Title
The Microstrip SQUID Amplier in the Axion Dark Matter eXperiment (ADMX)

Author
O'Kelley, Sean Robert

Publication Date
2019

Peer reviewed|Thesis/dissertation

The Microstrip SQUID Amplifier in the Axion Dark Matter eXperiment (ADMX)

by

Sean Robert O'Kelley

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Clarke, Chair
Professor Irfan Siddiqi
Professor Karl van Bibber

Summer 2019

The Microstrip SQUID Amplifier
in the
Axion Dark Matter eXperiment (ADMX)

Abstract

The Microstrip SQUID Amplifier
in the
Axion Dark Matter eXperiment (ADMX)

by

Sean Robert O'Kelley

Doctor of Philosophy in Physics

University of California, Berkeley

Professor John Clarke, Chair

We present the development and performance of low noise tunable Microstrip SQUID Amplifiers (MSAs) operating at frequencies near 1 GHz for use in ADMX. The MSA has been used in ADMX to produce published exclusions to the axion mass and density in our galactic dark matter halo. The exclusions published by ADMX and achieved with an MSA are the most stringent exclusions to date.

Hypotheses on the nature of dark matter vary, so that the designs and operating principles of experimental dark matter searches vary as well. Every search is similar in that it is an endeavor to discover some very small signal–if the signal were not currently beyond perception we would not call our quarry "dark matter". In ADMX the principle of operation is to convert axions into photons in a strong magnetic field, enhance that conversion rate with a resonant cavity tuned to the photon frequency, and record the presence of those converted photons. The trick in any direct detection experiment is to increase signal and decrease noise. In ADMX, the signal is increased by using a strong magnetic field, a high-Q cavity, and as large a volume as possible. The noise is reduced by operating at a low physical temperature and using a low-noise amplifier. That amplifier is the titular MSA. The MSA operates as a DC-Superconducting QUantum Interference Device (DC-SQUID) coupled to a RF microstrip resonator such that the magnetic field component of a standing wave on the microstrip couples to the SQUID, generating an amplified voltage signal across the SQUID terminals, which is transmitted to the rest of the detection chain. The MSA has a noise temperature of 50 to 200 mK, enabling a search rate about 56 times faster than would be possible using the best available commercial microwave amplifiers.

We present background motivations and theoretical underpinnings to the axion dark matter search at a level appropriate for the graduate student first entering the field. We also include a substantial discussion of practical considerations and methods intended as a "quick-start" guide for experimentalists. Our guiding principle is to provide the handbook

we wish we had owned as a graduate student in physics first approaching the dark matter search, axion physics, and MSA design.

We demonstrate MSAs operating at frequencies as low as 350 MHz and as high as 1,160 MHz with a typical Q of 60, typical gain of 20 dB, and typical tunability range of about about 20% of the center frequency. We use GaAs varactor diodes to tune the MSAs, and demonstrate a surprising transition in varactor capacitance between about 2 K and 200 mK, which we believe has not been reported to date and may shed light on the theory of shallow negative U centers in semiconductors.

Contents

List of Figures

List of Tables

Acknowledgments

This work represents much more than the effort of one person, and though it is not possible to give proper thanks and credit to everyone who contributed, I'd like to make an incomplete gesture of gratitude by thanking the following people:

John Clarke, my mentor through this project, who made everything I accomplished as a graduate student possible.

Irfan Siddiqi and his team at QNL, who were my reliable "lab next door" always available to help talk through a difficult idea or lend a helping hand.

Jørn Bindslev Hansen, who spent summers visiting from the Technical University of Denmark was always a joy to work with and is the very picture of a gentleman and a scholar.

Gene Hilton at NIST Boulder, who fabricated every device we designed, for free, no less. The entire project of creating MSAs for ADMX would not have been possible without his expertise and generosity.

In the spirit of writing this dissertation as a helpful guide to future graduate students[1], we would like to recommend the following free tools and acknowledge their creators. We would encourage anyone writing their own dissertation or other technical works to explore these resources.

- **Overleaf.** Overleaf is a web-based LaTex editor that allows almost-WYSIWYG ("what you see is what you get") composition, and includes sharing and collaboration tools. This dissertation was written in Overleaf.
- **CodeCogs Equation Editor.** This online LaTex equation editor (`https://www.codecogs.com/latex/eqneditor.php`) allows for quick equation editing, so you can write and evaluate even complex math equations, saving the need to re-compile your manuscript to see if it formats correctly.
- **Tables Generator.** Similar to equation editor, tables generator (`https://www.tablesgenerator.com/`) allows for easy and powerful table creation in LaTex, allowing you to previewed and edit tables outside of your main tex project, and be dropped in when ready.
- **Citation Machine.** Citation machine helps complete and format citations in the needed style. This is especially useful for book citations when you may have only partial information available.
- **Detexify** Detexify is a tool that translates your hand-drawn symbols into Tex math commands. (`http://detexify.kirelabs.org/classify.html`) Very useful for finding obscure symbol commands.

[1] A quixotic sentiment, to be sure.

Chapter 1

Introduction & Overview

More than eighty percent of the matter of the universe is something unknown to science. To date, everything we know about this matter is through indirect (though compelling) evidence. Since it has to date never been seen or otherwise directly detected, we call it "dark matter"[1]. As this manuscript is receiving its final revisions in the summer of 2019, a group of scientists at the University of Washington is preparing, refitting, and upgrading a decades-long experiment designed to pull a tone from the ether that could tell us exactly what dark matter is, and open a new chapter in how humans understand the universe. The project is called the Axion Dark Matter eXperiemnt (ADMX), and the detector (they prefer to call it a "haloscope") is remarkably similar to an AM radio, except more so–more massive, more sensitive, more expensive, and much higher maintenance. They tune carefully and slowly across the spectrum, not knowing where the signal might be found, or exactly what it will sound like, but some speculations bear listening[1].

The idea that we could measure the composition of the universe, much less make a scientific study of its origins, shape, and ultimate fate, would have likely sounded impossible to a learned person at the turn of the 20th century. The seeds of the contemporary search for dark matter might have been known to our imaginary scholar if they were reading contemporary accounts of the first observations of superconductivity, but we may forgive them for failing to see where that discovery might lead. ADMX is enabled by superconductivity both in the construction of the large magnet necessary for catching and converting dark matter and in the operation of its primary amplifier–both fantastic power and fantastic finesse is needed to push the boundaries of knowledge.

[1]Dark matter is not black, as in reflecting, scattering, and transmitting no light, but is transparent, or rather, invisible. In pursuit of accuracy, we could call it "transparent matter" or "invisible matter", though those names have less poetic appeal. We will follow convention in calling it dark matter.

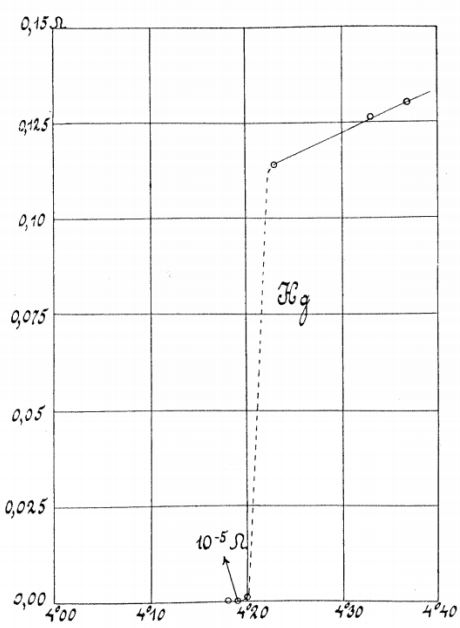

Figure 1.1: Superconducting transition of Hg observed in 1911. Reproduced from [3].

1.1 Superconductivity

On July 10, 1908 the Dutch physicist Heike Kamerlingh Onnes first liquefied helium and opened the door to a new era a low-temperature physics. Three years later on April 8, 1911 in the same lab in Leiden, Onnes and his associates observed a vanishingly small electrical resistance in mercury cooled to 3 K, initially placing an upper bound on the resistivity at 3 K of 10^{-4} times the extrapolated resistance of crystalline mercury at 0°C. [2] Despite the exciting nature of this result, Onnes was careful to avoid overstatement, writing in his first published work on the matter:

> The results just given for the resistance of mercury are, since they are founded upon a single experiment, communicated with all reserve.

In further experiments the transition to very low resistance proved to be quite abrupt, as shown in the famous figure published that same year in November [3]. (reproduced here as figure 1.1) This abrupt transition was not predicted by any theories of electrical conductivity at the time and was the first hint that this "supra-conductive" (as it was called at the time) behavior was not merely novel, but could be a window into wholly new physics.

To accurately measure just how low this resistance is, one can imagine forming a closed loop of superconductor supporting a long-lived current, and measuring the current indirectly and losslessly by measuring the induced magnetic field. A loop with resistance R and inductance L will exhibit an exponential decay in current with a time constant of $\tau = L/R$, so

with a known L and a measured τ, R could be determined. A normal metal loop of reasonable dimensions might have a $\tau = 10^{-10}$ seconds, but a loop with sufficiently low R could conceivably have a τ large enough to be measured with a stopwatch. Later experiments of exactly this nature were performed, first by Onnes in 1914 [4], and many years later by others, in one case observing a small loop of inductance around 5 nH for 3 hours [5], and in another case monitoring a current of several hundred amps over two and a half years [6]. These experiments showed zero decay to within the bounds of instrumentation error, and reduced the upper bound of the ratios of the "superconducting" resistance to the "normal" resistance R_S/R_N to less than 2×10^{-16}. One might forgive the scientist who terminates an experiment after 30 months of nothing happening, having bettered the upper bound set by the initial discovery by 12 orders of magnitude. Later experiments [3, 7] proved the result was no fluke, and in time both our theoretical understanding and experimental tests evolved to the understanding that the resistance is not just vanishingly small but is identically zero.

What was at the time an unexpected curiosity of experimental physics motivated a new and deeper understanding of condensed matter physics, and came to enable un-imagined technologies–to take a few examples, the power to non-invasively image the inside the living human body, to image both nano-scale magnetic phenomenon and kilometer-sized features deep inside the earth's crust, to re-create the nuclear reactions of the solar core in the lab, and to peer to the very birth of the universe. We use superconducting technology in *this* dissertation to investigate the constituency of the universe. In a very real sense, only 5% of the stuff of the universe is understood, and we seek to identify the rest. It is heartening to the experimentalist who has no idea how what their work could possibly impact science or humanity at large to think of Onnes and his low-temperature lab. He certainly had no idea what impact *his* work would have!

1.2 The Axion Dark Matter Candidate

Modern cosmology tells us that ordinary matter accounts for just less than 5% of matter-energy[2] (stuff) in the universe. About 27% of all stuff is made of something invisible that dominates gravitational interactions and kinematics at large scales.[3] At the very largest scales the motion of matter in the universe is dominated by an expansion that is *accelerating*, which is inconsistent with always-attractive gravity, but well explained by a pervasive energy with negative pressure associated with space itself. This pervasive energy is called dark energy, and accounts for 68% of the stuff of the universe. Figure 1.2 shows this concept of

[2] Anything that exists is made of matter, energy, or some combination of the two. Let's just call it "stuff".

[3] At risk of profound understatement, we assert that space is big, and mostly empty. In regions unusually rich in ordinary matter, like our solar system (even accounting for all the empty space between planets, our solar system is unusually rich in ordinary matter) ordinary matter dominates gravitational interactions and kinematics. Averaging over the scale of galaxies, galaxy clusters, and larger, dark matter is five times more common than ordinary matter.

the universe as small lamps of visible matter attached to invisible ships of dark matter afloat on a sea of dark energy.

The nature of dark matter is subject to a great deal of both speculation and scientific investigation. Most models propose some kind of exotic particle created in the early universe alongside ordinary matter that interacts gravitationally with itself and ordinary matter, but nearly lacks any other method of interaction. Of course if gravity were the *only* interaction between dark matter and ordinary matter, if would be hopeless to attempt a lab-based direct detection with instruments made of ordinary matter. Fortunately, particle physics, despite its renowned success, is not a hermetically self-contained and complete theory, but has its own open questions and mysteries. Some of the outstanding questions suggest extensions and/or modifications to existing theory that result in new particles that could be the dark matter we seek, and often those hypothetical particles have weak but extant interactions with ordinary matter. One such mystery and possible solution is the "strong CP problem" and the axion. This extension of existing particle theory tells us the properties of the hypothetical axion, which guides our design of an experiment to detect the axion particles that may be the dark matter continuously surrounding and streaming through us. That experiment is ADMX.

We reserve for others the glory of explicating and demonstrating the nature of dark energy.

1.3 The ADMX Haloscope

The ADMX haloscope is named because it is meant to give us a new kind of vision tuned to the dark matter halo of our galaxy. Much as a telescope unlocks the visibility of things very far away, or a microscope makes visible the very small, the haloscope is meant to make visible the otherwise invisible galactic dark matter halo in which we all reside.

Axions are detected by their conversion to photons. An axion decaying directly to photons is forbidden, but it may decay to two photons via the second-order process of an anomalous fermionic loop. Since direct decay to photons is not possible spontaneous axion decay is exceedingly rare, and the overwhelming fraction of any axions created in the big bang are still around today. To encourage conversion of axions to photons via the indirect fermionic loop process, one of the product photons may be supplied as a virtual photon associated with a strong static magnetic or electric field. In this kind of enhanced convertion, 100% of the axion mass-energy is converted to a single real photon, and the conversion rate is proportional to the square of the static field strength. Because the conversion rate is also proportional to the total density of available final states (Fermi's golden rule), if this conversion occurs in a cavity tuned to a frequency matching that of the emitted photon the conversion rate is enhanced by the Q of that cavity. ADMX is a resonant microwave structure immersed in a strong magnetic field (as strong as practicality and pocketbooks allow, about 7-8 T), and a very sensitive radio receiver poised to detect the occasional extra photon. To reduce the background level of thermal photons, the resonator is cooled to as low a temperature

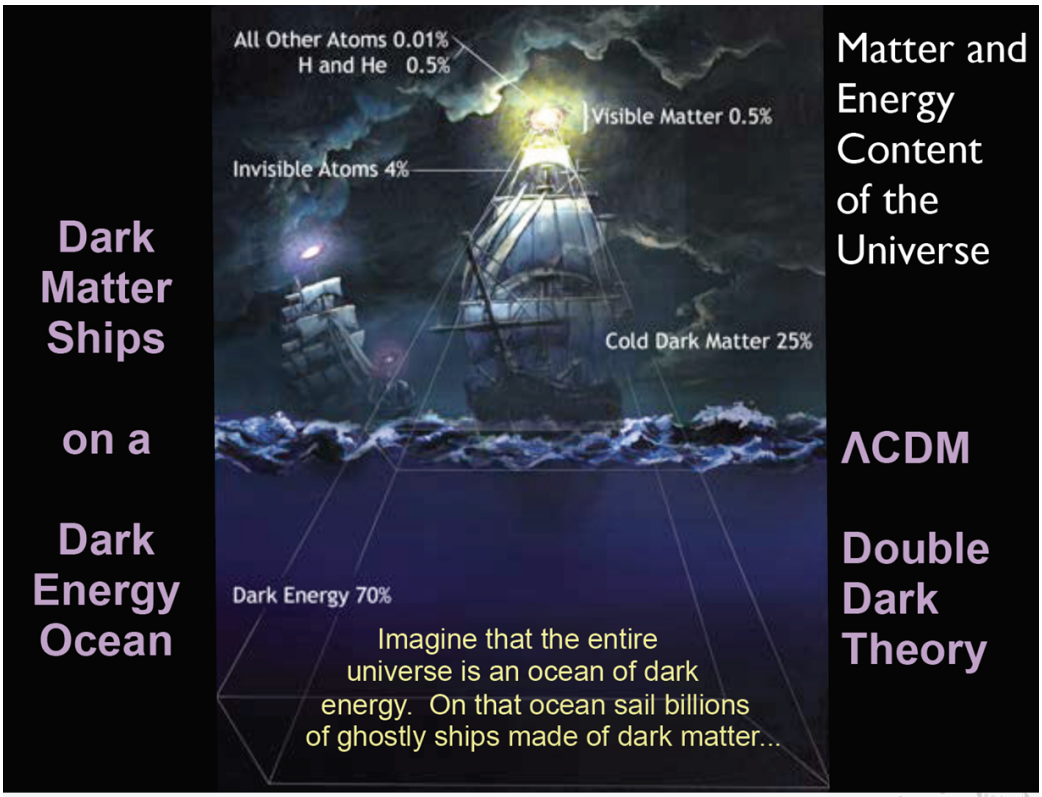

Figure 1.2: Composition of our "Dark" Universe. As we look out into the cosmos, what we see can be likened to looking over a dark ocean on a moonless night. Upon that dark ocean are otherwise invisible ships, each with a lantern on the topmast. Matter we can directly see (for example, stars, quasars, etc) account for about 0.5% of mass in the universe, corresponding to the lanterns on the ships. Other ordinary matter (mostly diffuse hydrogen gas) we can infer from absorption spectra or other subtle interaction with luminous matter, corresponding to the top sheets of the ships that may reflect or obscure light from the lanterns. All ordinary matter together adds to about 5% of the universe. The rest of the ship is utterly invisible, but we can tell from the way the lamps move that they are connected to a large mass rolling and bobbing on the ocean. By observing the movement of the lanterns, their typical rocking frequency, etc. we can infer the mass of the ships, much as we can measure the mass of a galaxy by studying the kinematics of the stars within it. The ships in this analogy are the dark matter of our cosmos, and account for about five times more mass than all the ordinary matter in the universe. All these islands of dark matter adorned with visible matter float on an ocean more massive than the ships themselves. The ocean is pervasive and continuous, filling all space without any regions of higher or lower concentration. This ocean corresponds to dark energy, the most massive and least understood component of the universe, accounting for about 68% of what exists. This model is called ΛCDM, where Λ is the traditional symbol for a static ocean of dark energy, and "CDM" means "cold dark matter", cold because it clumps into galaxies, rather than stream freely through space. Reproduced from [8].

as possible, again constrained only by cost and practicality. ADMX is cooled by a ^{3}He dilution fridge to a temperature of about 100 mK. The mass of the axion and frequency of the emitted photon are not known, so the cavity must be tunable over a wide range of frequencies. Cosmological and astronomical observations suggest the axion mass lies between 1 μeV and 1 meV, or 240 MHz to 240 GHz in photon frequency. ADMX is starting the search at the low-mass range, at frequencies around 1 GHz. The photon detector is a microwave amplifier coupled to the resonant cavity, and any amplifier adds noise comparable to random noise from a body at finite temperature. The use of an amplifier with a low noise temperature is as critical to the operation of ADMX as is operating at low physical temperature. The titular Microstrip SQUID Amplifier (MSA) is used for this purpose, and the development of the MSA for use in ADMX is the major achievement of this work.

1.4 The Microstrip SQUID Amplifier

The MSA was developed to fulfill the need for a low-noise amplifier for ADMX. The scan rate in ADMX is proportional to T^{-2}, where T is the sum of the physical temperature and amplifier noise temperature. The best commercially available amplifiers are based on High Electron Mobility Transistors (commonly abbreviated to HEMT) which have a noise temperature of about 2 K. The MSA is based on the DC SQUID, which is exquisitely sensitive to magnetic flux. By coupling the magnetic field of an electromagnetic wave to the SQUID, amplification at noise temperatures around 50 mK can be achieved. ADMX requires a tunable amplifier, and a tunable MSA that operates at a frequency around 1 GHz had not been demonstrated prior to the work described here.

The MSAs used in ADMX have enabled an axion search of unprecedented sensitivity and breadth in axion mass.

1.5 Structure of this Dissertation

- In chapter 1 we give a brief overview of the major concepts in this work.
- In chapter 2 we introduce the evidence and motivation for dark matter based on astronomical and cosmological considerations.
- In chapter 3 we introduce the strong CP problem and the axion as a possible solution. We show that the axion is also a strongly motivated candidate for the constituency of dark matter, and describe the ADMX haloscope currently operating at the University of Washington.
- In chapter 4 we introduce superconductivity, Josephson junctions, the principles of operation behind the DC SQUID and the MSA, and MSAs produced prior to our work.

- In chapter 5 we describe the design considerations and optimizations for the MSAs we developed.
- In chapter 6 we describe special techniques, custom hardware, and best practices developed in support of our goals of delivering MSAs to ADMX.
- In chapter 7 we present our experimental results, showing specific performance data of the MSAs we developed.
- In chapter 8 we give a summary and conclusion to the work described, and suggest directions for future work.

Chapter 2

Modern Cosmology and Dark Matter

Two among many of the great outstanding mysteries in science are the questions: "What is the composition of the universe?" and "Why is the strong nuclear force highly symmetric?" These two questions are commonly called the "Dark Matter Problem" and the "Strong CP Problem" respectively. This chapter explores the dark matter problem, and the strong CP problem is explored in chapter 3. After introducing both problems, we will show that a hypothesized particle called the Axion is a very strong candidate to solve both mysteries.

It is the task and pleasure of scientists to explore such questions and find explanations that withstand the closest scrutiny and also expand our understanding, whether by leaps or small increments, to previously unimagined horizons. The quality of explanations is measured first by agreement with observations, especially when predicting previously unobserved phenomena, and second by "conceptual economy" that is, the ability to explain much by saying little.

The question "What is the composition of the universe?" seems at first to be an uninteresting one that can be answered and dismissed with a casual gesture towards a wall hanging of the periodic table of the elements likely to be sharing company with those so touched as to ask such things. However, the observable universe is a big place, and it is not fair to extrapolate our direct experience localized to the immediate neighborhood of the surface of Earth to the whole cosmos. The task of accounting for the composition of the universe requires the tools of astronomy. There is a remarkable amount of evidence available to us in the form of visible light and other radiation from both near and places so distant in time and space that words are insufficient to express the magnitudes. The utterly surprising truth indicated by the evidence of astronomy is that any stuff you can place on the periodic table accounts for no more than 5% of the universe. No more than 30% of the universe is even *matter*, with the remainder conventionally called "dark energy". The portion of matter in the universe with no place on the periodic table (non-baryonic matter) we call "Dark Matter".

The term "Dark Matter" is a definite-sounding term invented to reflect the fact that the human scientific project to date can make no claim regarding what about 85% of the mass of the universe *is*. If one sees ignorance as shameful, and the mission of science as the quest for answers, then one might admit that despite the great success of science for the last few

hundred years, the current state of affairs is quite embarrassing indeed. On the other hand, if one sees ignorance as an opportunity to gain wisdom, and the mission of science as the quest for ever deeper, more challenging, and more interesting questions (and an endeavor that generates answers almost as a byproduct), then the current state of affairs, finding ourselves able to pose the question "What is the universe made of?" in a meaningful way accessible to current experiments, is truly an opportunity without historical precedent.

2.1 Early Cosmology

In the early 1900's it was widely accepted in western scientific circles that the universe was geometrically flat, infinite, and static. This hypothesis was consistent with available astronomical observations, but not well tested. The support for this cosmological model was at least as much due to aesthetic, philosophical, or metaphysical preference as support from evidence. Eventually Eisenstein's theory of general relativity and observations by Slipher[9], Wirtz [10], and Hubble[11] indicating a universal expansion would necessitate both re-examination of the nature of gravity and the steady-state universe, respectively. However, even before that time Newton's theory of gravity was incompatible with the steady-state universe, allowing only for a universe that is ever-expanding, *momentarily* static, or collapsing, in close analogy to a projectile escaping a gravitational body at greater than or equal to escape velocity, a projectile at its apoapsis, or a falling ballistic projectile. This inconsistency was well known to the scientist of the early 1900's who favored the static universe. One less than satisfactory resolution to this would be to suggest that we find ourselves at a unique time is history in which the universe is currently very close to static and at maximum extent (like a projectile near apoapsis), with collapse in the immediate future but at dynamical timescales inaccessible to current observational abilities. Another less than satisfactory resolution would be to suggest that Newton's laws apply to the motions of starts and planets, but not to universal evolution. In any case, debate about the state of the universe at times sufficiently far from the present, either in the past of future, was, in the absence of relevant observational data, rightly discussed as a metaphysical concern, rather than as the concern of science.

When Einstein formulated general relativity, describing the relationship of mass, energy, and the geometry of space and time, the nature of mass, energy, space, time, *and* gravity were all up for dramatic re-evaluation. Einstein's field equations can be written in tensor form as

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = \frac{8\pi G}{c^4} T_{\mu\nu} - g_{\mu\nu}\Lambda, \tag{2.1}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, $g_{\mu\nu}$ is the metric tensor, $\mathcal{R}$ is the Ricci scalar curvature, G is the gravitational constant from Newton's theory of gravity, c is the speed of light in vacuum, and $T_{\mu\nu}$ is the stress-energy tensor of all physical fields (mass, radiation,

etc). The final term Λ deserves special mention: it was named by Einstein the "cosmological constant".

The left side of equation 2.1 describes the curvature of space-time, which one would associate with the "force of gravity" in Newton's theory, and the right side describes the distribution of mass or energy (which can also appear as space-time stress), which one could associate with the distribution of gravitating mass in Newton's theory. The final term "Λ" in equation 2.1 has no analogy in Newton's theory, but might be called an "anti-gravity" term, a bit like upsidaisium (a rare mineral reported by [Rocky, Bullwinkle, 1960] that falls *up*), except that it pervades all space at constant density.

This Λ term was added by Einstein to resolve the static universe problem of Newtonian gravity. Einstein's equations of general relativity accurately reproduce Newtonian dynamics and gravity, explain phenomena unexplained by Newtonian gravity (ex: the precession of Mercury's orbit), and predict new phenomena, confirmed sooner or later experimentally (ex: gravitational lensing and frame-dragging), but without Λ the same old Newtonian problem of a universe only allowed to be expanding or contracting (not static) remains. Adding the "anti-gravity" or "self-repulsive" Λ term allows for an eternal static universe without impact on the other predictions of GR. However, the static universe allowed by the Λ term is unstable: self-gravitation correlates with mass-density but Λ is constant, so in Einstein's static eternal universe, in a slightly over-dense region self-gravity is stronger than self-repulsion so the region shrinks, increasing in density, and accelerating the collapse. In a slightly under-dense region, self-repulsion beats self-attraction, leading to slight expansion and lower density, thus lower self-attraction, and so asymptotically towards exponential expansion. The Λ term had no other motivation or consequences at the time. It was called a "blunder" by Einstein [12] when later observational evidence around 1930 showed that our universe is not static, but uniformly expanding. The Λ term was mostly abandoned until about 70 years later, when the observation that the universe is not only expanding, but *accelerating* (and a Nobel prize awarded in 2011) brought new reason to expect a self-repulsive term.

2.2 Modern Cosmology

The story of our universe as it is understood today is summarized neatly in figure 2.1. The universe today is expanding uniformly and about 13.7 billion years in the past, all of space was collected to a very hot, dense singularity. This origin is called the Big Bang. What happened before that is arguably not accessible to empirical science. The universe after the big bang can be categorized into 5 eras, Inflation, Q-G Plasma, Atomic Plasma, Dark Age, and the Modern Age.

The first 10^{-32} s after the big bang was a period of inflation, with the dominant mass-energy term being an "inflaton" scalar field that behaves similarly to the Λ term of equation 2.1, resulting in an exponential rate of universal expansion. The "inflaton" field is a false vacuum, and during this inflationary epoch the inflaton field "slow rolls" towards the true

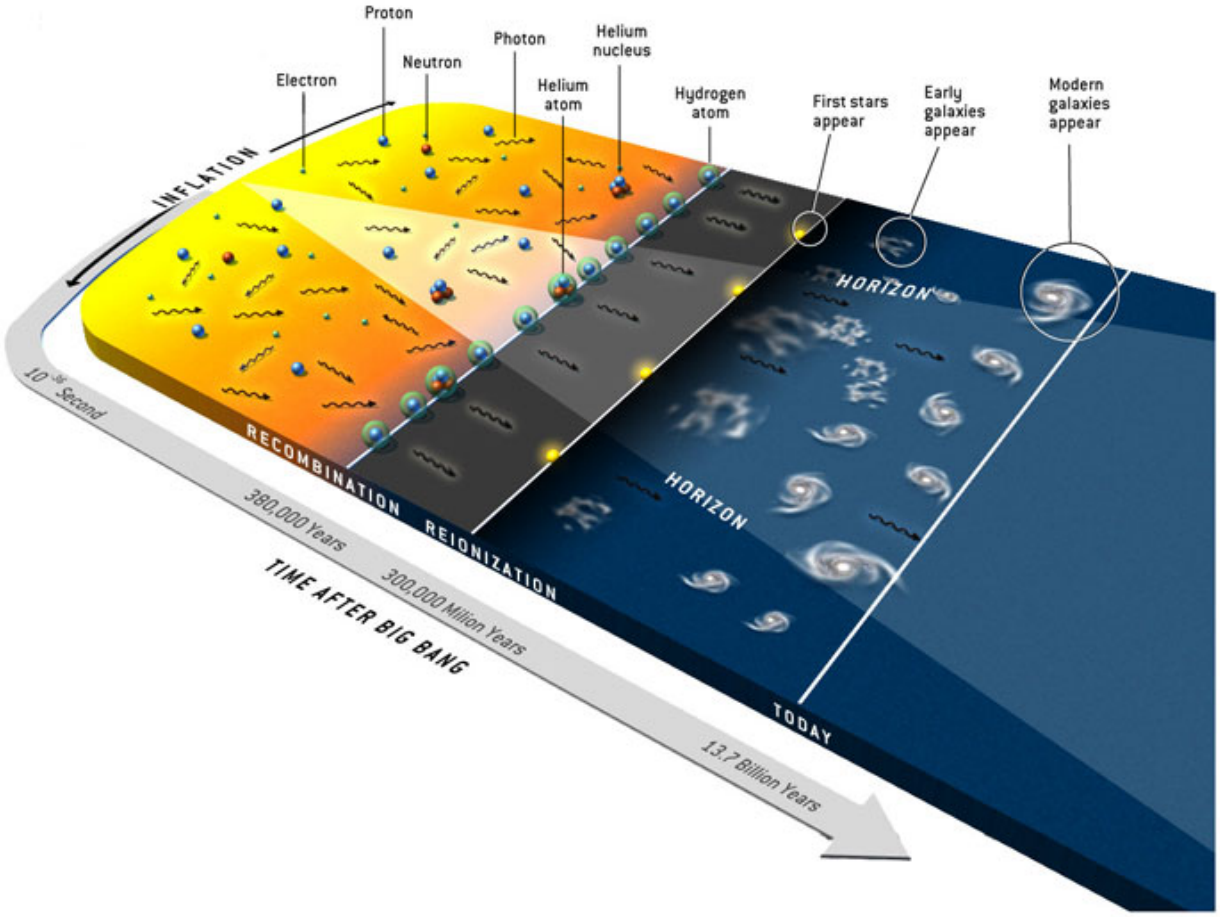

Figure 2.1: Summary History of the Universe. The story so far includes: the inflation age (leading edge at top left), the quark-gluon plasma age and atomic plasma age (yellow region), the Dark age (grey band), and the Modern age (blue). Reproduced from [13]

vacuum (lowest energy) state. Slow roll here means many exponential factors of expansion occur before the inflaton field reaches true vacuum. Any initial volume of space expands to something like e^{100} or 10^{43} times its original size before the inflaton field finds its true vacuum state and ends the era of exponential expansion. Any initial conditions existing before inflation are erased by the end of inflation, as any pre-inflation mass density has been diluted by a factor of 10^{130}, and any energy density or spatial inhomogeneity has been similarly diluted. When the inflaton field reaches, overshoots, and oscillates around the true vacuum state (energy minimum), those oscillations correspond to a spatially coherent bath of "inflaton particles". The inflaton particles decay quickly into every other fundamental particle and force carrier, leaving a perfectly smooth (and very hot) temperature distribution with tiny density variations resulting from quantum fluctuations of the inflaton field. The exhaustion of the inflaton field and reheating due to inflaton particle decay marks the end of the inflationary period.

The next era of the universe is the "Quark-Gluon plasma" era, lasting between 10^{-12} and 10^{-6} s. In this era, temperatures are on the scale of 100 MeV, too hot and dense for stable atomic nuclei to form. Most of the mass-energy of the universe is in the form of photons, with just a light dusting of quarks, leptons, and other fundamental particles. Universal expansion continues, but is slowing due the mutual gravitational attraction of all mass-energy.

Once temperatures and densities fall sufficiently to form stable protons and neutrons[1] there was never again a free quark, and thus began the age of the "Atomic Plasma", lasting 380,000 years. About 10 seconds into this age, the universe cooled sufficiently to allow the formation of stable light atomic nuclei, including ^{2}H, ^{3}He, ^{4}He, and ^{7}Li, in a process known as Big Bang Nucleosynthesis (BBN), which lasted about 20 minutes until further cooling and the decay of free neutrons made further nuclear fusion impossible. The next 380,000 years is relatively uneventful–during this era the universe is extremely homogeneous and non-dynamical, except for some tiny acoustic density-temperature (pressure) oscillations seeded by small density inhomogeneities left over from the inflationary era. The mass-energy of photons is greater than that of matter by a factor of around 10^9, but the universe was nonetheless opaque due to scattering from charged, free electrons and atomic nuclei. This era ended in an event called "recombination"[2] when the temperature cooled sufficiently to allow negative electrons and positive atomic nuclei to combine into stable, electrically neutral atoms (mostly hydrogen gas). Matter and light decoupled almost entirely, and the universe became transparent.

The 300,000 million years following recombination was truly a Dark Age. The only existing light was the left over radiation last scattered from the hot plasma of the previous age, now streaming through space and becoming dimmer and cooler due to continued universal expansion. Within 3 million years universal expansion had redshifted this light to

[1] If one is inclined to argue that an isolated neutron is not stable, we are inclined to note that the neutron half-life is about 10 minutes, whereas the universe at this point is no more than 10^{-6} s old. A particle with a half life greater than 10^8 times greater than the age of the universe may be called *stable*.

[2] This should perhaps be called just "combination" because it is the first time it had ever happened in the history of the universe, but here we will follow the usual convention and call it *re*combination.

the infrared, and the universe was devoid of visible light. This light exists today, redshifted beyond infrared to microwave, and is now known as the Cosmic Microwave Background (CMB). Expansion reduces the energy density of matter by R^3 (simple reduction in density) but reduces the density of radiation by R^4 (lower density *and* redshifting), so during this dark age the mass-energy of the universe came to be dominated by matter, rather than light. Silently, invisibly, and slowly, regions of space where the neutral hydrogen gas was slightly denser than average due to residue of the small acoustical vibrations of the Atomic Plasma age collapsed and densified under the influence of gravity. These regions would evolve in the next age into the galaxies and other large-scale structures of the universe. Though most gas remained unbound, the densest regions formed gravitational wells deep enough to allow the formation of the first stars. Because radiation from these first stars re-ionized the remaining neutral hydrogen gas, the end of this era is known as "reionization".

The time since re-ionization until today, 13.7 billion years since the big bang, might be called the "Modern Era". It is characterized by continuing star formation, and the formation of larger-scale structures, including galaxies and galaxy clusters, though most of the baryonic matter originating from the era of the Atomic Plasma still exists as interstellar or intergalactic gas. Elements heavier than Lithium now exist due to nuclear fusion in stars, and those heavier elements sometimes condense to form planets, interesting chemical compounds, and cosmologists.

The story to this point has focused on baryonic matter, because that's what we can see, either directly from starlight, or indirectly from its interaction with other light from stars, quasars, or CMB for example. The state and condition of baryonic matter inspires the delineation of eras of the universe used here, and though this is useful framework, it should not be mistaken for the while story. About 85% of mass in the universe has not been mentioned, but its effects can be seen from several independent lines of evidence, with each kind of evidence corresponding roughly to the dynamics of each era listed here.

2.2.1 Evidence for the Modern Cosmology

The seeds of modern cosmology were planted in the 1916 by the astronomical observations of Slipher[9], Wirtz [10], and Hubble [11], which showed that for objects outside our own galaxy, the retreating velocity is proportional to distance, as shown in figure 2.2a. Such an observation is consistent with a uniformly expanding universe. While Hubble's original estimate of the expansion rate (Hubble constant H_0) was too high by about 6 times (figure 2.2b), later and much more extensive surveys (figure 2.2c) have cemented the conclusion of an expanding universe.

The fact of universal expansion suggests a universal singularity at some point in the past. To see if this is actually so, one may take by convention a present-day scale factor of any size called a, (1 m or 100 Gpc, for example) and calculate from general relativity how a evolves in time, which we will do here.

The usual assumption is that on sufficiently large scales the universe is homogeneous and isotropic. This may have served merely as a useful simplifying assumption in the early

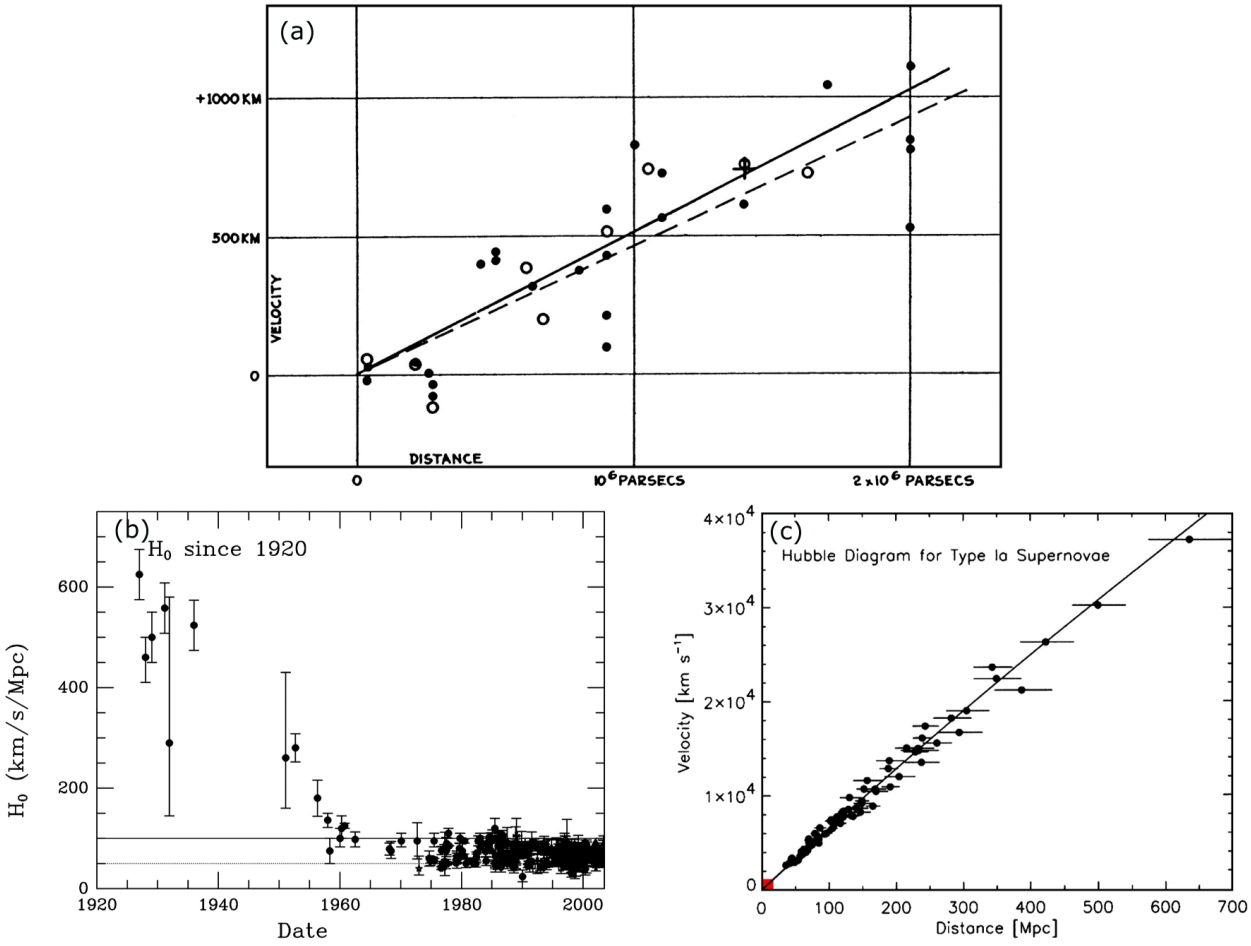

Figure 2.2: The Hubble constant: early and contemporary estimates. (a) Hubble's original diagram. Image reproduced from [11]. (b) Modern estimates of H_0 have converged to about 100 km/s/Mpc, from Hubble's original estimate of 600 km/s/Mpc. (c) Modern estimate of the Hubble constant. The small red square at the bottom left marks the span of Hubble's original diagram from 1929. (b) and (c) reproduced from [14].

days of cosmology, but it is well supported by all observational data to date. Applying this assumption to Einstein's field equations is known as the "Robertson-Walker" metric, and it does wonders to tame the dreaded right hand side of equation 2.1, reducing $T_{\mu\nu}$ to

$$T^{\mu}_{\nu} = \mathrm{diag}(\rho c^2, -p, -p, -p), \tag{2.2}$$

where ρ is the mass density (and/or energy density without the c^2 term) and p is the pressure. The $\mu = 0$ component of the conservation of stress energy ($T^{\mu\nu}{}_{;\nu} = 0$) gives the first law of thermodynamics,

$$d(\rho c^2\, a^3) = -p d(a^3), \tag{2.3}$$

that is, the change in energy in a co-moving volume element, $d(\rho c^2\, a^3)$, is equal to minus the pressure times change in volume, $-pd(a^3)$. The equation of state $p = w\rho c^2$ is sufficient to describe everything of interest to us, with $w = 0$ for ordinary matter, $w = 1/3$ for radiation or relativistic matter, and $w = -1$ for vacuum energy, resulting in:

$$\begin{aligned} \text{MATTER } (p = 0) &\Rightarrow \rho \propto a^{-3} \\ \text{RADIATION } (p = \frac{1}{3}\rho c^2) &\Rightarrow \rho \propto a^{-4} \\ \text{VACUUM ENERGY } (p = -\rho c^2) &\Rightarrow \rho \propto \text{const.} \end{aligned} \tag{2.4}$$

Equation 2.4 is a nice result because once we know the evolution of ρ and p in terms of the scale factor a, equation 2.1 allows us to solve for the time-dependent $a(t)$, which tells us nothing less than the fate of the universe.

From equations 2.1 and 2.2 the Ricci tensor has only the following non-zero components,

$$R_{00} = -3\frac{\ddot{a}}{a}, \tag{2.5}$$

$$R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + \frac{2k}{a^2}\right] g_{ij}, \tag{2.6}$$

where k is the curvature, normalized to a. The Ricci scalar is:

$$\mathcal{R} = -6\left[\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2}\right]. \tag{2.7}$$

Inspection of the $0-0$ components of equation 2.1 gives the first of the Friedmann equations,

$$\frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{8\pi G\rho c^2 + \Lambda}{3}, \tag{2.8}$$

while the $i - i$ component gives

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \Lambda - 8\pi G p. \tag{2.9}$$

Subtracting 2.8 from 2.9 gives the second Friedmann equation

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{4\pi G}{3}(\rho c^2 + 3p). \tag{2.10}$$

Equations 2.8 and 2.10 are usually re-written as

$$H^2 \equiv \frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho c^2 + \Lambda}{3} - \frac{k}{a^2}, \tag{2.11}$$

and

$$qH^2 \equiv -\frac{\ddot{a}}{a} = \frac{4\pi G}{3}(\rho c^2 + 3p) - \frac{\Lambda}{3}, \tag{2.12}$$

where H and q are purely kinematic observables, the "Hubble parameter" and "deceleration parameter" respectively. The Hubble parameter H and deceleration parameter q are understood to be time-dependent in general, and the notation H_0 and q_0 are used to refer to their present-day values. The "Hubble constant" H_0 is precisely what is measured in figure 2.2. The deceleration parameter q was defined with a minus sign because "everybody knew" that $\Lambda = 0$, which implies a negative acceleration, and it is more convenient to discuss a *positive* deceleration. Around 1970, some prominent cosmologists would characterize all of cosmology as the measurement of these two parameters (H_0 and q_0) [15]. The dynamical observables H_0 and q_0 remain key observables in cosmology, but today it is preferred to think in terms of a different two parameters: the energy density of matter and radiation Ω_M, and the energy density of Einstein's cosmological constant, or "dark energy" Ω_Λ.

2.2.2 Two Critical Densities: Ω_M and Ω_Λ

Note the curvature term k in equation 2.11: this term explicitly acknowledges that our universe might not be flat. Of course on the scales accessible to everyday human experience, the space we inhabit appears flat. That is, parallel lines never diverge or cross, the internal angles of a triangle of any size add to exactly 180°, and so on. However, any curved surface will appear nearly flat at small scales, so a belief in a flat universe may merely be provincial, and we must take seriously the possibility of a universal non-zero curvature. Equation 2.11 may be re-arranged in the form of normalized curvature

$$\frac{k}{H^2a^2} = \frac{8\pi G}{3H^2}\rho c^2 + \frac{\Lambda}{3H^2} - 1, \tag{2.13}$$

or

$$\frac{k}{H^2a^2} = \Omega_M + \Omega_\Lambda - 1, \tag{2.14}$$

with

$$\Omega_M = \frac{\rho c^2}{(3H^2/8\pi G)} \text{ , and } \Omega_\Lambda = \frac{\Lambda}{3H^2}. \tag{2.15}$$

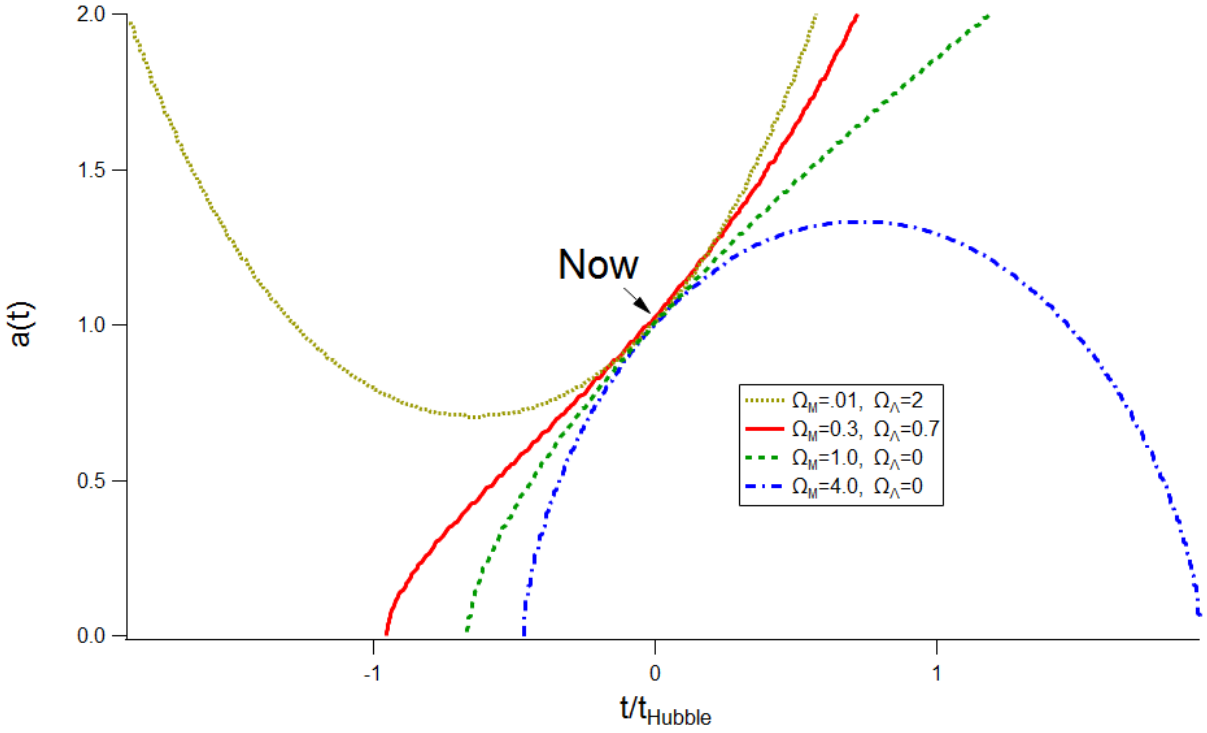

Figure 2.3: Possible histories and futures of the universe. Size factor a(t) is normalized to a(0)=1. Time t is defined such that now is t=0, and time is scaled by the Hubble time $H_0^{-1} \approx 14$ billion years.(ie a tangent line through "Now" intercepts the time axis at $t = -1$). Gold curve is a universe dominated by dark energy so the acceleration is always positive (no big bang). Red is a universe like our own with a 30-70 mix of matter and dark energy–it initially decelerates but expands to an extent that dark energy gains dominance leading to acceleration. Blue is a universe dominated by gravitating mass such that it is decelerating at all times, and has a "big crunch" in its future. Green is a universe with no dark energy and a matter density poised on the balance between eternal expansion and eventual re-collapse– a favorite model among cosmologists prior to the observation of dark energy.

Equation 2.14 shows explicitly that if the total energy density of the universe $\Omega_M + \Omega_\Lambda = 1$, the curvature k is zero and space is flat; $\Omega_M + \Omega_\Lambda > 1$ implies k is positive and the universe wraps back on itself like the surface of a sphere; and $\Omega_M + \Omega_\Lambda < 1$ implies k is negative and the universe is hyperbolic like the surface of a saddle.

Note that equation 2.12 implies that the deceleration term q is positive in a matter-dominated universe (small Λ). Imagine running the clock backwards on cosmic evolution: if H and q are positive now, then back in time both ρ and p were greater (from equation 2.4), so H and q were still greater in the past, and inevitably $a(t) = 0$ at some previous time. This possible "pre-history" is illustrated in figure 2.3 with the green and blue curves.

In figure 2.3 the time is normalized such that $t = 0$ at present and is parameterized in

units of the hubble time $H_0^{-1} \approx 14$ billion years. The scale factor $a(t)$ is normalized such that $a(0) = 1$. A universe that is expanding today with $\Lambda = 0$ inevitably has a "big bang" in its past. The future depends on the value of Ω_M: $\Omega_M > 1$ implies an eventual re-collapse, and $\Omega_M \leq 1$ implies eternal expansion. Large Ω_M means more gravity, slower acceleration, and maybe faster collapse.

In contrast to a $\Lambda = 0$ universe, a universe with a large value for Λ could have a *negative* q, meaning an accelerating expansion rate today and an expansion rate that is progressively smaller in the past, even to the point of $H \leq 0$, as shown with the gold curve in figure 2.3. In this large-Λ universe, there is no singularity and no "big bang". Large Ω_Λ means more "anti-gravity" and an accelerating expansion.

A relatively large value of Ω_Λ, ie $\Omega_\Lambda > \Omega_M$, can still be consistent with a big bang: in the past the universe was smaller, so ρ and p were greater and decelerate more strongly. With sufficient momentum, at some point in the past $4\pi G(\rho c^2 + 3p) > \Lambda$, so by equation 2.12 $q > 0$ and a singularity results, exactly as for a matter-dominated universe. The red curve in figure 2.3 illustrates this scenario, with the values $\Omega_\Lambda = 0.7, \Omega_M = 0.3$.

The geometry, past history, and future fate of a universe with for arbitrary values of Ω_M and Ω_Λ are summarized in figure 2.4. This figure assumes a present-day expanding universe ($H_0 > 0$).

The large-scale geometry of space is determined by $\Omega_M + \Omega_\Lambda$: for $\Omega_M + \Omega_\Lambda = 1$ (blue line) space is perfectly flat. Above the blue line ($\Omega_M + \Omega_\Lambda > 1$, regions **II**, **IV**, and **V**) space has positive curvature and curves back on itself like a sphere, and below the blue line ($\Omega_M + \Omega_\Lambda < 1$, regions **I** and **III**), space is hyperbolic.

The past history of the universe is delineated by the red line: above the red line, for $\Omega_\Lambda > 1$ *and* large values of Ω_Λ relative to Ω_M (region **V**), there is no big bang in the past, but instead a "soft turnaround" represented by the gold curve in figure 2.3. Most imaginable values for Ω_Λ and Ω_M (below the red line) predict a past singularity, or big bang.

The future fate of the universe is delineated by the green line: For positive values of Ω_Λ with not very large Ω_M (above the green line, regions **I**, **II**, and **V**) the universe expands forever, as illustrated by the gold, red and green curves in figure 2.3. For sufficiently large Ω_M and Ω_Λ small or negative (regions **III** and **IV**), the universe re-collapses in the future, illustrated by the blue curve in figure 2.3.

There is much confusion, especially in earlier literature on cosmology, due to unspoken assumptions, such as assuming the curvature k is zero, or assuming $\Lambda = 0$, or both. Some works on cosmology would declare "geometry is destiny" or other similar slogans meaning that a spherical universe will re-collapse and a hyperbolic universe will expand eternally, with a flat universe existing as an edge case. By implication only regions **I** and **IV** in figure 2.4 would be possible. Observations that the universal expansion is accelerating puts our universe well above the green line of figure 2.4, so we must also consider region **II** a real possibility. The nature of Ω_Λ is presently unknown, so negative values cannot be ruled out a priori, and region **III** (hyperbolic universe that will re-collapse) completes the set of possibilities for a universe that is presently expanding.

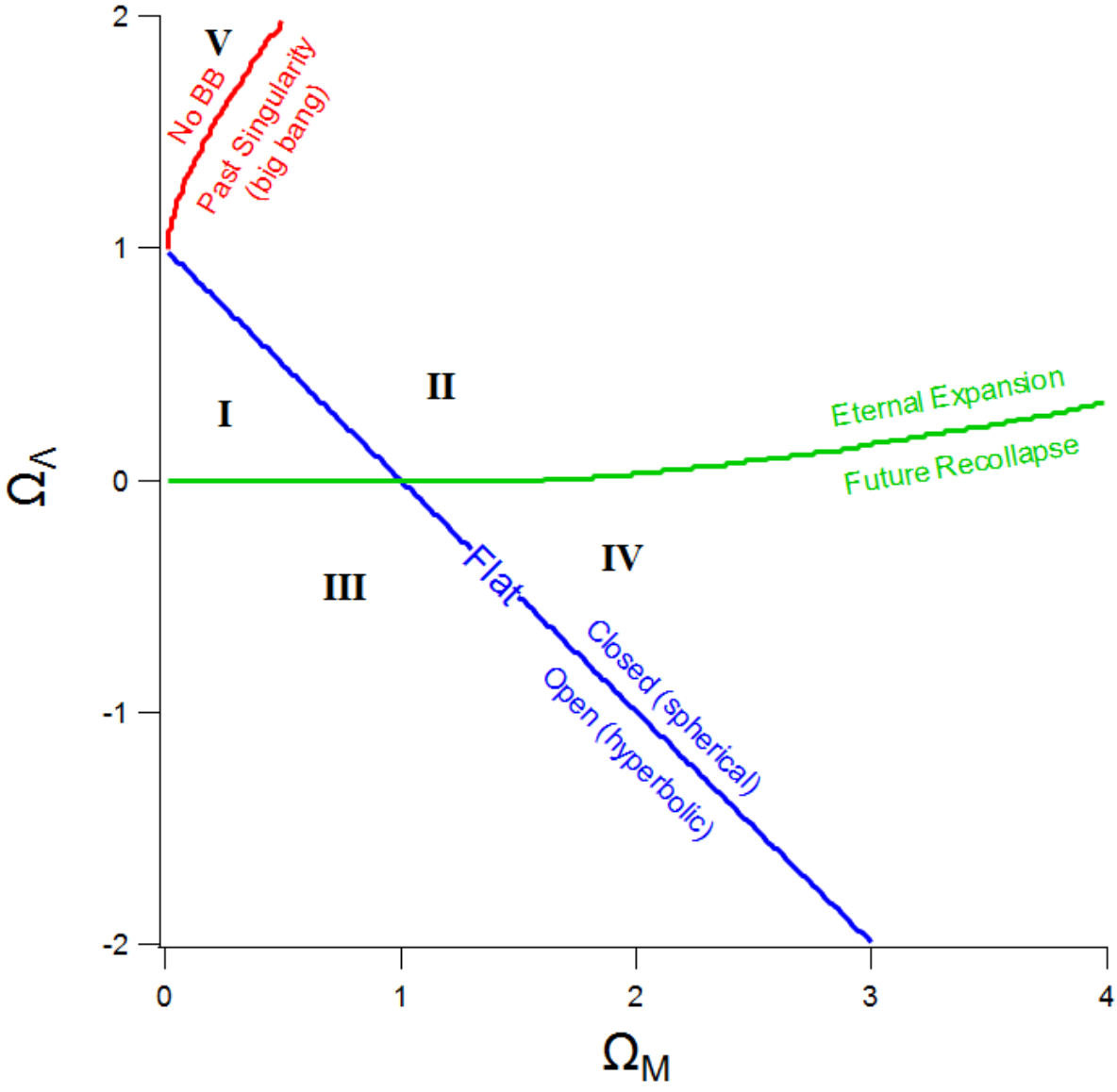

Figure 2.4: Map of Ω_M and Ω_Λ. Red line delineates the past history of the universe (past singularity or soft turnaround). Blue line delineates the shape of the universe (positive or negative curvature). Green line delineates the future fate of the universe (eternal expansion or eventual re-collapse).

2.2.3 Constraints on Ω_M and Ω_Λ from Observational Evidence

To measure what our universe is made of, one must first measure H_0, because that is a scale factor in every other measurement. Next, measure the deceleration parameter q_0–this approximately gives the difference between Ω_M and Ω_Λ. Finally, measure the curvature k, which gives the sum $\Omega_M + \Omega_\Lambda$. These last two measurements are nearly orthogonal, so together they constrain the values of Ω_Λ and Ω_M rather tightly. The deceleration parameter is measured by observations of type Ia supernova (SNe), and curvature is deduced from fluctuations in the cosmic background radiation (CMB). A summary of these measurements is given in figure 2.5, reproduced from [16], which also includes constraints from baryon acoustic oscillations (BAO). Note that figure 2.5 is plotted on a subset of the space of figure 2.4.

The totality of observational data today constrains dark energy to $\Omega_\Lambda = 0.729 \pm 0.014$, curvature to $k = 0 \pm 0.007$, and matter density to $\Omega_M = 0.271 \pm 0.005$ (68% confidence level including systematic errors). This is a remarkable degree of precision and a triumph of observational cosmology. A brief review of CMB, SNe, and BAO evidence follows.

2.2.4 Evidence from the Cosmic Microwave Background (CMB)

The orange region of figure 2.5 puts narrow constraints on the geometry of the universe (ie the sum of Ω_Λ and Ω_M), but says almost nothing about cosmic acceleration (ie the ratio of Ω_Λ and Ω_M).

Observational evidence overwhelmingly supports a cosmological model with a singularity in the past, at which all length scales reduce to zero and mass/energy density goes to infinity. The strongest signature of this Big Bang is the Cosmic Microwave Background (CMB) radiation, a nearly uniform thermal bath of photons with a temperature today of 2.7 K. Though the uniformity of the CMB is arguably its most remarkable feature, it is measurements of the structure of small anisotropies in the CMB (rms variation is 10 μK) that reveal a remarkable amount about the shape and composition of the universe.

The CMB is the afterglow of an early era of the universe about 380,000 years after the big bang, before the formation of either large-scale structure (galaxies) or small-scale structure (stars) when matter and energy were uniformly distributed as a plasma with a temperature of greater than 3,000 K. Expansion and cooling of the universe resulted in the plasma (mostly ionized hydrogen) recombining into neutral hydrogen gas and the mean free path of a photon diverging to infinity. In other words, the universe became suddenly transparent. The last thermal photons emitted from this era (what is called the "surface of last scattering") have been traveling unimpeded for the last 13.7 billion years or so, and except for the exceedingly small fraction that scatter off a bit of ionized gas or hit a solid body like a star or an astronomer's infrared telescope, will continue forever more without interaction.

Though the early universe was very close to isothermal, there existed variations in density due to quantum fluctuations that were "frozen out" at the end of cosmic inflation. Any initial density variation in the early hot plasma leads to oscillations: initially an over-dense

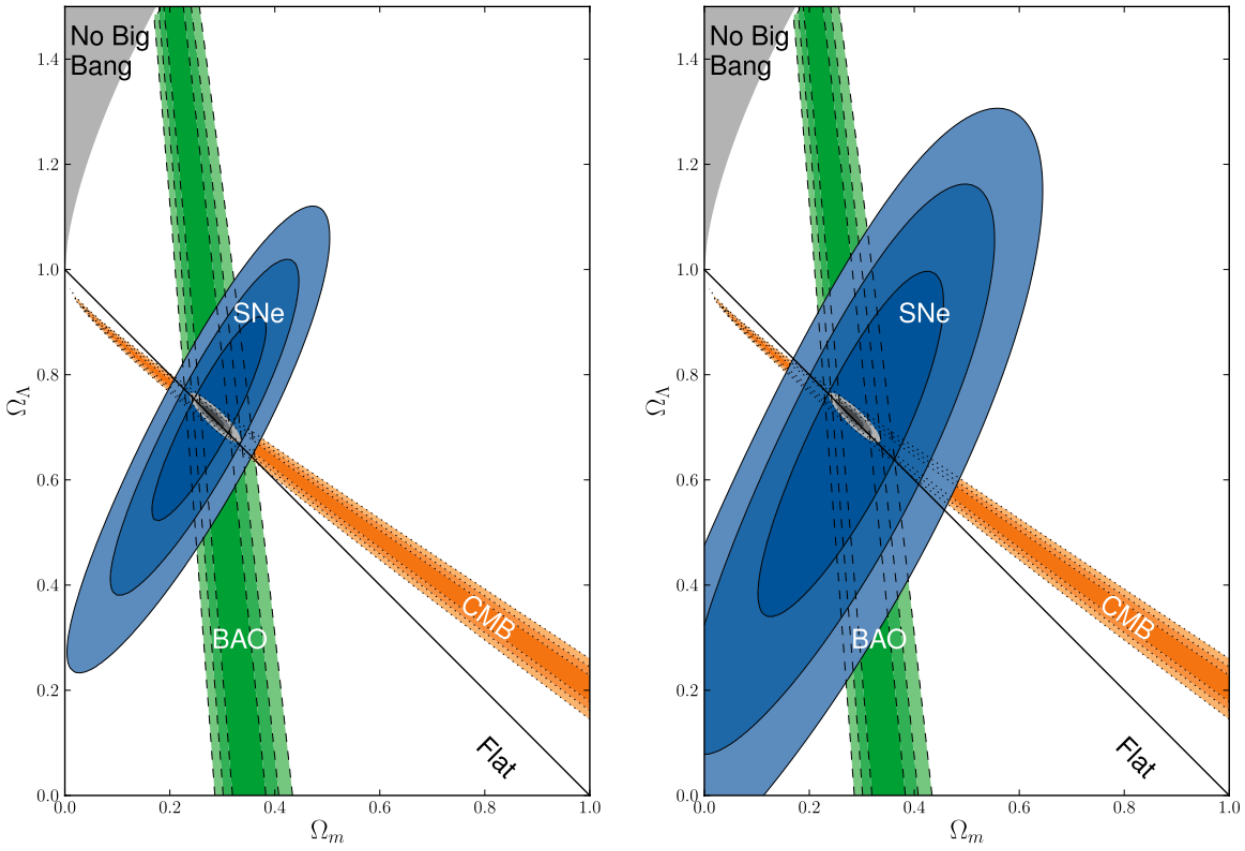

Figure 2.5: Observational constraints on Ω_M and Ω_Λ. Plot shows 68.3%, 95.4%, and 99.7% confidence regions of the $(\Omega_M, \Omega_\Lambda)$ plane from SNe Ia combined with the constraints from BAO and CMB. The left panel shows the SN Ia confidence region only including statistical errors, while the right panel shows the SN Ia confidence region with both statistical and systematic error estimates. Reproduced from [16].

region will collapse due to gravity, but in a plasma matter and radiation are tightly coupled so a collapsing region will get hotter, resulting in greater photon pressure and a restoring bounce, leading to an under-density, and so on. These pressure oscillations continue until the plasma recombines to a neutral gas and the universe becomes transparent. After decoupling matter and radiation, the matter collapses freely under gravity without any restoring force, evolving into the structure of stars and galaxies we know today, while the radiation propagates unchanged for all time, carrying a fossil imprint of density variations existing at the moment of last scattering. Figure 2.6 shows the temperature variations in the CMB visible today.

Note that there is a certain "granularity" to figure 2.6, with a few hundred "grains" from edge to edge. This is not an artifact of image compression or resolution of the telescope,

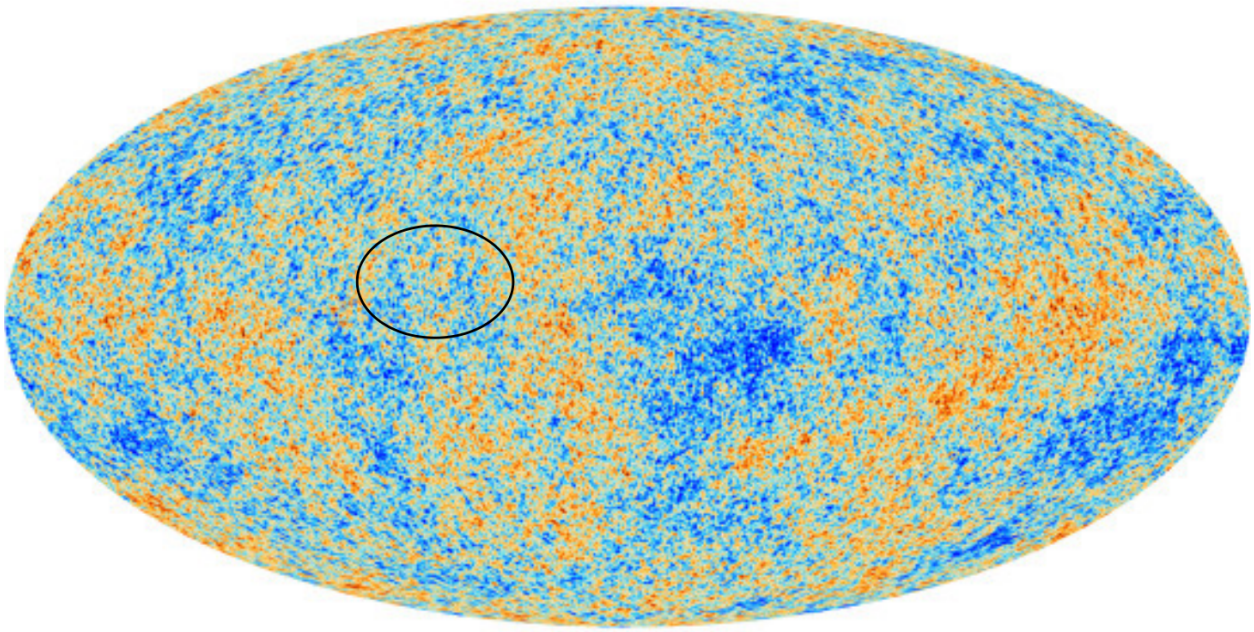

Figure 2.6: Temperature variation of the CMB. Data are from the ESA's Planck space telescope. Warmer regions are colored orange and cooler regions are blue. Stephen Hawking, who invented the universe, left his cool initials "S H" in the CMB, visible here in blue at a position indicated by a black oval. [17].

but represents the real structure of the CMB, and is of great importance. A full resolution image is available at [17]. The initial density variations due to quantum effects have an equal amplitude across all length scales. Divergences from this initial "white" spectrum arise from the acoustic dynamics of the cooling plasma. There was a finite time available for the evolution of these acoustic variations: only about 380,000 years from the big bang until the decoupling of light and matter. Acoustic waves with a period of twice the time until recombination will have just enough time to transform a slightly dense (underdense) region to a slightly hot (cold) one, maximizing the amplitude of variations of this scale imprinted on the CMB. The dynamics of a hot, uniform plasma and the expansion rate of the universe are very well known, so the physical size of this "sound horizon" is also very well known and translates to a particular angular size for temperature variations in the CMB today. A plot of the power spectrum of angular variations in the CMB is shown in figure 2.7.

In figure 2.7 ℓ is the multipole moment with larger ℓ meaning more "wiggles per sphere" with an approximate relation of angular size $\approx 180°/\ell$. Note the prominent peak in figure 2.7 around $\ell = 200$. This corresponds to the granularity of figure 2.6 or a typical feature size of about 0.9° wide (about 2 moon diameters, as seen from Earth). The ℓ-value of this peak puts a rather strict limit on the flatness of the universe, because it represents a

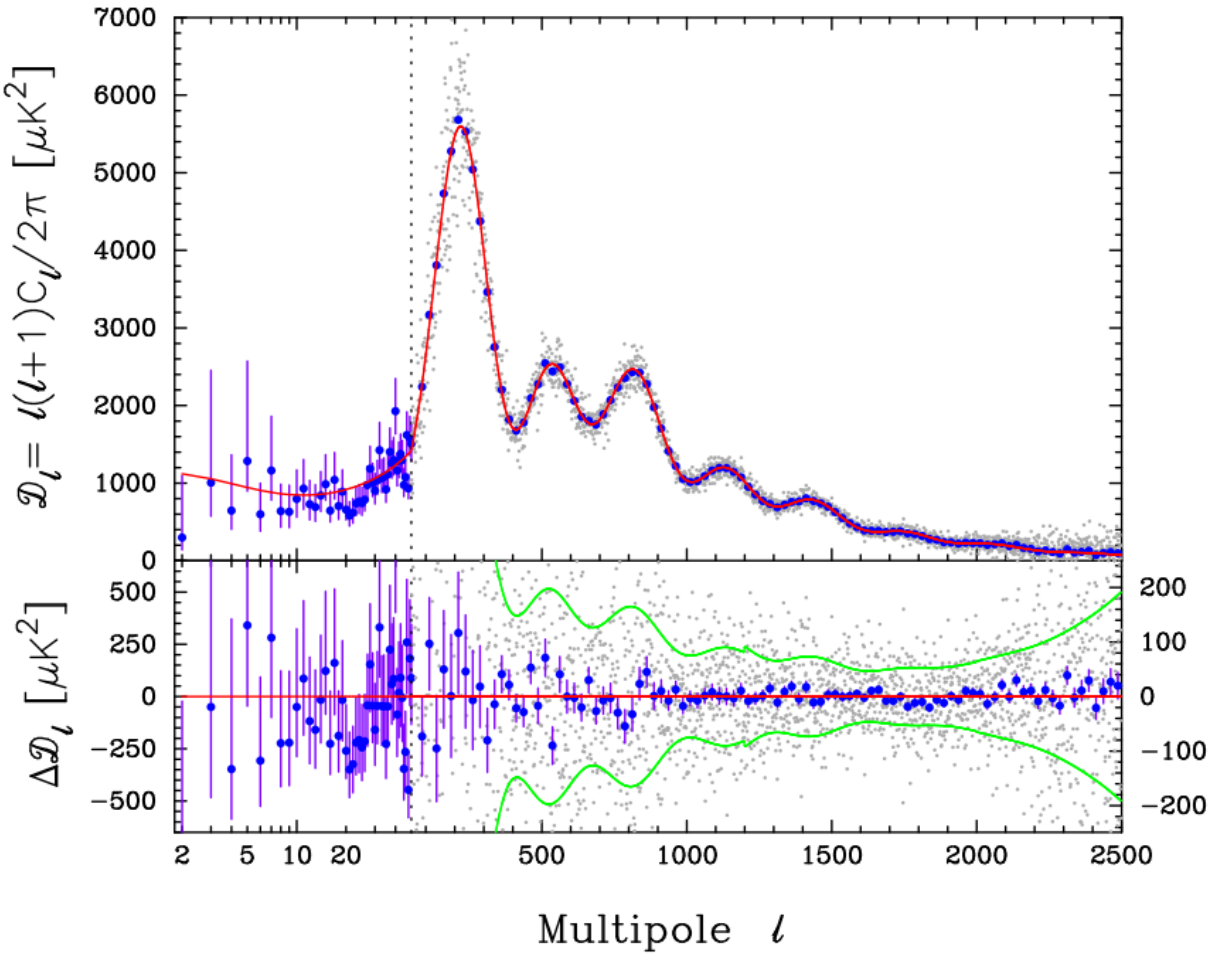

Figure 2.7: Power spectrum of angular temperature variations in the CMB. The position of the first peak indicates a flat universe. The position and height of the second and third peak carry information about the abundance of baryonic matter and dark matter, respectively. Reproduced from [18]

known physical size (the acoustic horizon) at a known distance (time since recombination and redshift). The region size is about 700 thousand light-years, expanded to about 700 million light-years today by cosmic expansion, and the distance is about 45 billion light years. (Although the photons have traveled for only about 14 billion years, the expansion of the universe has elongated their route.) In a flat universe, these are related by the simple equation $sin(\theta) = l/d$ but in a universe with positive curvature, θ will be greater for a given l/d (angles of a triangle add to more than 180°), moving the peak ℓ to lower values, and in a universe with negative curvature θ will appear smaller (angles of a triangle add to less than 180°), moving the peak ℓ to greater values. The observed peak ℓ value puts the flatness very close to zero, illustrated by the orange region of figure 2.5.

There is a great deal more to be learned from the data in figures 2.6 and 2.7. Notably, the relative density of baryonic matter to radiation is encoded in the second ℓ peak, corresponding to the amplitude of acoustic oscillations with period equal to the time until recombination. The ratio of "dark" matter (that is, matter that *never* had a strong coupling to radiation) to baryonic matter (matter that has a charge and scatters light) is encoded in the third ℓ peak, corresponding to acoustic oscillation with a period of 3/2 the time until recombination. This very clear third peak provides the first[3] hint that a large portion of matter in the universe is not baryonic. However, more sensitive measures for the density of baryonic and non-baryonic matter exist and the evidence from the CMB is valuable mainly as a consistency check against other work. An excellent resource for further study can be found at [19].

2.2.5 Evidence from Type Ia Supernova (SNe)

The blue region in figure 2.5 puts almost no constraint on the shape of the universe (curved vs flat), but limits rather sharply the ratio of Ω_Λ to Ω_M. This is possible because type Ia supernovae provide us with a reliable and very bright standard candle, so we have data on the distance and velocity from both near and very distant galaxies, allowing us to distinguish between different kinematic histories as illustrated in figure 2.3, and determine the actual kinematic history of the universe.

A type Ia supernova starts as a binary star system, one partner a white dwarf and the other a less dense star that accretes mass onto the white dwarf. The white dwarf is the core of a star that has exhausted or expelled its hydrogen and helium fuel and exists as a hot but non-fusing remnant composed mostly of carbon and oxygen. As the companion star transfers matter (mostly hydrogen and helium) the transfered matter is fused to oxygen and carbon, increasing the mass but not greatly altering the composition of the white dwarf. There is a critical mass, called the Chandrasekhar mass (not be confused with the Chandrasekhar limit), of about 1.44 solar masses at which point carbon fusion is initiated. Our sun and other main-sequence stars undergoing hydrogen fusion are supported against collapse by thermal pressure, which will react to high power output by expansion and cooling, resulting in a naturally self-regulating system, typically lasting billions of years. In contrast, a carbon

[3]In this work, not in the historical sense.

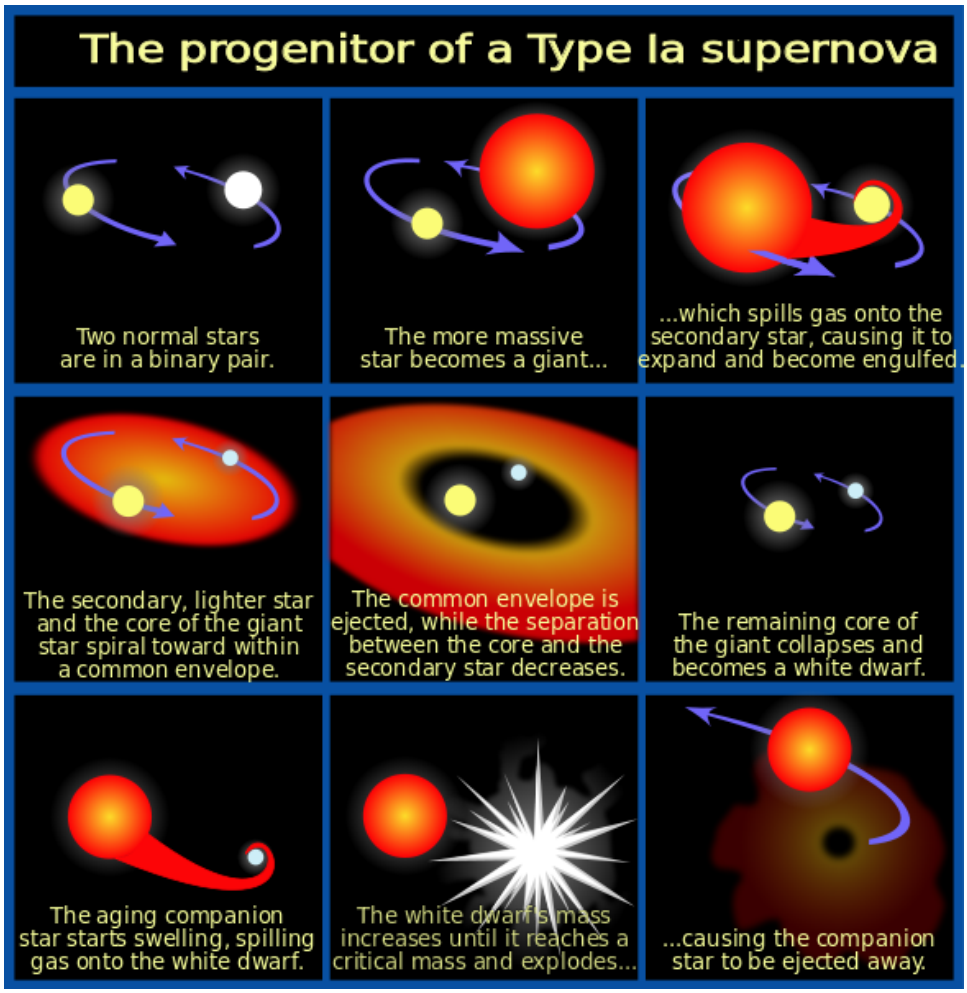

Figure 2.8: Sequence of a type Ia supernova. Fundamental limits on the stability of white dwarf stars and the common condition of a a gradual matter steam captured from a companion star ensure that all type 1a supernovae are similar. Credit: Space Telescope Science Institute

fusing star is supported against collapse only by degeneracy pressure, which is independent of temperature. The result is a runaway reaction that consumes the nuclear fuel in a matter of seconds and a momentary power output about 5 billion times greater than our sun. The supernova and its bright remnant typically outshine all the other stars in the supernova's host galaxy for several weeks. Figure 2.8 shows the sequence leading to a type Ia supernova.

Due to the tight mass constraint on type Ia supernova, there is little variability in their brightness. There exists some variability in type Ia brightness, but observations of nearby type Ia supernova reveal a simple and robust relationship between the absolute brightness and brightness decay time, as shown in figure 2.9. Thus, observation of the brightness

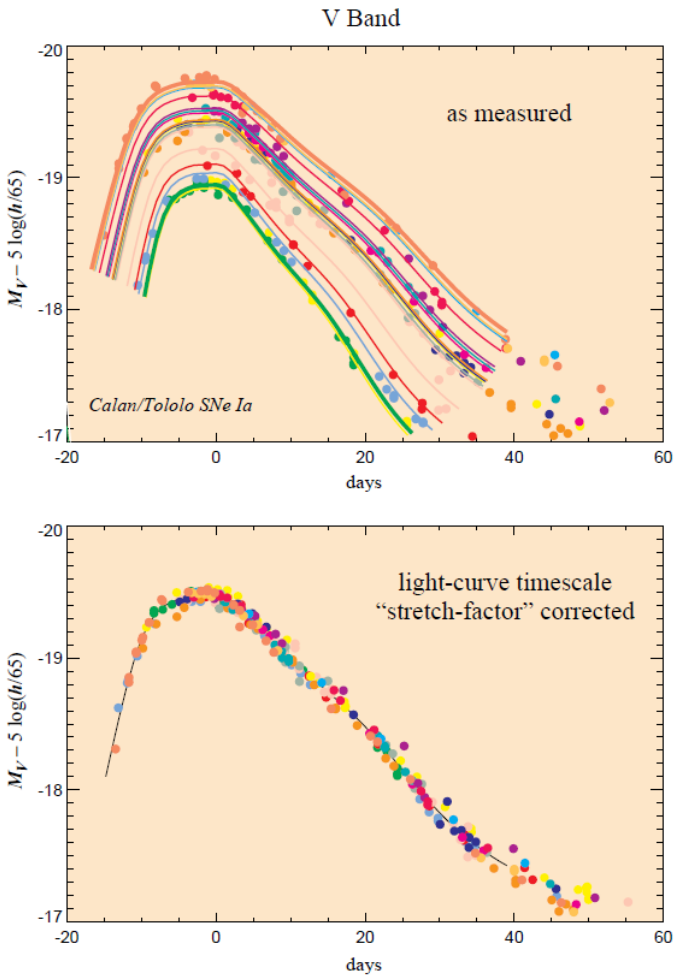

Figure 2.9: Light curves of nearby type Ia supernova. There is a simple and robust relation between absolute brightness and duration. Upper figure shows absolute brightness over time of type Ia supernovae sufficiently nearby to independently determine distance and thus absolute brightness. Lower figure shows the same data fit to the "master curve" demonstrating that absolute brightness can be known from duration alone. At very great distances, the time curve gives the absolute brightness, so apparent brightness can be used to determine distance. Image reproduced from [20].

over time gives the absolute brightness with high confidence, and the absolute distance is calculated by the simple $1/r^2$ reduction in apparent brightness with distance.

Spectral analysis, either of the short-lived signal of characteristic elements in the supernova remnant or of the persistent signal of the hydrogen gas of the host galaxy, gives the redshift. The apparent brightness and redshift of several type Ia supernovae are plotted in figure 2.10, along with fits for various (Ω_M, Ω_Λ) models.

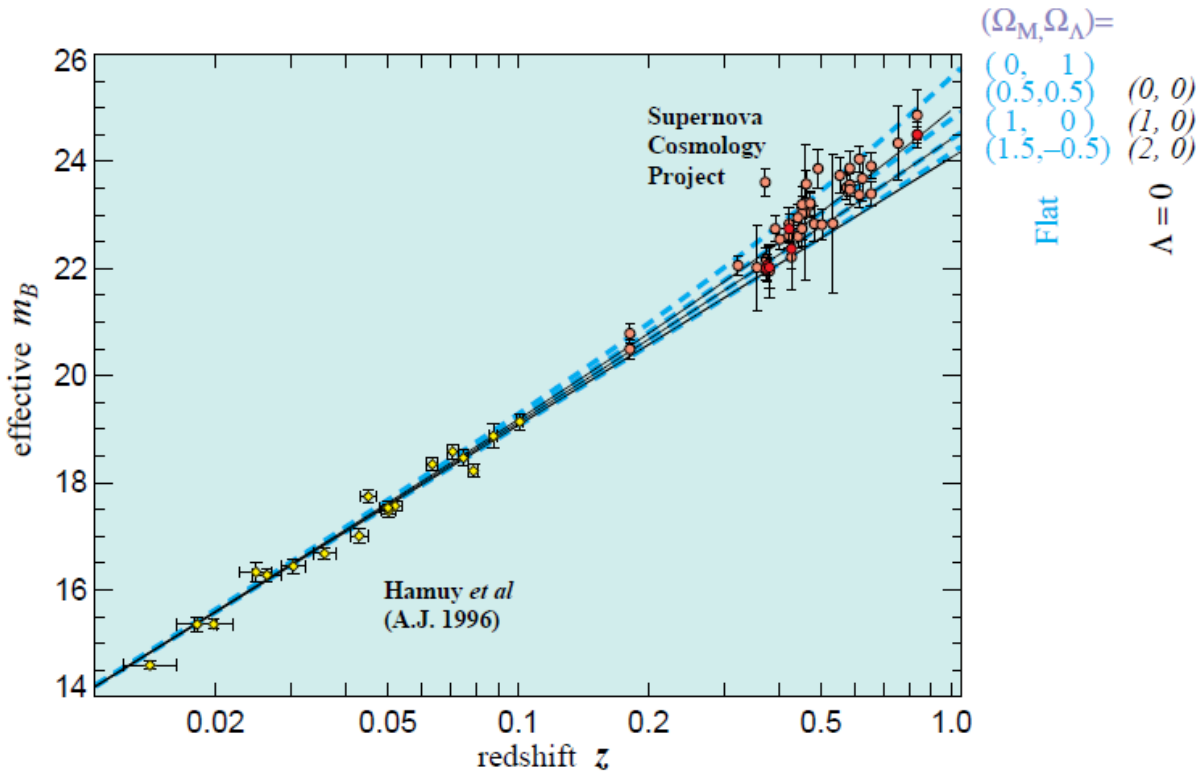

Figure 2.10: Hubble plot of high-z type Ia supernova. Redshift z is a measure of the total new space created between emission and detection of the light. Observed wavelength = original wavelength $\times(1+z)$. Effective m_B is a logarithmic measure of luminous intensity, accounting for energy lost due to redshift. Increasing brightness is in the direction of smaller m_B. Dashed blue curves are the expected m_B vs z relationship for various flat universe models. Solid black curves are the expected m_B vs z relationship for various universe models without dark energy. The data are most consistent with a universe with (Ω_M, Ω_Λ) between (0, 1) and (0.5, 0.5). Image reproduced from [20]

The data are inconsistent with a universe of $\Omega_\Lambda \leq 0$, and most consistent with a (Ω_M, Ω_Λ) closer to (0.5, 0.5) or (0, 1) a result that is very surprising, very well supported, and quite deserving of the 2011 Nobel prize in Physics.

