Digital and Microwave Superconducting Electronics and Experimental Apparatus

by

Brenden A. Butters

M.S., Massachusetts Institute of Technology (2018)
B. Engineering, University of Wollongong (2016)
B. Mathematics, University of Wollongong (2016)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

Author .....
Department of Electrical Engineering and Computer Science
January 20, 2022

Certified by .....
Karl K. Berggren
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Digital and Microwave Superconducting Electronics and Experimental Apparatus

by

Brenden A. Butters

Submitted to the Department of Electrical Engineering and Computer Science
on January 20, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

The lack of a high-performance and scalable superconducting memory has been a persistent issue in the field of superconducting computing for decades. There have been many attempts at addressing this issue; however, to-date no technology has been able to completely satisfy this demand. In this work we present a novel memory design based on superconducting nanowires controlled by localized thermal effects. Initial results from this design are very promising and suggest that with some further development, our design may satisfy the need for such a superconducting memory technology.

As superconducting nanowire electronics mature and become increasingly faster and more complex, the traditional reliance on off-chip microwave components has become unsustainable. In this thesis, we present the design and experimental results for a set of on-chip microwave devices, including bias tees, filters, detectors, couplers, and delay lines. In addition, by using the modeling developed for the memory, we make this set of microwave devices tunable through thermally controlling their kinetic inductance. To demonstrate the on-chip instrumentation that this library enables, a characterization of the thermal response of our tunable devices by means of an on-chip interferometer is presented. With the increasing complexity of our designs, we find ourselves in need of a new experimental apparatus to support our work. Finding no suitable solutions either commercially available or in literature, we developed a new versatile cryogenic experimental platform for nanowire electronics. The design presented here consolidates what was previously a number of discrete setups into one universal platform, while also greatly improving performance. Through the advances presented in this work, we have enabled the future realization of more complex nanowire-based superconducting electronics.

Thesis Supervisor: Karl K. Berggren
Title: Professor of Electrical Engineering and Computer Science

To Emily and my parents.

Acknowledgments

Much of the work presented in this thesis would not be possible without the efforts of many of my colleagues. I would like to acknowledge:

My advisor, Prof. Karl Berggren, for his insights, his encouragement, and his dedication to the development of his students. His enthusiasm for our work is what kept our work progressing – even when everything seemed to be going wrong.

Prof. Kevin O'Brien and Prof. Daniel Santavicca, for kindly agreeing to serve on my committee and for their many insights and helpful suggestions.

Owen Medeiros, for his efforts in fabricating the superconducting devices tested in this work. As well as his assistance in experiments, discussions, and upkeep of the lab.

Marco Colangelo, for his friendship and the many helpful discussions and resources that he provided over the years.

Dorothy Fleischer, for helping keep me organized, and for her ever-helpful and generous attitude.

Navid Abedzadeh and Marco Turchetti, for their assistance with photography and many interesting discussions.

All of the past and present members of the Quantum Nanostructures and Nanofabrication group. I could not have asked for a better group of people to have worked with over the past few years.

Jim Daley and Mark Mondol, for their upkeep of the NSL.

I would also like to acknowledge the major funding support that various aspects of this research received from IARPA, DARPA, NASA, and NSF.

The work that I conducted in pursuit of this thesis, and in other related projects, was only made possible by the support and advice that I received from my family and friends. In particular, I would like to thank:

Mark and Amy Belanger and their family, for their friendship and for all the support they have generously provided me over my time at MIT. I do not know what I would have done without them.

David Brunelle, for his friendship, generosity, and all the helpful technical discussions over the years.

Patrick McAtamney, for his friendship, the many interesting discussions, and his generous help.

Prof. Raad Raad, for supporting my research, all his help and encouragement, and the many hours of discussions we have had over the years. Without his influence, I would likely not have made it to where I am today.

Prof. Son Lam Phung, for his support in my early research. I might not have pursued this path if it were not for his encouragement.

Tony Missett and Joseph Larrivee, two of the best men I have met at MIT. Without their hard work and kindness, my time at MIT would not have been nearly as enjoyable. I have also learned more about MIT's history from Tony than from anyone else at MIT.

Dave Lewis, Anthony Pennes, and John Sweeney, for their work at EDS which proved an invaluable resource to both my research and personal projects.

Kenny Chen and Marie Shi Feng, for their friendship and all the great times we had while at the Warehouse.

My girlfriend, Emily Standefer, for her unending support and encouragement, and for the many hours she spent proofreading my work. I would also like to thank her family for their generosity and all the help that they have given me.

Finally, I would particularly like to thank my parents for their unwavering support and dedication. I am forever grateful for the sacrifices that they have made in order for me to pursue the path that has led to me completing my degree.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There has been a sustained demand for ever more efficient large scale computers. As such, there have been substantial efforts to develop cryogenic supercomputers as one possible solution [1, 2]. One of the main deficits in the field of cryogenic electronics at this point in time is the lack of a scalable memory technology. We have previously developed superconducting memory technologies compatible with these applications; however, there have been issues in realizing arrays of these memories [3–5]. In this work we will present a design of a memory cell specifically engineered for this application and designed explicitly for high density arraying.

In moving towards ever more complex nanowire devices operating at higher speeds, we find a need for on-chip microwave devices. In the memory work, the kinetic inductance changing with temperature is an undesirable effect that we needed to compensate for. To perform this compensation, we required the formation of an accurate model. In the microwave devices, we use this effect and modeling to realize tunability. These devices allow us to perform testing of nanowire devices that would not be possible otherwise.

The testing of our digital and microwave devices requires specific experimental apparatus, which must be custom-made for the purpose. In this work, we will cover the new superconducting memory device developments, the microwave devices that came out of the memory efforts, and the experimental apparatus required to test all these devices.

1.1 Superconducting switching devices

Broadly speaking, there are two groups of superconducting switching devices. The first group consists of those devices based on tunnel junctions, such as Josephson junctions (JJs), which we collectively refer to as "junction devices" [6]. The second group consists of those devices based on transitioning a region of the superconductor to its normal state. We collectively refer to this group of devices as "cryotrons" in reference to Dudley Buck's cryotron [7]. For more detailed descriptions of superconducting switching devices see [5].

For cryotrons, we exploit the fact that the superconducting phase can only exist for temperature below the critical temperature T_c, current densities below the critical current J_c, and magnetic fields below B_c for type I superconductors and B_{c2} for type II superconductors. The original cryotrons, as used by Buck, rely on magnetic fields to transition a superconducting wire to the normal state. Here, we use only current and temperature to switch our devices.

In this work we make use of the heater-cryotron (hTron) which is a four terminal switching device and is very much the thermal analog of the cryotron. The device utilizes a heater (which in the devices presented here is always a normal metal) placed in close proximity to a superconducting channel. When the heater current is increased, the local temperature increases and the channel switching current decreases. We have performed extensive analysis on the hTron, and the results are available in [8]. However, an effect that is central to the nature of cryotrons is that their inductance varies along with their switching current. This effect is often overlooked but is critical in this work. This effect must be mitigated in the memory work (see chapter 2) and is exploited for device tunability in the microwave work (see chapter 3).

It should be noted that all critical values depend on each other. In practice, we do not tend to observe variation in the critical temperature with the application of current; however, we do observe, and in fact rely on, the critical current changing with temperature. That is, we actively use increased local temperature to suppress the critical current – this is the basis of the hTron [8]. We can, with finite current

through a device, measure the critical temperature reliably. We can not easily measure the critical current. This difficulty arises from the fact that measuring a critical current involves flowing a current through a wire, which by its nature distributes the measurement over a spatially large region. The fact that we spread the measurement region over a wire means that a number of effects come into play, including inhomogeneities in the film, surface roughness, sensitivity to variation in cross-sectional area, current crowding effects, and a host of other effects [9]. Thus, we can only measure the "switching current" directly which is always less than the theoretical maximum current carrying capacity. With this in mind, throughout this work we use the terms "critical current" and "switching current" interchangeably as these are the observable values. When the theoretical critical current is meant, the term "depairing current" is used as it refers to the theoretical current at which Cooper pairs break down and "depair". In our devices the switching current is roughly 70% the theoretical depairing current [10].

1.2 Superconducting memory

There is a desire to improve the performance of supercomputers. Reducing power consumption while either maintaining or improving speed is highly desirable [1,2]. Thus, there is considerable interest in alternate computing paradigms and technologies. One such contender to the dominance of complementary metal-oxide-semiconductor (CMOS) devices are single flux quantum (SFQ) circuits. SFQ devices are based on Josephson junctions (JJ), and offer very high speed (around 1 ps switching time) with very low power dissipation (around 0.1 aJ per switching event). It would seem that these devices are the ideal successor to CMOS; however, it has been identified that there are three main technological shortcomings that are preventing SFQ from being competitive with CMOS. First, a lack of suitable interconnects between the superconducting device and room temperature, second, difficulty in the fabrication of the devices, and finally, the lack of a scalable memory technology [11].

1.2.1 Memory efforts to-date

There has been a consistent effort over the years to develop a suitable memory technology for SFQ-based computers. Naturally, cells based in SFQ logic would be the first option explored. These SFQ-based memories, while typically fast and energy efficient, are prohibitively large [12, 13]. There have even been calls for the use of CMOS memories due to their very high density, with the downside being their high power dissipation [14–16]. One proposed way around this issue is operating the CMOS memory at intermediate temperatures (77 K) where power dissipation is less of an issue [17].

There has been a focus on the development of magnetic and magnetic-superconductor hybrid memories [18–22]. Recent results of cryogenic spin-valves and magnetic tunnel junctions work have showed promise. These devices have demonstrated good performance. The combination of hTrons with magnetic tunnel junctions based on the spin-Hall effect has led to the development of a new memory design [11]. This new design, while addressing many of the shortcomings of the existing technology, still requires a relatively large area per bit and can be challenging to fabricate. Thus, there is still a substantial gap between the state of cryogenic memory now, and where it needs to be for SFQ and cryogenic computing in general, to be a competitive alternative to CMOS.

1.2.2 Cryotron-based memory cells

The majority of the memories currently being pursued are based on junction devices (with some using cryotrons for cell access [11]). In this work we present a memory design based exclusively on nanowire cryotrons. However, the concept of a cryotron-based memory is not new. In fact, cryotron-based memories date back to the 1950s and the invention of the cryotron. In Buck's original paper, he presents a number of flip-flop and register designs [7]. Extensive work was invested in investigating the cryotron as a storage element was reported in [23].

Central to these early cryotron storage elements was the cryotron-based flip-flop

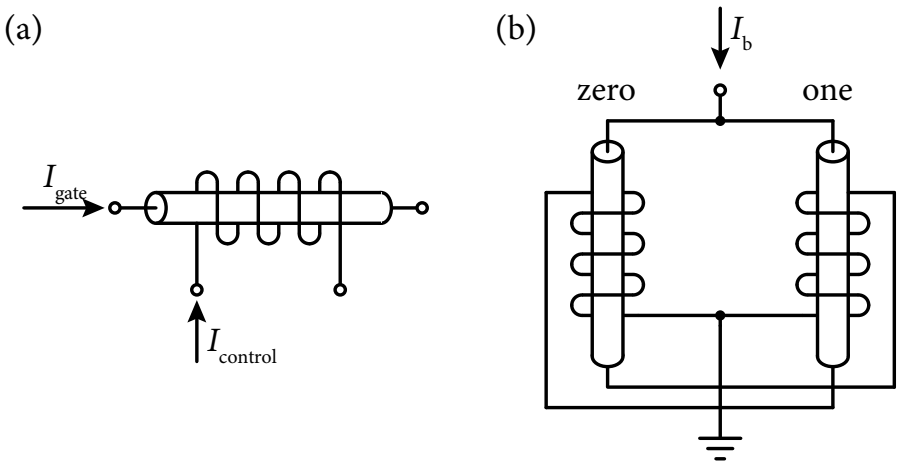

Figure 1-1: The cryotron-based flip-flop. (a) The schematic symbol of the cryotron, as proposed by Buck. This symbol is a simplification of the physical construction of the device where a thin niobium control wire carrying the control current I_{control} is wound around a thicker tantalum conductor (the gate) which carries a current I_{gate}. Note that the naming convention of Buck differs from our own: what Buck called the "gate" we call the "channel", and what he called the "control" we call the "gate". (b) The simple bistable element at the core of the original cryotron-based memory devices. When the majority of the bias current I_{b} flows through the gate of the "zero" branch, then the cell is said to be in the "zero" state. Conversely, when the majority of the current flows through the "one" branch, the cell is said to be in the "one" state. This cell is then built upon with additional cryotrons to provide readout, set/reset lines, and to form arrays.

shown in figure 1-1. This device is relatively simple in its basic operation. The bias current I_{b} splits between two branches, namely the "zero" branch and the "one" branch. Provided an appropriate bias current was selected, then one of the two sides would have its gate (what we would call the channel) switched. This would force the bias current into the opposing cryotron's control wire, further suppressing the superconductivity in the opposing gate. Thus, we have a positive feedback arrangement. Provided the bias is not removed from the cell, it will retain its state. This basic flip-flop is then extended by the incorporation of additional cryotrons to allow readout, setting and resetting the cell, and addressing when built into an array.

The first major downside of the original cryotron-based flip-flop was the size of the cryotrons. The cryotrons were typically composed of a 0.003" niobium control wire wound around a 0.009" tantalum gate wire with the whole device being an inch long.

For comparison, we could fit hundreds of our memory design as presented in figure 2-7 into the cross section of the control wire. However, this is not a fair comparison as the cryotron was competing against the vacuum tube. Even the smallest vacuum tubes of the era (which were commercialized a few years after the cryotron's publication), namely the nuvistor, were far larger and consumed orders of magnitude more power [24].

In efforts to downsize the cryotron, the "crossed-film cryotron" was developed. This is a planar analog of Buck's original device and operated in a similar manner [25, 26]. This device was used to make two major advances over the cryotron flip-flop. First, it miniaturized the design, and second, it removed the need for the quiescent current. The authors observed that a circulating current could be stored in the cell's loop, thereby keeping the state with no external bias applied. In these planar cryotron designs, the memory is organized into a shift register.

In order for a cryotron-based memory to be useful in modern designs, it must be small, energy efficient, and have addressable cells. The original cryotron flip-flops were addressable but dissipated quiescent power and were large. The planar cryotron cells reduced the cell size to $6\,\mathrm{mm}^2$ and eliminated quiescent power draw, but were slow and not addressable. Thus, a design that incorporates the best of both of these design approaches and further minimizes the cell size would be ideal. Unfortunately, the operating mechanisms of these designs are somewhat incompatible, making their merger non-trivial. Thus, a new memory design would be required.

1.2.3 Modern cryotron-based memory cells

We began development of the nanocryotron-based memory (nMem) with the non-destructive readout cells presented in [4] and discussed in detail in [5]. These cells were written to by breaking the loop with an in-plane hTron and read out by using a yTron. The hTron being the thermal equivalent to the cryotron, this design is somewhat reminiscent of the original cryotron-based storage elements. The yTron, on the other hand, is a relatively new device that allows for the memory state to be read non-destructively [27].

As we discussed at length in [5], the major downside of the non-destructive nMem is the fact that the cell has three terminals that are all connected by a superconductor – due to the yTron readout mechanism. This means that forming an array with this cell design is difficult, as sneak currents and undesired flux trapping must be addressed. To prevent the undesired flux from causing issues in the memory operation, resistive elements must be introduced to the array. These elements dissipate the flux trapped in these parasitic loops, but harm the speed, power, and scalability of the memory. Thus, an alternate solution is necessary for this memory technology to be competitive.

To realize a memory that operates without the need for the yTron, and hence requiring only two superconducting terminals for each memory cell, we developed the destructive readout nMem [3, 5]. This memory replaces the yTron with a second hTron, forming a superconducting quantum interference device (SQUID) from the memory. This greatly reduced the complexity of forming an array from the memory, as the cells can simply be tiled without thought to sneak currents or parasitic loops. The downside is that while the cell is very simple in construction and layout, the exact ratios of the branch inductances and switching currents make the cell somewhat complex to design.

In this work, we take the basic design of the destructive-readout memory presented in [3,5] and introduce a new readout operating scheme and greatly improved modeling. Through our modeling, we realized that the hTron is as much a variable inductor as it is a variable switching current device. Given the importance that inductance ratios play in the operation of the memory, this effect must be appropriately treated when designing the cell. With the advances we have made here we take a memory with a barely-usable error ratio of 1.5×10^{-3} of the original design and, through our improvements, demonstrate a design with an excellent error ratio of 7×10^{-10}. In addition, we perform modeling and characterization of the device that will enable its further development, and the development of similar devices in the future.

1.3 Superconducting microwave devices

The unique characteristics of superconductors lend themselves to application in creation of microwave devices impossible or otherwise much more difficult to realize in conventional materials [28]. In addition, many phenomena in superconductors are microwave in nature. These phenomena include the bandwidth of SFQ logic signals, which can operate at tens to hundreds of gigahertz [29]. The simple switching of a superconductor from the normal to the superconducting state produces very fast ramp rates [30]. These effects even include macroscopic quantum effects where we find that the energy difference between two states of a qubit are firmly in the microwave range [31, 32]. Non-linear effects in superconductors, and in particular JJs, are also used to make quantum-limited amplifiers [33].

Superconductors generally produce very low loss microwave devices. In a superconductor, according to the two-fluid model, there are both quasiparticles (normal electrons) and Cooper pairs (superconducting electrons) that participate in conduction. Within this model, finite loss in superconductors are due to the kinetic energy of the Cooper pairs creating a kinetic inductance effect, which results in non-zero impedance to AC signals. Thus, the quasiparticles will, as influenced by this electric field, participate in conduction, leading to loss. For this reason, when creating microwave superconducting devices, it is typical to use low kinetic inductance materials (unlike NbN). When larger inductances are needed in a low loss circuit, one must either use geometric inductance or Josephson inductance.

The microwave kinetic inductance detector (MKID) is a device that relies on the kinetic inductance changing with incidence of photons. The device is essentially a resonator that experiences a shift in resonant frequency when acted upon by incoming radiation [34]. It is one of the few devices that directly utilizes the variability of kinetic inductance for its primary operation.

The microwave treatment of a key nanowire device, the superconducting nanowire single-photon detector (SNSPD), by [35] led to a number of advancements in SNSPD readout techniques. The microwave treatment of large SNSPDs led to the develop-

ment of the superconducting nanowire single-photon imager (SNSPI) [36, 37]. The inclusion of on-chip impedance matching of SNSPDs led to improved performance of SNSPDs [30]. This treatment of nanowires has even led to the discovery of a new operating regime in which photon number resolution is possible [38]. By following this trend and leveraging the high kinetic inductance of our nanowires, we can physically compress designs that would otherwise be much larger with geometric inductance, as well as realize new devices such as on-chip forward couplers [39, 40].

Given the increasing treatment of our nanowire devices as intrinsically microwave in nature, we necessarily need to measure them as such. We have been utilizing microwave measurement techniques of our devices for some time now [5]. However, as the devices become more and more complex, we are beginning to run into issues with many microwave components being necessary for our measurements. This is compounded by nanowires (and superconductors in general) being sensitive to noise. Thus, we also need to incorporate filtering into our lines. For more complex devices, or when we test many devices, we either must resort to having these devices at room temperature, which brings many downsides, or installing them into the cryostat. The latter option is preferable for many reasons, not least of which is noise. However, as we make increasingly complex devices, we simply run out of space inside of the cryostat. To remedy this situation, we can incorporate our microwave devices on-chip.

The incorporation of microwave devices on-chip as discussed here is in contrast to that typically presented in literature. The majority of on-chip microwave devices published to-date are highly customized and central to the operation of the device for which they are testing. Here, we aim to develop more generic microwave devices that we can use with our nanowire devices. We need something akin to an on-chip equivalent of the ubiquitous Mini Circuits and similar interconnected microwave devices that find their way into nearly every experimental setup. That is to say, the devices are to be more generic, somewhat lower performance, but highly versatile. The circuit designer should not have to spend time simulating these structures. Rather, we need known devices can simply be incorporated into the design.

There are two methods that this "microwave toolbox" could be used by the circuit

designer. The designer could incorporate the microwave device into their design by simply importing the layout and wiring on-chip. Alternatively, these devices could be placed on the same chip (or a second chip on the same PCB) and configured with wire bonds. Either way, with the incorporation of these devices on-chip, we can achieve design density and complexity that is difficult or impossible with our current measurement techniques. The treatment of the hTron as a variable kinetic inductance device allows us to realize tunability in these microwave devices, further increasing their generalization and usefulness.

In this work, we lay the groundwork for further development of the microwave toolbox. We demonstrate an on-chip variable resonator, variable filters, a bias tee, a detector, a switch, and a demonstration of an interferometer for measuring the change in kinetic inductance in an hTron. The filters we demonstrate here, owing to their utilization of kinetic inductance, are physically very small. These devices are so small that we believe them to be the smallest multi-pole microwave filter for this frequency band ever realized. Furthermore, this device is widely tunable.

1.4 Cryogenic testing

In order to test our superconducting devices, we need to bring them to below their critical temperature. For the NbN devices presented in this work, the critical temperature is around $T_c = 10\,\mathrm{K}$. However, we need to be substantially below this temperature for our device to operate as intended. In practice, we need temperature below around $5\,\mathrm{K}$ for our devices to have enough headroom for the heater to effect a useful change in device behavior.

There are a number of cryostat designs that are used in superconducting device testing that would be suitable for our application [28, 41, 42]. Generally, the large variety of cryostat designs are divided into two main groups, namely "wet" and "dry" cryostats. A dry cryostat (also referred to as a cryogen-free system), comprises a closed system where fixed mass of refrigerant is cycled though some thermodynamic cycle to achieve cryogenic refrigeration. For the temperatures we are operating at,

this refrigerant is helium. All systems have some kind of cryocooler at the first stage. In cryocoolers, the refrigeration cycle is repeated continuously (as opposed to one-shot systems). Generally, a compressor is used to pressurize the working fluid, the high pressure fluid is then pre-cooled in a heat exchanger, and the high pressure fluid is expanded, thereby absorbing heat. The low pressure cold gas is then used in the exchanger to pre-cool the pressurized fluid, before the low pressure fluid returns to the compressor. There are a number of different types of cryocooler designs including the Stirling cooler, Gifford-McMahon cooler, pulse-tube refrigerator, and the Joule-Thomson cooler. Each of these designs have their advantages and disadvantages; however, all coolers are very limited in their cooling power when compared to wet systems. Where a wet system might take 30-minutes to cool down, a cryogen-free system might take from many hours to a number of days. Furthermore, a cryogen-free system requires careful design of the interconnects between the cold head and higher temperatures. This is again so that the cooling power of the cryocooler is not exceeded, thereby raising the ultimate temperatures of the system. Finally, all cryocoolers require some motors to drive the compressor and/or valves. These can be a source of electrical noise in the system, which can disrupt sensitive measurements – see section 5.2.

There are a number of wet system designs, the most notable of which are probes and flow cryostats. In a flow cryostat, the cryogen is transferred out of the dewar and into an external vessel where it cools the sample. After cooling the sample, the typically gaseous cryogen is either recovered for recycling or vented. For probe-style cryogen systems, there are two main designs. The first is where the sample is physically located outside of the dewar. In this configuration, cooling of the sample is achieved by means of a cold-finger that is thermally coupled to a high-thermal-conductivity rod that is exposed to the cryogen. The other design situates the sample within the dewar where either it is exposed to the cryogen (immersion testing) or isolated from the liquid cryogen but thermally coupled to the liquid. Due to the high turnaround and the ability to have many low-loss radio frequency (RF) lines to the sample, we rely on wet testing extensively for digital and microwave device testing –

as well as screening devices and performing T_c measurements when needed.

1.4.1 Immersion testing

Immersion testing is the original method of testing superconducting devices and has been used for testing croytrons back to the 1950s [23]. Immersion testing involves simply immersing the sample into a liquid cryogen to cool then sample. This was the primary method we used for testing digital devices and devices requiring any setup not available in our cryogen-free systems. While immersion testing is very simple, it is exactly this simplicity that makes it so useful. The fact that the dewar does not require a cryocooler means that it is free from mechanical vibrations and electrical interference due to the operation of a cooler. The fact that the cooling power of a dewar is very high means that it is possible to use low loss coaxial (coax) cables for interfacing the device under test.

For details on the process we used for immersion testing see section 2.4.1 of [5]. The basic process we follow for immersion testing begins with mounting our chip to a custom printed circuit board (PCB). We then wire bond to the chip and cover it with a plastic cap for protection. After wire bonding, we install the PCB onto the end of a number of coaxial cables. This assembly is covered in aluminum foil and slowly lowered into the dewar. All vented helium gas is lost and there is the possibility of ice forming in the dewar neck – possibly leading to the explosive failure of the dewar. Thus, there is room for improvement of our immersion testing protocols.

1.4.2 New apparatus

Given that our existing setup for immersion testing has a number of shortcomings, we set about searching for a new cryostat that meets our needs. There are a number of commercial offerings, including those from FormFactor, ICE Oxford, Oxford Instruments, and others. In addition, we examined a number of published designs for similar experimental setups [43–45]. However, we did not find any solutions that met our needs and were compatible with our existing equipment. Specifically, we required

optical illumination, the ability to expose the sample to magnetic fields, control the temperature accurately, have the sample in a vacuum, have many high-bandwidth low-loss microwave connections to the device, and rapidly cycle the sample. In addition, we wished for the sample to not require any expensive chip carriers but rather only require low-cost printed circuit boards (PCBs) for sample mounting. Thus, we elected to proceed with our own design of a probe that fulfills our needs, provides expandability in the future, is simple to construct, and fits within our existing processes and setups.

Our existing setup operates at 4.2 K, with up to 13 RF lines capable of around 6 GHz operation. We can install an optical fiber for flood illumination. Magnetic fields are difficult to obtain, and are limited to small values in our existing setup. We desire to improve all these specifications while retaining all functionality. In addition, as we expand to more complex devices we wish to increase the number of RF lines and their bandwidth. A summary of the desired specifications of the probe is shown in table 1.1. In addition to these specifications, we wish to add a temperature sensor, temperature control, and the option to operate the device in a vacuum or gaseous atmosphere (as opposed to immersed in the liquid helium). Finally, we need the system to enable us to recover the helium boiled off during cool-down and during the experiment.

One of the major issues we encountered while performing immersion testing was the reliability of the connectors [5]. We used SMP limited detent connectors which performed well under most normal circumstances. However, given the limited space in the dewar neck, the connectors would often become disconnected. A disconnected

Table 1.1: CryoProbe specifications

Specification	Minimum desired value	Design presented here
Maximum base temperature	4.5 K	4.287 K
Minimum hold time	24 hours	122 hours
RF cable bandwidth	12.5 GHz	20 GHz
Number of RF lines	16	28
Minimum magnetic flux density	0.1 T	0.41 T
Weight	< 30 lb	< 25 lb

cable required the sample to be warmed up, the connection reestablished, and the sample cooled again which is wasteful of helium and can lead to condensation on the sample. Additionally, the need to solder connectors onto sample PCBs for each sample became both costly and time-consuming. Thus, we wish to create a system that does not require connectors to be soldered to the sample PCB.

1.5 Thesis outline

This thesis is broken into six chapters, each of which is summarized in the following sections:

Chapter 2 – A scalable superconducting nanowire memory – the nMem

With the need for a better superconducting memory established, this chapter will cover a candidate high performance memory design. Of particular note in this chapter is the engineering of the kinetic inductance changing with temperature for device operation. The design process, simulation, parameter extraction, experimental setup, experiential results, and analysis will be covered.

Chapter 3 – Nanowire microwave devices – the microwave toolbox

This chapter presents a set of designs for microwave devices and discusses the design procedure for realizing on-chip microwave components and instrumentation. Of particular note is that this chapter takes the principal of kinetic inductance varying with temperature and utilizes it for the tuning of microwave devices.

Chapter 4 – A versatile cryogenic experiment platform – the CryoProbe

This chapter covers the design and development of a versatile cryogenic experiment platform. This probe was designed to enable measurements of superconducting nanowire devices that were previously difficult to perform. This platform was used to gather the majority of the experimental data presented in this work.

Chapter 5 – Experimental apparatus

This chapter provides details on the design and characterization of various experimental apparatus that have been developed for the experiments presented in this work, as well as for other projects not covered in this work. These apparatus include the nanoBERT, a bit error ratio test set for the nMem, a low noise motor driver for eliminating electrical motor noise in pulse-tube cryocoolers, the OmniBias, a device for measuring the current and voltage relationship in nanowires under diverse bias conditions, and other useful devices.

Chapter 6 – Conclusion and future work

This chapter provides concluding remarks and suggests topics that could be the subject of future work.

All mechanical drawings presented in this work are third angle projections with dimensions in inches unless otherwise noted.

Chapter 2

A scalable superconducting nanowire memory – the nMem

With the need for a scalable superconducting memory established in section 1.2, this chapter will present the advances in superconducting nanowire memories we have made since publishing [3].

2.1 Operating principal

The memory cell in essence is an asymmetric SQUID realized with kinetic rather than magnetic inductance. The schematic for a single memory cell is shown in figure 2-1. The state of the cell is encoded in the persistent super current I_p. This current is read out by operating the cell as a SQUID [46]. The important development of this memory over a simple asymmetric SQUID is the use of hTrons, which allow us to locally suppress the superconductivity in the cell, thereby selecting it from an array [8].

The SQUID nature of the cell would suggest it is sensitive to magnetic fields. This is not true as we rely almost exclusively on kinetic rather than magnetic inductance. This is due to our selection of materials, specifically NbN. For our devices, as we will discuss later, we realize kinetic inductances on the order of $20\,\mathrm{pH}/\square$, and each device consists of tens of squares. Our devices, for example the memory presented in figure

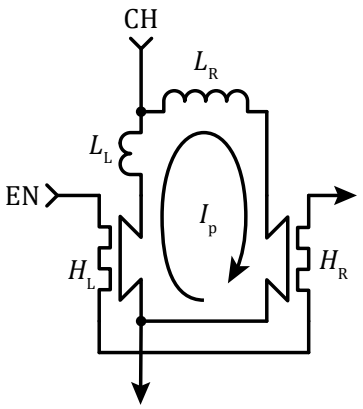

Figure 2-1: Simplified schematic of the memory cell. The channel terminal, CH, is used to write to and read from the memory cell. The enable line, EN, is used to select the cell when formed into an array. The inductor L_L is smaller than the inductor L_R, and each represents the branch inductance of each side of the memory cell. The state of the memory is encoded in the sign of the persistent current I_p. The persistent current will continue to flow indefinitely provided the loop is maintained in the superconducting state.

2-7, due to their small size possess a magnetic inductance of less than 10 pH [47]. Thus, the fact that the magnetic inductance is substantially smaller than the kinetic inductance means that we can safely ignore the magnetic inductance and only consider the kinetic inductance in our calculations. This being said, there is a measurable reaction to extremely large magnetic fields [5]. However, in a general application we do not need to be concerned with magnetic interference.

The memory cell is designed such that it can be tiled to form an array, as shown in figure 2-2. The ability to organize the cells in this manner without the need for external select logic is what sets this apart from other superconducting memory technology. The ability to form an array in such a simple manner makes the formation of arbitrary sized arrays simple. This ability, together with the ability to scale the cell size to almost any dimension (due to its reliance on kinetic inductance) is central to this memory technology's scalability [3, 5].

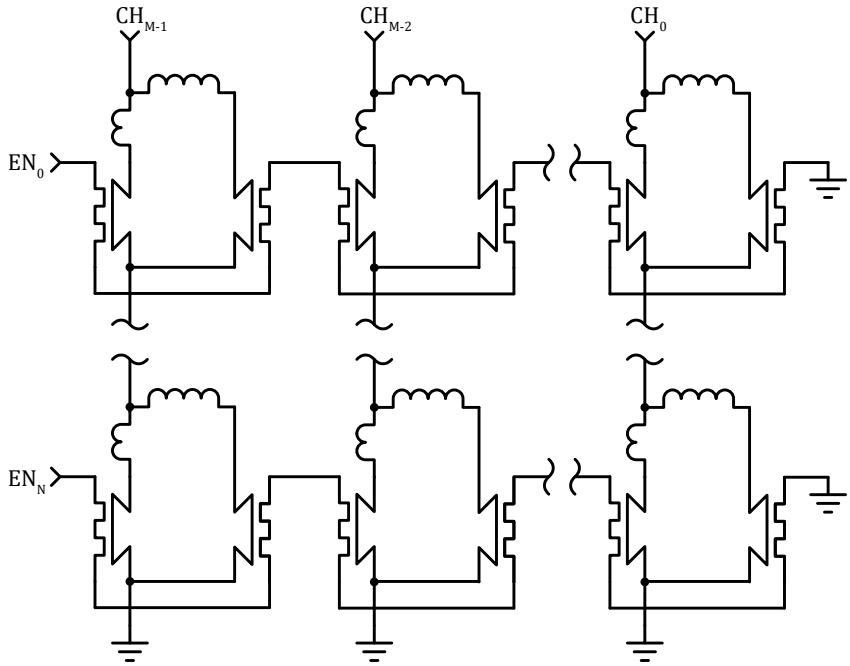

Figure 2-2: Simplified schematic of a memory array where cells are tiled to form an M-bit wide by N-bit deep array. To form virtually any sized array we simply tile the design and connect the heaters in rows and the channels in columns.

2.1.1 Writing to the memory cell

From a high level, writing to the memory consists of presenting the word to be written to the columns of the array, then asserting the enable line of the row to be written to. In terms of the cell operation, the state of the cell is encoded in the persistent current I_{p} which circulates in the loop. Thus, to write to the cell, we need a means of inducing a desired persistent current. This is achieved by applying a write bias I_{W} to the column channel line CH. This current flows through all the cells within the column but is chosen to have a small enough amplitude that it does not alter the state of any cells within that column. Each column is supplied a separate write bias depending on the desired state of the cell to be written to in that column. The row to be written to is selected by applying a write enable current I_{WE} to the desired row EN line. This weakens the channels of the hTrons within the row, thereby allowing the write bias to cause switching within the row. Thus, we have a means of writing to the array in a word-by-word access scheme.

We break the write process into three steps, as shown in figure 2-3. (1) The write procedure begins with applying the write bias I_{W} to the column line I_{ch}. The sign of this current determines the state to be written to the cell. This figure shows the procedure for writing a "1" to the cell. For writing a "0", a negative write bias is used. The write bias current flows through all the cells within the column. The amplitude of I_{W} is selected to be low enough that regardless of the state of the cell, no switching will occur, thereby preserving the state of unselected cells. The current splits between the right and left hTrons based on the inductance ratio $L_{\mathrm{L}} : L_{\mathrm{R}}$. (2) The second step involves three operations that occur in rapid succession based on the physical properties of the device. (2.1) The desired row is selected by applying the enable current I_{WE} to the row select line I_{en}. (2.2) With the induced suppression of the hTron channel switching current, the write bias through H_{L} is now sufficient to switch the channel, thereby redirecting the majority of the applied bias to the right hTron H_{R}. Thus, it is required that the right hTron's critical current, when suppressed, is high enough that the write bias will not cause it to switch when H_{R} carries the full current. (2.3) When the heater bias is removed and H_{L} returns to the superconducting state, flux is now trapped within the loop. (3) The write bias is removed, thus leaving a persistent current stored in the loop, the direction of which encodes the state of the cell.

Writing to the cell with a bipolar bias is the simplest means of obtaining two states within the cell (two distinct circulating currents). When a bipolar write bias is used, it is possible to operate an isolated cell without the need for the application of any heaters. However, operating the cell in this manner would prove particularly difficult when incorporated into an array. Thus, writing to the cell requires the application of two signals. One signal is the write enable I_{WE}, which is applied to the EN port and lowers the switching current of the hTrons. In an array, this signal spans the rows, thereby selecting all bits in the desired word. The other signal is the write bias I_{W} which is applied to the channel CH. When the write bias is applied to the cell it will inductively split between the left and right paths. In order to trap some current in the loop, we need one side of the loop to switch before the other. Such an asymmetry

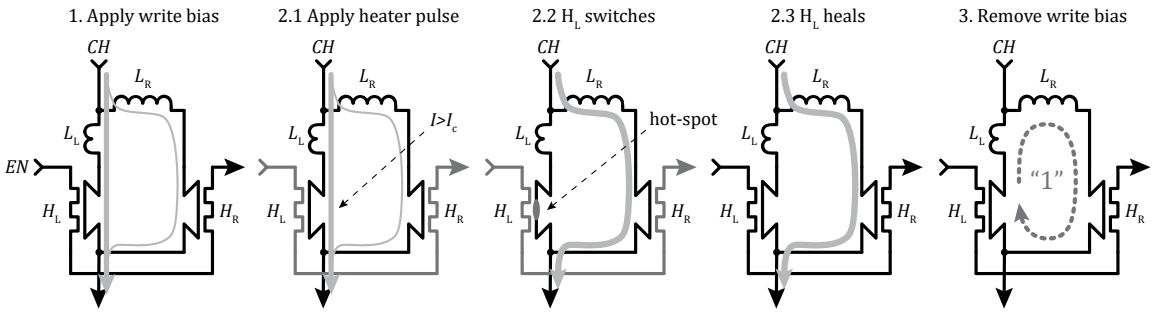

Figure 2-3: Write procedure for a cell within an array. The write process can be broken into three steps. Fist, the write bias is applied. Second, the enable line is asserted. Once this occurs, the cell undergoes a rapid evolution of the weaker hTron (in this case the left) switching, diverting current to the opposing hTron, and healing again. This process traps flux within the loop. Finally, all the applied signals are removed and the cell is left with a new persistent current I_{p}. In this figure the cell is written to the "1" state. In order to write the cell to the "0" state, the sign of the write bias is reversed, thereby trapping a current circulating in the counterclockwise direction.

can be achieved by either of the two hTrons H_{L} and H_{R} having dissimilar switching currents $I_{\mathrm{c,H_L}} \neq I_{\mathrm{c,H_R}}$, or having dissimilar left and right branch inductances $L_{\mathrm{L}} \neq L_{\mathrm{R}}$, or a combination of both. As we will see later, we will need both $I_{\mathrm{c,H_L}} \neq I_{\mathrm{c,H_R}}$ and $L_{\mathrm{L}} \neq L_{\mathrm{R}}$ in order to read out the cell's state.

Given that the loop branch inductances and hTron switching currents must be asymmetric, we can arbitrarily choose $L_{\mathrm{L}} < L_{\mathrm{R}}$. Therefore, when a bias is applied to the loop, the majority of the current will flow through L_{L}. Thus, we need H_{L} to have a lower switching current than H_{R}. This is because, during a write, we need H_{L} to switch, but H_{R} must remain in the superconducting state. Let us define $I_{\mathrm{c,H_L}}(I_{\mathrm{en}})$ to be the switching current of H_{L} as a function of the enable line current I_{en}, and likewise $I_{\mathrm{c,H_R}}(I_{\mathrm{en}})$ for H_{R}. Thus, in order to write to the memory, we need the selected cell switching currents to be $I_{\mathrm{c,H_R}}(I_{\mathrm{WE}}) > I_{\mathrm{c,H_L}}(I_{\mathrm{WE}})$. In order for the right constriction to not be switched during a write, and assuming a worst-case zero retrapping current, we need the right constriction to be able to carry the full write bias $I_{\mathrm{c,H_R}}(I_{\mathrm{WE}}) > I_{\mathrm{W}}$.

A memory array is formed by arranging the cells in a two-dimensional array as shown in figure 2-2 (c). Thus, the cell must be designed such that when the write bias is applied to a column in order to write to a selected cell, other cells in the

column must not be affected. For this to be the case, we need the superposition of the persistent current and the inductively divided bias current to be less than the unsuppressed hTron channel switching current.

For the sake of our analysis, we will arbitrarily define the "1" state to be when a positive write current is used, which results in the persistent current flowing in the clockwise direction, and the "0" state to be the opposite. There are four cases we must consider in this analysis, corresponding to a "1" write and a "0" write which occur while the cell is in either the "1" or "0" state. If we consider the case when the persistent current $I_\mathrm{p} = \pm I_\mathrm{P}$, and $I_\mathrm{w} = \pm I_\mathrm{W}$, then during a read, the current through the channel of H_R is $I_{\mathrm{H_R}} = I_\mathrm{p} + I_\mathrm{w} L_\mathrm{L}/(L_\mathrm{L} + L_\mathrm{R})$. We find that the magnitude of the hTron channel current expression simplifies to

$$|I_{\mathrm{H_R}}| = \left| I_\mathrm{W} \frac{L_\mathrm{L} \pm L_\mathrm{R}}{L_\mathrm{L} + L_\mathrm{R}} \right|. \tag{2.1}$$

The two extreme cases which could lead to an undesired switching of the hTron would be $|I_{\mathrm{H_R}}| = I_\mathrm{W}$. Clearly, we do not need to be concerned about this condition since I_W will pass through the right hTron during a write procedure. Since this current can be carried by the hTron when the enable signal is asserted, then the hTron must be able to carry this current when the gate is not asserted. Performing a similar calculation for the left-hand hTron H_L, leads to the expression $I_{\mathrm{H_L}} = -I_\mathrm{p} + I_\mathrm{w} L_\mathrm{R}/(L_\mathrm{L} + L_\mathrm{R})$, and again using the inductive splitting ratio, we find that the magnitude of the hTron current is

$$|I_{\mathrm{H_L}}| = \frac{2 I_\mathrm{W} L_\mathrm{R}}{L_\mathrm{L} + L_\mathrm{R}}, \tag{2.2}$$

along with the trivial solution $I_{\mathrm{H_L}} = 0$ which will not result in switching. Thus, we need the critical current of the left hTron with no bias to be $I_{\mathrm{c,H_L}}(0) > 2 I_\mathrm{W} L_\mathrm{R}/(L_\mathrm{L} + L_\mathrm{R})$.

The memory can be operated with only unipolar biases if desired. In such a system, one state of the memory would correspond to a circulating current of $I_\mathrm{p} \approx 0$. This state can be achieved by applying a zero write bias $I_\mathrm{W} = 0$, and applying a high write enable current. Doing so will reduce the switching current of both hTrons

to such a low value that nearly all the current stored in the loop will be dissipated. Thus, it is possible to avoid a bipolar write bias while also maintaining the ability to store two distinct states. The drawback of a unipolar scheme is that the margins are somewhat reduced, as instead of reading the difference between a high positive and negative I_p, we instead need to read the difference between a similarly positive (or negative if desired) I_p and $I_p \approx 0$. The upside of this approach is that it can be extended such that, to within flux quantization, noise, and system margins, an arbitrary number of states can be obtained – as demonstrated in [48, 49].

2.1.2 Reading from the memory cell

The mechanism by which the memory cell is read out is the main distinguishing feature between this memory and our prior work [3]. In order to simplify the integration of the new memory cell into an array, it was necessary to remove the yTron port, and only use what was originally referred to as the write port, for both write and read operations. In order to achieve this, the memory cell is read out in a similar manner as a SQUID is used to measure a magnetic field. That is, the switching current of the entire cell is measured, and used to determine the state of the cell.

It should be noted that unlike some SFQ circuits, the device is immune to the effects of low-level external fields due to the high kinetic inductance and simultaneously low geometric inductance (and as a result possible mutual inductances) of our devices. Additionally, flux trapping during cool-down is not an issue as the initial state of any cell does not affect the operation of the memory, as a cell must be written to before it is read out, and there are no non-storage superconducting loops in the design.

There are two operating regimes for the memory. The regime we find ourselves operating depends on which of the two hTrons switches first during a read. From a system point of view, this translates to whether the memory operates in an inverting mode or non-inverting mode – that is does a voltage present during a read indicate a "1" state (non-inverting) or a "0" state (inverting). Whether the memory is operating in the inverting or non-inverting regimes depends on the inductance and switching

current ratios. First, we will consider the inverting regime, as this was the first regime we designed the memory to operate in [3, 5]. After that, we will present the non-inverting regime, which is the regime we will utilize for the memory presented later.

Inverting readout

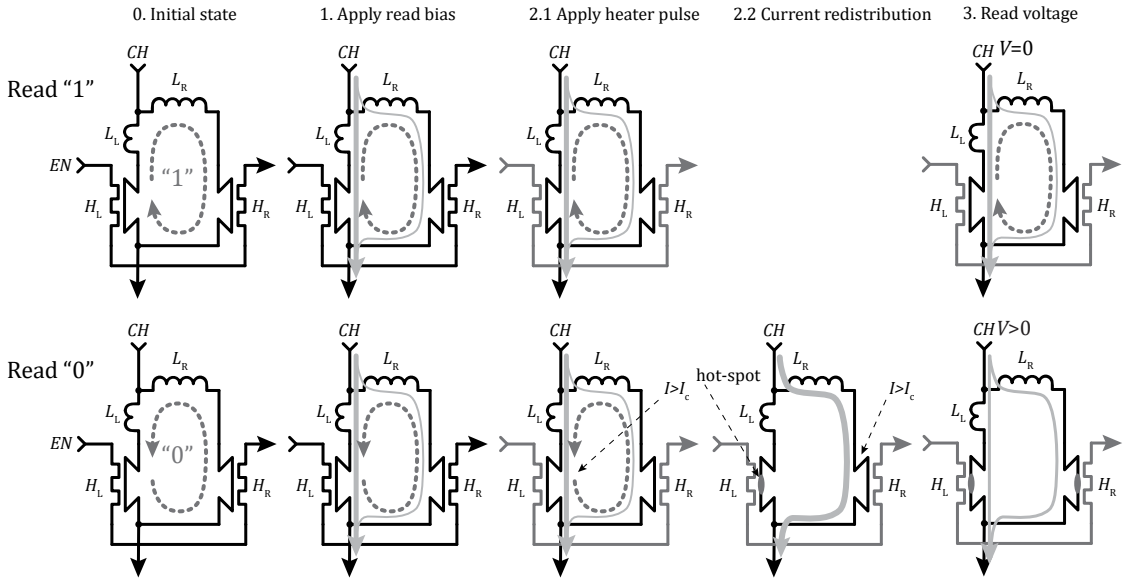

Figure 2-4: Read process for a non-inverting cell within an array. The basic read operation consists of the application of a read bias to the column, and selecting the row to be read by applying an enable signal to that row. If a selected cell is in the "1" state, then no voltage is seen at the top of that column. On the other hand, if the cell is in the "0" state, then the hTrons switch and a voltage is seen at the top of the column.

The read procedure involves three steps (using the numbering from figure 2-4): (1) application of the read bias, (2) application of the enable signal, (3) reading of the memory state. This process is shown in figure 2-4. First, we will assume that the cell is in the "1" state (read "1"). With a "1" as the initial state, the current circulates in the clockwise direction. To begin the process, the read bias I_{R} is applied to the column port CH. Similar to the write operation, this current is insufficient to cause switching in any unselected row. This current splits between H_{L} and H_{R} based on the inductance ratio $L_{\mathrm{L}} : L_{\mathrm{R}}$. This applied bias adds with the persistent current, which

in the case of a "1" results in a reduced current flowing through H_{L}, and a current that is slightly larger than the persistent current flowing through H_{R}. The desired row is then selected by applying a current I_{RE} to the row line EN, see (2) in figure 2-4. This current causes a suppression in the switching current of the hTrons within the row. However, due to the sign of the persistent current, neither the left or right hTrons will switch. To read the state of the cell, the voltage at the top of the column is read, see (3) in figure 2-4. In the case that the cell is in the "1" state, no switching occurs, so zero voltage is seen. For this state, the read bias is removed and there is no change in the cell state before and after the read has occurred. Now, we will consider what occurs when the cell is in "0" state. First, the same read bias I_{R} is applied to the column port CH, see (1) in figure 2-4. Again, this current splits between H_{L} and H_{R}; however, this time the persistent current sums to form an increased current through H_{L} and a reduced current through H_{R}. The desired row is then selected by applying a current I_{RE} to the row line EN, see (2.1) in figure 2-4.

In the "0" state case, the suppression of H_{L} combined with the increased current flow through the left side will result in the switching of H_{L}. The redistribution of the read bias results in nearly the full bias being applied to H_{R}, thus causing it to switch too, see (2.2) in figure 2-4. Provided the nanowires are latching, a stable state is entered where a constant voltage is sustained across the cell [50]. Reading the voltage at the top of the column, we find that it is non-zero, thus indicating the cell was in the "0" state, see (3) in figure 2-4. Upon the removal of the read bias, the cell is in a new state. Thus, a read causes the loss of the cell's prior state. Similar to the write procedure, it should be noted that if an isolated cell is operated, then no heater is required, but larger signal levels are needed.

The read out procedure using a pulsed access scheme (as would be done in a typical computer memory application) is summarized in figure 2-4. Consider a cell that is retaining one bit of information encoded as a persistent current I_{p}. Following our arbitrary decision on the definition of a "1" and a "0", if the persistent current is circulating clockwise, then the bit is in the "1" state and $I_{\mathrm{p}} = I_{\mathrm{P}}$. If the cell is in the "0" state, then the persistent current circulates counterclockwise, and $I_{\mathrm{p}} = -I_{\mathrm{P}}$.

During a read, the read bias is applied to the memory channel $I_{\mathrm{ch}} = I_{\mathrm{R}}$. The read bias is a positive current higher than the write bias, $I_{\mathrm{R}} > |I_{\mathrm{W}}| > 0$. Under these conditions during a read, the current in the channel of the left hTron is $I_{\mathrm{H_L}} = -I_{\mathrm{p}} + I_{\mathrm{R}} L_{\mathrm{R}}/(L_{\mathrm{L}} + L_{\mathrm{R}})$, and the current in the right hTron channel is $I_{\mathrm{H_R}} = I_{\mathrm{p}} + I_{\mathrm{R}} L_{\mathrm{L}}/(L_{\mathrm{L}} + L_{\mathrm{R}})$.

Since the loop is designed such that $I_{\mathrm{c,H_L}} < I_{\mathrm{c,H_R}}$, and from the above expression, we expect that the loop will only be able to withstand a smaller channel bias (read current) when the cell is in the "0" state. We expect this operation since the left constriction, which can withstand less current, must carry both the persistent current and the read bias, which in the "0" state both flow in the same direction. Thus, the left constriction will switch, and this leads to the read bias being diverted to the right constriction. The right construction cannot withstand the entire read bias and also switches. With both constrictions switched, a voltage will be present across the cell. For this to be the case, we need the suppressed hTron switching current to satisfy the inequality

$$-I_{\mathrm{P}} + I_{\mathrm{R}} \frac{L_{\mathrm{R}}}{L_{\mathrm{L}} + L_{\mathrm{R}}} < I_{\mathrm{c,H_L}}(I_{\mathrm{RE}}) < I_{\mathrm{P}} + I_{\mathrm{R}} \frac{L_{\mathrm{R}}}{L_{\mathrm{L}} + L_{\mathrm{R}}}, \tag{2.3}$$

where the read enable current I_{RE} is applied to the enable port, that is $I_{\mathrm{en}} = I_{\mathrm{RE}}$. The read enable current can be the same as the write enable current, and in all the experiments presented here $I_{\mathrm{RE}} = I_{\mathrm{WE}}$. Simultaneously, we require that the right hTron can carry the read bias during a "1" read and will be switched during a "0" read. That is, we require

$$I_{\mathrm{P}} + I_{\mathrm{R}} \frac{L_{\mathrm{L}}}{L_{\mathrm{L}} + L_{\mathrm{R}}} < I_{\mathrm{c,H_R}}(I_{\mathrm{RE}}) < I_{\mathrm{R}}. \tag{2.4}$$

Using these equations to govern the design of the memory, we have a space of valid designs that will, with varying margins, yield an operable cell.

Non-inverting readout

The second main operating regime of the nMem is the non-inverting readout. In this regime, we aim to switch the memory when it is in the "1" state and have the cell remain superconducting when in the "0" state. The read procedure, as far as

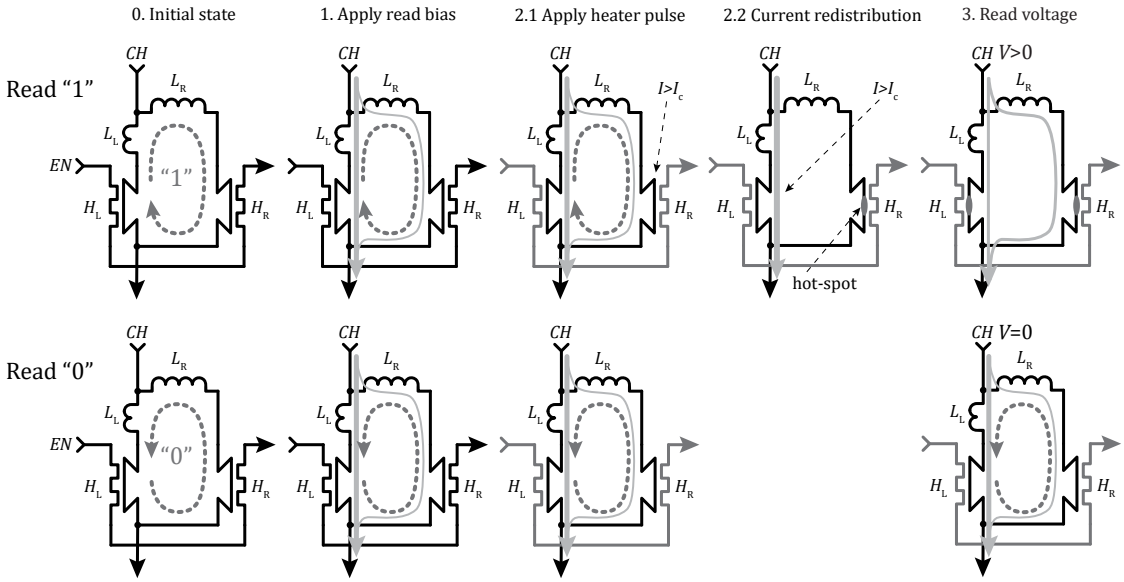

Figure 2-5: Read process for a non-inverting cell within an array. The process and evolution of cell state for the non-inverting memory cell is similar to that for an inverting cell (see figure 2-4). Again, to read the cell, a read bias is applied to the column, and the row to be read is selected by asserting the associated row-enable signal. If the selected cell is in the "1" state, then a voltage is seen on that column. On the other hand, if the selected cell is in the "0" state, then the cell remains superconducting and no voltage is seen at the top of the column.

the signals applied, is identical to the inverting readout. The only difference is the external circuitry no longer needs to invert the result of the read. While this is a trivial operation of the external controller to perform, the operation in the non-inverting regime alters the switching behavior of the memory and can lead to improved error ratios. The read procedure for the non-inverting regime is summarized in figure 2-5.

Let us consider when the cell is in the "1" state and $i_{\mathrm{p}} = I_{\mathrm{p}}$. Thus, after the application of the read bias I_{re} to the channel I_{ch}, the left and right branch currents are $i_{\mathrm{L}} = I_{\mathrm{re}} L_{\mathrm{R}}/(L_{\mathrm{L}} + L_{\mathrm{R}}) - I_{\mathrm{p}}$, and $i_{\mathrm{R}} = I_{\mathrm{re}} L_{\mathrm{L}}/(L_{\mathrm{L}} + L_{\mathrm{R}}) + I_{\mathrm{p}}$. In order for the memory to switch in this condition and not in the "0" state, we require the right hTron to switch $i_{\mathrm{R}} > I_{\mathrm{c,H_R}}(I_{\mathrm{RE}})$. Further, we require that after the right hTron switches the read bias, after redistributing, will switch the left hTron. That is, we need the read bias, less the retrapping current, to be sufficient to switch the left hTron, so $I_{\mathrm{re}} - I_{\mathrm{r,H_R}}(I_{\mathrm{RE}}) > I_{\mathrm{c,H_L}}(I_{\mathrm{RE}})$. We can combine these two expressions to find

the loop switching current in the "1" state as

$$I_{\mathrm{c,ch,\text{“}1\text{”}}} = \max\left(\left(I_{\mathrm{c,H_R}}(I_{\mathrm{RE}}) - I_{\mathrm{p}}\right) \frac{L_{\mathrm{L}} + L_{\mathrm{R}}}{L_{\mathrm{L}}}, I_{\mathrm{r,H_R}}(I_{\mathrm{RE}}) + I_{\mathrm{c,H_L}}(I_{\mathrm{RE}}) \right). \tag{2.5}$$

For this expression to be the critical current of the loop in the "1" state, it is necessary to prove that the memory will not switch by means of the left hTron switching prior to the right hTron. A more strict case of this requirement is to show that the memory will switch at a value higher than the above when in the "0" state, that is $I_{\mathrm{c,ch,\text{“}0\text{”}}} > I_{\mathrm{c,ch,\text{“}1\text{”}}}$. Thus, we will proceed to make this determination next.

When the loop is in the "0" state we have $i_{\mathrm{p}} = -I_{\mathrm{p}}$ giving us $i_{\mathrm{L}} = I_{\mathrm{re}} L_{\mathrm{R}}/(L_{\mathrm{L}} + L_{\mathrm{R}}) + I_{\mathrm{p}}$ and $i_{\mathrm{R}} = I_{\mathrm{re}} L_{\mathrm{L}}/(L_{\mathrm{L}} + L_{\mathrm{R}}) - I_{\mathrm{p}}$. Now, if we consider that we expect the left hTron to switch first, then we have $i_{\mathrm{L}} > I_{\mathrm{c,H_L}}(I_{\mathrm{RE}})$. After this switch occurs, we also need the current to be sufficient to switch the right hTron. Thus, we have the simultaneous requirement of $I_{\mathrm{re}} - I_{\mathrm{r,H_L}}(I_{\mathrm{RE}}) > I_{\mathrm{c,H_R}}(I_{\mathrm{RE}})$. So, the channel switching current will be

$$I_{\mathrm{c,ch,\text{“}0\text{”}}} = \max\left(\left(I_{\mathrm{c,H_L}}(I_{\mathrm{RE}}) - I_{\mathrm{p}}\right) \frac{L_{\mathrm{L}} + L_{\mathrm{R}}}{L_{\mathrm{R}}}, I_{\mathrm{r,H_L}}(I_{\mathrm{RE}}) + I_{\mathrm{c,H_R}}(I_{\mathrm{RE}}) \right). \tag{2.6}$$

Again, for this expression to be the critical current of the loop in the "0" state, it is necessary to prove that the memory will not switch by means of the right hTron switching prior to the left hTron. This and the prior condition are satisfied if $I_{\mathrm{c,ch,\text{“}0\text{”}}} > I_{\mathrm{c,ch,\text{“}1\text{”}}}$. In fact, in the condition that $I_{\mathrm{c,ch,\text{“}0\text{”}}} < I_{\mathrm{c,ch,\text{“}1\text{”}}}$ we will have an inverting readout memory.

2.1.3 Array architecture

The design of the nMem cell was specifically engineered to facilitate the formation of an array. This is achieved by tiling the cells such that all channels within the column are connected in series and all heaters in a row are connected in series. Thus, a word is accessed by enabling the heater on that row and reading from or writing to all columns simultaneously.

2.2 Physical construction of the memory cell

The physical construction of the memory is critical to the device operation. The memory operation is nearly entirely governed by the switching current and inductance ratio. These parameters are totally dependent on the layout of the cell. However, there is some flexibility in the actual layout, provided the correct ratios are achieved. This section begins with the basic device layout. We then cover the new methods of simulating inductive splitting we developed for the memory development. Finally, we present the device simulation and parameter extraction that allows us to assess the operating regime of the memory.

2.2.1 Device layout

The layout of the memory cell is where the device intricacies come into play. Conceptually, the basic memory cell shown in figure 2-1 is relatively simple to realize; however, in practice, there are many effects at play that must be given the appropriate treatment. As set out in section 2.1, the operation of the memory cell is almost entirely governed by the switching current ratio and the inductance ratio – with arrays relying on the unselected cells withstanding all bias currents to other cells in the column. The switching current and inductance ratios are governed by geometry and are modulated by the enable line heaters. This means that, for a cell to work, the layout must be very carefully constructed.

For the majority of the analysis of the memory to this point we considered the switching current and kinetic inductance to be constant value. However, in reality these values vary with current and temperature. For the array operation of the nMem we rely on the fact that a hTron channel switching current decreases with increased gate current (temperature). In our past work on superconducting memories we considered the varying in the kinetic inductance with temperature to be a negligible effect [3–5]. This assumption comes from the fact that the kinetic inductance varies with both the bias current and temperature. To-date the majority of devices that utilize variations in the kinetic inductance use the fact that the inductance varies

with current. This effect results in a relatively minor increase in kinetic inductance with increasing bias current – typically quoted to be around a 10 % to 20 % increase in kinetic inductance between zero bias and a bias just below the switching current. In theory, this effect should result in the kinetic inductance increasing exponentially as we approach the depairing current; however, since in practice we can never reach the depairing current, typically witnessing switching at around 80 % of that value, we see only minor increases in kinetic inductance. This is not true for increasing temperature. In terms of temperature, we can operate a device to within very close margins of the critical temperature. For the memory cell, this means that we can see large increases in the kinetic inductance in hTron channels at higher gate currents. It is important when designing with hTrons to consider that they are not simply a nanowire with variable switching current, but also a variable inductor. By harnessing this effect, we can realize devices that utilize variable inductors to perform interesting operations, such as microwave filters – see chapter 3.

For the memory cell, this means that when designing the memory we must take into account the fact that the hTrons on each of the left and right branch have different geometries, and thus will experience a different increase in kinetic inductance for the same gate current to the hTrons. Given that the temperature increase in a hTron channel is not perfectly localized to beneath the heater, we resort to 3-D simulations of the device to validate and iterate on the design. In order to generate a first design upon which we can, through insights from the simulation results, iterate on, we consider the heat to be only localized to beneath the heaters.

First, we will choose a loose set of initial parameters from which to start the design. Informed by our analysis from section 2.1 and our results from [3, 5], the initial cell was designed for a switching current ratio around $I_{c,H_L} : I_{c,H_R} = 1 : 2$, and an inductance ratio around $L_L : L_R = 4 : 9$. We will choose the smaller of the two channels to be 200 nm wide. We will design the heater to be the same for both sides, so then the channel of the larger hTron must be 400 nm. Now, this means that we have, for a given heater width, twice the area heated on the left (narrower channel) than on the right (wider channel). Thus, the inductance of the left hTron will

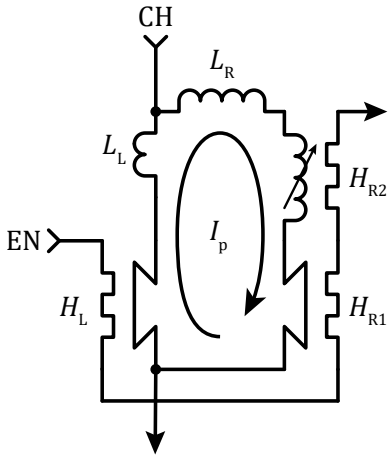

Figure 2-6: Simplified schematic of the memory cell with kinetic inductance compensation. The additional hTron H_{R2} is shown here as a variable inductor as it not intended to switch and it only contributes a variable inductance to the cell. The original right hTron that switches is renamed H_{R1} for clarity.

increase, with increasing temperature, at twice the rate of that of the right channel. There are a number of methods to address this issue. For example, we could use asymmetric heaters on each side, add more hTrons to one side, and/or change the shape of the heater. Here we elect to add a second hTron to the right-hand side to equalize the number of squares heated on both sides – see figure 2-6. This alone would result in a reduction in the inductance ratio $L_L : L_R$ as temperature increases, with $L_L : L_R$ approaching $1 : 1$ at high T. To counteract this effect and make the high branch inductance increase even faster than the left with increasing gate current, the heaters on the right are made slightly longer, so as to impart more heat to the channels, and hence see a higher temperature. This will result in a departure from our designed $I_{c,H_L} : I_{c,H_R}$ ratio. In order to satisfy both requirements a compromise between inductance ratio and switching current ratio was struck.

Since we have two hTrons on the right-hand side of the loop, we now have two possible locations for the device to switch. It is likely that, in a fabricated cell, one of these devices will have a lower switching current compared to the other, but the behavior of the device could be unpredictable. To eliminate any source of variance, a choke was added to one hTron channel. This choke is realized by adding two triangular

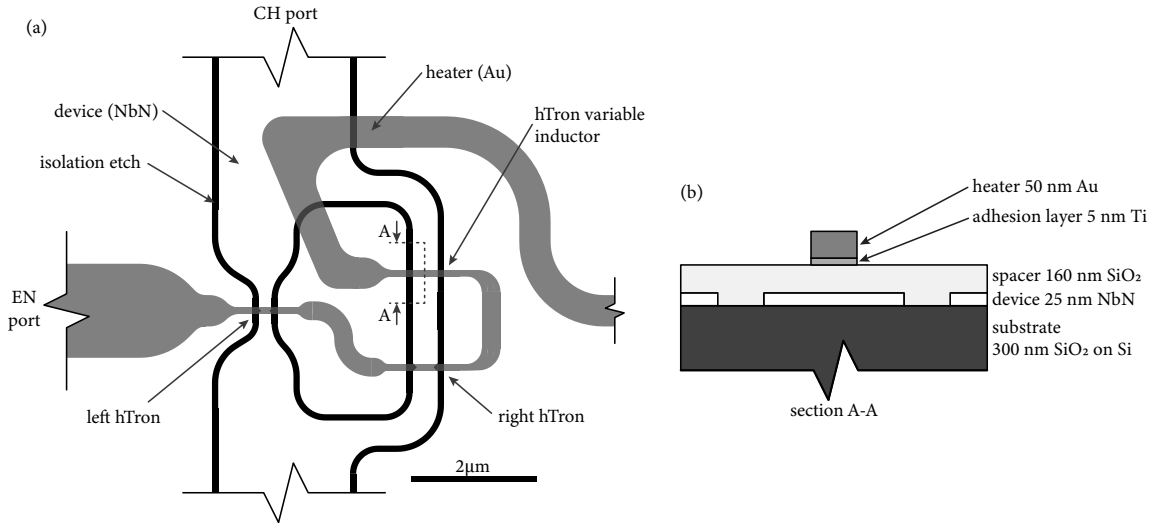

Figure 2-7: The layout and layer stack-up of the memory device. This is the exact layout and stack up of the final device simulated, fabricated, and tested in the following sections. (a) The layout of the memory cell with the black areas indicating where the NbN has been etched revealing the substrate below, while all other areas are left covered in NbN. The area around the device is grounded, and the island in the center of the memory is left unconnected. The gray layer is the heater material. The heater is shown as transparent such that the details of the hTron chokes can be seen. The heater is separated from the NbN layer by means of a silicon dioxide spacer. (b) The layer stack up (not to scale). The device is fabricated on a silicon wafer with a thermal oxide upon which NbN was deposited by the process presented in [51]. The NbN was patterned and etched to form the device layer. Then a plasma-enhanced chemical vapor deposition (PECVD) silicon dioxide spacer is deposited. On top of this spacer a thin titanium adhesion layer and thicker gold heater layer is fabricated by an evaporation and liftoff process.

notches to either side of the device, thereby reducing its width. It should be noted that in the layout the choke is drawn as a sharp point, but in a fabricated device the tips would be rounded due to a number of effects. The choke was selected to reduce the width by 40 %. This will throw off our switching current ratio $I_{c,H_L} : I_{c,H_R}$, so to compensate, the channel of the left hTron was similarly reduced to 40 % of its initial value. Thus, we have a choke width of 120 nm on the left and 240 nm on the right. The shape of both chokes was made to be similar so that the rounding of the tip in the fabricated device will result in a preservation of the switching current ratio.

The heater width over the channel was designed to be 100 nm and is made wider in non-active regions to reduce unwanted heating and power loss. Now, the fact that we have three hTrons on a single memory cell, which is in effect a loop, means that it is impossible to wire all the heaters in series without a fourth crossing of the loop. To make the fourth crossing without forming another hTron, the crossing was made in a location where the channel was wide (to eliminate the chance of switching the wire) and the heater was made wide (to reduce the temperature increase). With these design constraints in hand, and with standard concerns to current crowding, the cell layout shown in figure 2-7 was designed.

The interaction of the thermal with the electrical characteristics of the memory cell are difficult to predict accurately without the aid of a simulation. In particular, the effects of the temperature on the switching current and kinetic inductance of each branch require a simulation to accurately capture these effects. Thus, a method of simulating the thermal properties of the memory was required.

2.2.2 Simulating inductive splitting

Inductive splitting, which is central to the operation of the memory, is not well-defined in the static limit. This naturally stems from the fact that an inductor at DC presents no impedance. Thus, in a static scenario there are infinite solutions in a non-trivial inductive circuit. This fact presents an issue when modeling the memory, as a dynamical simulation would bring much more complexity and computation time than is necessary given that we are in fact looking at what, to an inductive circuit, is trivial

current splitting. To avoid dynamical simulations and their associated downsides, we take advantage of the fact that resistors at DC and inductors at frequencies $f \neq 0$ will split current in the same manner, albeit by different means. This substitution is enabled by the fact that the resistance of the material is presented as a resistivity and that for a given frequency the kinetic inductance can be expressed as a reactivity. That is, we can define a sheet resistance $R_{\square}$ for a normal film, and a sheet reactivity for a kinetic inductor at DC $X_{\mathrm{L}\square} = 2\pi f L_{\mathrm{K}\square}$ for a constant thickness film of sheet kinetic inductance $L_{\mathrm{K}\square}$.

For our simulation we chose the frequency f somewhat arbitrary selected since, save for parasitic capacitance and high frequency losses, there is no dependence on frequency in a purely inductive circuit in terms of current splitting ratios. So, we can say that $\tilde{R}_{\square} = 2\pi f L_{\mathrm{K}\square}$ is the pseudo sheet resistance. Given that we are performing simulations with multiple layers, we selected a full 3-D simulation. Thus, we need to determine a pseudo conductivity from our pseudo sheet resistance (that is an inverse inductivity). The pseudo conductivity is simply obtained as $\tilde{\sigma} = 1/2\pi f h L_{\mathrm{K}\square}$, where h is the film thickness.

There are a number of ways to consider this simulation's substitution of a resistivity for inductivity. First, we can simply think of it as replacing the resistance with a pseudo resistance that is equivalent to the inductance. Another way we could consider it is as us taking the complex portion of the Mattis-Bardeen equations with a scaling term [52]. Alternatively, one could consider the simulation to be conducted at a single-frequency. Regardless of how we consider the simulation, with our substitution, the currents are the same distribution (but different magnitude) than those that we would obtain by a full dynamical simulation. Since we are not concerned with the effects of bias current on inductance, we will gain the desired splitting and current distribution plots we need to assess the operation of the memory.

One downside of the use of this pseudo resistance approach is that the simulation package will take into account Joule heating within the nanowire during the superconducting phase, which is nonphysical as the inductors are (in the DC limit) dissipationless. This issue is easily mitigated by using either a scaling term to make

the pseudo resistance very small (that is reducing the frequency f), or reducing the bias current applied to the nanowire in the simulation, or both. In our simulation we performed both such that the currents in the heater (which are physical and of the correct magnitude) and the nanowire bias currents are within a few orders of magnitude. This similarity between these currents is useful for simulation convergence and accuracy. But in order to do this we need the resistance to be small such that the worse-case self heating of the nanowire is less than one one-thousandth of the heat from the heater, such that we can ignore it.

