integral is formulated for a 2D disordered metal while the electron-phonon and phonon-electron collision integrals assume the clean limit of a 3D normal metal. This form also assumes strong elastic scattering for the electronic distribution and isotropic scattering for the electron-phonon and phonon-electron interactions which removes the directional dependence of the phonon distribution function.

Using these equations, we can first confirm that our formulation is consistent with the results of Vodolazov. This is demonstrated by reproducing Figure 1 in [6], as shown below in Figure 2.1. There is agreement between our formulation and the published results. This validates the simplification of ignoring the superconducting state and treating the system as if it were a normal metal.
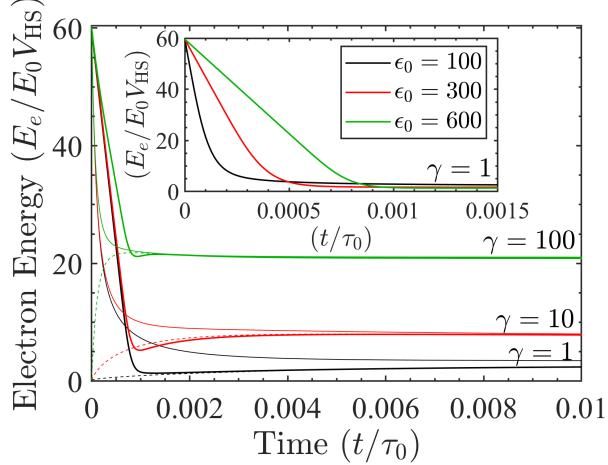
Figure 2.1: Energy in the electron system during downconversion for different initial conditions. The thick solid lines indicate an electron bubble with $\epsilon_0 = 600 k_B T_c$ and $\alpha_{e-e} = 0$, the dashed lines indicate a phonon bubble with $\epsilon_0 = 30 k_B T_c$ and $\alpha_{e-e} = 0$, and the thin solid lines indicate a thermalized electron system with $\alpha = 500$. In all cases, the total energy was kept the same at ~ 1.3 eV. The black curves use $\gamma = 1$, the red curves use $\gamma = 10$, and the green curves use $\gamma = 100$. The inset shows the energy of the electron system for different electron-bubble initial conditions with $\alpha_{e-e} = 0$ and $\gamma = 1$. There is agreement between these results and those of Figure 1 from Vodolazov [6].

2.3 Downconversion in NbN

In order to make a connection between the kinetic model of downconversion and experimental measurements, we must select an appropriate set of material parameters for our calculations. The most complete experimental dataset probing the intrinsic detection properties of NbN nanowires comes from a set of devices fabricated by MIT and tested at NIST and JPL [2], so we use the properties of those devices as the basis of our calculations. These are the same devices studied in Chapter 3. We take the parameters $T_c = 8.65$ K, $\tau_0 = 1.87$ ns, which comes from scaling $\tau_{ep}(10 \text{ K}) = 16$ ps to the appropriate T_c assuming a cubic power law and $\tau_0 = \tau_{ep}(T_c)\frac{720\zeta(5)}{\pi^2}$ where $\zeta(5) \approx 1.037$ is the Riemann zeta function evaluated at 5, and $\tau_{esc} = 10$ ps which is based on calculations of the acoustic mismatch model. For the phonon system, we assume an average sound velocity of $v_{avg} = 4912$ m/s, which comes from the sound velocity extracted from the elastic constants measured by neutron scattering of δ-NbN [74]. Using a mass density of 8.3 g/cm^3 [75], and an average atomic mass of $(92.9 + 14.0)/2$, we arrive at a Debye temperature of 664 K, and $\Omega_D = 76.7$ in units of $k_B T_c$. Using a sheet resistance $\rho_\square$ of $600\,\Omega/\square$, thickness d of 7 nm, and diffusion coefficient $D = 0.5$ cm^2/s, we arrive

at a single-spin density of states $N(0) = 1.49 \times 10^{22} \, \text{eV}^{-1}\text{cm}^{-3}$ using the Einstein relation $N(0) = 1/(2e^2 D \rho_\square d)$ and characteristic energy scale $E_0 = 33.0 \, \mu\text{eV/nm}^3$. This also leads to the phonon parameter $\gamma = 23.8$. These parameters lead to a zero temperature BCS coherence length $\xi(0) = \sqrt{\frac{\hbar D}{\Delta_0}} = 5.0 \, \text{nm}$, and a coherence length of $\xi = 6.2 \, \text{nm}$ at a substrate temperature of $T_{sub} = T_c/2$ [76], which we will use in all of the following calculations unless otherwise noted.

As a preliminary test, we model the downconversion for a hotspot with radius of $\xi = 6.2 \, \text{nm}$ and a photon energy of 1 eV deposited as an electron bubble for various electron energies. We model the electron bubble as a narrow Gaussian distribution in n with a total energy corresponding to the excitation energy. For an electron bubble with mean energy centered at $671 k_B T_c$, this corresponds to the excitation of a single electron-hole pair. For these simulations, the electron-electron collision integral is neglected. The results, shown in Figure 2.2, indicate that the choice of material properties for the NbN can have an impact on the qualitative conclusions reached from the simulation of downconversion. The choice of a larger Debye energy leads to significantly faster coupling of energy from the initial electron bubble to the phonons. This result suggests that the initial stages of downconversion are faster than predicted using the parameters of Vodolazov's work. Furthermore, because the initial electron-phonon coupling is so rapid, it suggests that neglecting the electron-electron collision integral is an acceptable simplification. The difference in the excitation spectrum of the energy deposition has only a small effect on the overall downconversion results. Figure 2.2b shows the first instants of the downconversion for the different numbers of electron-hole pair excitations, revealing that the rate of initial energy transfer is proportional to the number of excitations. However, this initial difference is quickly suppressed as the system equilibrates. The total amount of escaped energy also does not depend on the spectrum of the electron bubble.
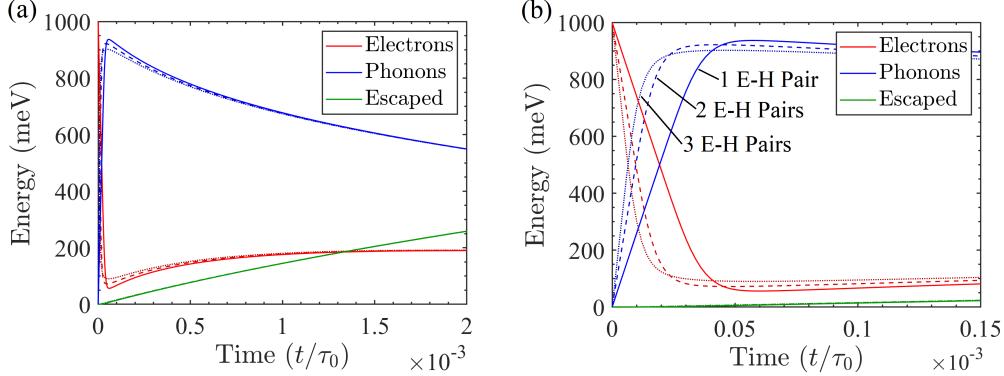
Figure 2.2: Simulated downconversion with NbN material parameters for a 1eV excitation. The energy was distributed as a single electron-hole pair (solid), two electron-hole pairs (dashed), and three electron-hole pairs (dotted). (a) On longer timescales, the difference in initial conditions becomes negligible. (b) By focusing on the shortest timescales, it is clear that the initial rate of energy transfer from the electron bubble to the phonon system is proportional to the number of electron-hole pairs, which is expected because the phonon emission rate does not depend on the electron energy when the energy remains greater than the Debye energy.

We can also probe the response of the system to different initial photon energies, when excited in the form of a single electron-hole pair. Figure 2.3 shows the system response for photon excitations of 0.5 eV, 1 eV, and 2 eV. The behavior is qualitatively consistent between all photon energies, but the initial transfer of energy from the electron to phonon system takes longer for higher energy photons. This result is as expected, but electron-electron collision scattering will become dominant as the photon energy increases [72], leading to a multiplication of initial electron excitations and in turn, a more rapid transfer of energy to the phonon system. This is much faster than the timescale of variation of the superconducting order parameter ($\tau_{|\Delta|} \sim \tau_0 \cdot 0.5 \times 10^{-3}$) and is not expected to play an important role in the timing response of the detector. By normalizing the response by the photon energy as in Figure 2.3b, we observe that as the photon energy increases, a greater fraction of the thermalized energy lies in the phonon system compared to the electron system. This is due to the $\sim T^3$ heat capacity of the lattice compared to the $\sim T$ dependence of the heat capacity of the normal state electronic system.
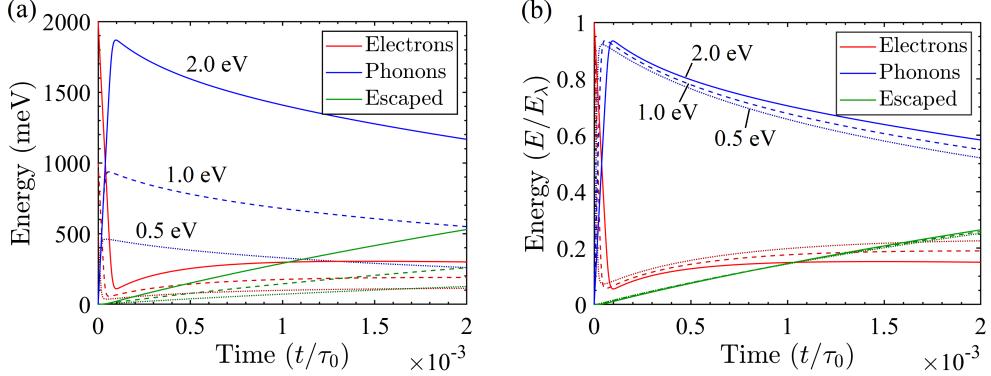
Figure 2.3: Simulated downconversion with NbN material parameters for (a) various photon energies. The energy was distributed as a single electron-hole pair for 0.5 eV (dotted), 1 eV (dashed), and 2 eV (solid) photon energies. The initial downconversion timescale is slower for higher energy photons. (b) When normalized by the photon energy, it is clear that more energy remains in the phonon system as the system thermalizes for higher energy photons.

2.3.1 Distribution Functions

The kinetic equations used throughout this section are not limited to tracking the total energy in electron and phonon systems. In fact, the strength of this set of equations is its ability to model the full distribution function of the electrons and phonons. This provides additional information about which modes are the dominant ones involved in the downconversion process.

Using the same simulation results as shown in Figure 2.2, we can track the evolution of the distribution function for both the electrons and phonons. Figure 2.4 shows the energy density distribution for the single electron-hole pair excitation with an energy of 1 eV in a hotspot with a radius of 6.2 nm. For comparison, the equilibrium distributions for the electron and phonon systems having the same energy at a given time are shown in the plot. Both the electron and phonon systems deviate from their equilibrium distributions, with the excitations being concentrated at higher energies compared to equilibrium. The phonon system is particularly far from the equilibrium distribution as shown in Figure 2.4b.
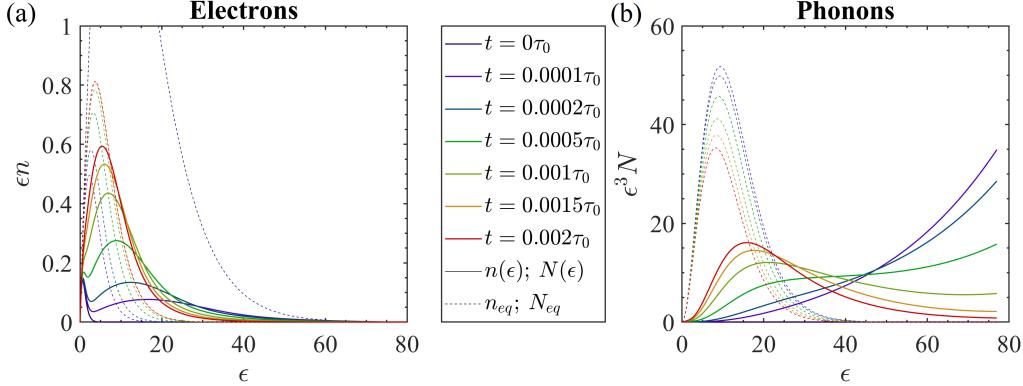
Figure 2.4: Energy density distribution functions during downconversion using NbN material parameters for the (a) electron and (b) phonon systems. The deposited energy is 1 eV for a hotspot radius of 6.2 nm, corresponding to a single electron-hole pair. The solid curves indicate the non-equilibrium distribution while the dashed curves are the equilibrium distribution for the same energy at the given time. At the final time ($t = 0.002\tau_0$), the effective temperature based on the energy of the electron and phonon systems is $2.92\,T_c$ for both the electron and phonon systems.

For most situations, analyzing the full distribution function is cumbersome, but we can estimate the number of excitations above the equilibrium level to provide a quick view of how the system is evolving. The expected number of quasiparticle excitations is given by

$$\bar{n} = 4N(0)k_BT_c \int dV \int_0^\infty d\epsilon \, (n - n_{eq}(T_{sub})) \qquad (2.12)$$

where the factor of 4 accounts for the two spins and the symmetry of electrons and holes (recall that ϵ is normalized by k_BT_c), while the expected number of phonon excitations is given by

$$\bar{N} = \frac{E_0}{\gamma k_BT_c} \int dV \int_0^{\Omega_D} d\epsilon\epsilon^2 (N - N_{eq}(T_{sub})). \qquad (2.13)$$

The average energy per excitation is simply the total energy exceeding the equilibrium energy at the substrate temperature divided by the expected number of excitations. The average excitation number and average energy per excitation are shown in Figure 2.5 for the same simulation. As energy is initially transferred from the high-energy electrons to phonon system, the number of phonon excitations rapidly increases, while the number of electron excitations only increases slowly. Once the phonon system is populated, low-energy electron excitations appear due to their interaction with the phonon system. The population of first-generation high-energy phonons is often called the phonon bubble. During the generation of the

phonon bubble, all modes are excited with equal probability. In this limit, we can calculate the average energy per mode analytically and confirm that the numerical results agree with the analytical calculation. For a constant occupation $N_{Ph.B.}$, the average excitation energy is

$$\bar{\epsilon}_{ph} = \frac{\frac{E_0}{\gamma} \int dV \int_0^{\Omega_D} d\epsilon \, \epsilon^3 \, N_{Ph.B.}}{\frac{E_0}{\gamma k_B T_c} \int dV \int_0^{\Omega_D} d\epsilon \epsilon^2 N_{Ph.B.}} = \frac{3}{4} \Omega_D k_B T_c = 43 \, \text{meV}, \qquad (2.14)$$

which corresponds exactly to the value calculated from the simulation during the initial transfer of energy from the electron to phonon system, as seen in Figure 2.5b.
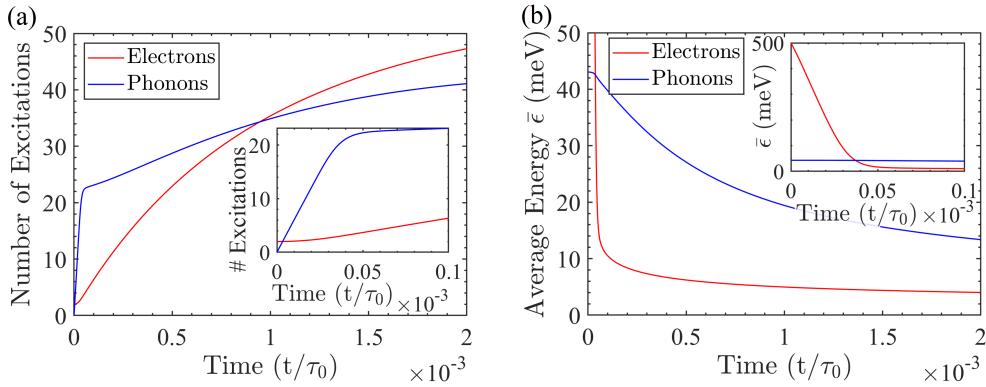
Figure 2.5: (a) Average number of excitations during downconversion using NbN material parameters. The inset shows the initial moments of downconversion and the linear rise in phonon number as the high-energy electrons couple to all phonon modes equally. (b) Average energy per excitation. The initial average phonon energy is 43 meV as predicted by (2.14). The inset shows the initial linear drop in electron energy on the same timescale as the number of phonon excitations increases.

2.3.2   Diffusion

In (2.3), diffusion was neglected, under the justification that on the timescales relevant to thermalization (on the order of $\tau_{th} \sim 1$ ps), the characteristic lengthscale of diffusion $l \sim \sqrt{D\tau_{th}}$ is on the order of 7 nm when assuming $D = 0.5 \, \text{cm}^2/\text{s}$. This puts the diffusive length scale at about the same distance as the coherence length. To study the implications of assuming a fixed hotspot size, we reinstate the diffusion term under the assumption of using the spectrum of a normal metal. The kinetic equation for the electron distribution function takes the form

$$\frac{\partial n}{\partial t} = D\nabla^2 n + I_{e-ph}(n, N) + I_{e-e}(n), \qquad (2.15)$$

and when posed in terms of cylindrical coordinates as is natural for thin films, we have the form

$$\frac{\partial n(\epsilon, r)}{\partial t} = D\left(\frac{\partial^2 n(\epsilon, r)}{\partial r^2} + \frac{1}{r}\frac{\partial n(\epsilon, r)}{\partial r}\right) + I_{e-ph}\left(n(\epsilon, r), N(\epsilon, r)\right) + I_{e-e}(n(\epsilon, r)).$$
(2.16)

For films where the thickness is comparable to or larger than the relevant diffusive length scale, it is informative to consider the 3D case by posing the problem in spherical coordinates. This leads to the form

$$\frac{\partial n(\epsilon, r)}{\partial t} = D\left(\frac{\partial^2 n(\epsilon, r)}{\partial r^2} + \frac{2}{r}\frac{\partial n(\epsilon, r)}{\partial r}\right) + I_{e-ph}\left(n(\epsilon, r), N(\epsilon, r)\right) + I_{e-e}(n(\epsilon, r)).$$
(2.17)

In reality, the true behavior of the system is likely to be somewhere between these two limits, with the initial diffusive behavior acting as a spherical expansion of energy until reaching the upper and lower boundaries of the nanowire. For this reason, we consider both cases with the intent of placing semi-quantitative bounds on the importance of diffusion during this process. The electron-electron collision integral is ignored in this formulation for simplicity.

Using the same NbN material parameters, 1 eV of energy is deposited in the system within a 1 nm hotspot radius as an electron bubble consisting of two electron-hole pairs. The choice of two pairs over one eases the computational burden by limiting the range of energies needed for the simulation, and is not expected to significantly alter the results. Phonon escape to the substrate is ignored because we are interested in the initial timescale of the formation of the phonon bubble which is much smaller than the escape time.

Figure 2.6 shows the fraction of energy contained within a certain radius from the center of detection for the electron and phonon systems when calculated using the two coordinate systems. On the timescale of the creation of the phonon bubble, energy diffuses through the electron channel to a radius of 2–3 nm as it transfers to the phonon system in a timescale on the order of $\tau_0 \cdot 3 \times 10^{-5}$. Beyond this time, the electron system continues to diffuse, but because there is little energy in the electrons at that point, the energy remains localized.

(a)

(b)
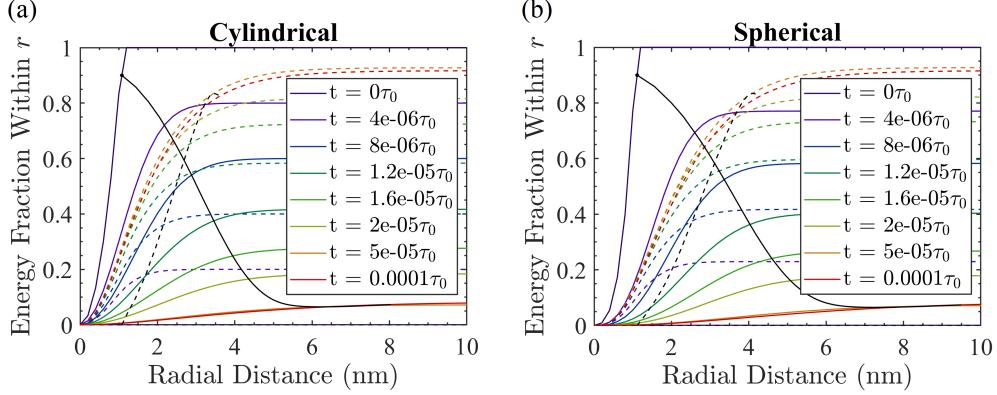
Figure 2.6: Fraction of the photon energy within the radial distance r at various times for (a) cylindrical and (b) spherical coordinates. The solid lines indicate the electron system and the dashed lines indicate the phonon system. The black lines indicate the radial distance where 90% of the electron (solid) and phonon (dashed) energy resides within that radius at a given time. At large values of time, the electron energy is smeared across a wide distance, while the phonon energy remains more localized, as seen by the extension of the solid black line to larger radii at longer times while the dashed black line remains nearly fixed. This is consistent with the analytical results of [73].

The localization of energy can be visualized by setting a threshold level as a fraction of the photon energy and plotting the radius at which the total energy within that radius equals the energy threshold. This is shown in Figure 2.7, and indicates that half of the energy remains within a radius of 2 nm for the spherical case on the timescale of these calculations. For a larger threshold value of 90% of the photon energy, the spherical expansion extends to a radius of 4–5 nm, exceeding the thickness of the film, at which point the cylindrical coordinate solution offers a more realistic estimate of the energy confinement at 3–4 nm.
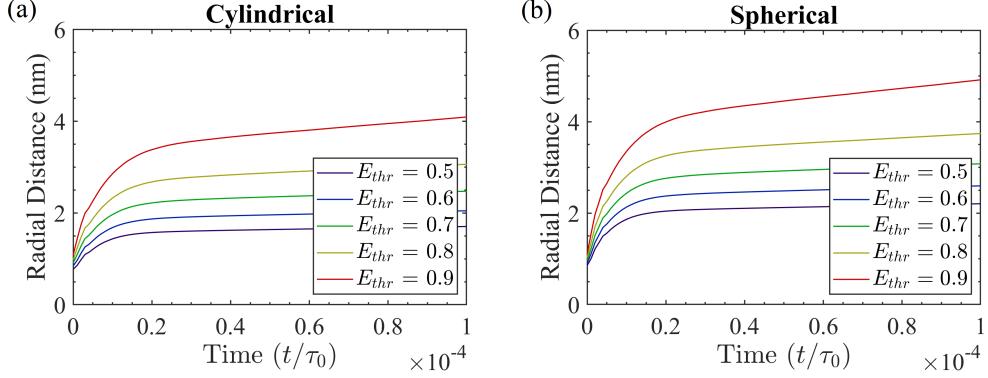
Figure 2.7: Radial distance containing certain fractions of the photon energy as a function of time for (a) cylindrical and (b) spherical coordinates.

A detailed view of the slow expansion of energy is provided in Figures 2.8 which shows the radial distances containing a certain fraction of the photon energy for both the electrons and phonons as a function of time. This shows the initial rapid transfer of energy from the electron system (solid lines) to the phonon system (dashed lines), at which point energy stops diffusing quickly because the phonon diffusion is neglected in this model on account of the low phonon group velocity. The slow decay of phonon energy at longer times is due to its coupling to the electron system where diffusion still occurs rapidly.
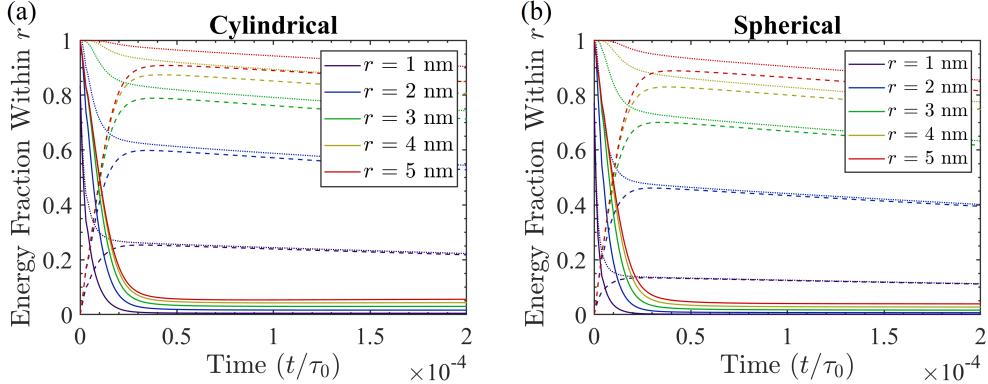
Figure 2.8: Fraction of energy within a given radius as a function of time for (a) cylindrical and (b) spherical coordinates. The solid line shows the electron system, the dashed line is the phonon system, and the dotted line is the combined system. The majority of the energy remains localized to within a few nanometers of the absorption location due to phonon localization.

These calculations neglect electron-electron scattering. With scattering added, the initial electron-hole pair scatters to create multiple excitations which couple to the

lattice more rapidly. This implies that the phonon distribution will be more localized than found in the current calculations. The initial electronic excitations created by photons with higher energy will undergo a greater amount of electron-electron scattering before interacting with the lattice, leading to a photon-energy dependence to the degree of localization of the phonon bubble, but this effect is not expected to be significant.

Overall, the calculations of energy diffusion suggest that the initial hotspot and phonon bubble may be smaller and more localized than the traditional length scale of ξ, which has implications for modeling the suppression of superconductivity following photon absorption. Part of the definition of the initial conditions of the TDGL simulations modeling the suppression of superconductivity is the size of the hotspot where the photon energy is deposited. If the initial hotspot is smaller than suggested by the ξ lengthscale, the energy density will be higher in the initial hotspot, which will alter the evolution of the system. This effect is studied in Chapter 3.

2.3.3 Electron-Electron Interactions

As observed in Section 2.3.1, both the electron and phonon distributions are far from their equilibrium values during the calculated time frame up to $0.002\tau_0 \sim 4\tau_{|\Delta|}$. However, these results neglected electron-electron interactions. The presence of these interactions, particularly at low energies, will help to restore the electron distribution to its equilibrium form. Solving the kinetic equations over the full energy range while including the electron collision integral is inefficient, but given that our goal is to determine the extent to which the collision integral restores the system to its equilibrium distribution at low energies, we can restrict solving the collision integral to energies below the Debye energy. This significantly improves numerical efficiency.

Using the same material parameters described throughout this chapter and a 1 eV electron bubble initial condition consisting of two electron-hole pairs, the downconversion response is modeled for various strengths of electron-electron interaction as defined by the parameter α_{e-e} (2.5). Diffusion is neglected in these calculations. As seen in Figure 2.9, the nominal $\alpha_{e-e} \sim 12.3$ based on the material parameters for NbN does not lead to significant thermalization of the electron system. Instead, the distribution remains far from equilibrium and centered at higher energy excitations. Only by increasing α_{e-e} by a factor of 10–50 does the distribution shift

toward its equilibrium distribution. Even in the case where the electron distribution approaches that of a Fermi distribution, the phonon system remains far from equilibrium. This is an indication that in subsequent modeling, treating the phonon system as being in equilibrium cannot be justified, even though such a simplification reduces computational complexity.
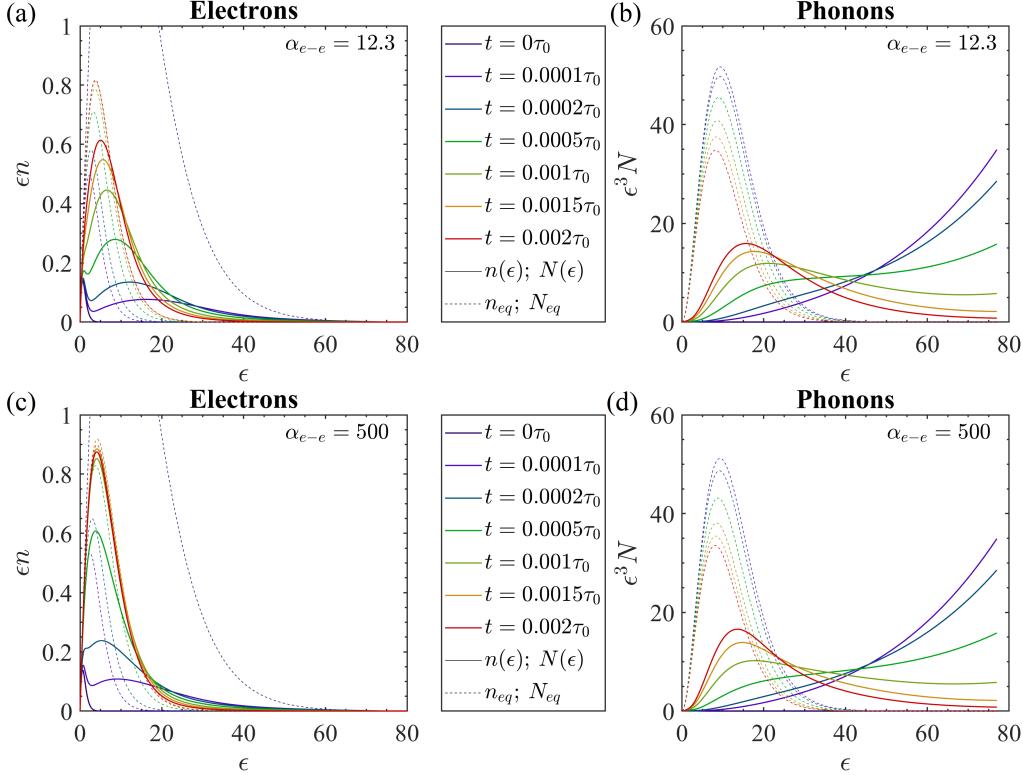
Figure 2.9: Comparison of the energy density distribution functions with the electron-electron collision integral included for (a,b) $\alpha_{e-e} = 12.3$ (corresponding to the value from (2.5)), and (c,d) $\alpha_{e-e} = 500$. Significant enhancement of scattering (α_{e-e} increased by factor >10) is needed for the electron distribution to approach the equilibrium distribution. The electron-electron scattering based on the theoretical estimate α_{e-e} does not significantly shift the distribution functions when compared to the results without scattering shown in Figure 2.4. For all cases, the phonon distribution does not approach its equilibrium form for times shorter than $0.002\tau_0$ as shown.

2.3.4 Phonon Plateau Initial Condition

During the initial stage of downconversion, energy is transferred from the electron system to the phonon system. This occurs on a timescale much faster than the timescale of order parameter variation in the superconductor. It is then useful to consider an approximation to the system where the energy starts in the phonon system

rather than an electron bubble. The rate of mode excitation is equal for all phonon modes during this initial stage of downconversion, so we can approximate the initial state of the system by the phonon plateau, which defines a uniform occupation $N(\epsilon) = N_{Ph.B.}$ for all energies such that the total energy is equal to the photon energy, and adding this uniform occupation to the equilibrium distribution at the substrate temperature. Within the Debye model, this corresponds to the condition $N_{Ph.B.} = \frac{E_\lambda}{V_{HS}E_0}\frac{4\gamma}{\Omega_D^4}$.

Using the same simulation parameters as above while neglecting diffusion and electron-electron interactions, we compute the response of the system for the phonon plateau and compare it to the results for the single electron-hole pair electron bubble in Figure 2.10. While there is a significant discrepancy between the two models at timescales below $t = \tau_0\, 1 \times 10^{-4}$ due to the initial transfer of energy from electrons to phonons within the electron bubble, for timescales relevant to the suppression of superconductivity, the total energy in the electron and phonon systems is approximately the same for the two initial conditions. This suggests that for the modeling the evolution of the superconducting state, the phonon bubble initial condition is sufficiently close to the true electron bubble at the relevant timescales that the change in initial conditions should not significantly alter the system response.
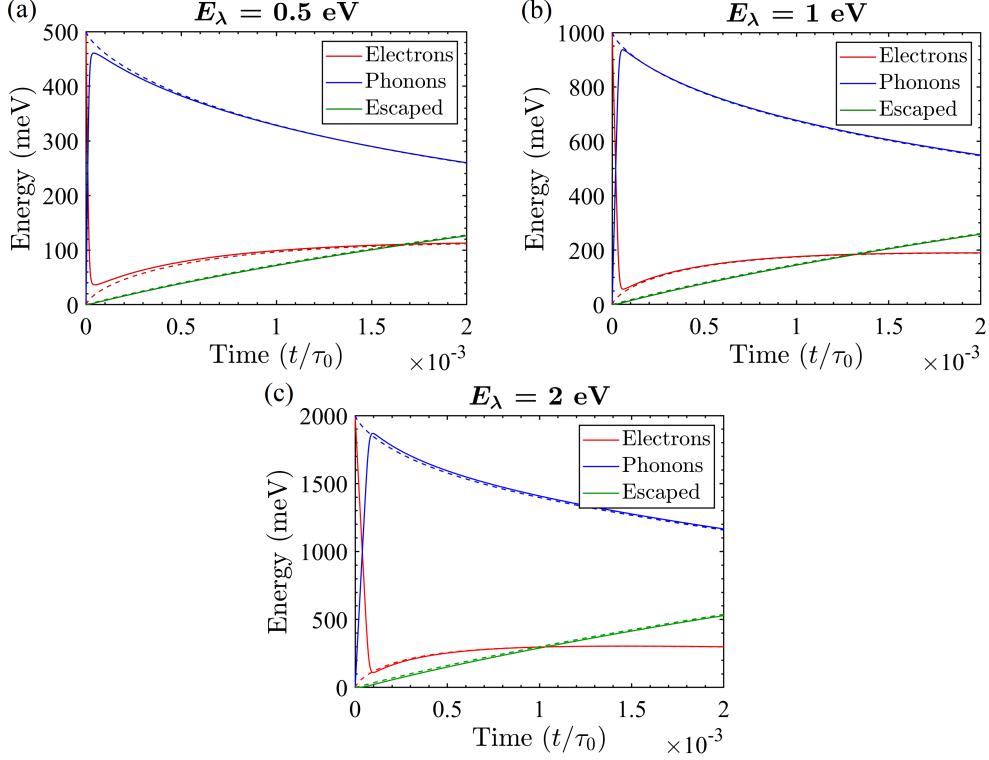
Figure 2.10: Comparison of downconversion calculations using the electron bubble (solid lines) and phonon plateau (dashed lines) initial conditions for photon energies of (a) 0.5 eV, (b) 1 eV, and (c) 2 eV.

While the energy balance within the phonon plateau is well matched to that of the electron bubble, this metric alone does not ensure that the phonon bubble is representative of the electron bubble. To confirm the agreement between the two initial conditions, we must verify that the excitation spectrum resulting from the two distributions is approximately the same. Figure 2.11 shows the number of excitations and average energy of the excitations for the two cases. The results are nearly identical, so we can safely conclude that the within the current kinetic model, the use of the phonon plateau initial condition is sufficient for simulating the behavior of the system on the timescale of $\tau_{|\Delta|}$ as needed for the TDGL formulation.
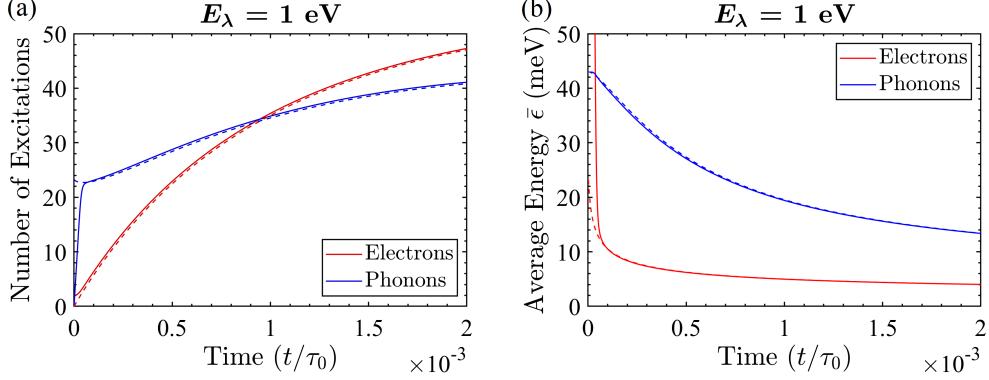
Figure 2.11: Comparison of downconversion calculations using the electron bubble (solid lines) and phonon plateau (dashed lines) initial conditions for a photon energy of 1 eV. The average number of excitations is shown in (a) while the average excitation energy is shown in (b).

As an additional check, we consider the impact of strong electron-electron interactions for the two initial conditions. By setting $\alpha_{e-e} = 500$, we simulate the system and confirm that the response of the system is consistent once the initial energy of the electron-hole pair is transferred to the phonon system. This can be seen in Figure 2.12.
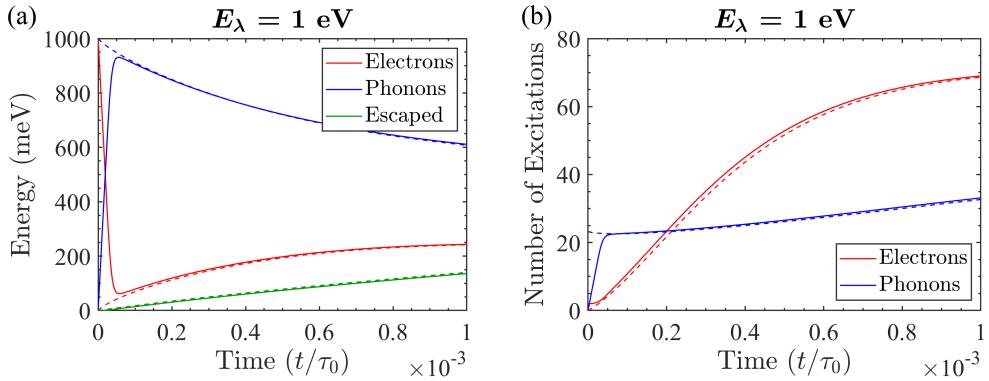
Figure 2.12: Comparison of downconversion calculations using the electron bubble (solid lines) and phonon plateau (dashed lines) initial conditions with strong electron-electron interactions where $\alpha_{e-e} = 500$. The energy partitioning is shown in (a) while the average number of excitations is shown in (b). There is minimal difference between the two results.

Finally, we compare the two initial conditions when diffusion is included. We consider the cylindrical coordinate system over a longer timescale to confirm that the energy partitioning between the electrons and phonons is consistent. A 1 nm hotspot

radius is used for both of the calculations, with the electron-bubble initial condition consisting of three electron-hole pairs. As shown in Figure 2.13, there is reasonable agreement between the two calculations, further supporting our conclusion that the phonon plateau is a suitable initial condition within the kinetic formulation.
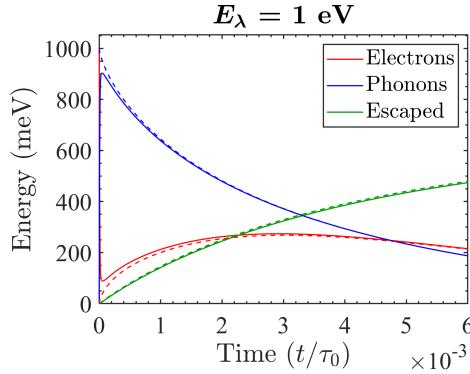
Figure 2.13: Comparison of downconversion calculations using the electron bubble (solid lines) and phonon plateau (dashed lines) initial conditions with diffusion included while using the cylindrical coordinate system. There is a larger difference between the two initial conditions when diffusion is included compared to the fixed hotspot results, but the agreement is still good.

2.3.5 Thermalization Time

As seen in Section 2.3.3, on the short timescales typical of the first stages of downconversion, both the electron and phonon systems remain far from their equilibrium distributions. Strong electron-electron interactions can help to thermalize the electron system, but this requires an enhancement of the interaction strength which is much larger than expected based on the theory of dirty 2D materials [77]. The TDGL formulation for describing the suppression of superconductivity, through its use of the two-temperature model, assumes that the electron and phonon systems are thermalized, being well described by the equilibrium temperatures T_e and T_{ph}. We can estimate the validity of this assumption by calculating the fraction of the energy which falls within the appropriate equilibrium distribution during the downconversion process as shown in Figure 2.14. Without enhanced electron-electron interactions, the time for the non-equilibrium electron distribution to coincide with the equilibrium distribution is several factors of $\tau_{|\Delta|}$. The phonon system is much further from equilibrium. Even after $20\tau_{|\Delta|}$, the phonon distribution remains out of equilibrium and skewed toward high-energy phonons. This is true even with enhanced electron scattering, which is consistent with the qualitative observations of Section 2.3.3. These calculations indicate that the assumption of thermalized

electrons within the TDGL formulation is not strictly valid during the first few picoseconds after photon absorption, and the thermalization assumption is never satisfied for the phonon system. The thermalization process is also relevant for understanding the validity of the hotspot initial conditions for the TDGL model. The hotspot assumes $T_e = T_{ph}$ within the hotspot volume, but as the inset of Figure 2.14 shows, when calculating the effective temperature of the electron and phonon systems based on their energy, it can take several factors of $\tau_{|\Delta|}$ for the electron temperature to reach that of the phonons.
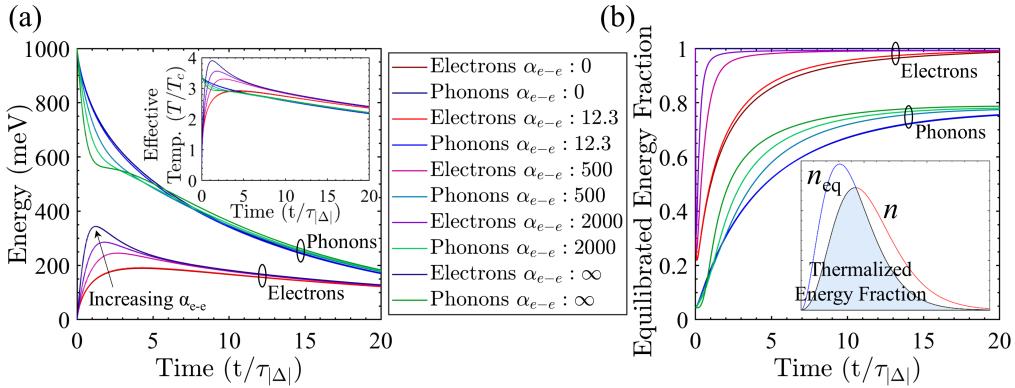
Figure 2.14: Thermalization time during downconversion for various α_{e-e}. The energy partitioning is shown in (a) while (b) indicates the fraction of the energy in the non-equilibrium distribution overlapping the equilibrium distribution having the same total energy. This thermalized energy fraction is the overlap fraction of the energy functions ϵn and $\epsilon n_{eq}(T_e)$ for the electron system and $\epsilon^3 N$ and $\epsilon^3 N_{eq}(T_{ph})$ for the phonon system, as shown schematically in the inset.

2.3.6 Summary

Within the framework of the current kinetic model, we can draw a number of conclusions about the nature of the downconversion process in SNSPDs. First, the downconversion process occurs on a timescale much faster than the order parameter can evolve. This means that the exact dynamics of the initial electron-hole pair and subsequent electron bubble are unlikely to be important in the suppression of the superconducting state of the nanowire. During this rapid downconversion stage, energy is transferred predominantly to the phonon system. Modes in the phonon system interact with the electronic system at an energy-independent rate, leading to equal occupation of all phonon states. Due to the higher density of states at higher phonon energies, the energy within the phonon system is skewed toward the Debye energy.

Based on this result, the second main conclusion is that we are able to use the phonon bubble initial condition in the form of a phonon plateau to faithfully reproduce the dynamics of the system at the timescales relevant to suppression of superconductivity. Within the model, this holds true both with and without enhanced electron-electron coupling. This is an important simplification as it allows the phonon system to be considered as two independent components: the initial phonon bubble and secondary phonons which are generated through interactions with the electronic system after the phonon bubble energy is transferred to the electronic system.

The rapid transfer of energy to the phonon system also ensures that diffusion plays a limited role in the expansion of the initial hotspot. The low group velocity of phonons, on the order of 3–5 nm/ps, means that once energy is transferred to the phonon system, it is effectively localized within a radius smaller than the coherence length, which is typically used to model the hotspot size. At the same time, diffusion in the electron system can be significant, and the energy of the electron system spreads beyond the traditional ξ radius hotspot on a time frame comparable to $\tau_{|\Delta|}$.

Finally, we observe that throughout the downconversion process, the distribution functions are far from their equilibrium values. While strong electron-electron collisions can reduce the deviation from equilibrium in the electronic system, the occupation of the phonon system remains far from that of a Bose-Einstein distribution. This conclusion raises concerns about the pervasive use of the two-temperature model to describe the evolving non-equilibrium system resulting from a strong perturbation, and encourages investigation of a more sophisticated model which accounts for the non-equilibrium behavior of the phonon system.

2.4 One-Temperature Model

In the limit of strong electron-electron coupling, the electron distribution function approaches that of a Fermi distribution at all times t during the evolution of the system. Within the kinetic-equation framework, assuming $n(t) = n_{eq}(T_e(t))$ greatly simplifies the complexity of calculations required to fully define the non-equilibrium phonon distribution function. This simplification even allows the description of the phonon system to extend beyond the formulation of (2.2) to incorporate the angle and z coordinate dependence of phonon occupation number.

A series of recent works by Bezuglyi and Shklovskij [78–80] investigates the response of the non-equilibrium phonon system in the limit of equilibrated electrons

for thin film normal metals. In this limit, the electron system is defined by a time-dependent temperature $T_e(t)$ and a full distribution function describing the phonons. The system response is governed by a kinetic equation defining the phonon system and an integro-differential equation to describe the electron temperature. We use this approach under the appropriate boundary and initial conditions to model the response of the thin film nanowire to photon absorption after the initial generation of a phonon bubble.

2.4.1 Formulation

Under the assumption of equilibrated electrons, Bezuglyj and Shklovskij solve the kinetic equation for phonons within a thin-film spanning $0 < z < d$ where d is the film thickness. When neglecting lateral diffusion, their equation takes the form

$$\frac{\partial N_{\vec{q}}}{\partial t} + c_z \frac{\partial N_{\vec{q}}}{\partial z} = -\frac{N_{\vec{q}} - N_{0,\vec{q}}(T_e(t))}{\tau_{pe}}, \tag{2.18}$$

where $N_{\vec{q}}$ is the phonon distribution, $c_z = c \cos(\theta)$ is the vertical component of the phonon group velocity, $N_{0,\vec{q}}(T_e(t)) = \frac{1}{e^{\epsilon/k_B T_e(t)} - 1}$ is the equilibrium phonon distribution at the electron temperature $T_e(t)$, and τ_{pe} is the phonon-electron relaxation time. Bezuglyj and Shklovskij solve the equation for the case of the electron system being heated by an external source, but we are interested in the case of downconversion just after the creation of the phonon bubble, so we augment their equation by explicitly introducing this phonon bubble. The revised equation becomes

$$\frac{\partial N_{\vec{q}}}{\partial t} + c_z \frac{\partial N_{\vec{q}}}{\partial z} = -\frac{N_{\vec{q}} - N_{0,\vec{q}}(T_e(t))}{\tau_{pe}} + N_{PB}(t=0)d\delta(t)\delta(z-z_0), \tag{2.19}$$

where N_{PB} is the distribution of the phonon bubble [81].

To further simplify our analysis, we assume that all phonons travel with the energy-independent mode-averaged sound velocity and can be represented by the Debye linear dispersion spectrum. This assumption is not valid for high-energy phonon modes, but should be representative of the group velocity for most of the modes excited by low-energy electrons. It is worth noting that in reality, the high-energy phonons of the phonon bubble will not have the same group velocity as the low-energy acoustic modes. Nonetheless, we will see that the structure of the equations permits solving for the decay of the phonon bubble distribution function independently, and through this mechanism, an altered group velocity can be implemented. The typical form of the phonon bubble places all energy in a number of modes located at the Debye energy, leading to $N_{Ph.B.}(t=0) = \frac{2\pi^2\hbar^3 c^3 E_\lambda}{3V_0 \Omega_D^3}\delta(\epsilon - \Omega_D)$, where

E_λ is the photon energy, V_0 is the hotspot volume, and Ω_D is the Debye energy. Alternatively, one can formulate the initial conditions as a phonon plateau, more consistent with the results of the previous downconversion calculations, leading to $N_{Ph.P.}(t=0) = \frac{2\pi^2 \hbar^3 c^3 E_\lambda}{3V_0}$.

Following Bezuglyj and Shklovskij, the kinetic equation is solved by taking a Fourier transform with respect to time, leading to

$$
N_{\vec{q}}(z, \omega) = C_{\vec{q}}(\omega) \exp\left(-\frac{\nu - i\omega}{c_z} z\right) + \frac{\nu N_{0,\vec{q}}(\omega)}{\nu - i\omega}
$$
$$
+ \frac{N_{Ph.B.}(t=0)d}{4\pi c_z} \exp\left(-\frac{\nu - i\omega}{c_z}(z - z_0)\right) \left(\Theta(z - z_0) - \Theta(z_0 - z)\right)
$$

(2.20)

where $\nu = \nu(\epsilon) = \tau_{pe}^{-1}(\epsilon)$ is the phonon-electron collision rate and the function $C_{\vec{q}}(\omega)$ must be found through the use of appropriate boundary conditions at the film interfaces. It becomes convenient to denote two distinct vectors $\vec{q} = (q_x, q_y, q_z)$ and $\vec{q}' = (q_x, q_y, -q_z)$ to indicate the upward and downward traveling phonon groups, while restricting $q_z > 0$. There are two natural choices for the interface boundary conditions. For films with an interface with the substrate at $z = 0$ but open to vacuum at $z = d$, Bezuglyj and Shklovskij use the boundary conditions

$$
N_{\vec{q}}(d, \omega) = N_{\vec{q}'}(d, \omega) \tag{2.21}
$$

at the vacuum interface and

$$
N_{\vec{q}}(0, \omega) = \alpha(\theta)\delta(\omega)N_{0,\vec{q}}(T_{sub}) + \beta(\theta)N_{\vec{q}'}(0, \omega) \tag{2.22}
$$

at the substrate interface. The first condition encodes specular reflection at the vacuum interface with phonons incident with a vertical velocity component q_z reflecting with a component $-q_z$. The second encodes the coupling of phonons from the film with the substrate, where $\alpha(\theta)$ is the angle-dependent transmission probability and $\beta(\theta)$ is the reflection probability, with $\alpha(\theta) + \beta(\theta) = 1$ for all θ.[4]

Solving for $C_{\vec{q}}(\omega)$ and $C_{\vec{q}'}(\omega)$ using (2.20)-(2.22), we arrive at

$$
C_{\vec{q}}(\omega) = \alpha \frac{\delta(\omega)N_{0,\vec{q}}(T_{sub}) - \frac{\nu N_{0,\vec{q}}(\omega)}{\nu - i\omega}}{1 - \beta \exp(-2ad)}
$$
$$
+ \Lambda \frac{\exp(az_0) + \beta \exp(-az_0) + 2\beta \exp(-ad)\cosh(a(d - z_0))}{1 - \beta \exp(-2ad)}
$$

(2.23)

---

[4]We have again made use of a single average phonon mode with this choice of boundary conditions. A more rigorous treatment would include mixing of the three acoustic branches of the phonon dispersion relation, and could be fully treated with the acoustic mismatch model. However, we do not expect the simplified treatment to significantly alter the conclusions of this analysis.

$$C_{\vec{q}'}(\omega) = \exp(-2ad)\left[\alpha\frac{\delta(\omega)N_{0,\vec{q}}(T_{sub}) - \frac{\nu N_{0,\vec{q}}(\omega)}{\nu - i\omega}}{1 - \beta\exp(-2ad)}\right.$$

$$\left. + \Lambda\frac{\exp(az_0) + \beta\exp(-az_0) + 2\exp(ad)\cosh(a(d-z_0))}{1 - \beta\exp(-2ad)}\right] \quad (2.24)$$

where $a = \frac{\nu-i\omega}{|c_z|}$ and $\Lambda = \frac{N_{Ph.B.}(t=0)d}{4\pi|c_z|}$. Finally, by taking the inverse Fourier transform, we arrive at the final equations for the phonon distribution

$$N_{\vec{q}}(z, z_0, t) = N_{\vec{q}}^{BS}(z, t) + N_{\vec{q}}^{Ph.B.}(z, z_0, t) \quad (2.25)$$

and

$$N_{\vec{q}'}(z, z_0, t) = N_{\vec{q}'}^{BS}(z, t) + N_{\vec{q}'}^{Ph.B.}(z, z_0, t) \quad (2.26)$$

where

$$N_{\vec{q}}^{BS}(z, t) = \frac{\alpha N_{0,\vec{q}}(T_{sub})\exp(-\nu z/|c_z|)}{1 - \beta\exp(-2\nu d/|c_z|)}$$

$$+ \nu\int_{-\infty}^{t}dt' N_{0,\vec{q}}(T_e(t'))\exp(-\nu(t-t'))\beta^{\lfloor\tau(t-t',z)+1\rfloor} \quad (2.27)$$

and

$$N_{\vec{q}'}^{BS}(z, t) = \frac{\alpha N_{0,\vec{q}}(T_{sub})\exp(-\nu(2d-z)/|c_z|)}{1 - \beta\exp(-2\nu d/|c_z|)}$$

$$+ \nu\int_{-\infty}^{t}dt' N_{0,\vec{q}}(T_e(t'))\exp(-\nu(t-t'))\beta^{\lfloor\tau(t-t',-z)\rfloor} \quad (2.28)$$

are the distribution functions obtained by Bezuglyj and Shklovskij with $\tau(t, z) = \frac{|c_z|t}{2d} - \frac{z}{2d}$ and $\lfloor x \rfloor$ indicating the floor function. The phonon bubble initial conditions modify the result from [80] by introducing the terms

$$N_{\vec{q}}^{Ph.B.}(z, z_0, t) =$$

$$4\pi\Theta(t)\Lambda\exp(-\nu t)\sum_{n=0}^{\infty}\beta^n\left[\delta\left(t - \frac{z - z_0 + 2dn}{|c_z|}\right) + \beta\delta\left(t - \frac{z + z_0 + 2dn}{|c_z|}\right)\right] \quad (2.29)$$

and

$$N_{\vec{q}'}^{Ph.B.}(z, z_0, t) =$$

$$4\pi\Theta(t)\Lambda\exp(-\nu t)\sum_{n=0}^{\infty}\beta^n\left[\delta\left(t - \frac{z_0 - z + 2dn}{|c_z|}\right) + \delta\left(t - \frac{2d - z - z_0 + 2dn}{|c_z|}\right)\right]. \quad (2.30)$$

There is a clear intuition about the nature of these terms. After time $t = 0$, as signaled by $\Theta(t)$, the initial distribution $N_{Ph.B.}$ decays through interactions with the

electrons at a rate proportional to the energy-dependent phonon-electron collision rate $v(\epsilon)$. Simultaneously, the sum encodes the cumulative result of n interactions of the phonons with the interface and the delta functions limit the phonons to time and coordinate combinations consistent with the initial location z_0 and propagation at constant speed $|c_z|$.

It is also common for SNSPDs to be embedded in an optical cavity. In this case, phonons can escape to the substrate at both the top and bottom interfaces. The boundary conditions become

$$N_{\vec{q}}(0,\omega) = \alpha_1(\theta)\delta(\omega)N_{0,\vec{q}}(T_{sub}) + \beta_1(\theta)N_{\vec{q}'}(0,\omega) \tag{2.31}$$

$$N_{\vec{q}'}(d,\omega) = \alpha_2(\theta)\delta(\omega)N_{0,\vec{q}'}(T_{sub}) + \beta_2(\theta)N_{\vec{q}}(d,\omega) \tag{2.32}$$

where the subscripts of the transmission coefficients indicate that the surfaces above and below the nanowire do not need to be the same material and could have different transmission properties. In this situation, solving (2.20) with the new boundary conditions leads to

$$C_{\vec{q}}^*(\omega) = \frac{1}{(1 - \beta_1\beta_2\exp(-2ad))}\Bigg[$$

$$(\alpha_1 + \alpha_2\beta_1\exp(-ad))\left(\delta(\omega)N_{0,\vec{q}}(T_{sub}) - \frac{vN_{0,\vec{q}}(\omega)}{v - i\omega}\right)$$

$$+ \Lambda\Bigg((\exp(az_0) + \beta_1\exp(-az_0))$$

$$+ \beta_1\exp(-ad)\left[\beta_2\exp(-a(d - z_0)) + \exp(a(d - z_0))\right]\Bigg)\Bigg]$$

$$\tag{2.33}$$

and

$$C_{\vec{q}'}^*(\omega) = \frac{\exp(-2ad)}{(1 - \beta_1\beta_2\exp(-2ad))}\Bigg[$$

$$(\alpha_1\beta_2 + \alpha_2\exp(ad))\left(\delta(\omega)N_{0,\vec{q}}(T_{sub}) - \frac{vN_{0,\vec{q}}(\omega)}{v - i\omega}\right)$$

$$+ \Lambda\Bigg(\beta_2(\exp(az_0) + \beta_1\exp(-az_0))$$

$$+ \exp(ad)\left[\beta_2\exp(-a(d - z_0)) + \exp(a(d - z_0))\right]\Bigg)\Bigg]$$

$$\tag{2.34}$$

where the $*$ simply denotes the use of the alternate boundary conditions. Under the transformation $\alpha_2 \to 0$ and $\beta_2 \to 1$, these equations recover $C_{\vec{q}}^*(\omega) \to C_{\vec{q}}(\omega)$ and

$C^*_{\vec{q'}}(\omega) \rightarrow C_{\vec{q'}}(\omega)$. Taking the inverse Fourier transform results in equations with the same structure

$$N^*_{\vec{q}}(z, z_0, t) = N^{BS*}_{\vec{q}}(z, t) + N^{Ph.B.*}_{\vec{q}}(z, z_0, t) \tag{2.35}$$

and

$$N^*_{\vec{q'}}(z, z_0, t) = N^{BS*}_{\vec{q'}}(z, t) + N^{Ph.B.*}_{\vec{q'}}(z, z_0, t) \tag{2.36}$$

where

$$N^{BS*}_{\vec{q}}(z, t) = \frac{(\alpha_1 \exp(-\nu z/|c_z|) + \alpha_2 \beta_1 \exp(-\nu(z + d)/|c_z|)) N_{0,\vec{q}}(T_{sub})}{(1 - \beta_1 \beta_2 \exp(-2\nu d/|c_z|))}$$
$$+\nu \int_{-\infty}^{t} dt' N_{0,\vec{q}}(T_e(t')) \exp(-\nu(t - t')) \beta_1^{\lfloor \tau(t-t',z)+1 \rfloor} \beta_2^{\lfloor \tau(t-t',z+d)+1 \rfloor} \tag{2.37}$$

and

$$N^{BS*}_{\vec{q'}}(z, t) = \frac{(\alpha_1 \beta_2 \exp(-\nu(2d - z)/|c_z|) + \alpha_2 \exp(-\nu(d - z)/|c_z|)) N_{0,\vec{q}}(T_{sub})}{(1 - \beta_1 \beta_2 \exp(-2\nu d/|c_z|))}$$
$$+\nu \int_{-\infty}^{t} dt' N_{0,\vec{q}}(T_e(t')) \exp(-\nu(t - t')) \beta_1^{\lfloor \tau(t-t',-z) \rfloor} \beta_2^{\lfloor \tau(t-t',d-z) \rfloor} \tag{2.38}$$

are the modified forms of the Bezuglyj and Shklovskij component. The modified phonon bubble terms are

$$N^{Ph.B.*}_{\vec{q}}(z, z_0, t) =$$
$$4\pi\Theta(t)\Lambda \exp(-\nu t) \sum_{n=0}^{\infty} \beta_1^n \beta_2^n \left[\delta\left(t - \frac{z - z_0 + 2dn}{|c_z|}\right) + \beta_1 \delta\left(t - \frac{z + z_0 + 2dn}{|c_z|}\right) \right] \tag{2.39}$$

and

$$N^{Ph.B.*}_{\vec{q'}}(z, z_0, t) =$$
$$4\pi\Theta(t)\Lambda \exp(-\nu t) \sum_{n=0}^{\infty} \beta_1^n \beta_2^n \left[\delta\left(t - \frac{z_0 - z + 2dn}{|c_z|}\right) + \beta_2 \delta\left(t - \frac{2d - z - z_0 + 2dn}{|c_z|}\right) \right]. \tag{2.40}$$

Energy flows between the electron and phonon systems, so the time-dependent electron temperature required to calculate the phonon distribution depends on the phonon dynamics. Therefore, the equations for the phonon distribution must be solved simultaneously for the energy balance equation of the electron system, leading to an integro-differential equation. The energy within the phonon system is the

integral over volume and energy of the phonon distribution function weighted by the density of states and the energy of each state. This takes the form

$$
\begin{aligned}
E_{ph}(z_0, t) &= \frac{3V_0}{4\pi^2\hbar^3 c^3} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} \int_0^{\Omega_D} d\epsilon\, \epsilon^3 \left[N_{\vec{q}}(z, z_0, t) + N_{\vec{q}'}(z, z_0, t) \right] \\
&= \frac{V_0 E_0}{2\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} \int_0^{\Omega_D} d\epsilon\, \epsilon^3 \left[N_{\vec{q}}(z, z_0, t) + N_{\vec{q}'}(z, z_0, t) \right].
\end{aligned}
\tag{2.41}
$$

The electron system is in equilibrium, so the electron energy maintains its equilibrium form, given by

$$
E_e = \frac{\pi^2}{12} E_0 V_0 \left(\frac{T_e}{T_c} \right)^2,
\tag{2.42}
$$

where we have again assumed that the superconductor behaves as a normal metal.

The energy flux between the electrons and phonons is given by the integral

$$
\begin{aligned}
\frac{\partial E_{e-ph}}{\partial t} &= \frac{V_0 E_0}{2\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} \int_0^{\Omega_D} d\epsilon \\
&\quad \epsilon^3 v \left[N_{\vec{q}}(z, z_0, t) + N_{\vec{q}'}(z, z_0, t) - 2N_{0,\vec{q}}(T_e(t)) \right].
\end{aligned}
\tag{2.43}
$$

The phonon flux between with the interface(s) is given by the difference in the upward and downward propagating phonon groups at the interface. When formulated as the energy flux of the phonon system, this takes the form

$$
\begin{aligned}
Q_{ph,1}(z_0, t) &= \frac{V_0 E_0}{2d\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} |c_z| \int_0^{\Omega_D} d\epsilon \\
&\quad \epsilon^3 \left[N_{\vec{q}}(0, z_0, t) - N_{\vec{q}'}(0, z_0, t) \right]
\end{aligned}
\tag{2.44}
$$

for the lower interface and

$$
\begin{aligned}
Q_{ph,2}(z_0, t) &= -\frac{V_0 E_0}{2d\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} |c_z| \int_0^{\Omega_D} d\epsilon \\
&\quad \epsilon^3 \left[N_{\vec{q}}(d, z_0, t) - N_{\vec{q}'}(d, z_0, t) \right]
\end{aligned}
\tag{2.45}
$$

for the upper interface. Combining these elements, we arrive at an integro-differential equation for the electron energy balance equation

$$
\begin{aligned}
\frac{\partial E_e}{\partial t} &= \frac{\pi^2}{6} E_0 V_0 \left(\frac{T_e}{T_c^2} \right) \frac{\partial T_e}{\partial t} = \frac{V_0 E_0}{2\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} \int_0^{\Omega_D} d\epsilon \\
&\quad \epsilon^3 v \left[N_{\vec{q}}(z, z_0, t) + N_{\vec{q}'}(z, z_0, t) - 2N_{0,\vec{q}}(T_e(t)) \right]
\end{aligned}
\tag{2.46}
$$

and an equation describing the rate of energy change of the phonon system

$$\frac{\partial E_{ph}}{\partial t} = \left(Q_{ph,1} + Q_{ph,2}\right) - \frac{V_0 E_0}{2\gamma} \int_0^d \frac{dz}{d} \int_0^c \frac{d|c_z|}{c} \int_0^{\Omega_D} d\epsilon$$
$$\epsilon^3 \nu \left[N_{\vec{q}}(z, z_0, t) + N_{\vec{q}'}(z, z_0, t) - 2N_{0,\vec{q}}(T_e(t)) \right].$$
$$(2.47)$$

These equations are solved within a fixed hotspot volume for a given set of phonon bubble or phonon plateau initial conditions in order to model the response of the system.

2.4.2 Downconversion Comparison

We are interested in the behavior of the superconducting system in the immediate time following the absorption of a phonon. The assumptions of the one-temperature model rule out the possibility of simulating the entire downconversion process because the initial electron state will not be in equilibrium. However, the use of a phonon bubble initial condition permits treating the low-energy excitations of the electron system as being equilibrated, as occurs in the presence of strong electron-electron scattering. We aim to compare the downconversion results calculated using this formulation to those calculated using the kinetic model for the timespan immediately following the excitation of the phonon bubble.

For a direct comparison, we use the same simulation parameters as used throughout this chapter for NbN films fabricated at MIT. The hotspot radius remains $\xi = 6.2$ nm. For simplicity, we assume a phonon bubble initial condition rather than the phonon plateau, but center the energy at the mean phonon energy within the hotspot plateau, given by $\frac{3}{4}\Omega_D$. We consider the case of the nanowire in contact with a single escape interface of SiO_2 and use the acoustic mismatch model (AMM) to describe the mode-dependent transmission probability of the phonons. These transmission probabilities are shown in Figure 2.15a for the NbN-SiO_2 interface . Details of the AMM calculation process are described in the appendix of Kaplan [82]. The main feature to note based on the AMM results is that NbN is reasonably well acoustically matched to SiO_2. The transmission probability remains high for all incident angles and no region of total internal reflection occurs. As a comparison, the AMM results for the acoustically softer metal WSi on SiO_2 are shown in Figure 2.15b. Above the critical angle of $\sim 47°$, transverse phonons are reflected and trapped within the WSi. Using these material parameters for NbN, we arrive at an escape time $\tau_{esc} = 9.4$ ps, assuming a single interface between the NbN and SiO_2.
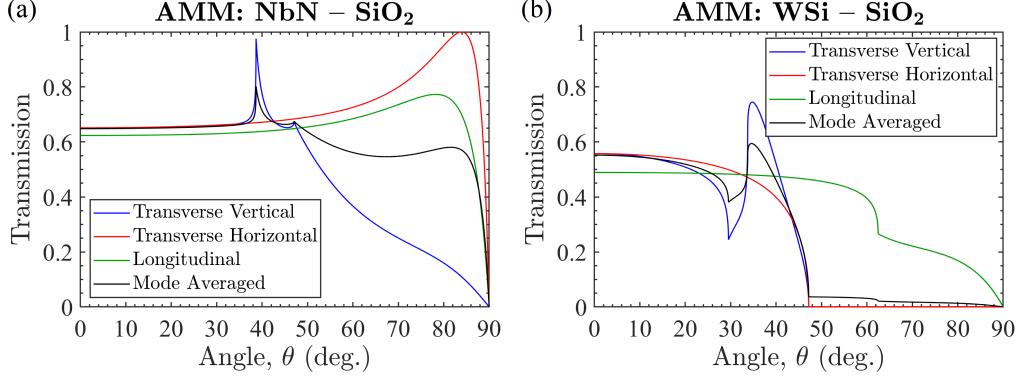
Figure 2.15: Phonon transmission probabilities for (a) NbN on SiO_2 and (b) WSi on SiO_2 within the framework of the AMM. The parameters for the calculations are NbN: $v_t = 4459\,\text{m/s}$, $v_l = 7137\,\text{m/s}$, $\varrho = 8.3\,\text{g/cm}^3$; WSi: $v_t = 3000\,\text{m/s}$, $v_l = 5400\,\text{m/s}$, $\varrho = 15.8\,\text{g/cm}^3$; SiO_2: $v_t = 4090\,\text{m/s}$, $v_l = 6090\,\text{m/s}$, $\varrho = 2.33\,\text{g/cm}^3$. In this notation, v_t is the transverse sound velocity, v_l is the longitudinal sound velocity, and ϱ is the density. The angle is measured from the direction normal to the interface.

The downconversion results are calculated using these parameters of NbN for various initial absorption coordinates z_0 and compared to the model of Section 2.2 as shown in Figure 2.16. For these calculations, we show the timescale in terms of the characteristic order parameter timescale $\tau_{|\Delta|}$ in order to emphasize that we are now interested in the timescales relevant to suppression of superconductivity. There are several features to notice from the new model. First, in the absence of enhanced electron-electron interactions, the one-temperature model deviates significantly from the kinetic model. More energy is initially transferred from the phonon bubble to the electronic system in the one-temperature model compared to the kinetic model. At later times, the difference between the two models shrinks. The second feature is that within the one-temperature model, the response of the system depends on the initial coordinate of absorption. For photons absorbed further from the interface, more energy is retained throughout the downconversion process. The implications of this result will be discussed in more detail in the following sections. Finally, when enhanced electron collisions are included in the kinetic model, the response shifts toward that of the one-temperature model. This is expected because the one-temperature model is formally accurate only in the limit of equilibrated electrons, as occurs for strong electron-electron collisions. Even in the limit of strong electron-electron interactions, the results of the one-temperature model do not coincide with those of the kinetic model due to the different phonon formula-

tions. The use of the idealized phonon bubble centered at $\frac{3}{4}\Omega_D$ rather than a proper phonon plateau also contributes to this discrepancy.
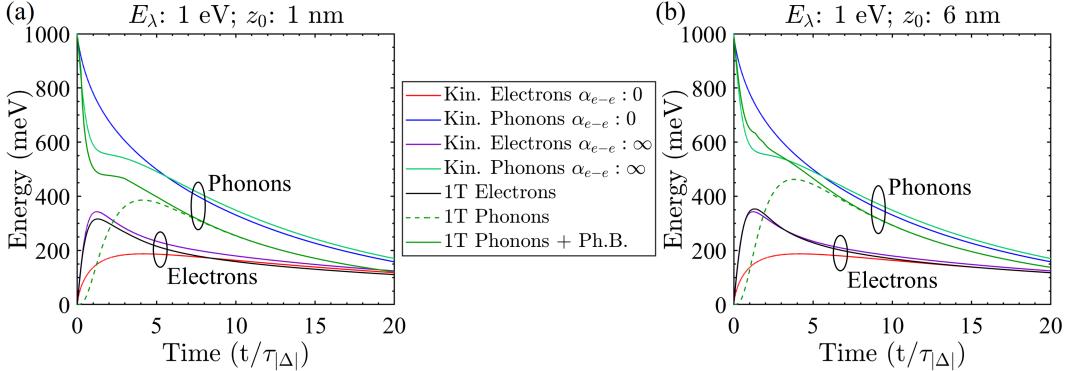
Figure 2.16: Comparison of the one-temperature and kinetic models of downconversion. Results are shown for absorption coordinates z_0 of (a) 1 nm and (b) 6 nm. As the absorption coordinate increases away from the interface ($z = 0$), more energy is retained within the one-temperature model. The response within the kinetic model of Section 2.2 does not depend on the initial absorption coordinate. As the strength of electron-electron interaction increases (increasing α_{e-e}), the kinetic model approaches the one-temperature model. The solid green lines show combined energy of the phonon bubble and phonon distribution within the one-temperature model while the dashed green lines indicates the phonon energy without the phonon bubble.

2.4.3 Escape of High-Energy Phonons

One of the strengths of the detailed phonon kinetic equation is that it permits simulating the response of the system under different initial coordinates of absorption z_0 along the thickness of the nanowire. As a result, different amounts of energy from the phonon bubble can escape from the film, which has a direct impact on the amount of energy which can break superconductivity and promote photon detection. There is a simple intuition behind this result. Photons absorbed closer to the nanowire-substrate interface generate a phonon bubble closer to the interface. On average, these phonons are more likely to escape before interacting with low-energy electrons. Within the kinetic model, the fraction of photon energy which is transferred from the first-generation phonons of the phonon bubble to the electronic system is independent of the initial photon energy. This is a simple consequence of the ballistic trajectories of individual phonon modes and the energy-dependent scattering time with the electronics system. Phonon-phonon interactions are neglected, so there is no non-linearity which could result from higher phonon densities.

Using these kinetic equations, we can calculate the expected energy loss due to downconversion for different device configurations. Using the same material properties as throughout this chapter, we simulate the evolution of the phonon bubble to calculate the fraction of energy transferred to the electronic system. The phonon bubble is described by phonons with the mean energy $\bar{\epsilon} = \frac{3}{4}\Omega_D$. The results are shown in Figure 2.17a for the case of a single NbN-SiO$_2$ interface while Figure 2.17b shows the result for two interfaces, corresponding to an embedded nanowire. Not surprisingly, the amount of escaped energy is larger for the absorption locations closer to the interface(s). For the single interface case, as $z_0 \rightarrow 0$, the escaped energy fraction approaches approximately half of the transmission probability, which corresponds to all phonons with negative vertical velocity encountering the interface exactly once before interacting with the electrons. The round-trip time to reflect off the top interface and interact with the substrate is long enough that nearly all of the energy is transferred to the electrons before this can occur.
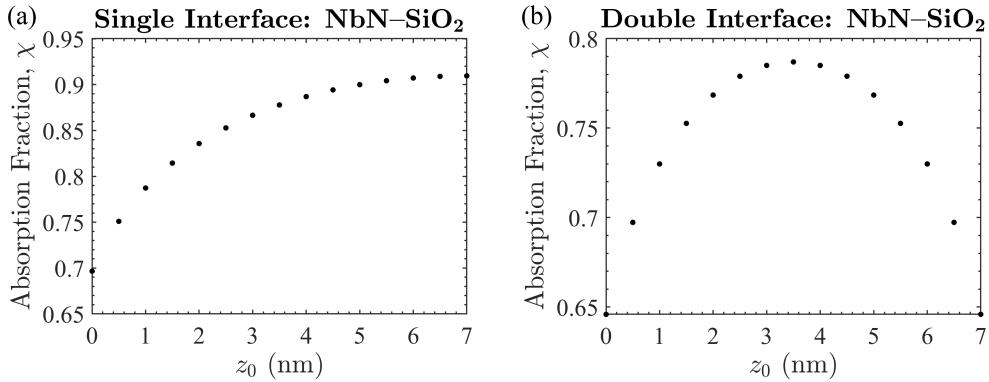
Figure 2.17: Average absorbed energy fraction in a NbN nanowire on SiO$_2$ for a (a) single interface and (b) double interface using the AMM.

We can calculate the average energy absorption $\bar{\chi}$ by integrating over the thickness of the film according to

$$\bar{\chi} = \int_0^d \chi(z)p(z)dz, \tag{2.48}$$

where $p(z)$ is the probability of absorption at the depth z. Estimation of the skin depth based on

$$\delta = \frac{\lambda}{2\pi}k(\lambda) \tag{2.49}$$

yields $\delta \approx 30$ nm for a photon wavelength (λ) of 500 nm, where $k(\lambda)$ is the imaginary part of the nanowire index of refraction, with larger skin depths at longer wavelengths. Based on this calculation, we assume that the probability of absorption

can be approximated as being uniform along the nanowire thickness and $p(z) = \frac{1}{d}$. Under this simplification, calculation yields $\bar{\chi} = 0.856$ for the single interface configuration and $\bar{\chi} = 0.747$ for the double interface configuration. This provides an estimate of the amount of energy which can be assumed to contribute to breaking superconductivity in the subsequent evolution of the non-equilibrium state.

As a comparison, we can perform the same analysis for the WSi material system. In this case, we use the material parameters which are based on films fabricated and characterized by NIST. We use $T_c = 4$ K and $\tau_0 = 5.15$ ns, which comes from averaging the two samples studied in [83, 84]. For the phonon system, we assume an average sound velocity of $v_{avg} = 3340$ m/s [85] and a Debye temperature of 390 K, and $\Omega_D = 97.5$ in units of $k_B T_c$. Using a sheet resistance $\rho_\square$ of $350 \, \Omega/\square$, thickness d of 7 nm, and diffusion coefficient $D = 0.74 \, \text{cm}^2/\text{s}$, we arrive at a single-spin density of states $N(0) = 1.72 \times 10^{22} \, \text{eV}^{-1}\text{cm}^{-3}$ and characteristic energy scale $E_0 = 8.2 \, \mu\text{eV}/\text{nm}^3$. This also leads to the phonon parameter $\gamma = 40.5$. These parameters lead to a zero temperature BCS coherence length $\xi(0) = \sqrt{\frac{\hbar D}{\Delta_0}} = 8.95$ nm, and a coherence length of $\xi = 11.2$ nm at a substrate temperature of $T_{sub} = T_c/2$ [76]. The AMM leads to an estimated τ_{esc} of 32.7 ps for a single interface with SiO_2.