2.2.6 Evidence from Baryon Acoustic Oscillations (BAO)

The green region in figure 2.5 illustrates bounds on the mass-energy density of the universe from observations of Baryon Acoustic Oscillations. The bounds here are almost entirely on the density of "ordinary gravitating" matter (Ω_M) and put almost no constraint on dark energy density (Ω_Λ). This line of evidence arises from observations of the large-scale distribution of matter observed today. While at the largest length scales, matter distribution is homogeneous and isotropic (the Robertson-Walker metric) at local length scales matter is organized by gravity into great voids, filaments, walls, galaxy clusters, and so on. Figure 2.11 shows the large-scale distribution of matter in a simulated universe similar to our own.[4]

This distribution can be compared to the "granularity" of the CMB. Whereas variations in the CMB reflect temperature differences across the oscillating plasma filling space at the moment of re-ionization, this structure arises from the velocity distribution of the same plasma at the same moment of reionization, as regions with an "infalling" velocity distribution collapsed inertially, felt no restoring force after matter and radiation decoupled, and the resulting density variations were amplified by gravitational attraction. The regions of higher matter density self-gravitated to form sheets and filaments of matter that experienced star formation, galaxy formation, and the genesis of all structures visible today. Our milky way galaxy has a size of about 0.03 Mpc, approximately 1 pixel on the most zoomed-in frame of figure 2.11.

The rate of gravitational collapse and "clustering strength" of large-scale structures today is set by the density of matter. A higher matter density from the beginning means more compact structures today. Note the second-to-bottom frame in figure 2.11: the "bubble size" of the quasi-periodic patterns is a little less than half the scale bar, or about 100 Mpc. Figure 2.12 shows the autocorrelation function of the observed matter distribution in our local universe. Black points are data, and the colored curves are various physical models.

Note the peak around 100 Mpc. The position of this peak is consistent with the granularity of the CMB, and the height of the peak tells us the degree of self-gravitation and thus matter density Ω_M of the observed universe. The matter density Ω_M is a sum of ordinary matter and dark matter, with the relative abundance affecting the details of the acoustic peak. The magenta line represents a model universe containing dark matter but no baryonic matter. The best fit model contains both baryonic matter (interacts with light) and dark

[4]Scales in figure 2.11 and 2.12 are given in Mpc h^{-1}, normalizing to the canonical reduced Hubble constant $h = H_0/100$ km/s/Mpc. Due to ever improving observations and evolution of the "scale ladder" used by astronomers, no contemporary measure of H_0 is considered definitive. A current best-guess value of H_0 is about 74 km/s/Mpc for h= 0.74.

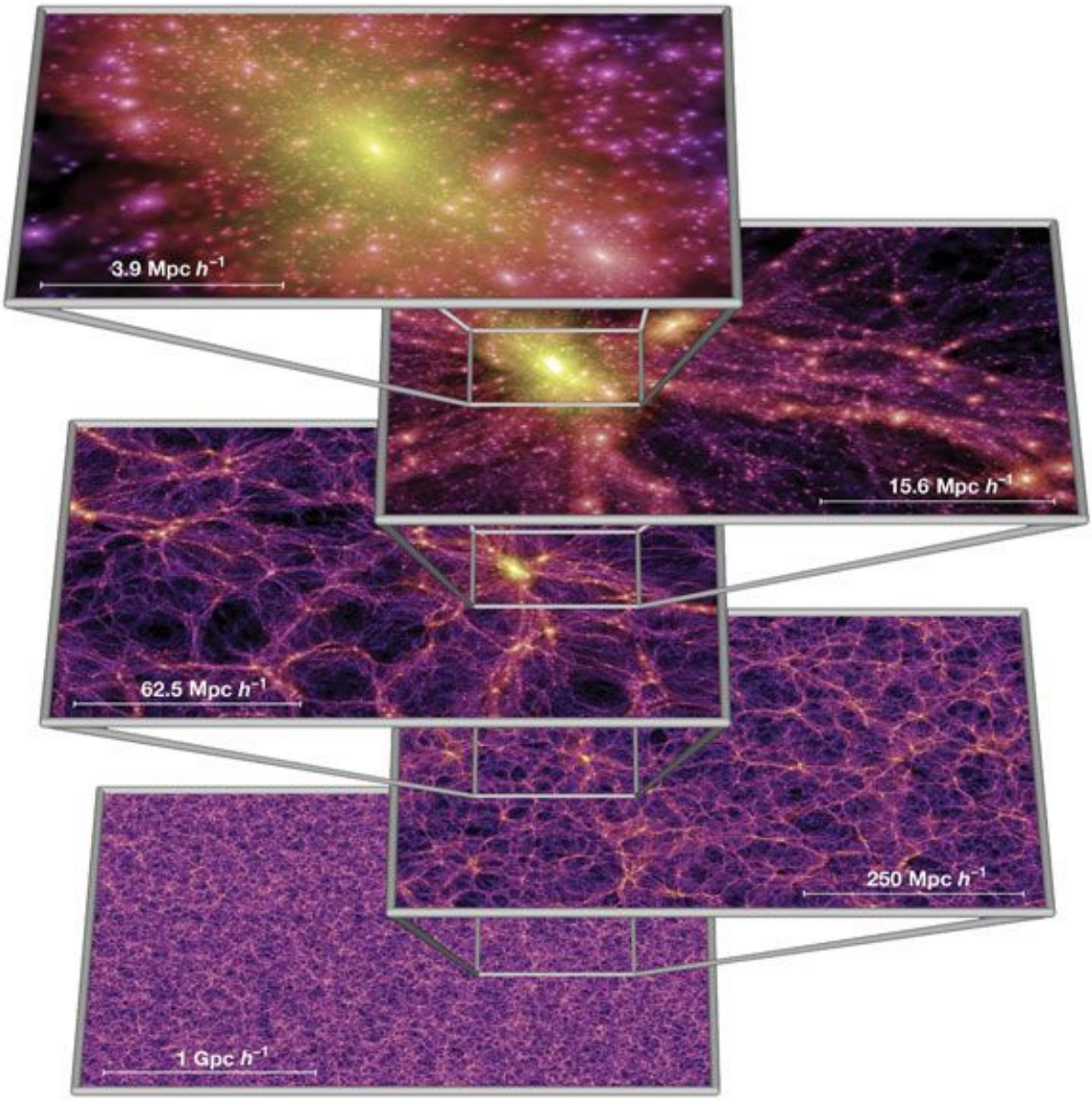

Figure 2.11: Large-scale matter distribution in a simulated universe. The Milky Way is approximately the size of one pixel in the top frame. Image reproduced from [21]

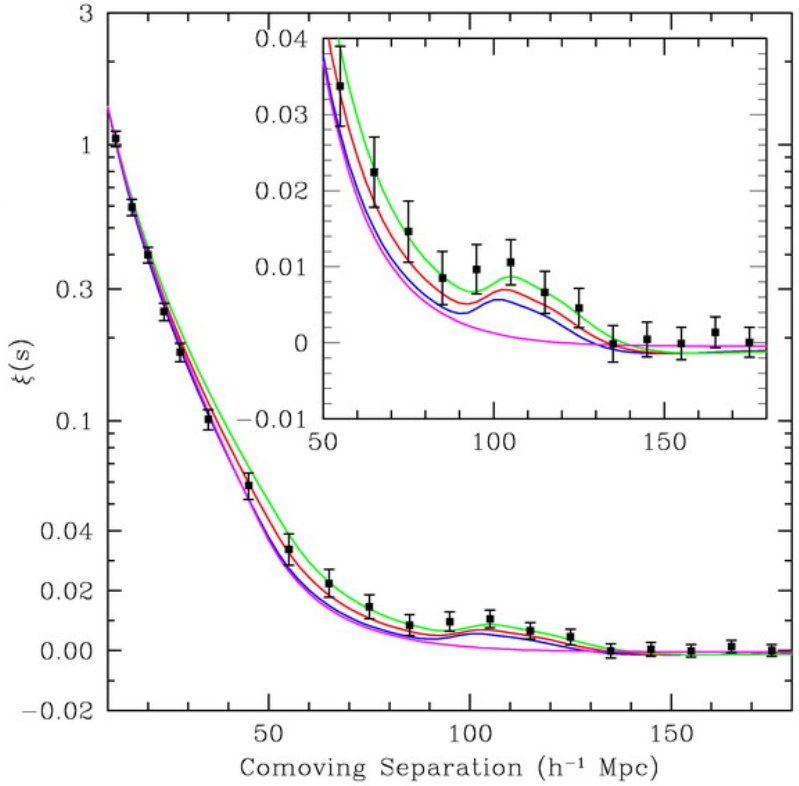

Figure 2.12: Acoustic peak observed in our local universe. Note the broad peak at a comoving separation of about 100 h^{-1} Mpc, consistent with the "cell size" in the second from bottom frame in figure 2.11. The height of the peak gives the fraction of matter that is baryonic. Magenta curve is a model with no baryonic matter, and the green curve is a model with a dark matter to baryonic matter ratio of about 5:1. Image reproduced from [22]

matter (no interaction with light) at a ratio of about 1:5. This is the second[5] hint that a large portion of matter in the universe is not baryonic.

2.3 The Observational Evidence for Dark Matter

Modern cosmology has established that (Ω_Λ, Ω_M) is close to (0.73, 0.27). There are indications from contemporary cosmological observations that matter (Ω_M) in the universe is mostly something other than baryonic, but there are strong lines of evidence tracing back to the 1930s [23], [24] that most matter in the universe is made from something other than protons, neutrons, electrons, and other particles that compose the material aspect of almost

[5]In this work, not in the historical sense.

everything we directly experience. Those lines of evidence include: the rotational curves of spiral galaxies, observations of gravitational lensing at galactic scales, and the primordial abundance elements arising from big bang nucleosynthesis (BBN). Each of these is explored presently.

2.3.1 Rotation Curves of Spiral Galaxies

Newtonian orbital dynamics provides a way of measuring mass: the period of an orbiting body depends on distance and the mass of the system. Kepler's 3rd law arises from applying the inverse-square law of gravity and Newtonian dynamics to a two-body system

$$\frac{a^3}{T^2} = \frac{G(M+m)}{4\pi^2}, \tag{2.16}$$

where a is the semi-major axis, T is the orbital period, G is the gravitational constant, and M and m are the masses of the two bodies. If one mass is much greater than the other, as in the case of Earth orbiting Sol or Sol orbiting the central bulge of the Milky Way, the smaller mass may be neglected and the greater mass may be calculated from the orbital parameters of the smaller body. In the case of a distributed central mass and circular orbits (a good approximation of a spiral galaxy) the semi-major axis a becomes r of the orbit, and $M = M(r)$ where $M(r)$ is the total mass inside a sphere or radius r,

$$M(r) = \frac{4\pi^2}{G} \cdot \frac{r^3}{T^2}. \tag{2.17}$$

Observationally, it is not usually practical to measure the period of a star orbiting a galaxy, but often the velocity can be measured by Doppler shift of spectral emissions. Relating period to velocity by $T = 2\pi r/v$ gives

$$M(r) = \frac{v^2 r}{G}, \tag{2.18}$$

or re-arranging to solve for $v(r)$ gives

$$v(r) = r^{-1/2}\sqrt{G \cdot M(r)}. \tag{2.19}$$

Equation 2.19 makes clear what observations of $v(r)$ tell us about a distributed central mass. If $M(r) =$ constant, as for a single point-like mass, then $v(r) \propto r^{-1/2}$. If $M(r) \propto r$, as for an axial rod of constant linear density, or spherical distribution with $\rho(r) \propto r^{-2}$ then $v(r) =$ constant. If $M(r) \propto r^2$, as for a disk of constant density, then $v(r) \propto r^{1/2}$.

Figure 2.13 shows a log-log plot of the rotation curve of our own Milky Way. At small r, $v(r) \propto r^{-1/2}$ because $M(r)$ is dominated by a massive central black hole. At greater distances from the center, the velocity increases slightly and becomes approximately constant, consistent with a gently decreasing density, but the $v(r) \propto r^{-1/2}$ behavior is never recovered, suggesting that the visible parts of the galaxy are never outside the bulk of the galactic mass.

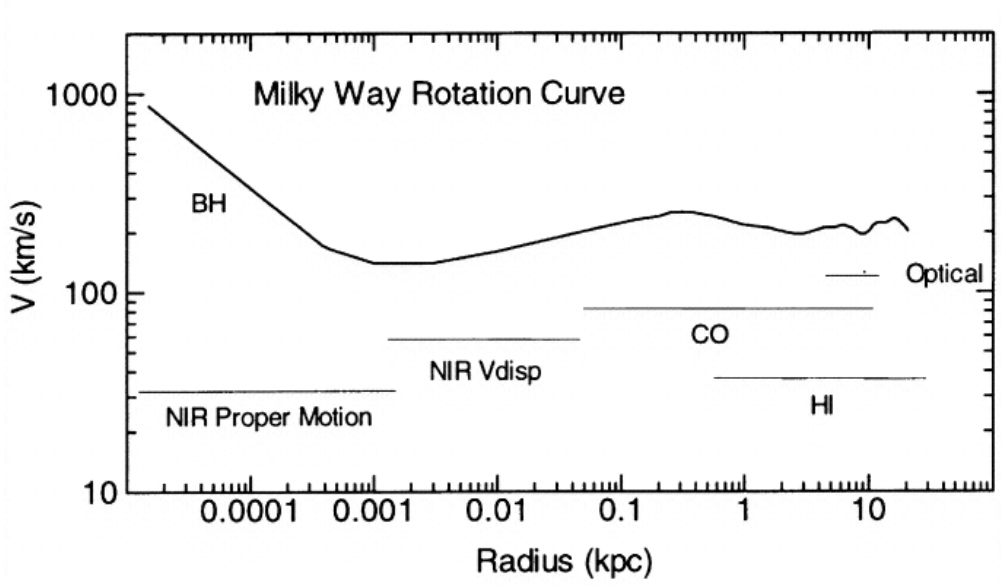

Figure 2.13: Milky Way Rotation Curve. At small r, $v(r) \propto r^{-1/2}$ because the mass is dominated by a massive central black hole. At greater r, $v(r)$ is approximately constant, indicating that no visible part of the Milky Way is outside the bulk mass of the Milky Way. Image reproduced from [25]

This $v(r) \approx$ constant behavior holds true even far outside the optical disk. Emission lines from diffuse hydrogen gas allows measure of $v(r)$ far outside the visible bulk of many galaxies. An example from the nearby Andromeda galaxy (known among astronomers by the less colorful moniker M31) is shown in figure 2.14a, and 2.14b shows rotation curves for a representative cross-section of galaxies.

The gravitational mass responsible for the flat $v(r)$ profiles is not diffuse gas or dust, which would be revealed by their own spectroscopic signatures against the bright background of luminous stars, but is seemingly invisible.

A similar analysis can be done with galaxy clusters. In this case there is no central, radial organization of mass, but several galaxies of comparable mass that are gravitationally bound to each other although not necessarily in dynamically stable orbits. As early as 1959 there was strong evidence that galaxy clusters were gravitationally bound by unseen mass "...the Local Group of galaxies can be dynamically stable only if it contains an appreciable amount of intergalactic matter." [27], though at the time the unseen mass was presumed to be diffuse hydrogen gas.

This evidence from Newtonian dynamics is the earliest indication that dark matter is a dominant part of mass in the universe, but the "dark matter interpretation" was not canonical until other independent lines of evidence were discovered.

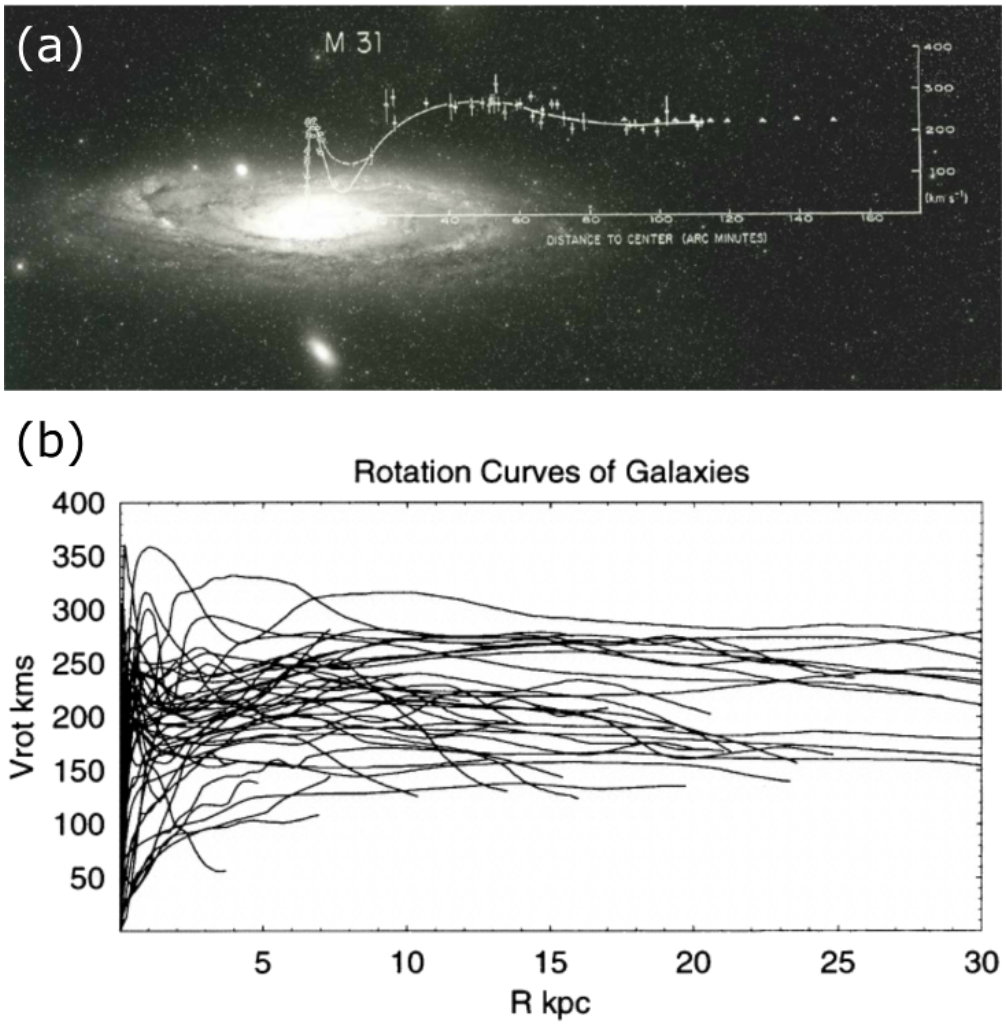

Figure 2.14: Galactic rotation curves. (a) Rotation curve of the Andromeda Galaxy (M31). Image reproduced from [26] (b) Overlaid Rotation Curves of several galaxies. In all cases, $v(r)$ is approximately flat at large r, indicating a total matter distribution inconsistent with the visible distribution of matter. Image reproduced from [25]

2.3.2 Gravitational Lensing and Galaxy Cluster Collisions

One independent line of evidence for dark matter comes from gravitational lensing and galaxy cluster collisions. A prediction of general relativity is that the path light will be bent by the space-time distortion cause by the presence of mass. [6] It is the first observation of this effect [28] that is widely considered the original empirical confirmation of GR. A beautiful example of gravitational lensing is shown in figure 2.15a. This image is made possible by the accidental co-linear alignment of a far-distant blue galaxy, a mid-distant massive elliptical red galaxy, and ourselves. Diverging light from the blue galaxy is bent by the gravitational distortion of the red galaxy to converge on our position, resulting in the blue galaxy appearing as a ring-like structure around the red galaxy. The size of the ring and relative distances of the two galaxies give an accurate measure of the mass of the foreground galaxy. A more massive foreground galaxy would result in light bending through a greater angle (larger ring) while a lighter foreground galaxy would result in a smaller ring. The mass indicated by the degree of lensing is much greater than the mass attributable to gas and stars, and consistent with the mass inferred by dynamical measurements.

Typical gravitational lensing events are not often so neatly aligned. Figure 2.15b shows a more typical example of observed lensing. Here there are several massive foreground galaxies, and many background galaxies distorted into partial arcs. Individual measure of the distance to, orientation, and degree of distortion of the background galaxies can give a detailed picture of the total mass and mass distribution of the foreground galaxies. The mass inferred from such lensing is consistent with the mass required to explain galactic stability, and also intergalactic mass required to explain the gravitational binding of galaxy clusters.

The measure of mass given by gravitational lensing does not specify the kind of matter present. It is not possible to rule out by gravitational lensing alone that the extra mass is composed entirely of diffuse neutral hydrogen gas. Our own local group, for example, contains a large amount of hydrogen gas not associated with any particular galaxy, but gravitationally bound to the group as a whole. Collisions of galaxy clusters provide a test to probe the composition of the intergalactic medium. When galaxy clusters fall into each other, only the intergalactic gas really "collides" with anything. The stars and other compact halo objects are so small relative to the surrounding space that collisions are very rare, though they do interact gravitationally. The dark matter, whatever its composition, also "falls straight through" because it has no interaction with light or other force-carrier other than gravity, so it also moves ballistically. The gas, on the other hand, behaves very much like an ordinary ideal gas with a mean free path much smaller than the interaction scale, exhibiting compression, heating, shock waves, etc.

Figure 2.16 shows a composite image of the bullet cluster, which is just such a galaxy cluster collision in progress. The orange and white parts are from an optical image from the

[6] This prediction is not strictly unique to GR. Newtonian dynamics (ND) also predicts deflection of a beam of light by a massive body. GR and ND differ in the fundamental underlying physical model and in the predicted degree of deflection–GR predicts (and measurement confirms) twice the deflection predicted by ND.

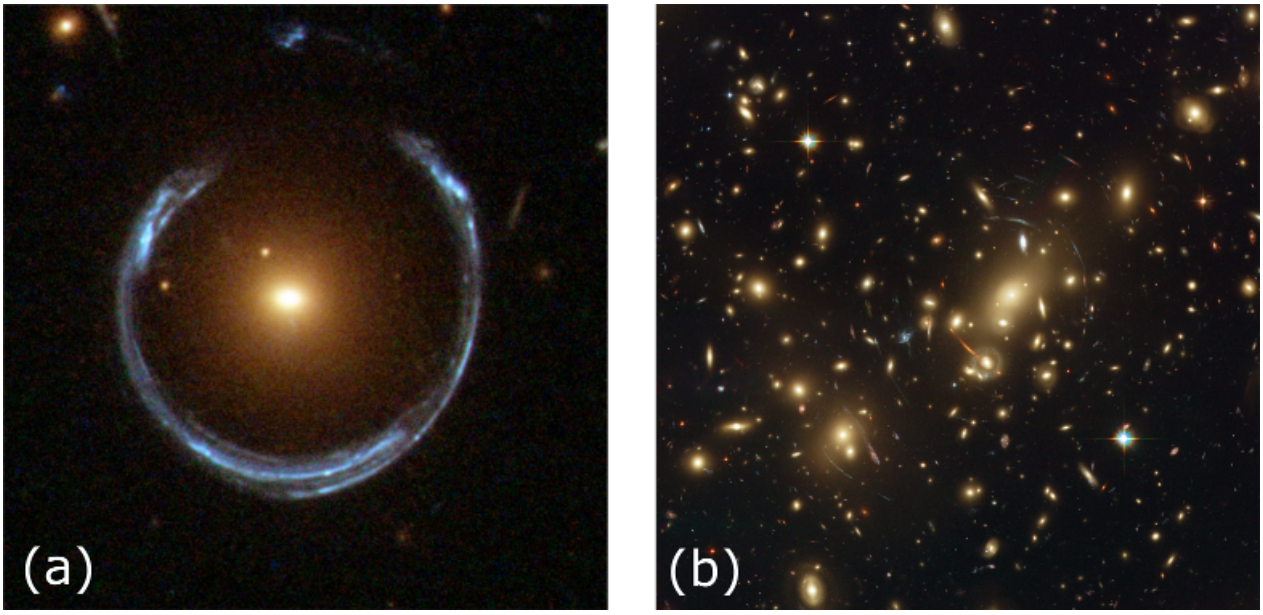

Figure 2.15: Gravitational Lensing. (a) A beautiful example of an Einstein ring due to gravitational lensing. (b) Multiple complex gravitational lensing. Image Credits: ESA/Hubble & NASA.

Hubble Space Telescope. A false-color overlay of the bullet cluster as seen by the Chandra X-ray telescope is shown in pink, and an estimate of the mass concentration inferred from gravitational lensing and distortion of background galaxies is overlaid in blue. This image shows two galaxy clusters that have recently[7] collided at a speed of 4500 km/s. The larger, left-hand cluster is currently moving leftwards and the smaller cluster is moving rightwards. The gas cloud of the smaller right-moving cluster is distorted into the eponymous "bullet" shape. This gas represents most of the ordinary, baryonic matter in either cluster. The intergalactic gas of each cluster has been compressed and heated by the collision (thus the brightness in the x-ray spectrum) and greatly slowed by drag effects. In contrast, the majority of the mass has not been slowed by the collision and exists today on either side of the collision center, along with most of the stars and non-gaseous visible matter.

This provides convincing evidence that (nearly-invisible) neutral hydrogen gas is not the primary mass of galaxy clusters, but that the great majority of mass in these clusters is something else entirely. Something for which we have no better name than "dark matter."

[7]About 150 million years ago.

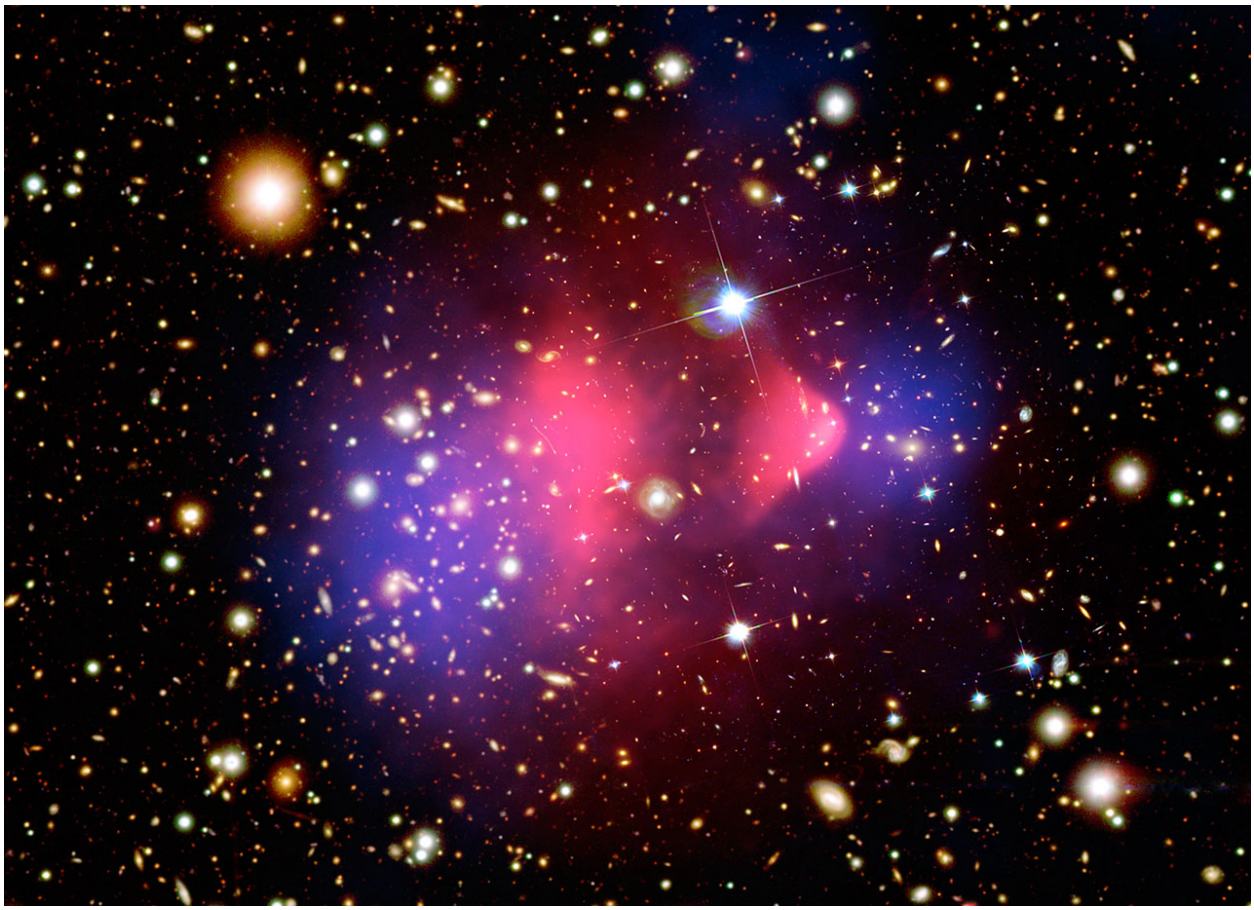

Figure 2.16: Composite Image of the Bullet Cluster. Image Credits X-ray (magenta): NASA/CXC/M.Markevitch et al. Optical (orange/white): NASA/STScI; Magellan/U.Arizona/D.Clowe et al. Lensing Map (blue): NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe et al.

2.3.3 Big Bang Nucleosynthesis (BBN)

Quite in contrast to observing current galactic dynamics to measure the density of matter, measuring the relative abundance of elements produced in the big bang gives us a total density of baryons, regardless of whether we can see or account for them today. The baryon density calculated this way is in good accordance with what we do directly estimate from star counting and optical estimates of intergalactic hydrogen gas, but falls far short of the density of matter indicated by cosmological evidence. The difference is then some matter that isn't visible and doesn't participate in nuclear reactions, a "dark matter" whose inferred density matches that indicated by galactic dynamics.

When the early universe was between 10 seconds and 20 minutes old it was dominated by radiation, with protons, neutrons, electrons, and their anti-particles comprising a tiny fraction of the total mass-energy. In a plasma cooling sufficiently slowly the free protons and neutrons would condense into heavier elements of lower nuclear potential keeping nuclear potential energy in thermal equilibrium, ending with every baryon incorporated into an atom of iron. Fortunately for us in the present, maintaining such a nuclear thermal equilibrium requires a steady supply of free neutrons, but free neutrons are unstable with a half-life of about 10 minutes. This leaves the mix of elements far out of equilibrium when nucleosynthesis halts, with the great majority of baryons existing as H, ^{2}H or ^{4}He, with a small amount of ^{4}He and ^{7}Li. [29] Elements heavier than Li exist today only by virtue of nuclear fusion in stars that started about 300,000 years after BBN and is ongoing.

The exact mix of light elements produced in BBN depends on the ratio of baryons to photons, as illustrated in figure 2.17a. The easiest ratio to measure (^{4}He/H) is the least sensitive to baryon density, but the observed abundance of ^{4}He (about 24% of baryonic matter) does serve as a robust confirmation of BBN as a model for the origin of elements. Deuterium (^{2}H) abundance provides the most sensitive measure of baryon density due to its low binding energy. A higher baryon density effectively provides more bullets that tend to break deuterium apart. At sufficiently high baryon density, ^{2}H is entirely eliminated.

The abundance of ^{2}H and other primordial elements can be observed from observing the light from quasars as filtered through clouds of intergalactic gas un-tempered by stellar nuclear reactions. Quasars are a very high intensity light source common in the early universe (probably accretion discs feeding super massive black holes). Observing the absorption lines gives the relative abundance of primordial gases sampled over a baseline spanning the intergalactic medium.

To an infinitesimal degree of accuracy, every baryon and photon existing in the era of BBN exist today. The baryons exist today mostly as hydrogen gas, either between galaxies, in galaxies, or in stars. Some have been transformed to heavier elements, but such reactions conserve baryon number. The photons have been scattered in the pre-recombination era and red-shifted to become today's CMB. Though we associate photons with visible light (sunlight, starlight, etc) this is a vanishingly small contribution to total photon number of the universe and it is almost exactly true to say every photon that exists today has its origin in the era of BBN. Thus the baryon to photon ratio of the BBN era is unchanged

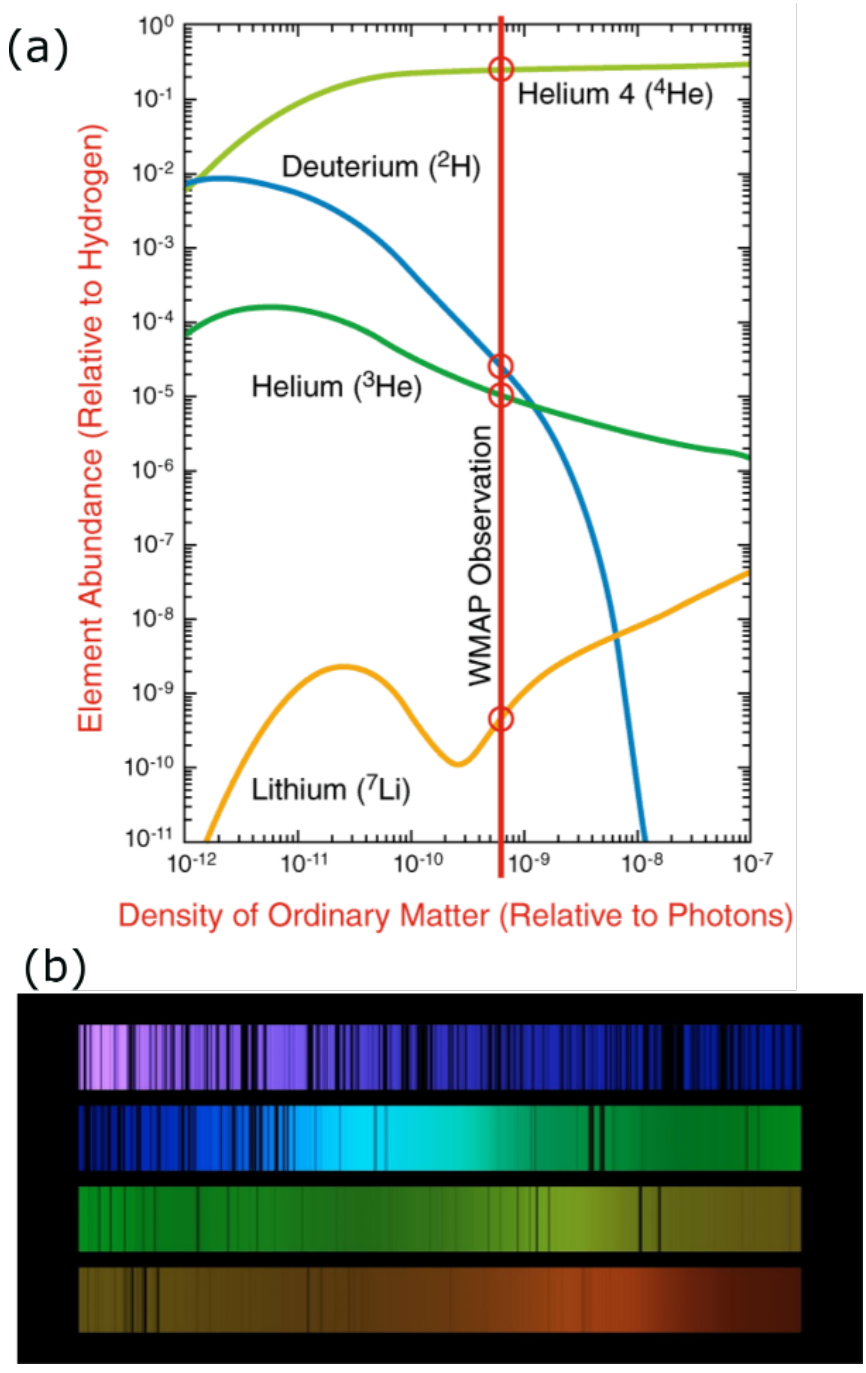

Figure 2.17: Relative abundances of primordial elements. (a) Relative Abundance of Primordial Elements vs. Density of Baryonic Matter. Image reproduced from [30]. (b) Quasar Spectrum with Absorption Lines. Credit: ESO and V. DOdorico (Osservatorio Astronomico di Trieste, Italy)

today. We have an excellent measure of photon density by way of the CMB, and so by the ratio indicated by BBN dynamics, we have an excellent measure of baryon density today as well. The baryon density indicated is about $\Omega_b = 0.046$, in excellent agreement with direct estimates.

2.3.4 Properties of Dark Matter

The observational evidence for dark matter is deep and convincing. While it is tempting and not strictly inaccurate to say "We have no idea what dark matter is.", the evidence we have that dark matter exists also puts rather strict bounds on its properties, which in turn puts strict bounds on viable models for its constitution. Any viable candidate for dark matter must have the following properties:

- **Non-interacting**

  Perhaps most obvious, dark matter has a minimal interaction with light and ordinary matter, and is detectable only by its gravitational influence. We would have seen it already otherwise. Many models of dark matter can be ruled out by the logic of "If it interacted that strongly, we would have seen evidence of it by now." Such limits are not necessarily set in the lab: energetic cosmic events like supernovae also can set bounds. During the stellar collapse immediately preceding a supernova, cooling is provided by EM radiation from the surface, and at the high densities and nuclear reaction rates at the core, neutrino radiation that carries away energy through the bulk of the star. A particle that comprises dark matter with sufficient interaction strength would provide another cooling channel, which would be observable in supernova dynamics. Other high-energy events can similarly put an upper bound on the interaction strength of any dark matter candidate. Perhaps less obviously, dark matter must have a minimal self-interaction. It is fun to posit a "shadow universe" of dark matter with dark stars emitting dark "light", dark planets, dark chemistry, and even dark life that is coincident with our own, but such a model is inconsistent with observed galactic and stellar evolution. Ordinary matter heats under gravitational collapse, and cools by radiation allowing further collapse. Our models of galactic and stellar evolution require a mix of ordinary matter with radiative cooling mechanisms and dark matter that is purely ballistic. There is no model of galactic and stellar evolution that does not include mostly non-interacting, non-clumping matter that is consistent with the observed state of the universe. A viable dark matter candidate must have a minimal interaction both with itself and ordinary light and matter.

- **Long-lived**

  Any dark matter candidate must be something with a lifetime that is long relative to the current age of the universe. There is a veritable particle zoo in the standard model, including particles with no electric charge and no color charge, for example the Higgs boson (H, fundamental) or the Lambda (Λ^0, compound). Lacking electric or color

charge, such a particle would behave like dark matter, but the Higgs has a lifetime of about 10^{-22} s, and the Λ^0 a lifetime of 10^{-10} s, so none created in the big bang exist today. Most proposed solutions to the dark matter problem posit an extension of the standard model that give rise to a new kind of particle. It is important that any such exotic particle be extremely stable to be a viable dark matter candidate.

- **Non-baryonic**

  One can imagine solving the dark matter puzzle by populating space with dark "ordinary" matter in the form of cold, non fusing "stars", rogue planets, or even massive alien civilizations with cloaking technology indistinguishable from magic. All such "hidden baryonic" mass would have its origin in the big bang, and been present in the era of BBN. The increased baryon number during BBN would have a clear signature in the relative abundance of ^{2}H, ^{3}He, and ^{7}Li in primordial gas we observe today. Furthermore, we have no model for cosmic evolution that results in the majority of baryonic mass collecting into "invisible pockets" while leaving the visible portion in the configuration we see it today.

- **Cold**

  Cold here means a thermal velocity distribution resulting in non-relativistic velocities. Of course a cold particle could exhibit a very large V_{RMS} if it were sufficiently light, but this would not be called "cold" in the context of dark matter because such fast particles would not be gravitationally bound to galaxies or galaxy clusters.[8] At first blush the neutrino appears to be an ideal dark matter candidate: it has no interaction with light, a very weak interaction with matter, is non-baryonic and long-lived, and best of all is known to actually exist. However, neutrinos are very light (until about the turn of the millennium thought massless, now thought bounded between 0 and 1 meV) and in the past at thermal equilibrium with the quark-gluon plasma and early BBN, so today they would have a thermalized kinetic energy of about $2.7K \cdot k_B \approx 0.2$ meV resulting in relativistic velocity. If most of the mass of the universe were composed of such "free streaming" matter, gravitational collapse and structure formation would be impossible except at the very largest scales. The large-scale structure of matter in the universe we observe is consistent with cold matter mixed with not more than a small fraction of free-streaming matter.[9] Any viable dark matter candidate must be rather slow-moving [31], due to either or both a large particle mass or a temperature much colder than the CMB.

[8] Astronomers use "cold" to mean *slow*, "metal" to mean *more protons than Helium*, and all too often, even "1" to mean *2*. We'll go along here, because they mean well and are fun at parties.

[9] Primordial big-bang neutrinos are indeed thought to exist, and to be the most common particle in the universe after photons. Due to their low energy they are yet to directly detected, and due to their low particle mass are only a small fraction of the total mass of the universe despite their abundance.

- **Common**

  Dark matter is common–much more common than the matter we can see. All evidence suggests that dark matter is 4 to 5 times more abundant than "ordinary[10] matter". Any theory of dark matter must allow for its creation in the correct abundance. A candidate like the neutrino has a very tightly constrained predicted primordial abundance from the equipartition theorem applied to the early hot universe, which is too small to account for dark matter, but consistent with neutrino observations to date. Other candidates, like any model of a very light Nambau-Goldstone boson emerging from a yet-to-be observed spontaneous symmetry breaking in the early universe might predict an over-production of dark matter and $\Omega_M > 1$, inconsistent with cosmological observations.

2.4 Notable Dark Matter Proposals

Most proposed solutions to the dark matter puzzle posit some kind of difficult-to-detect exotic particle, though they may also posit instead very massive objects, like otherwise undetected black holes. Figure 2.18 shows various dark matter proposals sorted by particle mass, covering some 80 orders of magnitude. The figure includes both candidates seriously considered in the scientific community (ex: "neutralinos", "black holes") and satirical candidates (ex: "bees"), perhaps as commentary on the profusion of presently viable dark matter models. We will comment here on a few prominent dark matter proposals, including our favorite, the axion, appearing on the lightest, left-hand side of figure 2.18.

2.4.1 WIMPs

WIMP is an acronym for Weakly Interacting Massive Particle, and is arguably the favored solution to the dark matter puzzle in the scientific community today.[11] First proposed in the 1970's based on he observation that a new particle with a self-annihilation cross-section associated with the weak scale would have about the right abundance today to account for dark matter, they remain a promising dark matter candidate. WIMPS arise from theories of "supersymmetry" that posit a fundamental symmetry between fermions and bosons, and a whole "mirror zoo" of particles to the standard model. Existence of such a supersymmetry in nature could explain the low mass of the Higgs boson, and is otherwise attractive for its ability to "beautify" aspects of particle theory. Examples of super-symmetric partners are the photino ($\tilde{\gamma}$), super-symmetric partner to the photon (γ), or the sneutrino ($\tilde{\nu_e}$), super-symmetric partner to the electron neutrino (ν_e). A WIMP dark matter candidate would be

[10] I would not be against calling baryonic matter "extraordinary" matter and dark matter "ordinary" matter. Are you with me?

[11] At least as measured by worldwide funding for direct detection experiments.

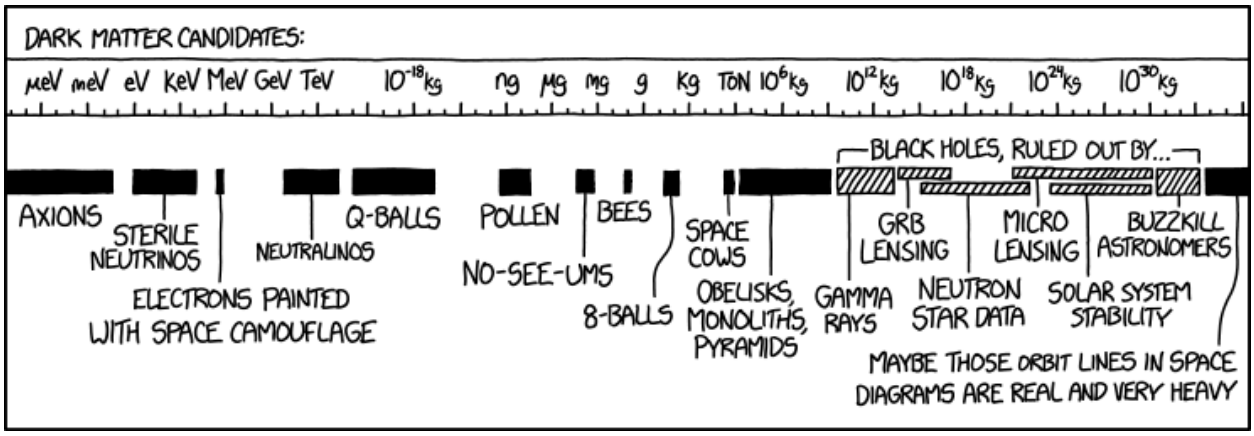

Figure 2.18: A satirical summary of dark matter candidates. This chart includes both "reputable" dark matter proposals (ex: our own beloved axion, "micro" black holes) and satirical elements (ex: "no-see-ums", "obelisks"). The artist's statement is: "My theory is that dark matter is actually just a thin patina of grime covering the whole universe, and we don't notice it because we haven't thoroughly cleaned the place in eons." Reproduced from [32].

the lightest stable particle of the model, with mass around 10-100 GeV. The WIMP is the "particle physicist's dark matter candidate", with a theory similar to existing particle theory and detection methods similar to existing particle detectors (scintillators, bubble chambers, etc). The WIMP models hit all the right points– they are non-interacting, long-lived, non-baryonic, cold, and common. There is much activity today on the front of WIMP-based dark matter searches, among them SuperCDMS, EDELWEISS, LUX, XENON, PICO, and others. The community has been very successful thus far in narrowing set of viable WIMP dark matter models, but has not yet made a direct detection. See [33] and [34] for excellent reviews of WIMPs as a dark mater candidate.

2.4.2 MACHOs

MACHO is an acronym for "Massive Compact Halo Object", playfully named as a riposte to the WIMP. Rather than arising from an extension of particle theory, the MACHO is the "astronomer's dark matter candidate". Any mass is easy to see if it is bright like a star, harder to see if it is non-luminous but dispersed like a gas or a dust cloud, and almost impossible to see if it collected into a tight ball, or compact halo, like a rogue planet or micro black hole. The MACHO proposal posits that the missing mass exists in such almost-invisible bodies, but makes no claim as to the composition: baryonic, non-baryonic, black holes, or otherwise. The search for MACHOs is a task of optical astronomy–any such object could be detected by a transient gravitational lensing event, making a star appear momentarily brighter as

the MACHO moves between ourselves and the star. By continuously monitoring millions of stars over long periods of time, a statistical model of MACHO mass and abundance can be developed. The Milky Way central bulge and the nearby Large Magellanic Cloud provide a backdrop on individually resolvable stars that has been observed for just such events by the MACHO project [35]. Years of observational data support the conclusion that MACHOs make up no more than 20% of the mass of our Milky Way galaxy. In some sense, MACHOs *are* a partial solution to the dark matter puzzle, but further dark matter models are required to fully explain the density of dark matter in our own galaxy, and the nature and origin of MACHOs in our own galaxy and in the greater universe remains an open question.

2.4.3 Axions

The Axion solution to dark matter is a particular extension of the standard model that posits a new exotic particle called the Axion. It is highly motivated because it was originally imagined as, and remains a viable solution to, the "Strong CP problem" to explain why the strong interaction appears highly symmetric while the weak interaction is demonstrably not. After its properties were fleshed out a bit by particle theorists, astronomers saw that it has all the right properties (section 2.3.4) to be a promising dark matter candidate. Interestingly and in contrast to WIMP models, axions are "cold" not because of a large mass, but because they were created in a non-thermal phase transition in the early universe, and so were "born cold" and remain cold today regardless of their mass. We favor the axion as a dark matter candidate, and will expound in greater detail in chapter 3.

2.4.4 MoND

MoND is an abbreviation for "Modified Newtonian Dynamics". This solution to the dark matter problem does not propose a new previously unseen matter, but instead proposes that our inference of mass from galactic-scale kinematics is flawed. The venerable relation $F = ma$ has not been directly tested for gravitating bodies at accelerations less than what is typical for bodies in the outer solar system (about 10^{-6} m/s^2), while the acceleration of Sol around the center of the Milky Way is about 10^{-12} m/s^2, and accelerations at greater distance or in smaller galaxies will be lesser still. MoND proposes that $F = ma$ becomes slightly non-linear at small a, accounting for the observed galaxy rotation curves. MoND is a disfavored hypothesis today because it is difficult to reconcile with General Relativity (both in terms of the theoretical framework and in terms of data from gravitational lensing), it leaves un-answered (or deepens the mystery of) the composition of matter that adds to $\Omega_M + \Omega_\Lambda = 1$, does not fit every galaxy rotation curve, and is not well motivated by theoretical reasoning, but is an ad-hoc theory made to to fit *some* available data. We share the general skepticism toward MoND theories, but include mention here in recognition of the importance of keeping one's mind flexible when considering any outstanding question in science. The quantization of light was, after all, originally just an ad hoc theory to explain

the spectrum of black-body radiation, but it was this idea that sparked the development of quantum mechanics.

Chapter 3

The Axion and ADMX

As chapter 2 was an endeavor to delineate the question "What is the composition of the universe?" (the Dark Matter Problem), this chapter approaches the question "Why is the strong nuclear force highly symmetric?" (the Strong CP Problem). Perhaps an esoteric concern about a mathematical abstraction concerning particle dynamics almost entirely divorced from direct human experience is less likely than musings on the ultimate origins, fate, and constitution of reality to capture the imagination, but it is truly no less fundamental or foundational to our understanding of material reality. Remarkably, a single idea[1] may offer a simultaneous solution to both mysteries.

The axion is a hypothetical particle that results from a modest modification to the Standard Model[2] of particle physics, first motivated as a solution to the Strong CP Problem. Here we both develop the motivation for the axion extension to the standard model and demonstrate that the natural properties of the axion make it an excellent dark matter candidate. The remarkable conceptual economy of the axion makes it a very promising and attractive field of study, but by no means does mere elegance make a theory correct. To make any claim that axions are "real" and not just a beautiful idea requires experimental evidence, and pursuit of direct experimental evidence is the raison d'être of ADMX. An overview of ADMX and other axion searches rounds out this chapter.

3.1 The Strong CP Fine-Tuning Problem

From the cosmological puzzle of dark matter and study of nature at the scale of 10^{26} m with dynamics governed by gravity, we turn to another outstanding problem of physics today: the strong CP problem, arising from study of nature at scales of 10^{-15} m or smaller with

[1] It's the axion, stupid.

[2] "The Standard Model" is a dull name for what is the most accurate scientific theory known to human beings. Perhaps it should be called "The Absolutely Amazing Theory of Almost Everything", instead of something that suggests if you can afford a few extra dollars a month you should buy the upgrade.

dynamics governed by the "strong nuclear force". It is this realm of physics that produced the idea of the axion.

The theory of the axion arises from Quantum Chromo-Dynamics (QCD) which describes the strong-force interactions of quarks and gluons within hadrons such as protons and neutrons. QCD is a very successful theory, proven reliable from decades of experimental science of the highest caliber, from the early days of the Stanford Linear Accelerator (SLAC) in 1967 to present-day work at CERN. Some predictions of QCD arise from the structure of the theory but some must be put in "by hand", for example the quark masses. QCD makes no prediction about what the quark masses should be, but once those masses are experimentally determined and fed into the theory, QCD makes very detailed and accurate predictions about real-world dynamics.[3] Another parameter that must be "put by hand" into QCD is called $\bar{\theta}$ which in the standard theory of QCD may have any value from 0 to 2π.[4] Experimentally, $\bar{\theta}$ has a value very, *very* close to 0, experimentally constrained to $\left|\bar{\theta}\right| < 10^{-10}$, set by measurements of the neutron electric dipole moment (eEDM) bounding the nEDM d_n to $(-7.0 < d_n < 5.0) \cdot 10^{-26} e \cdot \text{cm}$ [36]. Zero is a very particular, special value for $\bar{\theta}$, because only for $\bar{\theta} = 0$ is the strong nuclear force totally symmetric, exhibiting no CP violation. While there is a nice aesthetic to symmetry, there is no a priori reason to expect perfect symmetry in physical law, and indeed the weak nuclear force is notably asymmetric, as was first shown in 1957 [37]. Moreover, $\bar{\theta}$ is a sum of independent terms, so it seems possible though exceedingly unlikely they would sum to zero by lucky chance.

This is the "Strong CP problem". The strong nuclear force is symmetric only if $\bar{\theta} = 0$, and $\bar{\theta}$ is the sum of several non-zero terms. How is it that $\bar{\theta}$ is so close to zero, if not by chance? In this section we will give just enough background in QCD to to motivate the Pecci-Quinn solution to the strong CP problem, then introduce the Peccei-Quinn solution and show how the axion particle naturally arises from a particular modification to QCD.

3.1.1 A Very Brief Review of QCD

A minimal Lagrangian for Quantum Chromo-Dynamics (QCD) can be written as:

$$\mathcal{L}_{QCD} = \sum_{f=u,d,s,c,t,b} \bar{q_f} \left(i\gamma^\mu D_\mu - m_f \right) q_f - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{3.1}$$

where f runs over quark flavors, q_f are quark fields, γ^μ are the gamma matrices, D_μ is the covariant derivative, m_f are the quark masses, and $G^a_{\mu\nu}$ is the gluon field strength tensor. This form is good enough for calculating some perturbative solutions, but it is not complete. As written, equation 3.1 is symmetric under $U(1)_A$ global axial rotations, implying the existence of hadron parity doublets (like a parity-mirror proton) which are not observed in

[3] It might be more accurate to say not that quark masses are measured and the values fed into QCD, but that the quark masses are inferred by matching actual experimental results with predictions of QCD given some particular quark masses.

[4] Or $-\pi$ to π, or however else one prefers to define the limits of a full circle in radians.

nature. The $U(1)_A$ symmetry can be broken by adding an axial current in the style of the Adler-Bell-Jackiw anomaly first proposed in the context of pion decay [38][39], introducing a new term to the QCD Lagrangian [40] as in

$$\mathcal{L}_{QCD} = \sum_{f=u,d,s,c,t,b} \bar{q_f} \left(i\gamma^\mu D_\mu - m_f\right) q_f - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \frac{\bar{\theta}}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{\mu\nu a}. \tag{3.2}$$

This QCD Lagrangian is a marked improvement on equation 3.1 in that it predicts only particles actually observed in nature, but it is not entirely without consequence, and leads to a new prediction that the strong force (like the weak force) violates CP. The total derivative $G^a_{\mu\nu}\tilde{G}^{\mu\nu a}$ in the new rightmost term explicitly violates CP symmetry, and the degree of violation is set by $\bar{\theta}$. The factor $\bar{\theta}$ is the sum of two parts, as in

$$\bar{\theta} = \theta + arg\, det\, M, \tag{3.3}$$

where M is the quark mass matrix and θ describes the QCD vacuum. There is no particular relation between θ and $arg\, det\, M$, each of which could take on any value from 0 to 2π, so one would expect $\bar{\theta}$ to have a value of $\mathcal{O}(1)$. With $\bar{\theta}$ of $\mathcal{O}(1)$ the expected degree of CP violation results in a prediction of a neutron electric dipole moment of about $d_n = 8 \cdot 10^{-16} e$ cm [41], which would detectable with experimental technology available around 1980 [42][43]. This insight motivated experiments to measure the electric dipole moment of the neutron to set the degree of CP violation and determine the value of $\bar{\theta}$. Figure 3.1 shows a toy model of a neutron that violates CP.

In this toy model the neutron is imagined as two counter-rotating hemispheres with some charge distribution that is equal but of opposite sign, as depicted in the top-left of figure 3.1. In this model the total charge is 0, the angular momentum and charge distribution can be chosen to produce the experimentally observed magnetic moment $\vec{\mu}_n$, and the charge separation can be chosen to achieve an electric dipole moment $\vec{d_n}$. The axial currents implicit in the rightmost term of equation 3.2 are here shown explicitly with closed curves directed with black arrows. Applying a charge-conjugation operation (top right of figure 3.1), the baryon number is odd, but the product $\vec{\mu}_n \cdot \vec{d_n}$ is unchanged, equivalent to a rotation in space by π. Applying a parity operation to the original top-left state would flip either $\vec{d_n}$ (top-to-bottom mirror) or $\vec{\mu}_n$ (left-to-right mirror) resulting either way in negation of the product $\vec{\mu}_n \cdot \vec{d_n}$. Applying both charge-conjugation and parity transformations to the original model (bottom right of figure 3.1) also results in negation of the product $\vec{\mu}_n \cdot \vec{d_n}$, illustrating an explicit CP violation. A time-reversal operation (changing direction of the axial currents) is equivalent to a CP operation, consistent with universal CPT symmetry. This toy model shows the relation between the existence of a neutron electric dipole moment (EDM) and CP violation, but is a deliberate simplification. A deeper discussion of the topic is worthy of entire textbooks [45].

After about 20 years of research every measure of the neutron dipole moment has yielded results consistent with a value of d_n equal to exactly zero, with progressively tighter constraints on the upper bound. The latest, most stringent limit set on the neutron dipole

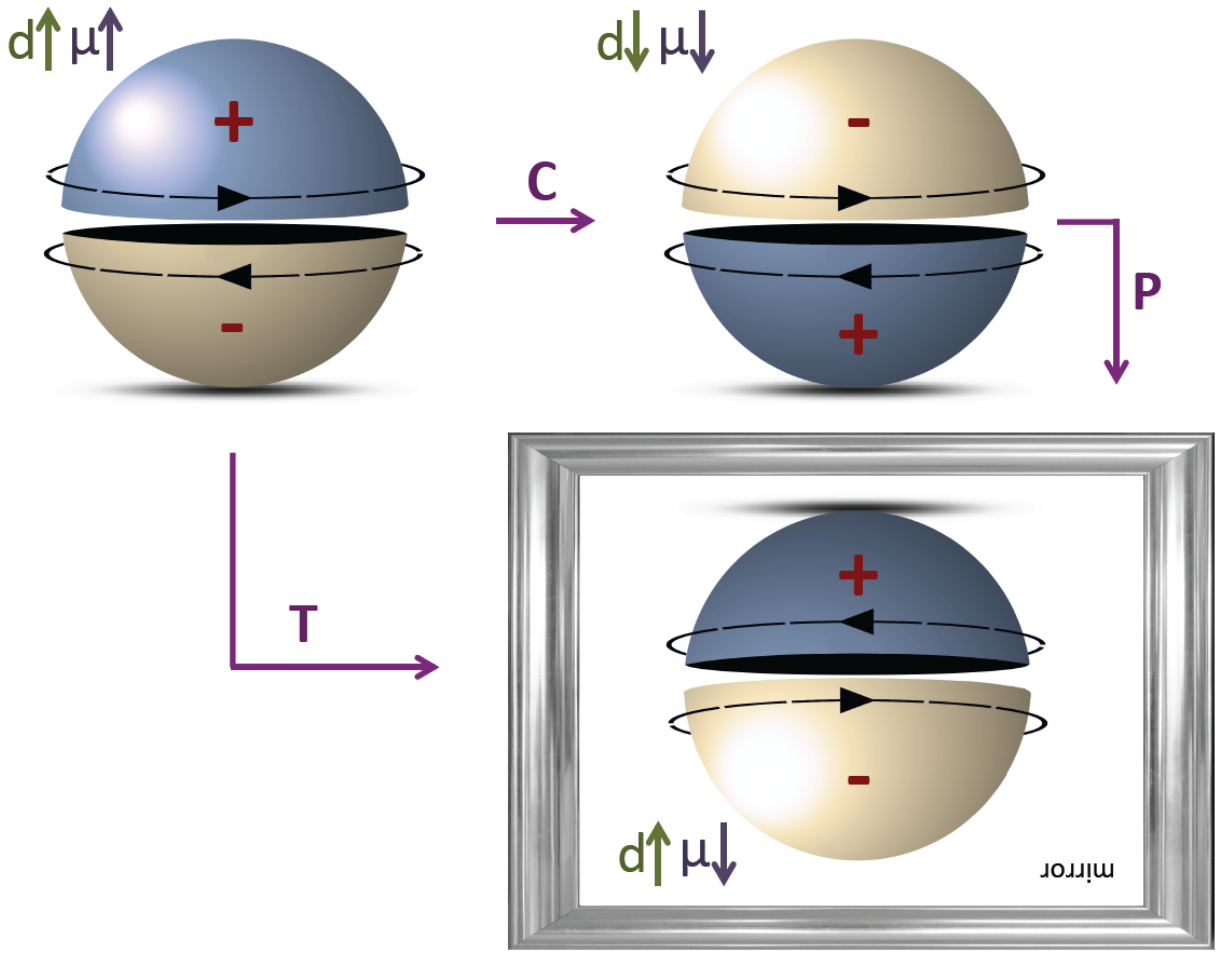

Figure 3.1: Toy model of a CP-violating neutron. Electric charge is indicated in red, electric dipole moment $\vec{d_n}$ is indicated with the green vector, and magnetic dipole moment $\vec{\mu_n}$ is indicated with the purple vector. The original toy neutron model is shown in the top-left, and symmetry operations are indicated with violet arrows and letters. "C" is charge-conjugation, "P" is parity (mirror flip in this example), and "T" is time-reversal. The transformation C + P is equal to T. Reproduced from [44].

moment is $(-7.0 < d_n < 5.0) \cdot 10^{-26} e \cdot \mathrm{cm}$ [36], some *ten billion* times smaller than expected, constraining $\bar{\theta}$ to $\left|\bar{\theta}\right| < 10^{-10}$. While no particular value of $\bar{\theta}$ is excluded by theory, the discovery that nature has chosen a value for $\bar{\theta}$ so extraordinarily close to the special value of 0 calls out for explanation. This is the "Strong CP Problem", one of the ranking outstanding problems in modern physics. One possible explanation is that one of the quark masses is equal to zero, which would make $det\, M$ undefined and $\bar{\theta}$ unobservable. However, it is well established experimentally that all quarks have a non-zero mass. Another solution is to introduce a spontaneous CP violation such that bare θ (from the QCD vacuum) vanishes and the quark mass matrix M is carefully re-worked such that $arg\, det\, M = 0$. [46][47] While this can't be discounted, it has the flavor of replacing one fine-tuning problem with two. The solution to the strong CP problem we will consider in more depth here is the Peccei-Quinn solution, which allows $\bar{\theta}$ to take on any value *dynamically*, rather than being a fixed parameter of the theory, and introduces a mechanism that naturally relaxes $\bar{\theta}$ to zero.

3.1.2 The Peccei-Quinn Solution to the Strong CP Problem

In 1977 Roberto Peccei and Helen Quinn proposed a solution to the strong CP problem by way of a new global U(1) chiral symmetry that was spontaneously broken in the early universe, in a theoretical treatment similar to that of the Higgs boson. [48] [49] In the very early universe at temperatures above Λ_{QCD} (about 200 MeV, or a few 10's of microseconds after the big bang) the $U(1)_{PQ}$ potential has the well-known "Mexican hat" form of

$$V(\phi) \propto \left(|\phi|^2 - f_{PQ}^2/2\right)^2, \tag{3.4}$$

where ϕ is a new complex pseudo-scalar field and f_{PQ} is the symmetry breaking scale. A cartoon of equation 3.4 is shown in figure 3.2. The pseudo-scalar ϕ is a dynamical variable and can in general take on any value at each point in space, just like (for example) the electric field. In the Peccei-Quinn solution, the "angle" $\bar{\theta}$ in equation 3.2 is the argument of the complex value ϕ. In a very hot, young, universe, the bump at the center of the hat is insignificant relative to the typical thermal energy scale, so the expectation value of $|\phi|$ is zero, and all values of $\bar{\theta}$ are equally likely. As the universe cools below f_{PQ}, the field tends towards the ring-shaped locus of lowest potential energy and ϕ develops a non-zero vacuum expectation value of $|\phi| = f_{PQ}/\sqrt{2}$, still with any value for $\bar{\theta}$ equally likely. This is known as a "spontaneously broken symmetry": the potential $V(\phi)$ remains symmetrical, but a particular value for $\bar{\theta}$ is spontaneously chosen from a continuum of possible values.

Now every point in space has a dynamical value of ϕ with equal magnitude and varied $\bar{\theta}$. One can imagine a constant $|\phi|$ and constant $\frac{d\bar{\theta}}{dt}$ corresponding to the blue ball in figure 3.2 making circles along the surface of lowest energy. Once can make an analogy to the time-evolution of a classical pendulum in zero gravitational field, which will just spin around its axle. Every point in space has a similar time-evolution of ϕ, but may vary in phase. This time-evolution of ϕ varying across space corresponds to a massless Nambu–Goldstone boson propagating at light speed.

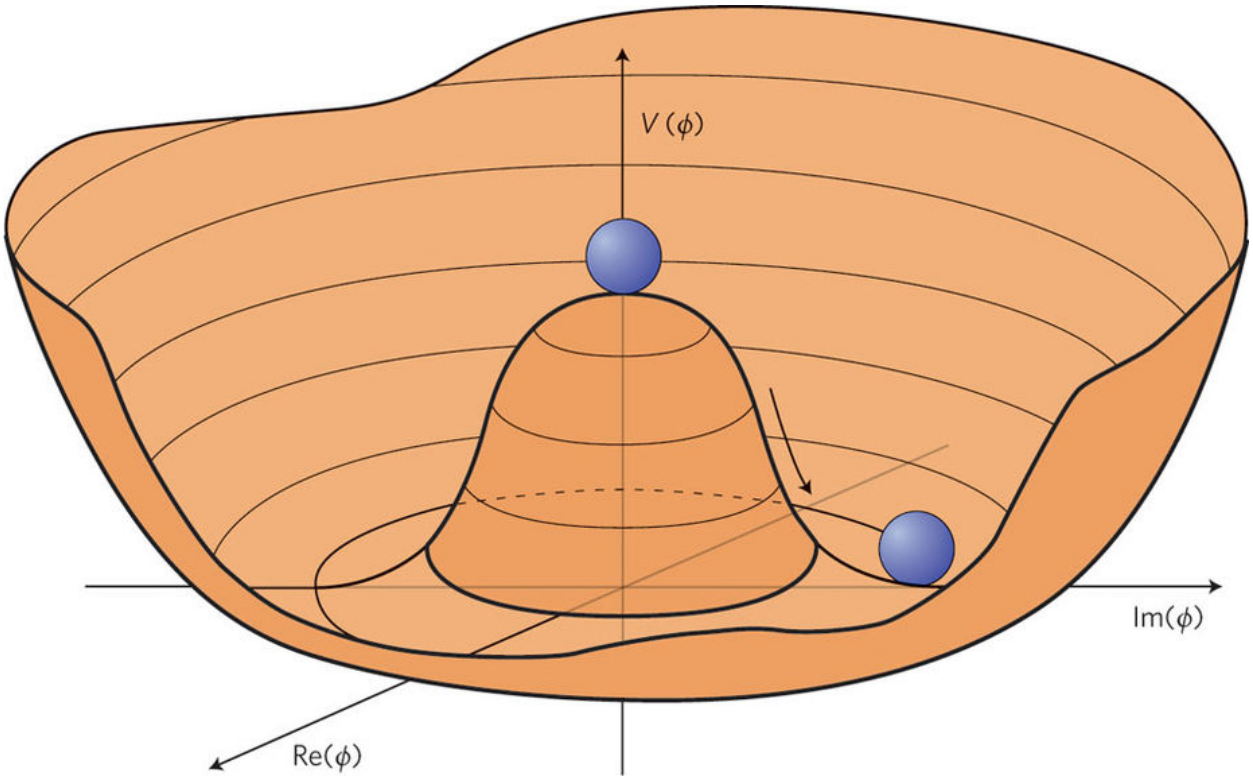

Figure 3.2: A cartoon of the Peccei-Quinn "Sombrero" potential. In the very young, hot universe, the center hump of $V(\phi)$ is thermally washed out and $\langle\phi\rangle = 0$. As the universe cools, ϕ settles to the ring-like locus at the minimum of $V(\phi)$. ϕ "running laps" around the flat bottom corresponds to a massless particle. If the hat acquires a tilt, perhaps due to quantum instanton effects, ϕ will oscillate around the new lowest point, with the natural frequency of oscillation corresponding to the rest mass of the quantized particle of the ϕ field (*ie* the Axion). Figure adapted from [50].

Our story is but microseconds old when the temperature of the universe cools further to around $T \propto \Lambda_{QCD}$, about the temperature where two and three quark bound states (mesons and baryons) are formed. At this temperature, the degree of CP-violation at different values of $\bar{\theta}$ results in quantum instanton effects that effectively tip the potential in a preferred direction, as illustrated in figure 3.3. The instanton effects introduce a new CP violating term to the Lagrangian as in

$$\mathcal{L}_{QCD} = \sum_{f=u,d,s,c,t,b} \bar{q_f}\left(i\gamma^\mu D_\mu - m_f\right) q_f - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \frac{\bar{\theta}}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{\mu\nu a} + C \frac{\phi}{32\pi^2} F^a_{\mu\nu} \tilde{F}^{\mu\nu a}, \quad (3.5)$$

which is the same as equation 3.2 but with one extra term where C is a model dependent

Figure 3.3: A Tilted Mexican Hat. Theorists disagree on whether there is actually a marble in the brim of the hat of famous mariachi artist Pedro Fernadez, but they agree that if one were present, it would surely execute simple harmonic motion around the lowest point near his right hand, very much like an axion. Photo credit: Apodaca Representaciones [51].

constant we will return to shortly.

By grouping like CP violating terms, the parameter $\bar{\theta}$ can be redefined in light of equation 3.5 to become

$$\bar{\theta} = \theta + arg\ det\ M - C\frac{arg\ \phi}{f_{PQ}}. \tag{3.6}$$

The direction of tilt $arg\ \phi$ due to instanton effects that minimizes $V(\phi)$ is

$$arg\ \phi = \frac{f_{PQ}}{C}\left(\theta + arg\ det\ M\right), \tag{3.7}$$

such that $\bar{\theta} = 0$ precisely where $V(\phi)$ is minimized. The very asymmetry that tilted the potential $V(\phi)$ ensures that a value of $\bar{\theta} = 0$ minimizes $V(\phi)$, so $\bar{\theta} = 0$ is the favored value.

Thus if one posits a universe in which $\bar{\theta}$ is a dynamic variable[5] rather than some fixed value, nature automatically favors $\bar{\theta} = 0$, and the paradox of strong CP symmetry is entirely resolved.

One might be bothered by the idea of $\bar{\theta}$ varying from moment to moment and place to place. Can the degree of CP violation in the strong force actually vary from place to place and moment to moment? The idea could make one feel a bit unmoored. We suggest an analogy to flat space-time: most physical models carry the implicit assumption of space-time that is homogeneous, isotropic, and everywhere flat. This is, clocks run at the same rate

[5]A dynamic $\bar{\theta}$ is a consequence of positing the new pseudo-scalar field ϕ and the PQ potential of equation 3.4.

Figure 3.4: Peccei, Quinn, and a sombrero potential in 2017. Source: Ryan Schude for Quanta Magazine [52].

everywhere and objects accelerate only when pushed. This perfect "space-time symmetry" is a reasonable approximation to reality for most purposes, but we know it is not exact. Clocks run slightly slower when near a large collection of mass, and spontaneously accelerate towards those same masses. Objects sitting still in space can even be shaken in the absence of interactions or forces from other objects by a passing gravitational wave. These phenomena could be considered consequences of a "breaking the symmetry of space-time" that varies in both time and location. If one can accept that the very firmament of space-time is dynamical, and revel in the richness of phenomena that results from this, then we hope the idea of a dynamic $\bar{\theta}$ can also be a welcome one. In any case, it doesn't seem to bother Peccei and Quinn, shown in figure 3.4 considering a Mexican hat potential.

3.1.2.1 The Axion Predicted in Peccei-Quinn Modified QCD

Later in the same year that Peccei and Quinn proposed their modification to QCD to solve the strong CP problem, Weinberg [53] and Wilczek [54] saw that the proposed spontaneously broken symmetry predicts that $\bar{\theta}$ will indeed naturally take an average value of $\langle\bar{\theta}\rangle = 0$, but $\bar{\theta}$ can also undergo simple harmonic motion (SHM) about its average value, implying the

existence of a new pseudo-Nambu-Goldstone boson with a rest mass corresponding to the frequency of SHM.

One can make an analogy between the dynamics of $\bar{\theta}$ of the PQ model and the angle θ of a classical pendulum at small angles, which behaves like a simple harmonic oscillator (SHO). A classical pendulum follows an equation of motion of the form $\theta(t) = \theta_0 e^{i\omega t}$. The amplitude θ_0 depends on initial conditions and the frequency ω depends on the length of the pendulum l and acceleration due to gravity g by $\omega = \sqrt{\frac{g}{l}}$. In this analogy, the "phase particle" represented by the blue ball in figure 3.2 rolling along the tilted circular track of the potential follows similar dynamics by the following analogies:

$$\begin{aligned}
&\theta_{SHO} \leftrightarrow \bar{\theta},\\
&\omega_{SHO} \leftrightarrow \omega_a,\\
&\omega_{SHO} = \sqrt{\frac{g}{l}},\\
&\quad g \leftrightarrow \text{tilt of hat potential} \propto \text{strength of instanton effects},\\
&\quad l \leftrightarrow |\phi| = f_{PQ}/\sqrt{2},\\
&\quad \text{and}\\
&\omega_a = \sqrt{\frac{\text{strength of instanton effects}}{f_{PQ}/\sqrt{2}}}.
\end{aligned} \tag{3.8}$$

The value ω_a gives us the axion mass. An axion (or any particle) at rest and at zero potential will have a time-dependent portion of its wave function equal to $e^{i\omega_a t}$ where the rate of phase evolution is given by the energy $\omega_a = E/\hbar$, and E is equal to the rest mass $m_a c^2$. Thus we have:

$$\begin{aligned}
\omega_a = E/\hbar = m_a c^2/\hbar &= \sqrt{\frac{\text{strength of instanton effects}}{f_{PQ}/\sqrt{2}}},\\
\text{or}&\\
m_a &= \frac{\hbar}{c^2}\sqrt{\frac{\text{strength of instanton effects}}{f_{PQ}/\sqrt{2}}}.
\end{aligned} \tag{3.9}$$

The realization that the PQ proposal predicts a new particle was of great importance because it created a testable prediction, raising the possibility of promoting the PQ modification from "interesting idea" to "scientific theory". This new particle was called either the "higglet" for its similarity in theoretical underpinnings to the Higgs boson, or the "Axion" for both the axial currents in the theory and its ability to clean up the strong CP problem, much like the kitchen soap by the same name can clean up tough, baked-on grease. The name "Axion" won out, and is in common use today (see figure 3.5).

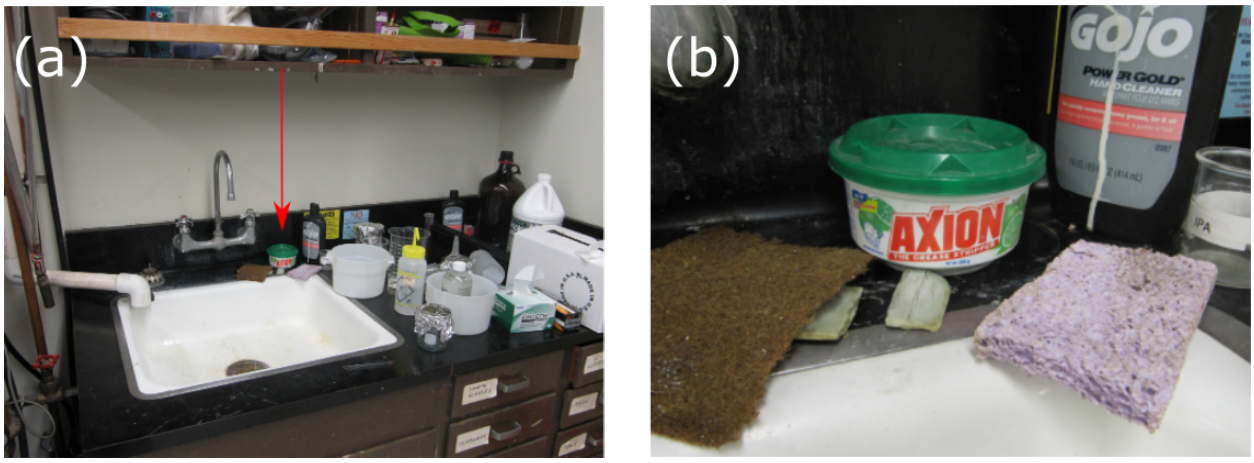

Figure 3.5: Axion observed in 2017 in a basement laboratory at UCB. The Axion has recently been observed at UC Berkeley, among a disused lab sink deep in the second basement of Birge hall. Initial data suggest a non-virialized velocity distribution and highly non-homogenous density, so universal abundance remains an open question and no competing DM candidates have yet been excluded. Initial experimental trials indicate that even 10 years after the expiration date, Axion remains both an excellent degreaser and a promising solution to the strong CP problem. (a) Axion and surroundings. (b) Detail of Axion.

Upon announcement of a new hypothesized particle, high-energy physicists are honor-bound to hunt for it with their particle accelerators. The axion as originally conceived (the Peccei-Quinn-Weinberg-Wilczek, or PQWW axion) was associated with the weak energy scale resulting in a predicted axion mass of 10 to 100 keV [55]. A heavy axion of this type would be evident in meson decays ($J/\Psi \rightarrow \gamma a$, $K^+ \rightarrow \pi^+ a$, $\Upsilon \rightarrow \gamma a$), nuclear de-excitations ($N^* \rightarrow Na$), and beam dump experiments ($a \rightarrow \gamma\gamma$, $a \rightarrow e^- e^+$). However, accelerator experiments failed to detect any of these generation or decay pathways, ruling out massive axions [56], [57].

Implicit in the first theories of the axion was a small symmetry breaking scale f_{PQ} resulting in a "small radius" for oscillations in $\bar{\theta}$, a higher frequency of SHM, and a large axion mass. The symmetry breaking scale f_{PQ} remains a free parameter of the theory, and by setting f_{PQ} to a large value the axion is made "invisible" (weakly coupled) with a mass given by

$$m_a = \frac{(m_u m_d)^{1/2}}{m_u + m_d} \frac{f_\pi}{(f_{PQ}/N)} m_\pi \approx 0.62eV \frac{10^7 GeV}{(f_{PQ}/N)}, \tag{3.10}$$

where m_u (m_d) is the mass of the up (down) quark, f_π is the pion decay constant, N is the PQ color anomaly and m_π is the pion mass. This "invisible" axion would indeed not be evident in particle accelerator experiments but some predicted interactions still remain. The interaction relevant to the work described here is given by the Lagrangian

$$\mathcal{L}_{a\gamma\gamma} = -g_{a\gamma\gamma}\phi_a \vec{E} \cdot \vec{B} \tag{3.11}$$

that describes the decay of an axion into two photons, where $g_{a\gamma\gamma}$ is the axion-to-two-photon coupling strength, ϕ_a is the axion field, $\vec{E}$ is the electric field and $\vec{B}$ is the magnetic field. The factor $g_{a\gamma\gamma}$ is given by

$$g_{a\gamma\gamma} = \frac{\alpha g_\gamma}{\pi(f_{PQ}/N)}, \tag{3.12}$$

where α is the fine structure constant and g_γ is a model dependent constant of $\mathcal{O}(1)$. If one is to detect the "invisible axion" via this interaction, it helps to calculate the interaction strength g_γ. The axion has no charge and cannot interact with photons directly, so to first order $g_\gamma = 0$. Another way to say this is that the interaction depicted in the left Feynman diagram in figure 3.6 is "suppressed at tree level". The axion-photon interaction is made possible by replacing the node with an "anomalous fermionic loop" shown on the right of figure 3.6 where "f" generically labels some fermion. The strength of the interaction depends on precisely which fermions can take part in this process.

There are two principle theoretical models for this process. The Kim [59], Shifman, Vainshtein, and Zakharov (KSVZ) [60] model is a "hadronic" model that allows coupling only to quarks and predicts $g_\gamma = -0.97$. The Dine, Fischler, Srednicki [61], and Zhitnitsky [62] (DF SZ) model allows coupling to both quarks and lighter leptons, and predicts $g_\gamma = +0.36$. These two models essentially define the extremes of "invisible" axion theory with only one remaining free parameter: the symmetry breaking scale f_{QP} and by extension, the axion mass m_a (equation 3.10) and double-photon coupling $g_{a\gamma\gamma}$ (equation 3.12). There are no constraints on f_{QP} from within the theory, but astrophysical observations give an upper limit on double-photon coupling, and cosmological considerations place a lower limit on the axion mass. We will explore these limits in due course.

3.2 The Axion as a Dark Matter Candidate

The Axion is a compelling theoretical candidate for explaining the "Strong CP Problem". Experimental evidence that axions exist (for example, creation of a few trace axions in a particle accelerator) would be sufficient to settle the question of the strong CP problem. However, any viable candidate for the constituency of dark matter must be not just a previously unknown exotic particle, but must also be extremely common (about 5 times greater total cosmological mass density than ordinary matter) and slow-moving.

To explore whether axions are a plausible solution to the dark matter problem, we must explore the nature of their cosmological origin, and whether the mechanisms of axion creation in the early universe could generate axions both in the correct quantities and in the correct spacial and momentum distribution to match the known properties of dark matter.

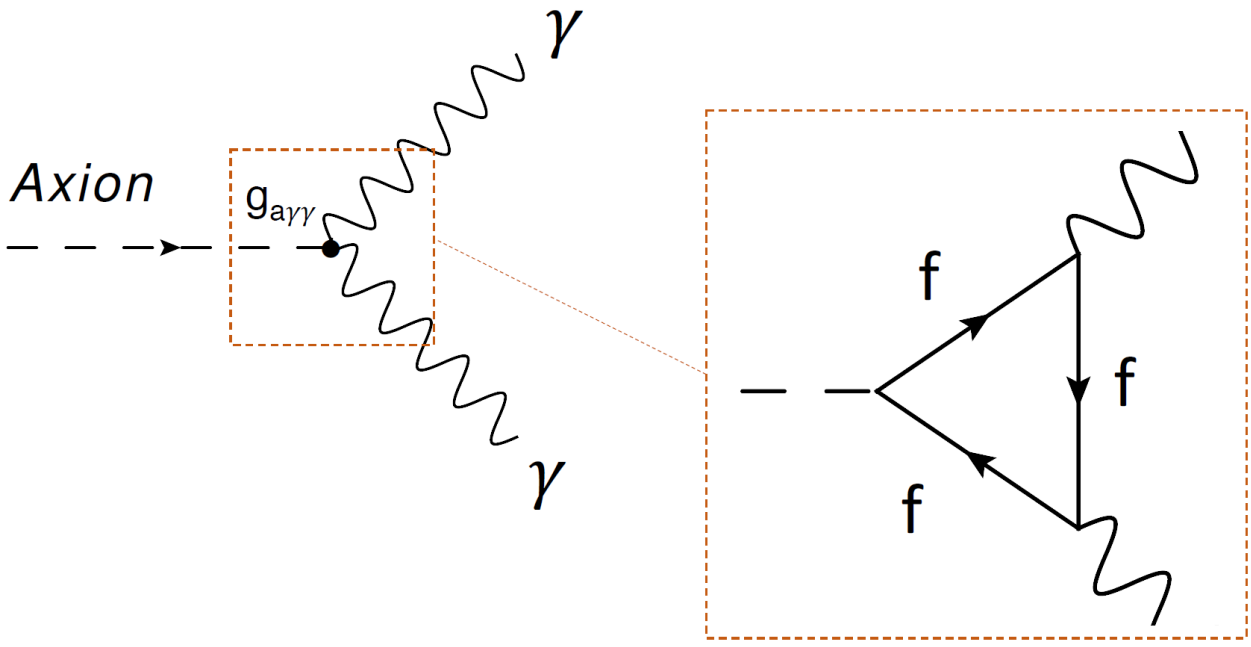

Figure 3.6: Feynman diagram for an axion decaying to two photons via a fermionic loop. An axion cannot directly decay into photons because it has zero charge (the decay is "suppressed at tree level"). Replacing the node with an "anomalous fermionic loop" allows an $a \rightarrow \gamma\gamma$ reaction, albeit at very low amplitude. The exact formulation of the allowed fermionic loops predicts the rate of decay. Particular models differ by about a factor of three. Reproduced from [58].