Now, to complete our pseudo resistance approach we need an expression for the kinetic inducitivty of the film. An equation for this is given in 3.1 as equation 3.14. This expression relies on knowledge of the superconducting gap Δ at the operating temperature T. However, there is no closed form solution for this value [53]. For high temperatures near T_c, [53] gives us the approximation $\Delta(T \to T_c) \approx 3.07 k_B T_c \sqrt{1-(T/T_c)}$. This approximation is good near T_c but begins to overestimate the gap as temperature decreases. In addition, we have that at $\Delta(T=0) = \Delta_0$, by definition. So, to form an approximation function for all $0 \leq T \leq T_c$, we need an expression that can transition between the high temperature approximation and the low temperature value of the gap.

Given that the high temperature approximation begins to grow much larger than the actual gap value as temperature decreases, it would seem that some kind of soft-clipping function would achieve our purpose. Other fields, such as signal processing and artificial neural networks, use various sigmoid functions to achieve this purpose [54, 55]. One common sigmoid function used for soft-clipping is the hyperbolic tangent. Now, to achieve accuracy at zero temperature, we need the argument of the hyperbolic tangent to be very large. The standard approximation, while larger than the gap for low temperature, does not grow particularly large. Thus, the introduction of a root in the denominator at zero temperature would achieve our goal, and if we ensure that our new term is unity at and near T_c, then we will not lose the accuracy of our approximation at higher temperatures. One such root could be $\sqrt{T_c/T}$, which neatly multiplies with our existing approximation to form

$\Delta(T \to T_c) \approx 3.07 k_B T_c \sqrt{(T_c/T) - 1}$, which in itself is a worse approximation to the gap, but when limited by the hyperbolic tangent we can form the approximation

$$\frac{\Delta(T)}{\Delta_0} \approx \tanh\left(\frac{3.07 k_B T_c}{\Delta_0}\sqrt{\frac{T_c}{T} - 1}\right) = \tanh\left(1.735\sqrt{\frac{T_c}{T} - 1}\right). \tag{2.7}$$

While this approximation is good and commonly used [56–59], we elected to enhance its accuracy by introducing two fitting parameters a and b to form the approximation

$$\frac{\Delta(T)}{\Delta_0} \approx \tanh\left(a\left(\frac{T_c}{T} - 1\right)^b\right). \tag{2.8}$$

To find the optimal values for our fitting parameters a and b, we need a numerical solution for the gap as a function of temperature.

From [60], the gap is the solution to the expression

$$\frac{1}{N(0)V} = \int_0^{\hbar\omega_D} \frac{1}{\sqrt{\Delta^2(T) + \xi^2}} \tanh\frac{\sqrt{\Delta^2(T) + \xi^2}}{2k_B T}\, d\xi, \tag{2.9}$$

where $N(0)$ is the density of states at the Fermi level, V is the interaction strength, ω_D is the Debye frequency, and ξ is the energy. By using the solution at $T = T_c$ we find

$$\frac{k_B T_c}{\hbar\omega_D} = \frac{2e^\gamma}{\pi} \exp\left(-\frac{1}{N(0)V}\right), \tag{2.10}$$

where γ is the Euler constant. Now, we can solve the system setup by these two equations numerically to give us values for $\Delta(T)$. The results of the numerical solution are shown in figure 2-8. The values are normalized to Δ_0 and T_c since for weakly-coupled systems, such as in our materials ($\hbar\omega_c/k_b T_c \approx 30 \gg 1$), we have the form of the gap normalized to Δ_0 and T_c is universal [60].

With the numerical solution determined, we can now find the value of a and b from equation 2.8 that will yield the best approximation. In order to determine these values, an optimizer implementing the downhill simplex method was used – see Chapter 5 of [5]. In optimization, the choice of cost function is critical to the resultant fit. Here, since we will be spanning temperatures from around $0.4T_c$ to around $0.8T_c$,

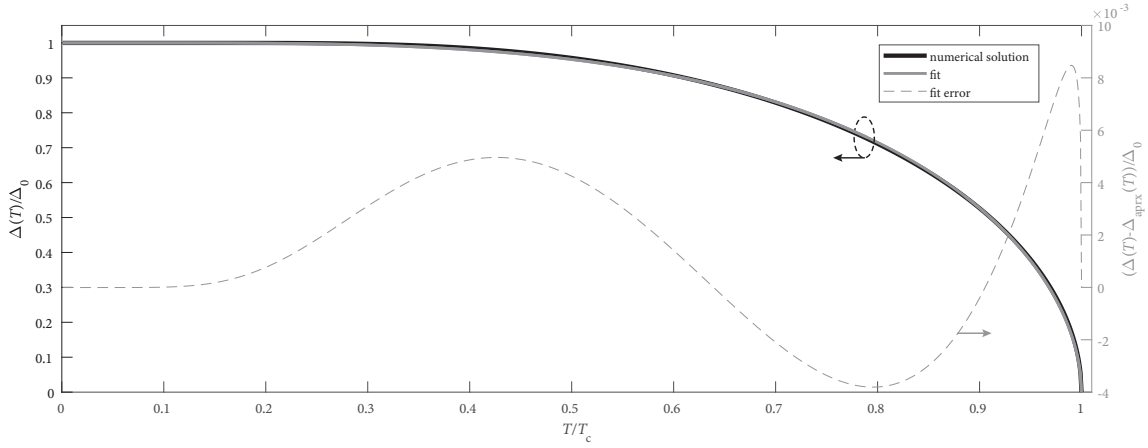

Figure 2-8: Plot of the numerical solution to the gap in our material as a function of temperature for a $T_c = 10\,\mathrm{K}$ along with a numerical fit given by equation 2.11. It can be seen that the numerical solution and the fit match extremely well with a mean square error of 10^{-5} and an absolute deviation of less than $9 \times 10^{-3}\Delta_0$ for all temperatures.

we opted to simply minimize the mean square error over the whole range from $T = 0$ to $T = T_\mathrm{c}$. Thus, the cost function $C(a,b) = \int_0^1 (\Delta(T/T_\mathrm{c}) - \Delta_\mathrm{aprx}(T/T_\mathrm{c}, a, b))^2$ was used. With the optimization complete, we find our new approximation

$$\frac{\Delta(T)}{\Delta_0} \approx \frac{\Delta_\mathrm{aprx}(T)}{\Delta_0} = \tanh\left(1.8602\left(\frac{T_\mathrm{c}}{T} - 1\right)^{0.5258}\right), \tag{2.11}$$

which we will use for all numerical simulations hereon. This approximation is plotted against the numerical solution in figure 2-8. It can be seen that the approximation gives an excellent representation of the gap, and will suffice for our simulations.

Finally, with our closed form expression of the gap, we can form our final expression for the pseudo conductivity,

$$\tilde{\sigma}(T) = \frac{\Delta_\mathrm{aprx}(T)\tanh\frac{\Delta_\mathrm{aprx}(T)}{2k_\mathrm{B}T}}{2fhR_\mathrm{sq}\hbar} \tag{2.12}$$

which combined with our approximation of the temperature-dependent gap allows us to, with high accuracy, predict the bias-current distribution within the memory. In the simulations presented here we used $f = 10^{-9}\,\mathrm{Hz}$.

2.2.3 Simulation results and cell parameter extraction

With the pseudo conductivity given by equation 2.12 in hand, it is now possible to simulate the current splitting in the memory using a static simulator. The layout and stack up from figure 2-7 were imported into COMSOL and, using the material parameters from [4], the device was simulated. In order to determine the current splitting ratio, we need to provide a probe current applied to the channel. However, as mentioned in the previous section, this will cause Joule heating, and so it must be small, but not so small that numerical errors become significant. With the low frequency selected for the pseudo conductivity, a bias of 1 nA was used and did not cause a heating in the film more than 1 mK. Thus, the contribution of the channel bias probe current to device temperature can be ignored.

Next, we needed to determine the heater bias I_{H} required to achieve our desired switching current suppression of around $I_{\mathrm{c,H}}(I_{\mathrm{H}}) = 0.25 I_{\mathrm{c,H_L}}(0)$. In order to determine this value, we need to first determine the temperature at which the channel of the hTrons is suppressed to the required value. From Ginzburg-Landau (GL) theory [53], we have that in a wire with small cross-sectional area the critical current suppression with temperature scales as

$$J_{\mathrm{c}} \propto \left(1 - \frac{T}{T_{\mathrm{c}}}\right)^{\frac{3}{2}}. \tag{2.13}$$

Thus, with a substrate temperature of $T_{\mathrm{sub}} = 4.2\,\mathrm{K}$ we need a hTron channel temperature T_{HC} that satisfies

$$\frac{J_{\mathrm{c}}(T_{\mathrm{HC}})}{J_{\mathrm{c}}(T_{\mathrm{sub}})} = \frac{\left(1 - \frac{T_{\mathrm{HC}}}{T_{\mathrm{c}}}\right)^{\frac{3}{2}}}{\left(1 - \frac{T_{\mathrm{s}}}{T_{\mathrm{c}}}\right)^{\frac{3}{2}}} = 0.25, \tag{2.14}$$

which for a nominal $T_{\mathrm{c}} = 10\,\mathrm{K}$ requires $T_{\mathrm{HC}} = 7.7\,\mathrm{K}$. Next, we varied the heater bias current I_{H} until this temperature was found in the channel. The resultant device temperatures are shown in figure 2-9.

With the simulation showing the hTron channels are at the desired temperatures, we can proceed with parameter extraction. Figures 2-10 shows the sheet kinetic

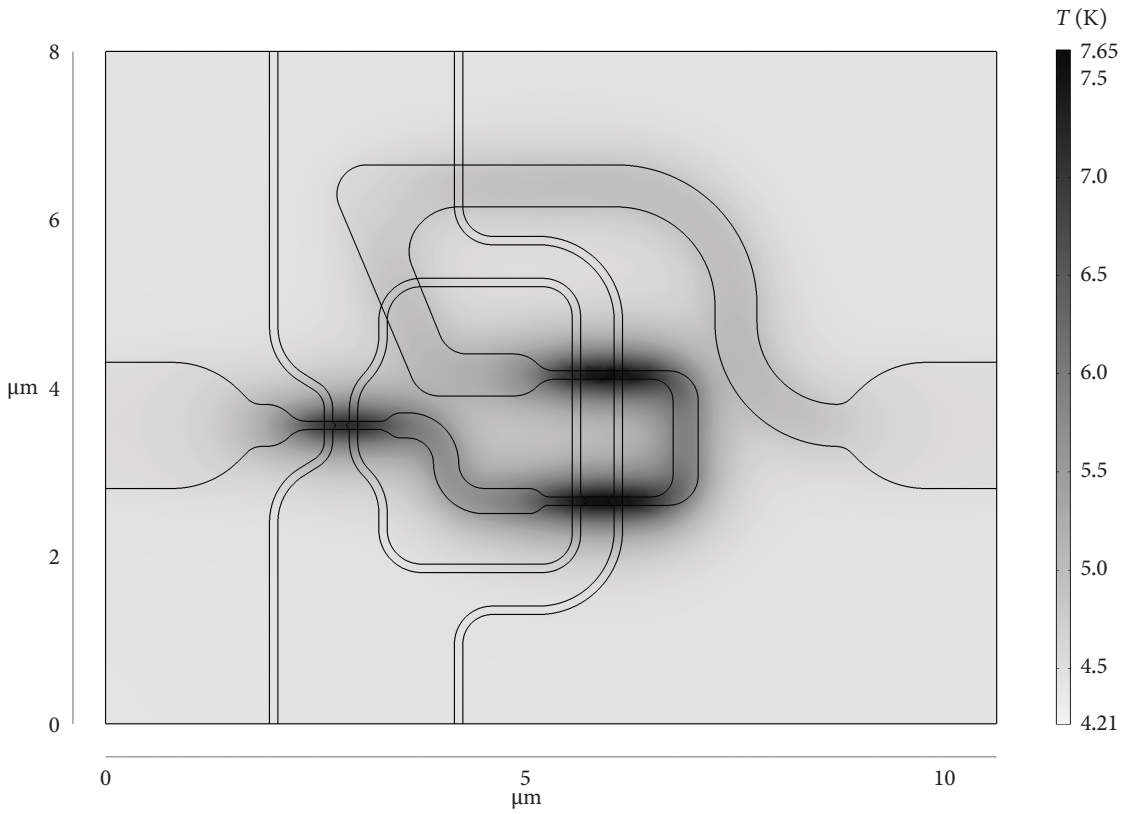

Figure 2-9: Simulation of the memory cell temperature with the enable line active. This plot shows a cross section of the memory at the device layer, and therefore the temperature of the NbN. It can be seen that the narrow portions of the heater are effective at localizing the heat to the hTrons and that the wider interconnects are heating the chip to a far lesser degree. In addition, we can see that the temperature increase on the right is slightly higher than that on the left due to the heaters being slightly longer on the right. However, there is very little overlap between the thermal response of the right-top and right-bottom heater, as desired. In this simulation the substrate temperature was $T_{\text{sub}} = 4.2\,\text{K}$.

inductance at each point in the memory. It can be seen that near the heaters the inductance is increased with a peak around twice the value elsewhere in the film. This change in the kinetic inductance, through our pseudo-conductance, allows the tool to accurately predict the current distribution assuming no flux is trapped in the loop. By integrating the current flowing through a cross section of each branch of the memory loop, we find the left branch current is $I_{\text{L}} = 0.65 I_{\text{b}}$ and the right branch current is $I_{\text{L}} = 0.35 I_{\text{b}}$, where I_{b} is the bias current applied to the channel of the memory. Rearranging these values in terms of our branch inductances we have that,

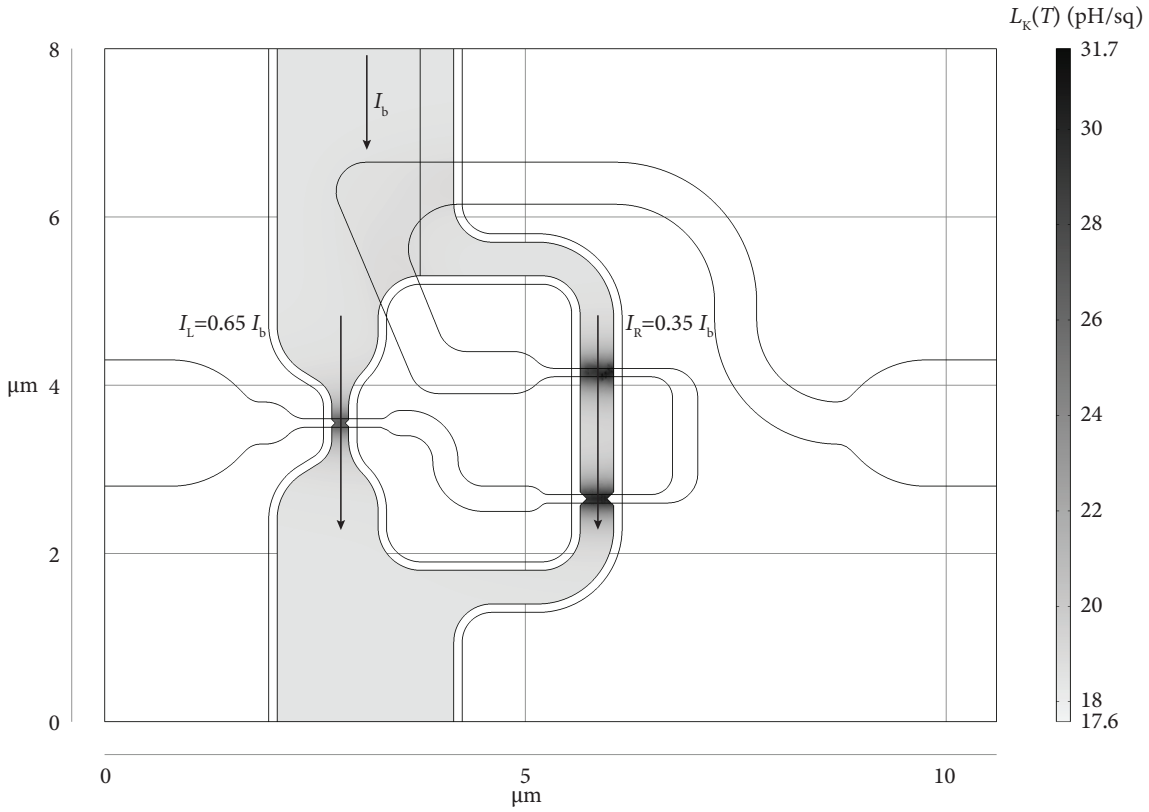

Figure 2-10: Simulation of the kinetic inductance within the memory cell with the enable line active. This simulation is a mapping of the temperature in figure 2-9 using equations 2.11 and 3.14. The current splitting ratio shown here is generated from the simulation by means of the pseudo conductivity $\tilde{\sigma}(T)$ function with a frequency of $f = 10^{-9}$ and bias current of $I_\mathrm{b} = 1\,\mathrm{nA}$. The currents in each branch were determined by splicing the memory along the horizontal axis through the loop and integrating the current density on each side. From these values we can determine the inductance ratio is $L_\mathrm{R} : L_\mathrm{L} = 13 : 7$ when the enable line is active.

when the enable line is active, $L_\mathrm{L}/(L_\mathrm{L} + L_\mathrm{R}) = 0.35$ and $L_\mathrm{R}/(L_\mathrm{L} + L_\mathrm{R}) = 0.65$, which is the equivalent of $L_\mathrm{R} : L_\mathrm{L} = 13 : 7$.

Finally, the switching current under heater suppression is extracted. This is achieved by applying equation 2.13 to the temperature data from figure 2-9. The resultant plot of switching current suppression is shown in figure 2-11. It can be seen that due to the heat spreading out as it conducts through the device, we have a relatively large region of suppression. This is another reason why we added the notches to localize the hot-spot formation to the left and right switching hTrons – as opposed to the variable inductor hTron. From these results we find that the suppression of the

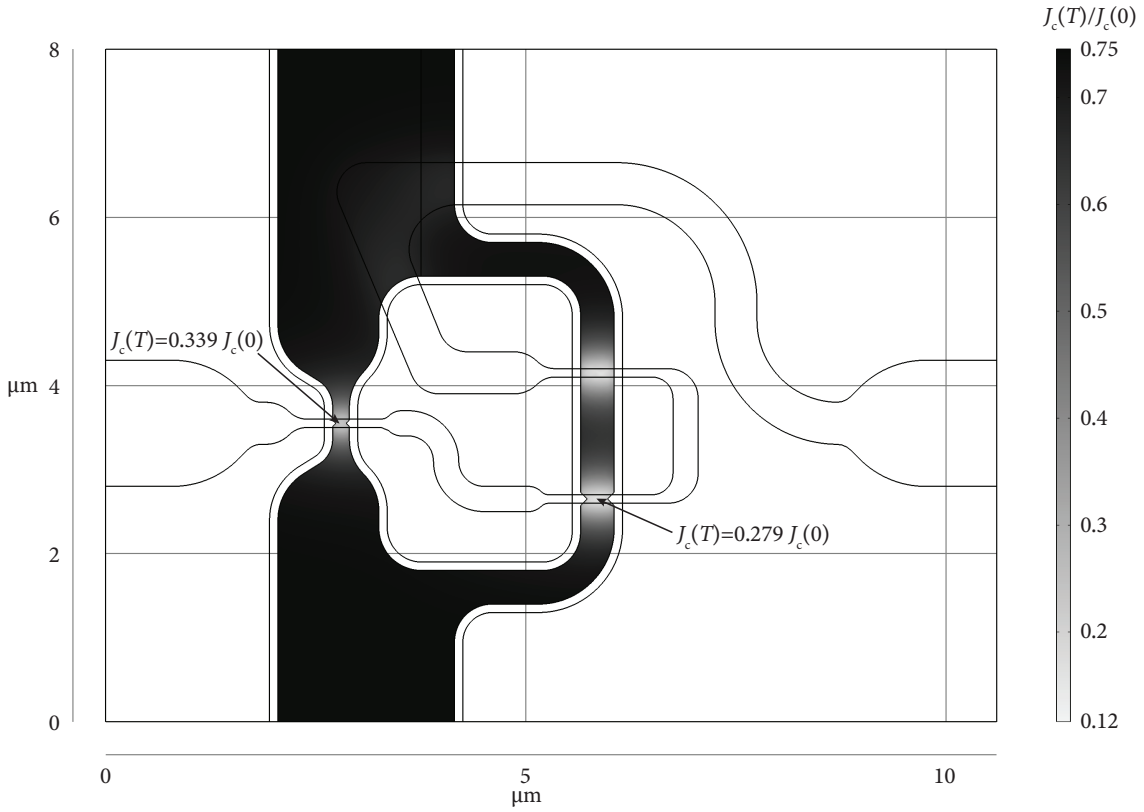

Figure 2-11: Simulation of the reduction in the critical current, or "suppression," of the critical current in the memory. It can be seen that the switching current is drastically reduced in the hTron channels. This high suppression allows for the selection of the cell, and for the read and write biases to pass the unselected cells without causing undesired switching, thereby resulting in errors. The suppression on variable inductor and right switching hTron on are almost identical. The width of the switching hTron being narrower leads it switching long before the variable inductor could switch.

left hTron is $J_{\mathrm{c,L}} = 0.339 J_{\mathrm{c}}(0)$, and for the right hTron is $J_{\mathrm{c,R}} = 0.279 J_{\mathrm{c}}(0)$. These values are close to our desired nominal suppression of $J_{\mathrm{c}} = 0.25 J_{\mathrm{c}}(0)$. In terms of the switching current ratios we find $I_{\mathrm{c,H_R}}/I_{\mathrm{c,H_L}} = 2 \times J_{\mathrm{c,R}}/J_{\mathrm{c,L}} = 2 \times 0.279/0.339 = 1.646$, with the factor of two resulting from the channel on the right hTron being twice the width of that on the left.

With the extracted parameters, we find ourselves with an operable memory in the non-inverting scheme. The operating regime is confirmed by evaluating equations 2.5 and 2.6, with the extracted parameters. From these evaluations we find that a "1" read occurs at 67 % of a "0" read, that is $I_{\mathrm{c,ch,"1"}}/I_{\mathrm{c,ch,"0"}} \approx 0.67$, when no noise or

variation is considered and the optimal persistent current is written to the loop for each state. This is a good operating point, and as such, this design was selected for the final layout of the chip.

2.3 Switching distribution

The read operation that we have considered thus far is what would be implemented in a final application – see section 5.1. The downside of this method of reading the cell is that little information about the operation of the cell can be obtained. Specifically, we can only determine the error rate, and no other statistics on the switching current distributions. In order to obtain this extra information, we need to make a measurement of the switching current during each read. This measurement can be made by using a ramp-based read out method [3, 5].

2.3.1 Experimental setup

This experiment is performed using the CryoProbe discussed in chapter 4. The memory chip was affixed to a sample PCB (see section 4.2.2) with GE varnish and baked at 70 °C for 30 minutes. Then the memory device was bonded to the pads on the PCB. The sample was then placed in the probe and the probe immersed in liquid helium. The memory operated at the base temperature of the probe (close to 4.2 K). The experimental setup for the switching distribution measurement is shown in figure 2-12.

In order to measure the switching of the memory channel, one would ideally use a current source and monitor the voltage generated across the memory; however, this is not trivial to accomplish experimentally. The difficulty arises from the fact that we are driving a relatively long coaxial cable with comparably fast signals. Thus, if we were to drive a current into the cable we would first measure the 50 Ω characteristic impedance of the cable until the reflection from the device has enough time to return to the instrument for us to see the load (device) impedance. To avoid this issue we use two cables, one to send our signal and another to detect when the switching

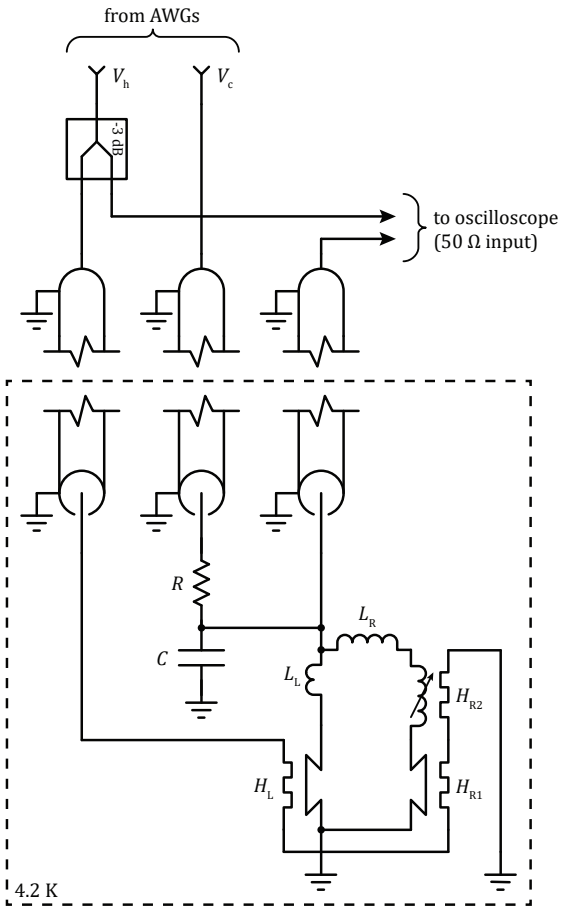

Figure 2-12: Schematic for the experimental setup for the ramp-based readout. The resistor $R = 50\,\Omega$ matches the input impedance of the channel bias line. The capacitor $C = 82\,\text{pF}$ shunts high frequency noise and effectively forms a low-pass filter with the bias resistor R. The resistor and capacitor are 0603 chip devices soldered onto the sample PCB.

occurs. The channel bias-voltage V_c is generated by an arbitrary waveform generator (AWG) and sent to the sample PCB on the cold head of the CryoProbe. On the sample PCB it is series terminated by a $50\,\Omega$ resistor. The memory is shunted by an $82\,\text{pF}$ capacitor which shunts high frequency noise – see section 5.3.3. The voltage at the top of the memory cell is then connected to a second cable that returns to room temperature and is terminated into a $50\,\Omega$ input of an oscilloscope. The heater pulse is monitored by splitting it at room temperature and sending one branch to the device and the other to the oscilloscope. No special attention is given to the heater signal at the sample-end of the cable as it is more robust to noise (being a resistive

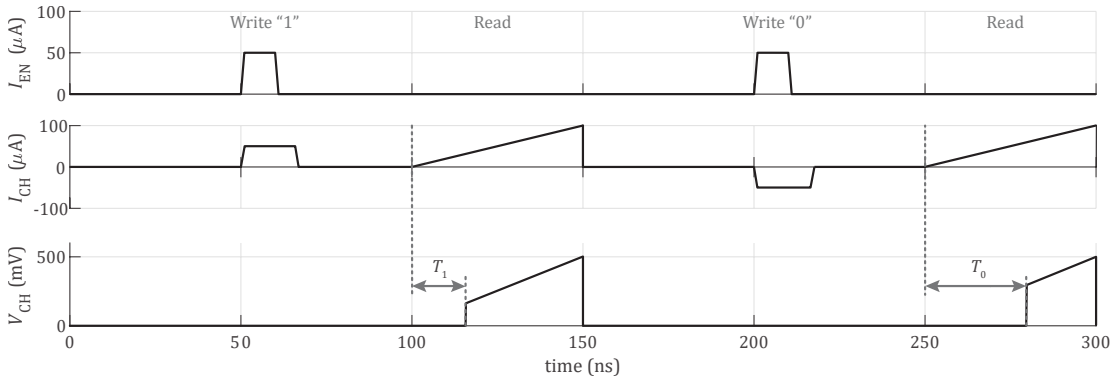

Figure 2-13: Schematic of the signals used in the cell experiments. Here the ramp-based read out scheme is used to determine the switching current distributions. This figure features two write/read cycles. The first operation, occurring at 50 ns, sets the cell to the "1" state. At 200 ns the memory is set to the "0" state. It can be seen that the write current is now a negative pulse, thus inducing negative current in the loop. The memory is read out by applying a ramping current to the channel, as seen at 100 ns and 250 ns. After the cell has been set to the "0" state, the switching current of the channel should be lower than when it was set to the "1" state, so we expect the skew times $T_0 < T_1$. These times can then be converted to a switching current measurement. Note that the amplitudes and timings shown in this figure are nominal and can be changed based on the cell design and how fast one chooses to run the memory.

heating element) and we do not need to monitor the signal at the device. All signals were de-skewed in the oscilloscope to compensate for the cable delays.

The cold bias resistor shown in figure 2-12 is a NiCr-based thin film 0603 chip resistor. This type was used as the thin-film resistors have good high frequency performance, stability, and precision. The NiCr-based composition was selected as it does not superconduct, unlike the other common compositions based on TaN. Additionally, NiCr has residual resistance ratio RRR of close to unity – meaning the resistance at our operating temperature will be very close to its room temperature value. The capacitor is an NP0-dielectric chip capacitor in an 0603 package. The NP0 dielectric was selected as it performs well down to below our operating temperature with little change in capacitance [61].

The waveforms used for the basic ramp-based read out method are summarized in figure 2-13. During the read operation, the switching current is measured by applying a ramp to the channel. At some point the channel switches and a voltage is seen at

the top of the column. The time delay between the start of the read operation and the switching of the channel is used to determine the current required to switch the channel. This was done as the time-resolution of the oscilloscope used in our experiments is greater than the voltage (effectively the applied current) resolution. Thus, we are able to determine an accurate measurement of the column's switching current. The variations in these times allows us to gain statistics on the switching current variation of the memory after being set to either state. Furthermore, the optimal value of the read current I_{R} does not need to be known prior to the experiment if the ramp-based read out is used. This method of operating the cell gives us switching current statistics that we can use to determine the operating margins of the cell.

While the waveforms given in figure 2-13 allow for the assessment of the operation of the memory, and make finding an initial operating point fast, they do not fully test the cell. In particular, the basic waveforms do not test the resilience of the cell to reads and writes from other cells with the column (if the memory were in an array). In order to determine if accessing other cells in the array results the cell to losing its state, or otherwise degrade performance, we use the array emulation waveforms shown in figure 2-14. These waveforms expose the cell to all conditions that it would be exposed to an array. This functionality is achieved by performing a "1" write, a read, a "0" write, and a final read without enabling the heater. These array-emulation signals are placed between the actual write and read from the cell. Thus, if they alter the cell performance in any way, we will see this as a change in the switching current distribution.

The waveforms shown in figure 2-14 are typical of those used in the prototype cell. The cell under test is written to and then read in an alternating fashion in order to determine the cell's basic performance. However, in the basic bit error ratio experiments, a simple alternating sequence of "1" and "0" was used, with the respective times-to-switch T_1 and T_0 used to determine switching currents, and hence error rates.

Due to limitations in our experimental setup, we could not perform cell access operations faster than 10 ns. As such, we could not determine the limits on the device operating speed. However, given the device is based on nanowire devices, we

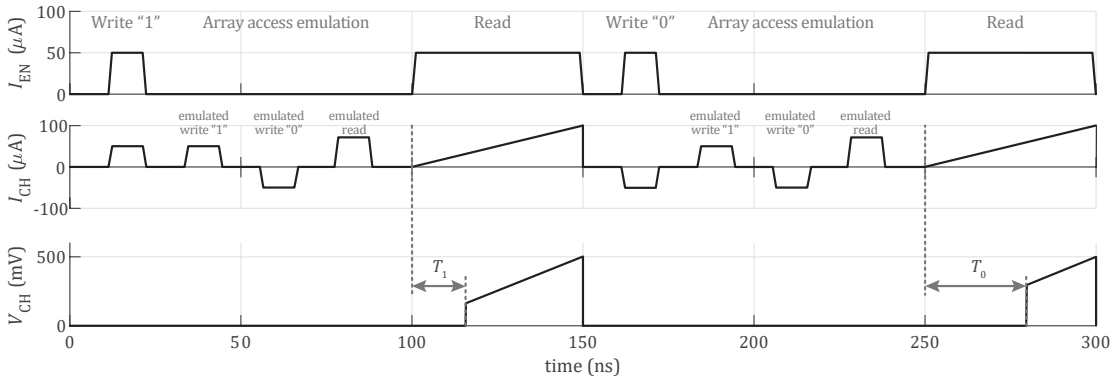

Figure 2-14: Schematic of the ramp-based read out with array emulation. This read out scheme is an enhancement of the scheme presented in figure 2-13. In this scheme the cell is subjected to the same conditions that it would be in an array. The major differences are the incorporation of a "1" write, a "0" write, and then a read to an emulated second cell. These signals are identical to those used to access the cell other than that the heater remains deactivated during the process. This access emulation routine is placed between the actual write and read to the cell. The level of the emulated read is selected to be higher than the maximum measured switching current (higher than the necessary read current). These waveform parameters ensure that the cell can operate in an array without being interfered with by accesses to other cells in the array.

believe speeds on the order of one nanosecond to be reasonable [3, 8, 11]. A detailed analysis of the device operating speed is covered in section 2.5.

2.3.2 Results

Starting out, it is difficult to accurately predict the operating biases that will yield an operable cell. In order to make a start at parameter optimization, the memory was operated in the ramp-based readout scheme without array emulation, as shown in figure 2-13. Once an initial operating point was found, the parameters were systematically varied to find an optimal operating point. At this point, the memory was operating with a lower error ratio than what we had achieved in our prior work. With the memory operating well, and close to optimal for an isolated cell, we progressed onto array emulation, as shown in figure 2-14. We found that switching distributions did not change appreciably between the experiment without array emulation and the experiment with array emulation – which indicated that we are operating at

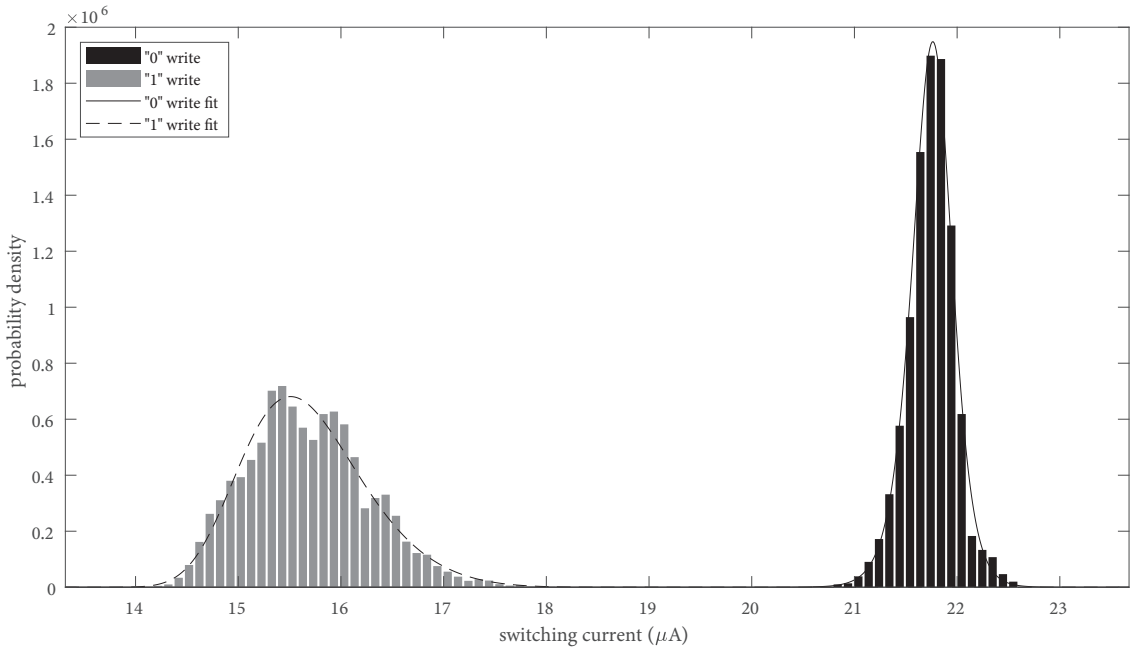

Figure 2-15: Switching distribution of the nMem along with the calculated fits. It can be seen that each distribution is distinct and that there is a region between the distributions where a threshold (the read bias I_{R}) could be setup. It can also be seen that the distributions are different for the "0" and "1" case. For the "1" write a GEV distribution was used, and for the "0" write a Burr distribution was used. Finally, it can be seen that in the "1" distribution there are peaks within the distribution spaced by around 504 nA.

sufficiently high suppression (enable line bias).

The results of the array emulated experiment are shown in figure 2-15. This figure shows the switching distribution of $20,000$ write and read cycles with alternating the data and with emulated array accesses between each write and read cycle. The ratio of the mode of the "1" distribution to that of the "0" distribution is 71 %, which is close to our predicted 67 % from section 2.2.3. The difference between the prediction and the experimental value is due to the noise and hence finite width distribution in the experimental result. It is notable that the shapes of the "0" and "1" distributions are different. This is due to the lower switching current distribution being dependent on the circulating current whereas the upper switching current is only dependent on the critical current of the wire. This effect is examined in further detail in the following section.

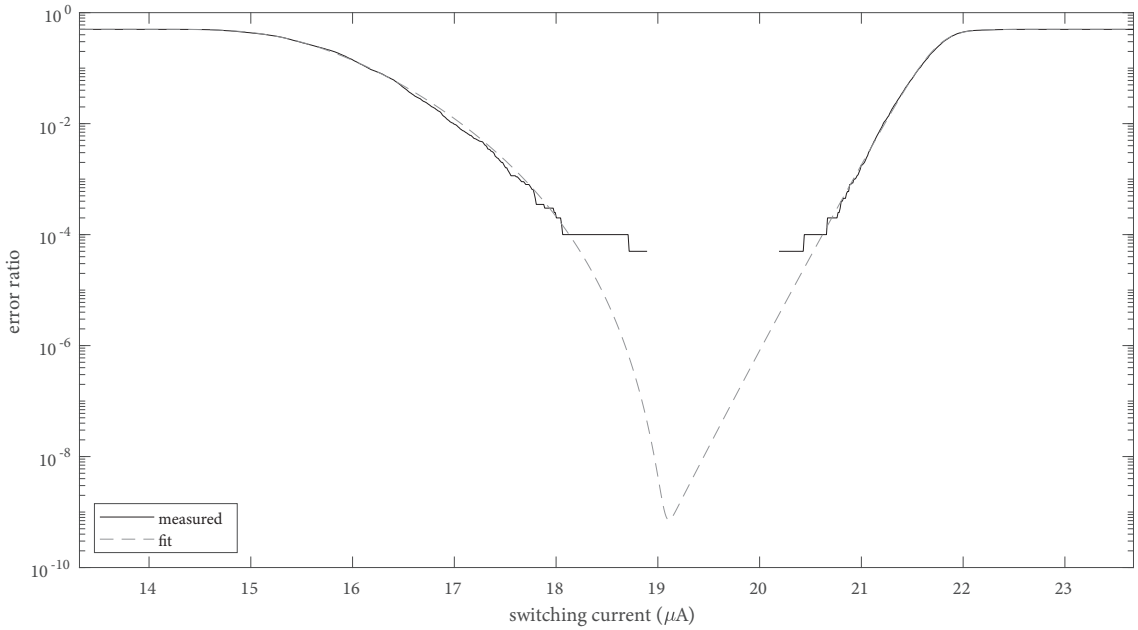

Figure 2-16: A plot of the error ratio of the memory in the ramp-based read scheme as a function of a threshold voltage. The data in this graph is the same data as shown in figure 2-15 and the same fits from that figure, but processed to form an error ratio plot. This shows how close the fits are over a number of orders of magnitude and allows us to use the fits to predict an ultimate error rate. Here, the ultimate error rate is 7×10^{-10}.

Given that the experiment only lasted $20,000$ cycles, we can only directly observe that the error ratio is likely better than one in $20,000$. In order to gain an estimate of the ultimate error rate, we must fit the data to known probability density functions and evaluate their overlap. As will be discussed in more detail in section 2.3.3, the nature of the experiment is that we are performing an extreme-value evaluation of the switching current when determining the "1" state switching current measurement. This type of evaluation leads to the generalized extreme value (GEV) distribution fitting the data well [62,63]. For the "0" data we use a Burr distribution, which fits the data well given that it has a skew towards lower switching currents (due to the slow experiment allowing more time for noise to suppress the switching current) [3,5]. The fits are shown in figure 2-15, and can be seen to match the data well.

The data from figure 2-15 was processed to form an error ratio plot as a function of a threshold current I_{th} – equivalent to a read bias I_{R} in the pulse-based readout

scheme. This processing was performed by means of counting, from the switching distribution data, the number of data points on the "incorrect" side of the threshold current. Functionally, this can be expressed as the formula

$$E(I_{\mathrm{th}}) = \frac{1}{2n}\left(\sum[I_{\mathrm{c,ch,\text{“}0\text{”}}} < I_{\mathrm{th}}] + \sum[I_{\mathrm{c,ch,\text{“}1\text{”}}} > I_{\mathrm{th}}]\right), \tag{2.15}$$

where $2n$ is the total number of write/read cycles captured and using Iverson bracket notation where $[P]$ is defied to be 1 if P is true, and 0 if P is false [64, 65]. Here we performed $2n = 20000$ writes and reads, with half being "0" and half being "1". For the fits (which are continuous data) we perform the continuous analog of this operation. By replacing the sums with integrals, we find that we are simply sampling the cumulative distribution functions. Plotting both the processed data and the processed fits, we generate the plot shown in figure 2-16. It can be seen in this figure that the ultimate error ratio is around 7×10^{-10}. This is an excellent error ratio and far exceeds what would be required for a final application.

2.3.3 Distribution simulation

The shape of the distributions in figure 2-15 are noteworthy. In particular, the shape of the distribution after the memory is written to the "1" state is wider than the distribution when it is written to the "0" state. In order to better understand these distributions, it would be ideal if we could isolate the noise sources. Experimentally this would be very difficult; however, through the development of a basic simulator we can gain the same insight.

In order to simulate the memory readout we took equations 2.5 and 2.6 and combined them to form

$$\begin{aligned} I_{\mathrm{c,ch}}(I_{\mathrm{p}}) = \min\Bigg(&\max\left(\left(I_{\mathrm{c,H_L}}(I_{\mathrm{RE}}) + I_{\mathrm{p}}\right)\frac{L_{\mathrm{L}}+L_{\mathrm{R}}}{L_{\mathrm{R}}}, I_{\mathrm{r,H_L}}(I_{\mathrm{RE}}) + I_{\mathrm{c,H_R}}(I_{\mathrm{RE}})\right), \\ &\max\left(\left(I_{\mathrm{c,H_R}}(I_{\mathrm{RE}}) - I_{\mathrm{p}}\right)\frac{L_{\mathrm{L}}+L_{\mathrm{R}}}{L_{\mathrm{L}}}, I_{\mathrm{r,H_R}}(I_{\mathrm{RE}}) + I_{\mathrm{c,H_L}}(I_{\mathrm{RE}})\right)\right), \end{aligned} \tag{2.16}$$

which determined the switching current for an arbitrary persistent current I_{p}. With

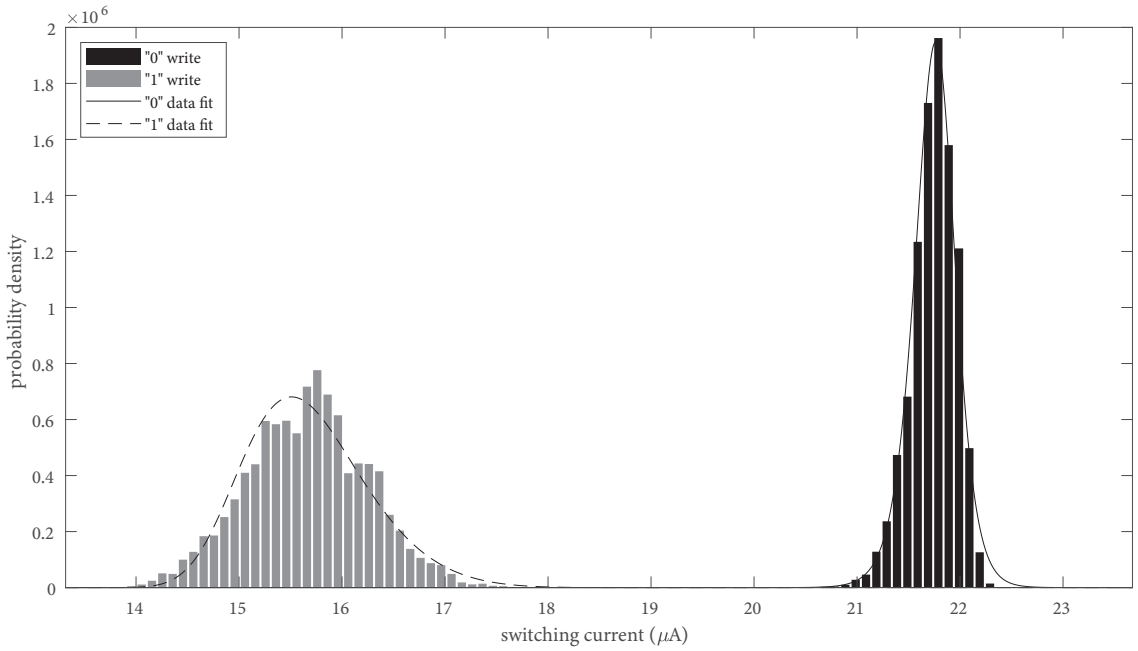

Figure 2-17: Simulation of the switching distributions of the nMem during a read. This simulation incorporates switching current noise, bias current noise, and variations in the persistent current (with flux quantization enforced). The fits shown in this figure are those from figure 2-16. It can be seen that the simulation and the data match extremely well.

this equation, we can then add noise into the system by means of allowing variations in the bias current, switching current, and the number of fluxons trapped in the loop (effectively I_{p}). Each of these variables are allowed to vary with a Gaussian noise (with the loop flux appropriately quantized). There are two fitting parameters (excluding the standard deviations of the random variables), namely the mean number of fluxons trapped in the loop, and the switching current. All other device parameters were taken directly from the simulations presented in section 2.2.2, specifically the switching current ratio and the inductance ratio. After the noiseless simulation was fit to the experiential data, the standard deviations of each random variable were adjusted until the simulation matched the experimental data. The resultant simulation is shown in figure 2-17. This simulation bares a striking resemblance to the experimental data.

The only notable discrepancy between the simulation and the experimental data is the skew in the "0" distribution. This skew is likely due to the fact that the simulation is effectively a slow experiment, where the bias current is increased very

slowly. In comparison, the experiment is a relatively fast measurement. The slower the sweep both in simulation and experiments, the more the switching current will skew to lower values. This is because the slower the experiment the more likely noise will result in a switching event to occur prematurely. In this simulation, the noise was assumed to be purely white, whereas in reality the noise would be pink as there are bandwidth limitations imposed by parasitics of the setup, and more the relatively large inductance of the device. To verify this hypothesis, the simulation was repeated with much faster ramp rate in comparison to the noise sample rate (effectively band limiting the noise). This did result in the "0" distribution skew being corrected; however, the downside of this is that the switching current resolution was degraded. Thus, only the slow simulation results are presented here.

Now that we have a simulation that matches our experimental results, we can investigate the source of the dissimilarity between the "0" and "1" distributions width and skew. As expected in the non-inverting regime, for the "0" state, our switching distribution is essentially just that of the left hTron. Thus, the distribution is simply that of a single nanowire with its intrinsic switching current variation along with noise on the bias line. Thus, assuming Gaussian noise (which matches with the experimental results) we get a Gaussian switching current distribution. Now, given that we are measuring the distribution by sweeping from a low to a high current, we expect a slight skew to lower switching currents, as we observe in the simulation and the experimental data.

For the "1" distribution we find that the operation is a little more complex. For us to see a voltage out of the memory we need that both the left and the right hTrons have switched. For our device parameters, we find that the right hTron switches only slightly after there is sufficient bias to switch both the left hTrons, see figure 2-18. Since there is some overlap between the higher end tail of the current required to switch both devices and the current required to only switch the right hTron, we see a compression of the low-end tail of the "1" distribution, effectively pushing it to higher switching currents. This truncation simply results from the fact that below the bottom edge of the distribution observed distribution, we find that there is a very low

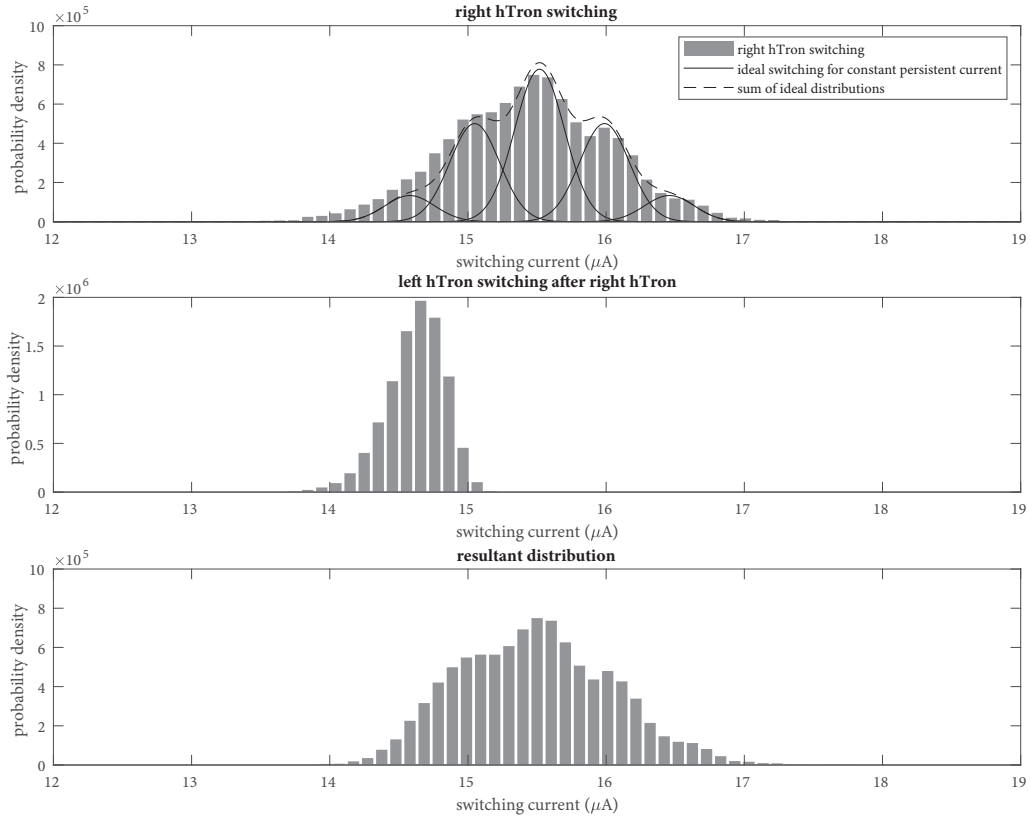

Figure 2-18: Plots of the distribution data extracted from the simulation for the reading of the memory when in the "1" state and non-inverting regime. In the top figure we have the distribution for the right hTron switching, that is $P(|I_{\mathrm{CH}}L_{\mathrm{L}}/(L_{\mathrm{L}}+L_{\mathrm{R}})+I_{\mathrm{p}}| = I_{\mathrm{c,H_R}})$, where I_{CH} is the channel current, along with the ideal distributions assuming a constant persistent current I_{p} for the nominal flux in the loop along with one and two extra and fewer in the loop. It can be seen that the variation in the number of fluxons in the loop is responsible for the local peaks and dips in the switching distribution. It can be seen that, apart from the natural skew towards zero due to the slow experiment, the distribution is not overly skewed. In the center we have the distribution of the left hTron switching assuming that the right has already switched, that is $P(|I_{\mathrm{CH}} - I_{\mathrm{r,H_R}}| = I_{\mathrm{c,H_L}})$, where $I_{\mathrm{r,H_R}}$ is the retrapping current of the right hTron. For the memory to give a voltage out, we need both of these statements to be true. That is, we need the channel current to be high enough that after the right hTron switches, we will see a switch in the left hTron; this essentially takes the maximum of the two random variables. The bottom figure is the final switching distribution for the memory in the "1" state. The combination of the distributions in the first and center figures is what gives rise to the final distribution resembling the first figure with its left-hand compressed rightwards. This effect explains why we see the skewness in the "1" distribution seen in the experiment and simulation results.

probability that we have sufficient bias current to switch both hTrons. This results in the distribution's asymmetry and being skewed to higher currents.

The second notable feature of the "1" distribution is its width in comparison to the "0" distribution. As we can see from the decomposition of the switching of the memory in the "1" state shown in figure 2-18, the greatest contribution to the width is from the variance in the persistent current. That is, because the memory readout current resolution is higher than the flux quantization, we can see the variation in the flux written into the loop in the write process. While the standard deviation in flux written is only around a single fluxon, this stacks on top of the switching variation we see in just reading a critical current, thereby making the distribution substantially wider, as seen in the sum of the distributions shown in figure 2-18.

The observation that the "1" distribution is the result of finding the maxima of multiple independent random values with identical distributions, then by the Fisher–Tippett–Gnedenko theorem, our resultant distribution tends to the GEV distribution [62, 63]. Thus giving a justification to our fitting curve from figures 2-16 and 2-15 being GEV distributions.

2.4 System emulation

The distribution measurements presented so far indicate that this memory would operate well in a final system; however, they do not directly operate the memory as it would be in a system-implementation. In order to operate the memory in such a manner, we developed a custom test setup, the nanoBERT, the design of which is detailed in section 5.1. This device allows us to perform large scale rests and sweeps on the memory with pulse-based waveforms akin to those that would be used in a computer.

The nanoBERT performs the same array-emulation operation as was presented in the previous section. In addition, instead of performing alternative "0" then "1" write/read cycles, the nanoBERT implements a pseudo-random binary sequence (PRBS) generator. The PRBS generator tests the memory with every order of 7-bits

(excluding all zeros due to the PRBS algorithm implementation – see section 5.1). This ensures that we are not biasing our data with our write order and allows us to better determine how the device would respond to real-world data.

The nanoBERT tracks what data was written and compares the data it reads. With this it gives us three data points, the number of write-"1"-read-"0" (W1R0) errors, the number of write-"0"-read-"1" (W0R1) errors, and the number of writes of each type performed. From this we can determine the error ratio as $P_{\mathrm{E}} = (N_{\mathrm{W0R1}} + N_{\mathrm{W1R0}})/(N_0 + N_1)$, where N_{W0R1} and N_{W1R0} are the number of W0R1 and W1R0 errors, respectively, and N_0 and N_1, are the total number of "0" writes and "1" writes, respectively.

Given that the test setup for the nanoBERT is different to that used for the testing presented in the prior section, some effort must be put into finding the correct operating point. One of the major disadvantages of operating the memory with pulse-based waveforms is that the only feedback we get is the error ratios. Thus, it can be difficult to determine which bias levels need to be increased or decreased to reach the operating point. This disadvantage is largely mitigated by the high speed of the nanoBERT which allows for sweeping large parameter spaces in orders-of-magnitude less time than with the AWG and oscilloscope based setup used in the previous section. So, to find a rough initial operating point we simply swept all parameters until we found the operating point.

A coarse sweep of the write bias is shown in figure 2-19. It can be seen in this figure that the memory transitions between three operating regimes. First, at low write biases we have insufficient bias to switch the memory, thereby leaving the memory in the state generated when the memory was last read. A read leaves the memory in the "1" state. Thus, we see a near unity W1R0 error ratio and a zero W0R1 error ratio. As the write level increases, we see the W1R0 error rate decrease to zero. There is then a regime for which the memory error ratio is zero. With increasing write bias, we find that the memory enters into an inversion regime. In this regime the memory state is the opposite of the intended result with near unity error ratio. This regime corresponds to when, during a write, both sides of the memory switch and latch.

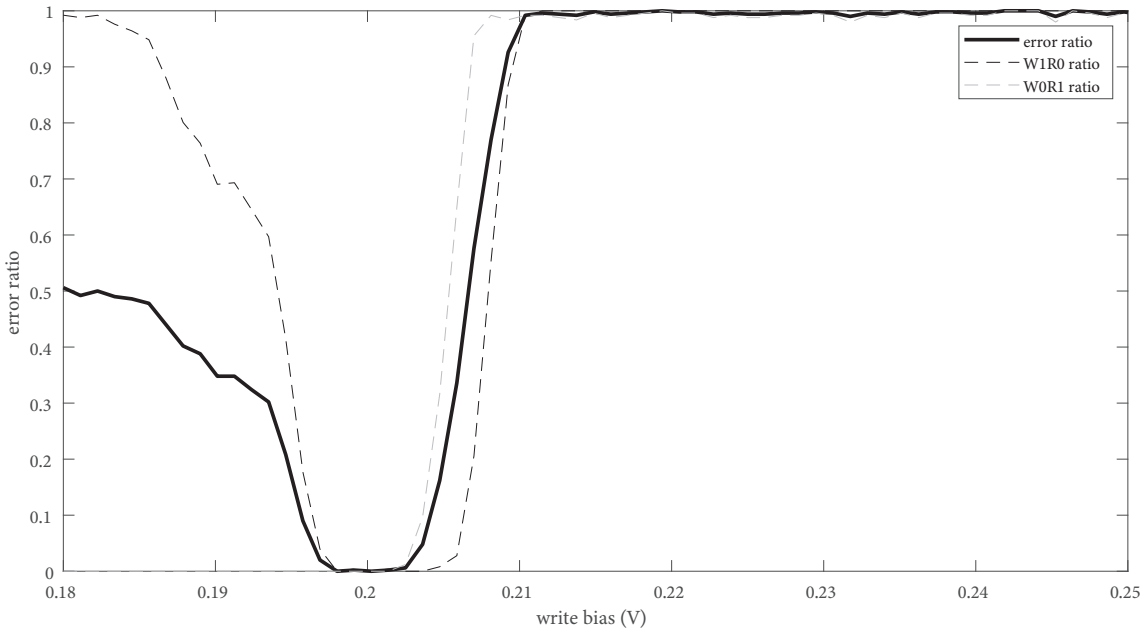

Figure 2-19: Error ratio of the nMem as the write bias is swept. The change in regime from not affecting the memory state below 0.18 V, to the expected operating regime from 0.198 V to 0.203 V, to the inverted regime from 0.210 V onward.

From figure 2-19 we find that we need to be operating with a write bias around 0.2 V with the given heater value. Similar sweeps were performed on all other parameters until a suitable operating point was found for each. This was then repeated a second time to ensure that we take into account the interaction between these parameters (to a first-degree at least). Once the appropriate operating point was found we could perform fine sweeps of the operating parameters to determine the system margins. The optimal points are summarized in table 2.1.

Sweeping the write bias we find the results shown in figure 2-20. It can be seen that below around 50 mV we have nearly all W1R0 errors. This follows from the fact that we are likely either not switching the constrictions during a write, or are only adding a small amount of flux when we do. Thus, the memory will stay in the "0" state no matter what was written. From this value to around 90 mV we have error ratios around 10^{-2}, with exclusively W1R0 errors. In this region we are not writing sufficient flux to overwhelm the variations due to noise in the device. Thus,

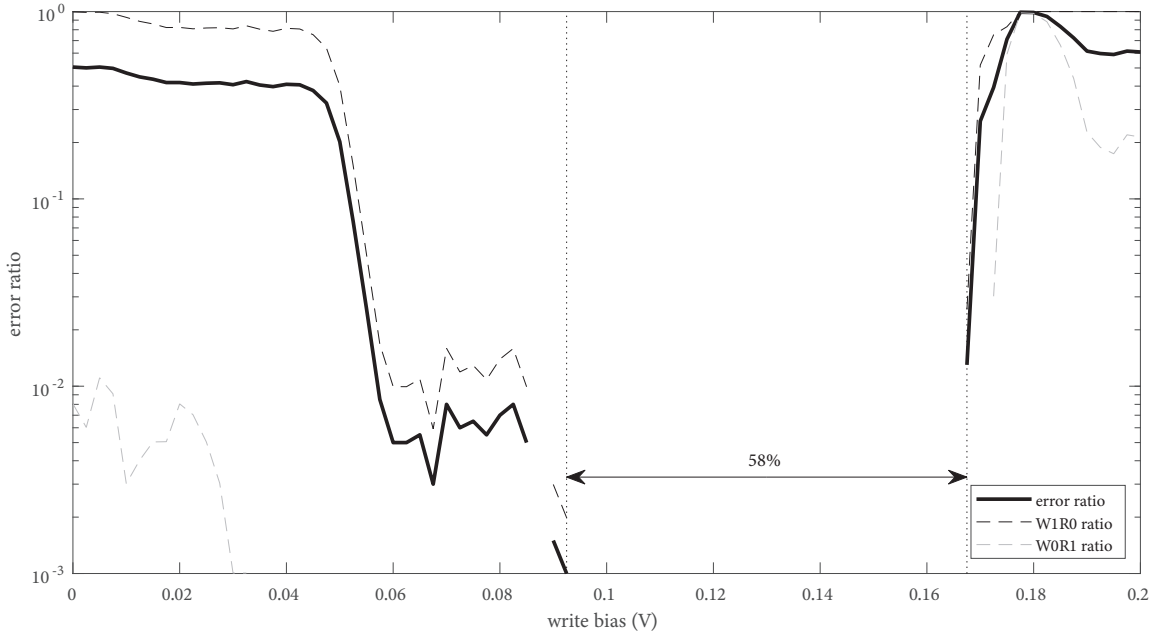

Figure 2-20: Plot of the error ratio observed in the memory cell as the write bias is swept. Each point in this plot represents 2,000 pseudo-random write-read cycles. The total error ratio along with the breakdown into W1R0 and the W0R1 error ratios are plotted. There is a region from around 0.09 V to 0.17 V where no errors were recorded. This represents a 58 % wide region with reference to optimal write bias (around the center of this region) for which there are no observed errors.

we still see W1R0 errors. From around 90 mV to around 170 mV we have no error observed. The width of this region is 58 % with respect to the optimal bias. From around 170 mV and up we have that during a write both sides of the memory are switching and, as such, we enter the inverted region.

Sweeping the read bias we find the results shown in figure 2-21. It can be seen that we start with exclusively W1R0 errors as we never switch the memory and as such only read "0". As we increase the read bias we see a decrease in error ratio. We see no errors for 273 mV to 291 mV. This region represents a 6.4 % wide operating region for which no errors are observed. From 291 mV upwards we see an inversion in the error memory operation. This is an interesting operating regime that would need further investigation to properly explain; however, it seems that the memory is operating in the inverting regime due to the memory transitioning to the inverting readout scheme (see section 2.1.2). It can be seen that the error ratio in this regime

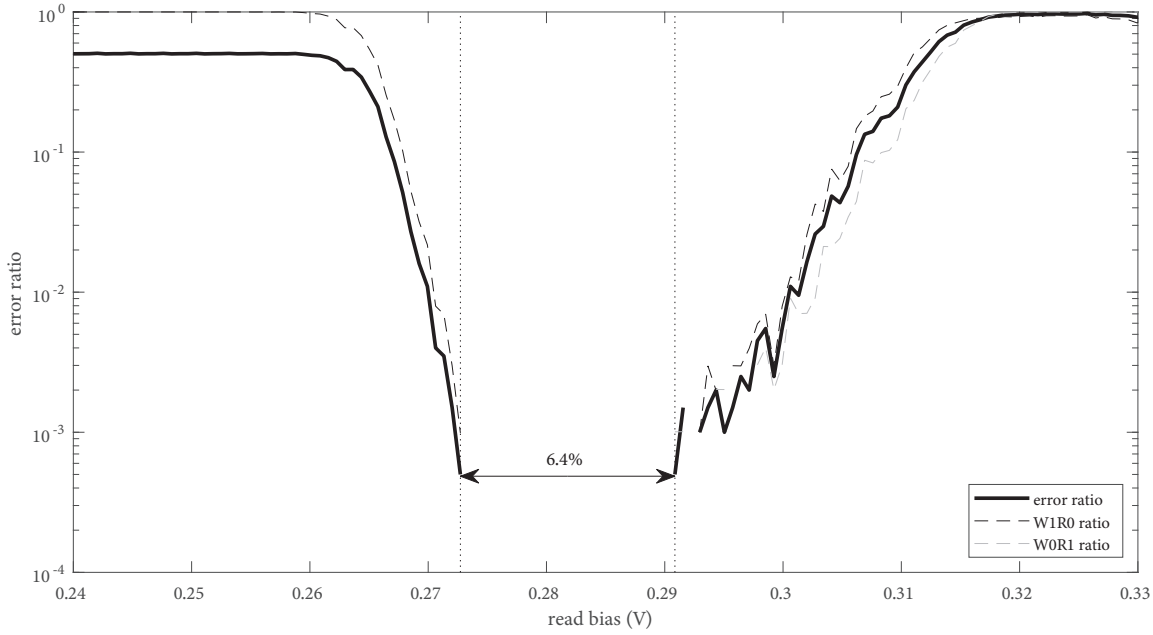

Figure 2-21: Plot of the error ratio observed in the memory cell as the read bias is swept. Each point in this plot represents 2,000 pseudo-random write-read cycles. The total error ratio along with the breakdown into W1R0 and the W0R1 error ratios are plotted. There is a region from around 0.273 V to 0.291 V where no errors were recorded. This represents a 6.4 % wide region with reference to optimal read bias (around the center of this region) for which there are no observed errors.

is worse, and thus we will continue operating in the intended non-inverting regime.

Finally, sweeping the heater bias during a read we have the results shown in figure 2-22. We find that at low heater-biases we have only W1R0 errors, up to around 0.98 V. We expect W1R0 errors since for insufficient suppression of the hTron critical current during a read we will fail to cause any switching, and so regardless of the memory state, we will read a "0". From 0.98 V to 1.18 V we see no errors; this region represents around 18 % of the optimal bias point. From 1.18 V on we see only W0R1 error. Again, this follows since we expect that for excessive suppression, regardless of the cell state, the memory will switch, giving an output of only "1".

In an attempt to characterize an ultimate error rate experimentally, we exercised the memory optimal operating for 1,000,000 write/read cycles. Table 2.1 summarizes the results of the experiment. Over the course of this experiment we found no errors occurred. Thus, we experimentally observe an error rate less than 10^{-6}, which is

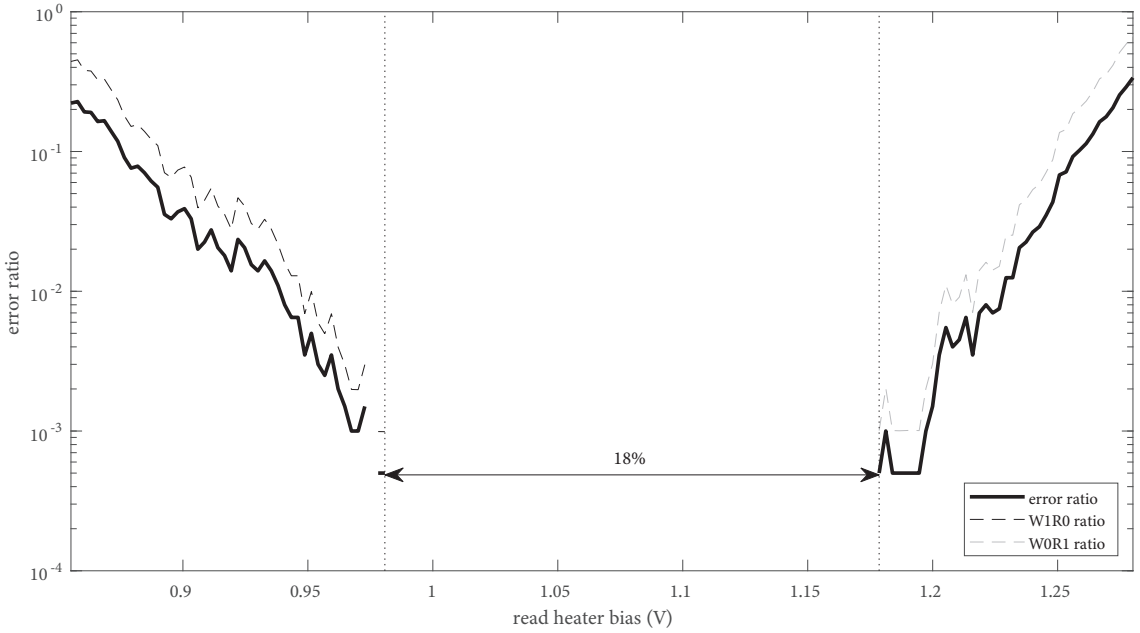

Figure 2-22: Plot of the error ratio observed in the memory cell as the read heater bias is swept. Each point in this plot represents 2,000 pseudo-random write-read cycles. The total error ratio along with the breakdown into W1R0 and the W0R1 error ratios are plotted. There is a region from around 0.98 V to 1.18 V where no errors were recorded. This represents a 18 % wide region with reference to optimal read heater bias (around the center of this region) for which there are no observed errors.

consistent with our prediction from section 2.3.2 of an ultimate error rate around 7×10^{-10}. Thus, the experimental performance of this cell is comparable to the state-of-the-art cells, and the predicted ultimate error rate is superior to that of alternate memory technology [11].

2.5 Speed and power estimates

For a cryogenic memory to be useful it must be, among other traits, fast and consume little power – as speed and power are the major driving forces behind using cryogenic and superconducting supercomputers. Due to the limitations in our experimental setup and available instruments, the fastest we have operated the memory is with 10 ns pulse widths, for which we see no degradation in performance. In order to gain an estimate of the speed and power limitations of the memory we will need to resort

Table 2.1: Memory operation in system emulation

Parameter	Value
write bias	178 mV
write heater bias	1.6 V
read level	279 mV
read heater bias	1.069 V
number of "0" written	496,069
number of "1" written	503,931
number of W0R1 errors	0
number of W1R0 errors	0
total number of writes	1,000,000
total number of errors	0
observed error ratio	$< 10^{-6}$

to non-experimental means. The exact speed and power limitations of the cell are a complex interplay between the electrical and thermal characteristics of switching nanowires. Thus, we need to perform some simplification of the system in order to develop an appropriate model.

To model the superconducting circuit side of the memory, a simple simulation tool was developed based on the simulation methodologies described in [66], [67] and [68]. Rather than implementing the simulator in SPICE as was done in [68], we elected to develop a simple simulator in MATLAB. This decision was made to avoid the issues with simulating superconductors in SPICE, namely the necessity for there to be some resistance in a loop. In addition, programming a custom simulator allowed us to avoid some of the convergence issues that we commonly encounter in SPICE models of nanowires. The major downside of this approach, the lack of SPICE-like user-friendliness, is that the simulator is comparably slower than SPICE. This downside is not noticeable in our simulations as we are only interested in relatively small simulations, and for the simulations presented here the processing time was less than a second.