The calculation results are shown in Figure 2.18 for the single and double interface scenarios for WSi on SiO_2. This yields $\bar{\chi} = 0.927$ for the single interface configuration and $\bar{\chi} = 0.873$ for the double interface configuration.
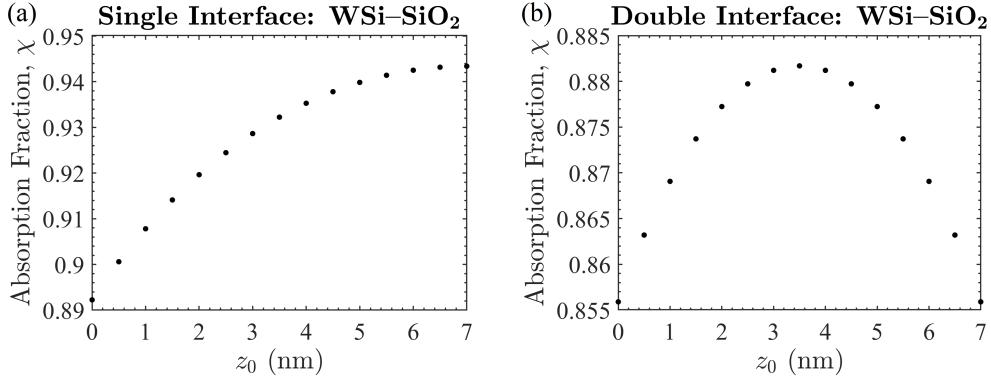
Figure 2.18: Average absorbed energy fraction in a WSi nanowire on SiO_2 for a (a) single interface and (b) double interface using the AMM.

The preceding calculations of energy escape rely on the AMM to determine the transmission of phonons from the superconducting film to the substrate. While this approach generally gives reasonable estimates at low temperatures, its accuracy

can suffer at higher energies where the wavelength of the phonons approaches the size of defects of the interface. Another approach, the diffuse mismatch model (DMM) [86], assumes that phonons scatter independent of incoming state upon reaching the interface. Within the DMM and assuming a Debye phonon spectrum, the transmission coefficient $\alpha_{DMM,1}$ from material 1 to material 2 takes the form

$$\alpha_{DMM,1} = \frac{\sum\limits_{k} c_{2,k}^{-2}}{\sum\limits_{k} c_{1,k}^{-2} + \sum\limits_{k} c_{2,k}^{-2}}, \tag{2.50}$$

where k indicates the phonon mode, and $c_{i,k}$ is the sound velocity of material i in mode k. When applied to the NbN-SiO$_2$ interface, $\alpha_{DMM,NbN} = 0.55$ which is only slightly smaller than the AMM estimate. Conversely, applying the DMM to the WSi-SiO$_2$ system yields $\alpha_{DMM,WSi} = 0.36$, which is larger than the AMM estimate when averaging over angles.

We can estimate the energy loss during the downconversion process within the DMM under the same assumptions and material parameters as above. Contrary to the AMM treatment, within DMM, scattering at the interfaces randomizes phonon state after a scattering event. Within this framework, the average absorption fraction for NbN is $\bar{\chi} = 0.873$ for the single interface configuration and $\bar{\chi} = 0.776$ for the double interface configuration. For WSi, the calculations yield $\bar{\chi} = 0.895$ for a single interface and $\bar{\chi} = 0.811$ for a double interface. These results are shown in Figure 2.19. While the NbN results are quite similar to those obtained with the AMM, much more energy escapes from the WSi film within the DMM compared to the AMM. This is explained by the significant phonon bottlenecking effect which occurs within WSi due to the total internal reflection of high incidence angle phonons within the AMM.
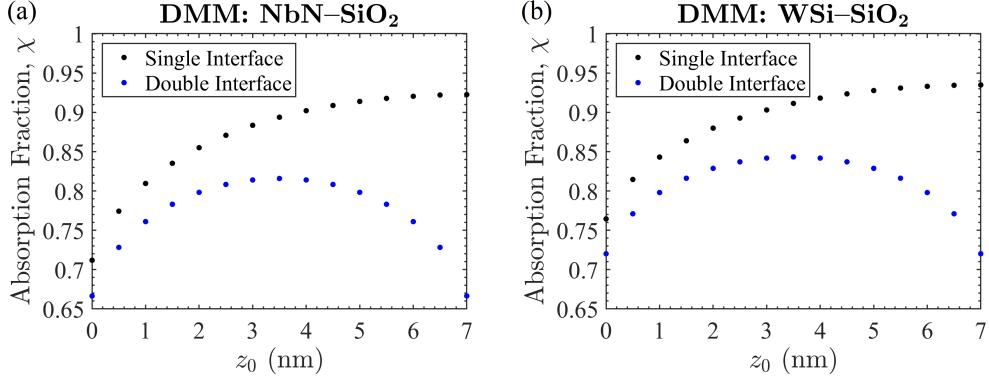
Figure 2.19: Average absorbed energy for (a) NbN and (b) WSi on SiO$_2$ within the framework of the DMM. Black indicates the case of a single interface with the substrate material while blue indicates a double interface.

2.4.4 Phonon Escape Fluctuations

The escape of high-energy phonons from the superconducting film to the substrate is a probabilistic process. Fluctuations during downconversion lead to variations in the total amount of energy retained by the superconductor once the system has thermalized [53, 72]. There are several sources which contribute to these fluctuations. Different numbers of high-energy phonons can be excited during downconversion for a fixed photon energy, and the distribution of energies of these phonons will change. The total amount of energy transferred from the electron-hole pair to the phonon bubble is not a fixed quantity and when fluctuations lead to more energy being transferred to the phonon bubble, more energy is available for escape. Furthermore, as we have just seen, the z_0 coordinate of the absorption also changes the energy retained by the system. Modeling all of the intricacies of the stochastic effects of downconversion is a non-trivial task involving precise knowledge of the interaction strengths of the electron and phonon systems, detailed understanding of the phonon escape process, and knowledge of the true electron density of states and phonon dispersion relations. Such a precise model is not practical for the NbN and WSi materials typically used in SNSPDs due to the large uncertainties in important material properties.

Rather than use a complex model beyond the realm where it can be assumed to be reasonably accurate, we adopt a simple way to estimate the fluctuations during downconversion. We assume that the dominant source of fluctuations is due to shot noise in the phonon escape process. Such an assumption is not unreasonable given that the mean number of phonons present in the phonon bubble is a few tens of

phonons, based on calculations of Section 2.3.1. Within this framework, we can further simplify our analysis by assuming that the phonon bubble consists of N phonons each with an energy $\bar{\epsilon}$. Each phonon has an escape probability of $1 - \bar{\chi}$ which was calculated in Section 2.4.3. Under these assumptions, the number of retained phonons takes the form of a binomial distribution and the variance of the retained energy takes the form

$$\sigma^2 = \bar{\epsilon}^2 \sum_{i=0}^{N} \left[(i - N\bar{\chi})^2 \binom{N}{i} \bar{\chi}^i (1 - \bar{\chi})^{N-i} \right] = \bar{\epsilon}^2 N \bar{\chi} (1 - \bar{\chi}), \qquad (2.51)$$

where the index i corresponds to the number of retained phonons out of the total N. This has the convenient form that $\sigma = \sigma_0 \sqrt{E_\lambda / E_{\lambda,0}}$ where $\sigma_0 = \sqrt{\bar{\epsilon} E_{\lambda,0} \bar{\chi} (1 - \bar{\chi})}$ and $E_{\lambda,0}$ is a reference photon energy. This is the same form as is commonly used to fit experimental data [2, 3, 53]. The parameter σ_0 is shown in Table 2.1 for the various interface models used in Section 2.4.3.

	σ_0 [meV] ($E_{\lambda,0} = 800\,\text{meV}$)			
	NbN		WSi	
	Single	Double	Single	Double
AMM	65.2	80.7	37.0	47.4
DMM	61.8	77.4	43.6	55.7

Table 2.1: Estimated standard deviation of fluctuations due to phonon escape. The parameter σ_0 is calculated for a reference wavelength of 1550 nm.

We can account for the z dependence of the escape probability by expanding the definition of the standard deviation to

$$\sigma_z^2 = \bar{\epsilon}^2 \frac{1}{d} \int_0^d dz \sum_{i=0}^{N} \left[(i - N\bar{\chi})^2 \binom{N}{i} \chi(z)^i (1 - \chi(z))^{N-i} \right] \qquad (2.52)$$

where we have again assumed equal probability of absorption throughout the thickness of the film. Using these two metrics and the results of Section 2.4.3, we can calculate the standard deviation σ for the different material and interface models for various phonon energies. We restrict the photon energy to be an integer multiple of the average phonon energy for simplicity. Figure 2.20 shows the calculated standard deviation according to equations (2.51) as lines and (2.52) as circles. At low photon energies where the phonon number is low, fluctuations are dominated by the binomial contribution and the two results converge. However, at large energies, the absorption depth dependence becomes significant and the total fluctuations are

larger than would be estimated by only considering the contribution due to fluctuations about the mean energy loss. It should be noted that the fluctuations due to the absorption coordinate are likely overestimated because the phonon bubble is assumed to be point-like in the current model. Our downconversion calculations which include the effects of diffusion indicate that the phonon bubble actually has an initial radius of a few nm due to the initial diffusion of electrons during the transfer of energy from the electron-hole pair to the phonon system. This behavior is expected to smear out some of the differences between initial absorption coordinates.
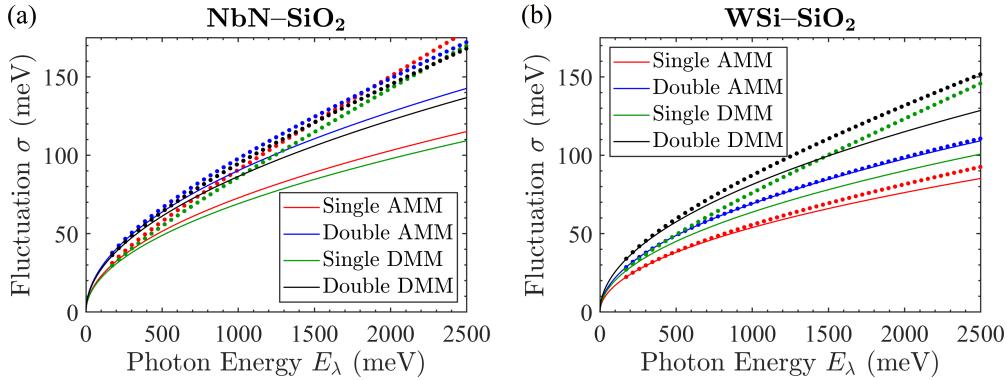
Figure 2.20: Fluctuations during downconversion for (a) NbN and (b) WSi on SiO_2. Circles indicate the calculation including the z contribution according to (2.52) while the lines show the $\sim \sqrt{E_\lambda}$ dependence according to (2.51).

2.4.5 Comparison with Two-Temperature Model

It is common for the response of a superconductor to an external excitation to be modeled within the framework of the two-temperature model. This model is also described in Chapter 4, but to summarize, both the electron and phonon systems are assumed to be well described by their equilibrium distributions according to an electron temperature T_e and phonon temperature T_{ph}. This is the model used in the TDGL simulations of the photon-detection process [6]. In the limiting case of a normal metal without diffusion as we have used throughout this chapter, the two-temperature model reduces to the coupled nonlinear partial differential equations

$$\frac{\pi^2}{6} E_0 \frac{T_e}{T_c^2} \frac{\partial T_e}{\partial t} = -E_0 \frac{24\zeta(5)}{\tau_0} \frac{T_e^5 - T_{ph}^5}{T_c^5} + \frac{1}{V_0} \frac{\partial E_{e-ph,Ph.B.}}{\partial t} \tag{2.53}$$

describing the electronic system and

$$T_{ph}^3 \frac{\partial T_{ph}}{\partial t} = \frac{90\gamma\zeta(5)}{\pi^4 \tau_0 T_c} \left(T_e^5 - T_{ph}^5 \right) - \frac{T_{ph}^4 - T_{sub}^4}{4\tau_{esc}}, \tag{2.54}$$

describing the phonon system where the definitions are the same as throughout the chapter. The third term of the electron equation describes the flow of energy from the phonon bubble to the electron system and takes the form of Equation (2.43) where the phonon distribution function is given by the phonon bubble terms (2.29–2.30) or (2.39–2.40) depending on the choice of boundary conditions. The presence of the phonon bubble is ignored in the phonon energy balance equation, in order to capture the behavior of only the lower energy phonon modes excited by electron recombination in the equilibrated T_{ph} distribution. In this way, the z_0 coordinate dependent losses of the phonon bubble can be accounted for explicitly in the $\frac{\partial E_{e-ph,Ph.B.}}{\partial t}$ term without needing to model the entire phonon distribution. The electron energy has the same form as (2.42)

$$E_e = \frac{\pi^2}{12} E_0 V_0 \left(\frac{T_e}{T_c}\right)^2 , \qquad (2.55)$$

while the phonon energy, excluding that of the phonon bubble, has the form

$$E_{ph} = \frac{B_{ph}}{4} V_0 T_{ph}^4 = \frac{\pi^4}{15} \frac{E_0 V_0}{\gamma} \left(\frac{T_{ph}}{T_c}\right)^4 . \qquad (2.56)$$

While assuming the phonon system is equilibrated is certainly not accurate based on the earlier results of this chapter, the two-temperature model substantially reduces the complexity of the system. Only two parameters: T_e and T_{ph} are required to fully define the state of the system, as opposed to the full distribution functions n and N for the kinetic model of Section 2.2 or the full time history of the electron temperature $T_e(t)$ as in the one-temperature model. Calculations of downconversion using the full kinetic or one-temperature models are tractable due to their limited spatial dimensionality, but extending them to a full 2D simulation is not practical. Therefore, it is useful to evaluate how accurate the two-temperature model is in the limited case without diffusion to determine how well it can be applied to more complex simulations where the description of a full distribution function is not feasible. This also provides intuition about the circumstances under which the model fails.

We compare the response of the two-temperature model to that of the one-temperature model using the same simulation parameters as used throughout the chapter. We assume a single escape interface between the NbN and SiO$_2$ with an escape time $\tau_{esc} = 9.4$ ps which corresponds to the average escape from the AMM. Figure 2.21 shows the energy partitioning between the electron and phonon systems for the two

models. Surprisingly, despite the simplicity of the two-temperature model, the two calculations produce similar results when describing the electronic energy over a range of photon energies and absorption coordinates. In contrast, the phonon energies show significant deviations. This is not surprising given the wide deviation from the equilibrium distribution in the phonon system during downconversion, as seen in Section 2.3.1. Despite the misrepresentation of the total energy, the coupling with the electronic system remains reasonably well approximated. This suggests that for the NbN system, the two-temperature approximation should still capture the main semi-quantitative behavior of the system, even if it will not accurately reproduce the finer details of the system's evolution.
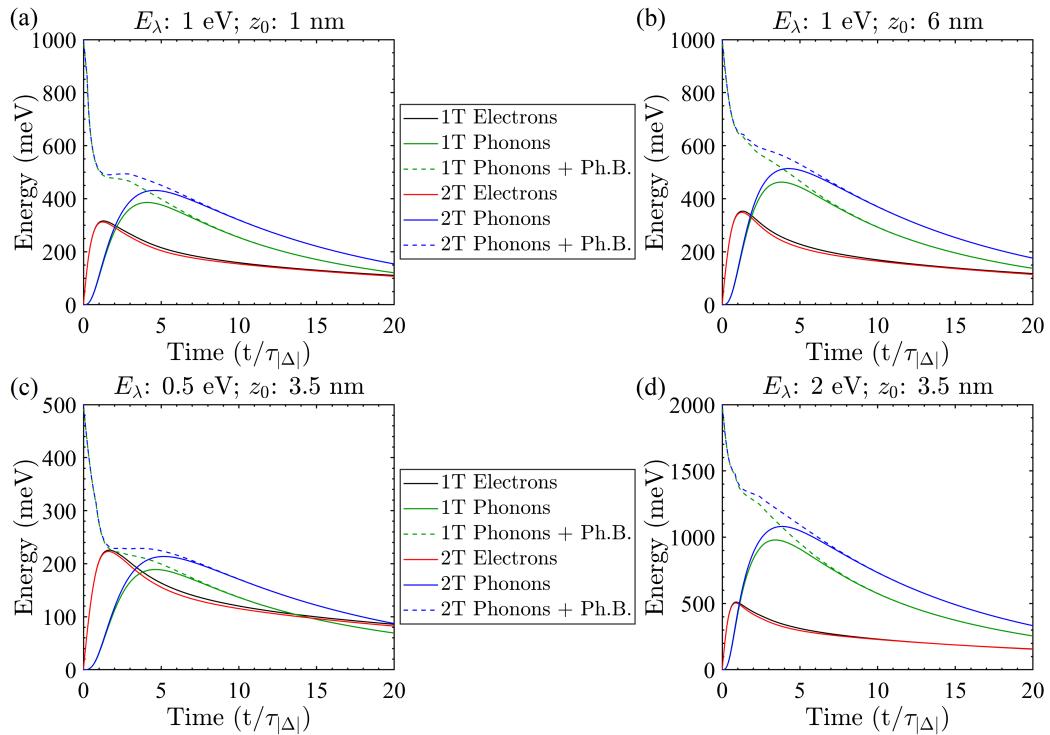
Figure 2.21: Comparison of the one-temperature and two-temperature models of NbN for a 1 eV photon energy at different absorption locations (a) $z_0 = 1$ nm and (b) $z_0 = 6$ nm, and different photon energies of (c) 0.5 eV and (d) 2 eV. The electron energies are comparable, but there are deviations in the phonon energies.

We can similarly check the applicability of the two-temperature model for the WSi material system. Using the material parameters listed above, we compare the downconversion results within the frameworks of the one-temperature and two-temperature models. As shown in Figure 2.22, there is a larger difference between the two models than in the NbN system. This is likely due to the strong phonon

bottlenecking effect which occurs in the thin WSi film. Phonons within the escape cone quickly thermalize with the bath within the one-temperature model, leading to a lower total phonon energy. At the same time, non-escaping modes remain populated, limiting recombination of electron-hole pairs to generate high-energy phonons. This transfer is the primary mechanism for cooling the electron system within this simplified system, so the electron energy remains higher in the one-temperature model compared to the two-temperature model.
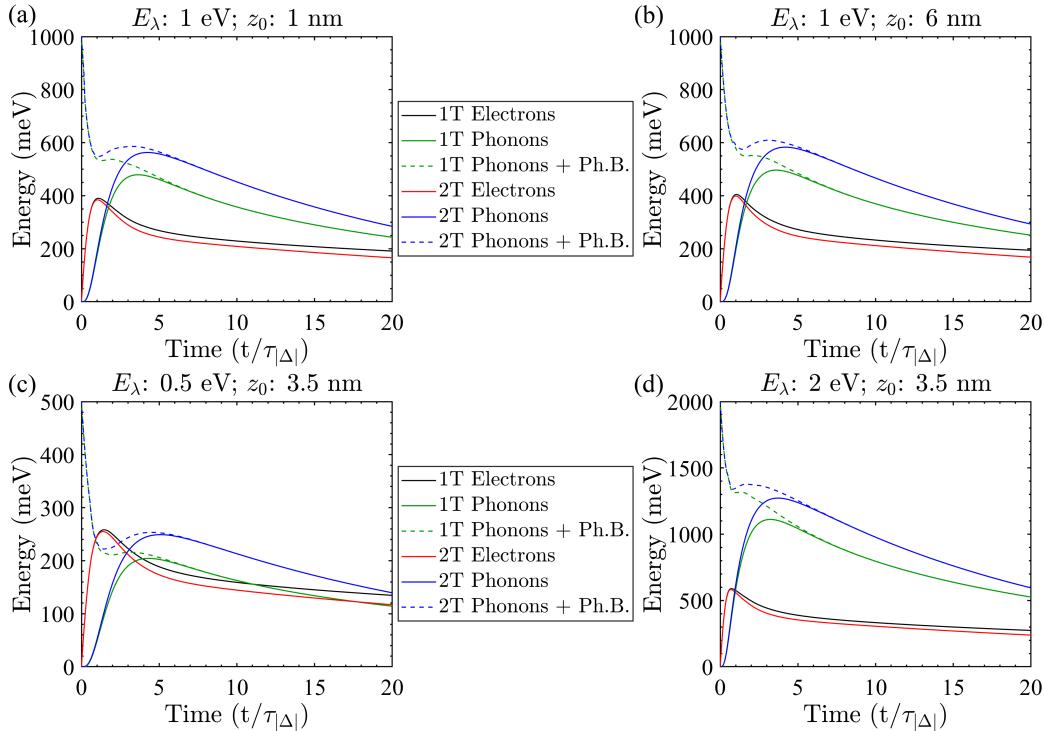
Figure 2.22: Comparison of the one-temperature and two-temperature models of WSi for a 1 eV photon energy at different absorption locations (a) z_0 = 1 nm and (b) z_0 = 6 nm, and different photon energies of (c) 0.5 eV and (d) 2 eV. Unlike in NbN, there are now significant deviations in the electron energy as well as the phonon energy. This is attributed to the strong bottlenecking effect in the material, which leads to a larger deviation in the phonon occupation from an equilibrium distribution. These deviations are larger at lower photon energies.

Due to its simplicity, it is much more convenient to use the two-temperature model rather than the one-temperature model for numerical simulations within the TDGL formulation. Given the discrepancies between two models for WSi, it would be useful to modify the two-temperature model to more accurately reproduce the one-temperature results. Because the primary cause of this discrepancy at long times

originates from the bottlenecking behavior of phonons in the more accurate one-temperature model, the resolution to this problem comes from the application of a three-temperature model [80], which is discussed in more detail in Chapter 4. Bezuglyj and Shklovskij show that in the limit of thin films where the phonon lifetime is much larger than the film thickness, the three-temperature model is able to reproduce the correct long-time cooling of the electron system after heating [80]. As shown in Figure 2.23, the three-temperature model more accurately reproduces the results of the one-temperature model. However, there is still a deviation immediately after the electron system achieves its maximum energy. Due to the different phonon formulations, there is also a large discrepancy in the phonon energies, but this is less important for understanding the dynamics of suppression of superconductivity which we will investigate in Chapter 3.
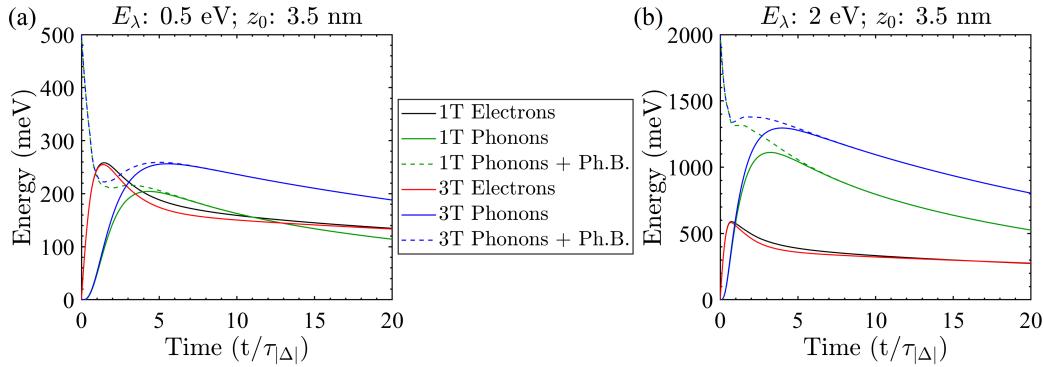
Figure 2.23: Comparison of the one-temperature and three-temperature models of WSi for a (a) 0.5 eV and (b) 2 eV photon energies. The separation of the phonon system into escaping and bottlenecked groups reduces the discrepancy between the two models at longer times, but there is still a deviation in the electron energy immediately after the electron system achieves its maximum energy.

2.4.6 Modified Phonon Bubble

As seen in Section 2.4.2, the one-temperature model is only representative of the non-equilibrium kinetic model of Section 2.2 in the limit of enhanced electron-electron coupling. Without this additional scattering, the energy which is present in the electron system is overestimated by the one-temperature or two-temperature models due to enhanced coupling between the high-energy phonons and the lower energy equilibrated electron distribution. Overestimating the electron energy in the hotspot will directly impact the suppression of superconductivity in subsequent TDGL modeling efforts. Therefore, it is useful to consider how the model of the phonon bubble energy can be modified in order to more accurately reproduce the energy

partitioning between the electron and phonon systems during the downconversion process.

The transfer of energy from the phonon bubble to the electronic system in the one-temperature and two-temperature models is governed by the decay time $\tau_{DC} = \frac{4}{3}\frac{\tau_0}{\gamma\Omega_D} = 1.4\,\mathrm{ps}$, which is based on the scattering time for phonons at the mean energy of the phonon plateau. This simple formulation does not account for the occupation of high-energy electronic states which balance this energy transfer by re-exciting high-energy phonon modes through electron-hole recombination. This energy transfer back to the phonon system occurs when the system is far from equilibrium and the electronic energy is predominately situated at high energies rather than the equilibrium distribution implied by the total energy in the system. By modifying τ_{DC}, we can slow the rate of energy deposition from the high-energy phonon bubble to the electronic system in a simplified manner which crudely approximates the complex dynamics of the non-equilibrium distribution functions. As shown in Figure 2.24, increasing τ_{DC} by a factor of four leads to a more reasonable agreement in the evolution of the energy of the electronic system for both the one-temperature and two-temperature models when compared with the non-equilibrium kinetic model. We will see in Chapter 3 that modifying this rate of energy transfer from the phonon bubble to the electronic system when formulating the initial conditions for the TDGL model as a phonon bubble alters the detection characteristics. Obviously, such a modification of the downconversion process has implications for estimating the loss of high-energy phonons and their fluctuations, but we leave the detailed analysis of this process for future work.
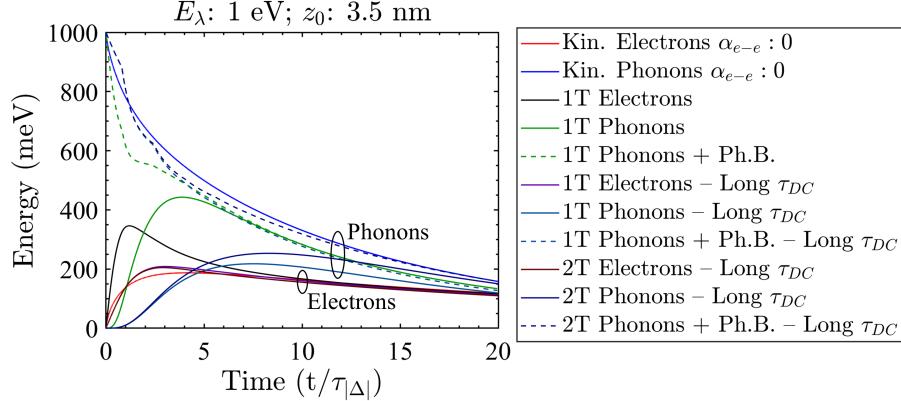
Figure 2.24: Comparison of the one-temperature and two-temperature models to the kinetic model when implementing the modified phonon bubble initial conditions with electron-phonon coupling slowed by a factor of four compared to the standard phonon bubble. The modification improves the agreement between the electron energies for all of the models.

The simple modification performed above improves the agreement of the electron energy between the one or two-temperature models and the non-equilibrium kinetic model, but there are some clear deviations in the evolution at small times. We can confirm that this is not due to our assumption of a phonon bubble centered at $\bar{\epsilon}$ rather than the phonon plateau. If we assume a phonon plateau, and neglect losses at the interface, the energy within the phonon bubble is transformed from

$$E_{Ph.B.} = E_\lambda \exp\left(-\frac{\gamma\bar{\epsilon}t}{\tau_0}\right) \tag{2.57}$$

to

$$E_{Ph.P.} = E_\lambda \frac{4}{\Omega_D^4} \int_0^{\Omega_D} \epsilon^3 \exp\left(-\frac{\gamma\epsilon t}{\tau_0}\right) d\epsilon, \tag{2.58}$$

but this leads to only small differences in the phonon bubble energy and is not responsible for our discrepancy.

2.5 Model Limitations

It must be emphasized that all of the work in this chapter considers a highly idealized model. There are a number of features which limit the quantitative reliability of such a model, and for this reason, the results should be considered as only a semi-quantitative guide for the expected behavior of downconversion in SNSPDs. Given the uncertainty in a number of material properties of the films commonly used in these devices and the difficulty in characterizing all of the relevant parameters of a

single device, theory and modeling can only provide a loose guide to understanding device performance at the current time.

Perhaps the weakest feature of the current model is the use of the Debye phonon spectrum. While accurate at low temperatures, the downconversion process is concerned with high-energy excitations where the linear dispersion relation is known to be inaccurate. Furthermore, the use of the linear dispersion relation implies a fixed group velocity for all phonon modes in the system, while in reality, higher energy modes have lower group velocities. This directly impacts the estimated energy loss of high-energy first-generation phonons. In this light, the estimated energy retained by the system in Section 2.4.3 should be considered only a rough estimate. The use of an appropriate group velocity is expected to reduce the estimated energy loss. At the same time, recent ab-initio calculations by Babu and Guo [87] provide an estimate of the phonon dispersion relations in various phases of NbN and can be used to assess the accuracy of the Debye approximation. By comparing their density of states to the Debye model, one finds that the Debye model actually underestimates a number of lower energy states near the Van Hove singularities. Such states would have longer lifetime than phonons at the Debye energy and could contribute to additional energy loss during downconversion. At the same time, optical modes are neglected in the Debye spectrum, and such modes have low group velocities, making them unlikely to contribute to significant energy loss from the phonon bubble. The treatment of phonon transmission at the film interface could also benefit from a more rigorous definition of the phonon spectrum. The DMM is readily able to extend to the treatment of more realistic phonon dispersion relations for both the film and substrate by using the energy-dependent density of states on both sides of the interface. Such an approach could leverage the calculations of Babu and Guo. While ab-initio calculations provide promising insight, undertaking a detailed analysis using those results cannot be justified given the lingering uncertainty over even basic properties such as electron density of states, diffusion coefficient, and degree of disorder in typical films.

Beyond the phonon system, there are weaknesses in the treatment of the electron-phonon coupling. Only the clean limit is considered here. In reality, dimensionality and the degree of disorder both affect coupling between the two systems. Dirty metals are expected to have electron-phonon coupling rates proportional to $\tau_{ep} \sim 1/T^4$ for strong disorder at high temperatures or $\tau_{ep} \sim 1/T^2$ if scattering is limited by large features such as grain boundaries [88]. As the system dimensionality decreases

from 3D toward 2D at lower temperatures and energies, the strong disorder power law is expected to decrease by another power toward $\tau_{ep} \sim 1/T^3$. There is some experimental evidence that the films may be in this limit due to the measurement of $\tau_{ep}(T) \sim 1/T^{3.21-3.77}$ in a recent experimental work [89]. However, the exact nature of the phonon modes in thin films embedded in dielectrics remains to be investigated, and the assumption of the 2D nature of these modes at low temperatures, as is often quoted in the literature, is likely inaccurate. Babu and Gao [87] also calculate the Eliashberg function in their ab-initio treatment of the NbN material system, and this could be used to refine estimates of electron-phonon coupling in the presence of the modified phonon density of states. While incorporation of a disorder-modified power law will change the quantitative results of these downconversion calculations, the qualitative picture remains the same.

2.6 Conclusions and Future Work

The qualitative analysis performed throughout this chapter offers a compellingly simple picture of the downconversion process in SNSPDs. For typical photon energies used in SNSPDs, after the generation of an electron-hole pair in the superconductor, rapid electron-phonon interactions transfer the majority of the energy to the phonon system in what is known as the phonon bubble (or plateau). These high-energy phonons then transfer energy back to the electronic system while some fraction of energy is lost due to the escape of these phonons to the substrate. Within the framework of the current models, the system can be represented by phonon bubble initial conditions rather than a single electron-hole pair, which greatly simplifies the analysis.

On timescales comparable to the suppression of superconductivity, the one-temperature and two-temperature models provide a reasonable description of the electronic energy in the limit of strong electron-electron interactions. However, the two-temperature model does not adequately describe the phonon system. Without strong electron interactions, only the full kinetic description can be considered valid. This is an important limitation to remember when using these models to describe the energy balance of the superconducting system within the TDGL framework.

The estimation of phonon loss and fluctuations in the phonon bubble provide a guide for what should be considered reasonable fitting parameters within the Fano fluctuation model [53]. The parameters estimated in this chapter show the correct order of magnitude as those used to fit recent experiments [2, 3], which lends credi-

bility to the fitting procedure. The calculations also show that phonon bottlenecking within the AMM can contribute to retaining photon energy and improving SNSPD sensitivity. For NbN, the choice of substrate material plays a role in this process, but it does not appear to be a dominant effect due to the fast transfer of energy from the phonon bubble to the electronic system. Optimizing the choice of substrate material to minimize phonon bubble losses could be expected to show gains only on the order of a few 10s of percent at most given that the phonon loss is only on the order of 20% on SiO_2, which is a reasonably well matched substrate. However, the current model only considers the loss of first-generation phonons and the cumulative loss of energy is the more relevant metric for determining the effect of substrate material on device sensitivity.

There are several of additional topics which are relevant for a complete understanding of downconversion in SNSPDs. Significant consideration must be given to the problem of the electron-electron collision integral, the extent of thermalization in the electron system during this process, and how non-equilibrium effects influence the system when modeling the subsequent suppression of superconductivity. Fluctuations during the downconversion process were calculated using the simplification of equal energy phonons. In reality, there will be a distribution of phonon energies and this distribution will change the characteristics of the resulting escaped energy distribution. Furthermore, a proper treatment would move beyond the Debye model and incorporate proper dispersion relations to describe the phonon density of states, group velocities, and coupling to the electronic system. Despite some limitations, the current model provides a framework within which to model the subsequent stages of photon detection in SNSPDs.

Chapter 3

DETECTION MECHANISM

This chapter contains published work from [3] and [2].

Text from Sections 3.1, 3.3, and 3.4.2 and Figures 3.19 and 3.20 are adapted with permission from: J. Allmaras, A. Kozorezov, B. Korzh, K. Berggren, and M. Shaw, "Intrinsic timing jitter and latency in superconducting nanowire single-photon detectors", Physical Review Applied **11**, 034062 (2019) `10.1103/PhysRevApplied.11.034062`. Copyright © 2019 by the American Physical Society.

Text from Section 3.2 and Figures 3.1a, 3.2, and 3.7 are adapted with permission from: B. Korzh, Q.-Y. Zhao, J. P. Allmaras, S. Frasca, T. M. Autry, E. A. Bersin, A. D. Beyer, R. M. Briggs, B. Bumble, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, A. E. Lita, F. Marsili, G. Moody, C. Peña, E. Ramirez, J. D. Rezac, N. Sinclair, M. J. Stevens, A. E. Velasco, V. B. Verma, E. E. Wollman, S. Xie, D. Zhu, P. D. Hale, M. Spiropulu, K. L. Silverman, R. P. Mirin, S. W. Nam, A. G. Kozorezov, M. D. Shaw, and K. K. Berggren, "Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector", Nature Photonics **14**, 1–6 (2020) `10.1038/s41566-020-0589-x`. Copyright © 2020 Springer Nature. © 2019 Springer Nature.

3.1 Background

Ever since the first demonstration of single-photon detection using a superconducting nanowire [7], significant theoretical efforts have been made to understand the detection mechanism of these new devices. This primarily has focused on understanding the process of breaking superconductivity in a current-carrying nanowire. The earliest of these works focused on what is known as the 'hotspot' model [7, 42, 90]. In this interpretation, the absorption of a photon in a current-carrying nanowire leads to the rapid excitation of quasiparticles, which then thermalize within a confined region known as a hotspot. Due to the excitation of quasiparticles, the material

in the hotspot is assumed to have its superconductivity fully suppressed such that it cannot support any supercurrent. Current is therefore diverted around the hotspot which leads to an increased supercurrent density in the regions around the hotspot, often called the sidewalks. If the bias current is sufficiently high, this redistribution of current will exceed the critical current density of the nanowire and the superconductor switches to the normal state.

While appealing for its qualitative simplicity, an improved model immediately emerged in order to account for discrepancies between model predictions and experimental observations. This first refinement of the model introduced the diffusion of excess quasiparticles in the hotspot to regions around the location of photon absorption, reducing the supercurrent carrying capacity of these surrounding areas [42]. Diffusion has the effect of reducing the detection threshold for a given photon energy, and was necessary to explain the spectral cutoff energy at near-infrared wavelengths in NbN detectors [43].

After the proposal of the original hotspot model, several works attempted to refine the definition of the detection threshold after photon absorption. These models focused on the importance of vortex formation and crossing as the primary mechanism of breaking superconductivity after the formation of an initial hotspot [45, 47]. Earlier works exploited the idea of thermal activation of vortices to explain intrinsic dark counts in SNSPDs, but these works did not attempt to explain photon detection using this concept [44, 91]. Vortex-assisted photodetection was modeled using the Ginzburg–Landau energy functional [46] to describe the energy barrier to vortex entry in the superconductor, providing a new threshold at which superconductivity was broken. Similarly, several works used the solution of the time-dependent Ginzburg–Landau equations to describe the deterministic evolution of the superconducting state while accounting for vortex formation [47, 92, 93]. This also introduced the idea of vortex-antivortex unbinding within the superconductor as a detection mechanism for photons absorbed in the center of the nanowire while previous work only considered vortex entry from the edge of the nanowire.

Based on the work regarding vortex entry, a further iteration of the quasiparticle diffusion model was introduced by Engel et al. [48] which investigated the criterion for photon detection. This work compared the detection thresholds within the quasiparticle diffusion framework when considering the suppression of superconductivity within the hard-core hotspot model, diffusive normal-core model, and vortex entry model. However, due to simplifications in the model, the current redistribution

due to the presence of quasiparticles was not solved exactly, and vortex entry was only considered from the edge of the nanowire, ignoring the possibility of vortex-antivortex unbinding. Furthermore, the equations describing quasiparticle diffusion ignored important details or the energy exchange within the superconductor, making them more phenomenological than quantitative.

The most recent advance in modeling was proposed by Vodolazov in 2017 which combined a modified TDGL approach with the solution of the electrothermal equations and Maxwell's equations in a 2D geometry [6]. This work captures the main ideas of quasiparticle diffusion, vortex formation, and thermal dissipation in a more rigorous approach which has its origins in the kinetic theory of superconductors. Much of this chapter is dedicated to investigating the accuracy of this model and proposing refinements which enable a more accurate comparison between experiment and theory.

Parallel to these theoretical developments, a number of experiments were performed in order to validate the various detection-mechanism models. These focused on two aspects of detector performance: detector efficiency [94, 95] and dark count rates [44, 96]. In a series of works, photoresponse was characterized using bowtie NbN detectors [94, 95], finding a linear dependence between bias current and detection energy. This was used to infer the importance of diffusion in the detection process due to the linear energy-current relation. Furthermore, the work concluded that vortex crossing also plays an important role in the detection process due to the temperature dependence of the detection threshold current. A more definitive verification of the importance of vortex dynamics came with the measurement of photoresponse in a magnetic field [51]. This work measured shifts in the normalized photoresponse count rate (PCR) curves in applied magnetic fields which were qualitatively consistent with the predictions of the vortex hotspot model [93].

Several works have measured the intrinsic dark count rates of SNSPDs under various conditions in order to reveal the dominant mechanism of these events. Early investigations immediately identified vortex motion as the primary cause of dark counts [44, 96], but ambiguity remained as to the exact mechanism for this motion. As measurements and models were refined, thermal activation of vortices was identified as a candidate capable of describing experimental measurements in meandered NbN devices, but the measurements could not distinguish between the mechanisms of unbinding of vortex-antivortex pairs and vortex entry from the edge of the nanowire [97]. A later work determined that thermal activation of vortex-antivortex unbind-

ing was the dominant source of dark counts in a 100 nm wide NbN device at the low temperatures where SNSPDs are typically operated [98]. This understanding is consistent with the observation of increased dark count rates in detectors with sharp corners where current crowding [99] leads to an increased current density and reduced barrier for vortex entry or vortex-antivortex unbinding [51, 100].

While this body of work represents remarkable progress in the qualitative understanding of single-photon detection in SNSPDs, there are a number of shortcomings which still must be addressed. The primary shortcoming of previous works is quantitative accuracy and predictive capability. A proper detection model should at least be capable of fitting experimental data, but ideally would be able to predict behavior under new operating conditions. Furthermore, such a model should be capable of describing all detector behavior rather than a single experimental dataset.

There are several areas of concern regarding the conclusions reached by previous experiments. In particular, the series of works conducted on bowtie detectors suffer from ambiguities emerging from the geometry of the devices. It is well known that non-uniform geometry leads to current crowding in SNSPDs [99]. Bowtie detectors are no exception. In this geometry, the leads of the detector sustain a lower current density than the active square, but additionally, there is current crowding at the corners where the square meets the triangular leads. Therefore, any measured photoresponse or dark count rates are a combination of the detections in the leads as well as the active square. Current crowding also means that the edges of the detector see a higher current density than the center, which would make the edges more photosensitive than the center compared to an identical device with uniform current distribution. Such a flaw skews the conclusions reached from detection modeling. Furthermore, the conclusions about photodetection use an onset threshold of 1% probability of detection, which is a regime where the behavior of fluctuations is expected to be strongly influential. This skews the data toward characterizing only the most photosensitive part of the detector rather than describing the whole detector active area. Due to these problems, the linear energy-current relation observed in those works has largely been dismissed in favor of a nonlinear energy-current relation, as observed in WSi [53], MoSi [101], and NbN [51].

It was only recently that attention turned from general investigations of the internal detection efficiency to the timing properties of these devices. Single-photon detection is not an instantaneous process. The primary timescale determining the

detector response is its latency. By definition, the latency of a detector, τ_{lat}, is the time interval between the photo-absorption event and the detector click. Unfortunately, the latency is not directly measurable so timing information about a detector has traditionally been characterized through its instrument response function (IRF), which is a measure of the response time of the nanowire compared to a reference. This IRF represents a distribution of detector response times, and the FWHM of this distribution is typically denoted as the timing jitter.

For many years, the jitter of SNSPDs was dominated by electrical and amplifier noise [102, 103]. Noise, when added to a signal with finite slew rate, leads to differences in the time at which the signal crosses a threshold level, as shown in Figure 3.1a. Local inhomogeneity in the nanowire can contribute to timing jitter by changing the dynamics of hotspot growth [104, 105], but these differences in the properties of the superconductor can also change the time required to suppress superconductivity, thus changing the latency. Any probabilistic change in latency contributes to timing jitter. More recently, the concept of longitudinal geometric jitter was introduced to account for differences in detector timing due to the finite propagation speed of microwave signals along the length of the nanowire [32, 70] as shown in Figure 3.1b. This mechanism can actually be used to infer the location of photon absorption by operating the device in a double ended readout scheme [32].
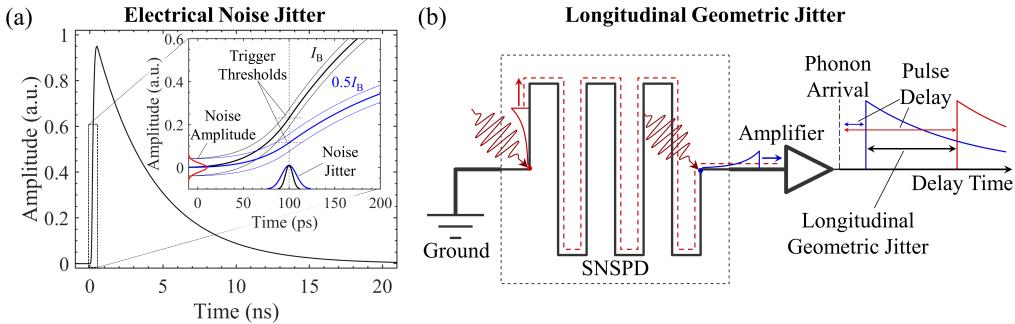
Figure 3.1: Jitter Overview. (a) Origin of the noise contribution to timing jitter. (b) Origin of the longitudinal geometric contribution to timing jitter. Differences in the distance microwave signals travel along the length of the SNSPD lead to different delay times in the onset of the detection pulse.

There are a number of other sources of jitter which are directly related to the microscopic physics of the detection process. These sources are often termed intrinsic jitter because they are inherent to the detection process itself. There has been intense interest in studying the intrinsic jitter of SNSPDs from both an experimental

and theoretical perspective. An experimental study of the jitter associated with meandered [106] and straight [107] nanowires found asymmetry in the jitter profile which was attributed to intrinsic effects. The same group also measured an increase in the timing jitter of straight nanowires in the presence of increasing magnetic field [108]. Unfortunately, these works rely on subtracting the electrical noise contribution to the timing jitter and do not provide sufficient experimental data to make any useful quantitative comparisons with theory. By reducing the classical contributions of jitter by employing low-noise cryogenic amplifiers and reducing the length of devices to 5 μm, a collaboration between the groups of MIT, JPL, and NIST Boulder demonstrated record low timing jitter of 2.6 ps at a wavelength of 532 nm [2]. Investigation of the wavelength, substrate temperature, and nanowire width dependence of the timing jitter confirmed that the jitter was dominated by intrinsic components. Furthermore, by studying a variety of devices and photon energies, the team produced the first dataset which can be used to test detection models on a quantitative level.

Several theoretical works have accompanied the recent experimental efforts. A number of these studied the effect of the coordinate of photon absorption along the width of the wire on timing performance. Differences in the response time for various absorption sites appear as timing jitter when averaging over all potential absorption locations, in an effect referred to as transverse geometric jitter. One work studied this jitter mechanism through the effect of vortex entry [54] based on the model of Engel [48], but its quantitative accuracy is questionable due to the limitations of the underlying model. A later work [55] studies the same effect using the TDGL model [6] which naturally handles the formation and crossing of vortex-antivortex pairs as well as vortices entering from the edge of the nanowire. This model was similarly used to describe the qualitative increase in timing jitter in a magnetic field observed experimentally [108].

Fluctuations during the detection process contribute directly to timing jitter. A theoretical work on quantum fluctuations estimated this effect on timing jitter and detection efficiency [109] based on the quasiparticle diffusion model [48], but the simplistic nature of the model used to describe the superconducting condensate in that work makes it difficult to draw any connections to experimental works. Recently, we discussed the implications of finite latency in the face of energy fluctuations either from spatial non-uniformity or Fano fluctuations during the downconversion process [3, 53]. It was found that the detector latency sets the scale for determining

the impact of energy fluctuations on the eventual timing response of the detector. We also determine that the intimate relationship between timing jitter and internal efficiency observed in recent experiments [2] is directly associated with the latency of the detector response. This will be discussed qualitatively in Section 3.3 and plays an important role in connecting experimental results with microscopic modeling as shown in Section 3.4.2.

This chapter discusses the efforts made to connect experimental measurements with the predictions of detection models for SNSPDs. It begins with a description of the new experimental techniques developed through the collaboration between MIT, JPL, and NIST which has enabled the measurement of the intrinsically-limited performance of current-carrying nanowire structures in the time domain. This new experimental window on detector performance provides an additional mechanism for probing the fundamental detection process of these detectors. Armed with this new technique, we apply the most advanced existing detection models in an attempt to fit experimental data and validate the model's performance, but find that the model is unable to match all experimental results. Given this failure, we seek to refine the theoretical model of the SNSPD detection mechanism in order to improve the qualitative understanding of detector performance and eventually develop a fully predictive model of SNSPD detection. The later sections of the chapter discuss modifications to the existing SNSPD detection model and their impact on the ability to fit experimental data. The chapter concludes with a discussion of the limitations of our modified detection model.

3.2 Experimental Techniques

Despite all of the theoretical work on the SNSPD detection mechanism, there have been few attempts to fit experimental data at a quantitative level. One reason is a lack of complete experimental measurements which simultaneously measure more than one performance metric. The experimental work of Korzh et al. [2] and subsequent enhancements to the techniques presented in that work have offered the first possibility of making a quantitative comparison between theory and experiment.

3.2.1 Intrinsic Timing Jitter

In order to use the timing information of the detector response to inform theoretical efforts, the timing jitter must be reduced to the level where it is limited by the intrinsic sources of jitter discussed above. As knowledge of the classical sources of jitter improved, specialized devices could be designed to reduce these effects and

probe the intrinsic sources of jitter. Devices were designed by the MIT-JPL-NIST collaboration to eliminate longitudinal geometric jitter by having only a short 5 μm long active area in series with a 96 nH inductor which was optimized to keep the rising edge of a detection pulse as fast as possible while being long enough to avoid latching [56, 68, 110]. The typical microwave signal transmission speed in NbN devices of these dimensions is on the order of 2% the speed of light, or about 6 μm/ps, so a 5 μm device would have geometric jitter below 1 ps.

The electrical noise contribution to the timing jitter was substantially reduced by using a low-noise cryogenic amplifier (Cosmic Microwave, CITLF1) operating at the 4 K stage of the cryostat. For devices with low bias currents, an additional amplifier stage at room temperature was used to improve the signal-to-noise ratio of the detector output. Jitter measurements were performed with a high bandwidth sampling oscilloscope measuring at 80 gigasamples per second, though the signal bandwidth was reduced to 6 GHz to optimize the signal-to-noise ratio of the resulting signal. Finally, the devices were free-spaced coupled to avoid dispersion of the optical pulse in fiber. This also enabled measuring over a wide range of photon energies. All of the measurements used in the following analysis were performed using 1550 nm, 1064 nm, 775 nm, and 532 nm light with the experimental setup shown schematically in Figure 3.2.[1]

---

[1] Additional experimental details can be found in [2].
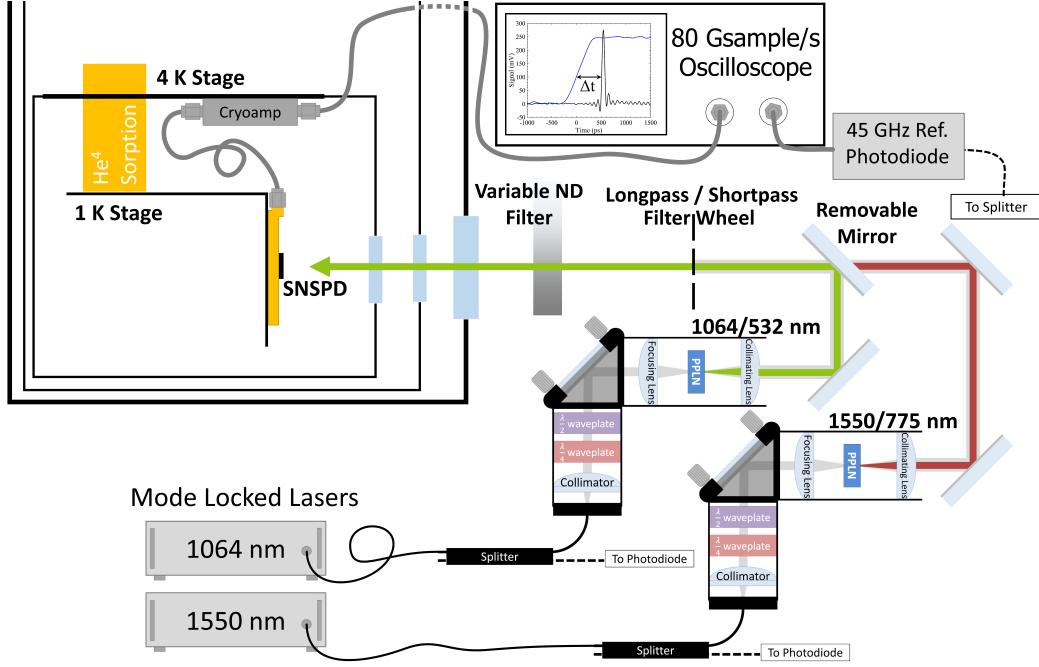
Figure 3.2: Low jitter experimental setup. Two mode-locked lasers (1550 nm and 1064 nm) are focused into second harmonic generation crystals to produce 775 nm and 532 nm light. After a wavelength is selected by appropriate use of a longpass or shortpass filter, light is free space coupled into the cryostat and onto the detector. The signal from the SNSPD is amplified using a cryogenic amplifier operating at 4 K before being recorded by a 80 gigasample per second oscilloscope.

3.2.2 Relative Latency

To understand the timescale of photon detection, an ideal experiment would measure the latency of the detection process directly. However, it is extremely difficult to calculate or measure exactly when a photon is absorbed in the nanowire. Even if the delay could be properly calibrated, such a measurement still only provides a measurement of the detector latency relative to a reference detector, which itself has some unknown latency. Devising a scheme which can accurately measure absolute latency is an open question.

The MIT-JPL-NIST work avoids this problem by introducing a new technique to measure the relative latency between detection events of photons with different energies. Using periodically-poled lithium niobate (PPLN) second harmonic generation (SHG) crystals, light from either the 1550 nm or 1064 nm mode-locked laser sources could be frequency doubled to generate synchronized 775 nm or 532 nm light, respectively. Because the optical pulses travel along the same optical path, they are synchronized in time and each wavelength can be selected by the appropriate choice

of a longpass or shortpass filter, as shown in Figure 3.2. The IRF of a detector can be measured for the two different energies of light referenced to a photodiode. As long as the electrical characteristics of measurement apparatus do not change, the IRF of the two wavelengths can be compared to calculate a relative delay time difference for the photons of different energies. It is important to recognize that the measurement relies on the electrical setup remaining unchanged as the two wavelengths are selected. Any change in the amplifier or trigger settings leads to a shift in the measured delay time and compromises the measurement. The technique also requires calibration of the optical delay through the different elements of the free space setup because the propagation speed of the two wavelengths is different due to the wavelength dependence of the material index of refraction. This new measurement, known as relative latency, is a direct measure of the timescale of the single-photon detection process in SNSPDs.

A comparison of the three main experimental metrics, internal efficiency, timing jitter, and relative latency, is shown for a 100 nm wide NbN device measured at a temperature of 1 K. A quick inspection shows that there is a clear connection between these device properties. For a given bias current, as the wavelength decreases, the timing jitter decreases. Furthermore, as the bias current increases, both the timing jitter and relative latency decrease for all of the measured wavelengths. The reduction in timing jitter and relative latency coincides with the transition toward saturated internal detection efficiency. It should also be noted that the relative latency and jitter FWHM have similar values, further hinting at the connection between these two metrics.
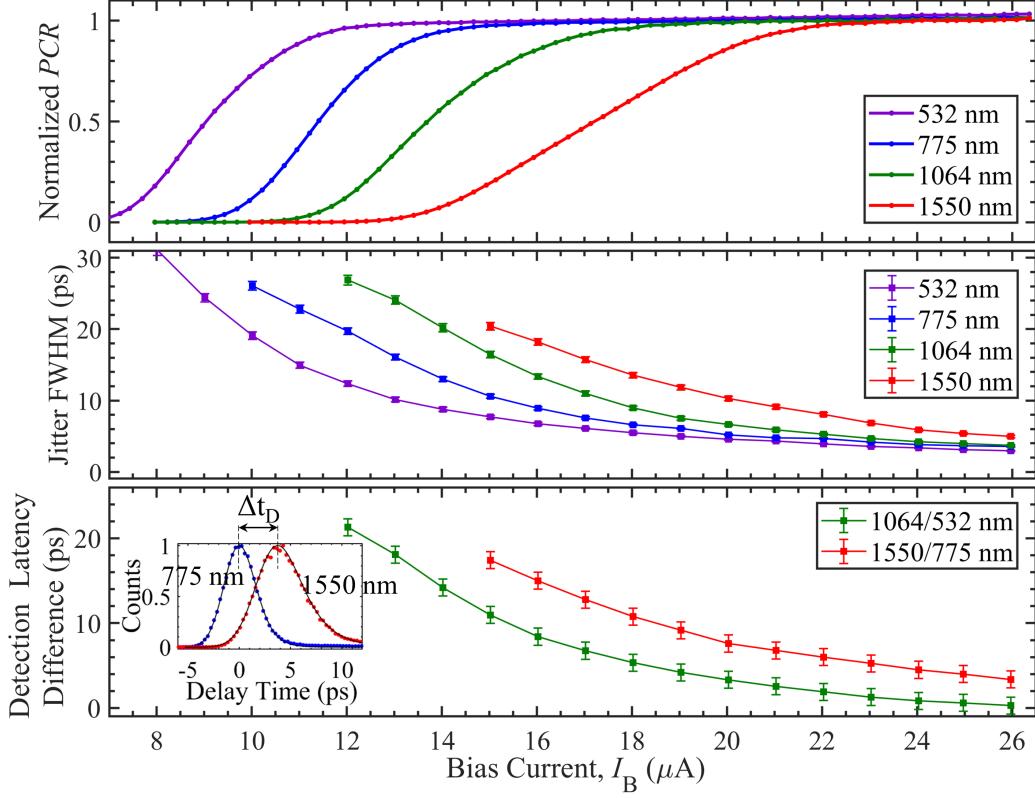
Figure 3.3: Summary of measurements for a 100 nm NbN device. There is a clear connection between the PCR (top), jitter (middle), and relative latency (bottom) for the different photon energies. As the bias current decreases from the switching current, the increase in timing jitter coincides with the transition away from saturated internal detection efficiency. The increase in relative latency (shown schematically in the inset as Δt_D) coincides with the increase in timing jitter for the longer wavelength of each photon pair.

3.2.3 Polarization-Dependent Detection Efficiency

The absorption of electromagnetic waves in SNSPDs is well described by classical electromagnetic theory with the absorption efficiency linked to the magnitude of the electric field in the active nanowire of the device. It is well known that the electric field distribution in SNSPDs is dependent on the polarization of the incident photon. In the transverse-electric (TE) polarization, the electric field is oriented parallel to the direction of the nanowire, while in the transverse-magnetic (TM) polarization, the electric field is perpendicular to the nanowire as shown in Figures 3.4a and 3.4b.

Calculating the behavior of the electric field is non-trivial, and generally requires the use of numerical techniques and specialized software. The periodic geometry of nanowire meanders enables the use of Fourier techniques through the rigorous

coupled-wave analysis (RCWA) method [63–65] to calculate the optical behavior of the nanowire. An alternative approach is to use finite-difference time-domain (FDTD) software to model the electromagnetic response of the detector, but this is more computationally demanding than the simple RCWA technique. While it may be difficult to predict the precise absorption behavior of nanowires without numerical modeling, the main qualitative feature of polarization dependence can be understood from simple intuitive arguments. When the electric field is oriented parallel to the nanowire in the TE polarization, charge oscillates along the length of the nanowire. For nanowires smaller than the diffraction limit, this oscillation is approximately uniform across the width of the nanowire. In contrast, in the TM polarization, charge is accelerated perpendicular to the nanowire. At the edges of the wire, the nanowire has an interface with an insulating dielectric, so the charge cannot continue its motion. Current is proportional to electric field, so the electric field must drop to zero. This implies that the electric field is expected to be larger in the center of the nanowire compared to the edges.

The change in electric field within the nanowire directly impacts the probability of photon absorption at different locations in the nanowire. Because the energy of optical photons is significantly larger than the superconducting gap, the optical absorption can be approximated as being that of a normal metal. The probability of generating a single electron-hole pair is proportional to $|E|^2$. Based on the transverse coordinate dependence of the electric field, it is possible to infer the probability of a photon being absorbed as a function of the transverse coordinate of the nanowire. If the detection process of the SNSPD changes with the transverse coordinate, as assumed with the hotspot model, then this effect should be evident by changes in detector performance with polarization.

Using both RCWA and FDTD techniques, the electric field profile was calculated for the NbN nanowires used throughout this chapter. Figure 3.4c shows the magnitude squared of the electric field, which is proportional to the probability of absorption, as a function of transverse coordinate along the width of the wire based on the FDTD method. The calculations agree with the simple intuition about the motion of charge in the nanowire. These calculations indicate that the TE polarization samples the entire nanowire width nearly uniformly, while the TM polarization samples the middle more heavily than the edges. Therefore, any differences in the internal efficiency or timing jitter of the detector response for the two polarizations can be attributed to the difference in the transverse coordinate dependence of the detection

mechanism. It is worth pointing out that this simple argument about the electric field profiles for TE and TM polarizations is only true for narrow nanowires which are much smaller than the wavelength of light. In wide nanowires, more complex electric fields can be generated.
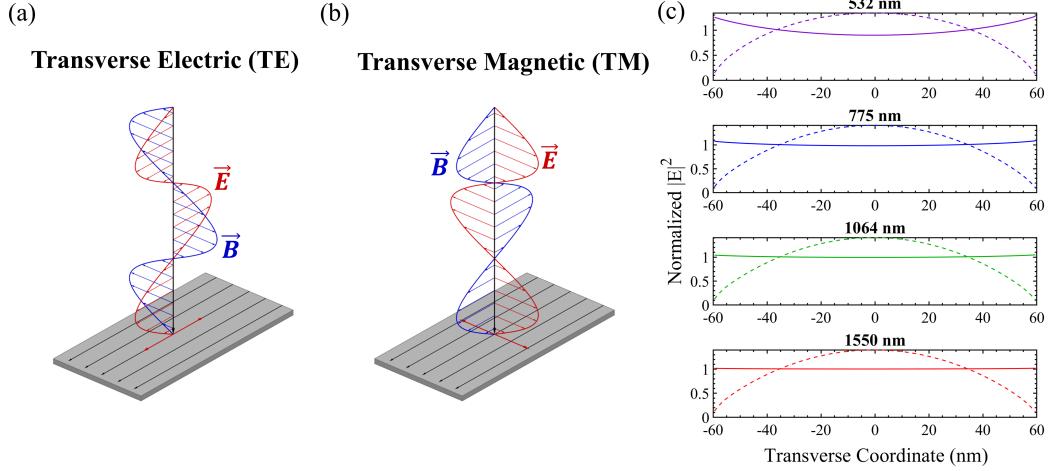
Figure 3.4: Orientation of the (a) TE and (b) TM polarizations with respect to the nanowire. (c) Electric field profile in a 120 nm NbN nanowire encapsulated in SiO_2 for the four wavelengths of interest for experimental measurements. Solid lines indicate the TE mode while dashed lines are the TM mode. Calculations were performed using FDTD software on a CPW structure with the same dimensions as the experimental devices.

The polarization-dependent internal efficiency, timing jitter, and relative latency were measured for four NbN devices with widths of 60 nm, 80 nm, 100 nm, and 120 nm at bath temperatures of 1 K and 4 K. Measurements were performed on the low jitter setup described in Section 3.2.1 with the addition of a half-wave plate located just before the windows of the cryostat. By rotating the half-wave plate by 45 degrees, linearly polarized light was shifted between the TE and TM polarizations. As mentioned in the relative latency section, calibration of the optical delay through all of the optical components was necessary to compare the latency of one photon energy to another. The addition of the waveplate meant that an additional calibration factor was needed to account for the change in propagation delay for the waveplates of the different wavelengths. This was performed by measuring the relative latency between the same photon energies with and without the waveplate in the system.

Photon count rates on the order of 10–50 kilo-counts per second were used for all of the measurements. The low count rate was used to ensure the detector fully

recovered between registered detection events. In order to gain useful statistics for the low count rate regions of the PCR curves, several minutes of integration were necessary. Furthermore, the measurements were performed by acquiring several PCR curves with shorter integration times and averaging the results together. This helped compensate for long timescale drift of the optical setup.

The polarization-dependent detection efficiency and timing jitter for the 120 nm wide device measured at 1 K are shown in Figure 3.5. There are a few immediate trends to observe. First, there is a clear distinction between the two polarizations both in the PCR curves and in the timing jitter. At high photon energies (low bias currents), the TM polarization saturates internal efficiency before the TE polarization, implying that the center of the nanowire is more photosensitive than the edge, based on the electric field profile. At low photon energies (high bias currents), the opposite is true. The TE polarization saturates before the TM polarization, indicating that the center of the nanowire is less photosensitive than the edge. There is a parallel behavior in the timing jitter. For 532 nm light, the TM polarization exhibits lower jitter than the than the TE polarization, while for the 1550 nm wavelength, the TE polarization shows lower jitter. Once again, we see a clear connection between the saturation of internal efficiency and the timing jitter of the device.
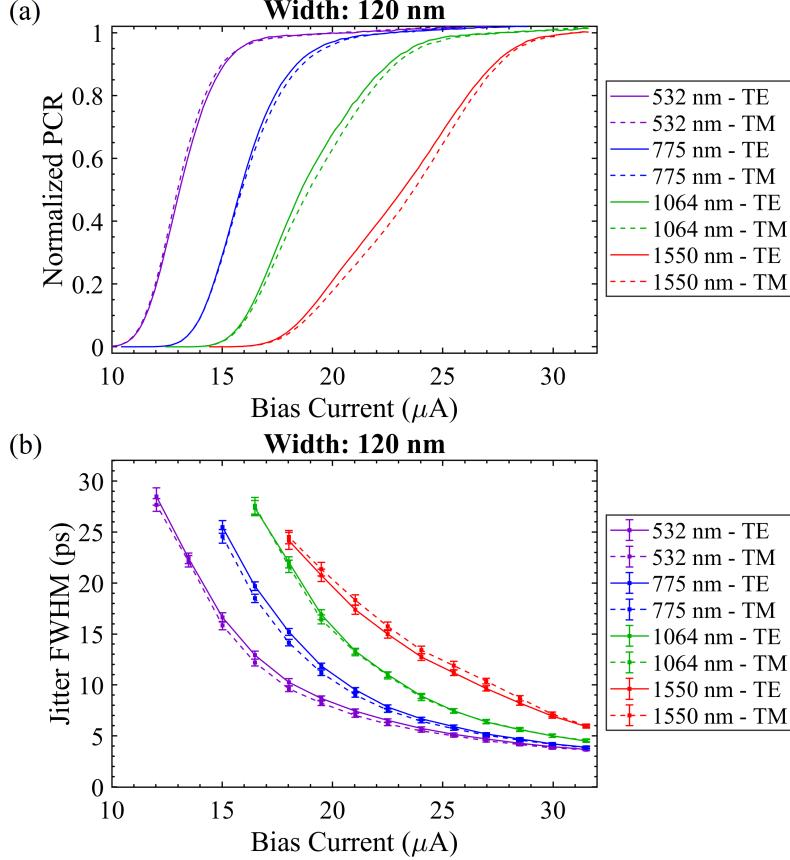
Figure 3.5: Polarization-dependent (a) PCR and (b) timing jitter for a 120 nm NbN device measured at a bath temperature of 1 K. Solid lines indicate the TE polarization and dashed lines show the TM polarization.

As the nanowire width decreases, the 2D nature of the hotspot detection process is expected to give way to a quasi-1D process. Given that the polarization-dependent response elucidates the 2D nature of the photodetection process in the 120 nm wide device, it is interesting to probe if the same is true for the narrowest device measured in this work. Somewhat surprisingly, the 60 nm wide device still shows a measurable polarization dependence in the PCR curves and timing jitter at 1 K as shown in Figure 3.6 for certain combinations of bias current and photon energy. The general trends are the same as for the 120 nm wide device. This result demonstrates the 2D nature of the detection mechanism even in relatively narrow NbN detectors. We will show in Section 3.4 that the polarization-dependent detector response enables a detailed comparison between theory and experiment and is an essential tool in refining models of the detection mechanism.
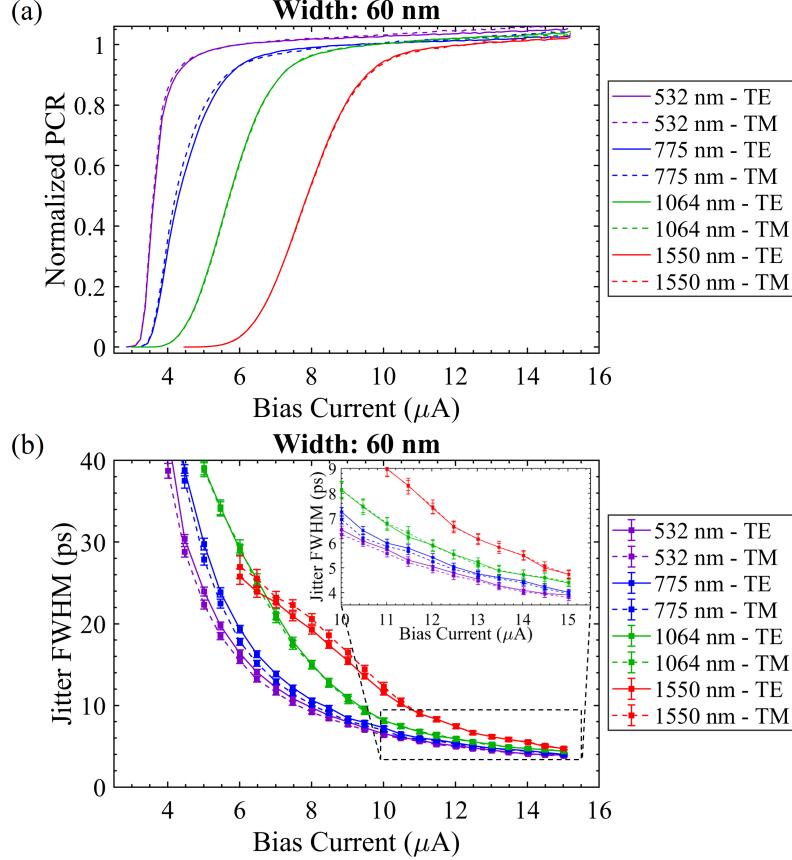
Figure 3.6: Polarization-dependent (a) PCR and (b) timing jitter for a 60 nm NbN device measured at a bath temperature of 1 K. Solid lines indicate the TE polarization and dashed lines show the TM polarization.

3.3 Detector Latency and Timing Jitter

The experimental demonstration of the connection between internal efficiency, timing jitter, and relative latency suggests that understanding the timing of the detection is a prerequisite for developing an accurate model of single-photon detection. Timing is governed by the latency of the detector response, and is a fundamental characteristic of any detector. Using straightforward qualitative arguments about the latency function, it is possible to demonstrate that all of the qualitative trends observed in experiment are expected based on the properties of the detector latency and energy fluctuations during the downconversion process [3].

We define the deterministic latency function $\tau_{lat}(I_B, T_{sub}, E, y, B)$, which encodes the time it takes for the detector to register a click for a given bias current I_B, substrate temperature T_{sub}, absorbed energy E, transverse coordinate y, and magnetic field B. The latency function does not explicitly encode the effect of fluctuations in the

system. General arguments allow us to make several qualitative statements about the shape of the latency function. Below some detection energy $E_{det}(I_B, T_{sub}, y, B)$, the detector does not register detection, so the latency must be $\tau_{lat} \to \infty$. For energies $E > E_{det}(I_B, T_{sub}, y, B)$ the latency must be finite and is assumed to decrease monotonically with increasing energy. The latency must also saturate as the energy increases, $\lim_{E \to \infty} \tau_{lat}(I_B, T_{sub}, E, y, B) = \tau_\infty(I_B, T_{sub}, y, B)$ because superconductivity cannot be suppressed immediately everywhere along the length of the nanowire after the absorption of a photon. These arguments alone suggest (but do not guarantee) that the latency curve should have positive curvature in order to smoothly match the integrable singularity at E_{det} with the limit at τ_∞. This is consistent with microscopic modeling of the 1D system [3]. It should be noted that in [55], Vodolazov argues against the presence of a singularity based 2D modeling results. However, this does not change the qualitative results based on our general statements of the shape of the detector latency, and the 2D modeling results [55] still exhibit a rapid increase in latency upon reaching the detection energy in a manner qualitatively similar to a singularity.

Having defined the latency, we can consider the effect of energy fluctuations on the response of the system. In Chapter 2, we explored the role of phonon escape in creating fluctuations in the amount of energy retained by the superconductor after downconversion. This is the mechanism used to describe fluctuations in [53] and [3]. It can also be shown that non-uniformity in the detector can be expressed as an energy fluctuation source with $\sigma_{NU} \sim E_\lambda$ [3]. The probability that a particular energy E is deposited in the nanowire is given by a normalized Gaussian distribution

$$P(E) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-\bar{E})^2}{2\sigma^2}}, \tag{3.1}$$

where $\bar{E}$ is the average deposited energy and $\sigma^2 = \sigma_F^2 + \sigma_{NU}^2$ is the quadrature sum of the fluctuation contributions due to Fano fluctuations and non-uniformity. Treating the fluctuation sources as independent is justified because the dynamic process of energy loss due to phonon escape is not strongly correlated with the small fluctuations in material properties of the film.

Within this framework, the normalized time-dependent photon-counting rate is given

by

$$\text{PCR}(t, I_B, T_{sub}, B, E_\lambda) =$$
$$\frac{1}{w} \int_{w/2}^{w/2} dy \, p(y) \int_0^{E_\lambda} dE \, P(E) \Theta \left[t - \tau_{lat}(I_B, T_{sub}, E, y, B) \right] \tag{3.2}$$

where w is the nanowire width, $p(y)$ is the probability of absorption at coordinate y, and Θ is the Heaviside function. This defines the probability of a registered detection on the time interval $[0, t]$, averaged over the transverse coordinate and potential deposited energies. The IRF defines the probability density of observing a detection at time t, which corresponds to the rate of change of the PCR. This leads to the form $\text{IRF}(t, I_B, T_{sub}, B, E_\lambda) = \frac{d\,\text{PCR}(t, I_B, T_{sub}, B, E_\lambda)}{dt}$. Defining the time-dependent detection energy $E^*(t, I_B, T_{sub}, y, B)$ according to the solution of $t = \tau_{lat}(I_B, T_{sub}, E^*, y, B)$, we can differentiate the expression for the PCR to arrive at a definition of the IRF

$$\text{IRF}(t, I_B, T_{sub}, B, E_\lambda) =$$
$$- \frac{1}{w} \int_{w/2}^{w/2} dy \, p(y) P\left(E^*(t, I_B, T_{sub}, y, B)\right) \frac{\partial E^*(t, I_B, T_{sub}, y, B)}{\partial t}. \tag{3.3}$$

Note that we have used the monotonic decrease of τ_{lat} with increasing energy to ensure E^* is single-valued and $E^* \to \infty$ for $E < E_{det}$, ensuring that $P(E^*) = 0$ for $E < E_{det}$. This form of the IRF is not Gaussian due to the non-linearity of the function E^*.

The general expression (3.3) for the IRF can be simplified for the case of small energy fluctuations. Defining the average delay time as $\bar{t} = \tau_{lat}(I_B, T_{sub}, \bar{E}, y, B)$, taking a series expansion of

$$E^*(t, I_B, T_{sub}, y, B) \approx E^*(\bar{t}, I_B, T_{sub}, y, B) + \left. \frac{\partial E^*(\bar{t}, I_B, T_{sub}, y, B)}{\partial t} \right|_{t=\bar{t}} (t - \bar{t}), \tag{3.4}$$

and ignoring the y dependence of the detection function leads to a Gaussian IRF

$$\text{IRF}(t, I_B, T_{sub}, B, E_\lambda) \approx \frac{1}{\sqrt{2\pi}\bar{\sigma}\left(I_B, T_{sub}, B, \bar{E}\right)} \exp \left[-\frac{(t - \bar{t})^2}{2\bar{\sigma}^2\left(I_B, T_{sub}, B, \bar{E}\right)} \right] \tag{3.5}$$

with

$$\bar{\sigma}\left(I_B, T_{sub}, B, \bar{E}\right) = \sigma \left| \left(\left. \frac{\partial E^*(\bar{t}, I_B, T_{sub}, B)}{\partial t} \right|_{t=\bar{t}} \right)^{-1} \right| = \sigma \left| \frac{\partial \tau_{lat}\left(I_B, T_{sub}, \bar{E}, B\right)}{\partial \bar{E}} \right| \tag{3.6}$$

From this expression, it is clear that in this limit of small fluctuations, the jitter FWHM is proportional to the slope of the latency function with energy. This is shown visually in Figure 3.7a. Reintroducing the dependence on y distorts the shape of the IRF and it will no longer be Gaussian as the IRF becomes an average of the coordinate dependent IRF(y) weighted by the probability of absorption at each coordinate.