3.2.1 Cosmological Origins of the Axion

There are three ways Axions could have conceivably been produced in the early universe: thermal production, cosmic string or domain wall decay, and the "misalignment mechanism".

3.2.1.1 Thermal Origins

The thermal process is the same process giving rise to all ordinary matter, in the cosmic particle accelerator that was the "Quark-Gluon Plasma" and "Atomic Plasma" eras of the early universe. In this model all matter and energy fields are in thermal equilibrium, so (for example) electron-positron annihilation is in equilibrium with electron-positron production from thermally polarized vacuum fluctuations. As the universe cools and becomes less dense, the matter and energy fields decouple and fall out of thermal equilibrium, resulting in the fortunate state of our present universe with abundant free energy. The axion coupling strength is proportional to its mass, so the requirement of sufficient axion coupling strength to maintain thermal equilibrium in the early universe sets a mass bound that $m_a > 1$

meV. However, astronomical observations of supernova (described in section 3.2.2) put a contradictory bound of $m_a < 1$ meV, so the existence of axions of thermal origin is ruled out. If on the other hand the axion mass is indeed $m_a < 1$ meV and, by some mechanism currently thought impossible, was and is near thermal equilibrium with the rest of the universe ($T_a \approx T_{CMB} = 2.7K$), then the axion V_{RMS} would be $0.6c$ or greater, inconsistent with the cosmological observation that the great majority of mass in the universe is "cold" in the astronomical sense of slow-moving.

3.2.1.2 Cosmic String and Domain Wall Origins

The cosmic string or domain wall decay model arises from the observation that any physical model of spontaneous broken symmetry allows for the formation of domain walls or cosmic strings. In the case of axions, where the broken symmetry is of the U(1) type (every point in space has a value of $arg\phi \in [-\pi, \pi]$) cosmic-sized discontinuities in $arg\phi$ resolve in to linear structures where $|\phi| = 0$ and $|\phi| \neq 0$ outside the string and with the value of $arg\phi$ winding around by 2π along any closed path around the string. Topologically, the strings must from closed loops, but the radius of these loops has no upper limit and could be expected to be larger than the observable universe. Such strings would form at the time of Pecci-Quinn (PQ) spontaneous symmetry breaking (SSB), which could have been before or after the era of cosmic inflation. If PQ SSB occurred before or during inflation, any resulting cosmic strings would be redshifted beyond our relativistic horizon and effectively banished to non-existence. If PQ SSB occurred after inflation, cosmic strings would exist in our universe and decay by emission of axions. The change in axion number density per entropy density is given by

$$\Delta\left(n_a/s\right) \sim \frac{\mu/t^2}{\omega T^3}\Delta H t, \tag{3.13}$$

where $\mu \sim f_{PQ}^2 \ln\left(f_{PQ} d\right)$ is the linear mass density of axionic cosmic strings and d is the typical string spacing, ω is the average energy per radiated axion, T is the temperature, t is time, and H is the Hubble parameter. The total number of axions produced from string decay is then[6]

$$\frac{n_a}{s} \sim \mu \int_{T_1}^{f_{PQ}} \frac{dT}{\omega m_{Pl}^2}, \tag{3.14}$$

where T_1 is the temperature at which the axion mass becomes comparable to the expansion rate, and $m_{Pl} = \sqrt{\hbar c/G}$ is the Planck mass. Integrating equation 3.14 requires defining a function for ω, and here theorists are not of one mind. Davis [63, 64] argues that emitted axions would have wavelengths on the order of the cosmic horizon, so $\omega(t) \sim t^{-1}$, while Harari and Sikivie [65] argue that the radiated axions have a $1/k$ spectrum, so $\omega(t) \sim \ln\left(f_{PQ} t\right) t^{-1}$, which is larger by a factor of about 100.

[6]It is correct to consider only the number of axions produced and not total axion energy because any axion momenta will have by today been entirely redshifted away.

In either case, the total contribution of axion mass to the cosmic mass density is

$$\Omega_S h^2 \simeq [1 \text{ or } \ln(f_{PQ}t_1)]\,(m_a/10^{-3}\text{eV})^{-1.18}, \tag{3.15}$$

and we know that the mass density of the universe (and by extension the axion) is less than 1, so if Davis is correct, $m_a \geq 1$ meV, while if Harari ans Sikivie are correct, $m_a \geq 1$ μeV. Astronomical observations (section 3.2.2 rule out $m_a > 1$ meV, so cosmic-string axions of $m_a > 1$ μeV from cosmic string evaporation may be present today.

The "misalignment mechanism" of axion production is the favored method and was briefly touched upon in section 3.1.2. In the very early universe, the pseudo-scalar axion field ϕ has an average value of $\langle\phi\rangle = 0$ and the effective potential $U(\phi)$ is symmetrical in $\arg\phi$ and quadratic in $|\phi|$. As the universe cools to $T \leq f_{PQ}$, the potential $U(\phi)$ resolves to the famous "sombrero potential" shown in figure 3.2 with the form of equation 3.4. In this potential landscape, the "ϕ particle" finds the potential minimized for $\left\langle\left|\vec{\phi}\right|\right\rangle = f_{PQ}/\sqrt{2}$, and with the free parameter of $\arg\phi \equiv \bar{\theta}$ evolving freely, ϕ "runs laps" around the bottom track of the potential, corresponding to propagation of a massless particle. As the universe cools further to $T \leq \Lambda_{QCD}$ instanton effects increasingly tilt the potential towards $\bar{\theta} = 0$, resulting in harmonic motion of $\bar{\theta}$ around $\bar{\theta} = 0$, with the oscillation frequency corresponding to the axion mass, and the oscillation amplitude corresponding to the axion density. The temperature dependence of axion mass on temperature is approximately [66]

$$m_a(T) \simeq 0.1 m_a(T=0)(\Lambda_{QCD/T})^{3.7}, \tag{3.16}$$

valid for $\pi T/\Lambda_{QCD} \gg 1$.

The equation of motion for $\bar{\theta}$ is then

$$\ddot{\bar{\theta}} + 3H\dot{\bar{\theta}} + m_a^2(T)\bar{\theta} = 0, \tag{3.17}$$

where H is the Hubble parameter so the dynamics of $\bar{\theta}$ are much like a damped harmonic oscillator where the Hubble expansion applies damping and the mass is a function of the universe temperature T. This system is overdamped and non-oscilatory for $m_a < 3H$, so the era of onset of axion oscillations (ie the era when the axion mass "turns on") is a function of both $m_a(T)$ and H(t). When coherent oscillations commence, we can replace $\dot{\bar{\theta}}^2$ with the average over one cycle: $\langle\dot{\bar{\theta}}^2\rangle = \rho_a$ where the axion density ρ_a is

$$\rho_a = \frac{f_{PQ}^2}{2}\left(\dot{\bar{\theta}}^2 + m_a(T)^2\bar{\theta}^2\right). \tag{3.18}$$

Multiplying equation 3.17 by $\dot{\bar{\theta}}$ and replacing $\langle\dot{\bar{\theta}}^2\rangle$ by ρ_a in equation 3.18 yields the expression

$$\dot{\rho_a} = \left[\frac{\dot{m_a}}{m_a} - 3H\right]\rho_a \tag{3.19}$$

which has the simple solution

$$\rho_a = \text{const}\frac{m_a(T)}{a(t)^3}, \tag{3.20}$$

where $a(t)$ is the size-scale of the universe. Thus the number of primordial axions in a comoving volume of space is constant even as the axion mass is settling to its final value. The initial axion number density is

$$n_a(T_1) \simeq m_a(T_1)\bar{\theta_1}^2(f_P Q^2)/2, \tag{3.21}$$

where T_1 is the temperature when $\bar{\theta}$ first starts oscillating and $\bar{\theta_1}$ is the "misalignment" value of $\bar{\theta_1}$ prior to the onset of oscillations. The value of T_1 is found by solving $m_a(T_1) = 3H(T_1)$ given what we know about H(t) and T(t) in the early universe from (for example) analysis of the CMB. Critically, T_1 scales as $m_a(T=0)^{0.18}$. This means that for a larger present day axion mass [$m_a \equiv m_a(T=0)$], the oscillation onset temperature T_1 is hotter and $m_a(T_1)$ is (by equation 3.16) smaller, for a smaller initial n_a and ρ_a. *The more massive the axion is, the less total axion mass-energy would be created in the early universe!* By our best estimates [67], the total mass-energy of axions created via the misalignment mechanism as a fraction of the critical density is

$$\Omega_{MIS}h^2 = 0.13 \times 10^{\pm 0.4}\Lambda_{200}^{-0.7} f(\bar{\theta_1})\bar{\theta_1^2}(m_a/10^{-5}\text{eV})^{-1.18} \tag{3.22}$$

where h is the reduced Hubble constant[7], $\Lambda_{200} \equiv \Lambda_{QCD}/200$ MeV, $f(\bar{\theta_1})$ is a function of $\mathcal{O}(1)$ to account for anharmonicities of a simple pendulum, and the factor of $10^{\pm 0.4}$ reflects theoretical uncertainties in (for example) $m_a(T)$ or H(t). As there is no preferred value for $\bar{\theta_1}$, we may draw values from a flat distribution of $\bar{\theta} \in [-\pi, \pi]$ for a typical value of

$$(\bar{\theta_1})_{rms} = \pi/\sqrt{3}. \tag{3.23}$$

This is a valid figure for an expected value of $\bar{\theta_1}$ regardless of whether f_{PQ} is greater or less than the inflationary energy scale. Our current observable universe is comprised of about 10^{30} volumes that were causally distinct at the time the axion field started oscillating, so $(\bar{\theta_1})_{rms}$ is truly a robust statistical average of the values of $\bar{\theta_1}$ just after inflation. If f_{PQ} is instead so large that the PQ symmetry was broken prior to the end of the inflationary era, then our entire observable universe has the same singular value of $\bar{\theta_1}$, and $(\bar{\theta_1})_{rms} = \pi/\sqrt{3}$ is then a typical expected value, rather than a hard predicted value. Plugging $\bar{\theta_1} = \pi/\sqrt{3}$ and $m_a = 10^{-5}$ eV into equation 3.22 yields $\Omega_{MIS} = 0.85$, only a few times greater than the known mass-density of dark matter, and within the bounds of theoretical uncertainty.

Axions produced by this misalignment mechanism in the early universe would today be very cold indeed. If f_{PQ} is greater than inflationary energy scales, then any pre-inflationary

[7]The reduced Hubble constant $h \equiv H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ is not to be confused with Planck's constant $h \approx 6.267 \times 10^{-34} J \cdot s$, or the Helen h, a unit of beauty sufficient to launch 1000 ships. The tension between the appearance of h and h (the first two) remains unresolved in the literature, while the tension caused by h should it happen to appear to the author would likely prove both unresolved and awkward.

region of space with a particular value of $\bar{\theta_1}$ was expanded far beyond the edge of today's observable universe during inflation so post-inflation $\bar{\theta_1}$ was homogenous everywhere causally connected to us, so $\nabla\bar{\theta}$ was nearly zero and the primordial axion momentum is vanishingly small. On the other hand, if f_{PQ} is smaller than the inflationary energy scale, then axions could have been "born hot" with a characteristic momentum $p_1 \sim H_1^{-1}$, (that is, moving fast enough to cross the visible universe in the age of the universe) and from this and the initial axion number density (from equation 3.21) we can estimate the axion phase space density as

$$f_a(\vec{p_a} = 0) \sim \frac{m_a(T_1) f_{PQ}^2}{p_1^3} \sim 10^{50}(m_a/\mathrm{eV})^{-2.7}, \tag{3.24}$$

which is a Bose-Einstein condensate by any measure. Alternatively, we can estimate the typical axion velocity today as

$$v_a/c \sim \left(\frac{p_a}{m_a}\right)_{today} \sim \frac{p_1(2.7K/T_1)}{m_a} \sim 10^{-22}(m_a/\mathrm{eV})^{-0.82}, \tag{3.25}$$

which, for $m_a = 10^{-6}$ eV is about equal to a typical human hair growth rate– sufficiently slow to be gravitationally bound to galaxies and galaxy clusters.

3.2.2 Bounds on the Axion Mass & Abundance

There is nothing in the fundamental theory of the axion to suggest what its mass should be—f_{PQ} remains a free parameter. However, based on observational evidence, we *can* put both upper and lower bounds on the axion particle mass. From equations 3.10 and 3.12 we see that a more massive axion has a stronger interaction, so analysis of certain high-energy events, such as supernova 1987a and the stellar dynamics of red giant stars, allow us to put an upper bound on the axion interaction strength, and thus mass. Equations 3.15 and 3.22 indicate that a lighter axion mass results in a greater total axion mass-density, so by cosmological arguments for a maximum universal mass density, we can put a lower bound on the axion mass. A summary of observational bounds to the axion mass is given in figure 3.7.

3.2.2.1 Observational Bounds (SN1987a)

A tried and true way of "discovering" a new particle is to generate it in a particle accelerator. No axion has ever been generated in a terrestrial particle accelerator, but nature provides a test laboratory with access to far greater energies than achieved by human technology.

About 160,000 years ago, a blue supergiant star in the Large Magellanic Cloud once known as Sanduleak-69 202 collapsed and exploded in a spectacular supernova. In 1987, the enormous flux of light and neutrinos from this event reached Earth. Had this occurred in our own stellar neighborhood, it would have put a definitive end to speculations on axions,

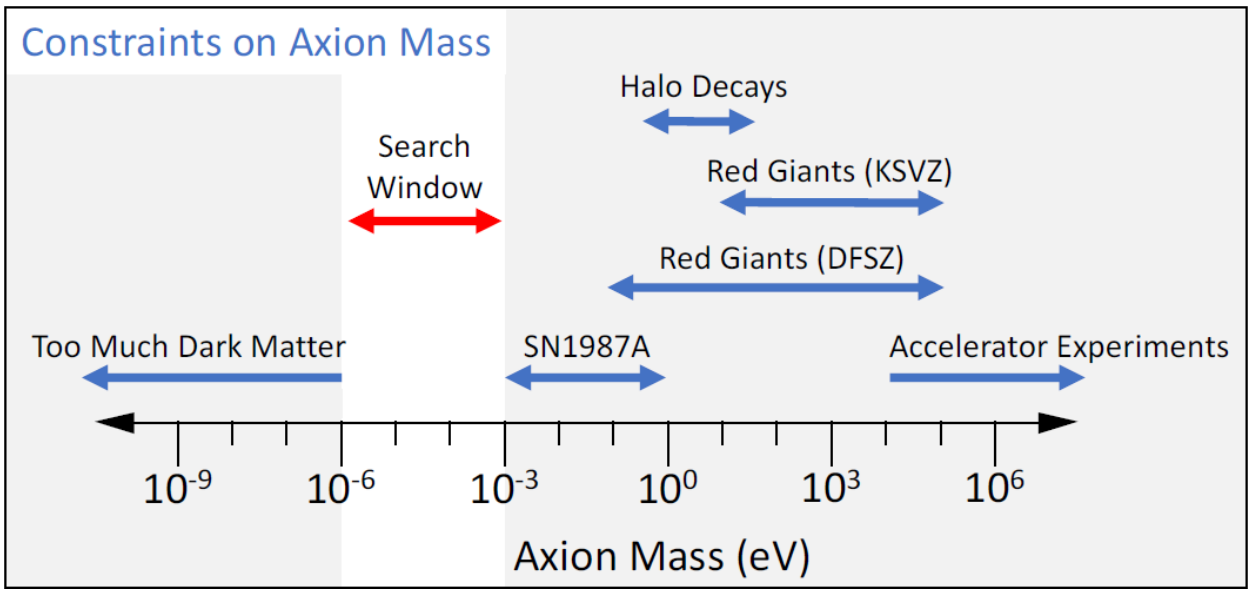

Figure 3.7: Axion Mass Constraints. The Axion mass is unknown, but various indirect observations allow us to establish some upper and lower bounds. The lighter the axion, the more total mass of axions would have been created in the early universe. The cosmological observation that $\Omega_M \leq 1$ tells us that $m_a \geq 1\ \mu$eV. Greater axion mass means greater interaction strength, so the existence of axions would be revealed in various high-energy dynamics. The most stringent bound is set by SN1987a, with progressively higher-mass axions ruled out by failure to observe their effects in progressively lower-energy events, including red giant stellar dynamics and terrestrial-based accelerator experiments. The "open" search space for axion mass ranges from 1 μeV to 1 meV. Reproduced from [58]

dark matter, and the strong CP problem by putting a definitive end to all life on Earth. Fortunately, the intensity of the flash was sufficiently diminished by spreading through space that it posed no danger, yet remained sufficiently intense that we could observe the light and even catch a few neutrinos in the Kamiokande II (KII) and Irvine-Michigan-Brookhaven (IMB) neutrino detectors. The event was named SN1987a and it was a boon to science.

Nineteen neutrinos from SN1987a were detected on Earth over a span of about 12 seconds, 8 in IMB [68] and 11 in the KII [69] water Cherenkov detectors. Combined with data from visible-light astronomy, these were sufficient to put strict bounds on the time vs. temperature function at the core of SN1987a with a time resolution of about a second. Knowing the time-temperature function at the core tells us a great deal about what particle species participate in the core dynamics. As the core collapses, the temperature spikes, and photons, neutrinos, and any number of other particles are thermally generated (via Compton scattering, Bremsstrahlung, etc.) in great quantities. The photons and even neutrinos thus generated cannot immediately escape the core because their mean free path is much shorter

than the core size due to the core's extreme mass density.[8] The "trapping" of these particles means the temperature stays high and collapse is "pushed back" by radiation pressure, at least to a degree. If thermal axions with a scattering length much longer than neutrinos were generated in the core, they would free-stream out, carrying away energy and cooling the core, resulting in a much shorter-lived temperature spike, and more tightly-spaced neutrino burst detected on earth.[9] The reconstructed time-temperature function at the core of SN1987a is entirely consistent with the dynamics of known particles (3 kinds of neutrinos, etc) with no indication of "hidden" cooling channels due to axions or another undiscovered particle. This then only allows axions with an interaction length so short (massive, strongly interacting) that they cannot escape the core, or so long (light, weakly interacting) that they are not generated in great numbers. This relation of axion mass to axion cooling flux in SN1987a is illustrated in figure 3.8, and the excluded mass range included in figure 3.7.

3.2.2.2 Cosmological Bounds

Precluding the positive detection of primordial axions[10] there is no lower bound on the cosmological axion mass density. Axions may comprise all of dark matter and thus most of the mass of the universe, only a fraction of dark matter, with the remainder composed of some other exotic particle, or axions may not exist at all outside of human imagination. However, there is an upper bound on axion mass density. Cosmological observations indicate that $\Omega = 1$ and Ω_M is not more than about 0.3, so the axion mass density in turn cannot be more than about 0.3. Note from equation 3.22 that the mass density of axions produced by the misalignment mechanism *increases* with diminishing axion mass, so a sufficiently light axion implies $\Omega_M > 1$, contrary to cosmological observations. The lightest axion allowed by this reasoning is about 1 μeV. This is not a hard lower bound on axion mass: note that equation 3.22 is proportional to the square of the initial misalignment angle $\bar{\theta_1}$ with $-\pi < \bar{\theta_1} < \pi$ and no preferred direction for $\bar{\theta_1}$. If the Pecci-Quinn spontaneous symmetry breaking occurred after the end of inflation, then $\bar{\theta_1}$ throughout space would be probabilistically selected from a flat distribution of values, and $\left\langle \bar{\theta_1}^2 \right\rangle = \pi^2/3$. On the other hand, if the Pecci-Quinn spontaneous symmetry breaking occurred before the end of inflation, our entire visible universe would have had a single homogeneous value of $\bar{\theta_1}$ at the end of the inflationary era, and $\left\langle \bar{\theta_1}^2 \right\rangle = \pi^2/3$ becomes merely a "reasonable guess" rather than a hard prediction. In this second case, any proposal for a very light axion can be rationalized with a concurrent small value for $\bar{\theta_1}$ to avoid an "over-massive" universe, but doing so merely re-creates a fine-tuning problem almost exactly like the one that originally

[8] In contrast, about 1011 out of 1012 neutrinos crossing through the center of our planet will pass unscattered, because the rock and iron constituting Earth are considerably less dense than the matter at the core of a supernova.

[9] An interaction length much smaller than the object means that cooling will be regulated by the slow process of diffusion, while an interaction length much larger than the object meas that the coolant interacts so weakly that its production rate is negligible.

[10] Fingers crossed!

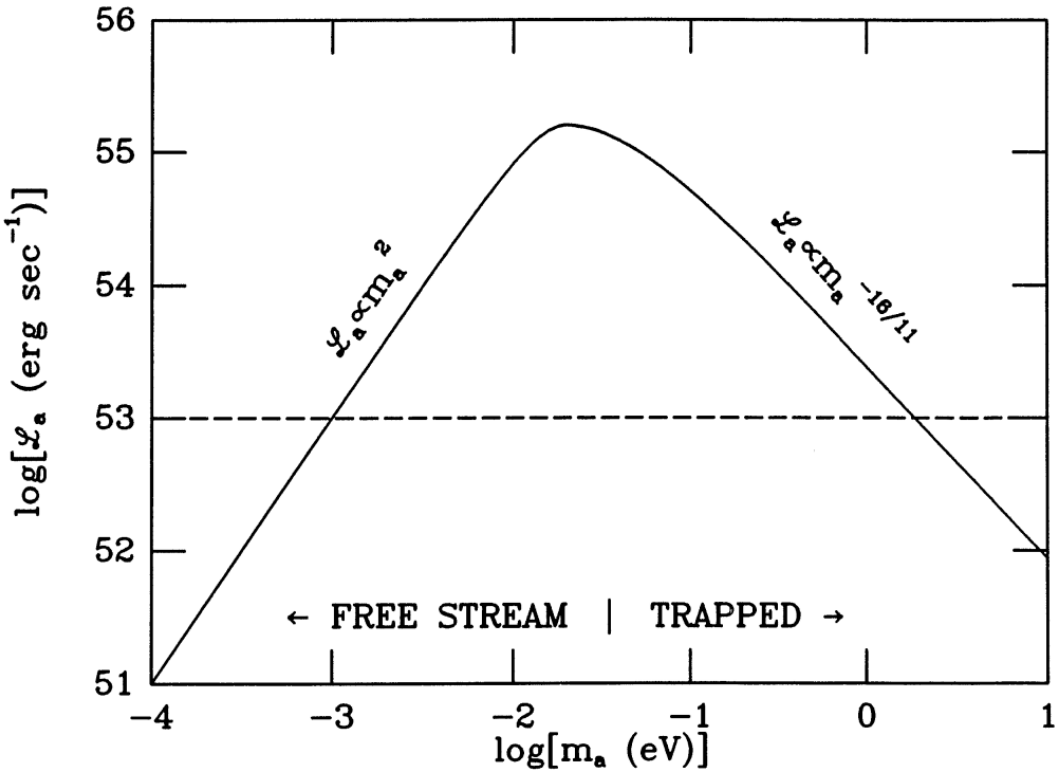

Figure 3.8: Axion luminosity from SN1987a as a function of axion mass. Astronomical observations of photon and neutrino flux set an upper bound on peak axion luminosity $\mathscr{L}_a \leq 10^{53}$ erg/s, and thus exclude the axion mass range from 1 meV to 1 eV. Axion interaction strength is proportional to axion mass. A low-mass axion created in the core of SN1987a would interact weakly with the stellar material and stream freely out of the star, acting as an auxiliary "cooling channel" to the core, much like neutrinos produced in the core of our own sun. A high-mass axion created in the core of SN1987a would interact strongly with the stellar material and diffuse only slowly to the surface, much like photons produced in the core of own sun, maintaining a high core temperature. A sufficiently light axion ($M_a < 1$ meV) would be produced at such low rates that total axion luminosity would be negligible. A sufficiently heavy axion ($M_a > 1$ eV) would be generated in great quantities but diffuse out so slowly that axion flux would again be negligible. Reproduced from [70].

inspired the idea of the axion![11] This cosmological bound on light axions is illustrated by the left hand exclution "Too Much Dark Matter" in figure 3.7.

3.2.2.3 Red Giant and Other Bounds

Cosmological considerations put a lower bound on the axion mass ($m_a > 1\ \mu$eV), and analysis of SN1987a puts the most stringent upper bound on the axion mass (1 meV$< m_a <$ 1eV), but neither excludes an absolute upper bound. A more massive axion has a stronger interaction and no axion has been directly observed in a laboratory to date, so the absolute upper exclusion range in axion mass ($m_a < 10$ keV) is set by their non-appearance in particle accelerator data. The mass range (10^{-1} eV $< m_a < 10^5$ eV), between and overlapping with the SN1987a and accelerator exclusion ranges, is excluded by astrophysical observations of phenomena far more common than SN1987a but at far greater energies and time-scales than available to terrestrial experiments–namely, stellar dynamics.

Stars exist in a dynamic equilibrium of gravitational in-fall against thermal pressure, with thermal energy radiated away at the surface matched by heat generated at the core by thermal fusion. By analyzing hydrostatic equilibrium and energy transport, we arrive at the luminosity formula originally described by Chandrasekhar [71]

$$\mathcal{L}_\gamma \propto (G\mu^7)M^5T_c^{1/2}, \tag{3.26}$$

where $\mathcal{L}_\gamma$ is the photon luminosity G is the gravitational constant, μ is the mean molecular weight per particle (mostly electrons, protons, and helium nuclei), M is the total stellar mass, and T_c is the core temperature. Nuclear reaction rates at the core are highly temperature dependent, with

$$Q_{nuc} \propto T^7. \tag{3.27}$$

In a star with no axion luminosity, the photon luminosity is equal to the rate of nuclear energy conversion: $\mathcal{L}_\gamma^0 = Q_{nuc}^0$, where 0's indicate the no-axion case.

If axions are generated at the core, they will carry energy away unimpeded by the overlying mass, and act as an 'auxiliary cooling channel" to the core. Perhaps counter-intuitively, this results in an *increase* in core temperatures and accelerated nuclear reaction rates. The argument is as follows: in the presence of an axion flux $\mathcal{L}_a$, the energy balance is modified to $\mathcal{L}_a + \mathcal{L}_\gamma = Q_{nuc}$, with $Q_{nuc} \equiv Q_{nuc}^0 + \delta Q_{nuc}$ and $\mathcal{L}_\gamma \equiv \mathcal{L}_\gamma^0 + \delta\mathcal{L}_\gamma$. From equation 3.26 $\mathcal{L}_\gamma/\mathcal{L}_\gamma^0 = 0.5(\delta T_c/T_c^0)$, and from equation 3.27 $\delta Q_{nuc}/Q_{nuc}^0 = 7(\delta T_c/T_c^0)$. Understanding that $\delta Q_{nuc} = \mathcal{L}_a + \delta\mathcal{L}_\gamma$ and defining $\epsilon \equiv \mathcal{L}_a/Q_{nuc}^0$, we have

[11] The over-production of light axions cannot be fixed by "dilution by inflation". Recall that axions have no mass before the "mexican hat" potential tilts, which occurs at a temperature/energy around Λ_{QCD}, an era well *after* inflation.

$$\begin{aligned} \delta T_c/T_c^0 &= \epsilon/6.5, \\ \delta \mathcal{L}_\gamma/\mathcal{L}_\gamma^0 &= \epsilon/13, \\ \delta Q_{nuc}/Q_{nuc}^0 &= 14\epsilon/13, \\ \delta R/R^0 &= -\epsilon/6.5, \end{aligned} \tag{3.28}$$

where the stellar radius $R \propto T_c^{-1}$ from hydrostatic equilibrium. Turning on an extra "axion cooling channel" would cause a star to contract, heat up, increase the core nuclear reaction rate, and become more luminous. This has the consequence of predicting a shorter stellar lifetime for any star, but due to the strong temperature dependence of the nuclear reaction rate, the greatest effect would be seen in red giant stars, which have much hotter, helium-burning cores than the hydrogen-burning cores of main-sequence stars like Sol. We do not observe stellar lifetimes directly, but we do observe relative stellar populations and infer lifetimes (shorter-lived stars are less common). Observation of the abundance of red giants relative to other star types in our local neighborhood allows us to set bounds on the axion production rate, and in turn, the axion mass. These bounds are indicated in the "Red Giants" exclusion regions of figure 3.7.

The "Halo Decays" exclusion band of figure 3.7 corresponds to Earth's atmospheric optical transparency window. At any axion mass, galactic halo axions would (very) occasionally decay into photons when interacting with ordinary matter. This would result in a faint but very narrow-band"background glow" of photons emanating from the galactic halo. Careful searches via optical astronomy have put limits (nominal non-existence) on the presence of any background halo glow in the frequency ranges accessible to optical techniques.

3.3 The ADMX Haloscope

The current exclusions to the axion mass leave a window 3 orders of magnitude wide, from about 1 μeV to 1 meV, or 240 MHz to 240 GHz in terms of a photon with equivalent mass-energy. There exist experiments that in principle respond to axions in this mass/frequency range, but most to date have insufficient sensitivity to either identify or rule out the existence of axions. To illustrate this, it is useful to create a plot like figure 3.7 but add a second axis for sensitivity, as in figure 3.9. This figure shows sensitivity to a given axion-photon coupling strength ($g_{a\gamma\gamma}$) on the vertical axis and axion mass on the horizontal axis.

The least stringent limits on the diagram are from "Light Shining through Walls" (LSW) experiments which are sensitive to axions and "Axion Like Particles" (ALPS)[12]. The LSW experiment [73] is "pure lab science" in that it does not depend on any outside source

[12]The designation "ALPS" covers a broad range of hypothetical particles that have a photon coupling mechanism similar or identical to the axion. We will not explore the world of ALPS models here, but suffice it to say that the difference between an ALP and an axion is that the axion solves the strong-CP problem, but various ALPS do not.

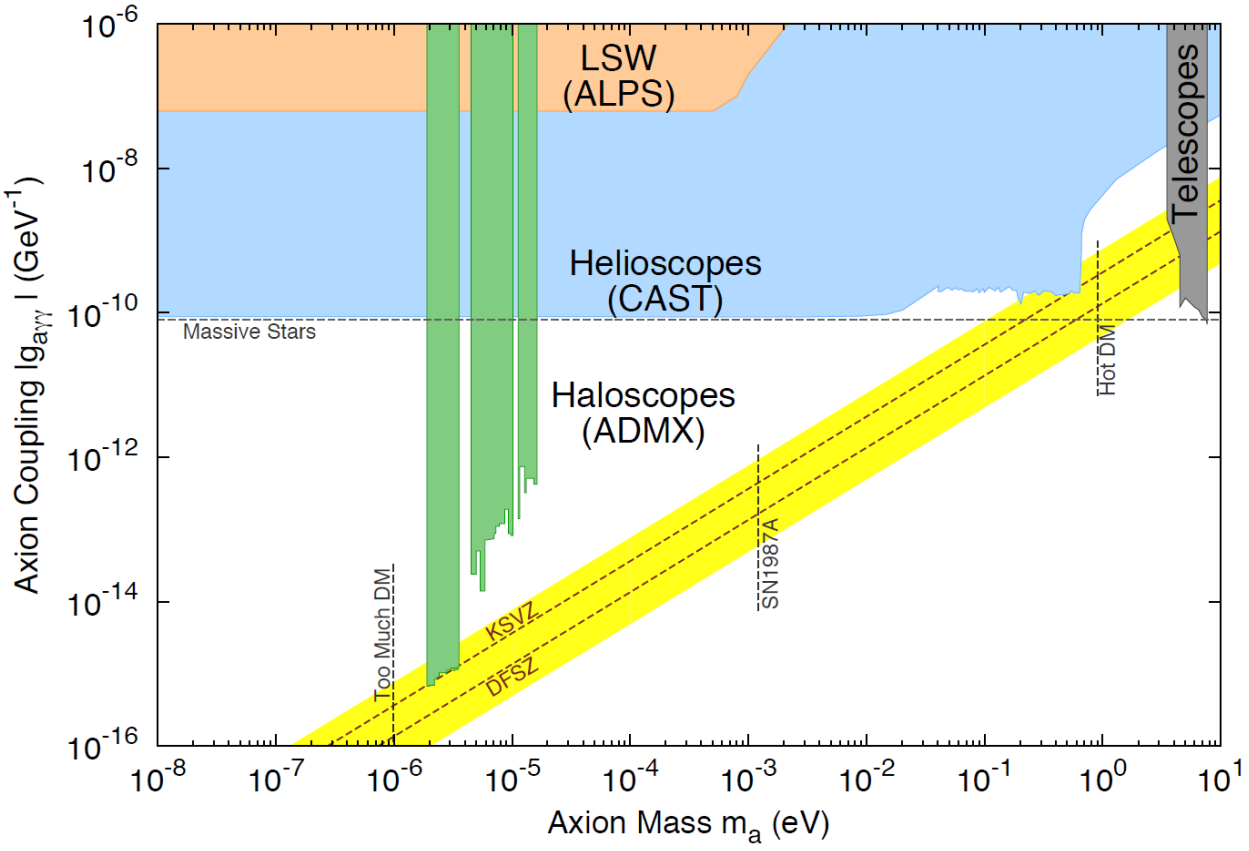

Figure 3.9: Axion Search Window. This figure serves a a "map of the territory" of the axion search space, with the axion mass on the horizonal-axis and axion coupling strength on the vertical-axis. All published results so far are negative, so the colored regions are excluded zones, marking that "no axion of this particular mass and coupling strength exists". To date, the most sensitive experiment is ADMX, followed by other haloscope experiments, albeit at a narrow mass range. The canonical theoretical mass-dependent coupling strength models (KSVZ and DFSZ) are indicated by the rising diagonal lines, with a yellow band indicating the range of proposed coupling strengths currently debated in the community. The most likely mass range for the axions (as asserted in this work) is from 10^{-6} to 10^{-3} eV, corresponding to a photon frequency of 240 MHz to 240 GHz, with the lower bound set by cosmological considerastions and the upper bound set by data from SN1987A. Adapted from [72].

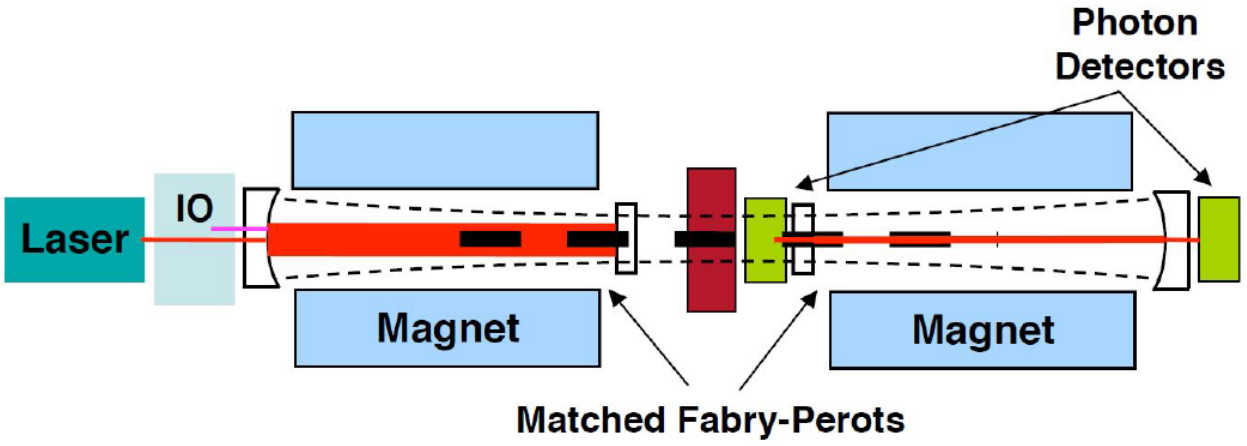

Figure 3.10: Light Shining Through Walls (LSW) experimental outline. A powerful laser beam is circulated in a region with a strong static transverse magnetic field, generating a small number of axions. The axions are not constrained by the cavity, and propagate freely "through walls" to a second, similar magnet. At the second magnet, the axions convert to photons in the reverse process as the first chamber (with the same small conversion rate). Because of the need for two axion-photon conversions, this approach is most challenged by a small value of $g_{a\gamma\gamma}$, but potentially provides a full axion laboratory, with freedom to adjust all relevant parameters. Adapted from [73].

of axions or ALPS, but endeavors to both generate *and* detect exotic new particles. The operating principle is to shine an intense laser beam through the bore of a dipole magnet, such that some conversion to axions or ALPS occurs. Those ALPS will escape the magnet on a trajectory aligned with the axis, and travel through a "brick wall" (and any other ordinary matter) into a second, identical magnet. In the second magnet, the ALPs are converted back into visible light and detected, as shown in figure 3.10. The photon-to-axion conversion rate is proportional to $g^2_{a\gamma\gamma}$, and the axion-to-photon conversion rate is again proportional to $g^2_{a\gamma\gamma}$, for a total factor of $g^4_{a\gamma\gamma}$, so this method has the greatest challenge in overcoming a weak axion coupling.

The next bound is set by axion helioscopes like the CERN Axion Solar Telescope (CAST). The principle here is similar to LSW, but instead of generating axions or ALPs in the lab, it converts particles from our sun, which is potentially a much brighter source of axions or ALPs. In this experiment, a large magnet is placed on a mount to track the sun, and heliogenic axions are converted to photons. The sensitivity of helioscopes depend of the rate of axion production in the sun (which can be through mechanisms other than $g_{a\gamma\gamma}$, including compton scattering, bremsstrahlung, etc) and the rate of axion-to-photon conversion in the experiment ($g^2_{a\gamma\gamma}$), so the sensitivity bound in figure 3.9 depends on solar physics models, and its total sensitivity is proportional to something smaller than $g^2_{a\gamma\gamma}$. A diagram of axion helioscopes is shown in figure 3.11.

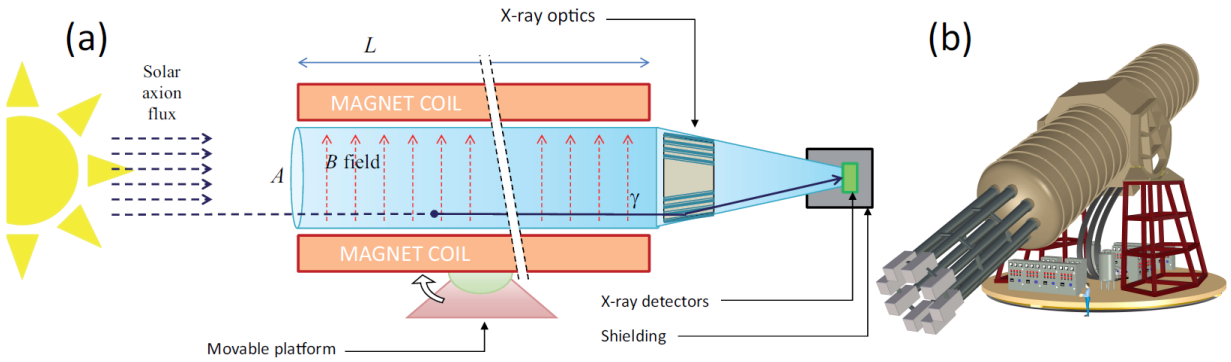

Figure 3.11: CAST helioscope diagram. (a) The magnet bore is aligned with the momentum vectors of heliogenic axions, and any photons generated are focused onto a photon counter. The configuration here is optimized for high-mass axions and x-ray photons. (b) Scale of the proposed CAST upgrade, the International Axion Observatory (IAXO). Note human figure in the foreground. Adapted from [73].

The most sensitive axion searches to date are axion haloscopes, so named because they are sensitive to the galactic halo of axions thought to constitute dark matter, with haloscope bounds shown in green in figure 3.9. In this case the axions are provided "for free" from dynamics of the early universe (the misalignment mechanism, or perhaps cosmic string decay) and are understood to constitute 100% of the local galactic dark matter density. The haloscope bounds shown here are the most sensitive, not least because they depend on only a single conversion factor of $g^2_{a\gamma\gamma}$. In perfect fairness, if axions prove to constitute only 10% of the dark matter in our local neighborhood (with the balance made up of, for example, neutralinos) the haloscope sensitivity should be revised to a lesser value (higher on figure 3.9) by an order of magnitude. The haloscope axion search was first proposed by Pierre Sikivie in 1983 [74], pictured in figure 3.12 some three decades later with his idea made manifest. Details of the Haloscope appear in the following sections.

The other bounds already discussed and appearing in figure 3.9 include the "Red Giant" limit (horizontal dashed line labeled "Massive stars", the "Halo Decay" limit (grey region), the lower mass limit from SN1987a (vertical dashed line labeled "SN1987A"), and the upper mass limit from cosmological considerations (vertical dashed line labeled "Too Much DM").

The proportionality between axion mass and interaction strength is illustrated by the two rising diagonal lines in figure 3.9 labeled "KSVZ" and "DFSZ", a representing the two "canonical" axion coupling models, named by their originator's initials [59, 60, 61, 62], with the yellow shaded region approximating the range of coupling strengths expected by the theoretical community. An experiment with a sensitivity plotted above the yellow region could be characterized as a "pathfinder" experiment, an experiment with sensitivity overlapping the yellow region could be characterized as "scientifically significant", and an experiment with sensitivity extending beyond the yellow region could be characterized as

Figure 3.12: Pierre Sikivie and Christian Boutan at the ADMX site. Sikivie is the originator of the axion haloscope concept, and Boutan is the originator of a great number of figures in this work. Photo taken in 2014.

"definitive" or "exhaustive".

3.3.1 Haloscope Overview and Principle of Operation

In one sentence, the ADMX haloscope is a resonant microwave cavity in a static magnetic field set to detect the resonant conversion of primordial axions to photons via an inverse primakoff effect. Equation 3.11 describes the relevant interaction, where $\vec{E} \cdot \vec{B}$ is the overlap of the electric and magnetic fields. A diagram showing the conversion and detection mechanism is shown in figure 3.13. In ADMX, the magnetic field is supplied by a large superconducting solenoid and the electric field is defined by the resonant mode of the cavity. The resonant cavity is a right circular cylinder with a diameter of about 42 cm, set by the desired resonant frequency range. Because the axion mass, and thus necessary cavity resonant frequency, is not known, the cavity is constructed with internal conductive tuning rods that can be moved from near the edge of the cavity (lowest frequency) to near the center (highest frequency).

The Feynman diagram of figure 3.6 is represented in figure 3.13 with a virtual photon provided by the static magnetic field, resulting in a total mass-energy conversion from an axion to a real photon. The axion momentum is assumed to be along a randomized vector, but with a velocity typical of any body gravitationally bound to the Milky way and at

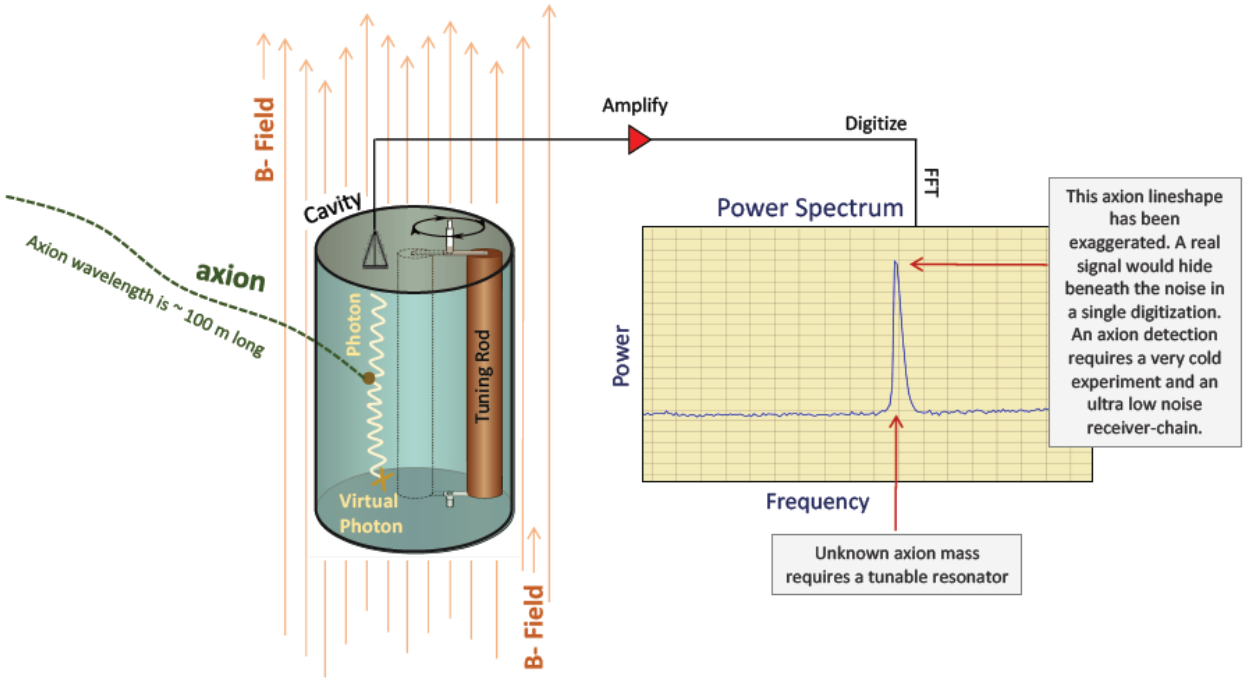

Figure 3.13: ADMX Haloscope concept. The axion haloscope is a microwave resonant chamber immersed in a strong static magnetic field. If the haloscope in located inside the galactic dark matter halo[†] there is a continuous flow of axions through the detector. The magnetic field is a source of virtual photons, allowing for the occasional 100% mass-energy conversion of an axion to a real photon. If the photon frequency matches the resonant frequency of the chamber, conversion rate is enhanced by a factor of the resonator's Q. An antenna critically coupled to the resonator feeds the extra photon to an amplifier and analysis chain, with the extra photons appearing as a narrow-band peak in the power spectrum. The experiment must be maintained at the lowest possible temperature to reduce blackbody radiation, and the resonator must be tunable because the axion mass is not known. Reproduced from [58].

[†]The people and facilities at the University of Washington were selected for ADMX because they are totally out of this world, yet still within the Milky Way dark matter halo.

our radial distance from the galactic center. This results in an axion wavelength some 100 m long, so a long-wavelength approximation is justified in context of ADMX, which has a detector size on the scale of 1 m. By Fermi's golden rule, the axion-to-photon conversion rate is proportional to both the density of virtual photons (B-field strength) and density of final states (cavity Q), thus two key design criteria are a large B and large Q. A coupling antenna in the cavity transfers photons in the cavity to an amplifier and analysis chain, ultimately resulting in a power vs frequency plot with an axion signal appearing as a narrow-band spike of power rising above the noise floor. The noise floor is set by the sum of blackbody radiation in the cavity and noise added by the detection and processing chain, so two more key design criteria are low physical temperature and low-noise electronics. A diagram of the ADMX insert is shown in figure 3.14 at left, with a photo of the insert just after lifting out of the cold space for maintenance on the right.

3.3.2 Things that Affect Scanning Rate (T_N, etc)

If the resonant cavity is tuned to match the axion energy, the power due to axion-to-photon conversion delivered to a transmission line coupled to the cavity is given by

$$\begin{aligned} P_{out} = 5.7 \times 10^{-27}\,\mathrm{W} \times & \left[\left(\frac{g_\gamma}{0.97}\right)^2 \left(\frac{f}{1\mathrm{GHz}}\right) \left(\frac{\rho_a}{0.45\,\mathrm{GeV/cc}}\right)\right] \\ & \times \left[\left(\frac{B}{\mathrm{Tesla}}\right)^2 \left(\frac{V}{\mathrm{liter}}\right) \left(\frac{C_{nmp}}{0.5}\right) \left(\frac{1-2S_{11}}{1-S_{11}}\right) \left(\frac{Q_L}{10,000}\right)\right] \end{aligned} \tag{3.29}$$

where the values grouped in the first brackets are fixed by nature and the values in the second brackets can be controlled by the experimenter. The coupling and mass terms are distributed between g_γ and f, and the power is proportional to f to reflect the mass-dependent coupling strength. The local galactic density of axions is ρ_a. The average magnetic field strength in the cavity is given by B in Tesla, and V is the volume of the cavity in liters. The term C_{nmp} is called the "form factor" and accounts for spatial variations in $\vec{E}$ and $\vec{B}$, as well as the integrated "parallelness" between $\vec{E}$ and $\vec{B}$ inside the cavity, by $C_{nmp} = \left(\int_V dV \vec{E}(x,t)\cdot\vec{B}(x)\right)^2 / \left(VB^2 \int_V dV \epsilon_r E^2\right)$, which has a maximum value of 1. The term S_{11} is the reflection coefficient of a signal on the output line reflecting off the cavity, and is equal to 0 for critical coupling. Finally, Q_L is the loaded quality factor of the cavity.

For typical parameter values at ADMX ($\rho = 0.45$ GeV/cc, $B = 7$ T, $V = 135$ liters, $f = 700$ MHz, $Q = 15,000$, and $C_{010} = 0.5$) the expected power output is 1.8×10^{-22} W (2.5×10^{-23} W) for KSVZ (DFSZ) axions. This is a remarkably low power, comparable to the power from an Earth-based cell-phone tower received on the surface of Mars[13]. ADMX

[13]This calculation assumes a 1kW transmission power from an omni-directional source, a typical Earth-Mars distance taken to be equal to the semi-major axis of Mars's orbit, and a receiving area of 100 cm^2, typical of a contemporary cell phone.

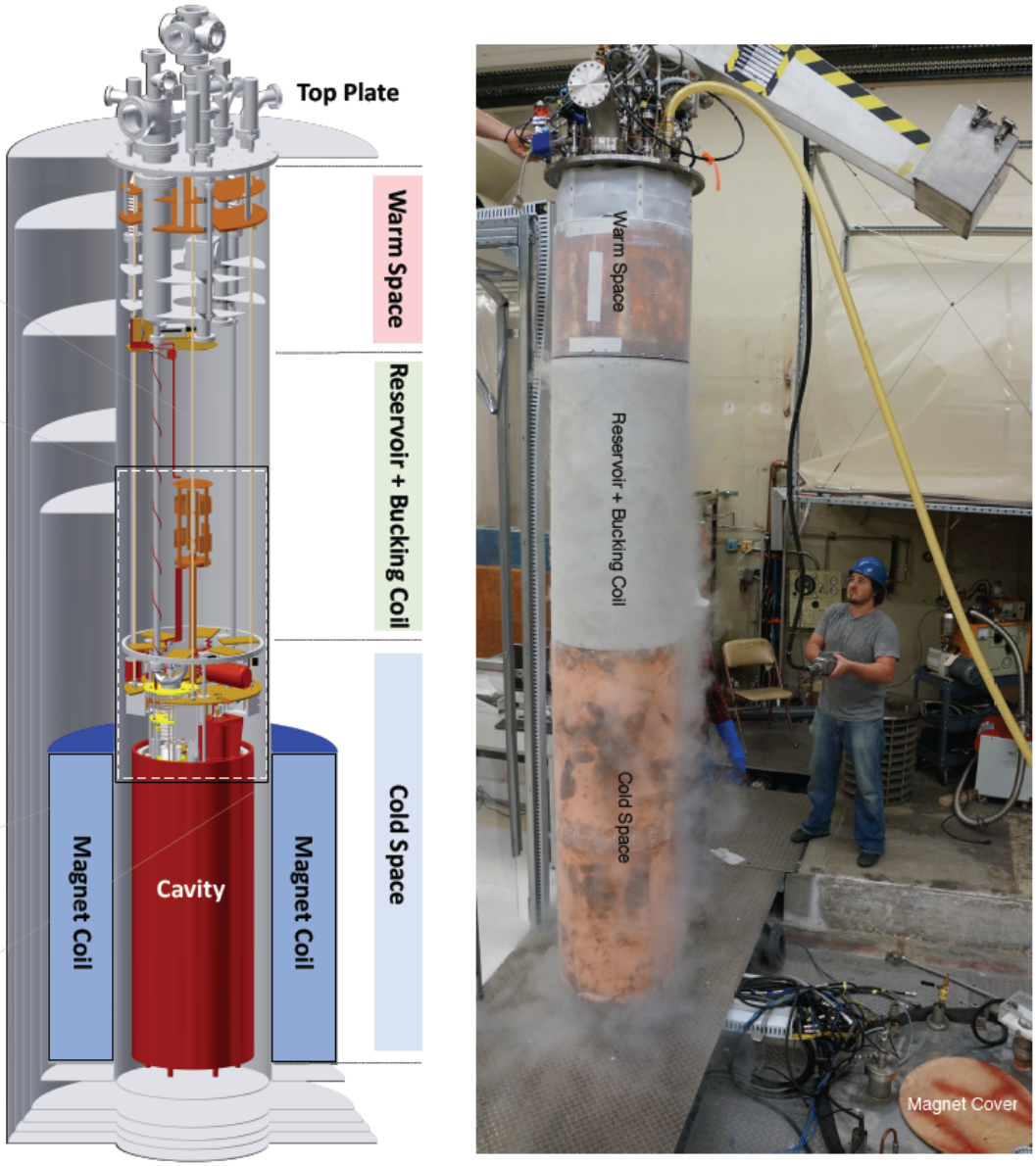

Figure 3.14: ADMX Insert. At left is a partly sectioned drawing showing the main magnet and cavity of the haloscope, along with heat shields, vacuum spaces, scaffolding for electronics, actuators, other cold accessories, and plumbing ports for gas handling. The "Top Plate" is at ground level of the operating site. At right is a photo of the insert (cavity, dilution fridge, cold electronics, etc) being lifted by crane clear of the main magnet bore for scheduled maintenance. Grad student for scale. Reproduced from [58].

strives for a "loud and clear" axion signal, motivating the unofficial rallying cry: "Four bars on Mars!"

Equally important to increasing the total power of converted axions is reducing the noise floor so that any increment of extra power in the cavity is detectable. A key performance measure is the signal to noise ratio (SNR) given by the radiometer equation

$$\begin{aligned} SNR &= \frac{P_{out}}{k_B T_{sys}} \sqrt{\frac{t}{BW}} \\ &= 4.1 \times 10^{-4} \frac{\left(\frac{g_\gamma}{0.97}\right)^2 \left(\frac{f}{1\text{GHz}}\right) \left(\frac{\rho_a}{0.45\ \text{GeV/cc}}\right) \left(\frac{B^2 V}{T^2 l}\right) \left(\frac{C_{nmp}}{0.5}\right) \left(\frac{1-2S_{11}}{1-S_{11}}\right) \left(\frac{Q_L}{10{,}000}\right) \sqrt{t}}{T_{sys}\sqrt{BW}}, \end{aligned} \tag{3.30}$$

where k_B is the Boltzmann constant, t is the integration time, BW is the bandwidth, and T_{sys} is the system noise temperature. The system noise temperature is the sum of physical temperature ($T_{physical}$) and noise added by the amplifier and detection chain, parameterized as an equivalent increase in physical temperature (T_N)

$$T_{sys} = T_{physical} + T_N. \tag{3.31}$$

Note the importance of T_{sys} is equation 3.30–ADMX will integrate at a particular frequency for sufficient time to achieve a target SNR, step to a new frequency through a small adjustment to the tuning rods, and repeat. If T_{sys} were to double, then *4 times* the integration time t would be necessary to achieve the same SNR. Similarly, halving T_{sys} would quarter the integration time t. In general, the total scan time necessary to explore any particular frequency interval (equivalently, axion mass interval) scales as T_{sys}^2. To make a concrete example of this, consider an early version of ADMX with T_{phys}=1.5 K (typical of a pumped ^{4}He fridge) and T_N=1.7 K (typical of the best available transistor-based amplifiers) for a T_{sys} of 3.2 K. The total integration time to explore the lowest octave of the axion search space (1 to 2 μeV, ie 240 to 480 MHz) with a sensitivity sufficient to reveal axions with the canonical DFSZ coupling strength would be 270 years.[14] In contrast, with T_{phys} = 50 mK (readily achievable with a ^{3}He dilution fridge) and T_N = 50 mK (as demonstrated for the best MSAs) for a T_{sys} of 100 mK, the time to to explore that same range is only 100 days.

In ADMX (as is typical) there is a chain of N amplifiers from the physical signal and the final recording device, each with its own gain G_i and noise temperature T_i. With the amplifier $i = 1$ the first to encounter the signal and amplifier $i = N$ the last in the chain, the system noise temperature is

$$T_{sys} = T_{physical} + T_N = T_{physical} + T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \frac{T_4}{G_1 G_2 G_3} + \cdots \tag{3.32}$$

[14]This is an unreasonably long time to wait, even for someone with a graduation timeline as dilated as the author's.

which shows that the physical temperature and T_N of the fist amplifier are on equal footing, while the T_N of components after the first amplifier are rather insignificant. Critically, to avoid a large T_{sys}, T_1 must be small or on par with T_{phys}, and G_1 must be large relative to T_2/T_1. These are among the critical performance goals of the MSA, used as the first amplifier in the ADMX receiver chain.

3.3.3 Current Operation and Short-Term Outlook

ADMX is the most sensitive axion search to date. Figure 3.9 incorporates limits published no later than 2013, but since that data was published, ADMX has undergone various upgrades to increase sensitivity, including fridge upgrades to operate at lower T_{phys} and amplifier development[15] achieving reduced T_N. The amplifier improvements in concert with several other critical upgrades have enabled unprecedented search sensitivity and faster scan rates, leading to the first published exclusion of galactic halo axions at the DFSZ coupling limit. This exclusion [75] was published in April 2018, with the central result shown in figure 3.15. The newest exclusion range (dark blue region in the inset of figure 3.15) covers only about 10% of the mass range of previously published haloscope searches [76, 77, 78, 79, 80, 81], but covered this range at greater sensitivity and in a relatively short period of operation. At the current operating rate, a DFSZ exclusion band of width comparable to the previous KSVZ exclusion range may be published in 2019.

3.3.4 Long-Term Outlook

It is easy to imagine a definitive axion search covering the mass range from about 1 to 25 μeV (240 MHz to 6 GHz) based on the haloscope detection principle as demonstrated by ADMX. Stronger magnets, colder fridges, refined amplifiers, improved cavity design (higher Q), and other reasonable extensions of the existing technology are likely to prove successful in expanding the explored parameter space in the vicinity of currently published limits. An excellent example of this approach is to be had in the Haloscope at Yale Sensitive To Axion CDM (HAYSTAC). This project is tackling these issues at mass ranges near the highest practical for a haloscope search [82], and is recently nudging the boundary between "pathfinder" and "scientifically significant" data production [83].

The search for axions at very high frequencies brings new challenges. As an example: equation 3.30 shows that SNR is proportional to the resonator cavity volume, but the resonator size scales inversely with frequency, creating an effective f^{-3} multiplier to SNR. Maintaining high Q at high f becomes difficult (due to the anomalous skin effect, etc.), adding further to the difficulties of higher frequency searches.

Exploration up to 1 meV (240 GHz) will require new novel cavity design (photonic band gap, coupled cavities, etc) for intermediate frequencies, and radical non-cavity designs for the highest frequencies, perhaps similar to the Orpheus design [84]. Higher frequencies also

[15] see: this work

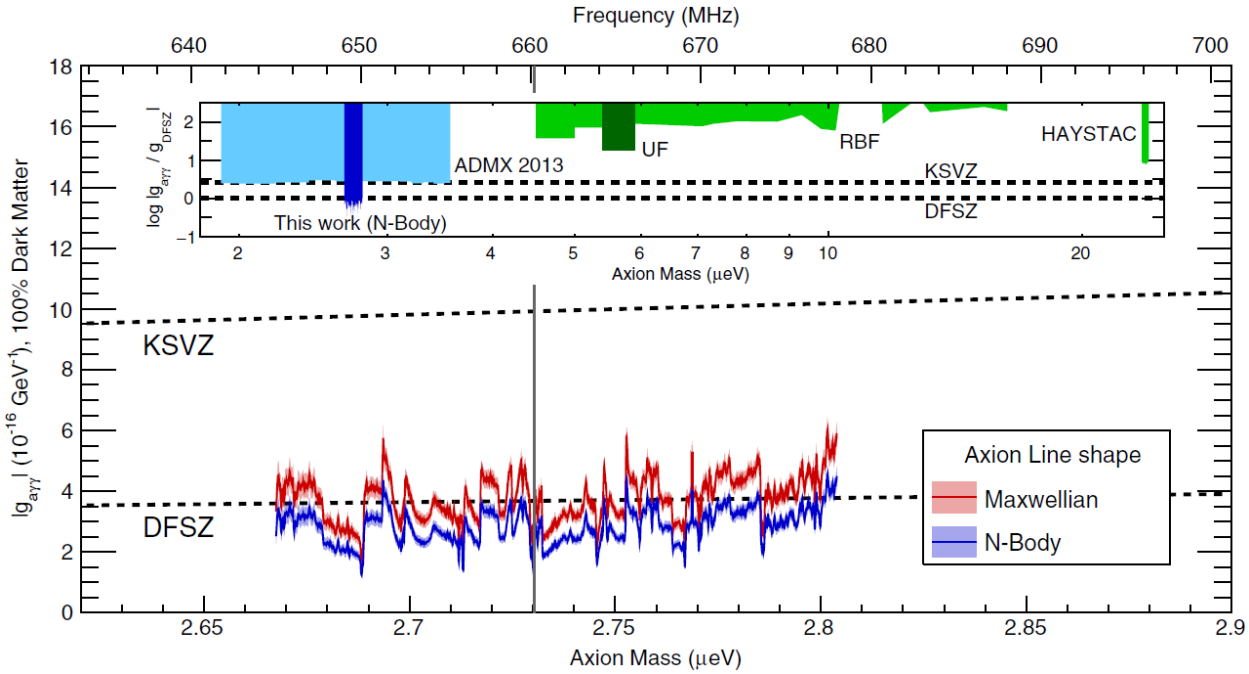

Figure 3.15: ADMX best published axion exclusion limits as of April 2018. Operation with an MSA as described in this work and other upgrades has enabled the first ever axion exclusion to the DFSZ coupling limit. The light blue area in the inset corresponds to the "deepest dive" green area of figure 3.9. The red and blue solid lines represent the exclusion limit based on two models of the axion linewidth. The red limit assumes a fully virialized velocity distribution of galactic axions, while the blue limit assumes a velocity distribution deriving from a particular model of galaxy formation. No exclusion could be made in the narrow gray region near 660 MHz due to some external electromagnetic interference. Reproduced from [75].

necessitate new detection methods/amplifier design. Any phase-preserving amplifier (like the MSA) has a minimum T_N set by the "standard quantum limit" $T_Q = hf/k_B$, about 50 mK per GHz. At frequencies above several GHz, even a "perfectly engineered" phase-preserving amplifier will have a prohibitively high T_N, so a phase insensitive amplifier must be used. Every amplifier is subject to the standard quantum limit, which arises from the fundamental fact of quantum indeterminacy, but it is possible to "beat" the standard quantum limit by engineering the "shape" of the indeterminacy. A phase-preserving amplifier spreads the indeterminacy equally around the available degrees of freedom (I and Q quadratures, number and phase, or the non-commuting pair of your choice), while a phase-insensitive amplifier can in principle put all the indeterminacy into one degree of freedom (rendering it useless as a source of data) to gain the substantial benefit of reducing the uncertainty of the remaining degree of freedom to an arbitrarily small spread. The Josephson Parametric Amplifier

(JPA)[85, 86] is an example of an amplifier that can in principle be used to squeeze the uncertainty into one quadrature (Q), leaving the other quadrature (I) with arbitrarily small uncertainty. The single-photon counter[16] is a kind of amplifier that puts all the uncertainty into phase, leaving arbitrarily small uncertainty in particle number.

Today as ever, the desire to explore fundamental scientific questions motivates the development of improved instrumentation, and the availability of improved instrumentation (and the data thus gathered) excites greater imagination and gives rise to previously unimagined avenues of inquiry. The remainder of this work is dedicated to the particular instrumentation improvements the author has contributed to ADMX, and one hopes, the greater human scientific project.

[16]There are a variety of ways to make a photon-counting amplifier. One is the venerable photon multiplier tube (PMT). A newer example is the phase qubit. The engineering details depend greatly on the frequency of interest, but the principle is the same.

Chapter 4

The Josephson Junction, DC SQUID, and Microstrip SQUID Amplifier (MSA)

The MSA[1] is a microwave amplifier inspired by the need for a detector of unprecedented sensitivity for use in ADMX. Developed by the Clarke group at UC Berkeley, the latest developments in this technology are the primary focus of this work. The MSA has enabled ADMX to push the axion dark matter search deeper then was previously possible, but is not limited to this application. As a general phase-preserving microwave amplifier, the MSA can be used in service of other scientific applications in need of low-noise measurement, for example in hybrid electro-mechanical cryogenic bulk acoustic wave devices, near-ground state penning traps, or any other quantum electrical system near the 1 GHz range.

At a UC Berkeley physics colloquium in 2015, Nobel Laureate Steven Chu gave the following career advice for young physicists: "Become a master of new technologies and methods, and focus on fundamental problems." I hope the MSA proves a useful tool that helps illuminate many new frontiers in physics, and am delighted that it is already instrumental in investigating the fundamental problems of Dark Matter and CP symmetry.

4.1 Introduction to Superconductivity

The sensitivity of the MSA can be attributed to its fundamentally quantum properties. For any measured quantity, a sensitive probe must operate and be engineered at the quantum level. The MSA achieves this by virtue of superconductivity, which is effectively a large-scale quantum-coherent state of electrons that forms spontaneously at sufficiently low temperature. This chapter endeavors to introduce the fundamental concepts behind operation of the MSA.

[1] "MSA" is a double-packed acronym, which fully unpacked becomes "Micro-stripline Superconducting Quantum Interference Device Amplifier"– an arrhythmic and eye-glazing abomination that must never be allowed to escape the bounds of this footnote.

4.1.1 Quantum Coherence in Superconductivity

The remarkable properties of superconductivity arise from quantum coherence across many bodies over a large span. In most other circumstances, pure quantum states are found only tiny, well isolated systems of few particles. For example, a helium atom might have two electrons in a coherent singlet state in the ground orbital, for N=2 particles over a span of about 1 Å. In contrast, a superconducting sample of aluminum has a coherence length of about 1600 nm, which at an atomic spacing of 404 pm contains about $6 \cdot 10^{10}$ atoms, or $1.8 \cdot 10^{11}$ conduction electrons, or (at low temperatures) about $9 \cdot 10^{10}$ Cooper pairs, depending on your preferred counting unit. In typical lab conditions (current, applied field, etc) the coherence length can be much larger than 1600 nm, up to th size of a typical MSA (about 1 mm) or even many meters.

The phenomenon of superconductivity owes its name to the original phenomenological observation of dissipation-free conduction of an electrical current. While this remains an important property for many applications, it would be misleading to characterize dissipation-free transmission of electrical current as the unifying core concept of superconductivity. Had H. Kamerlingh Onnes possessed on the day of his remarkable discovery the theoretical understanding developed in the following decades, perhaps he would have named the new phenomenon not "Superconductivity" but something like "Degenerate Electrical Quantum Coherence". Then again, there is a virtue in coining a short descriptive neologism that easily captures the imagination, despite the confusion it might one day cause for students first approaching the subject.

The Cooper pairs in a superconductor are spin-zero bosons and as such at sufficiently low temperature will all condense into the lowest energy quantum state. Thus a macroscopic number of conduction electrons bound as Cooper pairs can be described by a single wave function of the form

$$\psi = |\psi(\mathbf{r})| e^{i\theta(\mathbf{r})}, \tag{4.1}$$

where θ is the phase at location $\mathbf{r}$, and ψ is normalized such that $\psi^* \psi$ equals the number density of Cooper pairs $n_s(\mathbf{r})$ which we will usually take to be spatially constant.[2] The absolute phase θ of the wavefunction has no directly observable consequence, but nearly every interesting or useful property of superconductivity derives from the macroscopic coherence of θ. For instance, the gradient of θ can be related to the momentum of the whole Cooper pair field, as can be demonstrated by applying the quantum mechanical momentum operator $-i\hbar\nabla$ to ψ. The expectation value of $\mathbf{p}$ is then

[2]The astute reader will note that phase and number are non-commutating operators, and so are subject to the uncertainty relation $(\Delta n)(\Delta\theta) \geq 1/2$, in contradiction to the assumption that n_s is everywhere constant. This is resolved by noting that n_s is a very large number, for example in aluminum about $6 \cdot 10^{10}$ at a minimum, and often magnitudes greater. If the uncertainty $\Delta\theta$ is rather small, say 10^{-5} radians, then the uncertainty Δn_s is no less than $5 \cdot 10^4$, or $\Delta n_s / n_s$ of about a part in one million. The very large typical n_s in superconducting devices makes the practical consequences of Δn_s negligible.

$$n_s\mathbf{p} = \langle\psi| - i\hbar\nabla|\psi\rangle = -i\hbar(i\nabla\theta)\langle\psi|\psi\rangle = n_s\hbar\nabla\theta \tag{4.2}$$

or

$$\mathbf{p} = \hbar\nabla\theta. \tag{4.3}$$

That is: a Cooper pair condensate with a spatially constant phase is at rest with zero momentum, while one with a spatially homogeneous $d\theta/d\mathbf{r}$ has a spatially constant momentum in the $\mathbf{r}$ direction. A ψ with a phase that differs gently from place to place has a small momentum, while large, spatially abrupt changes in the phase θ are associated with a large momentum.

Two phenomenological consequences of this large-scale quantum coherence are flux quantization and the Josephson effect.

4.1.2 Flux Quantization

Macroscopic quantum coherence requires the phase θ of the superconducting wavefunction be everywhere continuous and single-valued. One consequence of this is that the amount of magnetic flux inside a superconducting ring is constrained to integer multiples of $h/2e$. This condition arises because the total phase change of ψ in one trip around the superconducting loop must be an integer value of 2π, ie

$$\oint \nabla\theta \cdot d\mathbf{l} = n \cdot 2\pi. \tag{4.4}$$

To relate this to the magnetic field, recall that the canonical momentum for a charged body is

$$\mathbf{p} = 2m\mathbf{v} - 2q\boldsymbol{A}, \tag{4.5}$$

where m is the mass, $\mathbf{v}$ is the velocity vector, q is the charge, and $\boldsymbol{A}$ is the vector potential. In the case of the superconducting condensate

$$n_s\mathbf{p} = n_s(2m_e\mathbf{v} - 2e\boldsymbol{A}), \tag{4.6}$$

where n_s is the Cooper pair density, m_e is the electron mass, e is the elementary charge and factors of two appear because the Cooper pairs are composed of two electrons. Relating the current density $\mathbf{J}$ to the Cooper pair velocity $\mathbf{v}$ by $\mathbf{J} = -2n_se\mathbf{v}$ and substituting equation 4.3, equation 4.5 becomes

$$\hbar\nabla\theta = -\mathbf{J}\frac{m_e}{n_se} - 2e\boldsymbol{A}. \tag{4.7}$$

Taking the loop integral and applying the single-valued requirement of 4.4 gives

$$\hbar \oint \nabla\theta \cdot d\mathbf{l} = \hbar n \cdot 2\pi = hn = -\oint \left(\mathbf{J}\frac{m_e}{n_s e} + 2e\boldsymbol{A} \right) \cdot d\mathbf{l}. \tag{4.8}$$

The supercurrents in the ring exist mostly near the surface and are exponentially suppressed in the bulk of the metal. This allows us to simplify equation 4.8 by taking the path of the loop integral to be deep inside the superconductor where currents are effectively zero. This gives us

$$hn = -2e \oint (\boldsymbol{A}) \cdot d\mathbf{l} \tag{4.9}$$

and by Stokes' theorem

$$hn = -2e \int_S (\nabla \times \boldsymbol{A}) \cdot d\mathbf{S} \tag{4.10}$$

or

$$hn = -2e \int_S \boldsymbol{B} \cdot d\mathbf{S} = -2e\Phi_{enc} \tag{4.11}$$

by the definition of the vector potential $\boldsymbol{A}$, the magnetic field $\boldsymbol{B}$, and enclosed magnetic flux Φ_{enc}. This is the origin of the flux quantization condition

$$\Phi_{enc} = n\frac{h}{2e}, \tag{4.12}$$

where the $(-)$ has been dropped with the understanding that n may have either sign. The flux quantum

$$\Phi_0 \equiv \frac{h}{2e} \approx 2.068 \times 10^{-15} \text{ Wb} \tag{4.13}$$

appears as a fundamental quantity in superconducting applications and arises as a result of large-scale quantum coherence.

4.1.3 The Josephson Effect

If two metals are separated by an insulator, then classically no current may flow between them. However, the Cooper pair condensate in a superconductor is a fundamentally quantum thing, and a supercurrent *may* tunnel through such a gap. This phenomenon is called the Josephson effect, which was predicted in 1962 [87][88]. At the interface between the superconductor and insulator the amplitude of ψ does not go abruptly to zero, which would represent an infinite second derivative and infinite energy, but instead has an evanescent component that decays exponentially outside the superconductor. If two superconductors are placed in close proximity such that their evanescent waves overlap, they will have an interaction that allows a supercurrent to flow through the forbidden region. This construction is called a Josephson junction, and is illustrated in figure 4.1.

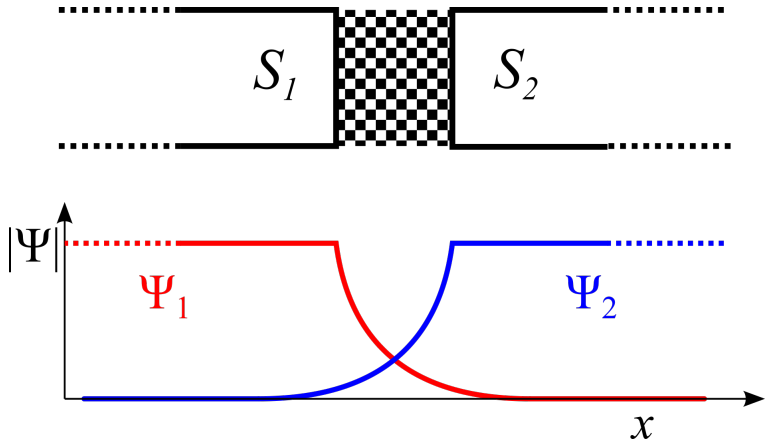

Figure 4.1: Model Josephson junction. The coherent superconducting condensates in two superconducting segments have overlapping wavefunctions in the "forbidden" insulator region. Coupling between the wavefunctions allows a classically forbidden supercurrent to flow.

The two superconducting elements, labeled S_1 and S_2 in figure 4.1, are separated by an insulating gap (checker pattern). The gap could be vacuum, but in practice is usually a thin oxide film. The coherent wavefunction amplitudes of the left and right hand Cooper pair condensates are drawn below, labeled Ψ_1 and Ψ_2. Inside either superconductor, the overlap between Ψ_1 and Ψ_2 is minimal, and can be ignored. In the insulator region, Ψ_1 and Ψ_2 decay, with an exact functional form that depends on the particular insulator properties. In this treatment we will abstract away the details of the overlap integral and say simply that there is a coupling energy K between Ψ_1 and Ψ_2. The dynamics of the two wavefunctions are then determined by the coupled Schrödinger equations

$$\begin{aligned} i\hbar\frac{\partial\Psi_1}{\partial t} &= \mu_1\Psi_1 + K\Psi_2 \\ i\hbar\frac{\partial\Psi_2}{\partial t} &= \mu_2\Psi_2 + K\Psi_1, \end{aligned} \tag{4.14}$$

where μ_1 and μ_2 are the ground-state energy of the condensate on either side of the gap.

Using equation 4.1, and assuming the Cooper pair density n and phase θ on either side of the insulator have negligible spatial variance, we find

$$\begin{aligned} \Psi_1 &= \sqrt{n_1}e^{i\theta_1} \\ \Psi_2 &= \sqrt{n_2}e^{i\theta_2}. \end{aligned} \tag{4.15}$$

Substituting equation 4.15 into 4.14 gives

$$\hbar\frac{\partial n_1}{\partial t} = -\hbar\frac{\partial n_2}{\partial t} = 2K\sqrt{n_1 n_2}\sin(\theta_2 - \theta_1) \tag{4.16}$$

and

$$-\hbar\frac{\partial}{\partial t}(\theta_2 - \theta_1) = \mu_2 - \mu_1. \tag{4.17}$$

Equations 4.16 and 4.17 can be re-written in a form more useful to the experimentalist by recognizing that $\partial n/\partial t$ represents a superconducting charge transport, ie current I_s, and a difference in ground-state energy indicates a voltage difference between S_1 and S_2, so $\mu_2 - \mu_1 = 2eV$. Defining $I_0 \equiv 2K\sqrt{n_1 n_2}$ and $\delta \equiv \theta_1 - \theta_2$ gives

$$I_s = I_0 \sin\delta \tag{4.18}$$

and

$$\frac{\partial\delta}{\partial t} = \frac{2eV}{\hbar} = \frac{2\pi}{\Phi_0}V, \tag{4.19}$$

which are the canonical DC and AC Josephson equations, respectively.

The behavior described by equation 4.18 was predicted in 1962 and observed the next year [89]. Equation 4.18 predicts that when no current flows through the barrier, the two superconductors have the same phase, and are at the same voltage. As the current increases, a static phase difference between S_1 and S_2 develops, up to a maximum of $\delta = \pm\pi/2$ for currents of $I_s = \pm I_0$, but with the relative phase constant and V=0. If $I > I_0$, a zero-voltage DC supercurrent cannot be sustained. If a constant voltage is applied between S_1 and S_2, the phase will evolve linearly and the current will evolve sinusoidally according to equation 4.19 as

$$I_s(t) = I_0 \sin\left(\frac{2\pi V}{\Phi_0}t\right), \tag{4.20}$$

that is, a constant voltage will excite a sinusoidal current with an amplitude of I_0 and frequency of $f \approx 483.6$ MHz$/\mu V$, where I_0 is the critical current and f is called the Josephson frequency. Relations analogous to equations 4.18 and 4.19 hold for other kinds of weak link between two superconductors, for example a narrow bridge that is short relative to the coherence length, but they are somewhat modified, for example by a non-sinusoidal but still periodic current-phase relationship.

4.1.4 The Resistively-Capacitively-Shunted-Junction (RCSJ) model

Any two conductors separated by a small gap will necessarily have a capacitance. This was neglected in the previous section and is irrelevant in the DC case, but must be included for

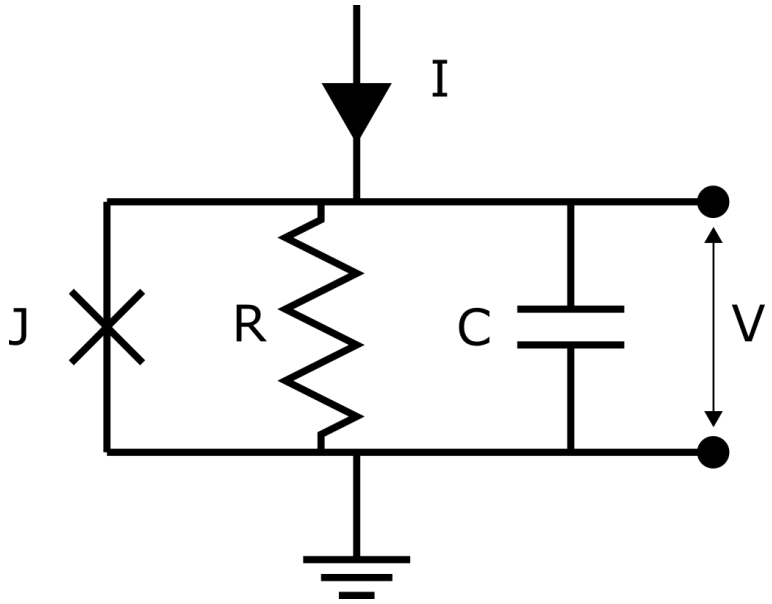

Figure 4.2: RCSJ model. J represents the "pure" Josephson junction, R represents the parallel sum of any conduction channels in the junction and external added resistance, and C represents the sum of the junction capacitance and any explicit added capacitance.

a complete dynamic analysis. In addition, there may be a resistance either due to non-ideal performance of the insulator, or deliberately added for reasons that will soon be apparent.

A junction with parallel resistance and capacitance is shown in figure 4.2. The current through this circuit is the sum of the Josephson current, Ohmic current, and displacement current

$$I = I_0 \sin(\delta) + \frac{V}{R} + C\frac{dV}{dt}. \tag{4.21}$$

To write the current I purely in terms of phase δ, we can apply equation 4.19 to give

$$I = I_0 \sin(\delta) + \frac{\Phi_0}{2\pi}\frac{1}{R}\dot{\delta} + \frac{\Phi_0}{2\pi}C\ddot{\delta}. \tag{4.22}$$

To gain insight into equation 4.22, we make an analogy to a particle of mass m moving in a 1-D potential $U(x)$ with a viscous drag ξ:

$$m\ddot{x} + \xi\dot{x} = -\frac{\partial}{\partial x}U(x). \tag{4.23}$$

Equation 4.22 then becomes

$$C\left(\frac{\Phi_0}{2\pi}\right)^2\ddot{\delta} + \frac{1}{R}\left(\frac{\Phi_0}{2\pi}\right)^2\dot{\delta} = -E_J\frac{\partial}{\partial\delta}\left[-\cos(\delta) - (I/I_0)\delta\right] \tag{4.24}$$

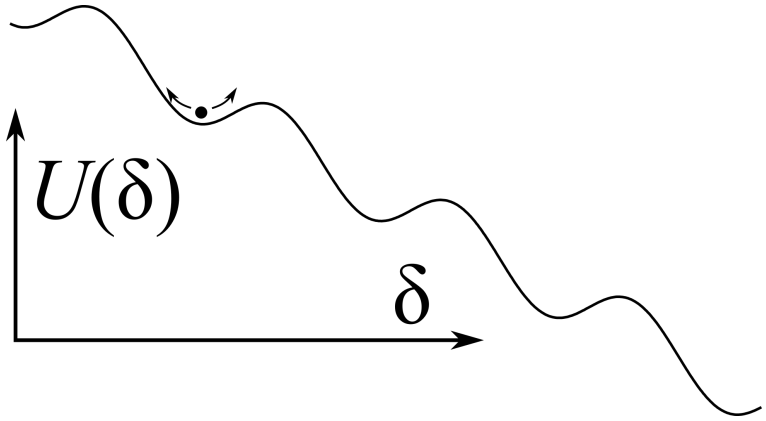

Figure 4.3: The tilted washboard potential. The ratio I/I_0 determines the tilt of the potential function $U(\delta)$. For $I < I_0$ there exist potential wells that can bind the phase particle. For $I > I_0$, the phase particle will roll downhill.

and we can make the analogy to the dynamics of a particle on a tilted washboard potential with viscous drag by

$$\begin{aligned} x &\leftrightarrow \delta \\ m &\leftrightarrow C\left(\frac{\Phi_0}{2\pi}\right)^2 \\ \xi &\leftrightarrow \frac{1}{R}\left(\frac{\Phi_0}{2\pi}\right)^2 \\ U(x) &\leftrightarrow E_J\left[-\cos(\delta) - (I/I_0)\delta\right], \end{aligned} \tag{4.25}$$

where $E_J \equiv \Phi_0 I_0/2\pi$ is the Josephson coupling energy. For a typical junction with $I_0 = 2.5\ \mu A$, $E_J \approx 5$ meV. The potential function $U(\delta)$ is shown in figure 4.3 for $I/I_0 = 1/2$.

The ratio I/I_0 determines the tilt of the potential function $U(\delta)$. If $I = 0$ there is no tilt and the phase particle is confined to one well centered at $\delta = 0$. The phase particle will oscillate at the bottom of the well at the plasma frequency $\omega_p \equiv \sqrt{E_J/m}$. For a junction typical of the work presented here, $I_0 = 2.5\mu A$ and $C = 312$ fF, so $\omega_p \approx 2\pi \times 25$ GHz. The phase particle describes a quantum mechanical system with a quanta of excitation of $\hbar\omega_p \approx 0.1$ meV. Comparing this to the well depth $2E_J \approx 10$ meV, the ground state energy is about 1% of the well depth, so a classical approximation of the dynamics is reasonable, and we can safely neglect, for example, tunneling to an adjacent well.