Within the memory cell there are three primary time scales at play, namely that related to the thermal reaction of the wire to the enable-line heater, that associated with the hot-spot formation and collapsing, and that associated with the current re-

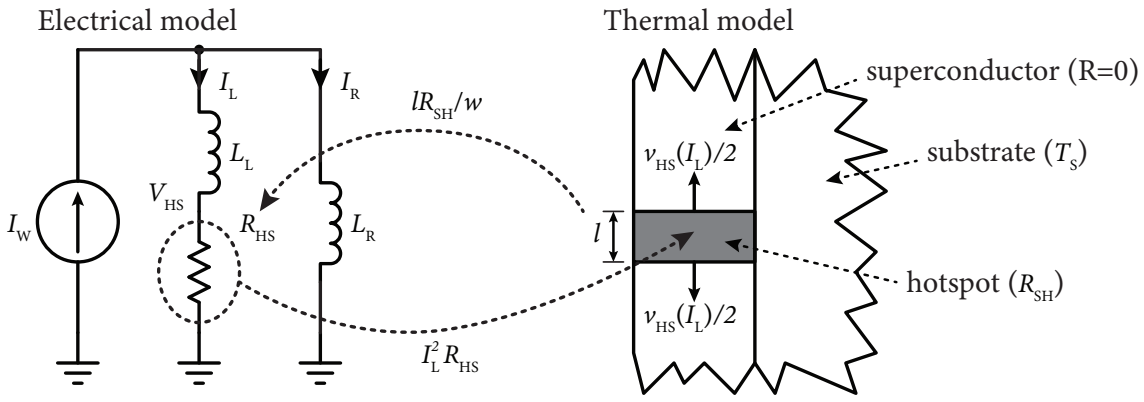

Figure 2-23: Schematic representation of the model of the nMem simulation for a write. The resistor R_{HS} fully encompasses the electrical portion of the nanowire dynamics by having its value continually changed in response to the thermal portion of the model. If the nanowire is in the superconducting state the resistor is set to exactly zero ohms, $R_{\mathrm{HS}} = 0$.

distribution. To simplify analysis we separate these dynamics into two sets. First, the "slow" thermal timescales associated with the heater and the "fast" timescales associated with the electro-thermal modeling of the memory loop and nanowire switching dynamics. This division is natural to make given that the heater experiences little feedback from the nanowire dynamics and the electrical circuits of the heater and the nanowire are isolated.

First, we will consider the "fast" dynamics of the nanowire. In order to model the device we reduce the system down to a 1-D electro-thermal model of a nanowire and a lumped model of the electrical circuit interfacing to the nanowire. A schematic of the modeled system is shown in figure 2-23. This model is based on those presented in [66–68]. Here we will only briefly go over the operation of this simulator. Central to the operation of the model is the treatment of the nanowire as an infinitely long 1-D strip of superconductor with zero resistance within which there is either no hot-spot or a single hot-spot. If a hot-spot exists then it has a resistance R_{sh}/w per unit length, with w being the nanowire width. This single hot-spot can grow and shrink with velocity $v_{\mathrm{HS}}(i)$, where i is the current through the nanowire. For writing to the memory, we have only a hot-spot in the left hTron which carries a current I_{L}. Within

this framework the hot-spot speed is given by

$$v_{\mathrm{HS}} = 2v_{\circ}\frac{\psi I_{\mathrm{L}}^2/I_{\mathrm{c}}^2 - 2}{\sqrt{\psi I_{\mathrm{L}}^2/I_{\mathrm{c}}^2 - 1}}, \tag{2.17}$$

where $v_{\circ} = \sqrt{h_{\mathrm{c}}\kappa/d}/c$ is a characteristic speed, with h_{c} being the thermal contact conductivity of the interface between the nanowire and the substrate, κ being the thermal conductivity of the nanowire, d being the thickness of the nanowire, c being the specific heat per unit volume of the nanowire, and

$$\psi = \frac{\rho I_{\mathrm{c}}^2}{h_{\mathrm{c}} w^2 d\,(T_{\mathrm{c}} - T_{\mathrm{S}})}, \tag{2.18}$$

is the Stekley parameter which is the ratio of the nanowire Joule-heating to thermal conduction in the normal region. Here we use the typical values for NbN either from our device or from the literature [68]. A plot of this hot-spot speed is given in figure 2-24. It can be seen that for increased bias currents, such as the critical current, the hot-spot increases very rapidly. The hot-spot will continue to increase until the bias through the device deceased to below a value "hot-spot plateau" current. For values below this current the hot spot will collapse extremely fast. At a current slightly below the hot-spot plateau current we have the device returning to the superconducting state. This current is the retrapping current I_{r}.

Now, when we run the simulation with the parameters of our film and the parameters we extracted from the device simulations in section 2.2.3, we find the dynamics shown in figure 2-25. The simulation was run with an initial persistent current of $I_{\mathrm{p}} = 0$, and a bias to the memory of I_{W}. The left nanowire switches at $t = 0$ and the dynamics evolve from there. From the point that the left hTron switches, the current through the left branch I_{L} progressively decreases and that through the right branch I_{R} simultaneously increases – of course maintaining $I_{\mathrm{W}} = I_{\mathrm{L}} + I_{\mathrm{R}}$. The dynamics terminate after the current through the left branch deceases to the retrapping current $I_{\mathrm{r,H_L}}$, which occurs in less than 50 ps. The total energy dissipated in the nanowire is the integral of the power dissipated in the nanowire $E_{\mathrm{W}} = \int I_{\mathrm{L}} V_{\mathrm{HS}}\,\mathrm{d}t$, where V_{HS} is

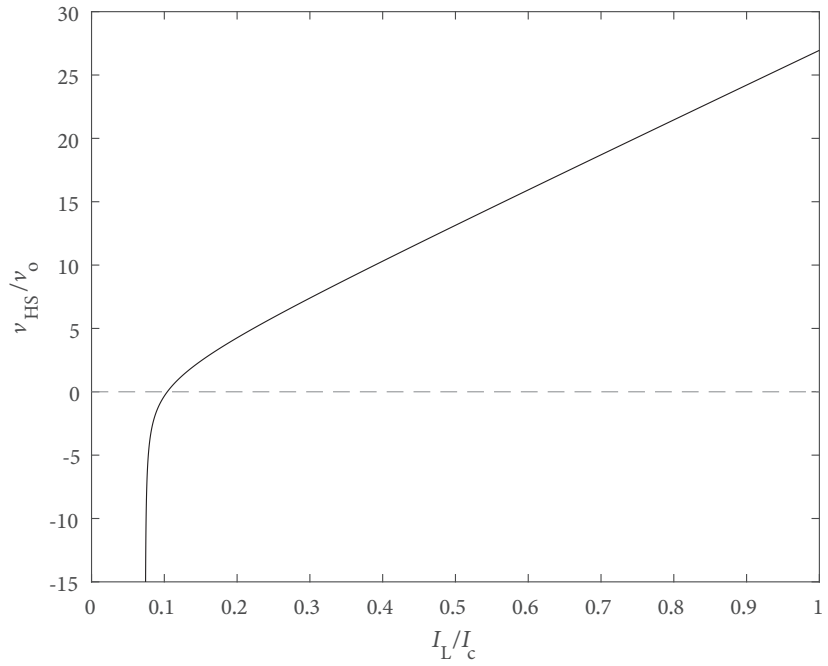

Figure 2-24: Plot of the phenomenological model of the hot-spot front propagation along a theoretical infinite length and constant width nanowire for the parameters of our particular device. Specifically, $v_{\mathrm{HS}} = dl/dt$, where l is the length of the hot-spot. It can be seen that at a bias of around $I_{\mathrm{c}}/10$ we see the hot-spot neither grows nor shrinks. This static point gives rise to the "hot-spot plateau" seen in the IV-curves of nanowires.

the voltage across the hot-spot. The plot of the instantaneous power is shown in the right of figure 2-25. The total energy dissipated in the nanowire for a write is 1.29 aJ. For different starting persistent currents, very similar timing and energy values will be found.

For a read things are a little more complex. In a write operation, the hot-spot never becomes particularly large, thus we do not need to worry about our assumption that the nanowire is a 1-D infinitely long strip of superconductor with a finite normal region. However, for reading we latch the nanowire; thus, we need both sides of the loop to have a hot-spot at the same time. This causes issues since our model has no limiting function on the size of the hot-spot under a constant current bias. In order to obtain estimates on the read speed and power, we must implement such a limit.

In reality the size of the hot-spot is limited by two main factors. First, the nanowire is not of constant width, so at some point the wire is going to be so wide

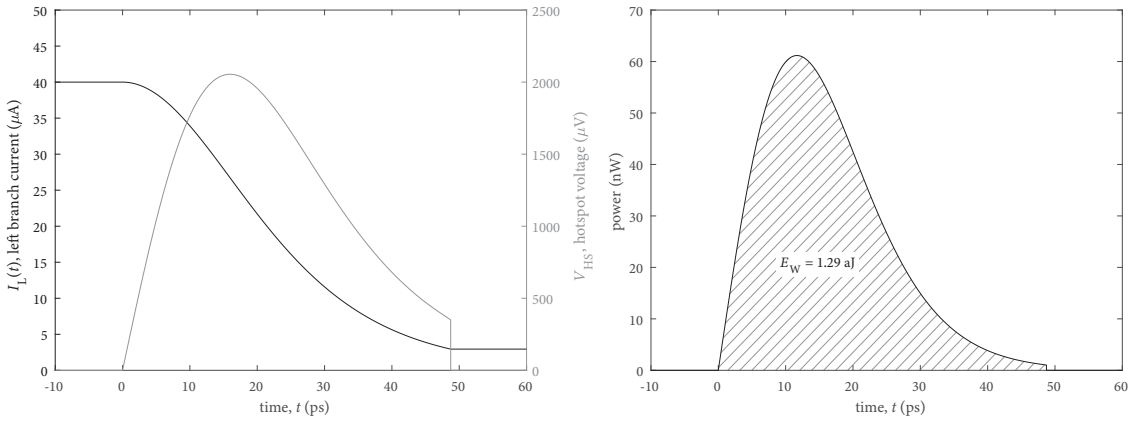

Figure 2-25: Simulation result of the memory during a write operation with $I_\mathrm{p} = 0$ at $t = 0$ and ended at $I_\mathrm{p} = I_\mathrm{W} - I_{\mathrm{r,H_L}}$, where I_W is the write bias and $I_{\mathrm{r,H_L}}$ is the retrapping current of the left hTron. (left) The current through the left hTron and the voltage across the hot-spot evolving as time progresses. The nanowire switches at time $t = 0$, and the simulation evolves from there. (right) The power dissipated in the hot-spot calculated as $I_\mathrm{L} V_\mathrm{HS}$ along with its integral, the total energy required for a write operation.

that it will stay superconducting. Second, the localization of the heat from the hTron heaters means that as we progress further from the heater location the wire temperature drops and so the J_c increases, thus again limiting the size of the hot-spot. Since these effects would require a full 2-D (or 3-D) electro-thermal simulation, which is difficult to perform, we implement a soft limiting function on the hot-spot length generated by our model. This does not provide us with an exact model of the behavior in the real device, but will allow us to give reasonable estimates on the power and speed limitations, yielding results in the same order of magnitude as we expect in the real device.

For our simulation we implement the soft limiter on the hot-spot length as

$$l = L_\mathrm{max} \tanh\left(\frac{\int v_\mathrm{HS}\,\mathrm{d}t}{L_\mathrm{max}}\right) \tag{2.19}$$

where L_max is the maximum length we allow the hot-spot to ever achieve – see section 2.2.2 for a discussion on soft limiting/soft clipping. For the simulations presented here we used $L_\mathrm{max} = 10w$ where w is the hTron channel width. This is generously

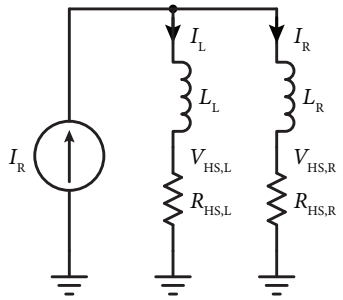

Figure 2-26: Schematic of the electrical circuit portion of the read simulation. The thermal portion of the model consists of two hot-spot models, one for each hTron, and are identical to those shown in figure 2-23.

overestimating the maximum hot-spot size that we would ever be able to expect in reality given the device geometry. With the soft limiter now implemented, we can simulate the read circuit shown in figure 2-26.

For a "0" read in the non-inverting readout we do not see any switching, and so we need not consider the energy and timing in this case. For a "1" read we will start our persistent current being $I_p = 0$ which will result in a "1" read for the bias current selected. This will yield a higher read power dissipation and long read time than is expected in an actual device. The results of the read simulation are shown in figure 2-27. It can be seen that the state of the memory is valid after less than 50 ps, and that the dynamics settle at around 100 ps. However, we do not need the result to settle, as once the result become valid we will not see them become invalid until the bias is removed.

Next, we will consider the "slow" dynamics of the heater. A complete treatment of the dynamics present in the hTron requires extensive modeling taking into account many factors that affect the operation of the device. Such a treatment is beyond the scope of the work presented here. We have published work on such modeling in [8]. As such we will just state that, based on this modeling, the hTron used in this device has a latency between the first application of the bias current and the channel becoming fully suppressed of around 1 ns. This is a relatively slow process here due to a number of factors including material selection and the thickness of our spacer. It is anticipated that this could be engineered to be reduced greatly.

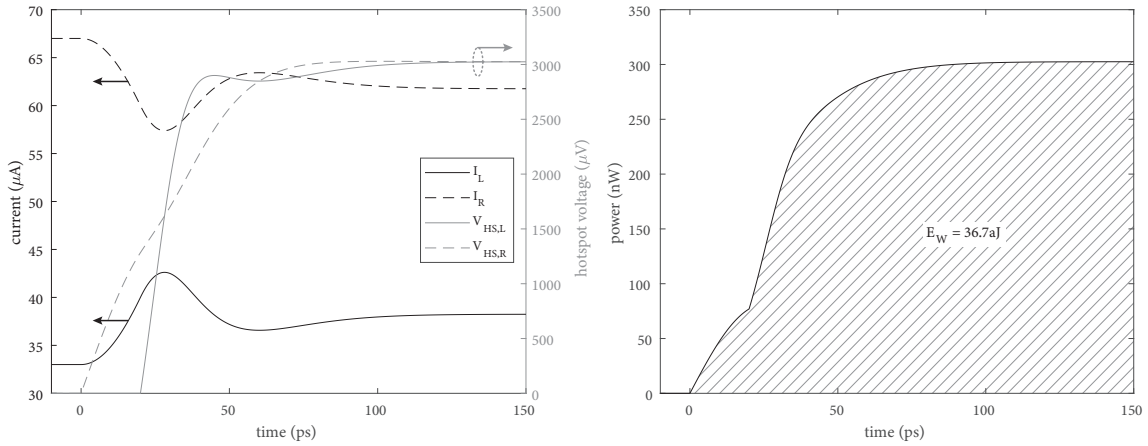

Figure 2-27: Simulation results showing the internal state evolution within the memory during a read operation. The first nanowire switching event occurs at $t = 0$ and evolves from that point on. (left) The current through each hTron along with the hotspot voltages. (right) The total power dissipation, and the integration result giving the total energy. The energy shown in the figure is only that for the period necessary for the memory to achieve a stable state. Since the memory is in the normal state at the end of the simulation period, it will continue to dissipate power until the read bias is removed.

The complete timing limits for this prototype cell are shown in the timing diagram in figure 2-28, and values listed in table 2.2. Here we present the timing for a write, and a read, as well as the timing for a read and write-back. The read and write-back operation is a common operation where, due to the destructive nature of the read, the state of the cell must be restored after a read. Since the cell is already selected, it follows that it will be faster for the memory controller to write the state of the cell back to it after a read without having to wait for another hTron propagation delay.

With the heater as simulated in section 2.2.3, we required $6\,\mu$W to achieve the required suppression of the channel. Thus, for a write, where we need the heater to be active for 1 ns, we have a write requires a minimum heater dissipation of 6 fJ. Similarly for a read, the heater is required to be active for the same period, and so the same energy requirement. For a read and write-back operation, we also have that the heater is on for around the same minimum time (roughly 50 ps longer), and so a similar energy dissipation. The heater dissipation being much larger than the nanowire dissipation (6 fJ in comparison to the less than 50 aJ dissipation from figure

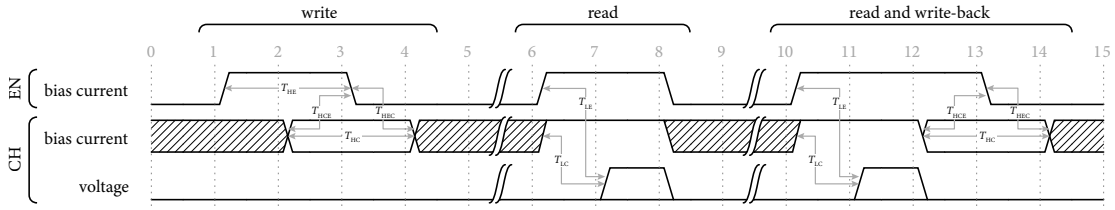

Figure 2-28: Timing diagram for writing to, reading from, and performing a read and write-back operation. This diagram summarizes the logical sequencing and timing requirements for the memory cell. The levels here simply represent logical states and not exact voltages and currents. The timing value estimates for the prototype cell are provided in table 2.2.

Table 2.2: Estimated timing requirements for the prototype memory cell

Parameter	Symbol	Value
write		
enable hold time	T_{HE}	1000 ps
channel-to-enable hold time	T_{HCE}	50 ps
enable-to-channel hold time	T_{HEC}	1000 ps
channel hold time	T_{HC}	50 ps
total write time	T_{W}	2000 ps
read		
enable latency	T_{LE}	1000 ps
channel latency	T_{LC}	50 ps
channel latency	T_{LC}	50 ps
minimum total read time	T_{R}	1000 ps
read and write-back		
minimum total read and write-back time	T_{RW}	2050 ps

2-27 and figure 2-28), we can simply consider the cell energy dissipation to be equal to the heater dissipation for this cell design. Like the heater latency, this energy requirement can be engineered to achieve improved performance. However, with the timing and power dissipation as they are, this device is comparable to, or better than, alternative state-of-the-art memory technology [11, 69].

With further engineering of the materials, specifically the substrate and spacer, as well as the heater geometry [8], improved performance will be achieved. With the realization of these improvements, and further optimization of the cell design, this memory could finally satisfy the need for a superconducting memory technology.

Chapter 3

Nanowire microwave devices – the microwave toolbox

In this chapter, we utilize the temperature dependence of the kinetic inductance that we had to compensate for in the nMem, and instead use it to realize tunable microwave devices. We will start with a discussion of kinetic inductance. Next, we will present a tunable on-chip tank resonator. Next, we will present a number of tunable filters. Finally, we will present a microwave detector.

3.1 Kinetic inductance

Kinetic inductance is a result of the finite number of carriers that participate in conduction. Given that a finite number of carriers exist, then there is a non-zero inertia of these carriers, thus, they exhibit a resistance to changing velocity. From a purely electrical point of view, the inertia is indistinguishable from an inductance. We can form an expression for the kinetic inductance by equating the kinetic energy of the Cooper pairs in a superconductor to the energy in an inductor as

$$\frac{1}{2}(2mv^2)(n_s l A) = \frac{1}{2}L_K I^2, \tag{3.1}$$

where $2m$ is the Cooper pair mass, v is their velocity, n_s is the density of Cooper pairs, l is the length of the superconductor, A is the superconductor cross-sectional area, L_K is our kinetic inductance, and I is the current through the superconductor. We determine the velocity based on the drift velocity expression $v = I/2en_sA$, where $2e$ is the charge of a Cooper pair. Combining equation 3.1 with the drift velocity we find that the kinetic inductance can be expressed as

$$L_K = \frac{ml}{2An_se^2}. \tag{3.2}$$

For a normal metal we have a similar expression $L_K = ml/Ane^2$ where n is the density of carriers. So one might expect a normal metal to in fact have a higher kinetic inductance than a superconductor. The reason we do not see this in most circumstances is that for most metals the density of electrons is extraordinarily high, and so the kinetic inductance is small. For metals with lower density of electrons, the resistance is correspondingly high, and so the inductance is not readily observed. In a superconductor, we can have a small density of Cooper pairs, and of course no DC resistance, thus making the kinetic inductance quite apparent. It should be noted that for normal metals, at very high frequencies $\sim$THz, the kinetic inductance is quite readily apparent, and in fact, put to use [70].

From examining our expression for kinetic inductance, we find that it depends on two types of parameters that can be changed, namely geometry (length l and cross-sectional area A), and the density of Cooper pars n_s. For a practical device, we can typically consider the geometry fixed (although a kinetic inductance-based strain sensor may be possible). So for practical purposes, this leaves us with kinetic inductance being inversely proportional to the density of Cooper pairs. Given the nature of superconductivity, the Cooper pair density is dependent on each of the axes in the superconducting phase diagram. That is, the Cooper pair density decreases with increased temperature, current, and magnetic field. For all of our devices we can ignore the dependence on the magnetic field as we are operating at very low fields, and they are relatively constant over our experiments. Temperature and current, on

the other hand, we vary from near zero to their respective critical values.

3.1.1 Ginzburg-Landau theory

There are a number of methods of deriving the dependence of the kinetic inductance on temperature and current. First, we will discuses GL theory, which is only valid near T_c [53, 71]. GL theory introduces a complex pseudo wave function ψ as an order parameter for the superconducting electrons such that the local density of superconducting electrons is

$$n_s = |\psi(x)|^2 . \tag{3.3}$$

For the temperature variation in $|\psi(x)|^2$ we will consider the GL free-energy. If we assume ψ is small and varies slowly in space then the free-energy density f can be expanded to

$$f_s = f_{n0} + \alpha |\psi|^2 + \frac{1}{2}\beta |\psi|^4 + \frac{1}{2m^*} \left| (\frac{\hbar}{j}\nabla - \frac{q^*}{c}\mathbf{A})\psi \right|^2 + \frac{\mathbf{B}^2}{8\pi}, \tag{3.4}$$

where α and $\beta > 0$ are two variables, $\mathbf{A}$ is the vector potential, $\mathbf{B}$ the magnetic field, m^* and q^* are the mass and charge of the particle, respectively, and c is the speed of light. We encapsulate the normal state contributions as $f_n = f|_{\psi=0}$. If we take the difference between the superconducting and normal free-energies, and with no fields or gradients we have

$$f_s - f_n = \alpha |\psi|^2 + \frac{1}{2}\beta |\psi|^4 . \tag{3.5}$$

This is essentially a series expansion of $|\psi|^2$ with only the first two terms. These terms are adequate for determining $|\psi|^2$ near T_c.

Plotting the superconducting excess free-energy we find, as shown in figure 3-1, that there are two sets of equilibrium points. One set for $T > T_c$ ($\alpha > 0$) for which $|\psi(x)|^2 = 0$, corresponding to the normal state. The second solution is for $T < T_c$ ($\alpha < 0$) for which $|\psi_\infty|^2 = -\alpha/\beta$. The symbol ψ_∞ is used to denote this equilibrium point since it is the case for infinitely far into the bulk where perfect screening occurs.

It follows that α changes sign as we cross over from $T < T_c$ to $T > T_c$. Taking the

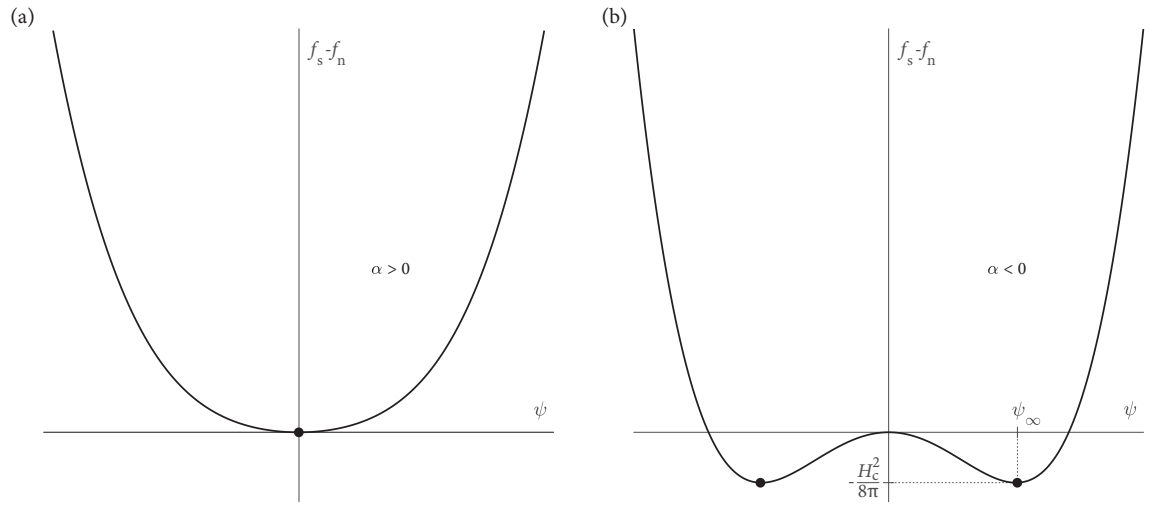

Figure 3-1: Sketches of the free energy for (a) the case when $\alpha > 0$ ($T > T_c$), and (b) the case when $\alpha < 0$ ($T < T_c$). It can be seen that there is only one equilibrium point in the $\alpha > 0$ case, whereas there are two in the $\alpha < 0$ case. These two equilibrium points are located at ψ_∞.

Taylor series expansion around $T = T_c$ that $\alpha(T) \approx \alpha'(T - T_c)$, where $\alpha' > 0$. Now, we need the temperature variation of β. Substituting our solution for the equilibrium into equation 3.5, we find that $f_s - f_n = -\alpha^2/2\beta$, which is, from the definition of the thermodynamic critical field H_c is equal to $f_s - f_n = -H_c^2/8\pi$. Now, since H_c varies linearly with $1 - T/T_c$ near T_c, and β is roughly constant near T_c. Thus, $|\psi|^2 \propto (1-T/T_c)$ for T near T_c with $T < T_c$. Taking this back to the superconducting electron density using equation 3.3, we find $n_s(T) \approx n_s(0)(1 - T/T_c)$. So, for zero bias current (or small currents) the kinetic inductance can be expressed as

$$L_K(T) = \frac{ml}{2An_s(0)(1 - T/T_c)e^2} = L_K(0)\frac{1}{1 - T/T_c}. \tag{3.6}$$

3.1.2 Mattis-Bardeen theory

The major drawback of GL theory and applying it to our device is that it only applies near T_c. We can find an expression for all temperatures from $T = 0$ to T_c through BCS and specifically Mattis-Bardeen theory [52,53]. The Mattis-Bardeen theory gives

us an expression for the complex conductivity $\sigma(\omega) = \sigma_1(\omega) - j\sigma_2(\omega)$, where

$$\begin{aligned}\frac{\sigma_1}{\sigma_\mathrm{N}} &= \frac{2}{\hbar\omega}\int_\Delta^\infty \frac{E^2+\Delta^2+\hbar\omega E}{\sqrt{E^2-\Delta^2}\sqrt{(E+\hbar\omega)^2-\Delta^2}}\left(f(E)-f(E+\hbar\omega)\right)\,\mathrm{d}E \\ \frac{\sigma_2}{\sigma_\mathrm{N}} &= \frac{1}{\hbar\omega}\int_\Delta^{\Delta+\hbar\omega} \frac{E^2+\Delta^2-\hbar\omega E}{\sqrt{E^2-\Delta^2}\sqrt{\Delta^2-(E-\hbar\omega)^2}}(1-2f(E))\,\mathrm{d}E\end{aligned} \tag{3.7}$$

where $f(E)$ is the distribution function for quasiparticles, which in thermal equilibrium is given by the Fermi-Dirac distribution $f(E) = 1/(\exp(E/k_\mathrm{B}T)+1)$. The quasiparticle density follows as

$$n_\mathrm{qp} = 4N_0\int_\Delta^\infty \frac{E}{\sqrt{E^2-\Delta^2}}f(E)\,\mathrm{d}E, \tag{3.8}$$

where N_0 is the single-spin density of states at the Fermi energy. As the temperature decreases $f(E)$ for $E \geq \Delta$ decreases exponentially, since for $f(E)/f(\Delta) \approx exp((\Delta-E)/k_\mathrm{B}T) \leq 1$. So as temperature decreases, the gap suppression $\Delta - \Delta_0$, the quasiparticle density n_qp, and the real conductivity σ_1 vanish exponentially. In the low frequency limit we can find an analytic solution to the variation in $\sigma_2/\sigma_\mathrm{N}$ from equation 3.7 as a function of temperature [53]. This analytic solution is

$$\frac{\sigma_2}{\sigma_\mathrm{N}} = \frac{\Delta}{2\hbar f}\tanh\frac{\Delta}{2k_\mathrm{B}T}. \tag{3.9}$$

We have that $\sigma = \sigma_1 - j\sigma_2$, which we can convert to an admittance by scaling by A/l, where A is the cross-sectional area of our nanowire and l is the length. Thus, we have

$$Y = \frac{A}{l}(\sigma_1 - j\sigma_2). \tag{3.10}$$

Or equivalently as an impedance $Z = 1/Y$, which expands to give

$$Z = \frac{l}{A}\frac{\sigma_1+j\sigma_2}{\sigma_1^2+\sigma_2^2}. \tag{3.11}$$

From inspection, this is equivalent to a resistor in parallel with an inductor $Z = R||j\omega L_\mathrm{K}$.

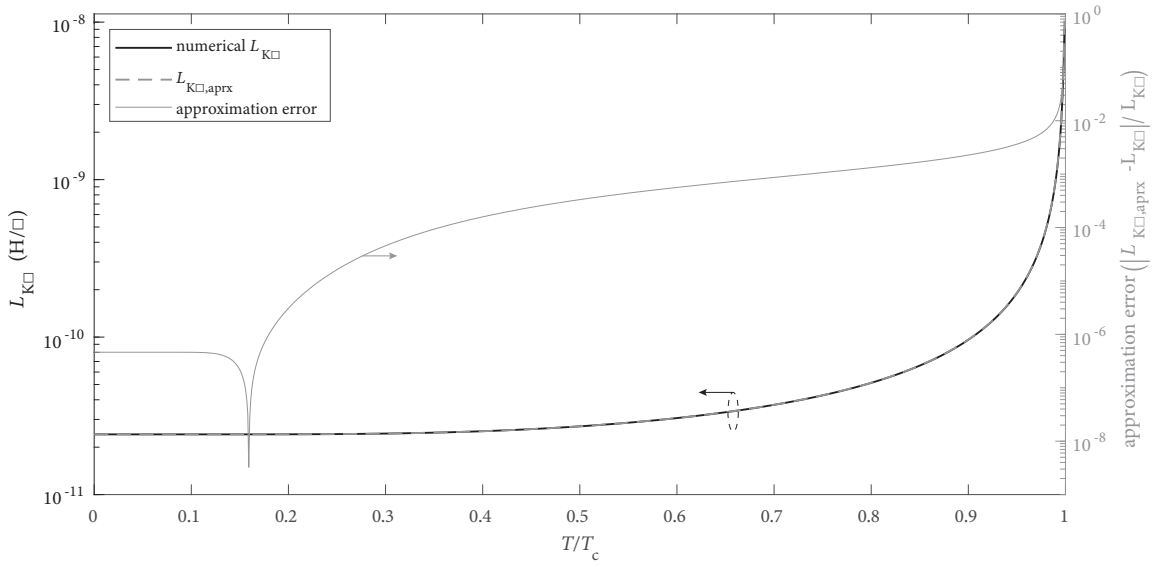

Figure 3-2: Comparison of the kinetic inductance per square as calculated numerically from the Mattis-Bardeen equations for complex conductivity, and the approximation $L_{\mathrm{K}\square,\mathrm{aprx}}$ given in equation 3.14. This simulation was conducted with typical parameters for NbN, $T_\mathrm{c} = 10\,\mathrm{K}$, $R_\mathrm{sq} = 175\,\Omega$, and at a frequency of $f = 1\,\mathrm{GHz}$. It can be seen that the error is low, less than 1 % for all temperatures less than $0.98T_\mathrm{c}$.

For this model, the inductance is

$$L_\mathrm{K} = \frac{l}{\omega A \sigma_2}. \tag{3.12}$$

Using our expression in equation 3.9, we have

$$L_\mathrm{K} = \frac{l}{A\sigma_\mathrm{N}} \frac{\hbar}{\pi\Delta \tanh\left(\frac{\Delta}{2k_\mathrm{B}T}\right)}. \tag{3.13}$$

We can handle the normal conductance by using $\sigma_\mathrm{N} = 1/R_\mathrm{sq}t$, where t is the film thickness. As we find that we will end up with l/At, we can convert our inductance into a per-square value as

$$L_{\mathrm{K}\square} = \frac{R_\mathrm{sq}\hbar}{\pi\Delta \tanh\frac{\Delta}{2k_\mathrm{B}T}}, \tag{3.14}$$

where $L_\mathrm{K} = L_{\mathrm{K}\square} l/w$ with w being the width of the nanowire.

Relating the expression in equation 3.13 back to our generic equation for a kinetic inductor, equation 3.2, we find that $n_\mathrm{s} \propto \Delta \tanh(\Delta/2k_\mathrm{B}T)$. Now if we consider T near

T_c, then this can be simplified to $n_s \propto \Delta^2$. Using the approximation $\Delta(T \to T_c) \approx 3.06 k_B T_c \sqrt{1-(T/T_c)}$ from section 2.2.2 we can further simplify this to $n_s \propto 1-T/T_c$. Thus, for T near T_c then $L_K(T \to T_c) = L_K(0)/(1-T/T_c)$, which is exactly the result from GL theory given in equation 3.6. Thus, since GL theory result is a special case of the full expression given in equation 3.13, we will proceed using the expression for kinetic inductance given in equation 3.14.

In order to confirm that the approximations we made in forming equation 3.14 are in fact accurate for our frequencies, we can solve the Mattis-Bardeen equations as stated in equation 3.7 numerically and compare the results to our analytic approximations. In this simulation, we used the gap approximation from equation 2.11. The results of the simulation are shown in figure 3-2. It can be seen in this figure that the results match extremely well up to within very near T_c, thus verifying our approximation's consistency with the original Mattis-Bardeen equations.

3.1.3 Direct measurement of nanowire kinetic inductance

There are a number of methods of measuring inductance, and in particular inductivity of a superconducting material [35, 71]. Many of these approaches use the resonance of either the structure or a capacitance forming either a distributed or lumped resonator. While these methods have their advantages, and can yield very high accuracy, there are some downsides. For resonant structures, simulations are often required to extract the inductance from the results, and for distributed or lumped resonators, the capacitance needs to be known accurately. To avoid these necessities, and since at the time of these experiments we did not have the process described in section 3.2, we opted for direct measurement of the inductance by treating it as a lumped inductor and fitting parasitic terms. In section 3.4, we will measure kinetic inductance by means of a lumped resonator.

We model the nanowire as an inductor to which we add the wire-bond inductance and a stray capacitance term. The model of the load is shown in figure 3-3 along with the experimental setup. In this model, the load impedance is $Z_L(f) = 2\pi f(L_K + L_s)j || 1/2\pi f C j$, giving rise to the reflection coefficient at the end of the transmission

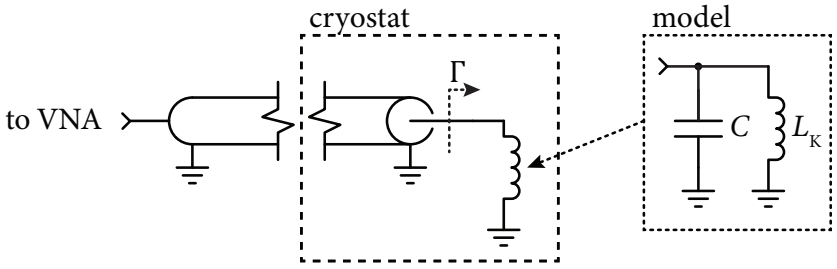

Figure 3-3: Experimental setup for measuring the inductance of nanowires. A VNA directly measures the phase response of the inductor connected to the end of a transmission line. The inductor is modeled as a parallel combination of a parasitic capacitor C, the kinetic inductance L_{K}.

line of $\Gamma(f) = (Z_L(f) - Z_0)/(Z_L(f) + Z_0)$. With this we can find that the phase of the S_{11} parameter is given as

$$\angle S_{11}(f) = \angle\Gamma(f) + 2\pi f \Delta t + \phi_0 \tag{3.15}$$

where Δt is the time delay incurred by the cable and ϕ_0 represents a phase calibration error.

With this setup, we simply take the phase data from the vector network analyzer (VNA) and fit it to our model given by equation 3.15 to determine the two fitting parameters Δt and ϕ_0, as well as our desired value of inductance L. We then repeat this process as the device temperature and bias current are swept. In order to avoid over-fitting, the value of Δt was allowed to vary in one run, the mean of the term found, and the mean used in the final determination of the inductance. The value of ϕ_0 was allowed to be free as the experiment took many hours and the phase drifted over this time.

In order to fit the extracted inductance-versus-temperature data to our model, we used the known sheet resistance, number of squares, and allowed the T_{c} to be varied. For the BCS fit at zero current, we used a cost function weighted towards matching low temperature values over higher temperature values. This is because we anticipate a deviation at higher temperatures, as will be discussed. The results of this fitting procedure are shown in the right of figure 3-4.

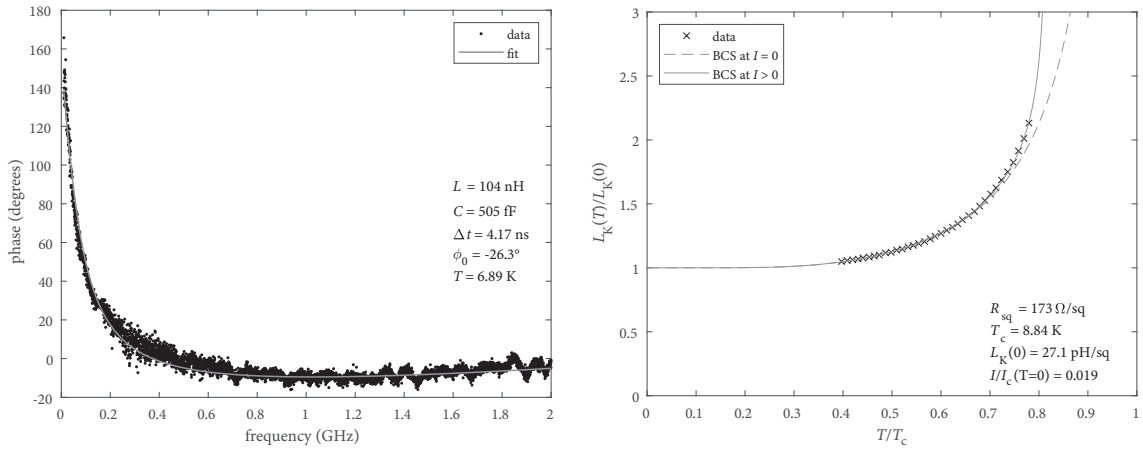

Figure 3-4: (left) One of the many S_{11} phase data sets and the fit calculated based on the model in equation 3.15 used to find each point on the right figure. (right) A comparison between the inductance as extracted in the method shown in the left figure, compared to the BCS predicted kinetic inductance at no bias current (dashed) and the BCS prediction at finite current (solid).

It can be seen that in figure 3-4 the BCS prediction of the kinetic inductance from equation 3.14 does not match well for higher temperatures. This is due to the stimulus from the VNA resulting in a non-zero current at the measurement time. With this in mind, when performing the experiment we set the VNA for the lowest output power that will give us good results with acceptable noise. However, as we increase the device temperature, we decrease the device switching current. Thus, the fraction of the device depairing current at which we are operating increases. We can compensate for this effect by using the expression

$$L_{\mathrm{K}}(T,I) \approx \frac{L_{\mathrm{K}}(T,I=0)}{\left(1-\left(\frac{I}{I_{\mathrm{c}}(T)}\right)^{n}\right)^{\frac{1}{n}}}, \tag{3.16}$$

where n depends on the fraction of T_{c} at which we are operating [35, 72]. Here we used a value of $n = 2.22$, which works well for temperatures around $0.6T_{\mathrm{c}}$ and varies little for higher temperatures, decreasing to $n = 2.11$ at T_{c}. For the switching current we use the approximation $I_{\mathrm{c}}(T) = I_{\mathrm{c}}(T=0)(1-T/T_{\mathrm{c}})^{3/2}$ from equation 2.13. Thus,

we have the complete approximation for the kinetic inductance as

$$L_{\mathrm{K}\square}(T,I) \approx \frac{\frac{R_{\mathrm{sq}}\hbar}{\pi\Delta(T)\tanh\frac{\Delta(T)}{2k_{\mathrm{B}}T}}}{\left(1-\left(\frac{I}{I_{\mathrm{c}}(T=0)\left(1-\frac{T}{T_{\mathrm{c}}}\right)^{\frac{3}{2}}}\right)^{n}\right)^{\frac{1}{n}}}. \tag{3.17}$$

In this we have an additional parameter, namely $I/I_{\mathrm{c}}(T=0)$. We fit this to our data giving us the result shown in figure 3-4.

3.2 Nanowire microwave device design elements

In order to design microwave devices, we need to pay special attention to the physical construction of the device as even small parasitic elements can have large effects. In this section, we will first present the layer stack-up used for the devices presented in the remainder of this chapter. We will then go on to discuss the design aspects of on-chip devices.

3.2.1 Device layer stack-up

There are two main options available for realizing on-chip transmission lines: the coplanar waveguide and the microstrip. For our applications, the main advantage of the former is that it does not require a ground plane. However, in order to form $50\,\Omega$ transmission lines, very wide structures are needed (on the order of $100\,\mu$m). On the other hand, microstrips can achieve the same impedance with much narrower line widths (on the order of $1\,\mu$m). The downside of the microstrip approach is that is requires a ground plane and a dielectric spacer, making fabrication more complex. Additionally, the presence of a ground plane allows for the formation of relatively high value capacitors. Thus, we proceed with the use of the ground plane and have a stack-up as shown in figure 3-5.

In selecting the dielectric spacer we are influenced by a number of factors. The thinner the spacer, the higher capacitance we can obtain, in a smaller space. Thus,

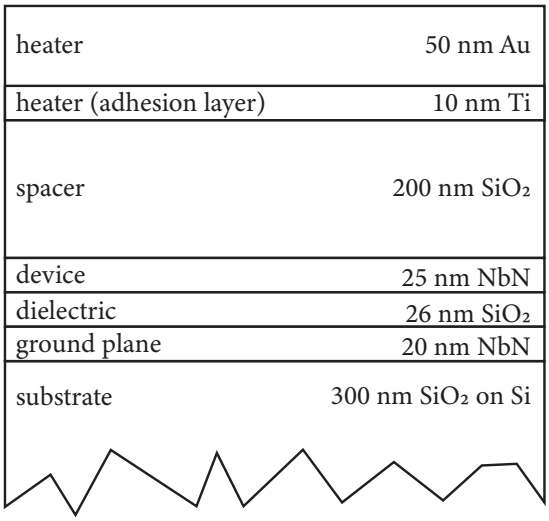

Figure 3-5: Nominal layer stack-up of the microwave toolbox chip. The ground plane and device layer are separated by a PECVD SiO_2 dielectric. On top of this arrangement is a thick oxide layer that separates the heater from the device layer. This stack-up is similar to that of the memory process with the addition of the thin dielectric and ground plane.

we can make microstrips narrower and shunt capacitors smaller. However, the smaller we make the spacer the more fabrication issues we are likely to run into. Possible issues include the existence of holes in the spacer, as well as the fact that the process variability increases with decreased thickness. We found that with our tools we could achieve a thickness of 26.5 nm relatively reliably. As such, we opted for this spacer thickness in our design.

The layer stack-up presented in figure 3-5 is heavily influenced by the memory work and the stack-up developed for that project – see section 2.2. In fact, the stack-up used here is nominally the same memory process with the addition of a dielectric spacer and ground plane. The heater spacer is slightly increased to reduce electromagnetic coupling to the heater.

3.2.2 Microstrip transmission lines

Now that we have the layer stack-up designed, we cam move onto determining transmission line dimensions. Our microstrip is of width W and height H and is embedded

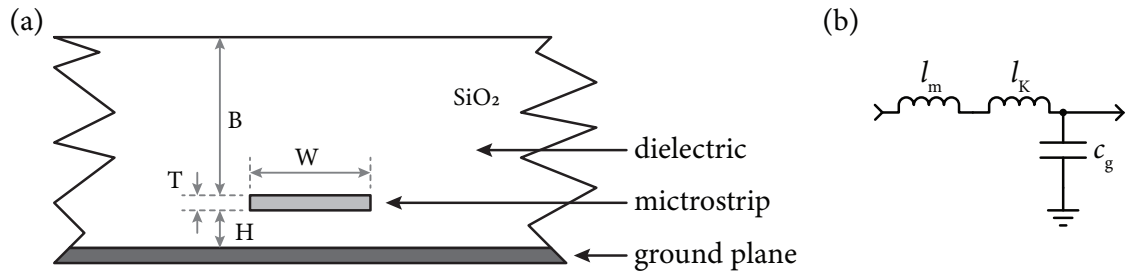

Figure 3-6: Schematic of the microstrip transmission lines. (a) The cross section of the transmission line with relevant dimensions shown. (b) Shows an element of the distributed model of the transmission line with per unit length magnetic inductance l_{m}, kinetic inductance l_{K}, and capacitance to ground c_{g}. For our transmission line $l_{\mathrm{K}} \gg l_{\mathrm{m}}$, and so we can normally ignore l_{m}.

in silicon dioxide (dielectric constant $\varepsilon_{\mathrm{r}} = 3.9$) and separated from the superconducting ground plane by a distance H. Further, the heater spacer, made from the same dielectric material, extends from the top of the microstrip a distance of B, as shown in figure 3-6. We will model our transmission line as lossless and with an elementary component given by the schematic shown in figure 3-6. If necessary, one could determine losses in the transmission line by using the calculations presented in section 3.2.3 and section 3.2.4.

Within the elementary component we have a capacitor representing the capacitance per unit length between the strip and ground c_{g}. Within the elementary component we also have two inductances, one being the kinetic inductance l_{K} per unit length, and the other being the magnetic inductance l_{m} per unit length. As discussed earlier in section 2.1, the kinetic inductance is much larger than the magnetic inductance. Here, the magnetic inductance will be around three orders of magnitude smaller than the kinetic inductance for the same length segment. As the magnetic inductance is negligibly small in comparison to the kinetic inductance, we will simply ignore the magnetic inductance term, $l_{\mathrm{m}} \approx 0$. Thus, we have only our kinetic inductance term $l_{\mathrm{K}} = L_{\mathrm{K}}/W$.

For the case when we have no heater above the microstrip we have an embedded microstrip structure for which we use the $\varepsilon_{\mathrm{e}} = \varepsilon_{\mathrm{r}}(1 - \exp(-1.55(H + B)/H))$ [73]. We find for our stack-up, the equation's solution is within 0.5 % of ε_{r}. Thus, we

simply use $\varepsilon_\mathrm{e} = \varepsilon_\mathrm{r}$. With this result in hand, we can now simply apply the standard microstrip capacitance equation

$$c_\mathrm{g} = \begin{cases} \frac{2\varepsilon_\mathrm{e}\varepsilon_0\pi}{\ln\left(\frac{8H}{W}+\frac{W}{4H}\right)} & \frac{W}{H} \leq 1 \\ \varepsilon_\mathrm{e}\varepsilon_0\left(\frac{W}{H}+1.393+0.667\ln\left(\frac{W}{H}+1.444\right)\right) & \frac{W}{H} \geq 1. \end{cases} \tag{3.18}$$

We will only need the expression for $\frac{W}{H} \geq 1$ since for all typical transmission line impedance and our stack-up we have $W \gg H$. In fact, our wires are so wide in comparison to the dielectric that we are able to avoid equation 3.18, resorting instead to the simple parallel plate equation $c_\mathrm{g} = \varepsilon_\mathrm{r}\varepsilon_0 W/H$, without significant error. For the design we will use for our $50\,\Omega$-line, equation 3.18 calculates less than $5\,\%$ more capacitance than the parallel plate approximation.

The impedance of the microstrip without the heater is given by

$$\begin{aligned} Z_0 &= \sqrt{\frac{l_\mathrm{K}+l_\mathrm{m}}{c_\mathrm{g}}} \\ &\approx \sqrt{\frac{L_\mathrm{K}}{\varepsilon_\mathrm{e}\varepsilon_0 W\left(\frac{W}{H}+1.393+0.667\ln\left(\frac{W}{H}+1.444\right)\right)}} \\ &\approx \sqrt{\frac{L_\mathrm{K}H}{\varepsilon_\mathrm{r}\varepsilon_0 W^2}}, \end{aligned} \tag{3.19}$$

where we can use either of the last two approximations. The benefit of the last approximation is that it is simple to solve for the width as $W \approx \sqrt{L_K H/\varepsilon_\mathrm{r}\varepsilon_0}/Z_0$, whereas for the slightly more accurate approximation the solution is more complex. For our stack-up and a system impedance of $Z_0 = 50\,\Omega$, we find that we need a width of $W_{50\,\Omega} = 2.35\,\mu$m for the more accurate solution or $W_{50\,\Omega} = 2.42\,\mu$m for our simple solutions.

With the microstrip designed, we can then determine the phase velocity from the expression $v_\mathrm{p} = 1/\sqrt{l_\mathrm{K} c_\mathrm{g}}$. For our stack-up, we find that the phase velocity is around $2\,\%$ the speed of light in a vacuum for practical widths – see figure 3-7. For our $Z_0 = 50\,\Omega$ microstrip of width $W_{50\,\Omega} = 2.42\,\mu$m we have $v_\mathrm{p}/c = 2.02\,\%$.

While in the designs presented here we only use $50\,\Omega$ transmission lines, it is of

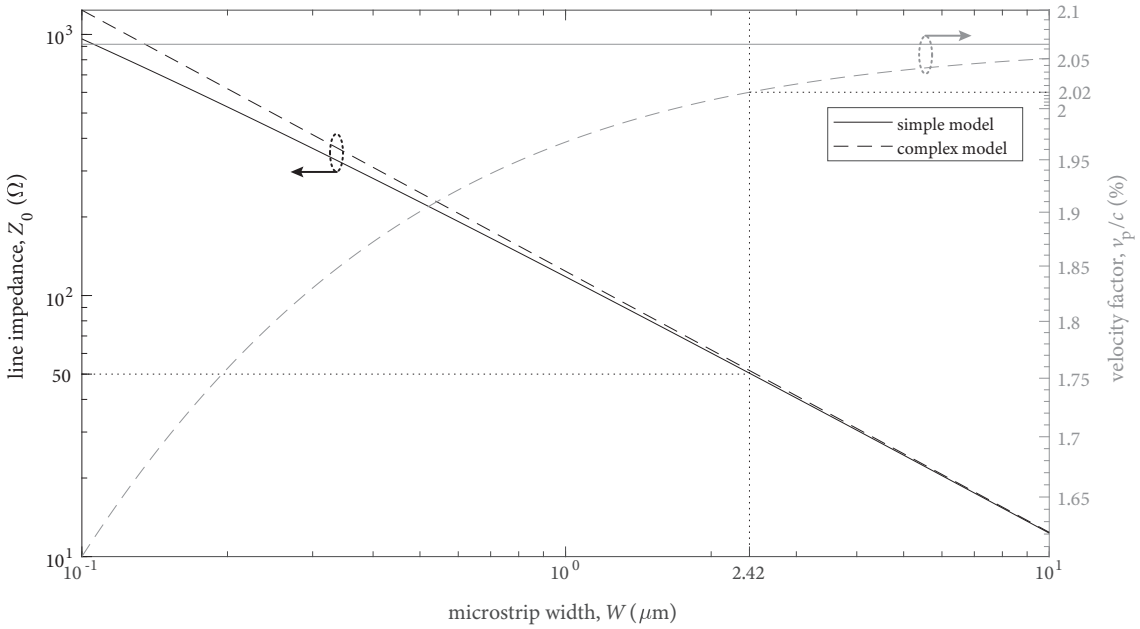

Figure 3-7: Plot of the characteristic impedance Z_0 and velocity factor v_p/c, for a microstrip of varying width (W from figure 3-6) for our layer stack-up. This calculation was performed with the simple and complex models. It can be seen that the simple model suffices for the line impedance. On the other hand the more complex model must be used for the velocity factor. The simple model's velocity factor is the asymptote to which the complex model is approaching, and at common values the difference is still relatively large. The line impedance decreases as the width increases as for wider widths we have smaller inductance and higher capacitance to ground. The values for a line impedance of $Z = 50\,\Omega$ is found to occur for a width of $W = 2.42\,\mu$m, which gives a velocity factor of $v_\mathrm{p}/c = 2.02\,\%$.

course possible to make any microstrips of other impedance. We are limited only by the minimum feature size we can fabricate reliably.

3.2.3 Inductors

We can form lumped inductors by simply removing the ground plane from beneath the nanowire. This is essentially equivalent to taking the microstrip model shown in figure 3-6 and removing the ground and hence the c_g term. We of course will have stray capacitance to other structures and the wire itself; however, these are very small in comparison to the inductance. Then, we can simply calculate the inductance of a nanowire by using 3.14.

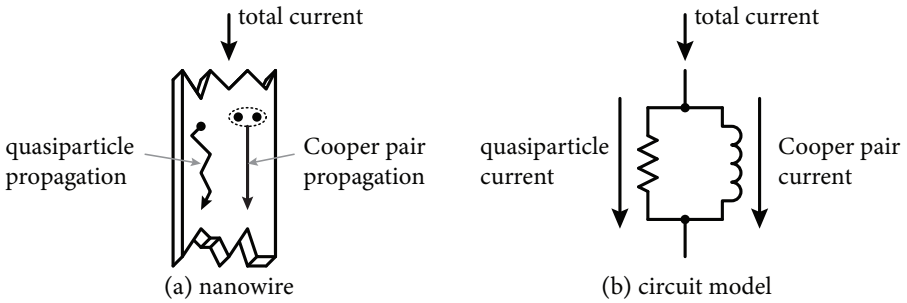

Figure 3-8: Model of the current through a superconducting nanowire following the two-fluid model. Within the nanowire both the quasiparticles (normal electrons) and Cooper pairs participate in the total current through the wire. (a) Shows a nanowire depicting an abstract concept of the propagation of Cooper pairs with their zero resistance and large kinetic energy, and quasiparticles with their non-zero resistance and small kinetic energy. (b) Shows the circuit model of the wire where the quasiparticle conduction is modeled as a resistor and the Cooper pair conduction is modeled as an inductor. If we were to model magnetic inductance, we would add that term in series with this model (but here we ignore this term).

In order to understand losses in our inductor, we can rely on the two-fluid model which follows from the Mattis-Bardeen equations for the complex conductivity presented in section 3.1 [74]. In section 3.1, we were only interested in the imaginary term of the complex conductivity σ_2, as that represented our kinetic inductance. When calculating the kinetic inductance, we assumed the real part of the conduction was so small in comparison to the imaginary part that we could make the assumption $\sigma_1 \approx 0$. While this assumption holds for most circumstances, when considering circuits that are sensitive to real losses, we need to treat this real term in our analysis. To the circuit designer we can introduce this loss as shown in figure 3-8, where the real conduction term is modeled as a resistor and the inductor represents our kinetic inductance.

For microwave devices, we are often very interested in the losses of a device as they have implications on system bandwidth and Q in general. Thus, an expression for the Q of the inductor would be very useful. Here, we will define the Q of an inductor as the Q of an RLC circuit with an ideal capacitor. This circuit is tuned for resonant frequency ω. Following this model, the Q of an inductor is given by $Q_{\mathrm{L}} = R/X_{\mathrm{L}}$, where X_{L} is the inductor reactance and R is the inductor parallel resistance. The

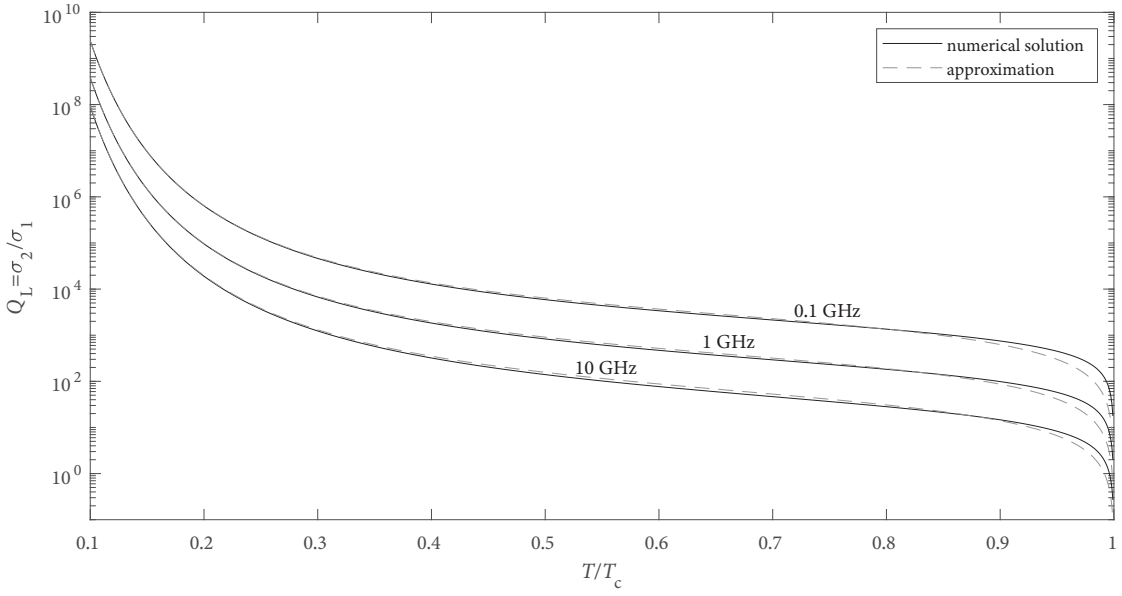

Figure 3-9: Plot of the inductor Q from the Mattis-Bardeen equations complex conductivity. Here we compare the approximation of the Q given by equation 3.22 with the numerical solution to the Mattis-Bardeen equations. It can be seen that up to around $0.95T_c$ the approximation and the numerical solution agree well.

parallel model is used to be consistent with the circuit given by the two-fluid model, as shown in figure 3-8. In section 3.1, we performed a decomposition of the two-fluid nanowire model impedance into an inductive reactance $X_L = l/A\sigma_2$, and resistance $R = l/A\sigma_1$. This, combined with our expression for the Q, yield

$$Q_L = \frac{\sigma_2}{\sigma_1}. \tag{3.20}$$

This ratio can be found from numerically solving the Mattis-Bardeen equations as shown in equation 3.7. However, an analytic approximation that will be good for most applications can be found.

From our treatment of kinetic inductance in section 3.1, we have an expression for σ_2/σ_N in equation 3.9. This expression assumes $\sigma_2 \gg \sigma_1$, which still holds here assuming our Q-factor is much larger than 1. For an expression for σ_1 we can use the

expression from [75] for low-temperatures $k_{\mathrm{B}}T < \Delta$, and low frequency $\hbar\omega < \Delta$,

$$\frac{\sigma_1(T)}{\sigma_{\mathrm{N}}} \approx \frac{4\Delta_0}{\hbar\omega} e^{-\frac{\Delta_0}{k_{\mathrm{B}}T}} \sinh\left(\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right) K_0\left(\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right), \tag{3.21}$$

where $K_0(\cdot)$ is the 0^{th}-order modified Bessel function of the second kind. Using this expression, we can find the low-temperature and low-frequency Q is

$$Q_{\mathrm{L}} \approx \frac{\pi\Delta \tanh\frac{\Delta}{2k_{\mathrm{B}}T}}{4\Delta_0 e^{-\frac{\Delta_0}{k_{\mathrm{B}}T}} \sinh\left(\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right) K_0\left(\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right)}. \tag{3.22}$$

Figure 3-9 shows a comparison between this approximation and the numerical solution to the system of equation 3.20 and equations 3.7.

3.2.4 Capacitors

To form capacitors, we simply make a patch to ground, or equivalently a wide line (much wider than the microstrip for the characteristic impedance). We must be conscious of the parasitic inductance, and any transmission line effects, such that we will have a device that operates as a capacitor over the frequency range of interest. All of our capacitors have dimensions greater than $100H$, where H is the spacing between the device layer and ground plane. So, we can simply use the parallel plate approximation and ignore fringing fields. As such, we simply have that we calculate our capacitance as

$$C = \varepsilon_{\mathrm{r}}\varepsilon_0 \frac{A}{H}, \tag{3.23}$$

where A is the area of our capacitor plate.

Given that we will be interested in the quality factor of the devices we construct, we need to develop some treatment for the losses. For a practical capacitor we can model the device as an ideal capacitor in series with a resistor. The equivalent series resistor R_{ESR} represents all forms of loss in the device lumped into a single term. We will assume that our terminals and ground plane have very little losses. For a generic dielectric, the material will be described by a conductivity σ_{d} and a complex

permittivity $\varepsilon_{\mathrm{d}} = \varepsilon'_{\mathrm{d}} - j\varepsilon''_{\mathrm{d}}$. The imaginary term in the permittivity represents the losses from bound charges and dipole relaxation, and the conductivity represents the conductive losses not included in the complex permittivity. Defining a wave as

$$\mathbf{E} = \mathbf{E}_0 e^{j\omega t}, \tag{3.24}$$

then by Maxwell's extension of Ampère's circuital law with $\mathbf{J} = 0$ we have

$$\nabla \times \mathbf{H} = j\omega\varepsilon'_{\mathrm{d}}\mathbf{E} + (\omega\varepsilon''_{\mathrm{d}} + \sigma_{\mathrm{d}})\mathbf{E}. \tag{3.25}$$

This equation models the loss in the real term, and as such we can formulate a measurement of this loss as a ratio to the real and imaginary components

$$\tan\delta = \frac{\omega\varepsilon''_{\mathrm{d}} + \sigma_{\mathrm{d}}}{\omega\varepsilon'_{\mathrm{d}}}. \tag{3.26}$$

We use $\tan\delta$ to represent this as that ratio is the tangent of the angle δ between the lossy reaction and lossless reaction in equation 3.25. As mentioned earlier, the Q-factor is 2π multiplied by the ratio of the energy stored over the energy lost per cycle. Given the definition of $\tan\delta$ as the inverse of this, we find that $\tan\delta = 1/Q_{\mathrm{C}}$. Thus, for our capacitor we can determine the Q simply from the dielectric tangent term for PECVD deposition silicon dioxide, $\tan\delta \approx 2.8 \times 10^{-3} \Rightarrow Q_{\mathrm{C}} \approx 350$ [76].

3.2.5 Current crowding

It is important to avoid sharp corners in superconducting structures biased at high fractions of their critical current, and for structures where non-uniform currents are possible (at least where the dimensions of these abrupt transitions are much larger than the coherence length) [9, 77, 78]. Bends cause the current to preferentially flow through the shorter inner path within the bend. If this bend is smooth (filleted) then, while the current density will increase, the reduction in the critical current will be limited. However, if the bend is sharp, then on the sharp point the current density can be much greater than throughout the rest of the film, possibly leading to premature

switching. This is simply mitigated by the use of rounded corners on the inside edge of bends. Through the use of optimally rounded curves, as presented in [77], the bends in a structure can result in no reduction in critical current.

In practice, if we are not trying to operate at high fractions of the bias current in the straight sections of the wire, then we can just fillet or "teardrop" transitions in width or shape with a smooth shape. While this transition will not be optimal, and will result in a higher current density, it is often more convenient when laying out a chip. In this work we will use circular fillets to form the transitions between sections that are non-critical.

For transmission lines, the typical method of making right-angle bends is to miter the corner [79]. The reason for this is that in a bend, say of 90°, the outer-most section of wire is effectively making the wire more capacitive than is required for the desired impedance. Thus, by cutting off the tip of the corner one can maintain a better approximation to a straight section of line with no bend. However, that leaves a sharp inner corner which would not be desirable here. A simple method of overcoming this issue is to make circular bends in our transmission lines. In order for these bends to be nearly indistinguishable from a straight line, we need the radius to be large. Typically, a bend of radius three times the line width is used and is nearly identical in performance to a straight line [80].

3.3 General experimental setup

For the measurements performed in this section we employ a VNA. Given that there are many errors intrinsic to any VNA measurement setup, we must perform a calibration procedure to compensate for these errors. In doing so, we (at least partially) de-embed the device from the experimental setup. Calibration typically involves the presentation of a number of different loads to the VNA and comparing the actual response to the expected response. For measurements in cryostats, the process is somewhat complicated by the fact that we cannot easily access the terminals where the device under test (DUT) is attached. Thus, careful experiment design is re-

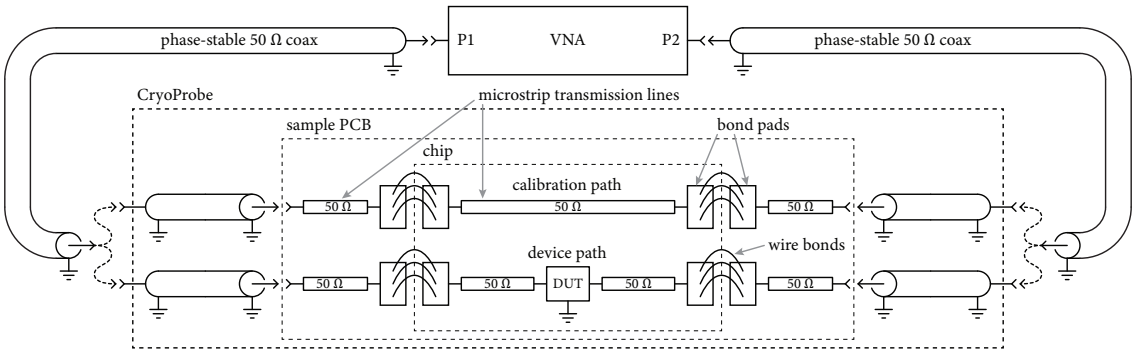

Figure 3-10: Schematic diagram of the experimental setup for microwave measurements. The calibration path and device path are identical with the only difference being the presence of the DUT in the device path. This allows for de-embedding of the apparatus and extraction of only the DUT response. The cables within the probe all belong to the same bank and as such are phase matched. The room-temperature cables are phase-stable to avoid errors due to cable flexure. Note that the ground of the chip is bonded to the ground of the PCB in many locations and that this is not shown in the figure.

quired in order to obtain only the DUT response and not the fixture DUT response embedded in the fixture response. However, regardless of how carefully we design our process, we will never be able to fully de-embed the setup given that we cannot switch the load. Furthermore, without using more advanced de-embedding methods that take into account higher-order terms, we will always see some fixture response in our measurements [81].

The experimental setup we used for the results in the section is shown in figure 3-10. Here we used phase stable low-loss coax cables attached to the ports on the VNA. These cables must be phase stable as we will need to move them between ports on the CryoProbe during the experiment. First, we calibrate to the end of these cables by performing a short open load through (SOLT) calibration. Now we have the reference plane at the end of the cables. Next, we need to compensate for the fixture up to the DUT.

Given that the cables within the CryoProbe are phase matched within a bank within the probe (and have a good phase match between banks as well), we can use a separate calibration path and device path. We capture the response of the calibration path and save the data. We then move the connectors at room temperature to the

device path and capture the device response embedded in the fixture response. In post-processing we de-embed the DUT response from the fixture response. This procedure relies on the both the calibration path and device path to be identical in all ways other than the presence of the DUT in the device path. This is not possible in reality. The cables, while phase matched, do not have differing responses. The wire bonds are completed by hand and, while the chip was designed with the same distance between the PCB pad the chip bond pad, they are inevitably going to be slightly different. However, without on-chip switching to calibration loads, we cannot remove these errors experimentally. Thus, we will see some fixture response in our measurements which without using more complex calibration procedures is unavoidable.

3.4 A tunable resonator

As a first demonstration of an on-chip integrated tunable device, we present a tunable resonator. The following sections cover the design, layout, and experimental results for this resonator. The experimental results, and in particular the fits that relate the heater current to device temperature will prove invaluable later in this work.

3.4.1 Design

The resonator design is a simple series LC circuit with the inductor realized by means of a tunable kinetic inductor, namely an hTron. We directly couple the resonator to a transmission line in a shunt configuration. In doing so, we essentially create a notch filter. We can then measure the insertion loss of this filter, and calibrating out the setup response, we can extract the resonator response. A schematic of this design is shown in figure 3-11.

The resonator was designed for a frequency of $f_c = 1\,\mathrm{GHz}$ at no heater current. Our film was specified for $L_{K\square} = 20\,\mathrm{pH}/\square$. To achieve our desired center frequency require $LC = 1/(2\pi f_c)^2$. We are free to choose any L and C that satisfies this equation; however, the resonator response will vary based on our selection. In particular,

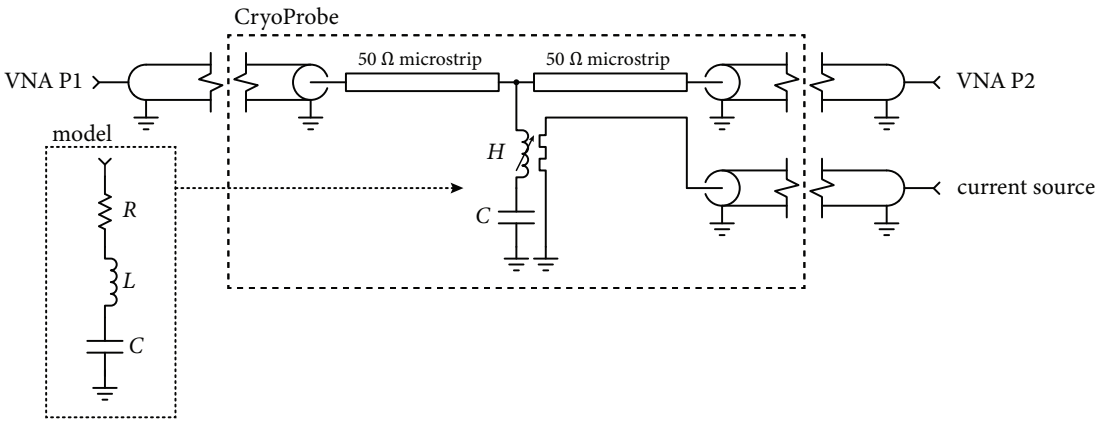

Figure 3-11: Schematic of the resonator and the experimental setup in which it was tested. The resonator is implemented as a series combination of an hTron variable inductor and a capacitor to ground. The resonator is coupled to a microstrip transmission line for measurement. Each end of the microstrip is connected to a port of the VNA. The inset shows the circuit model of the resonator for quality factor estimates.

the quality factor will be influenced by our choice. While we are not attempting to create a high-Q resonator here, rather we only wish to demonstrate tunability, any discussion of a resonator without addressing the quality factor would be incomplete. Thus, here we will present a brief treatment of the resonator Q.

Our device is present as a shunt element on a transmission line and is directly connected to it. Thus, we are strongly coupled to the transmission line, which means that energy can be coupled into and out of the resonator readily. This leads to a poor quality factor of the system. That is, the loaded quality factor will be

$$Q_{\mathrm{s}} = \frac{1}{\frac{1}{Q_{\mathrm{i}}} + \frac{1}{Q_{\mathrm{cp}}}}, \tag{3.27}$$

where Q_{i} is the unloaded (internal) quality factor, and Q_{cp} described the coupling. Which, we intuitively expect to satisfy the approximation $Q_{\mathrm{s}} \approx Q_{\mathrm{cp}}$, as our losses to the $50\,\Omega$ transmission line will dominate the losses within the resonator – we prove this result next.

One of the notable features of the expressions for the Q of our inductors and capacitors is that, to within our approximations, they are independent of the inductor and capacitor size. Thus, for the unloaded Q of our resonator, we are free to chose

any L and C that meets our resonant-frequency requirement. This freedom does not extend to the loaded-Q. First, we will consider the unloaded Q of our resonator. With the inductor and capacitor Q from section 3.2.3 and section 3.2.4, we can combine them to find the total unloaded Q, Q_{i}. This is achieved using the expression

$$Q_{\mathrm{i}} = \frac{1}{\frac{1}{Q_{\mathrm{L}}} + \frac{1}{Q_{\mathrm{C}}}}, \tag{3.28}$$

for which we find that $Q_{\mathrm{i}} \approx 220$, largely limited by our PECVD-deposited silicon dioxide dielectric spacer. So, for our unloaded Q we expect a value around 220, which is higher than those typically obtained for non-superconducting materials [82].