The general arguments about the shape of the latency function explain the main qualitative trends observed in experiment [2]. The saturation of the latency at high values of E leads to a smaller slope $\partial\tau_{lat}/\partial E$, which leads to smaller jitter, as observed with the reduced timing jitter at increased photon energies. This also ensures that photons with higher energy lead to detections before those of lower energies, as observed in the relative latency measurements. This difference in time decreases as the photon energies increase, as observed when comparing the 1550/775 nm pair to the 1064/532 nm pair shown in Figure 3.3. Finally, the positive curvature of the latency function leads to the appearance of a long-delay time tail in the IRF as observed in in experiments with NbN [2, 106–108], MoSi [111] and WSi [2]. This is also easily understood through visualization in Figure 3.7. Due to the positive curvature, detections below $\bar{E}$ are delayed to much greater times than the mean compared to detections with energies above $\bar{E}$ which occur only a short time before the mean delay.
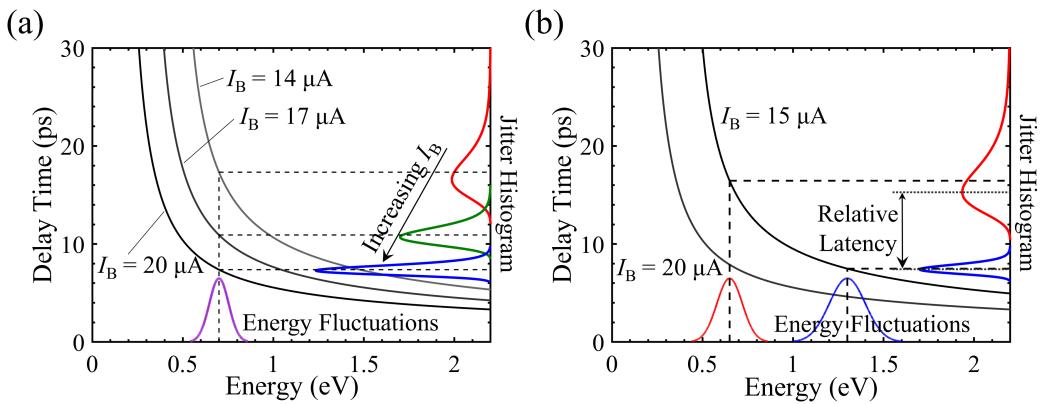
Figure 3.7: Connection between energy fluctuations and latency. For latency curves with different slopes (a), if the magnitude of the slope is larger, the magnitude of the jitter will also be larger for a given energy fluctuation distribution. For a given latency curve (b), increasing the photon energy leads to smaller jitter and a shorter delay time, resulting in a relative latency which can be measured experimentally.

Introducing the concept of detector latency provides a framework within which it

is easy to qualitatively explain the main observations of experiment. However, predicting and analyzing the details of this latency curve must come from a microscopic theory.

3.4   Time-Dependent Ginzburg–Landau Modeling

To begin formulating a model to describe the detection process in SNSPDs, we start from the most advanced microscopic model presented in the literature, that of Vodolazov [6]. This model employs a set of modified time-dependent Ginzburg–Landau (TDGL) equations to describe the evolution of the superconducting state coupled to a two-temperature model to describe the energy flow in the electron and phonon systems. These equations are solved together with Maxwell's equations describing the electric field in the superconductor. This constitutes a set of five coupled equations which must be solved over the relevant geometry of the nanowire for an appropriate set of initial and boundary conditions.

There are numerous advantages to using a Ginzburg–Landau formalism to describe the superconducting state. The equations can be described using a relatively simple set of partial differential equations using the complex order parameter $\Delta = |\Delta|e^{i\phi}$. The equations also naturally describe the nucleation and motion of vortices in the superconducting condensate. This is in contrast to the more formally accurate kinetic equations which require solving for distribution functions describing the excitations within the superconductor. Despite these advantages, there are severe limitations. It is well known that the TDGL equations can only be derived from the microscopic kinetic equations in the limit of $T_e \rightarrow T_c$ and in the limits of $\Delta \rightarrow 0$, small spatial variations of T_e and Δ, and slow temporal variations. These conditions are not met during SNSPD operation. Typical SNSPDs are operated at temperatures only a fraction (0.1–0.3) of T_c. Furthermore, during the initial process of suppression of superconductivity due to quasiparticles generated from photon absorption, the superconducting gap remains finite. The length scale of the suppression of superconductivity is on the order of the superconducting coherence length ξ, and local suppression leads to significant gradients in both temperature and order parameter.

Given all of these limitations, it is optimistic to assume that this formulation is capable of quantitatively describing the single-photon detection process in current-carrying nanowires. Nevertheless, we apply this model to attempt to explain the experimental measurements of relative latency, intrinsic timing jitter, and polarization-

dependent detection efficiency and jitter. We find that the model proposed by Vodolazov [6] fails to account for the timing performance of SNSPDs, though it qualitatively describes the polarization-dependent response of the detectors. Motivated by this failure, we make various modifications to the model to improve the connection between theory and experiment. Despite a range of modifications, the model is unable to fully reproduce experimental measurements, casting doubt on the quantitative accuracy of the TDGL formulation.

3.4.1 Application of the Vodolazov Model

For reference, we reproduce the formulation used by Vodolazov [6], which acts as the starting point for the investigation into the process of suppression of superconductivity in SNSPDs. The two-temperature model is used to describe the electron (T_e) and phonon (T_{ph}) systems and the linear Debye spectrum is used to describe the phonon density of states. The energy balance equations for the electron and phonon systems ((30) and (31) in [6]) are given by

$$\frac{\partial}{\partial t}\left(\frac{\pi^2 k_B{}^2 N(0)T_e^2}{3} - E_0 \mathcal{E}_s\left(T_e, |\Delta|\right)\right) = \nabla k_s \nabla T_e - \frac{96\zeta(5)N(0)k_B{}^2}{\tau_0}\frac{\left(T_e^5 - T_{ph}^5\right)}{T_c^3} + \vec{j}\vec{E}$$

(3.7)

and

$$B_{ph}T_{ph}^3\frac{\partial T_{ph}}{\partial t} = \frac{90\gamma B_{ph}\zeta(5)}{\pi^4\tau_0\,T_c}\left(T_e^5 - T_{ph}^5\right) - \frac{B_{ph}}{4\tau_{esc}}\left(T_{ph}^4 - T_{sub}^4\right),$$

(3.8)

where we have transformed the phonon equation to use the convention of the present work. In (3.7) and (3.8), k_B is the Boltzmann constant, $N(0)$ is the single-spin density of states at the Fermi level, $E_0 = 4N(0)k_B{}^2T_c^2$ is a characteristic energy density, $\mathcal{E}_s$ is the energy gain from being in the superconducting state, $|\Delta|$ is the magnitude of the order parameter, τ_0 describes the electron-phonon coupling strength, $\vec{j}$ is the current density, and $\vec{E}$ is the electric field. The thermal conductivity k_s is given by

$$k_s = \frac{2\pi^2 Dk_B{}^2 N(0)}{3}T_e\left(1 - \frac{6}{\pi^2}\int_0^{|\Delta|/k_B T_e}\frac{x^2 e^x dx}{(e^x + 1)^2}\right)$$

(3.9)

which was first derived by Bardeen [112]. The energy of the superconducting state is given by

$$\mathcal{E}_s\left(T_e, |\Delta|\right) = \int_0^{|\Delta|/k_B T_c}\tilde{\epsilon}n_{\tilde{\epsilon}}d\tilde{\epsilon}$$

$$-\int_{|\Delta|/k_B T_c}^{\infty}\tilde{\epsilon}\left(N_1\left(\tilde{\epsilon}\right) - 1\right)n_{\tilde{\epsilon}}d\tilde{\epsilon} + \left(\frac{|\Delta|}{2k_B T_c}\right)^2\left[\frac{1}{2} + \ln\left(\frac{\Delta_0}{|\Delta|}\right)\right]$$

(3.10)

where

$$N_1\left(\tilde{\epsilon}\right) = \frac{\tilde{\epsilon}}{\sqrt{\tilde{\epsilon}^2 - \left(|\Delta|/k_BT_c\right)^2}}\Theta\left(\tilde{\epsilon} - \left(|\Delta|/k_BT_c\right)\right) \tag{3.11}$$

is the spectral function describing the density of states in the superconducting state as solved using the Usadel equations for a dirty superconductor in the limit of no current flow. In the phonon equation, $B_{ph} = \frac{2\pi^2 k_B^4}{5\hbar^3 c_{avg}^3}$ describes the phonon heat capacity according to the Debye model with $C_{ph}(T) = B_{ph}T^3$, and c_{avg} is the mode-averaged sound velocity. The parameter $\gamma = \frac{4\pi^4}{15}\frac{E_0}{B_{ph}T_c^4}$ describes the ratio of heat capacities of the electron and phonon systems at T_c and τ_{esc} is the phonon escape time.

The complex superconducting order parameter Δ is modeled using a modified TDGL equation ((36) in [6])

$$\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\Delta}{\partial t} + \frac{2ie\varphi\Delta}{\hbar}\right) = \xi_{mod}^2\left(\nabla - i\frac{2e}{\hbar c}\vec{A}\right)^2\Delta + \left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2}\right)\Delta$$
$$+ i\frac{\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}}{|\Delta|^2}\frac{e\hbar D\Delta}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}. \tag{3.12}$$

In this modified form, the coefficients

$$\xi_{mod}^2 = \frac{\pi\sqrt{2}\hbar D}{8k_BT_c\sqrt{1 + T_e/T_c}} \tag{3.13}$$

and

$$\Delta_{mod}^2 = \frac{\left(\Delta_0\tanh\left(1.74\sqrt{T_c/T_c - 1}\right)\right)^2}{1 - T_e/T_c} \tag{3.14}$$

have a temperature dependence which is close to the true temperature dependence of the coherence length and order parameter magnitude, respectively. In (3.12), φ is the electrostatic potential, e is the elementary charge, c is the speed of light, $\vec{A}$ is the vector potential, and σ is the normal state conductivity. The supercurrent of the system is described using an approximation to the Usadel supercurrent [6]

$$\vec{j}_s^{Us} = \frac{\pi\sigma}{2e\hbar}|\Delta|\tanh\left(\frac{|\Delta|}{2k_BT_e}\right)\vec{q}_s, \tag{3.15}$$

where $\vec{q}_s = \hbar\left(\nabla\phi - 2e\vec{A}/\hbar c\right)$ is the superfluid momentum, and is valid in the limit of small $|\vec{q}_s|$ relative to its value at the depairing current. The Ginzburg–Landau supercurrent

$$\vec{j}_s^{GL} = \frac{\pi\sigma|\Delta|^2}{4e\hbar k_BT_c}\vec{q}_s \tag{3.16}$$

also appears in the final term of (3.12). This final term[2] is added to ensure that the stationary state leads to $\nabla \cdot \vec{j}_s^{Us} = 0$ rather than $\nabla \cdot \vec{j}_s^{GL} = 0$ as would occur without the correction. See Appendix B.1 for details. The normal current density $\vec{j}_n = -\sigma \nabla \varphi$ comes from the definition of the electrostatic potential, which is found through the current conservation equation

$$\nabla \cdot \vec{j} = \nabla \cdot \left(\vec{j}_s^{Us} + \vec{j}_n \right) = 0. \tag{3.17}$$

The total current density is defined as $\vec{j} = \vec{j}_s^{Us} + \vec{j}_n$. Based on Equation (3.17), the electrostatic potential is solved using $\nabla \cdot \vec{j}_s^{Us} = \sigma \nabla^2 \varphi$.

The boundary conditions for these equations ensure there is a constant supercurrent flowing in the bulk of the nanowire. At the longitudinal ends of the wire, $T_e = T_{sub}$, $|\Delta| = 0$, $\vec{j}_s^{Us} = 0$, and $-\sigma \nabla \varphi|_n = I/wd$, where $|_n$ indicates that the gradient is taken perpendicular to the boundary. These correspond to the injection of a fixed normal current I while keeping the order parameter suppressed and fixing the temperatures to be that of the substrate. At the transverse edges of the nanowire, the boundary conditions are $\nabla T_e|_n = 0$, $\nabla |\Delta||_n = 0$, $\vec{j}_s^{Us}|_n = 0$, and $\vec{j}_n|_n = 0$, which corresponds to a thermally and electrically insulating boundary. The phonon equation does not require boundary conditions because diffusion has been neglected.

We can compare the stationary state results of this model with the appropriate solution of the Usadel equations for the dirty limit of a current-carrying nanowire. Figure 3.8a shows the critical depairing current as defined by the modified TDGL equations, the Usadel equations, and the Bardeen temperature dependence $j_{dep}(\tilde{T}_e)/j_{dep}(0) = \left(1 - \tilde{T}_e^2\right)^{3/2}$, where $\tilde{T}_e = T_e/T_c$ is the electron temperature normalized by T_c. The solution of the Usadel equations follows the approach of Clem and Kogan [76]. The comparison is favorable for temperatures above $T_c/2$, but becomes worse at lower temperatures. Even with this reduced accuracy for low temperatures, the error between the Vodolazov TDGL formulation and the Usadel equation is less than 7% over the entire temperature range as shown in Figure 3.8b.

---

[2]The final term of (3.12) contains an extra factor of $e\, e^{i\phi}$ (elementary charge (e) multiplied by $\Delta/|\Delta|$), which corrects a typo present in [6].
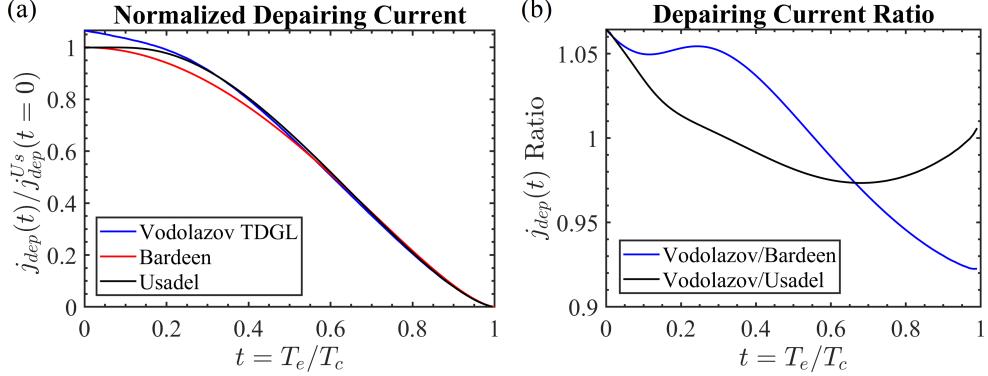
Figure 3.8: Comparison of the stationary current-carrying state within the Vodolazov TDGL model to the solution of the Usadel equations and Bardeen equation. The temperature-dependent depairing current (a) shows good agreement between all three models for $T_e > T_c/2$. The largest error in the ratio of Vodolazov's model to the other two (b) remains below 7% over the entire temperature range.

The stationary supercurrent is not the only factor to consider when describing the current-carrying superconducting state. The magnitude of the order parameter $|\Delta|$ and the superfluid momentum $\vec{q}_s$ both influence the properties of the superconducting state. Figure 3.9 shows a comparison of the magnitude of the order parameter in the stationary current-carrying state for the Usadel and TDGL formulations of the superconductor. At high temperatures and low bias currents, both formulations show agreement in the stationary state order parameter magnitude and phase gradient, but at lower temperatures and higher bias currents, the TDGL formulation becomes more inaccurate by underestimating the magnitude of the order parameter.
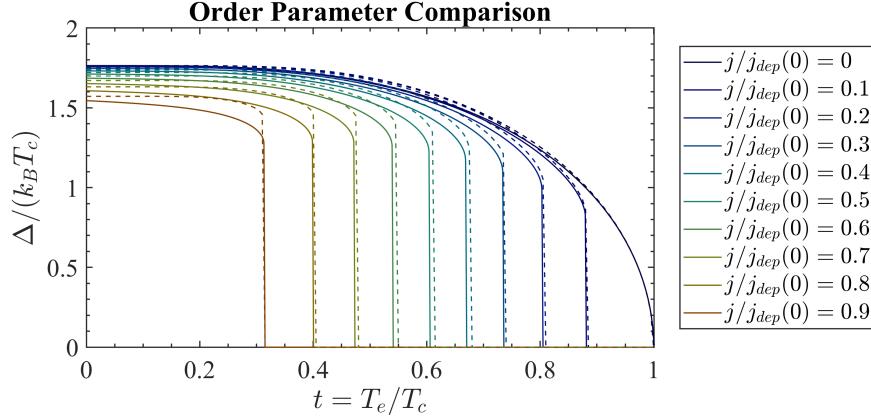
Figure 3.9: Comparison of the stationary state order parameter at various bias currents within the Vodolazov TDGL model (solid lines) and Usadel model (dashed lines). The vertical discontinuities correspond to the temperature where the current exceeds the critical depairing current.

In order to simulate the response of the superconducting nanowire to a photo-excitation, energy from the photon must be inserted into the nanowire. Vodolazov [6] introduces this energy in the form of equilibrated hotspot initial conditions. In this form, a square hotspot of size $2.5\xi_c \times 2.5\xi_c$ is excited to a hotspot temperature $T_e = T_{ph} = T_{HS}$.[3] The energy deposited in this hotspot volume is calculated numerically from the energy balance equation

$$\frac{E_\lambda}{E_0 V_{HS}} = \int_{T_{sub}}^{T_{HS}} \left(\mathcal{E}_e(T, I) + \mathcal{E}_{ph}(T) \right) dT \tag{3.18}$$

where E_λ is the photon energy, $V_{HS} = (2.5\xi_c)^2 d$ is the hotspot volume, $\xi_c = \sqrt{\frac{\hbar D}{k_B T_c}}$ is a characteristic length scale, d is the nanowire thickness, $\mathcal{E}_e(T, I) = \frac{\pi^2}{12} \frac{T^2}{T_c^2} - \mathcal{E}_s(T, |\Delta|(T, I))$ is the total electron energy normalized by E_0 at temperature T and order parameter $|\Delta|$, and $\mathcal{E}_{ph}(T) = \frac{\pi^4}{15\gamma} \left(\frac{T_{ph}}{T_c}\right)^4$ is the total phonon energy normalized by E_0. Note that the magnitude of the order parameter is dependent on the bias current of the nanowire. This set of initial conditions corresponds to the limit of fast thermalization of quasiparticles and phonons, valid when the characteristic thermalization time is much faster than the timescale of evolution of the order parameter.

---

[3]The text of [6] indicates the hotspot size is $2\xi_c \times 2\xi_c$, but this represents elevating T_e and T_{ph} within a 5×5 grid of simulation points on a spacing of $\xi_c/2$. Due to finite spacing, the effective area extends by an extra $\xi_c/4$ in each direction around the hotspot, so the true hotspot size would be $2.5\xi_c \times 2.5\xi_c$.

We apply this numerical model to understand the detection characteristics of short NbN detectors [2]. Our goal is to match the detection efficiency, relative latency, intrinsic jitter, and polarization dependence of the photoresponse. While the TDGL model is deterministic, fluctuations are added in the form of Fano fluctuations [3, 53]. This modifies the energy deposited in the system through the initial temperature of the hotspot. In the idealized Fano fluctuation model, the deposited energy takes the form of a Gaussian distribution with two fitting parameters: $\bar{\chi}$ and σ_F. The mean fraction of photon energy, $\bar{\chi}$, represents the fraction of the photon energy which remains in the superconductor after downconversion. This value is assumed to be independent of the photon energy and can be estimated based on the properties of the superconductor and substrate, as done in Chapter 2. The standard deviation of fluctuations, σ_F, is assumed to be due to the stochastic escape of high-energy phonons during the initial stage of the downconversion cascade. Because this process should be dominated by the early generation high-energy phonons, $\sigma_F \sim \sqrt{N_{ph}} \sim \sqrt{E_\lambda}$ as expected from shot noise in the number of high-energy phonons N_{ph} present during downconversion. Therefore, the probability distribution function for the energy deposited in the superconductor takes the form $p(E) = \frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{(E-\bar{\chi}E_\lambda)^2}{2\sigma_F^2}}$.

The remaining parameters are estimated from literature values or experimental measurement, and are approximately consistent with the parameters presented in Chapter 2. We take the parameters $T_c = 8.65$ K, $\tau_0 = 1.87$ ns (which comes from scaling $\tau_{ep}(10\,\text{K})$ to the appropriate T_c assuming a cubic power law and $\tau_0 = \tau_{ep}(T_c)\frac{720\zeta(5)}{\pi^2}$), and $\tau_{esc} = 9.4$ ps. For the phonon system, we assume an average sound velocity of $v_{avg} = 4912$ m/s, which comes from the sound velocity extracted from the elastic constants measured by neutron scattering of δ-NbN [74]. Using a mass density of 8.3 g/cm^3 [113], and an average atomic mass of (92.9 + 14.0)/2, we arrive at a Debye temperature of 664 K, and $\Omega_D = 76.7$ in units of $k_B T_c$. Using a sheet resistance $\rho_\square$ of $600\,\Omega/\square$ and diffusion coefficient $D = 0.5$ cm^2/s, we arrive at a single-spin density of states $N(0) = 1.49 \times 10^{22}$ eV^{-1}cm^{-3} and characteristic energy scale $E_0 = 33.0\,\mu$eV/nm^3. This also leads to the phonon parameter $\gamma = 23.8$, but for the current simulations, we approximate this as $\gamma = 20$. The small difference in γ does not significantly alter the results.

For ease of computation, the equations are non-dimensionalized, as demonstrated in Appendix B.2. Bias currents are defined according to a fraction of the temperature-dependent depairing current, as estimated by the Bardeen temperature dependence. The temperatures and order parameter are calculated using a local forward Euler

numerical method while the electric potential is solved simultaneously over the entire domain using direct sparse matrix factorization with the UMFPACK library. The numerical scheme was validated in MATLAB through a comparison of results with Vodolazov, but then ported to a MATLAB executable function (mex) written in C++ for faster performance, utilizing Intel's Math Kernel Libraries.

Qualitative Response

Before comparing with experiment, it is informative to first gain a qualitative understanding of the suppression of superconductivity within the TDGL model. The main feature of this model is the natural formation and motion of vortices within the superconducting film. Figure 3.10 shows the magnitude of the order parameter at a sequence of times following the deposition of photon energy in the hotspot region of the detector. The bias current is 60% of the temperature-dependent depairing current, the substrate temperature is 4 K, and the hotspot temperature is 2.53 T_c in a 15×15 nm area. After a period of time, a vortex/antivortex pair unbinds from the hotspot region, with each vortex traveling from the center to the edge of the nanowire. In the wake of the vortices, superconductivity is suppressed. After the crossing of the first vortex/antivortex pair, the crossing of subsequent vortices is so fast that they resemble 1D phase-slip lines rather than 2D vortices. After several phase-slip lines oscillate across the width of the nanowire, a normal domain is formed. As the normal domain grows, a voltage transient appears across the wire, which represents the output pulse from the detector.
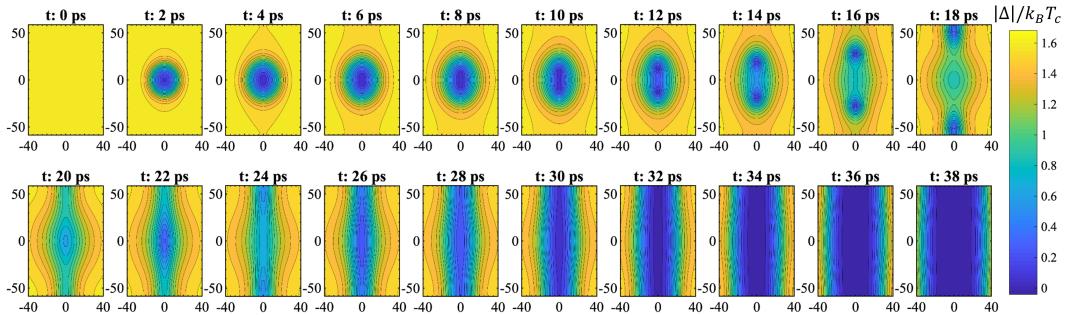
Figure 3.10: Evolution of the magnitude of the order parameter for a detection in the center of a 120 nm wire. The vertical scale is the transverse coordinate y while the horizontal scale is the longitudinal coordinate x with the bias current flowing from left to right. The formation of a vortex/antivortex pair and subsequent motion of the pair to the edge of the nanowire facilitates the suppression of superconductivity.

As a comparison, Figure 3.11 shows a similar sequence of frames of the order parameter for a detection event at the same bias current and photon energy occurring near the edge of the nanowire. In this case, a vortex appears from the edge of the wire and crosses the strip, suppressing superconductivity in its wake. Multiple vortices are required before the order parameter suppression resembles that of a phase-slip line rather than a vortex.
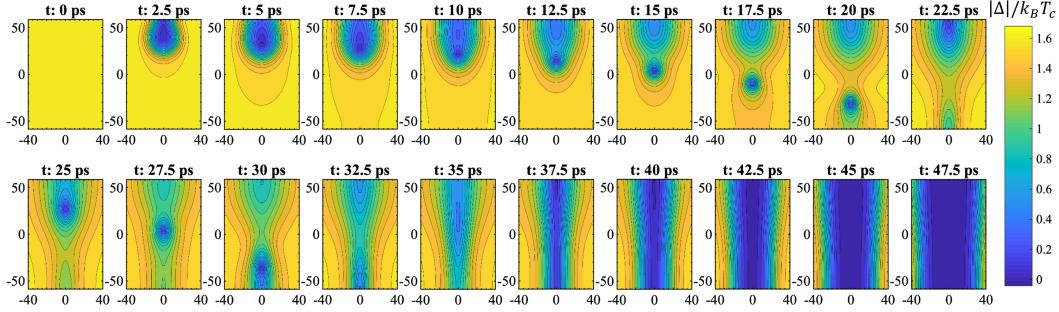
Figure 3.11: Evolution of the magnitude of the order parameter for a detection near the edge of a 120 nm wire. The vertical scale is the transverse coordinate y while the horizontal scale is the longitudinal coordinate x with the bias current flowing from left to right. The formation and subsequent motion of single vortices from one edge to another facilitates the suppression of superconductivity.

The information of the order parameter suppression can be conveyed compactly by forming a contour plot of a cut of the simulation space along the $x = 0$ axis and plotting the order parameter as a function of transverse coordinate y and time. The result of this operation is shown in Figure 3.12 for the detection in the center of the nanowire and Figure 3.13 for the detection near the edge. In this configuration, the motion of vortices along the width of the wire is clearly visible, and by measuring the slope of these vortex trajectories, once can extract the vortex velocity. Interestingly, as the vortex moves across the width of the film, its velocity is not constant. As the vortex nears the edge of the nanowire, it accelerates. Throughout this motion, the vortex suppresses superconductivity in its wake. Therefore, the supercurrent tends to flow around the leading edge of the vortex, and as the vortex progresses, more supercurrent is redirected into a smaller cross-section of the nanowire. This larger supercurrent density increases the speed of vortex propagation due to the Lorentz force. Figures 3.12 and 3.13 also show the temperature evolution at particular points ($y = 0$, 20, and 40 nm) within the superconductor in the bottom panels. Also shown is the voltage which develops as superconductivity is suppressed. The appearance

of voltage oscillations corresponds to the passage of phase-slip lines or vortices throughout the detection process.
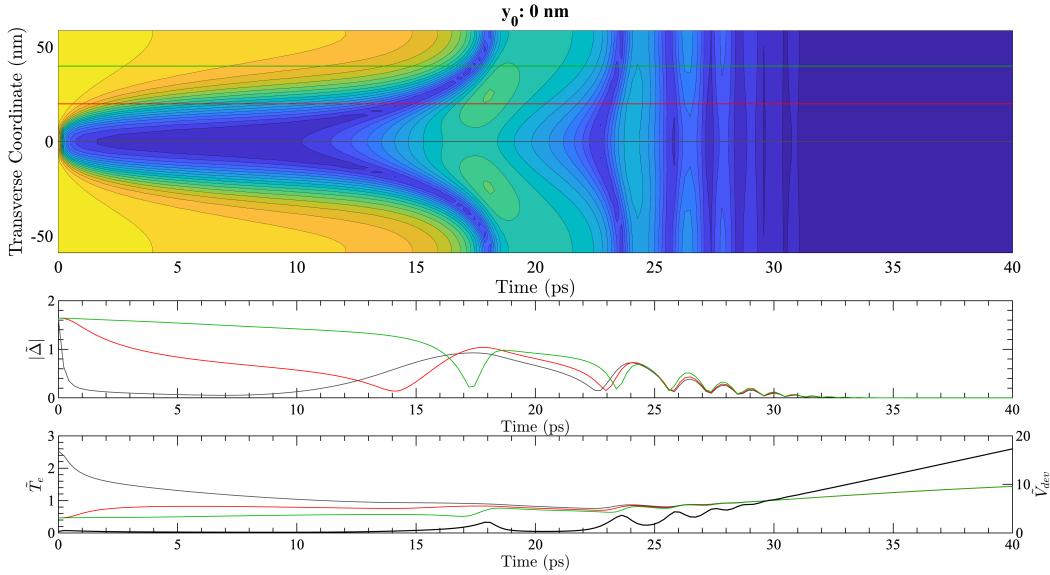
Figure 3.12: Overview of detection process for a photon absorbed at the center of a 120 nm wide NbN nanowire. The contour plot (top) shows the magnitude of the order parameter at $x = 0$ nm for all transverse coordinates y as a function of time. The grey ($y = 0$ nm), red ($y = 20$ nm), and green ($y = 40$ nm) lines indicate the coordinates of the slices shown for the order parameter (middle) and temperature (bottom) plots. The bottom plot also shows the voltage transient which occurs as a result of order parameter suppression.
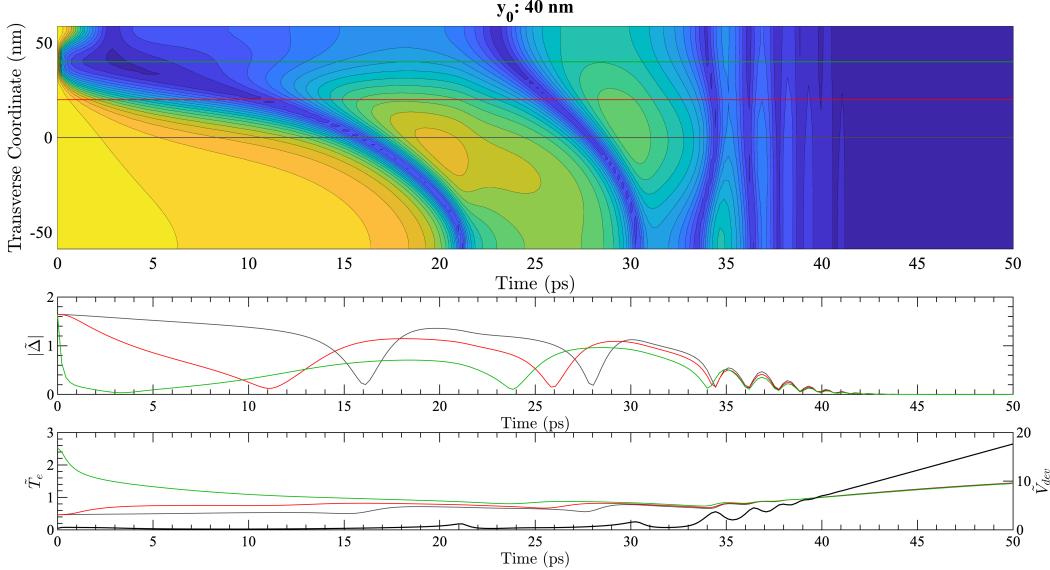
Figure 3.13: Overview of detection process for a photon absorbed near the edge of a 120 nm wide NbN nanowire. The contour plot (top) shows the magnitude of the order parameter at $x = 0$ nm for all transverse coordinates y as a function of time. The grey ($y = 0$ nm), red ($y = 20$ nm), and green ($y = 40$ nm) lines indicate the coordinates of the slices shown for the order parameter (middle) and temperature (bottom) plots. The bottom plot also shows the voltage transient which occurs as a result of order parameter suppression.

For a given bias current, we can set a voltage threshold which determines when the detector registers a 'click' due to photon absorption. By simulating the response of a nanowire for various initial hotspot transverse locations y_0 and energies $\bar{\chi} E_\lambda$, characteristic latency curves are generated, as shown in Figure 3.14 for the case of a 100 nm wide nanowire with the same material properties as listed above. This figure shows the detector response for bias currents of $10.5\,\mu\text{A}$ and $16.8\,\mu\text{A}$ for transverse coordinates ranging from 0 to 40 nm. Several qualitative features are immediately obvious. For a given bias current, the characteristic latency curves depend on both photon energy and the transverse coordinate of detection. At higher hotspot energies, detection events toward the edge of the wire have longer latency than those near the center. This will contribute to timing jitter because the delay before detection depends on the transverse coordinate of absorption, which is a probabilistic process. As the hotspot energy decreases to some lower threshold, the latency increases rapidly in an effect resembling a singularity. As the hotspot energy increases, the latency decreases monotonically with positive curvature and saturates at high photon energy. These features were predicted based on qualitative

arguments in Section 3.3. Finally, at the higher bias current shown in the figure, the energy threshold where the edge becomes photosensitive is lower than at the center, in contrast to the results at the lower bias current.
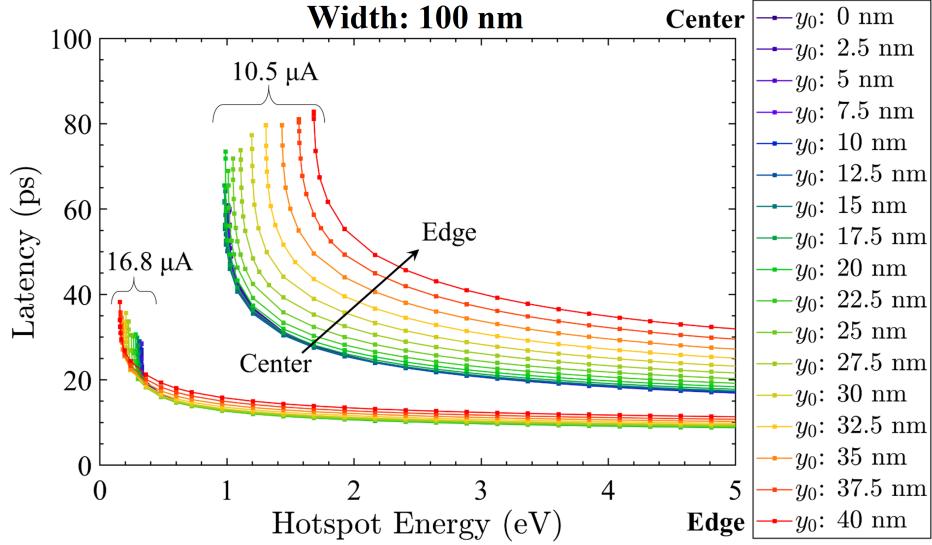
Figure 3.14: Simulated latency for bias currents of 10.5 µA (upper family of curves) and 16.8 µA (lower family of curves) for a 100 nm wide NbN nanowire.

By defining the detection energy $E_{det}(I_B, y_0)$ as the minimum energy required to generate a detection event, it is possible to quickly identify trends in the transverse coordinate dependence of detection properties, as shown in Figure 3.15. From this plot, there is a clear transition from a 'U' shaped detection profile at low bias currents, through a 'W' shape, and finally to a 'Bell' shape at high bias currents. This corresponds to the coordinate of minimum detection transitioning from the center of the nanowire at low bias currents to the edge at high bias currents.
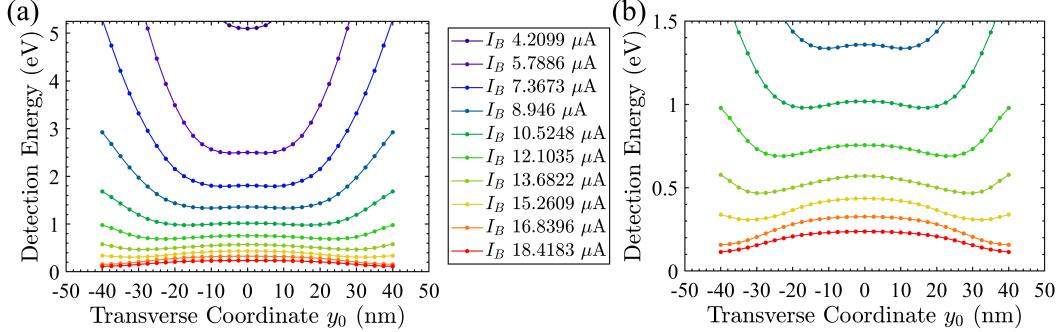
Figure 3.15: Detection energy as a function of transverse coordinate for a 100 nm wide wire at various bias currents. As the bias current increases, the detection energy decreases. A closer view (b) of the high bias current region shows the transition from a 'W' shape to a 'Bell' shape.

Comparison with Experiment

Using this model, we can attempt to fit experimental data from NbN devices designed to probe the intrinsic jitter and relative latency of the detection process. Measurements were performed at 4 K for the specific purpose of comparing to the model. Operating near $T_c/2$ is unfavorable for device performance, but is closer to the regime where the model is expected to be more representative of the device physics. The two fitting parameters $\bar{\chi}$ and σ_F account for variation of retained energy due to the downconversion process, but there is additional uncertainty in the fraction of depairing current reached experimentally. Recent work has demonstrated that the depairing current can be estimated by measuring the shift in kinetic inductance with increasing bias current for resonator structures [114], then fitting the result to a model based on the solution of the Usadel equations [76]. These measurements demonstrated that for thin films used for SNSPDs, the fraction of depairing current reached is typically between 0.5–0.8. Due to the non-dimensionalization of the problem, the fraction of the depairing current reached is modified through the parameter $\rho_\square$, which also influences E_0. However, the initial conditions are given in terms of T_{HS}/T_{sub}, so E_0 can be properly normalized when allowing $\rho_\square$ to vary as a free parameter. For the calculations shown here, $\rho_\square$ is estimated as $650\,\Omega/\square$, which corresponds to reaching 90% of the temperature-dependent depairing current at 1 K, which is reasonable given the short length of the device. This is consistent with the appropriate depairing current of resonator structures fabricated from the same superconducting film.

When PCR curves are fitted to the model, as shown in Figure 3.16a using the pa-

rameters $\bar{\chi} = 0.68$ and $\sigma_{F,1550} = 176$ meV, there are some immediate concerns. The energy-current dependence of the detector response does not match experimental results as evidenced by the failure of the model to match the spacing between the PCR curves of different photon energies. However, that is not the most glaring problem. The timing characteristics are off by a factor of $\sim 2 - 4$ both in terms of jitter (b) and relative latency (c). The fact that relative latency is not reproduced is significant because it should not be dependent on the main noise sources of the experiment.
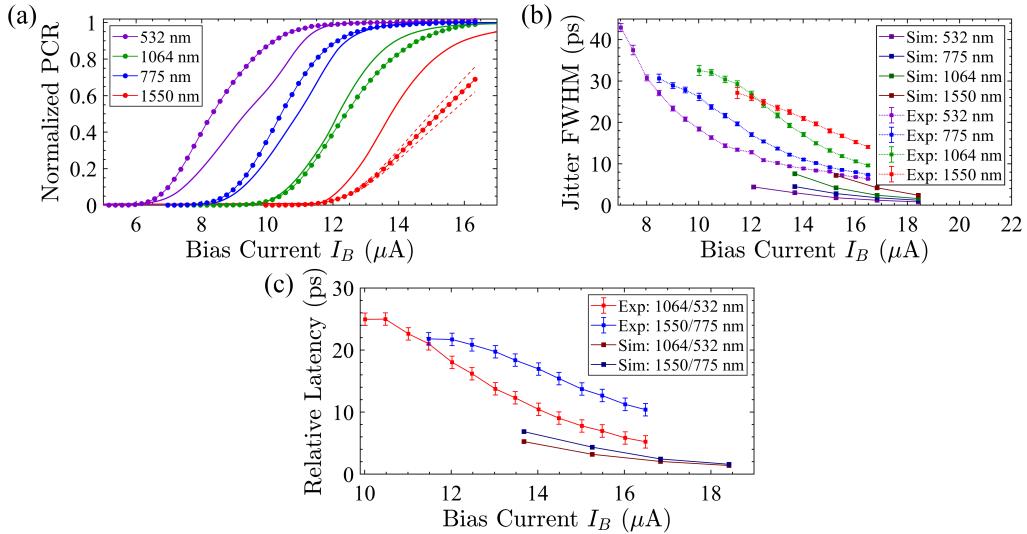
Figure 3.16: Comparison of the Vodolazov TDGL model with experiment for the case of a 100 nm wide NbN detector operating at 4 K. (a) The PCR results of the model (solid lines) are spaced closer together for the wavelengths of interest than the experimental results (circles with lines). The timing jitter (b) and relative latency (c) predicted by the model are significantly smaller than measured experimentally. This suggests that the model is not accurately capturing the timescale of the suppression of superconductivity.

If modification of the characteristic timescales τ_{ep} and τ_{esc} is allowed, justified by uncertainty in the material parameters, this changes the spacing of the PCR curves and the necessary $\bar{\chi}$ and σ_F needed to approximate the experimental results. If the phonon escape time is increased by an order of magnitude to 94 ps and $\rho_\square = 600\,\Omega/\square$ is used, a better fit to the experimental PCR curves can be achieved, as shown in Figure 3.17a, but this does not significantly change the relative latency as shown in Figure 3.17b. This suggests that the detector latency is dominated by the process of superconductivity suppression, not the details of the energy balance. Therefore, we

can conclude that the Vodolazov model [6] based on the modified TDGL equations is inadequate to describe experimental results at a semi-quantitative level.
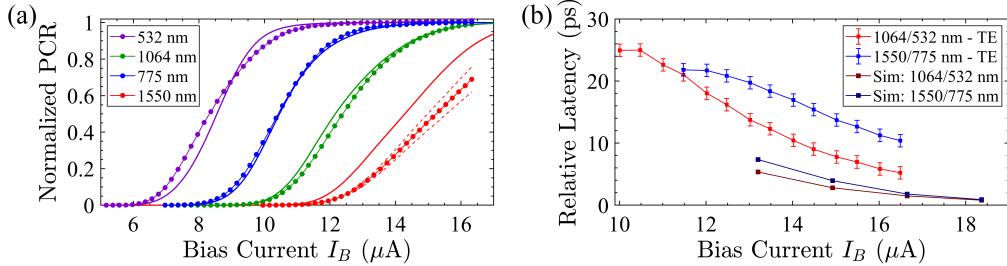
Figure 3.17: Comparison of the (a) PCR and (b) relative latency results of the Vodolazov TDGL model with experiment for the case of a 100 nm wide NbN detector operating at 4 K using a modified τ_{esc} of 94 ps. The fitting parameters are $\bar{\chi} = 0.44$ and $\sigma_F = 96$ meV.

Despite the quantitative failure of this model, there are a number of qualitative successes that suggest the foundation of this approach is a useful way of understanding the photodetection process. The model successfully predicts the increase in the width of the PCR transition when operated in magnetic fields [51], and also predicts an increase in timing jitter in magnetic fields [108]. Both of these features have been used as direct evidence of the 2D nature of the photodetection process, and the TDGL model naturally handles this effect through the asymmetric energy barrier to vortex or antivortex formation which occurs in magnetic fields.

As shown in Figure 3.16b, the model predicts decreasing timing jitter with increasing bias current and increasing photon energy, which is qualitatively consistent with experiment. Furthermore, the model predicts decreasing relative latency with increasing bias current and increasing photon energy, which is again consistent with experiment. These two results suggest that the general form of the latency curves produced by the TDGL model have the appropriate qualitative shape in order to understand experimental results.

The model also appears to capture the correct qualitative response to detections at different transverse coordinates y_0. The key metric for distinguishing this detail from experimental measurement is the shape and polarization dependence of the PCR curves at various wavelengths. Figure 3.18a shows the polarization-dependent PCR curves generated from the Vodolazov model over a wide range of deposited energies. The solid lines indicate the TE mode while the dashed lines indicate the TM mode. As the photon energy decreases, the PCR transition first becomes

sharper, then broadens at longer wavelengths. The relative polarization dependence also shifts from TM saturating first at high energies to TE saturating first at low energies. This is compared to Figure 3.18b which shows the experimental PCR curves for 532, 775, 1064, and 1550 nm light measured in a 120 nm wide device at 1 K.[4] The same PCR shapes emerge as the bias current increases. At low bias currents, the transition is sharper, with the TE and TM modes having approximately the same transition, similar to the curves at around 12 μA in the simulation. As the bias current increases, the shape transitions from a sigmoidal shape resembling an error function to a broader transition with inflection points throughout the transition. This exactly resembles the shapes shown for the 19 μA bias range within the model. The simulations showing this shift in polarization dependence can only be performed at high bath temperatures, so a direct comparison is not possible, but the correct trend in polarization dependence demonstrates that the model captures the correct qualitative trend in the transverse coordinate detection efficiency.
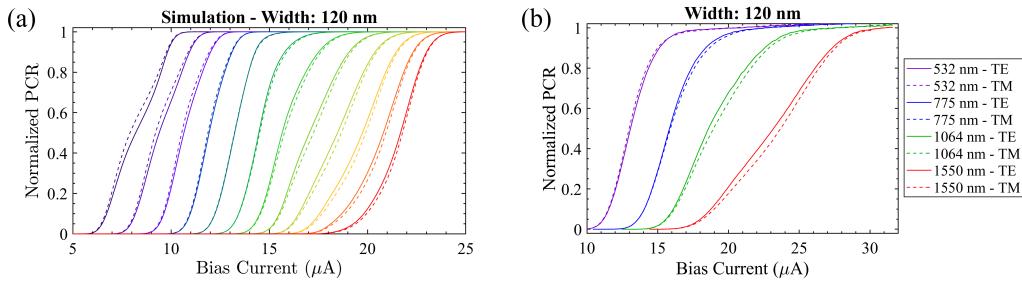
Figure 3.18: Comparison of polarization-dependent response of the Vodolazov TDGL model (a) with experiment (b) for the case of a 120 nm wide NbN detector operating at 4 K. Solid lines indicate the TE polarization while dashed lines are the TM polarization. The simulation parameters are $\gamma = 40$ and $\tau_{esc} = 30$ ps.

While this model captures a number of features of the SNSPD detection process, it is also important to emphasize its shortcomings. As stated earlier, the TDGL equations can only be justified from microscopic theory in the limits of $T \to T_c$, $\Delta \to 0$, small spatial gradients, and slow variation in time. These conditions are not met during the photon-detection process in SNSPDs. Furthermore, the initial conditions used to describe the photon energy are not rigorously justified by the kinetic equations describing the downconversion process. As shown in Chapter 2, the downconversion process does not lead to thermalized electron and

---

[4]For a wider range of photon energies without polarization dependence, see Figure 4b in [2]. From there, it is possible to observe wide PCR transitions at low bias currents and high photon energies, as observed in simulation.

phonon systems which can be well described by equilibrium temperatures. Despite the modifications to the TDGL equations, the current formulation is only able to simulate photoresponse at temperatures above $\sim 0.35\, T_c$. Below this temperature, the equations become unstable. This occurs because the immediate formation of a hotspot leads to an electric field which can be aligned with the supercurrent in some locations. Based on the Joule heating term of the energy balance equation, this actually leads to a cooling effect which drives the temperature to 0, causing a breakdown of the model. Such a failure hints at a problem in the formulation of the energy balance equation and its coupling to the TDGL equations. The overall qualitative success of the model suggests that if the shortcomings can be overcome through various corrections, a detection model based on the TDGL equations could be found to semi-quantitatively predict the main features of SNSPD operation. Such a tool would be extremely useful in optimizing device design and understanding the fundamental and practical limits of detector performance. Finally, we must remember that the current model is purely deterministic and is not sufficient for understanding intrinsic dark counts, or photon counts in the probabilistic regime.

3.4.2 Corrections to the Vodolazov Model

It is clear from the application of the Vodolazov model that substantial modification is required in order for the model to predict the timescales of detection observed in experiments. Furthermore, there are indications that the description of the energy balance equation contains some form of inconsistency due to the numerical instability observed at low bath temperatures. This section describes various corrections we applied to the Vodolazov model in an effort to improve the quantitative accuracy of the model.

Electron Energy Functional

The internal energy of the electron system is given by equations (3.10) and (3.11) within the Vodolazov model. These are well known and can be derived from the kinetic equations within the BCS framework while using the spectral functions from the Usadel equation in equilibrium. However, they neglect information about the non-equilibrium evolution of the system and can lead to significant deviations from the kinetic formulation.

Following [6], the kinetic equation describing the evolution of the electron distribu-

tion function $n(\epsilon)$ is given by

$$N_1 \frac{\partial n}{\partial t} = D\nabla \left(\left(N_1^2 - R_2^2 \right) \cdot \nabla n \right) - R_2 \frac{\partial n}{\partial \epsilon} \frac{\partial |\Delta|}{\partial t} + I_{ep}(n, N) + I_{ee}(n). \qquad (3.19)$$

Within the two-temperature model, we assume that the electron-electron interactions are instantaneous, so the electron distribution takes the form of an equilibrium distribution at the temperature T_e: $n = 1/\left(e^{(\epsilon/k_B T_e)} + 1 \right)$. This eliminates the electron-electron collision integral. Solution of the Usadel equations for in the absence of bias current defines the spectral functions N_1 (3.11) and

$$R_2 \left(\tilde{\epsilon} \right) = \frac{|\Delta|/k_B T_c}{\sqrt{\tilde{\epsilon}^2 - (|\Delta|/k_B T_c)^2}} \Theta \left(\tilde{\epsilon} - (|\Delta|/k_B T_c) \right), \qquad (3.20)$$

with $\tilde{\epsilon} = \epsilon/k_B T_c$. By integrating the resulting kinetic equation over energy and multiplying by the density of states, we arrive at the simplified energy balance equation

$$4N(0) \left\{ \int_0^\infty d\epsilon \; \epsilon \left[N_1 \frac{\partial n}{\partial t} - D\nabla \left(\left(N_1^2 - R_2^2 \right) \cdot \nabla n \right) + R_2 \frac{\partial n}{\partial \epsilon} \frac{\partial |\Delta|}{\partial t} - I_{ep}(n, N) \right] \right\}$$
$$= 0. \qquad (3.21)$$

The factor of $4N(0)$ is a result of a factor of $2N(0)$ to account for the full electron density of states at the Fermi level due to two electron spins and the additional factor of 2 comes from exploiting the symmetry about $\epsilon = 0$ to simplify the lower bound of integration from $-\infty$ to 0. By rearranging the terms of this equation, we arrive at a form

$$4N(0) \int_0^\infty d\epsilon \; \epsilon \left[N_1 \frac{\partial n}{\partial t} + R_2 \frac{\partial n}{\partial \epsilon} \frac{\partial |\Delta|}{\partial t} \right] =$$
$$4N(0) \left\{ \int_0^\infty d\epsilon \; \epsilon \left[D\nabla \left(\left(N_1^2 - R_2^2 \right) \cdot \nabla n \right) + I_{ep}(n, N) \right] \right\} \qquad (3.22)$$

where each term can be recognized by its counterpart in (3.7). The left hand side represents the rate of change of the energy of the electron system, the first term on the right describes diffusion, and the second term on the right describes electron-phonon coupling. The Joule heating term was not explicitly included in the kinetic equation and must be reinstated. Focusing on the left hand side of the equation, we apply the chain rule and separate the two terms into temperature and order parameter derivatives with time.

$$4N(0) \int_0^\infty d\epsilon \; \epsilon \left[N_1 \frac{\partial n}{\partial t} + R_2 \frac{\partial n}{\partial \epsilon} \frac{\partial |\Delta|}{\partial t} \right] =$$
$$4N(0) \left\{ \left[\int_0^\infty d\epsilon \; \epsilon N_1 \frac{\partial n}{\partial T} \right] \frac{\partial T}{\partial t} + \left[\int_0^\infty d\epsilon \; \epsilon R_2 \frac{\partial n}{\partial \epsilon} \right] \frac{\partial |\Delta|}{\partial t} \right\} \qquad (3.23)$$

Each of these two integrals can be computed numerically. It can be shown that within the BCS framework, (3.23) is equal to the left hand size of (3.7) for $|\Delta| = |\Delta|_{BCS}(T_e)$. However, when the system is not in equilibrium, the integral preceding the $\frac{\partial |\Delta|}{\partial t}$ term can be significantly different.

We argue that the form based on the kinetic equation is more general than that using the BCS self-consistency equation. There is no reason to expect that the BCS self-consistency relation should hold when the system is quickly perturbed from the equilibrium state following photon absorption. Use of (3.23) in the energy balance equation leads to non-negligible changes in the response of the system during simulation. The detection energy shifts as a results of the new formulation, which has immediate implications for fitting experimental data, but the revised energy balance equation does not solve the problem of the timescale of detection.

Generalized TDGL Equations

The principle failure of the standard TDGL model is its inability to predict the long latency observed experimentally through the relative latency measurements of synchronized photons of different energies. The relative latency predicted by the model is approximately a factor of 2–4 smaller than that observed experimentally. Therefore, we need to alter the equations governing the evolution of the superconducting state to account for this slower suppression of superconductivity. While we could insert fitting parameters to simply slow down this suppression of superconductivity, such a modification would not be justified from the microscopic model. Instead, we seek a refined model derived from the kinetic equations which has the effect of slowing the suppression of superconductivity.

The generalized TDGL equations fit this description. Introduced in 1981 by Watts-Tobin et al. [115], the generalized TDGL equations are a formulation of non-equilibrium superconductivity which is derived from the kinetic equations while relaxing some of the assumptions of the standard TDGL equations. Following the derivation of Kopnin [116], in a dirty superconductor with strong impurity scattering (mean free path $l \ll \xi_0$) and in the limits of $T \to T_c$ and small gradients in time and space ($\omega, Dk^2 \ll \tau_{sc}^{-1}$ where τ_{sc} is the characteristic inelastic scattering time), the kinetic equations take the form of a generalized TDGL equation [116]. Note that unlike the standard TDGL equations, the generalized equations do not require a gapless superconductor ($\Delta \tau_m \ll 1$ where τ_m is the magnetic impurity scattering time), making the equations better suited to handle the case of photon detection

where the system initially maintains a large gap. Following Vodolazov's approach [6], we modify the generalized TDGL equation with the parameters ξ_{mod} and Δ_{mod} in order to better represent the temperature dependence of the system away from T_c, and introduce the supercurrent correction term as done in (3.12), leading to

$$\frac{\pi\hbar}{8k_BT_c}\left(\varrho\left(T_e\right)\frac{\partial}{\partial t}\left|\Delta\right| + \frac{i\left|\Delta\right|}{\varrho\left(T_e\right)}\frac{\partial}{\partial t}\phi + \frac{2ie\left|\Delta\right|}{\varrho\left(T_e\right)\hbar}\varphi\right) =$$

$$\xi_{mod}\left(T_e\right)^2\left(\nabla + i\left(\nabla\phi - \frac{2e}{\hbar c}\vec{A}\right)\right)^2\left|\Delta\right| + \left(1 - \frac{T_e}{T_c} - \frac{\left|\Delta\right|^2}{\Delta_{mod}^2\left(T_e\right)}\right)\left|\Delta\right|$$

$$+i\frac{\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)}{\left|\Delta\right|}\frac{\hbar eD}{\sigma_n\sqrt{2}\sqrt{1 + T_e/T_c}}.$$

$$(3.24)$$

The parameter $\varrho\left(T_e\right) = \sqrt{1 + 4\left|\Delta\right|^2\tau_{sc}\left(T_e\right)^2/\hbar^2}$ modifies the rates of order parameter magnitude and phase evolution in the generalized TDGL equation. The inelastic scattering rate τ_{sc} incorporates both electron-electron and electron-phonon interactions according to $\tau_{sc}\left(T_e\right) = 1/\left(1/\tau_{ee}\left(T_e\right) + 1/\tau_{ep}\left(T_e\right)\right)$ where $\tau_{ee}\left(T_e\right)$ and $\tau_{ep}\left(T_e\right)$ are the electron-electron and electron-phonon inelastic scattering times respectively. The temperature dependence of these scattering rates is approximated by $\tau_{ee}\left(T_e\right) = \tau_{ee}\left(T_c\right)T_c/T_e$ and $\tau_{ep}\left(T_e\right) = \tau_{ep}\left(T_c\right)\left(T_c/T_e\right)^3$. In the limit of infinite scattering, $\tau_{ee}(T_c) \to 0$, $\tau_{sc} \to 0$, $\varrho \to 1$, and the generalized TDGL equation simplifies to the standard TDGL equation. The energy balance equations, current conservation equation, and boundary conditions remain the same as the standard TDGL formulation. As with the standard TDGL equations, the form of the equation must be altered in order to make computation feasible without encountering singularities. This form, in terms of the real and imaginary components of the order parameter, is derived in Appendix B.3.

The introduction of the generalized TDGL equation leads to a new fitting parameter $\tau_{ee}(T_c)$, which directly impacts the rate of order parameter evolution through $\varrho(T_e)$. To understand the impact of this parameter, we first model the system response in 1D while attempting to achieve the appropriate relative latency and timing jitter for narrow nanowires. In a recent publication, we show that the choice of $\tau_{ee}(T_c)$ dramatically changes the timing response of the system, and with it, the jitter and relative latency [3]. For these initial calculations, we used Vodolazov's form of the electron energy balance equation. In its 1D form, the initial condition takes the form of a 'hotbelt' where the electron and phonon systems are thermalized to a temperature T_{HB} across the width of the nanowire. To model the NbN films measured

by Korzh et al. [2], we use an 80 nm wide wire with $\rho_\square = 608\,\Omega/\square$, $T_c = 8.65\,\text{K}$, $\tau_{esc} = 20\,\text{ps}$, $\tau_{ep}(T_c) = 24.7\,\text{ps}$, and $\gamma = 60$. The initial hotbelt length was 40 nm and calculations were performed at a bath temperature of $T_{sub} = 2\,\text{K}$.[5] Using these parameters, we vary $\tau_{ee}(T_c)$ while adjusting $\bar{\chi}$ and σ_F to maintain appropriate fitting with the experimentally measured PCR curves. The introduction of the generalized TDGL equation has a significant impact on the response of the system. As $\tau_{ee}(T_c)$ increases, the nature of the voltage transient changes, as shown in Figures 3.19a and 3.19b. The modification of the rates of order parameter magnitude and phase evolution means that the phase of the order parameter can change more rapidly than the magnitude when compared to the limit of $\tau_{ee}(T_c) \to 0$. The result can be seen in the voltage traces in two ways. First, the time before the voltage transient appears is much longer in the case of non-zero $\tau_{ee}(T_c)$ due to the slower suppression of the magnitude of the order parameter. Once the superconductor can no longer support the current flowing through the nanowire as supercurrent, the phase-slip lines rapidly develop, leading to a sharp increase in voltage across the device and rapid oscillations in the voltage. The delay required for voltage to form for $\tau_{ee}(T_c)$ is directly tied to an increase in the detection energy for the nanowire, as shown in Figure 3.19c. This is intuitive because during the latency period, the photon energy diffuses away from the location of absorption, and simultaneously couples to the substrate, reducing the amount of energy which contributes to suppressing superconductivity.

---

[5]In the 1D geometry, the equations remain stable down to lower temperatures. The flow of current through the localized hotspot in the 2D case exacerbates the instability of the equations.
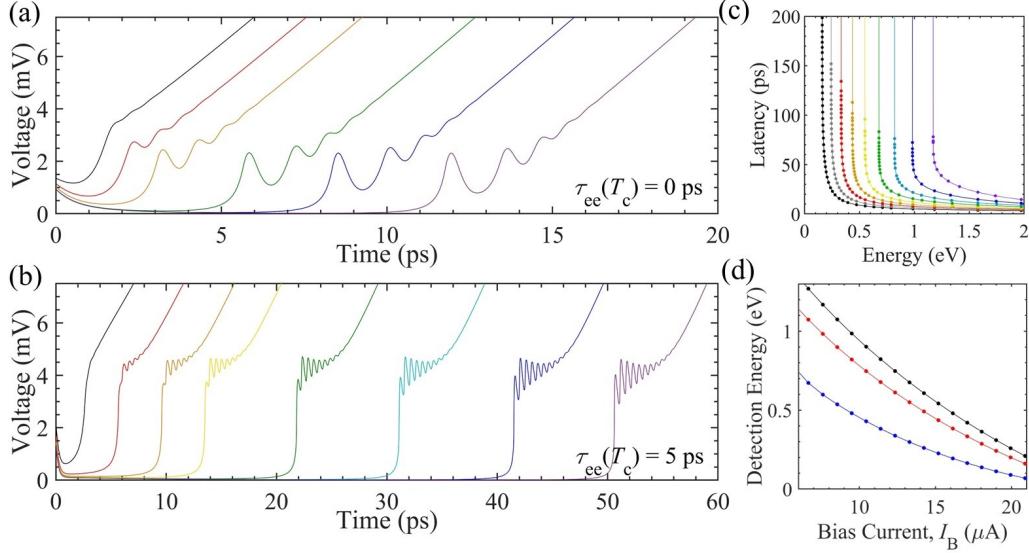
Figure 3.19: Example voltage transients for the (a) standard TDGL and (b) generalized TDGL models at a bias current of 12.35 μA in response to different amounts of deposited energy. For $\tau_{ee}(T_c) = 0$ ps, the deposited energies are 0.6316 eV (black), 0.4737 eV (red), 0.3947 eV (orange), 0.3454 eV (green), 0.3392 eV (blue), and 0.3385 eV (violet) while for $\tau_{ee}(T_c) = 5$ ps, the deposited energies are 1.5789 eV (black), 0.9868 eV (red), 0.7895 eV (orange), 0.7105 eV (yellow), 0.6447 eV (green), 0.6217 eV (teal), 0.6143 eV (blue), and 0.6127 eV (violet). (c) Latency curves for the case of $\tau_{ee}(T_c) = 5$ ps at bias currents of 5.7 μA (violet), 7.6 μA (blue), 9.5 μA (teal), 11.4 μA (green), 13.3 μA (yellow), 15.2 μA (orange), 17.1 μA (red), 19.0 μA (grey), and 20.9 μA (black). (d) Detection energy for $\tau_{ee}(T_c) = 0$ ps (blue), $\tau_{ee}(T_c) = 5$ ps (red), and $\tau_{ee}(T_c) = 10$ ps (black). Increased $\tau_{ee}(T_c)$ leads to a larger detection energy for a given bias current.

Using these results, we can make a qualitative comparison of the predictions of the simplified 1D model with the experimental results of [2]. The fitting parameters $\bar{\chi}$, σ_F, and σ_{NU} were selected to approximately match the PCR curves for 1550 nm and 775 nm light.[6] The values of $\bar{\chi}$ are 0.37, 0.65, and 0.79, σ_{Fano} are 0.064, 0.088, and 0.104 eV, and σ_{NU} are 0.024, 0.040, and 0.048 eV for $\tau_{ee}(T_c)$ of 0, 5, and 10 ps, respectively. A summary of the results is shown in Figure 3.20, which shows the

---

[6]The fitting procedure introduced an additional parameter σ_{NU} which represents fluctuations in energy due to changes in the material properties of the system with $\sigma_{NU} \sim E_\lambda$. This addition actually reduced the quality of the fitting for higher energy photons, but was necessary to avoid a sharp cutoff in the energy distribution function at E_λ. Fluctuations due to the Fano contribution cannot exceed E_λ because the energy distribution cannot exceed the photon energy, but fluctuations due to non-uniformity are not subject to the same restriction. The fitting for $\tau_{ee}(T_c) = 10$ ps ignored this constraint on the Fano fluctuation contribution in order to provide cleaner insight into the importance of $\tau_{ee}(T_c)$.

(a) PCR curves, (b) jitter histograms for $\tau_{ee}(T_c) = 5$ ps, (c) jitter FWHM, and (d) relative latency for these two wavelengths. There are several features which emerge from these results. When fitting to a set of PCR curves, increasing $\tau_{ee}(T_c)$ tends to separate the curves because more energy is lost during the detection process when the suppression of the order parameter is slower. This effect is most noticeable at low values of the bias current where Joule heating is smaller. The relative latency is directly correlated with $\tau_{ee}(T_c)$ as expected based on the form of the generalized TDGL equation. The value of $\tau_{ee}(T_c) = 5$ ps captures the appropriate scale of the relative latency for the pair of wavelengths. Simultaneously, this set of parameters also captures the correct order of magnitude for the timing jitter (see Figure 3 in [2] for a direct comparison).
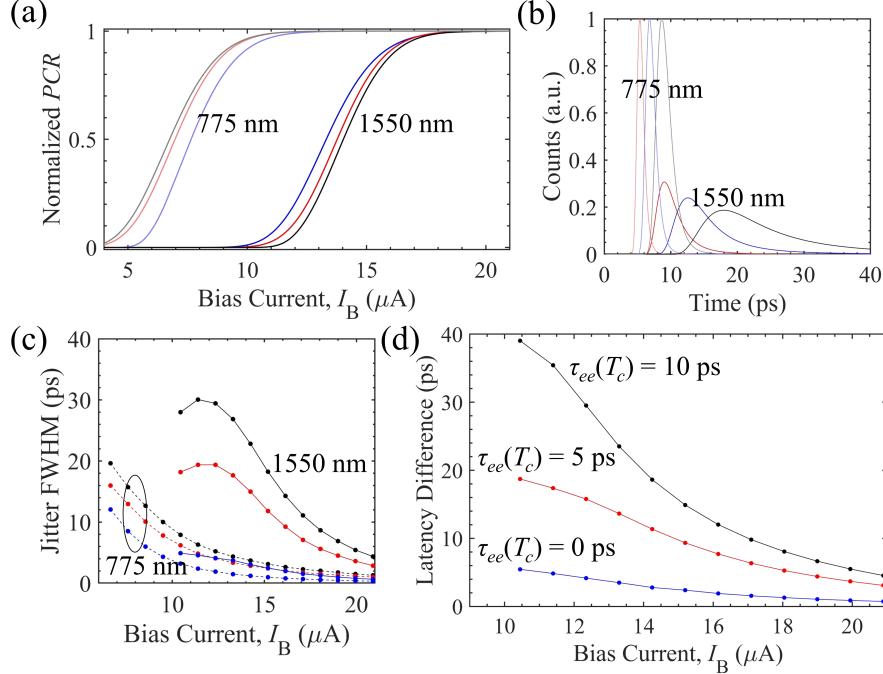
Figure 3.20: Detection characteristics based on the 1D model. (a) PCR curves for 775 nm (light) and 1550 nm (dark) photons for τ_{ee} (T_c) of 0 ps (blue), 5 ps (red), and 10 ps (black). The experimental PCR curves of [2] reasonably match the blue curves. (b) IRF for bias currents of 15.2 μA (black), 17.6 μA (blue), and 20.0 μA (red) for τ_{ee} (T_c) of 5 ps for 775 nm (light) and 1550 nm (dark) photon wavelengths. For each bias current, the 775 nm histogram is normalized to a unit maximum while the corresponding 1550 nm histogram is subsequently normalized to have the same area. (c) Jitter FWHM vs. bias current for 1550 nm (solid) and 775 nm (dashed) photon energies. The results are shown for τ_{ee} (T_c) values of 0 ps (blue), 5 ps (red), and 10 ps (black). (d) Latency difference between photons of 1550 and 775 nm wavelength for τ_{ee} (T_c) of 0 ps (blue), 5 ps (red), and 10 ps (black). Increasing τ_{ee} (T_c) leads to larger relative latency difference.

While the 1D model shows the correct qualitative behavior, its simplified geometry cannot be considered an accurate representation of the physics of the SNSPD detection process. That requires the use of a 2D model. In particular, understanding the polarization dependence of detector response requires a 2D formulation, and the 1D approximate fitting breaks down for a wider range of photon energies.

Using a similar set of material parameters with $T_c = 8.65$ K, $D = 0.5\,\text{cm}^2/\text{s}$, $\rho_\square = 600\,\Omega/\square$, $\tau_{ee}(T_c) = 5$ ps, and $\gamma = 60$, we apply the 2D model and attempt to fit the polarization-dependent experimental results at 4 K. We also use the kinetic form of the electron energy balance equation described in Section 3.4.2. For the 80 nm

wide device, the PCR curves and relative latency are shown in Figure 3.21 for the accessible wavelength range of the experimental setup using the fitting parameters $\bar{\chi} = 0.57$ and $\sigma_{F,1550} = 96\,\text{meV}$. In order to fit the experimental data, a larger escape time of $\tau_{esc} = 80\,\text{ps}$ is required. Use of a shorter escape time compresses the spacing of the PCR curves and therefore fails to reproduce the energy-current dependence of the system. As with the 1D model, the relative latency is more appropriately reproduced by the generalized TDGL model than the standard TDGL model.
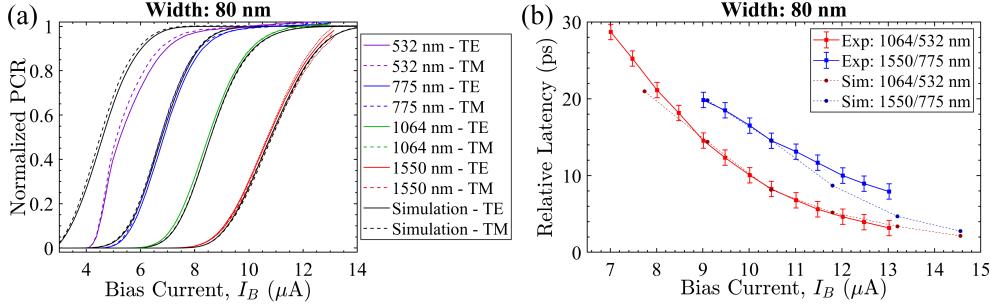
Figure 3.21: Comparison of the (a) PCR and (b) relative latency of the generalized TDGL model and experiment. For the PCR curves, the colored lines are the experimental data and the black lines are the model fit. The dotted red curves in the PCR plot show the approximate error bounds of the PCR curves for the 1550 nm wavelength. Solid lines are the TE polarization and dashed are the TM polarization. For the relative latency results, the solid lines are the experimental data while the dashed lines are the model fit. Both simulations use the same set of material and fitting parameters.

In order for this approach to be valid, the fitting parameters used for one nanowire width should be similar to those needed for other nanowire widths with the same material parameters. As a comparison, we apply the TDGL model to 100 nm and 120 nm wide wires as shown in Figure 3.22. The fitting parameters are $\bar{\chi} = 0.56$ and $\sigma_{F,1550} = 128\,\text{meV}$ for the 100 nm wide nanowire and $\bar{\chi} = 0.53$ and $\sigma_{F,1550} = 144\,\text{meV}$ for the 120 nm wide device. The fit quality suffers for longer wavelength photons at higher bias currents, but the overall results are promising. The fraction of energy retained by the nanowire stays nearly constant, which suggests that the model captures the main physics of the detection process. If we estimate the expected size of fluctuations due to energy escape based on Equation 2.51 of Chapter 2, for $\bar{\chi} = 0.57$, $\sigma_{F,1550} = 92\,\text{meV}$ which is the right order of magnitude for the 80 nm wide wire. However, there are clear limitations. The magnitude of the fluctuations as encoded in fitting parameter $\sigma_{F,1550}$ increases substantially as the width increases. This fitting should not change with nanowire width, so this is an indication that the

model is not appropriately capturing how the detection process scales with wider widths.
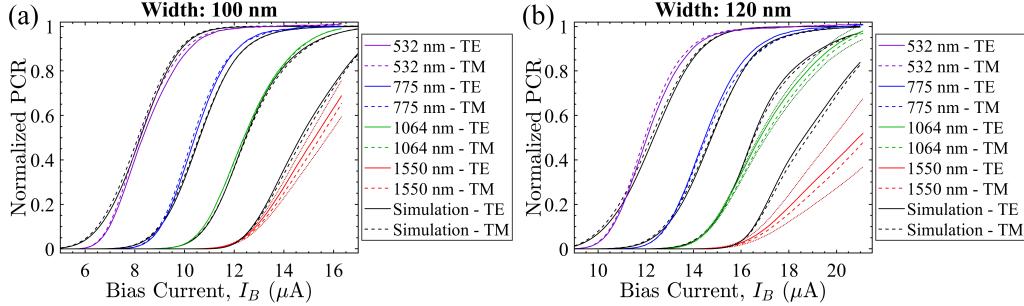
Figure 3.22: Comparison of simulated and experimental PCR curves for device widths of (a) 100 nm and (b) 120 nm. The colored curves are the experimental measurements while the black curves are the fit using the model. The dotted red and green curves show the approximate error bounds of the PCR curves for the 1550 nm and 1064 nm wavelengths, respectively. Solid lines indicate the TE polarization while dashed lines are the TM polarization.

The generalized TDGL model with the kinetic form of the electron energy balance equation shows great promise as a model of the detection mechanism, but there are several features which limit its quantitative accuracy. First, the material properties required to fit the experimental data are extremely favorable for photon detection. The use of $\gamma = 60$ is higher than most estimates of this parameter, which leads to a greater fraction of the photon energy contributing to breaking superconductivity in the electronic system. Additionally, the escape time of 80 ps is several factors larger than estimates based on the acoustic mismatch model or diffuse mismatch model, and larger than values extracted from experimental measurements on thin films [89]. Furthermore, while the model shows indications of the correct trends of polarization dependence in terms of which polarization saturates first, the model predicts a smaller difference between the polarizations than seen experimentally. This means that the model underestimates the importance of transverse coordinate effects. Larger amounts of fluctuations (σ_F) are required to fit the experimental data for wider nanowires, further suggesting that the importance of transverse coordinate effects is underestimated. Part of this comes from the use of a large escape time, which traps the phonon energy and allows more heat to diffuse through the electron channel. In this regime where diffusion dominates over coupling to the substrate, the transverse coordinate of absorption plays a reduced role in determining the detection energy at a given bias current because changes in the coupling of the electron system

to the phonon system do not significantly alter the total energy retained by the system over the latency period. This can be important if the energy is deposited near the edge of the nanowire where heat cannot diffuse through the boundary, leading to a higher electron temperature and stronger coupling to the phonon system where energy can escape to the substrate. Overall, this formulation of the generalized TDGL detection model is a significant step toward a semi-quantitative model of photon detection, but still has shortcomings which must be addressed before it can be considered accurate.

Initial Conditions: Modified Phonon Bubble

Within the TDGL model, the initial conditions of the hotspot dictate the response of the system for a given photon energy and bias current. Therefore, it is crucial that this energy deposition is a realistic representation of the downconversion process in SNSPDs. Chapter 2 was dedicated to investigating the downconversion process based on the solution of the kinetic equations within a fixed hotspot volume, with the intent of determining appropriate initial conditions for TDGL modeling and estimating energy losses due to the escape of high-energy phonons. Here, we summarize the conclusions and show the implications for TDGL modeling.

Based on the Chapter 2 calculations, the use of a thermalized hotspot defined by $T_e = T_{ph} = T_{HS}$ has some limitations. The fluctuation model based on the escape of high-energy phonons introduces the fitting parameters $\bar{\chi}$ and σ_F which modify the total energy deposited in the nanowire. In the thermalized hotspot model, this energy loss occurs only during the time it takes for the hotspot to initially form. However, a significant fraction of this energy is associated with the phonon system. These phonons then undergo escape based on the two-temperature model, increasing the total energy loss associated with high-energy phonons. Calculations using the kinetic model show that the electron and phonon systems do not reach an equilibrated temperature instantly. It takes several factors of $\tau_{|\Delta|}$ for the effective temperature of the electron system to reach that of the phonons. Furthermore, in the absence of enhanced electron-electron interactions, the kinetic model predicts that the electron and phonon systems cannot be described by their equilibrium distributions on timescales comparable to $\tau_{|\Delta|}$ as is assumed by the two-temperature model, so an equilibrated hotspot does not accurately represent the partitioning of energy between electrons and phonons during the downconversion process.

In an effort to improve the connection between downconversion and TDGL model-

ing, we introduce a modified form of the phonon bubble as a more appropriate initial condition for the current simulations. As discussed in the downconversion chapter, the use of the phonon bubble accurately reproduces the downconversion behavior on the timescales relevant to suppression of superconductivity. Our modified phonon bubble takes the form of a heat source for the electron system within the hotspot volume. This introduces the term

$$\frac{\partial E_{e-ph,Ph.B}}{\partial t} = \frac{E}{V_{HS}\tau_{DC}} \exp\left(-t/\tau_{DC}\right) \tag{3.25}$$

to the right hand side of (3.7) for the volume within the hotspot. This term corresponds to an exponential decay of high-energy phonons interacting with the electron system over a uniform volume, with the timescale $\tau_{DC} = \frac{4}{3}\frac{\tau_0}{\gamma\Omega_D} \sim 1.4\,\text{ps}$ coming from the average phonon-electron interaction time for the phonon plateau (see Section 2.4).