If the current is increased, the washboard tilt increases, the equilibrium point shifts right, and the plasma frequency is reduced. If the current is increased past $I = I_0$, there are no stable wells at all and the phase will evolve to monotonically greater values, meaning by equation 4.19 a voltage develops across the junction. If the current is then decreased to

$I < I_0$ there are two possible behaviors, depending on the damping strength: in the case of strong damping the particle will slow and eventually stop in a well, but in the case of small damping the energy gained by falling to a lower well may not be less than the energy lost to friction, the particle will continue "rolling downhill" even when $I < I_0$. This is called hysteretic operation, and in this work is always to be avoided. The critical damping parameter was described by Stewart and McCumber [90], [91] with a damping of

$$\beta_C \equiv \frac{2\pi I_0 R^2 C}{\Phi_0} < 1 \tag{4.26}$$

required to avoid hysteretic operation.

In general, the dynamics of equation 4.24 cannot be solved in closed form, but it is instructive to explore the limit of extreme over-damping ($\beta_C \ll 1$), which *can* be solved in closed form. In the over-damped limit, viscosity dominates momentum and by dropping the "momentum" term of equation 4.22 we have

$$\Phi_0 \frac{\partial \delta}{\partial t} = 2\pi I_0 R \left(\frac{I}{I_0} - \sin \delta \right). \tag{4.27}$$

For the case $|I/I_0| < 1$, the solution is easy: $\sin \delta = I/I_0$ and $\frac{\partial \delta}{\partial t} = 0$. For the case $|I/I_0| > 1$, we recognize that the solution will have a $\frac{\partial \delta}{\partial t}$ periodic in time with some period τ, and $\delta(t + \tau) = \delta(t) + 2\pi$. Separating variables and integrating then gives

$$\frac{2\pi I_0 R}{\Phi_0} \int_0^\tau dt = \int_0^{2\pi} \left(\frac{I}{I_0} - \sin \delta \right)^{-1} \tag{4.28}$$

for

$$\tau = \frac{\Phi_0}{I_0 R} \left[\left(\frac{I}{I_0} \right)^2 - 1 \right]^{-1/2}. \tag{4.29}$$

With this result, δ increases by 2π in each time period τ, so

$$\left\langle \frac{\partial \delta}{\partial t} \right\rangle = \frac{2\pi}{\tau} = \frac{2\pi I_0 R}{\Phi_0} \left[\left(\frac{I}{I_0} \right)^2 - 1 \right]^{1/2} = \frac{2\pi}{\Phi_0} \langle V \rangle \tag{4.30}$$

or

$$\langle V \rangle = R\sqrt{I^2 - I_0^2}. \tag{4.31}$$

Equation 4.31 is plotted in figure 4.4. Note the sudden onset of voltage when $I > I_0$ and an asymptote to $V = IR$ for $I \gg I_0$. These critical features are preserved even for a junction with finite damping.

Just for fun and since we do have an exactly solvable model in this overdamped case, the exact phase-time relation is

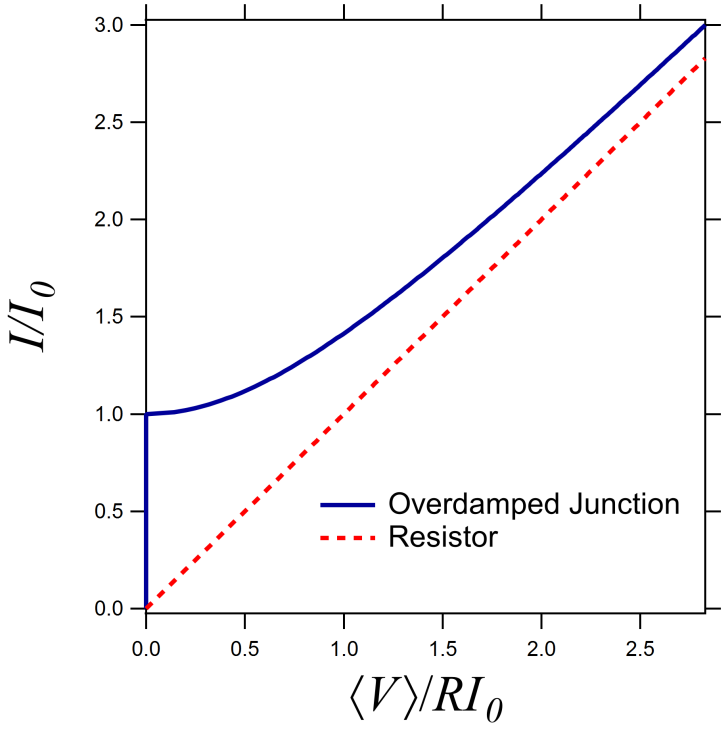

Figure 4.4: I-V characteristic of an overdamped junction. Solid blue is the RCSJ junction and dashed red is the shunt resistance. At large current bias ($I \gg I_0$), the curves converge.

$$\delta(t) = 2\arctan\left\{\frac{I}{I_0}\sqrt{\left(\frac{I}{I_0}\right)^2 - 1}\left[\tan\left(t\frac{\Phi_0\sqrt{(I/I_0)^2-1}}{4\pi I_0 R}\right) + 1\right]\right\}. \tag{4.32}$$

To illustrate the result of equation 4.32, figure 4.5 shows the same I-V characteristic for an overdamped junction as figure 4.4 in red, and overlays in blue two particular time-voltage curves indexed to points on the I-V curve marked **A** and **B**. In each case, the current is constant, and the voltage is a periodic function in time with a constant offset. At I just above I_0, the pulses are slow, and at greater I the pulses are faster. This general behavior is observed even for junctions that are not overdamped, though the exact shape of the pulses may differ. The frequency of the pulses is universal for any junction regardless of damping and is called the Josephson frequency

$$f_J \equiv V/\Phi_0 \approx 483.6 \text{ MHz}/\mu V. \tag{4.33}$$

As a practical matter when using a Josephson junction device in the "voltage state" ie $I > I_0$ to measure or deliver time-varying signals, the signal in question should contain only frequency components well below the Josephson frequency. The devices presented in this

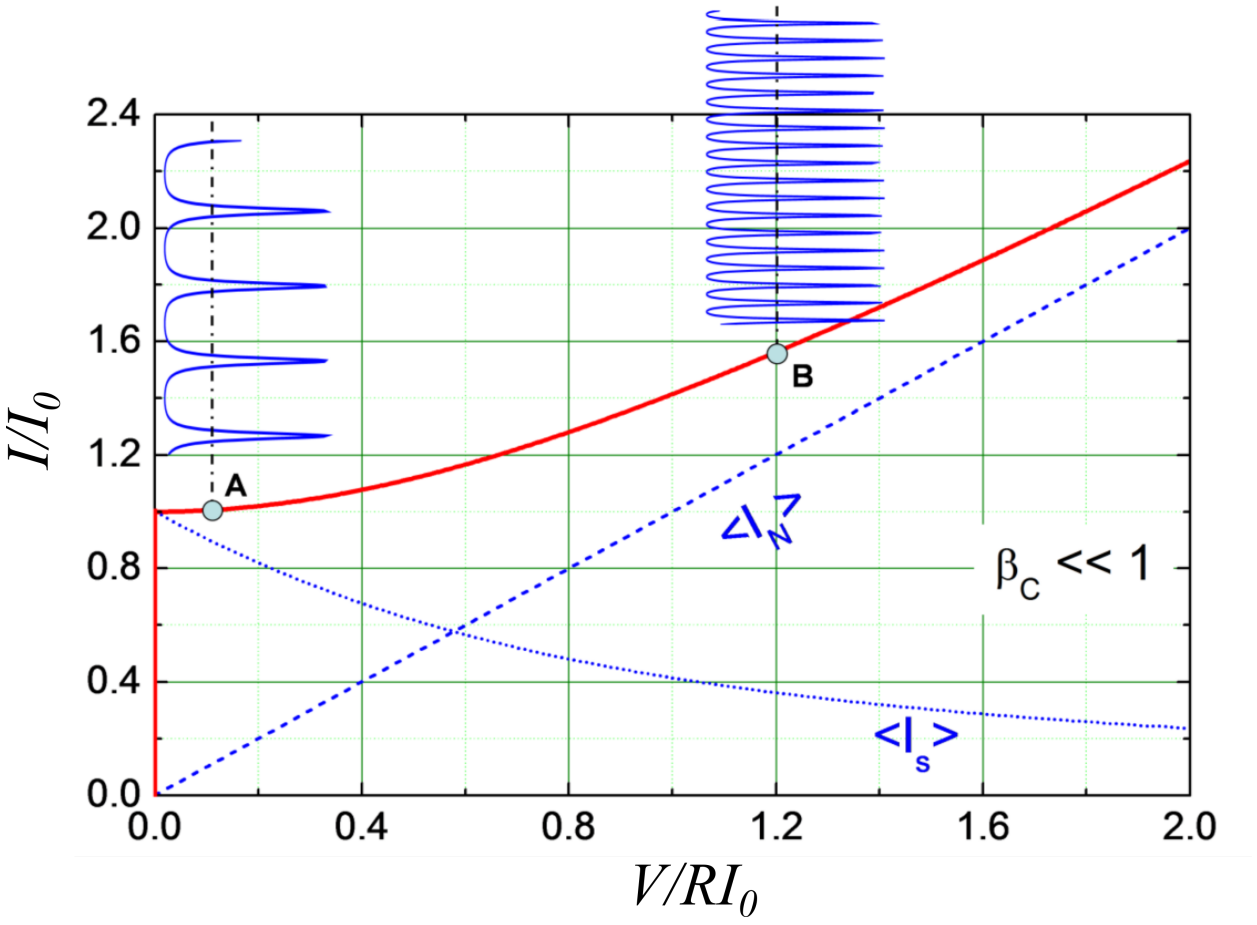

Figure 4.5: Voltage pulses in an overdamped junction. Red is the time-averaged RCSJ current and voltage, dashed blue is the normal (dissipative) current component and dotted blue is the Josephson supercurrent. Points **A** and **B** index two time-voltage curves in solid blue. At **A** I is just above I_0 and $\langle V \rangle$ is the average of occasional, sharp pulses. At **B** I is well above I_0 and $\langle V \rangle$ is the average of a constant V plus pulses of the same amplitude but at greater frequency.

work have a Josephson frequency of about 10 GHz and a signal frequency of about 700 MHz, satisfying this criterion, if only just.

4.2 The DC SQUID

We have seen that when a superconductor is fashioned into a ring, only certain discrete values of flux $\Phi = n\Phi_0$ are allowed to thread the ring. However, if the ring is interrupted by one or more Josephson junctions, the resulting interference effects allow creation of a device with exquisite sensitivity to magnetic fields called a Superconducting QUantum Interference Device, or SQUID. A ring with one junction is called an AC SQUID[3], and a ring with more than one junction is called a DC SQUID. The simplest DC SQUID has just two junctions and this case is presented here in detail.

Recall that from 4.12 the flux allowed inside a superconducting ring is quantized due to the requirement that the phase be single-valued

$$\oint \nabla\theta \cdot d\mathbf{l} = \frac{-2\pi}{\Phi_0}\int_S \boldsymbol{B}\cdot d\mathbf{S} = -2\pi\frac{\Phi_{enc}}{\Phi_0} = 2\pi n. \tag{4.34}$$

We now interrupt that ring with two junctions, as shown in figure 4.6. With the introduction of two Josephson junctions to the ring, there are two points along the integration contour where the phase may make a discrete jump of δ_1 and δ_2 and equation 4.34 is modified to

$$\oint \nabla\theta \cdot d\mathbf{l} = \delta_1 - \delta_2 = -2\pi\frac{\Phi_{enc}}{\Phi_0} = 2\pi n, \tag{4.35}$$

or

$$2\pi\frac{\Phi_{enc}}{\Phi_0}\,[\text{mod } 2\pi] = \delta_1 - \delta_2. \tag{4.36}$$

A bias current I_b is applied across sections of the ring separated by the junctions. The applied current splits to travel through the left and right junctions then recombines to exit the device. Any difference in current between the two junctions is parameterized as a circulating current J. The ring has inductance L. The total magnetic flux through the loop is the sum of any externally applied flux Φ_a and flux due to the circulating current J.

$$\Phi_{enc} = \Phi_a + \Phi_J = \Phi_a + JL. \tag{4.37}$$

For simplicity, we initially take the inductance L to be negligible so $\Phi_{enc} = \Phi_a$, and the junctions to be identical so $I_{0,1} = I_{0,2} = I_0$. Clearly, the maximum supercurrent that can

[3]Strictly speaking, it is not possible to exhibit interference in a device with only one junction. The so-called "AC SQUID" would be better described as a "non-linear inductor". This non-linearity can be used to measure a static magnetic field and is the basis of the flux Qubit, among many other applications, but it is not an "interference device". Nonetheless, the terminology "AC SQUID" is used here for the sake of convention.

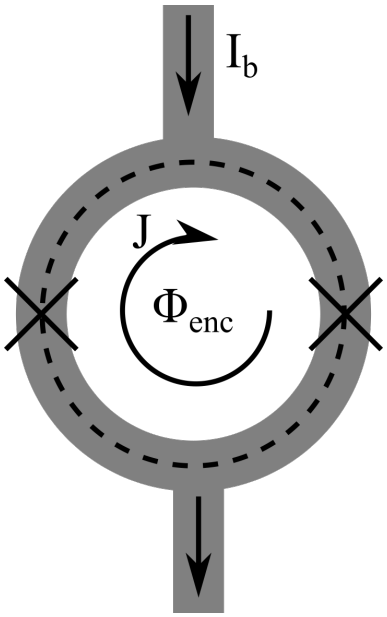

Figure 4.6: Schematic of the DC SQUID. The applied bias current I_b splits between the left and right branches and recombines at the bottom terminal. Any asymmetry between the left and right currents is encoded as a circulating current J. The flux through the loop Φ_{enc} is the sum of flux due to the circulating current and any applied outside flux. The relative phases between the two branches is a function of Φ_{enc}.

flow through the device is $2I_0$, with $\delta_1 = \delta_2 = \pi/2$, so by equation 4.36 Φ_a is an integer multiple of Φ_0. In contrast, if $\Phi_a = (n + 1/2)\Phi_0$, then $\delta_1 - \delta_2 = \pi$ and

$$\frac{I_{s,1} + I_{s,2}}{I_0} = \sin(\delta_1) + \sin(\delta_2) = \sin(\delta_1) + \sin(\delta_1 + \pi) = \sin(\delta_1) + \sin(\delta_1) = 0. \tag{4.38}$$

Evidently, when $\delta_1 - \delta_2 = \pi$ the device supercurrent is entirely suppressed and no current can flow without developing a voltage across the device. In the limit of $L \to 0$ the function of critical current as a function of applied flux is exactly solvable:

$$I_c = 2I_0 \cdot \left| \cos\left(\pi \frac{\Phi_a}{\Phi_0}\right) \right|. \tag{4.39}$$

This is a close analogy to the two-slit experiment with coherent light, where the we equate the phases δ_1 and δ_2 to the path lengths of the waves, the total supercurrent I_c to the beam intensity, Φ_a to the angle relative to the slit plane normal, and Φ_0 to the angle between fringes. At this point, the acronym Superconducting QUantum Interference Device should seem well chosen.

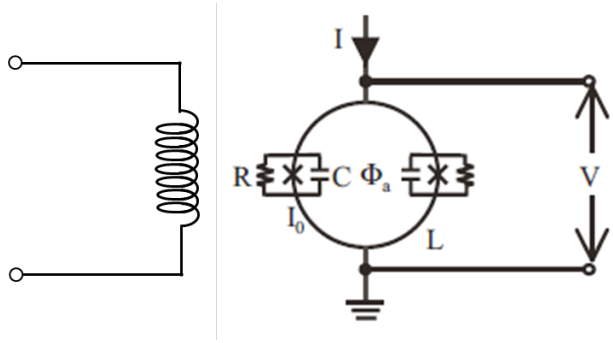

Figure 4.7: Schematic of a DC SQUID. Each junction is a RCSJ, the loop has inductance L, and flux Φ_a is applied by an external coil on the left.

In a more general case, L≠0, the applied flux of $\Phi_a \neq n\Phi_0$, the phase differences are not equal ($\delta_1 \neq \delta_2$), the circulating current J is non-zero, and flux due to the circulating current must be accounted for. Introducing the unitless inductance parameter

$$\beta_L \equiv \frac{2I_0 L}{\Phi_0} \tag{4.40}$$

equation 4.36 becomes

$$2\pi\frac{\Phi_{enc}}{\Phi_0} = 2\pi\frac{\Phi_a}{\Phi_0} + \pi\beta_L\frac{J}{I_0} = \delta_1 - \delta_2. \tag{4.41}$$

In a practical device the junctions will be resistively and capacitively shunted, so the current through junction ($k = 1, 2$) is

$$\frac{I_b}{2} \pm J = I_{0,k}\sin\delta_k + \frac{\Phi_0}{2\pi}\frac{1}{R_k}\dot{\delta_k} + \frac{\Phi_0}{2\pi}C_k\ddot{\delta_k}. \tag{4.42}$$

A schematic of a DC SQUID, including junction capacitance and resistance, loop inductance, applied flux and external flux coil, is shown in figure 4.7. A nonzero L has no effect on the maximum critical current ($I_c = 2I_0$) but the ring inductance has the effect of screening any applied flux and increasing the minimum critical current. The effect of finite L can be numerically calculated for the RCSJ SQUID model of figure 4.7 and and the result is shown in figure 4.8.

At $\beta_L \ll 1$ the modulation depth $\Delta I_c/I_{c,max}$ is approximately independent of L, at $\beta_L = 1$ the modulation depth is halved, and at $\beta_L \gg 1$ the modulation depth $\Delta I_c/I_{c,max} \propto 1/\beta_L$. Numerical analysis accounting for the effects of temperature show that performance is optimized for values of $\beta_L \simeq 1$ [93], which is typical for devices described in this work.

A DC SQUID behaves very much like a "tunable" junction, with the critical current I_c adjustable between $2I_0$ and (typically) I_0. Any current in excess of I_c results in a voltage appearing across the SQUID. Figure 4.9 shows V vs Φ_a for a device biased at $I_b = 2I_0$.

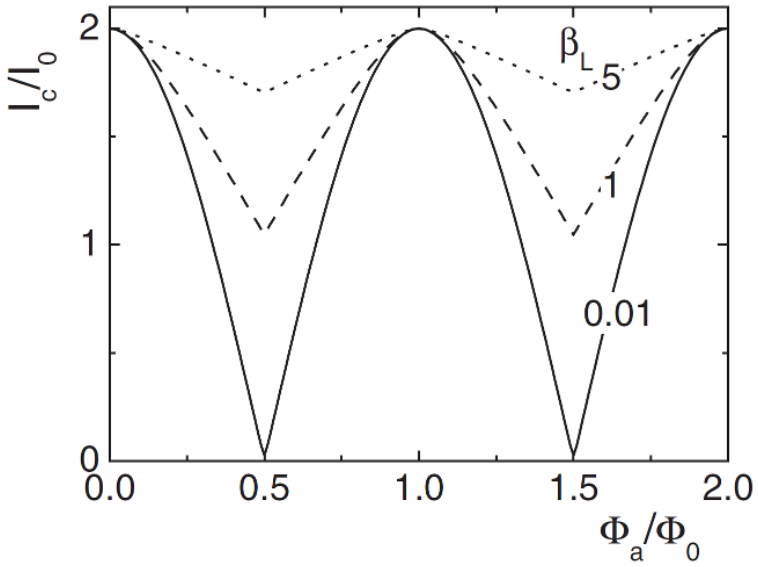

Figure 4.8: Critical current I_c as a function of applied flux Φ_a. For vanishing loop inductance L, I_c modulates to zero. For equal loop inductance and Josephson inductance ($\beta_L = 1$), I_c modulates to half its maximum value. Modulation depth reduces to zero for $\beta_L \gg 1$. Reproduced from [92].

When $\Phi_a = n\Phi_0$, no voltage develops, but for other values of Φ_a, $V > 0$. Near flux biases of $\Phi_a/\Phi_0 \approx n \pm 0.25$, $|dV/d\Phi_a|$ is maximized and the SQUID acts as a very sensitive flux-to-voltage transducer, with typical flux resolution of $10^{-6}\Phi_0/\sqrt{\text{Hz}}$. In figure 4.9 the red and blue annotations illustrate that if $\Phi \approx (n + 1/4)\Phi_0$, small variation in Φ can result in large variations in V.

Figure 4.10 shows the flux modulated I-V characteristic of an overdamped DC SQUID biased near $\Phi_a = 0.25$. The curves show I-V for $\Phi_a/\Phi_0 = 0.2525$ (solid blue) and $\Phi_a/\Phi_0 = 0.2475$ (dashed red), for $\Delta I_c/I_0$ of 0.005 and $\Delta V/RI_0 \approx 0.1$. In this example,

$$\frac{dV}{d\Phi} \cdot \frac{\Phi_0}{RI_0} = 20 \tag{4.43}$$

or with typical values of $R = 10\,\Omega$ and $I_c = 3\,\mu A$,

$$\frac{dV}{d\Phi} \equiv V_\Phi \approx 600\,\mu V/\Phi_0. \tag{4.44}$$

In practice, the knee on I-V transitioning from the superconducting to voltage state is not sharp as shown in figure 4.10, but is somewhat rounded due to thermal noise (see section 4.2.1), which reduces ΔV, but the damping parameter β_c is typically not $\ll 1$ as in figure 4.10 but is closer to 1, resulting in greater dV/dI, and greater ΔV in turn. The two effects

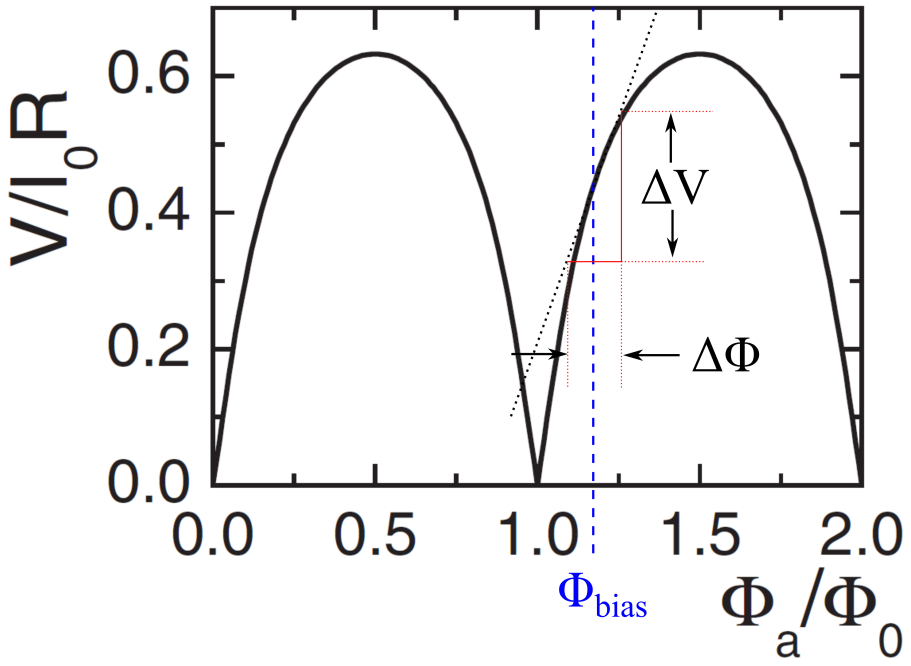

Figure 4.9: SQUID voltage as a function of applied flux Φ_a. Blue dashed line marks a flux bias near $\Phi_a/\Phi_0 \approx n \pm 0.25$, where $|dV/d\Phi_a|$ is large. At this flux bias, the DC SQUID is a very sensitive flux-to-voltage transducer. Adapted from [92].

somewhat cancel, and the value of $dV/d\Phi$ given in equation 4.44 is to an order of magnitude typical for practical real-world devices.

4.2.1 Thermal Noise in the DC SQUID

The treatment of the DC SQUID so far has neglected noise considerations. At sufficiently high temperatures, the quantum effects will be completely washed out by thermal noise. A good figure of merit is the ratio of the typical thermal energy k_BT to the Josephson coupling energy E_J

$$\Gamma \equiv \frac{k_BT}{E_J} = \frac{2\pi k_BT}{I_0\Phi_0}. \tag{4.45}$$

If $\Gamma \geq 1$, then the Φ_0-periodic potential of figure 4.3 becomes insignificant and the junction behavior becomes entirely classical. For small values of Γ, one can imagine a junction biased at I just below I_0, corresponding to a potential like figure 4.3 but with very little tilt. The particle will usually remain in a single well, with occasional thermally activated jumps to a neighboring well. Due to the tilt, the right hand barrier is lower than the left hand barrier, so jumps to the right are more common then jumps to the left, and there is an average trend

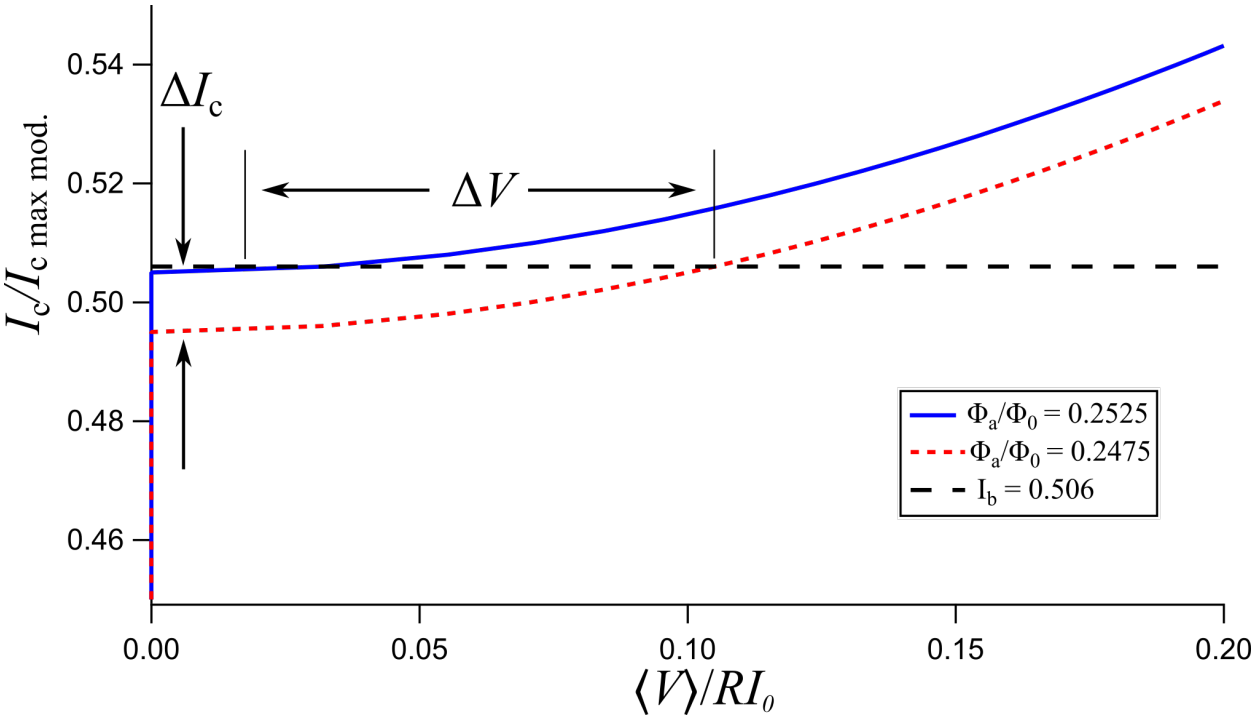

Figure 4.10: DC SQUID I-V, modulated by applied flux Φ_a. Blue solid curve is the I-V at Φ_a/Φ_0 just over 1/4 and the red dashed curve is the I-V at Φ_a/Φ_0 just under 1/4. A small change in applied flux modulates I_c, resulting in a large modulation in V.

of $d\delta/dt > 0$, or $\langle V \rangle > 0$, so the I-V knee is slightly rounded, and $V > 0$ even for $I < I_0$. Figure 4.11 illustrates this "knee softening" for different values of Γ. In practical operation, we require $\Gamma \leq 0.05$. At the highest temperatures and typical I_0 presented in this work (4.2 K and 3 μA, respectively), $\Gamma \approx 0.044$.

We must also ensure the thermal noise energy is small relative to the inductive energy per flux quantum

$$\frac{4\pi k_B T L}{\Phi_0^2} \ll 1, \tag{4.46}$$

lest the thermal noise utterly obscure any quantum interference. The largest inductance and temperatures presented in the work are T = 4.2K and L = 431 pH for $4\pi k_B TL/\Phi_0^2 \approx 0.07$, and the smallest inductance and temperatures presented in this work are T = 0.1K and L = 80 pH for $4\pi k_B TL/\Phi_0^2 \approx 3 \times 10^{-4}$.

To maximize the voltage-transfer coefficient ($dV/d\Phi \equiv V_\Phi$) in the presence of noise, the device parameters should be chosen such that $\beta_C = \beta_L = 1$[93], and then V_ϕ is proportional to R/L

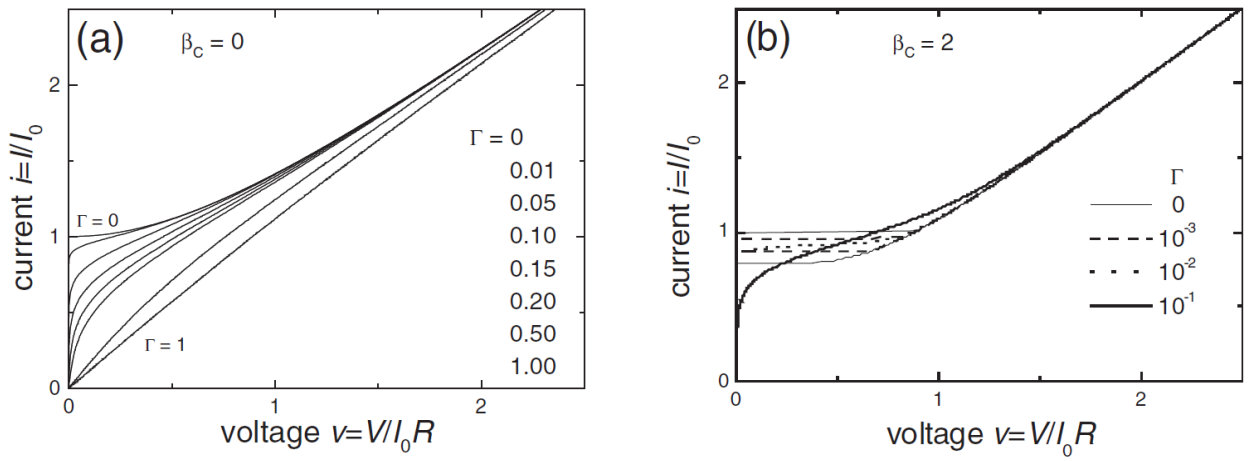

Figure 4.11: Junction I-V at finite temperature. (a) I-V for an overdamped junction. At $\Gamma = 1$ the junction is indistinguishable from a resistor. Even at $\Gamma = 0.01$, noise rounding is noticeable. (b) I-V for a junction with $\beta_c > 1$, exhibiting hysteresis. At $\Gamma = 0$ hysteresis is greatest, but the hysteresis is reduced as the temperature increases. At sufficiently high temperature, the hysteresis is entirely masked by thermal noise. A seemingly non-hysteretic device may be revealed to be hysteretic when measured at lower temperatures. Reproduced from [92].

$$V_\Phi \approx R/L \approx (\pi LC)^{-1/2}. \tag{4.47}$$

There will be Nyquist noise $S_v(f)$ at the SQUID terminals due to the shunt resistors. When using a SQUID as a flux-to-voltage transducer, it is natural to convert the voltage noise to an equivalent flux noise by $S_\Phi(f) = S_V(f)/V_\Phi^2$ to give

$$S_\Phi(f) \approx 16 k_B T L^2/R, \tag{4.48}$$

again assuming

$$\beta_C = \frac{2\pi I_0 R^2 C}{\Phi_0} = \beta_L = \frac{2 I_0 L}{\Phi_0} = 1. \tag{4.49}$$

4.2.2 Practical DC SQUID Fabrication

A SQUID is typically fabricated by depositing and patterning a thin film of Nb or other superconductor on an insulating substrate, as shown in figure 4.12. The main loop is formed as a large square washer with a thin slit. The junctions are usually pushed towards one side of the loop and reside very close together at the end of the washer slit, which helps reduce variation between the junctions due to unintended fabrication gradients. The junctions are shunted with deposited thin-film resistors, usually Au, Pd, or a Au-Cu alloy. The entire

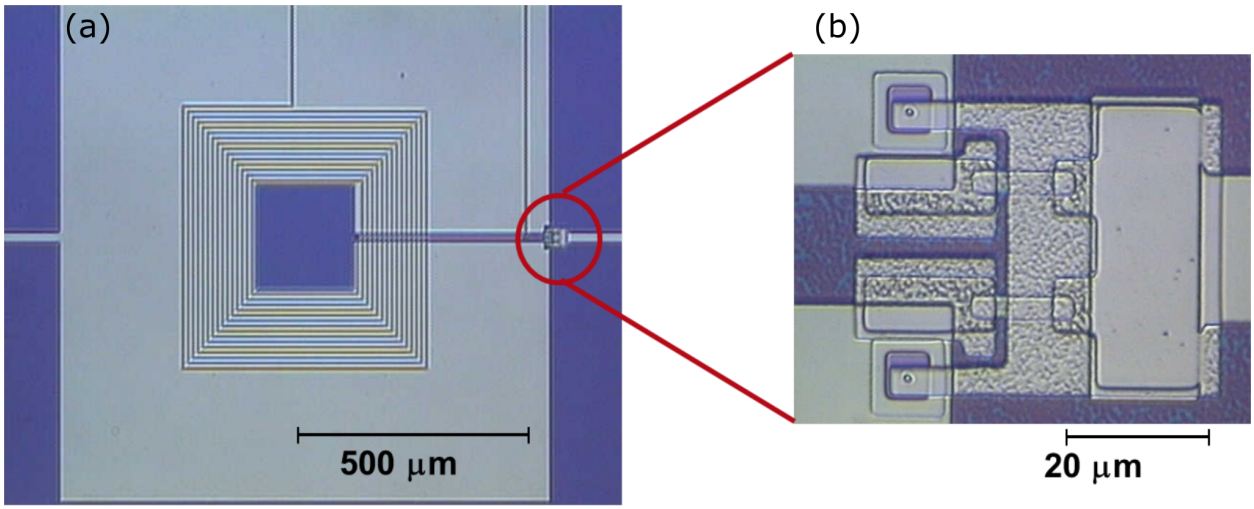

Figure 4.12: SQUID micrograph. A practical SQUID is composed of thin patterned layers of metals and insulators on a silicon or other inert substrate. (a) Overview of a typical SQUID. Grey is Nb, blue is the underlying substrate, and fine lines are a second Nb layer separated from the base Nb film by a thin transparent insulating layer, typically SiO_x. The superconducting loop is the large square grey washer, and the junctions are moved very close together on the far right. The bias current I_b enters from the left, travels around the top or bottom of the center hole, flows through the top or bottom junction, and exits on the right. Flux is coupled through an integrated input coil fabricated as a flat square spiral. A flux-coupling current flows from the top center, circles the central hole in turns of diminishing radius, crosses under itself making use of the slit in the main washer, and exits on the top right. (b) Detail of the junctions. The junctions are pinpoint-like dots near the edge of the washer, inside windows made in the oxide layer. The junctions connect to a common Nb counter-electrode on the right. Each junction is shunted by a parallel resistor to ensure non-hysteretic operation. (Courtesy Xiaofan Meng and Robert McDermott)

assembly is covered with an insulating layer, usually SiO_x, and a second layer of Nb or other superconductor patterned on top to create a spiral flux coupling coil. The two SQUID terminals and two coil terminals typically run to large bonding pads (not shown) allowing wirebonding to external circuitry.

4.3 Microstrip Coupling to the DC SQUID

A DC SQUID operates very well via flux coupling to an external source (antenna, material sample, etc) through its integrated input coil at frequencies from 0 to about 200 MHz[94]. At frequencies above 200 MHz, capacitive coupling between the input coil and SQUID washer

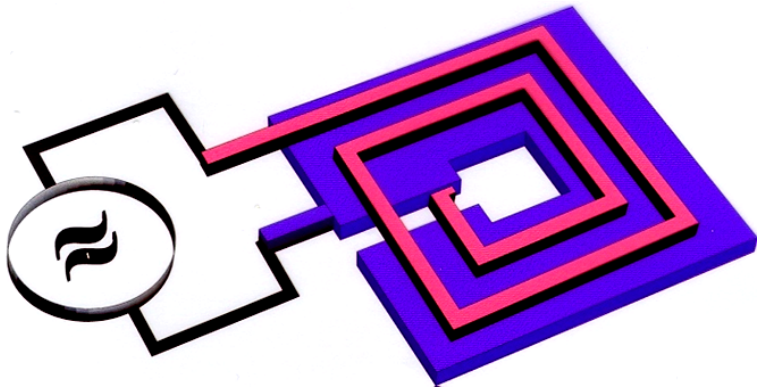

Figure 4.13: MSA coupling principle. One end of the input coil is left unconnected and the other is driven by the signal source. The resonant frequency is set by the input coil inductance and capacitance, which depend on geometric parameters.

tends to shunt the input signal, or even lead to undesired resonances. ADMX requires a microwave receiver in the range of 1 GHz, so a different coupling method is needed.

To operate at high frequencies, the ground-referenced input signal is fed to one end of the input coil, and the other end of the input coil is left unconnected. Figure 4.13 illustrates this concept. The spiral input coil now acts as a microstrip and the SQUID washer as the microstrip groundplane such that the once deleterious capacitive coupling becomes a key transmission element. If the input is weakly coupled to the microstrip input, the input coil forms a $\lambda/2$ microstrip resonator with a frequency set by the total input coil inductance and capacitance. The first MSA was demonstrated in in Clarke group in 1998 [95].

This then, is the titular Microstrip SQUID Amplifier (MSA) that has proved a capable low-noise microwave amplifier. The device as described to this point device works only over a narrow frequency range determined by the geometry of the SQUID and integrated input coil. Figure 4.14 shows the gain vs frequency of an early MSA. The lowest frequency corresponds to an MSA with a 31-turn input coil, and higher frequency performance is achieved by progressively removing turns from the input coil with a photomask and reactive ion etching (an inherently irreversible process). By reducing the number of turns from 31 to 6, the frequency is increased from 200 MHz to 625 MHz.

Operation as an MSA leads to a new configuration option not relevant to the DC SQUID: the possibility for positive or negative feedback. The input coil is capacitively coupled to the SQUID washer "ground plane", but the SQUID washer need not be microwave ground. One can instead connect the small section of the SQUID loop between the junctions (the counterelectrode) to microwave ground, and allow the SQUID voltage to appear on the washer. In this case, the output voltage will feed back to the input coil via the capacitive coupling. Note from figure 4.9 that the sign of $V_\Phi \equiv dV/d\Phi$ may be positive or negative depending of the choice of the static Φ_{bias} (with Φ_{bias}/Φ_0 typicaly a little more or an little

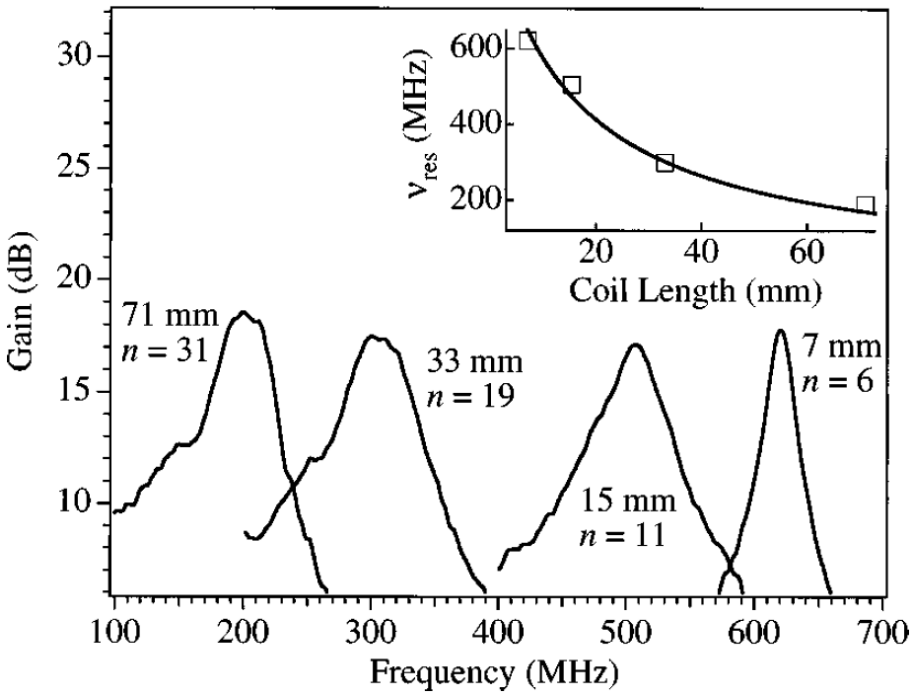

Figure 4.14: Early MSA results. Resonant frequency is set by the input coil geometry. A many-turn input coil (n = 31) yields a resonant frequency of 200 MHz, and a shorter, 6-turn coil yields a resonant frequency of 625 MHz. Reproduced from [95].

less than one), so in the case of a "grounded counterelectrode" (ie "active washer") MSA, the sign of the feedback may be positive or negative (V_Φ^+ or V_Φ^-, respectively). Figure 4.15 illustrates MSA performance differences for three different feedback regimes. In the case of a grounded washer, the gain is moderate (about 14 dB) and the Q is about 10. Feedback coupling is minimal, so the two feedback signs V_Φ^+ and V_Φ^- are nearly identical and only V_Φ^+ is shown. In the case of the grounded counterelectrode, there is a stark difference between V_Φ^+ and V_Φ^-. In the positive feedback case, the gain in increased to about 18 dB, and the Q is increased to about 31. In the negative feedback case, the gain is diminished to a few dB and no resonance is evident.

A schematic of the MSA coupling used to generate figure 4.15 is shown in figure 4.16. The input signal enters from the left-hand coaxial line and encounters a network of matching resistors. From the matching network, the signal continues to the microstrip input coil, shown as a distributed inductance and capacitance. The SQUID is the rectangle at center, with the bias current fed from the top, the junctions at the bottom, and the counterelectrode grounded. The voltage developed across the SQUID is coupled to the output coaxial line on the right through another resistive matching network. As drawn, the counterelectrode is grounded and the washer is "active", so the output signal is capacitively coupled to the

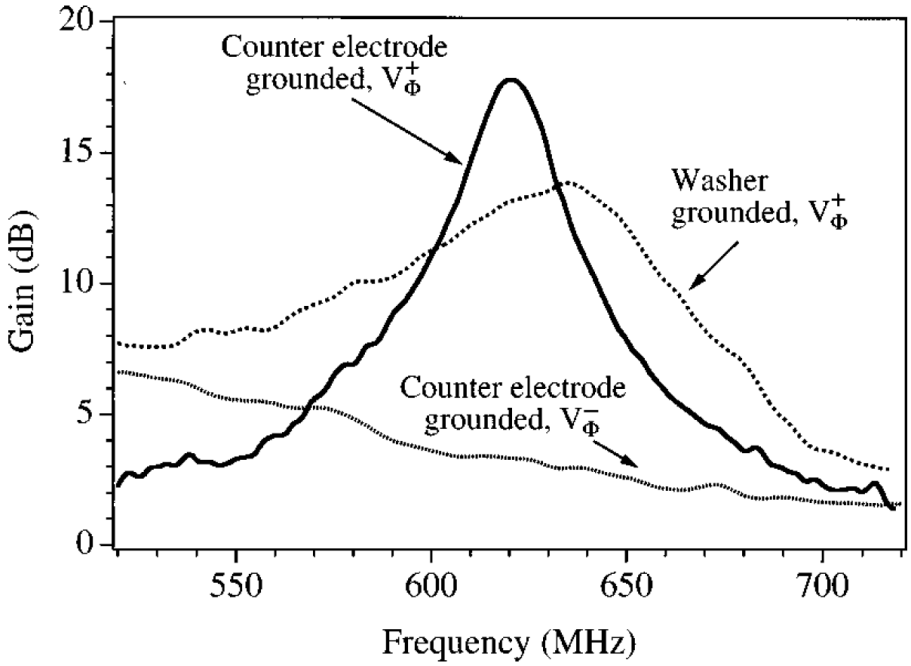

Figure 4.15: MSA Feedback. With the washer grounded, there is minimal feedback to the input coil, gain and Q are moderate, and V_Φ^+ and V_Φ^- are nominally identical so only V_Φ^+ is shown. With the counterelectrode grounded and the washer "active", the output voltage is capacitively coupled to the input coil, and the sign of feedback is set by the choice of static flux bias. In the case of positive feedback (V_Φ^+) the gain and Q are increased. In the case of negative feedback (V_Φ^-) the resonance is entirely suppressed and gain is diminished. Reproduced from [95].

input line. Choice of the flux bias current I_Φ determines the flux sensitivity $dV/d\Phi$, which may be of either sign, selecting for positive or negative feedback. Switching the connections at the points labeled "A" and "B" would result in a "grounded washer" configuration, with little to no capacitive coupling between the input and output, and thus little to no feedback.

While a fixed-frequency narrow-band amplifier is a wonderful thing, ADMX (and other applications) require operation over a wide range of frequencies. If a wide instantaneous bandwidth is not needed, the center gain frequency can be tuned by addition of additional external components.

4.4 The Tunable MSA

The termination of the input coil illustrated in figure 4.13 is an open circuit, resulting in a reflected phase of π and $\lambda/2$ resonance. If the end were shorted to ground, the reflected

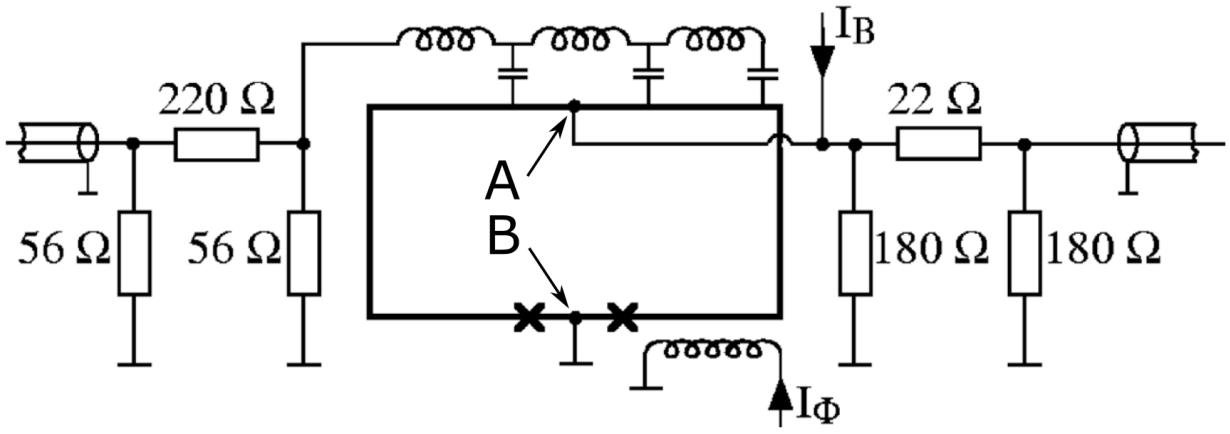

Figure 4.16: MSA coupling schematic. The input signal enters from the left on a 50-Ω transmission line. The left-land resistor network matches the transmission line impedance to the MSA input coil impedance (with not insignificant loss). The MSA resonant input coil is represented by a distributed inductance and capacitance. The flux component of the standing wave couples to the SQUID loop, which if tuned to a high $|dV/d\phi|$, develops a large output voltage that is coupled to the output line. The right-hand resistor network matches the SQUID output impedance to the output transmission line impedance. If the SQUID washer carries the output signal (as shown) the output signal is capacitively coupled to the input coil, allowing for either positive or negative feedback, depending on the chosen sign of $dV/d\phi$. If the electrical connections at points A and B are switched, there will be no feedback. Adapted from [95].

phase would be 0, the allowed resonance would be altered to $\lambda/4$ and the resonant frequency halved. If the end is instead connected to ground through a reactive element (for example, a capacitor), the resonance frequency will be shifted to a point between these two extremes. An expression for the phase or reflection ϕ is given in [96] as

$$\begin{aligned} \phi = \tan^{-1}[2Z_0\omega C/(Z_0^2\omega^2C^2-1)] + \pi,\ C < 1/Z_0\omega, \\ \phi = \pi/2,\ C = 1/Z_0\omega, \\ \phi = \tan^{-1}[2Z_0\omega C/(Z_0^2\omega^2C^2-1)],\ C > 1/Z_0\omega. \end{aligned} \tag{4.50}$$

However, we prefer the mathematically equivalent but cleaner and more intuitive expression

$$\phi = 2\tan^{-1}\left(\frac{1}{\omega C Z_0}\right), \tag{4.51}$$

valid for a terminating capacitance, or the generalized expression for an arbitrary termination load Z_L

$$\phi = 2\tan^{-1}\left(i\frac{Z_L}{Z_0}\right). \tag{4.52}$$

Adding a fixed capacitor is more convenient than physically altering the integrated SQUID input coil, but still results in a fixed resonant frequency. For live tunability we can terminate the input coil with a varactor instead a fixed capacitor. A varactor is a semiconductor junction reverse-biased such that the DC current flow is virtually zero. Under reverse-bias conditions, the two conduction layers (the N-doped and P-doped regions) are separated by a non-conducting depletion zone, forming a capacitor. The width of the depletion layer is a function of the applied voltage, so the applied voltage determines the capacitance. With a sufficiently large swing in terminating capacitance, the MSA can be tuned over up to an octave in frequency. Figure 4.17 shows a schematic of just such a setup. In the center in black is a square SQUID loop and DC flux bias coil. The integrated RF input coil is represented as a distributed inductance and capacitance (blue) coupled to the SQUID loop. The input coil is terminated by a varactor, shown in red. The RF input signal from the left is coupled to the resonant input coil by a resistive coupling network (green), and the RF output is coupled to the right by a similar resistive network. The standing wave on the resonant input coil couples magnetically to the SQUID, which when biased to a point of large $dV/d\Phi$ by the external DC flux coil develops a large voltage at the output terminal in phase with the standing wave, which is coupled to the output circuitry. Figure 4.18 demonstrates performance of an MSA tuned with a varactor. In this example the unterminated frequency was 215 MHz, and with attached varactors the frequency is tunable from 117 to 195 MHz.

The performance demonstrated in [96] is a compelling proof of concept, but it leaves the question of noise temperature (T_N) unaddressed. As demonstrated in chapter 3, an MSA must have a low T_N to be useful as the primary amplifier in ADMX. Outside of MSAs, the best available microwave amplifier at the frequencies of immediate interest for ADMX (about 560 MHz to 850 MHz) are based on Heterostructure Field-Effect Transistors, or HFET amplifiers [97]. An optimized HFET has a T_N of about 2 K. An MSA operating at about 500 MHz is described in [98] with a T_N calculated at 50 mK[4], albeit without tunability. Figure 4.19 Shows an MSA T_N as a function of physical temperature measured at a narrow band centered at 538 MHz.

ADMX ultimately would like to explore frequencies well above 1GHz, and MSAs have been demonstrated with operating frequencies of up to 7.4 GHz [99]. While the MSAs in this example are relatively low gain and of unknown T_N, it is a notable demonstration of MSA operation at rather high frequency. More recent work has demonstrated MSAs with gain above 24 dB at frequencies up to 2.5 GHz by clever use of higher-order modes in the

[4]This figure for T_N was calculated from a series of amplitude calibrations, depending on various gain and attenuation measurements measured as separate systems and algebraically combined in the final T_N calculation. The calculated MSA T_N was a relatively small contribution to the total system noise temperature, so there is substantial opportunity for the introduction of systematic errors. Even if this calculation for T_N is an order of magnitude too small, the result remains a compelling demonstration of low noise operation of an MSA.

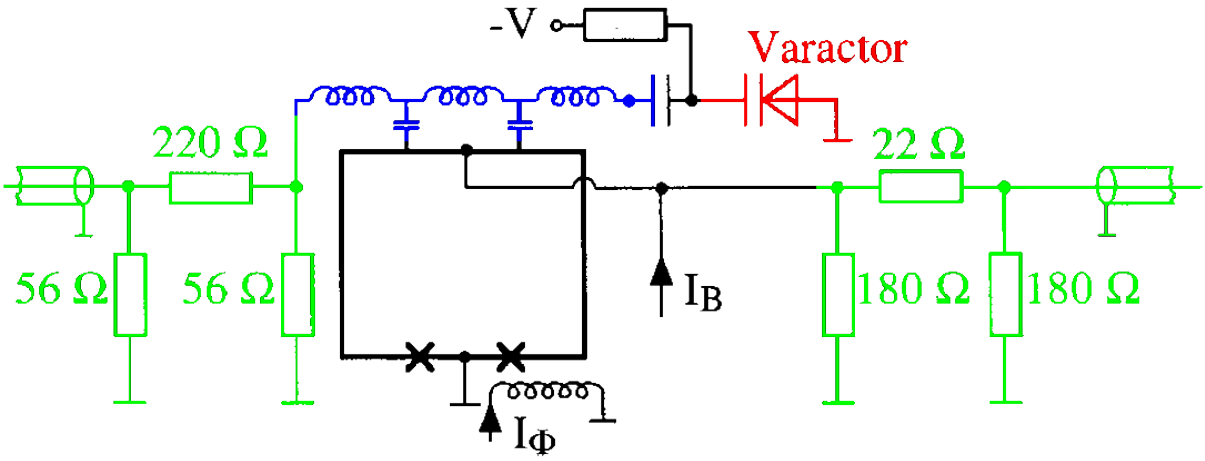

Figure 4.17: Schematic of a tunable MSA. Green components are the RF input (left) and output (right) matching networks. Center black rectangle and the SQUID loop with junctions. Blue is the integrated resonant input coil, and red is the tunable varactor, which sets the resonant frequency by adjusting the phase of reflection. Reproduced from [96], with color added.

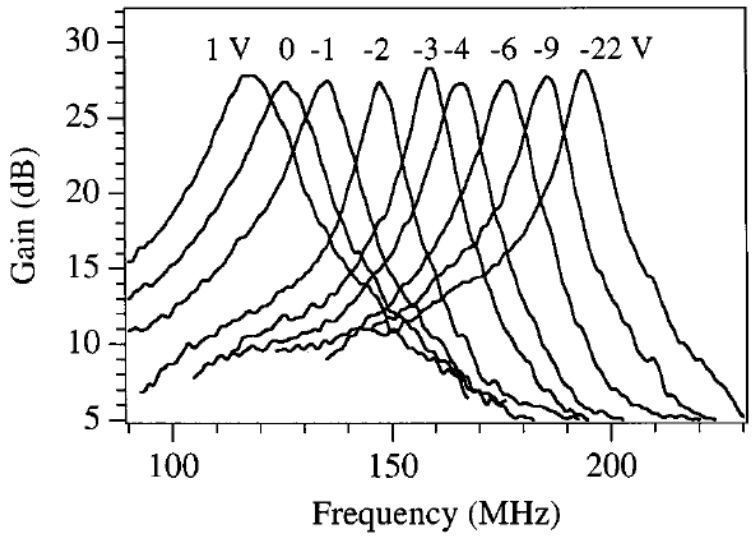

Figure 4.18: Early MSA varactor tuning demonstration. At very negative varactor bias, the bipolar junction is fully depleted, the terminating capacitance is minimized, and the resonant frequency is maximized. At zero or slightly positive voltages, the depletion region of the varactor is very thin, resulting in a large capacitance and the lowest frequencies. Reproduced from [96].

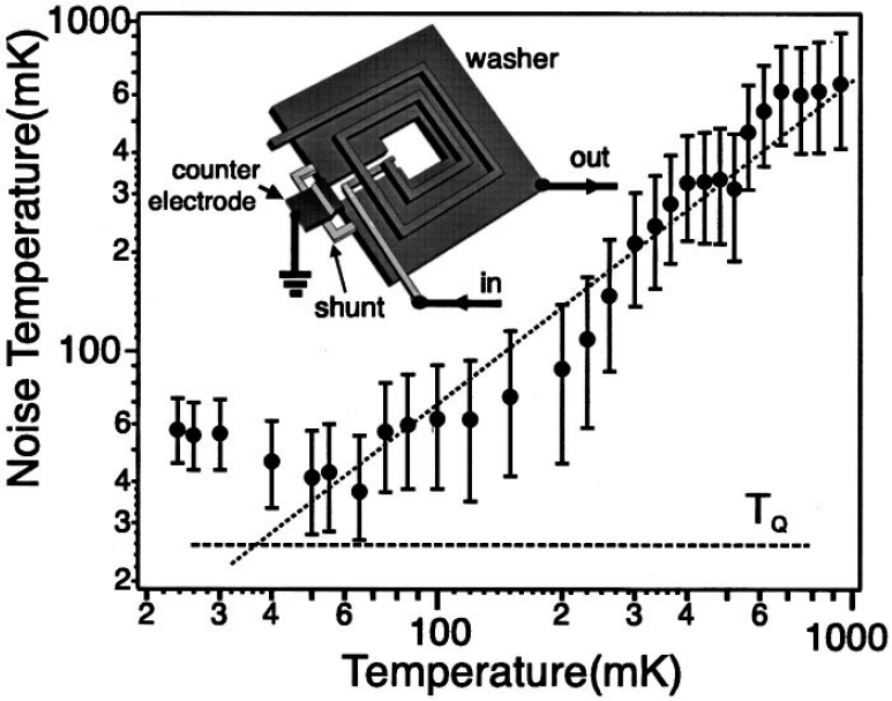

Figure 4.19: Fixed-frequency MSA noise temperature. Diagonal line corresponds to $T_N \propto T$, and the horizontal line represents $T_Q = hf/k_B$. Inset shows MSA configuration with grounded counterelectrode. T_N measurement calibrated through characterization of total attenuation through the entire signal path at 538 MHz. Reproduced from [98].

input coil [100]. Figure 4.20 shows a diagram of an MSA input coil with the winding sense reversed in two places to promote high gain at the $3\lambda/2$ mode. Figure 4.21 shows MSA gain profiles for a standard $\lambda/2$ resonant input coil, an input coil with a $3\lambda/2$ standing wave, and a $4\lambda/2$ standing wave.

Publications preceding this work has explored further details of MSA performance, for example characterization of S-parameters [101] or conditional stability criteria [102]. This work will re-open these issues with a new take on techniques and revised results of MSA performance in chapters 6, 5, and 7. In a very important pathfinding experiment in 2011, an MSA was demonstrated operating in a frequency range of interest to ADMX with sufficiently high gain (G$\approx$ 20 dB) and sufficiently low T_N ($T_N \approx$ 45 mK) to substantially increase the scan rate [103]. Data illustrating this result are shown in figure 4.22.

The key criteria needed in an amplifier for use in ADMX are:

- Operation in the ADMX frequency range (560 to 850 MHz).
- In situ tunability.

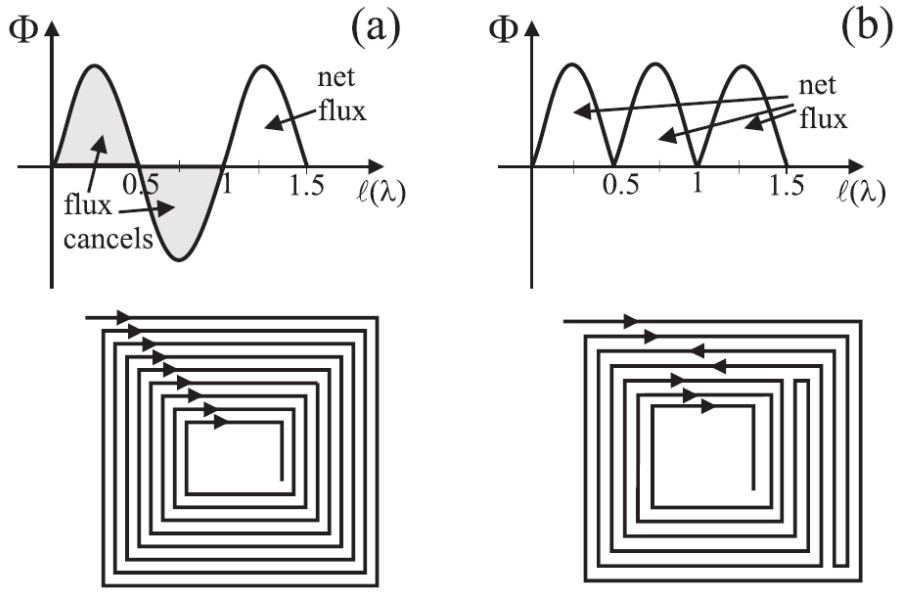

Figure 4.20: An MSA input coil with higher-order modes enabled by reversing the winding sense. (a) Higher-order modes can exist on a conventional input coil, but the gain will be low because the total flux coupled to the SQUID is largely canceled. (b) By reversing the winding sense of the input coil at the current nodes, a higher-order mode (here $3\lambda/2$) may have all the flux coupled to the SQUID without cancellation. Reproduced from [100].

- T_N substantially lower than 2 K, ideally measured by a hot-cold load.

An amplifier satisfying all three criteria has not been demonstrated in any prior work. The development, demonstration, and integration with ADMX of amplifiers satisfying all necessary criteria is the primary achievement described in this work. The MSA's described in this work represent a critical technology without which current progress in the Axion dark matter search would not have been possible.

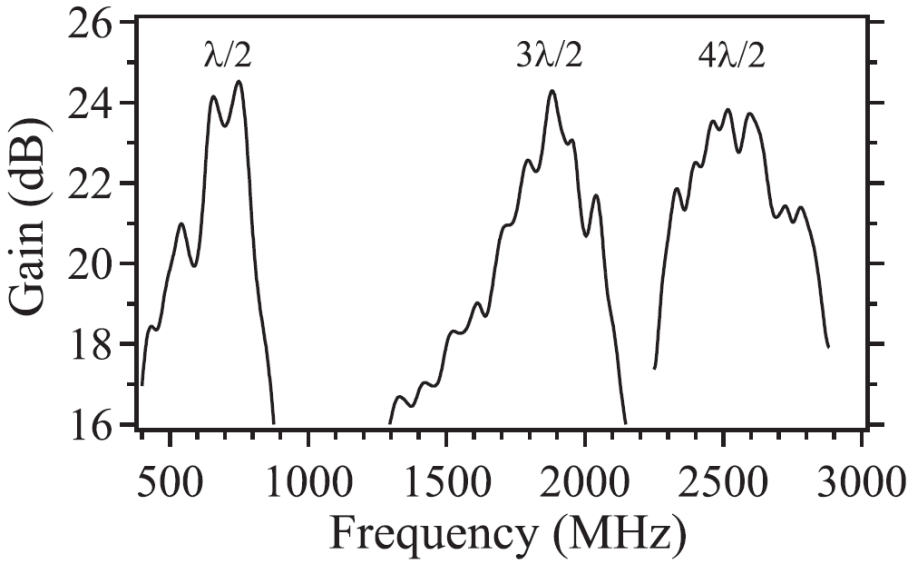

Figure 4.21: High-frequency MSA performance using a high-harmonic input coil. Lowest frequency gain peak is from an MSA with a "classic" half-wave resonance. By "folding back" the spiral input coil, higher-order resonant modes can be made to constructively couple each $\lambda/2$ anti-node of the standing wave to the SQUID. Reproduced from [100].

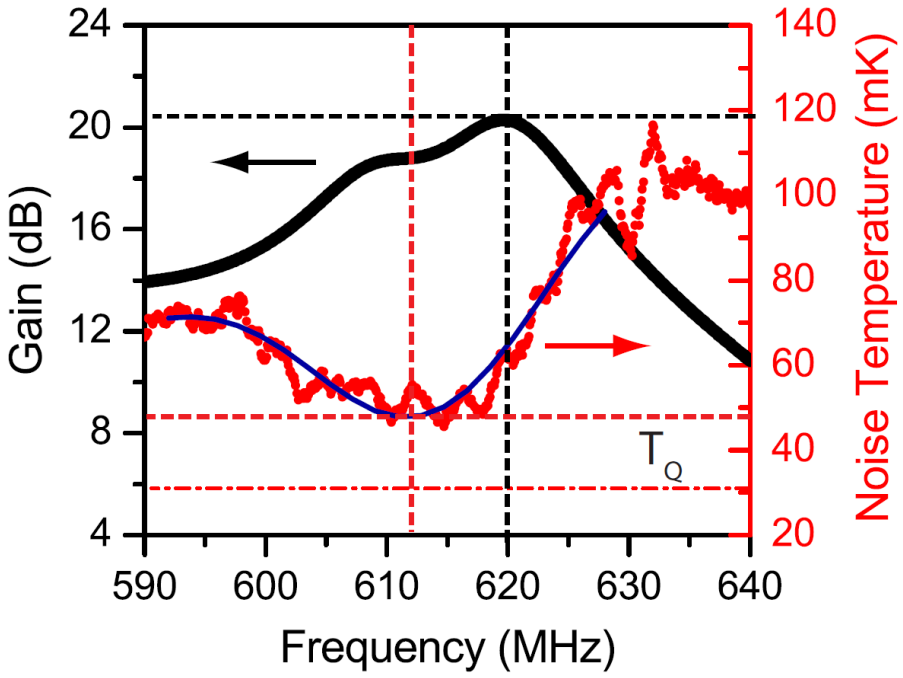

Figure 4.22: Low T_N operation of an MSA at frequencies relevant to ADMX. Lowest T_N is 45 mK at a frequency 612 MHz and gain of 18 dB. Dashed-dotted red line represents $T_Q = hf/k_B$. Reproduced from [103].

Chapter 5

Design and Optimization of the MSA

Here we approach the design considerations that went into the MSAs made particularly for ADMX. The ability to model a 1-D resonator terminated with an arbitrary complex load, and solve for any particular parameter (ex: f_0, Z_0) given any arbitrary set of other parameters (ex: C, ℓ) is a tool used throughout this process. The interrelated choices of DC-SQUID parameters, integrated microstrip design, and RF circuit design are discussed here.

5.1 Model of an arbitrarily loaded 1-D resonator

The tunable MSA makes a 1-D microstrip resonator of the integrated input coil and couples the magnetic field from this resonator to the SQUID. The resonator is tuned by varying the loading of the end terminal with one or more varactor diodes. It is necessary to develop a theoretical and numerical model of a 1-D resonator made from a transmission line with arbitrary termination as an aid to MSA design. With this goal we start with a model of a half-infinite transmission line.

5.1.1 Half-Infinite Transmission Line with Arbitrary Termination

A half-infinite transmission line with microwave impedance Z_0 is terminated at position $x = 0$ with an arbitrary impedance $Z_{x=0}$ as shown in figure 5.1. In practice, this termination impedance may be a varactor diode. The microwave signal on the transmission line is composed of a right-traveling wave $I^+(x,t)$ and a left-traveling wave $I^-(x,t)$, which we describe (up to a global phase) by equations (5.1):

$$\begin{aligned} I^+(x,t) &= I_0 e^{i(\omega t - kx)} \\ I^-(x,t) &= I_0 e^{i(\omega t + kx + \phi)}. \end{aligned} \tag{5.1}$$

We focus on the current component of the traveling wave rather than the voltage component because we couple the magnetic field of the wave to a SQUID in the MSA, and the

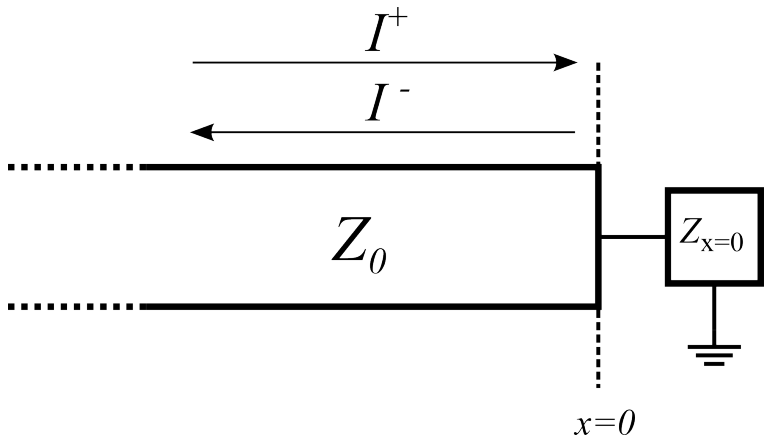

Figure 5.1: Model of a half-infinite transmission line with arbitrary termination.

magnetic field arises directly from the current. Identifying the current as the wave amplitude makes our analysis of the MSA more direct, but has the unfortunate consequence that some results here will be "exactly opposite" the usual results for microwave systems where the *voltage component* of the wave is identified with the amplitude.

The phase ϕ of the left traveling wave is determined by the boundary conditions set by $Z_{x=0}$. To determine ϕ, note that on the transmission line these waves have a voltage component determined by Z_0 and the propagation direction as in equation (5.2):

$$\begin{aligned} V^+(x,t) &= Z_0 I^+(x,t) \\ V^-(x,t) - &= -Z_0 I^-(x,t). \end{aligned} \tag{5.2}$$

Also note that both the current and voltage at $x = 0$ must be continuous with the values at $x = -\epsilon$ (just before the terminal load), so that at $x = 0$, $Z_{x=0}$ is the ratio of the sum voltage to sum current, expressed in equation (5.3) as:

$$\begin{aligned} Z_{x=0} &= \frac{V^+ + V^-}{I^+ + I^-} = \frac{Z_0 I^+ - Z_0 I^-}{I^+ + I^-} = Z_0 \frac{e^{i\omega t} - e^{i(\omega t+\phi)}}{e^{i\omega t} + e^{i(\omega t+\phi)}} = Z_0 \frac{1 - e^{i\phi}}{1 + e^{i\phi}} \\ &\text{and} \\ Z_{x=0} &= -iZ_0 \tan\left(\frac{\phi}{2}\right). \end{aligned} \tag{5.3}$$

This entirely determines the value of ϕ, given Z_0 and $Z_{x=0}$. Note that if $Z_{x=0}$ is purely imaginary, as for an ideal capacitor, ϕ is purely real. However, if $Z_{x=0}$ has a real component, as in the case of some resistive loss, ϕ will have an imaginary component. It is useful to further develop this expression such that the real and imaginary parts of ϕ are expressed in terms of the resistive and reactive components of $Z_{x=0}$ in a form that is amenable to numerical solution:

$$\tan\left(\frac{\phi}{2}\right) = \frac{\sin(a)}{\cos(a)+\cosh(b)} + i\frac{\sinh(b)}{\cos(a)+\cosh(b)} \quad (\text{with } \phi = a+ib). \tag{5.4}$$

For concreteness, suppose $Z_{x=0}$ is a series RC, so that $Z_{x=0} = R + \frac{1}{i\omega C}$. Then from equations (5.3) and (5.4):

$$\frac{iR}{Z_0} + \frac{1}{\omega C Z_0} = \frac{\sin(a)}{\cos(a)+\cosh(b)} + i\frac{\sinh(b)}{\cos(a)+\cosh(b)}. \tag{5.5}$$

Identifying real and imaginary parts gives:

$$\frac{1}{\omega C Z_0} = \frac{\sin(a)}{\cos(a)+\cosh(b)} \text{ (Real part)} \tag{5.6}$$

$$\frac{R}{Z_0} = \frac{\sinh(b)}{\cos(a)+\cosh(b)} \text{ (Imaginary part).} \tag{5.7}$$

To aid in solving these equations we would like an expression for $\cosh(b)$ in terms of a to solve equation (5.6) and an expression for $\cos(a)$ in terms of b to solve equation (5.7). To this end, combining equations (5.6) and (5.7) gives:

$$R\omega C\sin(a) = \sinh(b) \quad (\text{valid for all values of } a \text{ except } a = \pi). \tag{5.8}$$

With the help of the trigonometric identities $\sin^2(x)+\cos^2(x) = 1$ and $\cosh^2(x) - \sinh^2(x) = 1$, equation (5.8) can be re-arranged as needed to give:

$$\begin{aligned} \cosh(b) &= \sqrt{1+(R\omega C\sin(a))^2} \\ \text{and} \quad \cos(a) &= \sqrt{1-\left(\frac{\sinh(b)}{R\omega C}\right)^2}. \end{aligned} \tag{5.9}$$

Substituting the solutions in equation (5.9) into equations (5.6) and (5.7) as appropriate yields

$$0 = \cosh(b) + \sqrt{1+\left(\frac{\sinh(b)}{R\omega C}\right)^2} - \frac{Z_0}{R}\sinh(b) \tag{5.10}$$

and

$$0 = \cos(a) + \sqrt{1+(R\omega C\sin(a))^2} - \omega C Z_0\sin(a). \tag{5.11}$$

Equation (5.11) is quite amenable to numerical solution as it will always have a single solution for a in the range $[0,\pi)$ for any non-negative values of R, ω, C, and Z_0. The value solved for a may by applied to equation (5.8) (which is also numerically well-behaved) to determine b. In this way the complex phase of reflection can be numerically calculated for any complex impedance $Z_{x=0}$ terminating the waveguide.

It is instructive to check these equations in the familiar limits of a purely resistive and a purely reactive termination. From equation (5.1) the ratio of the reflected (left traveling) to incoming (right traveling) wave is

$$I^{-}_{x=0}(t)/I^{+}_{x=0}(t) = e^{i\phi} = e^{i(a+ib)} = e^{-b} \cdot e^{ia}\,, \tag{5.12}$$

that is, the imaginary part of ϕ (b) determines the magnitude of the reflection and the real part of ϕ (a) determines the phase. We expect $b \in [0, \infty^{+})$ if the reflection is not larger than the incoming wave and bound a to $a \in [0, 2\pi)$.

For a purely resistive termination, such that ϕ must be purely imaginary, equation (5.6) equals 0, $\sin(a) = 0$, and $a = 0$ or π. For $a = 0$ equation (5.7) becomes

$$\frac{R}{Z_0} = \frac{\sinh(b)}{1 + \cosh(b)} = \tanh\left(\frac{b}{2}\right) \tag{5.13}$$

and when $a = \pi$ equation (5.7) becomes

$$\frac{R}{Z_0} = \frac{\sinh(b)}{-1 + \cosh(b)} = \coth\left(\frac{b}{2}\right). \tag{5.14}$$

Plotting the two solutions shows that the case $a = 0$ gives valid solutions for any $0 < R/Z_0 < 1$, the case $a = \pi$ gives valid solutions for any $R/Z_0 > 1$, and $\lim_{R/Z_0 \to 1} b = \infty^{+}$, which by equation (5.12) means a reflection of zero amplitude. That is: when $R < Z_0$ the reflection returns with zero phase shift, when $R > Z_0$ the reflection returns with a π phase shift, and there is no reflection when $R = Z_0$. (This is opposite to the usual "voltage-centric" truism that an open line reflects with zero phase shift and a short circuit reflects with π phase shift.)

The termination resistance in this case could be replaced by a transmission line with characteristic impedance R without altering any argument or calculation above. In that case the well-known reflection coefficient Γ for mis-matched transmission lines [104] can be written as

$$\begin{aligned} \Gamma &= \frac{Z_L - Z_S}{Z_L + Z_S} \\ &= \frac{R - Z_0}{R + Z_0} = \frac{R/Z_0 - 1}{R/Z_0 + 1} = \frac{\coth\left(\frac{b}{2}\right) - 1}{\coth\left(\frac{b}{2}\right) + 1} = e^{-b}\,, \end{aligned} \tag{5.15}$$

showing consistency between this treatment and the canonical results.

Next consider a purely capacitive termination, as in the case of an ideal capacitor. This has been treated in prior work on tunable MSA's, as in Mück et al. [96], giving the result

$$\begin{aligned} &\phi = \arctan\left[2Z_0\omega C/(Z_0^2\omega^2C^2 - 1)\right] + \pi, && C < 1/Z_0\omega \\ &\phi = \pi/2, && C = 1/Z_0\omega \\ &\phi = \arctan\left[2Z_0\omega C/(Z_0^2\omega^2C^2 - 1)\right], && C < 1/Z_0\omega \end{aligned} \tag{5.16}$$

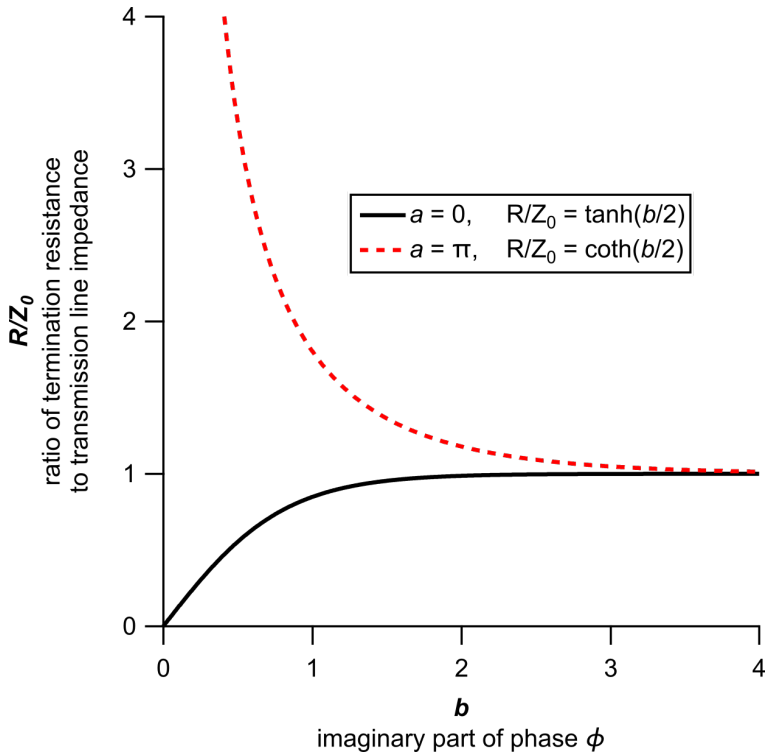

Figure 5.2: Exponent of reflection given a purely real termination impedance.

wheras equation (5.3) with $Z_{x=0} = 1/i\omega C$ becomes simply

$$\tan\left(\frac{\phi}{2}\right) = \frac{1}{\omega C Z_0}\,. \tag{5.17}$$

Equation (5.17) is equivalent to equation (5.16) but is somewhat more elegant and easier to use. The phase ϕ is purely real, so the reflected wave has the same amplitude as the incoming wave, and a phase that varies smoothly from 0 to π as $\frac{1}{\omega C Z_0}$ varies from 0 to ∞.

These results will be applied to a resonator, so it is convenient to derive an expression for the Q in terms of the reflection coefficient. In a resonator the Q-value may be defined as equation (5.18):

$$\begin{aligned} Q &\equiv \frac{\omega \cdot \text{stored energy}}{\text{rate of energy loss}} \\ &= \frac{2\pi f \cdot \text{stored energy}}{\text{rate of energy loss}}\,. \end{aligned} \tag{5.18}$$

To derive expressions for "stored energy" and "rate of energy loss" consider a steady-state interface between two half-infinite transmission lines coupled by a capacitance as shown in figure 5.3. There is a right-traveling wave incident on the interface (blue), a left-traveling wave reflected from the interface (orange), and a right-traveling wave transmitted through

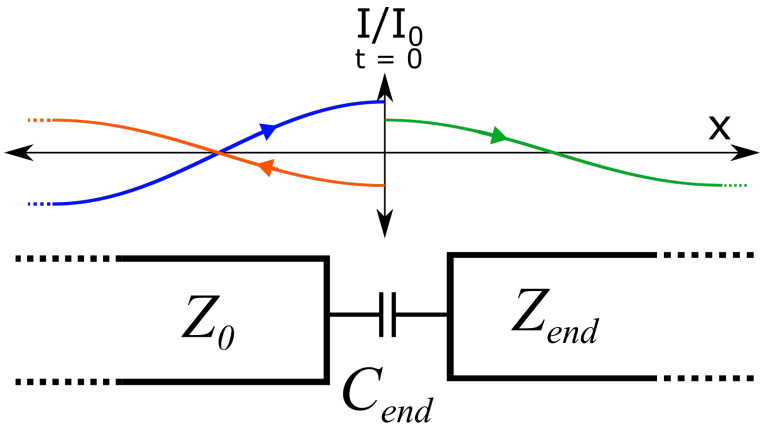

Figure 5.3: Two transmission lines coupled by a capacitor C_{end}.

the interface (green). The reflected wave will have an amplitude equal to or less than the incident wave given by equation (5.15), and the transmitted wave will have an amplitude set by conservation of energy. The equations for the three waves are given in equations (5.19).

$$\begin{aligned}
&\text{(blue)} && I^{+}_{incident}(x,t) &&= I_0 e^{i(\omega t - kx)}\,, && \\
&\text{(orange)} && I^{-}_{reflected}(x,t) &&= I_0 e^{i(\omega t + kx + \phi)}\,, && \\
&\text{and} && I^{-}_{reflected}(x,t) &&= I_0 e^{-b}\cdot e^{i(\omega t + kx + a)}\,, && (\text{with } \phi = a + ib) \\
&\text{(green)} && I^{+}_{transmitted}(x,t) &&= I_0\sqrt{1-e^{-2b}}\cdot e^{i(\omega t - kx + \theta)} && (\text{with } \theta \in \mathbb{R}).
\end{aligned} \tag{5.19}$$

From this we can construct Q from the rate of energy loss and the stored energy. The energy stored on the left-hand line is infinite for the half-infinite transmission line in this example, but will be finite for a resonator of finite length. Here we will parameterize to the number of wavelengths on the left line. The energy per wavelength on the left line is the sum of energies of the right and left traveling waves, as expressed in equation (5.20):

$$\text{stored energy on the left line} = \left(I_0^2 + I_0^2 e^{-2b}\right)\cdot(\#\text{ of wavelengths})\,. \tag{5.20}$$

The power transmitted to the right-hand line is the rate of energy loss and is equal to the energy contained in one wavelength times the frequency, as expressed in equation (5.21):

$$\text{rate of energy loss to right line} = f\cdot I_0^2\left(1-e^{-2b}\right). \tag{5.21}$$

By defining f_0 as the frequency corresponding to a $\lambda/2$ standing wave on the resonator, equations (5.18), (5.20), and (5.21) can be combined to give equation (5.22):

$$Q = \pi\frac{f}{f_0}\cdot\frac{I_0^2 + I_0^2 e^{-2b}}{I_0^2\left(1-e^{-2b}\right)} = \pi\frac{f}{f_0}\cdot\coth(b). \tag{5.22}$$

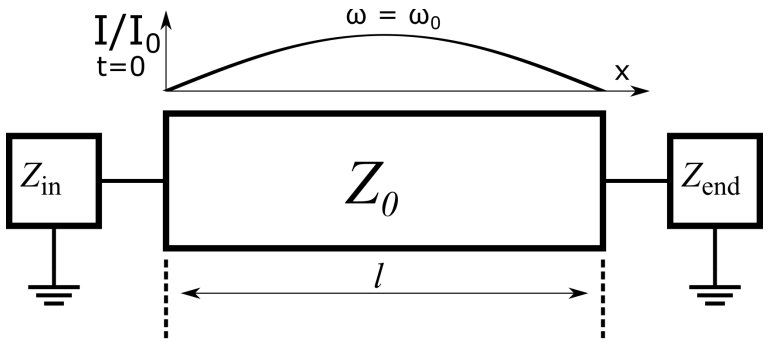

Figure 5.4: Transmission line with length l and arbitrary terminations. A model of a standing wave with $\omega = \omega_0$ at $t = 0$, consistent with $Z_{in} = Z_{end} = \infty$, is shown above the transmission line.

This expression for Q is due to loss at one end. The total Q will be given by the usual formula $Q_{total}^{-1} = \sum_i Q_i^{-1}$ for losses due to energy loss from two or more places.

5.1.2 Finite Transmission Line with Arbitrary Terminations

A transmission-line resonator is composed of a 1-D transmission line of length l with each side (the "input" side and "end" side) terminated by an impedance (Z_{in} or Z_{end}) that imposes a phase shift (ϕ_{in} or ϕ_{end}), respectively, on any wave impinging on it. In general, it is not possible for an arbitrary signal to propagate along the transmission line and also satisfy the boundary conditions set by Z_{in} and Z_{end}. However, there will be some discrete resonant frequencies that can satisfy the boundary conditions set by both Z_{in} and Z_{end}, and we will generally compare those frequencies ω to ω_0 with ω_0 defined as the frequency of the lowest possible frequency standing wave when $Z_{in} = Z_{end} = \infty$. A signal of frequency ω_0 will have a wavelength $\lambda_0 = 2l$, as shown in figure 5.4.