Now that we know the unloaded Q is high, we can address the coupling. Modeling the resonator in the transmission line setup, the loaded Q can be approximated by

$$Q_{\mathrm{s}} \approx \sqrt{\frac{L}{C}}\frac{2}{Z_0} = \frac{1}{\pi f_{\mathrm{c}} Z_0 C}, \tag{3.29}$$

where Z_0 is the system impedance which is assumed to be real and positive. We can make this approximation because $Q_{\mathrm{i}} \gg Q_{\mathrm{s}}$, which we know to be true since for the internal losses to match the coupling losses we would need $L = 4 \times 10^7 C$, which is not practical.

Thus, we are again in a situation where our selection of L and C is largely arbitrary as it will not greatly affect unloaded Q, and the loaded Q will always dominate. Thus, we select them to be of convenient sizes and with both inductor and capacitor large enough that the parasitics will not be significant. We select $L = 16.67\,\mathrm{nH}$, and so we have $C = 1.519\,\mathrm{pF}$. Thus, our loaded Q from equation 3.29 will be $Q_{\mathrm{s}}(f_{\mathrm{c}}) \approx 4.19 \times 10^9 / f_{\mathrm{c}}$, and of coarse our unloaded Q will remain unchanged at $Q_{\mathrm{i}} \approx 220$.

3.4.2 Layout

With the device values determined, we move onto the layout. The final layout for the device is shown in figure 3-12. It is a relatively simple implementation of the circuit

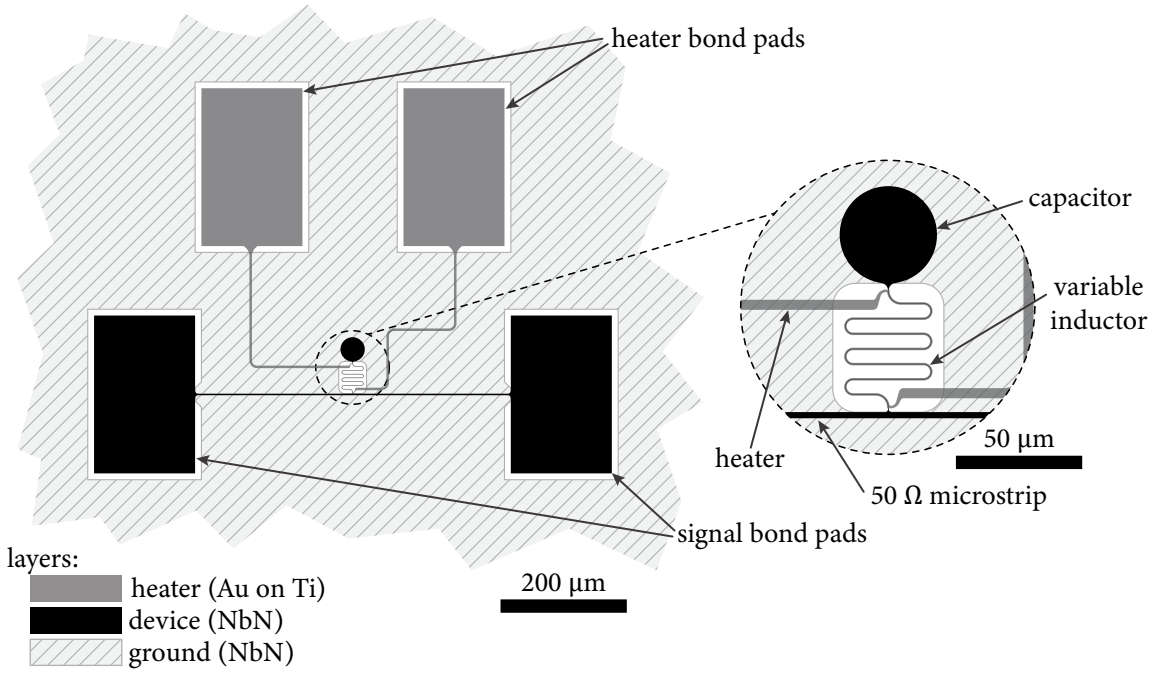

Figure 3-12: Layout of the resonator with its bond pads shown. The heater layer is shown in gray and transparent so that the features beneath it can be seen. The device layer is shown in black. The ground plane is shown light gray with a cross hatch pattern. In between each of these layers is a dielectric spacer which is not shown. Note that the ground plane is pulled back around the inductor and beneath the bond pads. The variable inductor is implemented as a 300 nm wide nanowire with a 1 μm wide heater over it.

shown in figure 3-11. To save space, the inductor is meandered. The capacitor is implemented as a circular patch with the ground plane beneath it. The connections to the microstrip and capacitor are filleted to reduce current crowding effects.

The inductor was selected to be 300 nm wide, which is somewhat arbitrary. Narrower width wires with the same kinetic inductance would lead to shorter inductors, but lower critical currents. Additionally, the fractional error in the width increases as the trace width is made more narrow. A width of $W = 300\,\text{nm}$ is a small but relatively reproducible width. We calculate the inductor length simply from $l_\text{L} = L_\text{K} W / L_{\text{K}\square}$. The heater over the inductor was selected to be 1 μm wide to allow for alignment tolerance between the layers. For the capacitor we simply used the parallel plate capacitor calculation. The result is shown in figure 3-12.

When wire bonding, it is relatively easy to bond through thin dielectrics. We

reliably bond through a silicon dioxide layer more than 100 nm thick to a 10 nm NbN layer below. Thus, to prevent shorts to ground, the ground plane must be removed from underneath all bond pads. Clearly, this must be done for the device layer as the capacitance of the pad would cause issues, but it is also necessary for the heater layer. This pullback of the ground plane around bond pads can be seen in figure 3-12.

The connections between the pads and the transmission line, the transmission line and the inductor, and the inductor and the capacitor are filleted. These fillets are an effort to reduce current crowding – see section 3.2.5. The heater is only narrow over the inductor and wide for the interconnect runs, thus reducing unnecessary heating. Additionally, the heater is kept away from structures that are not to be heated (such as the transmission line). Additional, more minor design considerations include the expansion of ground pullbacks to reduce parasitics and make the design tolerant to inter-layer misalignment, as well as the heater being wider than the nanowire to allow for alignment errors.

3.4.3 Results

With the device installed in the CryoProbe, and using the calibration procedure outlined in section 3.3, we performed two experiments. First, we performed a temperature sweep of the whole chip using the probe's temperate control features (see section 4.2). Next, we operated the device at base-temperature and swept the heater current. This is necessary as our theory is temperature-based, and so we need a method of relating the heater current to device temperature. As the materials are not simply modeled as would be done for room-temperature devices, this experiment is vital to match theory to the results. The expression for heater current to device temperature developed here will be relied on in the following sections.

Temperature sweep

The results of the temperature sweep are shown in figure 3-13. The resonator frequency is lower than designed at around 800 MHz compared to the designed value of

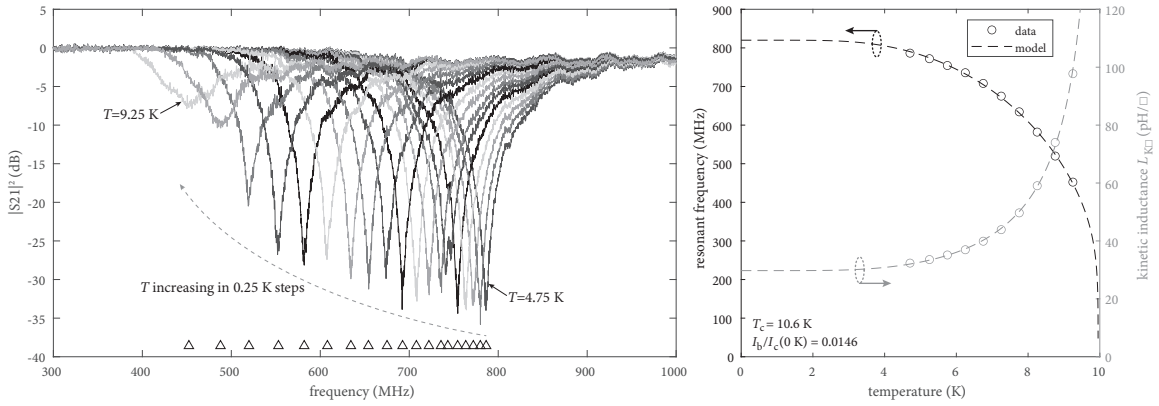

Figure 3-13: Response of the resonator to temperature variation. (left) The insertion loss of the resonator for various device temperatures. The markers show the extracted resonant frequencies. (right) Resonant frequency as a function of temperature (left axis) and the extracted kinetic inductance using the designed capacitance (right axis). A fit based on our kinetic inductance variation with temperature and current from equation 3.17, using the parameters shown is also provided. It can be seen that as the temperature increases, the resonant frequency progressively decreases over a large range of 340 MHz for a temperate change of 4.5 K.

1 GHz. This is due to either the kinetic inductance of the film being larger than expected, the capacitor having a higher value than designed, or a combination of both. For the comparisons to theory we perform here, it is not particularly important as we are only considering the ratio of changes; to these ends, we assume the capacitor value is as designed. However, as all devices investigated in this work were fabricated on the same chip, this means that all devices will not be operating at their designed frequency.

In order to determine the appropriate parameters for our later work, we needed to find a T_c and a bias current I_b that fit our expression in equation 3.17 to the results in figure 3-13. This was again performed by using the same optimizer used in section 2.2.2 and described in [5]. The fit was performed on the resonant frequency using $f_r(T_\mathrm{c}, I_\mathrm{b}/I_\mathrm{c}(0\,\mathrm{K}), \alpha) = 1/2\pi\sqrt{\alpha L_K(T_\mathrm{c}, I_\mathrm{b}/I_\mathrm{c}(0\,\mathrm{K}))C}$, where α compensates for the frequency shift due to the LC product error mentioned above. The value of α is not important and simply scales the results. The cost function used in the fits performed here is a basic mean square error calculated at each data point. The resultant fit and its parameters are shown in figure 3-13.

Fitting the basic series LCR circuit shown in figure 3-11, we find that the unloaded Q at $T = 4.75\,\mathrm{K}$ is approximately $Q_\mathrm{i} \approx 223$. This result matches well with our prediction of a Q around $Q_\mathrm{i} \approx 220$. Directly measuring the 3 dB of the dip from its minimum we find $f_\mathrm{c} = 803.2\,\mathrm{MHz}$, and $\Delta f = 4.02\,\mathrm{MHz}$, giving $Q_\mathrm{i} \approx 200$, which is again close to our prediction. For the loaded Q we measure the 3 dB bandwidth of the device as a notch filter for which we find $\Delta f = 184\,\mathrm{MHz}$ for $T = 4.75\,\mathrm{K}$. This bandwidth corresponds to $Q_\mathrm{s} = 4.37$. From our prediction in the design section, we expect $Q_\mathrm{s}(f_\mathrm{c} = 803.2\,\mathrm{MHz}) = 5.22$, which agrees well with our measurement.

Heater current sweep

Next, the current I_H sweep with the probe at temperature of $T = 4.28\,\mathrm{K}$ was performed. The results of this sweep are shown in figure 3-14. We sweep the heater current from zero to 1 mA in 100 μA steps. This range was selected as the device switched for heater currents much higher than $I_\mathrm{H} = 1\,\mathrm{mA}$. The heater bias was provided by a precision and low-noise current source with a floating output to avoid any ground loops.

In order to model the relation of the heater current to the kinetic inductance, we need a relation between the heater current and inductor temperature T_L. In comparing figure 3-13 and figure 3-14, we can see that T_L is not proportional to I_H^2, as one might expect. At these low temperatures, the thermal characteristics of the material, including thermal conductivity and heat capacity, are highly dependent temperature. From the modeling performed in [8], we form the modified relation

$$T_\mathrm{L} = \left(9.4407 \times 10^{12} I_\mathrm{H}^3 + T_\mathrm{sub}^4\right)^{1/4}, \tag{3.30}$$

which is the result of fitting the data from the temperature and current sweeps. This equation, while treated here as only a fitting model, can be considered to be a result of assuming a heater sheet resistance that is increasing with increasing temperature, and for our temperatures, proportional to the heater current, and the simplified model from [8]. This equation was used to generate the model shown in figure 3-14. The

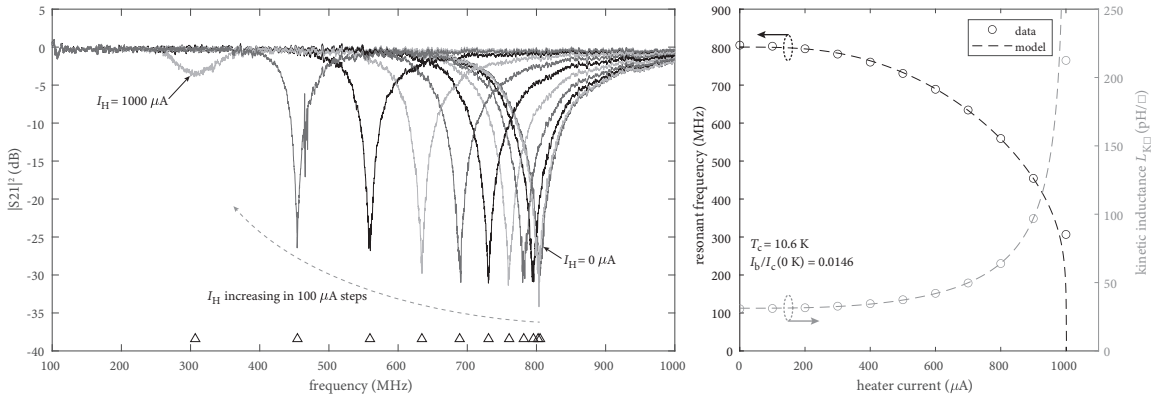

Figure 3-14: Response of the resonator to heater current variation. (left) The insertion loss of the resonator for various heater bias currents. The markers show the extracted resonant frequencies. (right) Resonant frequency (left axis) and the extracted kinetic inductance using the designed capacitance (right axis) as a function of bias currents. A fit based on our kinetic inductance variation with temperature and current from equation 3.17, and using the temperature sweep and the current-to-temperature relation in equation 3.30. It can be seen that as the heater current increases, the resonant frequency progressively decreases over a large range of 500 MHz for a heater current change of 1 mA.

data and model match very well at all heater currents other than the final point. The discrepancy in the final point is due to the resonator inductor switching at the higher bias currents which made extracting the resonant frequency for this point difficult. Given how well this model relates the results in this experiment, we will carry forward using equation 3.30 to relate heater current to inductor temperature.

3.5 Tunable low-pass filter

Now that we have established that an hTron can be used in microwave devices to realize tunability, we can move onto more complex devices. The first and most natural application would be tunable filters. Filter characteristics will change with varying only the inductor. We would ideally change the capacitance in addition to the inductance in order to maintain the filter response shape and impedance matching during tuning; however, here we only vary the inductor.

3.5.1 Design

We implement the filter in a Cauer (ladder) topology [83]. The parameters of the filter capacitors and inductors can be obtained by a number of methods. One could perform hand calculations based on prototype filters, or, as we did here, use filter design tools. For our first filter, we will implement a 9-pole Butterworth low-pass filter with a cutoff frequency of 1 GHz. The schematic for this filter is shown in part (a) of figure 3-15, and the component values are provided in table 3.1.

Within the Cauer topology and for an odd-order filter, we can select one of two designs, either one with a minimum number of inductors or a minimum number of capacitors. Here we opt for a minimum inductor design as it results in the first element loading the transmission line being a capacitor. For low-pass filters this can aid in high frequency rejection. Otherwise, the selection for our purposes is somewhat arbitrary.

The layout of the filter is fairly simple and follows similar practices to that discussed for the resonator. Here, the capacitors are implemented as radial stubs. Such designs are often used in more conventional microwave layouts to localize the loading on the line to which the capacitor is attached while also providing a broad-band response. The need for localized connections is less of an issue in high kinetic inductance materials. However, they are convenient for making compact filters and, provided that their resonant frequency is much higher than the operating frequency, will appear as a lumped capacitor. The layout of the filter is shown in figure 3-15.

The layout can be seen to have the same pad arrangement as was used in the resonator, as will be used for all other devices that fit within this structure. This allows us to use the same calibration line for all devices – see section 3.3. Similar to all other designs, the joins between elements are filleted to reduce current crowding. The heater is made wide for interconnects and kept away from structures not to be heated. Finally, the inductors are meandered to make the design more compact.

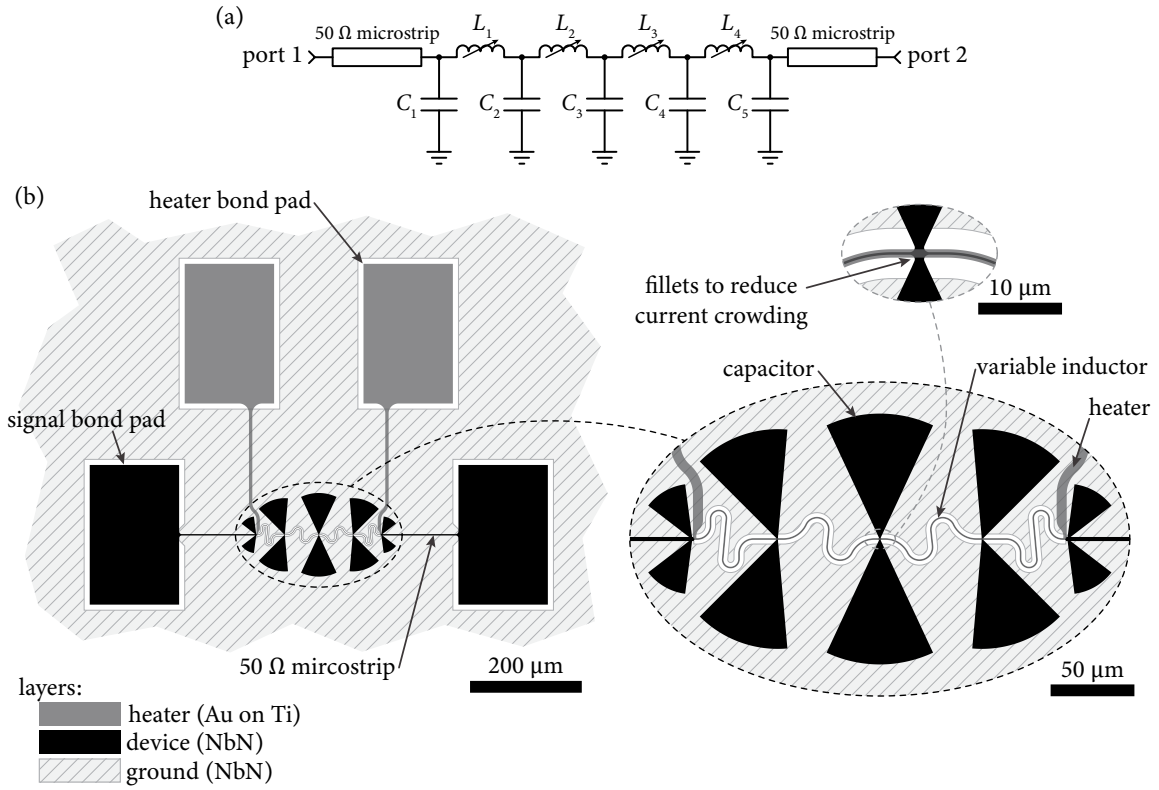

Figure 3-15: Schematic and layout of a 9-pole low-pass filter. (a) Schematic of the low-pass filter implemented here. The values of the components are provided in table 3.1. The heater is not shown in the schematic. (b) The layout of the filter. The bond pad configuration of is the same as that for the resonator, and the calibration structure. The radial stubs can be seen to be implemented in a symmetric manner with the inductors connecting them. To make the design more compact the inductors are meandered and the radial stubs angled. The inductor is 300 nm wide with a 1 μm wide heater over it. The connection between the stubs and the inductor is filleted to reduce current crowding.

Table 3.1: Butterworth 1 GHz 9-pole low-pass filter component values

Designator	Value
C_1, C_5	1.105 pF
C_2, C_4	4.877 pF
C_3	6.366 pF
L_1, L_4	7.958 nH
L_2, L_3	14.96 nH

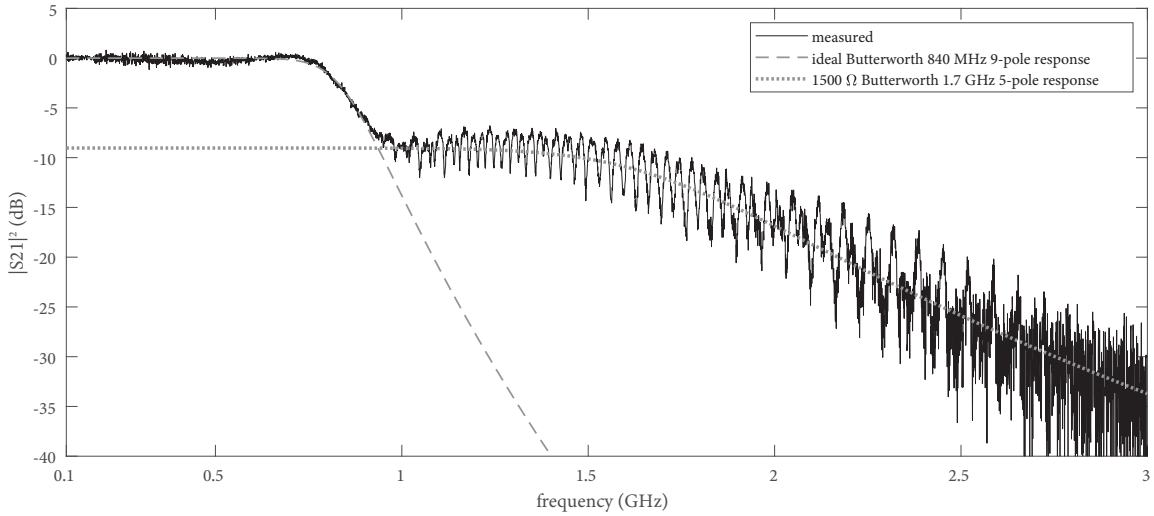

Figure 3-16: Response of the 9-pole low-pass Butterworth filter at the base-temperature of 4.28 K. In the pass-band the ripple is very small, and the cutoff frequency is 835 MHz. After the cutoff, the filter rolls off around the ideal 180 dB/dec to around −10 dB. We then see a second response, which we model here as a mismatched 1500 Ω 5-pole Butterworth filter at 1.7 GHz. This results in a roll off of 110 dB/dec. Additionally, the limitations of our calibration procedure are evident here. The periodic dips in the response are typical of a mismatch resulting in reflections and could be removed with more advanced de-embedding techniques.

3.5.2 Results

With the device installed in the CryoProbe, and using the calibration procedure outlined in section 3.3, we find the results shown in figures 3-16 and 3-17. Figure 3-16 shows the response of the filter at the base temperature of around 4.28 K. As this filter was fabricated on the same chip as the resonator, we also see the same shift in the response to a lower frequency. In the resonator we saw a shift from the designed resonance frequency of 1 GHz to a measured resonance frequency of 803 MHz, and here we see a similar shift from the designed cutoff of 1 GHz to a measured cutoff of 835 MHz.

In the base-temperature response, we see that the filter has little ripple in the pass band. After cutoff, the response begins to follow the ideal 9-pole Butterworth response to around −10 dB. At this point the response levels out and falls off at a higher frequency. To gain an insight into the shape of this response we compare it to a mismatched 5-pole Butterworth filter with a cutoff of 1.7 GHz. This is only in

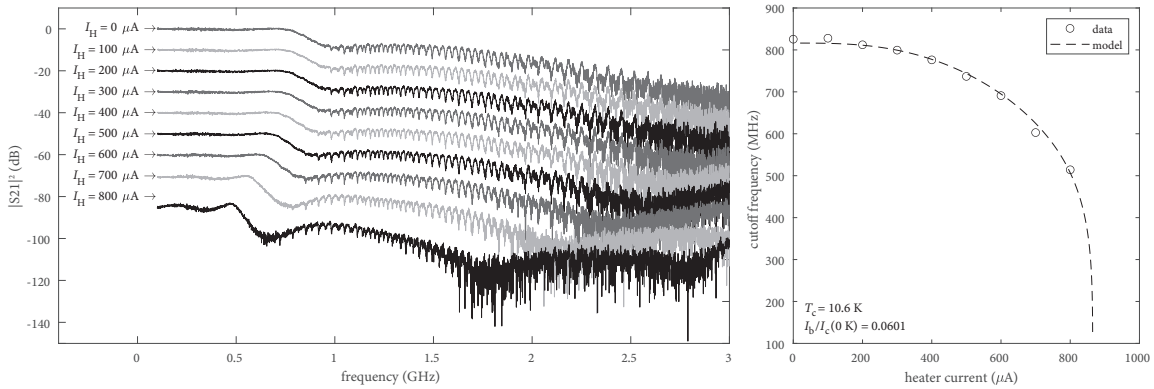

Figure 3-17: Response of the 9-pole low-pass Butterworth filter with changing heater current. (left) The filter response for various heater currents up to 800 μA after which the device switched. These responses are plotted with 10 dB of offset in each successive trace. The arrows on the current designator point to that curve's 0 dB baseline. (right) The cutoff frequency for the traces in the left plot as a function of heater bias current. A fit based on the variable inductance is also shown. The cutoff progressively decreases over 300 MHz for a heater bias current change of 800 μA. The response can be seen to be fairly flat for all heater biases until 800 μA at which point the insertion loss begins to increase.

order to gain an appreciation for the shape and rate that the filter rolls off, and not to suggest that this is in fact the correct model for the response of the filter. Using this fit we can see that the filter begins to roll off at 110 dB/dec after 1.7 GHz. The origin of this secondary response is suspected to be a result of undesired coupling between the capacitors, or coupling to the heater. Either issue is easily addressed in future designs by increasing the space between the capacitors, and increasing the heater spacer thickness or shifting the heater off to one side. Alternatively, both issues could be addressed by using a stripline approach where the device is implemented in a stripline fashion with a ground plane above and below the device.

The result of the heater tuning is shown in figure 3-17. The filter characteristic remains roughly constant for small tubing currents. As the tuning current increased the cutoff frequency begins to decrease the pass band ripple increases and the response rolls off faster. At a heater bias of 800 μA the filter loss in the pass band is high and there is likely at least one hot-spot present in the filter. Between the zero bias response and the highest bias, we see a shift in the cutoff frequency of over 300 MHz. Thus, we find that the filter response is widely tunable. The extent to which this tunability

is possible, suggests that one method of addressing the uncertainty in predicting the film's kinetic inductance, is to design the filter for a higher frequency than needed, and tune it down as necessary.

3.6 Tunable high-pass filter

With the demonstration of a low-pass filter, the next step would be a high-pass filter. The implementation of a Cauer topology high-pass filter is slightly more complex than a low-pass filter as we have a need for series capacitors and inductors to ground. This section will discuss the design of a high-pass filter, the physical implementation and experimental results.

3.6.1 Design

We again employ the Cauer topology for this filter and use a 9-pole Butterworth design. We design this high-pass filter with a cutoff frequency of 1 GHz. The schematic for this filter is shown in part (a) of figure 3-18, and the component values are provided in table 3.2.

The device being a Cauer topology high-pass filter, it requires inductors to ground. Given that we do not have vias available in this process, we implement the ground as another bond pad. The pad is made large for three reasons. First, the parasitic inductance will be lower for larger pad. The low inductance combined with the large capacitance to ground will provide a low-impedance to ground. Finally, we must bond to this pad, and so it must be large enough for this process. Like in the resonator, the inductors here are implemented as meandered nanowires of width 300 nm over a cutout in the ground plane.

A high-pass filter in the Cauer topology requires series capacitors. This might seem difficult to implement, especially for the relatively large values required. There are a number of options available, such as interdigitated capacitors; however, implementing such capacitors in a high kinetic inductance film leads to high parasitic inductances if a very large layout is not used. Here, we already have a patterned ground plane,

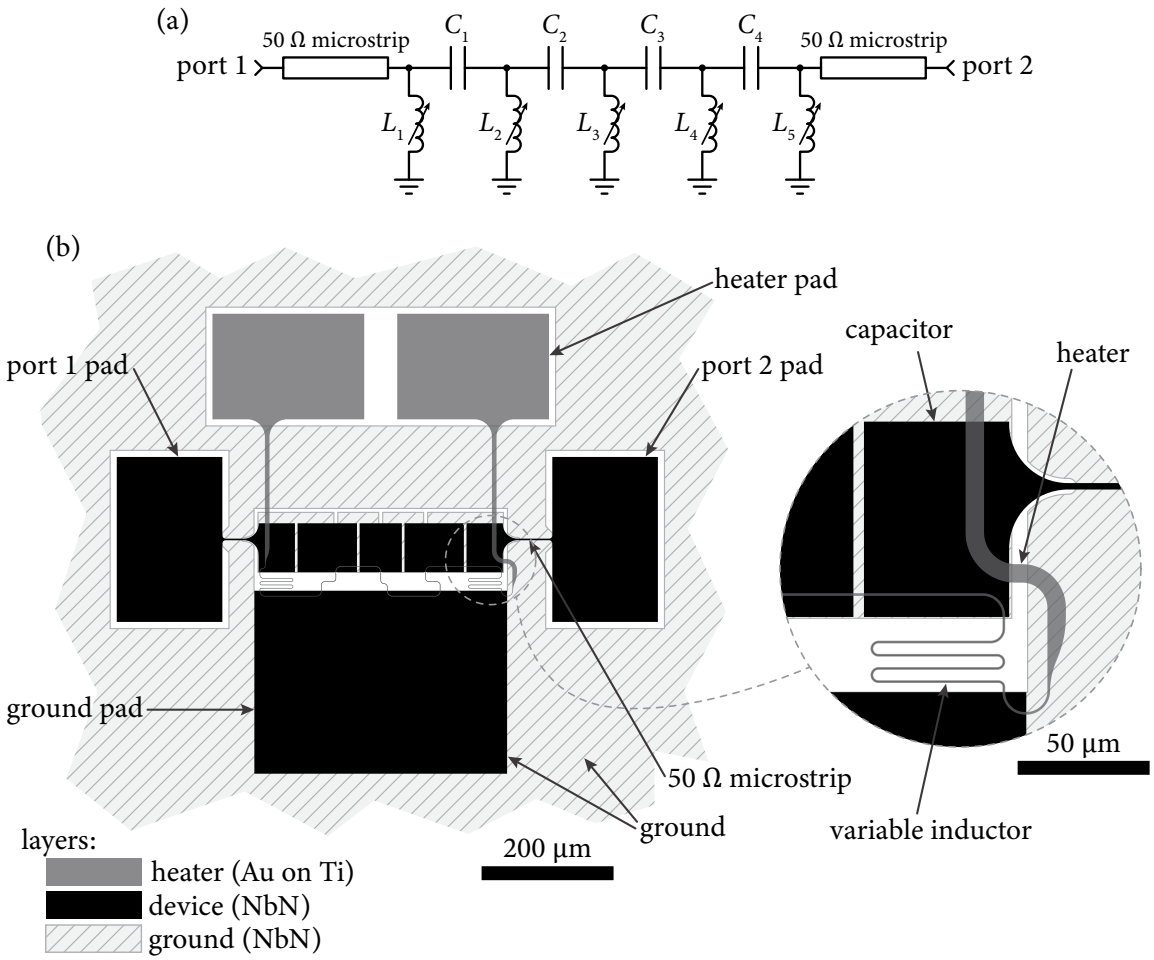

Figure 3-18: Schematic and layout of a 9-pole high-pass filter. (a) Schematic of the high-pass filter implemented here. The values of the components are provided in table 3.2. The heater is not shown in the schematic. (b) The layout of a 1 GHz cutoff frequency 9-pole Butterworth high-pass filter.

Table 3.2: Butterworth 1 GHz 9-pole high-pass filter component values

Designator	Value
C_1, C_4	3.183 pF
C_2, C_3	1.694 pF
L_1, L_5	22.91 nH
L_2, L_4	5.194 nH
L_3	3.979 nH

thus it follows that the most efficient means of implementing a series capacitor is by using isolated sections of the ground plane. In order to use the ground plane film as the capacitor, we break our capacitors into two series capacitors each of twice the value of the original. We then implement these capacitors as stubs to the ground plane. Finally, we cut the ground plane around the outline of the capacitor to form an isolated central terminal. This can be seen in figure 3.2.

3.6.2 Results

Again using the CryoProbe and the calibration procedure outlined in section 3.3, we find the results shown in figure 3-19. These results were obtained with the probe temperature varied as there was an issue with the fabricated device that prevented the use of the heater on this particular device. It can be seen that in the cut-off frequency range of the filter, the noise floor is only around $-20\,\mathrm{dB}$ to $-30\,\mathrm{dB}$; this is due to the VNA being operated at low stimulus powers. We operate at low power to avoid the stimulus of the VNA from interfering with the response of the filter. Furthermore, operating at low power allows us to operate at higher temperature without switching.

The filter response in figure 3-19 shows the filter is behaving close to its designed values but not exactly as designed. At base temperature the cut-off frequency is 900 MHz where it was intended to be 1 GHz. With an increase in the device temperature, we find that the cut-off frequency decreases to near its designed value. As we further increase temperature, we see a further decrease in the cut-off frequency of the filter. Over a temperature range of 3.4 K the cut-off frequency changes by 250 MHz or 25 % of the designed value.

It can be seen that for the curve at base temperature, the filter roll-off is shallower at around 140 dB/dec as opposed to the expected 180 dB/dec for a 9-pole Butterworth filter. This is due to the LC values being different to the designed value, and so the filter response shape is not exactly as intended. As we increase temperature, we find that he roll-off increases to the intended 180 dB/dec. This follows from the fact that we are moving the device values to their intended ratios, and as such, achieving the desired Butterworth characteristic.

3.7 Tunable band-pass filter

Following from the first two filters, we now preset a tunable band-pass filter. This design incorporates many of the aspects of the previous designs. In this section we will first present the filter design, then the layout, and finally the experimental results.

3.7.1 Design

In comparison to the Butterworth response used in the previous filters presented, we selected an elliptic response for this filter. The advantage of the elliptic response (and Chebyshev type-II) for our application is that it can be implemented in the shunt-notch topology. This means that the filter is essentially constructed for a series of series resonators.

We selected a center frequency of 1 GHz and bandwidth of 200 MHz. The stop band attenuation was set to 50 dB. For space considerations we reduce the filter

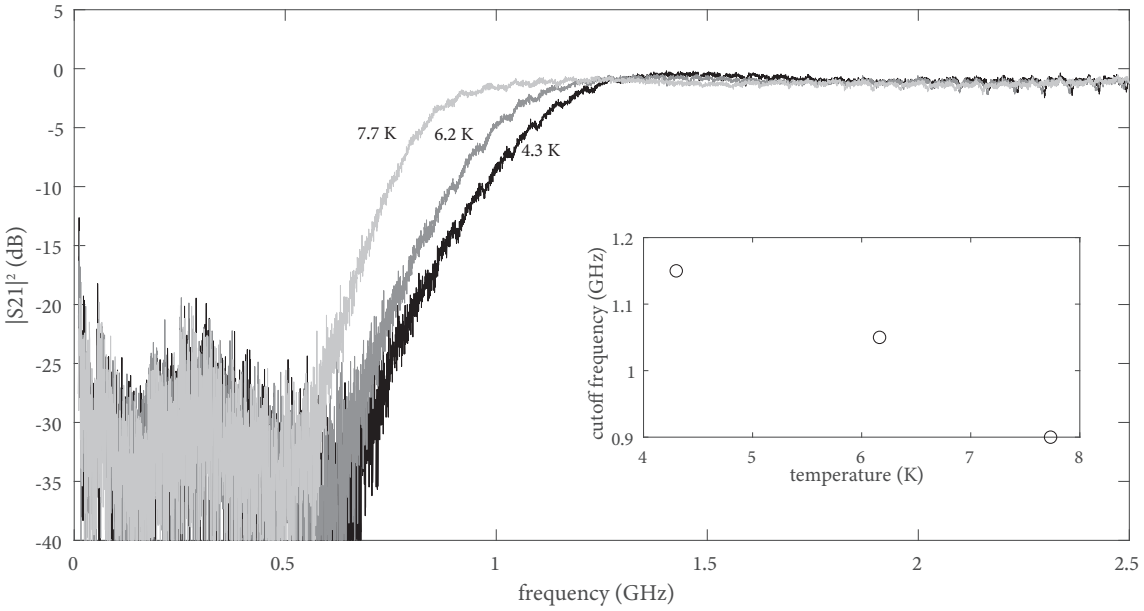

Figure 3-19: Response of the high-pass filter at three temperatures. It can be seen that the response is very nearly that of an ideal Butterworth filter with very little passband ripple. The inset shows the tuning of the cutoff frequency with temperature. Note that the high noise floor is due to the VNA operating at low powers and not the filter response.

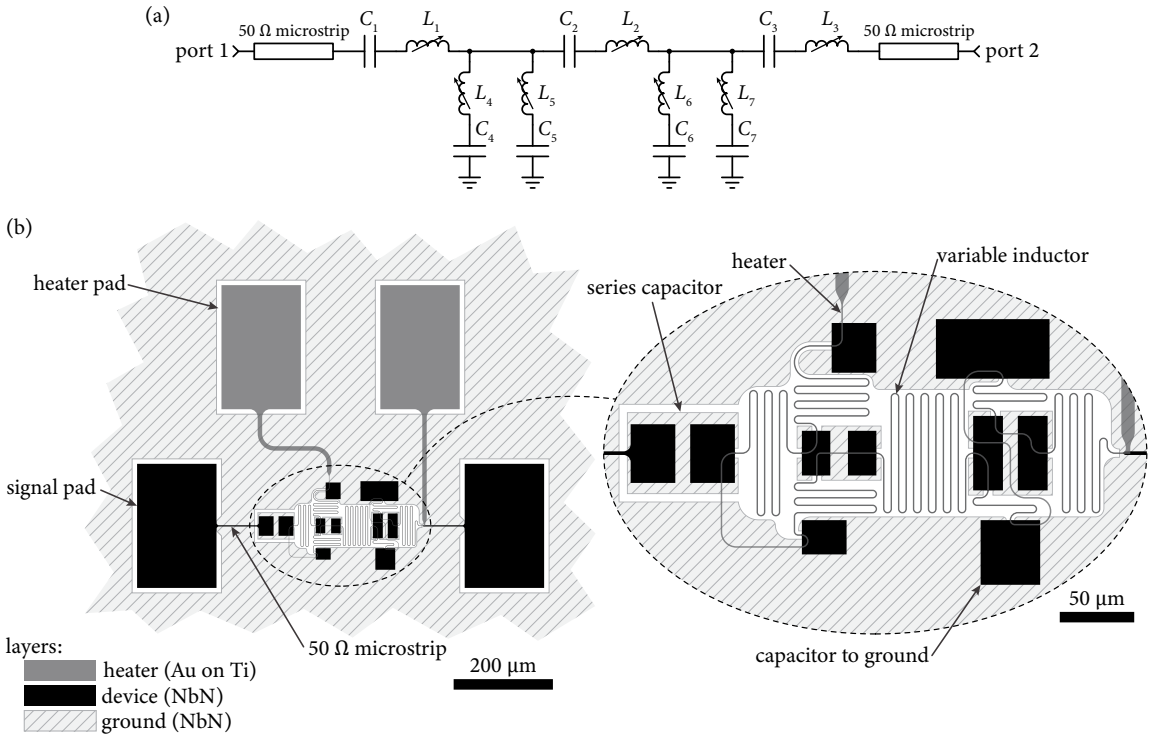

Figure 3-20: Schematic and layout of a 5$^{\text{th}}$-order elliptic 200 MHz band-pass filter centered at 1 GHz. (a) The schematic of the filter, which is essentially formed from seven resonators in series and parallel arrangements. The component values are provided in table 3.3. (b) The layout of the device. The capacitors and inductors are implemented as previously discussed. The heater can be seen to follow the inductors and take a path along all the inductors in series. The heater roughly follows from one port through the filter to the other port. Laying out the heater in this method reduces the deleterious effects of signals coupling to the heater.

Table 3.3: Elliptic 200 MHz band-pass filter centered at 1 GHz component values

Designators	Values	Resonance
C_1, L_1	816.7 fF, 31.02 nH	1 GHz
C_2, L_2	395.3 fF, 64.08 nH	1 GHz
C_3, L_3	690.0 fF, 36.71 nH	1 GHz
C_4, L_4	1.384 pF, 26.49 nH	831 MHz
C_5, L_5	956.1 fF, 18.30 nH	1.203 GHz
C_6, L_6	4.217 pF, 10.57 nH	754 MHz
C_7, L_7	2.397 pF, 6.01 nH	1.326 GHz

order to five. This sets the stop band width at 357 MHz. The schematic of the filter is shown in figure 3-20, and the component values (and their resonant frequencies) are provided in table 3.3.

3.7.2 Results

The response of the filter at base temperature with zero heater bias is shown in figure 3-21. It can be seen that the operating frequency is shifted down, again due to the higher than designed LC product. The center frequency is at 877 MHz, which is similar to the shift we found for the resonator in section 3.4. Additionally, the pass band of the filter is not flat, rather it is lower on the low side and higher on the high side. This is not consistent with an error in either the inductance or capacitance. Errors in these terms would increase the ripple but not produce a skewed response. Additionally, we only see one resonance in the upper stop band where we expect to see two – owing to the two high side shunt resonators, see table 3.3. These two effects are consistent with one of the shunt resonators being defective. We either have that the device is shifted in frequency or is open circuit due to some physical defect. Regardless, we will proceed with the characterization of this device, as even with the defect the filter would be useful for a number of possible applications.

In examining the filter response further, we find that our assessment of stop-band attenuation is limited by the noise floor of the VNA. As we are operating at relatively low stimulus power, we run into the noise floor at only around 30 dB of filter attenuation. However, from the results we can see that we expect the stop band attenuation to be at least around 35 dB, and likely better.

The results in figure 3-22 show the filter tuning with heater bias. The response shape can be seen to be nearly unchanged with tuning with only the upper and lower cutoff frequency changing. Similar to the low-pass filter, we find that at a heater bias of 800 μA, the response changes and the pass band attenuation increases, indicating possible switching. Over the full tuning range we see the upper and lower cutoff frequencies fitting well to our model. As expected, the upper and lower cutoff frequencies both decrease in the same proportion, and as such the filter bandwidth

narrows.

By using separate heaters for the different resonators in the filter, an arbitrarily programmed filter could be achieved. For example, the upper and lower cutoffs could be tuned independently, or disabled by switching the inductors. There are many opportunities to create filter designs that would have not previously been possible in other technologies.

3.8 Microwave switch

By using the hTron as the switching element, we can create a microwave switch. As demonstrated in [84], a superconducting wire (coplanar waveguide in their case) can be used as a microwave switch by transitioning a constriction in the wire to the normal state. The downside of the design presented in [84] is that it uses a DC bias on the RF line to switch a constriction. This necessitates the use of two bias tees to combine the RF and DC signal at the switch input and output. This complexity, along with

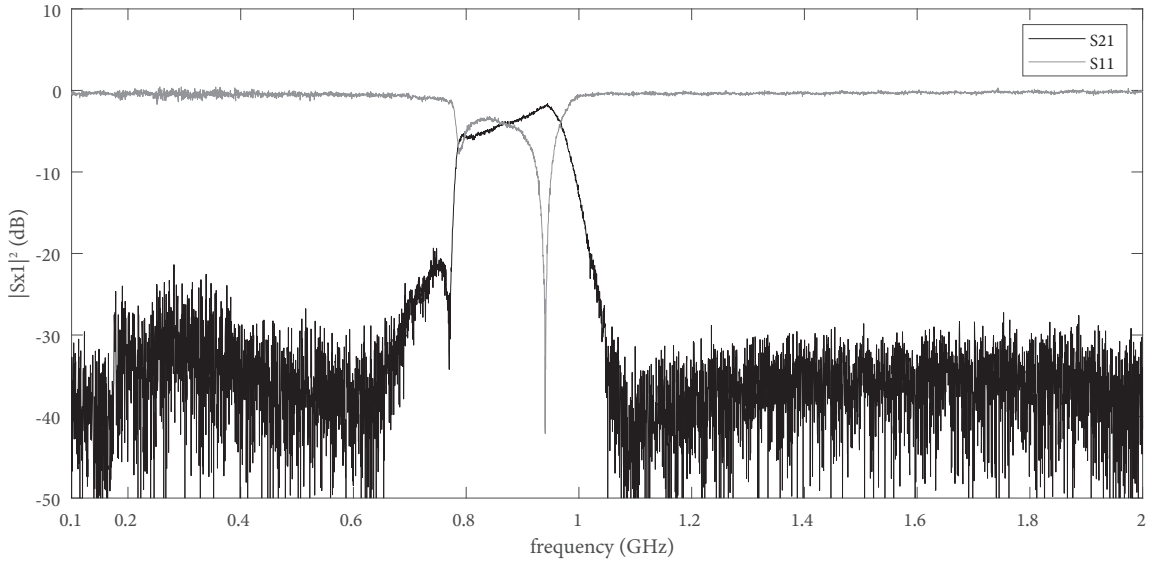

Figure 3-21: Response of the 5-pole elliptic band-pass filter at the base-temperature with no heater bias. The passband is centered at a frequency of 877 MHz with a bandwidth of 198 MHz. This is center frequency is lower than the designed frequency. It is apparent from the lack of a flat band-pass and a high side stop band resonance, that there is a defect in one of the shunt resonators.

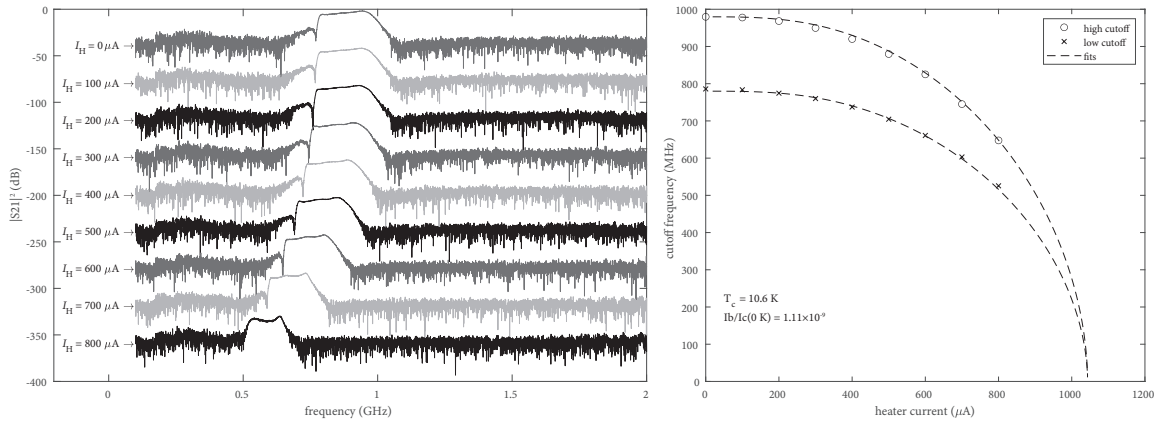

Figure 3-22: Response of the 5-pole elliptic band-pass filter with changing heater current. (left) The filter response for various heater currents up to 800 μA. These responses are plotted with 40 dB of offset in each successive trace. (right) The upper and lower cutoff frequency for the traces in the left plot as a function of heater bias current. A fit based on the variable inductance is also shown. The cutoff frequencies can decrease proportionally, and hence the center frequency and filter bandwidth both decrease with increasing heater bias. The filter begins at a center frequency of 877 MHz and a bandwidth of 198 MHz at zero heater bias and reduces to a center frequency of 674 MHz and a bandwidth of 142 MHz at a heater bias of 800 μA. The pass-band ripple can be seen to be fairly constant for all heater biases until 800 μA at which point the ripple increases.

the discontinuity formed by the constriction leads to losses when the switch is closed. In addition, the device cannot operate at DC, and if a low operating frequency is required, the bias tee must be physically large (see section 3.10). The design we present here uses only a microstrip and a heater with no need for a bias tee or a DC bias to switch the wire. This approach has the advantage that, when the switch is closed, it appears nearly transparent. In addition, there is no low frequency cutoff, which means that the switch can operate to DC.

3.8.1 Design

The design of the hTron switch is very simple. We take a section of microstrip transmission line and place a heater running over it. The transmission line, when in the superconducting sate, will simply carry the signal. However, when the heater is activated, the microstrip will transition to the normal state and present a very high resistance, here around 20 kΩ. The layout of the device is shown in figure 3-23.

Here, we again design for a system impedance of $Z_0 = 50\,\Omega$. As with all the other heaters presented, we increase the width of the heater for connection traces. The heater was designed to be slightly wider than the microstrip to allow for misalignment between the layers. We used a heater width of $4\,\mu$m, which gives us a tolerance of around ± 800 nm. If the heater were misaligned, it would not result in much of a change in performance, but it would mean more power needs to be dissipated for the same suppression of the microstrip. The length of the heated section was arbitrarily chosen to be $320\,\mu$m. It is likely a shorter section would perform just as well, and require less heat to switch. Alternatively, a heater running across the nanowire, such as that used in the nMem, may be sufficient, and would offer consume much less power. Additionally, for greater isolation, a shunt switch could be used, as was demonstrated in [84]; however, we found it unnecessary for our purposes.

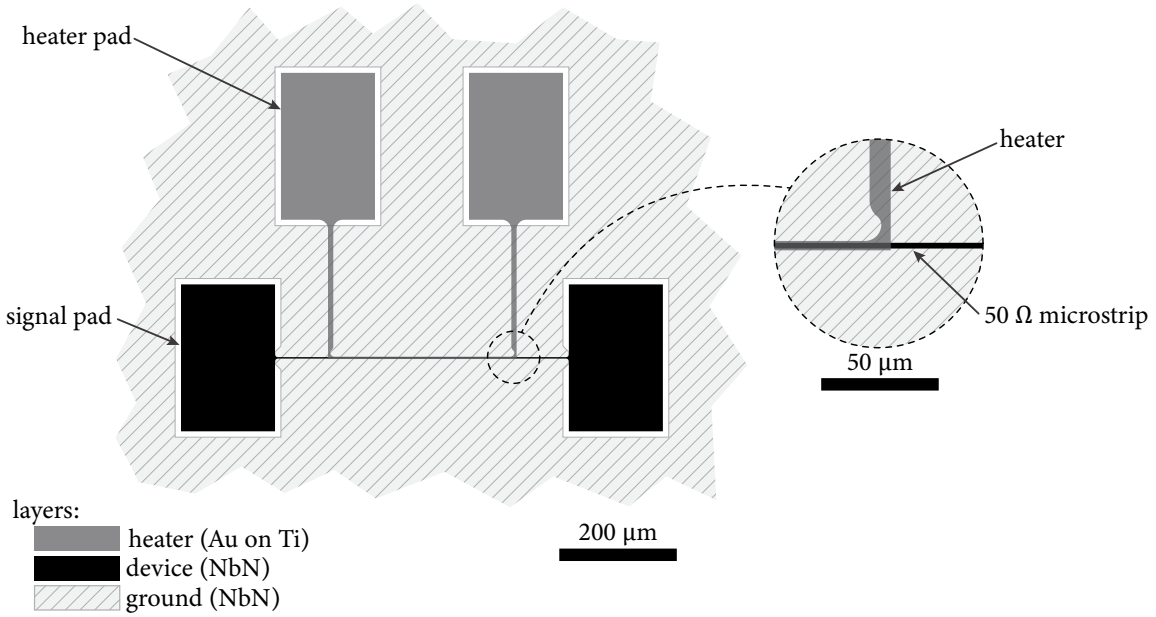

Figure 3-23: Layout of a hTron microwave switch. The device consists of a 50 Ω microstrip with a heater located on top of it. The heater runs along a length of the microstrip and and heats a relatively long section to increase isolation.

3.8.2 Results

The experimental results are shown in figure 3-24. These results were calibrated using the same procedure previously discussed. On the right of figure 3-24, the current-voltage relation of the device between the two signal pads is plotted. It can be seen that the wire has a switching current over 200 μA when the heater is disabled (switch closed). When we enable the heater (open the switch), the device becomes resistive with a resistance of over 20 kΩ. This high resistance, combined with the transmission line no longer having the inductance necessary to support the original impedance, leads to a large mismatch and hence good isolation.

In the left of figure 3-24 we have the insertion loss and isolation of the device in the open and closed states. In the closed state the loss is very minimal with the majority of the deviation from zero-loss being due to our limited calibration scheme – see section 3.3. In the switched state the device provides very high isolation for frequencies less than 1 GHz at more than 40 dB (where we are running into the noise-floor of the VNA). As we increase in frequency, we find that the isolation decreases.

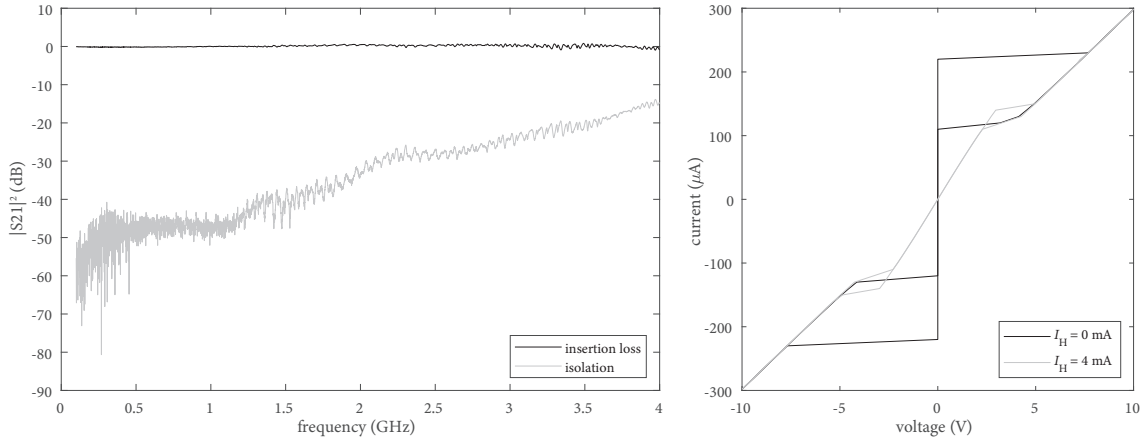

Figure 3-24: Microwave switch response. (left) The insertion loss and isolation when the switch is closed and open, respectively. The insertion loss is nearly zero for all frequencies tested. The isolation is very high for low frequencies and decreases at higher frequencies. (right) The current-voltage relation of the device in the open and closed states. The switch is considered open when the heater current I_{H} is high, and closed when I_{H} is low. Note that the switching visible in the open curve is due to only a portion of the line being heated.

This is likely due to some coupling between the ports of the device, through the heater, and with the largest contribution being due to coupling between the wire-bonds due to their proximity to each other. Regardless, the device provides more than 20 dB of isolation to over 3 GHz.

3.9 Hybrid coupler, variable delay line, and on-chip interferometry

A hybrid coupler is another relatively simple device that can be implemented in microstrips readily. The coupler is an interesting device as it allows for, amongst other applications, the realization of an on-chip interferometer. Such an interferometer allows for the phase-sensitive measurements to be performed on-chip with minimal parasitic components. This allows for measurements to be performed with simpler device models and hence fewer fitting parameters. This can be useful for measuring effects such as the change in kinetic inductance. Additionally, such devices can be used for other purposes such as in realizing quadrature modulators, and other such

devices.

As an example, we will introduce another new device, the hTron variable delay line. This device is simply a nanowire microstrip transmission line (or other transmission line) that, through the use of a heater, produces a variable velocity factor, and hence delay. The localized measurement allows for a very simple model with no need to take into account parasitic elements. Additionally, the continuous nature of the measurement allows for time-domain measurements to be performed. This enables, for example, the assessment of the thermal properties of the device to be characterized.

The following sections outline the design of a 90° hybrid coupler, and the design of a hTron delay line and its incorporation into an interferometer. With the design fabricated, we then present results from slow measurements matching theory to device operation, and high speed measurements to capture the thermal dynamics of the hTron.

During this work the pre-print article [85] was released, which presents a similar measurement architecture (use of hybrid coupler for on-chip interferometry). However, the work presented here differs in that the device presented in [85] is intended for different purposes, their device operates at different bands, and their device uses DC bias instead of hTrons. Additionally, the authors have yet to demonstrate a working device.

3.9.1 Coupler design

The coupler is a simple 90° hybrid (3 dB) implemented in a single branch-line configuration [86]. The branch-line design consists of four $\lambda/4$ transmission lines arranged in a rectangle, as shown in part (a) of figure 3-25. From [86] we have that for such a coupler we need the branch lines (the vertical transmission lines in figure 3-25) to have an impedance of Z_0/y_B, and the series lines (horizontal lines in figure 3-25), have an impedance of Z_0/y_A. For matching and isolation, we need the normalized admittances to be related by $y_\mathrm{A}^2 = 1 + y_\mathrm{B}^2$. Now, given that we wish to have equal power out of the two ports, we require a coupling of 3.01 dB, which of course gives

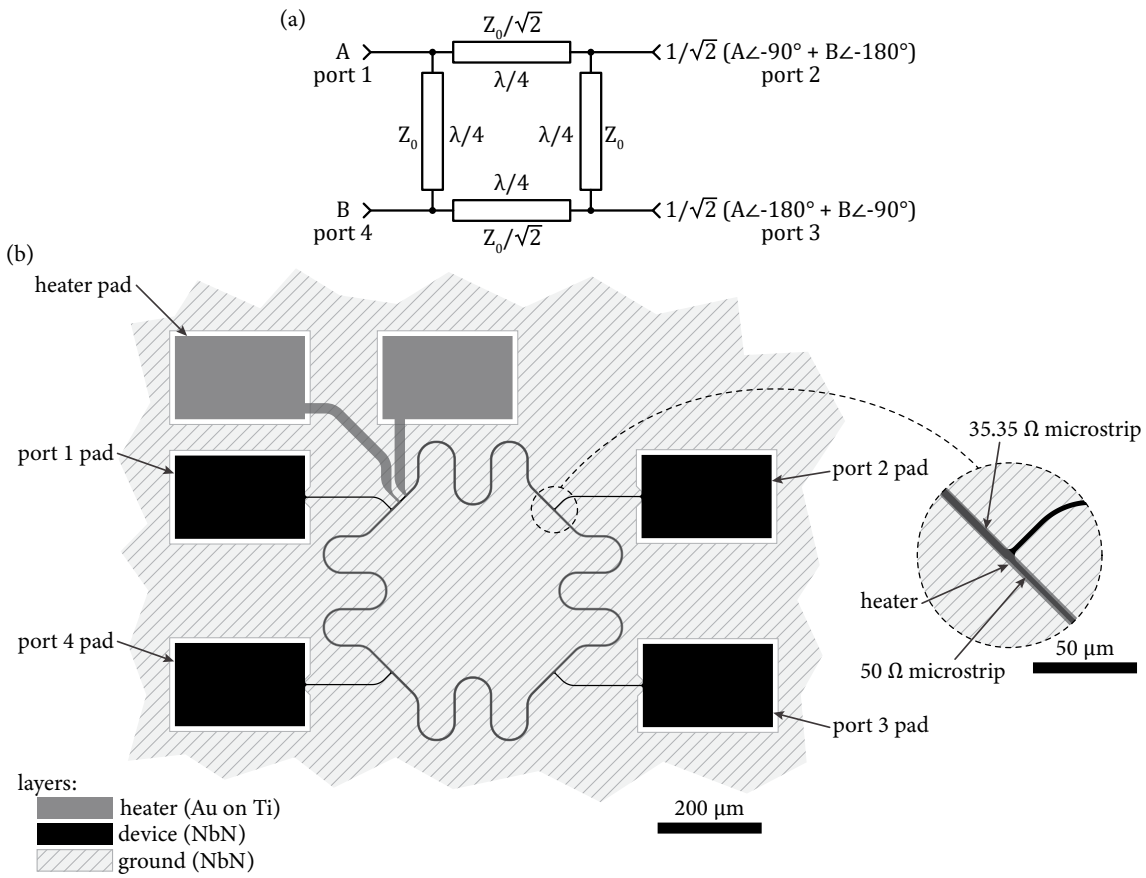

Figure 3-25: Schematic and layout of a 3 dB 90° hybrid coupler. The coupler is designed for an operating frequency of 2 GHz and a system impedance of $Z_0 = 50\,\Omega$. The transitions between the line widths and the joins are filleted to reduce current crowding.

us a coupling coefficient of $c = 10^{-3.01/20} = 1/2$. In order to achieve this, we require $y_\mathrm{B} = c/\sqrt{1-c^2} = 1$, and so $y_\mathrm{A} = \sqrt{2}$. Thus, giving us the line impedances shown in figure 3-25.

We designed for a center-frequency of $f_\mathrm{c} = 2\,\mathrm{GHz}$. In consulting figure 3-7 we find that for our $50\,\Omega$ lines we need a width of $2.42\,\mu\mathrm{m}$ at which we find our velocity factor is $v_\mathrm{p}/c = 2.02\,\%$. Thus, for quarter-wavelength $\lambda/4$ lines we need a length of $\lambda/4 = v_\mathrm{p}/4f_\mathrm{c} = 757.5\,\mu\mathrm{m}$. Similarly, for the remaining two lines, we need an impedance of $Z = 50/\sqrt{2}$, for which we need a width of $3.42\,\mu\mathrm{m}$ and have a velocity factor of $v_\mathrm{p}/c = 2.03\,\%$ ($\lambda/4 = v_\mathrm{p}/4f_\mathrm{c} = 760.5\,\mu\mathrm{m}$).

Using the calculated line lengths and widths we form the layout shown in figure 3-25. In order to reduce the layout size, we meander the lines in each branch. A heater

was run along the lines to allow for tunability of the coupler operating frequency. Transitions between the line two line widths and the join of the port transmission lines are filleted to reduce current crowding.

3.9.2 On-chip interferometer

As a demonstration of the coupler, we form an interferometer for on-chip measurement of phase differences between paths. For this we use two of the above described couplers and two variable-phase delay transmission lines. These devices are arranged as shown in figure 3-27.

From section 3.2.2, the phase velocity $v_{\mathrm{p}}(T)$ of a heated transmission line decreases with increased $L_{\mathrm{K}\square}(T)$ and hence the change in phase-delay $\phi(T)$ of a line will increase accordingly. Thus, a fixed length transmission line, when heated, will appear to change length according to

$$\frac{\lambda(T)}{\lambda(0\,\mathrm{K})} = \frac{\phi(T)}{\phi(0\,\mathrm{K})} = \frac{v_{\mathrm{p}}(0\,\mathrm{K})}{v_{\mathrm{p}}(T)} = \sqrt{\frac{L_{\mathrm{K}}(T)}{L_{\mathrm{K}}(0\,\mathrm{K})}}. \tag{3.31}$$

This relation is shown in figure 3-26.

Additionally, the transmission line impedance changed with temperature. This leads to the heated branch model being given by the product of three terms, the transmission coefficient from the unheated to heated line $(1+\Gamma_{\mathrm{FS}})$, the delay, and then the transmission coefficient from the heated $(1+\Gamma_{\mathrm{SF}})$ to the unheated line, $(1+\Gamma_{\mathrm{FS}}(T))(1+\Gamma_{\mathrm{SF}}(t))\exp(-j\phi(T))$. Thus, we have the model of the heated delay line is

$$H(T) = \frac{4\sqrt{L_{\mathrm{K}}(T)/L_{\mathrm{K}}(0\,\mathrm{K})}}{\left(1+\sqrt{L_{\mathrm{K}}(T)/L_{\mathrm{K}}(0\,\mathrm{K})}\right)^2} e^{-j\phi(0\,\mathrm{K})\sqrt{L_{\mathrm{K}}(T)/L_{\mathrm{K}}(0\,\mathrm{K})}}. \tag{3.32}$$

This function can then be used with the temperature dependent kinetic inductance expression to find the value of $H(I_{\mathrm{H}})$.

By using a coupler as a phase splitter by terminating the isolation port, we obtain a quadrature signal from a single input signal. We can then send these signals through the delay lines and apply it to a second coupler to obtain the interference between

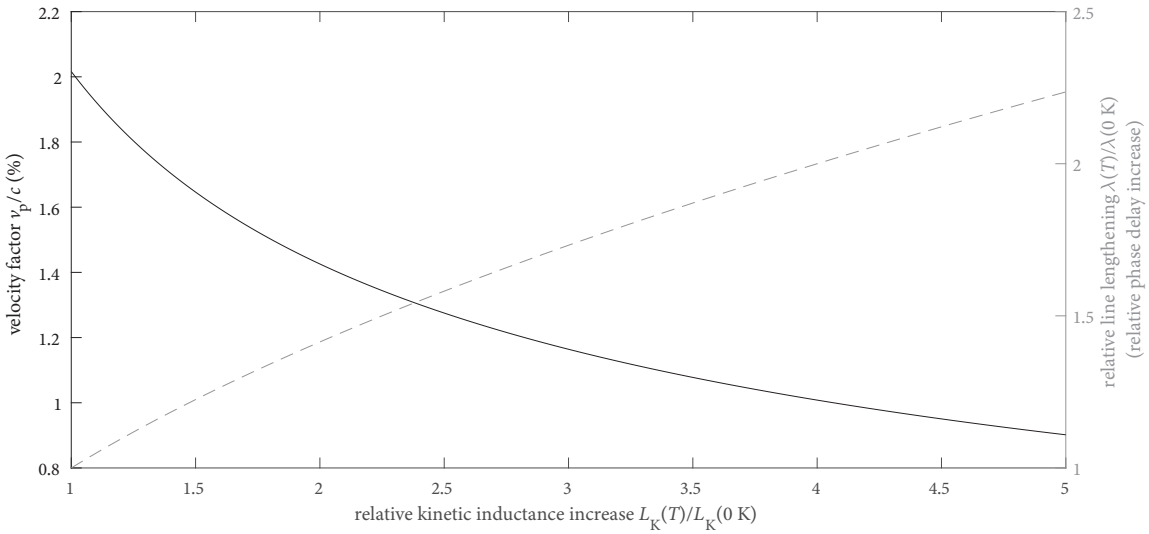

Figure 3-26: Velocity factor (left) and relative line lengthening (right) in a microstrip as the relative kinetic inductance increases. The left axis shows the velocity factor for a $2.42\,\mu$m wide ($50\,\Omega$) microstrip. It is evident that the velocity factor continues to decrease as the kinetic inductance increases. The right axis shows the relative line lengthening, or alternatively phase delay increase as the kinetic inductance of the line increases. The delay of the line can be seen to increase, doubling at a kinetic inductance increase of around four times its zero-temperature value.

the signals. We somewhat arbitrarily set the length to one branch at $\lambda/4$. In order to increase the sensitivity of our interferometer to the change in kinetic inductance, we select the length of the opposing branch to be longer than the reference branch. In doing so, we increase the $\phi(0\,\mathrm{K})$ coefficient in equation 3.32. Additionally, to resolve small phase changes at low bias we wish to operate close to but below a null on one port. This operation can be obtained by having the delays differ by $\pi/8$. To satisfy these requirements and not make the line too long to be practical, we select the long branch to be $4.5\times$ longer than the reference branch. We refer to the shorter reference branch as the "short path" and the longer measured path as the "long path".

With the selected configuration, as outlined schematically in figure 3-27, the two outputs of the final coupler are thus

$$\begin{aligned} O_1 &= \frac{1}{2}A\left(H_s(I_\mathrm{H})e^{-j\pi} + H_l(I_\mathrm{H})e^{-j2\pi}\right) = \frac{1}{2}A\left(-H_s(I_\mathrm{H}) + H_l(I_\mathrm{H})\right) \\ O_2 &= \frac{1}{2}A\left(H_s(I_\mathrm{H})e^{-j3\pi/2} + H_l(I_\mathrm{H})e^{-j3\pi/2}\right) = \frac{1}{2}Aj\left(H_s(I_\mathrm{H}) + H_l(I_\mathrm{H})\right), \end{aligned} \tag{3.33}$$

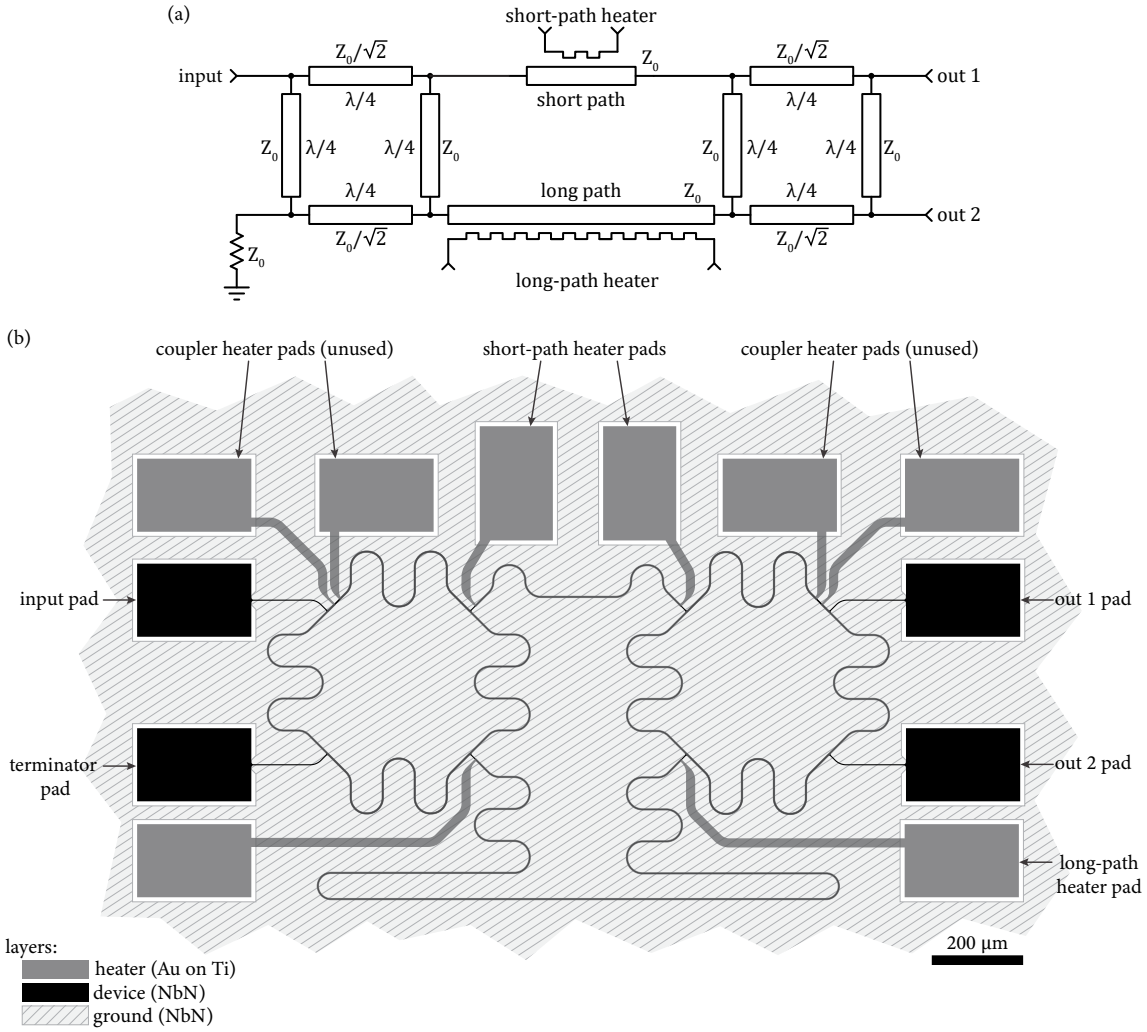

Figure 3-27: Schematic and layout of the on-chip 2 GHz interferometer for measuring microstrip delay lines. (a) The schematic of the interferometer. Note that the heaters on the couplers are not shown as they were not used. The short path is $\lambda/4$ long and the long path is 1.125λ long with no heater bias. The termination resistor was implemented at room temperature. (b) The layout of the device. The couplers and delay lines are implemented as microstrips that are meandered to make the design more compact. In the experiments, the couplers were not tuned, and so the coupler heaters were left unbonded.

where A is our input signal. Now, if we do not heat the reference branch then we have $H_s(I_\mathrm{H}) = \exp(-j\pi/2) = -j$. Which gives us the output signals

$$\begin{aligned} O_1 &= \frac{1}{2}A\left(j + H_l(I_\mathrm{H})\right) = \frac{1}{2}A\left(j + |H_l(I_\mathrm{H})|\, e^{-2.25\pi j\sqrt{L_\mathrm{K}(T)/L_\mathrm{K}(0\,\mathrm{K})}}\right) \\ O_2 &= \frac{1}{2}Aj\left(-j + H_l(I_\mathrm{H})\right) = \frac{1}{2}Aj\left(-j + |H_l(I_\mathrm{H})|\, e^{-2.25\pi j\sqrt{L_\mathrm{K}(T)/L_\mathrm{K}(0\,\mathrm{K})}}\right). \end{aligned} \tag{3.34}$$

This leads to the design shown in figure 3-27.