Implementing the phonon bubble initial condition leads to immediate quantitative changes to the response of the system within the generalized TDGL model. As expected based on the arguments above, the sensitivity of the system increases for a given amount of deposited energy E. The detection energy for the 80 nm wide wire with $\gamma = 20$, $\tau_{esc} = 9.4\,\text{ps}$, and $\rho_\square = 600\,\Omega/\square$ is shown in Figure 3.23. The detection energy for the thermalized hotspot is significantly higher than that for the phonon bubble initial condition for all bias currents and transverse coordinates.
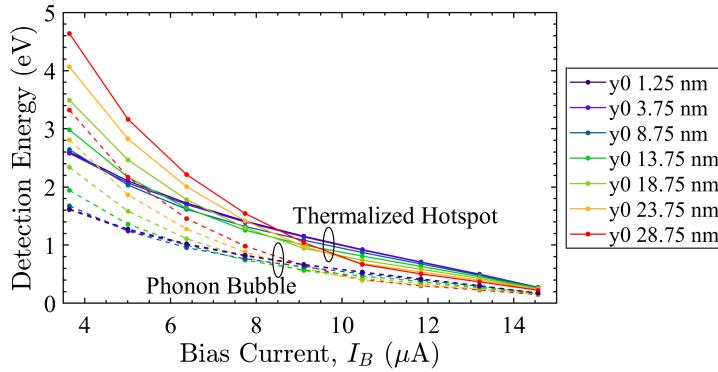
Figure 3.23: Comparison of the detection energy for the thermalized hotspot (solid lines) and modified phonon bubble (dashed lines) initial conditions for an 80 nm wide wire with the material parameters $\tau_{esc} = 9.4\,\text{ps}$, $\gamma = 20$, and $\tau_{ee}(T_c) = 5\,\text{ps}$.

The increased sensitivity using the phonon bubble initial condition allows for a more realistic fit to experimental data using the theoretical estimates of the escape time

$\tau_{esc} = 9.4$ ps and phonon heat capacity parameter $\gamma = 20$. This fit is shown in Figure 3.24a using the free parameters $\bar{\chi} = 0.61$ and $\sigma_{F,1550} = 128$ meV while maintaining $\tau_{ee}(T_c) = 5$ ps and $T_{sub} = 4$ K. As a comparison, the same simulation is performed while using the thermalized hotspot initial condition using the fitting parameters $\bar{\chi} = 0.92$ and $\sigma_{F,1550} = 176$ meV with the results shown in Figure 3.24b. Use of the more realistic material parameters does not lead to a high quality fit for either initial condition. Both fits suffer from problems in the shape of the PCR curves for 532 nm photons and the spacing between the simulated PCR curves does not match the spacing between the experimentally measured curves. Nevertheless, the result is still informative. The use of the modified phonon bubble initial condition enables fitting with the parameter $\bar{\chi} = 0.61$, which is plausible based on the calculations of Chapter 2. This also means that the distribution of energies (defined by $\bar{\chi}$ and $\sigma_{F,1550}$) required to fit the experimental data remains below E_λ as needed for a physically reasonable result. In contrast, the use of the thermalized hotspot initial condition leads to $\bar{\chi} = 0.92$ and the energy distribution exceeds E_λ. The thermalized hotspot fits shown here ignore the E_λ constraint on the energy distribution in order to allow for a comparison between the two models, but this is a clear indication that the thermalized hotspot initial condition does not capture the appropriate energy loss in the system when using literature estimates of the NbN material parameters.
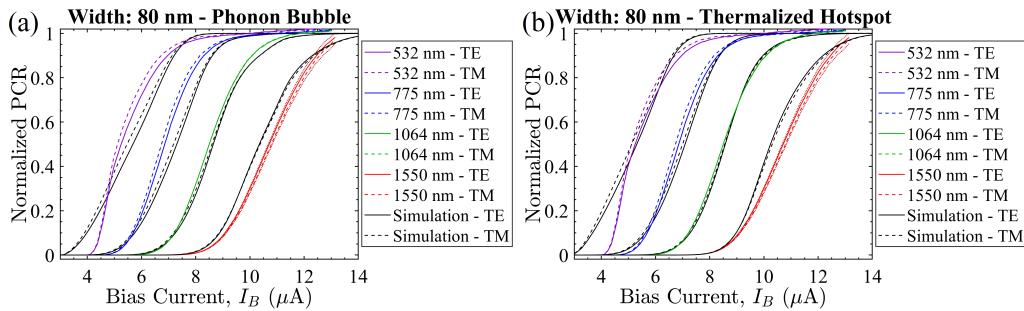
Figure 3.24: Comparison of model fitting for (a) phonon bubble initial conditions and (b) thermalized hotspot initial conditions.

Initial Conditions: Phonon Bubble Timescale

When considering the phonon bubble initial condition in Section 2.4, calculations based on the two-temperature model showed that the assumption of thermalized electrons leads to a higher concentration of energy in the electron system than can be expected based on the solution of the kinetic equations. This increased energy density alters the evolution of the two-temperature system compared to the kinetic

system, which will change the response of the TDGL system. We explored the possibility of modifying τ_{DC} in order to more appropriately match the energy density of the electron system for the two-temperature model compared to the kinetic model, finding that increasing τ_{DC} by a factor of ~ 4 leads to a significant improvement. This modification was motivated by the understanding that the energy density of the electron system directly impacts the evolution of the superconducting state within the TDGL formulation while the phonon system does not. To check the influence of this parameter, the response of an 80 nm wide device was modeled using various values of τ_{DC} in order to gauge the importance of this parameter. The detection energy as a function of transverse coordinate is shown in Figure 3.25 for a range of values for the phonon bubble time constant τ_{DC}. Increasing τ_{DC} initially leads to a small increase in sensitivity in the center of the nanowire until τ_{DC} is increased by a factor of ~ 3 compared to the nominal value of $\tau_{DC} = \frac{4}{3}\frac{\tau_0}{\gamma\Omega_D}$. This occurs because energy is transferred more slowly to the electron system. Energy diffuses through the electron channel during this time, leading to a lower electron energy density spread over a wider area. The lower electron energy density couples less energy to the phonon system, so less energy is subsequently lost due to phonon escape. For values of τ_{DC} greater than a factor of ~ 3 increase, the sensitivity actually decreases in the center of the nanowire. This occurs because energy in the electronic system diffuses enough that it no longer suppresses the order parameter within a confined hotspot. The edge of the nanowire shows slightly altered behavior. At low bias currents, the longest τ_{DC} actually leads to enhanced sensitivity. This occurs because phonon loss is a more significant factor near the edge of the wire where diffusion is limited based on geometry. With slower diffusion, more energy couples to the phonon system and can be lost to the bath. A longer τ_{DC} gives the electron system time to diffuse, leading to a slower buildup of energy, less transfer to the phonon system, and enhanced sensitivity. These calculations demonstrate that the transverse coordinate dependence of absorption, and therefore the polarization dependence, are strongly influence by the balance between diffusion and coupling to the substrate.
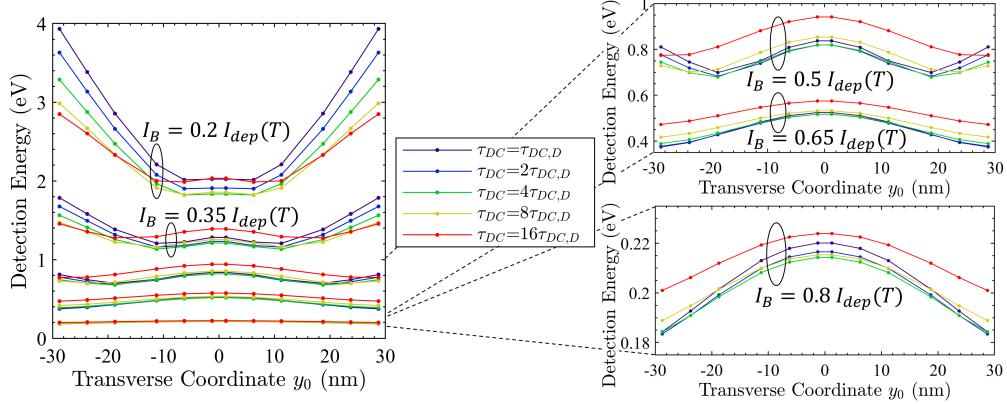
Figure 3.25: Simulated detection energy for various scaling factors to the value of $\tau_{DC,D} = \frac{\tau_0}{\gamma \Omega_D}$, which corresponds to the phonon-electron scattering time for Debye phonons. Colors indicate the correction factor, ranging from the nominal value (purple) to an increase by a factor of 16 (red). The plots on the right highlight the behavior at higher bias currents.

Based on the calculations of Chapter 2, we attempt the same modeling using τ_{DC} increased by a factor of 4 (compared to $\frac{4}{3} \frac{\tau_0}{\gamma \Omega_D}$) and using $\gamma = 25$, $\tau_{esc} = 9.4$ ps, and $\tau_{ee}(T_c) = 5$ ps. There is a small improvement in the quality of the fitting, as shown in Figure 3.26a, when compared to the results calculated using $\tau_{DC} = \frac{4}{3} \frac{\tau_0}{\gamma \Omega_D}$. The fit is obtained using $\bar{\chi} = 0.59$ and $\sigma_{F,1550} = 104$ meV. For a nanowire with 120 nm width, the fitting remains poor, particularly at high bias currents as shown in Figure 3.26b. For the wider wire, the fitting parameters are $\bar{\chi} = 0.56$ and $\sigma_{F,1550} = 160$ meV. For reference, application of Equation 2.51 of Chapter 2 based on high energy phonon escape suggests $\sigma_{F,1550} = 91$ meV for $\bar{\chi} = 0.59$, which is reasonably close to the fit value of $\sigma_{F,1550} = 104$ meV. However, application of the same formula for the 120 nm wide results leads to $\sigma_{F,1550} = 92$ meV for $\bar{\chi} = 0.56$, which is far from the fitting parameter of $\sigma_{F,1550} = 160$ meV. The relative latency for the 80 nm fit (not shown) is similar to that of Figure 3.21b, showing the correct timescale for the detection process when compared with experiment.
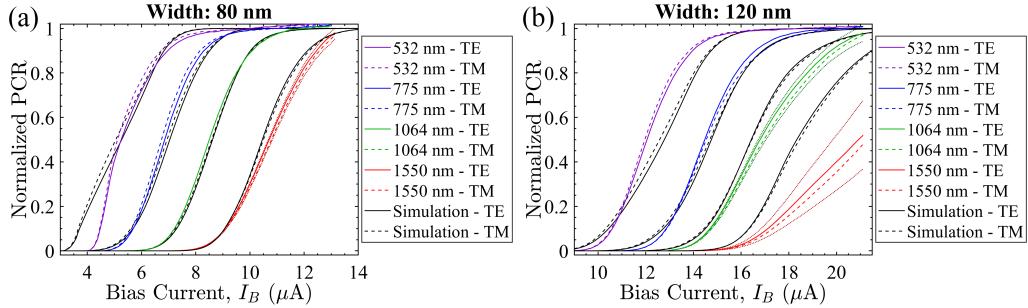
Figure 3.26: Fitting of the model to experiment for (a) 80 nm width and (b) 120 nm width devices while using τ_{DC} increased by a factor of 4 compared to the average value for phonons of the phonon plateau.

Initial Conditions: Diffusion During Downconversion

Application of the modified phonon bubble initial condition appears to bring the model predictions closer to experimental observations based on theoretical estimates of the material parameters of NbN, but it does not solve the problems associated with trying to fit the data for wider nanowires. In the chapter on downconversion, we noted that the incorporation of diffusion while solving the kinetic equations leads to additional localization of the phonon energy in a region smaller than the $2.5\xi_c \times 2.5\xi_c$ size used in the TDGL model. With this in mind, we check the influence of the initial hotspot size on the detection characteristics within the TDGL model. For the modified phonon bubble initial condition, the choice of hotspot size has only a minimal impact on the detection energy. To understand why, it is instructive to return to the problem of downconversion when diffusion is considered.

By comparing the results of the kinetic downconversion model to that of the two-temperature model, with diffusion, in the limit of a normal metal, we can gain insight into some of the limitations of the model. The following calculations use the same parameters as Chapter 2 with $\tau_{esc} = 10$ ps. Diffusion is modeled using the cylindrical coordinate system, and a fixed temperature boundary condition is used at the maximum simulation radius of 50 nm. The initial condition for the kinetic calculation consists of three equal energy electron-hole pairs within a hotspot radius of 1 nm. Figure 3.27 shows the comparison of the solution to the downconversion problem in cylindrical coordinates for the kinetic model and the two-temperature model using the phonon bubble initial conditions with the standard τ_{DC}. Unlike the case of a fixed hotspot size studied in Chapter 2, it is clear that the two-temperature model fails to capture the dynamics of the expanding hotspot when

diffusion is considered. Much more energy is coupled to the electron system in the two-temperature model than the kinetic model. This can be understood based on the non-equilibrium dynamics of the hotspot. In the kinetic model, when high-energy quasiparticles diffuse, they still rapidly interact with the phonon system because they have high energy, even though their density is low. In contrast, in the two-temperature model, the expanding hotspot cools rapidly due to the radial geometry of the expansion. This cooler electronic system does not couple as readily to the phonon system, so energy is retained by the diffusing electrons rather than coupling back to the phonon system. This also explains why the initial hotspot size does not strongly influence the detection characteristics within the TDGL model. Regardless of the hotspot size, the total energy in the electronic system is approximately the same over on the timescale of order parameter evolution, so the detection characteristics are not strongly influenced.
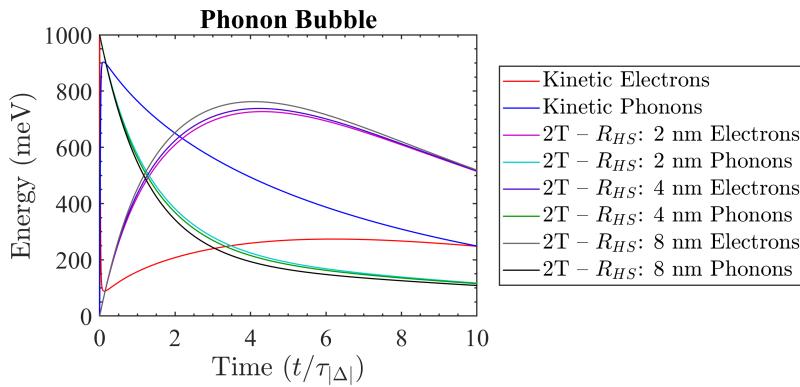
Figure 3.27: Comparison of the kinetic downconversion model with the two-temperature model in the case of cylindrical expansion of a hotspot using the modified phonon bubble initial conditions. Changes to the hotspot size of the modified phonon bubble do not strongly influence the energy evolution and partitioning of the system.

We can make a similar comparison to the thermalized hotspot initial condition, as shown in Figure 3.28. The use of the thermalized hotspot leads to a larger deviation in the electron energy in the system based on the hotspot size. However, no choice of hotspot size leads to a reasonable match between the two-temperature model and the kinetic model of downconversion. As with the phonon bubble, the two-temperature model overestimates the amount of energy present in the electronic system following downconversion. Compared to the modified phonon bubble, the thermalized hotspot leads to less energy coupled to the electron system. This occurs because a significant

fraction of the photon energy is initially deposited in the phonon system and this energy can escape to the substrate.
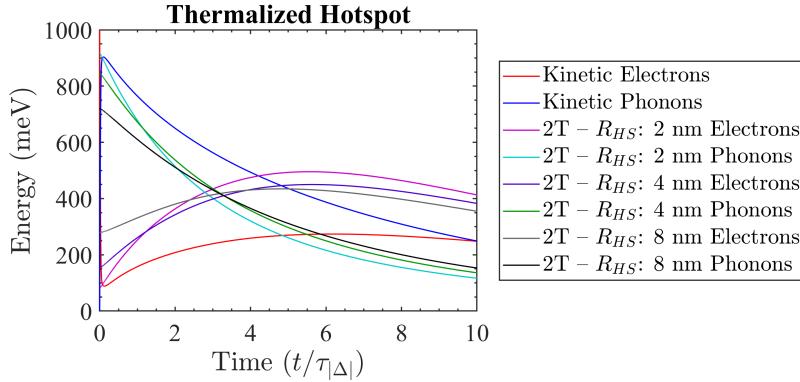
Figure 3.28: Comparison of the kinetic downconversion model with the two-temperature model in the case of cylindrical expansion of a hotspot using the modified phonon bubble initial conditions.

For the fixed-size hotspot studied in Section 2.4.6, increasing τ_{DC} by a factor of four reduced the error between the two-temperature and kinetic models. If we implement the same correction factor in the case of diffusion, there is still a significant difference between the two models, as shown in Figure 3.29. However, once implementing the energy loss factor $\bar{\chi}$ as done with the TDGL simulations, the amount of energy in the electron system becomes much closer to that predicted by the kinetic model. The purple line, representing the electron energy in the two-temperature model with $\bar{\chi} = 0.59$, has a shape which is qualitatively similar to that of the red curve representing the electron energy in the kinetic model. However, there is still a large discrepancy in the quantitative agreement between the two, with the two-temperature model showing a 25% larger maximum energy in the electronic system than the kinetic model.
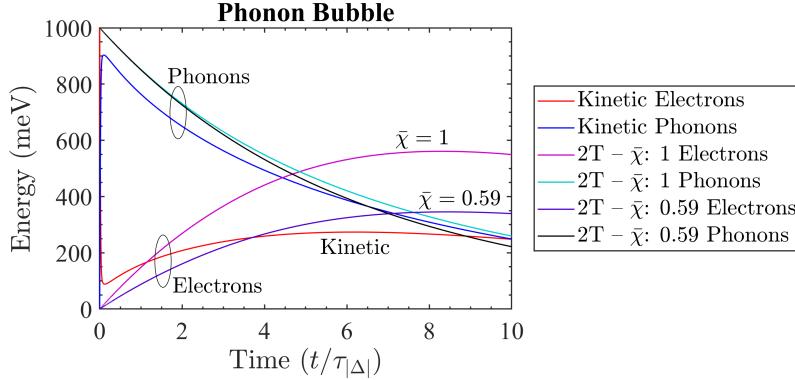
Figure 3.29: Comparison of the kinetic downconversion model with the two-temperature model for the case of cylindrical expansion of a hotspot using the modified phonon bubble initial conditions with τ_{DC} increased by a factor of 4.

The careful choice of $\bar{\chi}$ and τ_{DC} can certainly improve the agreement between the kinetic and two-temperature models to some extent, but any such modifications can only be considered an approximation to the true non-equilibrium behavior of the system. Our introduction of the modified phonon bubble initial condition is simply intended to be a way to alter the energy flow into the electronic system during downconversion in a way which crudely approximates the non-equilibrium dynamics of the electron and phonon systems. The use of this initial condition in the generalized TDGL equations leads to reasonable, but not perfect fitting of experimental data using fitting parameters which are plausible based on the model of high energy phonon loss during downconversion. The same fitting parameters lead to a qualitatively similar evolution of the electron energy during downconversion when comparing the two-temperature model with the kinetic model when considering diffusion during the downconversion process. Clearly, the non-equilibrium dynamics of the electronic system immediately following downconversion must be taken into consideration for a fully quantitative model of the detection process, but even the application of approximation methods, as demonstrated in this work, can bring model predictions much closer to semi-quantitative agreement with experiment.

One-Temperature Model

The details of the phonon interaction with the substrate can play an important role in determining the sensitivity of the nanowire within the TDGL model. The model of the energy balance equations used throughout this chapter relies upon the two-temperature model. However, it is known the two-temperature model does not

accurately reproduce the behavior of a system which is far from equilibrium [80]. Therefore, it would be useful to check that the more accurate one-temperature model based on the solution of the kinetic equation for the phonon system (see Section 2.4) produces results consistent with the two-temperature model for the case of the TDGL model.

The challenge with the one-temperature model is solving the integro-differential equation for every grid point of the simulation domain. In the form presented in Chapter 2, it is not feasible to solve the one-temperature model for a 2D simulation. Fortunately, the bulk of the calculations can be solved once prior to modeling the 2D dynamics to form a lookup table. Then, during the simulation, each element of the domain requires integration of a set of values from $t' = 0$ to t from the lookup table. While slow compared to the two-temperature model, this is still vastly more efficient than attempting to solve the standard form of the integro-differential equation and permits checking the consistency of the two approaches.

Using the modified phonon bubble initial condition, the latency results for the one-temperature model are compared to those of the two-temperature model in Figure 3.30. While there is a noticeable disagreement at low bias currents, the overall agreement between the two formulations is good. This is consistent with the results of Chapter 2 which demonstrated reasonable agreement between the one-temperature and two-temperature models for a fixed hotspot when considering downconversion.
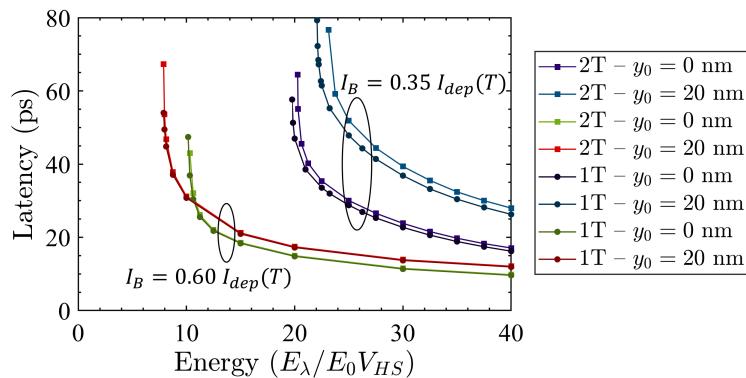
Figure 3.30: Comparison of the latency calculations for the one-temperature (dark circles) and two-temperature (squares) models for the energy balance within the TDGL equations. There is only a small difference between the two results. The material parameters are the same as throughout with $\gamma = 20$, $\tau_{esc} = 9.4$ ps, and $\tau_{ee}(T_c) = 5$ ps.

While the two-temperature model is sufficient for describing the NbN system (within the framework of thermalized electrons), modeling the WSi material system will likely require improvements. Phonon bottlenecking leads to a highly non-equilibrium distribution which must be considered using either the full one-temperature model or using a simplified three-temperature model (see Chapters 2 and 4). We have demonstrated that the solution to the one-temperature integro-differential equation is possible within the generalized TDGL framework, but even with the lookup table solution method, the computation is slow enough to make a wide sweep of transverse coordinates, photon energies, and bias currents impractical. Application of the three-temperature model, in contrast, would be relatively inexpensive. We do not attempt fitting to WSi in the current work, but any attempt to do so would require consideration of the bottlenecking effect.

3.5 Discussion

There has been growing interest in the possibility of using SNSPDs to detect single photons in the mid-infrared [41]. However, substantial technological development is required in order to achieve this goal. A recent effort has demonstrated near saturated internal efficiency for $9.9\,\mu\text{m}$ light in an ultrathin ($\sim 2.2\,\text{nm}$) and narrow ($50\,\text{nm}$) WSi nanowire with a stoichiometry tuned to create an alloy with $T_c = 2.8\,\text{K}$ [117]. Reduction of T_c is a promising strategy for increasing long-wavelength sensitivity because it reduces the energy gap, and therefore reduces the amount of energy required to suppress superconductivity.

While the current generalized TDGL model is not expected to make accurate direct quantitative predictions, its qualitative accuracy suggests it can be used as a tool to understand the scaling trends that can be expected as this avenue of T_c tuning is pursued. The characteristic diffusive length scale of the system is $l_D \sim \sqrt{D\,t_{|\Delta|}} \sim \sqrt{\frac{D}{k_B T_c}}$, and therefore increases as T_c drops, where we have assumed that the characteristic time for order parameter suppression $t_{|\Delta|} \sim \hbar/k_B T_c$. Therefore, for narrow nanowires where $l_D > w$, the system is expected to enter a quasi-1D regime which can be approximated by the 1D hotbelt model. Within the 1D model, a detection event is expected if the energy density in the hotbelt exceeds a certain threshold over the characteristic latency interval. Time scales according to $\tau_{T_c} = \hbar/k_B T_c$ for evolution of the superconductor, so one can expect the detection threshold to scale as $\sim V_{HB}\,E_0 = l_D dw E_0 \sim T_c^{3/2}$ for fixed geometry and diffusion coefficient. This general scaling argument is supported by detailed numerical simulation. The $T_c^{3/2}$ scaling is predicted by the standard TDGL model,

but this scaling only holds for a certain case of the generalized TDGL model. In the generalized TDGL model, the variation of the order parameter is modified by $\varrho\left(T_e\right) = \sqrt{1 + 4\left|\Delta\right|^2 \tau_{sc}\left(T_e\right)^2 / \hbar^2}$. In the current model, we assume that the scattering time is limited by electron collisions, which scale as $\tau_{sc} \sim 1/T_e$. When the equations are nondimensionalized, assuming the scattering time is not affected by T_c, $\varrho\left(T_e/T_c\right)$ remains unchanged, so the expected $T_c^{3/2}$ scaling remains valid. If the dominant scattering mechanism has a different power law, ϱ can scale with T_c, altering the response of the system. In general, assuming $\tau_{sc} \sim 1/T_e^n$, the detection threshold will scale faster than the 3/2 power for $n > 1$ and slower for $n < 1$. This general argument neglects the effect of phonon escape and electron-phonon coupling, both of which become slower with lower temperature. These scaling arguments implicitly assume that all of the other material properties of the system do not change as the T_c is altered. Scaling based on the generalized TDGL formulation supports the experimental evidence that reducing T_c is a viable strategy for improving mid-infrared sensitivity, but further analysis based on the kinetic equations is required to determine if there are potential problems due to fluctuations in the loss of high-energy phonons or delayed thermalization of the electron and phonon systems.

Another area of rapid development is the SNSPD community has been the prediction [6] and experimental confirmation [60–62] of single-photon detection in micron-wide SNSPDs. This result was first predicted by Vodolazov using the modified TDGL formulation and the experimental observation of this effect further supports the qualitative validity of the TDGL approach. However, as with narrow nanowires, there is still a gap between models and experimental results. The generalized TDGL model offers a promising framework for making a semi-quantitative connection to experimental results, but this still requires refinement of the model.

While the efforts to refine the quantitative accuracy of the TDGL model described in this chapter have dramatically improved the connection between experiment and theory, this model retains severe limitations. At a fundamental level, the generalized TDGL equations can not be expected to be quantitatively accurate. The requirements that the temperature be near T_c and that gradients in time and space be small are simply not satisfied over the entire latency period of detection. At first glance, this seems discouraging, but the suppression of superconductivity is dominated by the behavior of the hotspot. This area of the superconductor is often close to T_c over a large portion of the latency interval and therefore, the assumption about the temperature of the system might not be too bad.

At low substrate temperatures, an instability arises in the equations due to the definition of Joule heating in the electron energy balance equation. Several attempts were made to address this problem. The influence of the supercurrent on the electron energy density in the superconducting state was added to the system, but its effect on the system evolution was minimal and it did not resolve the instability at low temperatures. Further additions to the electron energy were made originating from the kinetic equations and electric potential, but these likewise had only insignificant effects. Because modifications to the energy balance equation failed to resolve this issue, it is plausible to assume that the failure comes from the use of the TDGL equations in the presence of strong gradients. Resolving this problem may require significant modification to the generalized TDGL formulation or the use of the more general Keldysh formalism.

Perhaps the most significant limitation of the TDGL formulation is the use of the two-temperature model. We have shown that the non-equilibrium dynamics of the system as calculated using the kinetic equations significantly deviate from those found using the two-temperature model. The presence of hot electrons immediately following downconversion leads to enhanced coupling between electrons and phonons as the electrons diffuse. This behavior is not captured with the two-temperature model, and the result is an overestimation of the energy in the electron system. Introduction of a modified phonon bubble initial condition improves the agreement of the two-temperature model with the kinetic model and leads to reasonable fitting of theory with experiment. However, this adjustment still fails to achieve high quality fitting to measurements of wider (120 nm) devices. This may be due to uncertainty in other material parameters. For instance, the diffusion coefficient changes the effective width of the nanowire when scaled in units of the coherence length, and is expected to have a strong impact on device performance. Further investigation into meshing the downconversion process with the TDGL simulations is necessary.

3.6 Summary

The SNSPD community's understanding of the detection mechanism of these devices has been driven by experimental progress. The latest experimental milestone, the measurement of intrinsic timing jitter and relative latency, has redefined our quantitative understanding of this process. The measurement of relative latency between phonon pairs on the order of 0 to 20 ps has demonstrated the need for the generalized TDGL formulation over the standard TDGL formulation which had been the widely accepted standard as the most accurate detection model for SNSPDs.

The generalized TDGL model predicts a much slower detection process than the standard TDGL model with the latency dominated by the scattering time τ_{sc}. This is theorized to be dominated by electron inelastic scattering, but the exact nature of this process has not been determined. Nonetheless, identifying this parameter of interest already suggests that one of the indicators of a sensitive SNSPD material is a fast scattering time. Finding or designing materials with enhanced scattering may provide an avenue to achieving lower timing jitter. Controlled measurement of the polarization dependence of the detection properties of SNSPDs has provided useful information about the transverse detection profile of narrow and standard NbN nanowires. By comparing these results with the predictions of the generalized TDGL model, we can determine the transverse detection profile of the nanowire and infer characteristics about its regime of operation. In this sense, theory and experiment mesh nicely at a qualitative level.

While the work presented in this chapter represents significant progress toward the goal of formulating a predictive model of the SNSPD detection mechanism, there are numerous limitations preventing accurate agreement between experiment and simulation. Though more general than the standard TDGL model, the generalized TDGL equations still require slow variation in time and space as well as temperatures close to T_c. These conditions are not strictly satisfied during the single-photon detection process. The instability of the formulation due to Joule heating is another limitation, and hints at the underlying problems of the TDGL set of equations in the presence of strong gradients in electric potential. The two-temperature model of the energy balance equations poses its own limitations. Neither the electron nor phonon systems can be accurately represented by equilibrium distributions immediately following downconversion. Use of the two-temperature model leads to significant deviations in the behavior of the system compared to the kinetic equations and ultimately alters the suppression of superconductivity and the detection properties. The generalized TDGL equations are powerful and relatively straightforward to solve numerically, but a truly quantitative model of SNSPD detection process appears to require a more rigorous treatment of the kinetic equations and non-equilibrium superconductivity.

C h a p t e r   4

ELECTROTHERMAL MODELING

This chapter contains published work from [4] and [5].

Text from Sections 4.1.2 and 4.2.4 and Figures 4.7(c,d) and 4.18 are adapted with permission from Springer Nature: J. P. Allmaras, A. G. Kozorezov, A. D. Beyer, F. Marsili, R. M. Briggs, and M. D. Shaw, "Thin-film thermal conductivity measurements using superconducting nanowires", Journal of Low Temperature Physics **193**, 380–386 (2018) `10.1103/PhysRevApplied.11.034062`. Copyright © 2018 Springer Nature.

Text from Sections 4.3 and 4.3.2 and Figures 4.30 and 4.31 are adapted with permission from: A. N. McCaughan, V. B. Verma, S. M. Buckley, J. P. Allmaras, A. G. Kozorezov, A. N. Tait, S. W. Nam, and J. M. Shainline, "A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors", Nature Electronics **2**, 451–456 (2019) `10.1038/s41928-019-0300-8`. Copyright © 2019 Springer Nature.

The temperature-dependent nature of superconductivity means there is complex interplay between the electrical and thermal energy within the system. There is a long history of attempts to better understand and model this behavior. Significant progress has been made in explaining features of the resistive transition [118–123], hysteretic effects [67], and microwave response [76], but there are still open questions about the response dynamics in SNSPDs.

Much of the early work in this field was aimed at understanding the DC behavior of superconductors. A typical I-V curve of a thin-film SNSPD is shown in Figure 4.1a. As the bias voltage increases, the current in the nanowire increases until it reaches the switching current (I_c or I_{sw}), at which point parts of the detector enter the normal state. In meander devices, this often leads to several segments of varying normal resistance corresponding to different legs of the meander going normal while others

remain superconducting. Finally, once the entire device is in the normal state, a fully resistive current vs voltage dependence is observed. When the bias voltage is slowly reduced, a hysteretic state appears at a fixed current, known as the hotspot current I_{HS} or retrapping current I_r. This effect was first explained by Skocpol, Beasley, and Tinkham (SBT) based on the formation of a self-heating normal domain within the nanowire [67]. In the SBT model, a linearized energy balance equation is solved with a circuit equation to describe the temperature profile within a superconducting nanowire with a normal domain. If the length of the nanowire is sufficiently long, a region of operation occurs where a fixed bias current leads to a Joule heating based self-heating hotspot of various lengths, as shown schematically in Figure 4.1b. While qualitatively sufficient to describe the hotspot current observed in SNSPDs, the linearized SBT model is insufficient for quantitatively describing the experimentally measured hotspot current over a range of bath temperatures. More recent works have introduced more appropriate descriptions of the heat transfer process in order to better explain this experimental observation, but predicting hotspot current behavior based on theory remains challenging due to uncertainties in material properties and interface thermal boundary conductance.
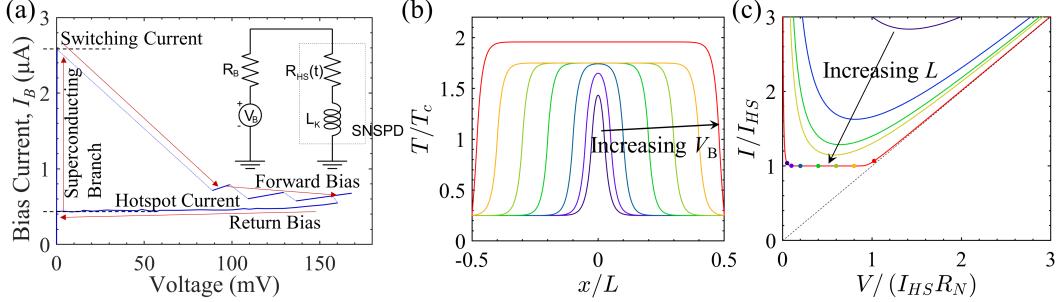
Figure 4.1: (a) Example IV curve of an SNSPD. As the bias voltage first increases, the device remains in the superconducting state with the device current increasing without the appearance of a voltage across the device. Once the current flowing through the nanowire exceeds the switching current, the device switches to the normal state and voltage appears across the nanowire. The linear segments in the forward bias branch occur when segments of the nanowire transition to the normal state. When the bias voltage is reduced, the nanowire enters a regime where the behavior is dominated by a self-heating hotspot. The current which sustains the self-heating hotspot is known as the hotspot current and appears as a flat region in the IV curve.[1] The inset shows the schematic circuit used to understand and model this DC behavior. A voltage source (V_B) in series with a bias resistor (R_B) acts as a current source while the superconductor is modeled as a variable resistor (R_{HS}) in series with an inductor (L_k). (b) Temperature profile within a nanowire using the SBT model. As the bias voltage increases, the hotspot expands until it covers the entire length of the nanowire. (c) IV curves of a nanowire within the SBT model. Different colored lines correspond to different lengths of nanowire L. With sufficiently long nanowires as is the case with SNSPDs, a hotspot plateau appears. The colored circles correspond to the curves of (b).

Building on the ideas of the SBT model, the SNSPD community has developed electrothermal models to describe the dynamic behavior of these detectors during the photodetection process and subsequent evolution of the system [56, 57, 68, 124]. The first of these models was proposed by Yang et al. in 2007 [68] which combined a one-dimensional heat transfer equation with a single temperature $T(x)$ describing a NbN SNSPD using a conventional AC readout circuit. While this model made a number of simplifying assumptions about the thermal properties of the NbN,[2] this

---

[1]This particular device was constricted and had a low switching current, making it easy to observe the linear segments in the forward branch. This makes it useful for showing the different regimes of an SNSPD IV curve, but the switching current shown is not representative of a the switching currents found in high performance SNSPDs.

[2]In the model, the thermal coupling to the bath was linearized, which is only valid in the limit of small temperature differences between the substrate and phonon system. The use of a single temperature to describe the electron and phonon systems is also only valid in the case of fast phonon-electron coupling, which is not justified in NbN.

was the first work to describe the role of electrothermal latching in SNSPDs. If the reset time of the SNSPD is too fast compared to the thermal relaxation time, a self-heating hotspot can form which prevents the detector from resetting back to the superconducting state as shown schematically in Figure 4.2. This thermally limited reset time ultimately determines the maximum count rate (MCR) an optimized detector can achieve for a given material system and is therefore of great interest when designing fast SNSPDs. While later work advanced these models to describe additional material systems [57], use a more appropriate two-temperature model [57, 124], and describe coupled parallel nanowires [124], no serious efforts have been made to extend these to include amorphous materials such as WSi or MoSi. One work considered the use of a simplified model to describe the growth dynamics of multiple hotspots in WSi [125], but the model used in that work, based on [56, 126], is not capable of accurately describing thermally-induced latching. Furthermore, there are still questions about the applicability of these equations to devices outside of the most idealized conditions for the model: polycrystalline materials such as NbN on crystalline substrates. It is not immediately clear if these models can be adapted for lower T_c materials and amorphous films such as WSi or MoSi on amorphous dielectric substrates.
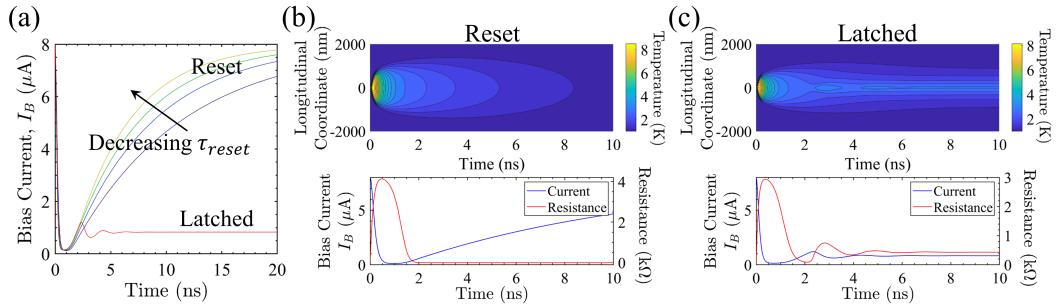
Figure 4.2: (a) Emergence of latching as the electrical reset time τ_{reset} decreases. In the case of (b) proper recovery the resistance of the device returns to zero after a detection and the temperature (shown in top) relaxes back to the substrate temperature. In contrast, for a (c) latching device, the resistance remains finite, and a segment of the nanowire remains at a temperature above T_c.

As SNSPDs were developed for various applications, it became important to scale the properties of single pixels to larger format arrays. For the Deep Space Optical Communication (DSOC) demonstration mission ground detector, a 64-pixel SNSPD array was developed using the WSi material system [27]. For this application, a large number of pixels were needed in order to reach the desired MCR and

active area, but in order to keep the optical coupling approximately constant on all of the pixels it was necessary to co-wind the individual pixels [27] (see Figure 4.8a). Unfortunately, during development, it was quickly found that co-wound pixels displayed a combination of thermal and electrical crosstalk if the pitch of the array was too small. A suitable balance between reduced fill factor for preventing crosstalk and high enough fill factor to achieve efficient optical coupling was reached through experimental iteration, but a theoretical understanding of thermal crosstalk would be beneficial for proper detector optimization. Thermal coupling has since been utilized to provide imaging capability in bi-layer devices (see Chapter 5) [1] and design heater-based multilayer superconducting switches [4, 58], but adequately describing thermal transport in these systems remains challenging.

In this chapter, we describe the application of electrothermal models to the WSi material system and describe experiments used to test these models. These experiments include characterizing the steady-state hotspot current, dynamic latching behavior, and crosstalk in arrays of nanowires. The standard electrothermal models described in the literature fail to account for all of the dynamics observed in the WSi system, and we therefore generalize the equations to more rigorously describe the electrical and thermal systems within the superconductor. Furthermore, in order to understand crosstalk behavior, we include the thermal transport in thin dielectric layers. Application of the generalized electrothermal equations is sufficient to qualitatively describe most behavior within the WSi system, but the amount of fine tuning required for the fitting parameters raises doubts about the current model's suitability for design prediction and optimization. Finally, we apply a similar set of electrothermal equations to describe the behavior of superconducting multilayer heater switches which are of interest for a number of applications.

4.1 Experimental Characterization – WSi

In order to develop an accurate electrothermal model for the WSi material system, it is important to understand how experimental data can be used to validate modeling efforts. A complete model would be able to describe the steady-state and dynamic response of WSi nanowires on various substrate materials. This includes understanding crosstalk between adjacent nanowires in arrays. Therefore, we must characterize heat transfer in the nanowire, coupling to the substrate, and also heat flow in the substrate material. To this end, three sets of experiments were conducted in order to measure the stationary and dynamic processes in both the nanowires and SiO_2 thin films.

4.1.1 Hotspot and Latching Current

A first set of devices was designed and fabricated in order to characterize the static and dynamic properties of WSi nanowires. These devices consisted of an array of WSi nanowires with different inductances in order to generate pixels with different electrical reset times ($\tau_{reset} \approx L_k/R_L$). Figure 4.3a shows a scanning electron microscope (SEM) image of the device design and Figure 4.3b shows an image of the CAD design with a single pixel highlighted. Each pixel consisted of a 50 µm long segment of active 160 nm wide nanowire in series with a 320 nm wide inductor. The active region was kept the same length in order to keep the probability of having a constriction the same for each of the pixels. This active length was also kept short in order to minimize the probability of having a constriction. The inductors used a wider width in order to ensure that the switching current was limited by the photosensitive region and was not dependent on the size of the inductor.

All samples were prepared by Dr. Andrew Beyer and Dr. Ryan Briggs using the nanofabrication facilities at JPL's Microdevice Laboratory (MDL) and Caltech's Kavli Nanoscience Institute (KNI). The fabrication process began with a 4 inch silicon wafer with ~ 240 nm of thermal oxide. WSi was sputtered from a compound target with a composition of 65% W and 35% Si.[3] Following the WSi deposition, a lift-off stencil was patterned by optical lithography (248 nm exposure light) to form the electrical leads. Niobium leads were deposited with a 15 nm gold passivation layer with ion mill cleaning steps between each deposition to remove any oxidized material which could introduce contact resistance between the leads and the nanowires. In addition, thin (~ 2 nm) titanium adhesion layers were deposited between the WSi, Nb, and Au films. Once the contacts were patterned, the nanowires were patterned with electron-beam lithography (EBL) using Ma-N 2401 negative tone resist. After development, the nanowires were etched using a dry CHF_3 and O_2 ICP reactive ion etch and the remaining EBL resist was stripped with acetone. Finally, the entire wafer was passivated with ~ 200 nm of SiO_2 to prevent oxidation and degradation of the WSi nanowires.

---

[3]Analysis using Rutherford Backscattering Spectroscopy (RBS) of films deposited from the same sputtering target under similar deposition conditions several years later showed that the final films had a stoichiometry of 87% W and 13% Si, meaning the films are much more tungsten rich than anticipated based on the stoichiometry of the sputtering target. However, it is not clear if the analysis performed on the recent films is representative of the properties of the films used in the devices described in this chapter.
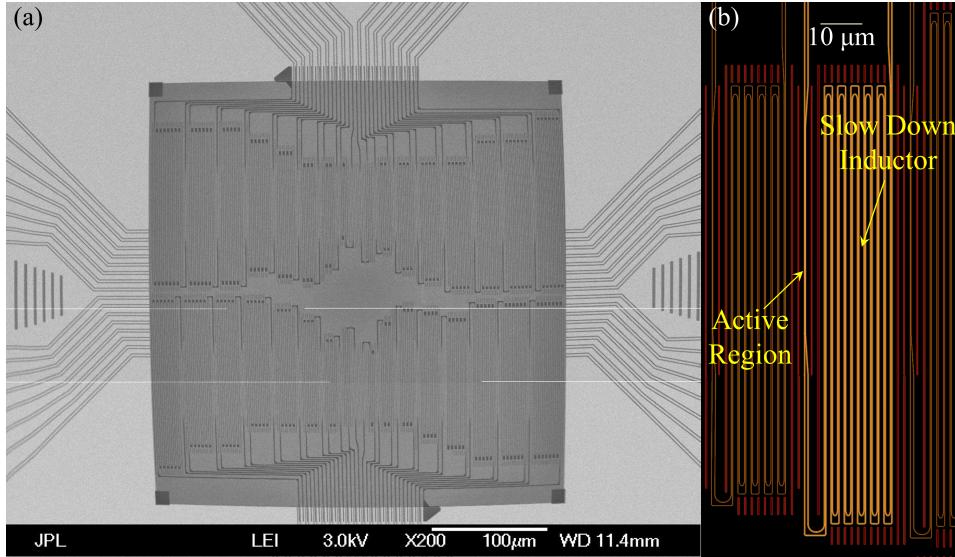
Figure 4.3: (a) SEM image of the latching devices fabricated to probe the reset characteristics of WSi nanowires. Each channel had a different length series inductor which controlled the reset time of the pixel. (b) CAD image highlighting a single pixel of the latching design. Each pixel consisted of a 50 μm long active segment with a wider series inductor.

Two series of measurements were performed on these devices. The first was a DC characterization which measured the hotspot current as a function of bath temperature. Measurements were performed using a Stanford Research Systems (SRS) SIM928 voltage source, a $10\,\mathrm{k\Omega}$ bias resistor, and SRS SIM970 digital voltmeter (DVM) as shown in Figure 4.4. Due to the high resistance of thin WSi nanowires in the normal state, it was necessary to account for the $10\,\mathrm{M\Omega}$ impedance of the DVM when extracting the hotspot current. From these measurements, one extracts the hotspot current as a function of bath temperature, as shown in Figure 4.4b, which acts as a metric for validating the electrothermal model for WSi.
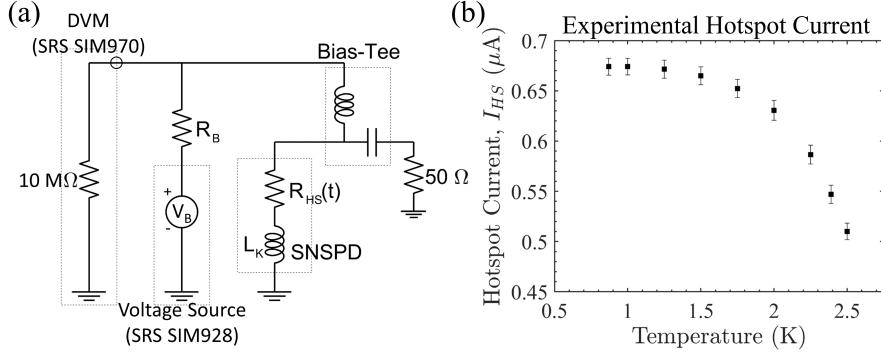
Figure 4.4: (a) Experimental setup for hotspot current and latching measurements. (b) Hotspot current vs bath temperature experimental results.

The second set of measurements characterized the switching current of the devices under different illumination conditions and temperatures. These measurements were used to characterize the dynamic reset characteristics in order to quantify the latching behavior of WSi nanowires. For the latching measurements, the switching currents were measured with a $50\,\Omega$ shunt on the RF port of the bias-tee to avoid reflections from the input of a readout amplifier. This effect is known to cause premature latching in SNSPDs. The experimental setup was the same as that for the hotspot current measurements.

The experimental latching current measurements are summarized in Figure 4.5. Each pixel of the array is indicated by its τ_{reset}, defined by the $1/e$ time constant, which was extracted from the electrical recovery time of pulse traces captured on an oscilloscope. There are several features to recognize. First, the switching behavior is different when devices are illuminated or dark as shown in Figure 4.5a. This is due to the low intrinsic dark count rates of WSi nanowires. Latching will only occur if the detector registers a click and then fails to recover, which can only occur if there is a source of detections. These detection events can be either due to intrinsic dark counts or true photon counts. As the bias current increases, if there is no source of counts, the bias current can exceed the latching current, with the detector entering an unstable regime where a single detection event will trigger the device to latch. This will appear as a switching current which exceeds the true latching current of the detector. Therefore, the switching currents measured when the detector was illuminated is the data relevant for understanding latching behavior. The second result of these measurements is that the latching behavior does not depend on bath temperature. Figure 4.5b shows the switching currents for

all devices and bath temperatures measured in this work. For devices with reset times >25 ns, decreasing the bath temperature leads to a continuous increase in the switching current. However, for pixels with shorter τ_{reset}, the switching current reaches an upper limit, approximated by the dashed black line in the figure. This indicates the point where latching limits the switching current. This latching current is independent of the bath temperature, as is most clearly shown in Figure 4.5c.
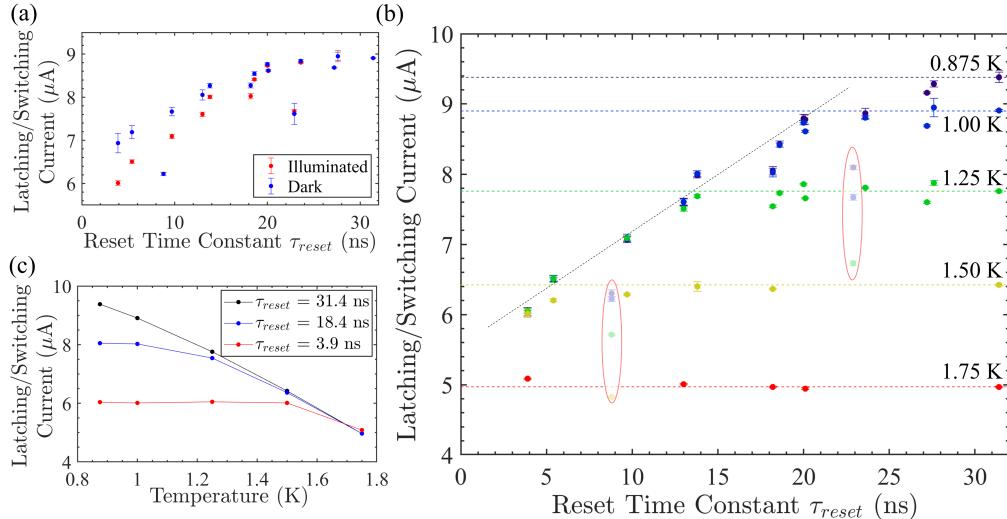
Figure 4.5: Summary of latching experimental results. (a) Latching current comparison when the device was illuminated or dark at a bath temperature of 1000 mK. Each pair of data points located at a given τ_{reset} corresponds to a single pixel of the array. Due to the low intrinsic dark count rate, latching can only be probed when the device is illuminated. (b) Latching behavior for all devices when illuminated. Vertical groups correspond to individual pixels while colors indicate the different bath temperatures (labeled to the right). Dashed colored lines indicate the nominal switching current without the presence of latching, while the dashed black line indicates the region where latching limits the switching current. The two circled pixels are constricted as evidenced by the low switching currents, and are therefore not considered in subsequent analysis. Not all pixels were measured at high bath temperature due to the limited hold-time of the fridge when heated to an elevated temperature. (c) Latching current vs temperature for three devices demonstrating the temperature independence of the latching current.

Higher temperature operation is generally not favorable for SNSPDs because they cannot reach as high a bias current due to the temperature dependence of the switching current. For devices which latch, high temperature operation can actually be advantageous. Increasing the bath temperature of an SNSPD reduces the bias current at which photodetection begins for a given wavelength (often termed the cutoff

current). Therefore, for a device which latches, increasing the bath temperature can improve photo-sensitivity without sacrificing switching current. This is demonstrated in Figure 4.6, which shows the PCR for three pixels with different reset times at 875 mK and 1500 mK. At the higher temperature, the device with a 3.9 ns recovery time constant shows saturation, despite not saturating at 875 mK. Such a device would be expected to have a MCR exceeding 100 Mcps, which is uncommon for WSi detectors. This is not an endorsement of WSi for applications requiring high count rates, but rather a demonstration that extra performance can be achieved by operating at higher temperatures for detectors that are limited by latching.
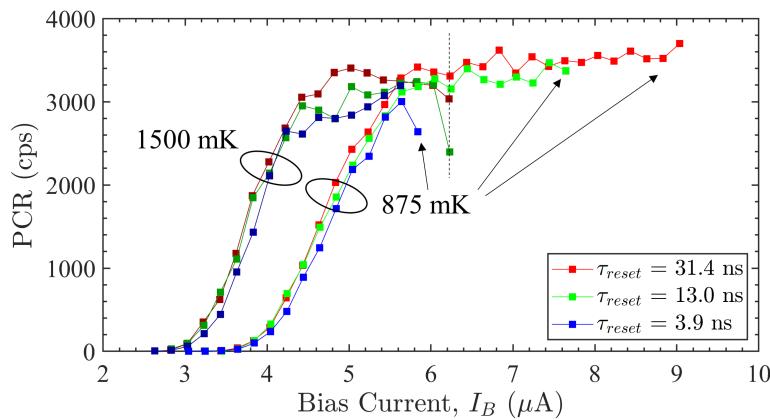
Figure 4.6: PCR measured at 1550 nm for three of the latching devices at 875 mK (light curves) and 1500 mK (dark curves). The emergence of saturated internal efficiency for the fastest device at higher bath temperature demonstrates a method of improving performance for devices which demonstrate latching. The dashed black line shows the switching current of the two slower devices at 1500 mK.

4.1.2   Nanowire Thermometry

While the hotspot current and latching experiments provide considerable information about the characteristics of the nanowires, they do not directly measure the heat transfer within the thin SiO_2 dielectric layers surrounding the nanowires. To address this shortcoming, a second series of devices was fabricated using the same process described above. These devices employed a new measurement technique, nanowire thermometry, which uses the superconducting properties of nanowires in order to perform local measurements of the temperature in thin dielectrics [5]. The scheme is outlined in Fig. 4.7. A series of parallel nanowires are patterned with a fixed separation distance. During measurement, one of the nanowires is biased at a sufficiently high current that the nanowire switches to the normal state. This heats

the surrounding nanowires as shown in Fig. 4.7c. Once the system reaches steady state, the switching currents of the surrounding probe nanowires are measured. By calibrating the switching current as a function of bath temperature as shown in Fig. 4.7d, it is possible to determine the local temperature of the nanowire due to the heating of the neighbor nanowire. This method enables measurement of the lateral heat propagation away from a heat source in thin dielectrics, as is necessary to better understand the thermal dissipation during operation of SNSPDs. The results, compared with modeling, are shown in Section 4.2.4.
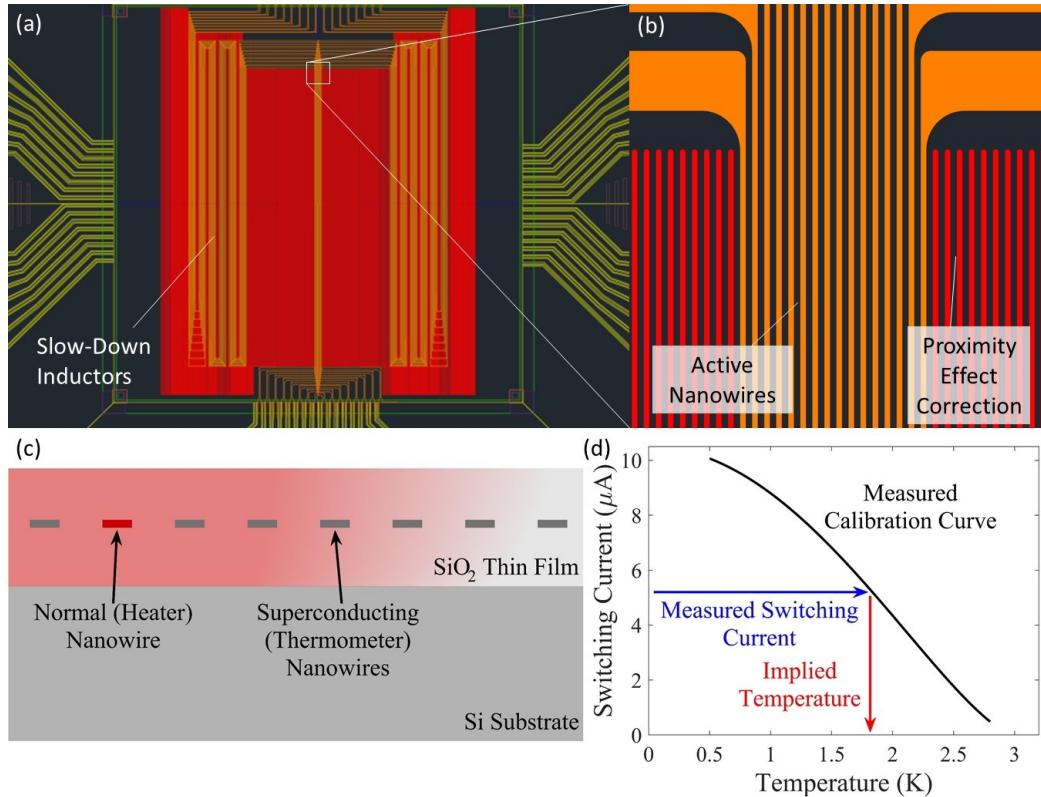
Figure 4.7: Nanowire Thermometry Scheme. (a) CAD image of the nanowire thermometry device design. Orange lines indicate the nanowire channels while the red lines show proximity effect correction nanowires used to improve the quality of the EBL. Each channel has an active region in the center of the device and a slow-down inductor far from the active region which is used to prevent latching. (b) Close view of the transition from active region to leads. The thick leads are necessary to confine the normal domain to the active region of the device. (c) Schematic illustration of the scheme. A single nanowire is biased in the normal regime while the remaining probe nanowires in the array remain superconducting. (d) Example calibration curve used to convert measured switching currents into local temperatures at the probe nanowires.

4.1.3 Crosstalk

The investigation into the thermal properties of WSi nanowires and the use of the nanowire thermometry scheme was motivated by the observation of crosstalk in arrays of WSi nanowires during the development of the ground detector for NASA's DSOC project. Measurements on prototype arrays during the development process exhibited correlated detection events on neighboring nanowires of co-wound arrays. The timescale of these correlations suggested that the main mechanism of crosstalk was thermal in nature because the delay between detections on neighboring wires was on the order of several nanoseconds. The series of measurements described in this section were performed by Dr. Francesco Marsili in order to better understand the nature of this crosstalk, but a suitable model was not developed to semi-quantitatively describe the results. The results of these measurements were never published, and are included here because they provide an additional experimental metric for evaluating the accuracy of electrothermal models. Unlike the nanowire thermometry scheme, which only operates in steady-state conditions, crosstalk measurements provide information about the dynamic timescales of thermal dissipation in these materials.

A series of devices was fabricated by Drs. Beyer and Briggs using the same method as described in Section 4.1.1. Arrays with 64 pixels were designed with pitches of 400 nm, 800 nm, and 1600 nm with 160 nm wide WSi nanowires covering an active area of ~320 μm diameter. Two types of measurements were performed to characterize the crosstalk between neighboring nanowires. The first characterized the crosstalk efficiency, by measuring the detector count rate (CR) of neighboring nanowires under different bias conditions. When a neighboring nanowire is grounded, the measured CR represents the true response of the measured pixel, which is equal to the photoresponse count rate (PCR) plus the dark count rate (DCR). However, when a neighboring nanowire is biased at a fixed current, the measured photoresponse is now the sum of the standard photoresponse and the crosstalk. The crosstalk efficiency η_{XTalk} is given by

$$\eta_{\text{XTalk}}(I_B) = \left[\text{CR}(I_B, I_{B,N}) - \text{CR}_0(I_B)\right]/\text{CR}_N(I_{B,N}) \qquad (4.1)$$

where $\text{CR}(I_B, I_{B,N})$ is the measured CR for a given bias current I_B and neighboring nanowire bias current $I_{B,N}$, $\text{CR}_0(I_B)$ is the CR with the neighboring nanowire grounded, and $\text{CR}_N(I_{B,N})$ is the CR of the neighboring nanowire at the bias current $I_{B,N}$. This is simply a statement that the crosstalk efficiency is the excess CR measured on the first pixel normalized by the total CR on the neighboring nanowire. The crosstalk efficiency is shown in Figure 4.8b for a 400 nm pitch array for various bias

conditions and bath temperatures. Similarly, the crosstalk efficiency for an 800 nm pitch array is shown in Figure 4.8c at a bath temperature of 630 mK. Fluctuations during the heat transfer process, inhomogeneity in the fabricated nanowires, and electrical noise contribute to broadening the transition between unity crosstalk, and no crosstalk. As the bias current on the triggering nanowire increases, the threshold for efficient crosstalk decreases toward lower bias currents on the receiving channel. As the bias current increases, more Joule heat is released during the detection process, which heats the neighboring nanowire to a higher temperature. For the 800 nm pitch array, more bias current is needed on both the triggering and receiving channels in order to observe efficient crosstalk. This is due to the longer distance between the two nanowires and the reduced kinetic inductance of each pixel, which reduces the potential energy of the inductor which can be released during a detection event. This potential energy ($\sim \frac{1}{2}L_k I_B^2$) becomes proportional to the length of the nanowire, and therefore inversely proportional to the pitch of the array because each design covered the same active area ($\sim 160 \, \mu m$ radius).
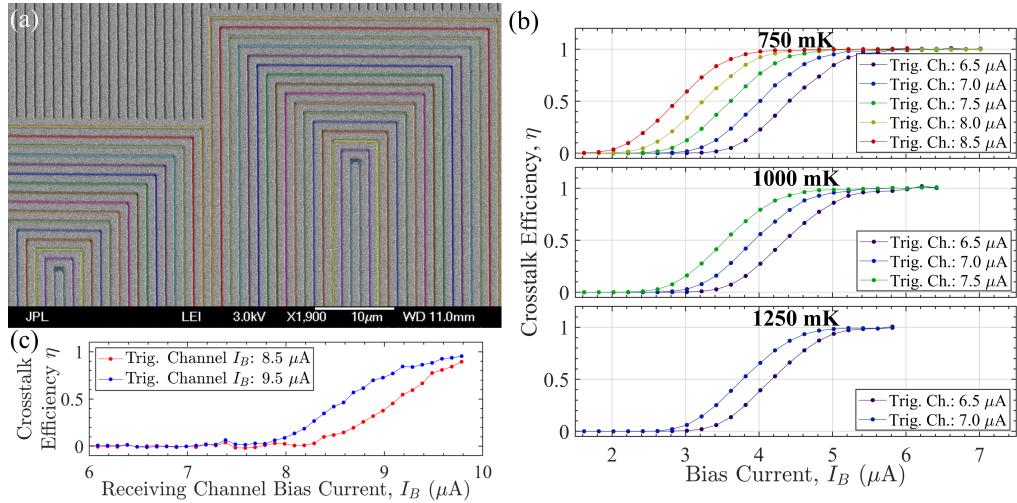
Figure 4.8: (a) False-color SEM image of a prototype DSOC array similar to the ones used in this study. The colored nanowires correspond to individual pixels, indicating the co-wound structure of these devices. The device in the image has a pitch of 1200 nm and exhibits no crosstalk, unlike the 400 nm and 800 nm pitch devices tested in this experiment. (b) Crosstalk efficiency for a 400 nm pitch array measured at bath temperatures of 750 mK, 1000 mK, and 1250 mK. As the bath temperature increases, the threshold current for crosstalk decreases to slightly smaller bias currents. (c) Crosstalk efficiency of an 800 nm pitch device measured at 630 mK.

Dr. Marsili performed a second series of crosstalk measurements which measured

the time delay between detection events on adjacent nanowires in the same devices. This time difference indicates the time required for the crosstalk mechanism to act on the adjacent the nanowire. Measurements were performed on the same 400 nm and 800 nm pitch arrays described above. The results of these measurements take the form of several distribution functions which show the probability of registering a crosstalk event at a certain delay time compared to the photodetection pulse on the triggering nanowire. The main results of these measurements are shown in Figure 4.9. For the 400 nm pitch array, measurements on nearest-neighbor pixels showed average delay times ranging from 2–2.75 ns based on the bias conditions (see Figure 4.9a) and did not display bath temperature dependence (see Figure 4.9b).[4] For the 800 nm pitch device, the average delay times were on the order of 5–6.5 ns based on the bias conditions, as shown in Figure 4.9c. When interpreting these results, it is important to remember that these histograms only represent the delay times given that crosstalk occurred, and should be interpreted parallel to the crosstalk efficiency curves of Figure 4.8 to understand the likelihood of any given photon detection triggering a crosstalk event.
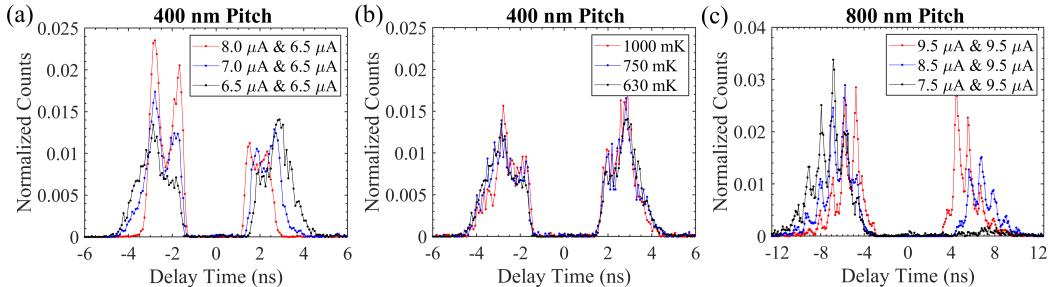
Figure 4.9: (a) Crosstalk histograms for the 400 nm pitch device measured at 630 mK. The different colors correspond to different combinations of bias currents on the two pixels. The left group of histograms corresponds to the variable bias current triggering the fixed 6.5 μA nanowire while the right group is the 6.5 μA nanowire triggering the nanowire with the varied bias current. (b) Crosstalk histograms at various bath temperatures for 6.5 μA bias currents on both channels. (c) Crosstalk histograms of the 800 nm pitch device measured at 630 mK for various bias combinations. The left group of histograms corresponds to the 9.5 μA bias current nanowire triggering the variable bias nanowire while the right group is the opposite.

---

[4]These measurements were performed in 2015. The details of the triggering process have since been lost, so there is the potential for some timing skew due to triggering at different fractions of the pulse amplitude when shifting bias currents. This skew is only expected to be a fraction of a nanosecond at most, so the main qualitative results remain valid.

The general features of the measured crosstalk can be explained qualitatively in terms of a combination of electrical and thermal crosstalk. The delay histograms, particularly for the 800 nm pitch device, show an oscillatory structure on top of a broad distribution. This was explained as a result of RF coupling leading to ringing in the adjacent nanowire. As thermal coupling slowly heats the nanowire, the nanowire will preferentially trigger during times of enhanced bias current due to the electrical ringing. Thus, the thermal crosstalk is responsible for the broad distribution while the electrical crosstalk is responsible for the oscillatory peaks within the distribution.

In principle, characterization of the thermal coupling efficiency provides information about the temperature rise caused by a detection event in a neighboring pixel. The threshold of 50% coupling efficiency can be considered the point where fluctuations and inhomogeneities equally cause or prevent a registered crosstalk event. Defining the temperature-dependent switching current according to the Bardeen temperature dependence

$$I_{sw}(T) = I_{dep}(0) \left(1 - \left(\frac{T}{T_c}\right)^2\right)^{3/2} \tag{4.2}$$

where $I_{dep}(0)$ is the zero temperature depairing current, we can approximate the maximum temperature reached. By inverting this equation using the bias current where the crosstalk efficiency is 50%, we estimate the maximum temperature reached during the crosstalk process. There is considerable uncertainty in the true depairing current of the nanowire due to uncertainty in material parameters. Furthermore, the electrical ringing means that the current in the adjacent nanowire during detection exceeds the nominal bias current, lowering the threshold for crosstalk. Given the uncertainties involved in this procedure (see Appendix C.1 for details), the estimated temperatures should be considered an estimate only. These values are shown in Figure 4.10 for the 400 nm and 800 nm pitch devices. We can make a further claim about the 800 nm pitch results. The histogram data for this device corresponds to a range of bias currents where the crosstalk efficiency is not saturated. This implies that the broad distribution corresponds to the time when the nanowire reaches its maximum temperature and fluctuations may or may not switch the device. The implication is that it takes approximately 6.5 ns for the nanowire to reach its maximum temperature when triggered by a nanowire biased at 9.5 μA. The combination of measurement and analysis provides approximate metrics which can validate the accuracy of an electrothermal model for crosstalk. While the experimental uncertainty makes it unreasonable to use these calculations as a precise

metric for comparison, they provide a general timescale of the thermal crosstalk mechanism which, as we will see, is already sufficient to provide feedback on the accuracy of basic models.
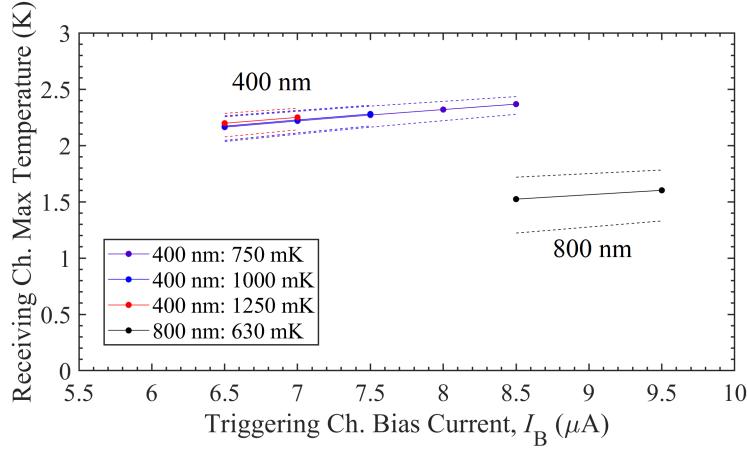
Figure 4.10: Estimated maximum temperature reached during crosstalk. The top group of curves shows the 400 nm pitch data for the three bath temperatures measured while the low curve show the 800 nm pitch data at 630 mK. The dashed lines show the approximate uncertainty bounds of the calculation.

4.2   Electrothermal Modeling

The goal of the preceding experiments was to gather the experimental data needed to generate an electrothermal model for WSi which is capable of informing the design process of SNSPDs. In particular, the electrothermal model aims to understand latching behavior, which dictates the MCR which can be achieved for an optimally designed detector, and crosstalk between adjacent pixels, which limits the maximum fill factor which can be achieved in arrays. Implicit in this goal is that the electrothermal model be as simple as possible so that it can be used as a design tool. For that reason, the following discussion begins with the simplest electrothermal description of the WSi nanowire and adds complexity as needed to fit experimental data. This begins with the standard two-temperature model describing the nanowire, which is then extended to the three-temperature model to describe the bottlenecking of phonons. The focus then shifts to describing the substrate, then finally, both elements are meshed together to provide a generalized model. The model predictions are compared to the experimental results throughout the section.

4.2.1 Two-Temperature Model

Introduced in 1956 by Kaganov, Lifshitz and Tanatarov [127], the two-temperature model is framework for describing the energy flow in the electron and phonon systems of a material when the system is perturbed from equilibrium. The model assumes that both the electron and phonon distribution functions can be represented by their equilibrium Fermi-Dirac and Bose-Einstein distributions at temperatures T_e and T_{ph}, respectively. The two-temperature model is often used in a linearized form to describe small perturbations from equilibrium, but it can be generalized into a set of coupled nonlinear partial differential equations which describe the energy exchange between electrons and phonons, diffusion, Joule heating, and phonon escape to a substrate. Building upon existing literature [56, 57, 68, 124, 128], we use the following 1D electrothermal equations to describe the WSi nanowire system, which are based on the two-temperature model. The energy balance equation for the electron system is given by

$$\frac{\partial E_e(T_e)}{\partial t} = C_{e,BCS}(T_e)\frac{\partial T_e}{\partial t} = \nabla \kappa_e(T_e)\nabla T_e - \Sigma_{e-ph}\left(T_e^n - T_{ph}^n\right) + j^2\rho\left(j, T_e\right) \quad (4.3)$$

while the phonon system is described by

$$\frac{\partial E_{ph}(T_{ph})}{\partial t} = C_{ph}F\,B_{ph}T_{ph}^3\frac{\partial T_{ph}}{\partial t} =$$
$$\nabla \kappa_{ph}(T_{ph})\nabla T_{ph} + \Sigma_{e-ph}\left(T_e^n - T_{ph}^n\right) - C_{ph}F\,\Sigma_{ph-sub}\left(T_{ph}^m - T_{sub}^m\right). \quad (4.4)$$

In these equations, $T_e(x,t)$ is the coordinate (x) and time (t) dependent electron temperature, $T_{ph}(x,t)$ is the coordinate and time-dependent phonon temperature, E_e is the electron internal energy, E_{ph} is the phonon internal energy, $C_{e,BCS}(T_e)$ is the BCS heat capacity, κ_e is the electron thermal conductivity which uses the Bardeen form [112] and the BCS value for the order parameter, $B_{ph} = \frac{2\pi^2 k_B^4}{5\hbar^3 v_{avg}^3}$ describes the phonon heat capacity in the Debye limit, κ_{ph} is the phonon thermal conductivity, Σ_{e-ph} describes the strength of the electron-phonon coupling, n describes the power law of this coupling, Σ_{ph-sub} describes the strength of the phonon coupling to the substrate, m describes the power law of this phonon coupling, $j(t)$ is the time-dependent current density, and $\rho(T_e, j)$ is the temperature and current density dependent resistivity. For the phonon equation, we have explicitly included a scaling factor $C_{ph}F$ which, when greater than unity, increases the phonon heat capacity to account for changes in the phonon density of states due to the amorphous character of the film [129–132]. In this formulation, we intentionally keep the electron-phonon coupling and phonon interface coupling in general exponential

terms in order to keep the set of equations as general as possible. This provides the flexibility to handle effects such as reduced dimensionality of the phonon system or the effects of strong disorder [88].

In the general form, the electron-phonon coupling coefficient has the form

$$\Sigma_{e-ph} = \frac{2\pi^2 k_B^2 N(0)}{3} \frac{T_c}{n T_c^{n-1} \tau_{ep}(T_c)} \tag{4.5}$$

which can be derived by taking a Taylor series expansion of the equation about the reference temperature T_c. This assumes the coupling is the same as in a normal metal rather than a superconductor, with $N(0)$ as the single-spin electron density of states at the Fermi level and $\tau_{ep}(T_c)$ as the electron-phonon coupling time at T_c. In the limit of a 3D normal metal without disorder and assuming a Debye spectrum of phonons, the electron-phonon coupling exponent is $n = 5$ and the coupling coefficient reduces to its familiar form

$$\Sigma_{e-ph} = \frac{2\pi^2 N(0) k_B^2 T_c^2}{3} \frac{1}{5 T_c^5 \tau_{ep}(T_c)}. \tag{4.6}$$

This form corresponds to the same as in the two-temperature models of Chapters 2 and 3 using the relation $\tau_0 = \frac{720\zeta(5)}{\pi^2} \tau_{ep}(T_c)$.

The general phonon-substrate coupling coefficient has the form

$$\Sigma_{ph-sub} = \frac{B_{ph} T_c^3}{m T_c^{m-1} \tau_{esc}(T_c)} \tag{4.7}$$

for arbitrary m while assuming the Debye model. Within the framework of the acoustic mismatch model (AMM) or diffuse mismatch model (DMM), the phonon coupling exponent is $m = 4$. This is the standard form of interface boundary resistance where the phonon-substrate coupling coefficient reduces to the familiar form

$$\Sigma_{ph-sub} = \frac{B_{ph}}{4\tau_{esc}}. \tag{4.8}$$

One of the crucial parameters which influences the electrothermal behavior of the two-temperature model is the way the superconductor transitions to the normal state. This is done primarily through the function $\rho(j, T_e)$ which defines the current and temperature-dependent resistivity of the material. Early works [68] defined this according to a step function, but it is now more common [128] to use a smooth function defining this transition for improved numerical stability. In this work, we

use a the form

$$j_c(T_e) = j_c(0)\mathrm{Re}\left(1 - (T_e/T_c)^2\right)^{3/2} \rightarrow T_{sw}(j) = T_c\mathrm{Re}\left(1 - (j/j_c(0))^{2/3}\right)^{1/2}$$
(4.9)

and

$$\rho(j, T_e) = \rho_0\left(\frac{1}{2} + \frac{1}{2}\tanh\left[\left(\frac{T_e - T_{sw}(j)}{a}\right)\left(\frac{1}{2} - \frac{1}{2}\tanh\left(\frac{T_e - T_{sw}(j)}{b}\right)\right)\right.\right.$$
$$\left.\left. + \left(\frac{T_e - T_{sw}(j)}{c}\right)\left(\frac{1}{2} + \frac{1}{2}\tanh\left(\frac{T_e - T_{sw}(j)}{b}\right)\right)\right]\right).$$
(4.10)

The first equation of (4.9) describes the temperature-dependent switching current according to the Bardeen temperature dependence while the second equation inverts this to describe a current density dependent switching temperature. $\mathrm{Re}()$ indicates taking the real part of the function. The resistivity defined by (4.10) scales the normal state resistivity ρ_0 based on the measured profile of the transition temperature of the WSi used in the latching and hotspot current measurements and we have replaced the zero current transition temperature with the current density dependent switching temperature $T_{sw}(j)$. The fitting parameters are $a = 58.5\,\mathrm{mK}$, $b = 194\,\mathrm{mK}$, and $c = 263\,\mathrm{mK}$. This approach is a simple phenomenological way to address the resistive transition without introducing additional complexity. The use of this transition width at low temperatures and high bias currents is not theoretically justified, but the main dynamics of the system are governed by the behavior near T_c, so we do not believe this simplification in the high bias current regime impacts the results significantly.