In the simple model of a finite transmission line with one end open and the other end terminated by an ideal capacitance, that is $Z_{in} = \infty$ and $Z_{end} = \frac{1}{i\omega C}$, $\phi_{in} = \pi$ and $\phi_{end} = 2\arctan\left(\frac{1}{i\omega C}\right)$ [equation (5.17)], the resonant frequency can be varied by a factor of two from $\omega = \omega_0$ to $\omega = \frac{\omega_0}{2}$ as ϕ_{end} varies from π to 0 and the wavelength varies from $\lambda = 2l$ to $\lambda = 4l$ as shown in figure 5.5.

This model served as inspiration for early MSAs, but there is utility in extending it. In practice, a signal must be coupled in through Z_{in} and the approximation that $Z_{in} = 0$ may not be valid. Furthermore, the assumption that the termination is purely reactive may not be valid in the case of lossy varactors, and it certainly not valid for Z_{in}, which must include the characteristic impedance of the input line, typically 50 Ω. For any pair of terminations on the transmission line, with phases ϕ_{in} and ϕ_{end} enforced at the boundaries, the resulting frequency ω is given by equation (5.23)

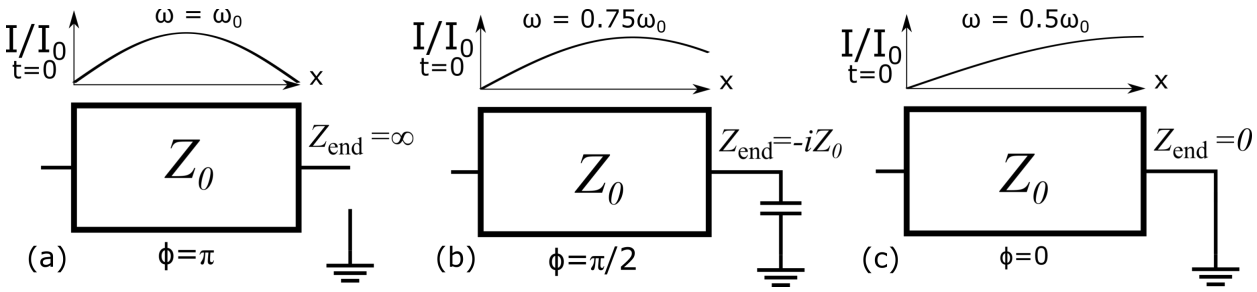

Figure 5.5: Tuning with input impedance $Z_{in} = \infty$ and an end impedance Z_{end} as ideal capacitor. (a) Tuning in the limit $C \to 0$. (b) Tuning at $C = 1/(Z_0\omega)$. (c) Tuning in the limit $C \to \infty$.

$$\omega = \omega_0 \frac{\phi_{in} + \phi_{end}}{2\pi} \tag{5.23}$$

which appears simple enough. However, ϕ_{in} and ϕ_{end} are complex functions of ω, R, and C by equations (5.11) and (5.8), and as a practical matter we may need to solve for any of ω, ω_0, C, R, Z_0, or some other parameter in terms of the others. To this end, in section 5.1.2.1 we describe a general numerical solver that can find a self-consistent solution to the resonant mode by varying any of the parameters $R_{in}, C_{in}, Z_0, \omega_0, \omega, R_{end}$, or C_{end}.

5.1.2.1 Numerical Solver for an Arbitrarily Terminated 1-D Resonator

We use a numerical solver of the generic "zero crossing" type. That is: vary a parameter of some multi-variable expression and observe a single output parameter. When the output parameter is sufficiently close to zero, the equation is for practical purposes "solved". Our question is whether a standing wave can be constructed from the parameters R_{in}, C_{in}, Z_0, ω_0, ω, R_{end}, and C_{end} on a 1-D resonator, and the output parameter is the "phase error" as described below.

The LabVIEW program "general phase mismatch.vi" takes as input R_{in}, C_{in}, Z_0, ω_0, ω, R_{end}, C_{end}, and a flag for the "variable parameter", and returns a "phase mismatch" parameter, as well as a graph and Q estimate. Though generally used only as a sub-vi, it includes graphical input and output to aid in understanding and debugging. (I hope no debugging is necessary at this point!) The reader is encouraged to download either the actual VI (requires LabVIEW) or the example program (Windows compatible, LabVIEW not required) to follow along. The programs are available on a github repository [105] and are also referenced by permanent DOI through Zenodo [105]. Figure 5.6 shows a sample result with the input parameters given in table 5.1.

The program "general phase mismatch.vi" calculates the complex ϕ_{in} and ϕ_{end} for the "input" (left side) and "end" (right side) of the transmission line from equations (5.11) and (5.8), and reports the numerical results on the left or right side respectively, of figure 5.6.

Table 5.1: Input values to "general phase mismatch.vi" used to generate Figure 5.6.

R_{in}	C_{in}	Z_0	f_0	f	R_{end}	C_{end}
50	0.1	120	1000	1050	0	3
(Ω)	(pF)	(Ω)	(MHz)	(MHz)	(Ω)	(pF)

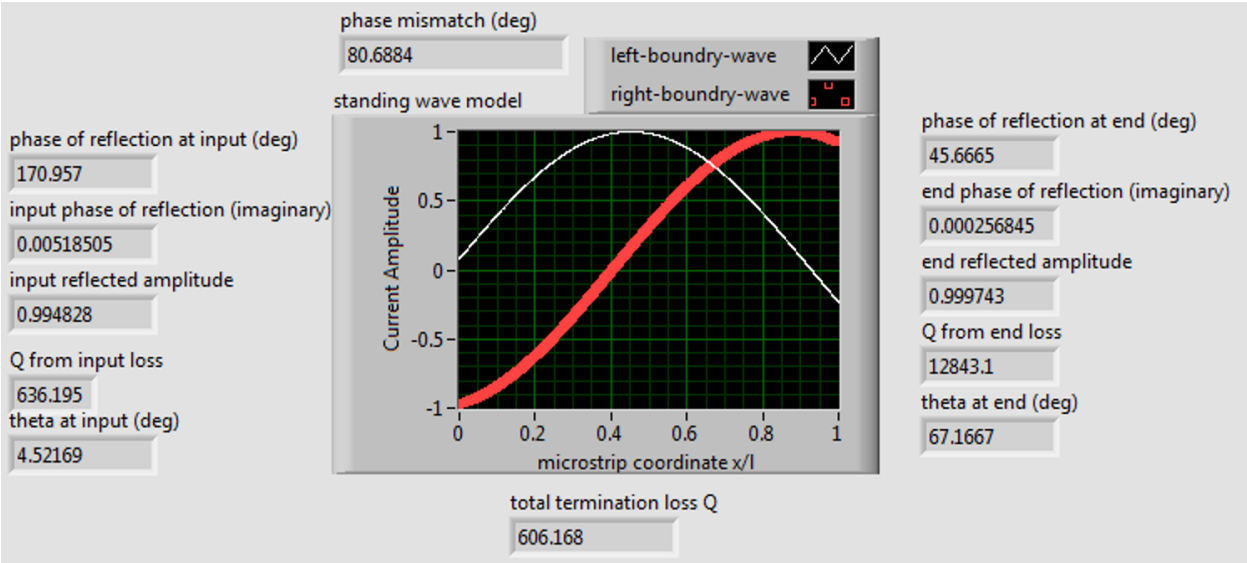

Figure 5.6: Sample output from the numerical solver "general phase mismatch.vi".

Based on those results, two trial standing waves of frequency f are plotted on the central graph labeled "standing wave model" in figure 5.6". The white curve represents a standing wave model consistent with the left (input) boundary conditions and the red curve represents a standing wave model consistent with the right (end) boundary conditions.

We can see that C_{in} is quite small at 0.1 pF which yields a left-side phase shift of 170.9°, and the white trace is indeed nearly at a current node at the left side. The right-side termination C_{end} is a moderate 3 pF which yields a right-side phase shift of 45.67°, and the red trace is closer to an anti-node on the right side. The imaginary parts of ϕ_{in} and ϕ_{end} are also accounted for, both listed explicitly and as a reflection amplitude. On the right boundary $R_{end} = 0\ \Omega$ so that the reflection should be lossless, but a reflection of 0.999743 is reported. This is due to an implementation of the solver chosen to guarantee robustness at the expense of a small numerical error. The left side reflection amplitude is 0.9948, or a slightly greater loss than a part in 1000 which is reasonably accurate. "Input" Q, "end" Q, and total Q are calculated as described in section 5.1.1

Both the red and white traces have a frequency of $f = 1050$ MHz, which is greater than $f_0 = 1000$ MHz, so that each presents a little more than a 1/2 wavelength on the resonator. Neither trace is a suitable model for a standing wave because they each satisfy only one boundary condition. The "amount of wrongness" is the phase difference between the two

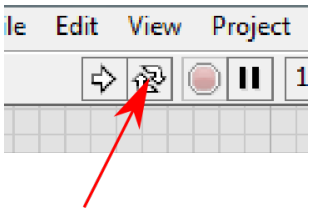

Figure 5.7: The"Run Continuously" button that enables live interaction with the numerical solver.

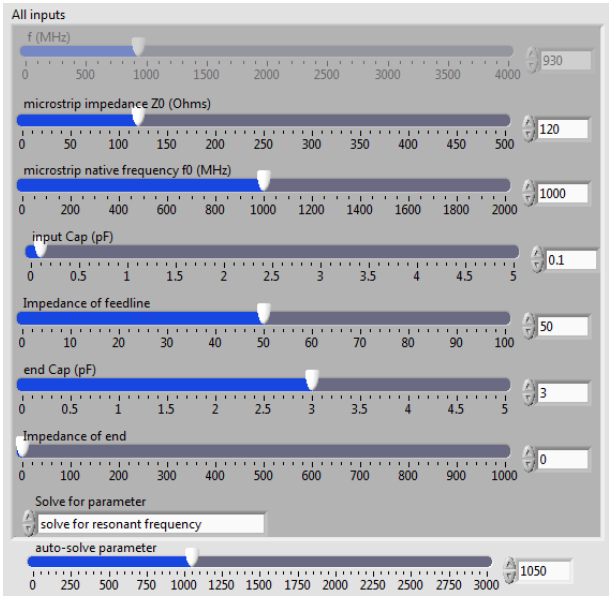

Figure 5.8: Sample input panel from the numerical solver "general phase mismatch.vi".

waves, reported as 80.69° in the box above the center graph. Indeed, the red and white traces appear almost like a *sin cos* pair.

We can informally find an "allowed" standing wave by adjusting the frequency of the simulated standing wave. If following with the software, note that in the input cluster shown in figure 5.8 the "frequency" slider is grayed out and disabled. This is because the "solve for parameter" flag (white box near bottom) is set to "solve for resonant frequency" so that the bottom slider labeled "auto-solve parameter" is the input for frequency. This is inconvenient as a human interface, but greatly simplifies the interface with other programs, which is the primary purpose of "general phase mismatch.vi". Click the "Run Continuously" button in the top right (figure 5.7) and try moving the bottom slider to change f until the two trial standing waves match and the phase mismatch is nearly zero. You can also type a number into the textbox to the right of the slider.

A value around $f = 669.66$ MHz will match up the two trial standing waves into a solution that satisfies both boundary conditions. Figure 5.9 shows a solution with a phase

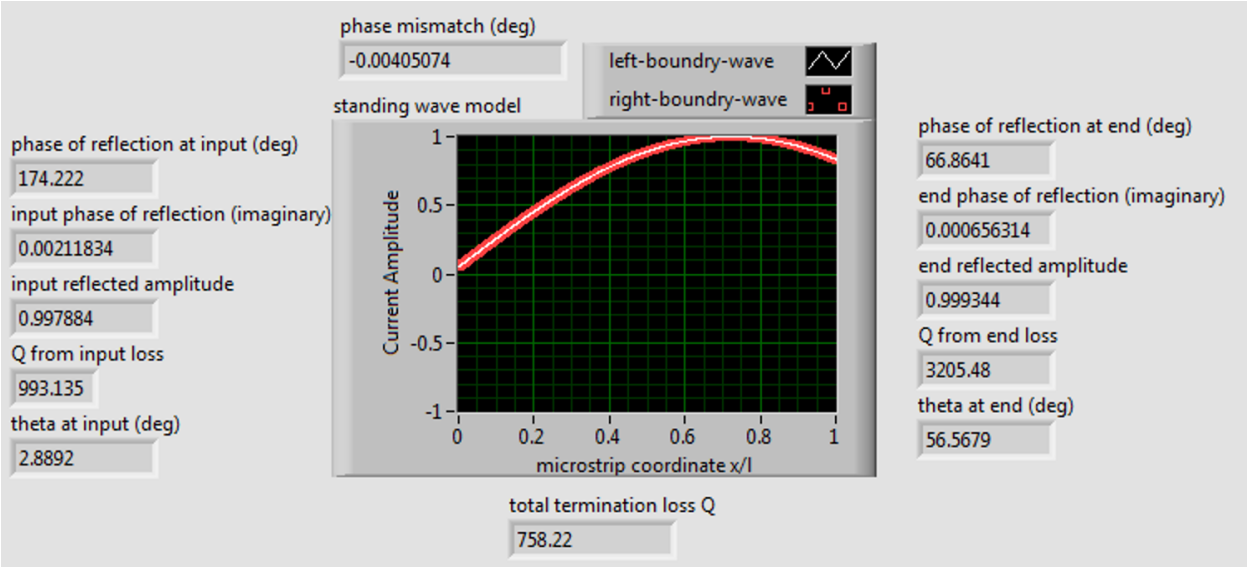

Figure 5.9: Sample output from "general phase mismatch.vi" with valid solution

Table 5.2: Input values to "general phase mismatch.vi" used to generate Figure 5.9

R_{in}	C_{in}	Z_0	f_0	f	R_{end}	C_{end}
50	0.1	120	1000	669.66	0	3
(Ω)	(pF)	(Ω)	(MHz)	(MHz)	(Ω)	(pF)

mismatch of just $-0.00405°$. Compare the results in figures 5.6 and 5.9: at the lower (fitted) frequency, the wavelength is of course longer, but the phases ϕ_{in} and ϕ_{end} have also changed– at a lower frequency the conductance of the capacitors is reduced, so that the reflected phases are closer to π (180°). The reduced capacitor conductivity also reduces transmission out of the resonator, so that Q is increased.

Note that there is more than one solution (actually infinitely many solutions) to the problem of a resonator with arbitrary terminations in the form of overtones. Figure 5.10 with table 5.1.2.1 show one such example. Note that the phase mismatch is 360.001°, so that this solution will not be arrived at by a zero-seeking algorithm. Higher-order solutions are not desired in this work, so this is exactly the behavior we want. However, if we did desire to solve for higher-order modes as in Mück, Schmidt, and Clarke [100] the algorithm could be trivially extended to search for a phase difference of 180°, 360°, 540°, or any other desired overtone.

The usual utility of "general phase mismatch.vi" is not to calculate the phase mismatch of a trial standing-wave solution, but to solve for one the the seven parameters R_{in}, C_{in}, Z_0, ω_0, ω, R_{end}, or C_{end} given the other six, that results in a valid standing-wave solution. For that purpose there is a wrapper called "match general parameter.vi" that uses "general phase mismatch.vi" as a zero-crossing oracle. The reader is encouraged to download either the

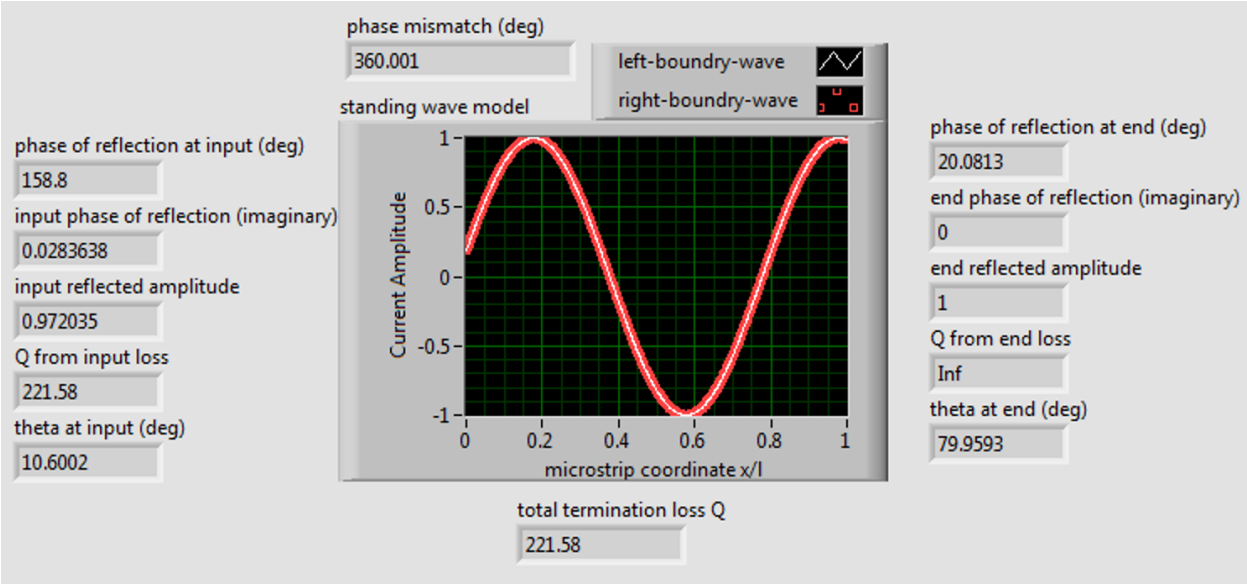

Figure 5.10: Sample output from "general phase mismatch.vi" with an overtone solution

Table 5.3: Input values to "general phase mismatch.vi" used to generate Figure 5.10

R_{in}	C_{in}	Z_0	f_0	f	R_{end}	C_{end}
50	0.1	120	1000	2496.9	0	3
(Ω)	(pF)	(Ω)	(MHz)	(MHz)	(Ω)	(pF)

actual VI (requires LabVIEW) or the example program (Windows compatible, LabVIEW not required) to get a feel for the program. Figure 5.11 shows a sample screen of "match general parameter.vi" in which the solution $f = 669.678$ MHz is given for the same parameters in the above examples. Also shown for debugging purposes is the "remaining phase error" (ie how close to zero did we get?) and "ticks" (ie how many iterations did it take?). This module is useful both for developing a quick intuition for the important parameters, such as sensitivity to C_{end} (good enough for our varactors) or sensitivity to the feedline impedance (not much), and more often as a subroutine in other programs to extract device parameters (ex: varactor capacitance) from experimental data when direct measurement is not possible.

5.2 Choice of DC SQUID Parameters

Design of an MSA starts with the choice of SQUID parameters: the critical current I_0, the junction shunt resistance R, the junction capacitance C, and the geometric loop inductance L. Equations 4.45, 4.46, 4.48, and 4.49 provide some useful relations guiding the choice of SQUID parameters, and for convenience they are reproduced here.

The ratio of thermal energy scale to the Josephson energy is known as the reduced

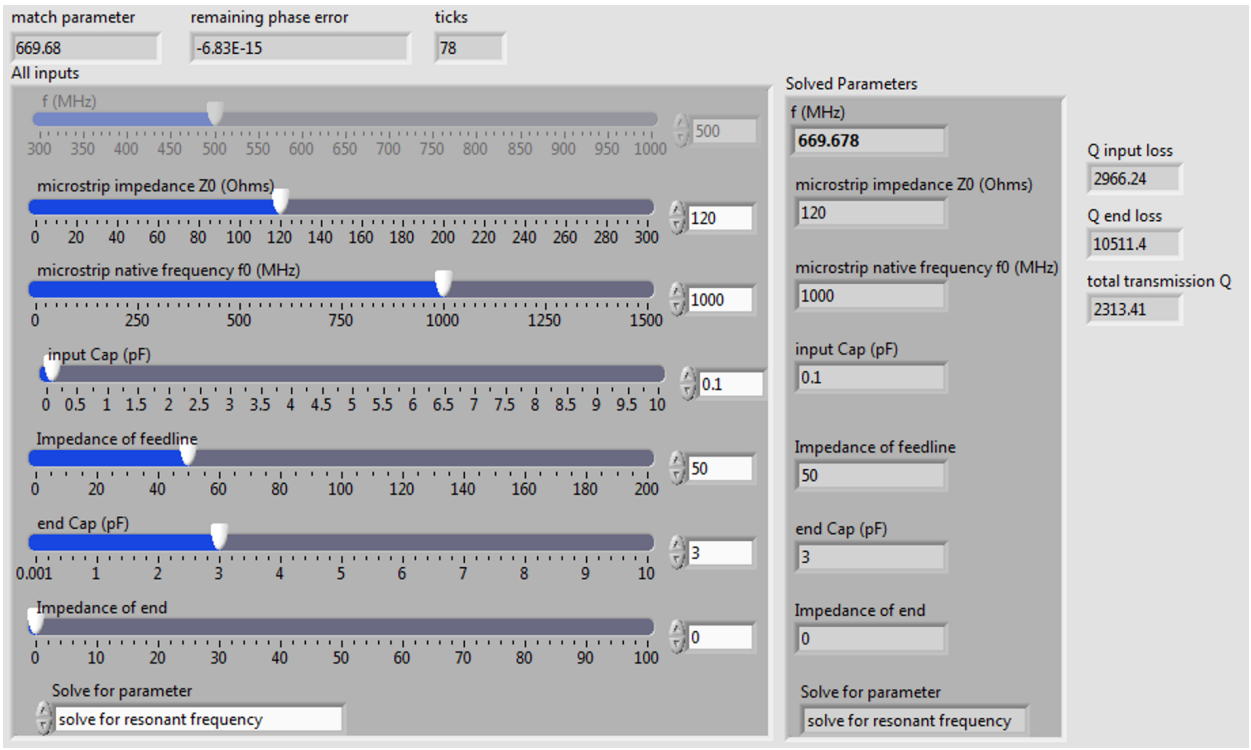

Figure 5.11: Sample output from "match general parameter.vi"

temperature, and is defined as

$$\Gamma \equiv \frac{k_B T}{E_J} = \frac{2\pi k_B T}{I_0 \Phi_0}, \tag{5.24}$$

the ratio of geometric inductance to Josephson inductance is defined as

$$\beta_L \equiv \frac{2 I_0 L}{\Phi_0}, \tag{5.25}$$

the Stewart-McCumber damping parameter is defined as

$$\beta_C = \frac{2\pi I_0 R^2 C}{\Phi_0}, \tag{5.26}$$

the ratio of thermal energy to inductive energy is

$$\Gamma \cdot \beta_L = \frac{4\pi k_B T L}{\Phi_0^2} \ll 1, \tag{5.27}$$

and the thermal flux noise is approximately

$$S_\Phi(f) \approx 16 k_B T L^2 / R \tag{5.28}$$

assuming certain optimizations.

Smaller Γ is always advantageous, suggesting we operate at the lowest possible physical temperature and use devices with the highest possible I_0. We achieve greater modulation depth for smaller β_L (recall figure 4.8) suggesting we make the inductance L as small as possible. However, a vanishingly small β_L yields only twice the modulation depth as $\beta_L = 1$, and flux is coupled into the SQUID via a mutual inductance between the input coil and SQUID inductance, so that a $\beta_L \approx 1$ gives optimal performance in practice. Reduced damping (increased β_C) allows for a greater voltage-transfer coefficient, but only to the point that operation remains linear. If $\beta_C > 1$, the I-V characteristic becomes hysteretic and the MSA cannot be used as a linear amplifier, thus we desire $\beta_C < 1$, but as close to unity as our process control allows.

Inspection of these relations suggests that for best performance we should operate at the lowest possible temperature, and construct a device with arbitrarily large I_0, small L (with $I_0 L = \Phi_0/2$), small C, and and R chosen to maintain $\beta_C \leq 1$.

The critical current I_0 can be increased by either increasing junction area A or reducing the barrier thickness d. Junction capacitance C is proportional to A, so if A is increased, R must be decreased by the same factor to maintain non-hysteretic operation, leading to increased flux noise by equation 5.28. A better approach to increasing I_0 is to reduce d. The critical current I_0 increases exponentially with $1/d$, while the capacitance is proportional to $1/d$, suggesting that smaller d is always better. In practice, a barrier that is too thin will succumb to slight fabrication variations, forming pinhole shorts or other inhomogeneities. The fabrication challenge is to make junctions with the greatest critical current density and smallest capacitance while maintaining high reliability. Our devices were fabricated by Gene Hilton at NIST Boulder, who determined a good compromise and reliable process for junctions with a capacitance of about 50 fF/(μm)^2, a critical current density of about 0.56 μA/(μm)^2, and an area of about 6.25(μm)^2, for a critical current of 3.5 μA and capacitance of 312 fF per junction. Given these I_0 and C, a shunt resistor of 11 Ω was chosen for $\beta_C = 0.4$. It may have been possible to chose a higher resistance for a β_C closer to unity, but the area and capacitance per junction are best guesses, rather than carefully measured, so that this choice of resistance reflects an abundance of caution considering that a second fabrication run cannot be ordered on a short timeframe. In any case our demonstrated performance was more than adequate.

With I_0, R, and C set, the inductance L should be about 300 pH to maintain $\beta_L \approx 1$. Rather than calculate an exact inductance from first principles or numerical simulations, we rely on prior semi-empirical results for thin-film SQUIDS with square hole geometry[106], giving the practical formula

$$L = 1.25 \times \mu_0 d + \ell \times 0.3 \text{ nH}/\mu m, \tag{5.29}$$

where d is the edge-length of the square hole, μ_0 is the permeability of free space, and ℓ is the slot length from the hole to the outer edge of the SQUID where the junctions attach.

The MSA operating frequency is set by the SQUID inductance and input coil design, and we designed MSAs with a wide variety of input coils and two SQUID inductances. The first

Table 5.4: Circulator frequency ranges in ADMX

Circulator	Frequency Range
1	560 - 640 MHz
2	644 - 714 MHz
3	715 - 850 MHz
4	850 - 1000 MHz

higher-inductance, lower frequency MSA design (the "LF" family) had a square hole 200 μm across and a 390 μm slit for $L \approx 430$ pH and $\beta_L \approx 1.5$. The second low-inductance, higher frequency MSA design (the "HF" family) was "slit only" with no enlarged center hole and a 240 μm long, 10 μm wide slit from center to edge for $L \approx 80$ pH and $\beta_L \approx 0.4$.

5.3 Design of the MSA Input Coil for Tunability

The MSA input coil design sets the baseline frequency of operation, although this may be altered by later loading to the input line or to microwave ground with variable varactors. Our frequency design targets are set by the tunable range of the ADMX main cavity, about 560 to 925 MHz. The ADMX detector chain relies on several cryogenic circulators which operate over the relatively narrow range of frequencies listed in table 5.4. These circulators cannot be changed during operation and but only during a total system warm-up, which is a considerable disruption to operation and performed as infrequently as possible. As such, any MSA delivered to ADMX need tune over only a frequency range matching the current circulator range, and the MSA can be replaced along with the circulators when a new frequency range is desired. This is fortunate because we found that in many instances the capacitance range of our varactors was insufficient to tune over a full octave.

An example of an input coil design is shown in figure 5.12. The resonant frequency of a particular design is estimated from the inductance per unit length L_ℓ, capacitance per unit length C_ℓ, and total length of the coil. The inductance of the input coil has three terms: the kinetic inductance, the geometric inductance, and the SQUID mutual inductance. The SQUID mutual inductance is $L_M = L_{SQ} \times N^2$ where L_{SQ} is the SQUID inductance and N is the number of turns on the input coil. The Meissner effect affords a very strong magnetic coupling of the the input coil to the SQUID washer so that L_M is the dominant term and the geometric and kinetic inductance terms are neglected in our calculations. The capacitance of the coil is estimated by a simple parallel plate approximation considering the area of the coil over the washer, the thickness of the oxide between the washer and coil, and the dielectric constant of SiO_x with thickness equal to 350 nm and $\epsilon = 3.9$. The length of the coil is taken to be the length of those portions over the washer. The length of the input coil could arguably include the "off-washer" length to the bonding pads and wire-bond lengths to the external circuit elements, but due to not being flux coupled to the SQUID loop and not running close over a ground plane in those regions, both L_ℓ and C_ℓ will be much smaller

Table 5.5: Summary design parameters of the 30 MSA designs in the first fabrication.

Base designation	L (pH)	coil pitch (μm)	Washer reference	Input coil # of turns
LFN	430	20	RF ground	7 to 14 (8 designs)
LFF	430	20	RF output	7 to 14 (8 designs)
HFN	80	6	RF ground	14 to 20 (7 designs)
HFF	80	6	RF output	14 to 20 (7 designs)

in those regions. As a result the signal propagation speed

$$\nu = 1/\sqrt{L_\ell C_\ell} \tag{5.30}$$

is much greater, and the electrical length of those regions is relatively small. The expected frequency of a λ/n standing wave is then

$$f_{\lambda/2} = \nu/(n\ell), \tag{5.31}$$

where n is expected to take values from 2 to 4 as the MSA is tuned, and the characteristic impedance of the input coil is

$$Z_0 = \sqrt{L_\ell/C_\ell}. \tag{5.32}$$

Recall from equation 4.52 that tuning effects from a variable termination will be greatest for a transmission line of high Z_0. This motivates making the linewidth of the input coil as narrow as possible to minimize C_ℓ. One could also increase the oxide thickness. Based on the advice of our fabricator Gene Hilton that a linewidth of 1 μm would start to invite problems with reliability, we settled on a linewidth of 2 μm and a standard process SiO_x thickness of 350 nm. We desired a line spacing such that direct turn-to-turn interactions would be minimized. To this end the line pitch was set to 20 μm (ten times the linewidth) in the "LF" devices. In the "HF" devices we had less room (a shorter slit) and needed more turns so we used a tighter pitch of just 6 μm, for only a 4 μm gap between 2 μm wide lines. We were willing to risk a tighter spacing on the "HF" devices because they were intended as a speculative experiment, while the "LF" devices were intended as a critical component in ADMX. In the end, both designs worked admirably. Figure 5.12 shows an example of an "LF" device and input coil, and figure 5.13 shows an example of an "HF" design and input coil.

We had limited access to fabrication runs, but could request about 300 MSAs in a single run. To make best use of this resource, we designed four SQUID base templates and 15 input coil designs, for 30 unique MSA designs. Table 5.5 lists the major design parameters differentiating the 30 unique designs. Our naming convention yielded device names such as "LFF-12H" where the first two letters are either "LF" or "HF" indicating either a 430 or 75 pH SQUID. The third letter is either "N" or "F" ("normal" or "flipped") to indicate a grounded or active washer. After the dash, the number indicates the number of turns on the

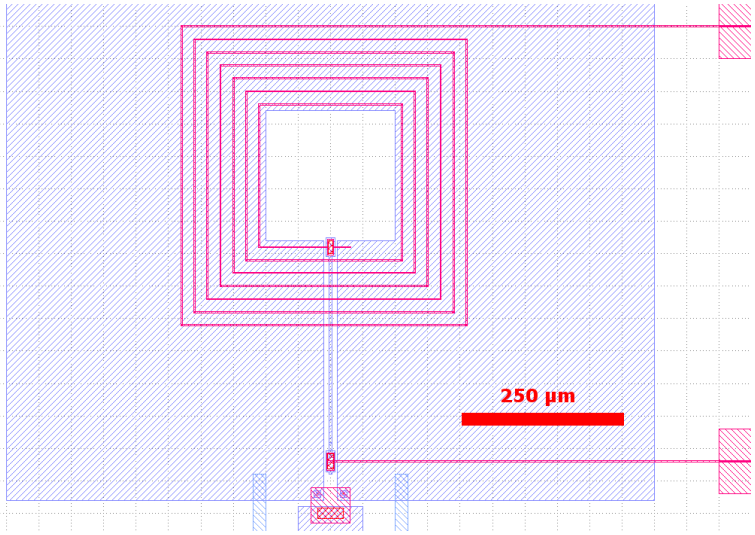

Figure 5.12: Detail of a design drawing for an "LF" MSA with a 7-turn input coil. Blue areas are the SQUID washer, and red areas are the input coil. There is a 350-nm thick insulating layer of SiO_x between the washer and input coil. The Josephson junctions are at center bottom.

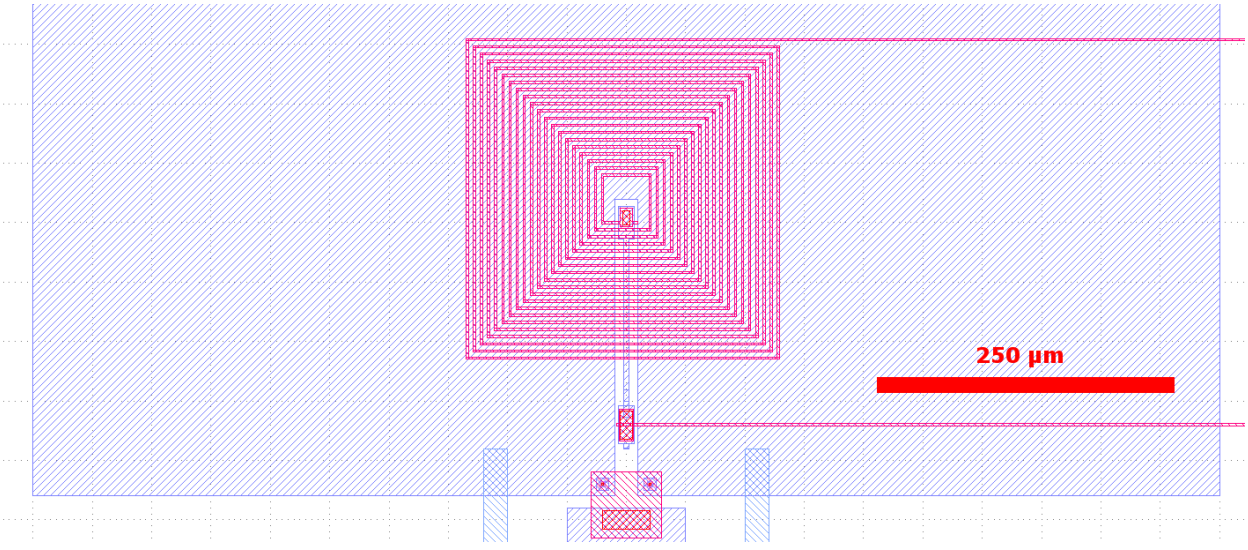

Figure 5.13: Detail of a design drawing for an "HF" MSA with a 20-turn input coil. Blue areas are the SQUID washer, and red areas are the input coil. There is a 350-nm thick insulating layer of SiO_x between the washer and input coil. The Josephson junctions are at center bottom.

Table 5.6: Detailed parameters of select MSA designs from the first fabrication run.

Name	LF-11	LF-12	HF-20
No. of turns	11	12	20
Coil pitch (μm)	20	20	6
Hole diameter (μm)	200	200	0
Slit length (μm)	390	390	240
Inductance (pH)	431	431	75
Coil inductance (nH)	52.2	62.1	30
Coil capacitance (pF)	3.93	4.46	2.1
$\lambda/2$ frequency (MHz)	1104	949	1994
Z_0 (Ohms)	115	118	119
Wave velocity %(v/c)	14.3	14.0	13.4

input coil, and the final letter is an arbitrary serial letter (A, B, C, D, etc.) differentiating nominally identical devices. Figure 5.14 shows design schematics for all four SQUID designs including the bonding pads used for attachment to external circuitry. Figure 5.15 shows the nominal $\lambda/2$ and $\lambda/4$ frequencies for every coil design for both "LF" and "HF" designs, with the four circulator frequency bands indicated by dashed lines for reference.

Table 5.6 shows detailed design parameters to select devices from the first fabrication run. Electrical and RF values are the nominal design values derived from the design geometery as described in the text and equations 5.30-5.32.

The grounded washer and grounded counter-electrode designs assume the same frequency, but without much justification. While the design guidelines described here are reasonable for a grounded washer, there is no reasonable expectation that models will be valid when the "groundplane" is active. Depending on the phase of the amplified signal applied to the washer, the effective capacitance could either be dramatically increased (decreasing both the frequency and characteristic impedance) or decreased (increasing the frequency and characteristic impedance). Recalling figure 4.9 our choice of the DC applied flux Φ_a during operation always allows us to choose the sign of $\partial V/\partial\Phi$, allowing a choice of two phases differing by 180° for the signal applied to the washer. Rather than model this in detail before deciding on an MSA design, we settled on the "shotgun" design approach, allowing ourselves the luxury of a wide range of coil designs and nominal design frequencies, with real-world performance determined by real-world testing. The practical considerations of the active washer operation are described in chapter 7. We found that although there are important differences between grounded-washer and active-washer operation, our initial best guess designs (the LFx-11 and LFx-12 series) proved the most useful to ADMX.

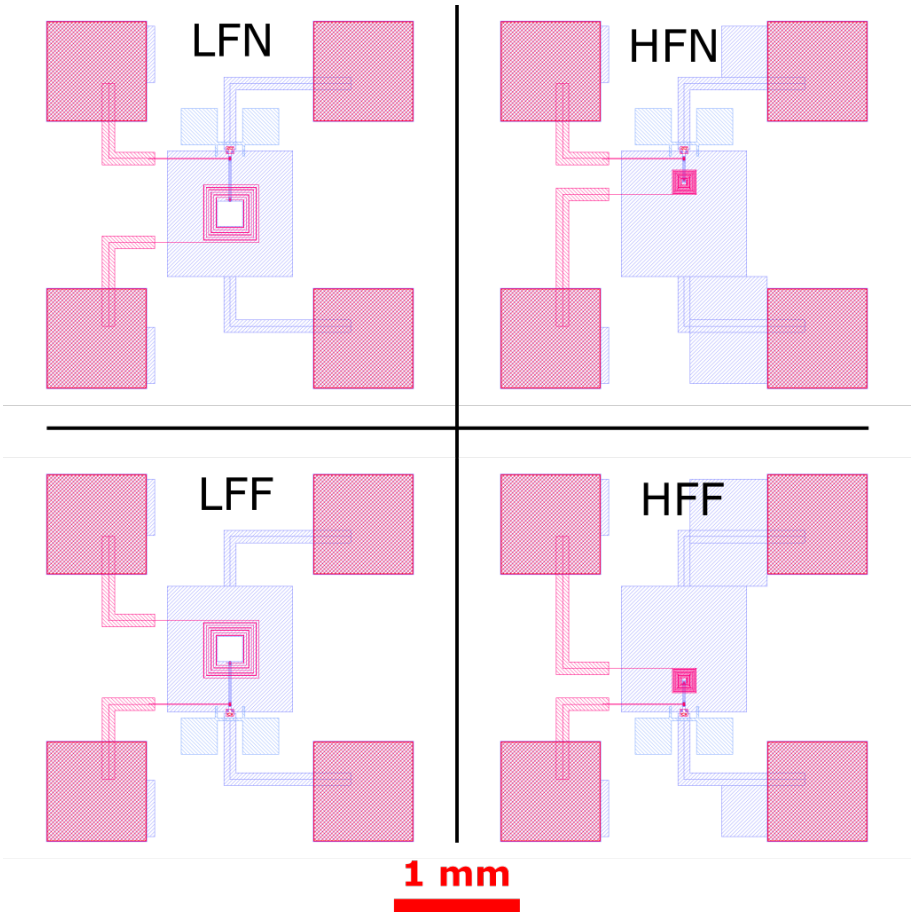

Figure 5.14: Four MSA types designed for the first fabrication run. Left hand column is the "low-frequency" devices (L = 431 pH) and right hand column is the "high-frequency" devices (L = 75 pH). Top row is the grounded-washer designs, and bottom row is the active washer (grounded counter-electrode) devices. Each design assumes the following bonding pad conventions: the top left pad connects to the 50-Ω line carrying the signal to be amplified, the bottom-left pad connects to RF ground through a varactor providing tuning, the bottom right pad connects to RF ground, and top right pad connects to RF output.

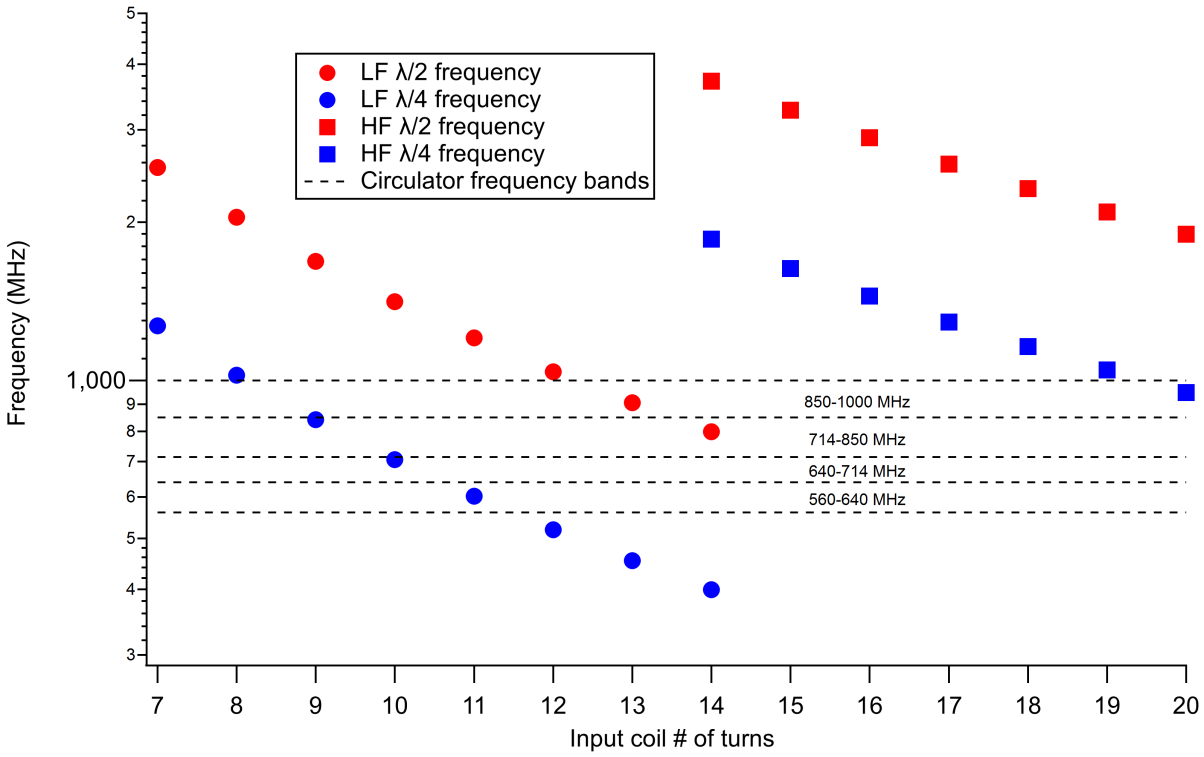

Figure 5.15: Nominal design frequencies of MSA designs for the first fabrication run. The LF series was designed to fill the immediate needs of ADMX, and the HF series was designed to explore higher frequency operation not previously attempted with an MSA. Consistent with design expectations, all devices delivered to ADMX were of the 11 or 12-turn LF designs.

5.4 Extended MSA Designs, Second Fabrication

All MSAs used in ADMX to date are from the "LF" line of devices, but we also designed and tested other designs made in a later fabrication batch, and many of the results reported here are from this second batch. After initial success with the first batch of MSAs, we duplicated the most useful designs from the first batch, made practical new designs focusing on the higher frequency range around 1330 MHz to 665 MHz anticipated to be of interest in future operation of ADMX, and a set of more extreme designs just to see how far we could push the technology, including designed frequencies up to 3 GHz and down to 230 MHz. The designs aimed at 1330 to 665 MHz we call the "range 3" designs, and the designs aimed at much higher frequencies (up to 3 GHz) or much lower frequencies (down to 230 MHz) we call the "reach" designs. All devices in the second fabrication had the same junction and resistive shunt design as the first batch.

We designed the second batch of devices with greater confidence in our ability to hit a target frequency, so each SQUID design was made with only three input coil designs–our

Table 5.7: MSA design parameters for "range 3" devices of the second fabrication batch.

Name	82-29	82-30	82-31	168-24	168-25	168-26	247-19	247-20	247-21
No. of turns	29	30	31	24	25	26	19	20	21
Hole diameter (μm)	0	0	0	0	0	0	68	68	68
Slit length (μm)	275	275	275	560	560	560	466	466	466
SQUID L (pH)	82.5	82.5	82.5	168	168	168	247	247	247
Coil L (nH)	69.4	74.3	79.3	96.8	105	114	89.2	98.8	109
Coil C (pF)	1.77	1.89	2.01	1.24	1.34	1.44	1.31	1.42	1.53
$\lambda/2$ freq. (MHz)	1426	1335	1252	1445	1335	1237	1462	1336	1226
Z_0 (Ohms)	198.0	198.3	198.6	279.7	280.4	281.0	260.4	263.7	266.8
Wave velocity %(v/c)	16.55	16.56	16.56	11.56	11.57	11.58	12.45	12.33	12.22

best guess design and an input coil with one more and one less turn. Table 5.7 shows the design parameters of the "range 3" devices. Each device in table 5.7 has a square input coil with a pitch of 5 μm, an input coil linewidth of 1 μm, and SQUID slit width of 20 μm. Note this smaller linewidth allows a smaller capacitance per unit length and higher operating frequencies. We were a little worried about the integrity of the input coil at this linewidth–it results in a structure 20,000 times longer than its width which must remain unbroken, but we saw no evidence of fabrication errors in operation. The device name (for example 168-25) is derived from the nominal SQUID inductance in pH (168 pH) and the number of turns on the input coil (25). For the "range 3" devices, we made three SQUID designs, and three coil designs per SQUID design. All "range 3" devices were aimed at the same frequency range, so that the lower inductance devices have more turns, and the higher inductance devices have fewer turns. The first three rows of table 5.7 show the physical layout parameters, the next three rows show the electrical parameters derived from the physical dimensions, and the last three rows show the nominal RF parameters derived from the electrical parameters. Each of these nine designs was made in both a grounded washer and grounded counter-electrode design, designated with an "N" (normal, grounded washer) or "F" (flipped, grounded counter-electrode) as in "168N-25" or "168F-25".

Table 5.8 shows the design parameters of the high frequency "reach" devices. Each device in table 5.8 has a square input coil, an input coil linewidth of 1 μm, SQUID slit width of 20 μm, and no central hole.

Table 5.9 shows the design parameters of the low frequency "reach" devices. Each device in table 5.9 has a square input coil, an input coil pitch of 5 μm, an input coil linewidth of 1 μm, and SQUID slit width of 20 μm.

The second fabrication run was composed of two wafers: the first used a new trilayer process, and the second used the same trilayer process as the first fabrication. From this we have the naming convention for devices from the second fabrication: an example would be "MSA2W2-168F-25B". The first part "MSA2W2" means this is an MSA from the second

Table 5.8: MSA design parameters for high frequency "reach" devices of the second fabrication batch.

Name	45-24	45-25	45-26	65-24	65-25	65-26	122-24	122-25	122-26
Coil pitch (μm)	4	4	4	5	5	5	5	5	5
No. of turns	24	25	26	24	25	26	24	25	26
Slit length (μm)	150	150	150	216	216	216	406	406	406
SQUID L (pH)	45	45	45	64.8	64.8	64.8	122	122	122
Coil L (nH)	25.9	28.1	30.4	37.3	40.5	43.8	70.3	76.3	82.5
Coil C (pF)	1.00	1.08	1.16	1.02	1.10	1.18	1.25	1.35	1.45
$\lambda/2$ freq. (MHz)	3066	2836	2631	2567	2374	2202	1686	1559	1445
Z_0 (Ohms)	159	160	160	192	192	193	237	237	238
Wave velocity %(v/c)	20.26	21.18	21.11	17.63	17.57	17.51	14.3	14.2	14.2

Table 5.9: MSA design parameters for low frequency "reach" devices of the second fabrication batch.

Name	289-24	289-25	289-26	454-24	454-25	454-26
No. of turns	24	25	26	24	25	26
Hole diameter (μm)	98	98	98	214	214	214
Slit length (μm)	451	451	451	393	393	393
Inductance (pH)	289	289	289	454	454	454
Coil inductance (nH)	70.3	76.3	82.5	262	284	307
Coil capacitance (pF)	2.18	2.32	2.46	3.28	3.46	3.65
$\lambda/2$ frequency (MHz)	829	772	720	540	504	472
Z_0 (Ohms)	276	279	282	282	286	290
Wave velocity %(v/c)	12.2	12.1	12.0	12.0	11.8	11.6

fabrication run, on the second wafer. The next part means this is a 168 pH SQUID with a grounded counter-electrode (active washer). The last part means this device has a 25-turn input coil and is the second such nominally identical device to be tested.

5.5 Design of the MSA RF Board, Package, and Support Components

The MSA design described to this point does not extend beyond thin metal films deposited on a 3.5-mm diced chip, but ADMX requires a "plug and play" amplifier with standard RF connectors and robust mechanical and electrical design. Delivery of a practical MSA requires the design of and integration to a package that makes use and operation by the

system integrators at ADMX as easy and smooth as possible.

The first part of the "full integrated" design is the design of the RF carrier board. The carrier board serves as a carrier for the MSA chip as well as the tuning varactors, RF connectors, DC bias tees and lo-pass filters. Figure 5.16 shows the front side of the carrier board, figure 5.17 shows the back side, and figure 5.18 is a schematic of the carrier board and MSA. The RF carrier is made from Rogers Corp. TMM6, and our friends at the Quantum Nanoelectronics Lab (QNL) were kind enough to manufacure our design in small batches. The RF signal pathway is a 50-Ω conductor-backed coplanar waveguide. Front-to-back ground vias along the signal pathway help to suppress spurious RF modes. Front-to-back ground vias are also placed near every DC ground reference to ensure a low-impedance path to ground. To improve both thermal characteristics and physical robustness, a 1-mm thick Cu plate is clamped and soldered to the back ground plane. The back ground plane is continuous except for 11 small holes made for DC control line clearance and one large hole directly behind the MSA. A ground plane present directly behind the MSA would introduce Nyquist noise via magnetic coupling of thermal eddy currents.

The DC paths incorporate series resistors with a value of 4.99 kΩ, visible as blue rectangles in figure 5.16. The resistors are made by Vishay Dale with part number TNPU12064K99BZEN00. These resistors have a room-temperature resistance tolerance of 0.1%, and although their performance is rated only to -55 degC (218 K), we found the measured resistance increased by only 0.08% on cooling from room temperature to 4 K (-269 degC) and the resistance did not vary by more than ±0.1% down to the lowest tested operating temperatures of ≈ 40 mK. These resistors are much larger (in physical size) and power handling capability than necessary for this application and could be reasonably substituted with physically smaller resistors, though we repeat our cautionary note that any component should be independently tested and qualified at planned operating temperatures before integration into the cryogenic experiment.

The DC filtering capacitors are 820 pF NPO ultra-low equivalent-series-resiststance (ESR) series 800C by American Technical Ceramics (ATC), part number ATC800B821KT, visible as tan-colored components about 2 mm square in figure 5.16. They are selected for high-frequency (low inductance) operation and excellent temperature stability. Like the resistors, they are specified for operation only to -55 degC, but we measured no deviation in capacitance between room temperature and 40 mK to within the limits of our measurement accuracy (about 1%). Size limits are a real concern for these capacitors– they occupy a large fraction of the total available space. It is tempting to use another ceramic type (for example X7R, popular in filtering applications) for higher capacitance/smaller size, but we note that X7R and other high-dielectric constant ceramic capacitors exhibit a dramatic reduction in capacitance at cryogenic temperatures as well as piezoelectric properties that could couple unwanted signals from mechanical vibrations. We have not discovered a more favorable capacitor type than this proprietary dielectric offered by ATC. The RC values used here give a knee frequency of about 40 kHz. The need to filter to considerably lower frequencies (below 60 Hz, for example) motivates the use of an external low-pass filter as described in section 6.1. A schematic representation of all the DC and RF components is shown in figure 5.18.

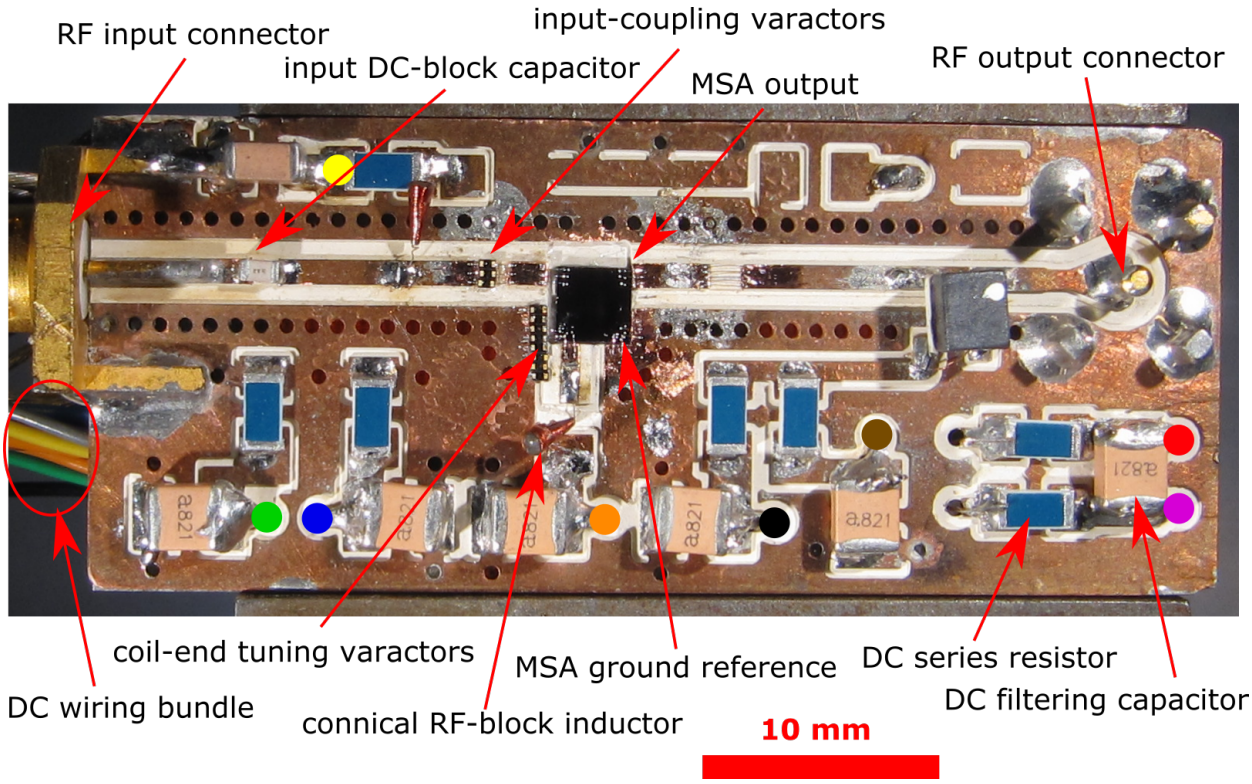

Figure 5.16: MSA carrier board front. Colored dots indicate vias where DC wiring connections are made from the back, and the color corresponds to the label colors for DC connections in schematic figure 5.18. One conical inductor is indicated, but three are present: one connects to the input waveguide, one (indicated) connects to the end of the MSA input coil, and one (inside black box with white dot) connects to the output waveguide. The DC filtering capacitors (tan squares) are American Technical Ceramics (ATC) ATC800B821KT, the DC resistors (blue rectangles) are Vishay-Dale TNPU12064K99BZEN00, the RF capacitors (white rectangles) are ATC series 600S, the varactors (small black rectangles with two gold pads) are M/A COM MA46H120, and inductors are Coilcaft BCR-652JL.

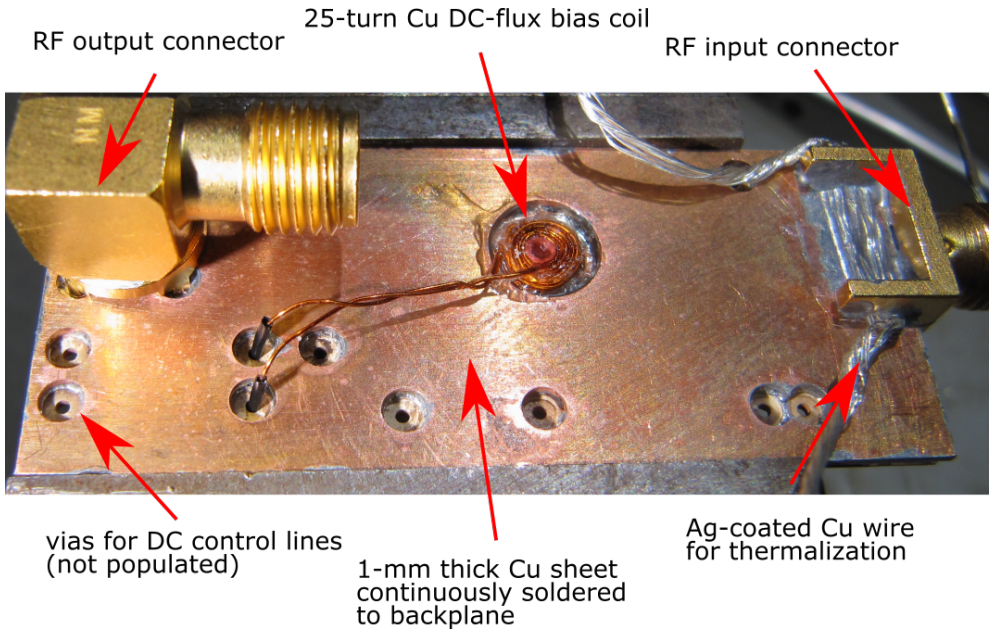

Figure 5.17: MSA carrier board back.

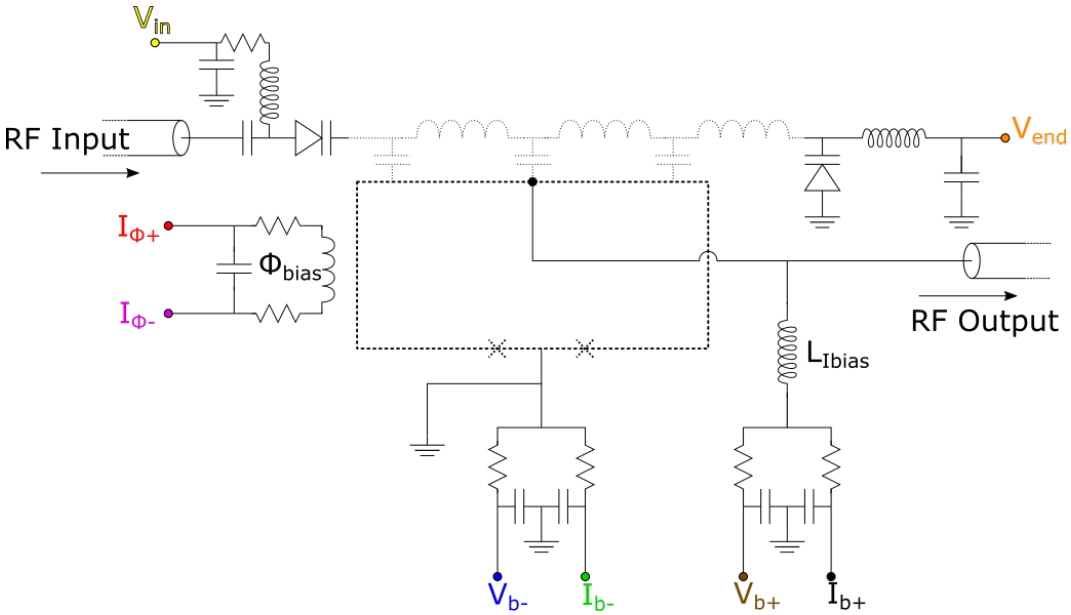

Figure 5.18: Schematic of MSA and immediate support electronics. Dashed sections indicate the MSA, and solid sections represent off-chip components. The color of the DC connection labels corresponds to the color of the solid circles in figure 5.16.

The RF capacitors used in the signal path are ATC 600S series capacitors of various values, depending on the particular MSA. An example part number for a 2.2 pF capacitor is ATC600S2R2BT. An example is visible in figure 5.16 on the input trace. Similar to the ATC 800C series capacitors, they are selected for very low ESR, excellent high-frequency operation (low inductance) and temperature stability well beyond the rated temperature range, down to 40 mK.

The varactors used both for coupling to the input signal and coupling the end of the MSA input coil to ground are made by M/A COM, part number MA46H120. The varactors are visible in figure 5.16 as black rectangles less than 1 mm long with two gold pads each. Figure 5.16 shows three varactors arranged in parallel as an element of the signal input path, and eight arranged near the coil-end bonding pad of the MSA, but only two of the eight electrically connected to the circuit. Figure 5.19 shows a detail view of and tuning varactors and MSA with lighting optimized to show the MSA features . In contrast to figure 5.16, figure 5.19 uses a fixed capacitor (an ATC 600S series) for the input coupling, has four varactors physically installed, and has only one varactor electrically connected. The varactors were selected for low inductance and cryogenic capability. Many available varactors (particularly those with an enticingly high capacitance) have an unacceptably high series inductance, perhaps attributable to the packaging. This model is a "flip-chip" package, intended for face-down surface-mount wave-soldering, which is the most minimal packaging and lowest inductance arrangement we could find. Rather than making electrical connections by soldering, we affixed the varactors to the carrier "face-up" with low-viscosity cyanoacrylate glue and made electrical connections with Al wirebonds to the Nb MSA bonding pads and Cu ground plane as needed. Care must be taken to use gentle wirebonding settings (low pressure, low power) to avoid damaging the varactors. Fortunately, the gold bonding pads are nearly ideal wirbonding surfaces so that low-power settings are sufficient for reliable bonding. Another critical property of these varactors is the over-doped GaAs formulation, which retains its semiconductor properties at low temperatures. This is in contrast to Si varactors, which have thermally promoted charge carries which freeze out at low temperatures. Unfortunately, the properties of these varactors are not constant at our operating temperatures–though they are largely unchanged from 300 K to 4 K, they exhibit a surprisingly large reduction in capacitance as the temperature falls below 2 K, with no further reductions observed below 200 mK. This low-temperature capacitance loss is an interesting phenomena unto itself and is covered in greater detail in section 7.1 and appendix A.

The inductors are Coilcraft BCR-652JL, and have a nominal inductance of 6.5 μH. The conical design is critical; a toroidal or solenoidal geometry would have a consistent turn-to-turn capacitance and mutual inductance resulting in a sharp resonant frequency, but a conical design has a distribution of turn-to-turn capacitances and mutual inductances that "spreads out" any possible resonance resulting in a flat frequency response up to 40 GHz. At our nominal operating frequency of 1GHz, the inductors have an AC resistance of 40 kΩ, so they reflect almost all of the signal present on the 50 Ω transmission lines or MSA input coil.

The DC bias coil is shown in figure 5.17 and is constructed of 25 turns of Cu wire five

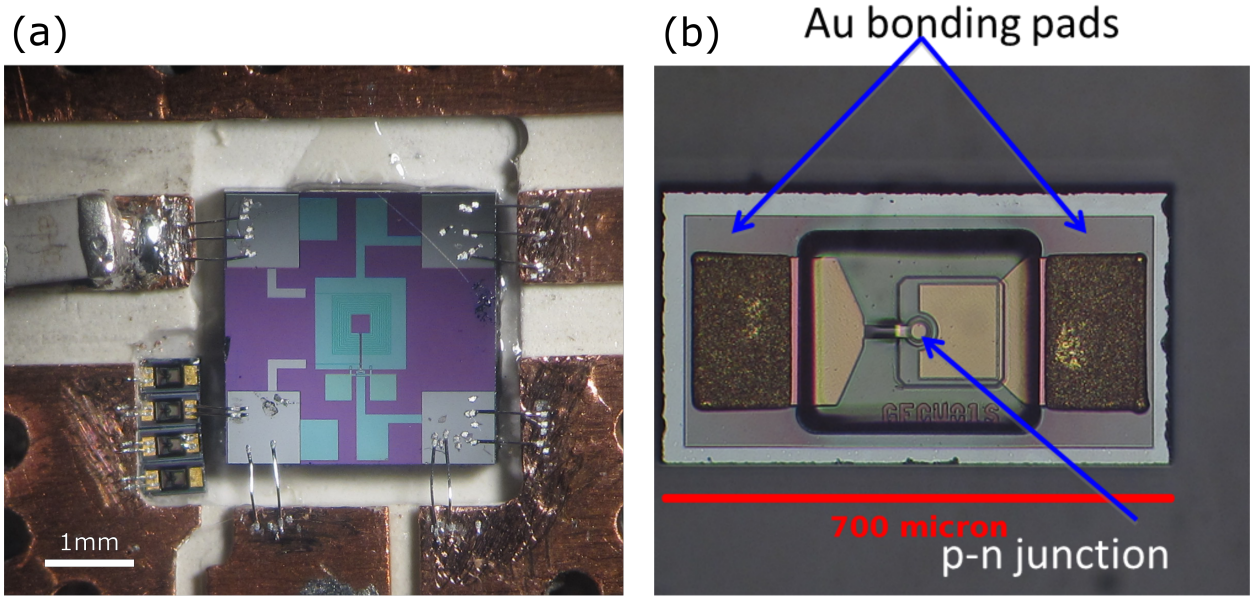

Figure 5.19: MSA chip and varactor. (a) An MSA chip bonded into the supporting circuit board, with all four bonding pads visible. The top left pad bonds to the signal input line, in this case coupled through a capacitor. Bottom left pad connects to varactors and the DC voltage line used to bias the varactors. Bottom right pad connects to the RF groundplane. The top right pad connects to the signal output, in this case a 50 Ω coplanar waveguide. The output waveguide also provides the DC bias current. (b) Detail of a tuning varactor.

turns thick and 5 turns deep infused with cyanoacrilate glue, and mounted to the carrier with rubber cement. The ID is about 1 mm, and the OD is about 3 mm (about the same size as the MSA chip), and the thickness is about 1 mm. Note the extra insulation applied where the wires mate with the board vias–it is a wise precaution to protect fine wires any place they touch a hard edge. The mutual inductance the to MSA is estimated by summing the on-axis field per unit of current from each circular loop for all 25 turns of the coil and multiplying by the effective area of the SQUID

$$M = A_{eff} \sum_{m,n} \frac{\mu_0 R_m^2}{2(R_m^2 + Z_n^2)^{3/2}}, \tag{5.33}$$

where R_m is the radius of the mth radial layer of wire and Z_n is the distance of the nth axial layer to the plane of the MSA. The effective area A_{eff} is the geometric mean of the SQUID hole area and washer coverage (200 μm × 200 μm and 1 mm × 1 mm, respectively, for the "LF" MSA design), due to Miessner-effect flux focusing. This predicts 0.28 ϕ_0/μA, (3.55 μA/ϕ_0) for the "LF" designs and empirical testing (subject to off-axis misalignment

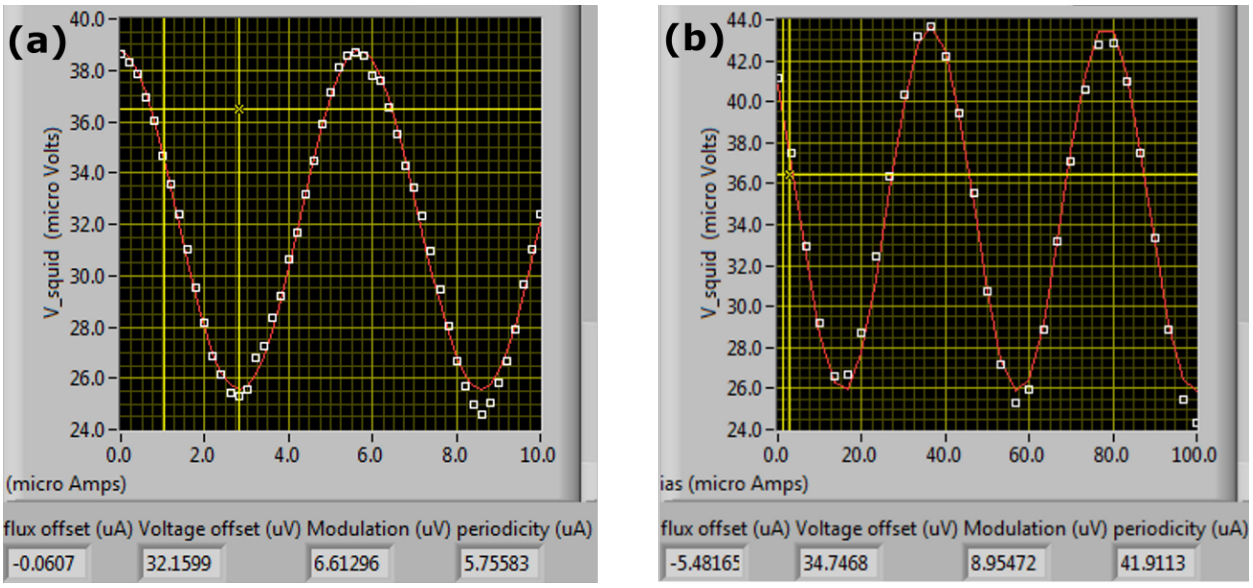

Figure 5.20: Periodic V vs. I_{coil} function, demonstrating the mutual inductance of the MSA and DC bias coil. Both plots show the DC SQUID voltage on the vertical axis and current applied to the DC bias coil on the horizontal axis in units of μA. The periodicity in V is ϕ_0, so the measure of μA per period is the reciprocal of the mutual inductance between the DC flux bias coil and MSA. In both cases the MSA is biased to the voltage state with $I_{bias} = 6\mu A$. (a) shows the V-ϕ for an "LF" series SQUID (inductance of 430 pH) demonstrating a mutual inductance of 5.75 $\mu A/\phi_0$. (b) shows the V-ϕ for an "HF" series SQUID (inductance of 80 pH) and a lower mutual inductance of 41.9 $\mu A/\phi_0$.

and fringing of the actual field) consistently showed a somewhat smaller mutual inductance close to 5.7 $\mu A/\phi_0$ as illustrated in figure 5.20.

Chapter 6

Experimental Measurement System and Methods

The delivery of a low-noise amplifier for ADMX is the key accomplishment of this work, but the path to that goal was not direct. The development of experimental methods and techniques sufficient to test, develop, and evaluate the MSA amounts to perhaps the largest portion of labor invested in this project. We include some portions of this less glamorous supporting work here in hopes it might prove useful to future experimenters. [1] This "supporting" work includes development of integrated inline filtering, the 4-K probe which allows for rapid thermal cycling between room temperature and 4.2K necessary for efficient testing and design iteration, dilution fridge wiring for both RF and DC signals, a custom heated load for accurate T_N measurement, and a fair amount of software allowing for efficient automation and unmonitored testing. The hardware design and considerations are covered here, and the software design is left to the online archive at [105].

6.1 Integrated Inline DC filter

The MSA package contains RC low-pass filters on the DC lines, but space limitations restrict the maximum capacitor size, and thus lowest cutoff frequency, that can be attained by filters inside the MSA package. Space constraints are relaxed outside the MSA package, so additional filtering can be added there. Both electrical and thermal considerations demand that any external filtering be placed as close as practical to the MSA package. This was accomplished by building an inline DC filter integrated to an external wiring cable, as shown in figure 6.1.

The external filter block consists of three independent floating bipolar passive RC filters and three independent ground-referenced passive RC filters. The PCB has two continuous internal Cu ground planes that ensure good thermalization. An external grounding wire

[1] When consulting past dissertations for reference, we found ourselves most often returning to the practical rather than theoretical sections.

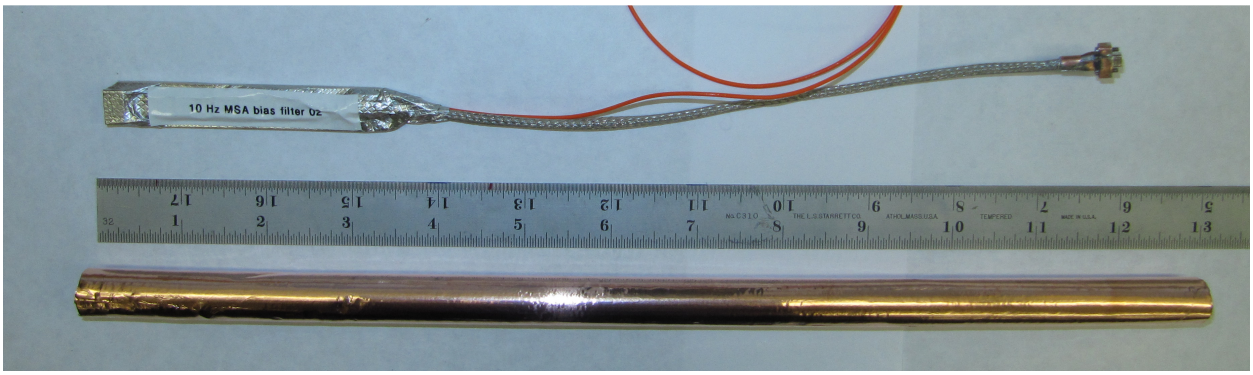

Figure 6.1: Inline DC filter alongside Cu-covered Pb shield.

(orange in figure 6.1) is soldered to the ground planes for use as a thermal anchor. The connectors are 9-pin nano-Ds, proven reliable[2] for cryogenic thermal cycling, and all wiring is Ag-clad Cu with PTFE insulation. The 9 "pigtail" wires are twisted and shielded inside a tinned Cu braid, which provides for both electrical shielding and strain relief. The braid is mechanically and electrically connected to the PCB by tightly twisted Cu wires and solder (figure 6.2) and mechanically and electrically connected to the male connector far from the PCB through a customized shroud made from 0.2-mm Cu sheet. The result is both flexible and robust against electrical interference and mechanical damage. The board is electrically insulated with one layer of kapton tape, then wrapped in tinned Cu tape with conductive adhesive which forms a nominally superconducting Faraday cage. For improved shielding against magnetic pickup, the entire assembly may be placed inside a thin-walled tube (bottom of figure 6.1) made from soldered Pb foil covered with Cu tape (the Cu tape serving both to improve handling safety and thermalization).

Identifying commercially available components that are stable and reliable from room temperature to 50 mK and survive repeated thermal cycling was no small task. It is rare to encounter a commercially available product rated for operation as low as 77 K, rarer still to find one rated to 4 K, and virtually unheard of to find anything with performance ratings at 50 mK or below.[3] We have never lost so much time as when we assumed a rating at 0 C could be extrapolated to 77 K, a 77-K rating could be extrapolated to 4 K, or 4-K performance could be extrapolated to mK temperatures. Figure 6.4 shows measured

[2]The nano-D (a.k.a. Bi Lobe) connectors made by Omnetics have become a favorite. We have never experienced an electrical failure of these connectors, despite frequent and sometimes aggressive thermal cycling. The screw-locking mechanism with captured screws also makes for very reliable mechanical stability.

[3]Nitrogen boils at 77 K, and liquid nitrogen is relatively common and inexpensive, so rating to this temperature by manufacturers are not too uncommon. Liquid helium boils at 4.2 K, requires greater care in handling to maintain in its liquid state, and is more expensive than LN_2 but allows for "passive dunk" cooling, so rating by manufacturers to this temperature is rare, but not unheard of. Temperatures below about 1 K require a dilution fridge, which cost about $0.5 M, so almost no one outside of a research lab can test at these low temperatures, and manufacturer ratings to such low temperatures are unheard of.

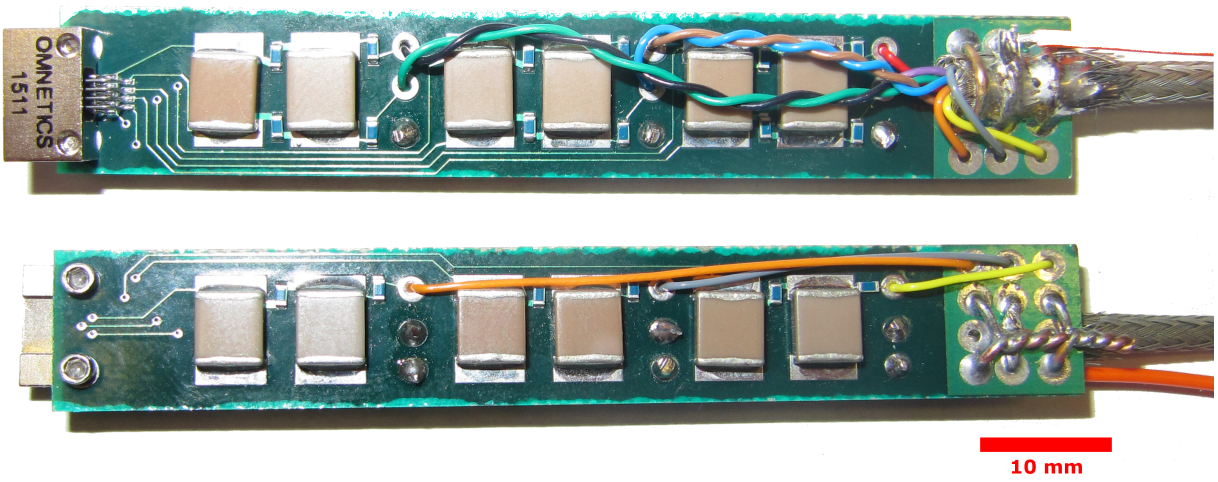

Figure 6.2: Inline DC filter circuit board. Top: top side of board with three 2-stage RC filters for flux bias, current bias, and voltage monitoring. Each filter is fully differential and without ground reference. Bottom: bottom side of board with three 2-stage RC filters for single-sided DC bias eg: varactor bias voltages. Each filter is ground referenced.

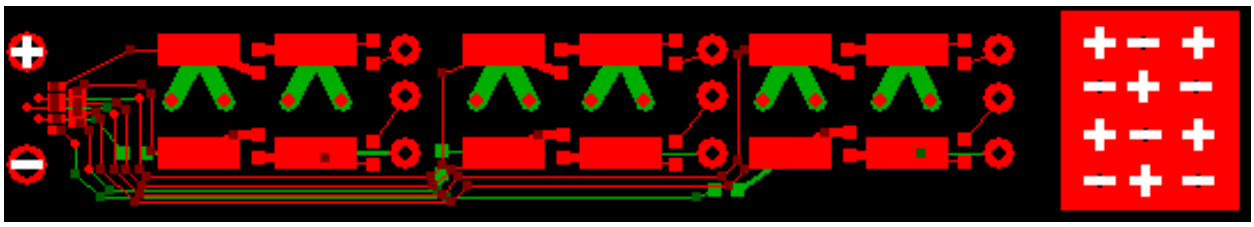

Figure 6.3: Board layout for in-line filter. Full layout file available at [105].

resistance vs. temperature for a particular Commercial Off The Shelf (COTS) SMT resistor commonly used in the Clarke group. This resistor has a specified temperature coefficient of 100 ppm/deg C, which comes to a 3% change in resistance between room temperature and dilution fridge temperatures. However, its ratings are only guaranteed from -55°C to 155°C, (218 to 428 K), and actual temperature performance from 20 mK to 300 K is shown in figure 6.4. Though resistance has changed by only a few % from 300 to 77K, the resistance has increased by 13 times at 20 mK. This is a useful feature in some circumstances[4] but would not be appropriate for this particular application. We have seen similar behavior from capacitors, connectors, and just about every other common component one would prefer to

[4]We have, in other experiments, created "health check" lines for verifying critical wiring in a dilution fridge at room temperature or early cool-down phases that transition to very high-impedance lines at mK temperatures for improved isolation during operation.

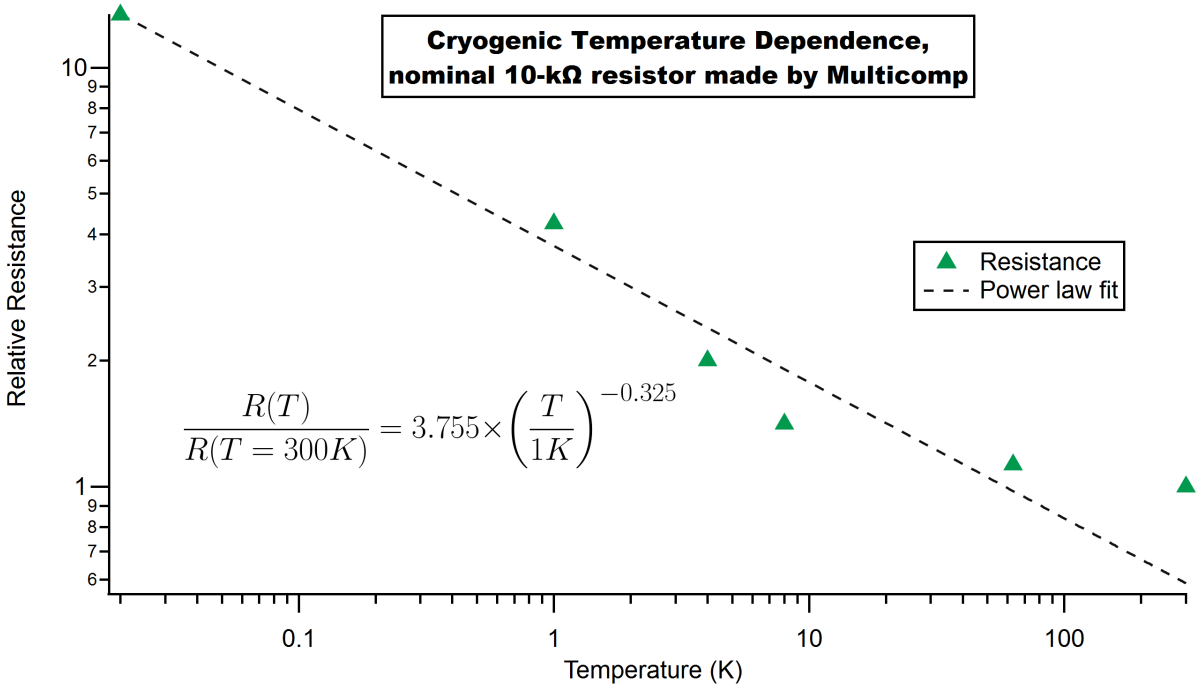

Figure 6.4: Surprising temperature dependence of a commercial-off-the-shelf resistor. Note that from 300 K to 77 K, the resistance is relatively unchanged, but by 20 mK, the resistance is 13 times higher, with an approximately power-law-like trend. The lesson here is: assume nothing, test everything!

take for granted. Though it seems tedious, we recommend every new component be tested at its service temperature before integrating it in any assembly. Some fruits of extensive testing are provided in table 6.1, a bill of materials for the inline DC filter.

The filter design may appear backwards on first inspection– "everyone knows" an RC low-pass filter has the capacitor in parallel with the load! This is true for the more common case of signals-as-voltage, with an ideally low impedance source and high impedance load. This application in contrast has signals-as current, (in the case of the current bias and flux bias) with an ideally high impedance source and low impedance load, and the filter is correct as built. The voltage monitor lines in contrast carry signals-as-voltage, but with the MSA as source and some lab pre-amp the load, the as-built "resistors-towards-MSA" orientation is correct.

The inline filter was designed with a mind to reduce 60 Hz pickup in the bias lines, and figure 6.6 (a) shows a simulation of the rolloff characteristic, with about a 10 dB reduction in power at 60 Hz. Figure 6.6 (b) shows measured performance at 4 K, which is slightly less effective than modeled at 60 Hz, but proved effective nonetheless. Future iterations of the design might attempt a lower frequency rolloff.

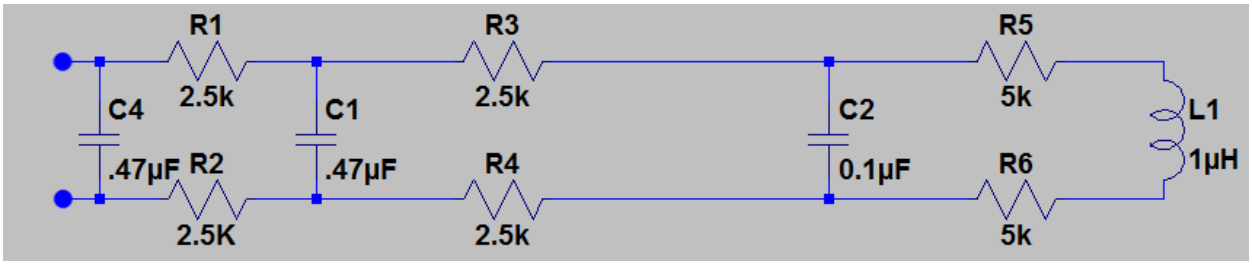

Figure 6.5: Schematic of a single differential pair filter. This schematic is for the DC flux bias in particular– the 1 μH inductor represents the DC flux bias coil. Right-hand components (C2, R5, R6, L1) are inside the MSA package, and the other components are part of the external inline filter. Electronic design files available at [105].