As with the coupler, the branches in the interferometer are meandered to make the design more compact. Both the short and long branches were heated to allow both sides to be tuned. Additionally, the couplers have heaters to allow for tuning; however, in the experiments presented here we did not tune the coupler.

3.9.3 Heater sweep results

The first experiment conducted with the interferometer simply involved sweeping the heater bias to the long branch. The experiment was setup using a VNA with port one connected to the input port on the interferometer. Port two of the VNA was connected to either output port with the opposing port terminated. The VNA was configured for a sweep around the expected operating point; however, like we found with the resonators, the values were not as expected due to the product LC being higher than expected (leading to all phase values being other than designed). It was found that we had destructive interference on output one at a frequency of 1.961 GHz, which while not our designed operating frequency, will suffice for our purposes.

With an operating point located, the VNA was configured for zero-span operation. This mode allows us to use the VNA as a vector receiver and signal generator. With the VNA in this mode, a programmable current source was used to sweep the heater bias on the long-branch heaters. The results of this sweep are shown in figure 3-28. The model can be seen to agree well with the data up to near switching. Around the switching current our modeling kinetic inductance change with bias current breaks down and increases faster than observed experimentally. This is in part due to our simplified modeling of the kinetic inductance dependence on bias current – see section

3.1.3.

In order to compensate for the shifted LC values for the device and for operating at a frequency other than the designed frequency, we introduced a new term to equation 3.34. All phase delays were scaled by a term $\lambda_{\text{actual}}/\lambda_{\text{design}} = 1.0339$, which was found from fitting to match the model to the measurement. With this term in place we see good agreement with the model and the experimental results.

There can be seen to be a shift in the phase data for one output. This is likely due to the experimental setup, in particular the cable lengths being different, rather than an effect associated with the device. For our purposes, the phase information is not of particular importance as one of the main reasons for using the interferometer is to convert phase to amplitude. We present the captured phase results here simple to demonstrate the agreement with our model.

Given that the heater is of different dimensions to that used in the resonator and filters, we need a new function to map heater current to temperature. We used the same equation from the resonator section, with a new coefficient. The equation is given in the inset of the first panel in figure 3-28.

The primary deviations from the model are due to higher order reflections in the delay line. As the mismatch increases with higher temperature, the contribution to the response of these reflections will also increase. It is possible to take these effects into account by modifying the model in equation 3.34, to include these terms. However, as our model matches well for low temperatures, we will proceed with the simplified model.

With the model fitting the experimental data well, up to near switching, we can now map between interferometer output and device temperature. This is achieved by using the results presented in the past plot in figure 3-28, which shows the device temperature required to achieve the kinetic inductance (right axis of the same plot) that gives the necessary propagation velocity for the phase response required to produce the output we see. From this result we find that there exists a monotonic relation between output one of the interferometer and the delay line temperature up to around 8.5 K. Thus, we can use this mapping to find the device temperature

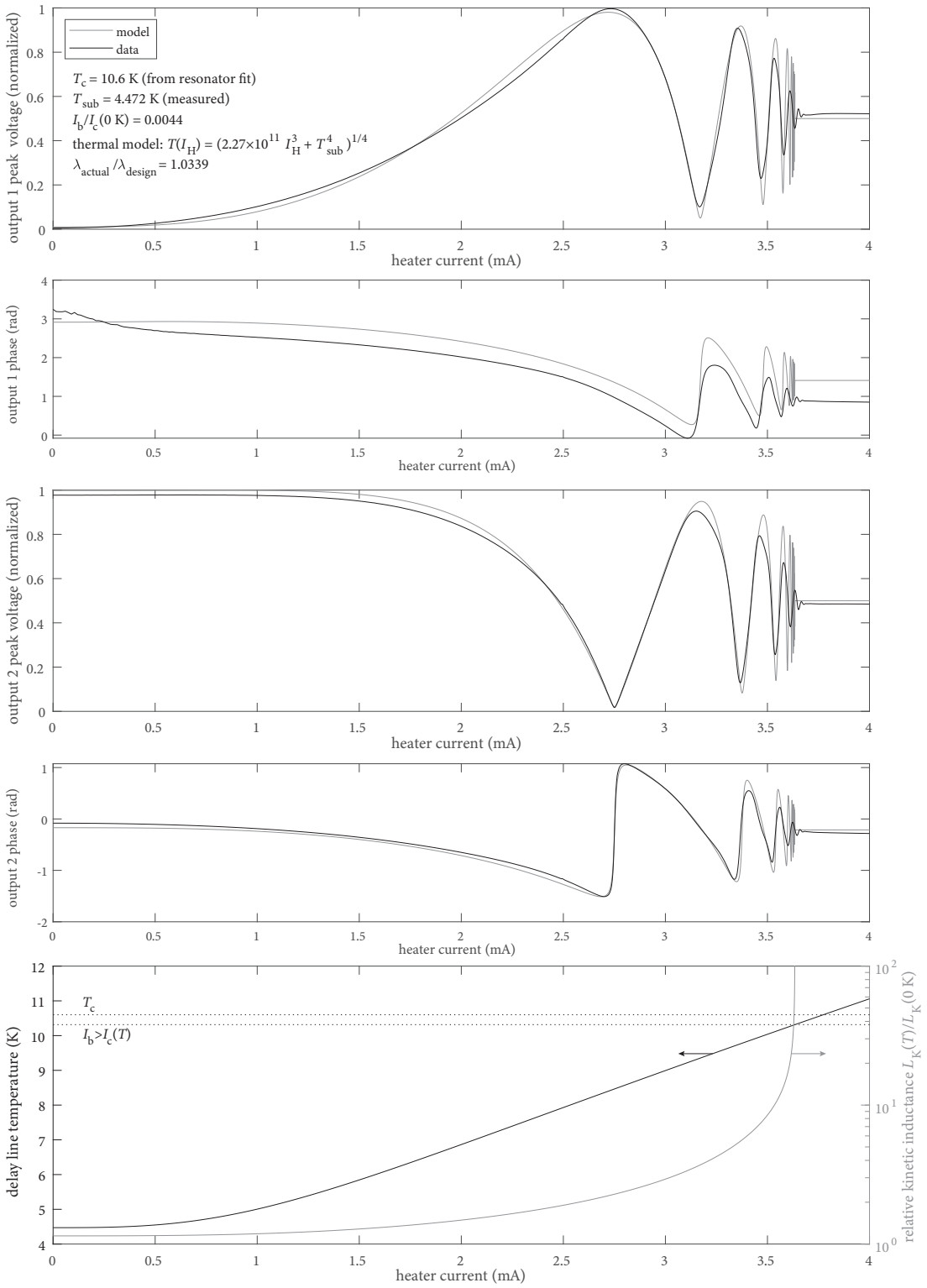

Figure 3-28: Experimental and model results for the on-chip nanowire interferometer. The top four plots show the detected voltage and phase from the two output ports. The model parameters are shown as an inset in the top plot. The final plot shows the modeled temperature and resultant kinetic inductance.

relatively accuracy using our model and a simple mapping procedure.

3.9.4 Pulse response results

With the delay line temperature relation extracted, we are now equipped to convert interferometer outputs to temperature. This relation allows us to perform time-domain measurements of the delay line temperature evolution. To this end, we perform a second experiment to assess the speed at which the heater can affect a change in the nanowire inductance. For this experiment, we setup an oscilloscope monitoring the two outputs and a signal generator providing the input signal. An AWG provides the heater pulse. For the heater, one end was attached to the AWG and the other to the oscilloscope's 50 Ω-terminated input. In doing so, we make all signals from the device to the oscilloscope travel the same length of cable, thus equalizing their propagation delays. Thus, no deskewing of the captured signals is needed.

In the first round of testing we provided a 1.5 V high and 200 ns wide pulse to the heater. The interferometer was fed form a signal generator with the same parameters as used by the VNA in the previous section. The resultant captures are shown the first two plots of figure 3-29. The first plot shows the heater bias pulse, which is only shown for reference. It is apparent that there is a small reflection, as evident by a roughly 15 ns repression in the high-level after the rising edge, and a similar reflection after the falling edge. However, due the amplitude of this reflection is small, and its contribution to heating is correspondingly small. For our analysis we will assume the pulse to be ideal with no reflections.

As the two outputs of the coupler contain the same information, we opt to only operate on the signal captured from output one. The 1.961 GHz output of the coupler was captured by the oscilloscope after triggering on the rising edge of the heater pulse. Given the small amplitude of the signal and that we are close to the sample rate of the oscilloscope, the trace is somewhat noisy. To combat this, first we use a digital band-pass filter with a pass band of 1 GHz to 2.5 GHz. The result of this filtering is shown in the capture in figure 3-29. We take this filtered signal and envelope detect it. We then perform this for one hundred successive captures and average the detected

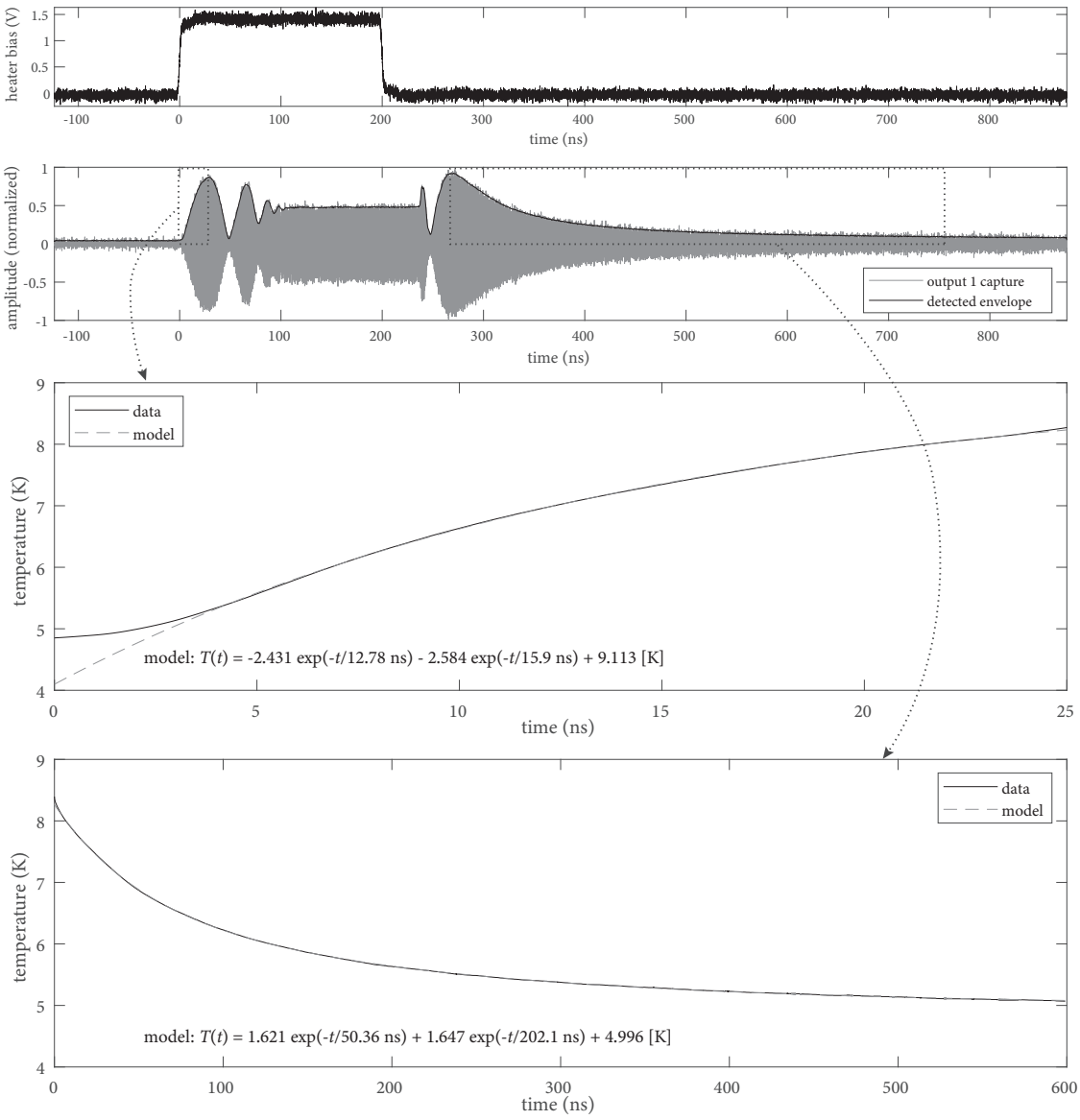

Figure 3-29: Experimental and model results for measuring the thermal response of a nanowire to a 1.5 V heater pulse. The top plot shows the heater pulse applied to the long path heater. The plot below that is the detected output signal on the oscilloscope along with an envelope detected from one hundred oscilloscope captures averaged. The bottom two plots show the output voltage mapped to nanowire temperature for the rising and falling edge of the interferometer response. Basic fits were applied to these signals and are shown in these plots.

envelopes together. The resultant average envelope is shown in the second plot of figure 3-29.

There is quite a lot of dynamics occurring in the response shown in figure 3-29. First, we have the device beginning at the base temperature. Then with the application of the heater the delay line temperature increases, resulting in a phase change detected by the interferometer. This results in the output tracing the response shown in figure 3-28. At some point the lower hTron switches, and we see no more change in the interferometer output. When the heater is deactivated, the device begins to cool. The delay line will not being superconducting until it cools off sufficiently. Given that the heater may have heated the device to well above T_c, this can take some time. When the device does begin superconducting again, it will rapidly retrace the response from 3-28 and then, close to exponentially, approach the substrate temperature.

With clean captures of the envelope, we can extract the rising and falling edge of the output. Next, we map this response onto the model from the previous experiment to extract the delay line temperature. The extracted temperature data is shown for both the rising and falling edges in figure 3-29. It can be seen that for the given input power, a change of over 3 K in the inductor temperature occurs in a period 25 ns. In order to gain some insight into the timescales at play, we attempted to fit an exponential to the data. However, owing to the material characteristics changing with temperature, we find that we must also include a second exponential term. The fits we found are shown superimposed on the extracted temperature data in figure 3-29. The initial response of the fit deviates from the experimental results, this could be due to a number of effect, but is likely primarily a consequence of the heater pulse rising edge being relatively slow at 3.3 ns.

In the rising edge of the data, we find that the response is largely characterized by two time constants, $\tau_1 = 12.78$ ns and $\tau_2 = 15.9$ ns. When we perform this same experiment again, but with twice the heater bias, the response is now largely dominated by a single time constant of $\tau = 24.35$ ns, as shown in figure 3-30. So it would seem that the response becomes slower for higher power dissipation. Of course, for

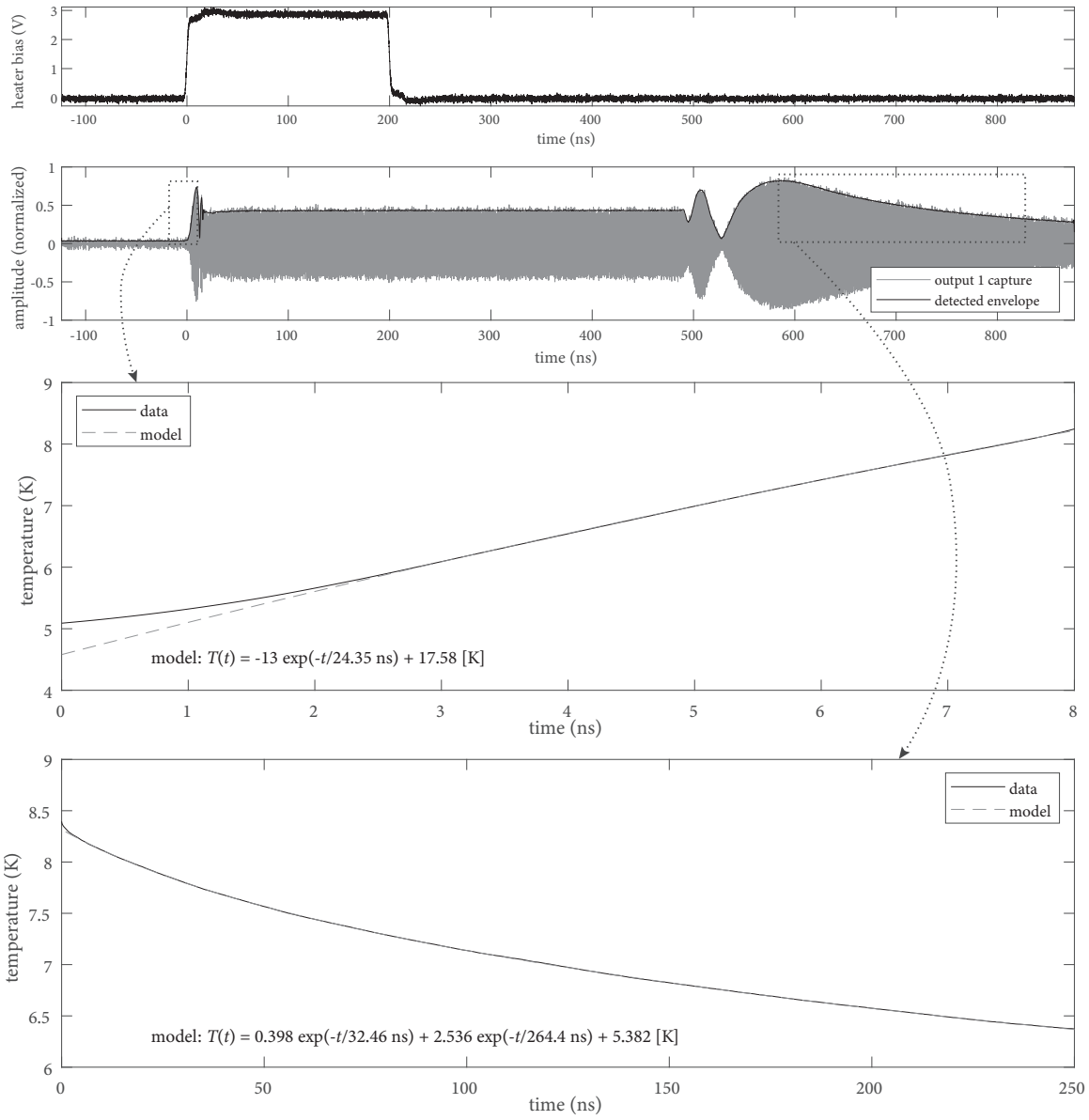

Figure 3-30: Experimental and model results for measuring the thermal response of a nanowire to a 3 V heater pulse. The top plot shows the heater pulse applied to the long path heater. The plot below that is the detected output signal on the oscilloscope along with an envelope detected from one hundred oscilloscope captures averaged. The bottom two plots show the output voltage mapped to nanowire temperature for the rising and falling edge of the interferometer response. Basic fits were applied to these signals and are shown in these plots.

the higher pulse height the device temperature increases faster as the coefficient of the exponential is larger, and so our slope is steeper. With the higher heater power dissipation we see a change of over 3 K in the inductor temperature occurs in a period 8 ns, as opposed to the 25 ns required for the response with half the pulse height.

For the cooling of the device, we see a similar combination of two exponential terms matching the data well. In the device cooling, we find that the timescale is substantially longer. In the smaller pulse experiment shown in figure 3-29, we see the cool down is characterized by the time constants $\tau_1 = 50.36$ ns and $\tau_2 = 202.1$ ns. Here it is interesting that there are two rather different time constants with one roughly four times the other. It seems that the device temperature cools somewhat rapidly and then begins to much more slowly approach the substrate temperature. This may be due to the region of the substrate heated taking longer to cool than the local area or other some other effects. For the experiment with twice the pulse height, shown in figure 3-30, we have the time constants $\tau_1 = 35.46$ ns and $\tau_2 = 264.4$ ns. It is interesting that the response of the initial cooling is faster, but the longer timescale response is slower. We now see a ratio between the short and time timescales of around eight, in comparison to the three for the half height pulse. Clearly, the dynamics of the hTron channel temperature with heater bias variation requires more investigation to fully understand.

One notable different between the response shown in figure 3-29, and that of the same experiment with twice the heater pulse amplitude, shown in figure 3-30, is the delay between the falling edge of the heater and the device returning to the superconducting state. This seems to be due to the device temperature in the higher amplitude experiment reaching a much higher value than in the first experiment. In order to investigate this effect, we performed a new experiment with a fixed height pulse, and varied its duration from 10 ns to 50 ns, as shown in figure 3-31. It can be seen in the results of this experiment that the time it takes for the device to return to the superconducting sate increases roughly linearly. For the shortest 10 ns heater pulse, the device needs around 70 ns to begin superconducting again, and for the longest heater pulse at 50 ns, the device needs around 275 ns to begin superconducting

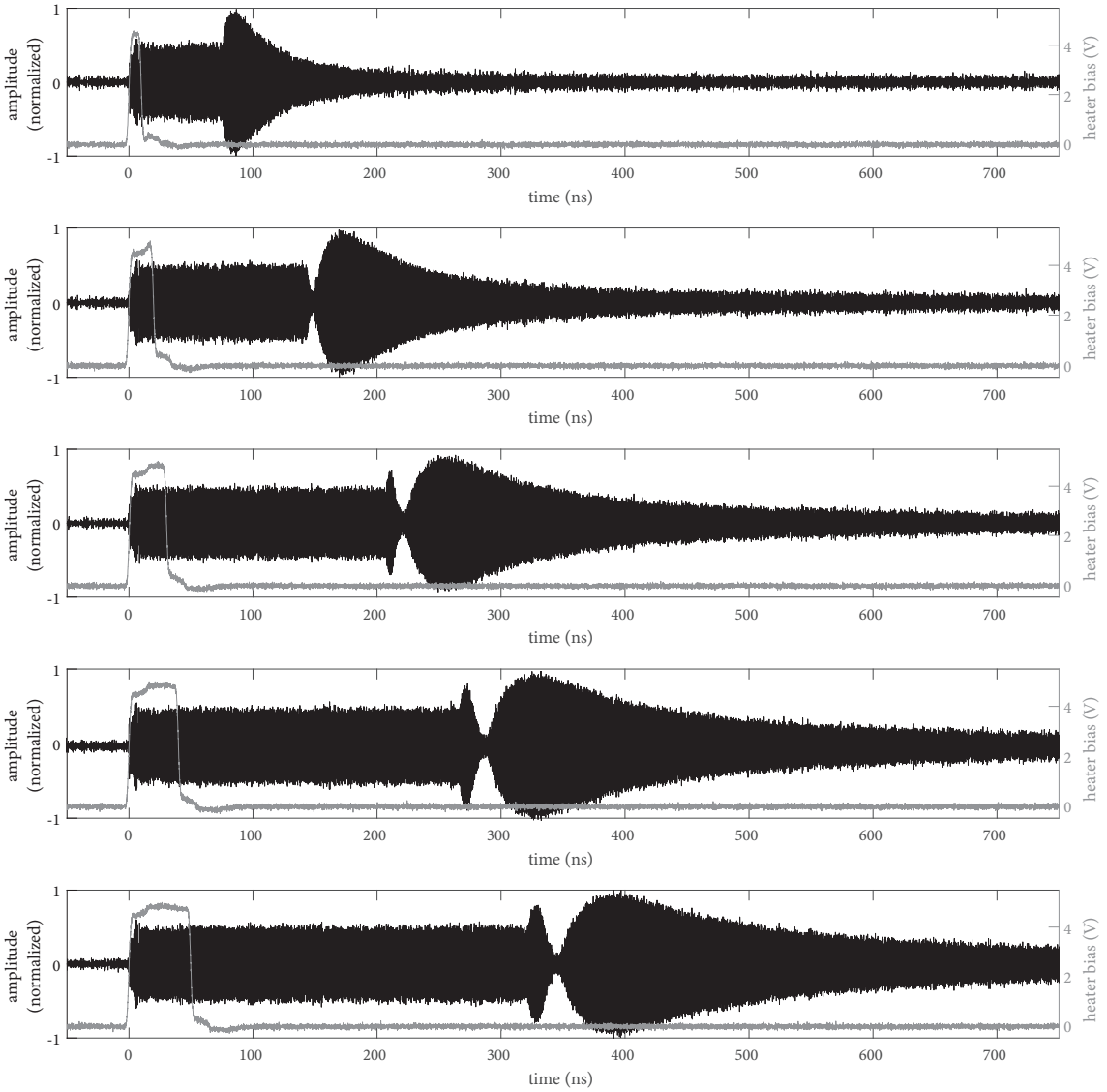

Figure 3-31: Experimental results showing the interferometer output response for 5 V pulses from 10 ns to 50 ns. The period necessary for the wire to return to the superconducting state can be seen to increase with increased pulse duration. For this high heater bias, the transmission from the superconducting to normal regime occurs very quickly and seems to be limited by the rising edge of the AWG output.

again. A further interesting observation is that as the pulse length grows, the time constant associated with the initial transition from the normal-to-superconducting transition becomes slower. This effect is evidenced by the change in shape of the envelope at the beginning of the superconducting regime. This seems to be consistent with the results we found in the previous pulse experiments where the time constants changed with heater bias (see figure 3-29 and figure 3-30).

With the use of this new instrument, we can now perform hTron characterization experiments that were previously difficult to perform without on-chip instruments. Through the results we have presented here, we find that the dynamics of the hTron are complex and need further investigation. However, we can say that for a fast response, one needs to apply a high pulse amplitude for a short duration. In doing so, one can achieve a fast switch of the channel and a fast cool down back to the superconducting state. We expect that we will be able to refine our hTron dynamical models by matching their predictions to experimental results obtained from this and similar on-chip instrumentation.

3.10 Bias Tee

One of the most ubiquitous devices utilized in microwave testing is the bias tee. The bias tee, as its name implies, is a three port device that allows for the injection of a DC bias to an "RF+DC" port where it is combined with an AC signal from the "RF" port, with there being good isolation between the RF port and the DC port. This operation is realized by a capacitor between the RF port and the RF+DC port, and an inductor between the DC port and the RF+DC port, as shown in figure 3-32.

3.10.1 Design

For a bias tee to achieve its purpose, we need the RF to DC+RF port to have a low insertion loss at the operating frequencies and the RF to DC+RF port to have a high insertion loss at these frequencies (high isolation). Assuming ideal inductors and capacitors, we need only worry about the lower frequency of the band f_c. For

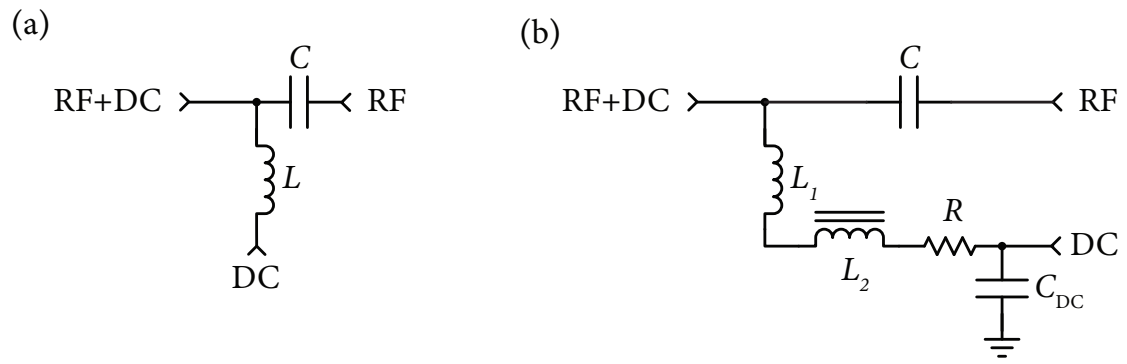

Figure 3-32: Schematics of a bias tee. (a) The schematic of this bias tee. (b) The schematic of a practical bias tee de-Qing as implemented to reduce undesired resonances. Possible methods of de-Quing include adding resistors and lossy cored inductors.

a system impedance of Z_0, we would select the component values of the inductor L and capacitor C such that

$$\begin{aligned} X_{\mathrm{C}} &= \frac{1}{2\pi f_{\mathrm{c}} C} \ll Z_0, \text{and} \\ X_{\mathrm{L}} &= 2\pi f_{\mathrm{c}} L \gg Z_0. \end{aligned} \tag{3.35}$$

A wideband bias tee can be difficult to realize in practice due to nonidealities in the components. In particular, practical inductors will often have self-resonant frequencies that lie in the frequency of interest. To reduce the effects of these resonances, components are often added to the DC coupling branch to "de-Q" the components. There are a number of methods, such as adding series or parallel resistors to the inductors, using multiple inductors, and using cored inductors that are lossy at higher frequencies among others – see figure 3-32.

For the bias tee we developed here, we did not need to introduce any de-Qing to our inductor. The layout was designed to minimize inter-winding capacitance, thereby pushing any possible self-resonances to higher frequencies. We designed the bias tee for an operating frequency of 1 GHz. Thus, for our design we can choose the inductor impedance to be ten time higher than our system impedance of $Z_0 = 50\,\Omega$ and our capacitor impedance to be ten times lower than Z_0 in order to satisfy equation 3.35. With these values we get $L = 80\,\mathrm{nH}$ and $C = 32\,\mathrm{pF}$.

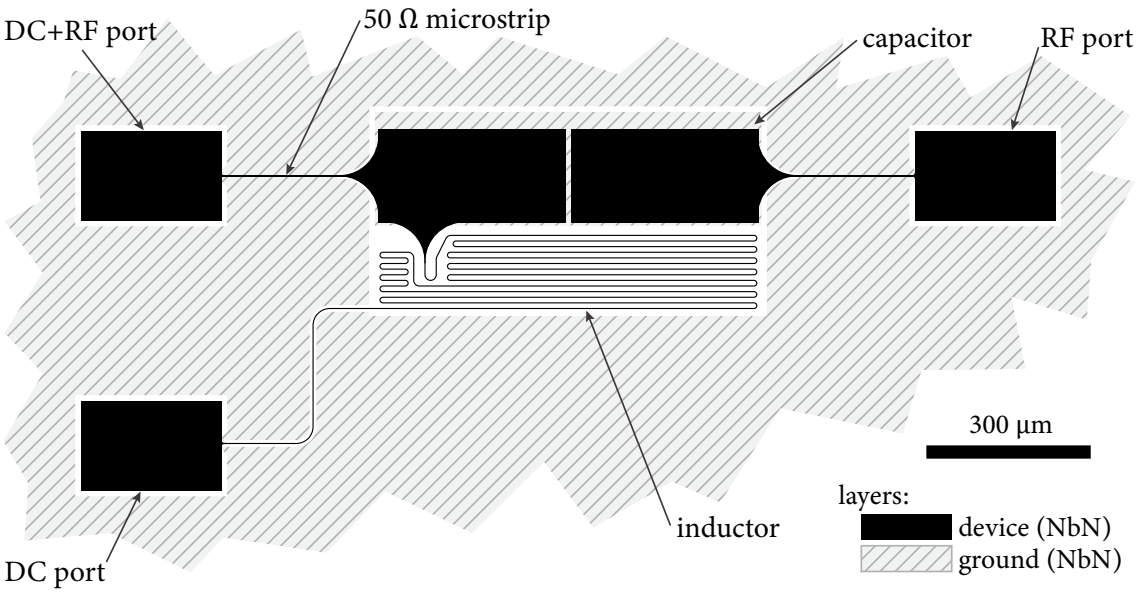

Figure 3-33: The layout of the bias tee. The capacitor, being realized by two series capacitors, has its center tap formed as an isolated island on the ground plane. The ground plane is removed from under the inductor and the lines leading to the inductor. The capacitor is formed from a series combination of two 64 pF parallel plate capacitors, and the inductor is $4,750\,\square \approx 95\,\text{nH}$.

As we need a capacitor between two terminals (as opposed to ground), we realize this capacitor as a series combination of two capacitors to an isolated section of the ground plane. Given that we need a total capacitance of $C = 32\,\text{pF}$, each of these two capacitors were designed for 64 pF. The inductor width was selected as $2\,\mu\text{m}$. This width is a trade-off between device size and current carrying capacity. The width of $2\,\mu\text{m}$ was selected as the transmission lines are around $2.4\,\mu\text{m}$ and need to carry the DC bias and RF bias, so making the inductor any wider than $2.4\,\mu\text{m}$ would not be useful. The inductor is laid out to maximize density while also minimizing coupling to the capacitor terminals and itself. With the actual device layout, the inductance was slightly larger than needed at $L = 95\,\text{nH}$, which will improve the low frequency performance of the tee.

The connections on the bias tee have large teardrops so that they can carry the higher current densities without issues from current crowding occurring. The bonding pads had the same teardrop size as was used for the calibration through so that we can de-embed the pads and only consider the performance of the bias tee. In a final

application, the designer would need to ensure that all lines that carry higher currents are appropriately treated to avoid current crowding.

3.10.2 Results

The experimental results along with the ideal response is shown in figure 3-34. It can be seen that the device performs very well over the measured range and matches the ideal response well. The 3 dB crossover between the isolation and insertion loss is at 91 MHz. For frequencies above 300 MHz we have an insertion loss less than 1 dB. At the designed frequency of 1 GHz we have excellent performance. The isolation can be seen to begin to decrease at around 1.5 GHz. This is more likely to be due to coupling between our bond wires than the device, but regardless, the isolation is sill within acceptable levels for a single stage DC line with no bypass capacitor.

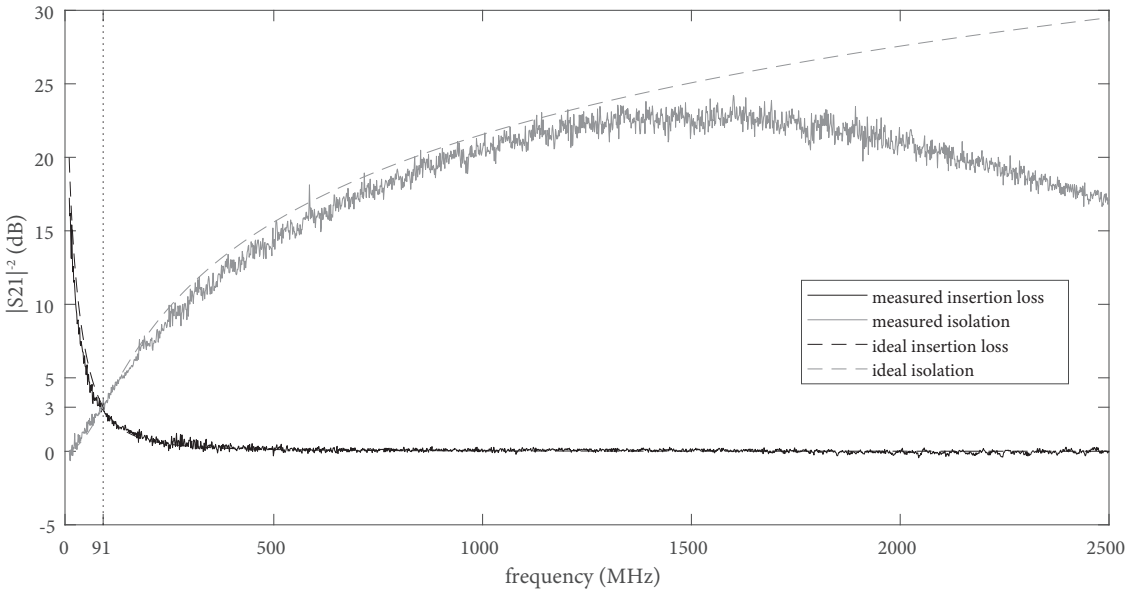

Figure 3-34: Plot of the insertion loss (RF port to DC+RF port) and isolation (DC port to DC+RF port) for the bias tee operating at 4.2 K. The ideal response is also shown using the designed values. The ideal and measured responses match well. The 3 dB frequency is 91 MHz and the device offers very low insertion loss beyond around 300 MHz.

3.11 A nanowire-based microwave detector

It has been long observed that superconductors and nanowires will react to microwave stimulus [87–89]. In particular, if one is to observe the low-frequency voltage current relation of a nanowire while simultaneously applying a microwave signal to the wire, it is readily apparent that the shape of the IV curve changes. In particular, the switching current is suppressed and the curve appears more and more shunted, taking on a sigmoid shape as bias is increased – see figure 3-35. For larger bias currents the residual current is seen to be unchanged along with the differential resistance. Near zero-bias the shape of the curve can be seen to vary dramatically.

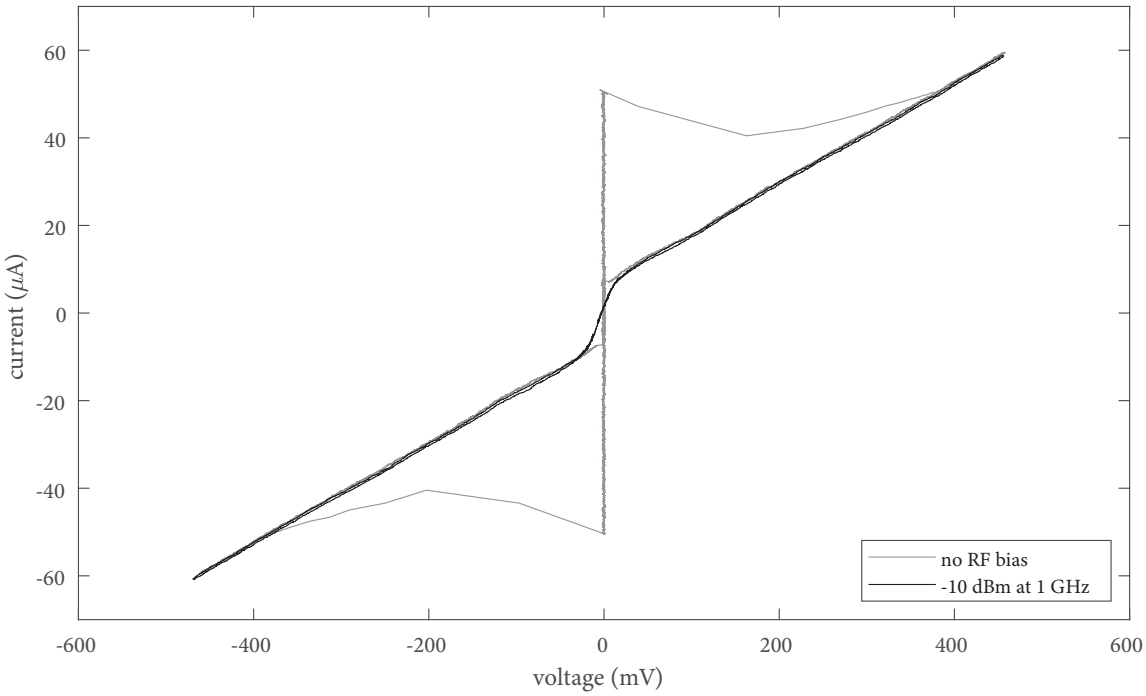

Figure 3-35: Current-voltage relation for a 100 nm nanowire with and without an RF bias applied to the nanowire by a bias tee. The switching current can be seen to be highly suppressed with the nanowire behaving resistively near zero bias but returning to its resistive with residual current relation at higher biases.

3.11.1 Design

The fact that IV curve of the nanowire changes to such a degree with the application of RF bias means that the nanowire lends itself particularly well to the detection of RF power. The measurement of the switching current would likely provide a very sensitive measure of RF bias – see section 5.3.3. However, here we will investigate the reaction of the nanowire to higher RF biases. To form our detector, we will voltage bias a nanowire near zero volts (here we used 7 mV) and apply an RF signal to the same wire while monitoring the current draw. The biasing and metering of the current was achieved with the instrument described in section 5.3.

The experiment is set up with a short 100 nm wide nanowire operating at 4.2 K in the CryoProbe. The device is supplied with a DC bias from the OmniBias set in the voltage-bias mode with a set-point of 7 mV, and the current gain set to 100 mV/μA. The current-sense output signal of the OmniBias is monitored by an oscilloscope. An RF signal is coupled into the line to the device through a room-temperature bias tee.

3.11.2 Model

We can model the suppression of the shunted nanowire behavior to a first order by simply integrating the idealized retrapping curve over a single cycle of an AC sine wave perturbation. This is essentially taking the load line of the nanowire with both a DC current bias I_b and an AC voltage bias $v_\mathrm{rf} = V_\mathrm{rf}\sin(\omega t)$ provided through a resistance R_b. The nanowire IV curve is modeled as a superconducting region up to an absolute current of I_r, which here is both the retrapping current and residual current, and extracted from the IV curve in figure 3-35 we have $I_\mathrm{r} = 7\,\mu$A. We then have that the differential resistance in the switched region is $R' = \mathrm{d}v/\mathrm{d}i$, which again from figure 3-35 we extract a value of $R' = 8.8\,\mathrm{k}\Omega$. We then perform the load line analysis as summarized graphically in part (a) of figure 3-36.

For a given AC signal v_ac and DC bias point I_b, after finding the operation point from the load line we have an output voltage signal $v(t)$ and current $i(t)$. Since we are only interested in the low frequency response of the nanowire to the signal, we

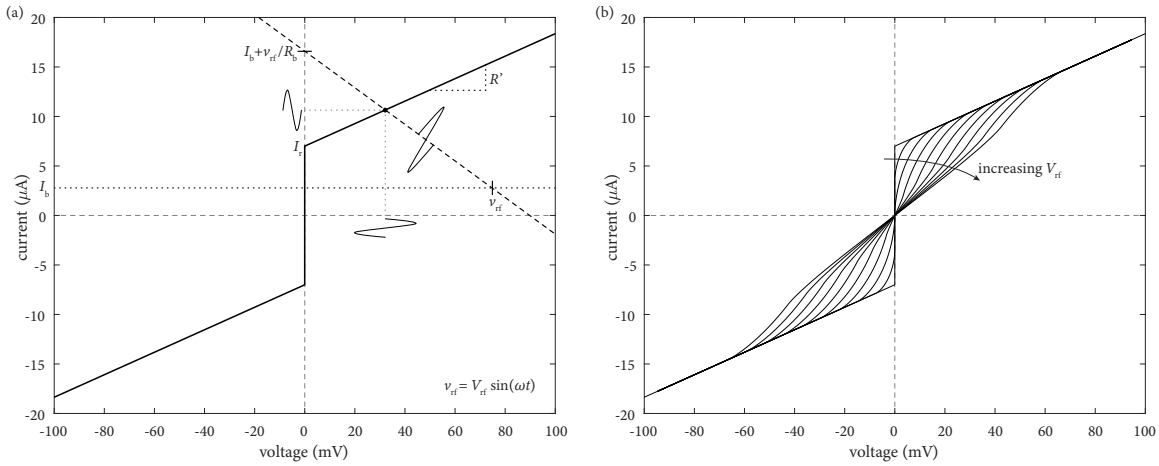

Figure 3-36: Current voltage relations for a nanowire. (a) Shows the retrapping portion of the IV curve for a nanowire. The nanowire curve parameters are those extracted from figure 3-35. The retrapping current $I_\mathrm{r} = 7\,\mu\mathrm{A}$ and is here equal to the residual current, and the differential resistance in the switched region is $R' = 8.8\,\mathrm{k}\Omega$. A DC current I_b offsets the load line for a voltage source with voltage v_rf and source resistance R_b. As the voltage source value varies with time according to $v_\mathrm{rf} = V_\mathrm{rf}\sin(\omega t)$, the intersection between the load line and the IV relation of the nanowire moves. If we integrate over a single cycle for various amplitudes V_rf, we would find that in the low frequency limit the IV curves are those shown in (b). The IV curves under RF bias shown in (b) are close to the shape of the experimentally obtained curves shown in figure 3-35.

low-pass filter these signals down to a single DC value by integrating over a single cycle as $V = \int v(t)\ \mathrm{d}t/2\pi$ and $I = \int i(t)\ \mathrm{d}t/2\pi$. Now, we can generate an IV curve for any given RF signal bias. A set of these curves is shown in part (b) of figure 3-36.

In order to generate the RF-suppressed IV curve model we found the load-line bias point as

$$
\begin{aligned}
v_\mathrm{rf}(\theta) &= V_\mathrm{rf}\sin(\theta) \\
I_\mathrm{n}(V_\mathrm{rf}, I_\mathrm{b}) &= \int_{\{0\le\theta\le 2\pi:|I_\mathrm{b}+v_\mathrm{rf}(\theta)|>I_\mathrm{r}\}} \frac{I_\mathrm{b}R_\mathrm{b} + v_\mathrm{rf}(\theta) + I_\mathrm{r}R'\mathrm{sgn}(I_\mathrm{b}R_\mathrm{b} + v_\mathrm{rf}(\theta))}{(R' + R_\mathrm{b})2\pi}\ \mathrm{d}\theta \\
I(V_\mathrm{rf}, I_\mathrm{b}) &= \int_{\{0\le\theta\le 2\pi:|I_\mathrm{b}+v_\mathrm{rf}(\theta)|\le I_\mathrm{r}\}} \frac{1}{2\pi}\left(I_\mathrm{b} + \frac{v_\mathrm{rf}(\theta)}{R_\mathrm{b}}\right)\ \mathrm{d}\theta + I_\mathrm{n}(V_\mathrm{rf}, I_\mathrm{b}) \\
V(V_\mathrm{rf}, I_\mathrm{b}) &= \int_{\{0\le\theta\le 2\pi:|I_\mathrm{b}+v_\mathrm{rf}(\theta)|>I_\mathrm{r}\}} \frac{1}{2\pi}\left(I_\mathrm{b}R_\mathrm{b} + v_\mathrm{rf}(\theta)\right)\ \mathrm{d}\theta - R_\mathrm{b}I_\mathrm{n}(V_\mathrm{rf}, I_\mathrm{b}),
\end{aligned}
\tag{3.36}
$$

where $I_n(V_{rf}, I_b)$ is simply a partial result representing the current flowing in the switched state. The bias current I_b is not the DC bias current through the wire. Rather, I_b is a free parameter that represents the offset in the RF source load line. Thus, in the IV curves shown in part (b) of figure 3-36, we fix V_{rf} and parametrically plot $I(V_{rf}, I_b)$ versus $V(V_{rf}, I_b)$ as we sweep I_b. In comparing the shape of the curves from our model (figure 3-36) to the measured data in figure 3-35, we can see a strong resemblance.

When operating the nanowire as an RF detector, we need to take the set of IV curves generated by equations 3.36 and convert them into a plot of DC bias current versus RF amplitude. To do this, we take a given value of V_{rf} and we solve $V(V_{rf}, I_b) = V_{DC}$, where V_{DC} is our DC voltage bias, to find I_b. With the appropriate value of I_b, we then simply solve $I(V_{rf}, I_b)$ to find the associated current. A result of this analysis is shown in figure 3-37, along with experimental results.

3.11.3 Results

The experiment was setup as described in section 3.11.1. A 6 GHz amplitude modulated (AM) signal was applied to the bias tee, and the OmniBias configured for a 7 mV constant voltage bias. The resultant bias current drawn from the OmniBias (the DC bias current) is shown in figure 3-37. Additionally, in this figure we plot the results of our simple model discussed in the previous section. It can be seen that our simple model matches well with the experimental results. There is likely a lot of room for improvement in this model; however, it does seem to relatively accurately capture the phenomenological behavior of the nanowire as an RF detector.

The only fitting parameter in this plot was the RF bias resistance R_b. We need to fit this value since we do not know the exact coupling of the RF source to the device. In the practical device this modeled resistance is in fact the absolute value of some impedance $R_b = |Z|$, where Z is influenced by device inductance, source impedance, and losses.

Due to the non-linearity in the detection mechanism, the detector can be used over a very large power range, while also giving high accuracy results. For low powers

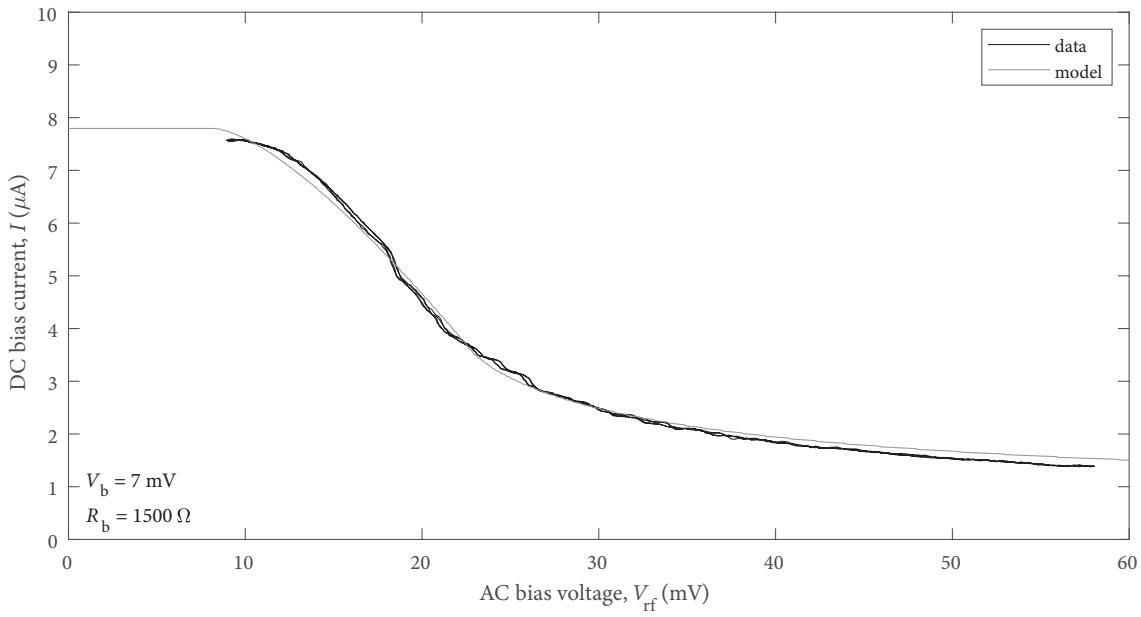

Figure 3-37: The response of a 100 nm wide nanowire voltage biased to 7 mV to a 6 GHz AM modulated signal. The results of the simple phenomenological model are shown on the same plot and can be seen to match well with the experimental data. In this experiment the voltage bias was generated by the OmniBias in the 100 mV/μA range for which the characteristics are presented in section 5.3.

we chose a voltage bias near 0 V, and for higher powers we chose a higher bias voltage. One could imagine implementing a control loop around the detector that varies the bias voltage to ensure the detector is operating in the most sensitive region of the response curve for the applied RF power.

A demonstration of the device being used as an AM radio receiver is provided in appendix A.

Chapter 4

A versatile cryogenic experiment platform – the CryoProbe

In order to test digital and microwave devices that we have developed in the proceeding two chapters, we needed a new experimental platform. Using the specifications we developed in chapter 1, we will present the design of a new CryoProbe. This device was used for all the experimental results presented in this work (unless otherwise noted).

This chapter will begin with an overview of the design. Then, specific aspects of the design will be covered in detail. Finally, the use and some performance metrics of the unit will be discussed.

4.1 Design overview

For our apparatus we elected for a "dipper probe" design [42]. This type of probe, while being the most simple cryostat design, offers the most versatility. The downside of the use of this style of probe is the requirement for liquid helium. As we already have a helium recovery system installed, this is not an issue for our purposes. The small quantity of helium that is not recovered is made up for by compressed helium deliveries. Given the low loss, the system running costs are similar to that of a cryogen-free system.

Conceptually the probe is very simple. The sample is mounted to the end of a low thermal conductivity tube and lowered into the liquid helium. For operating above the base temperature of around 4.2 K, the probe is withdrawn from the liquid and cooled only by the vapor.

An overall diagram of the probe is shown in figure 4-1. The probe can be divided into a number of sections. We have cold end of the probe which includes all the parts that interface to the sample. We have the probe-body which is the main structural connection between the cold end and the room temperature portion of the probe. Finally, we have manifold which is main room temperature section of the design and houses all connections to the probe. These portions of the design are covered in detail in the following sections.

4.1.1 Weight

The mechanical design of the probe was made with careful attention to weight, while still ensuring that the functionality of the probe is given the highest priority. The reason for keeping the weight of the probe low is twofold. On the cold-end the more mass there is to be cooled the more helium we will consume during cool down. Overall, the weight of the probe needs to be low so that we can easily maneuver it into and out of the dewar. With this in mind, the manifold was design with weight reducing strategies [90]. The final weight of the complete probe is under 25 lb, meeting our requirement from section 1.4.2.

4.1.2 Safety

When designing any experimental apparatus, safety must be kept in mind. This requirement is even more important for apparatuses that utilize cryogenic liquid. One liter of liquid helium expands to 780 liters of gaseous helium at room temperature [91]. Thus, it is critical that anywhere that could contain liquid helium has appropriate pressure relief valves installed. Here, there are two volumes that do or could potentially contain cryogenic liquid, the dewar and the probe itself. The dewar has built

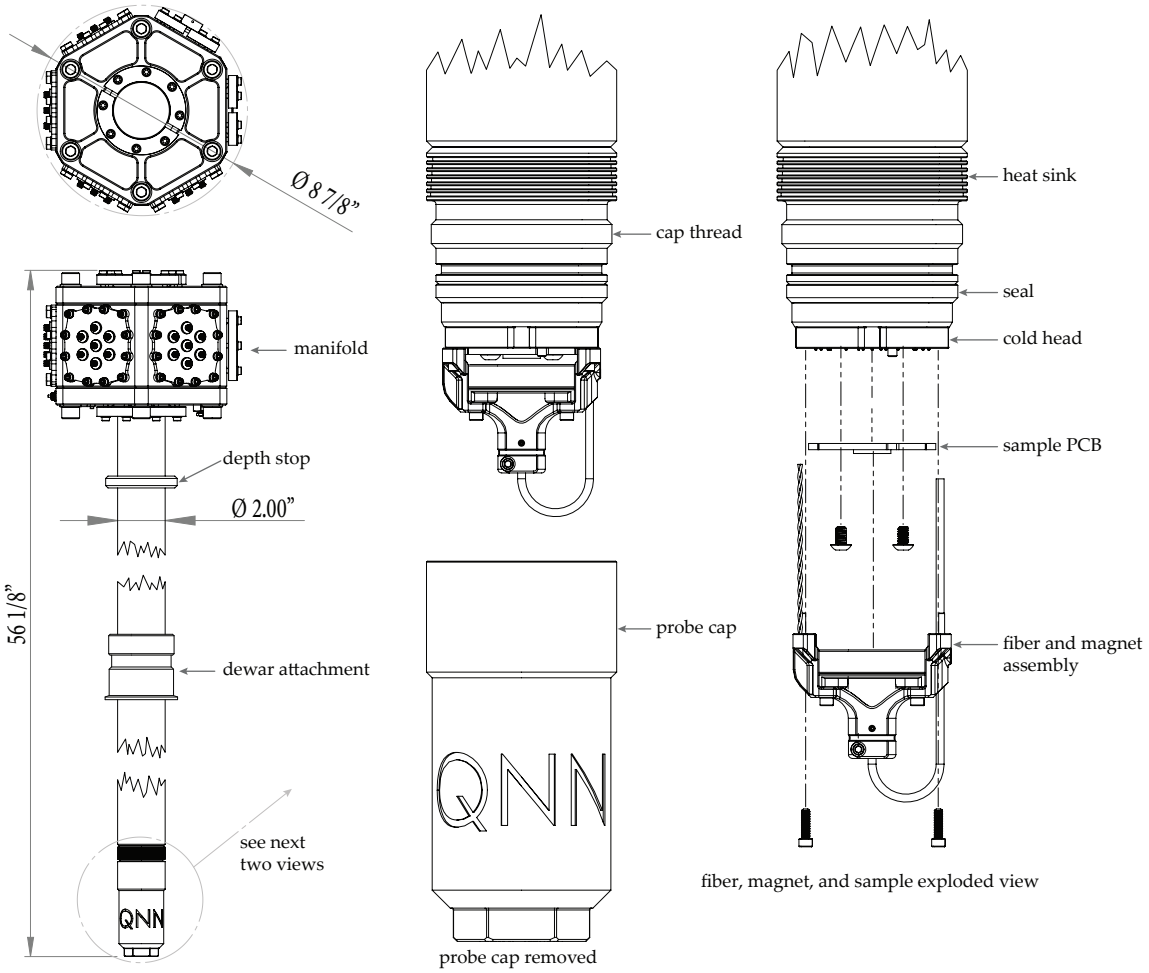

Figure 4-1: Overview of the new probe design. (left) The complete probe design viewed from the front and top. The probe is configured for an experiment with the cap installed. (center) The cold-end of the probe with the vacuum cap removed showing the internal cold head, sample, and optical fiber and magnet assembly. (right) Exploded view of the probe end with the fiber and magnet assembly as well as the sample removed.

in protection valves, and the probe body was designed to not interfere with these systems. For the probe, a pop-off relief valve is installed in the manifold. This valve actuates at a very low pressure around 5 psi, and prevents any kind of hazardous pressure buildup in the probe.

During a cool down, the top of the dewar can become cold due to the flowing gas. When the probe is removed from the dewar after an experiment, the body of the probe will be very cold. Contact with the probe or dewar neck during or after an experiment can be hazardous. Thus, appropriate personal protective equipment is required when operating the probe to avoid injury. Additionally, if the helium recovery system were to leak, or the dewar to vent, the helium could displace oxygen in the room leading to potential asphyxiation. For this reason, the room is equipped with an oxygen monitoring system.

4.2 Cold head assembly

The cold head assembly is the collection of parts that make up the sample interface. The cold head assembly is the most interesting and unique aspects of this probe design. An exploded view of the cold head assembly without the fiber and magnet assembly is shown in figure 4-2.

One of the major issues we encountered while performing immersion testing was the use of soldered-on connectors on the PCB. Thus, we decided to explore alternative methods of achieving high bandwidth connections to our samples without the need for connectors. We make extensive use of spring test probes or "pogo pins" in our DC setups for room temperature and cryogenic testing of films. An RF pogo pin that worked at cryogenic temperatures would be ideal for this application.

We investigated Samtec Bulls Eye connectors (BDRx series) and found that these connectors work well at liquid helium temperatures [92]. Thus, we selected these devices for our RF pogo pins. The use of these connectors allows all the connector design to be incorporated into the PCB and the CryoProbe. Thus, there is no need for soldering connectors onto the PCB. The special considerations for the PCB design

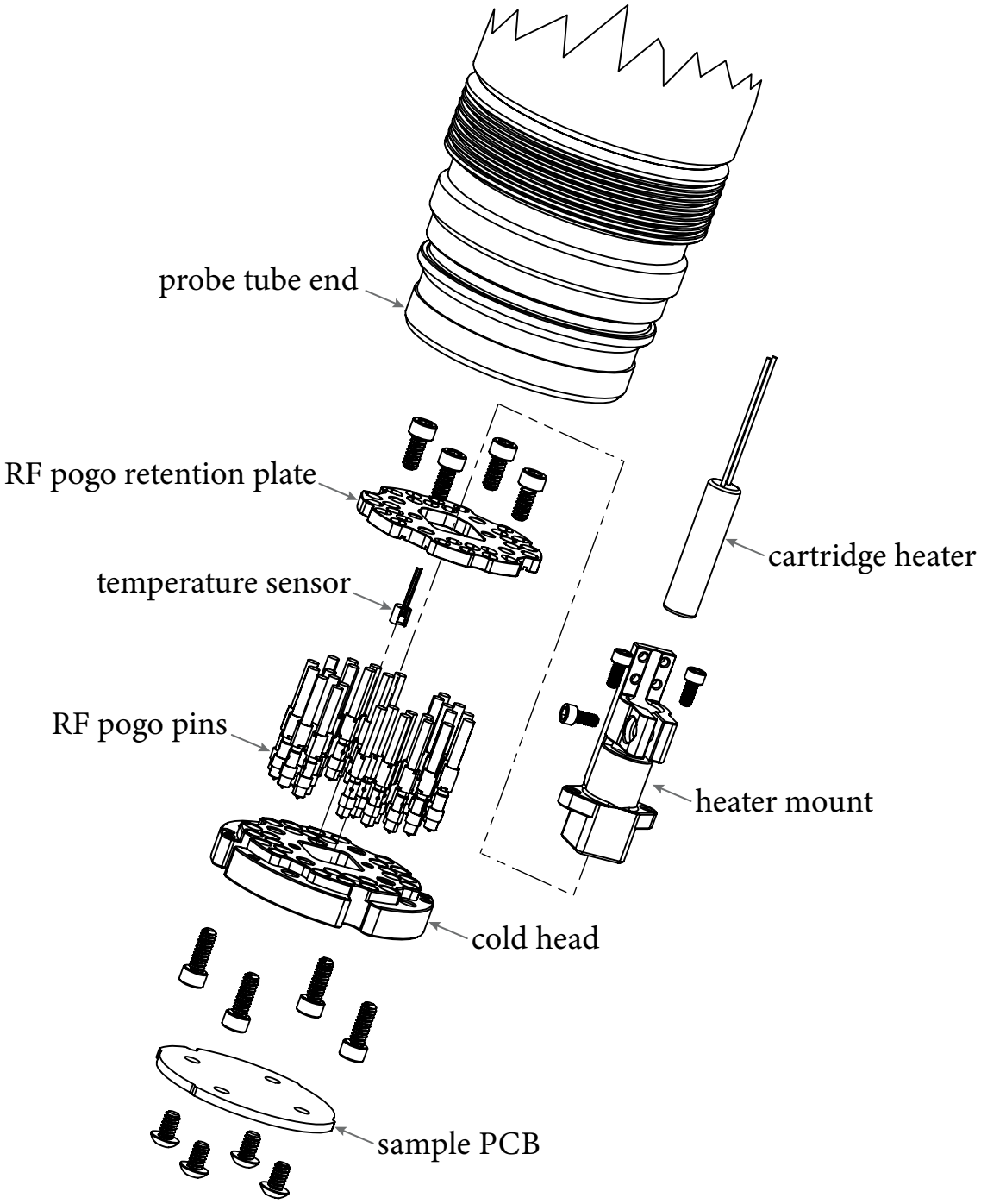

Figure 4-2: Exploded view of the sample mount and cold head assembly. The cold head is central to the assembly. The RF cables, temperature control hardware, fiber and magnet assembly (not shown), and sample all attach to the cold head. Note: Only the end of the RF cables are shown, and the wires for the heater and temperature sensor are not shown.

are covered in section 4.2.2.

The design of the sample mount is largely driven by the use of the Bulls Eye connectors. These connectors are retained by what amounts to a bayonet-style connection. The connector termination is inserted into the connector body. Once installed, a retention ring with two protruding radial tabs is inserted into the connector body and rotates one quarter turn. Once the connector is rotated, it springs into a locking groove preventing it from rotating and becoming loose. The details of the connector and the retention mechanism are shown in figure 4-3.

Samtec only sold sets of eight phase-matched cables at the time the probe was made. Thus, given that we have a total requirement of 28 cables, we divided the cables into four banks with 10 ps phase matching between the cables in these banks. These banks are named "A" through "D" with cables 1 to 7 being in bank "A", 8 to 14 in bank "B" and so on. The manifold was designed to group the connectors into these same banks, with all connectors in a bank being on the same side of the manifold – see section 4.4.3. The connectors on the cold head are divided in a similar manner. Thus, for a PCB designed for a square chip we have that all pads along one side can be within the same bank – see section 4.2.2.

A detailed view of the cold head and the dimensions of the connector pins and their locations is shown in figure 4-4. This figure gives all the dimensions required to design samples to interface to the probe. An example PCB is shown and discussed in section 4.2.2. There are a number of key features of the cold head that must be understood in order to design samples to interface to it. The full area of the cold head can be used; however, if the sample blocks all the radial notches, pumping the probe will be slow. Additionally, if the fiber and magnet assembly is to be installed, then the region shown in detail B of figure 4-4 must be left clear.

The four banks of connectors and the four mounting screws are rotationally symmetric. This symmetry could easily lead to a sample being installed in the incorrect orientation. There are two features that ensure the proper installation of the sample. First, there is a visual marker machined into the surface of the cold head indicating the pin 1 location. The second feature is a 1/8" dowel pin installed next to the pin 1

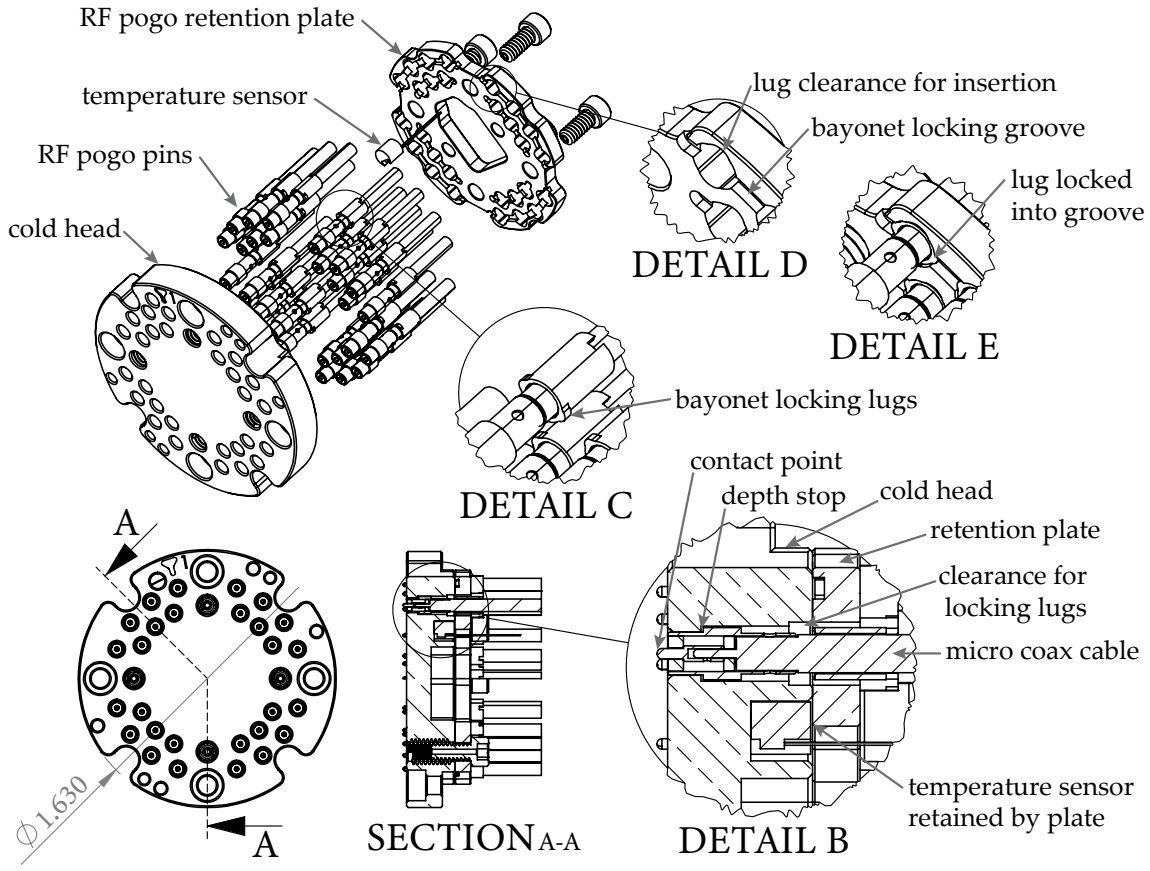

Figure 4-3: Details of the RF pogo pins within the cold head assembly. Section A-A cuts the assembled cold head revealing the details of the connector and temperature sensor pocket. Detail B shows the RF pogo pin, the retention geometry in the cold head and retention plate. Detail C shows the bayonet locking lugs on the RF pogo pins in the exploded view of the cold head assembly. Detail D shows the clearance and notch which allows for the installation of the bayonet portion of the RF pogo pins, as well as the locking groove. Detail E shows the same view as detail D, but in the assembled state highlighting the lug locking into the locking groove of the retention plate. Additionally, the details of the temperature sensor and the mechanism by which it is retained by the RF pogo pin retention plate are shown. Note: Only short sections of the micro coax cable are shown, the temperature sensor wires are not shown, and the RF pogo pin springs are not shown.

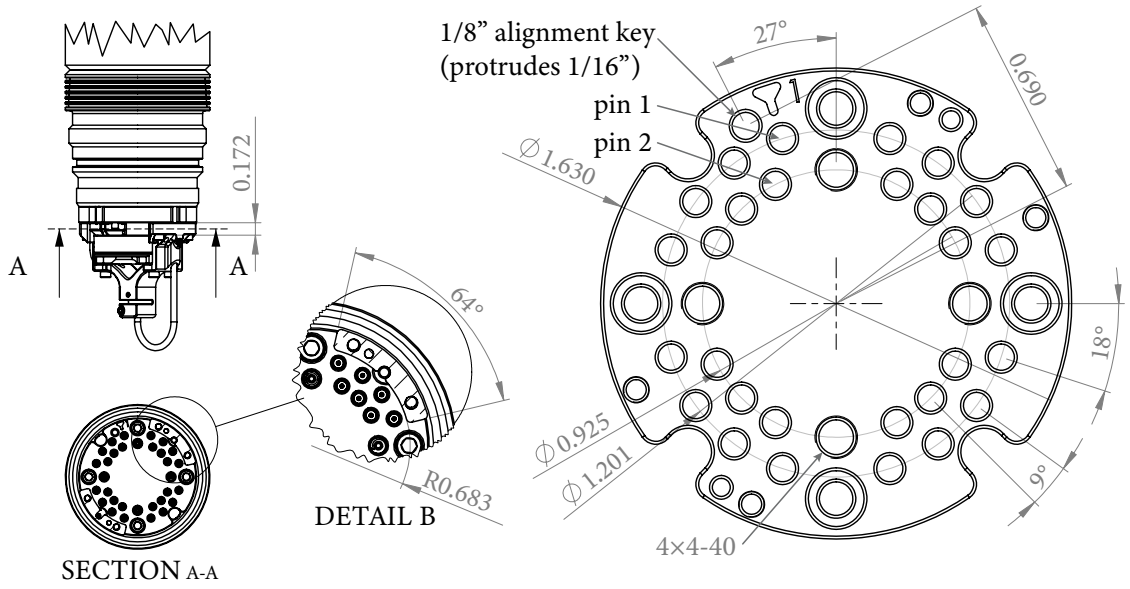

Figure 4-4: Detailed view of the cold head with the dimensions required to design a PCB, or other sample mount, to fit in the probe. (left) View of the probe with the fiber and magnet installed showing the regions with the sample must fit to not interfere with the magnet and fiber assembly mounting. Detail B shows the dimensions of the area that the fiber and magnet assembly occupy. (right) The dimensions of the cold head and pin numbering. Note that the 4-40 UNC holes appear larger on this drawing as they are installed with helical inserts which require larger threaded holes.

location. If the PCB is designed with a notch or hole in this location then the PCB can only be installed in one orientation. This pin only protrudes $1/16$" so as not to interfere with the sample space greater than the PCB thickness.