Fitting to Experiment

In order to simulate the response of the nanowire system, equations (4.3) and (4.4) must be solved together with a circuit equation and appropriate boundary conditions for the temperature equations. For simulating the latching behavior, the circuit equation is shown in Figure 4.11a, while for the hotspot current simulations, the circuit model is simplified to a voltage source in series with a $10\,\mathrm{k\Omega}$ bias resistor and the nanowire with a small series inductor shown in Figure 4.11b. The electrothermal equations are solved in a 1D geometry and the boundary conditions are $T_e = T_{ph} = T_{sub}$ at the ends of the nanowire. The total resistance of the nanowire is calculated by integrating the resistivity over the device length L according to

$$R_{HS}(t) = \frac{1}{wd}\int_{-L/2}^{L/2}\rho(j(t), T_e(t))dx.$$
(4.11)
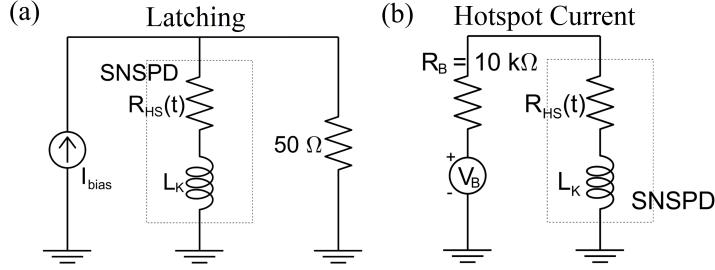
Figure 4.11: Circuit equations used for modeling (a) latching and (b) hotspot current behavior. A capacitor could be added to the output between the SNSPD and $50\,\Omega$ termination of the latching circuit to better represent the presence of a bias-tee, but for the reset time constants probed in this work, the capacitance of a typical bias-tee does not influence the dynamics.

Given the large uncertainty in the material parameters of these WSi films, there are a number of free parameters which could be adjusted to attempt to fit experimental data. The parameters T_c and $\rho_\square$ were measured experimentally, but the diffusion coefficient D was not because this measurement requires additional experimental hardware not present at JPL. Literature results for the diffusion coefficient range from 0.58 to 0.85 cm^2/s for thin WSi films [83, 84, 133]. The films used in these experiments have a lower T_c than those which report the diffusion coefficient [83, 84, 133], which may be an indication of stronger disorder and smaller diffusion coefficient in the current films. The electron-phonon coupling time $\tau_{ep}(T_c)$ and phonon escape time τ_{esc} are fitting parameters, but these can also be constrained based on published results. Once scaled for differences in T_c according to a $1/T^3$ power law, estimates for $\tau_{ep}(T_c)$ range from 182 to 210 ps [83, 84, 133]. The escape time can be approximated using the AMM or DMM to yield a τ_{esc} of 11.7 ps and 8.6 ps respectively for a WSi nanowire embedded in SiO$_2$ with two escape interfaces using the WSi density and sound velocities listed in Section 2.4.3.

One of the important metrics of nanowire quality is the ratio of the switching current to the depairing current achieved in experiment. A recent work [114] measured this ratio by inferring the depairing current in resonator structures through a bias current-dependent resonance measurement followed by fitting the result to a model based on the solution of the Usadel equations [76]. That work found that the ratio of experimental switching current to depairing current ranged from 0.5–0.8 depending on the material system and nanowire width, for measurements performed at 1 K. The depairing current for the devices used in the latching and hotspot current

measurements is estimated by the solution of the Usadel equations [134], leading to

$$I_{dep}(0) = 1.491 \frac{w}{2e\rho_\square} \sqrt{\frac{1}{D\hbar}} (1.764 k_B T_c)^{3/2}. \tag{4.12}$$

In (4.12), the only unmeasured parameter is the diffusion coefficient. Assuming a diffusion coefficient at the high end of the range of $0.85\,\mathrm{cm}^2/\mathrm{s}$ leads to $I_{dep}(0) = 12.1\,\mu\mathrm{A}$ and an experimental to theoretical depairing current ratio of 0.88, while a value at the lower end at $0.58\,\mathrm{cm}^2/\mathrm{s}$ leads to $I_{dep}(0) = 14.7\,\mu\mathrm{A}$ and an experimental to theoretical depairing current ratio of 0.73. The ratio of experimental to theoretical depairing current is more realistic for the smaller value of the diffusion coefficient. It should be noted that the thickness does not actually arise in the expression for the depairing current when expressed in terms of the sheet resistance. Therefore, the choice of diffusion coefficient fixes a certain fraction of the depairing current reached in experiment. In the current work, we use the estimated depairing current as the switching current for both the latching and hotspot current simulations. It would also be possible to consider two different switching currents. For latching simulations, the experimental switching current could be used as the nanowire switching current because the device is limited by the weakest point of the nanowire. In contrast, for hotspot current simulations, the full depairing current could be used because outside of the weakest constriction, superconductivity should be closer to its theoretical value. When measuring the hotspot current the normal domain extends well beyond the constricted region. Introduction of the additional switching current for the latching simulations did not improve fitting, so in all of the following results, we use the estimated depairing current as the switching current for simplicity.

Using these parameters as a starting point, it is not difficult to achieve an approximate fit to the experimental hotspot current, but reproducing the latching behavior is more challenging. There is no combination of $\tau_{ep}(T_c)$, τ_{esc}, and D which is able to match the latching behavior of the devices. Figure 4.12 shows the results of the fitting for $D = 0.6\,\mathrm{cm}^2/\mathrm{s}$ for two sets of fitting parameters. The first is the case where energy dissipation is limited by the electron-phonon coupling channel, with $\tau_{ep}(T_c) = 411\,\mathrm{ps}$ and $\tau_{esc} = 5\,\mathrm{ps}$ while the second is limited by phonon escape, using the literature estimate of $\tau_{ep}(T_c) = 200\,\mathrm{ps}$ and $\tau_{esc} = 46\,\mathrm{ps}$. The free parameters were chosen to achieve correct fitting to the hotspot current at a bath temperature of $1\,\mathrm{K}$, as shown in Figure 4.12a. Unfortunately, these fitting parameters fail to reproduce the latching behavior, as demonstrated in Figure 4.12b.
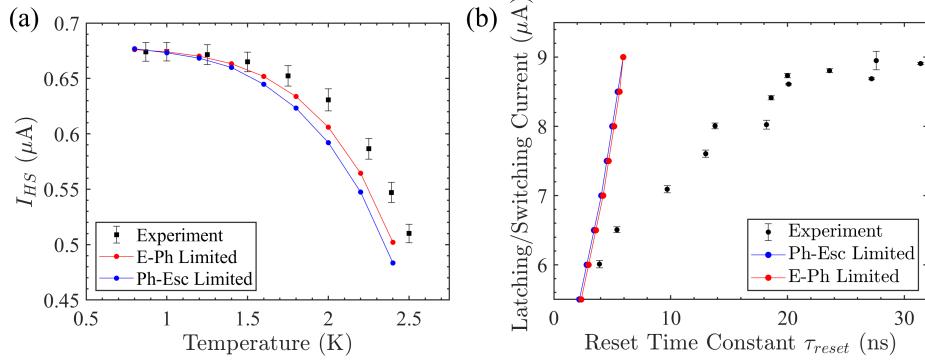
Figure 4.12: (a) Comparison of the experimental hotspot current (black symbols) and the predictions of the two-temperature model for $D = 0.6\,\text{cm}^2/\text{s}$ for two sets of fitting parameters. The first (red) is limited by the electron-phonon coupling channel, with $\tau_{ep}(T_c) = 411\,\text{ps}$ and $\tau_{esc} = 5\,\text{ps}$ while the second (blue) is limited by phonon escape, using the literature estimate of $\tau_{ep}(T_c) = 200\,\text{ps}$ and $\tau_{esc} = 46\,\text{ps}$. (b) Comparison with the experimental latching currents at a bath temperature of 1 K for the same fitting parameters shows poor agreement.

It was found by Sidorova et al. [133] that the estimated heat capacity of the phonon system based on fitting a linearized electrothermal model to experiment was higher than predicted by the Debye model by a factor of 5 to 6 [133]. If the phonon heat capacity is increased by a factor of $C_{ph}F$, a better fit to the latching data can be found in the limit of unrealistically large $C_{ph}F$, as shown in Figure 4.13a. For this fitting, the escape time was increased by an equivalent factor $C_{ph}F$ in order to maintain the same hotspot current behavior in the stationary state. It is this increase in τ_{esc} that is primarily responsible for the better fitting to the latching behavior with $\tau_{esc} = 3.68\,\text{ns}$ for the best fitting case, but the increased escape time alone is not sufficient to reproduce the experimental latching behavior. If τ_{esc} is increased without increasing the phonon heat capacity, the two temperature model is unable to produce the correct latching current trend, as shown in Figure 4.13b.
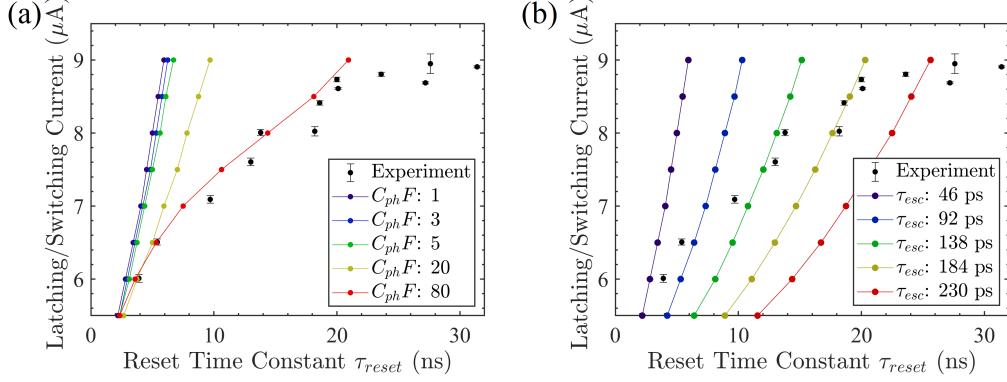
Figure 4.13: (a) Comparison of the experimental latching behavior (black symbols) and the predictions of the two-temperature model for $D = 0.6\,\text{cm}^2/\text{s}$ when varying $C_{ph}F$ as a free parameter for the phonon escape limited fitting parameters listed above. For these simulations, τ_{esc} was also scaled according to $C_{ph}F$. The hotspot current is the same as the phonon escape limited case shown by the blue curve in 4.12a. (b) Comparison of the experimental latching behavior (black symbols) and the predictions of the two-temperature model for $D = 0.6\,\text{cm}^2/\text{s}$ when varying τ_{esc} and keeping $C_{ph}F = 1$. The hotspot current for these parameter sets does not match experiment due to the modified combination $C_{ph}F\Sigma_{ph-sub}$.

There is a difference in the quality of the fit for the various parameter combinations for the two-temperature model when compared to the experimentally measured hotspot current. In the unrealistic case of an escape time of 5 ps, the rolloff of the temperature dependence of the hotspot current better matches experiment compared to the other parameter combinations. This is explained by the limiting thermal link in the system. With such a fast escape time, the limiting link is the electron-phonon coupling, which has a 5th-power coupling relation, compared to the 4th-power relation for a phonon escape limited link. A higher order power law limiting thermal link leads to a sharper rolloff upon approaching T_c, so the experimental results suggest that the channel responsible for bottlenecking thermal dissipation to the substrate is of a high order (> 5). Based on this preliminary fitting, the two-temperature model is insufficient to describe the dynamics of the WSi system and must be modified in order to match experimental data.

4.2.2 Three-Temperature Model

The two-temperature model can be extended in a straightforward matter for the case where phonons cannot be approximated as having a constant escape time for all phonon states. In particular, if phonons with a large incidence angle at the interface experience total internal reflection, these phonons may have a much longer lifetime

than those within the critical escape cone. Bezuglyj and Shklovskij discuss this case using the solution of the kinetic equation (see Chapter 2) and conclude that for the case where the characteristic phonon-electron scattering rate $\nu = 1/\tau_{pe}$ is slower than the time for phonons to escape the film $d/(c\alpha_0)$ when their momentum is within the critical cone (α_0 is the transmission probability of the interface), the two-temperature model is not accurate [80]. Instead, they show that a three-temperature model accurately captures the dominant pole in equations governing the energy exchange of a thin film with the substrate. In this limit, the fastest dynamics of the energy exchange are not captured accurately, but the main cooling effect is governed by the dominant pole. Thin WSi films in the vicinity of $T_c \sim 3\mathrm{K}$ operate in this limit where the two-temperature model is not accurate (see Section 2.4 of Chapter 2), so use of the three-temperature model is justified.[5]

We formulate the three-temperature model starting from its existing form [80], but modify it to include the possibility of excess heat capacity in the phonon system by a factor $C_{ph}F$ as was done with the two-temperature model. The model consists of a population of escaping phonons ($T_{ph,1}$) which interact with the substrate and a population of trapped phonons ($T_{ph,2}$), which undergo total internal reflection at the interface. We add the excess heat capacity $C_{ph}F$ to the trapped phonon population because it represents the additional states caused by the amorphous nature of the film, and these oscillator states will not have the large group velocity characteristic of low-energy acoustic modes. Thus, they are approximated as being trapped in the film in the same way as the bottlenecked phonons. In the simplified treatment here, these additional phonons maintain a Debye spectrum rather than a more realistic linear spectrum [129–132] and assume they have the same coupling strength to the electronic system. Furthermore, we add an additional scattering mechanism between the population of escaping phonons and trapped phonons using a Rayleigh-like energy dependence $\tau_S \sim \omega^{-4}$, leading to a T^8 power law coupling the energies of these two groups. This approach was used in [129] to qualitatively understand the temperature dependence of the phonon mean free path in amorphous dielectrics. This model is the generalized form of the linearized three-temperature model used by Sidorova et al. to account for shortcomings in the two-temperature model fitting of amplitude modulated absorption of radiation (AMAR) data [133] for thin WSi

---

[5]Even more accurate than the three-temperature model is the one-temperature model described in Chapter 2, but solving the integro-differential equation of the one-temperature model over long time periods becomes challenging. We will not consider it here under the justification that the relaxation process of the superconductor should be dominated by the slowest pole of the system, which should be accurately captured by the three-temperature model [80].

films.

Under these assumptions, the energy balance equation for the electron system is expressed as

$$\frac{\partial E_e(T_e)}{\partial t} = C_{e,BCS}(T_e)\frac{\partial T_e}{\partial t} = \nabla\kappa_e(T_e)\nabla T_e - \frac{\alpha}{C_{ph}F}\Sigma_{e-ph}\left(T_e^5 - T_{ph,1}^5\right)$$
$$-\frac{C_{ph}F - \alpha}{C_{ph}F}\Sigma_{e-ph}\left(T_e^5 - T_{ph,2}^5\right) + j^2\rho(j,T_e) \tag{4.13}$$

while the phonon system is described by

$$\frac{\partial E_{ph,1}}{\partial t} = \alpha B_{ph}T_{ph,1}^3\frac{\partial T_{ph,1}}{\partial t} =$$
$$\frac{\alpha}{C_{ph}F}\Sigma_{e-ph}\left(T_e^5 - T_{ph,1}^5\right) - \frac{\alpha B_{ph}}{4\tau_{esc}}\left(T_{ph,1}^4 - T_{SiO_2}^4\right) - \Sigma_s\left(T_{ph,1}^8 - T_{ph,2}^8\right). \tag{4.14}$$

$$\frac{\partial E_{ph,2}}{\partial t} = \left(C_{ph}F - \alpha\right)B_{ph}T_{ph,2}^3\frac{\partial T_{ph,2}}{\partial t} =$$
$$\frac{C_{ph}F - \alpha}{C_{ph}F}\Sigma_{e-ph}\left(T_e^5 - T_{ph,2}^5\right) + \Sigma_s\left(T_{ph,1}^8 - T_{ph,2}^8\right). \tag{4.15}$$

In the three-temperature model, α is the fraction of the Debye phonon heat capacity in the escaping population. For a given substrate material, the AMM can be employed to estimate this fraction of the phonon population in the critical escape cone. Using the same parameters for WSi as throughout this chapter, this defines the parameter $\alpha = 0.318$, corresponding to a critical angle of $\sim 47°$. In (4.14) and (4.15) $\Sigma_s = \frac{C_{ph,2}(T_c)}{8(T_c)^7\tau_s(T_c)}$ defines the coupling strength between the two phonon populations, and Σ_{e-ph} retains its definition from the two-temperature model (4.6). The critical current density, resistance, and circuit equations are handled using the same form as the two-temperature model.

Fitting to Experiment

Starting with reasonable estimates for D, $\tau_{ep}(T_c)$, and α while fixing $\tau_{esc} = 2\,\text{ps}$, we can attempt to fit the experimental hotspot current and latching data. However, as with the application of the two-temperature model, the three-temperature model is inadequate to simultaneously reproduce the experimental measurements of the hotspot and latching currents. The same general trends apply, but the results provide more insight into the ranges of material properties and thermal coupling relations that govern the hotspot and latching behavior. Application of the model using theoretical estimates of the electron-phonon coupling times and using the escape

time to fit the low bath temperature hotspot current, the fitting resembles that of the two-temperature model in the case of electron-phonon coupling limited loss. This is shown in Figure 4.14a by the purple curve. If the phonon heat capacity is increased substantially, the flow of energy in the system is altered to favor coupling from the electrons to trapped phonons to the escaping phonons, rather than directly from the electrons to escaping phonons. This is due to the factor $\alpha/C_{ph}F$ which determines the fraction of the electron-phonon coupling which couples directly to the escaping phonon group. With sufficient enhancement of $C_{ph}F$, the temperature dependence of the hotspot current reflects this shift in energy flow, leading to a sharper rolloff of the hotspot current as shown in Figure 4.14a. This corresponds to the thermal bottleneck becoming the T^8 coupling link between the phonon groups rather than the T^5 coupling between electrons and phonons. As with the two-temperature model, a large enhancement of the phonon heat capacity is required to shift the latching current toward the experimental results, as shown in Figure 4.14b, and is accompanied by an increase in τ_s in order to maintain fitting to the hotspot current results. Rather than increasing $C_{ph}F$, the same trend in hotspot current shifting can be achieved by reducing α, which then does not lead to a significant shift in the latching behavior. Unlike the two-temperature model, the three-temperature model is unable to reproduce the experimentally observed latching behavior, even for the case of increased $C_{ph}F$. Instead, at high bias currents, the system begins to afterpulse rather than latch. This feature is not observed experimentally, and will be discussed below.
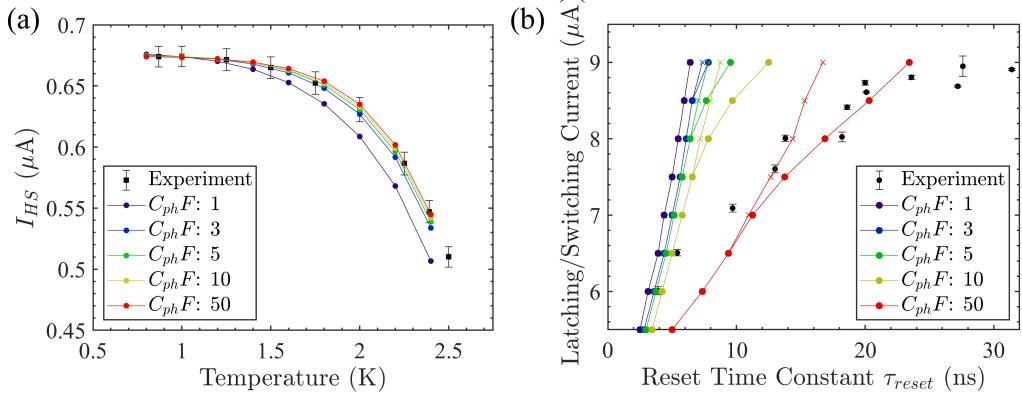
Figure 4.14: (a) Comparison of the experimental hotspot current (black symbols) and the predictions of the three-temperature model for $D = 0.6\,\mathrm{cm^2/s}$, $\tau_{ep}(T_c) = 150\,\mathrm{ps}$, $\alpha = 0.316$, and $\tau_{esc} = 2\,\mathrm{ps}$ for various combinations of τ_s and $C_{ph}F$. The combinations are $C_{ph}F = 1$, $\tau_s = 0.156\,\mathrm{ns}$; $C_{ph}F = 3$, $\tau_s = 0.075\,\mathrm{ns}$; $C_{ph}F = 5$, $\tau_s = 0.105\,\mathrm{ns}$; $C_{ph}F = 10$, $\tau_s = 0.187\,\mathrm{ns}$; $C_{ph}F = 50$, $\tau_s = 0.853\,\mathrm{ns}$, with each combination shown by a different color in the plots. A comparison to the latching experimental data is shown in (b) for the same parameter combinations. With this formulation, we see the appearance of afterpulsing (circles) in addition to latching (x's) at high bias currents.

4.2.3 Discussion on Latching and Hotspot Current

The two-temperature and three-temperature models fail to provide a reasonable match to experiment, but we can still use their predictions to draw conclusions about the nature of the thermal relaxation process in SNSPDs. Based on the hotspot current measurements as a function of bath temperature, we can conclude that the thermal relaxation pathway is dominated by a high order power law, with an exponent close to 6. This is based on the shape of rolloff of hotspot current as the bath temperature approaches T_c. The standard two-temperature model does not predict this type of power law, and the three-temperature model only recovers this dependence when energy flows predominantly from the electron system to the bottlenecked phonons before scattering into escaping phonon modes. Such a process is plausible for thin WSi films. The dominant phonon wavelength of the equilibrium energy distribution at a given temperature is given by $\lambda_{ph_E} \approx \frac{2\pi\hbar v_{avg}}{2.82 k_B T}$. For $T_c = 3\,\mathrm{K}$, this corresponds to $\lambda_{ph_E} \approx 19\,\mathrm{nm}$, or a factor of ~4 times the thickness of the film. At this temperature, the use of the 3D Debye model is highly suspect, and with it, the use of a phonon escape term with a 4$^{\text{th}}$-order power law. In this regime, phonons will be subject to dimensionality constraints within the film, and are likely defined by modes which extend into the surrounding dielectric substrate. In this context, the

idea that these modes scatter via impurities according to a Rayleigh-like frequency dependence leading to a high order power law governing the escape of energy from the phonon system seems plausible, but a more detailed investigation of the the phonon modes for thin films embedded in dielectrics is needed.

For both the two-temperature and three-temperature models, reasonable fitting to the latching experimental data can only be achieved when scaling the phonon heat capacity by a large multiple on the order of ~ 80. This has the effect of increasing τ_{esc} without changing the hotspot current. Such a significant scaling does not seem consistent with the additional states which present due to the amorphous structure of the film, or due to the reduced dimensionality of the film. Furthermore, if the lattice heat capacity were this large, WSi nanowires would not be sensitive to near-infrared photons due to trapping of heat in the phonon system, which is contrary to experimental observation. Instead, this result suggests that the dynamics of the dielectric substrate cannot be decoupled from the nanowire and play an important role in determining the reset characteristics of WSi SNSPDs. A factor of 80 times the estimated WSi heat capacity describes the bulk SiO_2 heat capacity of a cross-section equal width of the wire and a thickness of $\sim 300\,nm$, which is surprisingly close to the thickness of the dielectric at $\sim 350\,nm$. However, the estimated escape time is only $\sim 1.8\,ns$ for a $\sim 350\,nm$ film based on the AMM (scaling the Debye heat capacity by a factor of ~ 3 to include non-Debye modes based on bulk heat capacity measurements [129]), compared to $\tau_{esc} = 3.68\,ns$ needed to fit the experimental data. This is an oversimplification of the true heat transfer in this type of system, but it supports the idea that the interaction between thin dielectric layers and SNSPDs deserves further investigation.

Using the generalized form of the two-temperature model, we can attempt to fit both sets of data simultaneously using a modified form of the phonon escape term with $m = 6$ and $m = 7$ in (4.4). Using $D = 0.6\,cm^2/s$ and $\tau_{ep}(T_c) = 200\,ps$ based on literature results, the data can be reasonably matched using the parameters $C_{ph}F = 100$ and $\tau_{esc} = 2.31\,ns$ for $m = 6$ and $C_{ph}F = 100$ and $\tau_{esc} = 1.85\,ns$ for $m = 7$. These results are shown in Figure 4.15. While this model can technically reproduce the experimental results, we should remain skeptical about directly using it to infer properties about the WSi material system. Instead, it is a qualitative guide about the nature of the steady-state thermal dissipation and the existence of a thermal bottleneck with a time constant on the order of a few nanoseconds.
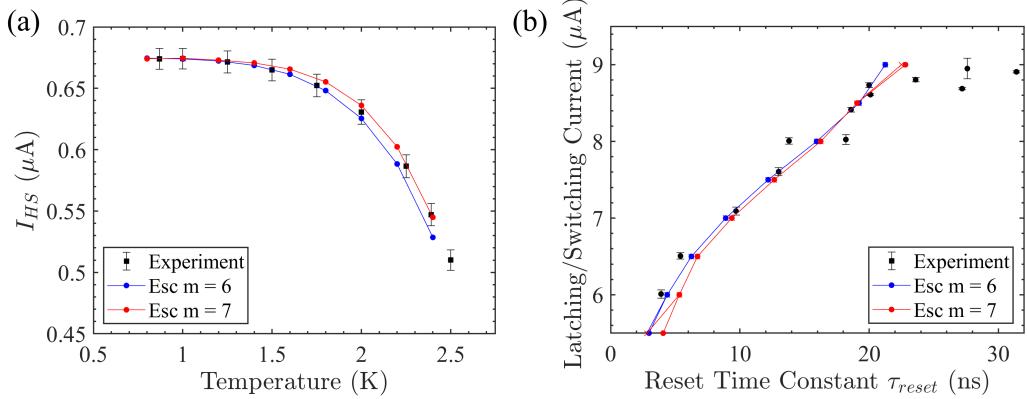
Figure 4.15: Comparison of the experimental (a) hotspot current and (b) latching behavior to the predictions of the modified two-temperature model for $D = 0.6\,\text{cm}^2/\text{s}$, $\tau_{ep}(T_c) = 200\,\text{ps}$, $C_{ph}F = 100$ and $\tau_{esc} = 2.31\,\text{ns}$ for $m = 6$ and $C_{ph}F = 100$ and $\tau_{esc} = 1.85\,\text{ns}$ for $m = 7$. The modification of the thermal bottleneck power law and the substantial increase in phonon heat capacity enable fitting the experimental data.

Latching, Paralysis, Electrothermal Oscillations, and Afterpulsing

Latching is not the only behavior which adversely affects the reset dynamics of SNSPDs. Afterpulsing is a process where the detector has a high probability of triggering a second time during the recovery process after a detection event, and can be caused by electrical reflections, or electrical recovery faster than the system can fully thermally recover. This effect, which has been observed in SNSPDs, can lead to spurious false counts correlated in time with true photon counts as well as contribute to the deadtime of the detector. Both of these effects are detrimental to overall performance.

Less studied or discussed in the literature is the process by which a nanowire takes an unusually long time to recover to the superconducting state. This is often also referred to as latching because it typically precedes true latching, but for clarity, we propose defining this behavior as paralysis and will refer to it as such through the remainder of this work.[6] Paralysis appears to be caused by thermal effects where thermal dissipation is initially too slow for the system to recover to the superconducting state before the current begins to return to the nanowire, but the dissipation is sufficiently fast to prevent a stable self-heating hotspot to form, so the detector does not fully latch.

---

[6]The term paralysis was used by Dr. Marsili to describe this behavior, in reference to an effect in SPADs which leads to a similar behavior of increased deadtime.

The WSi detectors studied in the latching devices demonstrated significant paralysis upon approaching the latching or switching current. This alone provides a metric for comparing simulation because it contains information about the nature of thermal relaxation, but more importantly, there is structure in the way the system recovers to the superconducting state. The devices exhibit electrothermal oscillations while in the paralyzed state, which makes the detector more likely to recover at times aligned with the oscillations. Figure 4.16 shows a persistence trace, which acts as a histogram for many pulse trace acquisitions, for a detector with $\tau_{reset} = 27.2$ ns. The preferred recovery paths are clearly visible with warm colors representing a higher density of traces. This particular trace is also informative because it shows how electrical reflections can provide feedback and influence the system. At a delay time of approximately 47 ns from the trigger, an electrical reflection from the input of the room temperature RF amplifier returns to the detector, leading to a sharp peak in the signal. This increases the probability of knocking the system out of its precarious paralyzed state, which is seen as an increase in trace density immediately following the reflection. The other feature of interest which emerges from the persistence traces is the dependence of the recovery pathway on the origin of the counts. Intrinsic dark counts suffer more paralysis effects than standard photon counts. This is consistent with the idea of constrictions being the source of these counts. In a constricted region, the bias current approaches the local switching or depairing current of the system, which is lower than elsewhere in the device. For example, if the constriction is caused by a reduced T_c in a given area, the detector must cool to a lower temperature before it can recover when compared to the rest of the device. Therefore, in a device exhibiting paralysis, the effect would be exacerbated in an area of reduced T_c.

(a) Dark Counts
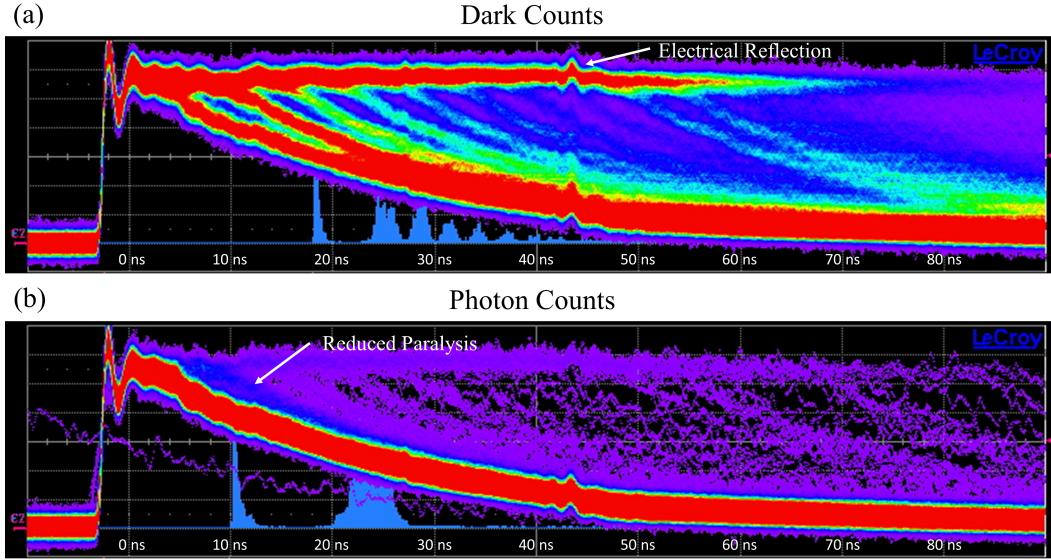
(b) Photon Counts

Figure 4.16: Paralysis in the WSi detector with an electrical recovery time constant of 27.2 ns for (a) dark counts (b) photon counts when biased at 8.6 μA at a bath temperature of 875 mK. The paralysis effect is much more pronounced for dark counts compared to photon counts.

Latching, paralysis, and afterpulsing are all predicted by the electrothermal model when operating in different conditions and with different material parameters. This acts as a metric for qualitatively checking the conclusions of the electrothermal model. For the standard two-temperature model, if the heat capacity factor remains small with $C_{ph} = 1$, paralysis does not occur over a wide range of biases or reset times. The system either latches or recovers. However, as the escape time increases, paralysis occurs over a wider range of bias conditions. Figure 4.17 shows the response of the system within the standard two-temperature model for $C_{ph}F = 80$. The results of the model are qualitatively similar to experiment. This suggests that the treatment of an electron system thermally coupled to a temperature bath with a long escape time is qualitatively consistent with experimental observations.

(a)

(b)
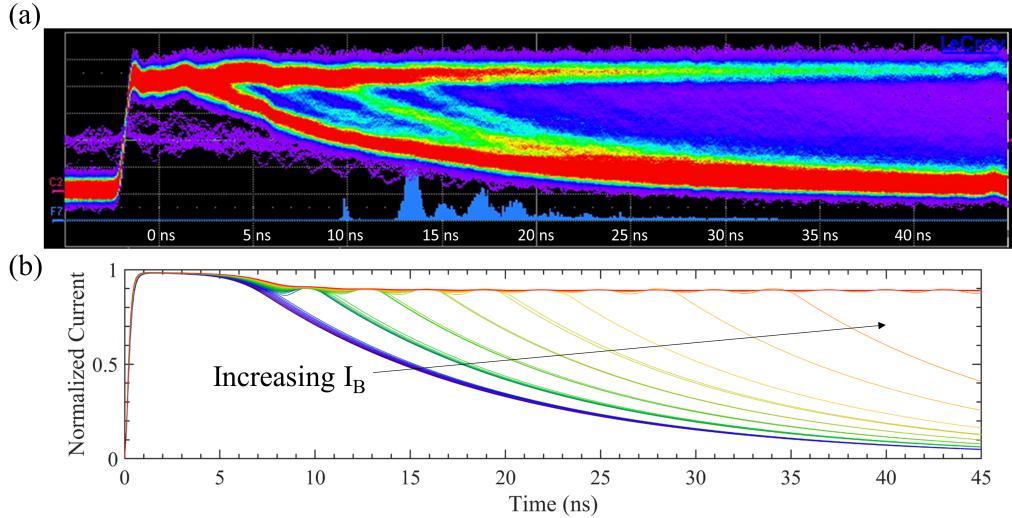
Figure 4.17: Comparison of experimental (a) and model (b) paralysis for a WSi detector with an electrical recovery time constant τ_{reset} of 13 ns at a bath temperature of 1 K. The device was biased at 7.3 µA for the measurement. For the simulation, the bias current was increased from 6.5 µA to 8.0 µA in order to show the transition through the various stages of paralysis, but in a real device, random noise leads to sampling the various relaxation pathways.

For the three-temperature model, increasing the heat capacity in order to match latching behavior leads to the appearance of afterpulsing in addition to paralysis. Afterpulsing was not observed in experiment, so this suggests that this form of the model is incorrect. This is actually reassuring because such a model would imply that the excess heat capacity is located entirely in the nanowire system, which is not compatible with our understanding of the WSi material system. A similar result occurs if an excess heat capacity is coupled only to the electronic system (not shown here). Afterpulsing in this context is due to slow cooling of the system at temperatures below T_c. Initially the phonon and electron systems cool quickly enough for superconductivity to return and for current to begin flowing in the nanowire. However, as the temperature of the system drops toward the substrate temperature, current returns faster than the system cools, and the temperature-dependent critical current is exceeded, leading to a second switching event. This behavior was observed more frequently in models which had a high order power law governing the thermal weak link, which could be due to the significantly reduced coupling strength when the temperatures are close together.

All of this qualitative analysis points toward the substrate as the source of the discrepancies between the latching and hotspot current results when analyzed in the context

of a generalized 1D electrothermal model. Paralysis becomes prominent when the two-temperature formulation features a thermal bottleneck with the substrate and electrothermal oscillations appear in a manner consistent with experiment. This result further demonstrates that a more detailed understanding of heat flow in thin dielectrics is needed in order to fully model the reset dynamics of low T_c SNSPDs.

4.2.4 Nanowire Thermometry

When considering the nanowire thermometry experiment, it does not seem necessary to simulate the full dynamics of the WSi nanowire coupled to the SiO_2 dielectric. Because the nanowire acts as a steady heater, the current and voltage characteristics can be used to determine the total amount of power dissipated by the nanowire. The only thermal sink for the nanowire is the SiO_2 dielectric, so the distribution of energy between the electrons and phonons is unimportant. The only relevant parameter is the total heat dissipated per unit length of the nanowire. Furthermore, because the nanowires are designed to be long compared to any thermal diffusion length scales, we can simplify the analysis to a 2D system and only consider the lateral heat flow in the dielectric layer.

It is known that heat flow in thin materials can be significantly different than that in bulk materials [135]. Scattering at surfaces can lead to reduced thermal conductivity by reducing the phonon mean free path, and the limit where phonons scatter diffusely at all surfaces is known as the Casimir limit. Experimentally, it has been observed that the thermal conductivity based on the kinetic equations can be even smaller than expected based on the Casimir limit [135]. This was attributed to macroscopic defects in the material. Based on these observations, we use the kinetic form of the thermal conductivity for the SiO_2 layer, describing it as

$$\kappa_{SiO_2}(T) = \frac{1}{3}c_{avg}l(T)C_D(T) \tag{4.16}$$

where T is the local SiO_2 temperature, $\kappa_{SiO_2}(T)$ is the thermal conductivity, $c_{avg} = \left(\frac{3}{\frac{1}{c_l^3} + \frac{2}{c_t^3}}\right)^{1/3}$ is the mode-averaged sound velocity, c_l is the longitudinal sound velocity, c_t is the transverse sound velocity, $l(T)$ is the phonon mean free path, and $C_D(T)$ is the heat capacity of Debye phonons. The mean free path is the average of a bulk component found using (4.16) for the bulk thermal conductivity, leading to $l_{bulk}(T) = \frac{3\kappa_{SiO_2}(T)}{c_{avg}C_D(T)}$, and a second mean free path l_0 which originates from an energy-independent scattering mechanism based on the geometry of the sample [135]. The average mean free path, found by averaging the scattering times using

Mattiessen's Rule, is given by

$$l(T)^{-1} = l_{bulk}(T)^{-1} + l_0^{-1}. \tag{4.17}$$

The boundary condition between the SiO_2 and silicon substrate is given by the general thermal boundary resistance form

$$\kappa_{SiO_2}(T)\nabla T \cdot \hat{n}|_{interface} = -R_1\left(T^{r_1} - T_{Si}^{r_1}\right) \tag{4.18}$$

where R_1 describes the magnitude of the heat flux while r_1 describes the exponential relation of the thermal coupling. Unlike during normal SNSPD operation, the amount of power dissipated along the entire length of the nanowire can be substantial, on the order of several mW. This is enough power to raise the temperature of the entire silicon chip, and must be accounted for in the model. It is expected that the thermal boundary between the silicon and the substrate holder limits this power dissipation, leading to a local increase in temperature of the silicon die. We account for this within the model by assuming an additional thermal boundary described by

$$T_{Si} = \left(\frac{P_{total}}{R_2} + T_{bath}^{r_2}\right)^{1/r_2} \tag{4.19}$$

where T_{Si} is the temperature of the silicon substrate, P_{total} is the total power dissipated by the nanowire, R_2 describes the magnitude of the interface coupling, and r_2 is the exponent of this power law coupling. We make the assumption that the T_{Si} is uniform throughout the entire chip. This is justified because the phonon mean free path in the Si is large compared to its thickness. Note that the factor of area from typical Kaptiza-like boundaries such as (4.18) has been absorbed in the coefficient R_2.

The heat flow was modeled using COMSOL Multiphysics by solving for the steady-state temperature profile of the SiO_2 layer, $T(\vec{x})$, for various amounts of dissipated power. Fitting was attempted for two devices. The first used a nanowire pitch of 1600 nm while the second used a nanowire pitch of 400 nm. Figure 4.18 shows the comparison between the experimental data and the model for the best set of fitting parameters. The fitting parameters are l_0 describing the thermal conductivity, R_1 and r_1 describing the SiO_2-Si interface, and R_2 and r_2 describing the silicon-bath interface. The primary goals were to understand the thermal conductivity and the interface boundary resistance between the SiO_2 and silicon. Because these devices were fabricated on the same wafer, we expect the interface properties R_1 and r_1 to be the same and restrict the fitting to enforce this condition. Intuition would suggest

that the mean free path l_0 should be the same for both devices, but a reasonable fit to experiment could only be achieved when different mean free paths were used. The 1600 nm pitch device operated at a bath temperature of 950 mK using fit parameters of $l_0 = 360$ nm, $R_2 = 3.80\,\mu W/K^4$, and $r_2 = 4$. The 400 nm pitch device operated at a bath temperature of 1000 mK and we use the fitting parameters $l_0 = 190$ nm, $R_2 = 3.0\,\mu W/K^4$, and $r_2 = 4$ to describe its behavior. Both devices use the common fit parameters of $R_1 = 66\,W/m^2K^5$ and $r_1 = 5$. It is not surprising that there is a difference between the parameters R_2 used to fit the devices. This boundary is dictated by how the silicon die was thermalized to the gold-plated copper mount of the cryostat, and was affixed using GE Varnish. There should be no reason to assume that the way these devices were mounted would lead to identical thermal boundary properties.
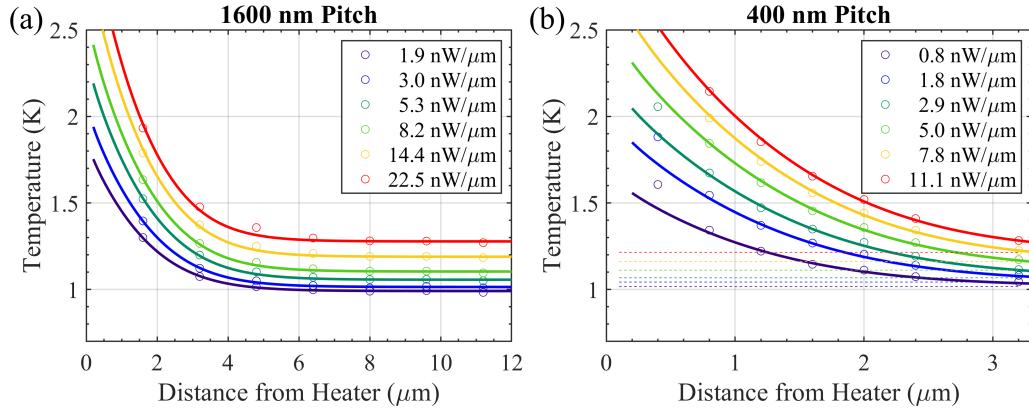
Figure 4.18: Comparison of nanowire thermometry experiment and model for (a) 1600 nm pitch and (b) 400 nm pitch devices. Circles represent the experimental data while the lines are the fit using the model. The dashed lines in (b) represent the Si temperature from the fitting. The 1600 nm pitch device operated at a bath temperature of 950 mK using fit parameters of $l_0 = 360$ nm, $R_2 = 3.80\,\mu W/K^4$, and $r_2 = 4$. The 400 nm pitch device operated at a bath temperature of 1000 mK using fit parameters of $l_0 = 190$ nm, $R_2 = 3.0\,\mu W/K^4$, and $r_2 = 4$. Both devices use a common fit of $R_1 = 66\,W/m^2K^5$ and $r_1 = 5$ to describe the interface between the SiO_2 and silicon because the devices were fabricated on the same wafer.

From this experiment, we can draw several conclusions about the heat transfer process in thin SiO_2 dielectric layers. First, heat is localized to within about 5 μm of a heat source in the thin dielectric. Beyond this distance, heating only occurs when the entire silicon die heats due to the large power dissipation from the entire device. The second interesting result is that the mean free path required to fit the 400 nm

pitch data is much smaller than expected even from the Casimir limit of surface scattering. This data also requires a smaller mean free path than required to fit the 1600 nm pitch results. This discrepancy suggests that the nanowires used to probe the temperature may affect the heat flow within the dielectric by causing additional scattering. This scattering would reduce the effective mean free path of the phonons and reduce the thermal conductivity, as needed to fit the experimental results using the diffusion-based model. Another conclusion is that the use of the standard AMM does not provide an adequate fit to the experimental data. The higher power of $r_1 = 5$ compared to the expected $r_1 = 4$ from the AMM could be due to the amorphous nature of the dielectric, or it could be an indication that the phonons of the SiO$_2$ are not well thermalized and therefore cannot be represented by an equilibrium distribution at T. There is good reason to fear that the phonons are not thermalized. Figure 4.19a shows the predictions of the AMM for an interface between SiO$_2$ and Si. Below $\sim 50°$, there two materials are acoustically well matched, but above that angle, phonons are totally internally reflected in the SiO$_2$. Looking at the phonon mean free path in the bulk material, shown in Figure 4.19b, it clear that for the temperatures of interest for WSi SNSPD operation, the mean free path of phonons is several microns, and therefore much longer than the thickness of the dielectric layer. We can conclude that the assumption of thermalized phonons used in the diffusion-based model is highly suspect.
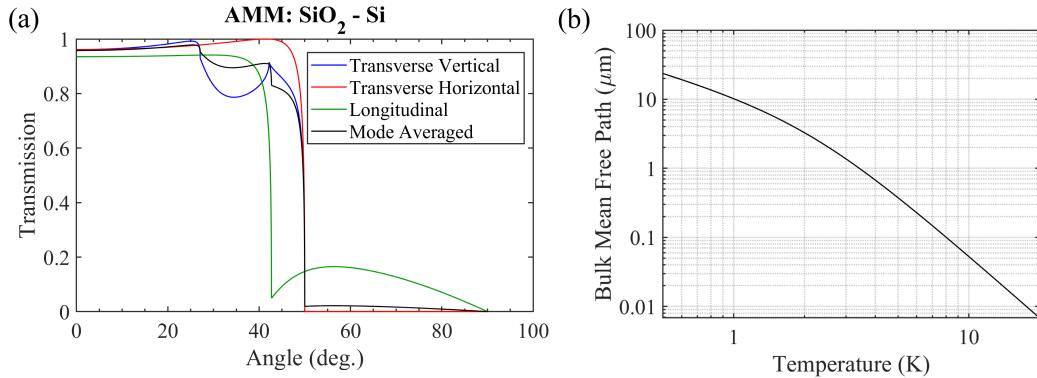
Figure 4.19: (a) AMM predictions of the phonon transmission for an SiO$_2$ - Si interface. (b) Bulk phonon mean free path for SiO$_2$.

Motivated by the concerns over the treatment of the SiO$_2$ and inspired by the approach of the three-temperature model, we construct a two-temperature model to describe the SiO$_2$ system. As with the three-temperature model, we define one population of phonons which is within the escape cone and capable of coupling

to the substrate while a second population is trapped. Coupling between the two groups of phonon occurs due to scattering within the SiO_2 and scattering due to coupling with the nanowires. In this simplified approach, we define the escaping phonons according to T_1 and the trapped phonons with T_2. The heat capacity and thermal conductivity of the T_1 group are scaled by the factor $\beta = 0.29$ which is the fraction of Debye phonons in the escape cone. The heat capacity and thermal conductivity of the T_2 group are scaled according to $(1 - \beta)$. We can define the scattering from the trapped group to the escaping group according to

$$\frac{\partial E_{2-1}}{\partial t} = \beta \frac{C_{ph}(T_2)}{\tau_{mfp}(T_2)} (T_2 - T_1) \tag{4.20}$$

where $\tau_{mfp}(T) = l(T)/c_{avg}$. This linearized form is only valid in the limit of small temperature differences between the two bodies of phonons and assumes that the scattering is isotropic. The phonon mean free path is defined according to (4.17) with l_0 taken as a fitting parameter. Under the assumption that scattering is limited by the presence of the probe nanowires, we enforce that l_0 for the 1600 nm pitch device is four times that of the 400 nm pitch device. The energy flux of the escaping group of phonons to the substrate is given by the AMM with the form of (4.18) with $r_1 = 4$ and $R_1 = 198 \, W/m^2 K^4$.

A second scattering contribution comes from interactions of phonons with the nanowires. This is modeled by assuming that these interactions are dominated by coupling of phonons to the electronic system leading to a boundary flux of the form

$$\kappa_{SiO_2}(T_1)\nabla T_1 \cdot \hat{n}|_{interface} = \beta \Sigma_{e-ph} d \left(T_{WSi}^5 - T_1^5 \right) \tag{4.21}$$

and

$$\kappa_{SiO_2}(T_2)\nabla T_2 \cdot \hat{n}|_{interface} = (1 - \beta) \Sigma_{e-ph} d \left(T_{WSi}^5 - T_2^5 \right) \tag{4.22}$$

for each of the phonon populations where Σ_{e-ph} is defined in the same way as the two-temperature model with (4.6). However, for realistic values of the electron-phonon coupling time $\tau_{ep}(T_c) \sim 200 \, ps$, this term does not significantly thermalize the two phonon populations of the SiO_2. Qualitatively, the behavior is dominated by the parameters l_0 and R_1.

If we use the same R_2 and r_2 parameters as with the diffusive modeling (as a reminder, these parameters determine the temperature of the silicon substrate), then we are left with a single fitting parameter l_0. Figure 4.20 shows the fitting to experiment using this alternative approach where l_0 is the only free parameter and l_0 for the 1600 nm

pitch device is four times that of the 400 nm pitch device based on the nanowire fill factor. Under these conditions, only a coarse fitting to experiment can be achieved, and it requires a mean free path of ~ 180 nm for the 400 nm pitch device (~ 720 nm for the 1600 nm pitch device) which is close to the value found for the 400 nm pitch device using the previous model. While better fitting can be achieved by fitting the R_1 and r_1 parameters defining the heat flux through the SiO_2 - Si interface, the current result is sufficient to show that the discrepancy between the fitting required for the two different pitch devices can be attributed to phonon scattering off the WSi probe nanowires. This scattering does not appear to be mediated by the electron system of the probes. While this result means that the nanowire thermometry technique is not ideal for extracting thermal conductivity properties of pristine thin films, it highlights the importance of the presence of nanowires and their fill factor when considering thermal dissipation. This is relevant for understanding thermal crosstalk between nanowires.
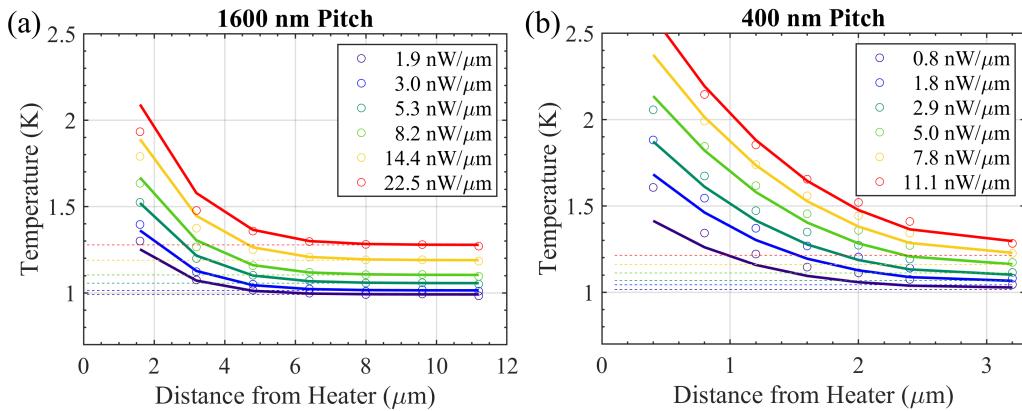
Figure 4.20: Fitting of nanowire thermometry experimental data using the two-temperature SiO_2 model for the (a) 400 nm pitch and (b) 1600 nm pitch devices.

4.2.5 Electrothermal Modeling in 3D

Simulating features of the more complex dynamics of SNSPDs such as thermal crosstalk requires combining the electrothermal model for the nanowire with a thermal transport model for the SiO_2. As we saw with the electrothermal modeling of the latching behavior of WSi devices, the only way to achieve a reasonable fit to experiment was to assume an unrealistically large additional heat capacity for the phonon system. This hinted at the importance of coupling to the dielectric when considering the timescale of thermal dissipation.

To model the dynamics of the nanowire coupled to a substrate, we perform finite

element simulations in a 3D geometry in COMSOL Multiphysics. Performing full 3D simulations over a geometry which includes the nanowire is not practical due to the small length scales involved. Instead, we treat the nanowire as a 1D domain and couple the phonon system to the 3D domain of the substrate and exploit the symmetry of the system to model only a single quadrant of the system. This coupling is achieved by converting the energy dissipation of the nanowire into a heat flux at the interface of the SiO_2 domain using the form

$$\dot{Q}(x, y) = \frac{d}{2} \Sigma_{ph-esc} \left(T_{ph}(x)^4 - T_{SiO_2}(x, y)^4 \right) \tag{4.23}$$

where the factor d converts the dissipation per unit volume to a flux per unit area and the factor of 1/2 accounts for the presence of two interfaces with the SiO_2. Similarly, the energy flow from the nanowire to the SiO_2 is treated as

$$\begin{aligned}
\frac{\partial E_{ph-sub}(x)}{\partial t} &= \frac{1}{2w} \int_{-w/2}^{w/2} \Sigma_{ph-esc} \left(T_{ph}(x)^4 - T_{SiO_2,lower}(x, y)^4 \right) dy \\
&+ \frac{1}{2w} \int_{-w/2}^{w/2} \Sigma_{ph-esc} \left(T_{ph}(x)^4 - T_{SiO_2,upper}(x, y)^4 \right) dy
\end{aligned} \tag{4.24}$$

which accounts for the heat flux at both the top ($T_{SiO_2,upper}(x, y)$) and bottom ($T_{SiO_2,lower}(x, y)$) interfaces. The coupling coefficient and exponents of the coupling can be modified as was done with the generalized from of the two-temperature model. The other elements of the 3D model are identical to those of the 2D treatment of the nanowire thermometry problem or the 1D treatment of latching and hotspot current simulations.

Latching

In the section on 1D electrothermal modeling, we found that a substantial increase in phonon heat capacity led to reasonable agreement between experiment and model. This suggests that the thin dielectric layer might be contributing to trapping heat near the nanowire and causing latching. To model the SiO_2, we adopt the single temperature model used for nanowire thermometry analysis and use the fitting parameters as listed for the 400 nm pitch device. Because the total heat dissipation is small, we can neglect the rise in temperature of the Si substrate.

We are then left with the parameters $l_0 = 180$ nm describing the mean free path of the SiO_2, and the combination of $r_1 = 5$ and $R_1 = 66$ W/m^{2}K^5 describing the SiO_2-Si thermal boundary. For the WSi nanowires, we adopt the three-temperature model,

using the parameters $D = 0.6\,\mathrm{cm}^2/\mathrm{s}$, $\tau_{ep}(T_c) = 150\,\mathrm{ps}$, $\alpha = 0.316$, $\tau_{esc} = 2\,\mathrm{ps}$, $C_{ph}F = 5$, and $\tau_s = 0.105\,\mathrm{ns}$.

All simulations were performed using a bath temperature of 1 K. This choice of WSi material parameters was motivated by the three-temperature model's ability to fit the correct temperature dependence of the experimental hotspot current, and the choice of $C_{ph}F = 5$ is reasonable based on experimental measurements [133].

Using the same 3D simulation parameters as above, we can compare the latching predictions of the 3D model with that of the 1D model. This is shown in Figure 4.21 for three values of the mean free path l_0. While the addition of the substrate does lead to additional latching compared to the 1D model, it alone is not sufficient to explain the difference between experiment and simulation. As the mean free path is reduced, corresponding to slower thermal conductivity, the simulated results shift closer to experiment. However, with a sufficiently small thermal conductivity, the diffusion of heat in the SiO_2 becomes the thermal bottleneck of the system and shifts the hotspot current away from the measured value. As we saw with the 1D model, fitting can be improved by increasing the heat capacity of the nanowire system and simultaneously increasing the phonon escape time, but there is a limit to the how much of the thermal bottlenecking can be directly attributed to the phonon system while keeping the WSi material parameters within reasonable limits.
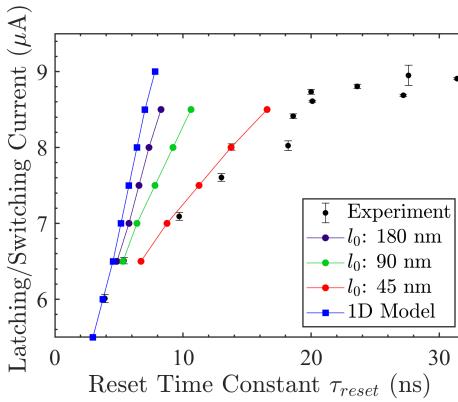
Figure 4.21: Comparison of latching in experiment and using the 3D three-temperature electrothermal model. Fitting to experiment can only be achieved by assuming a phonon mean free path l_0 four times smaller than that extracted from the nanowire thermometry measurements. The result of the 1D model with the same WSi material parameters is shown for comparison.

Crosstalk

To analyze the crosstalk behavior of the prototype DSOC arrays described in Section 4.1.3, we use the same material parameters as used in the latching and hotspot current electrothermal simulations. The films were sputtered from the same compound target on similar thermal-oxide wafers, so the properties should be similar. A more detailed characterization of the device properties would be needed to refine any modeling attempts, but small uncertainties in material parameters are not expected to be the limiting factor in accuracy of the following model. For the 400 nm pitch device, the electrical recovery time was 90 ns based on oscilloscope traces of output pulses, and the recovery time is estimated to be 45 ns for the 800 nm pitch based on the design and kinetic inductance of the material extracted from devices on the same wafer.

Using the material parameters for the 3D simulations listed above with $l_0 = 180\,\text{nm}$, we simulate the response of the system for different bias currents. There are two points of interest for comparison with experiment. First, we must confirm that the temperature increase experienced in a neighboring nanowire is within a reasonable range as expected from experiment. The second metric is checking that the time delay when the current-dependent critical temperature is passed is approximately the same as seen in experiment: 2–3 ns for the 400 nm device and 5–7 ns for the 800 nm device. Figure 4.22a shows the temperature response of the neighboring nanowire due to crosstalk from an adjacent nanowire for the 400 nm and 800 nm pitch cases and Figure 4.22b shows the maximum temperature reached due to crosstalk for the same simulations. There are several features to note. First the maximum temperature reached is within reasonable agreement with the experimental estimates based on the crosstalk efficiency. However, the timescale of thermal crosstalk is significantly shorter than the observed value for the 400 nm pitch simulations compared to experiment. The temperature of the neighboring nanowire in the 400 nm pitch simulations exceeds the temperature threshold for crosstalk of $\sim 1.5\,\text{K}$ at delay times of around a nanosecond. We do not need to consider the rise-time of the crosstalk-driven pulse because the same rise-time occurs in the triggering channel. For the 400 nm pitch simulations, the maximum temperature occurs at a time of $\sim 5.5\,\text{ns}$, which is the correct order of magnitude as found from crosstalk measurements.
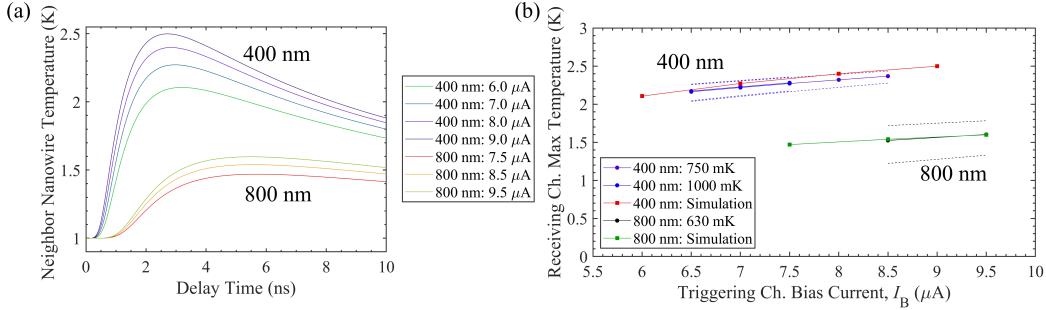
Figure 4.22: (a) Temperature rise in an adjacent nanowire for simulated detections in the 400 nm and 800 nm pitch devices. The timescale of the temperature rise is faster than observed experimentally for the 400 nm pitch device. (b) Comparison of maximum temperature reached in adjacent nanowires for simulation and experiment. There is reasonable agreement between the model and experiment.

The prediction of shorter crosstalk times compared to experiment demonstrates that the model does not capture the thermal transport physics correctly and must be modified. There are two effects which could lead to the observed longer delay times. First, a larger thermal boundary resistance between the SiO_2 and nanowire would lead to additional delay in the temperature rise of the neighboring nanowire. In order to match the hotspot current data, this would have to be accompanied by a proportional increase in the heat capacity of the phonon system of the nanowire, which would have the same effect of increasing the delay before the temperature rises in the neighboring nanowire. In essence, this keeps the heat flux constant while increasing the time constant of the temperature rise. However, this is also expected to decrease the maximum temperature reached by the nanowire because more energy is required to increase the temperature. A second hypothesis is that the thermal conductivity of the SiO_2 is smaller than expected. A smaller thermal conductivity would lead to longer delays before heat reaches the adjacent nanowire. It is difficult to distinguish between these two cases, but simulation can place bounds on the modifications required to bring the model in better agreement with experiment. Figure 4.23 shows the temperature rise results when the SiO_2 l_0 is reduced by a factor of four to 45 nm, representing a significant decrease in the effective thermal conductivity of the system. Such a significant drop in heat propagation has the effect of delaying the onset of temperature increase in the neighboring nanowires. For the 400 nm pitch device, the delay before reaching the current-dependent threshold shifts to approximately 2 ns, in better agreement with experiment. However, this also delays the increase in temperature for the 800 nm pitch device, leading to simulated

delay times that are slower than the measured delay times. Also, the maximum temperature reached based on the simulation falls outside the bound expected from experiment. These results suggest that reduced thermal conductivity in the SiO_2 is not solely responsible for the longer crosstalk times measured in experiment compared to simulation. This is reassuring because it would be difficult to explain a such a small mean free path in these films. At the same time, when fitting the nanowire thermometry results with the two-temperature model, we assumed that the phonon mean free path scales with nanowire fill factor. Implementing this effect here would reduce the onset time of temperature increase in the 800 nm pitch simulations. Despite this reservation, the crosstalk simulations suggest that it is more plausible that thermal bottlenecking effect which causes substantial latching in WSi nanowires is caused by the thermal boundary coupling between the nanowire and dielectric substrate.
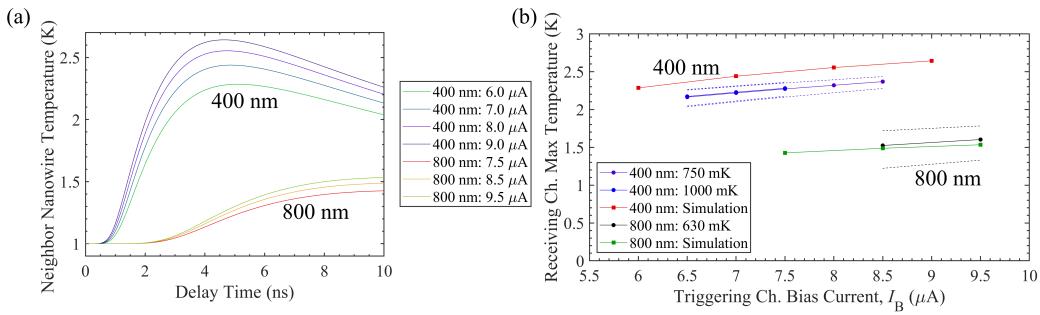
Figure 4.23: (a) Temperature rise in an adjacent nanowire for simulated detections in the 400 nm and 800 nm pitch devices with a reduced SiO_2 mean free path of $l_0 = 45$ nm. The timescale of the temperature rise is close to that observed experimentally for the 400 nm pitch device, but is slower than expected for the 800 nm pitch device. (b) Comparison of maximum temperature reached in adjacent nanowires for simulation and experiment. The maximum temperature falls outside reasonable bounds based on measurements for the 400 nm pitch device.

4.2.6 Discussion

Even with a 3D implementation which includes the basic thermal transport in the SiO_2, fitting all of the experimental metrics simultaneously remains a challenge. Within the current framework, the basic problem is one of thermal flux vs thermal time constants. The value of the hotspot current comes from a balance between Joule heating and dissipation through the substrate, which is typically limited by a particular thermal link. The coupling coefficients within the two or three-temperature models contain a coefficient of the form $\Sigma \sim C/\tau$ which governs the coupling. The

hotspot current defines the heat flux required to match experiment, so Σ ends up fixed. This leaves the time constant τ to fit the latching data, but comes at the cost of increasing the heat capacity to values much greater than reasonable based on the current understanding of the WSi material system. Implementation of diffusion in the SiO_2 dielectric shifts the behavior in the correct direction, but it alone is insufficient to describe the discrepancy between latching behavior in experiment and model when assuming realistic values of the phonon heat capacity.

There are several areas of concern which stem from the simplified treatment of the phonon system in the WSi and might be responsible for the failure of the current electrothermal model. The primary concern is the use of the Debye spectrum to describe the density of states and the associated coupling relations between the electron and phonon systems. For the thin WSi nanowires studied here, the nanowire cannot support standard bulk 3D phonon modes in the vertical dimension near T_c, which is the important temperature range for understanding hotspot current and latching behavior. In principle, we used the generalized form of the two-temperature model and could compensate for reduced dimensionality of the phonon system by adjusting the heat capacity to follow a T^2 power law and reduce the coupling coefficient power laws by a single power. Other effects such as disorder [88] may modify these power laws further, but even with these modifications, we cannot reconcile the hotspot current and latching behavior with reasonable estimates of the phonon heat capacity.

More likely, the importance of the reduced dimensionality of phonons comes from the way energy is coupled to the dielectric. If the reduced dimensionality of phonons resulted in phonon modes which behave as traditional 2D slab phonons, there would be significant implications for the coupling of energy to the substrate. Because the 2D phonons do not propagate with a velocity component perpendicular to the film, they do not escape from the film as readily as bulk 3D phonons. Instead, these vibrations would need to scatter into modes propagating away from the nanowire either through electron-mediated interactions or some form of impurity or boundary scattering. To complicate a numerical treatment of this problem, not all of the modes would be 2D. The higher energy phonons of the distribution would remain 3D and could escape by the standard escape process. A proper treatment would therefore have to consider both cases simultaneously. While the three-temperature model captures heat trapping ideas which are similar to those associated with phonon dimensionality reduction, the numerical treatment is founded on a different bottle-

necking mechanism. The three-temperature model separates phonons into trapped and escaping modes based on their momentum vector, but this is assumed to be energy-independent. In contrast, the reduction of modes and bottlenecking effects due to 2D phonons would be specifically tied to the energy of the phonons. Therefore, as the effective temperature of the system decreases, the three-temperature model predicts a persistent coupling of low-energy modes to the substrate while a 2D analysis would suggest that only the highest energy modes of the system would escape, with this occupation becoming suppressed as temperature drops. Unfortunately, this type of analysis is not easily treated with a thermalized phonon system approach as is assumed by assigning a phonon temperature to a particular population of phonons. Instead, investigating these ideas would likely require a non-equilibrium approach, increasing the computational and theoretical complexity of the model.

The preceding comments assume that reduced film temperature leads to a reduced phonon dimensionality, as is often discussed within the SNSPD community. However, it is not at all clear that this is the relevant case for SNSPDs. The use of 2D phonons typically refers to truly suspended slabs of material, but SNSPDs are typically embedded in a dielectric. It this context, it seems more likely that the low temperature phonon modes would simply extend beyond the thin film into the surrounding dielectric. This type of phonon would not face the same bottlenecking effect as is assumed with 2D phonons. It is expected, however, to lead to weaker coupling with the electronic system, which could contribute to latching behavior. A proper theoretical treatment of the phonon modes in embedded thin films is needed to address these questions.

The effects described above are true for all thin films. The WSi and SiO_2 systems considered here have the additional complexity that they are amorphous materials. It is well known that this alters the phonon density of states and introduces the presence of non-propagating low-energy vibrational modes [131, 132]. The current treatment only superficially considers their impact on the WSi system by introducing additional heat capacity and confining it to the bottlenecked phonon group in the three-temperature model. Their presence was included in the SiO_2 system by using the bulk heat capacity to describe the system rather than the Debye heat capacity, but their impact on heat propagation was ignored. A quantitative treatment of these vibrations is nontrivial, but may be necessary to capture all of the nanosecond timescale dynamics of the system.

Understanding the exact nature of the thermal bottlenecking problem is important

for engineering appropriate mitigation techniques. If the problem is exacerbated by the presence of a thermal boundary at the SiO_2-Si interface, proper dielectric and substrate choices could reduce this effect. Such engineering techniques would reduce the impact of thermal crosstalk because more energy could escape to the substrate before coupling to neighboring nanowires. The same is true for anti-reflection (AR) coatings. AR materials with similar acoustic impedances would allow phonons to propagate ballistically into the AR layers rather than reflecting back toward neighboring nanowires. In contrast, if the propagation of phonons is truly limited by a short mean free path due to material defects, these mitigation strategies would have a more limited effect.

Similarly, understanding the nature of the nanowire-dielectric boundary resistance may suggest strategies for improving thermal dissipation. A study of the normal modes of an embedded nanowire is likely to shed light on phonon confinement, the expected heat capacity of such a system, and the coupling strength with the electronic system of the nanowire. It is anticipated that these phonon modes will extend some distance into the surrounding dielectric, and understanding this mechanism will likely help explain the long escape times needed to explain latching behavior. Furthermore, by investigating the properties of these vibrations, one could use thin dielectrics with different acoustic impedances to modify the vibrational spectrum of the nanowire, providing an engineering tool for altering the heat flow in the system. There is an obvious tradeoff with adding additional thermal boundaries, but it would be worth investigation. While current electrothermal models have been useful in qualitatively understanding the main behavior of SNSPDs, even in modified form, they are not capable of accurately capturing all of the dynamics of WSi nanowires. As new applications call for thinner and lower T_c nanowires and scaling to large arrays, better understanding the electrothermal behavior of the WSi material system will be of increasing importance for engineering large-scale detectors and understanding fundamental limits of performance.

4.3 Multilayer Heater Switches

Superconducting platforms are currently of considerable interest for quantum and neuromorphic computing architectures. Some of these systems use superconducting elements as qubits [136, 137] while others use them as detectors [16, 138]. In either case, superconducting elements are typically interfaced with semiconductor logic either for control circuitry or signal amplification. One of the challenges with this process is that there is a large mismatch in the typical impedances of superconduct-

ing and semiconductor electronics. An ideal solution to this problem would be a superconducting switch which can interface between the low impedance superconducting devices and high impedance semiconductors. For more classical computing architectures, digital superconducting electronic circuits are based on the fundamental building block of high-speed switches. Therefore, an ideal superconducting switch would provide high switching speeds, low power dissipation, have a small device footprint, and exhibit flexibility in its input and output impedances.

In the past few years, there has been sustained development of superconducting switches based on electrothermal interactions in thin superconductors [128, 139–142]. The first of these approaches used a 3-terminal device known as the nanocryotron (nTron) [139–141]. An nTron consists of a narrow nanowire, known as the gate, connected to a relatively wide superconducting nanowire, the channel, through a constriction. When a current exceeding the switching current is applied to the gate, the constriction switches from the superconducting state to the resistive state. The normal domain expands, suppressing superconductivity across the entire channel, switching the channel to the normal state. This change in impedance of the channel diverts current which can be used to drive a load. While the nTron is capable of fast switching and has a simple planar geometry, it is plagued by impedance mismatch limitations, leakage current through the gate, and latching in the channel [139].

To overcome these problems, attempts were made to use thermal rather than electrical coupling as a switching mechanism [142]. A first attempt used a planar hTron device (P-hTron), where a narrow heater was patterned adjacent to the wider channel. When the current through the gate exceeded the critical current, Joule heat generated from the switched nanowire diffused to the channel, causing the channel to switch. While this approach has the advantage of keeping the gate electrically isolated from the channel, the thermal coupling is not efficient. Heat can diffuse in all directions from the gate, so only a small fraction of the energy contributes to breaking superconductivity in the channel.

To solve the thermal localization problem, multilayer thermal switches were developed [4, 58] where a heater element was patterned directly above the superconducting channel with a thin dielectric spacer providing electrical insulation. The heater can be patterned using either a superconductor or normal metal depending on the application. The advantage of this approach is that heat is directed toward the channel more efficiently because it cannot diffuse in all directions, but this comes at the cost of increased fabrication complexity. In particular, these devices require

high quality interfaces between materials and the dielectric must be free of pinholes which could short the gate (heater) to the channel. While the general approaches of McCaughan et al. [4] and Baghdadi et al. [58] are similar, the two works focus on different regimes of operation. McCaughan et al. considers the challenge of creating a high impedance switch by using a long meandering superconducting channel, while Baghdadi et al. focuses on a nanoscale high-speed switch, called the multilayer hTron (M-hTron).

The following section describes modeling efforts used to describe the behavior of these two similar thermal switch architectures. Due to the different operating conditions and materials systems, we find that a single framework is unable to account for all of the experimentally measured characteristics of the two devices. For this reason, two different models are used, corresponding to the two extreme limits describing phonon interactions in the system.

4.3.1 MIT NbN M-hTron

In an experimental effort conducted by Professor Berggren's group at MIT and led by postdoctoral scholar Dr. Reza Baghdadi, the team fabricated and tested a series of M-hTron devices based on a Ti heater and NbN channel separated by a thin SiO_2 dielectric. The basic structure of these devices is shown in Figure 4.24a. A normal-metal heater crosses perpendicular to a superconducting channel with the two films separated by the dielectric. When current passes through the normal metal, Joule heating increases the temperature of the electronic system. This energy couples to the phonon system, and these phonons couple to the dielectric and superconductor below. The phonon energy heats the superconductor until it switches to the normal state, triggering a large change in the impedance of the superconducting channel. The qualitative understanding of this process is straightforward, but to optimize future designs and understand fundamental limits of performance, a semi-quantitative model of the heat transfer process is needed.

Model Formulation

To study the heat transfer in M-hTron devices, we formulated and solved heat transfer equations in various device geometries. Several fundamental energy relaxation processes such as electron-electron, electron-phonon, phonon-electron, and phonon-phonon contribute to the heat transport in nanoscale devices at cryogenic temperatures. While in general these interactions can lead to complex dynamics,

to simplify our approach, we neglect the details of non-equilibrium electron and phonon distributions in favor of the quasi-equilibrium two-temperature model.

Heat transfer within the Ti heater is modeled using

$$C_e(T_e)\frac{\partial T_e}{\partial t} = -\Sigma_{e-ph}\left(T_e^5 - T_{ph}^5\right) + \nabla\kappa_e(T_e)\nabla T_e + \vec{j}\cdot\vec{E} \tag{4.25}$$

and

$$C_{ph}(T_{ph})\frac{\partial T_{ph}}{\partial t} = \Sigma_{e-ph}\left(T_e^5 - T_{ph}^5\right) + \nabla\kappa_{ph}(T_{ph})\nabla T_{ph}, \tag{4.26}$$

which form the basic two-temperature equations for the bulk of a normal metal. In (4.25) and (4.26), T_e is the electron temperature, T_{ph} is the phonon temperature, C_e is the electron heat capacity, Σ_{e-ph} describes the electron-phonon coupling strength, κ_e is the electron thermal conductivity, $\vec{j} = \sigma\vec{E}$ is the current density, σ is the conductivity, $\vec{E}$ is the electric field, C_{ph} is the phonon heat capacity, and κ_{ph} is the phonon thermal conductivity. Note that the temperature dependencies of the heat capacities and thermal conductivities are included in this formulation. The SiO$_2$ spacer layer is treated in a similar way to (4.26) but with electron-phonon interactions neglected and the heat capacity defined according the Debye model. The description of the superconductor system follows the approach of Vodolazov [6] using equations (30–32) therein, under the simplifying assumption that the magnitude of the superconducting order parameter is equal to its equilibrium value for the current value of the electron temperature $\Delta(t) = \Delta_{BCS}(T_e(t))$. Furthermore, we use the phenomenological approach of [128] to describe the current distribution in the superconductor and Joule heating. While this neglects the details of vortex formation and motion which could be captured by a time-dependent Ginzburg–Landau (TDGL) approach, the computational complexity of solving the TDGL equations makes them impractical for calculations of large 3D geometries over many nanosecond timescales. The phenomenological approach has been demonstrated to capture the main electrothermal physics of superconducting nanowires [56, 68, 124], which is the regime of interest in the current study.