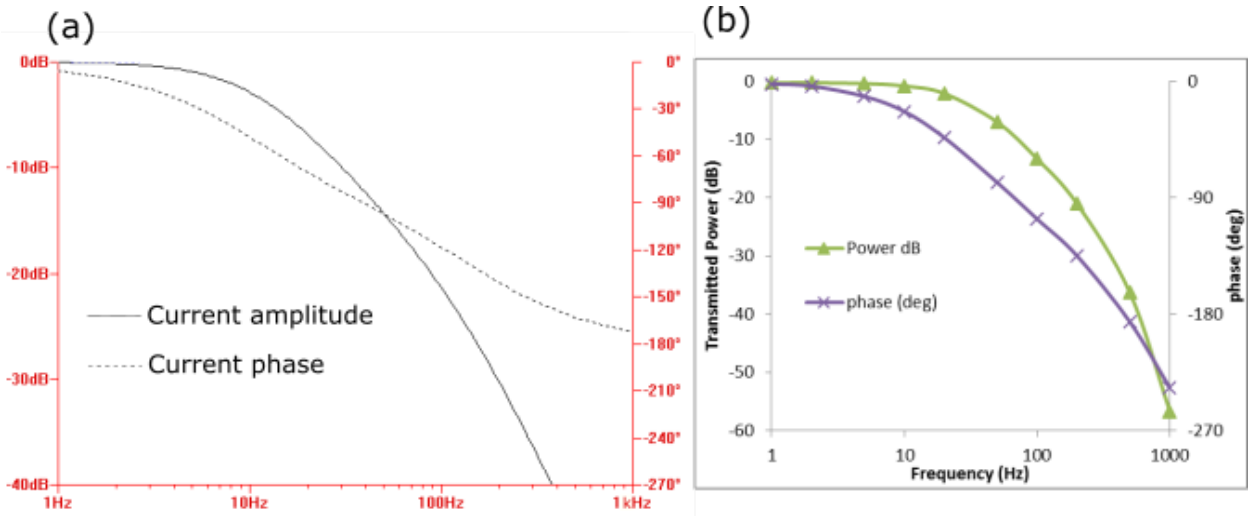

Figure 6.6: Inline DC filter performance, theory and experimental data. (a) Simulated performance based on schematic of figure 6.5. Relative magnitude and phase of current through the flux bias coil for a sinusoidal current into the filter. (b) Measured performance at 4 K. (Virtually identical to performance at room temperature.) Relative current power and phase of current through a low impedance load for a sinusoidal current applied through the filter.

Table 6.1: Inline DC filter Bill of Materials. electronic references for custom parts available at [105].

Description	P/N	Supplier	Quant.
Cryogenic cap, 0.47uF	399-5385-1-ND	DigiKey	12
Cryogenic resistor, 2.49kOhms	TNPU2.49KBACT-ND	DigiKey	18
Nano-D 9-pin male, flying leads	A28000-009	Omnetics	1
Nano-D 9-pin female, horiz. SMA	A29100-009	Omnetics	1
PCB circuit board	Custom, see .pcb file	Express PCB	1
.125" 0-80 socket cap screws			2
Tinned Cu braid	NEQ 24-5-36T	CoonerWire	10"
1" kapton tape			6"
Ag-clad Cu wire for thermalization			10"
1" Tinned Cu tape	3M11418-ND	DigiKey	6"
Braid strain relief	Custom, see cutting guide.		1

6.2 4-K Test Station Design

The "4-K probe" is a quick testing station that allows for quick cycling from room temperature to 4.2 K. It allows for testing of all relevant MSA performance, and with a cooling time of about 20 minutes and a warming time of about 40 minutes, allows for multiple cryogenic testing sessions and design revisions in a single day. In contrast a "proper" test at mK temperatures requires about 36 hours to cool down and 24 hours to warm up.

The probe connects room temperature components to the liquid helium bath, and the heat conducted to the bath must be considered both for economy (not boiling off too much helium) and practicality (not icing up the room temperature end). The standard 100-liter He storage dewars boil off about 1 to 2 liters per day without any probe installed, equivalent to 36 to 72 mW of heat power conducted from room temperature to the bath. This sets the benchmark for thermal conductivity of any component that extends from room temperature to the liquid helium.

The 4-K probe can be described in three sections: the room temperature junction box with amplifiers, the room-temperature to 4-K internal wiring, and the cold tube and mechanical supports, as shown in figure 6.7.

The top junction box with amplifiers shown in figure 6.8 incorporates a hermetic transition for signals from the helium space to lab atmosphere, room temperature amplifiers, connectors we found to be reliable for low-level signals, and at least one useless feature.[5] The hermetic via is a brass KF-40 stub with 9 DC lines and 2 RF lines potted in Stycast® 2850 epoxy. The RF lines are short stainless steel (more robust than copper line) connectorized on each end to minimize the chance of damaging the potted cables. The nine DC lines originate in a pre-wired nano-d connector (Omnetics P/N A29000-009) embedded in the epoxy as visible in figure 6.8c. The three floating low-level bipolar dc signals (current

[5]Useless but intentional.

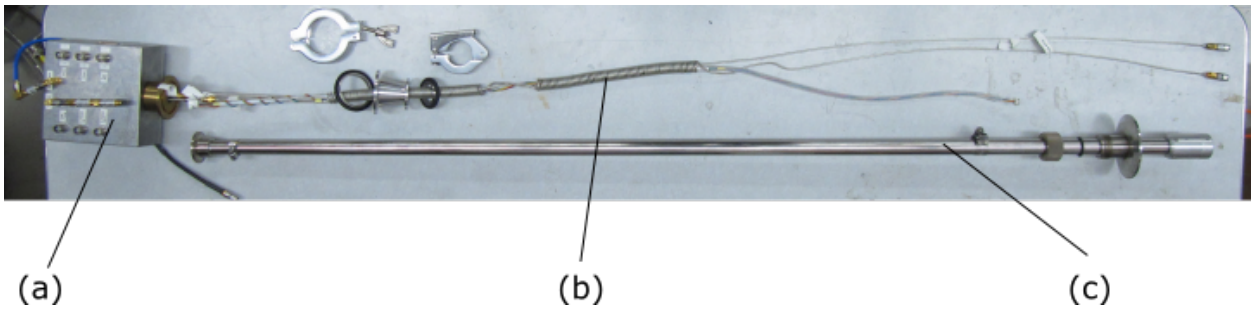

Figure 6.7: 4-K probe exploded view. Total length is about 1.5 m. (a) Room-temperature junction and amplifier box. (b) Internal room-temperature to 4-K wiring. (c) External tube for and adapters.

bias, voltage monitor, and flux bias) are coupled through BNC twinax connectors (Pasternack P/N PE4324) for attachment to shielded, twisted-pair cables at room temperature. Square flange connectors are much preferable to through-hole connectors because they will not twist when connecting cables. The single ended (referenced to chassis ground) signals are coupled through ordinary 50-Ω BNC connectors at room temperature. The SMA connectors are simple bulkhead thru connectors. Square-flange SMA bulkhead connectors are available, but are not necessary at attachment points like these that typically remain assembled during everyday use.

The DC wiring highlighted in figure 6.9 starts at a pre-wired nano-D connector (Omnetics P/N A28000-009) and the wires are extended with a twisted bundle of nine 30 AWG Ag-plated Cu 7-strand wire, MIL-W-16878 type ET[6] conductors. The DC wiring bundle is installed with extra length for tensile strain relief and the excess length wrapped loosely around a thin rod to keep it tidy. Near the cold end where the DC wires are regularly handled, the bundle has extra armor in the form of a polymer spiral wrap [figure 6.9(b)] and is terminated with an identical nano-D connector. The DC lines conduct heat to the liquid helium at a rate of about 37 mW, approximately equal to the heat flow through the wall of a better performing storage dewar.

The RF cables consist of two lengths of hand-formable 50-Ω coax (Amawave P/N UT-085C-FORM) and these cables directly support the weight of the MSA while it is under test. The RF lines are each tightly coiled about 75 times, forming mechanical springs and affording shock protection to protect the MSA and the RF connections. The coils add about 3 m to the total length of cable, so thermal conduction of heat to the liquid helium is reduced from 300 mW (1.5-m cables) to 100 mW (4.5-m cables), saving on boiloff costs. The RF line running to the MSA input is terminated by a -30 dB attenuator in the helium bath to reduce the broadband temperature noise of the incoming signal. Even a perfectly "noiseless" microwave generator will present a 300-K white noise at its output, which would be sufficient

[6]Mil-spec wire may seem overkill. Suffice to say the path to adopting "mil-spec" wiring is littered with failed and unreliable wiring solutions.

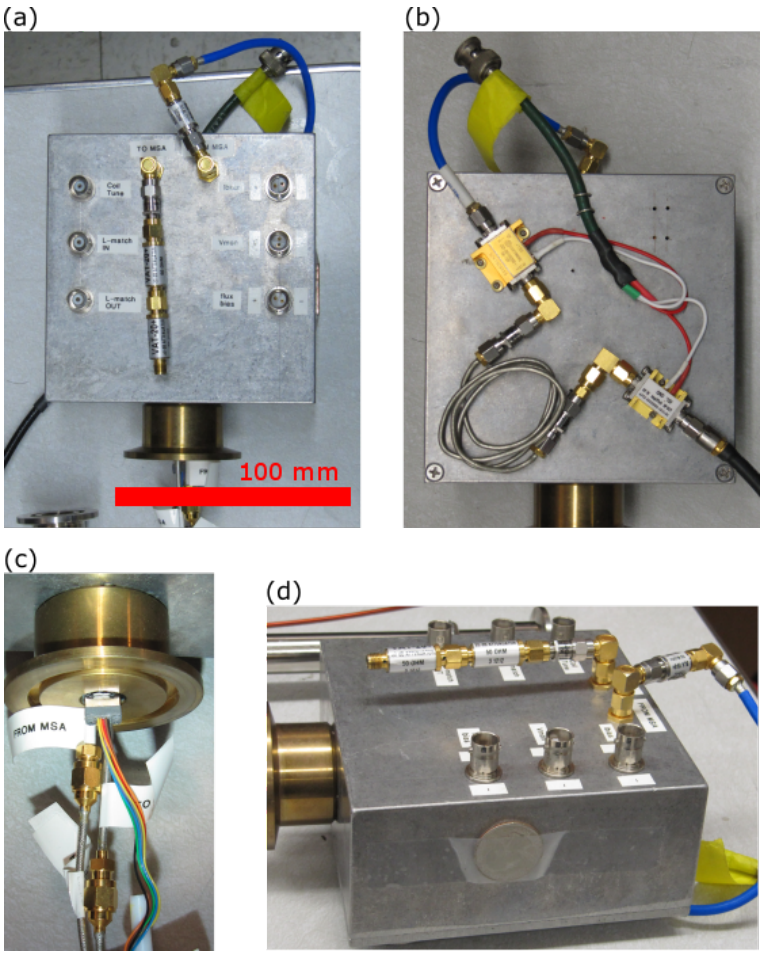

Figure 6.8: 4-K probe room-temperature junction box. (a) Room-temperature connectors. Left column: "Coil tune", "L-match IN", and "L-match OUT" BNC connectors for varactor control. Center: SMA RF connections, "TO MSA" with a total of -50 dB attenuation connected, and "FROM MSA" with a DC-block and mini-circuits cable (blue) running to the integrated amplifiers mounted on the opposite side. Right column: Twinax connectors for the floating bipolar DC signals "Current bias", "Voltage monitor", and "Flux bias". (b) Backside with RF amplifiers. The amplifiers are mounted to the Al junction box, which serves as a heatsink. Both are powered by +15V, delivered on the shared BNC line. The first amplifier is an LNA with $T_N = 50$ K, and the second is an ordinary, less expensive amplifier. They are joined by an assembly of 30 cm of 50-Ω transmission line and two -3dB attenuators, which reduces reflections and flattens the gain response. (c) All signals transition to the helium space through a KF stub potted with Stycast® 2850 epoxy for a nominally hermetic seal, and all lines (DC and RF) are connectorized for ease of repairability or reconfiguration. (d) A US 1/4 dollar coin is affixed the the side with clear adhesive tape, and serves no useful purpose. We found that some similar constructions built by past researchers unknown had coins attached. We continue the intentional practice of adding a purposeless feature as a humorous reminder against unintentional purposelessness in our designs.

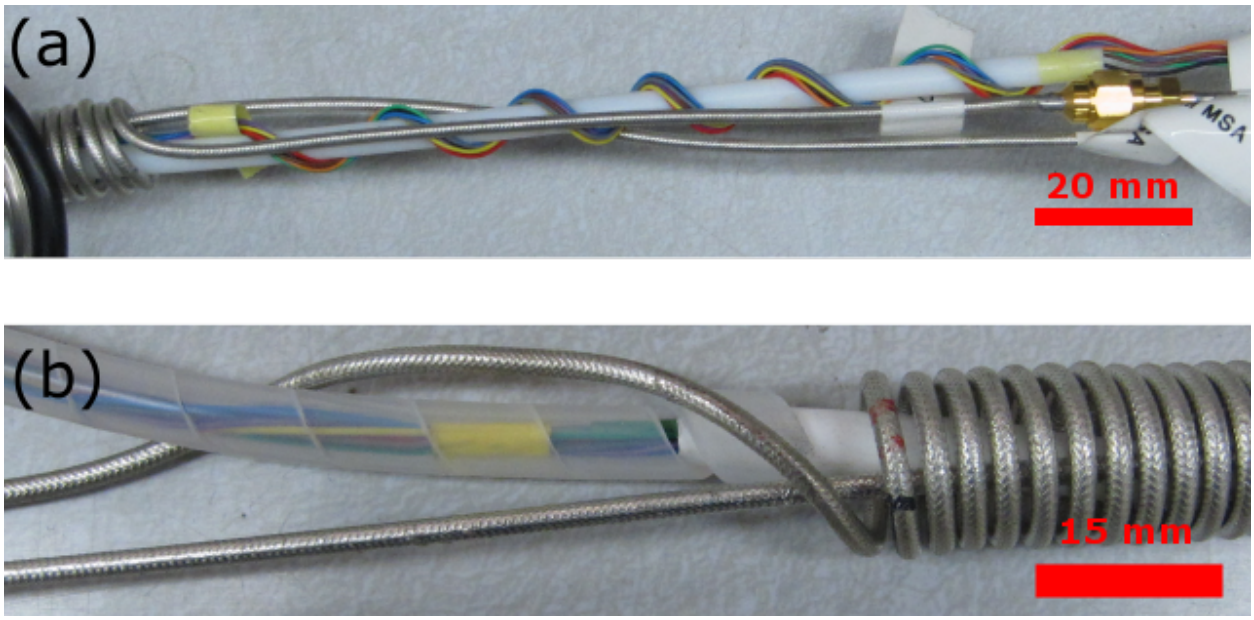

Figure 6.9: 4-K probe DC wires. The 4-K probe DC wiring consists of a twisted bundle of nine 30 awg Ag-plated stranded Cu lines with PTFE insulation. (a) The DC lines are rather delicate and should never have any tension applied. They are installed with extra slack, and wound loosely around a PTFE post near room temperature for cable management. There is no relative motion at this section so no armor is necessary. (b) At the cold end the DC wires may move relative to the RF cabling and are often handled during sample connection, so a polymer spiral wrap is used for extra mechanical protection.

to overwhelm the MSA with broadband power. The -30 dB attenuator will emit a thermal spectrum that is a weighted average of its own temperature and room temperature, namely

$$\mathrm{T}_{transmitted} = 0.999 \times 4.2 \text{ K} + 0.001 \times 300 \text{ K} = 4.5 \text{ K}, \tag{6.1}$$

which is sufficiently cool for the MSA to operate without deleterious effect. The output of the MSA is coupled to the "outgoing" coaxial line through a -3dB attenuator, which serves as a "poor mans circulator", reducing the SWR on the 4.5-m coax between the MSA and LNA, and "whitening" the response of the test system. The -3dB attenuator reduces the effective gain of the MSA, but his can be easily accounted for by calibration. Critically, it effectively doubles the T_N of the room temperature LNA, which must be accounted for in MSA T_N estimations.

The 4-K probe supporting tube starts at room temperature with a KF-40 flange on the breakout box (large enough to accommodate the hermetic potting and connectors) then transitions to a KF-25 flange welded to a SS 304 tube with a 19-mm OD and 16.3-mm ID (figure 6.11). The tube dimensions are chosen as a compromise between strength and thermal conductivity. The tube must mate to a standard He storage dewar, and an adapter

Figure 6.10: 4-K probe RF line shock coils. Note the right hand coil transitions to a straight segment transversing the center of the left coil, and the left coil transitions to a straight section transversing the center of the right coil. Each RF line acts as an independent spring and shock absorber while supporting the MSA sample. The RF cables are the greatest source of conducted heat from room temperature to the liquid helium bath, and the coiled sections add considerably to the length (and thus reduce the thermal conduction) of the RF cables.

is made from a 0.75" quick-connect (Lesker P/N S-075-K) welded to a 2.5-inch tri-clover cap (commonly available from dairy and beer equipment providers), shown in figure 6.12. The quick-connect allows the probe to be raised or lowered inside the storage dewar without breaking the helium-to-atmosphere seal. A warm probe may be lowered in small increments (over a time period of about 20 minutes) into the dewar to make best use of the available enthalpy of the warming helium gas. When it is time to warm the probe, it may be raised to the top of the dewar to warm in a helium atmosphere near room temperature before removal and exposure to atmosphere. The end of the tube is terminated in an Al shield holder, shown in figure 6.13. The shield holder is securely mounted to the tube by 8 screws (4 thru-hole screws for positive engagement, and 4 set screws for zero mechanical lash) and is fully removable to allow servicing of the quick-connect-tri-clover section. The heat conducted from the SS tube from room temperature to the liquid helium is about 160 mW, on par with the heat conducted by the Cu RF cables. Total estimated heat conduction from the DC wiring, RF cables, and SS tube from room temperature to the liquid helium is about 300 mW, or about 4 to 8 times the quiescent heat flow of the dewar. This might be improved by, say, reducing the wall thickness of the SS tube, but the current design only boils off about $2.30 of helium an hour, so further improvements are not highly motivated.

A high permeability magnetic shield shown in figure 6.14 terminates the probe. The MSA is a device exquisitely sensitive to magnetic fields, so it is important to cool and operate the MSA in a low magnetic field. To achieve this, a cylinder made of A4K® alloy (nominally equivalent to cryoperm) with a thin inner lead lining is assembled over the MSA package prior to cooling. The diameter of the shield is chosen to fit inside the throat of a He storage dewar with about 1 mm of assembly clearance. The shield is affixed to the shield adapter by 4 radial set screws, which are turned *in* to release the shield and turned *out* to engage the shield mounting holes. This attachment method eliminates any screw head height and allows for the maximum possible shield diameter and experiment volume.

A small but important feature of the magnetic shield is the drainage hole, shown in figure 6.15. The hole causes a small degradation in shielding performance, but allows helium to drain out when the probe is raised, conserving liquid helium and reducing the time to warm.

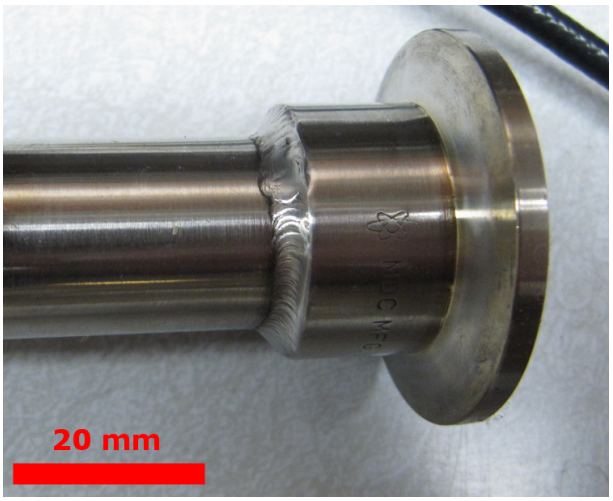

Figure 6.11: 4-K probe welded KF25 flange.

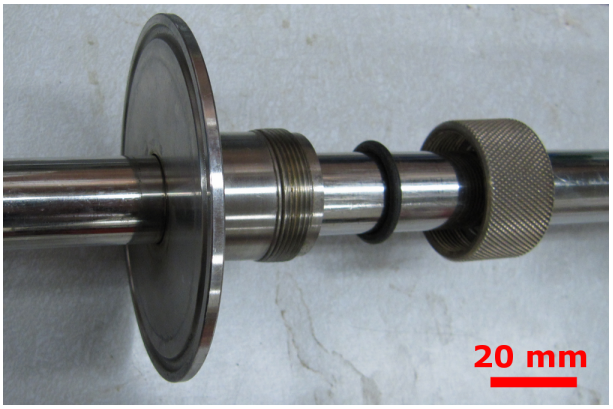

Figure 6.12: 4-K probe quick-connect to tri-clover flange.

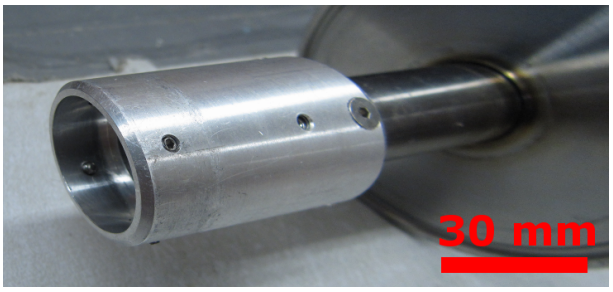

Figure 6.13: 4-K probe magnetic shield adapter.

Figure 6.14: 4-K probe magnetic shield.

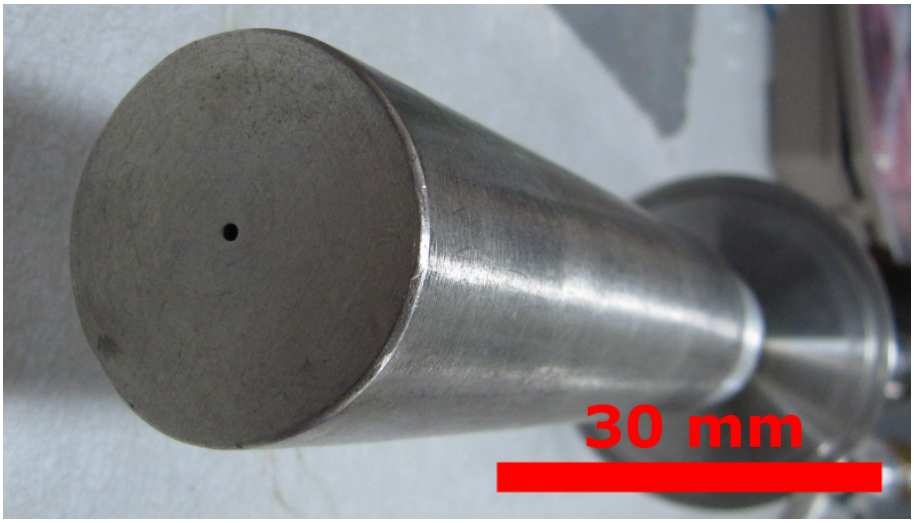

Figure 6.15: 4-K probe magnetic shield drainage hole.

The 0.2-mm thick lead liner inside the magnetic shield provides increased magnetic shielding because lead is a superconductor at 4 K. Though lead is nominally a type I superconductor, it is notorious for being riddled with both chemical and crystal defects that can trap flux. For this reason, the static magnetic field present when the temperature of the lead lining falls below T_c will in general be frozen into the lead foil, rather than expelled. However, any AC fields (for example, from 60 Hz power lines) are be greatly attenuated by the lead foil. The lead liner is shown in figure 6.16. Note that the seams are all continuously soldered to achieve an uninterrupted superconducting envelope.

Figure 6.16: 4-K probe magnetic shield and lead liner. The lead liner tends to "lock in" the static magnetic field present when it cools through T_c, but greatly attenuates and AC fields (for example, 60 Hz) inside the liner. Note the continuously soldered joints. The author has the manual skills to achieve soldered joins on 0.2-mm lead foil, but has insufficient writing skills to impart the skills with verbiage. We recommend practice, patience, and a hot-air solder re-work station.

6.3 Dilution Fridge Test Station and Design

To fully qualify the MSA for operation in ADMX, it must be tested at the ADMX operating temperature of about 100 mK. This necessitates preparing a ^{3}He dilution fridge with necessary cabling, and due to the low cooling power of a ^{3}He dilution fridge, the thermal conduction budget is several thousand times smaller than that of a 4-K dunk probe. Furthermore, the noise temperature T_N must be accurately assessed, which is best done with a heated load and Y-factor measurement. The heated load must be optimized to have excellent temperature stability, rapidly adjustable temperature, and must not overwhem the cooling power of the dillution fridge.

6.3.1 Thermal Considerations for Cables and Wires

Coupling signals between the room temperature lab environment to the experiment at mK temperatures necessitates running wires or cables between the two, and it is imperative to minimize the thermal conductivity of those connections. In contrast to the 4-K probe, with a thermal budget in the 100's of mW, the Oxford Triton 200 dilution fridge used in the experiments described in this dissertation has a nominal cooling power of 200 μW at a mixing chamber (MXC) temperature of 100 mK, with cooling power proportional to T^2. Thus at a MXC temperature of 60 mK, the sum of all heat conduction to the MXC must be less than 72 μW. By way of example, a single 40 AWG copper wire (0.005 mm^2 cross-section, about the size of a human hair) strung the 1 m distance between the top room temperature plate and MXC will conduct 600 μW of heat.[7]

A typical experiment run in a dilution fridge could have several coaxial lines for RF signals and dozens of twisted pairs for low-frequency signals. The sum cross-sectional area will be magnitudes greater than the preceding example of a single 40-AWG wire, so care must be taken to mitigate the thermal power conducted to the MXC by wires and cables. Heating of the MXC is mitigated in two ways: choice of low thermal conductivity materials, and through thermally sinking of wires and cables to intermediate temperature stages.

Metals can be broadly categorized into three categories based on thermal conductivity: high thermal conductivity (ex: copper), low thermal conductivity (ex: stainless steel), and extremely low thermal conductivity (ex: niobium-titanium superconductor). The high thermal conductivity metals are typically chemically pure and have low electrical resistance.

In any solid, heat is transferred almost entirely by conduction electrons and lattice vibrations (phonons).[8] For purposes of comparing the thermal conductivity of materials and

[7] This is a conservative estimate assuming a constant thermal conductivity k equal to the room temperature value. In truth, the conducted power is proportional to $\int_{T_C}^{T_H} k(T)dT$, where the thermal conductivity $k(T)$ is greater at low temperatures than at room temperature.

[8] Other excitations in a solid, for example plasmons, polaritons, polarons, magnons, or excitons, can also transfer heat, but in the applications considered here their contribution to the total thermal conductivity is small.

	low T	high T	difference between materials
v	const	const	Depends on stiffness and density. Varies from 100's of m/s to 12 km/s.
c_v	$\propto T^3$	const	About equal. $\approx$ 25 J/(K·mol) (high T limit)
λ	const	$\propto 1/T$	Highly dependent on order and purity. Varies from 100's of Å[107] to 10^6 Å[108]. (low T limit)

Table 6.2: Phonon temperature and material dependence of v, c_v, and λ.

their temperature dependence, the thermal conductivity k is proportional to the product of three factors:

$$k \propto v \cdot c_v \cdot \lambda \tag{6.2}$$

where v is the mean carrier velocity, c_v is the heat capacity, and λ is the mean free path. Each of these terms has a unique temperature dependence which goes into explaining the overall temperature dependence of thermal conductivities shown in figure 6.17.

In the case of phonons, the velocity v_{phonon} is a function of density and stiffness, neither of which is much dependent on temperature, so v_{phonon} is roughly constant with temperature. The value of v_{phonon} varies between materials from about 1 km/s in a soft, dense metal like lead, to about 12 km/s in diamond which has the greatest v_{phonon} known[9] due to it's very high stiffness and moderate density. The heat capacity of phonons c_v is proportional to T^3 at low temperatures and constant at high temperatures, as described by the Debye model. The border between "high" and "low" temperatures is given by the Debye temperature θ, which varies by material but is typically on the order of 100 K. The high and low temperature limits are reasonably accurate in the limits of $T < 0.1\,\theta$ and $T > \theta$. The mean free path of phonons λ_{phonon} is limited by two parallel scattering mechanisms: impurity scattering and phonon-phonon scattering. The density and scattering strength of impurities, whether they be chemical impurities or crystal defects, is independent of temperature, but the strength of phonon-phonon scattering is proportional to the phonon density, which is proportional to T^4 at low temperature and proportional to T at high temperatures. Thus λ_{phonon} is constant at low temperature (dominated by impurity scattering) and proportional to $1/T$ at high temperature (dominated by phonon-phonon scattering). These phonon temperature and material dependencies are summarized in table 6.2.

In the case of conduction electrons $v_{electron}$ is close to the Fermi velocity v_F, which depends on the lattice spacing and number of conduction electrons per atom, neither of which is much dependent on temperature, so $v_{electron}$ (like v_{phonon}) is constant with temperature. Because the atomic spacing and conduction electrons per atom vary only by a factor of 2 or 3 between materials, the value of $v_{electron}$ is within an order of magnitude for all metals, usually around $v_F = 10^6$ m/s. Since the conduction electrons are fermions, at low temperatures only a

[9]With the possible exception of carbon nanotubes.

	low T	high T	difference between materials
v	const	const	About equal. $\approx 10^6$ m/s.
c_v	$\propto T$	N/A	About equal. ≈ 2 mJ/(K^2 · mol)
λ	const	$\propto 1/T$	Highly dependent on order and purity. Varies from 100's of Å to 10^6 Å. (low T limit)

Table 6.3: Conduction electron temperature and material dependence of v, c_v, and λ.

small fraction, proportional to T/T_F, with $T_F \equiv (m_e/2k_B) \cdot v_f^2$ are thermally promoted out of the degenerate Fermi gas to a state that can conduct heat. The Fermi temperature T_F is typically about 30,000 K, so for purposes of conduction electrons at all our temperatures of interest, every temperature is a low temperature. The mean free path of conduction electrons $\lambda_{electron}$, like that of phonons, is limited by two parallel scattering mechanisms: impurity scattering and electron-phonon scattering. The density and scattering strength of impurities, whether they be chemical impurities or crystal defects, is independent of temperature, but the strength of electron-phonon scattering is proportional to the phonon density, which is proportional to T^4 at low temperatures and proportional to T at high temperatures. Thus $\lambda_{electron}$ is constant at low temperature (dominated by impurity scattering) and proportional to $1/T$ at high temperature (dominated by electron-phonon scattering). These conduction electron temperature and material dependencies are summarized in table 6.3.

With this understanding we can make sense of the data shown in figure 6.17. In metals, the thermal conductivity is generally greater than insulators. The thermal conductivity of conduction electrons is much greater than that of phonons, mostly because $v_{electron} \gg v_{phonon}$, so in metals often only the electron contribution is significant. Consider the thermal conductivity curves for copper: there are 5 curves (solid red lines) for 5 samples of Cu with residual resistivity ratios (RRR) ranging from 50 to 500. The RRR is a measure of the *electrical* conductivity, but here we can use it as a proxy for the chemical purity and microscopic crystal order, where higher RRR means greater purity and order. All 5 samples have similar conductivities at room temperature, where electron scattering is dominated by electron-phonon interactions, and each sample has the same density of thermally-excited phonons. As the temperature decreases, the phonon density decreases and λ increases until it is limited by impurity scattering. The higher RRR sample has fewer impurities and a greater maximum k. As the temperature decreases further, λ remains constant and the $c_v \propto T$ term dominates the temperature dependence of k.

Consider next the thermal conductivity for aluminum: the upper blue curve is for "commercially pure" Al 1100 ($\approx$ 99% pure), and the lower blue curve is for "all purpose" Al 6061 ($\approx$ 95% pure). The Al 1100 has a similar shape to the copper curves, but with less pronounced peak in k due to the enhanced scattering from the 1% impurity ratio, and a lower k overall. The 5% impurity ratio of Al 6061 has results in so much impurity scattering

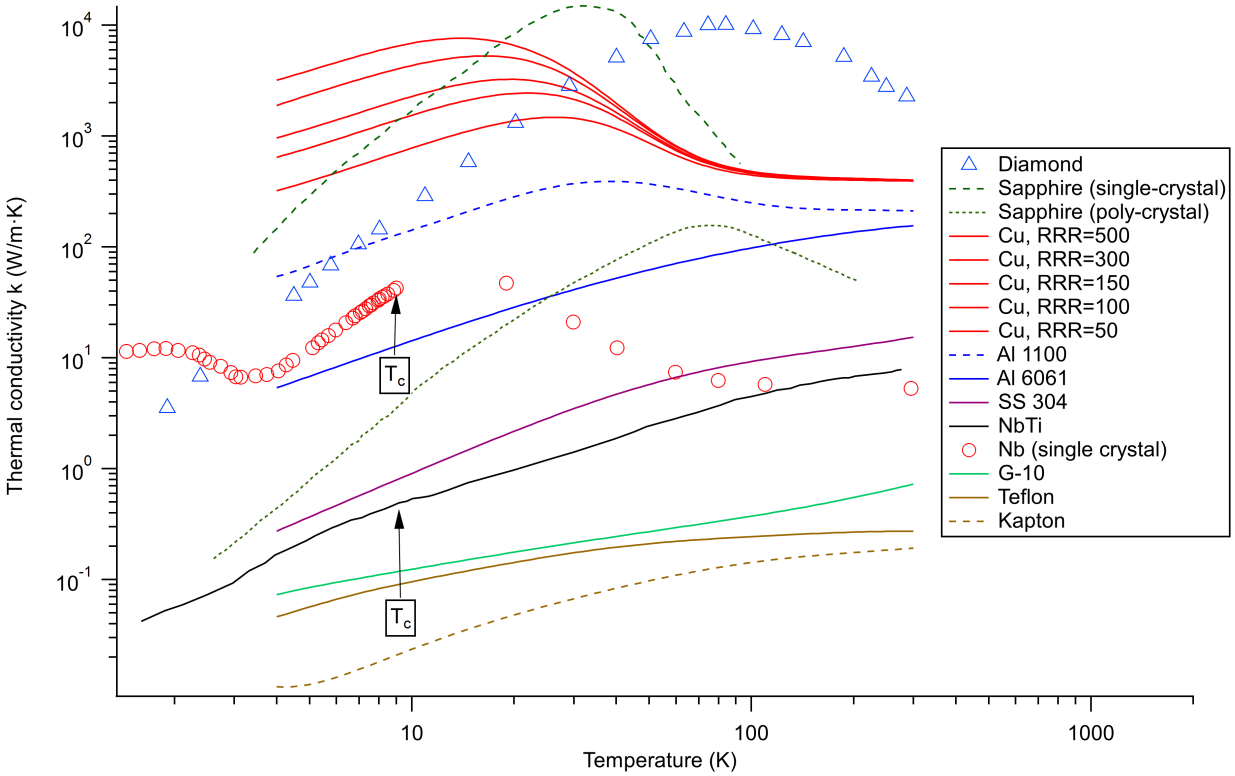

Figure 6.17: Thermal conductivity $k(T)$ as a function of temperature for some materials of interest. Data for copper, aluminum, stainless steel, G-10, teflon, and kapton are from [109], data for NbTi are from [110], data for diamond are from [111], data for sapphire are from [112], data for niobium ($T < T_c$) are from [113], and data for niobium ($T > T_c$) are from [114].

that the peak in k is altogether suppressed, with only a slight deviation from $k \propto T$ behavior at the highest temperatures due to additional phonon scattering.

The thermal conductivity of stainless steel (purple curve) is similar to Al 6061, but as a metal with a high percentage of alloying agents (about 20% chromium and 10% nickel) its thermal conductivity is an order of magnitude lower. The greatest amount of scattering due to chemical impurity occurs at a stoichiometric 50-50 mix of metals, as found in NbTi (black curve), CuNi, and other useful alloys. Every metal alloy with a ratio close to 50-50 has a thermal conductivity similar to that shown for SS 304 or NbTi. A thermal conductivity at these low values marks the first point where one would consider using these materials for making physical connections between temperature stages of a dilution fridge. The mechanical linkages between stages in a dilution fridge are often SS struts made hollow for maximum strength with minimum cross-sectional area, and the plumbing carrying gases between stages are typically made from SS as well.

There are several insulators on figure 6.17, including G-10 (a fiberglass reinforced phenolic, as commonly used in printed circuit boards) (green curve), Teflon (brown curve), and Kapton (dashed gold curve), all frequently used in cryogenic applications. All of these lack conduction electrons so phonons are the remaining dominant heat carrier, resulting in low k compared to metals. These three common non-conductors also have a disordered (non-crystalline) micro-structure, so phonon-disorder scattering is high, and k is small relative to insulators with a high degree of crystalline order. These materials are effectively thermal insulators in that one does not worry about heat loading caused by these materials bridging thermal stages in a dilution fridge. They all have diminishing thermal conductivity at lower temperatures, but a clean power-law dependence is rarely exhibited due to overlapping trends in c_v and λ.

An exception to the "electrical insulators are also thermal insulators" rule is found in diamond (blue triangles) and sapphire (green dashed) which have a very high phonon velocity and can be made as highly regular crystals with very little internal scattering. Diamond is not often used in cryogenic applications, but is included here because it is well known for having excellent thermal conductivity. The diamond data shown in figure 6.17 is for a rather pure single-crystal sample with very little impurity scattering. Sapphire is often thought of as "good, but not as good as diamond" in terms of thermal conductivity, but that heuristic is based on "human-level" experience with the materials at room temperature or somewhat hotter.[10] Sapphire is not quite as stiff as diamond so it has a lower Debye temperature (Θ_D) and lower thermal conductivity than diamond at room temperature, but at temperatures well below Θ_D sapphire can have a higher thermal conductivity than diamond. At low temperatures where phonon-phonon scattering is negligible and λ is constant, the $k \propto T^3$ behavior is readily apparent in diamond, single-crystal sapphire, and poly-crystal sapphire. Note that poly-crystal sapphire (aka alumina) has a considerably lower thermal conductivity due to its short mean free path λ set by the average internal crystal size. Diamond is not used in this work, but single-crystal sapphire can find use in cryogenic applications as a substrate for superconducting qubits, mechanical supports for the tuning rods in ADMX, or other cryogenic electronics when one wants efficient themalization through an electrical insulator.

Finally, there are two superconducting metals in figure 6.17, Nb and NbTi. They behave very much like other metals (pure or alloyed, respectively) above the superconducting transition temperature T_c (about 9 K in both cases). Below their transition temperature, c_v of the conduction electrons rapidly approaches zero because the conduction electrons condense to the superconducting ground state which carries zero entropy. One might expect a corresponding discontinuous drop in k, but this is not observed. Superconductivity is a result of a phonon-mediated electron-electron attraction, so superconducting materials typically have a strong phonon component to their thermal properties. Though the electron contribution to

[10]For example, the old gem-monger's saying 'Glass is a reluctant vampire.'[115] meaning that a true gem will feel cold in the hand, while glass will feel warm, only 'reluctantly' stealing heat from one's hand. Though glass, sapphire, and diamond have similar heat capacities (about 1J/(g·K) sapphire and diamond have a thermal conductivity about 100 to 1000 times higher than glass, respectively at body temperature.

k in a superconductor is exponentially suppressed at $T < T_c$, the total thermal conductivity is not greatly changed relative to a normal metal with a linearly suppressed k. Though the thermal conductivity of a superconductor below T_c is entirely due to phonons, k can still be quite large. In the case of pure Nb, k can be greater than that of common Al alloys! When selecting a superconductor for low k, it is always best to chose an alloy such as like NbTi or NbZr.

To reduce the thermal power conducted to the MXC of a dilution fridge, the importance of thermal sinking to intermediate temperature stages rivals that of the choice of low-k materials. As a general rule much higher cooling powers are available at the higher-temperature stages of a dilution fridge, so a favorable trade-off can be made: by increasing the heat conducted to warmer stages, the heat conducted to the coldest stages can be markedly reduced. Table 6.4 shows typical temperatures and cooling powers of the successive stages of a "dry" dilution fridge.

Table 6.4: Dilution fridge temperature stages

Fridge stage:	Mixing Chamber	100-mK plate	Still	4-K Plate	50-K Plate	Room temp.
Typical temperature:	60 mK	100 mK	700 mK	4 K	50 K	300 K
Approx. distance from top plate:	1 m	0.8 m	0.6 m	0.4 m	0.2 m	0 m
Approximate cooling power:	72 μW	–	10 mW	200 mW	1 W	–

The "50-K plate" and "4-K plate" are cooled by a pulse tube, and have substantial cooling power. The still is where ^{3}He is boiled off from the ^{3}He/^{4}He mixture and requires heat input to function. Most of that heat is supplied by electrical joule heating, so some passive heat dumping to the still actually increases total efficiency. The "100-mK plate" has no active cooling, and sits passively at the interface of the colder step heat exchanges and the warmer continuous flow heat exchangers. The "mixing chamber" (MXC) is cooled by the enthalpy of mixing nearly-pure liquid ^{3}He into the liquid ^{3}He/^{4}He mix, and has a cooling power proportional to T^2.

The advantage of good thermal anchoring is illustrated in figure 6.18, which shows the temperature profile and thermal power flow through a single strand of 40 AWG copper wire in two hypothetical cases: no intermediate thermal anchoring (blue), and ideal thermal anchoring at each stage (red). The blue traces show a constant thermal gradient[11] (solid) from 300 K to 60 mK and a constant heat flow of about 600 μW (dashed) from room temperature to the MXC, as in the earlier calculation. The red traces show a piecewise linear thermal gradient (solid), anchored to the (position, temperature) coordinates from table 6.4, and a power flow (dashed) decreasing stepwise with temperature at each anchoring point. Note the difference in power flow traces: the anchored line introduces about *4 times more* heat into the fridge from room temperature than the non-anchored line, and that power is transmitted to the 50-K stage. The power transmitted (a few mW) is an insignificant

[11]This illustration assumes a constant thermal conductivity, ie $k(T) = k(300\ K)$. This is not generally true, and the actual form of $k(T)$ can have important practical consequences in cryogenic design. However, the precise form of $k(T)$ does not impact this illustration, so accuracy is sacrificed for clarity.

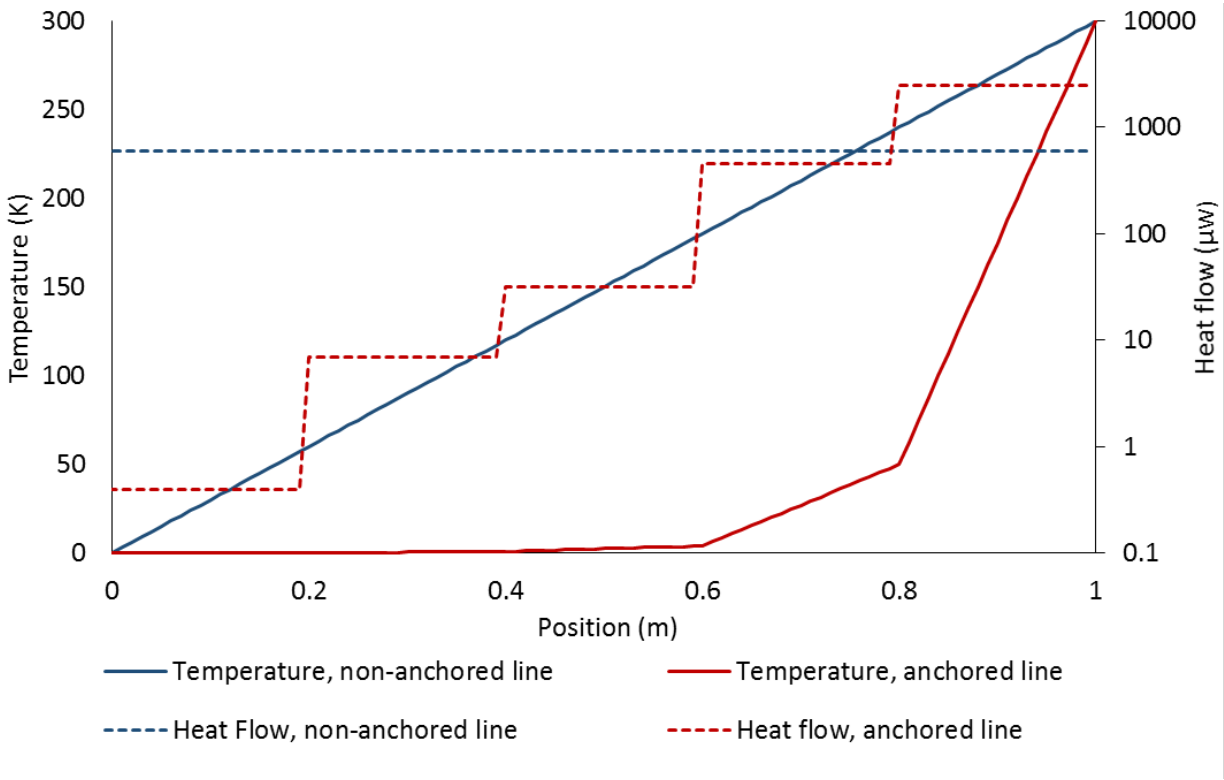

Figure 6.18: Hypothetical heat flow from a single 40 AWG copper wire, with and without thermal anchoring. Total length from room temperature to the MXC is about 1 m. Solid lines are temperature (linear) vs position. Dashed lines are power (log scale) vs position. The anchored line (red) conducts more heat into the fridge than the non-anchored line (blue) due to the greater thermal gradient near room temperature. However, the anchored line conducts progressively less power to each cooler stage, ultimately conducting about 1000 times less heat power to the MXC than the non-anchored line.

fraction of the 50-K stage cooling power (≈1 W), resulting in a negligible temperature rise on that stage. By anchoring to this dramatically lower temperature (50 K vs. 240 K for the the non-anchored line at this distance), heat flow is increased to the first stage but reduced to the next colder 4-K stage, and so on with each colder stage. Total heat transmitted to the MXC is reduced three orders of magnitude between the non-anchored and optimally anchored lines.

6.3.2 RF Cabling Design

The RF wiring diagram of the dilution fridge used for measuring MSA performance is shown in figure 6.19.

We can start analysis of this setup with the vector network analyzer (VNA) at the top

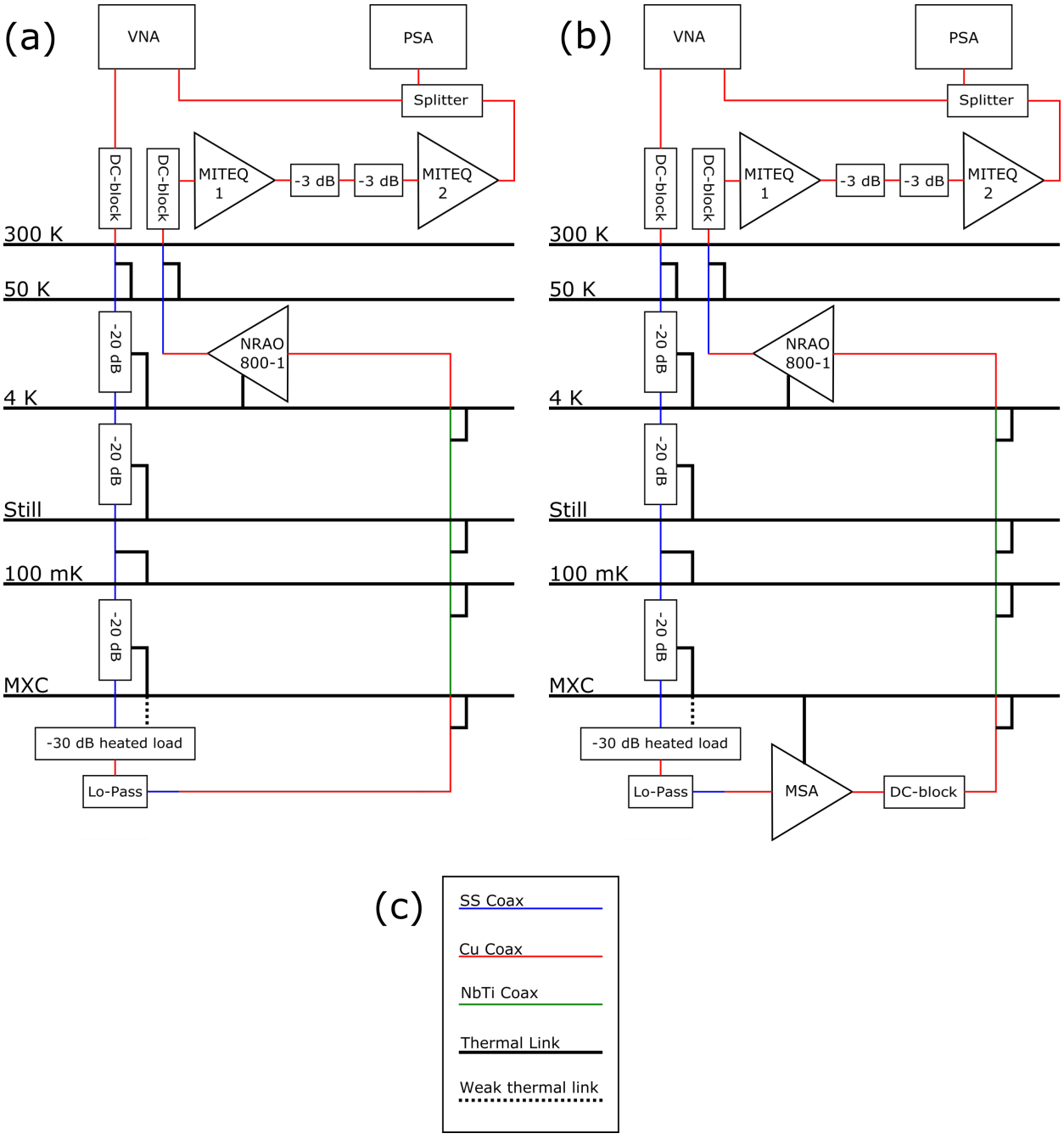

Figure 6.19: Dilution fridge RF wiring. (a) Wiring for baseline measurement. (b) Wiring for MSA characterization. (c) Wiring legend.

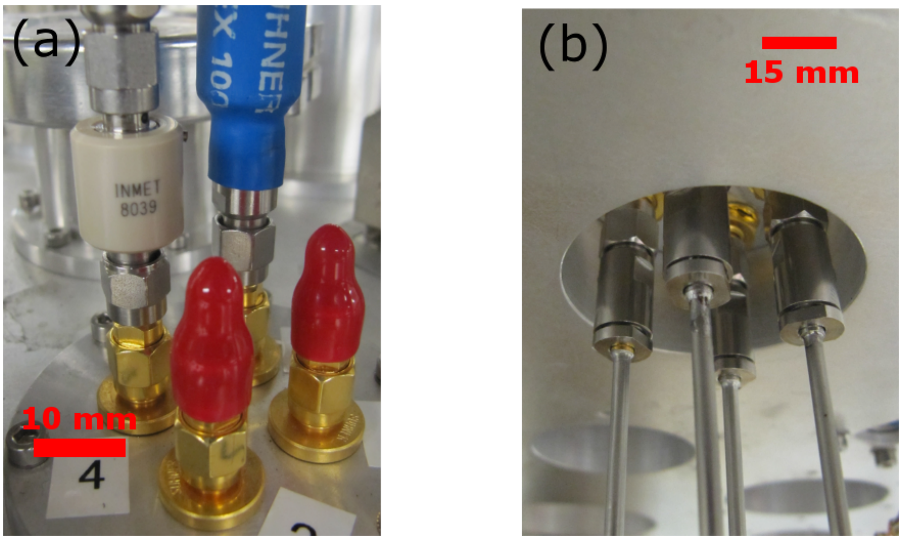

Figure 6.20: RF signal entering the fridge at the RT plate. (a) Top side with DC-DC block. (b) Bottom side inside vacuum space with 3.5 mm connectors

left of 6.19a. A Cu coax line runs about 3 m from the equipment rack to the top of the dilution fridge and an "inner-outer" DC block is placed in series where the signal enters the fridge. The DC block may not be critical, but is used for an abundance of caution to avoid possible 60-Hz ground loops. The DC-DC block (API Technologies P/N Inmet 8039), shown in figure 6.20a, has proven reliable for both room temperature and cryogenic applications down to 10 mK, despite its plastic construction. The signal passes though a hermetic connector to the vacuum space inside the fridge, carried on a SS-SS (stainless steel inner conductor and outer jacket) coax terminated with a 3.5-mm connector (part #200-36-20-850 by SGMC microwave). The 3.5-mm microwave cable terminations are inter-operable with SMA connectors, but more robust and bulky. For this reason, 3.5-mm cable connections are used inside the fridge wherever space allows.

The signal continues to a -20 dB attenuator (part #2782-6051-20 by XMA Omni Spectra) thermally anchored to the 4-K stage. This attenuator (and sister models with different attenuation values) is rated for 4-K operation, and has proven reliable to 10 mK. Though these attenuators exhibit excellent cryogenic microwave performance, the printing on the outer body is of very low quality and is easily removed during handling. For this reason, best practice includes secondary labeling with permanent ink or an engraver. Figure 6.21 shows two arrays of four input lines with -20 dB attenuators thermally anchored to their respective stages by a copper bracket. Wherever possible, thermal clamps are made from copper and assembled with brass screws. Brass has a greater coefficient of thermal expansion, so an assembly that is tight at room temperature will become slightly tighter at cryogenic temperatures. Where brass is not practical, SS hardware is acceptable.

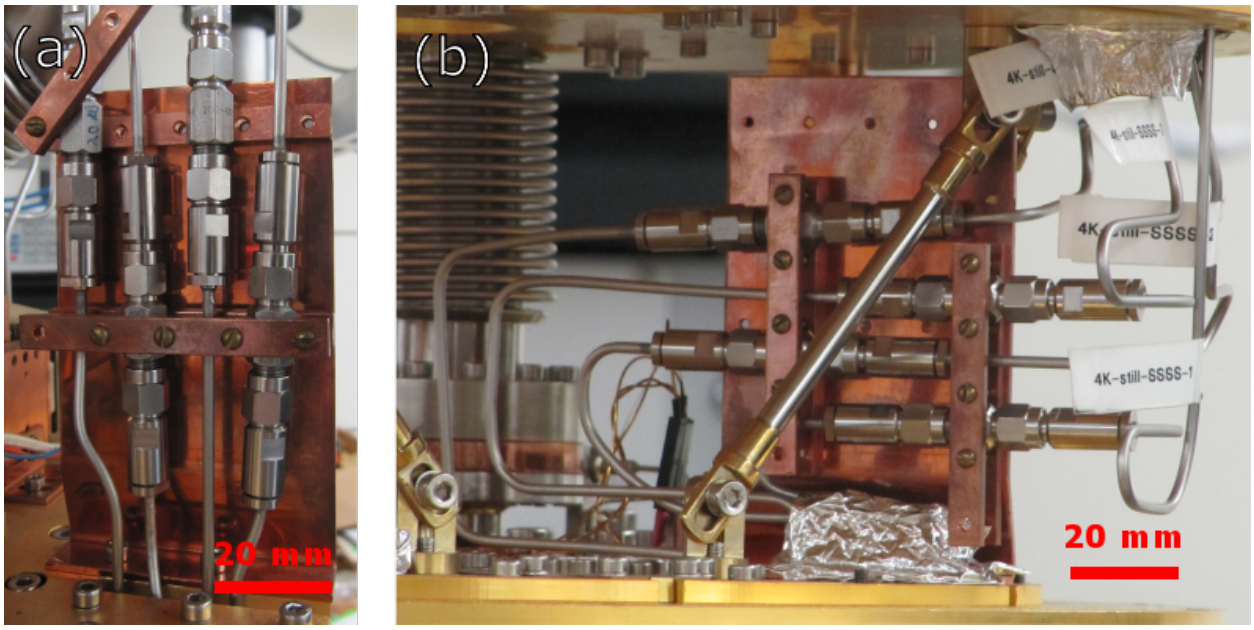

Figure 6.21: -20 dB attenuators on the dilution fridge. Though arrays of four attenuators are shown, only one input line is used in the work described here. (a) 4-K stage. The upper thermal anchoring clamp is partially removed to reveal the location of the -20 dB attenuators. (b) Still stage. Note the serpentine paths of the SS-SS coax lines, preferred for connections between temperature stages. 3.5-mm connectors are used throughout.

Between each temperature stage, the SS-SS coax is bent into a serpentine path, as illustrated in figure 6.21b. This serves three purposes: strain relief accounting for small thermal dimensional changes in the fridge, to increase the cable length diminishing thermal conductivity, and to diminish fiber-optic transmission of infrared light down the coax lines. On this last point, note that 50-K blackbody radiation has a peak intensity at about 1 THz, or about 288 μm in free space. Though this is well above the conventional TE01 cutoff frequency of RG 405 coax (around 61 GHz), the wavelength of 1 THz radiation in the PTFE dielectric is about 200 μm, small enough to propagate virtually loss-free[116] as a fiber-optic mode in the 600 μm thick PTFE between the center and outer conductors.

The outer conductor of the coax is thermally anchored by mechanical clamping, but because the dielectric (PTFE) has poor thermal conductivity, the center conductor remains thermally isolated. Since a -20 dB attenuator puts the center conductor of the coax in dissipative contact with the outer shell, it is an effective means to thermalize the center conductor. If attenuators are not used on a particular stage, it is still important to mechanically clamp the coax outer shell. Since the cross-sectional area of the outer conductor is about eight times the area of the inner conductor, clamping just the outer shell will accomplish most of the desired thermal anchoring. Figure 6.22 shows a thermal anchor on the outer

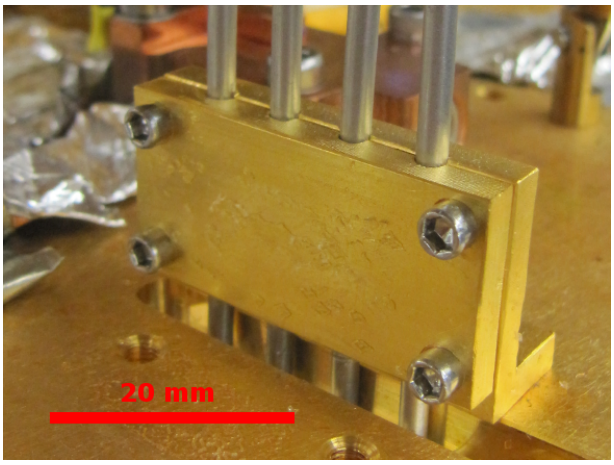

Figure 6.22: Thermal clamping of the coax outer conductor. Components in this figure are Cu coated with Au. Best practice is to gold-plate all thermal components to minimize corrosion and oxidation.

conductor of four SS coax lines.

A final -20 dB attenuator anchored to the mixing chamber stage (MXC), as shown in figure 6.23, finalizes the input path from RT lab equipment to the cold experimental space. Note that the attenuators on the upper stages are clamped by a copper bar with a brass screw on either side of each attenuator, ensuring good mechanical and thermal contact to each, (figure 6.21) but space constraints at the MXC preclude individual clamping screws for each attenuator (figure 6.23). Two methods are used to ensure good thermal contact to each attenuator on the MXC. First, a thick steel bar is used to back-up the copper bar to reduce bowing (figure 6.23a). Second, a short length of gold wire (about 1.5-mm dia.) is placed in a well in the copper bar adjacent to each attenuator. As the clamps are tightened, the gold wire is deformed to a circle about 3-mm dia. (visible in figure 6.23b), ensuring good thermal contact to each attenuator despite any small deviations from flatness in the copper bar.

The series of input attenuators is critical not just to reduce the heat load on the MXC, but to cool the RF blackbody radiation that reaches the MSA–without attenuation on the input line, the signal input to the experiment would include Johnson noise at a temperature of 300 K. In general, the Johnson noise after an attenuator at stage N is given by

$$T_{signal_N} = T_{signal_{N-1}} \times \text{trans}_N + T_{physical_N} \times (1 - \text{trans}_N) \tag{6.3}$$

where $T_{signal_{N-1}}$ is the signal temperature at the previous stage, trans_N is the transmission coefficient of the attenuator (ex: a 10 dB attenuator has a transmission cooeficient of 0.1), and $T_{physical_N}$ is the the temperature at the current stage. There is an obvious lesson in equation 6.3 that is sometimes not obvious to the neophyte cryogenic microwave engineer:

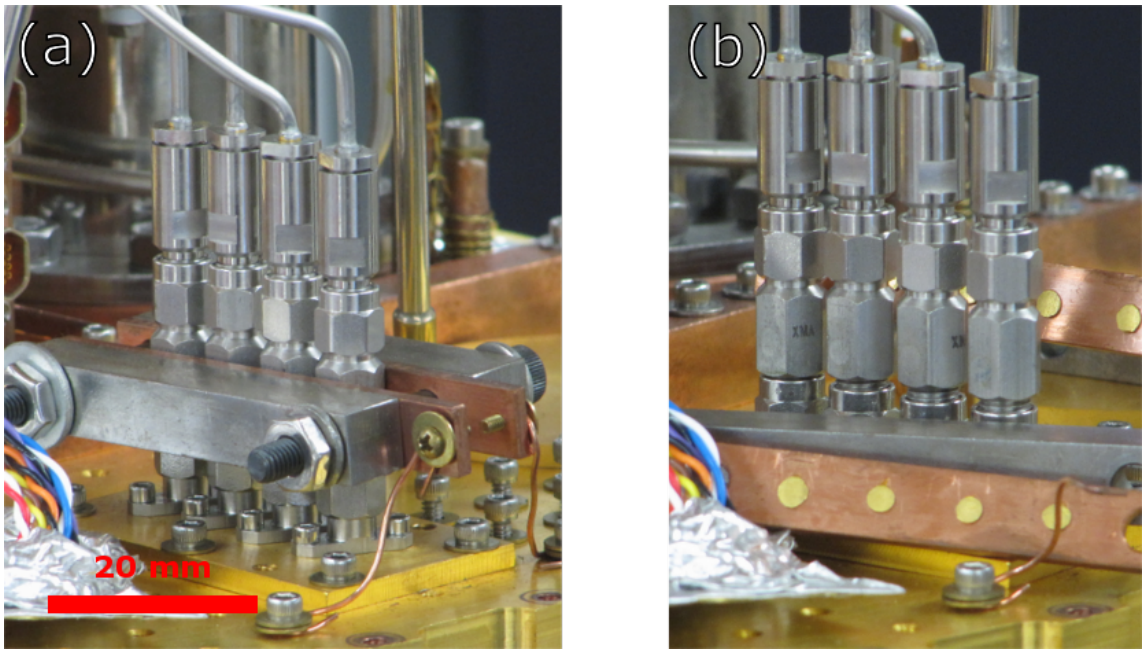

Figure 6.23: Final -20 dB input attenuators on the MXC. (a) Assembled for use. (b) Disassembled to show construction.

Table 6.5: Thermalization of the input signal.

stage	N	$T_{physical}$ (K)	att. (dB)	trans. (abs)	T_{signal} (K)
300 K	0	300.0000	0	1	300.0000
50 K	1	50.0000	0	1	300.0000
4 K	2	4.0000	-20	0.01	6.9600
Still	3	0.8000	-20	0.01	0.8616
100 mK	4	0.1000	0	1	0.8616
MXC	5	0.0500	-20	0.01	0.0581
heated load	6	0.5000	-30	0.001	0.4996

every component radiates Johnson noise, and that temperature is a linear combination of its own temperature and the temperature of outside signals transmitted or reflected from it.

An example of input line signal thermalization is given in table 6.5 with attenuators arranged as in figure 6.19 and typical physical temperatures. The heated load is the last section of the input line before reaching the MSA. If no heater power is applied to the heated load, its temperature approaches that of the MXC and the input signal temperature will be very close to that of the MXC. Use and design of the heated load are discussed further in section 6.4.

After the input signal has interacted with the experiment (typically an MSA) the signal

must return to room temperature. While input signals can be heavily attenuated and the loss made up by increasing the amplitude of applied probe signals, signals returning from the experiment are typically rather small, and lossy transmission lines cannot be used. For this reason, superconducting NbTi coax is used to deliver the signal from the MXC to the 4-K stage. Microwave losses in NbTi coax are very small, typically 0.3 dB/m, which over the 0.4 m from the MXC to the 4-K plate is negligible relative to typical connector reflection losses.

At the time of this writing, there are only two commercial sources of NbTi coax: "Coax Co, Ltd" and "Keycom". Our system uses cable made by Keycom, a decision made at the time based on immediate availability. Having had the opportunity to work with cable from both manufacturers, we recommend exclusive use of Coax Co cables. There are three differences between coax from these two sources: finish, dielectric quality, and bend radius. The finish of the Keycom cable is rather rough as shown in figure 6.24. The rough finish reduces the thermal conductivity of any external clamp. The surface finish is consistent with a highly worked and work-hardened metal, and may relate to the bend radius–Keycom specifies a minimum bend radius of 0.1 m, compared to 10 mm specified for the equivalent product by Coax Co. Such a large bend radius precludes forming to a serpentine path as in figure 6.21b and as a consequence, the benefits of strain relief and reduced optical transmission are lost. The dielectric of the NbTi cables from both manufacturers is PTFE, commercially known as teflon. The PTFE in cables by Keycom is opaque white and stringy, strongly reminiscent of teflon tape commonly used for assembling threaded pipe. This leads to difficulty in cutting the cable and assembling terminations, as pulling an incompletely cut "thread" of teflon will tend to pull more material out of the cable, emptying the cable of dielectric. Complete severing of the PTFE "threads" must be accomplished before mechanical removal, which is difficult to accomplish without also cutting into the center conductor. The PTFE in cables by Coax Co is vitreous and continuous, and appears similar to the PTFE found in most other conventional microwave lines (SS, Cu, etc). When prepping this cable for termination, it is sufficient to cut the dielectric about 2/3 the way through and twist off the remainder, resulting in a clean boundary and no chance of damaging the center conductor.

The NbTi superconducting lines cannot be soldered, so it is necessary to use crimp-on connectors. Unlike soldering, brazing, etc., crimping is a mechanical process that is vulnerable to failure due to differential thermal contraction as the assembly is temperature cycled between mK temperatures and room temperature. Through trial and error, we have found one crimp-on SMA connector (TE Connectivity P/N 1050611-1) that performs without fail. It is important to use the correct specialized crimping tool (Omni-Spectra P/N 2598-5200-54)[12] and follow the manufacturer's assembly instructions exactly, but for one point: we have found that adding about 0.020" to the center conductor length (0.090" rather than 0.070") before crimping leads to more reliable operation for cryogenically temperature-cycled applications.

[12]This tool costs about $3k-4k new and can be found for about $1k used. Either is a good price for the time and frustration saved.

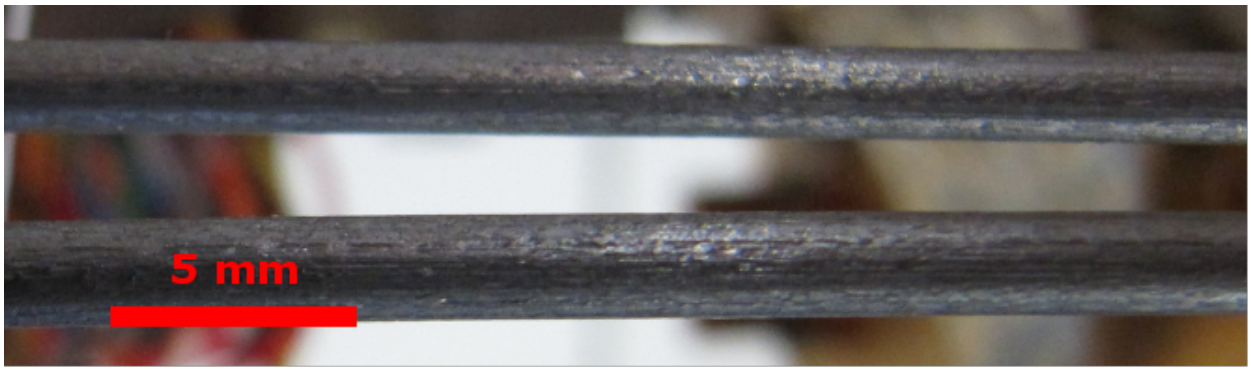

Figure 6.24: Surface texture of 0.085″ NbTi coax by Keycom.

6.3.3 DC Wiring Design

DC lines are used for MSA current and flux bias, SQUID I-V characterization, MSA varactor tuning, heaters and thermometry, and control of various cryogenic appliances (ex: RF crossover switch). These lines carry signals from room temperature to the dillution stage at some 10's of mK, and a great number of DC lines could be needed, so we are concerned with high-density, low thermal conductivity, high electrical conductivity, low dissipation (for high-current loads), and reliable operation. Typically, commercial-off-the-shelf (COTS) solutions are not suitable due to some combination of size, high thermal conductivity, and connectors unreliable for cryogenic cycling, so custom cabling is necessary.

We found a good solution is the use of 24-wire twisted-pair (12 pairs) Cryoloom® with micro-D connectors. Figure 6.25 shows a detail of the cryoloom. For connections between components on the same temperature stage, (where low thermal conductivity is not a requirement) we used Cu wire loom due to low cost and easy workability, and for connections between temperature stages (where low thermal conductivity is important) we used either resistance wire or CuNi-clad NbTi wire. For low current applications which do not present significant joule heating in the cabling (for example, the μA-level currents used for biasing the MSA) resistance wire is preferable due to its temperature stability, and for high-current applications, superconducting wire is preferable due to its low resistance. The CuNi cladding on superconducting wire allows for easy soldering without adding significantly to the thermal conductivity.

Without a shield and mechanical protection, the fine wires of the loom would be susceptible to damage and electrical interference. An initial design for the DC cables accounting for this is shown in figures 6.26 and 6.27. In this design, the cryoloom is soldered to the micro-D connector pins, the soldered joints are potted in Masterbond® EP51 epoxy, and the potted connection covered in a metal backshell. Exiting the backshell, the loom is threaded though a PTFE tube for bending strain relief and protection against abrasion from the outer braid. The tinned Cu braid assembled over the tube serves both as a Faraday shield and tensile strain relief.

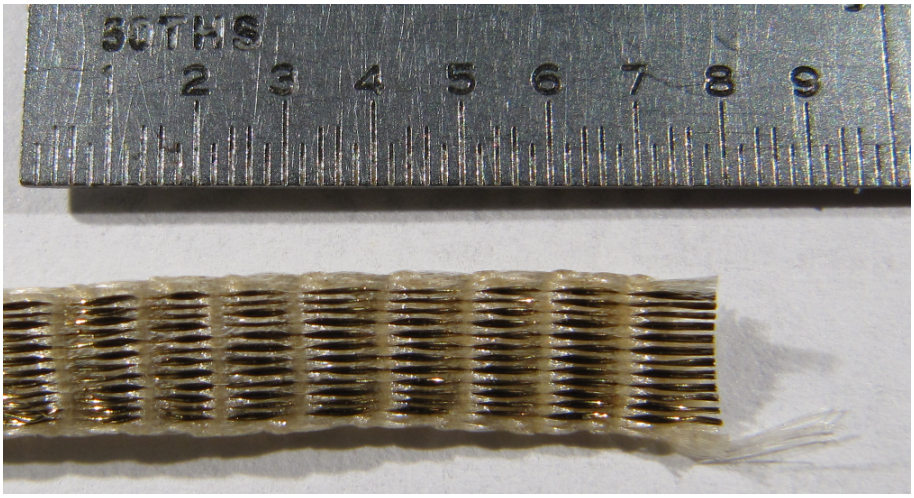

Figure 6.25: Cryoloom used in DC wiring. The ruler included for scale has smallest divisions of $1/50^{th}$ of an inch. The loom has 12 twisted pairs (24 conductors) insulated with HML, interwoven woven with PEEK fibers for strength.

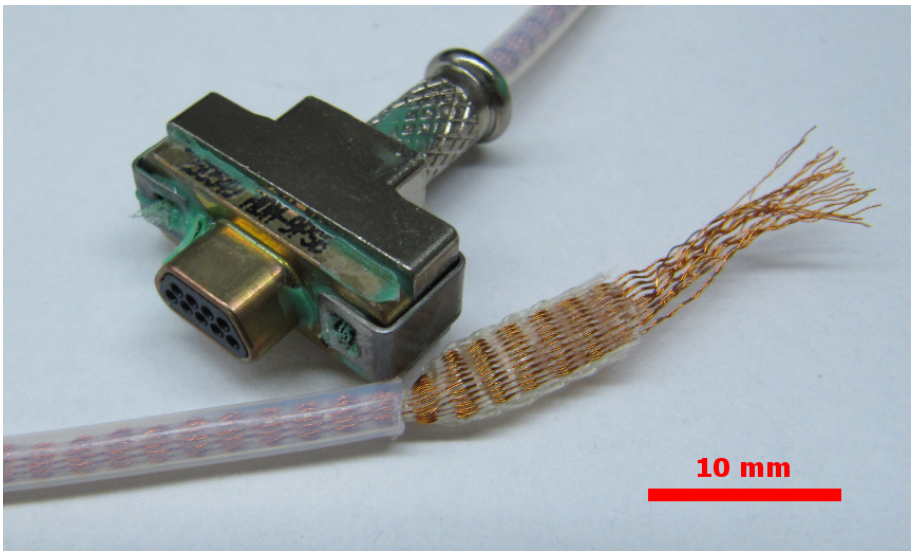

Figure 6.26: DC cable with potted backshell, inner construction. Residue of the EP51 epoxy (green) remains visible around the seams. Once constructed, the design is mechanically robust, but does not thermalize the conductors.

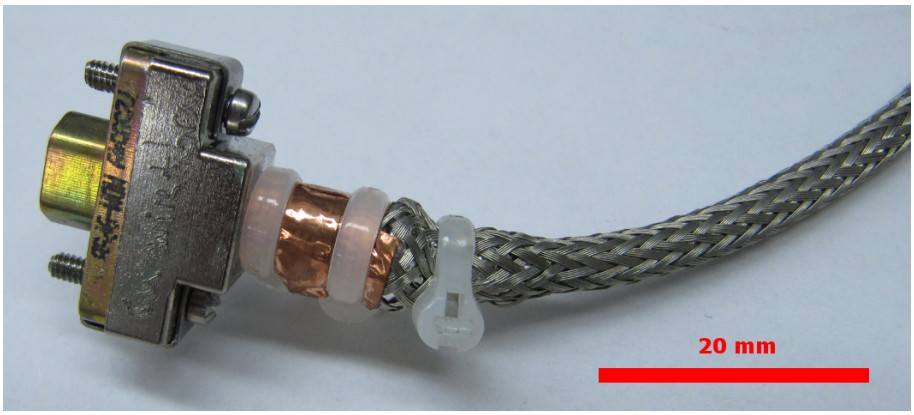

Figure 6.27: DC cable with potted backshell, final construction. The braid is mechanically and electrically connected to the backshell with Cu tape and zip ties. The final assembly does not allow for any tight bends near the connector plane, complicating cable routing inside the cramped space of the dillution fridge.

This design is very robust, but suffers from several drawbacks. The construction is fairly laborious and requires not insignificant skill to avoid wire breakage during the potting and assembling process. The mechanical tolerances on the loom vs. braid length are rather tight to achieve both effective strain relief and flexibility. The connector length is rather long (unable to make a tight bend near the connector plane). If the cable is to be used between thermal stages, the outer Cu braid must be interrupted to avoid a thermal short while maintaining electrical continuity (a solvable, but non-trivial problem), and perhaps most critically, this design affords almost no thermalization pathway for a superconducting loom, so the conductors may not cool below T_c on a practical timescale (in some cases, many days after the ridge has reached base temperature).

The drawbacks of the original design are largely addressed with the alternate design shown in figure 6.28. In this design, a tin-plated Cu adhesive tape (3M EMI Embossed Tin-Plated Copper Shielding Tape 1345) is used as backshell and cable shield. The solder pins are insulted from the outer Cu tape with a few wraps of PTFE tape, and an "S" bend in the loom near the connector prevents any tension from the cable transferring to the solder joints. Any tension on the cable is instead transferred to the connector body trough the Cu tape. In figure 6.28 the loom is shown uncovered, but in practice the entire cable would be covered with adhesive Cu tape.

A tinned Cu tape as shield is superior to the braid because it is continuous with no gaps, but remains flexible. It also increases the capacitance to ground ($\approx$ 130 pF/m wire-to-ground and 130 pF/m wire-to-wire, twisted pair) which is desirable for low-frequency signals. Most importantly, the Cu tape in close contact with the wire loom along its whole length assures efficient thermalization of the superconducting wire. In practice, cables of this design were observed to undergo a superconducting transition simultaneously with the relevant fridge

Figure 6.28: DC cable with tape backshell and shield.

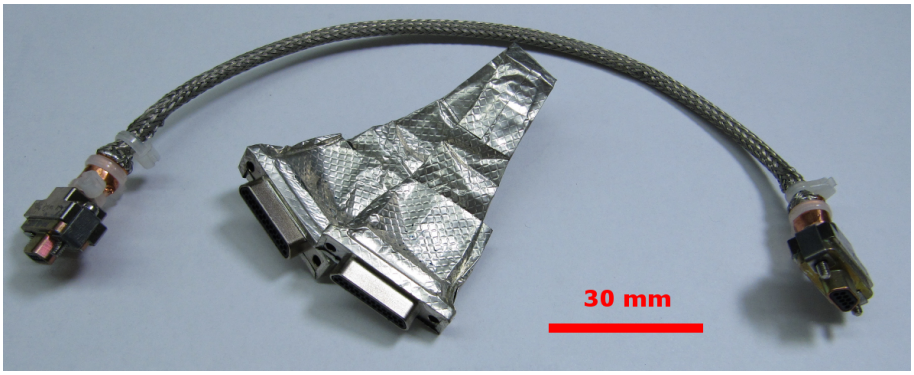

Figure 6.29: DC cable comparison. The conventional design (long arc) is certainly better looking, and also more mechanically robust. The "flat tape" design (center) is not beautiful, but is easier to manufacture, easier to work with, and is more reliable in practice.

stage passing through the critical temperature. Of course, it is not permissible to maintain a continuous length of Cu tape between thermal stages of the fridge. When a cable must traverse thermal stages, about 60 mm of Cu tape is replaced with a section of SS tape or Al tape. At the lower temperature stages (for example, between the MXC and still) Al tape is used because it is easy to work with and becomes a thermal break at temperatures below the aluminum T_c (about 1.2 K), while between higher temperature stages (for example, between the 50-K and 4-K plates) SS tape is used due to its consistently low thermal conductivity. SS tape tends to form razor-sharp edges, so the order of assembly must be chosen such that the SS tape does not cut into the wiring loom. This design remains flexible near the connector, so cable routing issues are less severe relative to the "braid and backshell" design. A final advantage of the "flat tape" design is ease of thermalization: clamping between any two flat features in the dilution fridge is sufficient to thermalize the cable. Figure 6.29 shows both cable types together.

6.4 Design of the Heated Load for Y-factor T_N Measurement

A critical performance specification of the MSA is its noise temperature T_N. To measure T_N by the Y-factor method, a known noise power is applied to the amplifier input and the output noise power recorded. A noiseless amplifier will present an output noise power strictly proportional to the input noise power, but any real amplifier will present both an amplified input power and a constant extra power–namely its own T_N. By measuring output noise power for two or more input noise powers and fitting the slope and intercept, both the gain and the "extra noise power" ie T_N can be determined. Figure 6.30 shows an example Y-factor measurement demonstrating a total system noise T_{sys} of 327 mK. A nice thing about this method is that the noise temperature T_N can be determined without any need to calibrate the amplifier gain beforehand. The concept of a Y-factor measurement is straightforward, but depends on a well-calibrated noise source. The noise source can be generated synthetically, which can be convenient, but in our application using a calibrated, room temperature noise source on our cryogenic amplifier would require exquisite characterization of the transmission lines running from room temperature to the MXC, and could potentially introduce large systematic errors. An alternate approach is to use a blackbody radiator as a noise source. If one can ensure an emissivity very close to one[13], the transmitted power is known as accurately and easily as one can measure the temperature. Though it means placing a physically hot object in the fridge we have worked so hard to keep cool, the "heated load" is our preferred calibrated noise source.

It is often desirable to send either noise or a narrow-band signal to the amplifier input, which can be accomplished by using an RF switch between inputs, or by way of a 3-port directional coupler as described in [103] and shown in figure 6.31. In this arrangement, the 50-Ω termination is heated, and the directional coupler serves to strongly couple the hot load and weakly couple a test input tone to the amplifier. We have found that 50-Ω terminators are unreliable with temperature cycling between room temperature and mK temperatures, often failing to an open circuit when cold, but appearing normal when warmed. The 3-port directional coupler is inherently a 4-port device with one port terminated by a 50-Ω cap, so this arrangement relies on two of the potentially faulty components. In contrast, we have never experienced an attenuator failing due to temperature cycling, so our heated load design is simply a temperature-controlled -30 dB attenuator, as indicated in figure 6.19. Any desired input signal can couple to the MSA through the -30 dB loss of the attenuator, and the heated attenuator has an emissivity of 0.999.

Figure 6.32 shows the heated load we used for Y-factor measurements, with the major components color-coded. A test tone can be coupled in through the gold SMA connector at the top-right, and the signal out to the MSA is coupled through the steel SMA connector at the bottom right. Red indicates the -30 dB attenuator, heater and thermometer. Blue indicates the coax lines coupling the attenuator to the signal input and output. Green

[13] A 50-Ω termination on a 50-Ω transmission line has an emissivity of 1, by definition.

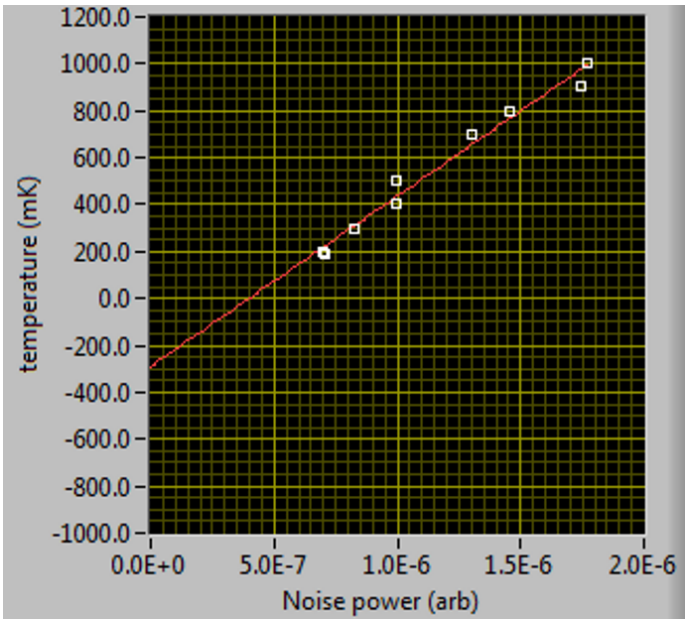

Figure 6.30: Example Y-factor measurement data. The X-axis is the noise power output at the end of the amplifier chain in arbitrary units. The Y-axis is the known temperature of a heated load applied to the first amplifier's input. White open squares are data points, and the red line is a linear regression. Note that extrapolating to zero input temperature the noise power output is non-zero due to finite noise power added by the amplifier chain. The input temperature necessary to bring the output power to zero (y-intercept) is negative, equal and opposite to the total system noise temperature T_{sys}. The noise temperature of the first amplifier T_N is bounded above by T_{sys}, and in a well-designed amplifier chain, T_{sys} will be dominated by T_N of the first amplifier, with only small additional contributions from the rest of the amplifier chain. In this example $T_{sys} = 327$ mK.

indicates copper wire used as a thermal shunt. Yellow indicates a "roach filter" lossy at high out-of-band frequencies (typically > 10 GHz). Each part of this design is informed by thermal, DC, and RF signal considerations.