All parts that make up the cold head assembly are machined from OFHC copper. This provides excellent thermal conductivity even at liquid helium temperatures [42]. The downside of using OFHC copper is that it is harder to machine, very soft, and susceptible to corrosion. The machinability was addressed by using appropriate speeds and feeds for the material. Corrosion of the parts is important to avoid as the oxidized surface will create poor thermal contact between parts. One method to passivate the surface is gold plating. The use of soft gold planting passivates the surface and enhance thermal conductivity between the surfaces (twenty times lower thermal resistance than copper to cooper [42]). Here, all mating surfaces that required good thermal conduction are made flat and smooth to allow for good mating and thermal conductivity. With the mating surfaces appropriately prepared, the parts were elec-

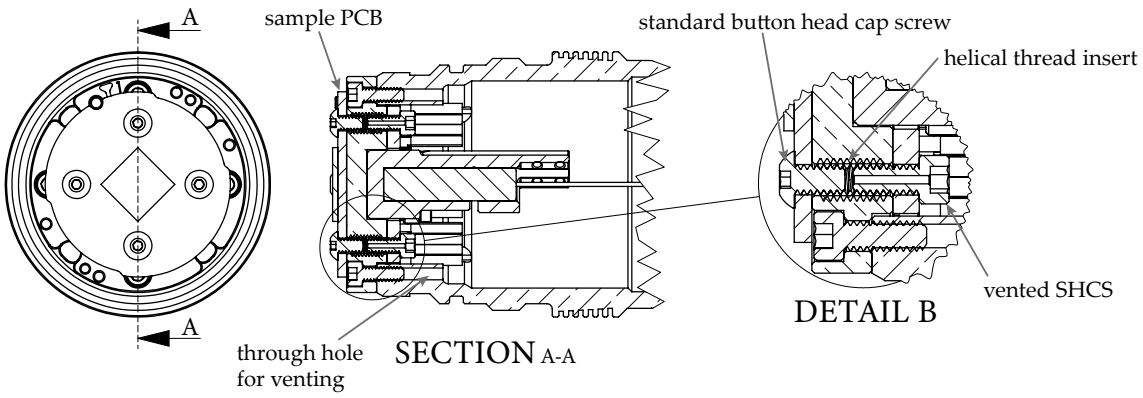

Figure 4-5: Cross section view of the cold head showing the details of the thread inserts. Additionally, the details of the venting of enclosed spaces within the probe are called out. A hole is drilled through the socket head cap screws (SHCS) that clamps the RF pogo pin connector plate onto the cold head. This hole allows trapped gas to escape and allows conventional button head screws to be used for mounting the sample.

troplated with gold to a minimum thickness of 100 μin. Many gold plating processes involve the deposition of a nickel (or other metal) on to the surface prior to the gold plating. This is not acceptable for our purposes as other interfaces will introduce increased thermal resistances. In addition, nickel is hard and is ferromagnetic, both of which are not desirable for our application. Thus, the plating process used here consisted of only electroplated gold and no other adhesion processes.

The softness of the OFHC copper used for the cold head makes the threads susceptible to damage after many cycles of the sample being installed and removed. In order to prevent such damage from ever occurring, stainless steel threaded inserts are used in the sample mount holes. In addition to wearing harder than the copper, the inserts spread the force from the screw over a larger portion of the base material owing to their larger outside diameter. A view of the cold head showing the thread insert details is shown in figure 4-5.

4.2.1 Ground loops

Ground engineering is critical in sensitive electrical measurements and even more so when the device presents a very low impedance – such as a superconductor. One of the

greatest sources for interference that we encounter are ground loops. At the core of the issue is that when a ground loop is formed, our assumption that all grounds are at the same potential breaks down. There are a number of methods to compensate for these effects, including using differential measurements. Ideally, when performing sensitive measurements, we would like the circuit between the source and the device to make a single loop which has a near-zero cross-sectional area (such as coax cable). Under this ideal setup, we can be sure that all current from our source will pass through our load. Furthermore, this setup is resilient to external interference as voltages coupled to the circuit are coupled to signal and ground equally, and magnetic pickup is nearly non-existent as our loop area is small.

Many cryostats do not avoid ground loops in their construction. The coax cables connect to the sample are connected to the ground at the sample. Then again, the coax shield is connected to ground at each stage in the cryostat up to room temperature. Finally, at room temperature the coax shield is again connected to the vacuum chamber of the cryostat. For cryostats that use cryocoolers, the shield is grounded at each stage of the cooler. This grounding is performed as higher stages have much higher cooling power than lower stages. Thus, it is desirable to thermalize cables at each state to limit the heat flux to the ultimate stage. These multiple ground loops within the instrument can cause multiple issues, not least of which are thermoelectric effects from the ground and signal taking different paths, microwave issues as grounding path impedances can vary at different frequencies, and pickup from croycooler motors.

The standard method to eliminate ground loops is by selecting a single point from which all signals are referenced. This point is referred to as the "star" ground as all other ground signals are connected to this point. The name is derived from the conceptual (and often physical) circuit that is formed with this setup where all ground connections radiate from this central point forming a star pattern. We elected for this approach in our design of the CryoProbe.

In the probe the star point is the cold head, as shown in figure 4-6. All coax cables are thermalized and electrically connected to the cold head. The cables are

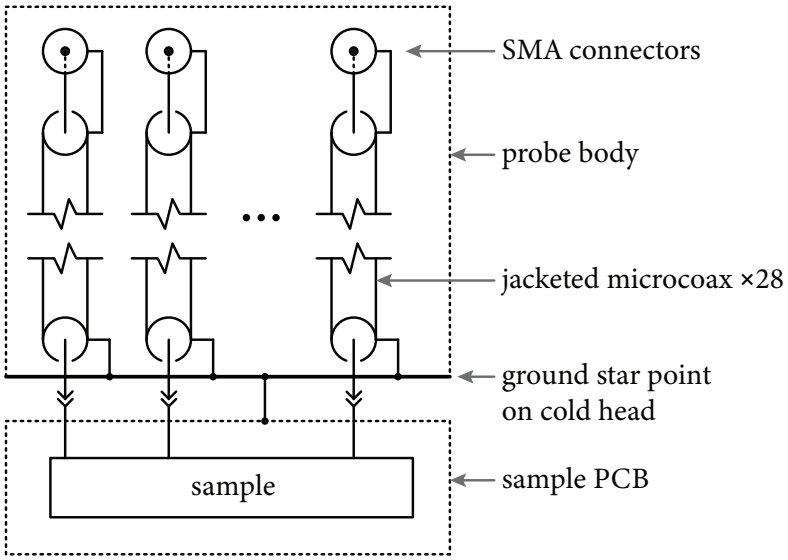

Figure 4-6: Simplified schematic of the grounding system employed in the CryoProbe. All cable grounds are connected to the copper cold head. The cables leading to the manifold are all insulated and their SMA terminators are mounded in an insulating bulkhead. The sample PCB ground plane is screwed directly onto the cold head star point. Thus, we introduce no ground loops within the probe. Note that no grounding is shown to the sample as it is up to the user how the sample PCB ground is connected to the sample.

jacketed and do not again come into electrical connection. They are mounted into the cold head with an insulating bulkhead to avoid grounding at room temperature. The heater and temperature sensor are floating and, as such, do not contribute to any grounding issues.

This leaves only the sample PCB and the probe user's experimental setup as possible sources for ground loops. For the sample PCB, such as the design presented in section 4.2.2, we utilize many ground planes to allow for the formation of microstrip transmission lines. While this does introduce ground loops, the impedance of these loops is vanishingly small and their cross-sectional area is virtually zero.

4.2.2 Sample PCB

The sample PCB is one of the main distinguishing features of this design. Owing to the RF pogo pins, we do not require any connectors on the PCB. Rather, we only

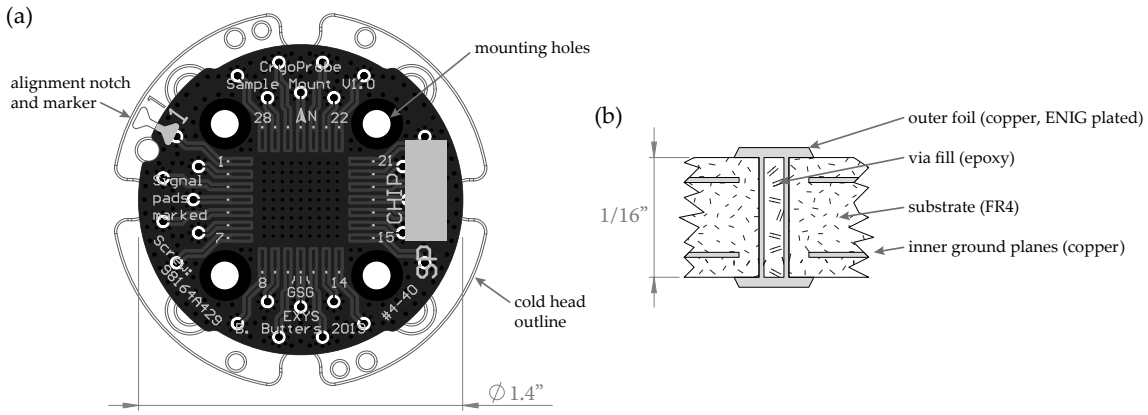

Figure 4-7: Details of the sample PCB design. (a) The sample PCB design. All PCB layers are shown stacked on top of each other with an outline of the cold head shown on the bottom. The annotations on the solder mask can be seen on top. Note the alignment marker, the pad labels, the region for chip number and the screw part number and type. The traces between the RF pogo pin pads and the chip bond pads are all designed for a characteristic impedance of 50 Ω. All other layers are simply ground fills. (b) The simplified layer stack-up showing a filled signal via. The layers and spacing in this figure are not to scale.

require contact pads on the reverse side of the PCB. A number of PCBs have been designed for the probe, some with pads for resistors and other components, and some simply for microwave connections to the sample, such as that shown in figure 4-7. The PCB is designed with contact pads and mounting hole patters from figure 4-4.

All of the PCBs designed for the CryoProbe have had similar construction. All boards have been designed with 50 Ω microstrips which largely dominated the layer stackup design. In order to achieve a 50 Ω microstrip with a standard 1/16” thick two layer PCB, we require a trace width of around 0.12”. This is far too wide to fit seven signals separated ground pads along one side of a 10 mm × 10 mm chip. To address this issue a four layer PCB was used instead. This allowed us to have the same PCB thickness but only require trace widths around 0.008”, while also using a standard low-cost fabrication process. Thus, we can fit all seven signals with ground pads between them along each chip edge.

Keeping track of many samples can become difficult when the samples and PCBs have no markings. To address this issue, the PCB is printed with an area to record the chip number. Thus preventing confusion and loss of samples. In addition, as losing

the mounting screws is an inevitability in a lab, the screw part number is printed on the PCB. This allows users in the future to find or order the correct screws without the need to search for the probe documentation – or worse use, the incorrect screw resulting in damage to the cold head.

For thermal contact, and to allow for wire bonding, the PCB is gold-plated. The reverse side of the PCB which mounts against the cold head is totally void of solder mask and silk screen markings. The entire surface is a single ground pour with only the area around the RF pogo pins removed for isolation. This high fill ratio ensures that the PCB surface is extremely flat. While initially expecting to require thermal grease to interface the PCB to the cold head, the PCB design along with the gold plating on the cold head provide excellent thermal conductivity. In experiments where we have placed temperature sensors on the PCB and measured the different between the PCB and the probe sensor we have seen no measurable difference in temperature under greater than normal thermal loading ($> 500\,\mathrm{mW}$). Thus, we find that no thermal interface material between the sample and cold head is necessary. This allows us to achieve both thermal and electrical grounding on the back side of the PCB effectively.

Many vias are placed under the sample to increase the thermal conductivity between the sample and the cold head. These vias were filled to prevent the varnish used to mount the sample onto the PCB from wicking to the underside of the PCB. Any contamination on the underside of the PCB would prevent it from making intimate contact with the cold head. Additionally, the use of filled vias allows for the placement of the signal vias for each of the RF pogo pins in the center of the pad. This is necessary since the reverse side of the PCB cannot have any solder mask to insulate traces as the thickness of the mask would again prevent the board from making intimate contact with the cold head.

To prevent the PCB from being installed into the probe in the wrong orientation, a locating pin was pressed into the cold head and a corresponding notch was designed into the PCB. The locating pin interferes with the PCB in all but the correct orientation, thus preventing the mounting holes from aligning in the incorrect rotation.

In addition, a "pin 1" marker is machined into the cold head and a corresponding marker is placed on the PCB silk screen. An incidental advantage of the locating pin and notch design is that it prevents the PCB from rotating when only one screw is installed. Fixing the PCB as such prevents possible damage to the sample and pogo pins should the user attempt to rotate the PCB to align the screw holes.

4.2.3 Fiber and magnet assembly

The fiber and magnet assembly allows for the exposure of the sample to both optical flood illumination and axial magnetic fields. An exploded view of the assembly is shown in figure 4-8. The assembly is made from four parts: the support yoke, the fiber tube, the optical fiber, and the magnet. The support yoke is central to this assembly and is the hub onto which all other parts are mounted.

The fiber and magnet assembly is registered to the cold head by two $1/16$" dowel pins located on opposing sides of the yoke. The dowel pins interface with holes on the cold head. The use of two dowel pins in this configuration results in the system of the yoke and the cold head being over-defined. To overcome this, one hole is a close fit and the other is a slightly looser fit. This results in the system being under-defined; however, the tolerances were selected such that the theoretical reproducibility in the fiber center location is around 0.001". This tolerance is more than is necessary for our application. The alignment dowel pins perform a secondary use when installing and removing the fiber and magnet assembly. It can be difficult to hold the probe, and the fiber and magnet assembly while installing the mounting screws without having the yoke move out of alignment. The dowel pins prevent the yoke from moving while the mounting screws are installed, thus making assembly easy.

The optical fiber is vulnerable at the end of the probe and having to undergo a round turn in a relatively small space. To prevent damage to the fiber, and to prevent it from being flexed after assembly, a protective metal tube is installed around it. This tube guides the fiber and ferrule through a gradual bend and into the yoke. A flexure clamp is used to secure the fiber tube into the yoke. The use of a flexure clamp has a number of advantages. It provides excellent thermal contact between the tube and

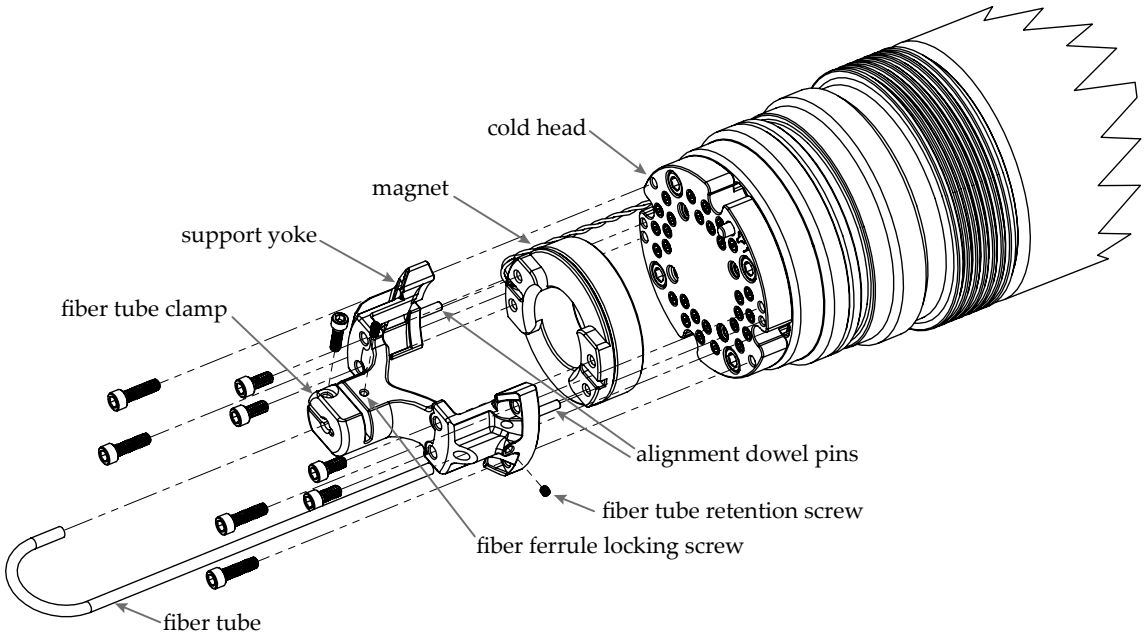

Figure 4-8: Exploded view of the fiber and magnet assembly along with its mounting location on the cold head. The support yoke is central the the assembly with all other parts of this assembly mounted to it. When the fiber and magnet assembly is installed on the probe, four mounting holes and two dowel pins locate and fasten the probe onto the cold head. The magnet is mounted to the yoke by four screws. The fiber tube is clamped to the yoke by a flexure clamp on one end and a set screw on the outside. The fiber ferrule is fixed in place with a single set screw. Note: The fiber is not shown in this figure.

the yoke as the complete circumference of the tube (less the area of the slot) is in contact. The clamping force is high in comparison to other methods. The location of the clamped tube will be highly repeatable. Finally, the whole clamping assembly only required a single screw and washer and no additional parts. On the outer portion of the yoke where the fiber tube enters into the body of the probe, there is insufficient space to use another flexure clamp. Instead, for this location the fiber tube is fed through a tight tolerance hole and a set screw secured the fiber tube.

On the cold end, the optical fiber is terminated into a 1.25 mm ceramic ferrule. This ferrule mounts into the pocket shown in detail B of figure 4-9. The pocket is oversize by 0.002" to allow for installation and removal without damaging the ferrule or its polished end surface. The other end of the fiber is terminated into a FC/PC connector. It would be ideal to install the fiber just on the inside of the probe body at the cold head. The issue with this approach is that the FC/PC connector would be exposed to cryogenic temperatures. In the past we have had issues with increased attenuation in cold FC/PC connections. In addition, 28 micro coax cables take up a large portion of the space available and are far more rigid than the somewhat more delicate fiber. So, we need the FC/PC connection to be made in the manifold at room temperature. The downside of this approach is that we need to feed the relatively large FC/PC connector from the cold head through the length of the probe to the manifold. Feeding the connector would be difficult with all the other cables within the probe. To make this process simple, and to protect the fiber, an inner tube was welded inside of the fiber body. This tube has an inner diameter slightly larger than the FC/PC connector and allows us to simply drop the fiber through the probe. The end of this inner tube can be seen in the top of section A-A in figure 4-9. The fiber tube on the fiber and magnet assembly enters inside of the inner tube of the probe to keep the fiber isolated from all the cables within the probe.

The superconducting magnet is wound onto a gold-plated OFHC copper coil form (bobbin) that provides heat sinking for the magnet as well as mechanical support. There are four mounting holes in the coil form that attach it to the yoke. The superconducting magnet is wound from formvar-insulated copper-clad NbTi wire with

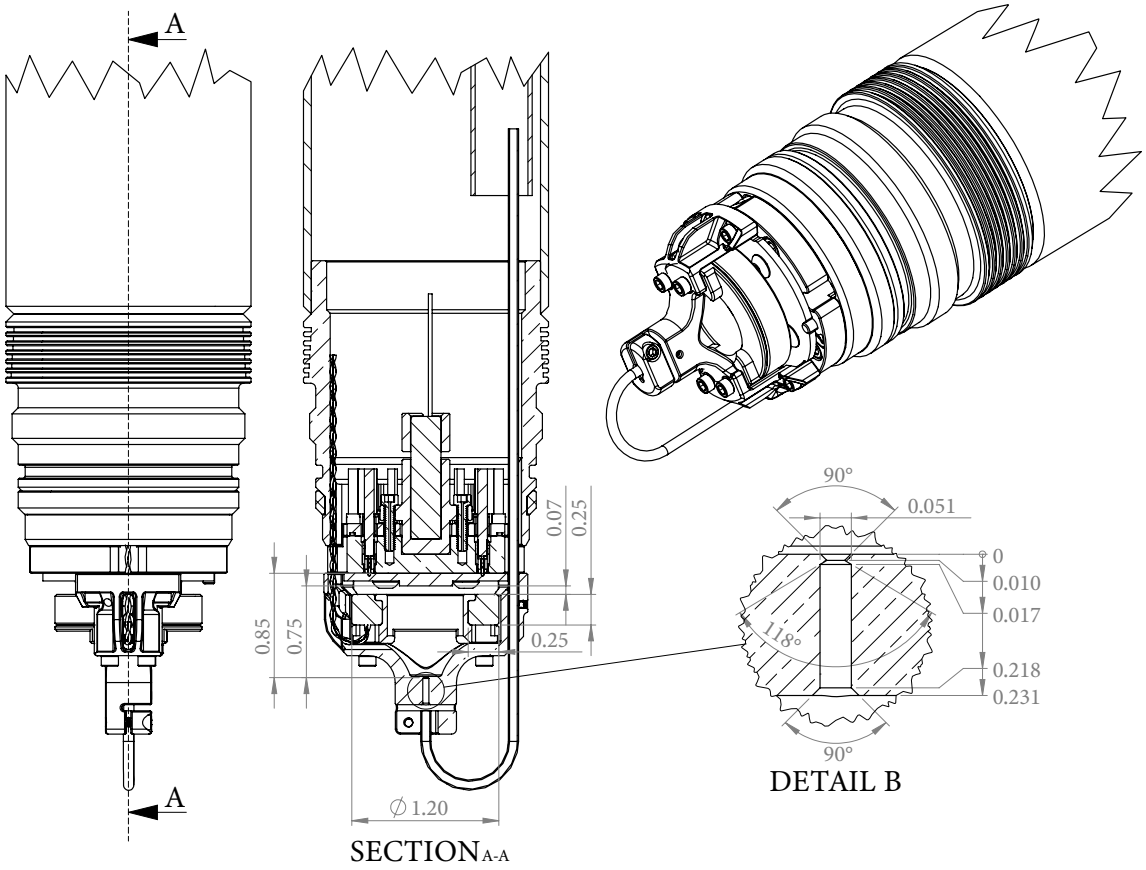

Figure 4-9: Drawing of the fiber and magnet assembly mounted onto the end of the probe. All of the major dimensions of the magnet and fiber are shown. Detail B shows the dimensions of the fiber ferrule mounting features. Note that the fiber and fiber ferrule are not shown in this figure.

a total diameter of 102 μm (Supercon model SC-T48B-M-0.079mm). The windings are thermalized and fixated by GE varnish. The winding area was made as large as possible within the constraints of the available area of the probe. Additionally, the coil was made to be as close to the surface of the sample without interfering with it. Thus, we maximize the field we can achieve on the surface of the sample.

The Biot-Savart law states that the magnetic flux density is given by

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \iiint_V \frac{(\mathbf{J}\, \mathrm{d}V) \times \mathbf{r}'}{|\mathbf{r}'|^3} \tag{4.1}$$

where $\mathbf{r}$ is our sample point, $\mathbf{r}'$ is the vector between $\mathrm{d}V$ and $\mathbf{r}$, $\mathbf{J}$ is the current density in the volume V which is the windings in our case [93]. For our case we can consider the axial field B_z at the surface of the same z_s to be given by

$$B_\mathrm{z}(z_\mathrm{s}) = \frac{\mu_0}{4\pi} \frac{IN}{|(r_2 - r_1)(z_2 - z_1)|} \int_0^{2\pi} \int_{r_1}^{r_2} \int_{z_1}^{z_2} \frac{r^2}{((z_\mathrm{s} - z)^2 + r^2)^{3/2}}\, \mathrm{d}z\, \mathrm{d}r\, \mathrm{d}\theta, \tag{4.2}$$

where I is the current into our magnet, N is the number of turns in the magnet, and r_1 and r_2 and the inner and outer radius of the magnet and z_1 and z_2 are the locations of the top and bottom of the magnet with z_s being the location of the surface of the chip. Our magnet has around $N \approx 1000$, and with the dimensions shown in figure 4-9, the magnetic field will be approximately $B_\mathrm{z}(z_\mathrm{s}) = 0.041I\,[T]$. With a maximum current through the wire, giving substantial de-rating of $I = 7\,\mathrm{A}$, the magnetic field can be up to $B_\mathrm{z}(z_\mathrm{s}) = 0.41\,\mathrm{T}$.

4.3 Probe Body

The probe body is a weldment composed of a number of parts. The completed weldment is shown in figure 4-10. The main body of the probe is a 2" diameter 316 stainless steel tube. As the tube size sets a limit on the size of the cold head, it is desirable to use the largest tube possible. We use "wide-neck" helium dewars which have a 2 1/2" sanitary flange, leading to a 2" diameter tube being the largest possible (using commonly available tubing) that will fit within the neck and provide sufficient

space for gas flow. It is critical to allow a region for gas to flow around the probe. Making the probe fit tightly within the dewar neck would block all routes for gas to escape the inner flask, leading to the internal pressure increasing, and possibly lead to the explosive failure of the dewar. It should be noted that this is extremely unlikely even if we were to block the neck of the dewar and defeat its over pressure protection system. The only thing that keeps the probe in the dewar is its weight, which is less than 25 lb. Given the probe has a 2" cross section, a pressure in the dewar higher than 8 psi could lift the probe out of the dewar (ignoring friction from the o-ring seal). Given that the dewar is rated for an operating pressure of 10 psi, and a failure pressure much greater than that, we are at little risk. Regardless, we need an annular space for gas flow to cool the probe and allow us to recover the helium boil-off.

Of particular note is that the fiber inner tube is only welded to the lower support. This leaves the upper support to only prevent lateral movement of the inner tube. As the outer body of the probe is likely to cool down and warm up much faster than the internals of the probe, it is expected there will be a differential thermal contraction and expansion between the inner and outer tube. If the inner tube were welded to the outer tube on both ends then there would be substantial forces on the welds and bowing of the inner and, to a lesser extent, outer tube. Thus, the inner tube is allowed to move independently of the outer tube, thereby generating no stress. Additionally, as the inner tube is simply to support the installation and removal of the optical fiber, it experiences no loading, so fully welding it is not necessary.

The cold end of the tube is composed of a machined section of OFHC copper which is soldered to the body of the probe. Performing this soldering with the complete length of tubing would be difficult as internal access to feed solder to the joint would be nearly impossible. In addition, heating the assembly with the long tube attached would be difficult. Thus, the OFHC copper part is only soldered to a small section of the tubing. After the soldering is complete, the end assembly is sent for gold plating and upon its return is welded to the remaining portion of the probe to form the complete assembly. An additional benefit of this manufacturing procedure is that it

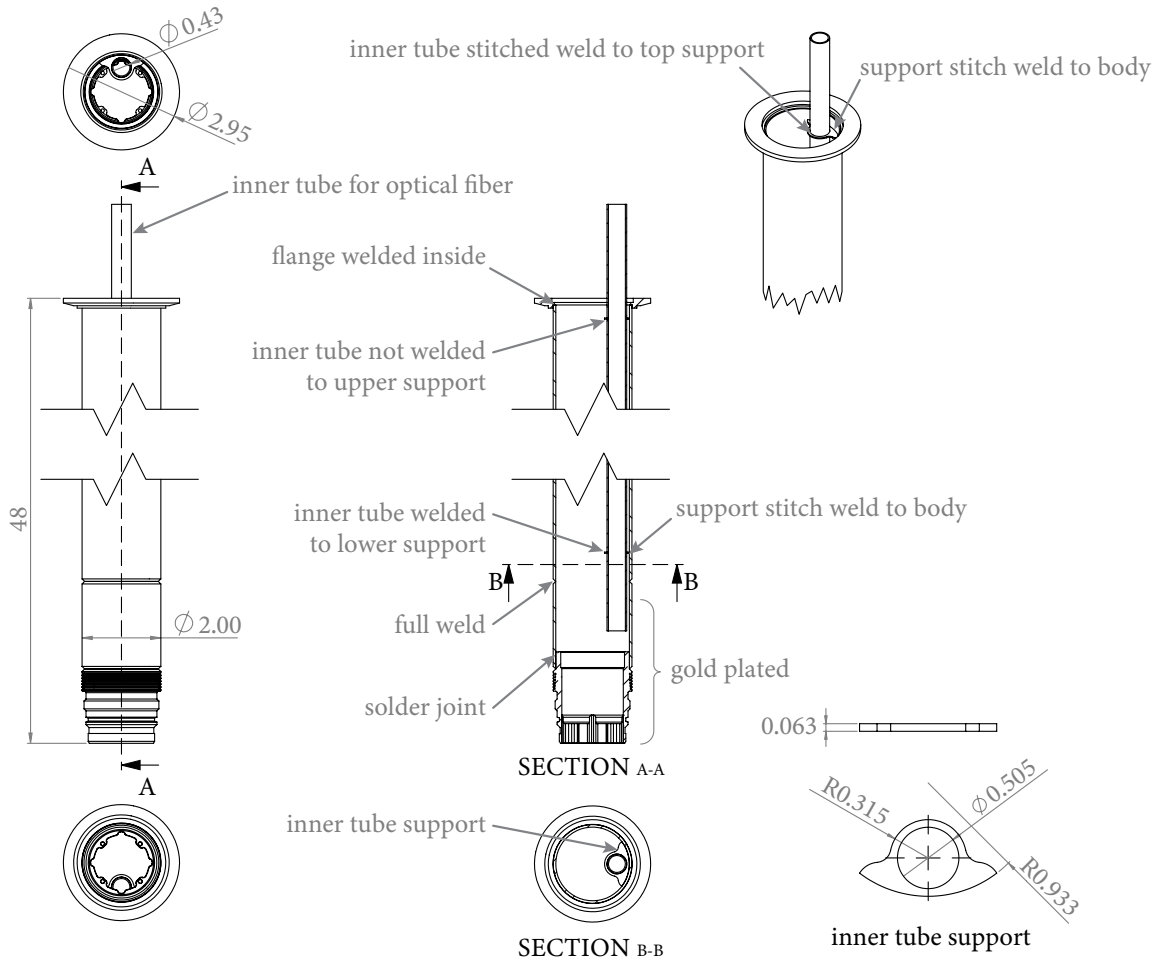

Figure 4-10: Details of the probe body. The probe body is a weldment composed of a number of components. The details of the inner tube and its support are shown including the weld specifications. The vacuum flange is welded onto the main body with a single internal weld. After this, the inner tube and its supports are welded in. The OFHC cold end of the tube is soldered into a short section of the 316 stainless steel tube. This end is then plated in gold. After this step, the plating is removed from the weld area to avoid contamination. Finally, the cold end is welded to the remainder of the probe body, taking care to ensure alignment of the inner tube.

avoids gold contamination of the solder joint, which can cause embrittlement of some solder alloys [94].

There are an additional two parts that interface to the probe body, namely the depth stop and the dewar attachment – see figure 4-1. The depth stop is a simple flexure clamp that prevents the probe cap from hitting the bottom of the dewar. This depth stop was set around 1/2" from where the cap would hit the bottom of the dewar. The dewar attachment is a relatively simple design of a sliding o-ring seal and 2 1/2" sanitary flange. The sanitary flange is clamped and sealed to the neck of the dewar when the probe is in use. A threaded cap is used to press the o-ring against the probe body to main a seal which keeps the helium from escaping the dewar. This cap can be loosened to raise and lower the probe within the dewar. The threaded cap, while allowing the o-ring preload to be adjusted, has only a relatively weak grip on the probe. This allows the pressure in the dewar to lift the probe out of the dewar in the case of some kind of failure in the pressure relief system.

4.4 Manifold

The manifold is the major room-temperature portion of the probe. It houses all the feedthroughs and ports for accessing the probe when in use. A drawing of the manifold is shown in figure 4-11 with all the notable sections highlighted. The major parts of the probe are the main body (comprising of the center section and two end plates), the optical feedthrough, the electrically isolated RF feedthroughs, and the DC feedthrough. Each of these elements of the manifold are covered in the following sections.

The design of the manifold was driven by a number of requirements. First, the manifold must allow for easy access to all RF, DC and fiber lines as well as the pump-down port. It must seal to all the required elements, maintaining the vacuum within the probe. It must also be robust whilst also being compact. Finally, it must be lightweight as a heavy construction would make the probe difficult to handle.

Here we elected to prioritize the east of accessing the RF and optical connectors.

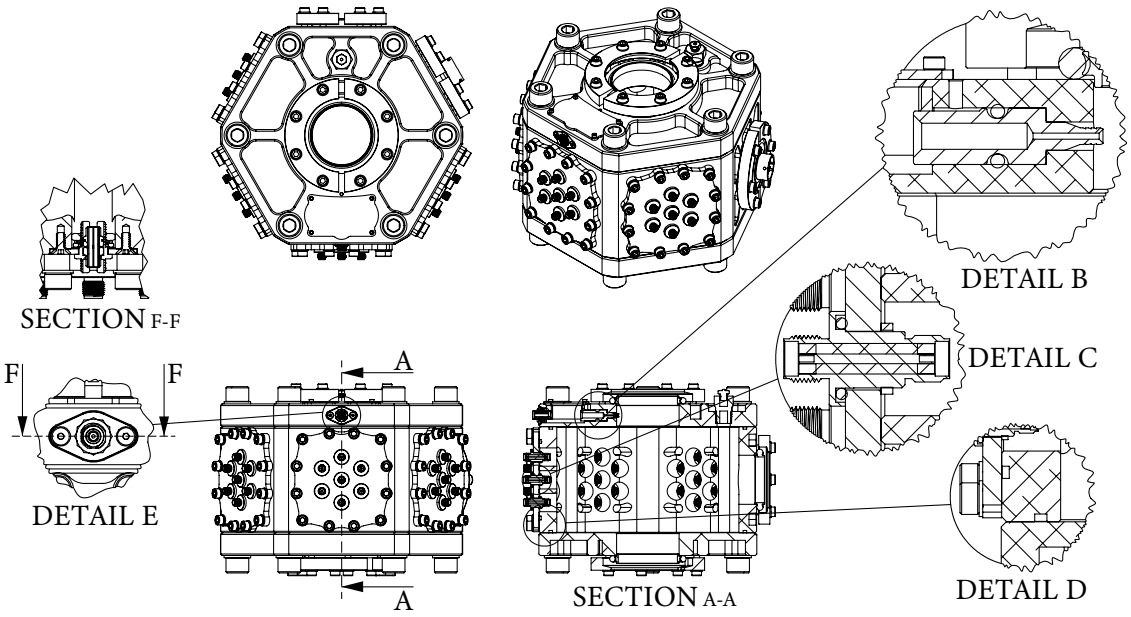

Figure 4-11: Drawing of the manifold with the probe portion removed. The key design elements of the manifold are called out in a number of details. Detail B shows the fiber feedthrough. Detail C shows the electrically isolated SMA connector feedthroughs. Detail D shows the o-ring glands designed to seal the manifold ends and RF feedthrough bulkheads. Detail E with section F-F shows the fiber mating sleeve and its mounting to the manifold top plate. Each of these elements is covered in detail later in this section.

With past cryostats the difficulty in accessing RF feedthroughs has lead to a number of issues. Thus, we chose to space the SMA connectors as far apart as possible within the given space on the manifold.

4.4.1 Main body

The body of the manifold is machined from three pieces of 6061 aluminum billet – see figure 4-12. The primary reason for this material selection was its high strength to weight ratio and high machinability. Given that the main body of the manifold is mostly hollow (see figure 4-13) it would be preferential to machine it from tubing; however, at the time of manufacture only round billet was available. This part, while being relatively simple, has the most setups in its machining. The part has features on each of its eight faces requiring eight mill setups plus one lathe setup for turning the outside diameter. Of these mill setups only three are unique. Thus, to save time

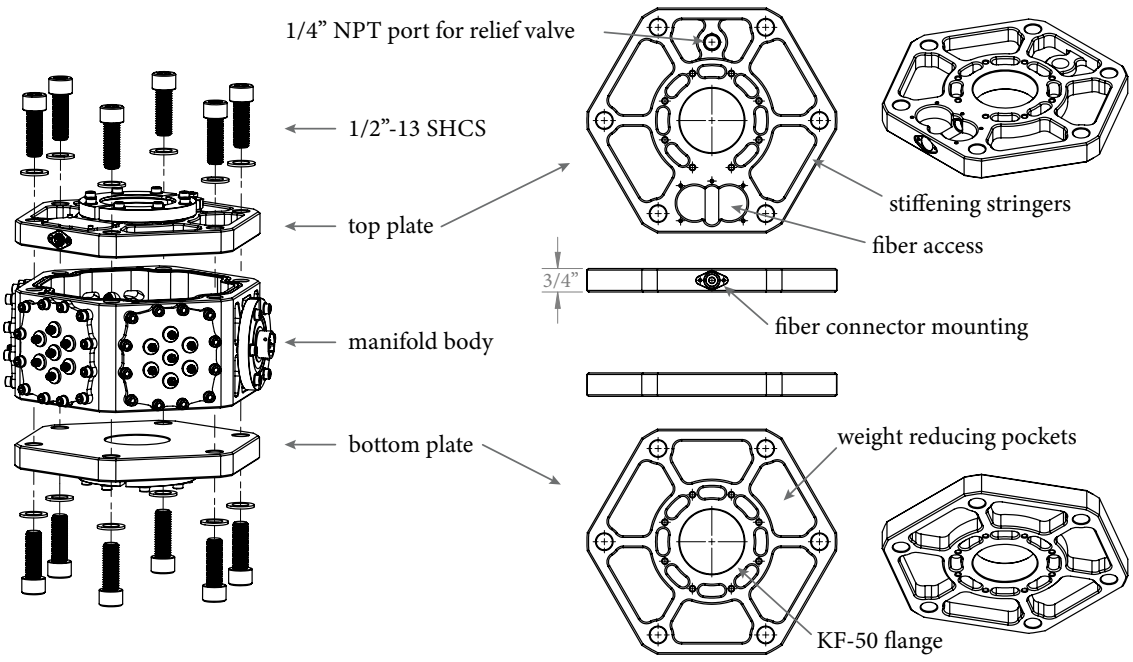

Figure 4-12: Exploded view of the three main parts that make up the mechanical structure of the manifold, namely the body and top and bottom plates. The details of the top and bottom plates are shown on the right. Note: the o-rings are not shown.

in machining, a fixture that allowed for easy indexing of the part was made. To locate the part on the fixture, the six $1/2" - 13$ threaded holes where used to mount the part, and dowel pin holes machined in to the body of the unit to allow for the location of the part in the fixture. The end plates, as shown in figure 4-12 are much simpler with only two setups for the bottom plate and three for the top plate. In terms of finishing the parts, all edges were chamfered in the machine and no hand finishing was performed other than the appropriate polishing of sealing faces. After machining all parts were solvent washed to remove oils from the machining processes.

As shown in figure 4-13, all sides of the manifold body have seal features on them, with six sides having o-ring glands. For a detailed description of the o-ring gland design see section 4.4.3.

To make the relatively large main body lighter we added weight reducing holes and pockets where possible. Stiffening stringers are run over the o-ring glands to reduce end plate deflection along sealing surfaces. From each bolt hole to the center of the part there is a stiffening stringer. These stringers both reduce deflection of the

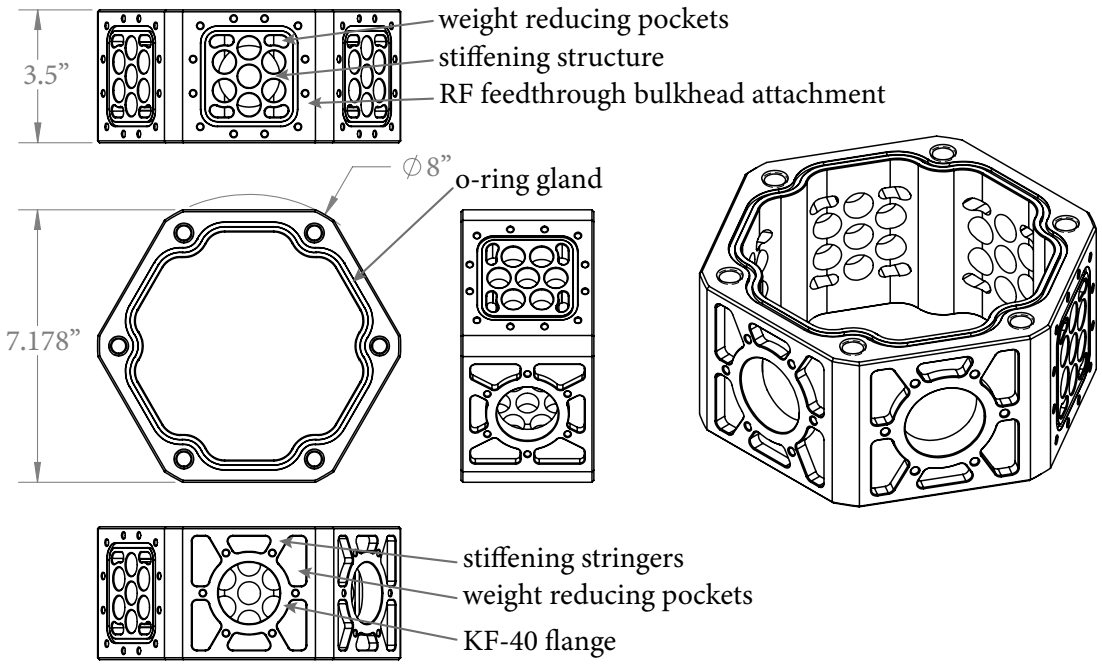

Figure 4-13: Drawing of the main body of the manifold. The details of the major features are called out.

part from vacuum forces as well as the forces applied to the probe body and manifold when the probe is inserted into and withdrawn from the dewar.

The top plate has a central KF-50 flange. In the probe design this port is unused and covered by a blanking plate. However, should there be a need for additional feedthroughs in the future, this provides a location for their mounting. The standard KF form-factor was selected as is easily adapted to without the need to machine any of the probe parts.

4.4.2 Fiber feedthrough

There are may commercial fiber feedthroughs available; however, most are expensive, limited in what type of fiber is used, and relatively large. To avoid these drawbacks, we developed our own simple feedthrough design as shown in figure 4-14. The seal between the fiber and the feedthrough is obtained by means of a low-outgassing epoxy (Torr-Seal). The seal between the feedthrough and the manifold is achieved by means of a FKM-based (Viton) o-ring.

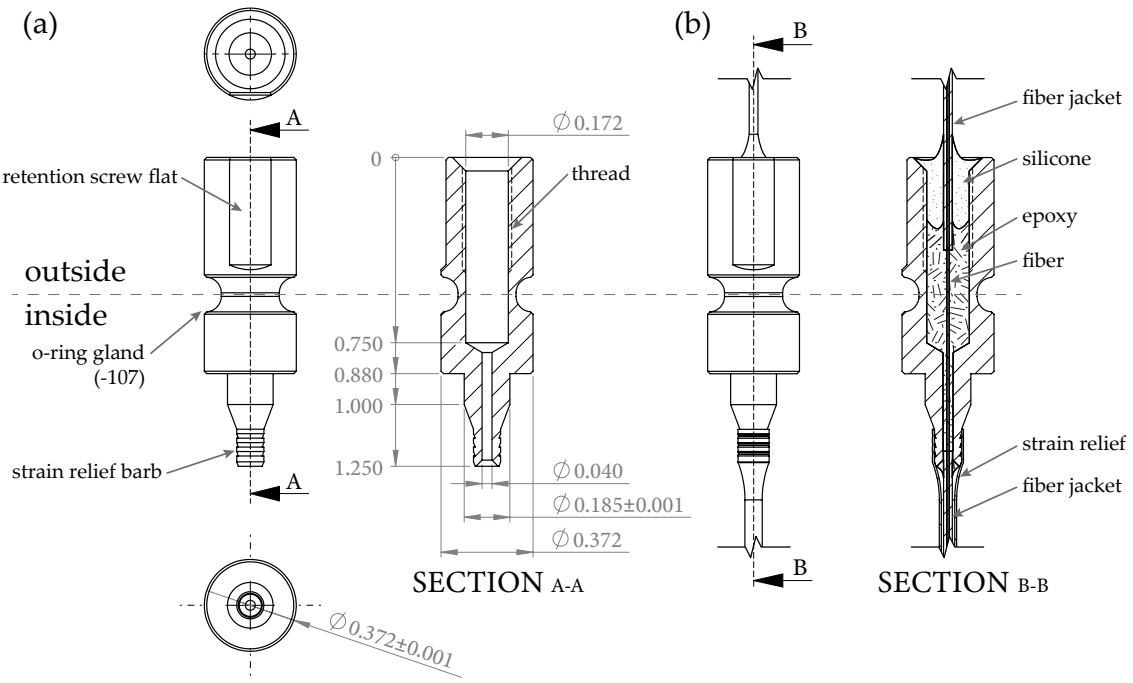

Figure 4-14: Drawing of the custom fiber feedthrough designed, machined, and assembled in-house. (a) The feedthrough body without fiber installed and showing the main dimensions of the unit. (b) The feedthrough fully assembled. Note the use of epoxy in order to achieve a gas-tight seal to the fiber and the use of silicone for strain-relief on the outside of the probe and a strain-relief tubing on the inside of the probe. A thread was machined into the inside of the feedthrough for the epoxy to key into.

Following Saint-Venant's principle, the body of the feedthrough is made almost three times longer than the diameter [95]. While this is less critical for an application like a feedthrough, the fact that the feedthrough is aluminum, as is the housing, and the fiber interface to the feedthrough is relatively sensitive, having the feedthrough jam while removing (which requires some force due to the friction from the o-ring compression) could result in damage to the fiber. Jamming of this interface, at least for the first stage of removal, is avoided by making the feedthrough body substantially longer than its diameter.

A setscrew is used to hold the feedthrough in place and prevent any movement with probe use that would translate to the fiber. A flat was milled onto the feedthrough upon which this set screw impinges. The purpose of this flat is to prevent the minor

marring that will occur from locking the feedthrough into the manifold and plate. As mentioned previously, the clearance between the feedthrough and the housing into which it is inserted is very small. The bur a set screw will produce when tightened against the feedthrough will likely be larger than this clearance, effectively preventing the feedthrough from being removed. By adding a flat region onto which the set screw can land, we need not be concerned about there being any difficulty in removing the feedthrough in the future.

With an epoxy seal we are left with an interface between a very flexible fiber and a very rigid epoxy. This leaves the fiber to take up all the strain when any force is applied to either part. To mitigate this to some extent we can insert the fiber jacket into the epoxy, thereby making the fiber slightly more rigid at the interface and encouraging a greater bend radius in the fiber. However, the jacketed fiber is still very flexible (especially the 900 μm jacket used here). To reduce the strain at the interface on the outside end of the fiber, a flexible silicone was cast around the fiber. This forces the fiber bend in a greater radius than it would have with only the jacket set in the epoxy, thereby making the fiber more resilient to damage. Here, a neutral-cure silicon (DOWSIL 738) was used to prevent corrosion to the parts. On the inside of the probe, silicone was not used as non-"space grade" silicones are not rated for out-gassing rate. Instead, a flexible tubing of greater diameter than the fiber jacket was used. This tubing is pressed onto a barb machined into the end of the feedthrough. This jacket extends the length of fiber over which strain will occur, thereby increasing the bend radius for the same force. When selecting the length of the silicone plug, and the strain relief tubing, for a first-order estimate, we again turn to Saint-Venant's principle and select something around five or more times longer than the characteristic dimension so that localized forces are minimal. As the silicone is within the body of the feedthrough, it is somewhat more rigid than it would be outside of an enclosure. Additionally, as seen in figure 4-14, we make the transition between the silicone-reinforced region and the fiber jacket smooth by filleting of silicone around the fiber. These measures collectively allow us to handle, install, and remove the fiber feedthrough without damage to the fiber.

As we have the strain relief tubing and fiber jacket within the vacuum, we need to be a conscious of the fact that we are making a very long path for trapped gas to escape. To avoid this, we make the jacket and tube only long enough to protect it from where the fiber enters the manifold to where it will be able to bend at a large and safe radius.

Regardless of what steps we take to avoid strain on the fiber, in a lab environment fibers fixed to feedthroughs are always vulnerable. Thus, to harden against possible damage, we terminate the outside end of the fiber into a bulkhead machined into the body of the manifold. With this design, and under normal use, the fiber will see no force from users of the instrument.

4.4.3 RF feedthrough

In order to maintain the probe vacuum and star grounding, a vacuum-compatible and electrically insulating bulkhead is necessary. There are a number of materials that could be used for this application; here we selected G-10 for its vacuum compatibility,

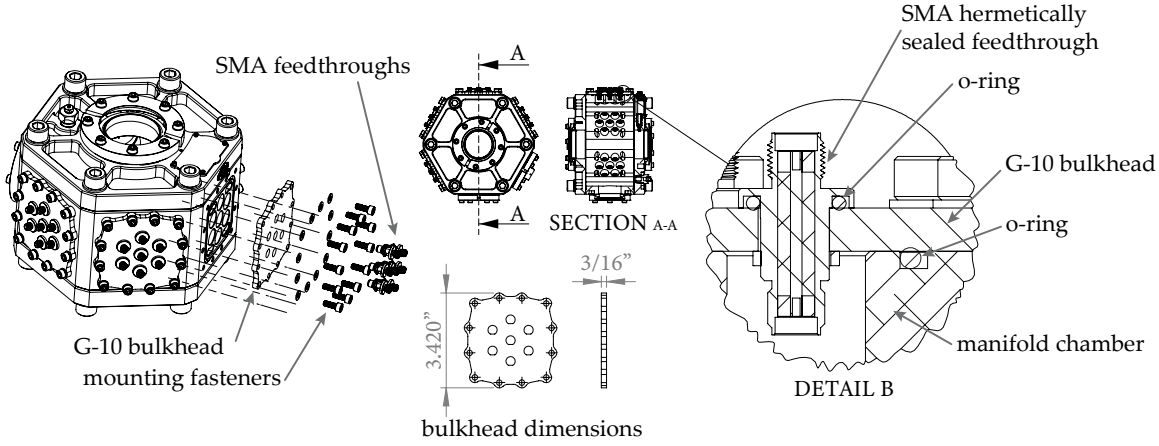

Figure 4-15: Details of the electrically insulated RF feedthroughs. The 3/16" thick G-10 bulkhead is mounted to the manifold by 12 screws spaced around the outside of the bulkhead. The bulkhead is sealed to the manifold by an o-ring (note that in the diagram the o-ring is shown in its uncompressed state). The bulkhead has seven D-shaped holes for the hermetically sealed SMA feedthroughs. Their D-shape prevents the connectors from loosening from the bulkhead in use. Each of these connectors is sealed to the bulkhead by means of an o-ring. The bulkhead is engraved with the bank letter and port numbers for reference.

strength, and excellent dielectric properties [42].

Proper o-ring gland (groove) design is critical for an o-ring to achieve a good seal [96, 97]. There are two main figures used when designing a gland, namely the compression ratio and fill ratio. The compression ratio is a measure of the amount of compression applied to the o-ring. For a face seal as we have here, the compression ratio is a ratio of the difference between the o-ring nominal diameter D_{nom} and the gland depth H_{g} divided by D_{nom}. The fill ratio is a measure of the ratio between the cross-sectional area of the o-ring A_{o} to the cross-sectional area of the gland A_{g}. The compression ratio must be sufficiently high to achieve a good seal while not damaging the o-ring, and the fill must be less than 100 % so there is enough space for the o-ring (and any swelling), but not so large that the o-ring might become dislodged. Additionally, for a face seal with vacuum on the inside, we should specify the perimeter of the inner surface to be slightly larger (around $S = 2\,\%$) than the inner diameter of the o-ring. This pre-loads the o-ring towards the inner surface and prevents movement with pressure cycling.

For the face seal we have designed here, we use a type –1XX o-ring with a nominal cross-sectional diameter of $D_{\text{nom}} = 0.103$". The gland depth was specified at $H_{\text{g}} = 0.075"^{+0.003}_{-0.000}$, giving us a compression ratio of $(D_{\text{nom}} - H_{\text{g}})/D_{\text{nom}} = 24\,\%$ to $27\,\%$. The gland width was specified at $W_{\text{g}} = 0.130"^{+0.003}_{-0.000}$ so with the cross-sectional area of the o-ring being $A_{\text{o}} = \pi H_{\text{nom}}^2/4$, and the gland cross-sectional area being $A_{\text{g}} = (1-S)D_{\text{g}}W_{\text{g}}$, we have a fill ratio of $A_{\text{o}}/A_{\text{g}} = 78\,\%$ to $84\,\%$. These two values are on the higher end of recommended compression and gland fill for conventional applications, but fall within the recommended range for vacuum applications at relatively constant temperature.

Another critical consideration when designing face seals is the bolt pattern. For a gasket with little energization, such as a copper flange gasket (such as CF flanges), it is recommended to overlap the pressure cone of each bolt/screw. This is why it is common to see bolt patterns consisting of many tens of fasteners on ultra-high vacuum systems. This is also why such flanges are so thick, as the thicker the flange the larger the pressure cone, the fewer bolts are required. In practice, a trade-off

between flange thickness and number of fasteners is struck. The same rule applies to all gaskets, although for more highly energized gaskets, such as o-rings, we can often get away with non-overlapping pressure cones. We are afforded this leeway by the relatively large compression ratio we used and the fact that we only elastically deform the o-ring. So, in areas where fastener pressure is low (outside the pressure cone), we have the o-ring pressing up against the two parts maintaining the seal. Of course, we can only push this effect so far – meaning we have to have sufficient fasteners so that we maintain pre-load on the seal.

To achieve suitable rigidity that we can maintain high o-ring compression ratio while not requiring too many fasteners, it is desired to make the G-10 bulkhead as thick as possible. However, commonly available hermetically sealed SMA connectors cannot be mounted onto a bulkhead thicker than 3/16". Thus, 3/16" thick G-10 material was used for the bulkheads. It should also be noted that for o-ring, as opposed to say, CF flanges, all the force from the fasteners (less the minor o-ring compression force) is going into pressing the two faces of the flange together.

Each of the four bulkheads corresponds to a bank, which are labeled from "A" to "D". The cables within a bank are all phase matched and connected to one side of the sample – see section 4.2 for more details.

4.4.4 DC feedthrough

The DC lines for the temperature control and magnet current are passed through a commercial feedthrough. This feedthrough is mounted to one of the sides of the manifold using bulkhead clamps and appropriate fasteners – see figure 4-16. The bore in the manifold was sized to be a close fit on the alignment ring to ensure the o-ring mated in the correct locations and did not become dislodged or pinched. The surface finish on the manifold around the bore is critical to sealing, and as such, was machined and then polished appropriately.

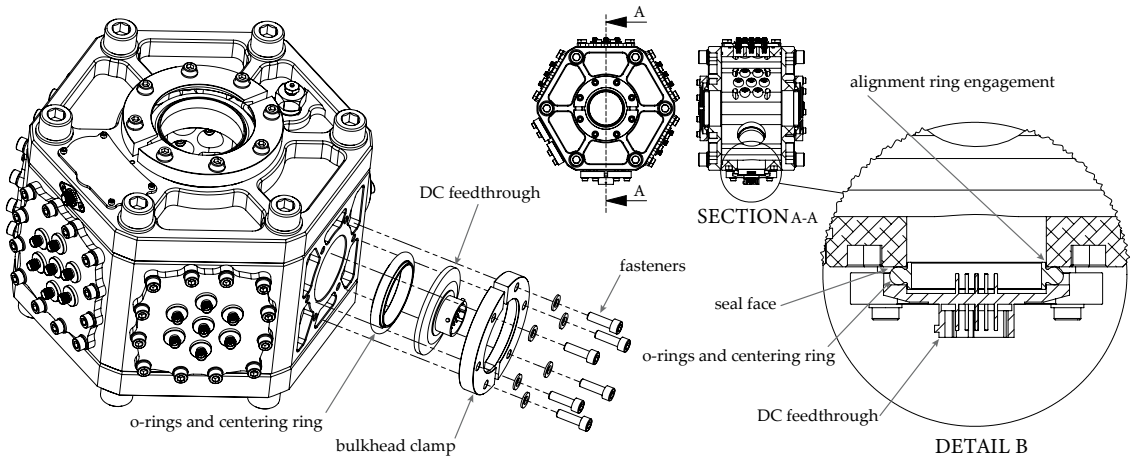

Figure 4-16: Exploded view of the DC feedthrough showing the clamping mechanism. The commercial feedthrough has a KF-40 flange which is mounted to the side of the manifold with an o-ring and alignment ring. The alignment with the manifold is achieved by an appropriately sized bore in the side of the manifold which engages with the alignment ring. The whole assembly is retained with a bulkhead clamping mechanism and fasteners.

4.5 Thermal design

Cooling is achieved by means of a heat sink that is exposed to liquid helium. This heat sink is then thermally coupled to the connector plate onto which the sample is mounted. A simplified model of the thermal circuit of the probe is shown in figure 4-17.

Ideally, the temperature of the chip would be measured; however, that would

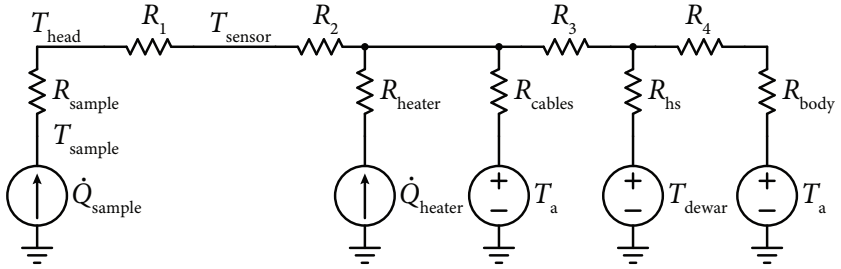

Figure 4-17: Schematic of the simplified thermal model for the cold head of the probe. $T_a = 300\,\mathrm{K}$ is the ambient room temperature, T_{dewar} is the temperature of the liquid helium or vapor (depending on if the probe is immersed in the liquid), and R_1 through R_4 represent the thermal resistances of the cold head and are very small values.

require a sensor on the sample. While an on-chip sensor is possible, for the vast majority of devices this would be rather inconvenient. The best alternative is a sensor on the cold head to which the sample is mounted. Here we embed a silicon diode temperature sensor within the cold head right behind the sample. This allows us to measure the temperature as close to the sample without requiring any special parts on the sample PCB.

In our thermal circuit, the thermal resistance between the sample and the cold head is governed by the user sample and their PCB design – see section 4.2.2. The thermal resistance R_1 between the cold head and the temperature sensor is trivially small. It represents the conduction through a roughly 1.5" diameter section of OFHC copper only 0.175" thick, thus we can say $R_1 \approx 0$. Along a similar argument, the material between the heater attachment point and the temperature sensor and cold head is very small and thick, thus we can again make the assumption that $R_2 \approx 0$. Now, as we provide the heater with a constant power from the controller, we can simply ignore the series resistor R_{heater} as it only plays a role in transient analysis.

The resistance R_3 represents the connection between the backside of the cold head and the heat sink fins on the probe. This section of the probe has a somewhat complex shape, and so a determination of this thermal resistance is hard to make exactly. One could perform finite element modeling; however, for our needs this is not necessary. We can simply model this section as a tube of diameter equal to the mean diameter of $D = 1.8$", a thickness of $t = 0.14$" and a length of $l = 0.8$". Thus, the resistance is around $R_3 \approx l/(\pi D t \kappa_{\text{Cu}}(T))$. If we assume OFHC copper of $RRR = 100$, then at around 4 K the thermal conductivity will be $\kappa(4\,\text{K}) = 642\,\text{W/Km}$ [98]. Under these assumptions $R_3 \approx 62\,\text{mK/W}$.

For the heat sink thermal resistance R_{hs} to the dewar or liquid helium (depending on if the probe is immersed in liquid or vapor only) temperature T_{dewar}, calculating this value is not trivial. Here we will assume it is relatively small as the heat sink has a large surface area and, when immersed in liquid helium, will present a very low thermal resistance. The section of copper between the heat sink and the body of the probe is short and made of much higher conductivity than the stainless steel

body, thus $R_4 \ll R_{\text{body}}$, where R_{body} is the resistance of the stainless steel body of the probe. As R_4 is in series with R_{body}, and $R_4 \ll R_{\text{body}}$, for our modeling we can set $R_4 \approx 0$.

This leaves the body of the probe R_{body} and the conductivity of the cables R_{cables}. For the probe body, we will consider the heat flux along the probe and convert this back to an equivalent resistance at the end. Now, our determination of the heat flux along the body of the probe is complicated by the fact that the thermal properties of materials tend to change, often drastically, as we approach liquid helium temperatures. We will encapsulate these effects in our temperature dependent thermal conductivity $\kappa(T)$.

For a constant cross section object with one end at T_1 and the other at T_2 and with a cross-sectional area A and length L, the heat flux along the object from T_2 to T_1 is given by

$$\dot{Q} = \frac{A}{L} \int_{T_1}^{T_2} \kappa(T)\, \mathrm{d}T, \tag{4.3}$$

where $\kappa(T)$ is the temperature dependent thermal conductivity of the material [42]. We perform this analysis for the probe body of diameter $D = 2$", thickness $t = 1/16$", and length $l = 45$". The thermal conductivity of 316 stainless steel $\kappa_{316}(T)$ at various values of T from 4 K to 300 K is presented in [98]. As this data is readily available, we will assume the cold end of the probe is at 4 K (close to our actual 4.2 K), and the warm end is at 300 K. To perform our calculation we numerically integrate along the length of the probe. The results of this analysis are shown in figure 4-18. From this analysis, the heat flux along the probe is $\dot{Q}_{\text{body}} = 672\,\text{mW}$. We can convert this to a mean thermal resistance of $R_{\text{body}} = 440\,\text{K/W}$. It should be noted that this ignores any cooling from the vapor in the dewar or the flow of gas to the recovery system. Thus, the heat flux will be lower in reality, and so the apparent value of R_{body} will be larger.

For the cables we perform a similar analysis. The micro coax cable is composed of a conductive copper core and shield with an insulator separating them and an undulating jacket. For our thermal analysis, we will ignore the thermal conduction

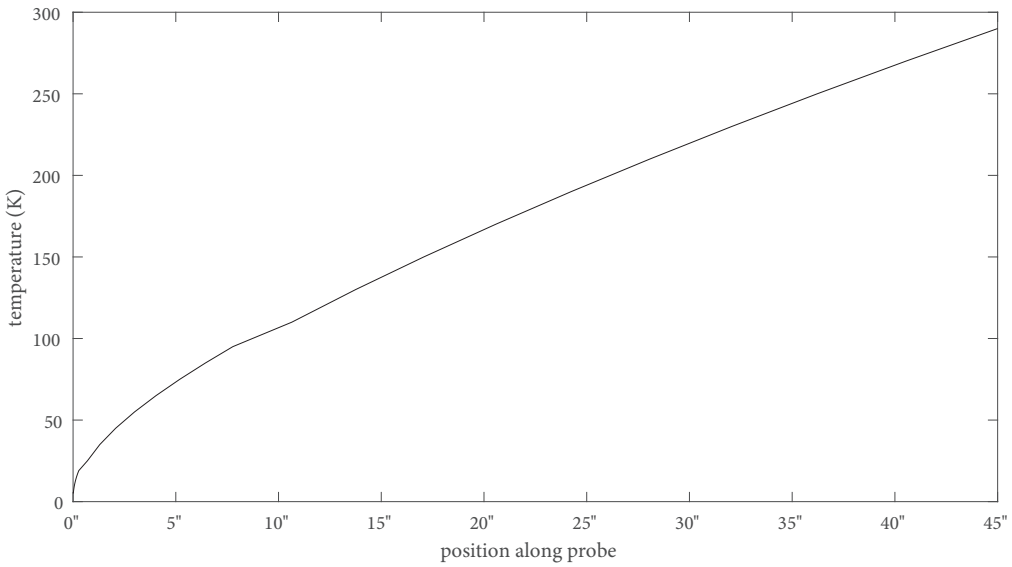

Figure 4-18: Plot of the temperature along the probe body assuming no convective cooling. The cold end of the probe is held at 4 K and the warm end at 300 K. It can be seen that the temperature drop increases as we reach the cold end of the probe due to the decreasing thermal conductivity of the 316 stainless steel at low temperatures. The values for temperature dependent thermal conductivity of the material were taken from [98].

along the insulators as they are much lower than the conduction along the conductors. With the 23 AWG micro coax cable we have that for the 1.5 m long cables $\dot{Q}_{\text{cables}} = 1.48\,\text{W}$, which gives $R_{\text{cables}} = 200\,\text{K/W}$.

With the heater off, and with the power dissipated by the sample being small, the combination of the cables and the thermal resistance R_3, the sample will be at most $\dot{Q}_{\text{cables}} R_3 = 87\,\text{mK}$ above the heat sink temperature T_{dewar}. So for immersion in liquid helium, the probe base temperature will be $T_{\text{base}} = 4.287\,\text{K}$. Thus, we meet our specification as set in section 1.4.2. Under these conditions, the total heat flux into the dewar is $\dot{Q}_{\text{total}} \approx \dot{Q}_{\text{body}} + \dot{Q}_{\text{cables}} \approx 2.15\,\text{W}$. In experiments, we find that with the dewar connected to the recovery system we obtain a base temperature of $T_{\text{base}} = 4.28 \pm 0.01\,\text{K}$, which agrees extremely well with our predicted value.

4.6 Cryogenic vacuum seal

Designing a seal to operate in liquid helium is difficult [42]. Sealing a vacuum from liquid helium is even more difficult. Here we need a vacuum seal at the cold end of the probe between the end of the probe and the probe cap. Generally, metallic seals such as copper or indium gaskets are preferred at low temperatures. The downside of designs with this style of seal is that they require high clamping force and hence space. Given that we are working with a 2" outside diameter of the can, and we need a reasonable area cold head, we require a very compact seal.

For a regular room-temperature application we would either use o-rings or polytetrafluoroethylene (PTFE) seals. The downside is that these seals lose their effectiveness at low temperatures. In particular, the seal materials lose elasticity and more importantly shrink at low temperatures. One method around this is to use spring-energized seals that combine materials with good seal performance such a PTFE with a spring that maintains the necessary preload on the seal [99]. This offers a very compact seal that performs well at low temperatures. The downside of this type of seal is that it is very sensitive to surface finish and contamination.

To seal the probe we used an unfilled PTFE seal with a helical elgiloy spring. The seal is shown in figure 4-19. The PTFE body of the seal is retained by a small step that ensures it remains in place during removal of the can. The minor diameter of the thread is slightly larger than the seal to prevent damage to the seal during installation or removal of the can. The seal journals are highly polished to ensure a good seal.

In figure 4-19 is can be seen that there is a 1 1/2" hex machined into the end of the probe cap. This is only used to remove the cap in the event that it were to get jammed, and is not to be used for tightening the cap onto the probe. Only hand tightening is needed for the type of seal used here.

It was found that even with this seal design, the probe would leak very slowly if the surfaces are not impeccably clean and relatively new seals are used. This leaves the user three options. If they desire their sample to be under vacuum with no helium

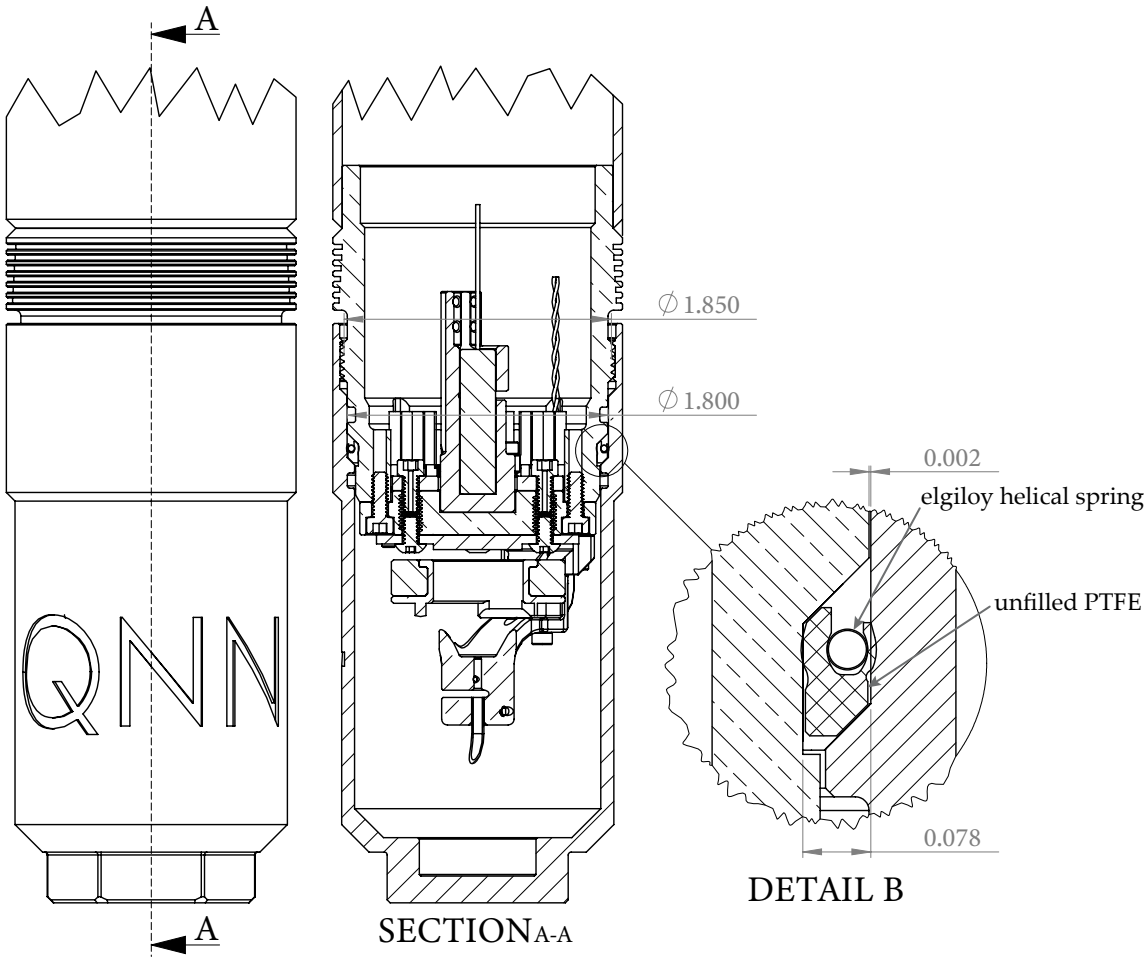

Figure 4-19: Cross section of the probe cold end with the probe cap installed. Detail B shows the seal cross section and its relation to the sealing surfaces. The seal is shown in the uncompressed state to show the amount of interference and hence preload in the seal.

gas present they can either replace the seal and ensure it is very clean prior to cool-down, or leave a slightly leaky seal and leave the probe attached to the vacuum pump during the experiment. Alternatively, if the sample can be allowed to be tested in a low pressure helium atmosphere the seal can be allowed to leak. The leaked helium has the added advantage of acting as a transfer gas and aiding in cooling of the sample and internal components of the probe.