A constricted region is included in the model by defining a length of 40 nm where the switching current is given by a fit to the experimentally measured switching current. Elsewhere in the NbN channel, the switching current is given by the theoretical depairing current which is approximated using the Bardeen temperature dependence and the zero temperature value calculated from the Usadel equations as used in (4.12) [134].

Heat fluxes at the interfaces between materials are determined using the AMM, following the approach of Kaplan [82], leading to boundary conditions of the form

$$\kappa_1(T_1)\nabla_\perp T_1 = -\frac{G_{12}}{4}\left(T_1^4 - T_2^4\right), \tag{4.27}$$

where T_1 and T_2 are the local phonon temperatures of the two materials at the interface, $\kappa_1(T_1)$ is the local phonon thermal conductivity at the interface, $\nabla_\perp$ indicates the outward gradient in the direction normal to the interface, and G_{12} describes the boundary conductance. Interfaces with the vacuum are assumed to be perfectly insulating, and the substrate is assumed to remain fixed at the bath temperature (T_{sub}). In the notation of Section 4.2, $G_{12} = 4d\Sigma_{ph-sub}$. A block diagram of this thermal system is shown in Figure 4.24.
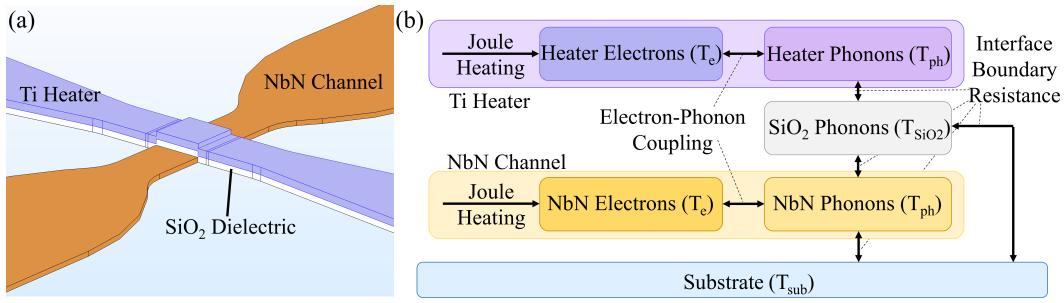
Figure 4.24: (a) Model of the M-hTron geometry. The NbN channel is shown in orange, the SiO$_2$ dielectric is white, and the Ti heater is purple. (b) Block diagram of the electrothermal model. The electron and phonon systems of the Ti heater and NbN channel are coupled by electron-phonon interactions. Each of these electron systems can be heated through Joule dissipation. The phonon systems of the heater, insulator, and channel are coupled by boundary resistances. The insulator and channel are also coupled by boundary resistances to the substrate which remains at a fixed temperature.

Simulation Parameters

The thermal model requires knowledge or estimates of a variety of physical properties of the superconductor, heater, and dielectric materials. Given that many of these parameters are difficult to measure experimentally, we use a combination of literature estimates and measured quantities in our modeling.

The thickness of the titanium heater was estimated to be 30 nm based on deposition time and rate, the sheet resistance was measured to be 24.8 Ω/□, and the electron thermal conductivity was modeled according to the Weidemann-Franz relation. The electron heat capacity is given by $C_{e,Ti}(T_{e,Ti}) = \gamma_{Ti}T_{e,Ti}$ with $\gamma_{Ti} =$

$320\,\mathrm{J/m^3K^2}$ [143]. The titanium phonon system is treated using the Debye model with $C_{ph,Ti}(T_{ph,Ti}) = \alpha_{Ti}T_{ph,Ti}^3$ with $\alpha_{Ti} = 2.47\,\mathrm{J/m^3K^4}$ [143]. The Ti phonon thermal conductivity is estimated as $\kappa_{ph,Ti} = \alpha_{Ti}D_{ph,Ti}T_{ph,Ti}^3$, with the phonon diffusion coefficient $D_{ph,Ti} \sim 0.27\,\mathrm{cm^2/s}$, which represents the Casimir limit. The electron-phonon coupling constant is $\Sigma_{e-ph,Ti} = 1.3 \times 10^9\,\mathrm{W/m^3K^5}$ [144].

The heat capacity of the SiO$_2$ spacer is estimated using the Debye model. The SiO$_2$ thermal conductivity is estimated by extracting the bulk phonon mean free path from bulk values of the thermal conductivity [145] and Debye heat capacity [146], averaging this mean free path with a wavelength independent lower bound equal to the film thickness, then using the kinetic relation to estimate the reduced thermal conductivity for a thin film [135].

The NbN film was 20 nm thick based on the deposition time and rate and the T_c was 8.4 K based on the temperature-dependent switching current. The sheet resistance for simulations was estimated to be $470\,\Omega/\square$ based on the average I-V characteristics of measured devices at cryogenic temperatures and the geometry of the devices. The electron diffusion coefficient of NbN is approximated as $0.5\,\mathrm{cm^2/s}$. These properties lead to an estimated zero temperature depairing current of $237.4\,\mu\mathrm{A}$ [134] for the 600 nm wide devices, and the Bardeen temperature dependence is used at elevated temperatures to approximate the temperature-dependent depairing current. The switching current is approximated by $I_c = 126.6\left(1 - (T/T_c)^3\right)^{2.1}\,\mu\mathrm{A}$ which comes from fitting experimental measurements of the switching current for the 600 nm wide channel device over the temperature range of interest. This represents the switching current of the device at the smallest constriction, which is assumed to be at the center of the nanowire for the 600 nm wide channel device. Outside of this constricted area, the switching current is given by the temperature-dependent depairing current. These values lead to a constriction factor of approximately 0.58 ± 0.05 over the range of $0.4 \leq T/T_c \leq 0.8$, which is consistent with measurements of the switching current to depairing current ratio in thin NbN films [114]. It is important to note that the use of a constriction underneath the heater and depairing current outside of the heater region is necessary to simultaneously reproduce the switching and hotspot current characteristics of the device within the model. If the experimentally measured switching current relation is used over the entire nanowire domain, the simulated hotspot current is significantly smaller than the experimentally measured results. For the NbN phonon system, we use the Debye model and the average phonon sound velocity $c_{avg,NbN} = 4912\,\mathrm{m/s}$ based on measurements of the elastic

Material 1	c_l [m/s]	c_t [m/s]	ϱ [g/m^3]	Material 2	c_l [m/s]	c_t [m/s]	ϱ [g/m^3]	G_{12} [W/m^{2}K^4]
NbN	7137	4459	8.25	Sapphire	10900	6450	3.98	516.2
NbN	7137	4459	8.25	SiO$_2$	5832	3712	2.2	738.1
SiO$_2$	5832	3712	2.2	Sapphire	10900	6450	3.98	452.5
Ti	6100	3120	4.50	SiO$_2$	5832	3712	2.2	1504

Table 4.1: AMM parameters. For a given material, c_l is the longitudinal sound velocity, c_t is the transverse sound velocity, and ϱ is the density.

properties of the material [74], which leads to $C_{ph,NbN}(T_{NbN}) = \alpha_{NbN}T^3_{ph,NbN}$ with $\alpha_{NbN} = 1.03\,\text{J/m}^3\text{K}^4$. The phonon thermal conductivity is treated in the same way as the Ti heater, using the phonon diffusion coefficient $D_{ph,NbN} \sim 0.33\,\text{cm}^2/\text{s}$.

The thermal boundary conductance at each of the interfaces is calculated using the AMM following the approach of Kaplan [82]. The relevant sound velocities and densities used for these calculations along with the boundary conductance parameters are listed in Table 4.1.

Comparison with Experiment

In the model, two current sources determine the bias conditions of the M-hTron as shown schematically in Figure 4.25a. A time-dependent current source ($I_H(t)$) governs the current flow in the heater. The superconducting channel is electrically connected in series with an inductor (L_s) and the combined inductor and channel are connected in parallel with a load resistor (R_L). This circuit is driven by the bias current source (I_B).

The nonlinear set of coupled partial differential equations is solved in a 3D geometry using finite element methods. The geometry of the system is defined to match the experimentally fabricated devices. The insulator and heater cross perpendicular to the superconducting channel as shown in Figure 4.24a and maintain the topography of the step edge at the channel. Maintaining this step edge necessitates the use of a 3D finite element solver. At interfaces of the channel, dielectric, or heater which are not in contact with each other, the material is in contact with vacuum and assumed to be perfectly insulating. As current passes through the heater, Joule heating first heats the electron system of the Ti. This heat couples to the phonon system and couples to the phonons of the dielectric spacer. Eventually, this heat is coupled to the phonon system of the superconductor and through electron-phonon interactions, to the electron system, which suppresses superconductivity. When the temperature of the superconductor surpasses the current-dependent critical temperature of the superconductor, the superconductor switches to the normal state, and Joule heating

expands the normal domain along the length of the channel. Figure 4.25b shows an example simulation of the time evolution of the temperature of the system for a 1 ns current pulse in the heater. The delay caused by heat passing through the multiple thermal boundary resistances and due to the heat capacities of these layers causes a lag between the onset of current in the heater and the formation of resistance in the superconducting channel, as seen in Figure 4.25c. The reset properties of the superconducting channel are determined by the series inductance, load seen by the device, and duration of the current pulse through the heater. For short pulsed heat sources, self-resetting can be achieved through proper selection of the L_s/R_L time constant or biasing the device below the hotspot current.
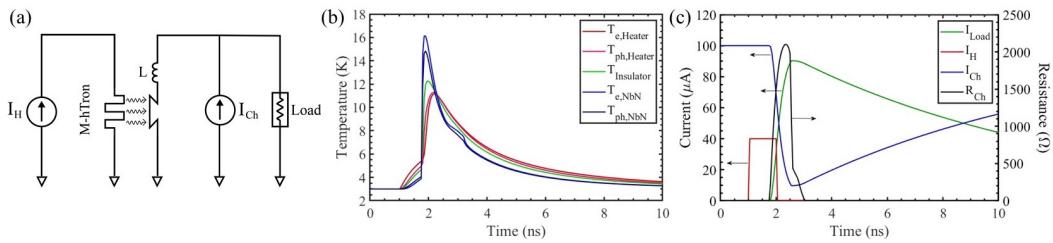
Figure 4.25: (a) Schematic circuit for simulating M-hTron performance. The heater current source I_H dissipates energy in the M-hTron resistor, increasing the temperature of the superconducting channel. When this heat drives the M-hTron channel to the normal state, the appearance of a finite resistance diverts current to the load, modeled as a $50\,\Omega$ resistor. Time-dependent (b) temperature response and (c) electrical response of the M-hTron. Once the heater current activates, the temperature of the system increases. Thermal boundary resistances and heat capacities lead to a lag between when the NbN temperature increases compared to the heater. Once the NbN electron temperature exceeds the current-dependent critical temperature, the channel switches to the normal state, and the Joule heat from the NbN rapidly increases the temperature of the device. The heater width is $500\,\text{nm}$, channel width is $600\,\text{nm}$, heater current is a 1 ns long $40\,\mu\text{A}$ pulse, channel bias current is $100\,\mu\text{A}$, and substrate temperature is 3 K.

The predictions of the electrothermal model are compared with experimental measurements for a device with a channel width of 600 nm and a heater width of 500 nm. Details of the experimental work can be found elsewhere [58]. The switching and hotspot currents of the superconducting channel were measured for different heater currents. As the heater current increases, the switching current of the channel is suppressed, as shown in Figure 4.26a. For simulations, we define the switching current of the constricted region of the superconducting channel by fitting the experimentally measured temperature-dependent switching current using the function

$I_c = 126.6 \left(1 - (T/T_c)^3\right)^{2.1}$ µA, shown in Figure 4.26b. The hotspot current simulations were performed with a time-dependent voltage source in series with the superconducting channel, series inductor, and a 10 kΩ bias resistor. The supply voltage is slowly dropped, allowing an established hotspot to shrink to its minimum size before relaxing to the superconducting state. The hotspot current is evaluated as the minimum current which sustains a hotspot. There is surprising agreement between the model and experimental measurements. The temperature-dependent switching current was fit to the experimental measurements, but the correct dependence of switching current on heater current naturally emerges without requiring additional fitting. This suggests that the AMM provides a reasonable estimate of the boundary conductance between the NbN and sapphire substrate.
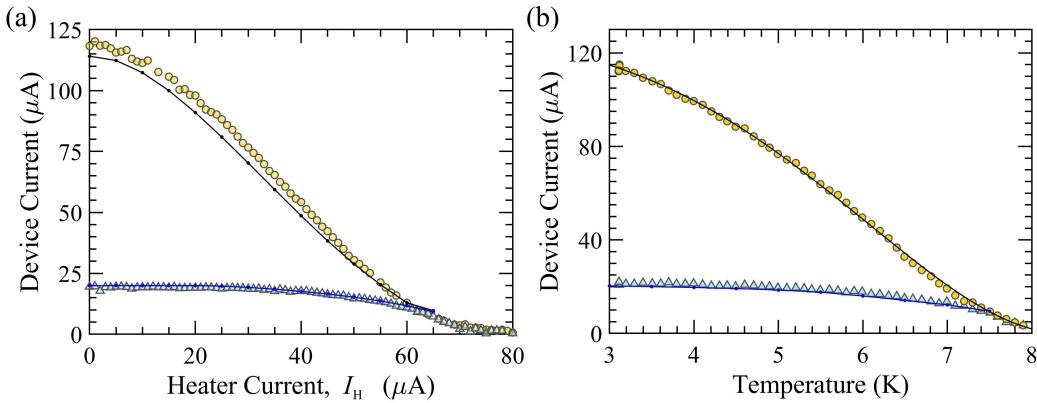
Figure 4.26: The switching current (yellow circles) and the hotspot current (blue triangles) of an M-hTron with a channel width of 600 nm and heater width of 500 nm as a function of (a) the heater current, and (b) temperature. The model predictions of the switching current (black line) and hotspot current (blue line) are shown for comparison. The model's temperature-dependent switching current is a fit to experiment.

While the fitting for the device of Figure 4.26 looks promising, not all of simulated devices immediately agree with the experimental data. Figure 4.27 shows the comparison of experiment and simulation for three devices with various heater widths and a channel width of 200 nm. There is a systematic shift in the results. Part of this might be the use of the same channel switching current dependence as was used in the device of Figure 4.26. Also, if the thickness of the heater changed from one deposition to the next, the total resistance of the heater would change, leading to a change in the total power dissipated.
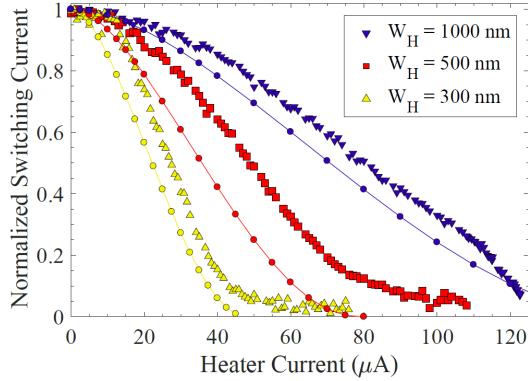
Figure 4.27: Normalized switching current of hTrons (symbols) with heater widths of 300 nm, 500 nm, and 1000 nm as a function of the heater current. The channel width is kept at 200 nm for these devices. Predictions from the electrothermal model (lines) are shown for comparison.

One potential application of the M-hTron is its use as a preamplifier for SNSPD readout. When operated in this configuration, the heater element of the M-hTron acts as the load of an SNSPD circuit while the channel of the M-hTron carries current which is diverted into a traditional RF amplifier as shown in Figure 4.28a. The M-hTron is capable of increasing the amount of current diverted into the amplifier, which can increase the signal-to-noise ratio of the electrical pulse and reduce the noise contribution to timing jitter. The use of the M-hTron for this application is simulated using the same material parameters as used throughout this work for a channel width of 1 μm. As shown in Figure 4.28b, when a 300 nm wide heater is used, a current amplification factor of 7 can be achieved for a current pulse resembling the output of an SNSPD with bias current of 25 μA. Optimization of the material properties and device geometry can improve this value. In practice, the amplification factor which can be attained is limited by the reset dynamics of the SNSPD. Due to the finite resistance of the normal domain generated during photon detection (typically a few kΩ), the heater resistance must be a small fraction of the normal domain resistance in order to ensure shunting of current from the nanowire during photon detection. Furthermore, the heater resistance contributes to determining the electrical reset of the SNSPD ($\tau_{\mathrm{reset}} = L_k/R_H$), and too large a resistance may induce latching [56]. Here, L_k is the kinetic inductance of the SNSPD and R_H is the heater resistance, which acts as the load impedance of the device.
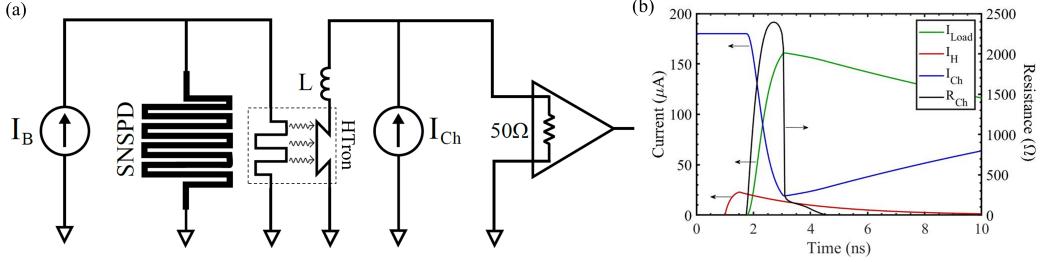
Figure 4.28: (a) Schematic circuit using the M-hTron as a preamplifier for SNSPD readout. The heater resistor of the M-hTron acts as the load resistor of the SNSPD. When the M-hTron switches, the channel current is diverted into an RF amplifier. (b) Simulated M-hTron response when operated as an SNSPD preamplifier. The channel width is 1000 nm, heater width is 300 nm, and channel bias current is 180 μA. The total resistance of the heater, including contacts, is 220 Ω and the series inductance of the M-hTron channel is 1 μH. The red curve indicates the heater current and represents an SNSPD-like pulse, the blue curve shows the current through the M-hTron channel, and the green curve shows the current diverted to the 50 Ω load of the M-hTron channel. The black curve shows the channel resistance. The simulated device achieves a current gain of 7.

Simplified Thermal Model

All of the simulation results shown above use the full 3D model. However, implementing such a model is computationally intensive and is often not necessary to gain useful information for making design decisions. The computational complexity of the model can be significantly simplified by reducing the model to a lower-dimensional form. In the limits of wide superconducting channel, wide heater, thin heater, thin dielectric spacer, and thin channel, the system of equations can be expressed as a 0D set of coupled PDEs. While this neglects the effects of diffusion and cannot be used to describe the electrothermal evolution of the superconducting channel, it does provide a simple means of estimating the thermal dissipation required to achieve switching of the superconductor. For the heater, the reduced set of equations is expressed as

$$C_{e,H}(T_{e,H})\frac{\partial T_{e,H}}{\partial t} = -\Sigma_{e-ph,H}\left(T_{e,H}^5 - T_{ph,H}^5\right) + \frac{I_H^2 \rho_H}{(w_H\,d_H)^2} \tag{4.28}$$

and

$$C_{ph,H}(T_{ph,H})\frac{\partial T_{ph,H}}{\partial t} = \Sigma_{e-ph,H}\left(T_{e,H}^5 - T_{ph,H}^5\right) - \frac{G_{H-D}}{4d_H}\left(T_{ph,H}^4 - T_D^4\right), \tag{4.29}$$

where $T_{e,H}$ is the heater electron temperature, $C_{e,H}$ is the electron heat capacity, $\Sigma_{e-ph,H}$ is the heater electron-phonon coupling strength, I_H is the heater current,

ρ_H is the heater resistivity, d_H is the heater thickness, w_H is the heater width, $T_{ph,H}$ is the heater phonon temperature, $C_{ph,H}$ is the phonon heat capacity, G_{H-D} is the heater-dielectric boundary conductance, and T_D is the dielectric temperature. The dielectric equation takes the form

$$C_{ph,D}(T_{ph,D})\frac{\partial T_{ph,D}}{\partial t} = \frac{G_{H-D}}{4d_D}\left(T_{ph,H}^4 - T_D^4\right) - \frac{G_{D-S}}{4d_D}\left(T_{ph,D}^4 - T_{ph,S}^4\right), \quad (4.30)$$

where $C_{ph,D}$ is the dielectric heat capacity, d_D is the dielectric thickness, G_{D-S} is the dielectric-superconductor boundary conductance, and $T_{ph,S}$ is the channel superconductor phonon temperature. Finally, the reduced equations for the superconducting channel become

$$C_{e,S}(T_{e,S})\frac{\partial T_{e,S}}{\partial t} = -\Sigma_{e-ph,S}\left(T_{e,S}^5 - T_{ph,S}^5\right) \quad (4.31)$$

and

$$C_{ph,H}(T_{ph,H})\frac{\partial T_{ph,H}}{\partial t} = \Sigma_{e-ph,S}\left(T_{e,S}^5 - T_{ph,S}^5\right) - \frac{G_{D-S}}{4d_S}\left(T_{ph,S}^4 - T_D^4\right) \\ - \frac{G_{S-Sub}}{4d_S}\left(T_{ph,S}^4 - T_{Sub}^4\right), \quad (4.32)$$

where $T_{e,S}$ is the electron temperature of the channel, $C_{e,S}$ is the electron heat capacity, $\Sigma_{e-ph,S}$ is the electron-phonon coupling constant, $C_{ph,H}$ is the phonon heat capacity, G_{S-Sub} is the channel-substrate boundary conductance parameter, d_S is the channel thickness, and T_{sub} is the substrate temperature.

The full 3D simulation results are compared with the simplified equations of (4.28–4.32) in Figure 4.29 for the time-dependent temperature response of the system to a current pulse in the heater. There is only a small difference in the transient response of the system, which is attributed to the thermal conduction across the thicknesses of the films.

Within the simplified model, calculation of the steady-state thermal coupling from the heater to the superconducting channel can be reduced to the analytical expression

$$T_{e,S}(I_H) = \left[\frac{4\rho_{\Box,H}}{G_{S-sub}}\left(\frac{I_H}{w_H}\right)^2 + T_{sub}^4\right]^{1/4}, \quad (4.33)$$

where $\rho_{\Box,H}$ is the heater sheet resistance. The heating induced switching current is found by inverting the temperature-dependent switching current expression. Comparison with the full 3D model is shown in Figure 4.29c.
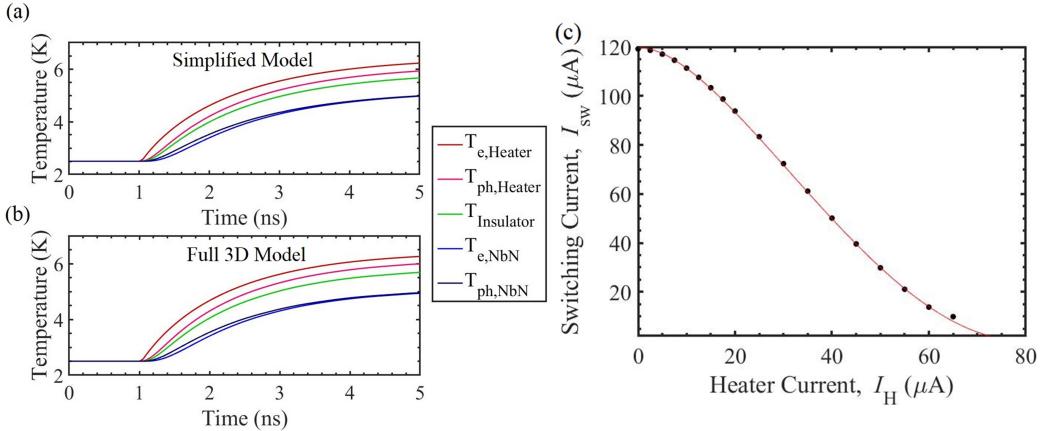
Figure 4.29: Comparison of the transient response of the (a) 0D simplified model to the (b) full 3D model. A heater current of $30\,\mu\text{A}$ turns on at a time of $1\,\text{ns}$. The channel width is $600\,\text{nm}$ and the heater width is $500\,\text{nm}$. The temperatures of the 3D model are sampled at the center of the intersection of the heater and nanowire and at a depth of half the film thickness. There is good quantitative agreement between the simplified and full models. (c) Comparison of the simulated switching current for the full 3D model (circles) and the analytical expression (4.33) (red line). The two models agree in this limit of wide heater and wide channel. The deviation in the switching current of the full model at the highest heater current is due to ambiguity in the definition of switching in this regime.

The modest success of the 3D model and the reasonable agreement of between the 0D and 3D models indicate that the 0D model could be a useful tool for designing the thermal coupling properties M-hTron devices. However, a full model is needed in order to understand the more complex electrothermal coupling which occurs in superconducting circuits.

4.3.2 NIST WSi High Impedance Switch

In an experimental effort led by Dr. Adam McCaughan and the team at NIST Boulder, low-input-impedance and high-output-impedance multilayer thermal switches were developed to help bridge the typical impedance gap existing between superconductor and semiconductor electronics. The devices here have the same basic structure as the M-hTron described in the preceding section, but rather than a single nanowire constituting the superconducting channel, the NIST device features a WSi nanowire meander spanning several tens of μm^2. The heater consists of a $15\,\text{nm}$ thick PdAu film and the dielectric spacer is $25\,\text{nm}$ thick SiO_2. While similar in general structure, there are several features which differentiate the high impedance switch and M-hTron. The nanowire has a non-unity fill factor beneath the heater.

Therefore, not all of the heat is forced to travel through the nanowire before escaping to the substrate. Second, the high impedance switch uses a WSi superconductor rather than NbN which impacts the amount of energy required to break superconductivity. Finally, the NIST device uses a lower resistivity heater which can impact the heat transfer from heater to dielectric. Additional details of the device design and characterization can be found elsewhere [4].

The goal of the following work was to augment the experimental work with a simple model which could be used to describe the experimental results. The main measurement of interest was turn-on delay between when an electrical pulse was applied to the heater and when a finite resistance appeared on the superconducting channel. The experimental team found that the turn-on delay for a given power density aligned well with the curve describing a fixed energy per unit area. A theoretical model should predict this general behavior, and for that we developed a simple model based on the ballistic propagation of phonons through the SiO_2 and WSi.

Ballistic Model

Given that the bulk mean free path of phonons in SiO_2 is on the order of $2\,\mu m$ at 2.5 K, we assume that phonons escaping from the heater travel through the dielectric without scattering and either interact with the nanowire or continue unimpeded to the substrate. Within this model and under the simplifying assumption of equilibrated electron and phonon systems in the nanowire at temperature T_{WSi}, the energy balance equation of the nanowire is given by

$$\left(C_e\left(T_{WSi}\right) + C_{ph}\left(T_{WSi}\right)\right) d\, f \frac{\partial T_{WSi}}{\partial t} = f \chi_{abs} P_{heater} - \Sigma\left(T_{WSi}{}^4 - T_{sub}{}^4\right) \quad (4.34)$$

where $C_e\left(T_{WSi}\right)$ is the BCS electron heat capacity, $C_{ph}\left(T_{WSi}\right)$ is the lattice heat capacity, d is the nanowire thickness, f is the nanowire fill factor, χ_{abs} is the fraction of energy incident on the nanowire which is absorbed, P_{heater} is the power dissipated by the heater per unit area, and Σ describes the magnitude of phonon energy flux from the nanowire to the substrate per unit area. For a WSi device with $d = 4.5\,nm$, $f = 0.5$, diffusion coefficient $D = 0.74\,cm^2/s$ [147], sheet resistance $\rho_\square = 590\,\Omega/\square$, $T_c = 3.4\,K$, and $C_e\left(T_c\right)/C_{ph}\left(T_c\right) \sim 1$ [133], the parameters χ_{abs} and Σ are chosen to fit the turn-on delay vs dissipated power results as shown in 4.30. The curves show the turn-on delay for temperature thresholds of 2.5 K and 3 K, $\chi_{abs} = 0.02$, and $\Sigma = 0.7\,W/m^2K^4$ which are the estimated switching

temperatures for the given bias currents. These are significantly smaller than the theoretical depairing current because the system is expected to experience severe current crowding due to the 50% fill factor of the device. The value of Σ corresponds to nonbolometric phonon bottlenecking at the WSi/SiO$_2$ interface with a conversion time from the non-escaping to escaping group of phonons with a magnitude over 1 ns, which is consistent with experiment [133] and aligns with the investigation of thin WSi films from Section 4.2. For comparison, the estimation of the parameter χ_{abs} based on the solution of ballistic phonon transport in the nanowire is provided in the next section. It should be noted that the first attempts to model this system used a two-temperature model similar to that used to describe the WSi material system or the M-hTron but were unable to reproduce experimental observations. This is further evidence that the two-temperature model has limited applicability when considering thin-film superconductors with T_c below 4 K as was observed in Section 4.2.
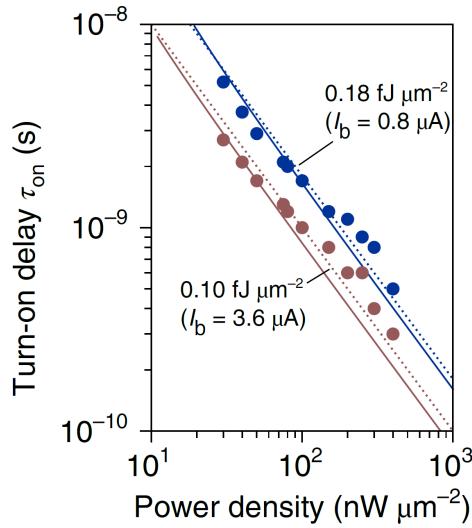
Figure 4.30: Turn-on delay for various applied power densities on the heater. The two colors represent different bias conditions on the superconductor. The solid lines are the fit using the ballistic transport model while the dashed lines correspond to curves of constant energy per unit area.

Estimation of χ_{abs}

We estimate the parameter χ_{abs} based on ideal ballistic propagation of phonons from the heater to the nanowire through the thin dielectric. To simplify the analysis, we assume that the heater instantaneously reaches the stationary response due to the dissipated power at the time when an electrical pulse is applied. Phonons in the

WSi travel along ballistic trajectories with the average sound velocity c_{avg} with an angle θ with respect to normal of the interface and we consider only the absorption of phonons, neglecting re-emission in this treatment. This is consistent with the formulation above which treats the re-emission of phonons in a separate term. The electron system of the nanowire is assumed to have an equilibrium distribution described by the temperature T_e. Under these assumptions, the kinetic equation describing the phonon distribution in the nanowire $N_{WSi}(\omega, z, \theta, t)$ with depth in the z direction from $0 \leq z \leq d$ is given by

$$\frac{\partial N_{WSi}}{\partial t} + c_{avg} \cos(\theta) \frac{\partial N_{WSi}}{\partial z} = -\frac{N_{WSi} - N^0(T_e)}{\tau_{ph-e}} \tag{4.35}$$

where $N^0(T_e)$ is the Planck distribution at the electron temperature of the nanowire and τ_{ph-e} is the phonon-electron interaction time. Under the assumption of a stationary input phonon distribution $N^H(T_H)$ from the heater at $z = 0$, the solution to this partial differential equation becomes

$$N_{WSi}(\omega, z, \theta) = N^0(T_e) + \left[N^H(T_H) - N^0(T_e)\right] e^{-\frac{z}{l_{ph-e}\cos(\theta)}} \tag{4.36}$$

where $l_{ph-e} = c_{avg}\tau_{ph-e}$ is the energy-dependent phonon-electron interaction length. The energy transferred to the WSi is expressed as $Q_{WSi}^{in} - Q_{WSi}^{out} = P\chi_{abs}\text{ff} = Q_{WSi}^{in}\chi_{abs}$. The energy flux is evaluated according to

$$Q(z) = \int_0^{\omega_D} d\omega \rho(\omega)\hbar\omega v \int_0^{\theta_m} d\theta \sin(\theta)\cos(\theta)N(\omega, z, \theta) \tag{4.37}$$

where $\rho(\omega)$ is the phonon density of states and ω_D is the Debye frequency. Using this form, we evaluate χ_{abs} by rearranging terms and evaluating Q_{WSi}^{in} at $z = 0$ and Q_{WSi}^{out} at $z = d$, leading to

$$\chi_{abs} =$$

$$\frac{\int_0^{\omega_D} d\omega \rho(\omega)\hbar\omega v \int_0^{\theta_m} d\theta \sin(\theta)\cos(\theta)\left[N^H(T_H) - N^0(T_e)\right]\left(1 - e^{-\frac{d}{l_{ph-e}\cos(\theta)}}\right)}{\int_0^{\omega_D} d\omega \rho(\omega)\hbar\omega v \int_0^{\theta_m} d\theta \sin(\theta)\cos(\theta)N^H(T_H)}$$

$$\tag{4.38}$$

We take a series expansion of the exponential function because $\frac{d}{l_{ph-e}\cos(\theta)} \ll 1$ for thin WSi nanowires. The interaction length is given by $\frac{1}{l_{ph-e}(\omega)} \approx \frac{\gamma}{v\tau_0}\left(\frac{\hbar\omega}{k_B T_c}\right)$ for $\hbar\omega \geq 2\Delta$ and 0 otherwise. This occurs because only phonons with energies greater than 2Δ are capable of breaking Cooper pairs. The parameter $\gamma = \frac{8\pi^2}{5}\frac{C_e}{C_{ph}}\Big|_{T_c}$ describes the ratio of electron to lattice heat capacity at T_c. We can neglect the $N^0(T_e)$

term in (4.38) because the WSi temperature is significantly less than the heater temperature for the power dissipation levels measured experimentally, and we confirm that this approximation is valid by numerical calculation of the full and approximate expressions. Under these simplifications, and assuming low temperature, we arrive at the expression

$$\chi_{abs} = \frac{1}{\cos\left(\frac{\theta_m}{2}\right)^2} \frac{\gamma d}{c_{avg}\tau_0} \left(\frac{T_H}{T_c}\right) \frac{\int_{2\Delta/k_B T_H}^{\infty} dx\, \frac{x^4}{e^x-1}}{\int_0^{\infty} dx \frac{x^3}{e^x-1}}. \tag{4.39}$$

To determine χ_{abs}, we estimate $\tau_0 = 5000$ ps based on measurements of τ_{ep} [133] and the parameters listed in the previous subsection. The heater temperature is determined using thermodynamic properties of Au and Pd and an estimated phonon escape time of 100 ps.
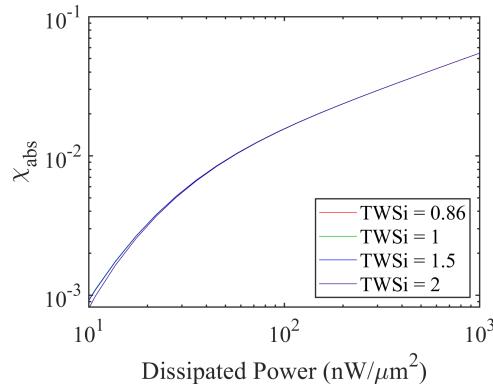
Figure 4.31: Estimation of χ_{abs}.

This calculation leads to a value of χ_{abs} which is between 0.01 and 0.03 over the dissipated power range measured in the turn-on delay experiment, which is the same range needed to match experiment based on the simple fixed χ_{abs} model. The absorption fraction increases as the dissipated power increases due to the increased number of high-energy phonons present in the heater radiation. These high-energy phonons have a shorter $l_{ph-e}(\omega)$, leading to a higher fraction of energy absorbed in the nanowire. However, this simplified model neglects the influence of SiO_2 scattering, which also increases as phonon energy increases and is expected to limit the magnitude of this change in χ_{abs}. Consequently, the ballistic propagation assumption used in this model is violated for these higher energy phonons. A fully quantitatively accurate model of phonon transport must also consider the reflection of phonons off both the SiO_2/WSi and SiO_2/Si interfaces. While transmission

from SiO_2 to WSi is estimated to be approximately 70% based on the AMM for most incident angles, total internal reflection of acoustic modes is predicted at the SiO_2/Si interface for incidence angles greater than 50 degrees, which will increase the number of phonons available for absorption in the nanowire layer.

4.4 Conclusion

Electrothermal modeling has provided a general framework for understanding the basic operation of SNSPDs. The simple lumped element models used to describe the normal domain growth and electrical reset of these devices have proved useful for understanding the electrical characteristics of detector output and have informed engineering decisions about amplifier and nanowire biasing requirements. As SNSPD technology matures and is expanded to large-format arrays or considers the use of thermal coupling as a form of multiplexing [1], the details of heat dissipation become important for optimizing device designs. Therefore, there is a need for quantitatively accurate electrothermal models which can predict the minimum spacing needed between adjacent pixels to prevent (or facilitate) crosstalk or predict the minimum allowable reset time which avoids latching. However, as demonstrated throughout this chapter, the standard electrothermal treatment based on the two-temperature model typically used to describe SNSPD dynamics fails for the technologically important WSi material system.

There are several weaknesses of the two-temperature model which are exacerbated in thin WSi nanowires. The two-temperature model does not accurately capture the phonon bottlenecking effect which occurs in films with total internal reflection for high incidence angle phonons. This is remedied by adopting a three-temperature model [80], but the three-temperature model does not address further shortcomings. The relevant temperature scale for the electrothermal dynamics in SNPSDs is T_c. For WSi, the low value of T_c between 3 K and 4 K means that standard bulk 3D phonons are not supported in the film near T_c. Not only does this alter the heat capacity of the lattice, but it is expected to alter the coupling with the substrate. Calculation of the resulting phonon spectrum is an essential step for understanding the coupling dynamics, making it necessary for understanding latching behavior. Due to the complication of phonon bottlenecking, a full non-equilibrium treatment is likely necessary to produce a fully quantitative model of WSi nanowires.

The heat transfer in the thin-film dielectrics plays a central role in array crosstalk [27], thin-film heater switches [4, 58], and new array architectures [1], but little work has

been done to incorporate this heat transfer in models of SNSPDs. The 3D modeling of latching suggests that these dielectric films contribute to thermal bottlenecking in WSi devices, but this effect does not seem to be sufficient to explain the full extent of latching observed in experiment. The simple diffusion-based model is also unable to predict the correct time delay for crosstalk in nanowire arrays. The nanowire thermometry measurements conducted in this work suggest that the presence of nanowires leads to substantial phonon scattering, decreasing the effective mean free path of phonons in the system and reducing the spread of heat. Understanding the extent of this scattering would help to enable the ability to quantitatively predict thermal coupling in devices with multiple nanowires in close proximity. Finally, the amorphous nature of both WSi and SiO_2 adds an additional level of complication to the quantitative treatment of the system, but the importance of this effect is not yet clear.

For years, the SNSPD community has relied upon an iterative trial-and-error approach to discover the constraints thermal effects place on device performance. Modeling efforts are used to understand these effects at a qualitative level, but such models are not seriously used as design tools for optimizing new detectors. This has not prevented progress, but as array size scales and as new techniques rely on thermal coupling to multiplex SNSPD readout, the iterative approach becomes increasingly costly. There is a need for a quantitatively accurate electrothermal model for WSi and other low T_c materials which focuses on the coupling of heat to and from the dielectrics. We have identified several weaknesses with the current methods for performing electrothermal analysis, but serious theoretical and experimental effort is needed to advance these models to the level where they can be used to efficiently optimize next-generation SNSPD arrays.

Chapter 5

THERMAL ROW-COLUMN

5.1 Background

Single-pixel superconducting nanowire single photon detectors (SNSPDs) [7] have achieved remarkable performance including >90% detection efficiency [11, 12], timing jitter below 12 ps at telecom wavelengths [2, 25, 111], and dark count rates below 0.5 counts per hour [14, 15]. Emerging applications such as focal plane arrays for MIR astronomical spectroscopy [41], detectors for dark matter search efforts [15], telecom LIDAR [148], and quantum imaging would benefit from SNSPDs' high efficiency and low intrinsic dark count rate performance for low photon-flux signals over wide wavelength ranges. However, these applications also require kilopixel to megapixel arrays covering millimeter-scale active areas. Traditional SNSPD readout uses high bandwidth RF cables for each channel of an array, but this places an unsustainable heat load on the cryogenic stage as the number of cables scales beyond a few tens of lines. Arrays of 64 pixels have been demonstrated using direct readout [27] but scaling to the kilopixel range is challenging. To overcome this limitation, various multiplexing schemes have been demonstrated which enable imaging capability while reducing the number of RF cables. One approach uses a row-column biasing scheme [28, 29, 31] and correlations between the detections on row and column readout channels to determine which pixel registered a detection. This architecture is able to read an $N \times N$ array with $2N$ readout lines. However, this architecture suffers from current redistribution among the pixels of each bias line, limiting the maximum count rate and degrading the timing jitter [29]. Multiplexing based on SFQ readout has been demonstrated in both standard [30] and row-column

[31] arrays, but this approach still struggles with the heat load placed on the cryogenic stage. A third approach fashions the active nanowire element as a delay line and uses differential readout to infer the location of absorption along the length of the nanowire [32, 33]. This design requires the use of large impedance matching tapers to improve signal quality [71], but this adds significant kinetic inductance to the device which reduces the maximum count rate (MCR). Furthermore, the spatial resolution is ultimately limited by the timing jitter of the system, necessitating the use of high performance amplifiers and readout electronics. As an alternative, the microwave kinetic inductance detector (MKID) community has used frequency multiplexing to develop detector arrays with several to tens of kilopixels using few microwave feedlines at the cryogenics stage [149–151]. While this type of frequency multiplexing has been demonstrated in SNSPD arrays using both DC [34] and AC [35] nanowire biasing, significant challenges remain in successfully scaling these architectures to the kilopixel scale.

We propose and demonstrate a new method of multiplexing, the thermally-coupled row-column (thermal row-column or TRC), which uses thermal coupling between two active SNSPD layers with channels arranged in rows and columns. By measuring coincidence events between row and column channels, the location of absorption is inferred to be the intersection of those channels, forming a pixel. The channels are electrically isolated at low frequencies, so there is no current redistribution and loss of electrical signal as occurs in the electrically coupled row-column architecture [28, 29]. The TRC has the additional advantage of not requiring biasing resistors or wiring within the device active area, as needed for electrical row-column arrays, which increases the maximum active fill factor which can be achieved.

5.2   Thermal Row-Column Concept

After the initial formation of a normal domain during photodetection, the inductance of the nanowire combined with Joule heating leads to dissipation of thermal energy. Typically, this thermal energy is a nuisance to design and can limit the minimum allowable reset time of a detector before latching occurs [56]. In multi-element devices, this thermal energy can lead to crosstalk between adjacent pixels [27] which limits the maximum achievable fill factor of the nanowires. The TRC utilizes this energy as a means of coupling row and column channels in an array. N nanowires are patterned into rows on one layer and N nanowires are patterned as columns on the second layer, with each nanowire having its own readout. The heat generated during a photon detection on one nanowire layer raises the temperature

of the dielectric and second nanowire layers, causing the nanowire on the second layer to switch, as shown in Figure 5.1a. This effect was exploited in a single-pixel multilayer device where thermal coupling was used to trigger an avalanche between two nanowires in parallel, increasing signal output and reducing electrical noise jitter [152]. While the dielectric spacer between the nanowires was only 12 nm in that work [152], crosstalk in arrays [27] suggests that thermal coupling should be observable over several hundred nanometer distances. By using the correlations between the detections on rows and columns, the location of the photon absorption is determined. Groups of channel detection events are considered correlated when they occur within a chosen time-span, known as the coincidence window. As with the standard row-column architecture, $2N$ readout channels are required to operate an $N \times N$ array. The thermally-coupled architecture is not limited to row-column style imaging arrays. As long as a unique overlap mapping exists between the channels of the layers, the same degree of multiplexing can be achieved, making the architecture viable for linear array spectrometers.

The nanowire orientation can be tailored to suit the needs of a particular application. In the parallel-polarization design shown schematically in Figure 5.1b, nanowires on both layers are oriented in the same direction and patterned directly above one another. This is ideal for ensuring efficient and fast thermal coupling between the layers, but leads to increased capacitance between pixels which can increase electrical crosstalk. When embedded in a typical optical cavity, the parallel design maintains a strong polarization sensitivity. For applications where this is undesirable, a cross-polarized design can be employed where the nanowires of the two layers are oriented perpendicularly as shown in Figure 5.1c. This can lead to polarization insensitive optical coupling, and the multilayer design enables broadband absorption with the appropriate optical stack design [153]. However, with the cross-polarized design, nanowires are not directly aligned, so the distance heat must travel to trigger the second layer depends on the absorption location. This leads to additional jitter in the thermal coupling time, requiring a larger coincidence window which can impact the MCR (see Section 5.3.5). Despite this limitation, the overall timing jitter does not suffer in this configuration because the detection jitter is linked only to the first detector to 'click.'

5.2.1 Design and Fabrication

Devices were fabricated on a 4 inch silicon wafer. A 60 nm thick Au back reflector was patterned via optical lithography and lift-off before a 155 nm thick film of SiO_2
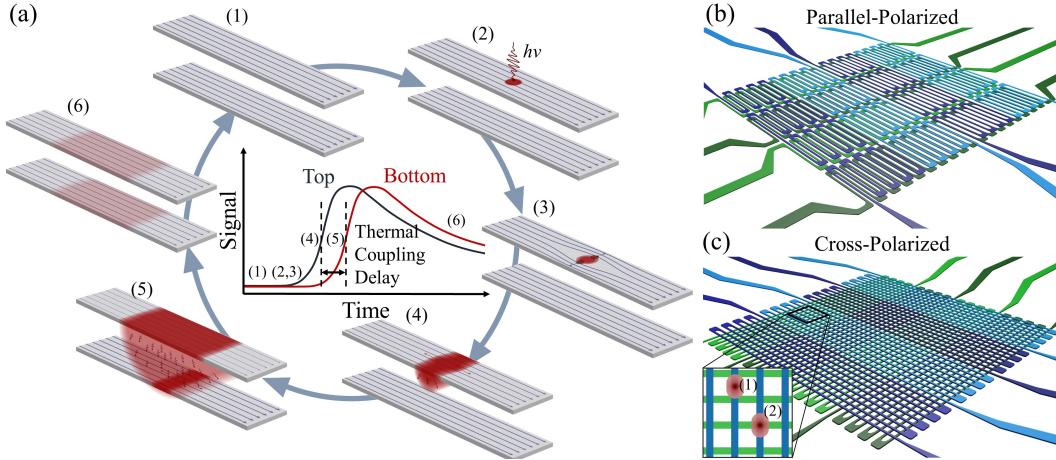
Figure 5.1: (a) Detection process in TRC arrays. The process begins with both nanowires current biased (1) until a photon is absorbed in one nanowire, creating a hotspot of excited quasiparticles (2). As the hotspot evolves, a vortex-antivortex pair unbinds (or vortex entry from the edge for photons absorbed near the edge of the nanowire) (3), leading to the formation of a normal domain. As the normal domain grows along the length of the nanowire, (4) heat is coupled to the dielectric spacer, eventually increasing the temperature of the second nanowire and causing it to switch (5). After current is diverted, both nanowires relax back to the superconducting state (6). Schematic illustration of 4×4 (b) parallel-polarized and (c) cross-polarized thermal row-column devices. The top layer channels (blue) are formed as columns while the bottom layer channels (green) are designed as rows. The area where a column overlaps a given row forms a pixel, so the 4×4 arrays shown here have 16 pixels. The inset of (c) shows a top view of the overlap of the top and bottom nanowires for the cross-polarized design, highlighting how the distance heat must travel to trigger the second layer is greater for absorption location (1) compared to absorption location (2). As the nanowire fill factor increases, this variability decreases.

was sputtered to form an insulating layer. An 8 nm thick WSi film was sputtered from a compound target, after which Au contact pads and leads were patterned via lift-off. The bottom nanowire layer was pattered using electron-beam lithography and etched using an ICP RIE dry-etch of CHF_3 and O_2. After the first nanowire layer was patterned, a buffer layer of ~190 nm SiO_2 was sputtered and smoothed using angled incidence Ar ion milling. The sputtering and ion milling process was necessary to planarize the surface sufficiently to yield the second layer of nanowires without constrictions. The second nanowire layer and leads were patterned using the same process as the lower layer and a final capping layer of ~ 65 nm of SiO_2 was sputtered for passivation. Optical efficiency could be enhanced by designing and depositing antireflection coatings to form an optical stack.
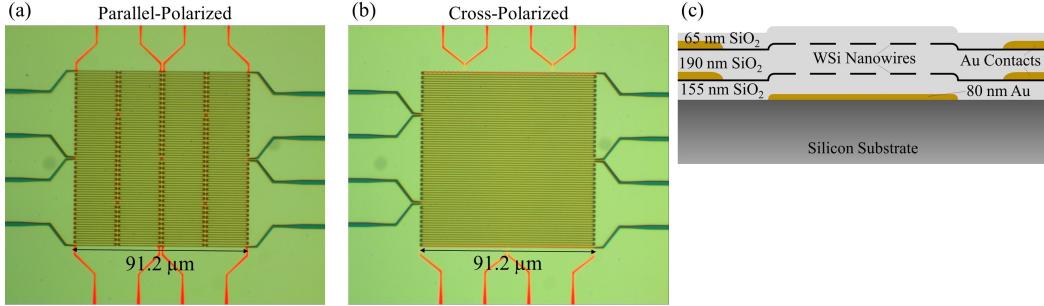
Figure 5.2: Optical microscope images of (a) parallel-polarized and (b) cross-polarized devices. Blue channels are the top layer while red channels are the bottom layer. (c) Schematic cross-section of the TRC devices showing layer thicknesses.

Two designs were fabricated using this procedure. The first was a 4×4 array using parallel nanowires on the top and bottom layers as shown schematically in Figure 5.1b and in the optical micrograph of Figure 5.2a. The total active area was $91.2\,\mu m$ by $91.2\,\mu m$ with each row and column consisting of four $22.8\,\mu m$ by $22.8\,\mu m$ units connected in series. The second design was a 4×4 cross-polarized array using perpendicular top and bottom nanowires as shown schematically in Figure 5.1c and in the microscope image of Figure 5.2b. The total active area was $91.2\,\mu m$ by $91.2\,\mu m$. Both devices use $160\,nm$ wide WSi nanowires with $1200\,nm$ pitch. The low fill factor was selected to avoid crosstalk between adjacent pixels [27] while maintaining a uniform nanowire fill throughout the entire active area. A higher nanowire fill factor could be achieved while avoiding crosstalk by increasing the fill factor within a pixel but leaving additional guard space between adjacent pixels.

Design of TRC arrays requires a balance between the electrical and thermal crosstalk mechanisms in neighboring nanowires. During a detection event in these two-layer devices, both mechanisms couple energy from the detecting nanowire to adjacent channels on both the original and second layers [152]. Capacitive coupling is the dominant mechanism of microwave coupling in this geometry, so as the thickness of the spacer layer between the nanowire layers decreases, the strength of the electrical coupling between neighboring nanowires increases. Electrical crosstalk of this form is undesirable because it is distributed across all channels of the second layer rather than localized to the detecting pixel. Thermal coupling is inherently local to where the detection occurs, but the timescale of this coupling is slow compared to electrical crosstalk. The time difference between the generation of normal domains in the two detecting nanowires, defined as the thermal coupling delay time (see Section 5.3.3), is on the order of a few nanoseconds for the geometries we demonstrate in this

work. An optimal design would minimize the spacer thickness while still preventing electrical crosstalk.

The need to balance electrical and thermal crosstalk dictates the acceptable spacing between the two nanowire layers. Previous experiments on planar arrays of WSi nanowires demonstrated that ~2.5 mm long nanowires experience near-unity thermal crosstalk and noticeable electrical crosstalk at a pitch of 400 nm with an estimated capacitance of 110 fF [27]. For the TRC, we conservatively desired to reduce the estimated electrical crosstalk by a factor of 3 which would reduce it to the level of the amplifier noise. For the spacer thickness of 190 nm chosen in this work, the channel to channel capacitance is estimated to be approximately 31 fF for the parallel-polarized and 24 fF for the cross-polarized devices, meeting the target capacitance requirement. The lack of evidence for electrical crosstalk in our experimental results justifies this choice of spacer thickness, but thinner dielectric layers could likely be used while still avoiding undesired electrical crosstalk and improving the thermal coupling performance.

5.3 Device Characterization

For the two devices which were characterized, all measurements were performed using a temperature-controlled 900 mK stage of a He-3 sorption fridge. Each of the 8 channels of the arrays was individually biased and measured using a DC coupled cryogenic amplifier chain with an additional room temperature low noise amplifier. Readout was performed using 8 channels of a 64 channel time-to-digital converter (TDC) with a comparator front end [27, 29]. Timetags were saved and analyzed in post-processing, but a real time streaming system could be developed using FPGA based readout.

5.3.1 Optical Efficiency

Both styles of device were characterized at a wavelength of 1550 nm in a free-space coupled cryostat. The cryogenic windows of the cryostat include blackbody filters with a total transmission estimated to be greater than 96% at 1550 nm [27]. The channels of the bottom layer exhibit a lower switching current than the top layer due to inconsistencies in the fabrication of the two nanowire layers. Despite this, both layers show saturation of internal efficiency for all channels as shown in Figure 5.3. The number of counts is not distributed evenly between the top and bottom layers due to the properties of the optical cavity, but this distribution is reasonably consistent with rigorous coupled wave analysis (RCWA) modeling of the devices.
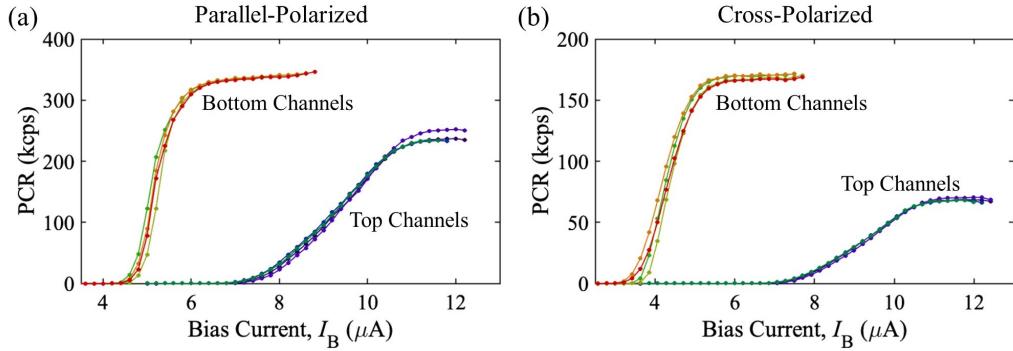
Figure 5.3: Flood illumination photoresponse count rate (PCR) curves at 1550 nm for (a) parallel-polarized and (b) cross-polarized devices when illuminated in the TM polarization with respect to the top nanowire layer. The lower switching current of the bottom layer channels is attributed to inconsistencies in the fabrication of the two layers. Measurement of the photoresponse was conducted with only the top or bottom channels biased at a time to avoid coupling between layers. The small increase in count rate of the bottom layer of the parallel-polarized device near the switching current is attributed to increased photo-sensitivity of the meander bends, which occupy approximately 18% of the active area. Note that this is not present in the cross-polarized device where bends are only located at the edge of the active area.

The optical efficiency of each device was measured using a 1550 nm CW laser focused to a Gaussian spot with a FWHM diameter of approximately 33 µm. Only counts which exhibit unambiguous coincidence groups are included in the efficiency measurement. This includes triple-coincidences (defined below in Section 5.3.2) for the cross-polarized device. We find that the RCWA technique provides a reasonable prediction of device efficiency. The parallel-polarization device exhibits 34% TE and 11% TM efficiency while the RCWA calculation predicts 34% TE and 14% TM efficiency. In the TM polarization, the model predicts a 64:36 bottom to top layer absorption ratio which is close to the experimentally measured 58:42 ratio. The cross-polarized device has 30% TE and 26% TM efficiency while the RCWA calculation predicts 30% TE and 29% TM efficiency for polarization with respect to the top layer orientation. In the TM polarization, the model predicts a 69:31 bottom to top ratio compared to the measured 72:28 ratio. The parallel-polarized device incurs additional optical coupling losses compared to the cross-polarized design due the presence of sections of non-photosensitive meander bends comprising approximately 18% of the active area. This advantage of the cross-polarized architecture is significant for small pixel sizes, but becomes less important

as the pixel size increases or the illuminating spot is not focused on an area of the array where the bends are present.

5.3.2 Thermal Coupling Efficiency

The TRC concept relies upon efficient thermal coupling between the active layers in order to extract the location where a detection occurred. This requires that every detection on one layer triggers a detection on the other layer. We define the coupling efficiency $p_i\left(I_{\mathrm{B,Top}}, I_{\mathrm{B,Bot}}\right)$ as the probability that a 'click' generated on a channel i leads to an unambiguous corresponding 'click(s)' on the second layer. An ideal device would only generate a single corresponding detection on the second layer, making the readout as straightforward as possible. If the additional complexity can be handled by the readout processing, triple coincidences (events with detections on three channels within a coincidence window) can be tolerated with the understanding that the detection occurred at the region where the pixels meet. This makes the implicit assumption that the probability of two photons being detected at the same time is negligibly small.

The coupling probability for the two array designs was characterized using a 1550 nm mode-locked laser for a range of bias currents on both top and bottom layers. All four channels of a given layer were biased at the same current and the photon flux was kept sufficiently low to make the probability of two photons being detected in the same optical pulse negligible. For each channel, a bias dependent calibration delay was applied to all timetags such that the mean delay for photons detected on each channel is zero based on the timebase of the mode-locked laser. The coupling probability is shown as a function of bias current in Figure 5.4 for representative channels on both the bottom and top layers. The parallel-polarization device exhibits near-unity coupling efficiency for a wide range of bias current combinations $\left(I_{\mathrm{B,Top}}, I_{\mathrm{B,Bot}}\right)$ as is expected based on the direct overlap between the nanowire regions of the two layers (see Figure 5.1b). The coupling efficiency increases to near unity as both the detecting and receiving nanowire bias currents increase. The explanation is straightforward. As the detecting nanowire bias current increases, more Joule heat is released during detection which heats the dielectric layer and second nanowire layer, increasing the coupling efficiency. As the bias current on the receiving layer increases, a smaller change in nanowire temperature is required to trigger a switching event, making the nanowire more sensitive to small heat pulses generated by the source nanowire.
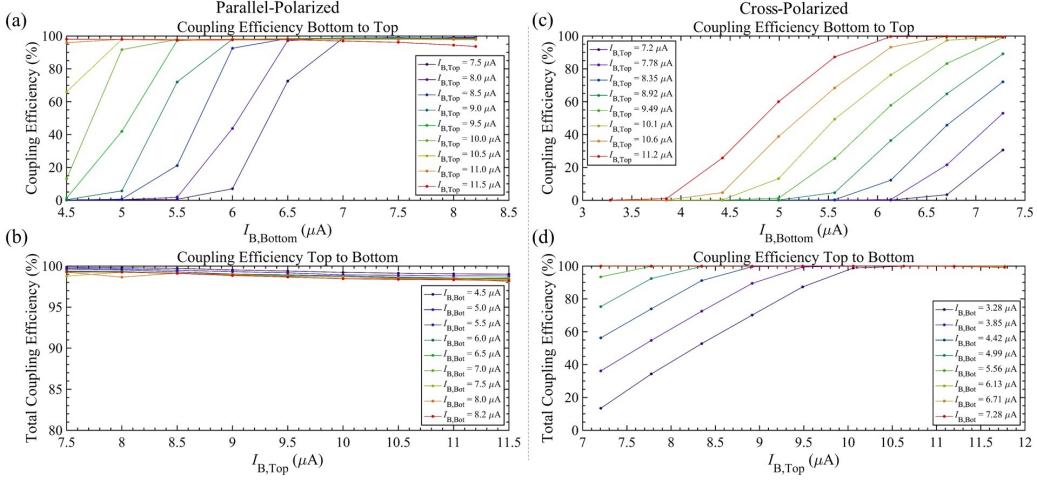
Figure 5.4: Coupling efficiency for representative channels of the parallel-polarized (a, b) and cross-polarized arrays (c, d). The coupling efficiency increases as both the source and receiving channel bias currents increase. Coupling from the bottom to top layer is less efficient than top to bottom due to the lower overall bias current and lower Joule heat generated during a detection. The transition from no coupling to efficient coupling is more gradual in the cross-polarized device compared to the parallel-polarized device due to the non-uniform overlap between the nanowires of the two layers. Lines are to guide the eye.

The cross-polarized device displays more varied behavior due to the nonuniform overlap between the two layers (see Figure 5.1c). Compared to the parallel-polarization device, higher bias currents are required to ensure near unity $p_i(I_{\text{B,Top}}, I_{\text{B,Bot}})$, but when both layers are biased near their switching currents, unity detection can be achieved. Furthermore, the transition region between no coupling and efficient coupling occurs over a wider range of bias currents. This is consistent with the varying overlap between the two nanowire layers.

The total coupling efficiency between layers in the cross-polarized device increases monotonically as the bias current increases on either the detecting or receiving channels, but the fraction of double coincidences (two detection events within a coincidence window) and triple coincidences (three detection events) is not constant. As seen in Figure 5.5a, as the bias current on the bottom channel increases, the coupling efficiency for double coincidences, shown schematically in Figure 5.5c, initially increases due to the increased Joule heat generated. However, this coupling efficiency actually decreases at the highest bias currents of the top channel. This occurs because the large amount of heat generated by the combination of the two channels switching can be sufficient to switch a neighboring channel which is part
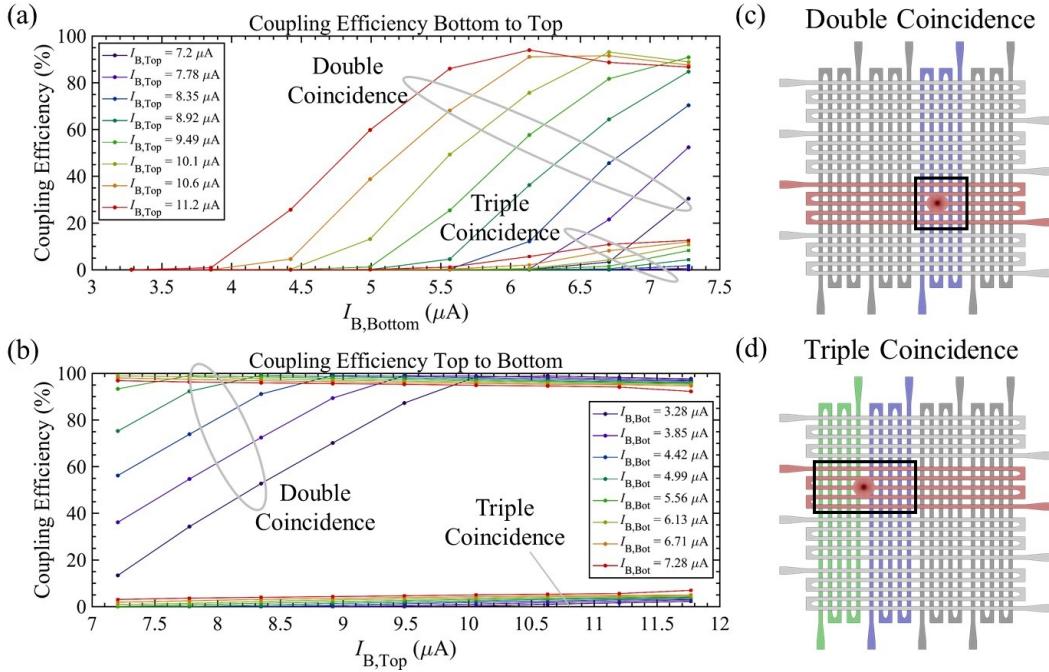
Figure 5.5: Cross-polarized array coupling efficiency by coincidence number for (a) a bottom channel triggering a top channel and (b) a top channel triggering a bottom channel. The number of triple coincidences is non-negligible for high bias currents on the either the top or bottom channels. Unambiguous double-coincidence and single-photon triple-coincidence detection events are shown schematically in (c) and (d), respectively. The colored channels indicate those which register a detection during the photon-detection event, and the overlap of the detecting channels indicates the determined location.

of a different pixel. The number of these triple coincidences, shown schematically in Figure 5.5d, increases as both the top and bottom bias currents increase, but the bias current of the top channel plays a more significant role than that of the bottom channel due to the larger overall bias currents involved. Triple coincidences are only considered to be valid if the two detecting channels on the same layer are adjacent and the probability of two photons being absorbed within a coincidence window is negligibly small. Analysis of the correlations confirms that at high bias current combinations, $> 98\%$ of the triple-coincidence events occur between two adjacent channels on one layer and a single channel on the second layer.

Triple coincidences are not inherently problematic in determining the location of the photon detection because at high bias current operating points, $>99.5\%$ of triple coincidences occur with the first receiving channel belonging to the second layer of the device. Thus, a proper detection location can be assigned based on the detecting

and first receiving channels alone. However, to simplify an FPGA based readout system, it is desirable to have only double-coincidence events. Furthermore, the possibility of single-photon triple-coincidence detection events makes it impossible to distinguish between blocking caused by two photon detections and a single-photon triple-coincidence event. In a cross-polarized device, single-photon triple coincidences can be minimized by leaving guard space between adjacent pixels. In the parallel-polarized array, single-photon triple coincidences make up a negligible fraction of the detection events for the large pitch arrays tested in this work.

5.3.3 Thermal Coupling Time

Thermal coupling between layers is not instantaneous. There is a delay on the order of several nanoseconds between associated detection events due to the finite time required for heat to propagate from one nanowire to the second layer. Furthermore, fluctuations in the heat transfer process lead to additional jitter in the detection timing of the second event compared to the first. The coupling delay time is defined as the time difference between the initial detection event and the detection on the second layer. The thermal coupling delay is a function of the bias currents on both the detecting and receiving channels. Coupling delay time histograms are shown in Figure 5.6 and average delay times in Figure 5.7 for a variety of bias conditions for the two arrays. As can be expected based on the geometry, the parallel-polarized array demonstrates faster coupling and a narrower distribution of coupling times between layers than the cross-polarized device. Differences in the distance heat must propagate to switch the second nanowire layer in the cross-polarized device lead to the appearance of a shoulder and extended long-delay time tail in the coupling behavior. While this effect is diminished at higher bias currents, it is not eliminated completely. Consequently, when compared to the parallel-polarized device, the cross-polarized design requires a wider coincidence window when defining correlated detection events.

Figure 5.7 also shows the bounds containing 90% of the thermal coupling delay distribution, indicating the amount of thermal coupling jitter. This corresponds to approximately 3.29σ for a Gaussian distribution. For the parallel-polarized device biased near the switching current, the top-to-bottom thermal coupling jitter is comparable to the photon-detection jitter of the system at approximately 260 ps (see Figure 5.8), but for lower bias currents, thermal coupling adds significant jitter. For the cross-polarized array, even at high bias currents, for the top nanowire triggering the bottom, significant thermal coupling jitter of 600 ps persists due to
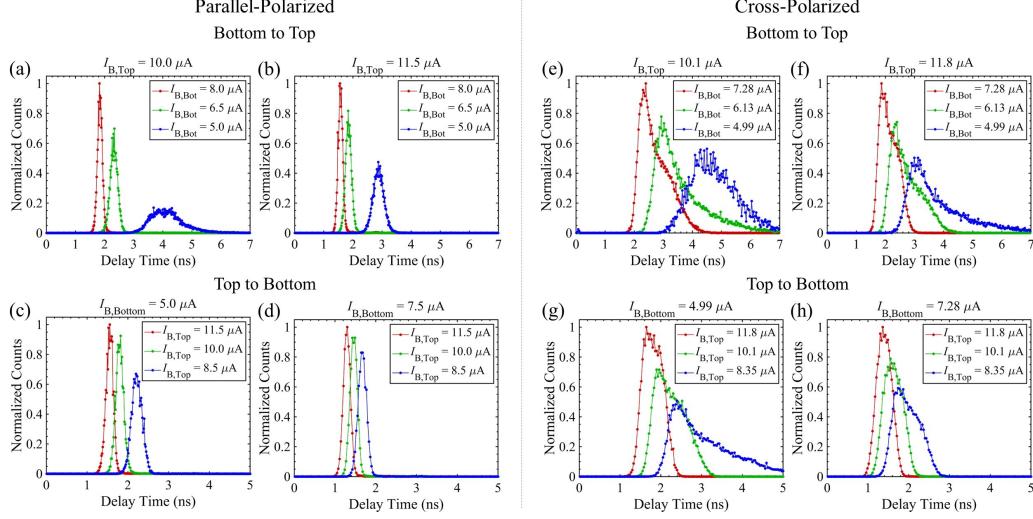
Figure 5.6: Thermal coupling delay histograms for parallel-polarized (a-d) and cross-polarized devices (e-h). The coupling efficiency increases as both the source and receiving channel bias currents increase. Coupling from the bottom to top layer is less efficient than top to bottom due to the lower overall bias current and lower Joule heat generated during a detection.

the non-uniform geometry as discussed throughout the text.

As the bias currents increase, the delay time decreases and the distribution width becomes narrower. Despite separation distances of only a few hundred nanometers between layers, the delay times are on the order of several nanoseconds. This is significantly slower than expected based on ballistic propagation of phonons and slower than expected based on thermal modeling using a Casimir limited thermal conductivity for thin-film dielectrics. There is experimental evidence that this type of suppressed thermal conductivity in thin-film SiO_2 is typical [5, 135], but more advanced modeling is needed to fully characterize the behavior, then use the results to optimize the thermal design of TRC arrays. Specialized devices can be designed which use a variety of nanowire widths, inductances, and overlap spacings in order to build a detailed experimental dataset of the coupling delay times between nanowires under different geometries. This type of device offers a new means of studying this physics and improving both engineering and fundamental models of this process. Previous work demonstrated that a complete electrothermal model of the reset dynamics of WSi nanowires requires an accurate model of the heat transfer in the surrounding dielectric material [5], but the TRC architecture offers a way of experimentally studying this process in order to develop and validate such a model.
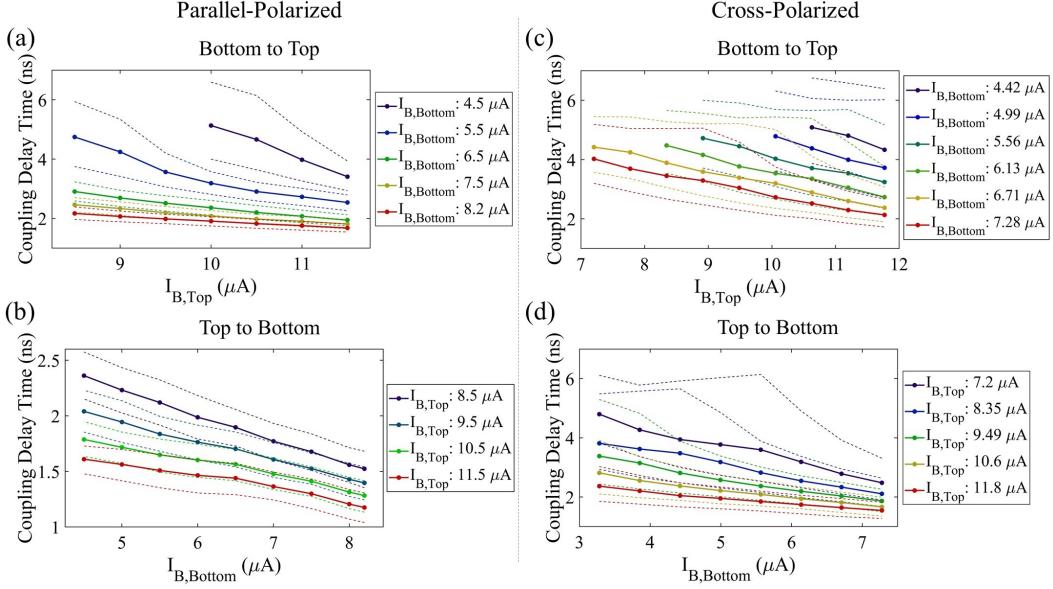
Figure 5.7: Thermal coupling delay times for (a, b) parallel-polarized and (c, d) cross-polarized representative channels of the arrays. Solid lines with symbols indicate the mean delay time while dashed lines indicate the 90% distribution bounds to show the width of the coupling delay time distribution. The coupling times and the width of the coupling distribution both decrease as the bias current increases on either the source or receiving channels. The cross-polarized channels show a wider distribution and longer mean coupling time than the parallel-polarized channels for comparable bias conditions.

This understanding is necessary to fully optimize TRC arrays.

5.3.4 Timing Jitter

The timing jitter of the system was characterized using a 20 MHz repetition rate 1550 nm mode-locked laser. A phase-locked loop circuit converted the electrical sync of the laser to a 10 MHz clock which acted as the timing source for the TDC. The laser was focused to a Gaussian spot with a diameter of approximately 85 μm FWHM. Due to non-Gaussian instrument response function (IRF) of the array channels, the timing jitter is defined as 2.355σ where σ is the sample standard deviation of the photocount timetags modulo 50 ns. Figure 5.8 shows the IRF and the bias current-dependent timing jitter for representative channels of both layers of the array, demonstrating that the overall jitter is less than 300 ps at optimal bias regardless of which nanowire absorbs the photon. The jitter of a given pixel of the array is the combination of the two histograms of the channels forming the pixel weighted by the relative absorption efficiency of the two channels because the
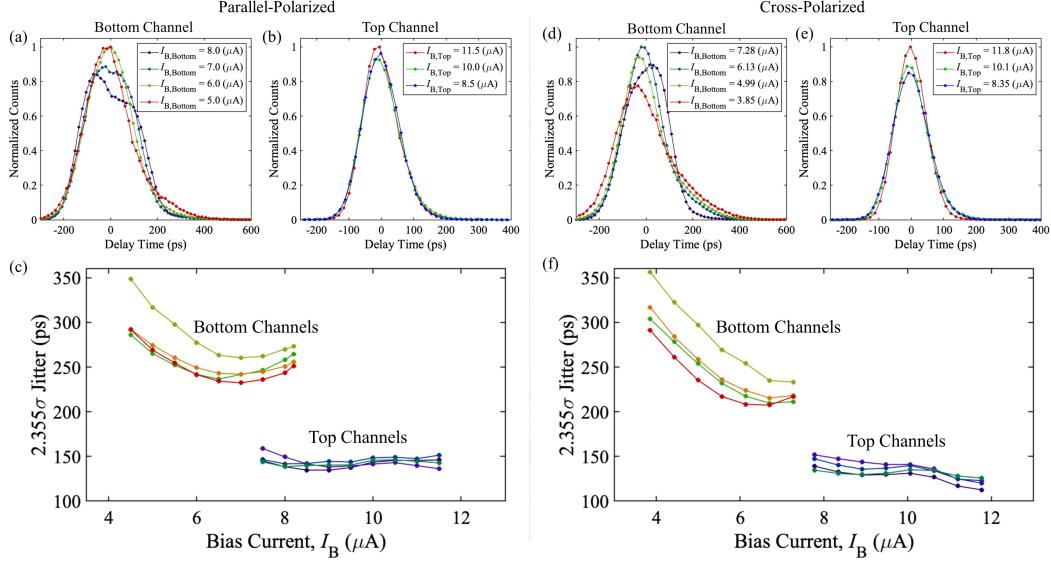
Figure 5.8: Jitter characterization for channels of the (a-c) parallel-polarized and (d-f) cross-polarized arrays. Jitter histograms for a representative channel of the bottom layer (a, d) show non-Gaussian IRF with a distinct asymmetry in the distribution at high bias currents. We attribute this to electrical coupling between the nanowire layers and to the back reflector which reduces the signal propagation velocity and leads longitudinal geometric jitter[32]. In contrast, the IRF of the top channels (b, e) is nearly Gaussian. To accommodate the non-Gaussian behavior of the lower layer channels, the total jitter is defined by 2.355σ (c, f), and is significantly larger for the bottom channels than the top, even when biased at the similar currents where the electrical noise contribution to the jitter from the readout electronics is similar. A device optimized for improved timing performance must manage the electrical coupling and signal propagation of the bottom layer to take full advantage of the low timing jitter of SNSPDs.

measured arrival time of the photon is given by the detection time of the first channel to detect the photon.