6.4.1 Thermal Considerations

The heat conducted from the hot element to the MXC must not be too great, and the thermal time constant of the heated load must not be too long. The MXC is maintained at typically 50 to 100 mK, with a cooling power of about 50 to 200 μW, and one might desire to raise the temperature of the heated load to several K. To avoid exceeding the cooling power of the fridge, it is important the physical connection between the heated load and MXC have a fairly low thermal conductance, not greater than about 10 μW/K. On the other hand, some finite thermal conductance is necessary because a typical experiment involves

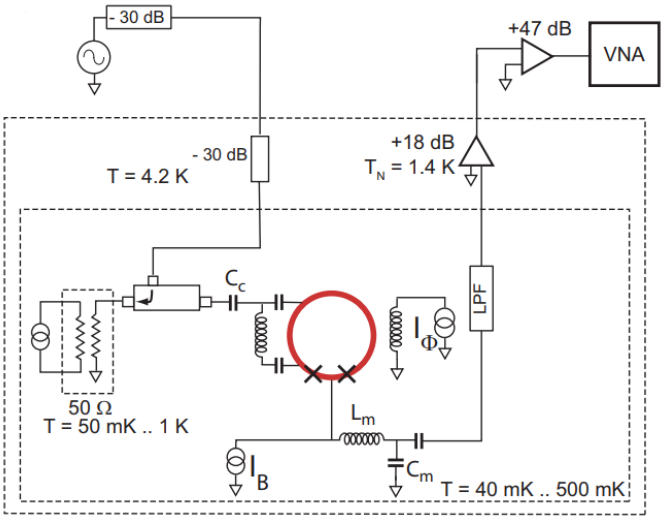

Figure 6.31: Heated load made from a 50 Ω terminator and directional coupler. Reproduced from [103]

both raising and lowering the temperature of the heated load, and while the temperature may be raised by resistive heating, it can only cool by thermal contact with the MXC. In practice, the thermal link is provided by the input and output microwave lines, colored blue in figure 6.4b. These coax lines are about 0.1 m long, with a cross sectional area of 1.78 mm^2 of SS (counting both the center and outer conductor), and 2 mm^2 of PTFE dielectric. Consulting figure 6.17, we see that the thermal conductivity k of SS is about 10 times greater than PTFE, so the PTFE can be safely neglected in the calculation of total k. The thermal conductivity of SS is a function of temperature, but we can take a value of about 0.2 W/(m·K) as typical at the temperatures of interest, for a thermal conductance of $0.2\ \mathrm{W/(m \cdot K)} \times 1.78\ \mathrm{mm^2/0.1m} \approx 3.5\ \mu\mathrm{W/K}$, within our budget of 10 μW/K. The SMA microwave connectors are electrically reliable, but *cannot* be relied on to provide a consistent mechanical and thermal coupling. To ensure consistent and predictable thermal coupling to the coax lines, short lengths of bare 0.84 mm copper wire (colored green in figure 6.4b) are tightly wound and then soldered to the outer SS coax, with the other end mechanically clamped with a brass screw to a Cu component.

The thermal time constant is determined by the quotient of the thermal mass and thermal conductivity, favoring a high thermal conductance and low thermal mass for faster temperature excursions. Since fridge cooling power sets an upper bound on the thermal conductance, it is necessary to minimize the mass of the heated load components colored in red in figure 6.4b. The red components in figure 6.4b are: the -30 dB attenuator, a copper clamp made to fit the attenuator, a heater assembly, and a thermometer assembly. The heater assembly and thermometer assembly bolt to the clamp with brass screws. The heater assembly in figure 6.32 is several meters of resistance wire wound on a 2.1-mm diameter copper post,

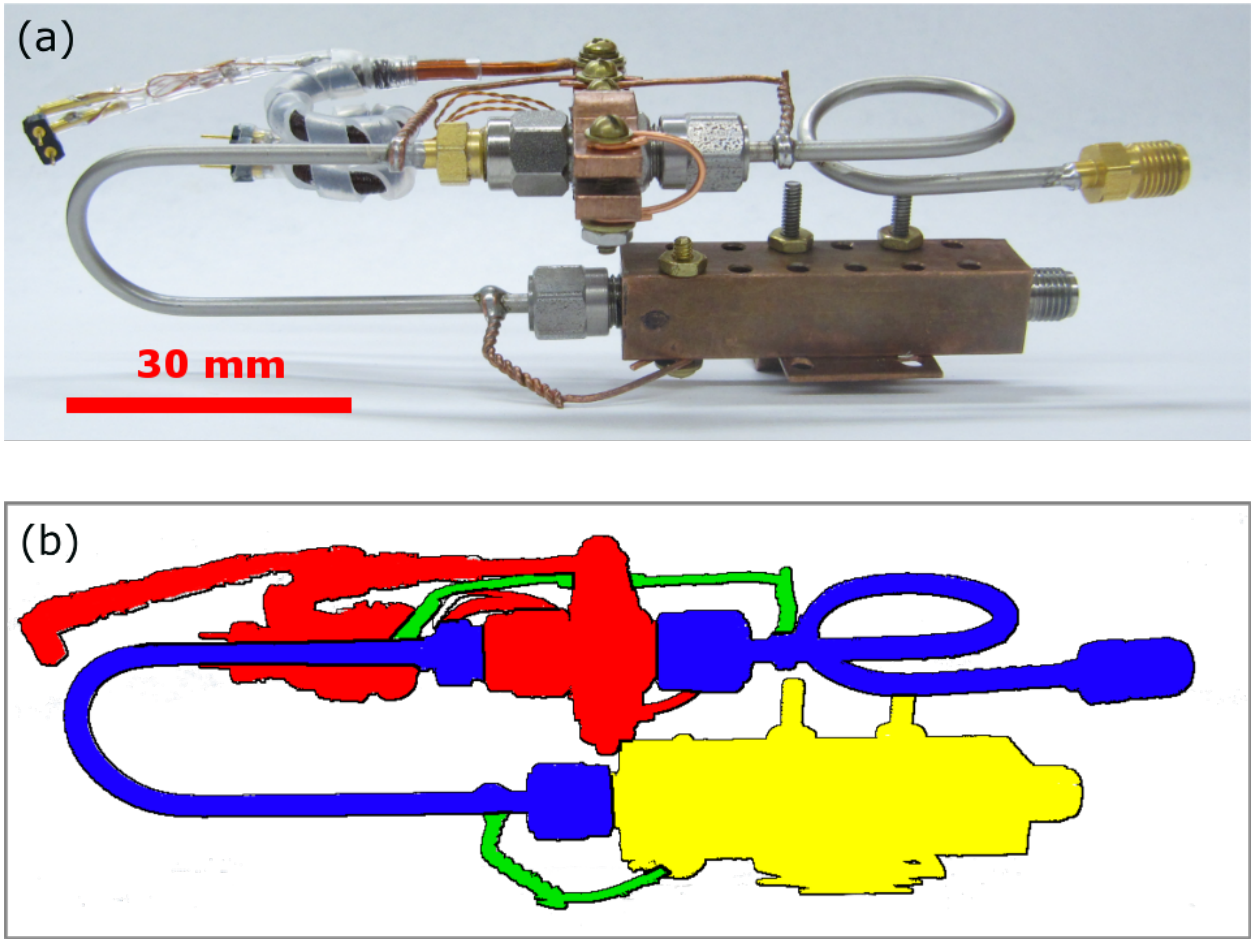

Figure 6.32: Heated load used for Y-factor measurements. (a) Photo of heated load. (b) Color-coded diagram of heated load. Red: -30dB attenuator, heater, and thermometer. Blue: SS-SS coax lines for carrying signals in and out, also serving as a weak thermal link. Green: Thermal shorts for stable thermal conductivity. Yellow: the "roach" filter for thermalizing high-frequency out-of-band noise.

and the thermometer assembly is a packaged RTD (P/N RX-AA 102A from Lakeshore) with additional copper blocks for clamping. The heater and thermometer assemblies proved too massive, and the long length of tightly wound resistance wire was prone to short-circuit, so a smaller, lighter, more reliable assembly was devised, as described presently.

To assemble an improved thermometer assembly, a short length of 2.1-mm diameter copper wire was flattened at the end and the insulation removed (fig. 6.33a). A new insulating layer is made on one side of the flat section by adhering a patch of cigarette paper saturated in GE varnish (fig. 6.33b). The ruthenium oxide RTD (P/N RX-BR 102A from Lakeshore) is affixed to the insulating layer with a dab of GE varnish (fig. 6.33c). The electrical leads are attached with conducting epoxy (fig. 6.33d) as soldering temperatures would damage the RTD. Final assembly (not shown) involves placing a short length of plastic tubing over the end assembly to avoid damage during handling, winding the fine copper leads around the 2.1-mm post and glazing with GE varnish. The far end of the 2.1mm post (not pictured) was curled into a circle, flattened, stripped of insulation, and affixed to the attenuator clamp with a 2-56 brass screw and washer. The final thermometer assembly has a total mass of about 1.5 g including mounting hardware, less than half the mass of the packaged sensor without mounting hardware.

An improved heater assembly was assembled by a very similar method. Instead of a single RTD, four-2.5 kOhm resistors (P/N TNPU06032K49BZEN00 by Vishay Dale) were wired in series and affixed with GE varnish. This particular part number would be a good starting point for anyone wishing to design their own heated load, but we hesitate to guarantee its suitability.[14] The electrical connections to the heater resistors were made with ordinary Pb-Sn solder, which is more reliable than epoxy.

6.4.2 DC Wiring Considerations and Phantom Power

Precise measurement of the RTD resistance is necessary for accurate temperature readings, so 4-wire lock-in method is used with a Lakeshore 372 temperature conroller. A typical excitation power is a few pW and the frequency is 13.7 Hz. The two copper wires visible in figure 6.33 transition to four low-thermal conductivity, high resistance wires at the first connector, a few cm from the RTD. Typical resistance of the RTD is a few 10's of kΩ, and temperature sensitivity is about 100,000 Ω/K so the added series resistance of a few mΩ due to the short 2-wire configuration is negligible.

The 10-kΩ resistance heating element is driven by a DC current generated by the Lakeshore 372, in a simple 2-wire configuration. Initial tests indicated a surprisingly large "phantom" power, evidenced by a floor temperature about 220 mK higher than the MXC temperature with nominally 0 W applied to the heater. The "phantom power" responsible for this ele-

[14]Our experience has shown that every manufactured batch of resistors must be independently verified for reliable cryogenic performance. The resistor referenced here is rated by the manufacturer to $-55°$ C (218 K) and performs well to our lowest tested temperatures around 50 mK. However, prior or subsequent batches made by the manufacturer may meet spec to the $-55°$ C rating but could very well use materials or manufacturing methods unsuited to lower temperatures.

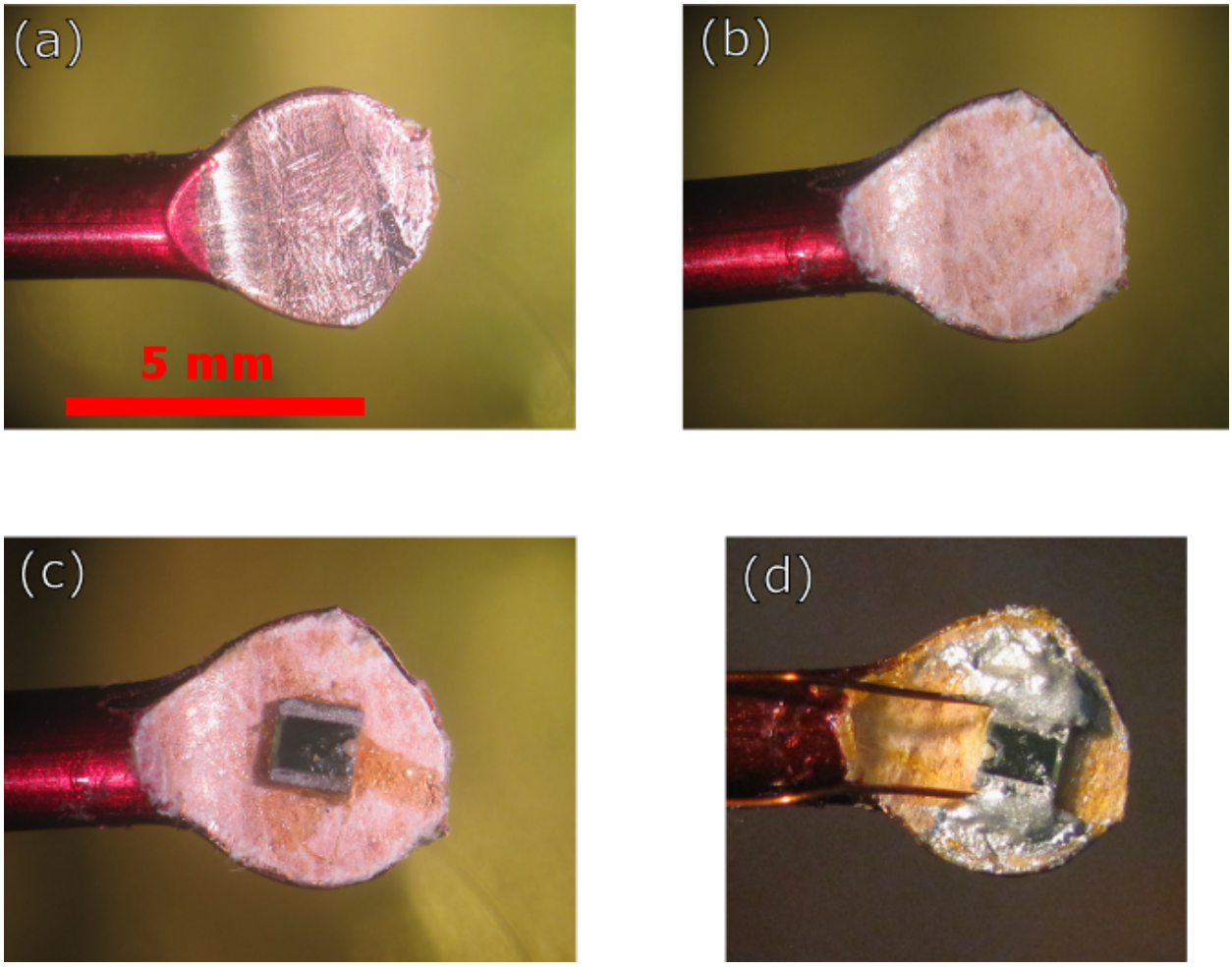

Figure 6.33: Low-mass temperature sensor assembly. (a) End of wire flattened. (b) Cigarette paper with GE varnish applied for electrical insulation. (c) Chip thermistor attached with GE varnish. (d) Cu Wire added with conductive epoxy.

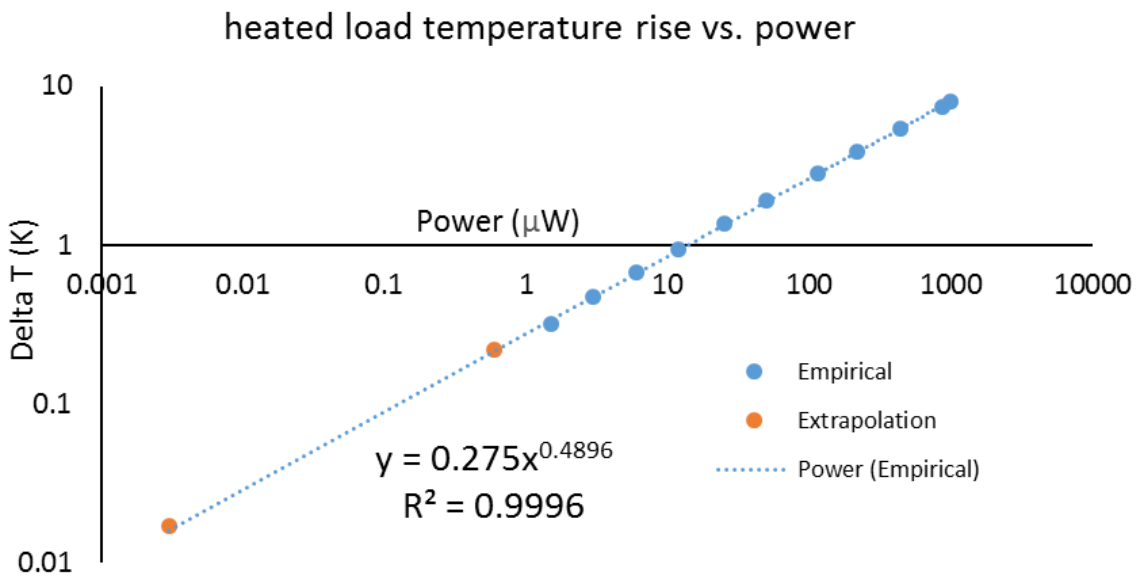

Figure 6.34: Empirical power law of heated load delta-T vs. applied power. Blue points are empirical data. Blue line is a power law fit to the empirical data. Orange points are a measured temperature rise with 0 W applied in two different experimental configurations, with the "phantom power" (x-coordinate) in each case extrapolated by fit to the blue power law function.

vated temperature could not be measured directly, but can be estimated from the remarkably good empirical power-law fit of heated load temperature rise vs. applied power at heater powers ranging from 1.5 μW to 1 mW. This empirical power-law relation is shown in blue in figure 6.34.

Surprisingly, a temperature rise of 220 mK, corresponding to about 0.6 μW of excess power, was observed even with the Lakeshore heater power set to "0". Even more surprisingly, disconnecting the Lakshore heater leads (replacing the heater power source with a open circuit) resulted in a *greater* temperature rise, and shorting the room temperature cable (replacing the heater power source with short circuit) *also* resulted in an increased temperature rise, suggesting that the origin of the excess heater power was some combination of AC E-field and B-field coupling to the heater power leads somewhere between the heater and room-temperature.[15] To test this "AC pickup" hypothesis, we disconnected the

[15]Short-circuit excess power suggests AC B-field coupling, and open-circuit excess power suggests AC E-field coupling. Reduced heating with the Lakeshore 372 connected suggests the coupled interference is absorbed by both the Lakshore's output circuitry *and* the cryogenic heater.

heater leads at the MXC, and observed no discernible temperature rise at the heated load, eliminating any other "phantom power" source, including thermometer mis-calibration or real joule heating due to thermometer excitation energy.

Given the strong circumstantial evidence of AC-coupled "phantom power", and understanding that the Lakeshore heater signal is DC, we applied a capacitive shunt to the heater leads on the cryogenic stage near the heated load. This shunt is made from a parallel arrangement of two 0.47 μF capacitors [16] and two 470 pF capacitors[17]. The -3dB knee due to the 0.47 μF capacitors and 10k heater resistance is 16 Hz, and the -3 dB knee from the 470 pF capacitors is 16 kHz. The larger capacitors are used to achieve a lower frequency knee, and the small capacitors are used for their known excellent high-frequency performance (first self-resonance $\geq$ 1 GHz). Both types are used because the frequency of the phantom power was unknown. After applying the shunt capacitors, the anomalous temperature rise was reduced to 17 mK, for an extrapolated "phantom power" of only 3 nW. This improved performance is illustrated by the lower orange point in figure 6.34.

Figure 6.35 illustrates the practical consequences of the above thermal mass and phantom power design considerations. The red dashed line is the cooling trend of the original heated load with zero power applied. Due to the excessive mass, it takes about 8 hours to cool from 18 K to the final temperature of 288 mK; the final temperature is high due to an unmitigated phantom power. The solid blue line is the actively controlled temperature after design improvements. Due to the reduced mass and heat capacity, it is trivial to execute three excursions from about 200 mK to 1 K and back in about 8 hours, and the lowest temperature possible is reduced to $<$ 100 mK due to reduced phantom power. The many flat regions are periods when the heated load is held at a constant temperature for periods of data acquisition. Since the total time is dominated by data acquisition periods (flat regions) rather than by heating or cooling periods, there is little to gain from further thermal optimizations.

6.4.3 Microwave Considerations

We desire that the heated load transmit thermal white noise to the amplifier with minimal attenuation in the signal band ($\approx$ 200 to 1000 GHz), but have minimal emission near the Josephson frequency ($\approx$ 30 GHz). This is achieved through combination of a homemade lossy "roach filter" and a commercially available low-pass filter. The "roach filter" (so named because "out of band signals check in, but they don't check out", much like the commercial pest control "roach motel") is composed of a 50-Ω waveguide with a lossy dielectric, with two versions shown in figure 6.36.

The coaxial roach filter (fig. 6.36a) is made from a bored copper block and center copper post, with the dielectric made from a cast-able microwave absorber (Eccosorb CR), with the dimensions chosen to ensure a 50-Ω environment. This design has the advantage of excellent

[16] P/N C2220C474J5GACTU by Kmet

[17] P/N ATC 800B 471JT by American Technical Ceramics

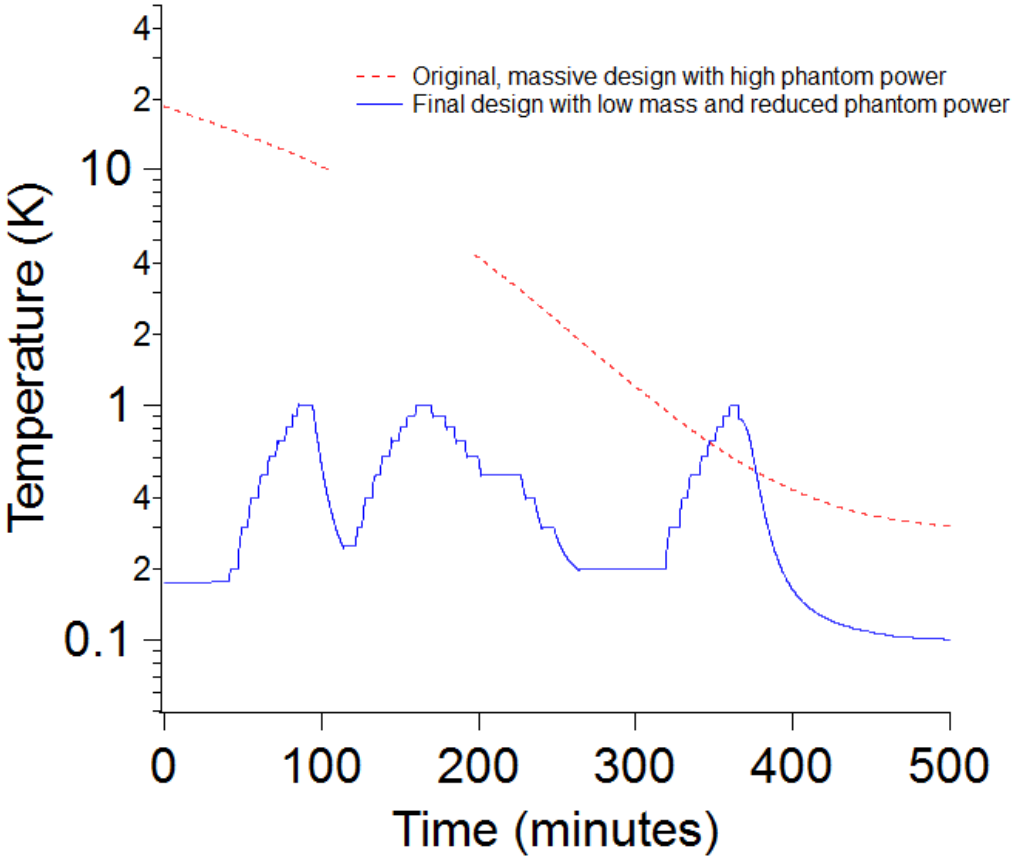

Figure 6.35: Thermal agility before and after optimization of the heated load. MXC is at 60 mK for both curves. Red data shows performance before optimization: a cooling trend with 0 power applied to the heater. It takes approximately 8 hours to cool from 20 K to 288 mK. (Gap in red data is due to a malfunction of the data logger.) Due to high final temperature and slow response time, this design is not practical. Blue data shows temperature vs time for about 8 hours of active temperature control. Rising sections have high applied power, falling sections have zero applied power, and flat sections are feedback-controlled to hold a constant temperature. With fast temperature response and a final temperature < 100 mK, the revised design is practical for performing T_N measurements.

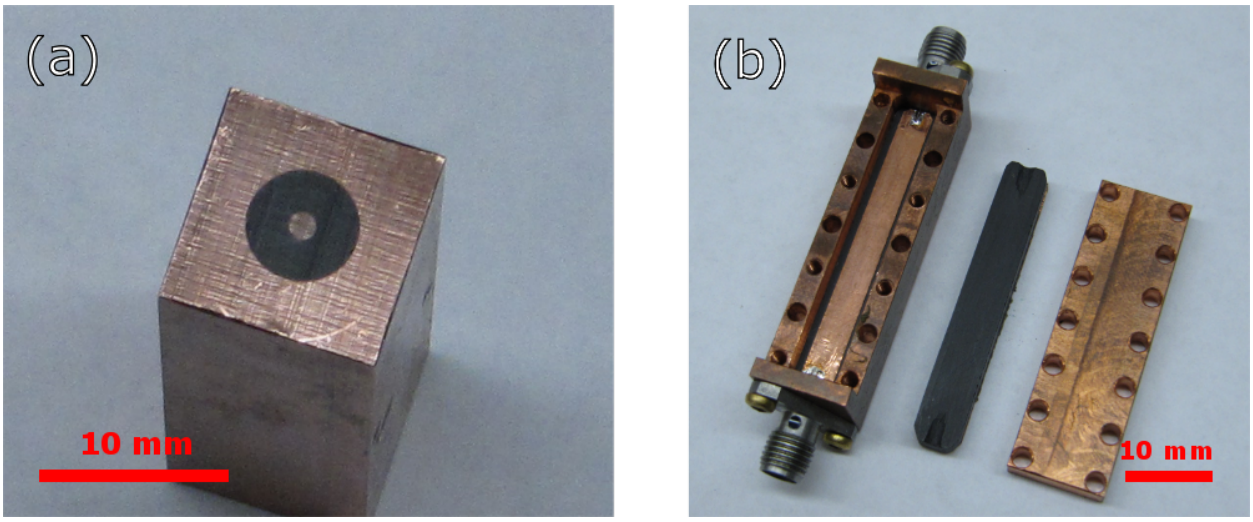

Figure 6.36: Two roach filter designs. (a) Cross-section of a coaxial "roach" filter. (b) Exploded microstrip "roach" filter.

dimensional stability, physical robustness, and simple, ideal, microwave design. It suffers from difficulty in attaching connectors and achieving a void-free casting in a narrow space. The microstrip roach filter (fig. 6.36b) is made from a machined copper cavity, two sheets of a machinable microwave absorber (Eccosorb MF) and a thin copper foil center conductor with the dimensions again chosen to ensure a 50-Ω environment. This design has the advantage of ease of manufacture and connectorization. It suffers from non-ideal microwave performance due to physical gaps and manufacturing tolerances. On balance we prefer the microstrip design. Either design has low loss of about -0.3 dB at frequencies up to 1 GHz, rising to about -25 dB at 18 GHz. This ensures that around the Josephson frequency the microwave environment remains a low-temperature thermal bath even as the temperature at signal frequencies is adjusted. The Ecccosorb materials have negligible loss below 3 GHz, so to further reduce out-of-band thermal radiation the roach filter is followed by a commercially available reactive low-pass filter. It is rare for a commercial product to be rated to cryogenic temperatures, so direct testing remains the best way to evaluate performance. We used both a relatively inexpensive 1.35 GHz low-pass filter (P/N VLFFX-1350 by Mini-Circuits) and a premium 2 GHz "clean up" filter with impressive high-frequency rejection (P/N 5L205-2000/T9000-0/0 by K& L Microwave).

If the lines between the heated load and amplifier have microwave loss, the effective temperature presented to the amplifier input will be something between the actual heated load temperature and the temperature of the lossy transmission line, in accordance with equation 6.3, introducing a systematic error to the T_N calculation. Unfortunately, low microwave loss correlates with high thermal conductivity in all normal metals. The high electrical and low thermal conductivity of a superconducting transmission line would be ideal, but this is not

practical for our purposes, as the T_c of a NbTi line is about 9 K, and we require temperatures as high as 20 K. A reasonable compromise is found with a few tens of millimeters of SS-SS coax.

Total loss from the SS coax line and roach filter in the signal band is shown in figure 6.37. "Filter A" is the roach filter, and "cable U" is the U-shaped SS coax cable. Total error amounts to about 8%, and is well characterized by a linear loss vs. frequency relation, which can be incorporated into final T_N calculations.

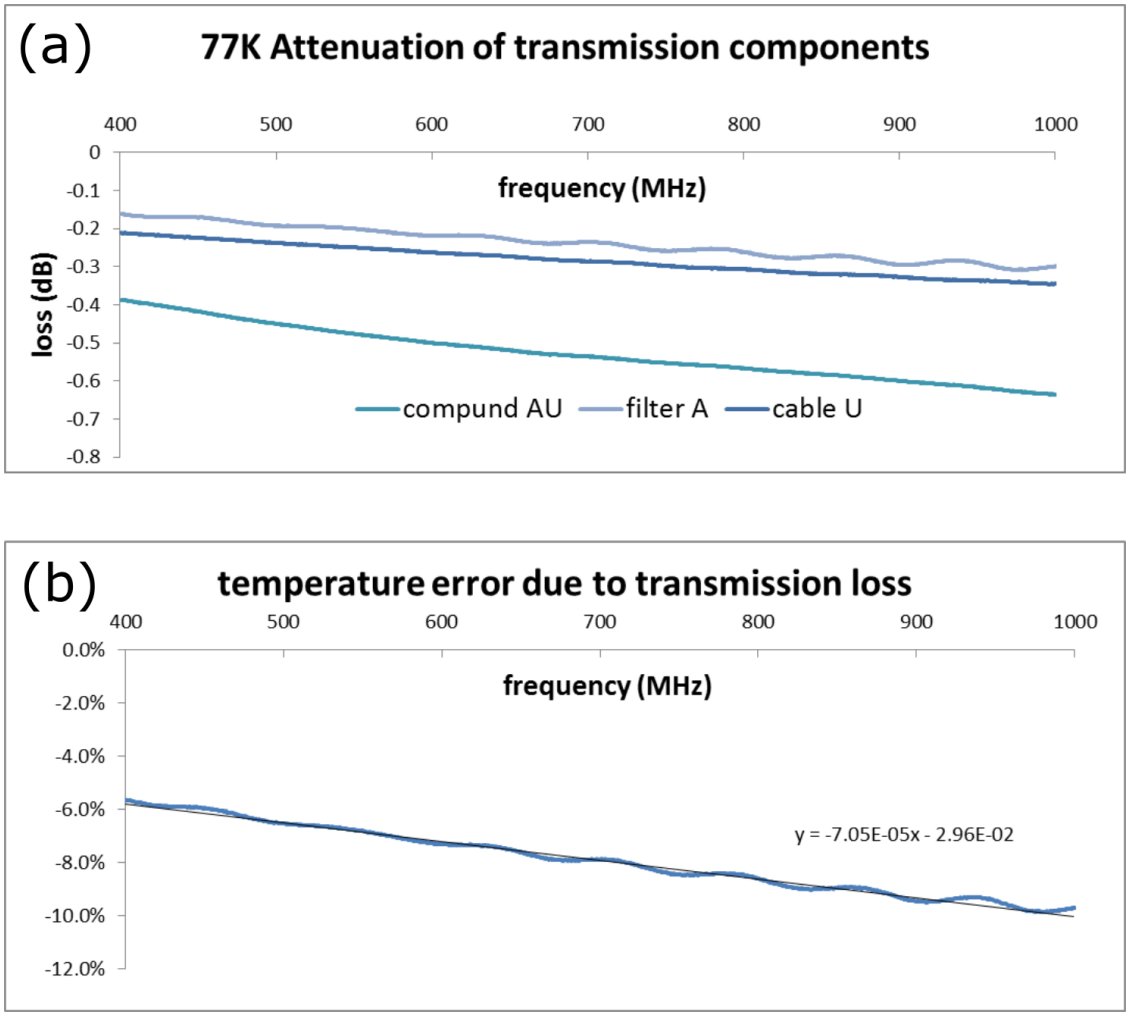

Figure 6.37: Microwave losses from transmission components of the heated load. (a) Losses in dB by component. "filter A" refers to the roach filter (yellow section in figure 6.32b, with detail in figure 6.36), "cable U" refers to the U-shaped coaxial cable (blue section in figure 6.32b between the heated attenuator and roach filter.), and "compound AU" refers to the series arrangement of "filter A" and "cable U". (b) Total temperature error due to transmission loss. The thermal power spectral density of the heated load is reduced by 6 to 10 percent from 400 MHz to 1 GHz due to transmission losses. The temperature error is well approximated by a linear fit, and can be accounted for in T_N calculations.

Chapter 7

Experimental Results

7.1 Varactor Temperature Dependence

Our first tests of MSA tunability in a dilution fridge at temperatures of 10's of mK showed that the tunable range was significantly diminished compared to the range accessible at 4 K. The high frequency tuning limit was nominally unchanged, but the low frequency tuning limit was much higher, suggesting a reduced maximum available capacitance at mK temperatures. This was surprising and conflicted with the naive expectation that a physical property virtually unchanged between 300 K and 4 K would remain constant at still lower temperatures.

We created a testbed to evaluate the varactor properties at low temperatures and found surprising results. We speculate that new solid-state physics may be demonstrated in these data, and propose the outlines of some possible theoretical models to explain the results.

7.1.1 Testbed for Investigating Temperature Dependence of Varactor Parameters

To investigate the phenomena of reduced tunable range at mK temperatures and to separate varactor performance from MSA performance, it was necessary to construct a simple linear resonator that could be terminated and tuned by the same varactors used to tune the MSA. By measuring the resonant frequency and Q as a function of both temperature and varactor voltage, we could investigate the varactor properties directly. Figure 7.1 shows the linear resonator used for the varactor tests. The resonator was constructed from available components, including sample boxes with internal coplanar waveguides originally made for flux qubit experiments and various SMA barrel connectors.

The left-hand Cu enclosure shown in detail in figure 7.1b contains the capacitor for input coupling and DC blocking, the DC bias inductor, and connectors for DC bias voltage and RF reflection measurement. The RF signal path starts at the lower right of figure 7.1b where a 50-Ω coax line transitions to a 50-Ω coplanar waveguide. The center trace is interrupted

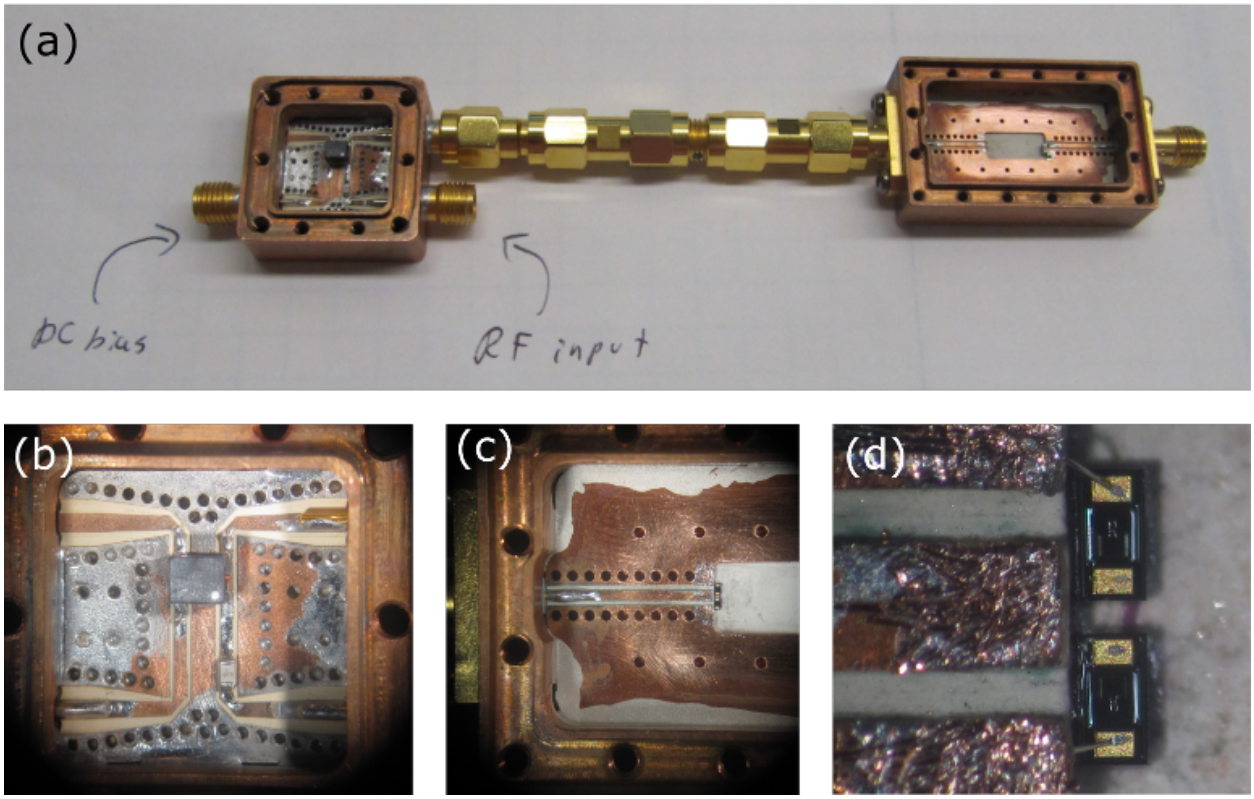

Figure 7.1: 1 GHz linear resonator used for varactor tests. This testbed was assembled from parts made for flux qubit experiments and stock SMA components. (a) Overview. (b) Left hand input box detail, showing coupling for test tone and DC bias. (c) Right hand termination box detail, showing coplanar waveguide. (d) Right hand termination varactor detail, showing end of coplanar waveguide and varactors at greater resolution.

by a capacitor, then connected to a DC coupling inductor and continues to a 50-Ω coax line exiting at the top right. The majority of the resonator by length is composed of a sequence of SMA barrel connectors selected for a total length of about 0.1 m and a resonant frequency of about 1 GHz, chosen to be close to the MSA operating frequency. The resonator line ends at another box shown in detail in figures 7.1c and 7.1d, in which the 50-Ω coax line transitions to a 50-Ω coplanar waveguide, which is terminated by two varactors to ground. (In figure 7.1d the varactors are disconnected from the center trace for a "open" resonator performance test. They are connected via Al wirebonds in later tests.) The SMA connector on the far right of figure 7.1a is unused, and is not connected to the resonator. Figure 7.2 shows a schematic of the linear resonator with a VNA attached.

A Vector Network Analyzer (VNA) (Agilent E5071B) is used to measure S_{11} of this tunable resonator, from which the complex impedance of the varactors can be extracted. To extract the varactor complex impedance from an S_{11} measurement, it is necessary first

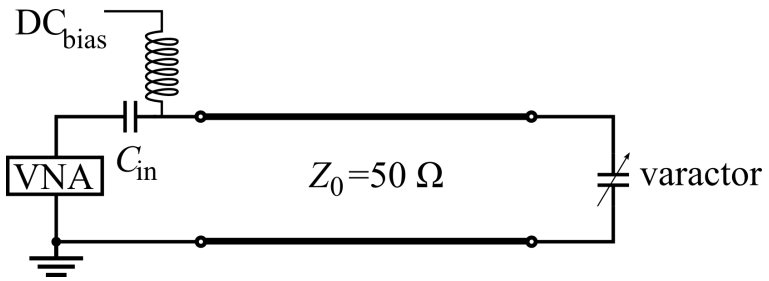

Figure 7.2: Schematic of 1 GHz linear resonator used for varactor tests. The length is chosen such that the $\lambda/2$ frequency is about 1 GHz, C_{in} decreases the frequency by a small amount due to finite conductivity, and C_{varactor} has a large effect on the frequency. C_{varactor} is varied by means of the applied DC bias, coupled in through an RF inductor to minimize RF communication between the resonator and DC bias line.

to characterize the impedance Z_0 and the unloaded $\lambda/2$ frequency f_0 of the resonator. By design $Z_0 = 50\,\Omega$, and in principle f_0 can be measured by using a vanishingly small C_{in}. In practice a moderate value of C_{in} is necessary to achieve coupling strong enough to be visible, so the loading from C_{in} must be accounted for. By measuring S_{11} of the "bare" resonator with a variety of values for C_{in}, f_0 may be extrapolated to $C_{in} = 0$ and with these same data, a self-consistency check made for Z_0. Figure 7.3 shows S_{11} of the "bare" resonator with five different values for C_{in}. For greatest consistency, these data were collected with the resonator mounted to the mixing chamber of the dilution fridge (at room temperature), using the same cabling and signal path used later at mK temperatures. As a consequence of the long cable lengths and many opportunities for slight impedance mismatches at connectors, there is a fair amount of ripple in these data. By subtracting the average ripple and fitting to a Lorentzian function, the resonant frequency can be extracted.

Figure 7.4 shows the measured resonance frequency with a variety of input coupling capacitors from 0.2 pF to 1 pF (blue dots). The uncertainty in capacitance is taken from the manufacturer's specification of $\pm$ 0.1 pF. The uncertainty in frequency is small, so the vertical error bars are suppressed for clarity. A linear regression on these data extrapolates to an unloaded frequency f_0 of 1005 MHz when $C_{in} = 0$. As a consistency check, the numerical model described in section 5.1 is applied assuming assuming an unloaded frequency $f_0 = 1005$ MHz, $Z_0 = 50\ \Omega$, $C_{end} = 0.001$ pF, and $R_{end} = 0$ to determine f as a function of C_{in}, and the result is plotted in orange. The close agreement between model and data suggests the resonator parameters Z_0 and f_0 are now sufficiently well characterized to model the varactor capacitance to $\pm$ 0.1 pF accuracy.

As a final check before testing varactors, the testbed must be checked for temperature dependence. With an input capacitor of 0.3 pF, the resonator is installed in the dilution fridge, and the resonant frequency and Q of the resonator measured at various temperatures. Figure 7.5 shows the resonant frequency and Q of the resonator at temperatures from 300 K

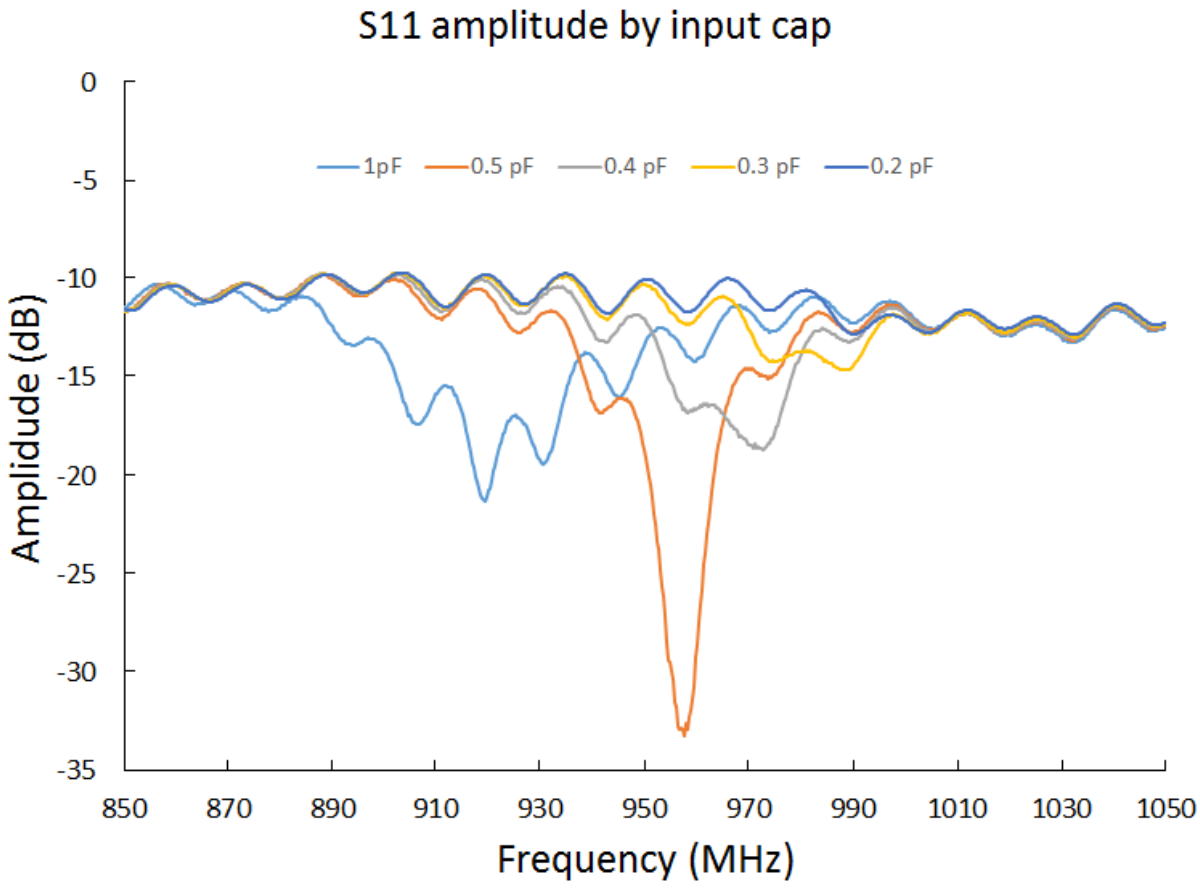

Figure 7.3: S_{11} of bare resonator with 5 different values of C_{in}. The trend to higher frequency with smaller C_{in} is clear, as well as an optimal coupling around $C_{in} = 0.5$ pF. An input capacitance of $C_{in} = 0.3$ pF (slightly under coupled, usable visibility) was used in subsequent tests.

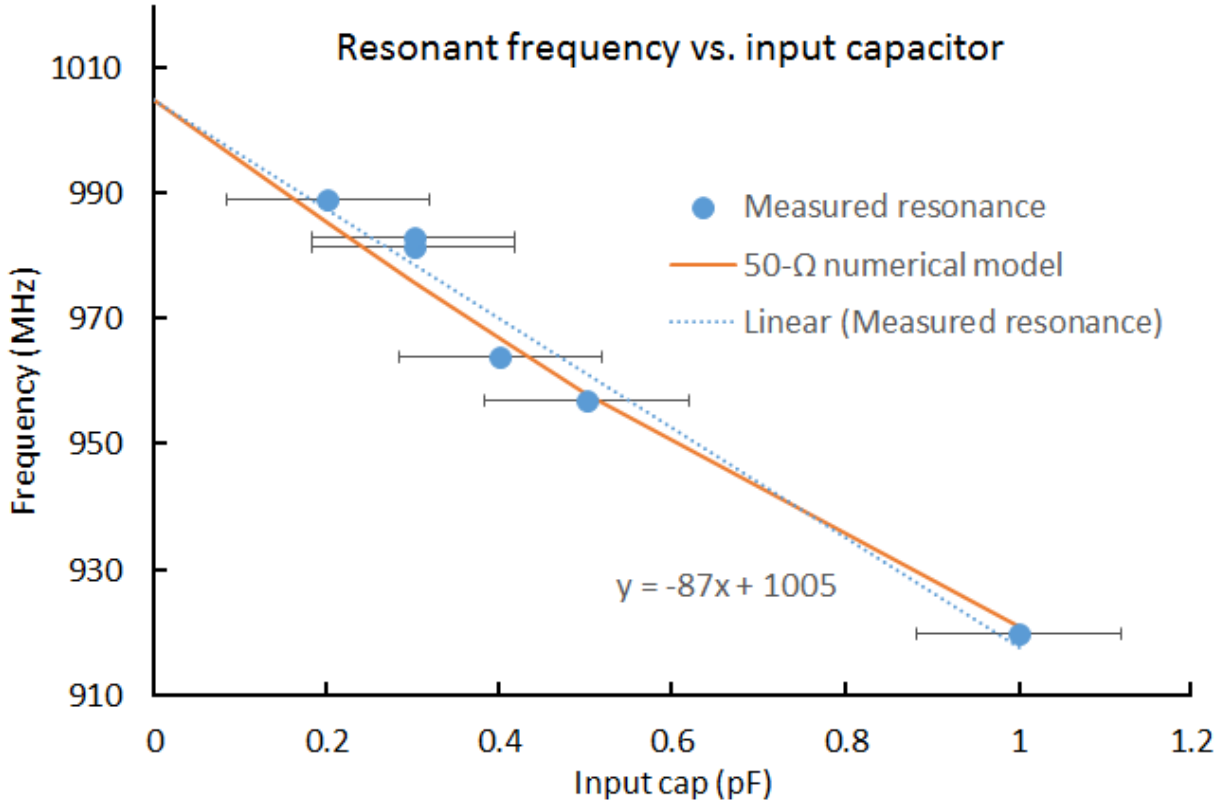

Figure 7.4: Data and numerical model of a 50-Ω, 1005-MHz resonator as loaded by C_{in}. Blue data points are frequency vs. C_{in}. Error bars represent the manufacturer's listed tolerance. Dashed blue line is a naive linear fit indicating $f = 1005$ MHz in the unloaded limit of $C_{in} = 0$. Orange line is a numerical simulation of a 50-Ω transmission line coupled through C_{in} to a 50-Ω resonator with a $\lambda/2$ frequency of 1005 MHz. This numerical model is what is assumed in following tests that incorporate varactor termination.

to $<$ 50 mK. Data are recorded at room temperature and in the two stable temperature regimes available with the dilution fridge: 1 K to 40 K (no ^{3}He$-^4$He phase separation) and <100 mK (full dilution cycle). The temperature dependence of frequency is negligible, with $f_0(300\text{ K}) = 981.0$ MHz and $f_0(30\text{ mK}) = 981.8$ MHz, and a smooth trend of f vs T at intermediate temperatures. The total frequency change (about 1 MHz) is equivalent to about 0.011 pF change in C_{in} by the linear fit in figure 7.4. The temperature dependence of Q is substantial, ranging from $Q = 166$ at $T = 300$ K to $Q = 561$ at $T = 30$ mK, with almost no temperature dependence below 10 K.

When the varactors are installed, provided the loaded Q is small relative to 500 and capacitance changes are large relative to 0.011 pF, this testbed is proven suitable.

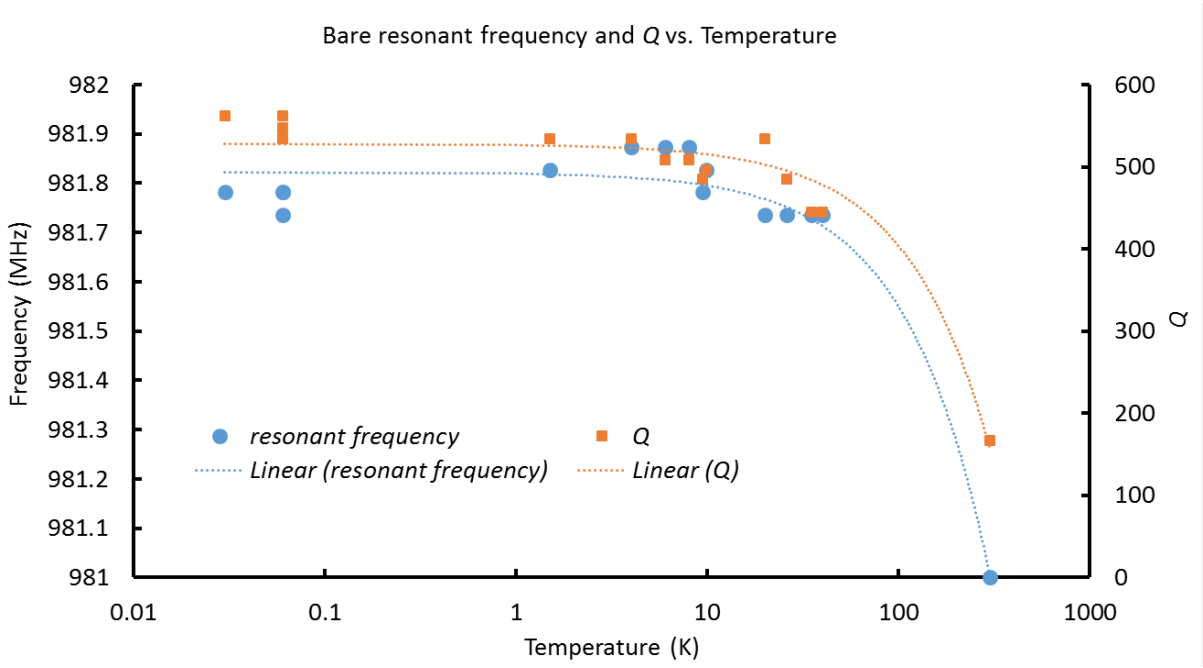

Figure 7.5: Un-loaded linear resonator: frequency and Q as a function of temperature. Both frequency (blue) and Q (orange) exhibit roughly linear trends in temperature between 300 K and 50 mK. (Linear fits shown as dashed lines) Both frequency and Q are largely constant below 10 K, so the testbed is proven stable for $T < 10$ K.

7.1.2 Data Collection Method

With the testbed parameters Z_0 and f_0 well characterized at every temperature of interest, it is possible to acquire data on the varactor performance. Two varactors are wirebonded between the center trace and ground at the end of the linear resonator as suggested by figure 7.1d. Two varactors are used rather than one both to create a greater tuning effect and for the aesthetic appeal of physical symmetry. An input capacitor of $C_{in} = 0.3$ pF is chosen as a reasonable compromise between visibility and loading. With the resonator installed on the fridge, the amplitude and phase of S_{11} are measured with the VNA. We need S_{11} as a function of frequency, varactor voltage, and temperature; frequency sweeps are very fast and are run internally by the VNA, voltage sweeps are achieved by controlling a Keithley 2400 sourcemeter and the VNA from the same custom software, and temperature sweeps (which are quite slow) are achieved by manual temperature changes and recording. At any single temperature, about 100 varactor voltages are applied and an equal number of complex S11 spectra are saved. Typically two sweeps are performed at a given temperature: an initial sweep is performed with a uniform steps in varactor bias voltage (typically from +1V to -12V), then with the first sweep as a guide, the second sweep uses a non-uniform progression

in voltage designed to produce uniform tuning in frequency.

Extracting the varactor capacitance as a function of voltage and temperature from raw S_{11} resonator data is not entirely straightforward, because the power reflected from the resonator is rather small and spurious reflections from various non-ideal components in the signal pathway between room temperature and the mK stage of the fridge tend to dominate the raw signal. The reflected signal is small because several 10's of dB of attenuation is explicitly added between room temperature and the mK stage to avoid illuminating the sample with 300 K microwave radiation. This attenuation diminishes both the illuminating power and reflected power. The spurious signal is large because a very high quality SMA connector on a microwave line may have a reflection coefficient of -30 dB, but a value of -20 dB or greater is more typical, and there are necessarily many couplings between the room temperature and mK stages of the dilution fridge. Even with great care, the reflected signal is small relative to the spurious reflected power. Fortunately, the spurious reflections are static while the signal of interest varies with varactor voltage, allowing us to see motion against a "bright, static" background. Appendix A details the method of removing the spurious background signal and extracting varactor parameters from the cleaned signal. Section 7.1.3 takes the signal analysis as given and presents the temperature dependence of the varactor parameters.

7.1.3 Varactor Parameters as a Function of Temperature

Plots of capacitance vs. voltage (like figure A.15c) at every measured temperature are shown overlaid in figure 7.6. Note that there is negligible temperature dependence at voltages near -12 V where the varactors are largely depleted and the remaining capacitance is strictly geometric, but there is a spread of values at voltages near 0 V, indicating a temperature dependence of charge carrier density.

The temperature dependence of these same data is better illustrated by plotting it as C vs. T at constant voltage as in figure 7.7. The top trace is a varactor bias of +1 V and the bottom trace is a bias of -12 V. The top trace has open squares marking the data points to illustrate the temperatures at which data were taken.

Note variation of the maximum capacitance (at most positive bias) as the varactor is cooled from room temperature: there is a small drop in capacitance of about 10% from 300 K to 100 K, and it remains nearly unchanged down to the experimentally important temperature of 4 K (temperature of liquid helium). From about 3 K to 400 mK, the capacitance drops to about half its room temperature value. From 400 mK to the lowest tested temperature of 60 mK, no further changes are observed. This has two important implications:

- For applications at temperatures of 4 K and higher the room-temperature varactor performance is a reasonable predictor of actual performance, but applications at dilution-fridge temperatures like those in ADMX should plan for about half as much capacitance per varactor.

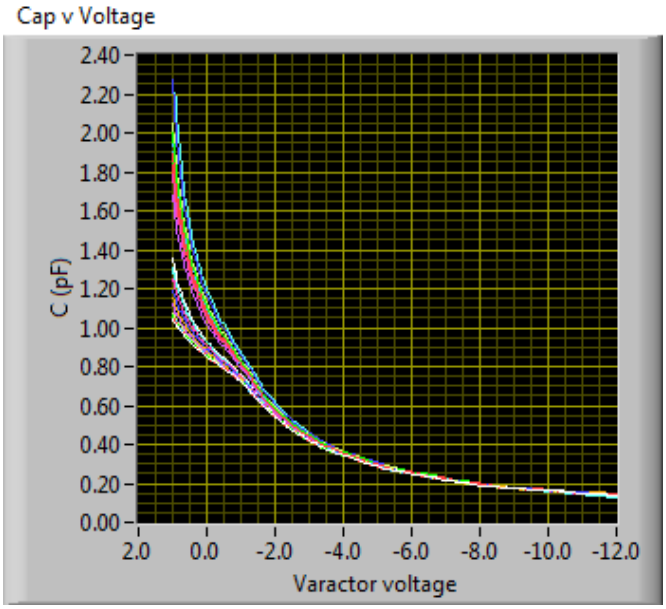

Figure 7.6: Map of all C vs V functions with T from 300 K to 60 mK. The upper curve is at $T = 300$ K, and the lower curve is at $T = 60$ mK. Note that C is independent of T at extreme negative voltage bias, but is a strong function of T at biases near 0.

- Any device tuned near the lower frequency range (ie max capacitance) will exhibit a strong temperature dependence at temperatures between 400 mK and 3 K, but will be relatively immune to temperature fluctuations below 400 mK.

We also have data of Q as a function of voltage and temperature, which proves to be rather surprising. These data are shown in figure 7.8. Figure 7.8a shows a Q that is oscillatory with V (at least for some temperatures) and very low Q for the most positive voltages at some temperatures. Figure 7.8b shows a narrow range of Qs at room temperature, a universal increase in Q as T drops initially, and a dramatic broadening of Qs across bias voltages at the lowest temperatures.
The many curve crossings in figure 7.8 make the data difficult to follow so they are is represented in figure 7.9 as a colored surface with voltage and temperature as the Y and X coordinates, respectively, and Q represented by color.
The left edge of the surface corresponds to room temperature: here Q is about 30 with some small periodic variations with voltage (frequency). As the temperature falls to about 30 K, the Q increases to a maximum of around 75 and the periodic variations become more pronounced. The Q-increase is entirely consistent with reduced resistance in Cu and the other materials making up the resonator, and the increased ripple could be explained by (non ideal) thermal dependence of couplings and connectors in the test setup. At voltages near -12 V with the varactor totally depleted, this may be a sufficient explanation for the increase in Q. However, at voltages near 0 V the Q drops dramatically for temperatures

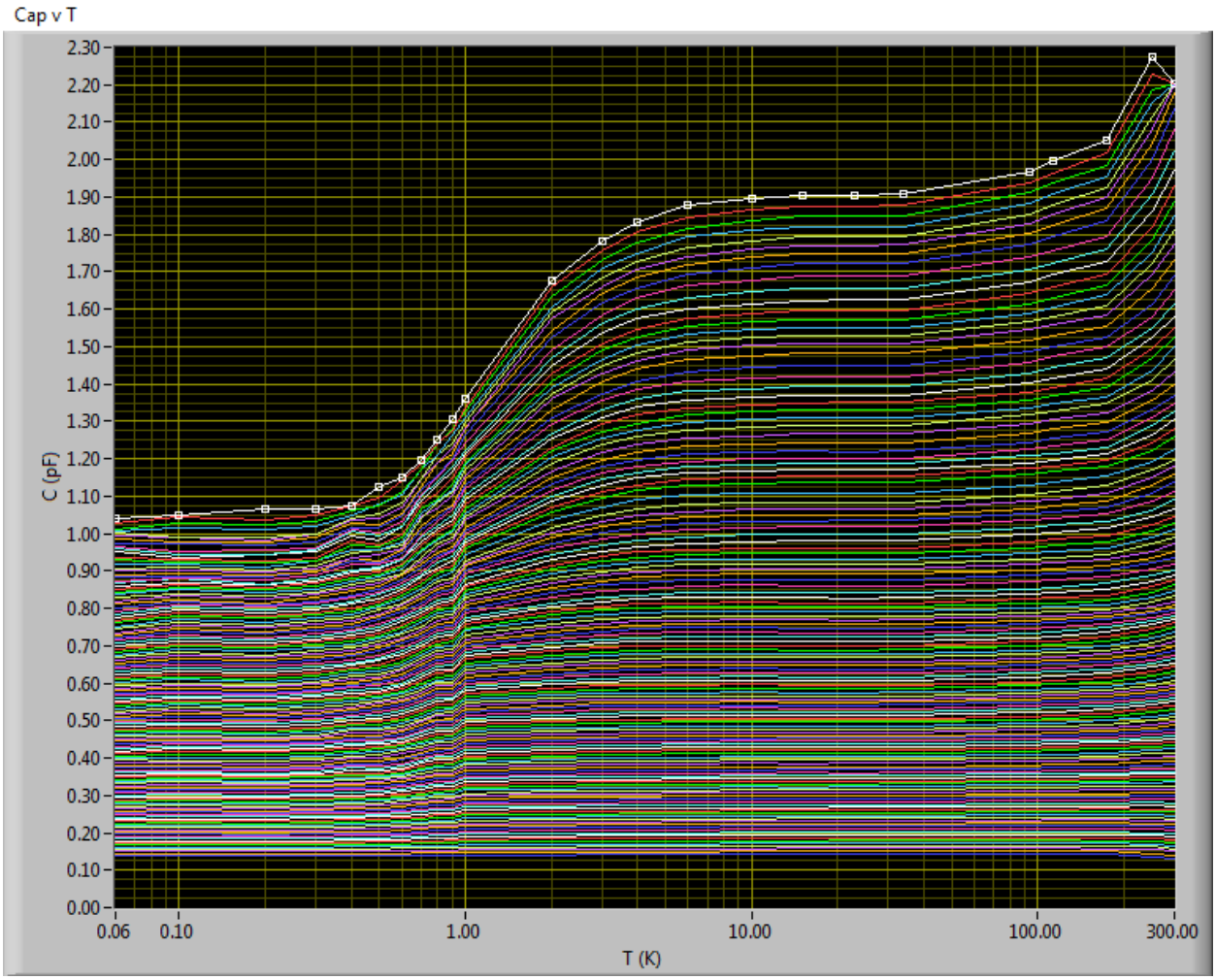

Figure 7.7: Map of all C vs. T functions with V from +1 V (top curve) to -12 V (bottom curve). After a small drop in C as the temperature drops from T = 300 K to T = 100K, the capacitance is fairly constant down to a few K. Remarkably, the capacitance drops again between 3 K to 400 mK. Below 400 mK, capacitance is insensitive to temperature.

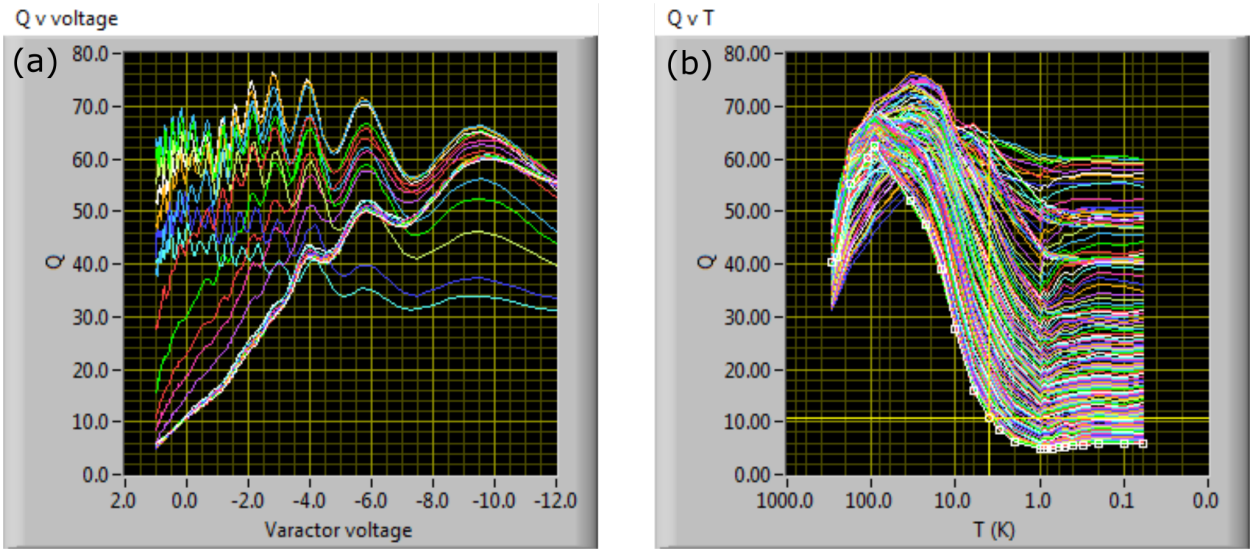

Figure 7.8: Q as a function of T and V. Q has a strong dependence on T, but the trend is more complex than for capacitance. (a) Overlaid Q vs. V at various temperatures. (b) Overlaid Q vs. T at various voltages.

below about 10 K to a low of about $Q = 5$. We currently have no theory for this behavior but maintain it reflects the properties of the varactor and not the experimental setup. Why the varactors become very lossy at temperatures below 10 K, and only when moderately biased remains unknown. The periodic variations appear closer together at more positive voltages, but are approximately equally spaced in frequency. To illustrate this, figure 7.9 is re-plotted in terms of frequency instead of bias voltage in figure 7.10.
In this representation the "Q-ripples" are clearly regular in frequency, and the boundaries of the surface are reminiscent of figure 7.7.

7.1.4 Theoretical Speculation on the Causes of Varactor Low-T Properties

The drop in varactor capacitance at temperatures around 1K was rather unexpected and remains unexplained. It suggests some semiconductor physics with a relevant energy scale of 10^{-4} eV, which is an energy scale not extensively explored in the field to date. Our proposed explanations are presented here not as solutions, but as a starting point for other interested parties to improve upon, either by expansion of the ideas presented here, or through a thorough refutation.

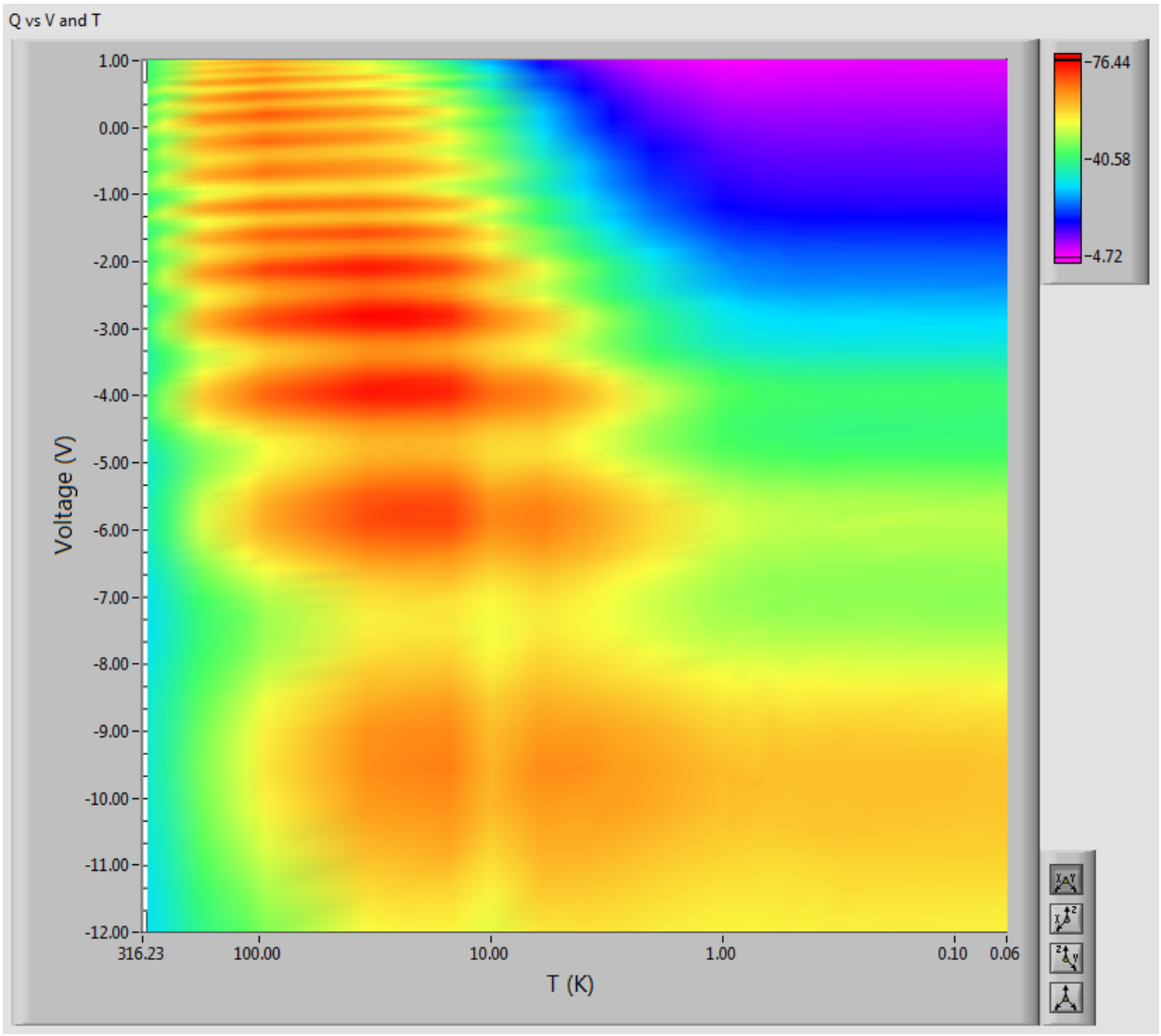

Figure 7.9: 3-D surface of Q, V, and T dependence.

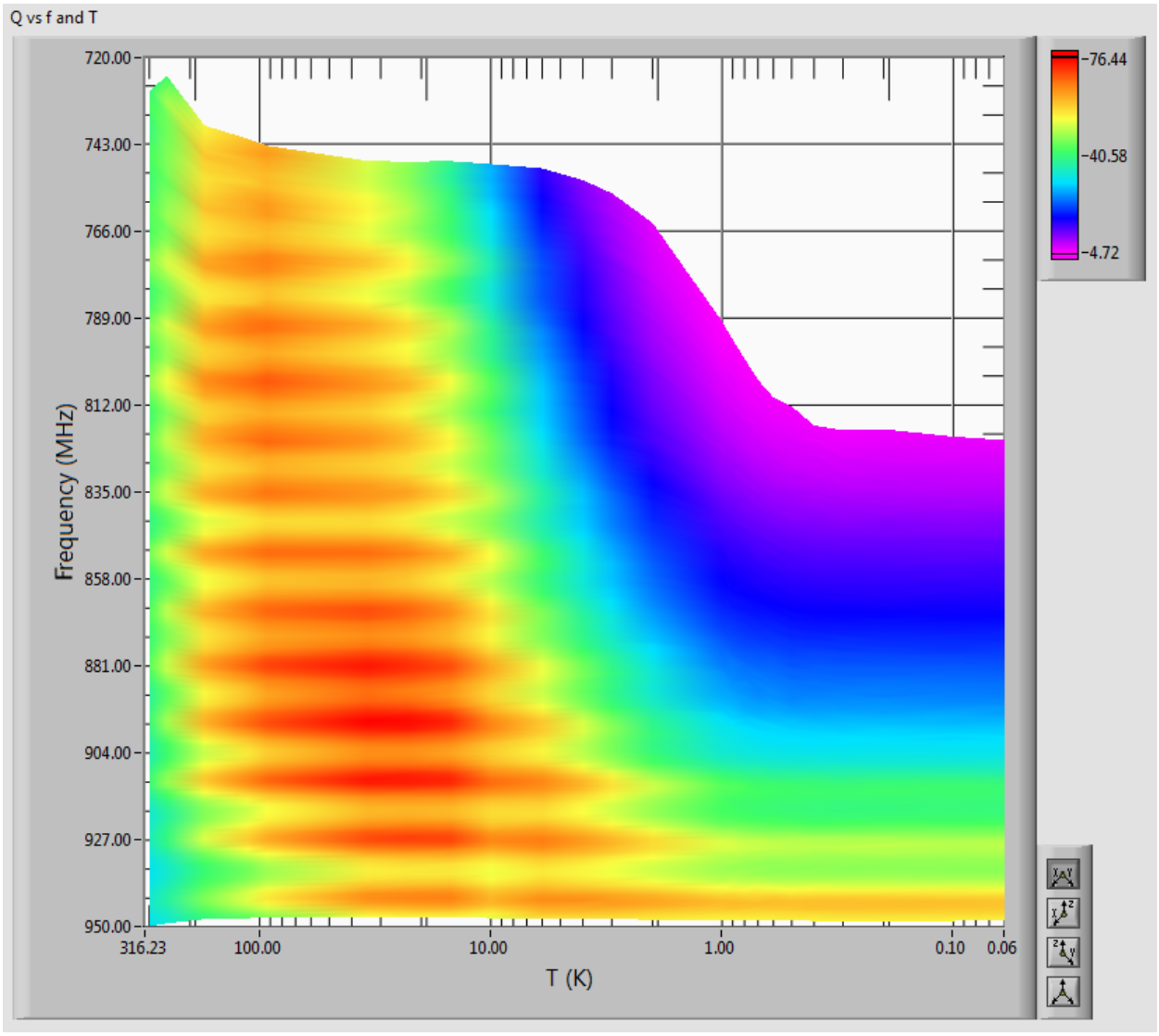

Figure 7.10: 3-D surface of Q, f, and T dependence.

7.1.4.1 Conventional Semiconductor Model

In the high-temperature regime ($T \ll 50$ K) defined by the typical GaAs biding energy (≈ 5 meV), we make the following simplifying assumptions:

- The donor and acceptor sites are uniformly distributed and the electrons and holes are unbound.
- Approximating a P-N junction as an infinite plane (ie ignoring edge effects) we can solve the 1-D Laplace equation to solve for the potential V of the charge carriers as a function of z, and find that V varies quadratically with z.
- The electron quasi-Fermi energy (μ_e) is a function of the donor level E_d and bandgap E_g. (The $E = 0$ reference is taken at the edge of the valence band, and $|E_d|$ is assumed small relative to E_g.) At $T = 0$ the Fermi energy is pinned to the donor edge ie $\mu_e(T = 0) = E_g - E_d$. As $T \to \infty$, the Fermi energy decreases towards the middle of the band gap ie $\mu_e(T = \infty) = E_g/2$. The same holds for the hole quasi-Fermi energy in that $\mu_h(T = 0) = E_h$ and $\mu_h(T = \infty) = E_g/2$.

Based on this model, at any temperature $T < \infty$ there would be a discontinuity between the μ_e and μ_h Fermi levels at the P-N junction (z=0). The discontinuity is smoothed out by electrons migrating from the n-type region to the p-type region until the Fermi energy is constant throughout the sample. This space-charge shifts the valence and conduction band edges as necessary to maintain a constant Fermi level. In the vicinity of the band-bending (near z=0) the donors and acceptors are stripped of their charges, creating a depletion layer of width w. Applying an external voltage to the two regions modifies the total amount of band bending, and thus modifies the width w of the depletion layer. The capacitance C of the junction is a function of w by $C \propto 1/w$. Since both Fermi energies E_d and E_h approach $E_g/2$ at high temperatures, the degree of band bending and the depletion width w decrease with T, and C will increase with T.

7.1.4.2 Neutral Donor Model

A conventional semiconductor model composed of charged donor sites with typical binding energies of several meV or more would be entirely bound, or "frozen out" at our temperatures of interest. To attempt an explanation for our observed low-temperature transition, we require donor sites with a binding energy on the order of 10^{-4} eV, so this end we propose neutral trapping centers arising from neutral atom substitution (for example Al replacing Ga), or resulting from crystal stress or dislocations.

If the donor density is high, we have A 3-D lattice of neutral hydrogen-like donor "atoms" (D^0) in the varactor substrate, each with several overlapping orbitals with its neighbors. This may form a "Donarium" solid, with a band structure and Fermi level that depends on T. Depletion layer theory may be sufficient to calculate the width of the depletion layer, and thus C as a function of T, but at the characteristic temperatures of this phenomenon, the

"donarium" may form a quantum solid, requiring a full quantum mechanical treatment. Whether zero-point motion of the GaAs host atoms would re-normalize the donarium band-gap remains an open question.

If the donor density is smaller, it is conceivable that only the single closest neighboring donor site has significant overlap, with interactions from all more distant sites exponentially suppressed. Given a random spatial distribution of donor sites with an average spacing approximately equal to or larger than the Bohr radius of the donor electrons (about 100 Å) , each donor site is likely to have only one significant neighbor, forming H_2 like "molecules". These molecules will have both bonding and anti-bonding orbitals, with the relative population of each a function of temperature. At lower temperatures more electrons will be in the bound bonding state, resulting in fewer free charges, a wider depletion layer, and smaller C.

At a still lower density, each donor site will act as a solitary hydrogen-like atom. The binding energy for neutral monatomic hydrogen is 13.6 eV, but monatomic hydrogen can also form a stable negative ion H^- with two electrons in a singlet spin state bound to a single proton. This negative hydrogen ion has a binding energy of 0.754 eV. At sufficiently low T we may have the following reaction: $D^0 + D^0 \rightarrow D^- + D^+$ where both D^- and D^+ represent a site with no associated free charge. A donor site in GaAs has a binding energy of about 5 meV, so by simple analogy to the hydrogen atom model, D^- would be expected to have a binding energy of about 0.3 meV, which is the right order of magnitude to explain our transition near 1 K (0.1meV). Such a binding site with two electrons is called "Negative U Center". A conventional negative U center is formed by a IV-III substitution, for example substituting Si for Al or Ga, and form a deep donor with a binding energy around 100 meV. If the "shallow donor negative U centers" model is correct, it represents the first demonstration of a negative U center at ultralow T.

7.2 Initial Demonstration of MSA Tunability

Our first tunable MSA tests used devices made by ez SQUID[1] and a large terminating varactor. This SQUID was manufactured for DC operation, but mounted to a carrier with 50-Ω waveguides and attached with short (< 1 mm) wirebonds. Our initial result, based on tests in a 4.2 K liquid helium bath, is shown in figure 7.11.

This was a very promising early result, and this ez SQUID design became the starting point for our own MSA designs. Note that the tuning range is almost an entire whole octave, consistent with the theoretical model of the standing wave ranging from a $\lambda/2$ to $\lambda/4$ mode. Unfortunately, devices made by ez SQUID ultimately proved unsuitable for our application because they became hysteretic at mK temperatures. We speculate the the manufacturing process left trace amounts of Al in parallel with the junctions. At temperatures around 1 K, the device performance would change drastically and deleteriously, consistent with the development of pinhole superconducting shorts across the junctions.

[1] ez SQUID is a a company founded and run by Michael Mück, formerly a member of the Clarke group at Berkeley. `www.ez-squid.de`

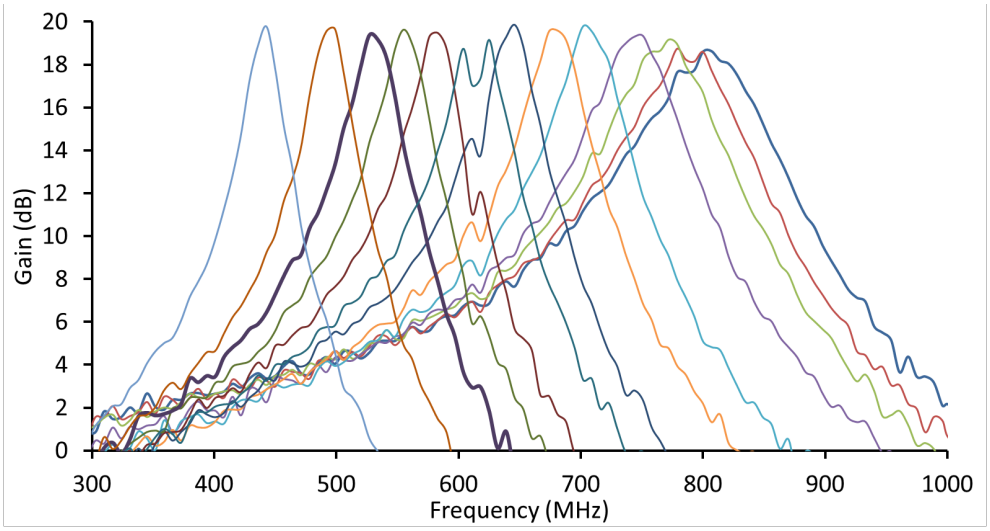

Figure 7.11: Tuned gain peak of an MSA. A gain of about 20 dB is tuned between 415 MHz and 800 MHz.

7.3 Dynamic Range

The MSA has a dynamic range limited by the periodicity of the $\partial V/\partial\phi$ function. A high gain is achieved when the DC flux is biased at a point of large $|\partial V/\partial\phi|$, and so long as the input signal is small, $\partial V/\partial\phi$ is practically constant. (Recall figure 4.9 for a model $\partial V/\partial\phi$ function.) Presuming the input flux is centered around $\phi_0/4$ by the DC bias coil and the amplitude exceeds about $\phi_0/4$, $\partial V/\partial\phi$ will first flatten out and even change sign at the instantaneous extremes of the input signal, resulting in a dramatic decrease in gain. Figure 7.12 shows normalized gain spectra for various input signal powers. At the lowest input power (-125 dBm), the gain is highest. As the input power is increased, the on-resonance gain is suppressed but the off-resonance gain is not affected because the absolute magnitude of flux on the input coil is small and the amplification remains nearly linear. At about -83 dBm, the positive-going and negative-going parts of the $\partial V/\partial\phi$ curve supply about equal power, and the on-resonance gain is reduced to its minimum of about -50 dB or so, which is the passive coupling throughput. As the input power is increased further, the center-frequency gain increases again because the "negating wings" of the non-linear $\partial V/\partial\phi$ function provide *more* power than the "positive center" section. Note that the -80 dBm input power trace shows "fully cancelled" nodes on either side of the center frequency. At the highest input power tested (-74 dBm) there are two "total negation" turnaround points visible.

This dynamic gain measure provides an estimate of the absolute signal flux coupled to the MSA. Figure 7.14 shows the measured gain at the center frequency as a function of input power and a theoretical model of the same. For the theoretical model we assume a transfer

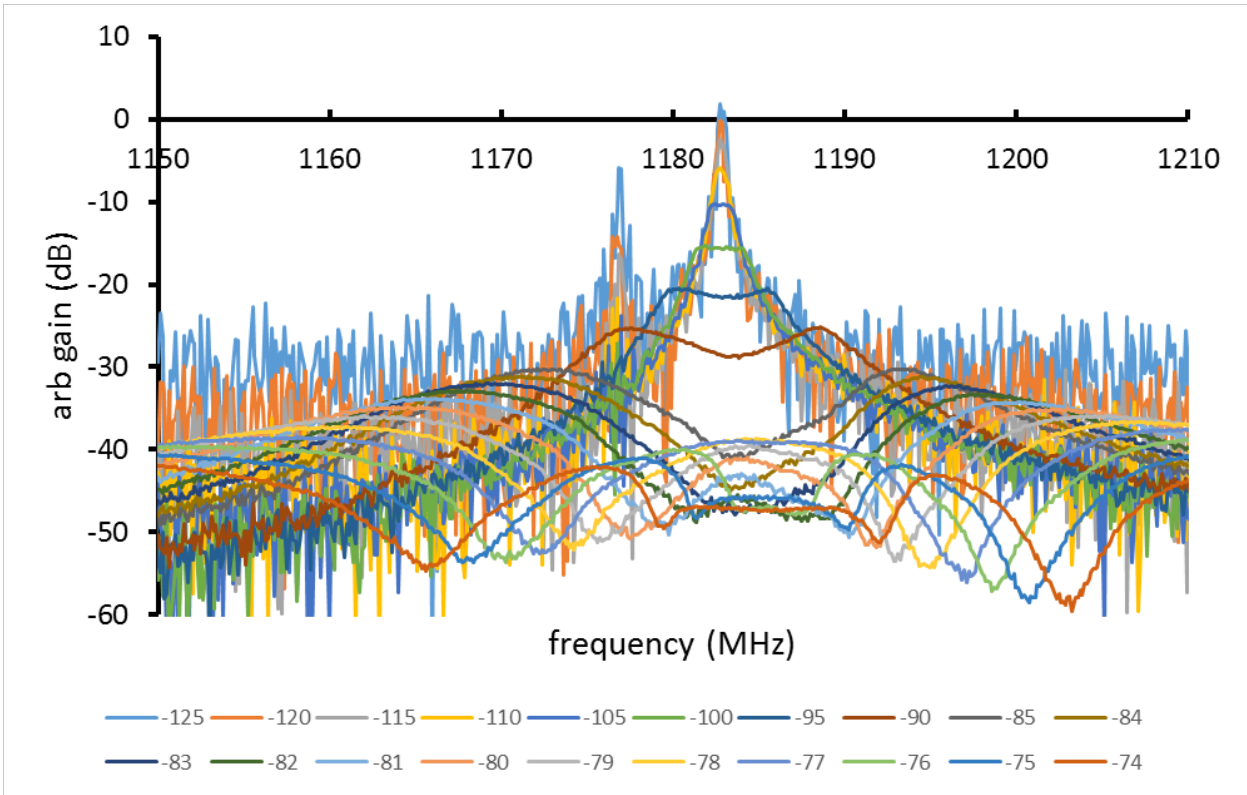

Figure 7.12: Gain spectrum of an MSA for input powers ranging from -125 dBm to -74 dBm. Gain is greatest for an input power of -120 dBm or less, gain is reduced by about 10 dB at -105 dBm of input power, and the gain is lower than the measurement floor at an input power of -83 dBm.

function

$$V_{out} = V_0 \left|\sin\left(\pi A_0 \cos(\omega t) + \pi/4\right)\right|, \tag{7.1}$$

where V_0 is the total peak-to-peak voltage modulation available from the SQUID at our particular current bias, and A_0 is the amplitude of the coupled input signal in units of ϕ_0, consistent with the model in figure 4.9. Figure 7.13 shows the application of this model for three amplitudes.