4.7 Performance

Finally, we briefly cover some performance of metrics for the probe that have not been discussed in previous sections.

4.7.1 Helium usage

In order to assess the helium usage of the probe, we tracked the dewar weight before and after measurements. This data was captured for 20 measurement cycles which were conducted by a number of different users. Performing a least-squares regression fit we find that the helium usage is given by $V_{\mathrm{LHe}} = 0.73t + 0.72\,[\mathrm{L}]$, where the time t is in hours. From this we can estimate that it takes, at roughly the average insertion rate, 0.72 L to cool the probe to base temperatures. With these values, we can estimate that for a full 100 L dewar we can measure for a total of 122 hours and leave 10 L of helium to keep the dewar cold once the measurement is complete.

4.7.2 RF line phase matching

There are four banks of cables that connect to the device. Each bank has seven cables that are phase matched to within 10 ps. The cable manufacturer makes no guarantees of the phase match between batches of cables that are the same length. Thus, we had to characterize this. To do this we simply selected a channel from each bank and compared the phase response of each bank to obtain a value for the phase match between banks. This experiment was performed cold with the PCB attached and the

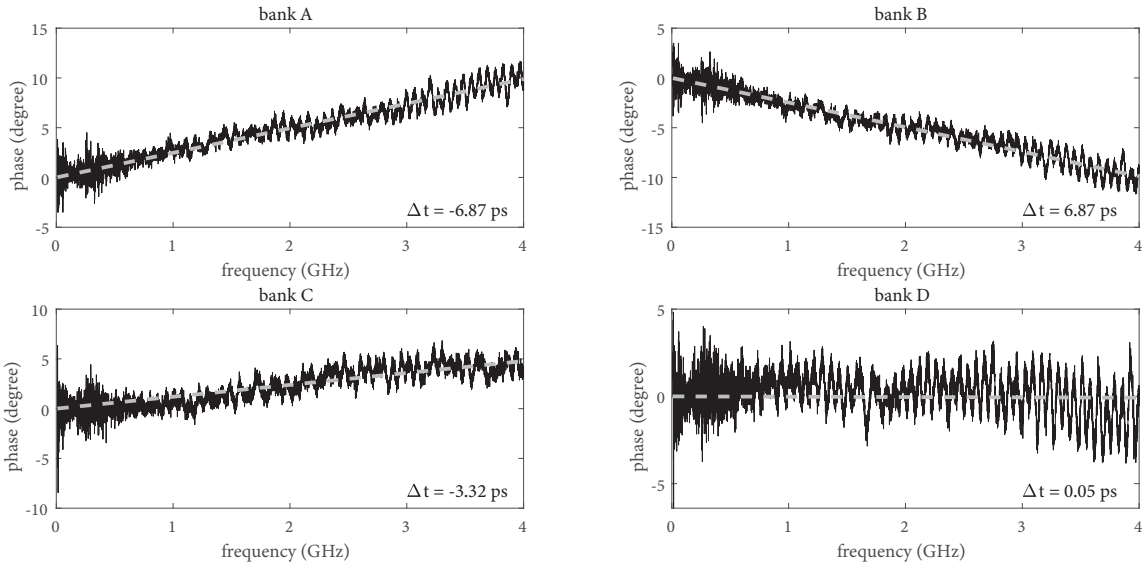

Figure 4-20: Plots of the phase response from a port in each bank of the probe. A linear fit is made for each response which allows us to estimate the time difference between each port's length. It can be seen that all the tests ports match with 7 ps of the median of the slowest and fastest port.

VNA test cables calibrated at the probe connector plane. The results are shown in figure 4-20.

4.7.3 Utility

The probe, owing to its versatility and ease of use, has replaced all other immersion testing with our group. The probe was used for nearly all experimental results that were presented in this work, some of which would have been difficult to obtain by other means. The greatest metric for the utility of any piece of equipment is how difficult it is to gain time on the equipment. Given that the author has had difficulty in obtaining time on the probe for upgrades and testing, it is safe to say that the utility of the probe, at least to our group, has been proven. Furthermore, we intend to fabricate a second probe due to how invaluable this apparatus has been to our work.

Chapter 5

Experimental apparatus

Following from the discussion of the CryoProbe, this chapter presents some details of the custom instrumentation required to support our experimental efforts.

5.1 Custom bit error ratio test set – the nanoBERT

With the memory performing well, we reached the limitations of our existing test setup rather rapidly. Specifically, we could not perform more than one hundred acquisitions per second with our existing setup. This meant that we could no longer gain an accurate estimate of the error ratio (BER), as such measurements would take far too long to be practical. Thus, there was a need for a new setup that could perform bit error ratio tests (BERT) rapidly while also allowing us to program custom read and write levels and waveforms. This was a problem we encountered when working on the original non-destructive read out (NDRO) nMem [4, 5]. To address this issue, we designed an FPGA-controlled test set that could read and write to a NDRO nMem array. However, this hardware was never used as we decided against further pursuing NDRO nMem arrays due to their shortcomings – see [5] for details. Instead of designing a new board, the original board was adapted to be compatible with the current nMem design in order to allow for us to perform the required experiments. This section briefly outlines the design of this test set, the nanoBERT.

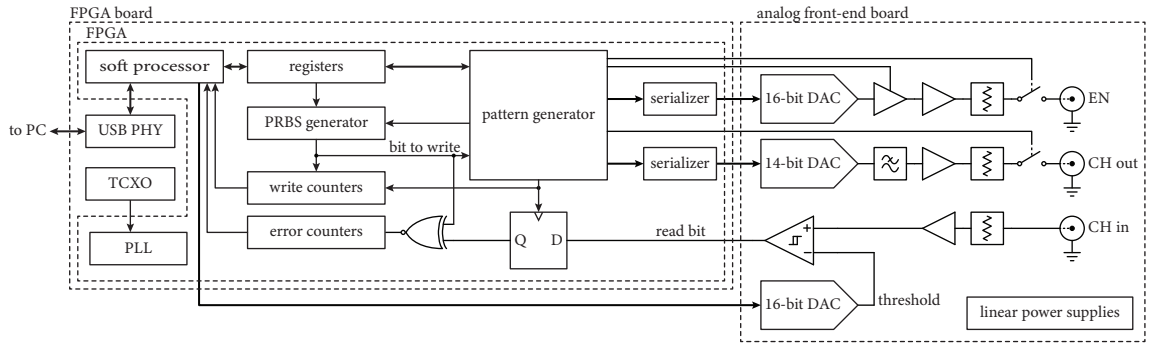

Figure 5-1: A simplified block diagram of the nanoBERT. The design is broken into two main sections: the digital section implemented on the FPGA, and the analog front-end section implemented on its own PCB. Note that in this figure, the power, reset, and clock lines are not shown. The whole test set is powered be a single floating supply, and communicated with a personal computer (PC) is implemented by means of a Universal Serial Bus (USB) interface.

5.1.1 System design

While the nMem is a digital device, we require high resolution control over the bias levels for the prototype cell. In a final application, and with appropriately mature fabrication processes, such control would not be necessary. Thus, the nanoBERT is a mixed-signal device, with digital control performed on an FPGA, and analog electronics to interface to the nMem. Given that the nanoBERT is only used for testing prototype cells, we wished to minimize the development time and cost associated with its design. For this reason, we opted to base the design on a commercial FPGA development board (the Terasic DE2-115) and focus our efforts on a custom analog front-end board specifically designed to interface to the nMem. This interface board generates the nMem programming, reading, and row-select waveforms, as well converts signals from the nMem to a form compatible with the FPGA. A simplified block diagram of the nanoBERT is shown in figure 5-1. The digital and analog sections of this block diagram are discussed in the following two sections.

The mechanical design of the nanoBERT is very simple. The mechanical assembly consists of the FPGA development board with a custom laser-cut acrylic cover upon which the interface board is mounted with standoffs. Mechanical stability of the test setup is required to ensure reliable long-term test results as it, amongst other factors,

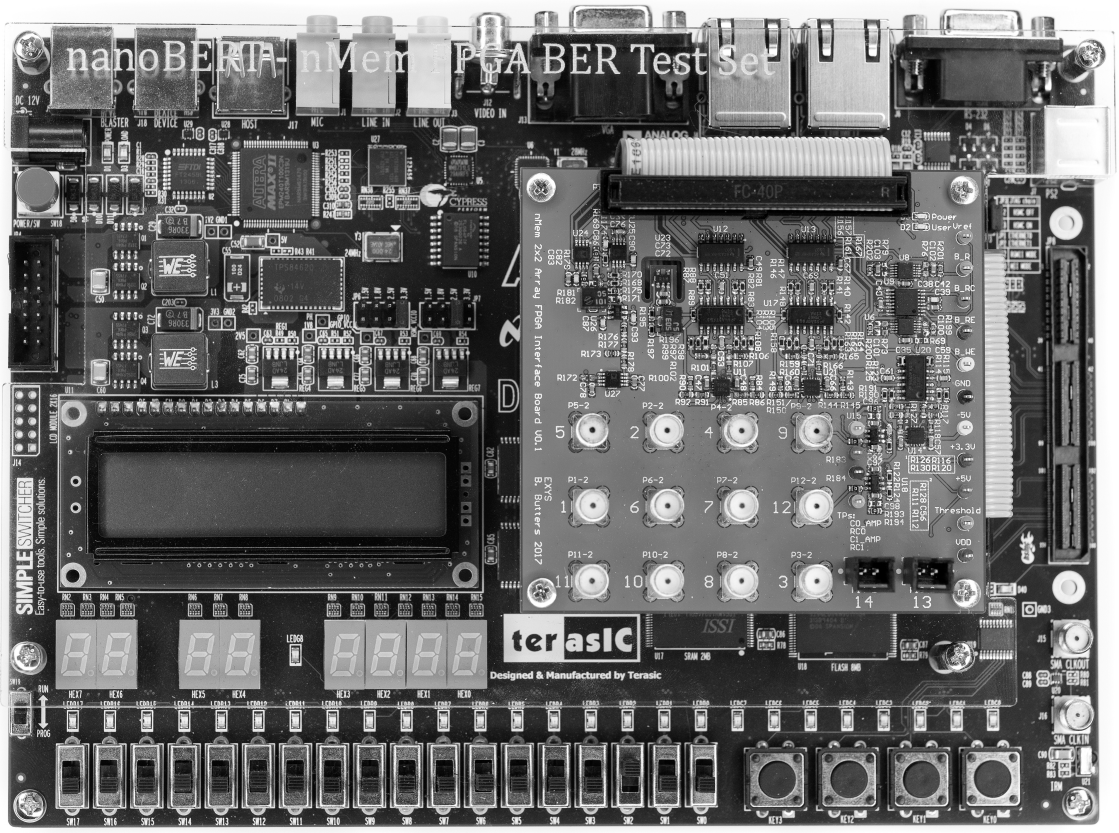

Figure 5-2: Photograph of the nanoBERT – the nMem FPGA BER test set. This set consists of a custom interface board and a DE2-115 FPGA board. The interface board is connected to the FPGA board by means of a 40-pin ribbon cable. There are a number of SMA connectors on the interface board, many of which can be attached to an nMem under test. In addition, there are test points along the right-hand edge of the board for monitoring signals internal to the interface board.

prevents the cables from flexing and the boards from moving. Mechanical stability is an often overlooked aspect of electrical test setups and is particularly important when high frequencies are involved. A photograph of the complete test set is shown in figure 5-2.

5.1.2 Digital design

The nanoBERT is controlled by a soft processor which acts as an interface between the device and the attached PC. The PC control is implemented through a USB interface realized by a USB PHY chip. The soft processor, under instruction from the PC,

sets the registers that control the BERT hardware, and reads back the counters. Additionally, since the threshold for the comparator need not change through the test, it is set by the soft processor before it initiates the BERT.

Central to the digital aspect of the design is the pattern generator. This block implements a state machine that controls nearly all aspects of the test. The state is controlled by a clock generated by the phase-locked loop (PLL), which is in turn driven by a temperature-controlled crystal oscillator (TCXO). The pattern generator is reset by the soft processor. The parameters from the register set control the operation of the pattern generator. The output levels as well as the timings of the waveforms are all parameterized in the registers. The pattern generator writes back to the register set when the BERT is complete.

The pattern generator takes a pseudo random bit from the PRBS generator and uses that to generate the sign of the output waveform. The PRBS generator is capable of implementing any linear-feedback shift register with polynomial up to 32-bits long. However, here we only used PRBS-7 with the polynomial $x^7 + x^6 + 1$. Thus, with this PRBS-7 algorithm, we will test every possible sequence of 7-bits written to and read from the memory (other than the all-zero sequence). While our PRBS generator can implement longer larger polynomials, they would not be of any benefit here (in fact a PRBS-7 sequence is longer than necessary).

With the required signals generated from the register parameters and the bit to be written from the PRBS generator, the pattern generator can create the required signals for the channel and enable lines. These signals are converted to serial data for the digital-to-analog converters (DACs) by means of a serializer. The serializer simply takes the parallel data out of the pattern generator, adds formatting, and generates a serial clock and data signal.

The pattern generator creates an additional clock signal. This signal increments the write counters that keep track of the number of "0" and "1" writes that have occurred. This signal also clocks the data read from the memory into a D-type flip-flop. This data is compared with the written data by means of an XNOR gate. When the data read does not match the data written, the appropriate error counter

is incremented. At the completion of the BERT, the soft processor reports back the counter values and resets the system.

5.1.3 Analog design

The analog front-end is powered by the FPGA board. This simplifies the design of the board; however, the switching regulators that supply the FPGA produce far too much noise for them to be used directly to power our analog electronics. Additionally, the FPGA board only provides positive biases, whereas we require a bipolar supply for the bipolar write biases. To generate a native supply rail, a switched-capacitor inverter was employed. The FPGA board supplies and the inverted bias are all linearly regulated to generate a set of low-noise voltage rails suitable for our analog electronics. In addition, a precision voltage reference is utilized to provide the reference voltage for the DACs. A board cutout is placed around three sides of the voltage reference to prevent it from being stressed due to thermal expansion and contraction of the board, but more importantly strain on the SMA connectors that lead to the CryoProbe. In all, the board generates $\pm 5\,\mathrm{V}$, $3.3\,\mathrm{V}$, and a precision reference voltage of $3.0\,\mathrm{V}$.

Two outputs from the nanoBERT lead to the prototype memory cell. One output provides the enable (EN) signal, and the other output provides the channel (CH) bias. The enable line level is generated by a 16-bit DAC. The output of the DAC feeds a switchable buffer amplifier. The buffer either provides a copy of the input when active, or provides a zero voltage output when deactivated. The output of the buffer feeds a unity gain amplifier that provides the output drive necessary to source to a $50\,\Omega$ load. The output feeds a π-pad which attenuates the signal to the low levels required by the memory. Finally, an analog switch isolates the nanoBERT from the load when the test is not active.

The channel output is driven in a similar manner to the enable signal, except for that the output must be bipolar and a fast DAC is used. Since the output must be inverted for writes, and different write and read levels are needed, we must update the output bias rapidly during the test. To achieve this operation, a lower precision 14-bit, but much faster DAC is used. The output of the DAC must be filtered to

remove undesired transients from code updates. From this point on, the signal path is the same as that for the enable line. An amplifier capable of driving 50 Ω loads buffers the signal. The signal is attenuated to the correct levels for the nMem, and an analog switch isolates the device when not being tested.

A high speed comparator is used to determine if the memory has switched or remains in the superconducting state. The comparator was designed with a small amount of hysteresis to prevent any possible noise issues. The inverting input is driven by a 16-bit DAC that sets the threshold voltage. The DAC is programmed by the soft processor before the start of the BERT. The signal from the memory channel is matched to 50 Ω and attenuated by a π-pad. The output of the attenuator drives an amplifier that then drives the non-inverting input of the comparator. The amplifier brings the memory channel signal to within the range of the DAC output. Centering in this range maximizes the resolution of the DAC-comparator system. Additionally, the amplifier matches the buffered output of the DAC, thereby feeding the comparator with the same impedance on both input, minimizing any possible offset due to the comparator input bias current.

5.1.4 Test algorithm

The nanoBERT can be configured for multiple tests including performing parameter sweeps. This is all achieved by configuring the soft processor as required. Within a single BERT, all the biases and thresholds are held constant. At the start of the BERT, the soft processor configures the threshold, waits for the output to settle, and then enables the output buffers. Once the initialization procedure is complete, the soft processor sets the BERT start flag to initiate the BERT procedure.

The FPGA implements a PRBS algorithm to generate the data to be written-to and later read-from the memory. Following a similar array emulation regime as discussed in section 2.3.1, we perform emulated array operation in between the write and read procedures. The waveforms used to perform the test are shown in figure 5-3. The result of this write is then compared to what we expect and the appropriate tally is updated (W0R1 error, W1R0 error, or no error). In addition, we keep track of the

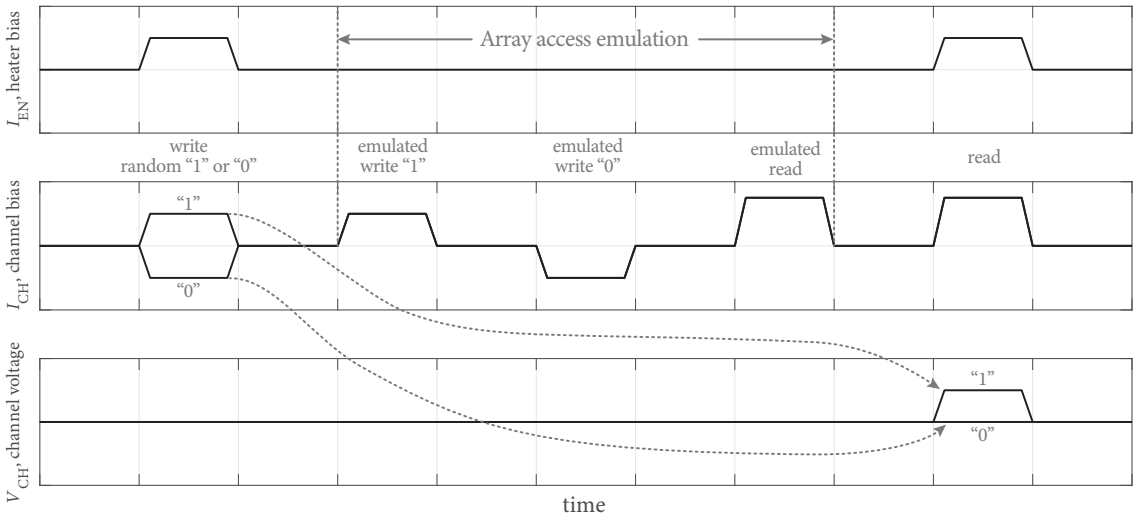

Figure 5-3: A sketch of the waveforms that the nanoBERT provides to the nMem when testing. A pseudo-random bit is written to the memory with the enable line active. The outputs are disabled and a cycle allowed to pass. We then perform the array emulation routine. During this routine the enable signal is not active and no change in the cell state should occur. The first emulated access is a "1" write, followed by a "0" write, and finally a read. After the array access emulation is complete, and the cell exposed to all biases it would need to pass in an array implementation, we read the state of the cell with the enable line active. If the memory cell performed as expected, we should read back the same data we wrote. So, when the memory is in the non-inverting regime, we expect that if a "0" were written, then we would see zero voltage on the channel during the read, and if a "1" were written, then we would see a voltage on the channel during the read.

total number of zeros and ones written to the memory such that we can determine W0R1 and W1R0 error ratios correctly.

At the completion of the BERT, the soft processor reads the tallies and uploads them to the attached computer. The soft processor then disables the outputs and resets the hardware. If the soft processor is configured for a sweep, the next operating point parameters is transferred to the control registers, and the next BERT is initiated. At the completion of the configured sweep the soft processor again disables all output and waits for further commands. The results of sweeps performed with this device are shown in section 2.4. By using the nanoBERT we were able to perform these large sweeps that would not have been possible previously.

5.2 A low noise motor driver – the infiniStep driver

Noise in cryostats is an ongoing issue and there are many approaches to solving this issue [100]. For pulse tube cryostats, even with remote motors, the noise from the motor switching supply is an issue. This is often improved by moving to commercial "linear motor drivers". When we encountered substantial noise in our pulse tube cooled cryostat we took similar steps. We sourced a linear motor driver from the pulse tube manufacturer, specifically the Cryomech Linear Driver model EPS002. The EPS002 is in fact a Precision Motion Controls model LN3 in a Cryomech enclosure. This is the same type of driver used in [100]. In comparison to the internal driver (the driver built into the compressor), the noise spectrum with the LN3 is different. However, the voltage provided to the motor terminals is not sinusoidal and exhibits substantial noise. In addition, the waveform form the output terminal to ground exhibited substantially more noise than the differential voltage across the motor. We found that with the LN3 driver, the noise coupled to our sample more readily than the noise from the internal driver. Thus, this driver in the configuration in which Cryomech provided it, is not suitable for our application.

An ideal driver for this application would provide a perfect sinusoidal drive to the motor. This would be free from any switching noise or crossover distortion. This driver would provide regulate the motor drive current to a specified level. In contrast to many other motor applications, this current need only be regulated on a relatively long time-scale. The valve which the motor drives is a relatively light load and operates at around 1 Hz while the motor is drive is operating at 35 Hz. So, we do not expect to need to drive much of a dynamic torque, and so the current drawn by the motor from a fixed voltage should be relatively constant. The only real need to regulate the current to the motor is, provided it can support a sufficient torque, is for thermal reasons.

The motor driver presented in this section is specifically designed for this application. It provides an exceptionally low distortion output voltage and performs current regulation on a long time-base. In contrast to most stepper drivers, which as their

name imply, are still operating in a stepping fashion, even with micro-stepping enabled, this driver is nearly continuous. Hence the name, "inifniStep" driver, as at the operating load, the output is practically indistinguishable from a perfect sine wave with infinite samples.

5.2.1 System design

From a high level, the motor driver simply needs to take mains power, a status signal from the cryocooler's compressor, and provide a low noise sinusoidal signal to the motor with sufficient power to drive the valve. This can be achieved by a number of means. The design we present here is influenced by a number of additional features we chose to incorporate based on our needs. In particular, we wished to be able to program the drive parameters digitally, modulate the output, have robust load protection circuitry, and perform soft starting and stopping of the motor.

The control system regulates the average motor current while providing a quadrature sinusoidal voltage drive to the motor windings. This ensures that the voltage across the motor is sinusoidal, while also ensuring the motor is driven with the necessary current for its load. To allow for the digital configuration of the drive parameters and advanced features such as spread spectrum modulation, we use a direct digital synthesis methodology. The output is encoded and sent to the DAC of each channel of the driver. The analog electronics amplify this signal and perform analog voltage regulation to ensure the load voltage matches the programmed voltage. In this regime we have maximum versatility in out output waveform control while also being able to perform constant-current control of the load.

A block diagram for the infiniStep driver is shown in figure 5-4. There are many aspects of this design that warrant in-depth explanation; however, they are beyond the scope of this work. Instead, we will only provide some details of the key aspects of the design. In reference to the block diagram, we can see that the design can be broken into three sections: the processor, the analog front-end, and the support electronics. The processor algorithms are discussed in section 5.2.5, and the analog front-end is discussed in section 5.2.4. The power supply portion of the support

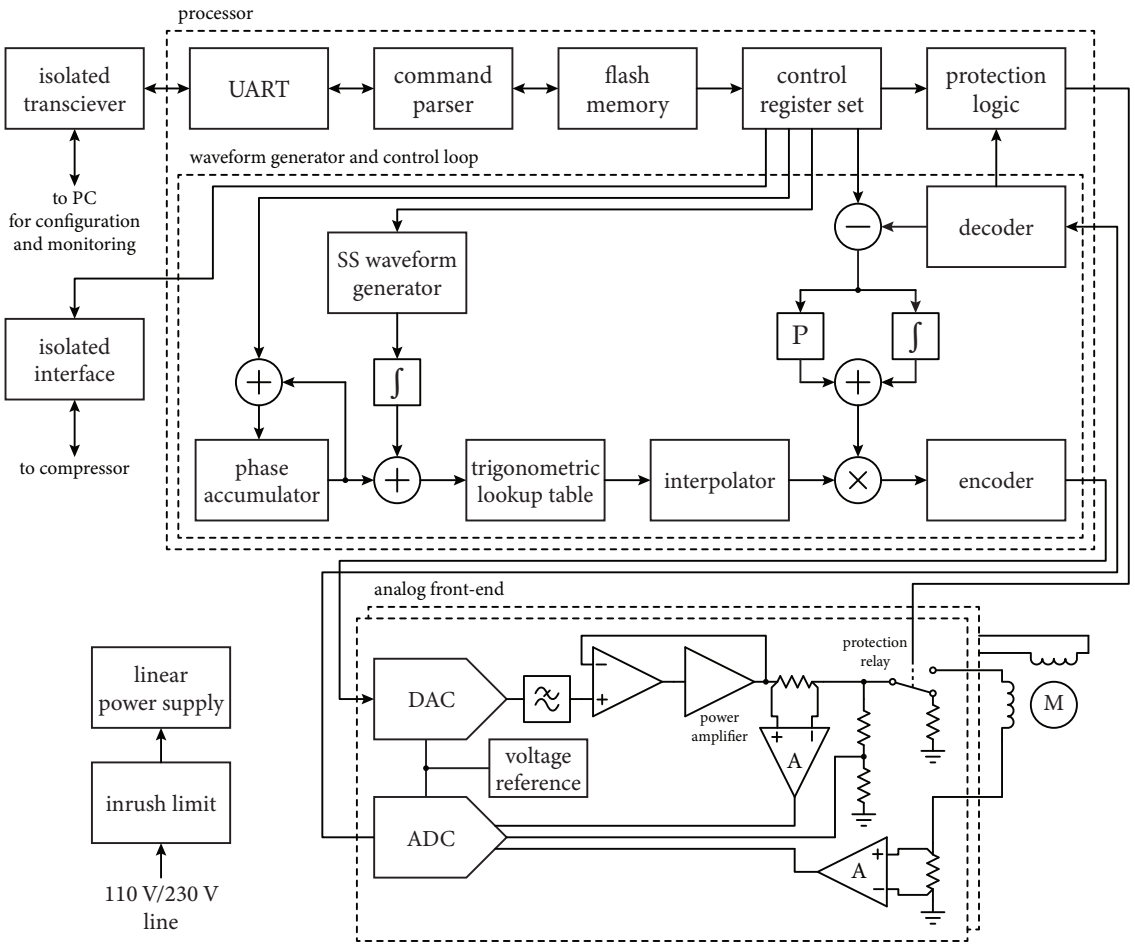

Figure 5-4: The block diagram of the infiniStep driver. The design is broken into three main sections. The processor implements all control and waveform generation as well as configuration of the unit. The analog front-end takes the digital waveform data from the processor and generates the requires signals and measures and digitizes the output voltage and supplied and returned current. Finally, the various support systems such as the isolated interfaces and the linear power supply.

electronics is covered in section 5.2.3. This leaves only the isolated transceiver and isolated interface.

The isolated transceiver is used to configure and monitor the unit through an RS232 interface that is galvanically isolated from the unit's power supply. The RS232 standard was selected to be consistent with the compressor's RS232 interface. Additionally, the simplicity of the standard make implementation and configuration simple. The transceiver is interfaces to a universal asynchronous receiver/transmitter (UART) integrated into the processor. The isolated interface consists of optically-isolated inputs to control the driver, and relay outputs to indicate the driver's status. There are two input signals, one to turn on and off the driver, and a second which can be configured to change driver parameters. There are two outputs, one indicates when the motor is energized, and a second which is active only when a fault occurs. With these interfaces, the infiniStep driver is transparent to the compressor, as when the compressor is switched on, the motor driver will detect this and begin driving the motor. In the event that the infiniStep driver detects a fault, it will disable the compressor to prevent damage.

5.2.2 Mechanical design

The unit was designed to be rack-mountable in order to fit in with existing equipment and make the device easy to integrate into existing laboratory environments. A 2U 19" enclosure was selected for the application. The internal layout of the unit is shown in figure 5-5. The transformer, being the heaviest internal component, is placed in the front of the unit. This location reduces the moment acting on the flanges when installed in a rack with only front supports.

All airflow is in through the front panel and passes through the unit to two variable speed fans down the heat sinks to the rear panel. Between the heat sinks and the rear panel are the emitter resistors which dissipate heat. A laser cut vulcanized fiber (fishpaper) box is places over and around the heat sinks and extends to the exhaust vents in the rear of the unit. This forces air to flow over the emitter resistors, keeping them cool, and forces the air driven by the fans, now warm from the heat sink, to

exit the unit and not recirculate within the unit.

5.2.3 Power supply

In order to eliminate the generation of switching noise on the power rails that would lead to noise in the output a linear power supply was used. The downside of such a supply is its weight and heat dissipation. However, since we area already dissipating substantial power in using linear amplifiers to drive a motor, the additional overhead from the power supply is not significant. The driver here is designed for a load of up to $\pm 10\,\mathrm{A}$ with an output voltage swing of $\pm 25\,\mathrm{V}$. Thus, substantial bulk capacitors were required in order to avoid the supply drooping too much between mains cycles. This lead to the need for inrush protection. The inrush protection is implemented with negative temperature coefficient thermistors. As the infiniStep driver can be powered from either a nominal $110\,\mathrm{V}$ or a $230\,\mathrm{V}$ mains supply, two thermistors are used for each primary winding on the transformer. Once the capacitors charge to near their operating voltage, relays short out the thermistors. The relays prevent a constant power dissipation in the thermistors leading to unnecessary losses or possible overheating of the thermistors.

While all the analog electronics and sensitive digital devices are supplied by linear regulators, there was a need for two switching regulators that are used less critical applications. The first switching regulator was used to generate the supply rails for the LCD and status LEDs. The reason for this decision was that the LEDs and the LCD back light require a relatively high current at a low voltage, which would lead to unnecessarily high losses if they were powered form the linear regulators. Additionally, the switching regulator is on the LCD and LED status PCB mounted to the front panel, which isolates the interference from the main board. The second switching regulator provides an isolated $15\,\mathrm{V}$ supply for the interface. A switching regulator was required for this application since generating an isolated supply by other means, such as a second mains transformer, would be needlessly expensive and the inclusion of this supply yields no appreciable increase in noise. Further alleviating noise concerns from this supply, a π-filter isolates switching noise from the linear

Figure 5-5: Photograph of the linear motor driver with the top cover removed. The unit is consists of a 2U 19" rack-mounted enclosure. The device status and configuration is shown by three LEDs and an LCD screen on the left of the font panel. On the right of the front panel is the air intake, and the exhaust is at the back of the unit. The power input module (power connector, fuse, and voltage selection switch) is mounted on the back panel. An isolated panel houses the serial port and interface port. Finally, a circular connector on the pack panel connects to the motor.

supply rails.

5.2.4 Analog front-end

The analog front-end performs data conversion and generates amplifies the signals form the DACs to drive the motor windings. The analog front-end is duplicated for each of the two output channels A and B for motor phases 1 and 2.

Power amplifier

A simplified schematic of the power amplifier portion of the infiniStep driver is shown in 5-6. The signal into the power amplifier is provided by the 16-bit DAC that is programmed over serial peripheral interface (SPI) by the processor. The output of the sense amplifiers and the output voltage are all captured by an ADC and sent to the processor. Both the ADC and DAC are provided with a precision reference voltage, which is shared between channels.

The output of the DAC is filtered through a Sallen-Key topology low-pass filter implemented using the operational amplifier (opamp) U2. As a unipolar DAC was used here we need to shift the result to obtain a bipolar output. This is achieved in the negative feedback path of the amplifier U2. The output of this filter is fed to the amplifier U1. R17 and C3 provide high-frequency feedback for the amplifier U1 to prevent instability in the larger power amplifier. The opamp U1 controls the input to the power amplifier stage and attempts to make the output of the amplifier (load voltage) equal to the output of the DAC filter. The output of the power stage is sampled through a resistor divider formed by R15 and R16. Sampling the divided output, in combination with the power stage accepting a current input referenced to near ground potential, allows for the use of a low-voltage opamp for U1. This reduces cost and greatly increases the flexibility when selecting an amplifier for this design. The amplifier feedback is fed to the non-inverting input of U1 as the power stage is itself inverting. The frequency response of the closed-loop system is shown in figure 5-7.

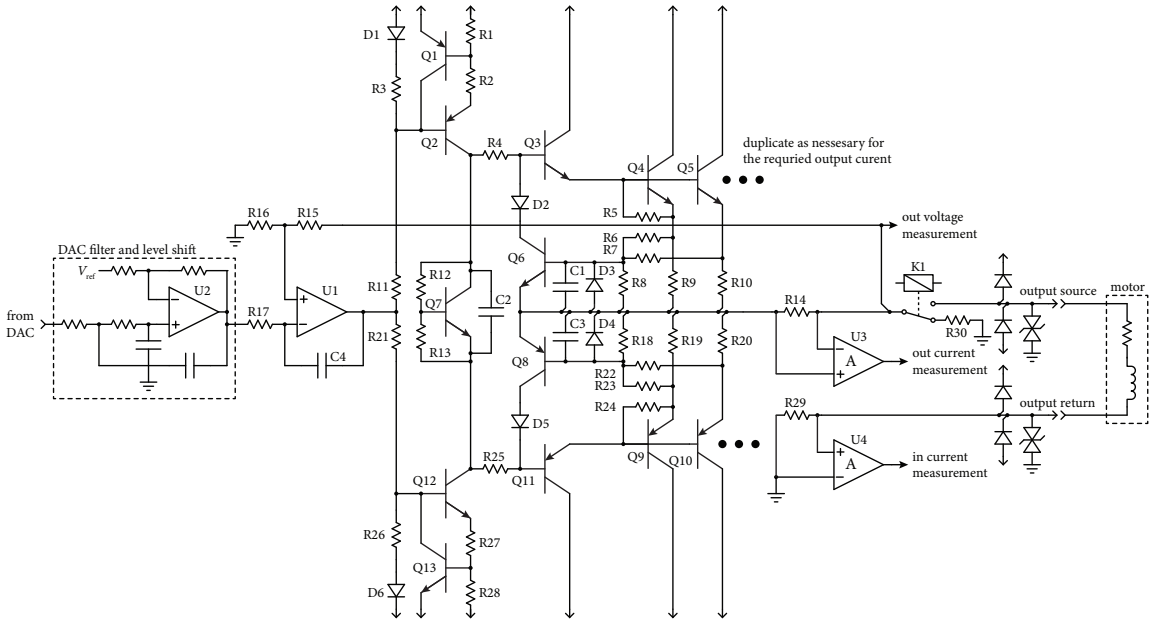

Figure 5-6: Simplified schematic of the power amplifier portion of the motor driver, along with the DAC output filter, output sensing, and protection electronics. This circuit is repeated twice within the infiniStep driver for the two channels.

The input of the power stage is a current from the amplifier U1. This current is driven into a resistor divider setup between the power rails (resistors R3, R11 and their complements R21 and R26), thereby setting up the offset for a pair of complementary current-limited common emitter amplifiers (Q2 and Q12). These amplifiers are current-limited by the additional transistors (Q1 and Q13) that pulls the common emitter towards lower currents when the main transistor emitter current is too high. These current-limited amplifiers drive the output class A-B stage which is set up as a Darlington configuration with one driver transistor (Q3 and its complement Q11) and up to four output transistors (only two are drawn with Q4 and Q5 on the high-side and Q9 and Q10 on the low-side) on each side of the complementary stage. The bias for the output stage is realized by means of a V_{BE}-multiplier (Q7). Current limiting is implemented in hardware by means of a pair of transistors (Q6 and Q8) that monitor the voltage dropped across the emitter resistors (R9, R10, R19, and R20) of the output transistors. If this voltage, after division (R6||R7 and R8, and their complements for the low-side), is sufficient to turn on the current limiting transistor,

then the output of the common emitter amplifier is sunk to the output potential, thus, beginning to turn-off the output transistors. The current limit in the common emitter amplifier is included such that the closed loop nature of the amplifier does not drive such high currents through this transistor or the current limit transistors that damage could occur.

As in all designs, we need to ensure thermal stability in the amplifier. Here this is accomplished by two features. First, the V_{BE}-multiplier transistor (Q7), the emitter follower diver transistors (Q3 and Q11), and the output transistors (Q4, Q5, Q9 and Q10) were all mounted on the same heat sink. Thus, we that the V_{BE}-multiplier transistor roughly compensates for the drops in the other devices. For the current-limited common emitter, we introduce a diode in the bias branch that ensures its operating point remains stable. Thus, under non-fault conditions, the amplifier should have a broadly stable operation over temperature.

As the motor outputs lead directly to the output transistors, we need some method of protecting the output and load. The load is isolated form the amplifier during power-on by the isolation relay K1. This has the added benefit of the normally closed path of the relay being used to connect the output to a known load R30, for calibration. The driver is protected from high voltage transients, such as static discharge, or form voltage transients due to the opening of the protection relay K1, by diodes to the supply rails and a transient voltage suppressors. These devices are present on both the supply and return connections to the motor. It should also be noted that the supply rails are protected from over-voltage conditions by additional circuity not shown here.

Feedback

On each output phase there are three measurements that are used for control and monitoring. The voltage supplied to the load, the current supplied to the load and the current returning to the unit are measured. Only the measured output current is used in the control of the load with the return current and the measured voltage used to monitor for malfunctions in either the driver or the load. Specifically, the

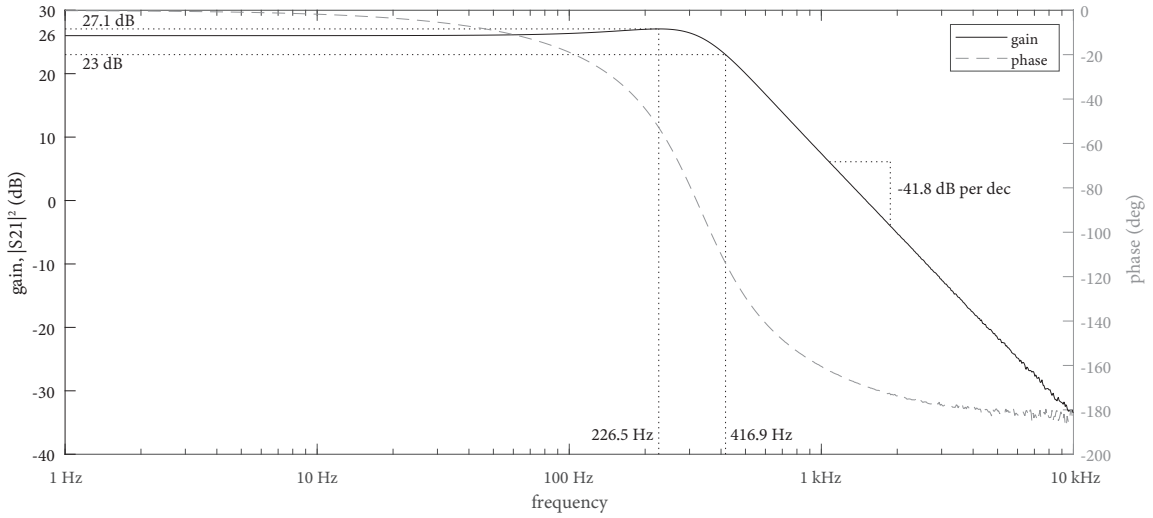

Figure 5-7: Response of the power amplifier including the DAC output filter. The 416.9 Hz cutoff is a result of the DAC output filter. The response is very flat with a small ripple before roll off. The response is near ideal, with only small errors due to part tolerances. The predicable gain fo the amplifier for the frequencies of interest is needed for good output voltage limiting in software.

output voltage is measured and compared to the compliance limit. In the event that the measured voltage is greater than the compliance (with some margin) the unit will throw a fault and the load protection relays will de-energize, thereby isolating the load. Thus, in the event that there is a issue with the analog output stage – for example a pass transistor fails short – then the load is protected. The return current is measured so as to be able to ensure that the load is connected properly and there is no substantial leakage. Specifically, the winding connections are ensured by measuring the load provided to a channel and the load returning from that channel. In the event that here is a short between a winding and ground or some other potential, the imbalance between provided current and return current will be detected and will result in a fault being thrown.

5.2.5 Software and interface

The main processor used in the infiniStep driver is a PIC32MK MCJ family device. This family was selected for its high performance analog peripherals, floating-point support, high clock rate, and low cost.

All timing sensitive operations are interrupt driven. Interrupt priorities are used to ensure that the timing-critical operations occur when needed and lower priority operations are performed when time permits. The highest priority interrupts deal with fatal faults. After the fault interrupts, the generation and processing of the waveforms and DAC update are given the next highest priority, thus guaranteeing that they occur when required. The lower level interrupts perform system monitoring and processing or interface operations and non-fatal faults. The main loop implements the user interface update routines (LCD updates) and serial interface processing.

A basic overview of the main interrupt service routine (ISR) is shown in algorithm 1. In this algorithm the logic that governs the detection of faults, their timing, and their relations to the output signal and protection relays are shown. The various aspects of the algorithm and in particular the waveform generation function is outlined in this section.

Control overview

The motor driver runs in a constant-controlled regime. Thus, the controller, which is implemented digitally, samples the output current and determines the required output voltage to achieve the programmed output current. A proportional-integral (PI) controller is used for this application.

The motor can be modeled simply as a resistor in series with a variable inductor. In order to form this model, number of assumptions are made. First, the mutual inductance between phases is ignored, the magnetic circuit is assumed to be linear (free from saturation), and all other parasitics (such as capacitance) are ignored. Using these simplifications, we can model the impedance of the motor winding (or phase) $\phi \in \mathbb{N}^+$ as

$$Z_{\mathrm{m},\phi} = R_\phi + L_\phi(\theta), \tag{5.1}$$

where R_ϕ is the winding resistance, and $L_\phi(\theta)$ is the winding inductance at the motor angle θ. We arbitrary chose the reference for the motor angle $\theta = 0$ to be when the first motor winding $\phi = 1$ is aligned (at its highest inductance). With this formulation

Algorithm 1 Interrupt algorithm.

```
1: global variables
2:     Fault                                          ▷ Flag indicating fault type
3:     State                                          ▷ Flag indicating system state
4: end global variables
5: procedure ISR
6:     debouncedOptoInput ← FALSE
7:     if fault ≠ NONE AND fault ≠ POWER_FAIL then
8:         Open protection relays
9:         Set DAC to 0 V
10:        Close fault relay
11:        Open run relay
12:        State.running ← FALSE
13:    else
14:        Close protection relays
15:        Open fault relay
16:        Close run relay
17:        if SPI transmit failure then
18:            Throw SPI fault
19:        else if ADC result not ready then
20:            Throw ADC fault
21:        else
22:            if State.running then
23:                Load code into DAC
24:                DACCodes ← WaveGen(debouncedOptoInput)
25:                Send DACCodes to DAC
26:                Start ADC conversion cycle
27:            else
28:                Open protection relays
29:                Open fault relay
30:                Open run relay
31:                Reset WaveGen parameters
32:                if debouncedOptoInput then
33:                    State.running ← TRUE
34:                end if
35:                Set DAC to 0 V
36:            end if
37:            debouncedOptoInput ← Debounced input from opto-isolator
38:        end if
39:    end if
40: end procedure
```

we can approximate the motor inductance as zeroth and first Fourier terms as $L_1(\theta) = L_0 + L_1\cos(N_\mathrm{r}\theta)$, where N_r is the number of teeth on the motor rotor, L_0 is the average motor inductance, and L_1 is the deviation from this inductance. For a two pole motor, such as that for which this controller is designed, $L_2(\theta) = L_0 + L_1\cos(N_\mathrm{r}\theta + \pi/2)$, for an arbitrary number of poles we have $L_\phi(\theta) = L_0 + L_1\cos(N_\mathrm{r}\theta + \pi(\phi - 1)/(N_\mathrm{p}))$, where N_p is the number of poles. Thus, we have that the current through the motor is not simply a linear relation to the output voltage. This means that out controller must either take these dynamics into account, or to remove them from the system. As the motor is slow and the load is constant the simplest means to address this issue is to filter the returned current waveform to eliminate the harmonics caused by the motor inductance variability. This method was used here.

With the given motor model, the motor current is

$$I_\phi(s) = \frac{V_\phi(s)}{Z_{\mathrm{m},\phi}(s)}, \tag{5.2}$$

where $I_\phi(s)$ and $V_\phi(s)$ are the current and voltage in the motor winding $\phi \in \{1, 2\}$. Thus, we can simplify this expression to $I_\phi(t) = YV_\phi(t)$, where $Y \in \mathbb{C}$ is the load admittance. Further, there is a 90° phase shift between $V_1(t)$ and $V_2(t)$. As Y is largely independent of the winding ϕ, we have the same phase shift between $I_1(t)$ and $I_2(t)$. Thus, to simplify the control algorithm, we instead control the current magnitude $|I_\phi(t)|$. However, estimating the current magnitude at any instance requires knowledge of the past, and hence a model of the load. To avoid this, we further simply by using the known phase difference between the two channels to give an estimate of the current as

$$I = \sqrt{I_1^2(t) + I_2^2(t)}. \tag{5.3}$$

In practice, the measurement of the output current must be completed upon every cycle of the control loop. The square root algorithm is a relatively intensive task. Thus, it is desirable to avoid any square root evaluations if possible. In order to achieve this, we instead control the value I^2 rather than I, thereby avoiding the square root evaluation.

The output voltage is $V_\phi(t) = A_c \cos(f(t)t \pm \pi(\phi - 1)/2)$, with the sign of the last term depending on the direction of motor rotation. In our PI controller we vary A_c in order to achieve $I^2 \rightarrow I^2_{prog}$, where I_{prog} is the programmed output current. Since there is no need for a fast controller as the loads are, over short timescales, constant, we can implement a relatively slow controller, which has the advantage of resulting in relatively slow transitions in the output current. It should also be noted that in order to prevent the load voltage from growing too large or clip, we perform a limiting function on the programmed voltage. This is simply achieved by setting A_c to be the minimum of a compliance voltage and the output of the PI controller. In order to allow for the controller to exit voltage-limited mode in a reasonable time, the integrator is adjusted when in voltage-limit mode.

Waveform generation and spread spectrum modulation

As previously discussed, the controller performs control evaluations at discrete time intervals. The waveform generation routine is called at a rate of 10 kHz. The role of the waveform generation function is to, based on the state of the output, generate the required DAC codes. The function, as summarized in figure 5-4, consists of a simple phase accumulator that increments a phase term with every cycle. When enabled, an additional spread spectrum term is also added.

With the motor running and no spread spectrum enabled, the accumulator simply adds a phase increment $\delta\theta$ to an accumulator θ every cycle. To prevent overflow, we evaluate at every cycle $\theta \rightarrow \mathrm{mod}_{2\pi}(\theta + \delta\theta)$. The phase accumulator value is then passed to the trigonometric functions (covered later in this section) to generate a $\sin(\theta)$ and $\cos(\theta)$, which are scaled by A_c, converted to DAC codes and returned by the function.

The basic phase accumulator approach as described above works well once the motor is running, but does not work well for starting the motor. A soft start system is necessary for this to ensure the motor starts smoothly. In addition, the motor frequency can be changed while the motor is running. To make these speed changes smooth, we need some method of varying $\delta\theta$ with time. We simply implement an

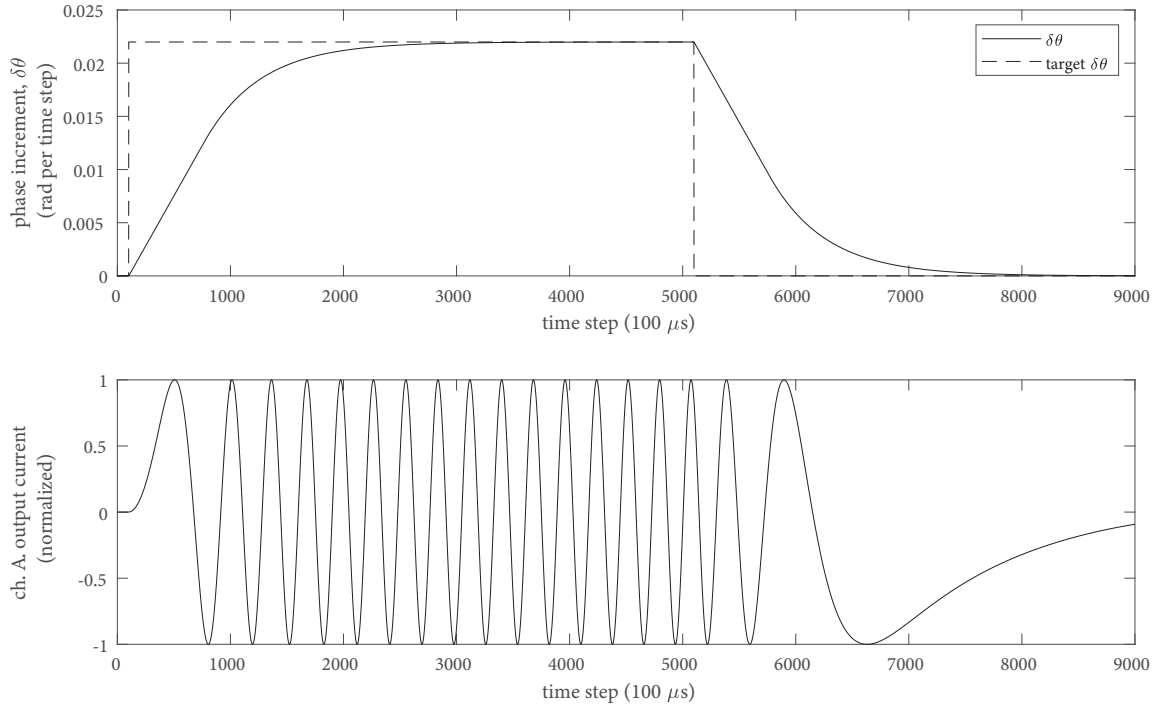

Figure 5-8: Simulation of motor driver with the output speed step response. The target $\delta\theta$ is stepped from zero to the desired increment, and then back to zero. This causes the waveform generator to vary the $\delta\theta$ in order to match the target value. The soft start and stop behavior of the motor can be seen, as well as the rate limit on the rising and falling edge of the phase increment response. Only one output of the two channels of the motor drive is shown. The output amplitude is normalized to remove the PI-loop's influence on the output signal.

algorithm the performs a rate-limited low-pass function to the value of $\delta\theta(t)$. This is realized by a simple infinite impulse response filter with the output change limited on each cycle. The result of this process is the soft start and stop behavior shown in figure 5-8.

The infiniStep driver is capable of spread spectrum modulation to reduce the power at any single frequency [101]. This reduces interference that could occur due to the fundamental and any harmonics. The spread spectrum modulation is implemented as a triangle frequency modulation on the driver frequency. Again we must complete these computations in minimal time such that we can complete the calculations in time for the next waveform point. The output frequency is

$$f(t) = f_{\mathrm{c}} + f_{\Delta}\Lambda(f_{\mathrm{m}}t), \tag{5.4}$$

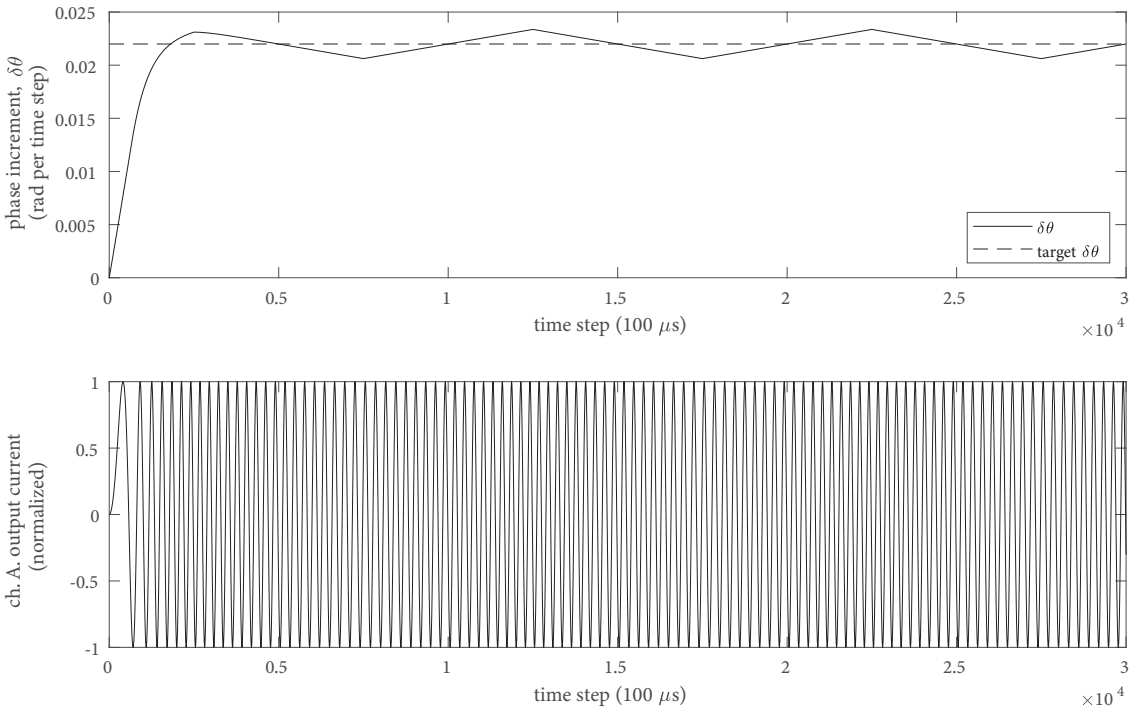

Figure 5-9: Simulation of motor driver with spread spectrum modulation enabled. Here an exaggerated modulation of 25 % at a frequency of 1 Hz is shown. The values are exaggerated in this plot for clarity as at the practical values, the deviation is difficult to visualize.

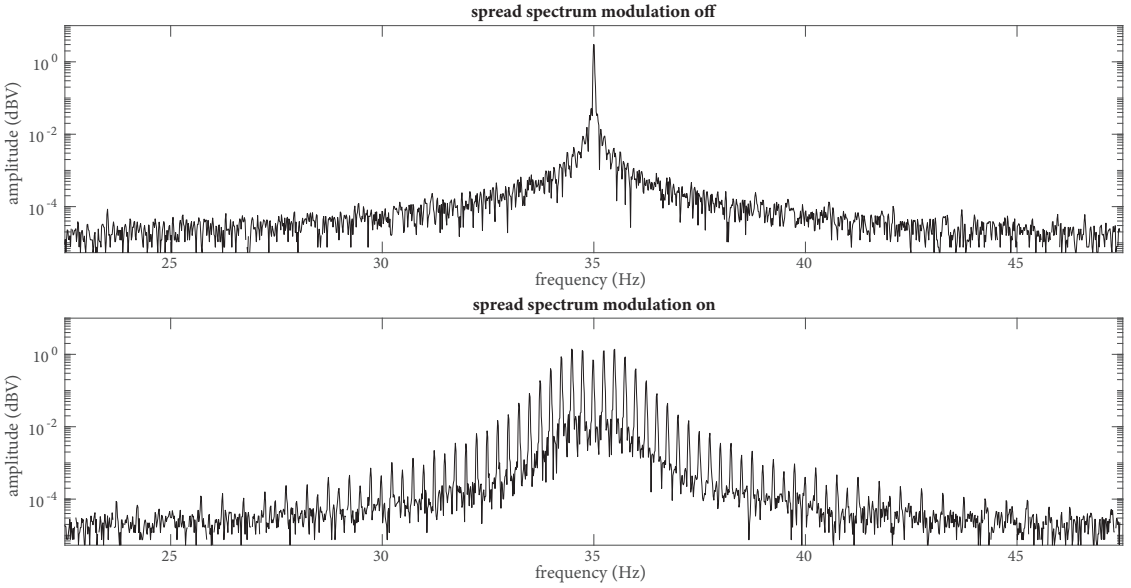

Figure 5-10: Measured output spectrum of the infiniStep driver with spread spectrum modulation disabled (top), and enabled (bottom). It can be seen that with spread spectrum modulation enabled the maximum amplitude of the signal decreases. A modulation of 2.5 % and frequency of 250 mHz was used.

where t is time, f_c is the center frequency, f_m is the modulation frequency, f_Δ is the peak modulation deviation, and $\Lambda(x) = 1 - 2\,|1 - (x + 0.5 \bmod 2)|$. To find our output signal $y(t)$ from this, we simply use

$$\begin{aligned} y(t) &= A_\mathrm{c}\cos\left(2\pi \int_0^t f(\tau)\,\mathrm{d}\tau\right) \\ &= A_\mathrm{c}\cos\left(2\pi f_c t + 2\pi f_\Delta \int_0^t \Lambda(f_m \tau)\,\mathrm{d}\tau\right), \end{aligned} \tag{5.5}$$

where A_c is our peak amplitude. From this equation, the control logic shown in figure 5-4, the infiniStep driver block diagram, is readily apparent. A simulation of the controller showing the spread spectrum modulation for exaggerated parameters is shown in figure 5-9. For our application, we find a percentage modulation, defined as f_Δ/f_c of 2.5 %, and a modulation frequency of $f_\mathrm{m} = 250\,\mathrm{mHz}$ works well. The output spectra with spread spectrum disabled and enabled are shown in figure 5-10 as measured on the real hardware. It is apparent in this figure that the modulation is effective at spreading the signal power over a larger bandwidth while also not interfering with the motor operation.

Trigonometric function implementation

The unit must generate a sin and cos waveforms with high fidelity to not introduce distortion in the output waveform. Given that the frequency of the output can be continually varied from DC to the desired upper frequency (around 400 Hz with the filters used in the prototype unit), these functions must be calculated on the fly and at high speed. A lookup table approach with interpolation was selected for this application. This approach was selected as it provides a good speed at the cost of either accuracy or program memory. Given the relatively large flash available on this processor, and the comparatively small application implemented here, there was a surplus of memory which allows for a large lookup table to be implemented.

A simple implementation of a lookup table could be implemented by simply evaluating a number of points of each function. This evaluation would give us the discrete cosine function $\cos_\mathrm{d}[n] = \cos(2\pi n/N)$, where $n \in \mathbb{N}^0$ and $0 \leq n < N$ is the index and

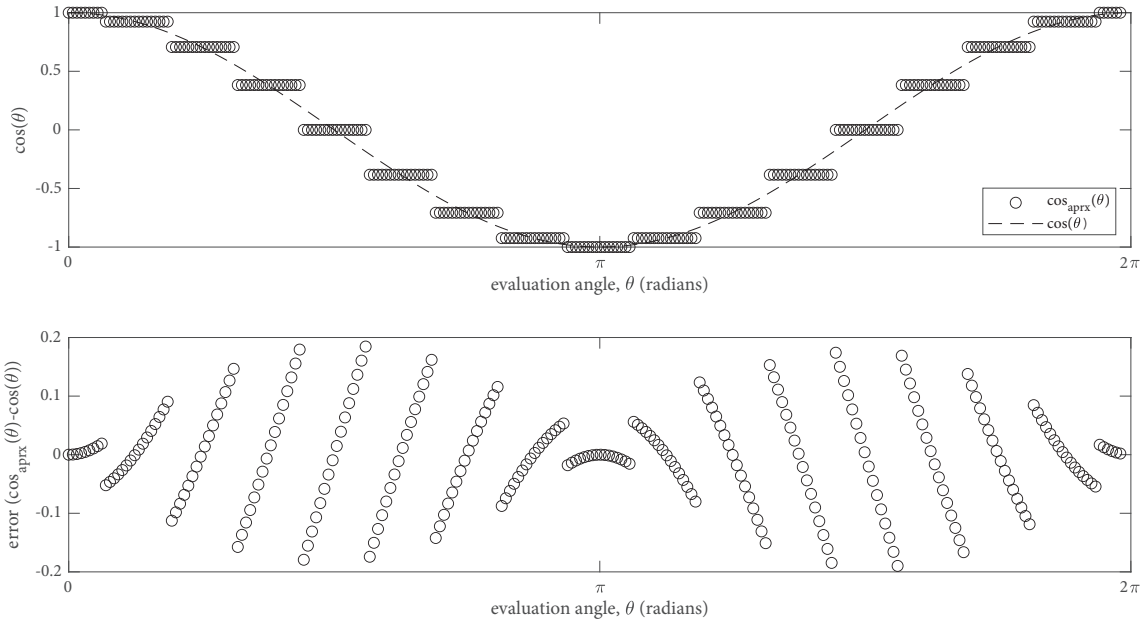

Figure 5-11: The output and error of a simple rounding implementation of a 16-value lookup table. The top plot shows the output of the approximation function and the ideal result. The bottom plot shows the difference between the approximation and actual value. It can be seen that by using this implementation, the error is high near the transition between values in the table.

$N \in \mathbb{N}^+$ is the number of points in the table, say for our example $N = 16$. Thus if given a phase θ we can calculate the best possible $n = \mathrm{mod}_N(\lfloor \theta N/2\pi - 1/2 \rfloor)$. As shown in figure 5-11, this gives us a reasonable approximation for our desired function; however, the error is still quite large. To reduce this error we can increase the size of the table or add interpolation – or both.

To expand our trigonometric function implementation to include interpolation we can simply implement the simple function

$$\cos_{\mathrm{i}}(\theta) = \cos_{\mathrm{d}}[k](1 - N\frac{\theta - 2\pi k}{2\pi}) + \cos_{\mathrm{d}}[k+1]N\frac{\theta - 2\pi k}{2\pi}, \tag{5.6}$$

where $k = \lfloor \theta N/2\pi \rfloor$. Such an implementation yields an approximation with roughly ten times smaller error when compared to the simple implementation, but this has come at the cost of many extra operations, some of which are floating-point. It is possible to optimize the calculation to reduce cycle-time. The method we use

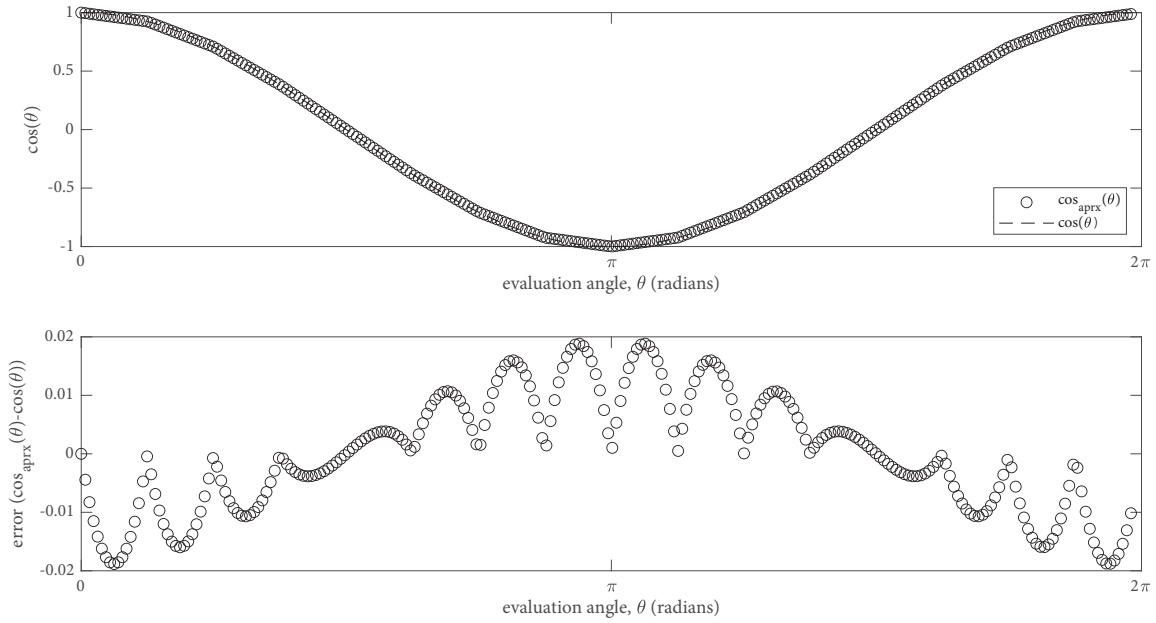

Figure 5-12: The output and error of an interpolating implementation of a 16-value lookup table. The top plot shows the output of the approximation function and the ideal result. The bottom plot shows the difference between the approximation and actual value. It can be seen that by using interpolation the error in the output signal is much smaller than it is in the simple rounding approach shown in figure 5-11.

here is to first, in the preprocessor, calculate a value of $2\pi/n$ for each $0 \leq n < N$ and store it in memory. Next we calculate the table position as a floating-point number as $p = \theta N/2\pi$. We then keep this position p and type-cast to the value $k = \lfloor p \rfloor$. We can now calculate $m = N\theta - 2\pi k/2\pi$ simply as $m = p - k$. Next we group all the terms that are multiplied by m together to that we can express $\cos_{\mathrm{i}}(\theta) = \cos_{\mathrm{d}}[k] + m(\cos_{\mathrm{d}}[k+1] - \cos_{\mathrm{d}}[k])$, to avoid two floating-point multiplications at the expense of an additional (and very fast) access to the lookup table. Thus, we can perform linear interposition in two floating-point multiplications and three floating-point additions. The results of this operation is shown in figure 5-12. It can be seen that the error is around ten times lower for this method than for the simple lookup approach.

The final method of improving trigonometric function approximation implemented here is to simple increase the size of the lookup table. The simplest way to do this is

to just calculate the function for higher values of N and store the result in memory. There is however a simple method to increase the number of effective points without increasing the memory demand. This is to take advantage of the symmetries of the sine and cosine functions. First, we can see that for $\sin(\theta) = -\sin(\theta + \pi)$, and $\sin(\theta) = \sin(\pi - \theta)$, and that $\sin(\theta) = \cos(\theta - \pi/2)$. Thus, my means of a handful of floating-point additions and sign changes, we can compress a full lookup table for both the sine and cosine functions by a factor of eight. Thus, for the same memory allocation we can have eight times as many points in the table. This folding of the lookup table is may not be suitable for low-performance processors, but here we have ample cycle time to perform all these orations. Thus, we are able to generate a sine and cosine wave to within the resolution of the 16-bit DAC with an error less than half a count using a small lookup table of only 256-points.

Calibration

The system was designed with extensive internal diagnostic and calibration capability. Upon power up the system begins measuring the power supply rails and waits for them to stabilize before proceeding. If a timeout occurs prior to the rails becoming stable, then a fault state is entered. One the rails stabilize, the dummy load resistors are used to exercise the full signal path from the DAC output, through the analog front-end, and back to the sense amplifiers and ADCs. The output voltages and currents are compared to the expected values. If the measured values deviate from the expected value by more than the specified tolerance, then the unit will throw a fault. Finally, from these measurements, the unit performs an automatic alignment such that any errors in the component values are appropriately compensated for. Only once all calibration and alignment is complete will the unit enter the ready state.

In addition to the startup calibration procedures, the unit continuously monitors the output voltage and current. The supplied current on a given phase is compared to the return current on that phase. If there is an imbalance between the supplied and return current, then the unit will throw a fault. The imbalance protection ensures that the motor has been correctly wired and that there is no leakage in the motor

or wiring. Additionally, the output current is compared to the programmed current limit. If the output current exceeds the limit by some margin (to compensate for control loop step response) then the unit will throw a fault. Finally, if the output voltage exceeds the programmed output voltage compliance, then the unit will throw a fault. The combination of these processes ensure that the driver and motor are not damaged during fault conditions. Such fault tolerance is vital as the cryocooler is expensive to repair or replace, but more importantly, the equipment down-time is extremely costly to the group.

The infiniStep driver has been in use within our lab for a year. It has run for many hundreds of hours continuously with no issues or faults. The reduction in the noise in the cryostat has proven invaluable for nearly all experiments performed with this apparatus.

5.3 Biasing superconductors – the OmniBias

The OmniBias is a simple device that allows us to generate current-voltage relation curves and bias nanowires with a current source, voltage source, or user-selectable source impedance. The device incorporates a current sense amplifier with switchable gain to provide current monitoring. A servo control loop can be enabled or disabled. This servo can be configured to provide an output voltage equal to the provided input voltage (voltage biasing) or provide a constant output current which will be equal to the input voltage divided by the current-sense amplifier gain selection (either $10\,\mathrm{mV}/\mu\mathrm{A}$ or $100\,\mathrm{mV}/\mu\mathrm{A}$).

5.3.1 Design

A simplified schematic of the unit is shown in figure 5-13. The instrumentation amplifier U2 measures the current to the device under test (DUT), and presents the result at the "I measured" output of the OmniBias. The switch S3 selects between two current-sense resistors Rg1 and Rg2. Two current gain selections were incorporated into the device primarily so that the user can make the trade-off between output

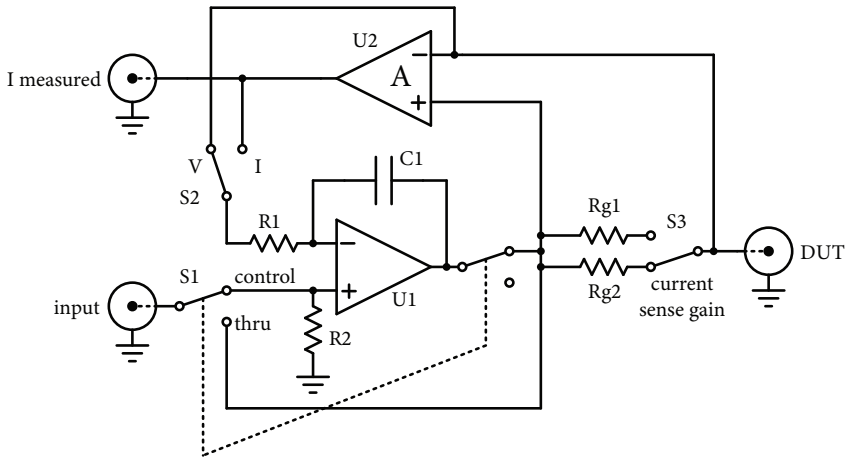

Figure 5-13: Simplified schematic of the OmniBias. Switch S1 selects between open-loop ("thru") where the OmniBias simply measures bias current, and the internal controller operating ("control"). When S1 is in the control-position, then S2 selects between voltage or current control. Switch S3 selects between the current sense resistors Rg1 and Rg2, which combined with the instrumentation amplifier U2, selects the current "I measured" output gain between 10 mV/μA and 100 mV/μA. The combination of R1 and C1 perform loop compensation to ensure stability or U1 over our expected loads. Resistor R2 provides a small load to the input signal and prevents the input from floating if disconnected.

current compliance and output resolution. An additional advantage to the use of two current-sense resistors is we can select the maximum possible current to the DUT so as to avoid damaging it by sourcing excessive current. With the OmniBias provided with a nominal $\pm$15 V supply, in the 100 mV/μA range it is capable of sourcing or sinking around 150 μA, and in the 10 mV/μA range it can source or sink around 1.5 mA.