5.3.5 Maximum Count Rate

The MCR was characterized using a 1550 nm CW source with variable attenuators and a focused spot with a FWHM diameter of 85 nm on the array. The MCR is defined as the count rate where the efficiency of the detector drops by 3 dB compared to the low count rate efficiency. The MCR curves are shown below in Figure 5.9 for both array designs. The parallel-polarized array demonstrates an MCR of 14.6 Mcps while it is only 10.8 Mcps for the cross-polarized array. The parallel-polarized array was biased at 11.5 μA and 8.2 μA for the top and bottom channels, respectively, while the cross-polarized array was biased at 11.8 μA and 7.28 μA for the top and
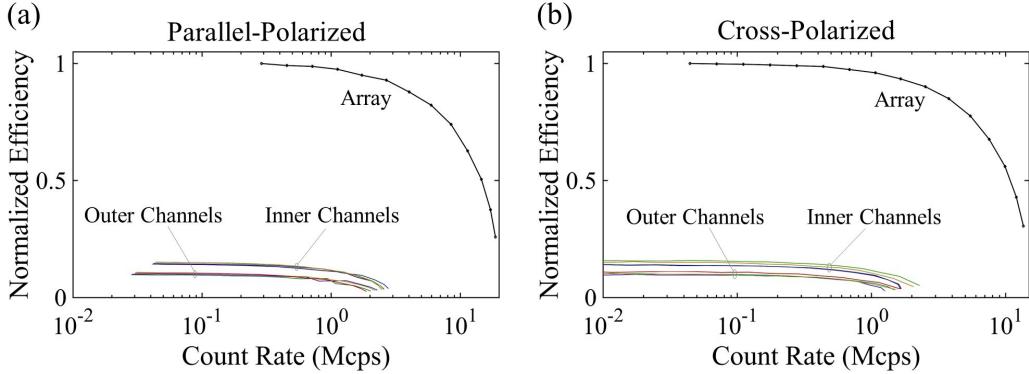
Figure 5.9: MCR curves for (a) parallel-polarized and (b) cross-polarized arrays. The parallel-polarized device was illuminated with TE polarized light while the cross-polarized array was illuminated with linear polarization rotated 45° with respect to the TE and TM modes. For both arrays, the two inner channels of each layer show higher count rates than the two outer channels due to unequal illumination with the laser spot. In these figures, the parallel-polarized device uses a coincidence window of 3.5 ns while the cross-polarized array uses a coincidence window of 4.5 ns.

bottom channels, respectively. Due to the polarization-dependent relative efficiency of the channel layers, the MCR of the cross-polarized device is slightly polarization dependent. When photons are preferentially absorbed in either the top or bottom layer, the array experiences elevated blocking loss, resulting in a reduced MCR. The MCR depends on the coincidence window chosen to analyze the timetag data. For the parallel-polarized data, the optimal window is 3.5 ns while for the cross-polarized array, the optimal window is 4.5 ns. The larger optimal window for the cross-polarized array is consistent with the longer thermal coupling delay time for these bias currents. In the MCR analysis of the cross-polarized array, triple coincidences were not included as unambiguous counts because the photon flux is not sufficiently low to eliminate the possibility of multiple photons being detected within a coincidence window. Single-photon triple coincidences lead to additional blocking loss and are the primary cause of the lower MCR in the cross-polarized array compared to the parallel-polarized device.

The thermal coupling delay times can have a direct impact on the MCR which can be achieved over the array without position ambiguity. Two factors contribute to position ambiguity. The first is due to blocking loss. During the deadtime of a channel, detections which occur on the other layer but overlapping the dead channel will not have a correlated click. In the limits of equal illumination of pixels, ideal

thermal coupling, identical channel properties, and Poisson distributed photons, the array MCR (MCR_A) scales according to $MCR_A \sim N^2 \, MCR_i / (2N - 1)$ due to this deadtime, where MCR_i is the channel MCR. The second factor is associated with the timing uncertainty of correlated detections. Electrical and thermal timing jitter lead to a range of thermal coupling times between the photon-induced detection and the thermally-coupled detection. When analyzing the channel detection times to determine the coincidence events, one must define a range of times during which two events can be correlated, known as a coincidence window. Two detection pairs which occur within the same coincidence window lead to ambiguity as to which combination of pixels were the source of detection events because two rows and two columns have four potential pixel locations. As the array size becomes large, the coincidence window of correlated detections limits the counting rate. A large thermal coupling delay indicates that the rate of temperature increase in the second nanowire is slow. Therefore, fluctuations during thermal coupling process lead to larger fluctuations in the coupling delay time, analogous to the dependence of electrical noise jitter on signal slew rate. This requires setting a larger coincidence window, which is only acceptable for low count-rate applications. Reducing the thickness of the spacer layer quickens thermal coupling and potentially leads to a smaller coincidence window, but leads to additional electrical coupling. An optimized design would reduce the thickness of the dielectric spacer to the minimum thickness where the electrical crosstalk can still be tolerated.

5.3.6   Imaging

To demonstrate the imaging capability of the TRC devices, we translated a laser spot on the active area of the device using a fast steering mirror. For the cross-polarized device, the spot was directed in a square pattern on the device with varying periods, and the location of the spot was extracted using the centroid of the measured counts. The spot tracking was reliably demonstrated up to limits of the steering mirror speed with a period of 4 ms using 250 µs frames and an average of 11.7 photon detections per frame. Figure 5.10 shows representative sample frames demonstrating this tracking capability, but based on the MCR performance of the device, a similar level of tracking accuracy could be achieved at a 100 µs period with a photon flux increase by a factor 50 with minimal blocking loss.
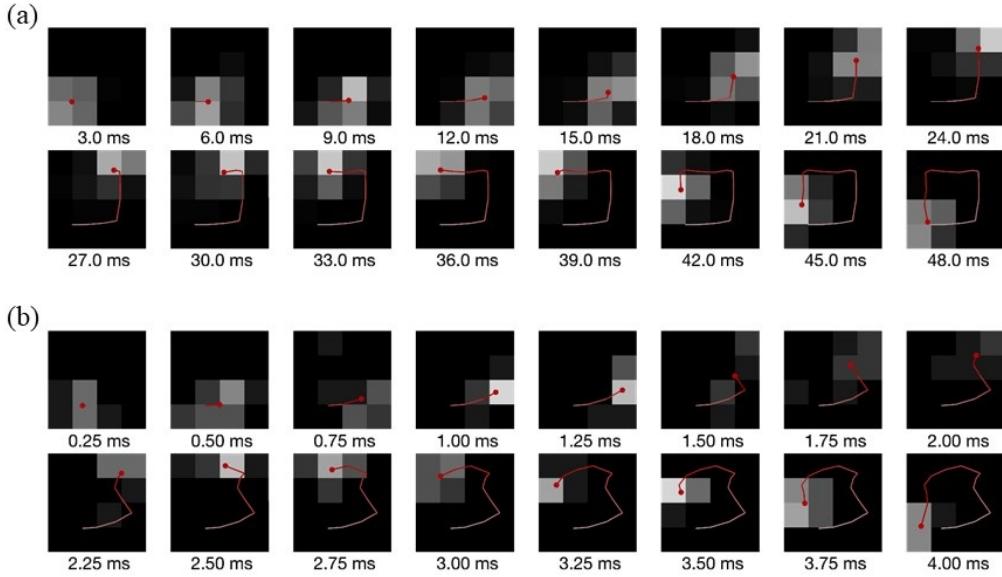
Figure 5.10: Laser spot tracking in a square pattern using the cross-polarized array. (a) Tracking using a 48 ms period and 3 ms frames with an average of 146.5 photon detections per frame. (b) Tracking using a 4 ms period and 250 µs frames with an average of 11.7 photon detections per frame. The red circle indicates the centroid of the detections for a given frame and the line shows the path of the centroid for the sequence of frames.

5.4 Discussion

The TRC architecture is not limited to the WSi material system. For devices which require lower jitter and higher count rates, NbN or NbTiN arrays are possible. The challenge with using these higher T_c materials is ensuring sufficient thermal coupling between the two layers to generate thermally-induced switching. The greater T_c requires that the temperature of the second nanowire layer must be elevated to a higher temperature to switch the device when biased at a given fraction of the nanowire depairing current. While this is not expected to be a problem for the parallel-polarized design, cross-polarized devices may require a high fill factor in order to ensure that there is sufficient overlap between the layers for efficient coupling. A better ability to model heat transfer in thin dielectrics and superconductors is needed to reliably predict the thermal coupling between superconducting nanowire layers in TRC devices.

Apart from the standard fabrication challenges of yielding large arrays of SNSPDs, there are no roadblocks to scaling the TRC to the kilopixel scale. Readout systems with 64 channels have been demonstrated with both direct [27] and SFQ [31]

architectures enabling 32×32 format arrays. To approach the megapixel scale, consideration must be given to the thermal load of the $2N$ RF cables required to operate an $N \times N$ array and the associated electronics. This can be overcome by using lower bandwidth cables which have a smaller thermal load at the expense of timing resolution of the array. For low count rate applications, a more practical solution would employ additional multiplexing through time-domain or standard electrical row-column schemes to improve the imaging resolution while maintaining a small number of high bandwidth cables. Such an approach would require determining four-fold coincidence events, but this is entirely within the capability of current FPGA technology.

In conclusion, we have demonstrated a row-column style multiplexing architecture for SNSPD imaging arrays using thermal coupling between two active layers. This new architecture eliminates the current-redistribution and limited-fill-factor shortcomings of electrically coupled row-column SNSPD arrays. Similar device structures using this platform provide a new means of experimentally probing low temperature heat transfer properties of thin dielectric structures and can be used to produce optimized thermal row-column arrays in the future. The 16-pixel devices characterized in this work are immediately scalable to the kilopixel size for use with existing readout systems and, when integrated with optical cavities, are expected to achieve optical efficiency exceeding 85%. When combined with an additional multiplexing scheme through time-domain or standard electrical row-column mechanisms, it will be possible to reach megapixel sized arrays with a manageable number of readout cables. The thermal row-column promises to be a key enabling technology for achieving large-area, low intrinsic dark count rate, and high-efficiency photon-counting imaging systems for low photon-flux signals.

C h a p t e r 6

DETECTOR OPTIMIZATION FOR DSOC

As the scientific instruments of spacecraft become more advanced, they require sending larger amounts of data back to Earth. Data is transmitted and received using the deep space network (DSN) which currently operates using S, X, K, and K_a band radio. However, as the number of operating spacecraft continues to grow, bandwidth on the DSN is becoming a limiting factor for the transmission of scientific data. The flood of spacecraft scheduled to launch to Mars in 2020 will only exacerbate this problem. Alternative methods are needed to communicate with spacecraft, and in particular, downlink data from these craft.

For deep space communication, optical communication links show promise over traditional radio frequency communication (RF) because they can provide an expected 10 times higher data rate for the same size, weight, and power (SWAP) on the spacecraft. The Deep Space Optical Communication (DSOC) demonstration mission is a technology demonstration payload scheduled to launch aboard the Psyche 2022 mission which will carry an optical transceiver for both uplink and downlink demonstrations. The downlink operates using 1550 nm wavelength light. Therefore, the ground detectors for this and future photon-starved optical communication links require large-area, high-efficiency infrared photon-counting detector arrays with time resolution in the hundreds of picoseconds in order to utilize efficient pulse position modulation (PPM) encoding schemes. With their high efficiency in the near-infrared [11–13], low dark count rates [11, 14, 15], and low timing jitter [2], SNSPDs offer a promising technology for these ground detectors. Historically, SNSPDs have often struggled to achieve high performance over large areas due to problems of yield in polycrystalline superconductors [154–156].[1] For this reason, the team at JPL has developed large-area SNSPD arrays from the amorphous WSi material system [19] which have demonstrated yield over larger areas. Compared to the more traditional NbN material, WSi has drawbacks due to its lower operating temperature and slower timing performance, so designing an array which meets

---

[1]When the DSOC project was proposed and the technology development was performed, it was commonly accepted that material quality and uniformity of thin-film polycrystalline superconductors was not sufficient to yield arrays with areas spanning the 320 μm diameter required by the DSOC project. Subsequent development has improved film quality and yield [157], and NbN and NbTiN arrays are being considered for future optical communication ground detectors.

the performance requirements of the DSOC program requires careful optimization. This chapter describes some of the work performed to design and optimize the performance of the prototype ground detector for the DSOC project which was carried out between 2015 and 2017.

6.1 Detector Design

The detector for the DSOC project is required to have low timing jitter (<235 ps FWHM), high efficiency at 1550 nm (> 70% targeted), a maximum count rate (MCR) approaching 1 gigacount per second, and fill a 320 µm diameter active area while maintaining a quad-cell structure to provide centroiding information for tip-tilt adjustments. Naturally, adding more and more pixels to an array increases the timing performance of the system, but this comes at a cost. Each pixel requires its own high bandwidth RF readout channel, which adds to the heat load on the cryogenic stage. Therefore, the cooling power of the cryostat places a limit on the number of pixels which can be operated. In the direct readout approach, each channel also requires its own dedicated amplifier and time-to-digital conversion hardware, which leads to additional cost and complexity as the number of channels increases. The JPL team committed to a 64-pixel approach early in the design process. This was chosen in order to keep the heat load on the cryostat manageable while minimizing the amount of specialized hardware development required to achieve a functioning time-tagging systems.

There are several considerations which make the design of optical communication ground detectors somewhat different than standard arrays. Ideally, all of the pixels of the array would be equally illuminated by the incident light in order to minimize the blocking loss experienced by the detector. For an SNSPD, this means the nanowires should be interleaved as shown in Figure 6.1 to minimize spatial variations. Another consideration is the nanowire fill factor, or the fraction of the active area which is covered by photosensitive nanowires. Increasing the fill factor improves optical absorption and a high fill factor is typically required to achieve high detection efficiency [11]. For a fixed active area, this introduces a tradeoff in timing performance. Implementing a higher fill factor while maintaining the same number of pixels increases the length of each nanowire, which increases the kinetic inductance proportionally. The recovery time of the detector is proportional to this kinetic inductance, so the MCR is limited by this factor. Also, the rise-time is linked to the kinetic inductance, with a slower rise-time contributing to the timing jitter due to the presence of electrical noise. Furthermore, with sufficiently long nanowires,

geometric effects become important [70] which add additional timing jitter. Therefore, there is a tradeoff between optical efficiency and timing properties which must be addressed by the appropriate choice of nanowire fill factor. A further constraint is placed on the system by the occurrence of electrical and thermal crosstalk between adjacent co-wound nanowires as discussed in Chapter 4.
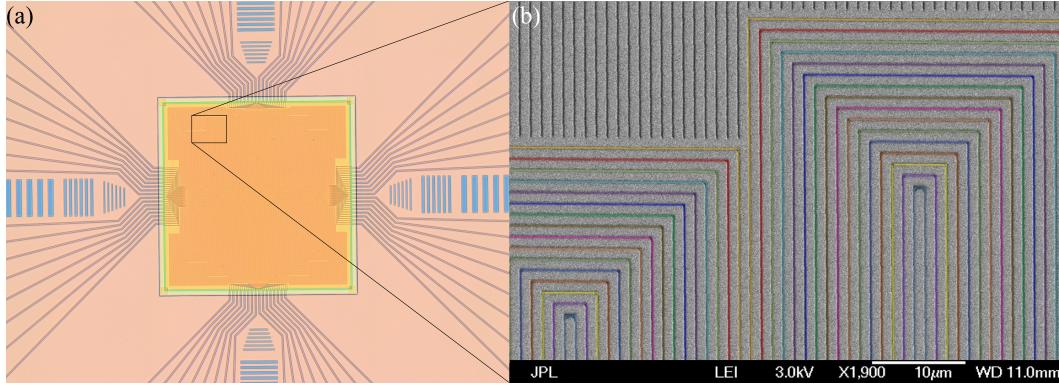
Figure 6.1: (a) Optical micrograph of the detector. The visible leads form $50\,\Omega$ matched CPW structures to route the nanowire signals to the device bonding pads (not shown). (b) False color scanning electron microscope (SEM) image of a section of one quadrant of the array. The 16 pixels of each quadrant are co-wound to minimize the illumination variation between pixels. The four quadrant structure allows for beam centering though an active feedback tip-tilt mirror which equalizes the count rates recorded by each quadrant.

The goal of the present work was to examine the problem in terms of the optical coupling which can be achieved for various anti-reflection (AR) coating designs and nanowire fill factors. Standard optical cavities for WSi detectors use a bilayer of SiO_2 and TiO_2 as an AR coating and a metal back reflector underneath the nanowire layer to enhance the optical absorption [11], but this design leads to poor optical coupling at low fill factors. Standard designs also only considers absorption at normal incidence. In order to efficiently couple to a 5 m telescope while maintaining a relatively small active area, the optical system uses a cryogenic lens with a large numerical aperture. This means the detector must be able to couple efficiently to light incoming at angles up to ∼25 degrees.

The rigorous coupled-wave analysis (RCWA) [63–65] technique was used to calculate and optimize the expected optical absorption in various SNSPD designs. Standard RCWA software only calculates the transmission and reflection of the optical system for a given incident plane wave. For SNSPDs, we are interested only

in light which is absorbed in the nanowire itself. Any light which is absorbed in the mirror layer or dielectrics does not contribute to photon counting and should be excluded when estimating the expected performance of a design. Therefore, I enhanced the capabilities of the available RCWA code to calculate the absorption which occurs in each layer of the optical stack. This becomes increasingly important as the fill factor of detector decreases.

Several features of optical cavities were investigated to determine a practical solution for the DSOC detector. First, the number of AR bi-layers (high index – low index pairs) was altered, finding that an increased number of layers leads to higher absorption at the design wavelength of 1550 nm. However, this comes at the cost of a narrower cavity. For focal planes operating at a single wavelength with a narrow field of view, this might not be a problem, but because the DSOC detector must accept many incidence angles, the narrower spectral cavity corresponds with a decrease in absorption efficiency as the incidence angle increases. This is intuitive because the dielectric layers appear to be thicker relative to the wavelength for light propagating at an angle relative to the vertical direction of the optical stack. The high index material of the AR layers was also altered by investigating α-Si, TiO$_2$, and SiN which are all available within the JPL fabrication facilities. As expected based on the higher index of refraction contrast, AR coatings based on α-Si show the highest performance at small incidence angles with improved efficiency and requiring fewer bi-layers to achieve a given optical coupling efficiency. However, these designs show a sharper drop in efficiency at larger incidence angles. Finally, devices with a dielectric mirror were modeled in order to understand the importance of the losses in the gold mirror. While the dielectric mirror leads to substantial improvements in performance for low fill factor devices at normal incidence, the dielectric mirror cavities show a stronger angular dependence, making them less suitable for applications requiring high efficiency at a wide range of incidence angles.

Based on ideal modeling, a device using α-Si as the high index AR coating material seems to provide the best overall performance, but an appropriate design also requires consideration of the fabrication feasibility. During fabrication, the deposition of thin-film dielectrics is not perfect. Small errors in thickness occur due to drift in the deposition conditions and deposition rate. To account for this potential problem, random noise was added to the layer thicknesses of an ideal stack design and simulated for several iterations of these modified stacks in order to understand the sensitivity of the design to fabrication drift. Not surprisingly, designs with a

narrower bandwidth cavity were more sensitive to these fluctuations. Thus, cavity designs with more AR bi-layers or a dielectric mirror were more susceptible to fabrication problems. The design which appeared to produce the most reasonable compromise between ideal efficiency and robustness to fabrication uncertainty consisted of two bi-layers of SiO_2 and TiO_2 forming the AR coating with a Au mirror forming the back reflector.

Crosstalk measurements (see Chapter 4) demonstrated that 20% fill factor devices demonstrated significant crosstalk while 13.3% fill factor devices did not. Therefore, the final result of this optimization and fabrication process was a prototype DSOC array based on 160 nm wide WSi nanowires on a 1200 nm pitch (13.3% fill factor) with two bi-layers of SiO_2 and TiO_2 forming the AR coating. A schematic cross-section of the device design is shown in Figure 6.2a and the modeled optical absorption is shown in Figure 6.2b. This detector was fully characterized to determine its suitability in the other performance metrics as described below.
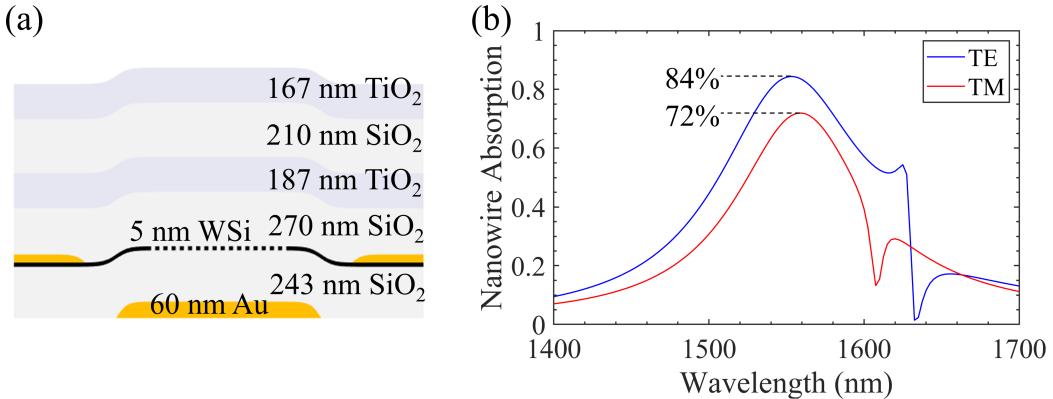
Figure 6.2: (a) Cross-section of the detector. The design consists of two bi-layers of SiO_2 and TiO_2 forming the AR coating and a Au mirror forming the back reflector. (b) Model of the optical absorption in the nanowire layer calculated using RCWA software.

6.2 Detector Characterization

The SNSPD array was operated in a closed cycle ^{3}He sorption refrigerator (Chase CRC10 038) mounted on the second stage of a two-stage pulse tube (Cryomech PT415). The temperature of the first stage pulse tube was 40 K, the temperature of the second stage pulse tube was 4 K, and the base temperature of the cold plate was approximately 770 mK during standard operation.

Pixels of the array were electrically monitored using a combination of cryogenic

and room temperature amplifiers. Each quadrant of the array was connected to a 16-channel brass stripline flex-print circuit (Samtec) at the cold plate. The flex tapes transmitted the signals from the cold plate to the 40 K stage where the electrical signals were amplified using cryogenic PHEMT (Avago Technologies ATF 35143) and SiGe (RFMD SGL0622Z) amplifiers. A low-noise amplifier (Minicircuits RAM 8A+) at room temperature was used to improve the gain. The input of the PHEMT stage was DC coupled with a $50\,\Omega$ termination to avoid the re-biasing challenges associated with AC coupled amplifiers operating at high count rates [69].

Light was free space coupled to the SNSPD array using three optical windows through the radiation shields of the cryostat. For the measurements described here, the window through the 300 K stage of the fridge was 2 inch diameter 12 mm thick AR-coated BK7 (Thorlabs WG12012-C). The 40 K window was 0.5 inch diameter and 0.5 inch thick AR-coated BK7 (Thorlabs) while the 4 K window is 0.5 inch diameter and 0.25 inch thick AR-coated BK7 (Thorlabs). The thick BK7 windows were necessary to help filter black body radiation from the room temperature environment because WSi SNSPDs are sensitive in the mid-infrared [17]. Light was coupled using a CW tunable laser (New Focus Venturi TLB 6600) at λ=1550 nm to measure the efficiency, dark counts, and MCR properties of the array while a mode-locked laser emitting λ=1550 nm light with a repetition rate of 10 MHz and pulse width of ~100 fs was used to measure the timing jitter.

Pixel pulse count rates (CR_P) were measured using a 64-channel programmable threshold comparator followed by a 64-channel time to digital converter (TDC). The pixel background count rate (BCR_P) was defined as the count rate measured with the light source blocked by a shutter in the free space optics beam path and the pixel photoresponse count rate (PCR_P) was defined as the difference between the count rate measured with the light source coupled to the detector and the BCR_P. The pixel system detection efficiency (SDE_P) was estimated as the ratio of PCR_P to the number of photons incident on the array. Using the measurements of SDE_P, the array system detection efficiency (SDE_A) was then defined as the sum of the SDE_P for all active pixels. The timing properties of the array were measured using a LeCroy WaveMaster 8600 oscilloscope with a 6 GHz bandwidth and a 20 Gsample/s acquisition rate.

Each pixel was individually biased with a voltage source and a $100\,\mathrm{k\Omega}$ resistor at room temperature. A resistive bias-tee coupled the bias current to the detector at the input of the cryogenic amplifier. Series resistance between the nanowire and

input of the cryogenic amplifier led to current splitting between the device and the $50\,\Omega$ input termination of the amplifier. The splitting ratio was determined by measuring the PCR_P for all channels in one quadrant of the array using a standard AC coupled scheme with a room temperature bias-tee and amplifiers, then fitting the splitting ratio to match the associated PCR_P curves measured with the cryogenic amplifier chain. The switching current (I_{sw}) of a pixel is defined as the current above which the nanowire switches to the normal state and is no longer photosensitive. The switching current depends on the amplifier chain used for the measurement. The AC coupled room temperature amplifier chain has a lower switching current because noise can induce latching at currents much lower than the depairing current of the nanowire. The DC coupled cryogenic amplifier chain provides a $50\,\Omega$ shunt resistance to ground which delays the onset of latching due to noise and electrical reflections caused by the impedance mismatch at the input of the readout amplifier. Due to the higher switching currents measured while using the DC coupled chain, we define the switching current based on those measurements.

6.2.1 Optical Characterization

Basic optical characterization was performed at 1550 nm using the free-space optics system shown in Figure 6.3. These optics were mounted on an optical breadboard which slides underneath the fridge to illuminate the detector through the cryogenic windows located on the bottom of each radiation shield of the cryostat. An optical fiber connected the optics of the sliding breadboard to the laser source and optical attenuators which were used for changing the illumination conditions on the detector.
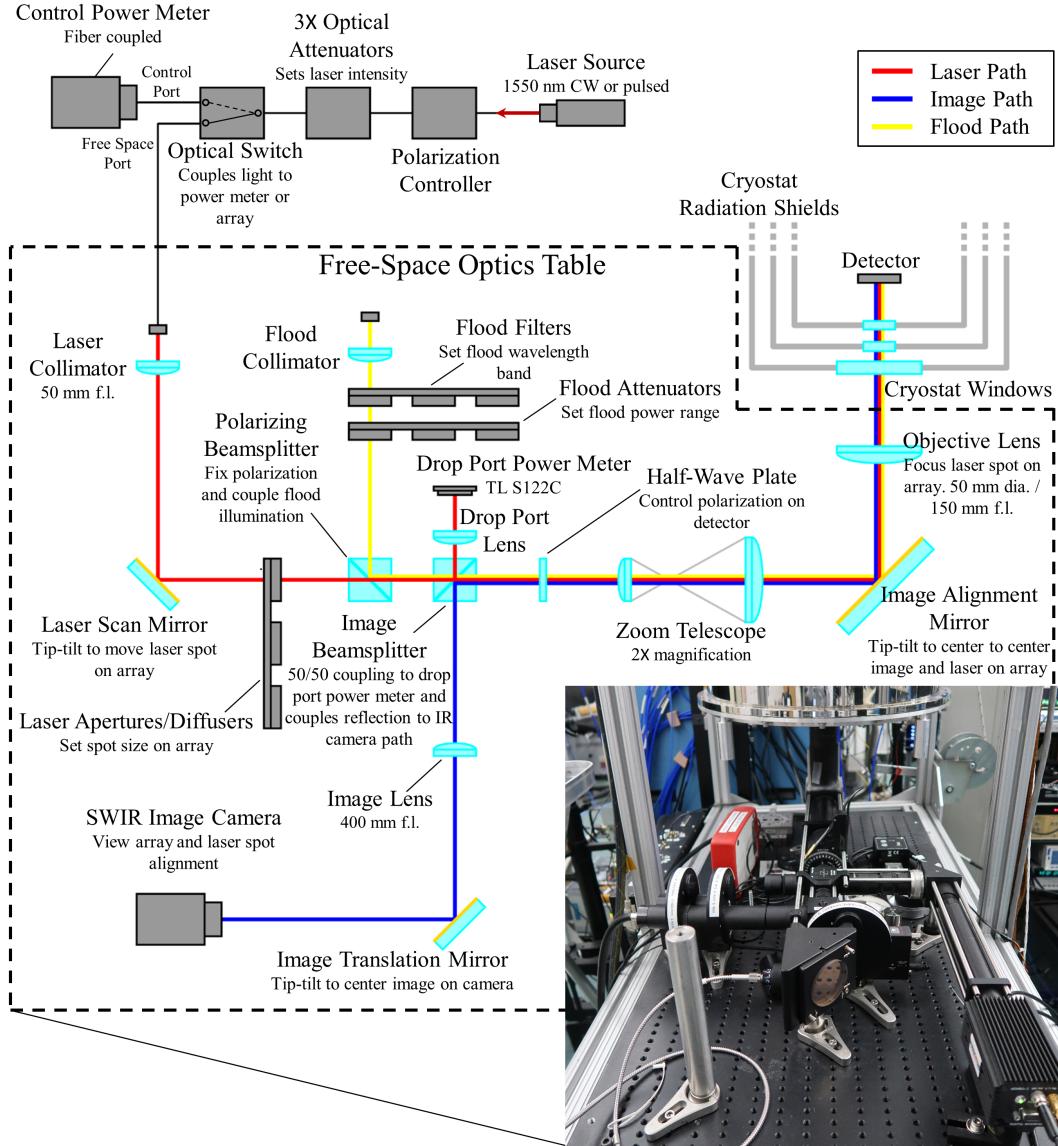
Figure 6.3: Optics system diagram. Black lines indicate optical fiber connections while the colored lines indicate the free-space beampath. When the breadboard is moved from underneath the cryostat, a power meter is mounted at the focal point of the objective lens to calibrate the transmission of the free space optics system (not shown). A picture of the free-space optics is shown for reference.

Efficiency Calibration

Calibrating the efficiency of the free-space optics poses several challenges. We would like to measure the amount of power entering the cryogenic windows in order to characterize the overall efficiency of the cryostat and detector system. This poses a challenge because a power meter cannot be positioned at the focus of the

objective lens when the optics are positioned underneath the cryostat, but changing the position of the optical breadboard shifts the position of the fiber connecting the free-space optics to the source. This shift in fiber position can change the polarization and attenuation of the light entering the free-space system and therefore must be considered when performing calibrated measurements. The following calibration procedure was devised to account for this problem.

The objective lens of the free-space setup was first focused on the array using flood illumination. With the lens locked in place, the optics tray was removed from underneath the cryostat windows and a Thorlabs S122C power meter head was mounted at the focus of the objective lens. With the flood illumination turned off and the flood port blocked, the CW laser was directed on the free-space power meter. The polarization was scanned with the polarization controller to maximize the power measured by the drop port S122C power meter head. With the polarization matched to the polarizing beam splitter, the position of the free-space power meter was fine tuned to maximize the measured power. Once the power was maximized, the power reading was simultaneously recorded and averaged for 60 seconds on the free-space and drop port power meters. The ratio of the free-space power (P_{FS}) to the drop port power (P_{DR}) was recorded as the drop port ratio (R_{DR}). With the drop port ratio measured, the free-space power meter was removed from the system and the optics were re-positioned under the cryostat. The process of sliding the optics under the cryostat could change the position of the optical fiber which would alter the polarization state leaving the fiber and entering the free-space system. Not only was the total polarization altered, but the fiber attenuation could also change. To account for this change, the power reaching the drop port was maximized by tuning the polarization to match the polarizing beam splitter, then the drop port power was averaged for 60 seconds and this value was recorded as the cryostat drop port power ($P_{DR,Cryo}$). The laser was then immediately directed into the control fiber coupled power meter and measured as the control port power (P_C). The ratio of ($P_{DR,Cryo} R_{DR})/P_C$ is the cryostat free-space coupling ratio ($R_{FS,Cryo}$) and represents the fraction of the power measured at the control port that is incident on the cryostat windows.

The laser was attenuated to the single-photon level by setting two of the three attenuators in the optical system to 40 dB. Calibration of the attenuators was performed individually in order to maintain an acceptable signal-to-noise ratio at the control power meter. The transmission of the second attenuator (T_2) was found by setting

the first attenuator to 0 dB then measuring the power at the control port power meter when the second attenuator was set at 0 dB and 40 dB. With this calibration complete and the second attenuator set to 0 dB, the power transmitted through the first attenuator when it was set to 40 dB was recorded (P_1). The second attenuator was set to 40 dB again and the estimated cryostat free-space coupled power ($P_{FS,Cryo}$) was calculated as $P_1\, T_2\, R_{FS,Cryo}$. This coupled power was converted to a photon flux using the Planck-Einstein relation.

Figure 6.4a shows the PCR_P as a function of bias current (I_B) for all 64 pixels of the array at a temperature of 770 mK for $\lambda = 1550$ nm. An efficiency plateau, indicating saturated internal detection efficiency, is present for 62 / 64 pixels. The remaining two pixels are not photosensitive. The PCR_P as a function of BCR_P as shown in Figure 6.4b indicates that the pixels exhibit < 75 kcps BCR_P. These false counts were limited by room temperature blackbody radiation and were later improved to < 100 cps with more aggressive filtering with the cryogenic windows. To measure the efficiency, the array was illuminated using a CW laser coupled to the detector. The laser spot size had an approximately 30 μm FWHM diameter, which is smaller than the size of a single quadrant. The spot was centered on the parallel nanowire region between two adjacent quadrants and the efficiency measured with both TE and TM polarizations. The measured efficiency, as shown in Figure 6.4c, was 75.0 ± 5.0% under TE polarized light and 65.6 ± 3.7% under TM polarized light. The region of bends in the center of the array has a nanowire orientation perpendicular to the majority of the device. This area, combined with the region of bends around the outer perimeter of the device, corresponds to 12% of the active area of the device. Scaling the measured efficiencies by the fraction of parallel and perpendicular nanowires and the number of photosensitive pixels provides an estimate of 71.5% as the realizable efficiency of the array under uniform TE illumination.
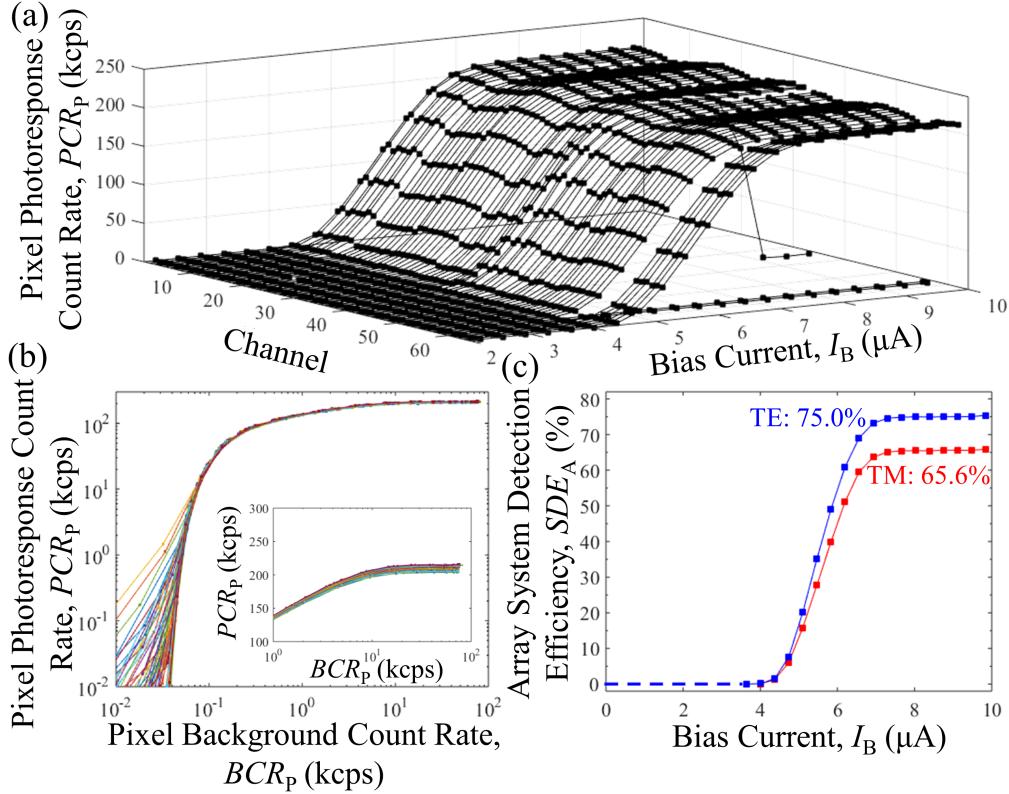
Figure 6.4: (a) PCR_P vs I_B for all pixels. The pixels exhibited saturated internal detection efficiency for 62/ 64 channels. Optical coupling was achieved by flood illuminating the entire array with a filtered white light source as shown in Figure 6.3. The small variation in PCR_P was due to varying nanowire lengths within the active area due to geometric design constraints. (b) PCR_P vs BCR_P for all pixels. The BCR_P was limited by room temperature blackbody radiation and was < 75 kcps per pixel. The inset shows the plateau region of this curve. (c) Array system detection efficiency at given operating currents. For TE polarized light, SDE_A was 75.0 ± 5.0% while SDE_A is 65.6 ± 3.7% for TM polarized light.

The uncertainty of the efficiency measurements can be separated into an absolute uncertainty coming from the power meter calibration and a statistical uncertainty related to the reproducibility of efficiency measurements. The absolute power incident on the detector is referenced to a single Thorlabs S122C power meter at the free space port with a stated error bound of ±5%. When the S122C power meter heads at the free space and drop ports are swapped, the measured R_{DR} changes by 2%. This implies a difference of 1% between the absolute calibrations of the two power meter heads. For our measurements, we reference the average calibration of these two and use the quoted ±5% error bound. There are numerous contributions to the reproducibility uncertainty of these efficiency measurements. Drift in the laser

power, attenuator transmission, power meter readings, and stability of the free space optics all contribute to noise in our efficiency measurements. Isolating the different contributions from the optical components shows that the dominant contribution to reproducibility error comes from the stability of the free space optics. Simultaneous measurement of the power at the free space and drop ports shows that the total free space power can drift up to 2% on the timescale of tens of minutes while R_{DR} remains nearly constant. To combat this drift, $P_{DR,Cryo}$ is measured immediately before taking an array efficiency measurement in order to minimize the effect of this drift. A second effect of importance is the positioning of the spot on the array. When the spot is not optimally positioned, light in the fringes of the spot is not coupled to the active area of the device and efficiency is underestimated. Low frequency vibration of the cold stage due to pulse tube operation also makes reproducibly positioning the optical spot on the detector difficult because the spot moves on the array due to these vibrations. This effect is difficult to isolate from the effects of drift in the free space optics because measuring the array efficiency requires using the optics. Due to these challenges, we find that the best way to estimate the uncertainty associated with the reproducibility of measurements is to take multiple measurements and calculate the statistical spread. Ten efficiency measurements were performed to determine this uncertainty. The average value for TE illumination was 75.0% with a sample standard deviation of 1.1% while the average value for TM illumination was 65.6% with a sample standard deviation of 0.5%. The total error of the measurement is then stated as the quadrature sum of the power meter 5% uncertainty scaled by the device efficiency and three times the sample standard deviation of the ten efficiency measurements. This leads to a TE efficiency of 75.0% ± 5.0% and a TM efficiency of 65.6% ± 3.7%.

Angle of Incidence Dependence

The basic characterization of the array described above was performed with the objective lens outside the cryostat. While this provides information about the efficiency for near normal incidence angles, it does not show how the detector will perform with a fast cryogenic lens. It would be useful for both RCWA model validation and link budget design to have experimental measurements of the array efficiency at different illumination angles. For a cryogenic system, this is quite challenging. With limited space available in the cryostat, mounting the detector at various angles with respect to the window was not practical. An alternate measurement technique was needed to characterize the angular response of the detection efficiency given

the constraints of the cryogenic system.

To solve this problem, I developed an optical technique to measure the detector response at various incidence angles without requiring special mounting of the detector in the fridge. A cryogenic lens was installed in the fridge and focused on the detector. The 0.5 inch diameter cryogenic windows were replaced by 1 inch windows so the full clear aperture of the lens was visible to the outside environment. Then, as shown schematically in Figure 6.5, a narrow collimated beam was translated on the cryogenic lens by moving a fold mirror on x-y micrometers on the optical breadboard outside the cryostat. This had the effect of changing the angle of incidence on the detector without requiring any moving parts inside the cryostat. Implementing this system required careful selection of optical components to avoid unnecessary aberrations and ensure the optical spot would remain centered on the detector as the system was translated. This was made especially difficult because a comparison to the RCWA model would only be possible if the optical spot remained on the parallel nanowires without interacting with the perpendicular segments at the center and edge of the active area (see Figure 6.1a).
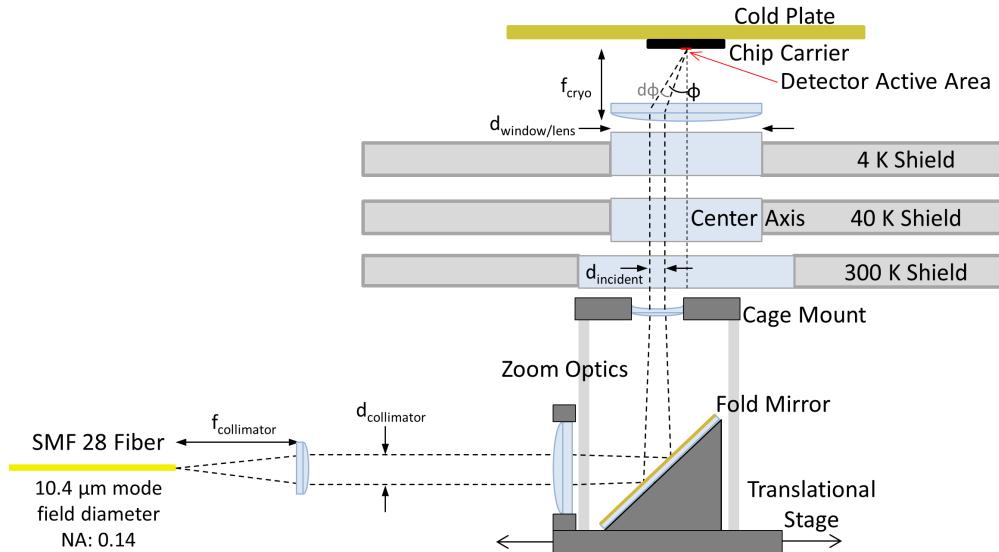
Figure 6.5: Schematic of the angled illumination experimental setup. Collimated light from a fiber was condensed to a small diameter collimated beam using a set of zoom optics mounted on x and y translation stages. This collimated beam entered the cryostat through the windows and was focused on the detector using a cryogenic lens. Translating the zoom optics changes the position where the collimated beam illuminates the cryogenic lens, which changes the angle of incidence at the detector.

Using this setup, relative efficiency measurements were performed on the detector and compared to the predictions of the RCWA model. The results were scaled such that the TE efficiency at normal incidence corresponds to the 75% measured in the calibrated experiment. The results, shown in Figure 6.6, show good qualitative agreement. The results show the change in efficiency as a function of incidence angle (defined with respect to the interface normal) for three azimuthal angles (defined with respect to the nanowire orientation). The experimental measurements show sharp drops in efficiency at particular angle combinations in qualitative agreement with the model, and the polarization trends are also consistent. While qualitatively accurate, there is a quantitative discrepancy between the model and experiment which may be due to misalignment of the optics, leading to error in the estimated position of the beam on the cryogenic lens. This leads to an error in the estimated angle of illumination. Also, changes in the index of refraction of the lens due to the cryogenic temperature may alter the focusing characteristics, leading to further error in the estimated angle of incidence. Due to time constraints and hardware problems with the TDC, only a coarse measurement was performed, but these results were deemed sufficient to validate the accuracy of the RCWA model.
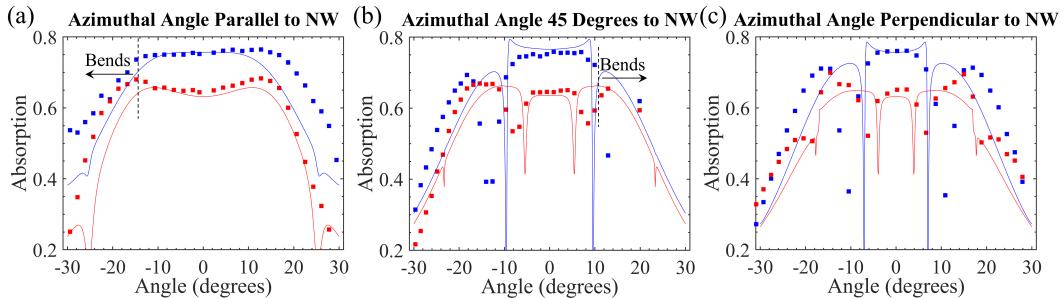
Figure 6.6: Comparison of the measured detector efficiency (squares) and RCWA model (lines) as a function of incidence angle for azimuthal angles (a) parallel, (b) 45 degrees, and (c) perpendicular to the nanowire orientation. The incidence angle is defined as the angle away from normal incidence with the detector. Due to imperfect alignment of the cryogenic lens, the beam can drift as the angle changes, leading to illumination of the bends of the detector. This alters the absorption characteristics. The regions where this occurred are labeled in (a) and (b).

6.2.2 Timing Characterization

The saturation properties of the system were probed using a single quadrant of the detector due to limitations in the counting speed of the TDC. Figure 6.7a shows the normalized quadrant system detection efficiency (SDE_Q) as a function of quadrant photoresponse count rate (PCR_Q) for various bias currents. The quadrant MCR

(MCR_Q) is defined as the PCR_Q where the SDE_Q drops by 3 dB compared to the low photon flux efficiency. The measured MCR_Q was 228 Mcps at a 8.0 µA bias current ($I_B = 0.79I_{sw}$) and 300 Mcps at a 9.5 µA bias current ($I_B = 0.93I_{sw}$). A histogram of the detector inter-arrival time (IAT), shown in Figure 6.7b, indicated there was no afterpulsing at a 9.5 µA bias current. The reset time (t_R) is defined as the time delay after a detection event required before subsequent photons can be measured and is quantified as the time for the detector to reach 50% of the nominal count rate in the IAT histogram. The IAT histogram indicates that this recovery time was approximately 28 ns at a 9.5 µA bias current. Scaling this performance to the full array, even considering the non-functioning pixels, leads to an MCR exceeding 1 Gcps, and therefore meets the DSOC requirements.
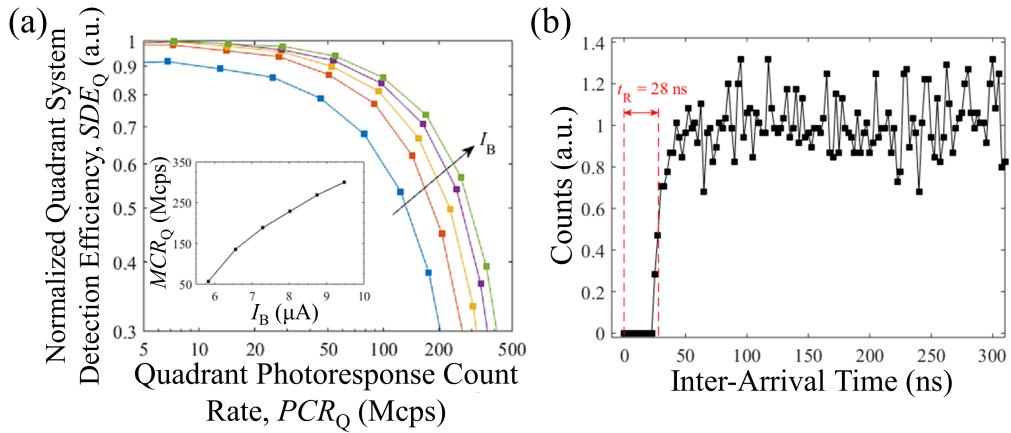
Figure 6.7: (a) MCR_Q curves through cryogenic amplifier chain for I_B ranging from 6.5 µA to 9.5 µA. For a 5.8 µA bias current ($I_B = 0.57$Isw) the MCR_Q was 56 Mcps, for a 7.3 µA bias current ($I_B = 0.71I_{sw}$) the MCR_Q was 188 Mcps, and for a 9.5 µA bias current ($I_B = 0.93I_{sw}$) the MCR_Q was 300 Mcps. (b) IAT histogram of one pixel at 9.5 µA bias current. The IAT measurements are referenced to the trigger time when the rising edge of a detection pulse crossed the 50% of amplitude the average detection waveform. The reset time was measured to be 28 ns.

The timing uncertainty was also characterized by measuring the instrument response function (IRF). Figure 6.8a shows the IRF of a pixel acquired at a 5.8 µA bias current ($I_B = 0.57I_{sw}$) and at 9.5 µA bias current ($I_B = 0.93I_{sw}$). The IRF was measured by calculating a histogram of the time delay between the rising edge of a fiber coupled InGaAs photodetector (Thorlabs DET08CFC) response pulse and the SNSPD detection pulse from an excitation of a femtosecond-pulse laser with a splitter coupling light into both detectors. The resulting histogram of time delays

was fitted with an exponentially modified Gaussian distribution. The system jitter is defined as the FWHM of the IRF fit and was measured to be 121 ps at a 5.8 μA bias current ($I_B = 0.57I_{sw}$) and 74 ps at a 9.5 μA bias current ($I_B = 0.93I_{sw}$). The jitter decreased as I_B increased as shown in Figure 6.8b. This result does not include the additional jitter due to the comparator and TDC, but meets the DSOC requirements with substantial margin for these additional factors.
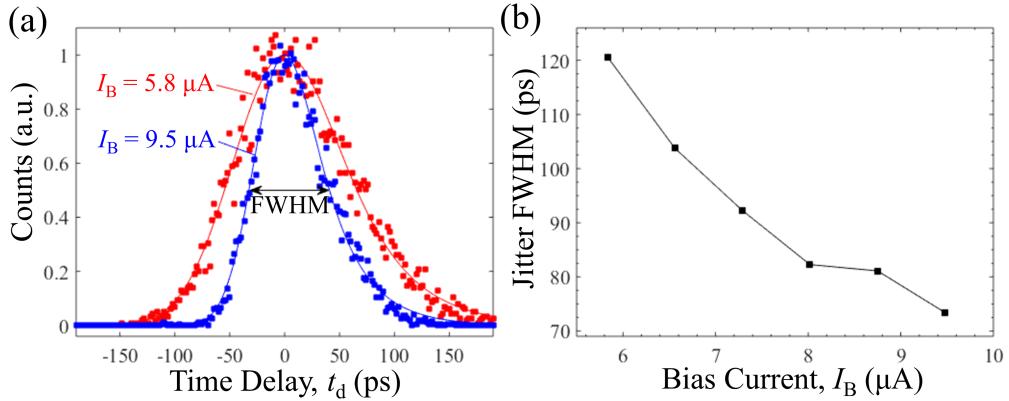
Figure 6.8: (a) IRF of one pixel through the cryogenic amplifier chain. An exponentially modified Gaussian distribution is fitted to the time delay histogram and shown as the solid lines. The jitter of the pixel, as quantified by the FWHM of the exponentially modified Gaussian fit, is 121 ps at 5.8 μA bias current ($I_B = 0.57I_{sw}$) and is 74 ps at 9.5 μA bias current ($I_B = 0.93I_{sw}$). (b) Timing jitter as a function of I_B. The jitter is below 100 ps for bias currents corresponding to saturated internal detection efficiency.

A unique challenge of designing SNSPD arrays with co-wound nanowires is the elimination of crosstalk. Due to its low fill factor, the DSOC prototype showed no evidence of crosstalk as demonstrated in Figure 6.9. The count rate as a function of bias current was measured under illumination when all pixels were biased and when only a single pixel per quadrant was biased. The count rate of the each nanowire did not increase when the bias current of the neighboring wires was increased, indicating that the operation of the neighboring pixel did not affect the count rate of the original nanowire. If crosstalk were present, there would be a difference between the two curves, as shown in the inset for an array with 800 nm pitch (20% fill factor).
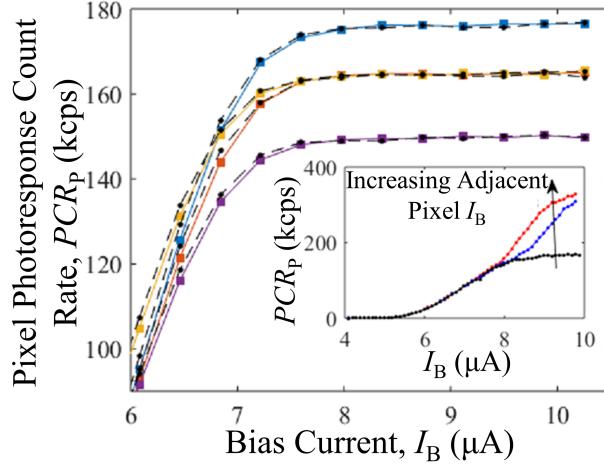
Figure 6.9: Demonstration of the absence of crosstalk in neighboring pixels. The PCR_P vs I_B curves of a single pixel per quadrant were measured when all other nanowires were biased (color squares, solid line) or no other nanowires were biased (black circles, dashed line). As neighboring pixels were biased, there was no change in the PCR_P vs I_B characteristics of the original nanowires indicating there was negligible crosstalk between neighboring pixels. If there were crosstalk between the wires, the count rate of a pixel biased at a fixed current would increase as a neighboring wire was biased in the active regime. This is demonstrated in the inset by a similar device with 800 nm nanowire pitch which did exhibit crosstalk at high biases. In the inset, the black data points show the PCR_P vs I_B curve when the neighboring pixel was not biased while the blue and red curves show the PCR_P vs I_B curve when the adjacent pixel was biased at 8.5 μA and 9.5 μA, respectively. For the current device, the offset in the sets of curves at low biases was due to changes in the true bias voltage when only four nanowires were biased compared to 64. When fewer nanowires were biased, the voltage output by the voltage source was slightly higher, leading to a larger current through the device. This shifted the PCR_P curves toward lower bias currents.

6.3 Summary

The detector described in this chapter marked a significant step in the development of SNSPD arrays for free-space coupled applications where large active areas are required. With a 320 μm diameter active area spread across 64 pixels, an efficiency of 71.5% was expected at a wavelength of 1550 nm under uniform illumination. This was the largest continuous active area array of infrared SNSPDs demonstrated at the time of development in 2017. By extrapolating the counting performance of a single quadrant, the array was expected to have an MCR_A of $\sim$ 1 Gcps, timing jitter $\sim$100 ps, and no measurable crosstalk. This demonstration was an important step in demonstrating the feasibility of the DSOC system. Subsequent arrays which

were fabricated based on this design have demonstrated proper performance in all 64 pixels, and TDC improvements now enable full readout of the system. The unconventional approach of using a low fill factor SNSPD array to simultaneously achieve impressive timing performance and coverage over a large active area has only been possible due to the more advanced AR design guided by precise optical modeling.

Chapter 7

OUTLOOK AND CONCLUSION

Even after nearly twenty years, the SNSPD community continues to make significant strides toward understanding the fundamental physics of these devices and perfecting the design of this remarkable technology. There are several main areas which are expected to be the focus of development in the coming years, all of which would benefit from a more quantitative understanding of SNSPD physics. The first area is the development of SNSPDs with widths on the order of several microns. The possibility of single-photon detection in micron wide wires was first predicted by Vodolazov [6] and has been experimentally verified in the main material systems used for SNSPDs [60][61][62]. The detection process for these devices is qualitatively well understood based on the process of photon assisted vortex-antivortex unbinding [6]. However, as we saw in Chapter 3, the current models can only give a qualitative prediction. Our attempts to improve the rigor of these models has moved predictions much closer to experimental observation based on realistic estimates of material parameters, but these models are not yet at a point where they can be considered even semi-quantitative. We have not made a serious attempt to apply the generalized TDGL model to the case of wide nanowires to fit the available experimental data, but it is expected that the generalized TDGL model, while superior to the standard TDGL model, would fail to provide predictive capabilities as is needed for understanding the fundamental limits of these detectors. Understanding the intrinsic timing jitter which can be achieved with these devices and determining the maximum width of a nanowire which can be used to achieve saturated internal efficiency for a given photon energy are of particular interest. Taking a more optimistic outlook, wide nanowires may provide experimental data in an operating regime previously unexplored and could aid in the development of a quantitative model.

Another topic of growing interest is understanding how SNSPDs can be designed to achieve high efficiency further into the mid-infrared. As thin-film deposition and fabrication techniques have advanced, teams have been able to demonstrate superconducting films with thicknesses approaching 2 nm. This has enabled the recent demonstration of near saturated internal efficiency of 9.9 μm wavelength light in a WSi detector [117]. Both experiment and modeling suggest that lowering T_c is a

viable strategy for improving long-wavelength sensitivity, but as the photon energy decreases, energy fluctuations are expected to play an increasingly important role. The current treatment of fluctuations is superficial, and only considers losses during the first-generation of phonons from the phonon bubble. In reality, these losses continue throughout the evolution of the hotspot, and fluctuations will continue to influence the behavior of the system. However, even the superficial treatment which only considers first-generation phonons highlights serious concern about the use of SNSPDs in the mid-infrared. As the photon energy decreases, very few Debye phonons will be generated from the initial electron-hole pair. This means that the escape of even a small number of phonons could represent the loss of a substantial fraction of the photon energy. This energy loss might place a limit on the long-wavelength limit where devices can achieve saturated internal efficiency.

In Chapters 2 and 3, we highlighted a number of weaknesses of the current model of the SNSPD detection process. Addressing these weaknesses presents an opportunity for future theoretical progress. There are two main features to address. The first is the non-equilibrium nature of the SNSPD detection process, from downconversion through electrothermal evolution. Solving the kinetic equations during downconversion while including the effects of diffusion demonstrated that careful attention must be given to the appropriate way to transition from modeling the kinetics of downconversion to the suppression of superconductivity with the TDGL model. Changing the initial conditions makes a substantial impact on the calculated detection energy of the nanowire, and any attempt to develop a quantitative model of the detection process must pay close attention to this issue. The two-temperature treatment of energy within the TDGL model certainly needs modification to account for these non-equilibrium effects as well, and it does not currently appear that the one-temperature model is a sufficient improvement. The TDGL formulation itself is not rigorously justified, so it would not be surprising if a more sophisticated model based on the Keldysh formalism is required, though it would be surprising and impressive if this technique was implemented at any point in the near future. Non-equilibrium effects may even be important in the late stages of electrothermal evolution in thin superconductors where phonon bottlenecking and dimensionality reduction become important. Given that both wide nanowires and mid-infrared detection rely upon ultrathin nanowires using low T_c materials, understanding these dynamics will be of increasing technological importance.

The second area of focus for theoretic development is that of fluctuations. Timing

jitter and intrinsic dark counts, both important characteristics of detector performance, are dominated by fluctuation effects. While great strides have been made in qualitatively describing both of these processes, experimental results are only fitted by models. Models are not capable of predicting these detector metrics from first principles. Studying the fluctuations beyond the first-generation phonons during downconversion will provide valuable information about the sources of timing jitter. A more ambitious task is to implement a proper treatment of thermal and quantum fluctuations of the superconductor during the detection process itself, but it is not yet clear if these contributions will be significant when compared to the downconversion fluctuations. A better understanding of these fluctuations in the superconductor is needed to be able to predict intrinsic dark count rates, which will be of considerable importance as nanowires detectors are considered for applications requiring low dark counts [15][41].

Finally, the development of large-format SNSPDs will be an area of intense research for the near future. Several advancements in multiplexing techniques have been demonstrated recently [32][29], one of which was discussed in Chapter 5. By employing these techniques and adding some simple brute force, SNPSD arrays have the potential to scale well beyond the kilopixel range toward the megapixel level for high-efficiency photon-counting imaging systems. However, there is a limit to how much multiplexing will be able to scale practical SNSPD arrays. We have approached the point where readout electronics, not focal plane arrays, limit the capability of state of the art systems. Creating functional systems will be increasingly dependent on a multidisciplinary approach which simultaneously addresses the problems of heat load on cryogenic systems, optimized optical coupling, readout electronics, and data processing. Due to these challenges, SNPSD imaging arrays will likely be limited to the several kilopixel scale unless a serious and coordinated effort is made.

I look forward to seeing how the SNSPD community addresses the challenges of practical device design when scaling to larger systems, and I hope that the work presented in this thesis helps guide future theory efforts toward a more quantitative understanding of SNSPDs.

BIBLIOGRAPHY

[1] J. P. Allmaras, E. E. Wollman, A. D. Beyer, R. M. Briggs, B. A. Korzh, B. Bumble, and M. D. Shaw, "Demonstration of a thermally coupled row-column SNSPD imaging array", Nano Letters **20**, 2163–2168 (2020) `10.1021/acs.nanolett.0c00246`.

[2] B. Korzh, Q.-Y. Zhao, J. P. Allmaras, S. Frasca, T. M. Autry, E. A. Bersin, A. D. Beyer, R. M. Briggs, B. Bumble, M. Colangelo, G. M. Crouch, A. E. Dane, T. Gerrits, A. E. Lita, F. Marsili, G. Moody, C. Peña, E. Ramirez, J. D. Rezac, N. Sinclair, M. J. Stevens, A. E. Velasco, V. B. Verma, E. E. Wollman, S. Xie, D. Zhu, P. D. Hale, M. Spiropulu, K. L. Silverman, R. P. Mirin, S. W. Nam, A. G. Kozorezov, M. D. Shaw, and K. K. Berggren, "Demonstration of sub-3 ps temporal resolution with a superconducting nanowire single-photon detector", Nature Photonics **14**, 1–6 (2020) `10.1038/s41566-020-0589-x`.

[3] J. Allmaras, A. Kozorezov, B. Korzh, K. Berggren, and M. Shaw, "Intrinsic timing jitter and latency in superconducting nanowire single-photon detectors", Physical Review Applied **11**, 034062 (2019) `10.1103/PhysRevApplied.11.034062`.

[4] A. N. McCaughan, V. B. Verma, S. M. Buckley, J. P. Allmaras, A. G. Kozorezov, A. N. Tait, S. W. Nam, and J. M. Shainline, "A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors", Nature Electronics **2**, 451–456 (2019) `10.1038/s41928-019-0300-8`.

[5] J. P. Allmaras, A. G. Kozorezov, A. D. Beyer, F. Marsili, R. M. Briggs, and M. D. Shaw, "Thin-film thermal conductivity measurements using superconducting nanowires", Journal of Low Temperature Physics **193**, 380–386 (2018) `10.1007/s10909-018-2022-0`.

[6] D. Y. Vodolazov, "Single-Photon Detection by a Dirty Current-Carrying Superconducting Strip Based on the Kinetic-Equation Approach", Physical Review Applied **7**, 034014 (2017) `10.1103/PhysRevApplied.7.034014`.

[7] G. N. Gol'tsman, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, K. Smirnov, B. Voronov, A. Dzardanov, C. Williams, and R. Sobolewski, "Picosecond superconducting single-photon optical detector", Applied Physics Letters **79**, 705 (2001) `10.1063/1.1388868`.

[8] C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, "Superconducting nanowire single-photon detectors: physics and applications", Superconductor Science and Technology **25**, 063001 (2012) `10.1088/0953-2048/25/6/063001`.

[9] E. A. Dauler, M. E. Grein, A. J. Kerman, F. Marsili, S. Miki, S. W. Nam, M. D. Shaw, H. Terai, V. B. Verma, and T. Yamashita, "Review of superconducting nanowire single-photon detector system design options and demonstrated performance", Optical Engineering **53**, 081907 (2014) `10.1117/1.OE.53.8.081907`.

[10] I. Holzman and Y. Ivry, "Superconducting nanowires for single-photon detection: progress, challenges, and opportunities", Advanced Quantum Technologies **2**, 1800058 (2019) `10.1002/qute.201800058`.

[11] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Detecting single infrared photons with 93% system efficiency", Nature Photonics **7**, 210 (2013) `10.1038/nphoton.2013.13`.

[12] W. Zhang, L. You, H. Li, J. Huang, C. Lv, L. Zhang, X. Liu, J. Wu, Z. Wang, and X. Xie, "NbN superconducting nanowire single photon detector with efficiency over 90% at 1550 nm wavelength operational at compact cryocooler temperature", Science China Physics, Mechanics & Astronomy **60**, 120314 (2017) `10.1007/s11433-017-9113-4`.

[13] D. V. Reddy, R. R. Nerem, A. E. Lita, S. W. Nam, R. P. Mirin, and V. B. Verma, "Exceeding 95% system efficiency within the telecom c-band in superconducting nanowire single photon detectors", in Conference on Lasers and Electro-Optics (2019), FF1A.3, `10.1364/CLEO_QELS.2019.FF1A.3`.

[14] E. E. Wollman, V. B. Verma, A. D. Beyer, R. M. Briggs, B. A. Korzh, J. P. Allmaras, F. Marsili, A. E. Lita, R. P. Mirin, S. W. Nam, and M. D. Shaw, "UV superconducting nanowire single-photon detectors with high efficiency, low noise, and 4 K operating temperature", Optics Express **25**, 26792–26801 (2017) `10.1364/OE.25.026792`.

[15] Y. Hochberg, I. Charaev, S.-W. Nam, V. Verma, M. Colangelo, and K. K. Berggren, "Detecting sub-GeV dark matter with superconducting nanowires", Physical Review Letters **123**, 151802 (2019) `10.1103/PhysRevLett.123.151802`.

[16] D. H. Slichter, V. B. Verma, D. Leibfried, R. P. Mirin, S. W. Nam, and D. J. Wineland, "UV-sensitive superconducting nanowire single photon detectors for integration in an ion trap", Optics Express **25**, 8705–8720 (2017) `10.1364/OE.25.008705`.

[17] F. Marsili, V. B. Verma, M. J. Stevens, J. A. Stern, M. D. Shaw, A. J. Miller, D. Schwarzer, A. Wodtke, R. P. Mirin, and S. W. Nam, "Mid-infrared single-photon detection with tungsten silicide superconducting nanowires", in CLEO: 2013, ISSN: 2160-8989 (June 2013), pp. 1–2, `10.1364/CLEO_SI.2013.CTu1H.1`.

[18] L. Chen, D. Schwarzer, V. B. Verma, M. J. Stevens, F. Marsili, R. P. Mirin, S. W. Nam, and A. M. Wodtke, "Mid-infrared laser-induced fluorescence with nanosecond time resolution using a superconducting nanowire single-photon detector: new technology for molecular science", Accounts of Chemical Research **50**, 1400–1409 (2017) `10.1021/acs.accounts.7b00071`.

[19] B. Baek, A. E. Lita, V. Verma, and S. W. Nam, "Superconducting a-W_xSi_{1-x} nanowire single-photon detector with saturated internal quantum efficiency from visible to 1850 nm", Applied Physics Letters **98**, 251105 (2011) `10.1063/1.3600793`.

[20] V. B. Verma, A. E. Lita, M. R. Vissers, F. Marsili, D. P. Pappas, R. P. Mirin, and S. W. Nam, "Superconducting nanowire single photon detectors fabricated from an amorphous $Mo_{0.75}Ge_{0.25}$ thin film", Applied Physics Letters **105**, 022602 (2014) `10.1063/1.4890277`.

[21] V. B. Verma, B. Korzh, F. Bussières, R. D. Horansky, S. D. Dyer, A. E. Lita, I. Vayshenker, F. Marsili, M. D. Shaw, H. Zbinden, R. P. Mirin, and S. W. Nam, "High-efficiency superconducting nanowire single-photon detectors fabricated from MoSi thin-films", Optics Express **23**, 33792–33801 (2015) `10.1364/OE.23.033792`.

[22] J. P. Sprengers, A. Gaggero, D. Sahin, S. Jahanmirinejad, G. Frucci, F. Mattioli, R. Leoni, J. Beetz, M. Lermer, M. Kamp, S. Höfling, R. Sanjines, and A. Fiore, "Waveguide superconducting single-photon detectors for integrated quantum photonic circuits", Applied Physics Letters **99**, 181110 (2011) `10.1063/1.3657518`.

[23] C. Schuck, W. H. P. Pernice, and H. X. Tang, "Waveguide integrated low noise NbTiN nanowire single-photon detectors with milli-Hz dark count rate", Scientific Reports **3**, 1893 (2013) `10.1038/srep01893`.

[24] O. Kahl, S. Ferrari, V. Kovalyuk, G. N. Goltsman, A. Korneev, and W. H. P. Pernice, "Waveguide integrated superconducting single-photon detectors with high internal quantum efficiency at telecom wavelengths", Scientific Reports **5**, 10941 (2015) `10.1038/srep10941`.

[25] I. E. Zadeh, J. W. N. Los, R. B. M. Gourgues, G. Bulgarini, S. M. Dobrovolskiy, V. Zwiller, and S. N. Dorenbos, "A single-photon detector with high efficiency and sub-10ps time resolution", arXiv:1801.06574 (2018).

[26] M. Colangelo, A. Beyer, B. Korzh, J. Allmaras, A. Mueller, R. Briggs, B. Bumble, M. Runyan, D. Zhu, S. Smith, W. Becker, L. Navaez, J. Bienfang, S. Frasca, E. Ramirez, A. Walter, E. Schmidt, E. Wollman, C. Peña, M. Spiropulu, K. Berggren, and M. Shaw, "Practical high-efficiency low-jitter photon counting using a differential superconducting nanowire single-photon detector", In Preparation (2020).

[27] M. D. Shaw, F. Marsili, A. D. Beyer, R. M. Briggs, J. P. Allmaras, and W. H. Farr, "Superconducting nanowire single photon detectors for deep space optical communication (conference presentation)", Proc. SPIE, 100960J (2017) `10.1117/12.2255836`.

[28] M. S. Allman, V. B. Verma, M. Stevens, T. Gerrits, R. D. Horansky, A. E. Lita, F. Marsili, A. Beyer, M. D. Shaw, D. Kumor, R. Mirin, and S. W. Nam, "A near-infrared 64-pixel superconducting nanowire single photon detector array with integrated multiplexed readout", Applied Physics Letters **106**, 192601 (2015) `10.1063/1.4921318`.

[29] E. E. Wollman, V. B. Verma, A. E. Lita, W. H. Farr, M. D. Shaw, R. P. Mirin, and S. Woo Nam, "Kilopixel array of superconducting nanowire single-photon detectors", Optics Express **27**, 35279 (2019) `10.1364/OE.27.035279`.

[30] S. Miyajima, M. Yabuno, S. Miki, T. Yamashita, and H. Terai, "High-time-resolved 64-channel single-flux quantum-based address encoder integrated with a multi-pixel superconducting nanowire single-photon detector", Optics Express **26**, 29045–29054 (2018) `10.1364/OE.26.029045`.

[31] S. Miyajima, M. Yabuno, S. Miki, and H. Terai, "Single-flux-quantum based event-driven encoder for large-pixel superconducting nanowire single-photon detector array", IEEE Transactions on Applied Superconductivity **29**, 1–4 (2019) `10.1109/TASC.2019.2899845`.

[32] Q.-Y. Zhao, D. Zhu, N. Calandri, A. E. Dane, A. N. McCaughan, F. Bellei, H.-Z. Wang, D. F. Santavicca, and K. K. Berggren, "Single-photon imager based on a superconducting nanowire delay line", Nature Photonics **11**, 247 (2017) `10.1038/nphoton.2017.35`.

[33] D. Zhu, Q.-Y. Zhao, H. Choi, T.-J. Lu, A. E. Dane, D. Englund, and K. K. Berggren, "A scalable multi-photon coincidence detector based on superconducting nanowires", Nature Nanotechnology **13**, 596–601 (2018) `10.1038/s41565-018-0160-9`.

[34] A. K. Sinclair, E. Schroeder, D. Zhu, M. Colangelo, J. Glasby, P. D. Mauskopf, H. Mani, and K. K. Berggren, "Demonstration of microwave multiplexed readout of DC biased superconducting nanowire detectors", IEEE Transactions on Applied Superconductivity **29**, 1–4 (2019) `10.1109/TASC.2019.2899329`.

[35] S. Doerner, A. Kuzmin, S. Wuensch, I. Charaev, F. Boes, T. Zwick, and M. Siegel, "Frequency-multiplexed bias and readout of a 16-pixel superconducting nanowire single-photon detector array", Applied Physics Letters **111**, 032603 (2017) `10.1063/1.4993779`.

[36] H. Takesue, S. W. Nam, Q. Zhang, R. H. Hadfield, T. Honjo, K. Tamaki, and Y. Yamamoto, "Quantum key distribution over a 40-dB channel loss

using superconducting single-photon detectors", Nature Photonics **1**, 343–348 (2007) `10.1038/nphoton.2007.75`.

[37] K. Takemoto, Y. Nambu, T. Miyazawa, Y. Sakuma, T. Yamamoto, S. Yorozu, and Y. Arakawa, "Quantum key distribution over 120 km using ultrahigh purity single-photon source and superconducting single-photon detectors", Scientific Reports **5**, 14383 (2015) `10.1038/srep14383`.

[38] A. Boaron, G. Boso, D. Rusca, C. Vulliez, C. Autebert, M. Caloz, M. Perrenoud, G. Gras, F. Bussières, M.-J. Li, D. Nolan, A. Martin, and H. Zbinden, "Secure quantum key distribution over 421 km of optical fiber", Physical Review Letters **121**, 190502 (2018) `10.1103/PhysRevLett.121.190502`.

[39] D. M. Boroson, B. S. Robinson, D. V. Murphy, D. A. Burianek, F. Khatri, J. M. Kovalik, Z. Sodnik, and D. M. Cornwell, "Overview and results of the Lunar Laser Communication Demonstration", in Free-space laser communication and atmospheric propagation xxvi, Vol. 8971, edited by H. Hemmati and D. M. Boroson (International Society for Optics and Photonics, 2014), pp. 213–223, `10.1117/12.2045508`.

[40] S. Crain, C. Cahall, G. Vrijsen, E. E. Wollman, M. D. Shaw, V. B. Verma, S. W. Nam, and J. Kim, "High-speed low-crosstalk detection of a 171yb+ qubit using superconducting nanowire single photon detectors", Communications Physics **2**, 97 (2019) `10.1038/s42005-019-0195-8`.

[41] The OST mission concept study team, "The origins space telescope (OST) mission concept study interim report", arXiv:1809.09702 (2018).

[42] A. D. Semenov, G. N. Gol'tsman, and A. e. A. Korneev, "Quantum detection by current carrying superconducting film", Physica C Superconductivity **351**, 349–356 (2001) `10.1016/S0921-4534(00)01637-3`.

[43] A. Semenov, A. Engel, H.-W. Hübers, K. Il'in, and M. Siegel, "Spectral cut-off in the efficiency of the resistive state formation caused by absorption of a single-photon in current-carrying superconducting nano-strips", The European Physical Journal B **47**, 495–501 (2005) `10.1140/epjb/e2005-00351-8`.

[44] A. Engel, A. D. Semenov, H.-W. Hübers, K. Il'in, and M. Siegel, "Fluctuation effects in superconducting nanostrips", Physica C: Superconductivity and its Applications **444**, 12–18 (2006) `10.1016/j.physc.2006.05.079`.

[45] L. N. Bulaevskii, M. J. Graf, C. D. Batista, and V. G. Kogan, "Vortex-induced dissipation in narrow current-biased thin-film superconducting strips", Physical Review B **83**, 144526 (2011) `10.1103/PhysRevB.83.144526`.

[46] L. N. Bulaevskii, M. J. Graf, and V. G. Kogan, "Vortex-assisted photon counts and their magnetic field dependence in single-photon superconducting detectors", Physical Review B **85**, 014505 (2012) `10.1103/PhysRevB.85.014505`.

[47] A. N. Zotova and D. Y. Vodolazov, "Photon detection by current-carrying superconducting film: a time-dependent Ginzburg–Landau approach", Physical Review B **85**, 024509 (2012) `10.1103/PhysRevB.85.024509`.

[48] A. Engel and A. Schilling, "Numerical analysis of detection-mechanism models of superconducting nanowire single-photon detector", Journal of Applied Physics **114**, 214501 (2013) `10.1063/1.4836878`.

[49] D. Y. Vodolazov, "Current dependence of the red boundary of superconducting single-photon detectors in the modified hot-spot model", Physical Review B **90**, 054515 (2014) `10.1103/PhysRevB.90.054515`.

[50] A. Engel, J. Lonsky, X. Zhang, and A. Schilling, "Detection Mechanism in SNSPD: Numerical Results of a Conceptually Simple, Yet Powerful Detection Model", IEEE Transactions on Applied Superconductivity **25**, 1 (2015) `10.1109/TASC.2014.2371537`.

[51] D. Y. Vodolazov, Y. P. Korneeva, A. V. Semenov, A. A. Korneev, and G. N. Goltsman, "Vortex-assisted mechanism of photon counting in a superconducting nanowire single-photon detector revealed by external magnetic field", Physical Review B **92**, 104503 (2015) `10.1103/PhysRevB.92.104503`.

[52] A. G. Kozorezov, C. Lambert, F. Marsili, M. J. Stevens, V. B. Verma, J. A. Stern, R. Horansky, S. Dyer, S. Duff, D. P. Pappas, A. Lita, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Quasiparticle recombination in hotspots in superconducting current-carrying nanowires", Physical Review B **92**, 064504 (2015) `10.1103/PhysRevB.92.064504`.

[53] A. G. Kozorezov, C. Lambert, F. Marsili, M. J. Stevens, V. B. Verma, J. P. Allmaras, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Fano fluctuations in superconducting-nanowire single-photon detectors", Physical Review B **96**, 054507 (2017) `10.1103/PhysRevB.96.054507`.

[54] H. Wu, C. Gu, Y. Cheng, and X. Hu, "Vortex-crossing-induced timing jitter of superconducting nanowire single-photon detectors", Applied Physics Letters **111**, 062603 (2017) `10.1063/1.4997930`.

[55] D. Y. Vodolazov, "Minimal timing jitter in superconducting nanowire single-photon detectors", Physical Review Applied **11**, 014016 (2019) `10.1103/PhysRevApplied.11.014016`.

[56] A. J. Kerman, J. K. W. Yang, R. J. Molnar, E. A. Dauler, and K. K. Berggren, "Electrothermal feedback in superconducting nanowire single-photon detectors", Physical Review B **79**, 100509 (2009) `10.1103/PhysRevB.79.100509`.

[57] A. J. Annunziata, O. Quaranta, D. F. Santavicca, A. Casaburi, L. Frunzio, M. Ejrnaes, M. J. Rooks, R. Cristiano, S. Pagano, A. Frydman, and D. E. Prober, "Reset dynamics and latching in niobium superconducting nanowire

single-photon detectors", Journal of Applied Physics **108**, 084507 (2010) `10.1063/1.3498809`.

[58] R. Baghdadi, J. P. Allmaras, B. A. Butters, A. E. Dane, S. Iqbal, A. N. Mc-Caughan, E. A. Toomey, Q.-Y. Zhao, A. G. Kozorezov, and K. K. Berggren, "Heater nano-cryotron: a superconducting nanowire-based thermal switch", In Preparation (2020).

[59] F. Marsili, F. Najafi, E. Dauler, F. Bellei, X. Hu, M. Csete, R. J. Molnar, and K. K. Berggren, "Single-photon detectors based on ultranarrow superconducting nanowires", Nano Letters **11**, 2048–2053 (2011) `10.1021/nl2005 143`.

[60] Y. P. Korneeva, D. Y. Vodolazov, A. V. Semenov, I. N. Florya, N. Simonov, E. Baeva, A. A. Korneev, G. N. Goltsman, and T. M. Klapwijk, "Optical single-photon detection in micrometer-scale NbN bridges", Physical Review Applied **9**, 064037 (2018) `10.1103/PhysRevApplied.9.064037`.

[61] J. Chiles, S. M. Buckley, A. Lita, V. B. Verma, J. Allmaras, B. Korzh, M. D. Shaw, J. M. Shainline, R. P. Mirin, and S. W. Nam, "Superconducting microwire detectors with single-photon sensitivity in the near-infrared", arXiv:2002.12858 (2020).

[62] I. Charaev, Y. Morimoto, A. Dane, A. Agarwal, M. Colangelo, and K. K. Berggren, "Large-area microwire MoSi single-photon detectors at 1550 nm wavelength", arXiv:2002.09088 (2020).

[63] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction", Journal of the Optical Society of America **71**, 811 (1981) `10.1364/JOSA.71.000811`.

[64] M. G. Moharam, T. K. Gaylord, E. B. Grann, and D. A. Pommet, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings", Journal of the Optical Society of America A **12**, 1068 (1995) `10.1364/JOSAA.12.001068`.

[65] M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach", Journal of the Optical Society of America A **12**, 1077–1086 (1995) `10.1364/JOSAA.12.001077`.

[66] A. M. Kadin and M. W. Johnson, "Nonequilibrium photon-induced hotspot: a new mechanism for photodetection in ultrathin metallic films", Applied Physics Letters **69**, 3938–3940 (1996) `10.1063/1.117576`.

[67] W. J. Skocpol, M. R. Beasley, and M. Tinkham, "Self-heating hotspots in superconducting thin-film microbridges", Journal of Applied Physics **45**, 4054–4066 (1974) `10.1063/1.1663912`.

[68] J. K. W. Yang, A. J. Kerman, E. A. Dauler, V. Anant, K. M. Rosfjord, and K. K. Berggren, "Modeling the Electrical and Thermal Response of Superconducting Nanowire Single-Photon Detectors", IEEE Transactions on Applied Superconductivity **17**, 581 (2007) `10.1109/TASC.2007.898660`.

[69] A. J. Kerman, D. Rosenberg, R. J. Molnar, and E. A. Dauler, "Readout of superconducting nanowire single-photon detectors at high count rates", Journal of Applied Physics **113**, 144511 (2013) `10.1063/1.4799397`.

[70] N. Calandri, Q.-Y. Zhao, D. Zhu, A. Dane, and K. K. Berggren, "Superconducting nanowire detector jitter limited by detector geometry", Applied Physics Letters **109**, 152601 (2016) `10.1063/1.4963158`.

[71] D. Zhu, M. Colangelo, B. A. Korzh, Q.-Y. Zhao, S. Frasca, A. E. Dane, A. E. Velasco, A. D. Beyer, J. P. Allmaras, E. Ramirez, W. J. Strickland, D. F. Santavicca, M. D. Shaw, and K. K. Berggren, "Superconducting nanowire single-photon detector with integrated impedance-matching taper", Applied Physics Letters **114**, 042601 (2019) `10.1063/1.5080721`.

[72] A. G. Kozorezov, A. F. Volkov, J. K. Wigmore, A. Peacock, A. Poelaert, and R. den Hartog, "Quasiparticle-phonon downconversion in nonequilibrium superconductors", Physical Review B **61**, 11807 (2000) `10.1103/PhysRevB.61.11807`.

[73] A. G. Kozorezov, J. K. Wigmore, D. Martin, P. Verhoeve, and A. Peacock, "Electron energy down-conversion in thin superconducting films", Physical Review B **75**, 094513 (2007) `10.1103/PhysRevB.75.094513`.

[74] X.-J. Chen, V. V. Struzhkin, Z. Wu, M. Somayazulu, J. Qian, S. Kung, A. N. Christensen, Y. Zhao, R. E. Cohen, H.-k. Mao, and R. J. Hemley, "Hard superconducting nitrides", Proceedings of the National Academy of Sciences **102**, 3198–3201 (2005) `10.1073/pnas.0500174102`.

[75] Y. Zou, X. Qi, C. Zhang, S. Ma, W. Zhang, Y. Li, T. Chen, X. Wang, Z. Chen, D. Welch, P. Zhu, B. Liu, Q. Li, T. Cui, and B. Li, "Discovery of superconductivity in hard hexagonal epsilon-NbN", Scientific Reports **6**, 22330 (2016) `10.1038/srep22330`.

[76] J. R. Clem and V. G. Kogan, "Kinetic impedance and depairing in thin and narrow superconducting films", Physical Review B **86**, 174521 (2012) `10.1103/PhysRevB.86.174521`.

[77] B. L. Altshuler and A. G. Aronov, in *Electron-electron interactions in disordered systems*, edited by A. L. Efros and M. Pollak (Elsevier Science, Amsterdam, 1985).

[78] A. I. Bezuglyi and V. A. Shklovskii, "The kinetics of low-temperature electron-phonon relaxation in a metallic film following instantaneous heating of the electrons", Journal of Experimental and Theoretical Physics **84**, 1149–1163 (1997) `10.1134/1.558253`.

[79] A. I. Bezuglyj and V. A. Shklovskij, "Energy relaxation times in metal films from the response of electrical conductivity to periodic heating", Physical Review B **89**, 214303 (2014) `10.1103/PhysRevB.89.214303`.

[80] A. I. Bezuglyj and V. A. Shklovskij, "Kinetics of electron cooling in metal films at low temperatures and revision of the two-temperature model", Journal of Physics: Condensed Matter **30**, 295001 (2018) `10.1088/1361-648x/aacb72`.

[81] A. G. Kozorezov, private communication, 2020.

[82] S. B. Kaplan, "Acoustic matching of superconducting films to substrates", Journal of Low Temperature Physics **37**, 343 (1979) `10.1007/BF00119193`.

[83] X. Zhang, A. Engel, Q. Wang, A. Schilling, A. Semenov, M. Sidorova, H.-W. Hübers, I. Charaev, K. Ilin, and M. Siegel, "Characteristics of superconducting tungsten silicide W_xSi_{1-x} for single photon detection", Physical Review B **94**, 174509 (2016) `10.1103/PhysRevB.94.174509`.

[84] X. Zhang, A. E. Lita, M. Sidorova, V. B. Verma, Q. Wang, S. W. Nam, A. Semenov, and A. Schilling, "Superconducting fluctuations and characteristic time scales in amorphous WSi", Physical Review B **97**, 174502 (2018) `10.1103/PhysRevB.97.174502`.

[85] Y. Pan and W. M. Guan, "Probing the balance between ductility and strength: transition metal silicides", Physical Chemistry Chemical Physics **19**, Publisher: The Royal Society of Chemistry, 19427–19433 (2017) `10.1039/C7CP03182C`.

[86] E. T. Swartz and R. O. Pohl, "Thermal boundary resistance", Reviews of Modern Physics **61**, 605 (1989) `10.1103/RevModPhys.61.605`.

[87] K. R. Babu and G.-Y. Guo, "Electron-phonon coupling, superconductivity, and nontrivial band topology in NbN polytypes", Physical Review B **99**, 104508 (2019) `10.1103/PhysRevB.99.104508`.

[88] A. Sergeev and V. Mitin, "Electron-phonon interaction in disordered conductors: static and vibrating scattering potentials", Physical Review B **61**, 6041–6047 (2000) `10.1103/PhysRevB.61.6041`.

[89] M. Sidorova, A. Semenov, H.-W. Hübers, K. Ilin, M. Siegel, I. Charaev, M. Moshkova, N. Kaurova, G. N. Goltsman, X. Zhang, and A. Schilling, "Electron energy relaxation in disordered superconducting NbN films", arXiv:1907.05039 (2019).

[90] J. Zhang, W. Słysz, A. Pearlman, A. Verevkin, R. Sobolewski, O. Okunev, G. Chulkova, and G. N. Gol'tsman, "Time delay of resistive-state formation in superconducting stripes excited by single optical photons", Physical Review B **67**, 132508 (2003) `10.1103/PhysRevB.67.132508`.

[91] J. Kitaygorsky, I. Komissarov, A. Jukna, D. Pan, O. Minaeva, N. Kaurova, A. Divochiy, A. Korneev, M. Tarkhov, B. Voronov, I. Milostnaya, G. Gol'tsman, and R. R. Sobolewski, "Dark counts in nanostructured NbN superconducting single-photon detectors and bridges", IEEE Transactions on Applied Superconductivity **17**, Conference Name: IEEE Transactions on Applied Superconductivity, 275–278 (2007) `10.1109/TASC.2007.898109`.

[92] Y. Ota, K. Kobayashi, M. Machida, T. Koyama, and F. Nori, "Full numerical simulations of dynamical response in superconducting single-photon detectors", IEEE Transactions on Applied Superconductivity **23**, Conference Name: IEEE Transactions on Applied Superconductivity, 2201105–2201105 (2013) `10.1109/TASC.2013.2248871`.

[93] A. N. Zotova and D. Y. Vodolazov, "Intrinsic detection efficiency of superconducting nanowire single photon detector in the modified hot spot model", Superconducting Science and Technology **27**, 125001 (2014) `10.1088/09 53-2048/27/12/125001`.

[94] J. J. Renema, R. Gaudio, Q. Wang, Z. Zhou, A. Gaggero, F. Mattioli, R. Leoni, D. Sahin, M. J. A. de Dood, A. Fiore, and M. P. van Exter, "Experimental Test of Theories of the Detection Mechanism in a Nanowire Superconducting Single Photon Detector", Physical Review Letters **112**, 117604 (2014) `10.1 103/PhysRevLett.112.117604`.

[95] J. J. Renema, Q. Wang, R. Gaudio, I. Komen, K. op 't Hoog, D. Sahin, A. Schilling, M. P. van Exter, A. Fiore, A. Engel, and M. J. A. de Dood, "Position-dependent local detection efficiency in a nanowire superconducting single-photon detector", Nano Letters **15**, 4541–4545 (2015) `10.1021 /acs.nanolett.5b01103`.

[96] A. Engel, A. Semenov, H.-W. Hübers, K. Il'in, and M. Siegel, "Fluctuations and dark count rates in superconducting NbN single-photon detectors", Physica Status Solidi C **2**, 1668–1673 (2005) `10.1002/pssc.200460809`.

[97] H. Bartolf, A. Engel, A. Schilling, K. Il'in, M. Siegel, H.-W. Hübers, and A. Semenov, "Current-assisted thermally activated flux liberation in ultrathin nanopatterned NbN superconducting meander structures", Physical Review B **81**, 024502 (2010) `10.1103/PhysRevB.81.024502`.

[98] T. Yamashita, S. Miki, K. Makise, W. Qiu, H. Terai, M. Fujiwara, M. Sasaki, and Z. Wang, "Origin of intrinsic dark count in superconducting nanowire single-photon detectors", Applied Physics Letters **99**, 161105 (2011) `10.1 063/1.3652908`.

[99] J. R. Clem and K. K. Berggren, "Geometry-dependent critical currents in superconducting nanocircuits", Physical Review B **84**, 174510 (2011) `10.1 103/PhysRevB.84.174510`.

[100] M. K. Akhlaghi, H. Atikian, A. Eftekharian, M. Loncar, and A. H. Majedi, "Reduced dark counts in optimized geometries for superconducting nanowire single photon detectors", Optics Express **20**, 23610 (2012) `10.1364/OE.20.023610`.

[101] M. Caloz, B. Korzh, N. Timoney, M. Weiss, S. Gariglio, R. J. Warburton, C. Schönenberger, J. Renema, H. Zbinden, and F. Bussières, "Optically probing the detection mechanism in a molybdenum silicide superconducting nanowire single-photon detector", Applied Physics Letters **110**, 083106 (2017) `10.1063/1.4977034`.

[102] Q. Zhao, L. Zhang, T. Jia, L. Kang, W. Xu, J. Chen, and P. Wu, "Intrinsic timing jitter of superconducting nanowire single-photon detectors", Applied Physics B **104**, 673 (2011) `10.1007/s00340-011-4574-4`.

[103] J. Wu, L. You, S. Chen, H. Li, Y. He, C. Lv, Z. Wang, and X. Xie, "Improving the timing jitter of a superconducting nanowire single-photon detection system", Applied Optics **56**, 2195 (2017) `10.1364/AO.56.002195`.

[104] J. A. O'Connor, M. G. Tanner, C. M. Natarajan, G. S. Buller, R. J. Warburton, S. Miki, Z. Wang, S. W. Nam, and R. H. Hadfield, "Spatial dependence of output pulse delay in a niobium nitride nanowire superconducting single-photon detector", Applied Physics Letters **98**, 201116 (2011) `10.1063/1.3581054`.

[105] Y. Cheng, C. Gu, and X. Hu, "Inhomogeneity-induced timing jitter of superconducting nanowire single-photon detectors", Applied Physics Letters **111**, 062604 (2017) `10.1063/1.4985226`.

[106] M. Sidorova, A. Semenov, H.-W. Hübers, I. Charaev, A. Kuzmin, S. Doerner, and M. Siegel, "Physical mechanisms of timing jitter in photon detection by current-carrying superconducting nanowires", Physical Review B **96**, 184504 (2017) `10.1103/PhysRevB.96.184504`.

[107] M. Sidorova, A. Semenov, A. Kuzmin, I. Charaev, S. Doerner, and M. Siegel, "Intrinsic Jitter in Photon Detection by Straight Superconducting Nanowires", IEEE Transactions on Applied Superconductivity **28**, 1–4 (2018) `10.1109/TASC.2018.2836989`.

[108] M. Sidorova, A. Semenov, H.-W. Hübers, A. Kuzmin, S. Doerner, K. Ilin, M. Siegel, I. Charaev, and D. Vodolazov, "Timing jitter in photon detection by straight superconducting nanowires: effect of magnetic field and photon flux", Physical Review B **98**, 134504 (2018) `10.1103/PhysRevB.98.134504`.

[109] S. Jahani, L.-P. Yang, A. B. Tepole, J. C. Bardin, H. X. Tang, and Z. Jacob, "Probabilistic vortex crossing criterion for superconducting nanowire single-photon detectors", Journal of Applied Physics **127**, 143101 (2020) `10.1063/1.5132961`.

[110] A. J. Kerman, E. A. Dauler, W. E. Keicher, J. K. W. Yang, K. K. Berggren, G. Gol'tsman, and B. Voronov, "Kinetic-inductance-limited reset time of superconducting nanowire photon counters", Applied Physics Letters **88**, 111116 (2006) `10.1063/1.2183810`.

[111] M. Caloz, B. Korzh, E. Ramirez, C. Schönenberger, R. J. Warburton, H. Zbinden, M. D. Shaw, and F. Bussières, "Intrinsically-limited timing jitter in molybdenum silicide superconducting nanowire single-photon detectors", Journal of Applied Physics **126**, 164501 (2019) `10.1063/1.5113748`.

[112] J. Bardeen, G. Rickayzen, and L. Tewordt, "Theory of the thermal conductivity of superconductors", Physical Review **113**, 982–994 (1959) `10.1103/PhysRev.113.982`.

[113] Y. Zou, X. Wang, T. Chen, X. Li, X. Qi, D. Welch, P. Zhu, B. Liu, T. Cui, and B. Li, "Hexagonal-structured ε-NbN: ultra-incompressibility, high shear rigidity, and a possible hard superconducting material", Scientific Reports **5**, 10811 (2015) `10.1038/srep10811`.

[114] S. Frasca, B. Korzh, M. Colangelo, D. Zhu, A. E. Lita, J. P. Allmaras, E. E. Wollman, V. B. Verma, A. E. Dane, E. Ramirez, A. D. Beyer, S. W. Nam, A. G. Kozorezov, M. D. Shaw, and K. K. Berggren, "Determining the depairing current in superconducting nanowire single-photon detectors", Physical Review B **100**, 054520 (2019) `10.1103/PhysRevB.100.054520`.

[115] R. J. Watts-Tobin, Y. Krähenbühl, and L. Kramer, "Nonequilibrium theory of dirty, current-carrying superconductors: phase-slip oscillators in narrow filaments near T_c", Journal of Low Temperature Physics **42**, 459 (1981) `10.1007/BF00117427`.

[116] N. B. Kopnin, *Theory of Nonequilibrium Superconductivity* (Oxford University Press, Oxford, 2001).

[117] V. B. Verma, In Preparation (2020).

[118] J. S. Langer and V. Ambegaokar, "Intrinsic resistive transition in narrow superconducting channels", Physical Review **164**, 498–510 (1967) `10.1103/PhysRev.164.498`.

[119] D. E. McCumber and B. I. Halperin, "Time scale of intrinsic resistive fluctuations in thin superconducting wires", Physical Review B **1**, 1054–1070 (1970) `10.1103/PhysRevB.1.1054`.

[120] W. J. Skocpol, M. R. Beasley, and M. Tinkham, "Phase-slip centers and nonequilibrium processes in superconducting tin microbridges", Journal of Low Temperature Physics **16**, 145–167 (1974) `10.1007/BF00655865`.

[121] C. N. Lau, N. Markovic, M. Bockrath, A. Bezryadin, and M. Tinkham, "Quantum phase slips in superconducting nanowires", Physical Review Letters **87**, 217003 (2001) `10.1103/PhysRevLett.87.217003`.

[122] M. Sahu, M.-H. Bae, A. Rogachev, D. Pekker, T.-C. Wei, N. Shah, P. M. Goldbart, and A. Bezryadin, "Individual topological tunnelling events of a quantum field probed through their macroscopic consequences", Nature Physics **5**, 503–508 (2009) `10.1038/nphys1276`.

[123] D. Pekker, N. Shah, M. Sahu, A. Bezryadin, and P. M. Goldbart, "Stochastic dynamics of phase-slip trains and superconductive-resistive switching in current-biased nanowires", Physical Review B **80**, 214525 (2009) `10.1103/PhysRevB.80.214525`.

[124] F. Marsili, F. Najafi, E. Dauler, R. J. Molnar, and K. K. Berggren, "Afterpulsing and instability in superconducting nanowire avalanche photodetectors", Applied Physics Letters **100**, 112601 (2012) `10.1063/1.3691944`.

[125] K. L. Nicolich, C. Cahall, N. T. Islam, G. P. Lafyatis, J. Kim, A. J. Miller, and D. J. Gauthier, "Universal model for the turn-on dynamics of superconducting nanowire single-photon detectors", Physical Review Applied **12**, 034020 (2019) `10.1103/PhysRevApplied.12.034020`.

[126] R. F. Broom and E. H. Rhoderick, "Thermal propagation of a normal region in a thin superconducting film and its application to a new type of bistable element", British Journal of Applied Physics **11**, 292–296 (1960) `10.1088/0508-3443/11/7/310`.

[127] M. I. Kaganov, I. M. Lifshitz, and L. V. Tanatarov, "Relaxation between electrons and the crystalline lattice", Zh. Eksp. Theor. Fiz. **31** (1956).

[128] A. N. McCaughan, "Superconducting thin film nanoelectronics", PhD thesis (Massachusetts Institute of Technology, Sept. 2015).

[129] R. C. Zeller and R. O. Pohl, "Thermal conductivity and specific heat of noncrystalline solids", Physical Review B **4**, 2029–2041 (1971) `10.1103/PhysRevB.4.2029`.

[130] R. B. Stephens, "Low-temperature specific heat and thermal conductivity of noncrystalline dielectric solids", Physical Review B **8**, 2896–2905 (1973) `10.1103/PhysRevB.8.2896`.

[131] P. W. Anderson, B. I. Halperin, and C. M. Varma, "Anomalous low-temperature thermal properties of glasses and spin glasses", The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics **25**, 1–9 (1972) `10.1080/14786437208229210`.

[132] W. A. Phillips, "Tunneling states in amorphous solids", Journal of Low Temperature Physics **7**, 351–360 (1972) `10.1007/BF00660072`.

[133] M. V. Sidorova, A. G. Kozorezov, A. V. Semenov, Y. P. Korneeva, M. Y. Mikhailov, A. Y. Devizenko, A. A. Korneev, G. M. Chulkova, and G. N. Goltsman, "Nonbolometric bottleneck in electron-phonon relaxation in ultrathin WSi films", Physical Review B **97**, 184512 (2018) `10.1103/PhysRevB.97.184512`.

[134] M. Y. Kupriyanov and V. F. Lukichev, "Temperature dependence of pair-breaking current in superconductors", Soviet Journal of Low Temperature Physics **6**, 210 (1980).

[135] D. G. Cahill, H. E. Fischer, T. Klitsner, E. T. Swartz, and R. O. Pohl, "Thermal conductivity of thin films: measurements and understanding", Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films **7**, 1259–1266 (1989) `10.1116/1.576265`.

[136] H. Zhang, C.-X. Liu, S. Gazibegovic, D. Xu, J. A. Logan, G. Wang, N. van Loo, J. D. S. Bommer, M. W. A. de Moor, D. Car, R. L. M. Op het Veld, P. J. van Veldhoven, S. Koelling, M. A. Verheijen, M. Pendharkar, D. J. Pennachio, B. Shojaei, J. S. Lee, C. J. Palmstrøm, E. P. A. M. Bakkers, S. D. Sarma, and L. P. Kouwenhoven, "Quantized Majorana conductance", Nature **556**, 74–79 (2018) `10.1038/nature26142`.

[137] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Fréchette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato, A. Y. Smirnov, L. J. Swenson, M. H. Volkmann, J. Whittaker, J. Yao, E. Ladizinsky, M. W. Johnson, J. Hilton, and M. H. Amin, "Observation of topological phenomena in a programmable lattice of 1,800 qubits", Nature **560**, 456–460 (2018) `10.1038/s41586-018-0410-x`.

[138] J. M. Shainline, S. M. Buckley, R. P. Mirin, and S. W. Nam, "Superconducting optoelectronic circuits for neuromorphic computing", Physical Review Applied **7**, 034013 (2017) `10.1103/PhysRevApplied.7.034013`.

[139] A. N. McCaughan and K. K. Berggren, "A superconducting-nanowire three-terminal electrothermal device", Nano Letters **14**, 5748–5753 (2014) `10.1021/nl502629x`.

[140] Q.-Y. Zhao, A. N. McCaughan, A. E. Dane, K. K. Berggren, and T. Ortlepp, "A nanocryotron comparator can connect single-flux-quantum circuits to conventional electronics", Superconductor Science and Technology **30**, 044002 (2017) `10.1088/1361-6668/aa5f33`.

[141] K. Sano, M. Suzuki, K. Maruyama, S. Taniguchi, M. Tanaka, A. Fujimaki, M. Inoue, and N. Yoshikawa, "Thermally assisted superconductor transistors for Josephson-CMOS hybrid memories", IEICE Transactions on Electronics **101**, 370–377 (2018) `10.1587/transele.E101.C.370`.

[142] Q.-Y. Zhao, E. A. Toomey, B. A. Butters, A. N. McCaughan, A. E. Dane, S.-W. Nam, and K. K. Berggren, "A compact superconducting nanowire memory element operated by nanowire cryotrons", Superconductor Science and Technology **31**, 035009 (2018) `10.1088/1361-6668/aaa820`.

[143] R. Corruccini and J. Gniewek, "Specific heats and enthalpies of technical solids at low temperatures. a compilation from the literature", (1960).

[144] A. J. Manninen, J. K. Suoknuuti, M. M. Leivo, and J. P. Pekola, "Cooling of a superconductor by quasiparticle tunneling", Applied Physics Letters **74**, 3020–3022 (1999) `10.1063/1.124051`.

[145] T. L. Smith, P. J. Anthony, and A. C. Anderson, "Effect of neutron irradiation on the density of low-energy excitations in vitreous silica", Physical Review B **17**, 4997–5008 (1978) `10.1103/PhysRevB.17.4997`.

[146] N. Bilir and W. A. Phillips, "Phonons in SiO_2: the low-temperature heat capacity of cristobalite", The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics **32**, 113–122 (1975) `10.1080/14786437508222809`.

[147] F. Marsili, M. J. Stevens, A. Kozorezov, V. B. Verma, C. Lambert, J. A. Stern, R. D. Horansky, S. Dyer, S. Duff, D. P. Pappas, A. E. Lita, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Hotspot relaxation dynamics in a current-carrying superconductor", Physical Review B **93**, 094518 (2016) `10.1103/PhysRevB.93.094518`.

[148] A. McCarthy, N. J. Krichel, N. R. Gemmell, X. Ren, M. G. Tanner, S. N. Dorenbos, V. Zwiller, R. H. Hadfield, and G. S. Buller, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection", Opt. Express **21**, 8904–8915 (2013) `10.1364/OE.21.008904`.

[149] J. Van Rantwijk, M. Grim, D. van Loon, S. Yates, A. Baryshev, and J. Baselmans, "Multiplexed readout for 1000-pixel arrays of microwave kinetic inductance detectors", IEEE Transactions on Microwave Theory and Techniques **64**, 1876–1883 (2016) `10.1109/TMTT.2016.2544303`.

[150] S. R. Meeker, B. A. Mazin, A. B. Walter, P. Strader, N. Fruitwala, C. Bockstiegel, P. Szypryt, G. Ulbricht, G. Coiffard, B. Bumble, G. Cancelo, T. Zmuda, K. Treptow, N. Wilcer, G. Collura, R. Dodkins, I. Lipartito, N. Zobrist, M. Bottom, J. C. Shelton, D. Mawet, J. C. van Eyken, G. Vasisht, and E. Serabyn, "DARKNESS: a microwave kinetic inductance detector integral field spectrograph for high-contrast astronomy", Publications of the Astronomical Society of the Pacific **130**, 065001 (2018) `10.1088/1538-3873/aab5e7`.

[151] P. Szypryt, S. R. Meeker, G. Coiffard, N. Fruitwala, B. Bumble, G. Ulbricht, A. B. Walter, M. Daal, C. Bockstiegel, G. Collura, N. Zobrist, I. Lipartito, and B. A. Mazin, "Large-format platinum silicide microwave kinetic inductance detectors for optical to near-IR astronomy", Optics Express **25**, 25894–25909 (2017) `10.1364/OE.25.025894`.

[152] V. B. Verma, A. E. Lita, M. J. Stevens, R. P. Mirin, and S. W. Nam, "Athermal avalanche in bilayer superconducting nanowire single-photon detectors", Applied Physics Letters **108**, 131108 (2016) `10.1063/1.4945595`.

[153] V. B. Verma, F. Marsili, S. Harrington, A. E. Lita, R. P. Mirin, and S. W. Nam, "A three-dimensional, polarization-insensitive superconducting nanowire avalanche photodetector", Applied Physics Letters **101**, 251114 (2012) 1 `0.1063/1.4768788`.

[154] A. J. Kerman, E. A. Dauler, J. K. W. Yang, K. M. Rosfjord, V. Anant, K. K. Berggren, G. N. Gol'tsman, and B. M. Voronov, "Constriction-limited detection efficiency of superconducting nanowire single-photon detectors", Applied Physics Letters **90**, 101110 (2007) `10.1063/1.2696926`.

[155] E. A. Dauler, A. J. Kerman, B. S. Robinson, J. K. Yang, B. Voronov, G. Goltsman, S. A. Hamilton, and K. K. Berggren, "Photon-number-resolution with sub-30-ps timing using multi-element superconducting nanowire single photon detectors", Journal of Modern Optics **56**, 364–373 (2009) `10.1080 /09500340802411989`.

[156] T. Yamashita, S. Miki, H. Terai, K. Makise, and Z. Wang, "Crosstalk-free operation of multielement superconducting nanowire single-photon detector array integrated with single-flux-quantum circuit in a 0.1 W Gifford–McMahon cryocooler", Optics Letters **37**, 2982 (2012) `10.1364/OL.37.002982`.

[157] A. E. Dane, A. N. McCaughan, D. Zhu, Q. Zhao, C.-S. Kim, N. Calandri, A. Agarwal, F. Bellei, and K. K. Berggren, "Bias sputtered NbN and superconducting nanowire devices", Applied Physics Letters **111**, 122601 (2017) `10.1063/1.4990066`.

Appendix A

ENERGY DOWNCONVERSION CASCADE

A.1   Electron-Electron Collision Integral

Beginning with the simplified electron-electron collision integral (2.4)

$$
I_{e-e}(n) = -\frac{\alpha_{e-e}}{\tau_0} \int_0^\infty \int_0^\infty d\epsilon_1 d\epsilon_2
$$

$$
a\Bigg[\frac{1}{|\epsilon - \epsilon_1|}\Big[n_\epsilon(1 - n_{\epsilon_1})(1 - n_{\epsilon_2})(1 - n_{\epsilon-\epsilon_1-\epsilon_2})
$$

$$
- (1 - n_\epsilon)n_{\epsilon_1}n_{\epsilon_2}n_{\epsilon-\epsilon_1-\epsilon_2}\Big]\Theta(\epsilon - \epsilon_1 - \epsilon_2)
$$

$$
\left(\frac{1}{|\epsilon + \epsilon_1|} + \frac{2}{|\epsilon - \epsilon_2|} \right)\Big[n_\epsilon n_{\epsilon_1}(1 - n_{\epsilon_2})(1 - n_{\epsilon+\epsilon_1-\epsilon_2}) \tag{A.1}
$$

$$
- (1 - n_\epsilon)(1 - n_{\epsilon_1})n_{\epsilon_2}n_{\epsilon+\epsilon_1-\epsilon_2}\Big]\Theta(\epsilon + \epsilon_1 - \epsilon_2)
$$

$$
\left(\frac{1}{|\epsilon - \epsilon_1|} + \frac{2}{|\epsilon + \epsilon_2|} \right)\Big[n_\epsilon(1 - n_{\epsilon_1})n_{\epsilon_2}n_{-\epsilon+\epsilon_1-\epsilon_2}
$$

$$
- (1 - n_\epsilon)n_{\epsilon_1}(1 - n_{\epsilon_2})(1 - n_{-\epsilon+\epsilon_1-\epsilon_2})\Big]\Theta(-\epsilon + \epsilon_1 - \epsilon_2)\Bigg],
$$

we observe that potential singularities occur as $\epsilon_1 \to \epsilon$ and $\epsilon_2 \to \epsilon$. We assume that $n(\epsilon)$ is a smoothly varying function of ϵ with a continuous derivative everywhere. In the region of $\epsilon_1 \to \epsilon$, $\delta\epsilon = \epsilon_1 - \epsilon$ and $n_{\epsilon_1} = n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon}\delta\epsilon$. Focusing on the two terms with the potential singularity, we arrive at a term proportional to

$$
\int_{\epsilon-\delta\epsilon}^{\epsilon+\delta\epsilon} d\epsilon \int_0^\infty d\epsilon_2 \frac{1}{|\delta\epsilon|}\Bigg[\Big[n_\epsilon\Big(1 - \Big(n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon}\delta\epsilon\Big)\Big)\Big)(1 - n_{\epsilon_2})(1 - n_{-\delta\epsilon-\epsilon_2})
$$

$$
- (1 - n_\epsilon)\Big(n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon}\delta\epsilon\Big)n_{\epsilon_2}n_{-\delta\epsilon-\epsilon_2}\Big]\Theta(-\delta\epsilon - \epsilon_2)
$$

$$
\Big[n_\epsilon(1 - n_{\epsilon_1})n_{\epsilon_2}n_{\delta\epsilon-\epsilon_2} \tag{A.2}
$$

$$
- (1 - n_\epsilon)\Big(n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon}\delta\epsilon\Big)(1 - n_{\epsilon_2})(1 - n_{\delta\epsilon-\epsilon_2})\Big]\Theta(\delta\epsilon - \epsilon_2)\Bigg].
$$

Quick inspection shows that the two terms will only contribute to the integral in the limit of $\epsilon_2 \to 0$. In this limit, $n_{\epsilon_2} \to \frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2$. Given the restrictions of integration, it is natural to break the integral into positive and negative components

of $\delta\epsilon$. This takes the form

$$\int_{-\delta\epsilon}^{0} d(\delta\epsilon') \int_{0}^{|\delta\epsilon'|} d\epsilon_2 \frac{1}{|\delta\epsilon'|}$$

$$\left[\left(n_\epsilon - \left(n_\epsilon^2 + n_\epsilon \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \right) (1 - n_{\epsilon_2}) (1 - n_{-\delta\epsilon'-\epsilon_2}) \right.$$

$$\left. - (1 - n_\epsilon) \left(n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) n_{\epsilon_2} n_{-\delta\epsilon'-\epsilon_2} \right] \tag{A.3}$$

$$+ \int_{0}^{\delta\epsilon} d(\delta\epsilon') \int_{0}^{|\delta\epsilon'|} d\epsilon_2 \frac{1}{|\delta\epsilon'|}$$

$$\left[\left(n_\epsilon - \left(n_\epsilon^2 + n_\epsilon \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \right) n_{\epsilon_2} n_{\delta\epsilon'-\epsilon_2} \right.$$

$$\left. - (1 - n_\epsilon) \left(n_\epsilon + \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) (1 - n_{\epsilon_2})(1 - n_{\delta\epsilon'-\epsilon_2}) \right].$$

in the limit as $\delta\epsilon \to 0$. We can recognize a certain symmetry between these two equations. Upon substitution $n_{\epsilon_2} \to \frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2$ and using the relations $n_{-\delta\epsilon'-\epsilon_2} = \frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2)$ and $n_{\delta\epsilon'-\epsilon_2} = \frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2)$ in the first and second integrals, respectively, the equations become

$$\int_{-\delta\epsilon}^{0} d(\delta\epsilon') \int_{0}^{|\delta\epsilon'|} d\epsilon_2 \frac{1}{|\delta\epsilon'|}$$

$$\left[\left(n_\epsilon - n_\epsilon^2 - n_\epsilon \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \left(\frac{1}{2} - \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2 \right) \left(\frac{1}{2} - \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2) \right) \right.$$

$$\left. - \left(n_\epsilon - n_\epsilon^2 + (1 - n_\epsilon) \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \left(\frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2 \right) \left(\frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2) \right) \right]$$

$$+ \int_{0}^{\delta\epsilon} d(\delta\epsilon') \int_{0}^{|\delta\epsilon'|} d\epsilon_2 \frac{1}{|\delta\epsilon'|}$$

$$\left[\left(n_\epsilon - \left(n_\epsilon^2 + n_\epsilon \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \right) \left(\frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2 \right) \left(\frac{1}{2} + \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2) \right) \right.$$

$$\left. - \left(n_\epsilon - n_\epsilon^2 + (1 - n_\epsilon) \frac{\partial n(\epsilon)}{\partial\epsilon} \delta\epsilon' \right) \left(\frac{1}{2} - \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}\delta\epsilon_2 \right) \left(\frac{1}{2} - \frac{\partial n}{\partial\epsilon}|_{\epsilon=0}(|\delta\epsilon'| - \delta\epsilon_2) \right) \right]. \tag{A.4}$$

By observing that the terms of $O(1)$ in the square brackets cancel for each of the two integrals, and using the inequality $\delta\epsilon_2 \leq |\delta\epsilon'|$, we have demonstrated that there is no singularity in the integral of the electron-electron collision integral as $\epsilon \to \epsilon_1$.

We perform a similar exercise for $\epsilon_2 \to \epsilon$. In the region of $\epsilon_2 \to \epsilon$, $\delta\epsilon = \epsilon_2 - \epsilon$, and $n_{\epsilon_2} = n_\epsilon + \frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon$. Substituting into the term of the integral proportional to $\frac{1}{|\epsilon-\epsilon_2|}$, and considering only the local area around $\epsilon_2 \to \epsilon$ we arrive at

$$
\int_{-\delta\epsilon}^{\delta\epsilon} d(\delta\epsilon') \int_0^\infty d\epsilon_1 \left(\frac{2}{|\delta\epsilon'|}\right) \left[n_\epsilon n_{\epsilon_1} \left(1 - \left(n_\epsilon + \frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon'\right)\right)(1 - n_{\epsilon_1 - \delta\epsilon'}) \right.
$$
$$
\left. -(1 - n_\epsilon)(1 - n_{\epsilon_1})\left(n_\epsilon + \frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon'\right) n_{\epsilon_1 - \delta\epsilon'} \right]\Theta(\epsilon_1 - \delta\epsilon').
$$

(A.5)

For small $\delta\epsilon'$, we note that $n_{\epsilon_1 - \delta\epsilon'} = n_{\epsilon_1} - \frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon'$. Substituting and rearranging the terms leads to the simplified form

$$
\int_{-\delta\epsilon}^{\delta\epsilon} d(\delta\epsilon') \int_0^\infty d\epsilon_1 \left(\frac{2}{|\delta\epsilon'|}\right) \left[(n_\epsilon - n_\epsilon^2)\frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon' - (n_{\epsilon_1} - n_{\epsilon_1}^2)\frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon' \right.
$$
$$
\left. +(1 - n_\epsilon - n_{\epsilon_1} + n_\epsilon n_{\epsilon_1})\left(\frac{\partial n(\epsilon)}{\partial \epsilon}\delta\epsilon'\right)^2 \right]\Theta(\epsilon_1 - \delta\epsilon').
$$

(A.6)

All of the terms within the square brackets of $O(1)$, have canceled, so the integral does not exhibit a singularity. This demonstrates that the electron-electron collision integral does not contain a singularity, and this permits us to use it for numerical calculations.

Appendix B

DETECTION MECHANISM

B.1 Derivation of Modified TDGL Final Term

The final term in Vodolazov's modified TDGL equation originates from enforcing the conservation of the approximate Usadel supercurrent in the stationary state rather than the conservation of the Ginzburg–Landau supercurrent. In the modified TDGL formulation before addition of the final term, the complex order parameter Δ can be expressed in terms of its magnitude $|\Delta|$ and phase ϕ. Under this substitution, and expanding the gradient terms, this equation becomes

$$
\frac{\pi\hbar}{8k_B T_c}\left(e^{i\phi}\frac{\partial|\Delta|}{\partial t} + i|\Delta|e^{i\phi}\frac{\partial\phi}{\partial t} + \frac{2ie\varphi|\Delta|e^{i\phi}}{\hbar}\right) =
$$
$$
\xi_{mod}^2\bigg((e^{i\phi}\nabla^2|\Delta| + 2ie^{i\phi}\,(\nabla|\Delta|\cdot\nabla\phi) + i|\Delta|e^{i\phi}\nabla^2\phi - |\Delta|e^{i\phi}\,(\nabla\phi\cdot\nabla\phi)
$$
$$
-i\frac{2e}{\hbar c}\left(\nabla\cdot\vec{A}\right)|\Delta|e^{i\phi} - 2i\frac{2e}{\hbar c}\left(\vec{A}\cdot\nabla|\Delta|\right)e^{i\phi} - 2i\frac{2e}{\hbar c}|\Delta|e^{i\phi}\left(\vec{A}\cdot\nabla\phi\right)
$$
$$
-\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)|\Delta|e^{i\phi}\bigg) + \left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2}\right)|\Delta|e^{i\phi}.
$$
(B.1)

This can be separated into the real

$$
\frac{\pi\hbar}{8k_B T_c}\left(e^{i\phi}\frac{\partial|\Delta|}{\partial t}\right) =
$$
$$
\xi_{mod}^2\bigg((e^{i\phi}\nabla^2|\Delta| - |\Delta|e^{i\phi}\,(\nabla\phi\cdot\nabla\phi) - \left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)|\Delta|e^{i\phi}\bigg) \qquad\text{(B.2)}
$$
$$
+ \left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2}\right)|\Delta|e^{i\phi}
$$

and imaginary

$$
\frac{\pi\hbar}{8k_B T_c}\left(i|\Delta|e^{i\phi}\frac{\partial\phi}{\partial t} + \frac{2ie\varphi|\Delta|e^{i\phi}}{\hbar}\right) =
$$
$$
\xi_{mod}^2\bigg(2ie^{i\phi}\,(\nabla|\Delta|\cdot\nabla\phi) + i|\Delta|e^{i\phi}\nabla^2\phi \qquad\qquad\qquad\text{(B.3)}
$$
$$
-i\frac{2e}{\hbar c}\left(\nabla\cdot\vec{A}\right)|\Delta|e^{i\phi} - 2i\frac{2e}{\hbar c}\left(\vec{A}\cdot\nabla|\Delta|\right)e^{i\phi} - 2i\frac{2e}{\hbar c}|\Delta|e^{i\phi}\left(\vec{A}\cdot\nabla\phi\right)\bigg)
$$

parts of the equation. In the absence of a magnetic field $\left(\vec{A} = 0\right)$, and eliminating the common factors of $e^{i\phi}$, the equations are further reduced to the following:

$$\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial|\Delta|}{\partial t}\right) = \xi_{mod}^2\left((\nabla^2|\Delta| - |\Delta|\left(\nabla\phi\cdot\nabla\phi\right)\right) + \left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2}\right)|\Delta| \quad \text{(B.4)}$$

and

$$\frac{\pi\hbar}{8k_BT_c}\left(|\Delta|\frac{\partial\phi}{\partial t} + \frac{2e\varphi|\Delta|}{\hbar}\right) = \xi_{mod}^2\left(2\left(\nabla|\Delta|\cdot\nabla\phi\right) + |\Delta|\nabla^2\phi\right). \quad \text{(B.5)}$$

In the stationary state, the equation describing the imaginary terms leads to $\frac{\partial\phi}{\partial t} + \frac{2e\varphi}{\hbar} = 0$ provided $\xi_{mod}^2\left(2\left(\nabla|\Delta|\cdot\nabla\phi\right) + |\Delta|\nabla^2\phi\right) = 0$. This can be related to the conservation of the Ginzburg–Landau current

$$\xi_{mod}^2\left(2\left(\nabla|\Delta|\cdot\nabla\phi\right) + |\Delta|\nabla^2\phi\right) = \xi_{mod}^2\frac{4ek_BT_c}{\pi\sigma}\frac{\nabla\cdot\vec{j}_s^{GL}}{|\Delta|} = 0. \quad \text{(B.6)}$$

If we wish to instead enforce that the approximation for the Usadel supercurrent is conserved, we must modify the equation by the difference between the two supercurrent definitions, leading to

$$\begin{aligned}\frac{\pi\hbar}{8k_BT_c}\left(|\Delta|\frac{\partial\phi}{\partial t} + \frac{2e\varphi|\Delta|}{\hbar}\right) &= \xi_{mod}^2\left(2\left(\nabla|\Delta|\cdot\nabla\phi\right) + |\Delta|\nabla^2\phi\right) \\ &+ \xi_{mod}^2\frac{4ek_BT_c}{\pi\sigma}\frac{\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)}{|\Delta|} = 0\end{aligned} \quad \text{(B.7)}$$

for the imaginary term. By reinstating the factor of $ie^{i\phi}$ which was factored out, and substituting the definition for ξ_{mod}^2, we arrive at the term

$$+i\frac{e\hbar D}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}\frac{\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)}{|\Delta|}e^{i\phi} \quad \text{(B.8)}$$

which appropriately modifies the TDGL equation to match the approximate definition of the Usadel supercurrent in (3.12). This calculation was done in the absence of $\vec{A}$ for clarity of notation, but the result is identical if $\vec{A}$ is reinstated with the appropriate definition of the supervelocity.

B.2 Nondimensional Vodolazov TDGL Equations for Computation

In order to use finite-difference methods to calculate the evolution of the superconducting state using the modified TDGL equation, it is convenient to nondimensionalize the equations in order to minimize the risk of overflow and roundoff errors.

More importantly, the equations must avoid any singularities which occur during the natural evolution of the system. The TDGL equations permit such calculation, but not every representation is allowed.

In Appendix B.1, we demonstrate that the modified TDGL equations can be expressed as two coupled partial differential equations describing the magnitude $|\Delta|$ and phase ϕ of the order parameter. These appear to be a very convenient form of the equations, particularly for 1D calculations, because the supercurrent is naturally expressed in terms of $|\Delta|$ and $\nabla\phi$. However, the imaginary term (prior to adding Vodolazov's correction term) yields the nonlinear partial differential equation for ϕ

$$
\frac{\pi\hbar}{8k_BT_c}\left(e^{i\phi}\frac{\partial\phi}{\partial t} + \frac{2e\varphi e^{i\phi}}{\hbar}\right) =
$$
$$
\frac{\xi_{mod}^2}{|\Delta|}\Bigg(2e^{i\phi}\left(\nabla|\Delta|\cdot\nabla\phi\right) + |\Delta|e^{i\phi}\nabla^2\phi \tag{B.9}
$$
$$
- \frac{2e}{\hbar c}\left(\nabla\cdot\vec{A}\right)|\Delta|e^{i\phi} - 2\frac{2e}{\hbar c}\left(\vec{A}\cdot\nabla|\Delta|\right)e^{i\phi} - 2\frac{2e}{\hbar c}|\Delta|e^{i\phi}\left(\vec{A}\cdot\nabla\phi\right)\Bigg)
$$

which exhibits a singularity as $|\Delta| \to 0$. This singularity is independent of the modification to the temperature dependence of $\xi_{mod}^2(T_e)$ and $\Delta_{mod}^2(T_e)$ and is inherent to the TDGL formulation. While this form might be convenient to describe the response of the system to small perturbations, it cannot be used when describing the full suppression of the order parameter as occurs during single photon detection.

As an alternative, we can represent the complex order parameter as the sum of a real and imaginary part such that $\Delta = \psi_1 + i\psi_2$. The complex modified TDGL equation (3.12) becomes the two equations

$$
\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\psi_1}{\partial t} - \frac{2e\varphi\psi_2}{\hbar}\right) =
$$
$$
\xi_{mod}^2\left(\nabla^2\psi_1 - \left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\psi_1 + 2\frac{2e}{\hbar c}\left(\vec{A}\cdot\nabla\psi_2\right) + \frac{2e}{\hbar c}\left(\nabla\cdot\vec{A}\right)\psi_2\right) \tag{B.10}
$$
$$
+ \left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2}\right)\psi_1 - \frac{\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}}{|\Delta|^2}\frac{e\hbar D\psi_2}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}.
$$

and

$$\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\psi_2}{\partial t}+\frac{2e\varphi\psi_1}{\hbar}\right)=$$

$$\xi_{mod}^2\left(\nabla^2\psi_2-\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\psi_2-2\frac{2e}{\hbar c}\left(\vec{A}\cdot\nabla\psi_1\right)-\frac{2e}{\hbar c}\left(\nabla\cdot\vec{A}\right)\psi_1\right)\quad\text{(B.11)}$$

$$+\left(1-\frac{T_e}{T_c}-\frac{|\Delta|^2}{\Delta_{mod}^2}\right)\psi_2+\frac{\nabla\cdot\vec{j}_s^{Us}-\nabla\cdot\vec{j}_s^{GL}}{|\Delta|^2}\frac{e\hbar D\psi_1}{\sigma\sqrt{2}\sqrt{1+T_e/T_c}}.$$

which do not exhibit singularities as $|\Delta|\to 0$. The final $|\Delta|^2$ of the denominator of the final terms is canceled by a $|\Delta|$ term within the supercurrent definitions and the ψ term in the numerator. Within this set of definitions, $|\Delta|^2=\psi_1^2+\psi_2^2$, and $|\Delta|^2\nabla\phi=\psi_1\nabla\psi_2-\psi_2\nabla\psi_1$, with the latter being important for calculating the supercurrent.

To nondimensionalize equations the full TDGL equations described in Section 3.4, we must determine appropriate unit scaling. It is natural to normalize temperature by T_c and energies, such as Δ, in units of k_BT_c. Furthermore, it is convenient to chose the characteristic timescale $\tau_{T_c}=\frac{\hbar}{k_BT_c}$ and the characteristic length scale $\xi_c=\sqrt{\frac{\hbar D}{k_BT_c}}$ following [6]. From the TDGL equations, it is clear that φ can be normalized by $\frac{k_BT_c}{2e}$ and $\vec{A}$ by $\frac{\hbar c}{2e}\sqrt{\frac{k_BT_c}{\hbar D}}$. This leads to the following parameterization: $\tilde{T}_e=T_e/T_c$, $\tilde{T}_{ph}=T_{ph}/T_c$, $\tilde{\Delta}=\Delta/k_BT_c=\tilde{\psi}_1+i\tilde{\psi}_2=(\psi_1+i\psi_2)/k_BT_c$, $\tilde{\varphi}=2e\varphi/k_BT_c$, and $\tilde{\vec{A}}=\frac{2e}{\hbar c}\sqrt{\frac{\hbar D}{k_BT_c}}\vec{A}$. We also normalize the other functions in the equation according to $\tilde{\nabla}=\xi_c\nabla$, $\tilde{\xi}_{mod}^2=(\xi_{mod}/\xi_c)^2$, and $\tilde{\Delta}_{mod}^2=\Delta_{mod}^2/(k_B^2T_c^2)$. Substituting these parameters into the TDGL equations yields

$$\frac{\partial}{\partial\tilde{t}}\left(\frac{\pi^2}{12}E_0\tilde{T}_e^2-E_0\mathcal{E}_s\left(\tilde{T}_e,|\tilde{\Delta}|\right)\right)=$$

$$\frac{\pi^2}{6}E_0\tilde{\nabla}\tilde{k}_s\tilde{\nabla}\tilde{T}_e-\frac{24\zeta(5)E_0}{\tilde{\tau}_0}\left(\tilde{T}_e^5-\tilde{T}_{ph}^5\right)\quad\text{(B.12)}$$

$$-\frac{1}{8}E_0\left(\pi|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi-\tilde{\vec{A}}\right)-\tilde{\nabla}\tilde{\varphi}\right)\cdot\tilde{\nabla}\tilde{\varphi},$$

for the electron energy balance equation and

$$\tilde{T}_{ph}^3\frac{\partial\tilde{T}_{ph}}{\partial\tilde{t}}=\frac{90\gamma\zeta(5)}{\pi^4\tilde{\tau}_0}\left(\tilde{T}_e^5-\tilde{T}_{ph}^5\right)-\frac{1}{4\tilde{\tau}_{esc}}\left(\tilde{T}_{ph}^4-\tilde{T}_{sub}^4\right),\quad\text{(B.13)}$$

for the phonon energy balance equation. The thermal conductivity is defined by

$$\tilde{k}_s=\tilde{T}_e\left(1-\frac{6}{\pi^2}\int_0^{|\tilde{\Delta}|/\tilde{T}_e}\frac{x^2e^x dx}{(e^x+1)^2}\right)\quad\text{(B.14)}$$

where we have used the Einstein relation $\sigma = 2e^2 D N(0)$. The two equations of the modified TDGL formulation become

$$
\frac{\pi}{8}\left(\frac{\partial \tilde{\psi}_1}{\partial \tilde{t}} - \tilde{\varphi}\tilde{\psi}_2\right) = \frac{\pi}{8}\frac{\sqrt{2}}{\sqrt{1 + \tilde{T}_e}}\left(\tilde{\nabla}^2\tilde{\psi}_1 - \left(\tilde{\vec{A}} \cdot \tilde{\vec{A}}\right)\tilde{\psi}_1 + 2\left(\tilde{\vec{A}} \cdot \tilde{\nabla}\tilde{\psi}_2\right) + \left(\tilde{\nabla} \cdot \tilde{\vec{A}}\right)\tilde{\psi}_2\right)
$$

$$
+ \left(1 - \tilde{T}_e - \frac{|\tilde{\Delta}|^2}{\tilde{\Delta}_{mod}^2}\right)\tilde{\psi}_1
$$

$$
- \frac{\pi}{8}\frac{\tilde{\nabla} \cdot \left(2|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right) - \tilde{\nabla} \cdot \left(|\tilde{\Delta}|^2\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right)}{|\tilde{\Delta}|^2}\frac{\sqrt{2}}{\sqrt{1 + \tilde{T}_e}}\tilde{\psi}_2
$$

<div align="right">(B.15)</div>

and

$$
\frac{\pi}{8}\left(\frac{\partial \tilde{\psi}_2}{\partial \tilde{t}} + \tilde{\varphi}\tilde{\psi}_1\right) = \frac{\pi}{8}\frac{\sqrt{2}}{\sqrt{1 + \tilde{T}_e}}\left(\tilde{\nabla}^2\tilde{\psi}_2 - \left(\tilde{\vec{A}} \cdot \tilde{\vec{A}}\right)\tilde{\psi}_2 - 2\left(\tilde{\vec{A}} \cdot \tilde{\nabla}\tilde{\psi}_1\right) - \left(\tilde{\nabla} \cdot \tilde{\vec{A}}\right)\tilde{\psi}_1\right)
$$

$$
+ \left(1 - \tilde{T}_e - \frac{|\tilde{\Delta}|^2}{\tilde{\Delta}_{mod}^2}\right)\tilde{\psi}_2
$$

$$
+ \frac{\pi}{8}\frac{\tilde{\nabla} \cdot \left(2|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right) - \tilde{\nabla} \cdot \left(|\tilde{\Delta}|^2\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right)}{|\tilde{\Delta}|^2}\frac{\sqrt{2}}{\sqrt{1 + \tilde{T}_e}}\tilde{\psi}_1,
$$

<div align="right">(B.16)</div>

and the current conservation equation is expressed as

$$
\pi\tilde{\nabla} \cdot \left(|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right) = \tilde{\nabla}^2\tilde{\varphi}.
$$

<div align="right">(B.17)</div>

These equations are in a form which can be easily implemented with finite-difference schemes.

B.3   Nondimensional Generalized TDGL Equations for Computation

As discussed in Appendix B.2, computational methods require that the order parameter be expressed in terms of a real and imaginary components in order to avoid the singularity associated with the magnitude and phase representation. However, the generalized TDGL equations as commonly expressed are not in such a convenient form due to the asymmetry between the timescales of evolution of the magnitude and phase. We can manipulate these equations in order to produce a form which is amenable to standard computational methods. Beginning with the generalized

TDGL equation

$$\frac{\pi\hbar}{8k_B T_c}\left(\varrho\frac{\partial}{\partial t}|\Delta| + \frac{i|\Delta|}{\varrho}\frac{\partial}{\partial t}\phi + \frac{2ie|\Delta|}{\hbar\varrho}\varphi\right) = \xi_{mod}^2\left(\nabla + i\left(\nabla\phi - \frac{2e}{\hbar c}\vec{A}\right)\right)^2|\Delta|$$

$$+\left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2(T_e)}\right)|\Delta| + i\frac{\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)}{|\Delta|}\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}},$$

$$\tag{B.18}$$

we express the order parameter in terms of real and imaginary components $|\Delta| = \sqrt{\psi_1^2 + \psi_2^2}$ and $\phi = \mathrm{atan}\left(\frac{\psi_2}{\psi_1}\right)$, leading to the complex equation

$$\frac{\pi\hbar}{8k_B T_c}\left(\varrho\frac{\partial}{\partial t}\sqrt{\psi_1^2 + \psi_2^2} + \frac{i\sqrt{\psi_1^2 + \psi_2^2}}{\varrho}\frac{\partial}{\partial t}\mathrm{atan}\left(\frac{\psi_2}{\psi_1}\right) + \frac{2ie\sqrt{\psi_1^2 + \psi_2^2}}{\hbar\varrho}\varphi\right) =$$

$$\xi_{mod}^2\left(\nabla + i\left(\nabla\phi - \frac{2e}{\hbar c}\vec{A}\right)\right)^2\sqrt{\psi_1^2 + \psi_2^2}$$

$$+\left(1 - \frac{T_e}{T_c} - \frac{\psi_1^2 + \psi_2^2}{\Delta_{mod}^2(T_e)}\right)\left(\sqrt{\psi_1^2 + \psi_2^2}\right) + i\frac{\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)}{\sqrt{\psi_1^2 + \psi_2^2}}\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}.$$

$$\tag{B.19}$$

With straightforward manipulation, this can be expressed as

$$\frac{\pi\hbar}{8k_B T_c}\left(\left(\varrho\psi_1 - i\frac{1}{\varrho}\psi_2\right)\frac{\partial\psi_1}{\partial t} + \left(i\frac{1}{\varrho}\psi_1 + \varrho\psi_2\right)\frac{\partial\psi_2}{\partial t} + \frac{2ie\left(\psi_1^2 + \psi_2^2\right)}{\hbar\varrho}\varphi\right) =$$

$$\sqrt{\psi_1^2 + \psi_2^2}\,\xi_{mod}^2\left(\nabla + i\left(\nabla\phi - \frac{2e}{\hbar c}\vec{A}\right)\right)^2\sqrt{\psi_1^2 + \psi_2^2}$$

$$+\left(1 - \frac{T_e}{T_c} - \frac{\psi_1^2 + \psi_2^2}{\Delta_{mod}^2(T_e)}\right)\left(\psi_1^2 + \psi_2^2\right) + i\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}.$$

$$\tag{B.20}$$

We seek a form which separates the time derivatives of ψ_1 and ψ_2 in a form similar to the standard TDGL equations. This is achieved by separation of variables, substitution, and some algebra. We divide (B.20) by $\left(\varrho\psi_1 - i\frac{1}{\varrho}\psi_2\right)$ and separate the

equation into its real and imaginary components, leading to the complex equation

$$
\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\psi_1}{\partial t} + \frac{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2 + i\left(\psi_1^2 + \psi_2^2\right)}{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2}\frac{\partial\psi_2}{\partial t} + \frac{\varrho\psi_1 + i\frac{1}{\varrho}\psi_2}{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2}\frac{2ie\left(\psi_1^2 + \psi_2^2\right)}{\hbar\varrho}\varphi\right) =
$$

$$
\frac{\varrho\psi_1 + i\frac{1}{\varrho}\psi_2}{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2}\sqrt{\psi_1^2 + \psi_2^2}\xi_{mod}^2\left[\frac{\left(\psi_1\nabla^2\psi_1 + \psi_2\nabla^2\psi_2\right)}{|\Delta|}\right.
$$

$$
+ 2\left(\frac{2e}{\hbar c}\vec{A}\cdot\left(\frac{\psi_1\nabla\psi_2 - \psi_2\nabla\psi_1}{|\Delta|}\right)\right) - \left(|\Delta|\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\right)
$$

$$
\left. + i\left[\frac{\left(\psi_1\nabla^2\psi_2 - \psi_2\nabla^2\psi_1\right)}{|\Delta|} + \left(-2\frac{\psi_1\nabla\psi_1 + \psi_2\nabla\psi_2}{|\Delta|}\frac{2e\vec{A}}{\hbar c} - |\Delta|\frac{2e}{\hbar c}\nabla\vec{A}\right)\right]\right]
$$

$$
+ \frac{\varrho\psi_1 + i\frac{1}{\varrho}\psi_2}{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2}\left(1 - \frac{T_e}{T_c} - \frac{\psi_1^2 + \psi_2^2}{\Delta_{mod}^2\left(T_e\right)}\right)\left(\psi_1^2 + \psi_2^2\right)
$$

$$
+ i\frac{\varrho\psi_1 + i\frac{1}{\varrho}\psi_2}{\varrho^2\psi_1^2 + \frac{1}{\varrho^2}\psi_2^2}\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}.
$$

$$\text{(B.21)}$$

with the imaginary equation

$$
\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\psi_2}{\partial t} + \psi_1\frac{2e}{\hbar}\varphi\right) = \xi_{mod}^2\left[\left(\left(\frac{1}{\varrho} - \varrho\right)\frac{\psi_1\psi_2}{|\Delta|^2}\nabla^2\psi_1 + \frac{\left(\varrho\psi_1^2 + \frac{1}{\varrho}\psi_2^2\right)}{|\Delta|^2}\nabla^2\psi_2\right)\right]
$$

$$
+ \xi_{mod}^2\left[2\frac{2e}{\hbar c}\left(\frac{-\left(\varrho\psi_1^2 + \frac{1}{\varrho}\psi_2^2\right)\vec{A}\cdot\nabla\psi_1 + \left(\frac{1}{\varrho} - \varrho\right)\psi_1\psi_2\vec{A}\cdot\nabla\psi_2}{|\Delta|^2}\right)\right.
$$

$$
\left. - \varrho\psi_1\frac{2e}{\hbar c}\nabla\cdot\vec{A} - \frac{1}{\varrho}\psi_2\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\right]
$$

$$
+ \frac{1}{\varrho}\psi_2\left(1 - \frac{T_e}{T_c} - \frac{|\Delta|^2}{\Delta_{mod}^2\left(T_e\right)}\right) + \frac{\varrho\psi_1}{|\Delta|^2}\left(\nabla\cdot\vec{j}_s^{Us} - \nabla\cdot\vec{j}_s^{GL}\right)\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1 + T_e/T_c}}.
$$

$$\text{(B.22)}$$

Similarly, dividing (B.20) by $\left(\frac{1}{\varrho}\psi_1 - i\varrho\psi_2\right)$ generates the complex equation

$$
\frac{\pi\hbar}{8k_BT_c}\left(\frac{\left((\psi_1^2+\psi_2^2)+i\left(\varrho^2-\frac{1}{\varrho^2}\right)\psi_1\psi_2\right)}{\left(\frac{1}{\varrho^2}\psi_1^2+\varrho^2\psi_2^2\right)}\frac{\partial\psi_1}{\partial t}+i\frac{\partial\psi_2}{\partial t}\right.
$$

$$
\left.+\frac{\frac{1}{\varrho}\psi_1+i\varrho\psi_2}{\left(\frac{1}{\varrho^2}\psi_1^2+\varrho^2\psi_2^2\right)}\frac{2ie\left(\psi_1^2+\psi_2^2\right)}{\hbar\varrho}\varphi\right)=
$$

$$
\frac{\frac{1}{\varrho}\psi_1+i\varrho\psi_2}{\left(\frac{1}{\varrho^2}\psi_1^2+\varrho^2\psi_2^2\right)}\sqrt{\psi_1^2+\psi_2^2}\xi_{mod}^2\left[\frac{(\psi_1\nabla^2\psi_1+\psi_2\nabla^2\psi_2)}{|\Delta|}\right.
$$

$$
+2\left(\frac{2e}{\hbar c}\vec{A}\cdot\left(\frac{\psi_1\nabla\psi_2-\psi_2\nabla\psi_1}{|\Delta|}\right)\right)-\left(|\Delta|\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\right)
$$

$$
+i\left[\frac{(\psi_1\nabla^2\psi_2-\psi_2\nabla^2\psi_1)}{|\Delta|}+\left(-2\frac{\psi_1\nabla\psi_1+\psi_2\nabla\psi_2}{|\Delta|}\frac{2e\vec{A}}{\hbar c}-|\Delta|\frac{2e}{\hbar c}\nabla\vec{A}\right)\right]\right]
$$

$$
+\frac{\frac{1}{\varrho}\psi_1+i\varrho\psi_2}{\left(\frac{1}{\varrho^2}\psi_1^2+\varrho^2\psi_2^2\right)}\left(1-\frac{T_e}{T_c}-\frac{\psi_1^2+\psi_2^2}{\Delta_{mod}^2\left(T_e\right)}\right)\left(\psi_1^2+\psi_2^2\right)
$$

$$
+i\frac{\frac{1}{\varrho}\psi_1+i\varrho\psi_2}{\left(\frac{1}{\varrho^2}\psi_1^2+\varrho^2\psi_2^2\right)}\left(\nabla\cdot\vec{j}_s^{Us}-\nabla\cdot\vec{j}_s^{GL}\right)\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1+T_e/T_c}}.
$$

$$\tag{B.23}$$

with the real equation

$$
\frac{\pi\hbar}{8k_BT_c}\left(\frac{\partial\psi_1}{\partial t}-\psi_2\frac{2e}{\hbar}\varphi\right)=\xi_{mod}^2\left[\frac{\left(\frac{1}{\varrho}\psi_1^2+\varrho\psi_2^2\right)\nabla^2\psi_1}{|\Delta|^2}+\frac{\left(\frac{1}{\varrho}-\varrho\right)\psi_1\psi_2\nabla^2\psi_2}{|\Delta|^2}\right]
$$

$$
+\xi_{mod}^2\left[-2\frac{2e}{\hbar c}\left(\frac{\left(\frac{1}{\varrho}-\varrho\right)\psi_1\psi_2\vec{A}\cdot\nabla\psi_1-\left(\frac{1}{\varrho}\psi_1^2+\varrho\psi_2^2\right)\vec{A}\cdot\nabla\psi_2}{|\Delta|^2}\right)\right.
$$

$$
\left.+\varrho\psi_2\frac{2e}{\hbar c}\nabla\cdot\vec{A}-\frac{1}{\varrho}\psi_1\left(\frac{2e}{\hbar c}\right)^2\left(\vec{A}\cdot\vec{A}\right)\right]
$$

$$
+\frac{1}{\varrho}\psi_1\left(1-\frac{T_e}{T_c}-\frac{|\Delta|^2}{\Delta_{mod}^2\left(T_e\right)}\right)-\frac{\varrho\psi_2}{|\Delta|^2}\left(\nabla\cdot\vec{j}_s^{Us}-\nabla\cdot\vec{j}_s^{GL}\right)\frac{\hbar eD}{\sigma\sqrt{2}\sqrt{1+T_e/T_c}}.
$$

$$\tag{B.24}$$

Equations (B.22) and (B.24) provide the first order partial differential equations which are convenient for numerical simulation. It is easily confirmed that in the limit of the standard TDGL equations, $\tau_{sc}\to 0$, $\varrho\to 1$, and equations (B.22) and (B.24)

reduce to (B.10) and (B.11) respectively. These equations are nondimensionalized using the same procedure as listed in Appendix B.2. Using this procedure, the nondimensionalized TDGL equations are expressed as

$$
\begin{aligned}
\frac{\pi}{8}\left(\frac{\partial \tilde{\psi}_1}{\partial \tilde{t}} - \tilde{\varphi}\tilde{\psi}_2\right) &= \xi_{mod}^2 \left[\frac{\left(\frac{1}{\varrho}\tilde{\psi}_1^2 + \varrho\tilde{\psi}_2^2\right)\tilde{\nabla}^2\tilde{\psi}_1}{|\tilde{\Delta}|^2} + \frac{\left(\frac{1}{\varrho} - \varrho\right)\tilde{\psi}_1\tilde{\psi}_2\tilde{\nabla}^2\tilde{\psi}_2}{|\tilde{\Delta}|^2}\right] \\
&+ \frac{\pi}{8}\frac{\sqrt{2}}{\sqrt{1+\tilde{T}_e}}\left[-2\left(\frac{\left(\frac{1}{\varrho} - \varrho\right)\tilde{\psi}_1\tilde{\psi}_2\tilde{\vec{A}}\cdot\tilde{\nabla}\tilde{\psi}_1 - \left(\frac{1}{\varrho}\tilde{\psi}_1^2 + \varrho\tilde{\psi}_2^2\right)\tilde{\vec{A}}\cdot\tilde{\nabla}\tilde{\psi}_2}{|\tilde{\Delta}|^2}\right)\right. \\
&\left.+ \varrho\tilde{\psi}_2\tilde{\nabla}\cdot\tilde{\vec{A}} - \frac{1}{\varrho}\tilde{\psi}_1\left(\tilde{\vec{A}}\cdot\tilde{\vec{A}}\right)\right] \\
&+ \left(1 - \tilde{T}_e - \frac{|\tilde{\Delta}|^2}{\tilde{\Delta}_{mod}^2}\right)\frac{\tilde{\psi}_1}{\varrho} \\
&- \frac{\pi}{8}\frac{\tilde{\nabla}\cdot\left(2|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right) - \tilde{\nabla}\cdot\left(|\tilde{\Delta}|^2\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right)}{|\tilde{\Delta}|^2}\frac{\sqrt{2}}{\sqrt{1+\tilde{T}_e}}\varrho\tilde{\psi}_2
\end{aligned}
\tag{B.25}
$$

and

$$
\begin{aligned}
\frac{\pi}{8}\left(\frac{\partial \tilde{\psi}_2}{\partial \tilde{t}} + \tilde{\varphi}\tilde{\psi}_1\right) &= \frac{\pi}{8}\frac{\sqrt{2}}{\sqrt{1+\tilde{T}_e}}\left[\left(\left(\frac{1}{\varrho} - \varrho\right)\frac{\tilde{\psi}_1\tilde{\psi}_2}{|\tilde{\Delta}|^2}\tilde{\nabla}^2\tilde{\psi}_1 + \frac{\left(\varrho\tilde{\psi}_1^2 + \frac{1}{\varrho}\tilde{\psi}_2^2\right)}{|\tilde{\Delta}|^2}\tilde{\nabla}^2\tilde{\psi}_2\right)\right] \\
&+ \xi_{mod}^2\left[2\left(\frac{-\left(\varrho\tilde{\psi}_1^2 + \frac{1}{\varrho}\tilde{\psi}_2^2\right)\tilde{\vec{A}}\cdot\tilde{\nabla}\tilde{\psi}_1 + \left(\frac{1}{\varrho} - \varrho\right)\tilde{\psi}_1\tilde{\psi}_2\tilde{\vec{A}}\cdot\tilde{\nabla}\tilde{\psi}_2}{|\tilde{\Delta}|^2}\right)\right. \\
&\left.- \varrho\tilde{\psi}_1\tilde{\nabla}\cdot\tilde{\vec{A}} - \frac{1}{\varrho}\tilde{\psi}_2\left(\tilde{\vec{A}}\cdot\tilde{\vec{A}}\right)\right] \\
&+ \left(1 - \tilde{T}_e - \frac{|\tilde{\Delta}|^2}{\tilde{\Delta}_{mod}^2}\right)\frac{\tilde{\psi}_2}{\varrho} \\
&+ \frac{\pi}{8}\frac{\tilde{\nabla}\cdot\left(2|\tilde{\Delta}|\tanh\left(\frac{|\tilde{\Delta}|}{2\tilde{T}_e}\right)\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right) - \tilde{\nabla}\cdot\left(|\tilde{\Delta}|^2\left(\tilde{\nabla}\phi - \tilde{\vec{A}}\right)\right)}{|\tilde{\Delta}|^2}\frac{\sqrt{2}}{\sqrt{1+\tilde{T}_e}}\varrho\tilde{\psi}_1.
\end{aligned}
\tag{B.26}
$$

The energy balance and current conservation equations remain the same.

A p p e n d i x   C

ELECTROTHERMAL MODELING

C.1   Approximate Nanowire Temperature from Crosstalk

With incomplete information about the material used in in the crosstalk measurements, it is necessary to make estimates in order to place reasonable bounds on the maximum temperature reached during the crosstalk process for various biasing conditions. The two devices are from the same wafer, so they are considered to have the same material properties. The devices are designed to have a width of 160 nm, thickness of 4.8–5 nm, T_c of ~2.9–3.1 K, sheet resistance of 350–450 Ω, and diffusion coefficient of $\sim 0.75\,\text{cm}^2/\text{s}$, leading to a zero temperature depairing current of approximately 14.3 ± 3.3 µA using (4.12). This corresponds to an experimental switching current to depairing current ratio of 0.79 at 630 mK for the 800 nm pitch device, which is reasonable for the WSi material system, but could easily be an underestimate of the true depairing current.

Using this mean estimate for $I_{dep}(0)$, we can extract a maximum temperature T_{max} by determining the switching temperature for the 50% crosstalk efficiency value using

$$T_{max}(I_B) = T_c \left(1 - \left((I_{50} + I_{XT})/I_{dep}(0)\right)^{2/3}\right)^{1/2}, \tag{C.1}$$

where I_{50} is the receiving channel bias current which has 50% crosstalk efficiency for the triggering channel bias current I_B, and I_{XT} is the additional current forced through the nanowire due to electrical crosstalk. In the current calculations, I_{XT} is estimated to be ~5% of the bias current based on oscilloscope measurements of the electrical coupling between adjacent nanowires on the 400 nm pitch device. Lacking experimental measurements, this effect is ignored for the 800 nm device. This is the procedure used to generate the temperature values shown in Figure 4.10 with the dashed lines corresponding to the limits of maximum and minimum $I_{dep}(0)$ estimated in this section. An example of this inversion process is shown in Figure C.1 for bias currents of 6.5 µA (red) and 8.5 µA (blue). As the bias current decreases, the uncertainty in the associated switching temperature decreases because estimates of the depairing current converge to T_c.
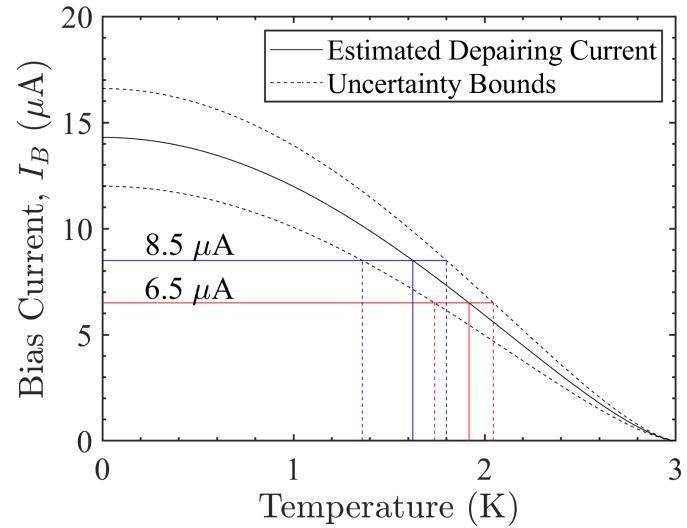
Figure C.1: Temperature-dependent depairing current estimates.