The operational amplifier U1 performs the servo control of the output when enabled by S1. This amplifier can be configured by S2 to either control for constant output voltage equal to the input voltage, or constant output current equal to the input voltage divided by the current gain setting. The amplifier is compensated by C1 and R1, which ensures that it remains stable into all loads we expect in our experiments. As will most amplifier stabilization networks, this has the downside of reducing the amplifier bandwidth. In addition to the circuit shown in figure 5-13, there are a number of protection circuits to prevent damage to the device from gross

overloads and electrostatic discharge, as well as power supply bypass and protection electronics.

5.3.2 Performance

It is important to understand the limitations and systemic errors of your tools when performing any sensitive measurements. Here we performed two main sets of characterization of the OmniBias, namely DC performance and AC performance in select operating modes.

DC performance

The DC performance of the OmniBias was analyzed by means of a 6 ½-digit multimeter and a source measure unit (SMU). The main analyses performed were to characterize the current measurement error (see figure 5-14), and the voltage bias error (see figure 5-15) as these are the two main modes used in this work. For this experiment the load current was swept and the output of the I measured port was monitored. Ample time between load steps was provided to ensure the output had time to settle. The load was set to be zero volts (as we are mainly interested in the current to a superconductor).

The DC current measurement errors of the OmniBias are shown in figure 5-14. It can be seen that in the expanded 10 mV/μA the output current error is worse than in the 100 mV/μA range. The current measurement error for the 10 mV/μA range is within 150 nA for loads between -1.3 mA and 1.3 mA. In the fine range, where the majority of the results presented in this work were taken, we have a roughly constant error of 20 nA with the peak being less than 60 nA over the range of load currents between $-130\,\mu$A and $130\,\mu$A. For the purposes of the measurements in this work, these errors are more than acceptable.

The voltage-bias errors of the OmniBias are shown in figure 5-15. In this experiment the device was configured in the voltage control mode with a sensitivity of the 10 mV/μA range. The input programming voltage was swept from -15 V to

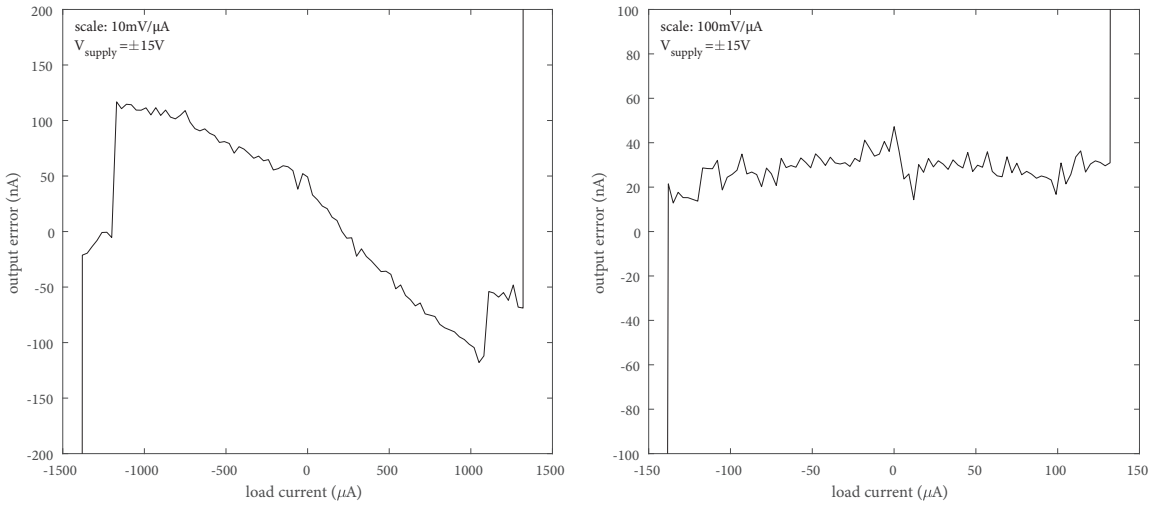

Figure 5-14: Plot of the output current errors for the OmniBias. It can be seen that in the expanded 10 mV/μA range, there is a peak error of less than 150 nA, and in the fine range the peak error is less than 60 nA. These are the errors that are baked into all current measurements of nanowires presented in this work.

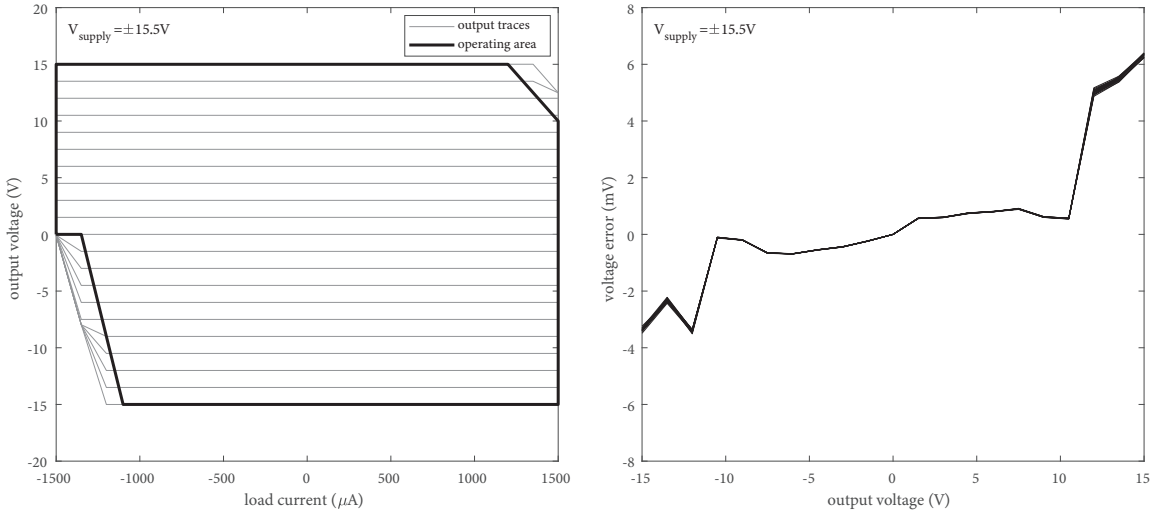

Figure 5-15: Plots of the performance of the OmniBias in the voltage bias mode. (left) The output voltage of the OmniBias as the load varies for various programmed output voltages. The thick trace shows the operating area for which the output regulation is limited. (right) The output voltage error, as the programmed output voltage varies. This plot is a superposition of all the load currents and output voltages within the operating area. The output error can be seen to be less than 1 mV for output voltages within ±10 V.

+15 V. The SMU was attached to the output and swept from −1.5 mA to 1.5 mA while the output voltage was monitored. With these parameters and for a supply bias of ±15.5 V, the operating area is that shown in the left of figure 5-15. Over this operating area we have the error shown in the right of figure 5-15. It can be seen that for voltages in the ±10 V range the voltage error is less than 1 mV, and is less than 7 mV over the full range. We operated the device mainly near zero volts where we have an error of less than 100 μV for voltages in the ±1 V range. For our purposes, this minor error is more than sufficient.

AC performance

For the AC performance, we first analyzed the device with a 100 mV stimulus at the input, with the unit in "thru" mode, and in the fine 100 mV/μA range. A load impedance of 1 kΩ was used and the VNA input set to a 1 MΩ input impedance. The results of the analysis are shown in figure 5-16. It can be seen that the gain of the device is very flat out to around 10 kHz, and the device has a 3 dB bandwidth of 38 kHz. The same analysis was performed for the device in the 10 mV/μA range, and the results are nearly identical to those shown in figure 5-16, thus this plot was omitted. Given that the amplifier U2 is limiting the bandwidth, and we are simply measuring the performance of that instrumentation amplifier, it follows that the current-sensing gain is the same for both ranges.

The AC performance of the OmniBias in the voltage bias mode was also analyzed. The same setup was used but with the OmniBias in the voltage control mode and the VNA input attached to the load. The results for the device in the 100 mV/μA range are shown in figure 5-17. Only the response for the 100 mV/μA is provided as the 10 mV/μA was not used in the experiments presented in this work. In the 10 mV/μA range, the bandwidth increased to 1.32 kHz with a similar roll-off and phase relation to that shown in figure 5-17.

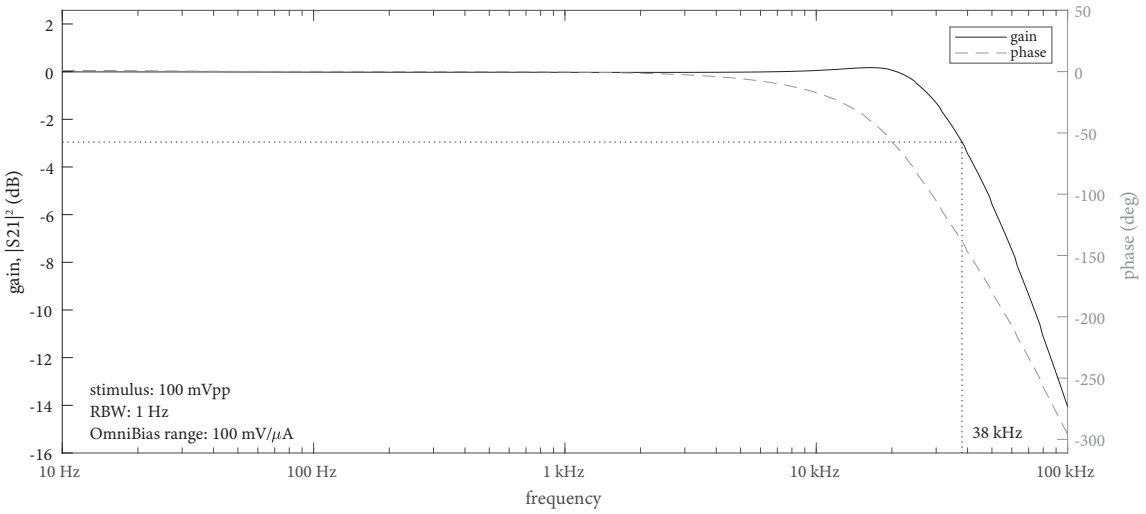

Figure 5-16: Bode plot of the AC performance of the OmniBias current measurement. The gain can be seen to be very flat out to around 10 kHz, with the 3 dB bandwidth being 38 kHz. This plot was captured for the device operating in the 100 mV/μA range; however, the performance is nearly identical for the 10 mV/μA. The output was loaded with a 1 kΩ resistor. The VNA resolution bandwidth was set to 1 Hz and the stimulus amplitude to the OmniBias in "thru" mode was set to 100 mV.

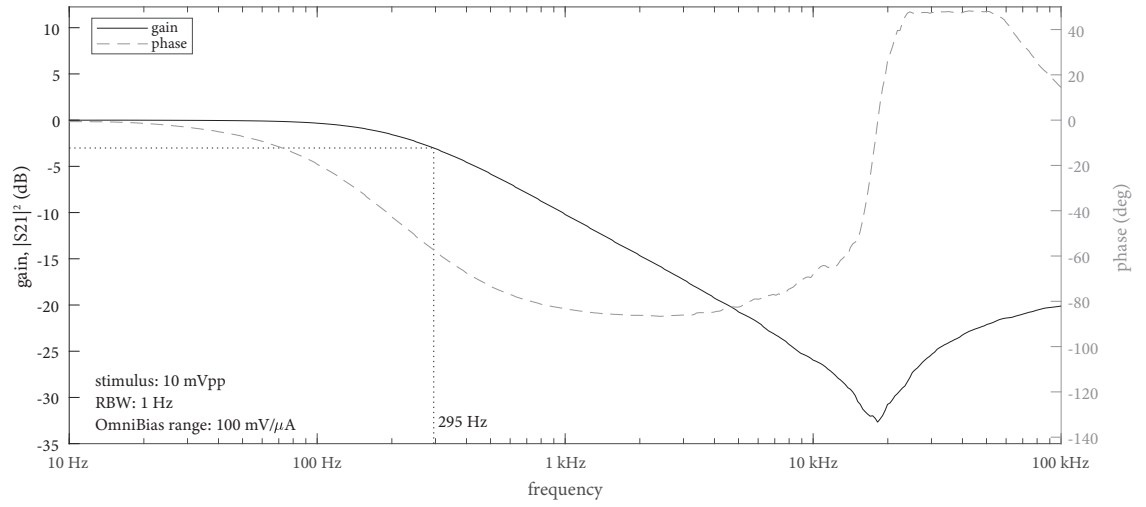

Figure 5-17: Plots of the magnitude and phase response of the OmniBias output to input with the unit in the voltage-control mode. The current sensitivity was set to 100 mV/μA, with a 10 mVpp voltage at the OmniBias input and a 1 kΩ load on the output. The VNA was configured for the sweep range shown with a resolution bandwidth of 1 Hz. It can be seen that the magnitude rolls off with a 3 dB frequency of 295 Hz.

5.3.3 Demonstration 1: capacitive shunting of nanowires

As a simple demonstration of the OmniBias, we will demonstrate the increase in device switching current when the device is shunted by a capacitor. It is well-known that shunting a nanowire can reduce noise in the nanowire current, thereby increasing the apparent switching current [102]. By creating a low impedance environment for high frequencies, we can obtain the higher switching currents that are possible with resistive shunting without the DC resistance and associated change in the retrapping behavior. This is simply achieved by shunting the device with a suitable capacitor.

The setup for this experiment is a single nanowire of width 100 nm and relatively short (around one square), with suitable tapers to avoid current crowding. The device was prepared on a PCB with a pad for a shunt capacitor. The PCB was installed into the CryoProbe and cooled down to the base temperature of around 4.28 K. The IV curve of the nanowire was then captured using an oscilloscope attached to the OmniBias in the constant-current mode with a gain of 100 mV/μA. The stimulus waveform to the OmniBias was a 100 Hz triangle wave with a peak-to-peak amplitude of 1.2 V_{pp}. The resultant IV curve is shown in figure 5-18.

Once the waveform was captured, the probe was removed from the dewar and allowed to warm up to near room temperatures. The sample PCB was removed and a 1 nF NP0 dielectric 0603 package capacitor was installed shunting the nanowire to ground. This was performed without affecting the wire bonds or changing the sample in any way. The PCB was then installed back into the probe and the same test performed again. The resultant IV curve is shown in figure 5-18. It can be seen that the critical switching current increased from 37.3 μA for the unshunted case to 51.0 μA for the shunted case. To ensure that nothing happened to the sample in the process of installing the capacitor, it was removed and the device tested again and the same unshuted switching current was observed.

There are a number of sources of noise that serve to reduce the switching current of a nanowire including external noise, as well as more intrinsic noise sourced such as thermal fluctuations. Many of these sources will see the shunt as a preferable path

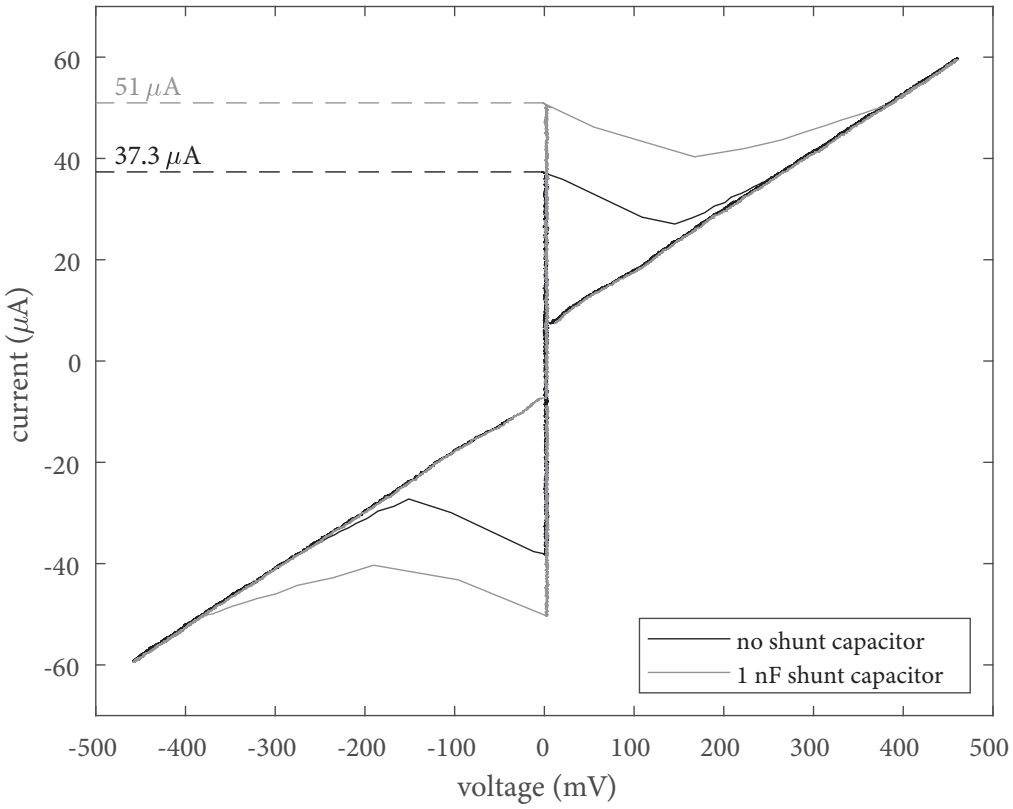

Figure 5-18: Comparison of the same short 100 nm wide nanowire tested at the same temperature and using the OmniBias with and without a 1 nF shunt capacitor installed across it. It can be seen that with the shunt installed, the switching current increased, and without it, the switching current is suppressed.

to take rather than keeping energy in the nanowire that leads to it switching early. As mentioned above, resistive shunting has been presented as a method of addressing these noise sources. However, using small shunts comes with a number of negative side effects, most notably being that it changes the retrapping behavior and simply making the signals very small in amplitude. Given that we see comparable improvements in switching currents by using a capacitive shunt, it follows that in such applications where the downside of resistive shunting prohibit their use, capacitve shunting may be an alternative. That being said, capacitive shunting has its downsides as well. The most notable downside being that it presents a decreasing impedance as frequency increases, which can lead to issues with fast measurements. In addition, non-nonidealities of the capacitor can become an issue in some applications.

5.3.4 Demonstration 2: the nanoADC

The construction of a crude ADC is another interesting demonstration of the OmniBias. Here we voltage-bias a nanowire at such a small voltage that it goes into relaxation oscillation. We can then measure the period of the oscillation to determine the voltage bias.

It might not be immediately obvious why a nanowire with a voltage bias will oscillate. When the nanowire transitions from the superconducting state to the normal state, we see a large increase in resistance. At the OmniBias output, the device is sinking current into what looks like a short circuit until it suddenly transitions to a resistor. At this transition point, the output must rapidly decrease its output current to maintain the programmed output voltage. However, the output cannot settle fast enough as the nanowire is operating orders of magnitude faster than the OmniBias. Thus, we end up undershooting the required bias briefly which resets the nanowire to the superconducting state. This can be seen in figure 5-19, where the nanowire cyclically charges and discharges.

From a circuit analysis point of view we can model output of the OmniBias, albeit somewhat crudely, as an ideal voltage source in series with the parallel combination of an inductor L_t and a resistor. This model will give a similar behavior to the transfer

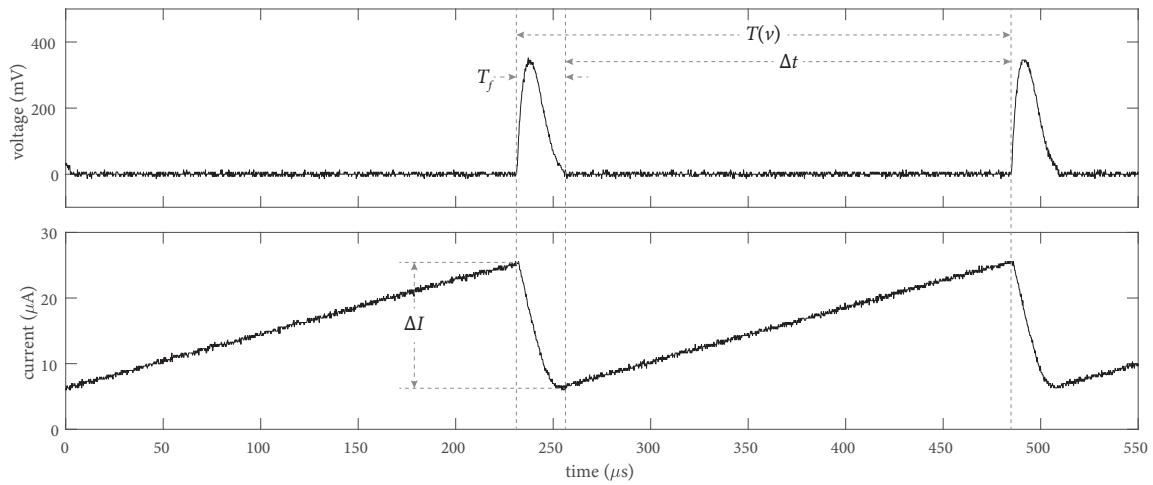

Figure 5-19: An annotated view of the conversion cycle of the nanoADC. In this figure we see the current through the nanowire, as well as the voltage across the nanowire with a constant voltage bias. It can be seen that the nanowire-OmniBias system is oscillating. This oscillation is divided into two regions: the Δt charging period, and the T_f discharging period. The current through the nanowire sweeps a region ΔI between the retrapping current and the critical current. One complete cycle of the nanoADC is composed of one charge and one discharge and takes a period of $T(v) = \Delta t + T_\mathrm{f}$, and depends on the voltage bias.

function of the OmniBias in voltage mode shown in figure 5-17. For the OmniBias in the $100\,\mathrm{mV}/\mu\mathrm{A}$ range we have $L_\mathrm{t} \approx 13\,\mathrm{mH}$. Now, when we have a constant voltage charging the inductor in series with the nanowire in the superconducting state, we will see a linear increase in current according to the inductor constitutive equation $v = L_\mathrm{t}\mathrm{d}i/\mathrm{d}t$. This charging period is evident when we plot the current through the nanowire, as shown in figure 5-19. The current through the nanowire will continue to increase linearly until it switches, at which point we will have the inductor in series with a resistor and the circuit will relax, and for a large enough inductor, it will reset to the superconducting state. Thus, we enter into relaxation oscillation. This is similar to the mechanism that resets SNSPDs [67, 103, 104].

Following from our simplistic circuit model, current to the nanowire over the charging period is given by

$$i(t) = \int \frac{v(t)}{L_\mathrm{t}}\,\mathrm{d}t + I_\mathrm{r}. \tag{5.7}$$

The nanowire current increases from the re-trapping current I_r to the switching current I_c, thus we have a $\Delta I = I_\mathrm{c} - I_\mathrm{r}$ – see figure 5-19. We can form an approximate solution to equation 5.7 by assuming that the inductor charge time (approximately the oscillation period) is much faster than the signal frequency. Under this approximation we can form the difference equation $\Delta I = (v/L_\mathrm{t})\Delta t$, where Δt is the charging time. For the period over which the nanowire is switched, our expression becomes a little more complex to determine as we need to take into account the non-linearities in the nanowire normal resistance – see section 2.5. Here, we will just call the time T_f and state that it is largely proportional to L_t and independent of bias voltage. For our nanowire and setup, this term was $T_\mathrm{f} \approx 25\,\mu\mathrm{s}$. Thus, for low frequency signals $v(t)$, the full period is given by $T = \Delta t + T_\mathrm{f}$ and can be expressed as

$$\begin{aligned} T(v) &= \frac{(I_\mathrm{c} - I_\mathrm{r})L_\mathrm{t}}{v} + T_\mathrm{f} \\ \Rightarrow v(T) &= \frac{(I_\mathrm{c} - I_\mathrm{r})L_\mathrm{t}}{T - T_\mathrm{f}}. \end{aligned} \tag{5.8}$$

Equation 5.8 does not take into account three effects. First, as discussed, it only assumes a slow signal $v(t)$. Second, the equation ignores the effect of noise. As mentioned previously in chapter 2, the slower the measurement, the lower the apparent switching current. This effect results in non-linearity in the ADC result as at lower v the slower the effective measurement of the switching current is. Finally, it does not take into account that the nanowire will eventually latch. Experimentally, we find that this occurs at around $T < 5T_\mathrm{f}$.

Experimentally we setup a short 100 nm wide nanowire biased from the OmniBias in voltage mode with the current sensitivity set to $100\,\mathrm{mV}/\mu\mathrm{A}$. The OmniBias input signal was provided by an AWG and the output of the OmniBias and the voltage across the nanowire were monitored on an oscilloscope. A single conversion cycle of the nanoADC is shown in figure 5-19.

We first performed a linearity measurement of the nanoADC. Here we swept the bias voltage from 0 mV to 2.2 mV over the period of one second. The output period of the nanoADC pulses were captured and are plotted on the left axis of figure 5-20.

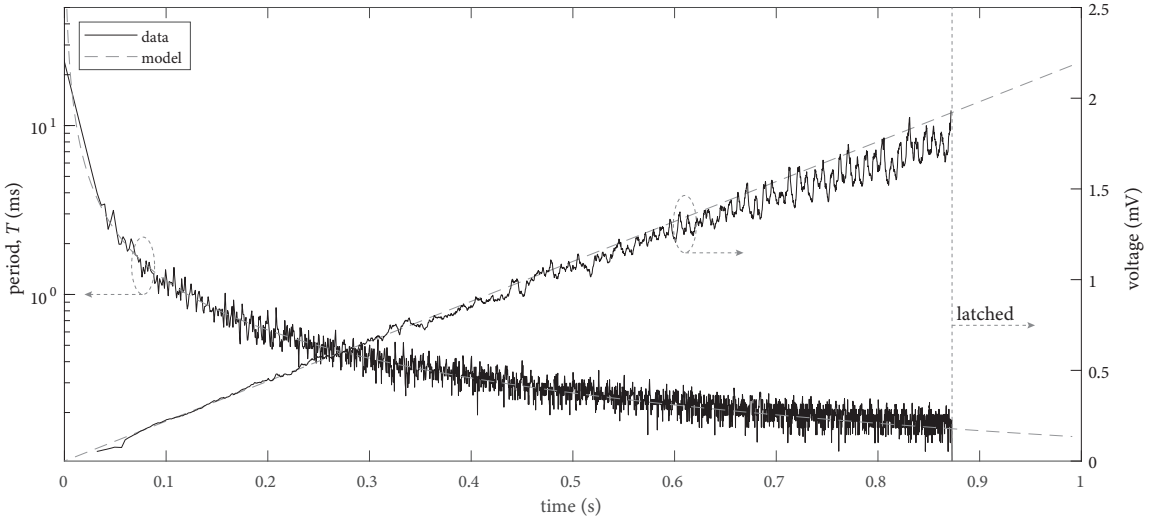

Figure 5-20: Plot of the output period and voltage for the nanoADC with a linear input voltage sweep. On the left axis we plot the period of the voltage pulses from the nanowire (solid) and the ideal period according to our model (dashed). On the right axis we plot the periods converted to voltages by equation 5.8 (solid) and the ideal response (dashed). The result of the conversion, after being converted to a voltage is then moving-averaged with a 25-sample wide window. It can be seen that for voltages above around 2 mV the nanowire latches, and we see no output pulses.

For comparison, the ideal model curve is also shown dotted. It can be seen that the model and the ideal curve agree up to around a voltage bias of 2 mV at which point the nanowire latches. The measured periods of the signal were captured and processed according to equation 5.8, to give the results shown on the right axis in figure 5-20. It can be seen that for small voltages, the output is very smooth and linear. As the voltage increases the noise increases and the output reading begins to slightly deviate towards lower values.

Next, we applied a sinusoidal signal to the nanoADC and the results are shown in figure 5-21. It can be seen that the output of the nanoADC agrees well with the applied signal. We find similar results to the linear results shown in figure 5-20 in terms of the signal noise increasing as the bias increases. A number of other signals were applied to the nanoADC and the results gave similar agreement. Due to the conversion mechanism, the bandwidth of the device is only limited to the relaxation oscillation frequency that occurs at the signal's minimum value. Thus, the step response of the ADC is as fast as a single conversion cycle with no settling

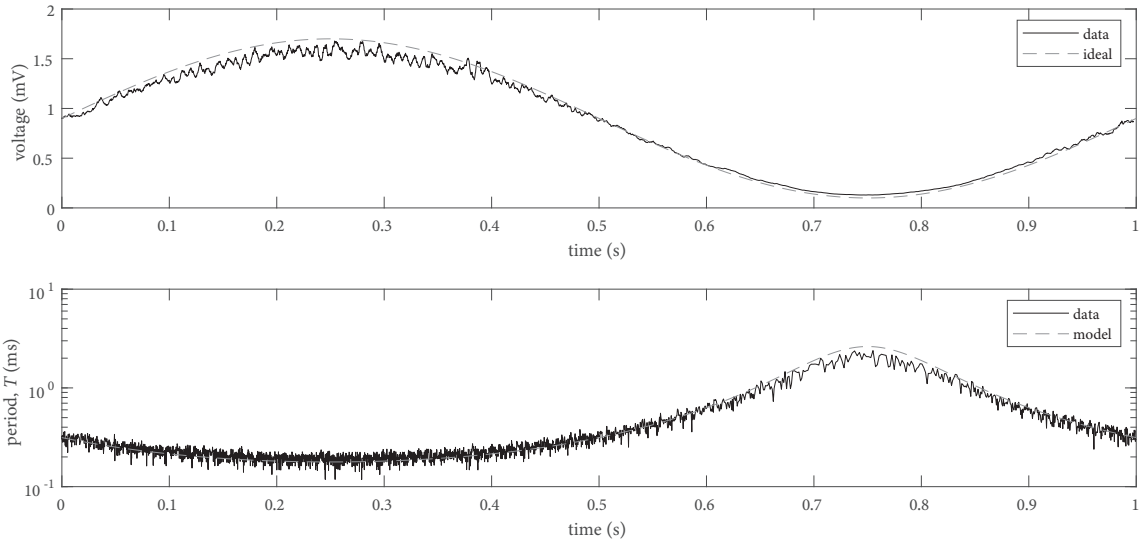

Figure 5-21: Plot of the response of the nanoADC to a 1 Hz sine wave signal. (top) The converted output voltage (solid) and the ideal output voltage (dashed). The result of the conversion to a voltage is moving averaged with a 25 sample wide window. (bottom) The measured period of the output pulses from the nanoADC (solid) and the model output signal (dashed).

time required. The main limitation is the noise in the output. The noise could be improved by moving the biasing circuitry closer to the chip and filtering high frequency interference. In addition, the output signal of the nanoADC could be improved by performing more oversampling.

This method of using a nanowire as a voltage-to-frequency converter is somewhat similar to DC-to-SFQ converters [105]. The nanowire method could be used for similar purposes or for small converting on-chip signals to frequencies that can be easily measured at room temperature. The downside of this system is that it requires either a low-impedance voltage source or a revision of equation 5.8 to suit the actual source impedance.

5.4 Membrane support

There are ongoing efforts to fabricate superconducting photon detectors on thin membranes. These devices are of interest due to an anticipated increase in performance obtained by limiting the escape of hot phonons into the substrate. However, fabrica-

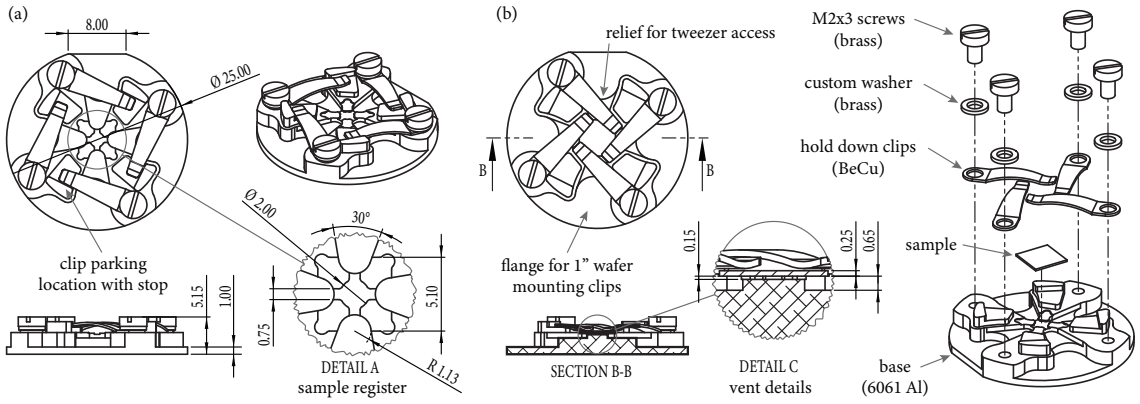

Figure 5-22: Details for the membrane support (all dimensions in mm). (a) The sample mount with the clips in their parked location and no chip present. (b) The sample installed into the support with the retaining clips retaining the sample within the membrane support. (right) An exploded view of the chip installed in the membrane support.

tion with these devices is non-trivial as simply handling the membranes are difficult without rupturing the membrane. In order to perform basic fabrication steps such as material deposition, resist spinning, and etching, a support mount was developed. This mount holds the membrane frame by its edges and leaves an open void below the membrane (located in the center of the chip). A figure of the membrane support for a 5 mm × 5 mm chip is shown in figure 5-22. A second design was created for 10 mm × 10 mm. This design for larger chips is functionally identical to that shown in these figures with only some minor changes to the geometry. Thus, drawings for this larger design are not presented here.

The membrane support was designed to have the profile of a 1" wafer with a single flat – the configuration of a 111-orientation p-type wafer. There are four regions around the outside of the base where clips can be used to retain the larger mount in fabrication equipment. In equipment with a vacuum chuck, the sample mount is simply installed as one would a wafer as the bottom side is simply a flat surface. When machining the part, careful attention was made to ensure the sample mount plane, and the bottom surface of the support were parallel. This is critical for fabrication processes that rely on the sample to be flat on the chuck.

The sample is mounted into a small recess in the top side of the membrane support.

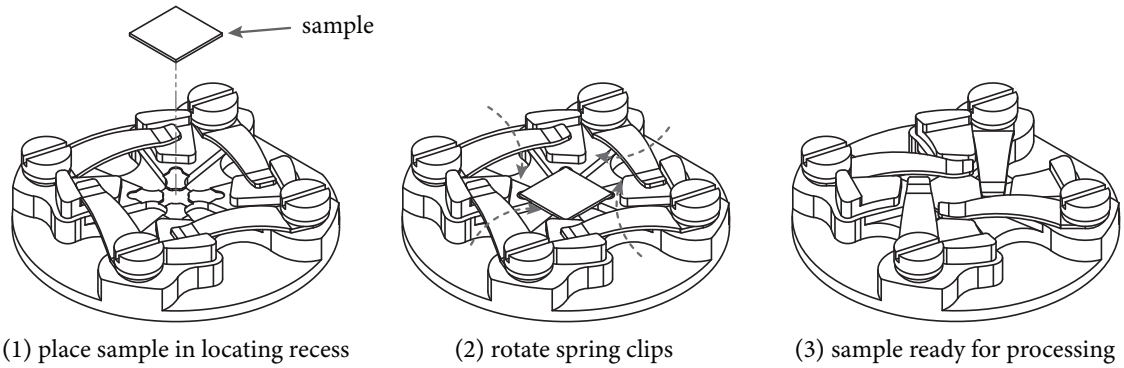

Figure 5-23: Steps for mounting a sample into the membrane support. (1) With the clips in their parked location, the support is cleaned and the sample installed into the central locating register. (2) With the chip located, the clips are rotated such that they secure the chip into the sample register. (3) With all the clips retaining the chip, the sample is ready for processing. To remove a sample after processing, the same steps are performed in reverse.

This constrains the sample in the horizontal plane. A small allowance of 0.1 mm was made to accommodate for sample size variation. A passage under the central region of the chip allows for air to escape. This passage prevents a pressure differential from forming and applying excessive stress to the membrane during installation and removal as well as pumping and venting of fabrication equipment chambers.

Four clips retain the sample in the membrane support. These clips are similar to those used to mount electron-microscopy samples. These clips are mounted by means of four 2 mm brass screws. The clips are extremely thin. To allow for the fact that the screw thread is not fully formed right up to the underside of the head of the screw, small washers are used, thus preventing the thread binding before the clip is clamped. In addition, the washers limit the torque imparted on the clip by the screw. These washers were custom machines such that the commercial screws did not need modification as the available lengths were either too long or too short for our application. These washers allowed us to use the longer screw and take up some length to prevent it from bottoming out.

The steps to mount and unmount a sample are shown in figure 5-23. When installing a sample, the clips are first positioned in their parking location. The clips are very small and can be difficult to maneuver. To prevent them from moving past

their parking location small end stops were machined into the body of the mount. With the clips pushed against their mount, there is ample space to install the chip. If the user desires, it is possible to remove one clip to allow access to the sample register such that tweezers can be used to manipulate the chip from the side. Alternatively, the sample can be installed by using tweezers from the top with all clips installed. Once the sample is located in the mount, the clips are rotated to press onto the sample. With all clips installed, they each interfere with each other to prevent one from sliding over the central area of the chip and damaging the membrane. At this point the screws can be tightened and the sample is now robustly fixated in the mount ready for fabrication. This device has been used to successfully fabricate a number of devices without damage to the membrane.

Chapter 6

Conclusion and future work

In this work we have presented a number of technological advances that will further the development of superconducting nanowire digital and microwave devices. Additionally, we have made a number of infrastructure advances that will enable us in the future to make measurements that were not previously possible. In the following sections we summarize the main results and discuss possible future research directions.

A scalable superconducting nanowire memory – the nMem

By applying the new modeling techniques presented here, we have made further progress on a scalable superconducting memory cell and array design. Here, we have demonstrated a scalable nanowire memory technology that has achieved measured bit error ratios of better than 10^{-6}, and a predicted ultimate error ratio of 7×10^{-10}. The power and speed of the memory is competitive with other superconducting memory technologies.

Fabrication and testing of nMem arrays based on the single cell presented here would be the next major step towards a large-scale superconducting memory technology. In addition, efforts to reduce the propagation delay of hTrons and their power dissipation would be helpful towards making the nMem more applicable to fast-cache applications. Although not strictly necessary, it should also be possible to engineer the memory structure such that the memory cell is non-destructive. This would be

achieved by ensuring that as the read bias is removed, we have that the two hTrons retrap such that a positive persistent current is left in the memory cell (right hTron retraps prior to the left hTron).

Ultimately, with additional research and development, the memory design presented here could allow for the construction of high performance superconducting supercomputers.

Nanowire microwave devices – the microwave toolbox

We have demonstrated on-chip microwave devices implemented with superconducting nanowires. These devices include common components such as bias tees, filters, switches, couplers, variable delay lines, and detectors. As a demonstration of an integrated on-chip measurements, an interferometer was constructed to measure the phase shift in a variable delay line. These devices exploit the tunability of kinetic inductance in hTron channels and achieve high tuning ratios greater than 40 %. Additionally, the filters implemented here are extremely small and are likely the smallest electrical multi-pole filter for their operating frequency – furthermore, these filters are tunable.

New filter designs that can be defined by tuning different elements into and out of the band of interest, or by switching elements in and out, are now possible. By investigating these devices further, it is expected that a new class of tunable filters can be realized. Such filters would find many applications as a single device layout could have its response programmed in the field by means of DC bias currents.

The on-chip interferometer presented here allows for the characterization of thermal parameters in hTrons that were not previously possible. Using this same technique, it is possible to perform other similar measurements. For example, by applying magnetic fields to the sample the screening current could be means of the change in kinetic inductance of the loop formed by the reference and measurement branches of the interferometer. This, in contrast a nanoSQUID, can perform continuous measurements of the magnetic field with no switching required [106, 107]. Similar measurements can be performed with the application of light to the sample and measuring

the associated change in Cooper pair density, similar to an MKID [34].

Given that we have demonstrated a tunable resonator and an on-chip microwave detector, it is now possible to realize a nanowire AM transceiver. By using the coupler presented here, it is possible to perform quadrature amplitude modulation (QAM) and demodulation. Such AM and QAM transceivers could find many applications, including realizing multiplexed communication between devices at low temperature and room temperature electronics.

With further development, modeling, and experimentation, a complete library of on-chip microwave devices implemented in nanowires can be realized. Should this investment be made, it will greatly expand our ability to implement many high-speed nanowire devices without the need for external microwave support components.

A versatile cryogenic experiment platform – the CryoProbe

We have presented the design of a new versatile experimental platform, the CryoProbe. This device allows for temperature controlled measurements down to 4.3 K. Four phase matched banks allow for a total of 28 RF lines rated to 20 GHz to be used for interfacing to the sample on a low-cost connector-less PCB. Measurements requiring axial magnetic fields up to 0.4 T can be performed utilizing the built-in superconducting magnet. Additionally, a fiber optic cable provides flood illumination of the sample. The sealed design of the probe makes it possible to operate the sample in either a helium atmosphere or a vacuum. Finally, the attachment of the probe to the dewar allows for the recovery of all helium boil-off during cool-down and prolonged experiments.

The addition of the CryoProbe to our experimental infrastructure has greatly accelerated the rate at which we can test our devices and enabled measurements that were either difficult or not possible to perform before. The probe has proven itself to be extremely useful and versatile and was used for nearly all experimental results presented in this work. Furthermore, the CryoProbe has replaced all other immersion testing in the group.

Experimental apparatus

Finally, we have documented and characterized a number of devices for experimental support of our research. The nanoBERT was used extensively for characterizing the performance of memory developed in chapter 2. The infinitStep motor driver has proven invaluable for performing sensitive experiments in various research projects not covered here. The OmniBias has been used for characterizing all manner of nanowire devices. The ability of the OmniBias to monitor the current through a voltage biased nanowire allowed us to realize a power microwave detector in chapter 3. Finally, the membrane support has allowed for the fabrication of nanowire devices on thin membranes without rupturing the membrane during the process. With the devices presented in this work, measurements of nanowire device that were previously not possible are now easily achieved.

Appendix A

An AM radio realized by a nanowire detector

As a simple demonstration of the nanowire detector as an AM receiver, a small AM transmitter was setup. For increased novelty, the AM transmitter is realized by means of a vacuum tube circuit, the schematic for which is shown in figure A-1.

The tube selected for this circuit is not optimal; however, it is what was available on-hand at the time of the experiments. The circuit for the transmitter comprises two main sections. The first being a grid-leak biased class-A amplifier using the triode of the 6CX8. This amplifier performs two main functions; it amplifies the input signal, and by means of the diodic nature of its interaction with the cathode, clamps the signal. Thus, we now have an amplified signal with a positive DC offset. The downside of this approach is that, as we will use this later on a grid of the pentode, we will have a modulation level dependent on the signal dynamic range. But given that the purpose of this circuit is not to provide quantitative results, we can accept this shortcoming. The second section of the circuit is a modulated crystal Colpitts oscillator realized using the pentode portion of the 6CX8. Again, this circuit was designed to be as simple as possible. Thus, a crystal was chosen to set the carrier frequency as it eliminated the need to tune the circuit for adequate performance.

The signal from the modulator shown in figure A-1, was coupled to a line leading to the detector, as discussed in section 3.11.1. In the experiment, we connected the

output of the detector to a speaker, allowing us to hear the demodulated signal. It was apparent in listening to the signal that the audio was effectively low-pass filtered, as expected. The audio signal, AM signal, and output of our nanowire detector were captures and processed to form the results shown in figure A-2. It can be seen in this figure that the modulator performs as expected, and that out detector was able to detect the signal effectively. It is also apparent that the frequency response of the detector to the base band signal is not flat. This issue could easily be addressed by increasing the bandwidth of bias network.

The nanowire detector, as demonstrated here, can be used as a practical RF power detector. In the main text we showed the detector operating at 6 GHz, and here down to 6.25 MHz. We expect that this operation can be extend to higher and lower frequencies. With the instrumentation discussed in this work, combined with this detector, and an oscillator, we can perform totally on-chip measurements with DC outputs.

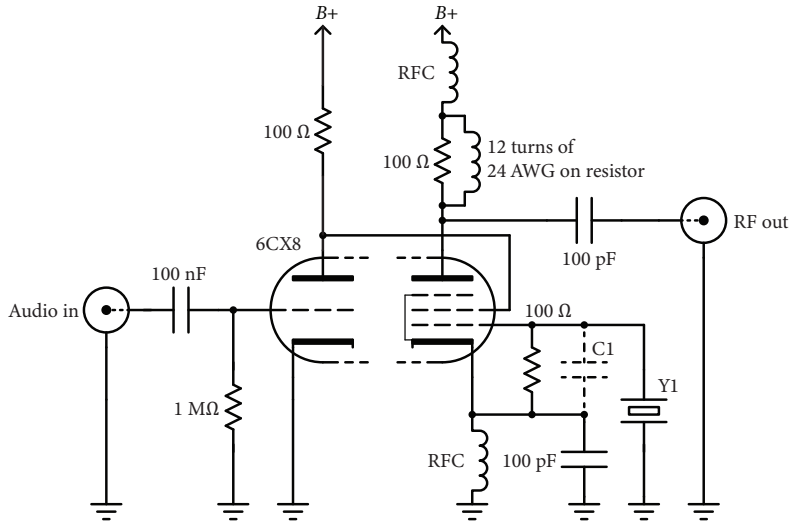

Figure A-1: Schematic of a very simple one-tube AM modulator. Note that C1 is not fitted in reality as the grid-to-cathode capacitance was sufficient such that an external capacitor was not required. Crystal Y1 was socketed, and for the results presented here, a 6.25 MHz crystal was used.

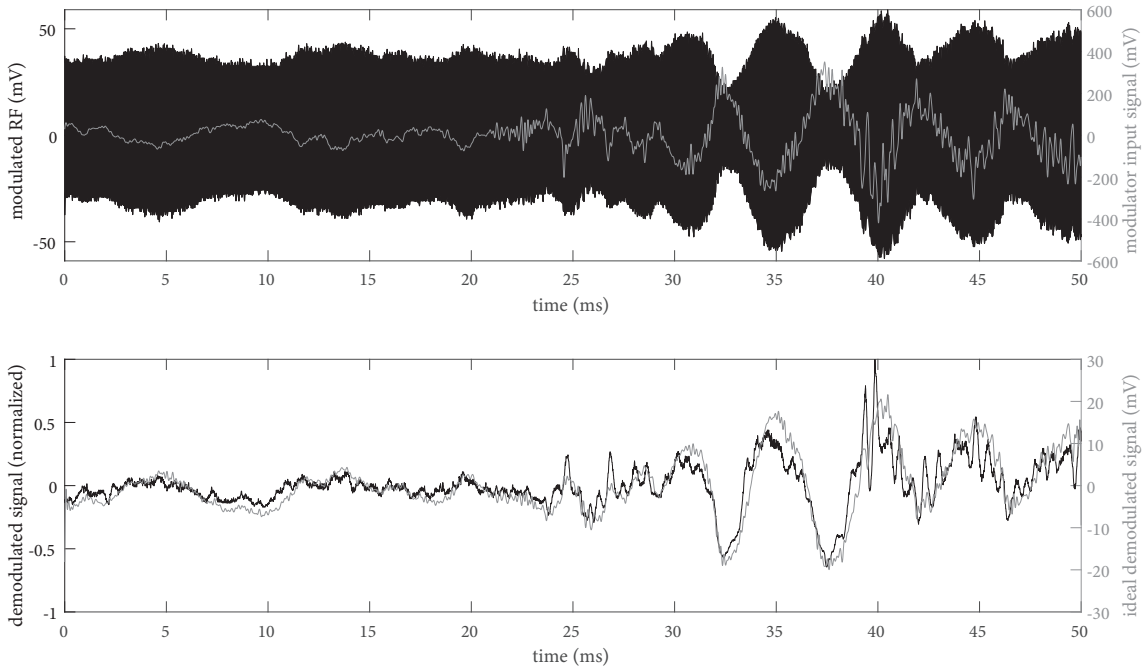

Figure A-2: Experimental results of the nanowire detector operating as an AM radio. In the top plot we have the audio and AM signal. In the bottom plot we have the detected AM envelope and the ideal detected signal. It can be seen that the nanowire detector output roughly matches the ideal envelope.

Bibliography

[1] D. S. Holmes, A. L. Ripple, and M. A. Manheimer. Energy-efficient superconducting computing—power budgets and requirements. *IEEE Transactions on Applied Superconductivity*, 23(3):1701610–1701610, June 2013.

[2] Marc A Manheimer. Cryogenic computing complexity program: Phase 1 introduction. *IEEE Transactions on Applied Superconductivity*, 25(3):1–4, 2015.

[3] Brenden A Butters, Reza Baghdadi, Murat Onen, Emily A Toomey, Owen Medeiros, and Karl K Berggren. A scalable superconducting nanowire memory cell and preliminary array test. 34(3):035003, jan 2021.

[4] Qing-Yuan Zhao, Emily A Toomey, Brenden A Butters, Adam N McCaughan, Andrew E Dane, Sae-Woo Nam, and Karl K Berggren. A compact superconducting nanowire memory element operated by nanowire cryotrons. *Superconductor Science and Technology*, 31(3):035009, February 2018.

[5] Brenden A Butters. Development of a scalable superconducting memory. Master's thesis, Massachusetts Institute of Technology, 2018.

[6] B.D. Josephson. Possible new effects in superconductive tunnelling. *Physics Letters*, 1(7):251 – 253, 1962.

[7] D. A. Buck. The cryotron-a superconductive computer component. *Proceedings of the IRE*, 44(4):482–493, April 1956.

[8] Reza Baghdadi, Jason P Allmaras, Brenden A Butters, Andrew E Dane, Saleem Iqbal, Adam N McCaughan, Emily A Toomey, Qing-Yuan Zhao, Alexander G Kozorezov, and Karl K Berggren. Multilayered heater nanocryotron: A superconducting-nanowire-based thermal switch. *Physical Review Applied*, 14(5):054011, 2020.

[9] HL Hortensius, EFC Driessen, TM Klapwijk, KK Berggren, and JR Clem. Critical-current reduction in thin superconducting wires due to current crowding. *Applied Physics Letters*, 100(18):182602, 2012.

[10] S Frasca, B Korzh, M Colangelo, D Zhu, AE Lita, JP Allmaras, EE Wollman, VB Verma, AE Dane, E Ramirez, et al. Determining the depairing current in superconducting nanowire single-photon detectors. *Physical Review B*, 100(5):054520, 2019.

[11] Minh-Hai Nguyen, Guilhem J Ribeill, Martin V Gustafsson, Shengjie Shi, Sriharsha V Aradhya, Andrew P Wagner, Leonardo M Ranzani, Lijun Zhu, Reza Baghdadi, Brenden Butters, et al. Cryogenic memory architecture integrating spin hall effect based magnetic memory and superconductive cryotron devices. *Scientific reports*, 10(1):1–11, 2020.

[12] Vasili K Semenov, Yuri A Polyakov, and Sergey K Tolpygo. Very large scale integration of josephson-junction-based superconductor random access memories. *IEEE Transactions on Applied Superconductivity*, 29(5):1–9, 2019.

[13] Shuichi Nagasawa, Kenji Hinode, Tetsuro Satoh, Yoshihiro Kitagawa, and Mutsuo Hidaka. Design of all-dc-powered high-speed single flux quantum random access memory based on a pipeline structure for memory cell arrays. *Superconductor Science and Technology*, 19(5):S325, 2006.

[14] Masamitsu Tanaka, Masato Suzuki, Gen Konno, Yuki Ito, Akira Fujimaki, and Nobuyuki Yoshikawa. Josephson-cmos hybrid memory with nanocryotrons. *IEEE Transactions on Applied Superconductivity*, 27(4):1–4, 2017.

[15] Q. Liu, K. Fujiwara, X. Meng, S. R. Whiteley, T. Van Duzer, N. Yoshikawa, Y. Thakahashi, T. Hikida, and N. Kawai. Latency and power measurements on a 64-kb hybrid josephson-cmos memory. *IEEE Transactions on Applied Superconductivity*, 17(2):526–529, 2007.

[16] W Chakraborty, KA Aabrar, J Gomez, R Saligram, A Raychowdhury, P Fay, and S Datta. Cryogenic rf cmos on 22nm fdsoi platform with record f t= 495ghz and f max= 497ghz. In *2021 Symposium on VLSI Technology*, pages 1–2. IEEE, 2021.

[17] Fred Ware, Liji Gopalakrishnan, Eric Linstadt, Sally A McKee, Thomas Vogelsang, Kenneth L Wright, Craig Hampel, and Gary Bronner. Do superconducting processors really need cryogenic memories? the case for cold dram. In *Proceedings of the International Symposium on Memory Systems*, pages 183–188, 2017.

[18] Ian M Dayton, Tessandra Sage, Eric C Gingrich, Melissa G Loving, Thomas F Ambrose, Nathan P Siwak, Shawn Keebaugh, Christopher Kirby, Donald L Miller, Anna Y Herr, et al. Experimental demonstration of a josephson magnetic memory cell with a programmable π-junction. *IEEE Magnetics Letters*, 9:1–5, 2018.

[19] GE Rowlands, CA Ryan, L Ye, L Rehm, D Pinna, AD Kent, and TA Ohki. A cryogenic spin-torque memory element with precessional magnetization dynamics. *Scientific reports*, 9(1):1–7, 2019.

[20] Laura Rehm, Volker Sluka, Graham E Rowlands, M-H Nguyen, Thomas A Ohki, and Andrew D Kent. Sub-nanosecond switching in a cryogenic spin-torque spin-valve memory element with a dilute permalloy free layer. *Applied Physics Letters*, 114(21):212402, 2019.

[21] Graham E. Rowlands, Minh-Hai Nguyen, Sriharsha V. Aradhya, Shengjie Shi, Colm A. Ryan, Robert A. Buhrman, and Thomas A. Ohki. Nanosecond reversal of three-terminal spin-hall-effect memories sustained at cryogenic temperatures. *Phys. Rev. Applied*, 15:L021004, Feb 2021.

[22] Igor V Vernik, Vitaly V Bol'ginov, Sergey V Bakurskiy, Alexander A Golubov, Mikhail Yu Kupriyanov, Valery V Ryazanov, and Oleg A Mukhanov. Magnetic josephson junctions with superconducting interlayer for cryogenic memory. *IEEE transactions on applied superconductivity*, 23(3):1701208–1701208, 2012.

[23] John O. Morin. The cryotron as a storage element, 1956.

[24] GC Hermeling. A nuvistor. low-noise vhf tuner. *IRE Transactions on Broadcast and Television Receivers*, (3):52–56, 1960.

[25] V. L. Newhouse, J. W. Bremer, and H. H. Edwards. The crossed-film cryotron and its application to digital computer circuits. In *Papers Presented at the December 1-3, 1959, Eastern Joint IRE-AIEE-ACM Computer Conference*, IRE-AIEE-ACM '59 (Eastern), page 255–260, New York, NY, USA, 1959. Association for Computing Machinery.

[26] V. L. Newhouse, J. W. Bremer, and H. H. Edwards. An improved film cryotron and its application to digital computers. *Proceedings of the IRE*, 48(8):1395–1404, 1960.

[27] Adam N McCaughan, Nathnael S Abebe, Qing-Yuan Zhao, and Karl K Berggren. Using geometry to sense current. *Nano letters*, 16(12):7626–7631, 2016.

[28] Steven M. Anlage. Microwave superconductivity. *IEEE Journal of Microwaves*, 1(1):389–402, 2021.

[29] W Chen, AV Rylyakov, Vijay Patel, JE Lukens, and KK Likharev. Rapid single flux quantum t-flip flop operating up to 770 ghz. *IEEE Transactions on Applied Superconductivity*, 9(2):3212–3215, 1999.

[30] Di Zhu, Marco Colangelo, Boris A Korzh, Qing-Yuan Zhao, Simone Frasca, Andrew E Dane, Angel E Velasco, Andrew D Beyer, Jason P Allmaras, Edward Ramirez, et al. Superconducting nanowire single-photon detector with integrated impedance-matching taper. *Applied Physics Letters*, 114(4):042601, 2019.

[31] Joseph C. Bardin, Daniel H. Slichter, and David J. Reilly. Microwaves in quantum computing. *IEEE Journal of Microwaves*, 1(1):403–427, 2021.

[32] Diego Ristè, Spencer Fallek, Brian Donovan, and Thomas A. Ohki. Microwave techniques for quantum computers: State-of-the-art control systems for quantum processors. *IEEE Microwave Magazine*, 21(8):60–71, 2020.

[33] Chris Macklin, K O'brien, D Hover, ME Schwartz, V Bolkhovsky, X Zhang, WD Oliver, and I Siddiqi. A near–quantum-limited josephson traveling-wave parametric amplifier. *Science*, 350(6258):307–310, 2015.

[34] Jonas Zmuidzinas. Superconducting microresonators: Physics and applications. *Annu. Rev. Condens. Matter Phys.*, 3(1):169–214, 2012.

[35] Daniel F Santavicca, Jesse K Adams, Lierd E Grant, Adam N McCaughan, and Karl K Berggren. Microwave dynamics of high aspect ratio superconducting nanowires studied using self-resonance. *Journal of Applied Physics*, 119(23):234302, 2016.

[36] Qing-Yuan Zhao, Niccolò Calandri Di Zhu, Andrew E Dane, Adam N McCaughan, Francesco Bellei, Hao-Zhu Wang, Daniel F Santavicca, and Karl K Berggren. Superconducting nanowire single-photon imager.

[37] Qing-Yuan Zhao, Daniel F Santavicca, Di Zhu, Brian Noble, and Karl K Berggren. A distributed electrical model for superconducting nanowire single photon detectors. *Applied Physics Letters*, 113(8):082601, 2018.

[38] Di Zhu, Marco Colangelo, Changchen Chen, Boris A. Korzh, Franco N. C. Wong, Matthew D. Shaw, and Karl K. Berggren. Resolving photon numbers using a superconducting nanowire with impedance-matching taper. *Nano Letters*, 20(5):3858–3863, 2020.

[39] Marco Colangelo, Di Zhu, Daniel F Santavicca, Brenden A Butters, Joshua C Bienfang, and Karl K Berggren. Compact and tunable forward coupler based on high-impedance superconducting nanowires. *Physical Review Applied*, 15(2):024064, 2021.

[40] Daniel F Santavicca, Marco Colangelo, Carleigh R Eagle, Maitri P Warusawithana, and Karl K Berggren. 50 ohm transmission lines with extreme wavelength compression based on superconducting nanowires on high-permittivity substrates. *arXiv preprint arXiv:2111.01006*, 2021.

[41] Francesco Bellei, Alyssa P Cartwright, Adam N McCaughan, Andrew E Dane, Faraz Najafi, Qingyuan Zhao, and Karl K Berggren. Free-space-coupled superconducting nanowire single-photon detectors for infrared optical communications. *Optics express*, 24(4):3248–3257, 2016.

[42] Jack Ekin. *Experimental techniques for low-temperature measurements: cryostat design, material properties and superconductor critical-current testing*. Oxford university press, 2006.

[43] MS Wire, DJ Durand, AH Silver, MK Wagner, and CA Hamilton. A multiple pin, flip-chip system for microwave and gigabit per second cryogenic device testing at variable temperatures. *Review of Scientific Instruments*, 72(2):1542–1547, 2001.

[44] B Naser, J Heeren, DK Ferry, and JP Bird. 50-ω-matched system for low-temperature measurements of the time-resolved conductance of low-dimensional semiconductors. *Review of scientific instruments*, 76(11):113905, 2005.

[45] MS Jun, SW Hwang, DY Jeong, and D Ahn. Microwave design and characterization of a cryogenic dip probe for time-domain measurements of nanodevices. *Review of scientific instruments*, 75(7):2455–2460, 2004.

[46] A. H. Silver and J. E. Zimmerman. Quantum states and transitions in weakly connected superconducting rings. *Phys. Rev.*, 157:317–341, May 1967.

[47] S.S. Mohan, M. del Mar Hershenson, S.P. Boyd, and T.H. Lee. Simple accurate expressions for planar spiral inductances. *IEEE Journal of Solid-State Circuits*, 34(10):1419–1424, 1999.

[48] Murat Onen, Brenden A Butters, Emily Toomey, Tayfun Gokmen, and Karl K Berggren. Design and characterization of superconducting nanowire-based processors for acceleration of deep neural network training. *Nanotechnology*, 31(2):025204, 2019.

[49] E. Toomey, M. Onen, M. Colangelo, B. A. Butters, A. N. McCaughan, and K. K. Berggren. Bridging the gap between nanowires and josephson junctions: A superconducting device based on controlled fluxon transfer. *Phys. Rev. Applied*, 11:034006, Mar 2019.

[50] Adam N. McCaughan and Karl K. Berggren. A superconducting-nanowire three-terminal electrothermal device. *Nano Letters*, 14(10):5748–5753, 2014. PMID: 25233488.

[51] Andrew E Dane, Adam N McCaughan, Di Zhu, Qingyuan Zhao, Chung-Soo Kim, Niccolo Calandri, Akshay Agarwal, Francesco Bellei, and Karl K Berggren. Bias sputtered nbn and superconducting nanowire devices. *Applied Physics Letters*, 111(12):122601, 2017.

[52] D. C. Mattis and J. Bardeen. Theory of the anomalous skin effect in normal and superconducting metals. *Phys. Rev.*, 111:412–417, Jul 1958.

[53] Michael Tinkham. *Introduction to superconductivity*. Courier Corporation, 2004.

[54] Sean Enderby and Zlatko Baracskai. Harmonic instability of digital soft clipping algorithms. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, 2012.

[55] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *towards data science*, 6(12):310–316, 2017.

[56] David Pines. Spin excitations and superconductivity in cuprate oxide and heavy electron superconductors. *Physica B: Condensed Matter*, 163(1):78–88, 1990.

[57] F Gross, BS Chandrasekhar, D Einzel, K Andres, PJ Hirschfeld, HR Ott, J Beuers, Z Fisk, and JL Smith. Anomalous temperature dependence of the magnetic field penetration depth in superconducting ube 13. *Zeitschrift für Physik B Condensed Matter*, 64(2):175–188, 1986.

[58] Masashi Tachiki, Tomio Koyama, and Saburo Takahashi. Superconducting properties of layered cuprate oxides. *Progress of Theoretical Physics Supplement*, 108:297–311, 1992.

[59] Hartmut Monien and David Pines. Spin excitations and pairing gaps in the superconducting state of $yba_2cu_3o_{7-\delta}$. *Phys. Rev. B*, 41:6297–6305, Apr 1990.

[60] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.

[61] F Teyssandier and D Prêle. Commercially available capacitors at cryogenic temperatures. In *Ninth International Workshop on Low Temperature Electronics-WOLTE9*, 2010.

[62] Laurens De Haan, Ana Ferreira, and Ana Ferreira. *Extreme value theory: an introduction*, volume 21. Springer, 2006.

[63] Alexander McFarlane Mood. Introduction to the theory of statistics. 1950.

[64] Kenneth E Iverson. A programming language. In *Proceedings of the May 1-3, 1962, spring joint computer conference*, pages 345–351, 1962.

[65] Donald E Knuth. Two notes on notation. *The American Mathematical Monthly*, 99(5):403–422, 1992.

[66] Andrew J. Kerman, Joel K. W. Yang, Richard J. Molnar, Eric A. Dauler, and Karl K. Berggren. Electrothermal feedback in superconducting nanowire single-photon detectors. *Phys. Rev. B*, 79:100509, Mar 2009.

[67] Joel K. W. Yang, Andrew J. Kerman, Eric A. Dauler, Vikas Anant, Kristine M. Rosfjord, and Karl K. Berggren. Modeling the electrical and thermal response of superconducting nanowire single-photon detectors. *IEEE Transactions on Applied Superconductivity*, 17(2):581–585, 2007.

[68] Karl K Berggren, Qing-Yuan Zhao, Nathnael Abebe, Minjie Chen, Prasana Ravindran, Adam McCaughan, and Joseph C Bardin. A superconducting nanowire can be modeled by using SPICE. *Superconductor Science and Technology*, 31(5):055010, April 2018.

[69] Farzaneh Zokaee and Lei Jiang. Smart: A heterogeneous scratchpad memory architecture for superconductor sfq-based systolic cnn accelerators. In *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 912–924, 2021.

[70] Matteo Staffaroni, Josh Conway, Shantha Vedantam, Japeck Tang, and Eli Yablonovitch. Circuit analysis in metal-optics. *Photonics and Nanostructures-Fundamentals and Applications*, 10(1):166–176, 2012.

[71] Anthony J Annunziata, Daniel F Santavicca, Luigi Frunzio, Gianluigi Catelani, Michael J Rooks, Aviad Frydman, and Daniel E Prober. Tunable superconducting nanoinductors. *Nanotechnology*, 21(44):445202, 2010.

[72] John R Clem and VG Kogan. Kinetic impedance and depairing in thin and narrow superconducting films. *Physical Review B*, 86(17):174521, 2012.

[73] Reinaldo Perez. *Wireless Communications Design Handbook: Interference Into Circuits: Aspects of Noise, Interference, and Environmental Concerns*. Academic Press, 1998.

[74] John Bardeen. Two-fluid model of superconductivity. *Physical Review Letters*, 1(11):399, 1958.

[75] Jiansong Gao, J Zmuidzinas, A Vayonakis, P Day, B Mazin, and H Leduc. Equivalence of the effects on the complex conductivity of superconductor due to temperature change and external pair breaking. *Journal of Low Temperature Physics*, 151(1):557–563, 2008.

[76] Aaron D O'Connell, M Ansmann, Radoslaw C Bialczak, Max Hofheinz, Nadav Katz, Erik Lucero, C McKenney, Matthew Neeley, Haohua Wang, Eva Maria Weig, et al. Microwave dielectric loss at single photon energies and millikelvin temperatures. *Applied Physics Letters*, 92(11):112903, 2008.

[77] John R Clem and Karl K Berggren. Geometry-dependent critical currents in superconducting nanocircuits. *Physical Review B*, 84(17):174510, 2011.

[78] Dagmar Henrich, P Reichensperger, M Hofherr, JM Meckbach, Konstantin Il'in, Michael Siegel, A Semenov, Anna Zotova, and D Yu Vodolazov. Geometry-induced reduction of the critical current in superconducting nanowires. *Physical Review B*, 86(14):144504, 2012.

[79] George Matthaei. Microwave filters, impedance-matching networks and coupling structures. *Artech House Book*, pages 775–809, 1980.

[80] Harlan Howe. Stripline circuit design, artech house. *Dedham, Massachusetts*, 1974.

[81] Errikos Lourandakis. *On-wafer Microwave Measurements and De-embedding*. Artech House, 2016.

[82] Olebogeng Bone Kobe, Joseph Chuma, Rodrigo Jamisola Jr, and Matthews Chose. A review on quality factor enhanced on-chip microwave planar resonators. *Engineering Science and Technology, an International Journal*, 20(2):460–466, 2017.

[83] Protap Pramanick and Prakash Bhartia. *Modern RF and microwave filter design*. Artech House, 2016.

[84] Andrew Wagner, Leonardo Ranzani, Guilhem Ribeill, and TA Ohki. Demonstration of a superconducting nanowire microwave switch. *Applied Physics Letters*, 115(17):172602, 2019.

[85] R Basu Thakur, A Steiger, S Shu, F Faramarzi, N Klimovich, PK Day, E Shirokoff, PD Mauskopf, and PS Barry. Development of superconducting on-chip fourier transform spectrometers. *arXiv preprint arXiv:2111.06558*, 2021.

[86] E Henry Fooks and Ramutis A Zakarevicius. *Microwave engineering using microstrip circuits*. Prentice-Hall, Inc., 1990.

[87] K. K. Likharev. Superconducting weak links. *Rev. Mod. Phys.*, 51:101–159, Jan 1979.

[88] P. W. Anderson and A. H. Dayem. Radio-frequency effects in superconducting thin film bridges. *Phys. Rev. Lett.*, 13:195–197, Aug 1964.

[89] Sidney Shapiro, Andre R. Janus, and Sandor Holly. Effect of microwaves on josephson currents in superconducting tunneling. *Rev. Mod. Phys.*, 36:223–225, Jan 1964.

[90] Paul Slysh, JE Dyer, JH Furman, and JE Key. Isogrid structural tests and stability analyses. *Journal of Aircraft*, 13(10):778–785, 1976.

[91] Michael George Zabetakis and Michael George Zabetakis. *Safety with Cryogenic Fluids*. Springer, 1967.

[92] Samtec. Precision micrwave/millimeter wave interconnect solutions guide. Technical report, 2020.

[93] Nathan Ida. *Engineering Electromagnetics*. Springer International Publishing, Cham, 4th ed. 2021 edition, 2021.

[94] John H Lau. *Solder joint reliability: theory and applications*. Springer Science & Business Media, 1991.

[95] Richard A Toupin. Saint-venant's principle. *Archive for Rational Mechanics and Analysis*, 18(2):83–96, 1965.

[96] Parker Seals. Parker o-ring handbook (ord 5700). *Parker Seal Group, Lexington, KY*, 1992.

[97] Apple Rubber Products. Seal design guide. *Lancaster, New York*, 1998.

[98] Peter E Bradley, Ray Radebaugh, et al. Properties of selected materials at cryogenic temperatures. *NIST Publ*, 680:1–14, 2013.

[99] Marco Rubber & Plastics. personal communication.

[100] A D'Addabbo, C Bucci, L Canonica, S Di Domizio, P Gorla, L Marini, A Nucciotti, I Nutini, C Rusconi, and B Welliver. An active noise cancellation technique for the cuore pulse tube cryocoolers. *Cryogenics*, 93:56–65, 2018.

[101] A Santolaria, J Balcells, D Gonzalez, J Gago, and SD Gil. Emi reduction in switched power converters by means of spread spectrum modulation techniques. In *2004 IEEE 35th Annual Power Electronics Specialists Conference (IEEE Cat. No. 04CH37551)*, volume 1, pages 292–296. IEEE, 2004.

[102] Matthew W. Brenner, Dibyendu Roy, Nayana Shah, and Alexey Bezryadin. Dynamics of superconducting nanowires shunted with an external resistor. *Phys. Rev. B*, 85:224507, Jun 2012.

[103] Anthony J Annunziata, Orlando Quaranta, Daniel F Santavicca, Alessandro Casaburi, Luigi Frunzio, Mikkel Ejrnaes, Michael J Rooks, Roberto Cristiano, Sergio Pagano, Aviad Frydman, et al. Reset dynamics and latching in niobium superconducting nanowire single-photon detectors. *Journal of Applied Physics*, 108(8):084507, 2010.

[104] Andrew J Kerman, Danna Rosenberg, Richard J Molnar, and Eric A Dauler. Readout of superconducting nanowire single-photon detectors at high count rates. *Journal of Applied Physics*, 113(14):144511, 2013.

[105] J Kimble. Advanced computing. *JSR*, 94:130, 1820.

[106] Aico GP Troeman, Hendrie Derking, Bert Borger, Johannes Pleikies, Dick Veldhuis, and Hans Hilgenkamp. Nanosquids based on niobium constrictions. *Nano Letters*, 7(7):2152–2156, 2007.

[107] Adam N McCaughan, Qingyuan Zhao, and Karl K Berggren. nanosquid operation using kinetic rather than magnetic induction. *Scientific reports*, 6(1):1–5, 2016.