The exact deviation from linearity depends on the particular form of $\partial V/\partial\phi$, but due to the quantization of flux condition any possible $\partial V/\partial\phi$ function will be periodic in flux with a period ϕ_0. Absolute cancellation[2] of the input signal occurs when the input signal has a peak-to-peak amplitude of about one ϕ_0 (actually closer to 1.2402 ϕ_0 p-p amplitude, or 0.438 ϕ_0 rms, assuming a sinusoidal input signal) regardless of the particular voltage transfer function, so long as that function has a period ϕ_0.

[2]To be precise, transferring all the input signal power to higher harmonics.

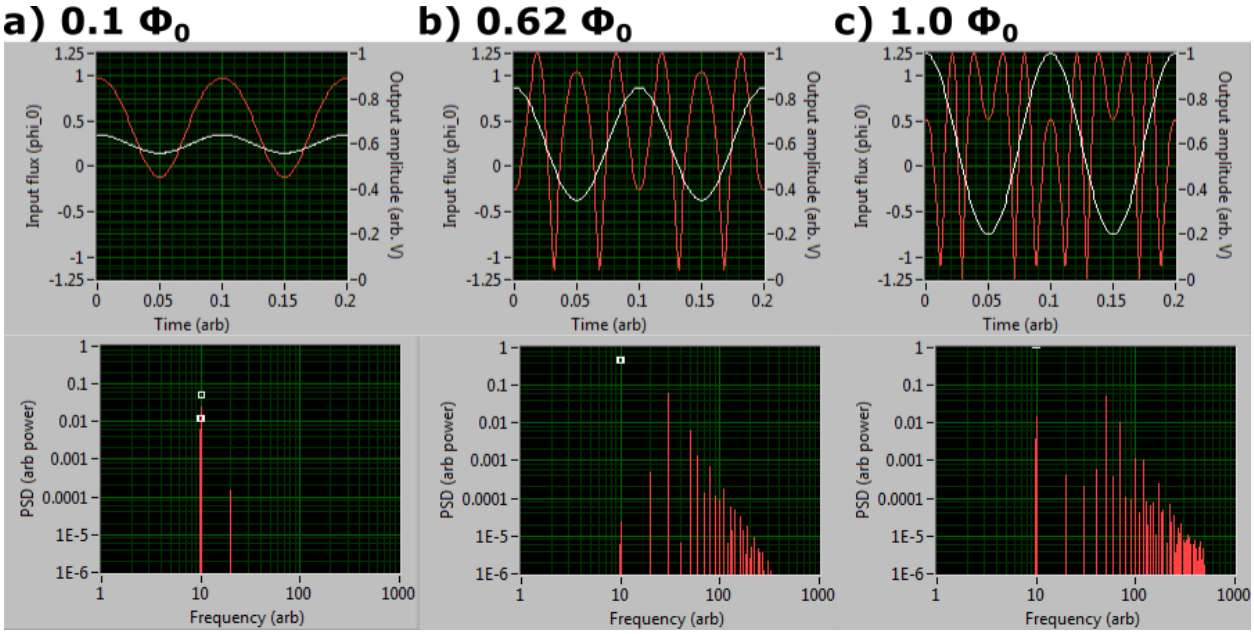

Figure 7.13: Examples of distortion due to the non-linear and periodic $\partial V/\partial\phi$ function. Top row shows input flux-vs time (white) and output voltage-vs-time (red). Bottom row shows the input power spectrum (white points) and output power spectrum (red bars). In each case, the transfer function is modeled as equation 7.1, and the DC flux bias is $\phi_0/4$. Column a) shows an input amplitude of 0.1 ϕ_0, corresponding to an input power of -101 dBm. The output voltage is only slightly distorted, and the power spectrum shows some leakage to the first harmonic. Column b) shows an input amplitude of 0.62 ϕ_0, corresponding to an input power of -83 dBm. The first harmonic has been almost entirely suppressed, and this input power corresponds to the first node in the gain vs. input power function. Column c) shows an input amplitude of ϕ_0, corresponding to a power of -78 dBm. The power in the fundamental frequency is somewhat higher than in column b, but most of the power is in the 4th harmonic. Note that in all cases, the output voltage amplitude is strictly limited due to the periodic nature of $\partial V/\partial\phi$.

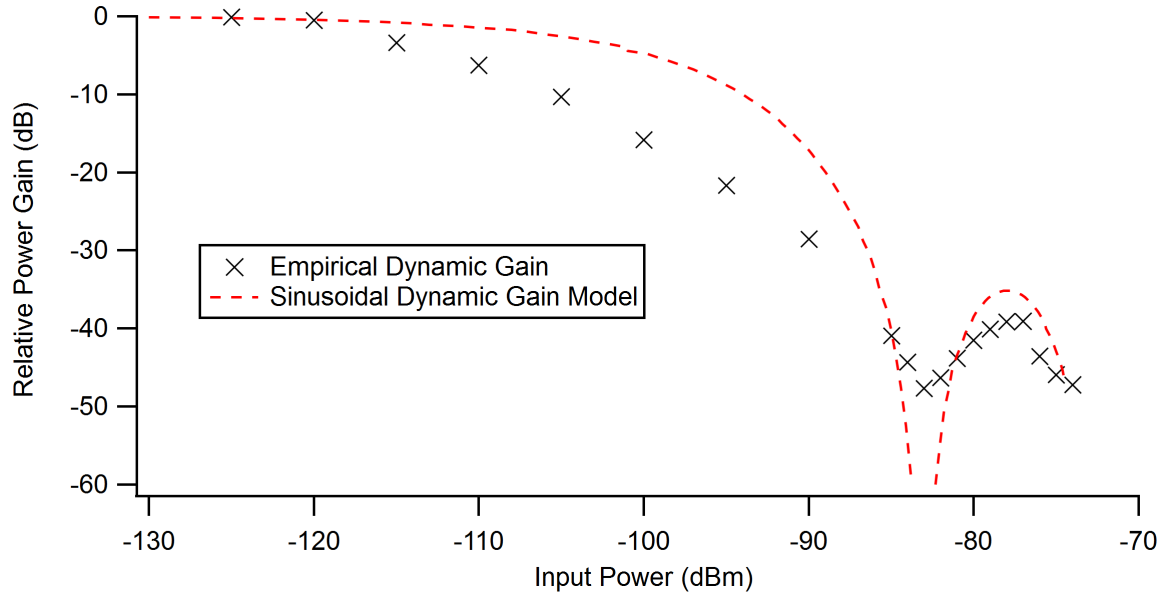

Figure 7.14: Empirical and theoretical dependence of gain on input power. The sinusoidal $\partial V/\partial\phi$ model predicts a gain reduction of about -3 dB at an input power around -95 dBm, or 0.21 ϕ_0 amplitude coupled to the SQUID. We see that the empirical data exhibit a more rapid drop in gain as the input power is increased, with a gain reduced by 3 dB at an input power around -110 dBm, or 0.04 ϕ_0 amplitude coupled to the SQUID. This tells us that the true $\partial V/\partial\phi$ function has deviations from a linear transfer characteristic that become significant at a scale of about 0.04 ϕ_0.

We can calibrate the absolute flux coupled to the SQUID by identifying the first minimum in the measured gain vs. input power function (at -83 dBm) with the first node of the theoretical model, corresponding to an input amplitude of about 0.620 ϕ_0. This single-parameter fit tells us that we can expect about a mean 0.074 ϕ_0^2 per pW of microwave power coupled to the MSA. Recall from section 3.3.2 that the expected signal power coupled to the MSA in ADMX is about 10^{-22} W, or 10^{-10} pW, so even assuming a rather horrendous S/N ratio, the MSA is in no danger of suffering from a lack of dynamic range in this application.

We can learn a bit more from this data: note that the empirical gain drops faster with increasing input power than does our theoretical model. The theoretical model gain falls by 3 dB at a power of -95 dBm, or about 0.21 ϕ_0 in amplitude. This stands to reason– at 0.25 ϕ_0 amplitude, our signal will be just large enough to reach the turn-around points in $\partial V/\partial\phi$, and distortion will be significant. In contrast, the empirical gain falls by 3 dB around -110 dBm, or 0.04 ϕ_0 amplitude. This tells us that the deviations from linearity in $\partial V/\partial\phi$ of this particular device become significant at a scale of about 0.04 ϕ_0.

The dynamic gain behavior of this device (MSA2W2-168F-25A) is typical of all MSAs studies in this work. RF gain was typically measured at -120 dBm input signal power.

Almost without fail, devices would show diminished gain around -110 dBm input power, and around -90 to -80 dBm, gross non-linear effects would become apparent.

7.4 Input Coil Characteristic Impedance Z_0

In our MSAs, the input signal is coupled from a 50-Ω transmission line to the resonant microstrip over the SQUID through a capacitor (or varactor), and the end of the microstrip is coupled to ground though a variable capacitance to achieve tuning. In both cases, the capacitance for optimal coupling and tuning depends on the characteristic impedance Z_0 of the microstrip resonator line. Sections 5.3 and 5.4 cover how the nominal input microstrip impedance Z_0 is calculated. The reduced varactor capacitance range described in section 7.1 results in reduced tunability. Recall from equation 4.52 that an increased Z_0 will allow a greater tunability for a given capacitance range. This motivates the increase in design impedance seen in tables 5.6, 5.7, 5.8, and 5.9. Early designs had a nominal design impedance of about 120 Ω, while later designs had a nominal design impedance of about 160 to 290 Ω.

Here we empirically test the impedance on the MSA input coil by coupling the MSA to the input line with a series of fixed capacitors, recording the frequency of peak gain and Q for each coupling capacitor, and performing a numerical fit. The end of the input coil was left floating, with an estimated 0.1 pF capacitance to ground through the un-connected bonding pad. Figure 7.15 shows empirical data for a device with a nominal design Z_0 of 281 Ω and $\lambda/2$ frequency of 1237 MHz in white and a numerical model of a linear resonator with a Z_0 of 84 Ω and unloaded $\lambda/2$ frequency of 1166 MHz in red, calculated as per section 5.1. The error bars in capacitance are derived from the manufacturer's tolerance of ± 0.1 pF. The error bars in frequency and Q reflect the variance in values due to MSA tuning parameters–for each fixed coupling capacitor tested the current bias and flux bias was adjusted for both maximum gain and lowest T_N, and those measured values plotted as the error bar endpoints.

The fit in frequency-vs-input capacitor is remarkable–the extrapolated unloaded frequency differs from the design frequency by only 6%, so that the product $L_\ell C_\ell$ differs by only 3%. In contrast to the nearly spot-on frequency, the empirical Z_0 is about *3.3 times smaller* than the nominal design value. Given the close match in frequency, this tells us the actual L_ℓ is about 3.3 times smaller, and the actual C_ℓ is about 3.3 times larger, than the nominal design values. This is not a trick of fine-tuning a fit–figure 7.16–shows the same data and a numerical model with f_0 and Z_0 equal to the nominal design values. There is no arguing *this* model fits the data. We can start to unpack this apparent high C and low L–in this test the MSA has an active washer and is operated in a positive-feedback mode. The feedback is capacitive between the SQUID washer and input coil, so during a positive-going swing (net positive charge) in the input coil voltage, the SQUID washer will also increase in voltage, reducing the electric field between the input coil and washer, effectively increasing the capacitance. We can estimate the size of this effect by considering the input and output impedances of the MSA and the measured power gain. The input power is $P_{in} = V_{in}I_{in} = V_{in}^2/Z_{in}$, the output power is $P_{out} = V_{out}I_{out} = V_{out}^2/Z_{out}$, and the gain is

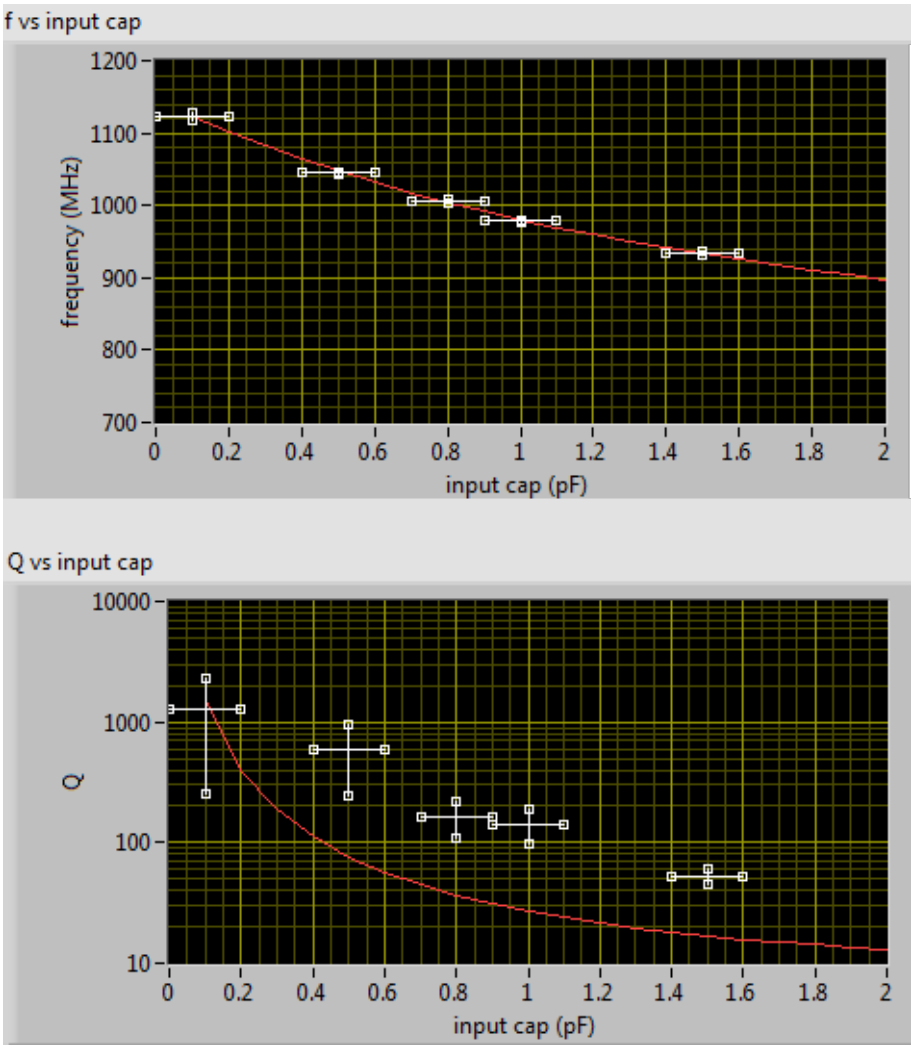

Figure 7.15: Resonant frequency and Q of the MSA input coil as a function of C_{in}. White points are the measured peak gain frequency and Q as a function of the input coupling capacitor. Error bars in capacitance are derived from the manufacturer's tolerance of $\pm$ 0.1 pF. Error bars in frequency and Q are derived from variation in values due to choice of SQUID bias parameters. This device has a nominal designed $\lambda/2$ f_0 of 1,237 MHz and Z_0 of 281 Ω. The fit (red) is a model of a 1-D linear resonator with an f_0 of 1,166 MHz and Z_0 of 84 Ω.

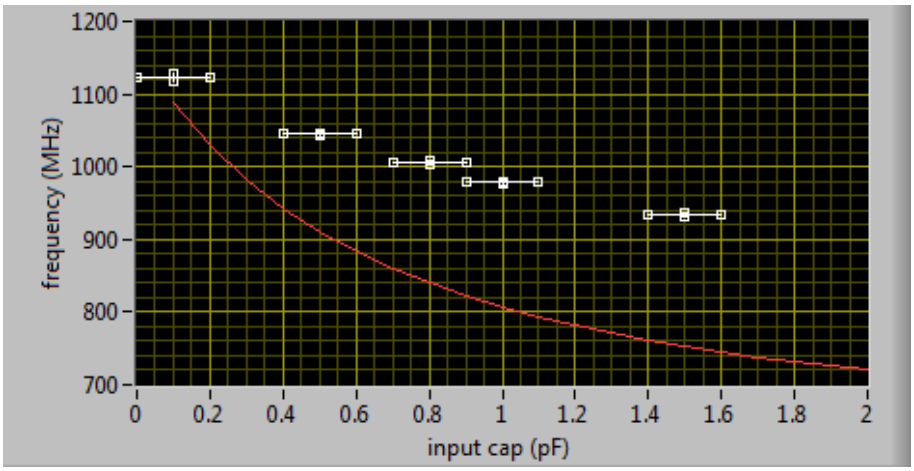

Figure 7.16: Resonant frequency of the MSA input coil as a function of C_{in}, expected and actual performance. White points are the measured peak gain frequency as a function of the input coupling capacitor. This device has a nominal designed $\lambda/2$ f_0 of 1,237 MHz and Z_0 of 281 Ω. The model (red) is a model of a 1-D linear resonator with an f_0 and Z_0 equal to the nominal design values. It is clear that the unloaded design frequency is well-matched, but the impedance Z_0 is far from the design value.

$G = P_{out}/P_{in}$. The ratio of output voltage to input voltage is then

$$\frac{V_{out}}{V_{in}} = \sqrt{G\frac{Z_{out}}{Z_{in}}}, \tag{7.2}$$

and with a typical gain of 20 dB (G = 100), Z_{in} equal to the measured 84 Ω, and Z_{out} measured from the SQUID I-V at about 10 Ω, the voltage ratio comes to 3.45, close to the correct value to account for a 3.3 times increase in the effective input capacitance.[3] The reduced effective L is not so readily explained. We only note that our nominal L is derived assuming perfect flux coupling $\alpha = 1$ between the input coil and SQUID washer. While the Meissner effect certainly helps focus magnetic flux and achieve a greater α than would be observed with a similar coil-and-washer geometry made with normal metals, ($\alpha = 1$) is always an upper bound, and the actual coupling constant will necessarily be $\alpha < 1$.

The effects of positive feedback are also seen in the measure of Q shown in the lower graph of figure 7.15. Here the only loss modeled is through coupling to the real impedance of the 50-Ω input line. Q approaches ∞ as the coupling capacitor approaches zero, and decreases as coupling to the 50-Ω environment becomes stronger. Notably, the measured Q is five to ten times greater than can be accounted for by losses to the 50-Ω line. This shows that positive feedback adds power back to the input coil at a rate equal to 80% to 90% the rate of power loss through the input line. An input capacitor of 0.1 pF linking an 84-Ω

[3]Admittedly this is a crude model that doe not account for, for example, the relative phase of V_{in} and V_{out}, or the contributions of the 50-Ω input and output lines.

transmission line to a 50-Ω transmission line will reflect about 99.6% of the incident power, for $Q \approx 1500$ in the absence of other loss mechanisms. Our data show a rough concordance of measured and predicted Q when the input capacitor is 0.1 pF, so that we conclude that all internal loss mechanisms (radiation, shunt losses, near-field coupling, etc.) siphon out about five to ten times the power lost through a 0.1 pF input capacitor, or about 2-4% of the resonant power per cycle, or set a max passive Q of about 155 to 310. As the input coupling capacitor in increased, the loss through the input line increases until it is the dominant loss mechanism. Optimal coupling occurs when the internal losses are equal to the input coupling losses. Figure 7.17 shows gain and noise temperature T_N for various input coupling capacitors, and the highest gain is observed with a 0.5 pF coupling capacitor. Interestingly, the lowest T_N is observed when the device is slightly over-coupled, with an input coupling capacitor of 0.8 pF.

7.5 Gain Corrections to the Y-factor T_{sys} Measurement

The Y-factor method of measuring T_{sys} presumes a power output from the amplifier in question that is a sum of some intrinsic noise and a noise power proportional to the heated load temperature. This characterization presumes that the gain (dP/dT) is constant, as is the internal instrinsic noise of the amplifier. Any deviation in gain for different temperature points will manifest as an error in the calculated T_{sys}. Our RF setup for characterizing MSAs (recall figure 6.19) allows us to simultaneously and independently measure the gain and noise power of our MSA, and we quite serendipitously observed gain drifts that affect our calculation of T_{sys}.

Here we describe the effects of those gain drifts in detail and how we compensated for them in our "gold standard" T_{sys} measurements. Figure 7.18 shows the recorded raw gain (a) and noise power spectra (b) for a device LFF-12H recorded at five heated load temperatures (111 mK, 514 mK, 1000 mK, 1460 mK, and 2002 mK). Note that the gain reported in the figure is "raw" gain, without a proper calibration applied it is not the absolute gain. Since the Y-factor measurement is not sensitive to absolute gain, it is not necessary to include, and even preferable to omit, any application of baseline calibration or other "mixing in" of other datasets. The absolute gain is about 20 dB. At first blush all appears well: the gain spectra appear nearly identical for each temperature, and the noise power spectra are clearly separated by temperature. (The small spikes in the power spectra are bits of the "gain sweep" signal from the VNA leaking into the PSA. We endeavor to arrange the sweep parameters of the VNA and PSA such that this visible overlap is avoided, but so long as the "contamination spikes" are well clear of the region of interest near the peak gain they do not affect the analysis.) The gain variations, though apparently small, are enough to lead to inaccurate estimations of T_{sys} if not accounted for.

Figure 7.19a shows a detailed view of the gain spectrum of figure 7.18a, and figure 7.19b

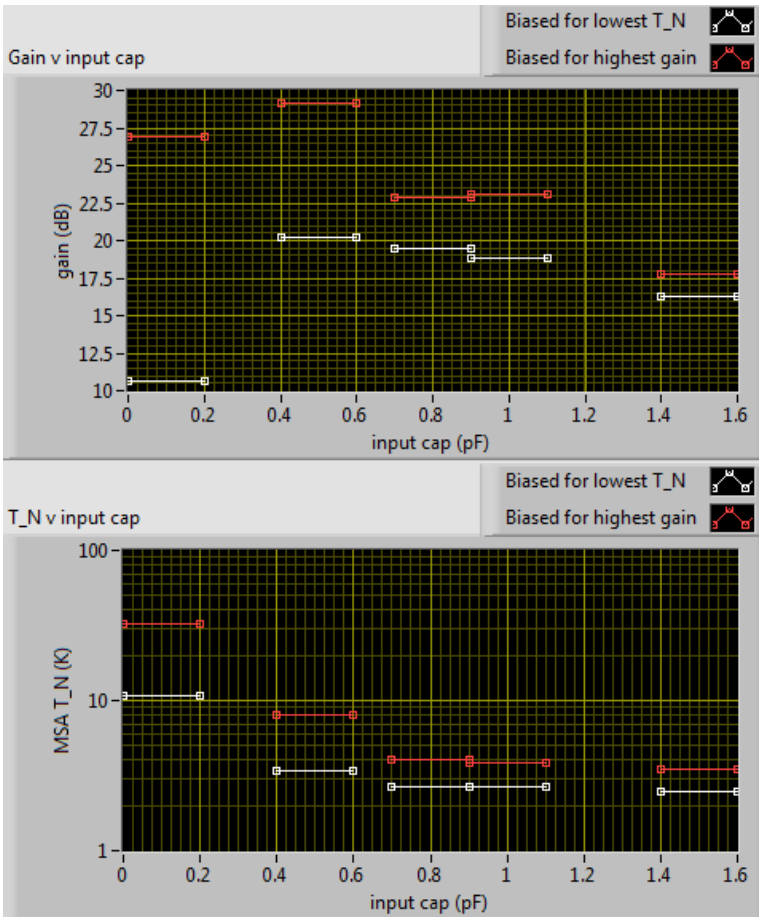

Figure 7.17: Gain and T_N as a function of input capacitor. Red points are data collected for device MSA2W2-168F-26B biased for maximum gain, and white points are data collected for the same device biased for minimum T_N. The highest gain is observed where expected, optimally coupled with C_{in}=0.5 pF, but the lowest noise temperature is measured with the device slightly over-coupled.

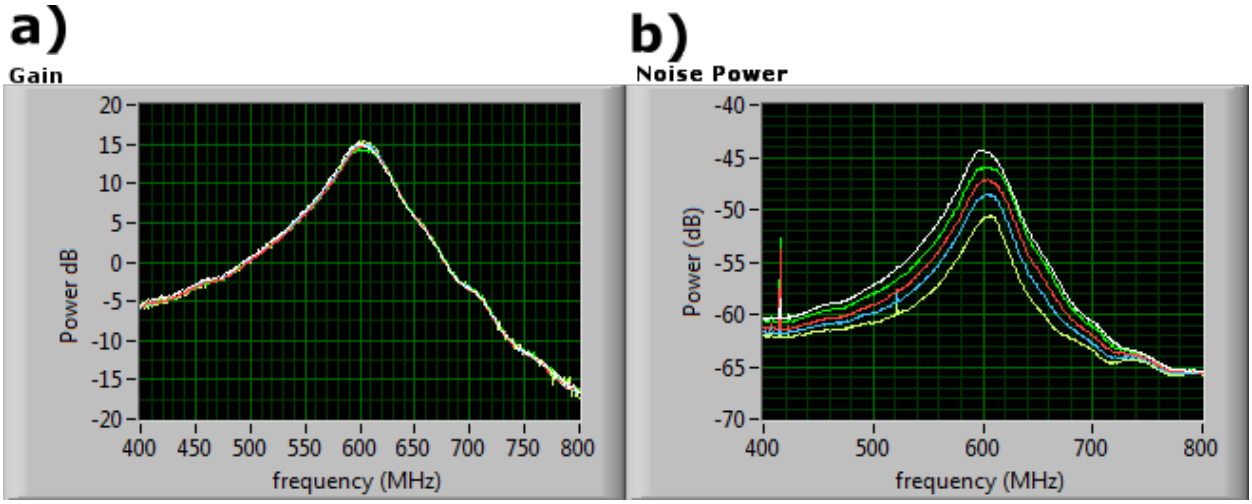

Figure 7.18: Gain and Noise power traces recorded at 5 heated-load temperatures. a) The five gain traces, which appear nearly identical. b) The five noise power spectra, which are clearly separated. The small variations in gain, barely visible here, can lead to inaccurate estimations of T_{sys} if not accounted for.

shows deviations from the averaged gain spectrum. This more clearly shows variations in gain, which vary by as much as 1.2 dB, or 32%, enough to significantly skew any T_{sys} calculation. Note for example that the highest-temperature gain trace (white in figure 7.19b) is about 0.5 dB higher than average gain at 590 MHz.

The effect of the gain drift on the power spectrum is shown in figure 7.20, along with the results of a nominal correction. In the ideal case of constant gain and T_N independent of frequency, each noise spectrum will have an identical shape and differ only by a shift in noise power. Figure 7.20a in contrast is a bit "wobbly", exhibiting a general trend in temperature, but with each spectra taking a somewhat different shape. To correct this, the gain variance spectra of 7.19b are subtracted from the power spectra of 7.20a, mathematically scaling the noise power by the relative gain, to give the corrected noise power spectra of 7.20b. The corrected spectra are much more self-similar, and are generally more symmetrical.

The top row of figure 7.21 shows the two power spectra, uncorrected (left) and gain-corrected (right), here on a linear power scale. The cursors indicate a frequency of 590 MHz. In each case (corrected and uncorrected), a linear regression is fit to load temperature vs. power at each frequency. This linear regression fit is shown explicitly for the data at 590 MHz in the second row of figure 7.21. The system noise temperature is identified with the Y-intercept of the linear regression. Note that in the uncorrected noise power data, the highest temperature trace (white) shows a "too-high" power at 590 MHz (data point too far right in the second row), which flattens out the linear fit and results in a "too-low" noise temperature. Contrast this to the linear fit of the right-hand column, where after applying the gain correction all the noise power points fall much closer to the best fit line. The bottom

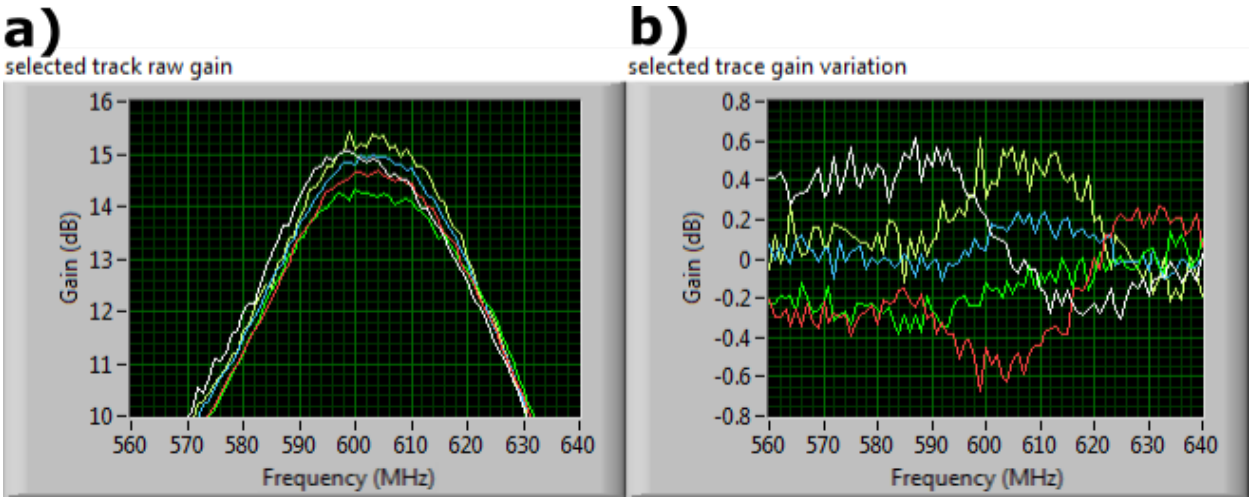

Figure 7.19: Detail of gain variations in a Y-factor measurement. a) Detail of the peak gain region. b) Deviations from the average gain. A 1.2 dB variation in gain translates to a 32% variation in absolute power.

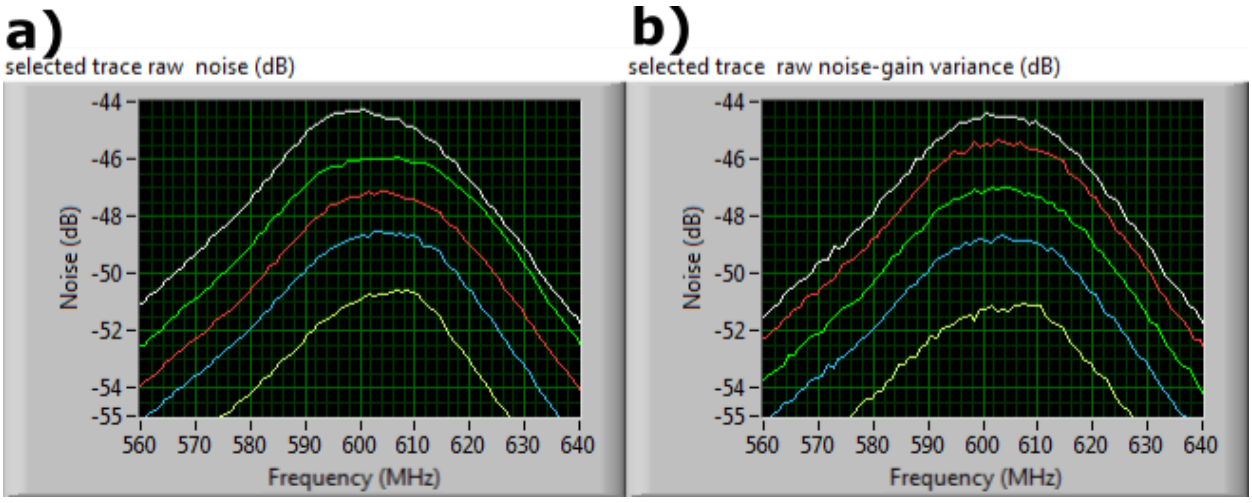

Figure 7.20: Detail of noise power spectra with and without gain correction applied. Left shows the raw noise power spectra recorded for 5 different heated load temperatures. Right shows the same data after applying the gain correction.

row of figure 7.21 shows the system noise temperature fit by this method for each frequency bin between 560 and 640 MHz. The gain-corrected T_{sys} spectrum is much flatter, and has a minimum near the peak gain frequency as expected.

We have not discovered any systematic trend in gain drift with heated load temperature. The gain is generally stable over time if the heatable load is not heated. We speculate that broadband thermal noise with a temperature of a few kelvin may be sufficient to move trapped flux in the MSA, or perhaps move trapped charges in the varactors. Regardless of the origins of the observed gain drift, we are confident that it can be corrected to deliver an accurate measure of T_{sys}.

This method of "gain-correcting" the noise power spectrum is not standard practice for a Y-factor measurement, but in the limit of zero variation in gain reduces to the canonical method. The standard Y-factor method also often relies on recording noise power from just two temperatures. If the gain is non-linear or drifting, a two-point fit will not reveal this. We propose that a more general and robust method of Y-factor measurement include multiple temperatures, so that linearity is explicitly checked and can be used as a measure of confidence in the extrapolated T_{sys}. Throughout this work it is understood that any reported Y-factor measurement relies on not less than four temperature points and has been calculated after gain-drift correction.

7.6 Comparison of T_N Measurements by S/N-ratio and Y-factor Methods

The noise temperature T_N is a critical performance parameter, and we have two primary ways of measuring it: with a variable heated load Y-factor measurement and by a signal-to-noise ratio. The Y-factor method is most accurate, but slow, and the S/N ratio method is much faster, but subject to systematic error. We wish to compare the methods and estimate the relative accuracy.

The variable heated load Y-factor measurement is the "gold standard"; its accuracy is limited only by the accuracy with which one can measure the temperature of the heated load and estimate the losses between the heated load and MSA, and depends on the linearity, but not absolute accuracy, of one's Power Spectral Amplifier (PSA). The temperature accuracy and linearity of the PSA are well within 1% according to the manufacturer's calibration certificates, and the losses between the heated load and MSA, after linear correction, are also significantly less than 1% (recall figure 6.37b). The T_{sys} calculated by the Y-factor method is entirely independent of the absolute gain and loss of components after the MSA (for example, from HEMT amplifiers or lossy cables) but is dependent on the linearity of the gain and losses after the MSA. This necessary linearity is taken for granted at the low signal powers we use in this application (typically not more than -20 dBm after all amplification stages). The Y-factor method suffers from slow data acquisition rate. The heated load must be adjusted to several temperatures and held at a constant temperature as

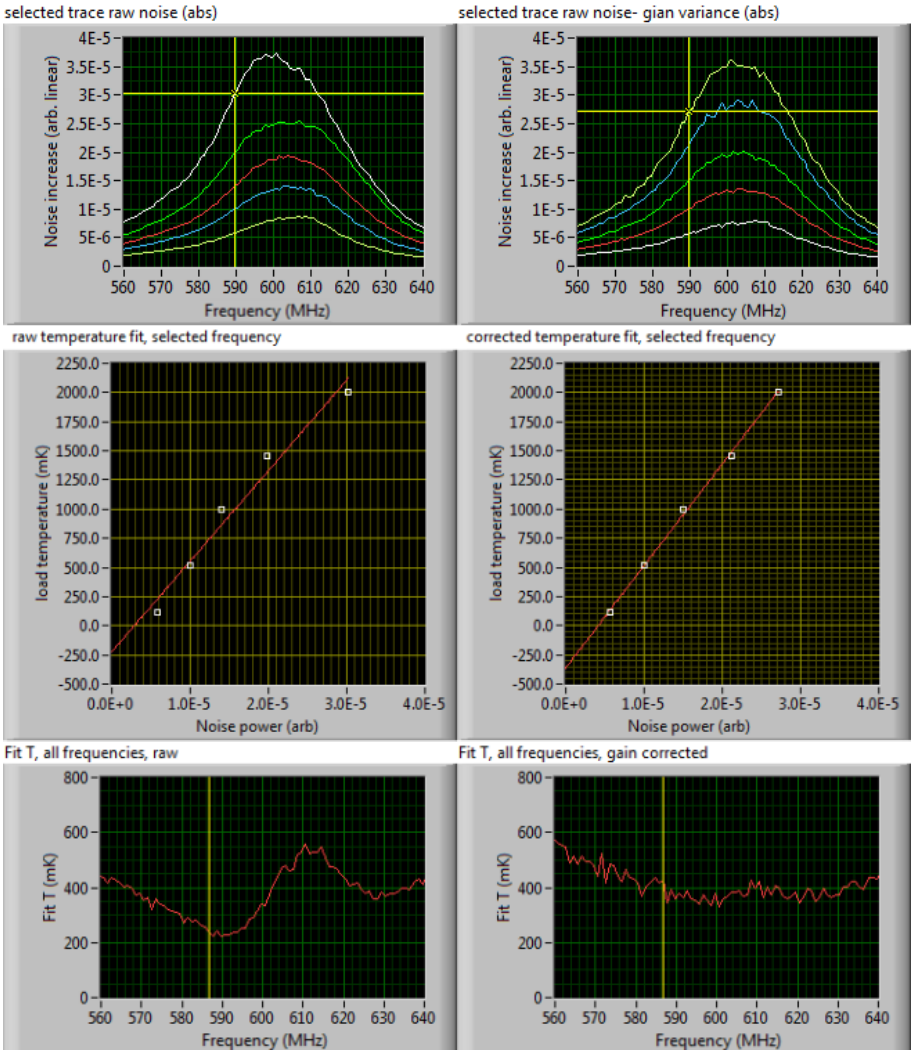

Figure 7.21: T_N calculated from a Y-factor measurement with and without gain correction. The left column shows at top the noise power spectra at five heated load temperatures on a linear scale, at center an example fit to five noise temperatures and five noise powers recorded at 590 MHz, and at bottom the system noise temperature spectrum extrapolated from all such fits. The right column is similar, but instead uses noise power data adjusted by the observed gain differences at each heated load temperature. The use of uncorrected noise power data results in relatively poor linear fits and a "wavy" noise temperature, due to underestimating T_{sys} when the gain positively correlates with temperature, and overestimating T_{sys} when the gain negatively correlates with temperature. The fits based on the gain-corrected power spectra result in a more uniform T_{sys} spectrum and, we believe, a more accurate T_{sys} calculation.

power spectrum must be integrated at each point. Each temperature point and integration takes several minutes, so a complete Y-factor measurement takes on the order of one hour. Recall figure 6.35 showing a "fast" data-acquisition example with a "nimble" heated load that demonstrates four temperature excursions of the heated load in about six hours. This is acceptable for a final check or global calibration, but is prohibitively slow if one wishes to, for example, measure the T_N of an MSA over a sweep of 625 different choices of current and flux bias.

The S/N ratio method is very fast, but is subject to more systematic errors. This method works by comparing a gain and noise power spectrum between a setup with the MSA present and active and one with the MSA replaced with an RF short. The noise power measured by the PSA with an RF short in place of the MSA is equated with the noise temperature of the HEMT immediately following the MSA (or short). With the MSA in place and amplifying, the gain spectrum will increase as will the noise spectrum. If all is well and the MSA has a lower T_N than the HEMT, the noise power will not increase as much as the gain. With this data, the system noise temperature is given by

$$T_{N\,SYS} = T_{N\,HEMT} \times \frac{\text{Noise Power Ratio (lin)}}{\text{Gain Power Ratio (lin)}}, \tag{7.3}$$

and the MSA noise temperature is given by

$$T_{N\,MSA} = T_{N\,HEMT} \times \frac{\text{Noise Power Ratio (lin)-1}}{\text{Gain Power Ratio (lin)}}, \tag{7.4}$$

where the (lin) label is a reminder that noise and gain are usually reported in dB or dBm but here we work with linear ratios, and equation 7.4 recognizes that some portion of the total system noise temperature is attributable to the HEMT noise contribution. We hope that the HEMT noise contribution is not large, but in some cases we are not blessed with a HEMT of sufficiently low-noise to justify disregarding its noise contribution. So long as the noise power ratio is about two or larger, this is sufficient for estimates of the MSA noise temperature. Figure 7.22 shows example data from one particular measure of T_N by a S/N ratio. At 600 MHz, the gain is about 25 dB and the noise increase is about 17 dB, for a S/N improvement of about 8 dB, or 6.31 times. The HEMT has a noise temperature of about 4 K, so $T_{sys} \approx 4\text{K}/6.31 \approx 0.6$ K.

The HEMT T_N used in calculating the right two panels of figure 7.22 was measured by the heated-load Y-factor method, and is shown in figure 7.23. Note the local maximum in noise power increase and T_{sys} around 670 MHz–this corresponds to a local maximum in the HEMT noise temperature, and is entirely absent from the bottom-left graph of figure 7.22, which endeavors to show only the noise contribution of the MSA.

We now compare the system noise temperature T_{sys} measured by the S/N ratio to the system noise temperature measured by a heated-load Y-factor measurement. We compare T_{sys} rather than T_N of the MSA because T_{sys} is what each method "really measures", and T_N of the MSA is attained only by further data manipulation. Figure 7.24 shows T_{sys} as measured by the two methods.

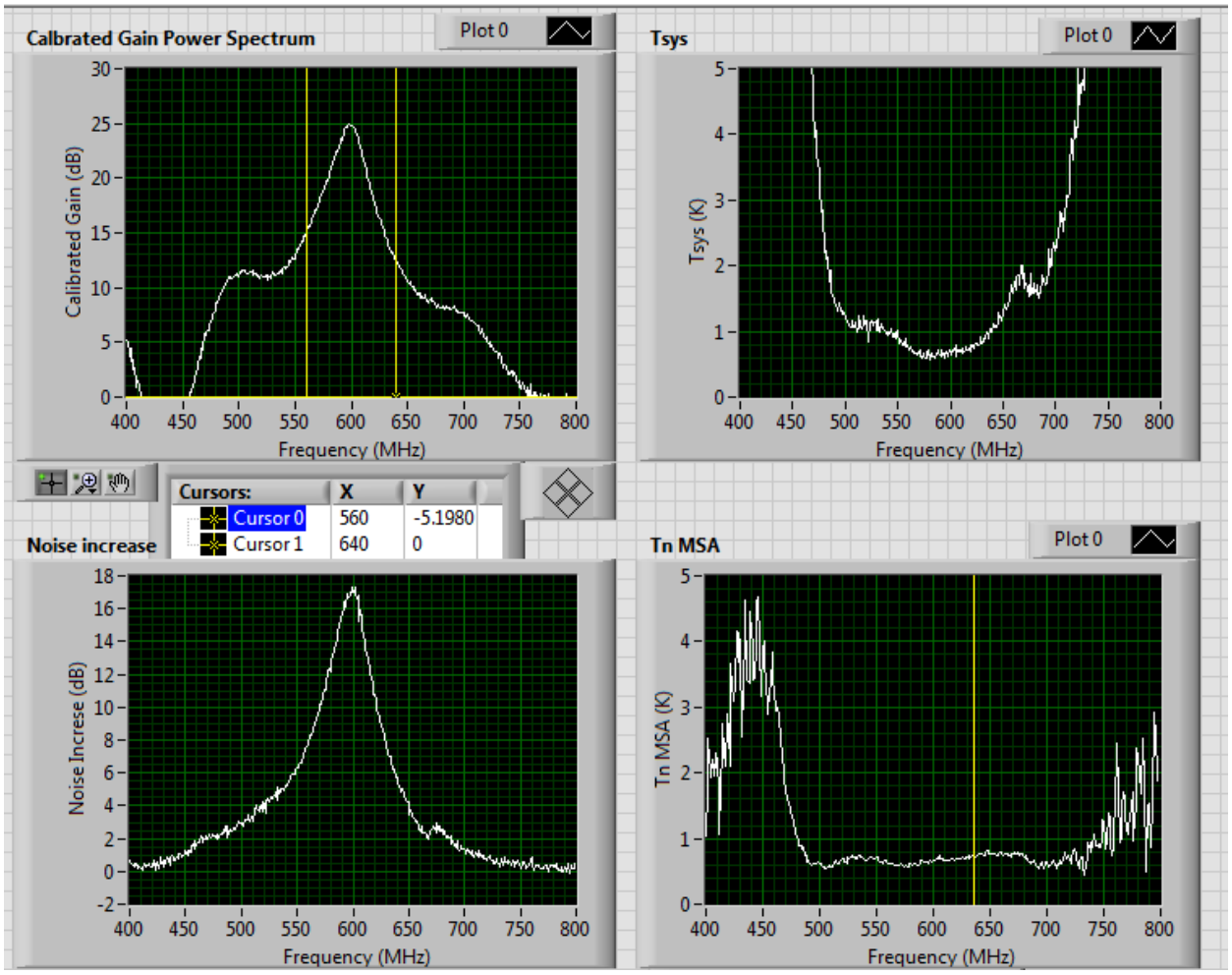

Figure 7.22: Example of a T_N estimate made by measurement of Gain and Noise Power increase. Top left shows the relative signal power of MSA active vs RF short (ie MSA gain). The peak gain is about 25 dB at 600 MHz. Bottom left shows relative noise power of MSA active vs RF short. The peak noise increase is about 17 dB. Top right is the system noise temperature calculated by equation 7.3. Bottom right is T_N of the MSA calculated by equation 7.4.

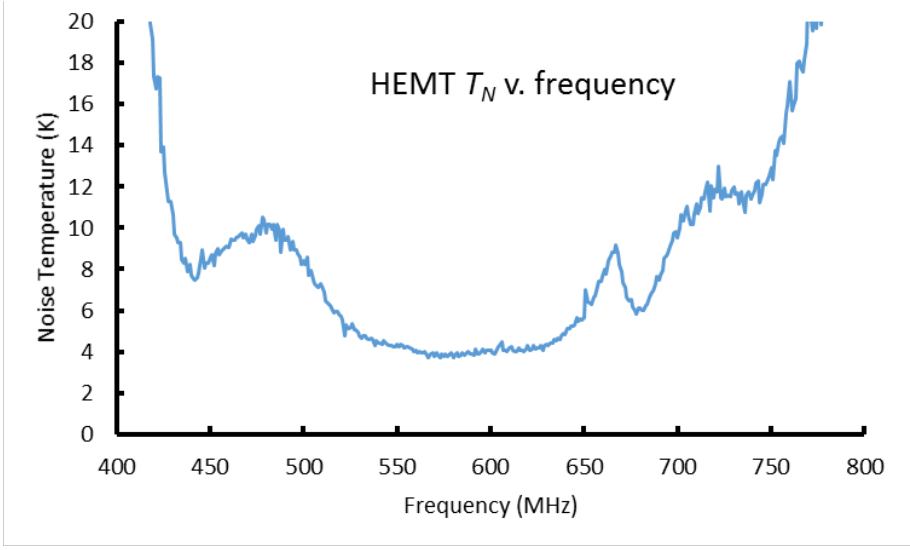

Figure 7.23: Noise temperature of HEMT "NRAO 600-7" as measured by a heated load Y-factor measurement. Strictly speaking this is the system noise temperature with contributions from the HEMT and all subsequent parts of the RF chain, with the HEMT presumably contributing the greatest portion. While it might be uesful to the HEMT engineer to separate out the contributions of the HEMT from the other components, our interest as MSA engineers is only in the noise contribution of "everything after the MSA", so the data set shown here is exactly what we find most useful.

Both methods yield a similar spectrum in T_{sys}, but the Y-factor method shows a mininmum T_{sys} of about 500 mK, while the S/N method shows a minimum T_{sys} of about 600 mK. This supports our practice of using the more convenient and much faster S/N method in most cases, particularly when exploring trends in T_{sys} as a function of, for example, RF coupling or SQUID bias parameters. We can be confident that a local minimum in T_{sys} measured by the S/N method will correspond to a a local minimum in T_{sys} measured by a Y-factor method, and that we are likely over-estimating the true T_{sys} by about 20%. We can thus learn a lot about MSA optimization quickly by relying on the S/N method, and invoke the Y-factor method only when we need a "hard, report-able" figure.

7.7 DC Characteristics

The SQUID DC characteristics are the starting point for RF performance. In operation the SQUID is biased with a current greater than the critical current, so a voltage appears across the terminals. That voltage is a function of applied flux (recall figure 4.9) and the applied bias current (recall figure 4.10). The partial derivative of voltage with respect to flux $\partial V/\partial\phi$ is the flux sensitivity, which may be positive or negative, but in general we desire a

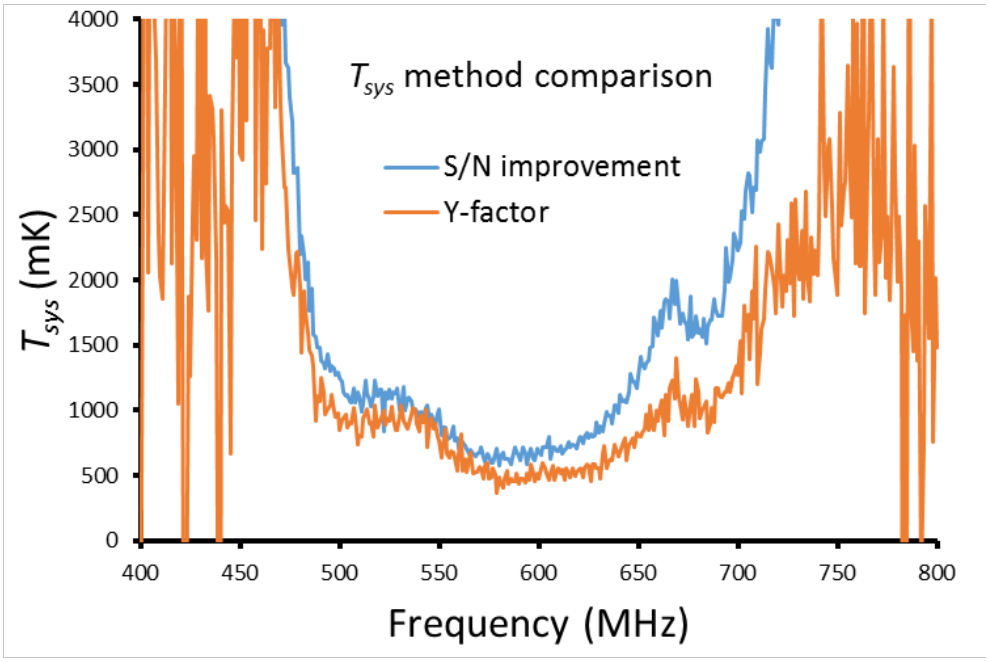

Figure 7.24: Comparison of the system noise temperature T_{sys} as measured by the Y-factor and S/N improvement methods. Both methods give similar results, but the S/N method reports a somewhat higher T_{sys} (600 mK vs 500 mK). With this result we may proceed with confidence that the much faster and convenient S/N measure can be used as a reliable figure of merit, and only invoke the Y-factor method when we need to report a "hard, reliable" result.

large magnitude for an MSA with high gain. The partial derivative of voltage with respect to bias current $\partial V/\partial I$ is the dynamic resistance, or signal output impedance, which is of importance when coupling to the 50-Ω output line. By making a map of the DC voltage landscape $V(I_{bias}, \phi_{bias})$ we can find those regions with a desirable $\partial V/\partial \phi$ and $\partial V/\partial I$ to identify promising bias parameters for RF amplification.[4] The advantage of generating a DC map is speed. We are ultimately interested in the gain and noise temperature as a function of bias parameters, but obtaining those data can take from several seconds to a minute of signal averaging, depending on the signal strength. In contrast, the measurement of a DC voltage is limited only by the bandwidth of our physical DC filters. Since our DC filters have a knee at about 10Hz, we can take data at about 3 samples/s without loss of accuracy. With a detailed map of DC performance in hand, we can narrow our search space

[4] One may object that $\partial V/\partial \phi$ and $\partial V/\partial I$ measured at quasi-DC frequencies may not be relevant to operation around 1 GHz, but we remind that person that to a SQUID, any frequency much lower than the Josephson frequency and junction plasma frequency is "quasi-DC". The Josephson inductance is $L_J = \phi_0/2\pi I_c \approx 47$ pH, and the junction capacitance is about 300 fF for a plasma frequency of about 42 GHz, and the Josephson frequency is V/ϕ_0, or about 5 GHz at a typical V of 10 μV. Since the signal frequency is 1 GHz or lower, we may be confident that our quasi-DC measurements are still valid signal frequencies.

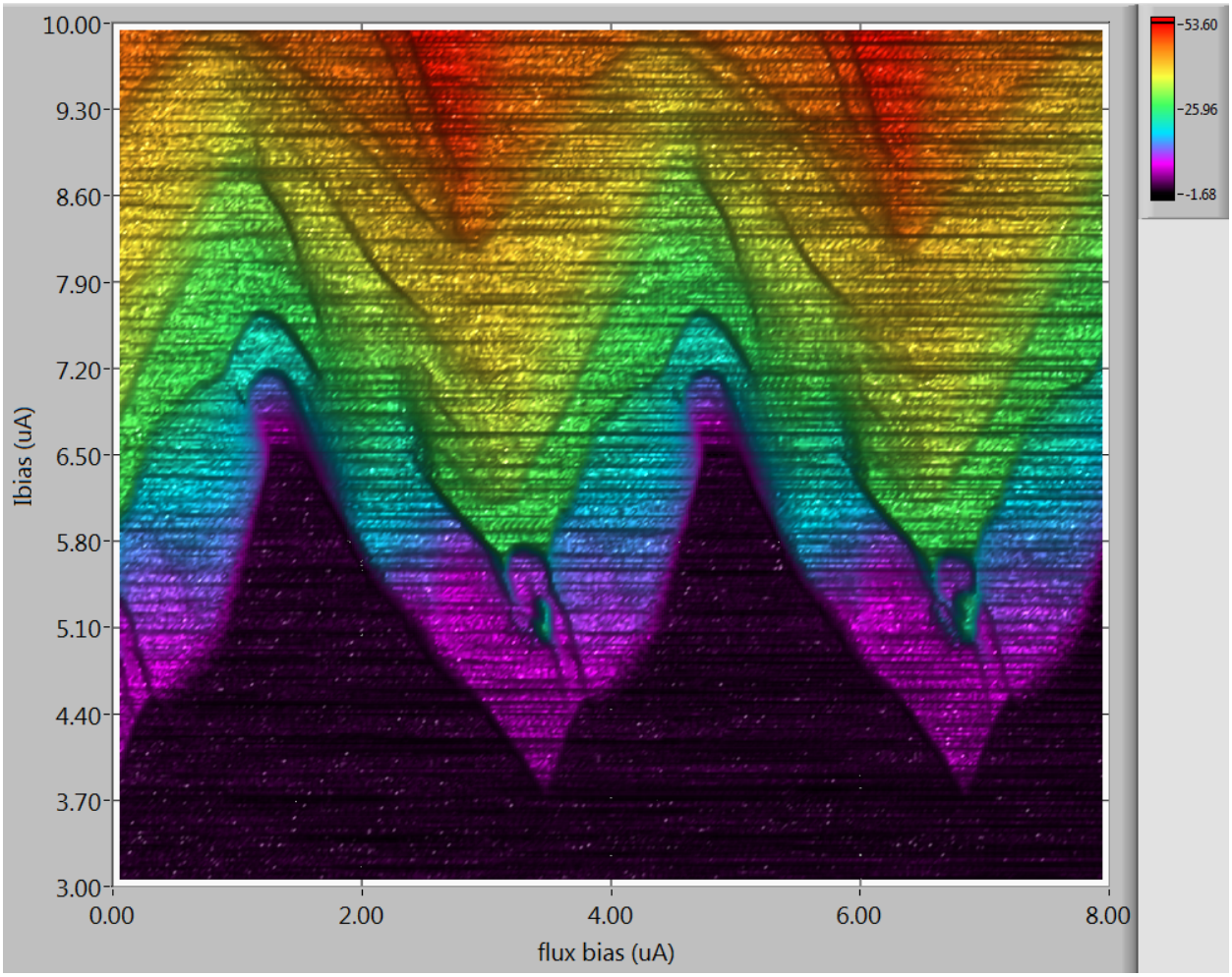

Figure 7.25: Raw DC voltage landscape of an MSA. The DC voltage of the MSA (Z and color) is plotted against bias current (Y) and flux bias (X). The flux bias ranges over a little more than two ϕ_0, the current bias ranges from 3 to 10 μA, and the voltage ranges from 0 to 53 μV. The periodicity in flux is clear from the translational symmetry, and the modulation of critical current with flux is visible in the zig-zag boundary between the flat black region and the raised colored area. As a general trend, the voltage increases with bias current. There are deep, perfectly straight horizontal lines at a rather high spatial frequency throughout the data, which are not characteristic of the MSA but are a measurement artefact.

of biasing parameters when measuring actual RF performance. An example of a detailed DC landscape is shown in figure 7.25.

Figure 7.25 represents the SQUID voltage as a color from black to red mapping 0 to 54 μV over 350 current bias points ranging from 3 to 10 μA and 400 flux bias points covering a little more than 2 ϕ_0. This map is composed of 140,000 voltage points, and represents

Table 7.1: Derivative Convolution Kernels

R_{dyn}

1	1	1
0	0	0
-1	-1	-1

V_ϕ

1	0	-1
1	0	-1
1	0	-1

about 12 hours of data-collection time as 3 points per second. (It is unusual to record a DC landscape at this level of detail, but the prettiest pictures tend to make the reports.) The voltage landscape shows a clear periodicity, and we can see a little more than two periods. The periodicity allows us to set an absolute flux scale, where one period is equal to one ϕ_0. The transition from the black area (0V) to the colored region (voltage state) is the classic I_c modulation curve, idealized in figure 4.8. We can see that the critical current modulates from about 6.5 μA to 3.7 μA, nearly a factor of two, so our β_L is reasonably close to one. There is a general trend of increasing voltage as the bias current is increased, and many plateaus and edges that modulate with flux even well above the critical current. At the highest visible spatial frequency are many hard horizontal edges–these are artifacts of the instrumentation. The voltage pre-amp used to read these voltages suffers from 1/f drift[5], and the landscape is scanned in lines of constant current bias. If the landscape were scanned from bottom to top, there would be considerable systematic error at large I_{bias}. To minimize the effect of 1/f preamp drift, the data is recorded in a boustrophedonic interlaced pattern, so sudden jumps in bias parameters are avoided and drifts at low time-frequencies are translated to high spatial-frequencies which can be filtered out by a spatial low-pass filter. Smoothing of the surface is especially important in the next step, when we consider the derivatives.

Figure 7.26 shows the data of figure 7.25 after a smoothing procedure. The smoothing procedure proceeds by first subtracting linear trends, applying a 2-D FFT, multiplying by an oblong Gaussian kernel centered at zero frequency with a width equal to 5% the Nyquist frequency in current bias and 10% the Nyquist frequency in flux bias, then applying a reverse FFT and adding back the linear trends. The sharp horizontal bands are largely suppressed, but the detailed voltage landscape of the MSA remains clear. From here the derivatives are calculated by convolving the data shown in figure 7.26 with the 3×3 kernels shown in table 7.1, and scaling by an appropriate factor calculated from the particular scale of Δ current bias per bin and Δ flux bias per bin. This method of calculating derivatives has the advantage of being unambiguous about which element to assign the calculated derivative, and acts as a final low-pass filter, averaging the derivatives of neighboring rows or columns.

A map of the derivatives is shown in figure 7.27, this time with only one period of the function shown to reduce redundancy. The sensitivity to flux ($\partial V/\partial\phi$, or V_ϕ) is shown in 7.27a and the dynamic impedance is shown in 7.27b. For a high gain MSA, we would like to chose bias parameters with a high absolute V_ϕ, which in figure 7.27a ranges from -820 to 963 $\mu V/\phi_0$, and for efficient RF output coupling we would like a dynamic impedance not too far

[5]Please let me know if you have a voltage pre-amp with stability much better than 1 μV over 12 hours!

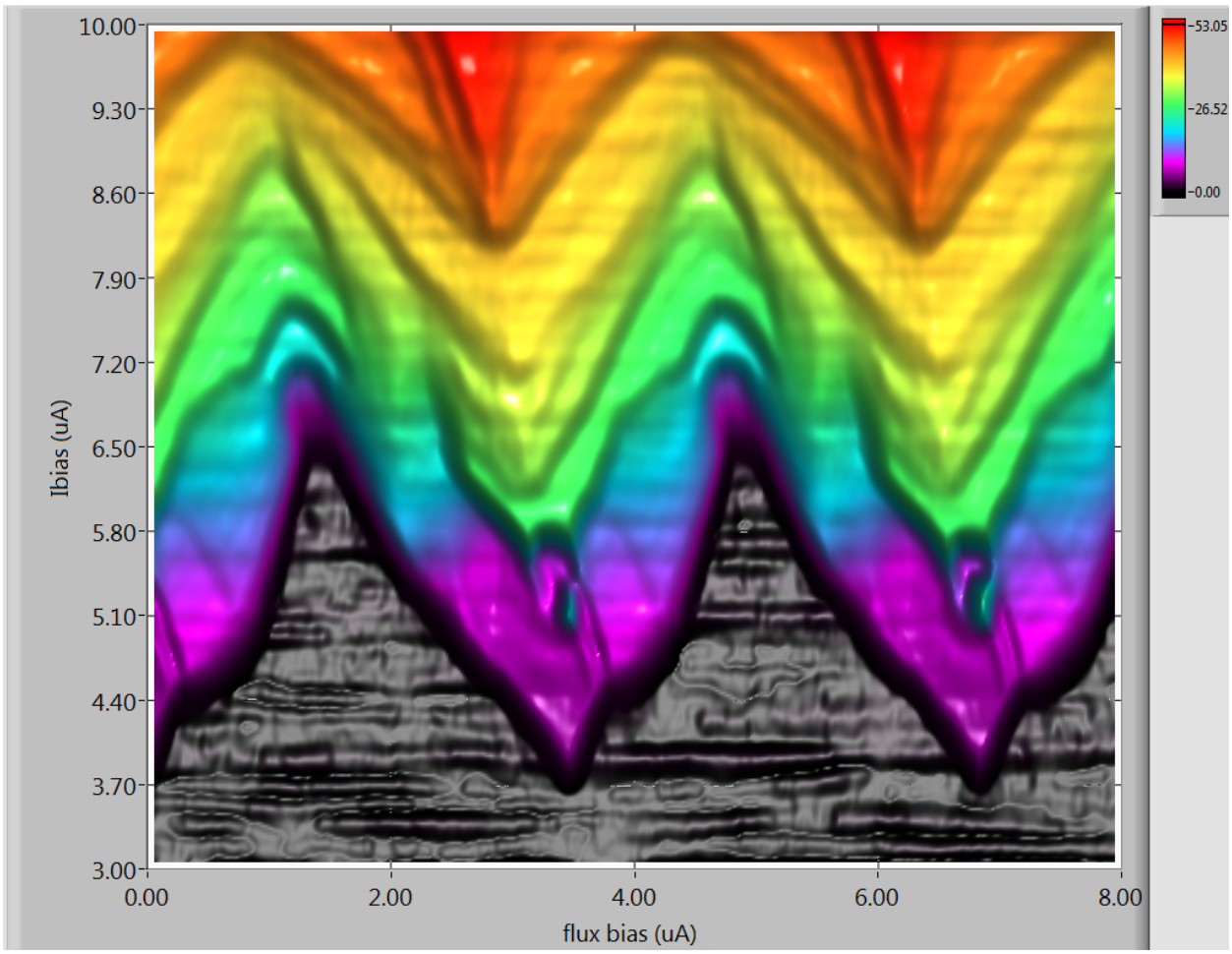

Figure 7.26: Smoothed DC voltage landscape of an MSA. The harsh horizontal lines are mostly suppressed, and the structure of the voltage landscape is more clear.

from 50 Ω. The dynamic impedance in figure 7.27b ranges from -128 Ω[6] to 231 Ω. A cursor marked with ortoganal red lines and black dot at the intersection indicates a promising bias point on both derivative plots. At a current bias of 5.84 μA and flux bias of 1.89 μA to the DC bias coil corresponding to 0.227 ϕ_0 in the SQUID, the flux sensitivity is 534 $\mu V/\phi_0$ and the dynamic impedance is 78.7 Ω. These figures are indicated near the cursors. The region of optimal bias parameters is rather narrow, and the immediate neighborhood is typically non-monotonic. One can imagine the relative futility searching for the ideal bias parameters with access to the RF performance data alone.

A more traditional way of visualizing the DC performance is shown in figure 7.28. This shows two cuts across the DC voltage landscape, one at a constant bias current of 5.84 μA, showing the sensitivity to flux, and the other at a constant flux bias of 0.227 ϕ_0, showing the

[6]For real. This DC landscape has regions with *negative* $\partial V/\partial I$.

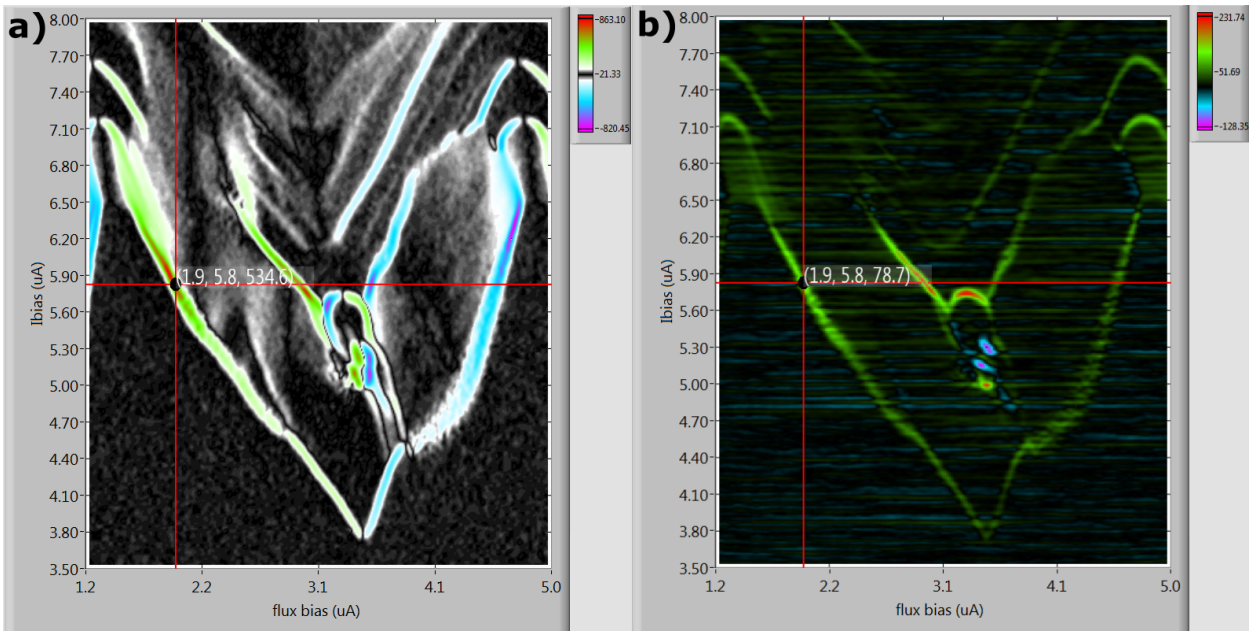

Figure 7.27: Derivatives of the DC voltage landscape. The red cursor lines indicate a promising bias point at (current to the flux bias coil, SQUID bias current) = (1.9 μA, 5.8 μA), giving V_ϕ = 534 μV/ϕ_0 and R_{dyn} = 78 Ω. The domain is restricted to a single flux quanta to reduce redundancy. (a) The derivative of voltage with respect to flux, V_ϕ. Black indicates zero flux sensitivity, and both slightly positive and slightly negative sensitivity is colored white. Moderately positive V_ϕ is green, and very positive V_ϕ is red. Moderately negative V_ϕ is blue and very negative V_ϕ is violet. Flux sensitivity ranges from 863 μV/ϕ_0 to -820 μV/ϕ_0. (b) The voltage derivative with respect to bias current, the dynamic resistance. Zero resistance is black, positive resistance ranges from green to red, and negative resistances range from blue to violet. Resistance ranges from -128 to 231 Ω.

traditional I-V plot. These are included to acknowledge the traditional way reporting DC performance parameters, and are naturally generated from the dataset of the full DC voltage landscape. This traditional way of reporting DC performance does not easily illustrate how an I-V will vary at another flux bias, or how a V-ϕ trace will vary at another current bias.

The DC performance data used in figures 7.25 through 7.28 were selected because hey are high-resolution and visually interesting. We have found that interesting DC landscapes are often also interesting in the sense of the apocryphal Chinese curse "May you live in interesting times."[7] Our most successful devices have considerably less interesting (smoother) DC landscapes. The origin of such a highly detailed landscape bears contemplation. The regions of nearly constant voltage separated by sudden steps in this landscape are reminiscent

[7]This "Chinese curse" cannot be attributed to any eastern source, but most likely originated with Sir Austen Chamberlain of the British diplomatic service in 1936. Perhaps the idea is more in line with western sentiments?

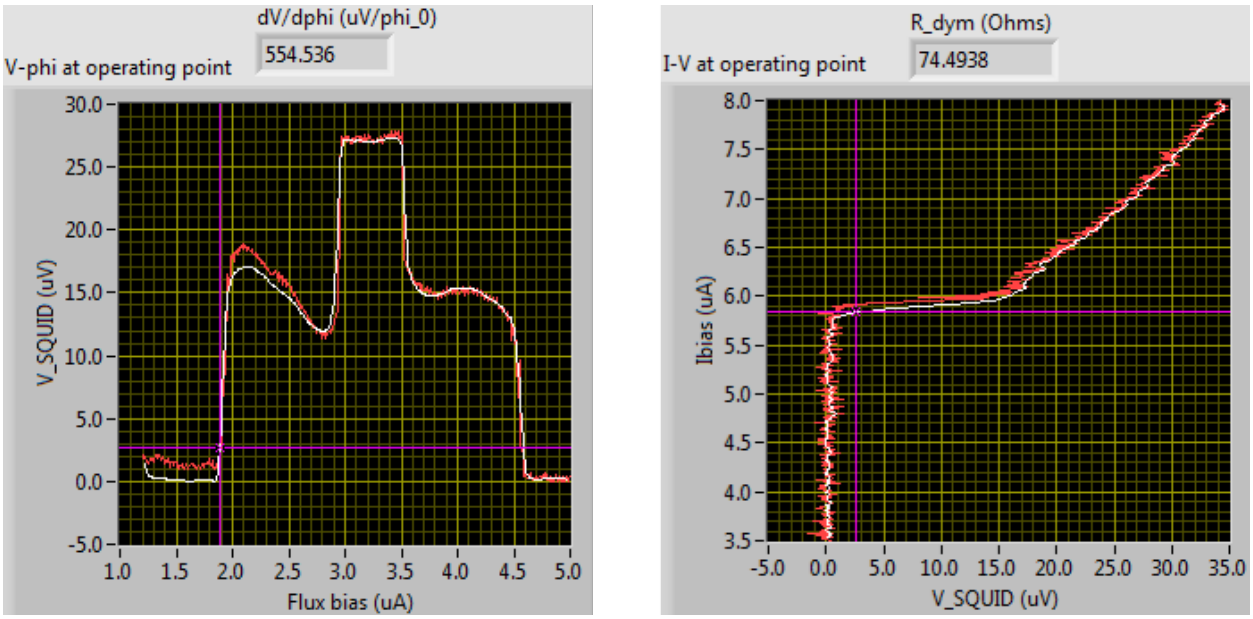

Figure 7.28: Cuts from the DC $V(I_{bias}, \phi_{bias})$ map. A cut of constant I_{bias} generates a $V - \phi$ curve, and a cut of constant flux bias generates an I-V curve. The operating point marked with the violet lines on these graphs is identical to the operating point marked by red lines in figure 7.27. The red traces are cuts taken from the the un-smoothed data set illustrated in figure 7.25, and the white traces are cuts taken from the smoothed data set illustrated in figure 7.26.

of "Shapiro steps" observed in a single Josephson junction radiated with RF energy. The voltages at which these steps appear are related by $V = \phi_0 f$, or 2.07 μV/GHz for a pure tone. Figure 7.29 shows Shapiro steps in a single Nb-Nb junction subject to two tones, with steps at voltages corresponding to the direct tones, their overtones, the difference tones, and other arithmetic combinations. The step size between plateaus in our DC landscape is about 5 to 10 μV, corresponding to frequencies of 2.4 to 5 GHz. These frequencies correspond to the first or second overtone of a standing wave on the MSA input coil, and also correspond to the typical interconnection cable length (100 mm or so) between devices near the MSA. It is not hard to imagine accidentally assembling an oscillator out of an RF environment with several available resonances coupled to a non-linear out-of-equilibrium device with a power source (the SQUID). If we are in fact generating standing waves in and around the MSA at a frequency of two or three times our operating frequency, it would not be visible to our room-temperature equipment, because those oscillations would be out of the transmission band of our cryogenic amplifiers, filters, circulators, and other equipment. This particular DC landscape was observed with the MSA connected to 6-way RF switches and circulators on the input and output. The circulators may very well be highly reflective at out-of-band frequencies, and the 6-way switches seemed to often be associated with poltergeists. Recall

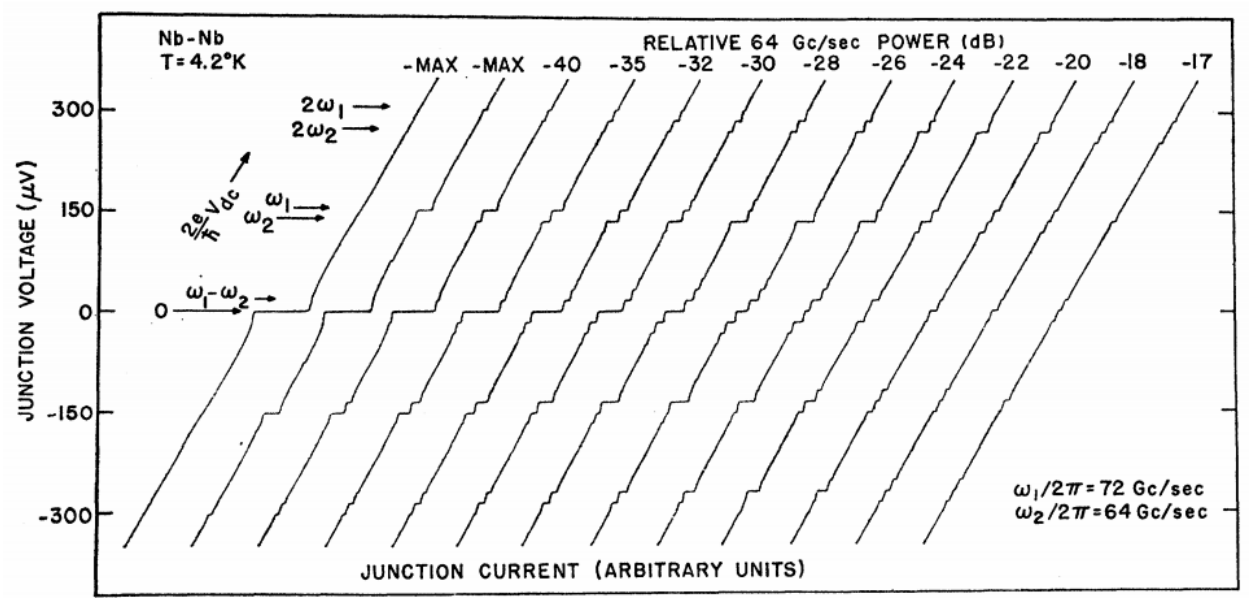

Figure 7.29: Shapiro steps in the I-V curve of a single Josephson junction due to simultaneous illumination at two RF frequencies. Note that steps appear at many (all?) combinations and multiples of the sum an difference frequencies. We speculate that the a stepped structure in the DC voltage landscape reflects simmilar mixing of multiple standing waves in the immediate RF environment of the MSA. Reproduced from [117].

that figure 6.19 of the fridge wiring shows no switches or circulators. Our best results were achieved in their absence. We did observe that the DC landscape could change with varactor tuning, (see figure 7.30) supporting the "local harmonic oscillator" model, but we have not extensively explored this connection. Rather, when we chose coupling parameters between the MSA and local RF environment we expected to be close to optimal for operation in our signal band, the DC landscapes were less interesting, and we focused our energies elsewhere.

7.8 RF Characteristics

The DC landscape of an MSA gives us a "map of the territory", including a region of likely favorable bias parameters and precise scaling of current to the DC bias coil to flux coupled to the SQUID. With that initial DC survey complete, we can turn to the RF performance of the MSA as a function of the current bias, flux bias, and varactor tuning voltage(s). Figure 7.31 shows the gain, noise increase, and S/N improvement spectra for a device at one particular set of bias parameters. We wish to explore the RF performance over a range of bias parameters (typically 625 discrete bias points per varactor tuning voltage, sometimes more, depending on resolution, and about 10 voltage tuning points), and require a more compact visualization than 6,250 or so iterations of figure 7.31.

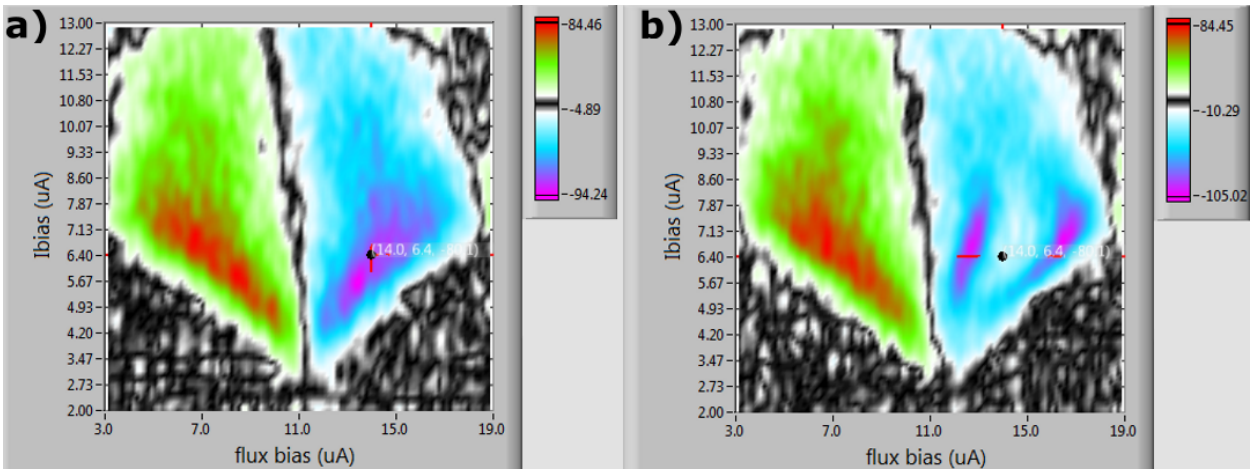

Figure 7.30: Change in the DC voltage landscape due to RF varactor tuning. Both maps show the flux sensitivity V_ϕ of device MSA2W2-168F-26B as a function of current bias and flux bias. The RF input was coupled via a fixed 0.8 pF capacitor and the end of the coil was coupled to ground through four parallel varactors. All measurements in a liquid He bath. (a) The flux sensitivity with the varactor voltage set to 0 V (lowest frequency, about 566 MHz). (b) The flux sensitivity with the varactor voltage set to -12 V (highest frequency, about 848 MHz). There is no accounting for this behavior solely in terms of the DC SQUID. This also illustrates that the optimal bias points may not be stationary as the MSA is tuned.

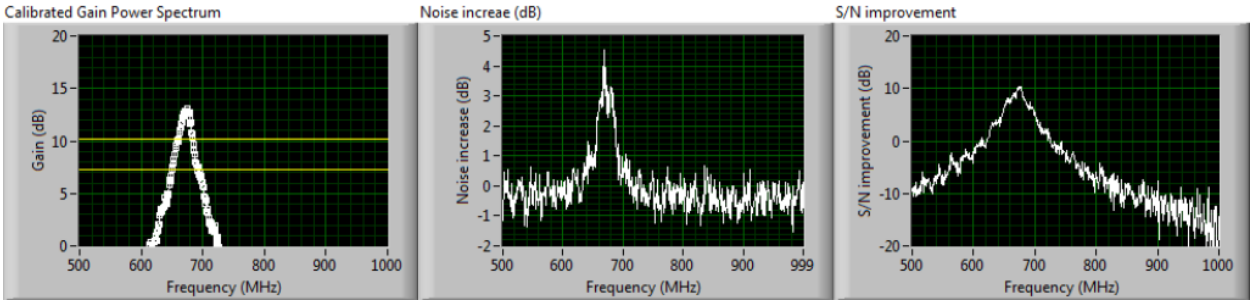

Figure 7.31: Gain, noise power, and S/N improvement spectra for one particular set of bias parameters. An exploration of RF performance as a function of bias parameters typically includes 625 current and flux bias points per varactor voltage, or about 6250 spectra data sets similar to what is shown here.

To more compactly represent the RF performance over a field of bias parameters, we extract key performance measures from each spectra set, including maximum gain, frequency of maximum gain, maximum S/N improvement, and S/N improvement at the frequency of maximum gain. These four scalar quantities can each be represented as the z-coordinate on its own field of flux and current bias points. Just such a representation is shown in figure 7.32. The four quadrants of figure 7.32 show peak gain (top left), best S/N ratio at any frequency (top right), best S/N ratio at frequency of highest gain (bottom left), and frequency of highest gain (bottom right). The flux bias is indexed on the x-axis and the current bias is indexed on the y-axis. Each point in the (flux bias, current bias) field references a full gain spectrum measured by the VNA and noise power spectrum measured by the PSA. There are 625 gain, noise power, and S/N improvement spectra indexed in figure 7.32, and the point indicated by the cursors (red orthogonal lines with a black dot at the intersection) corresponds to the spectra of figure 7.31.

We call this visualization of RF performance as a function of bias parameters the "spider-fang" plot, due to the shape reminiscent of arachnid mandibles (figure 7.33). The domain of this plot is usually zero to one flux quantum and just below the minimum I_c to about twice the maximum I_c to cover the full range of relevant flux biases and plausible bias currents.

The peak gain and S/N improvement are obviously important performance parameters to track on the spider-fang plot. The best S/N improvement at any frequency and S/N improvement at the frequency of best gain may seem like redundant measures, and indeed appear very simmilar in figure 7.32, but the utility in plotting both is to quickly identify cases when they *don't* match, so that the source of the disparity (for example a gross wiring failure) can be dealt with before performing further analysis. If all is well, the frequency spider-fang plot is flat in any region with measurable gain, indicating that the frequency is not much affected by DC bias parameters. (Outside areas of measurable gain, the peak frequency is chosen at random from a spectrum of white noise.) In figure 7.32 the peak gain, best global S/N improvement, and best S/N improvement at peak gain all occur at the same bias parameters, but this is not always the case. At the bias points indicated by the cursors in figure 7.32, the peak gain is 13.1 dB with a S/N improvement of 10.4 dB on a post-amplifier with a T_N of 50 K, for an MSA T_N of 2.42 K, which is respectable but less than extraordinary performance for an MSA in a 4.2 K liquid helium bath.

The data of figure 7.32 correspond to the exact device and setup measured to create the DC plot of figure 7.30a, with the tuning varactor set to 0 V. It is instructive to consider the spider-fang plot corresponding to the DC plot of figure 7.30b, with the tuning varactor set to -12 V. These RF data are shown in figure 7.34. Note that with the MSA tuned to its highest frequency, the "best gain" bias parameters are not much changed, but the asymmetry in the gain is now more pronounced. (The right fang of figure 7.34 is brighter than the left fang) Even more interestingly, the S/N improvement plots are "hollowed out" around the best-gain bias point, and the bias point for best S/N improvement has moved considerably. Figure 7.35 shows the gain and noise spectra corresponding to the bias parameters indicated by the cursors in figure 7.34. At the bias points indicated by the cursors in figure 7.34, the peak gain is 28.3 dB with a S/N improvement of 5.5 dB on a post-amplifier with a T_N of

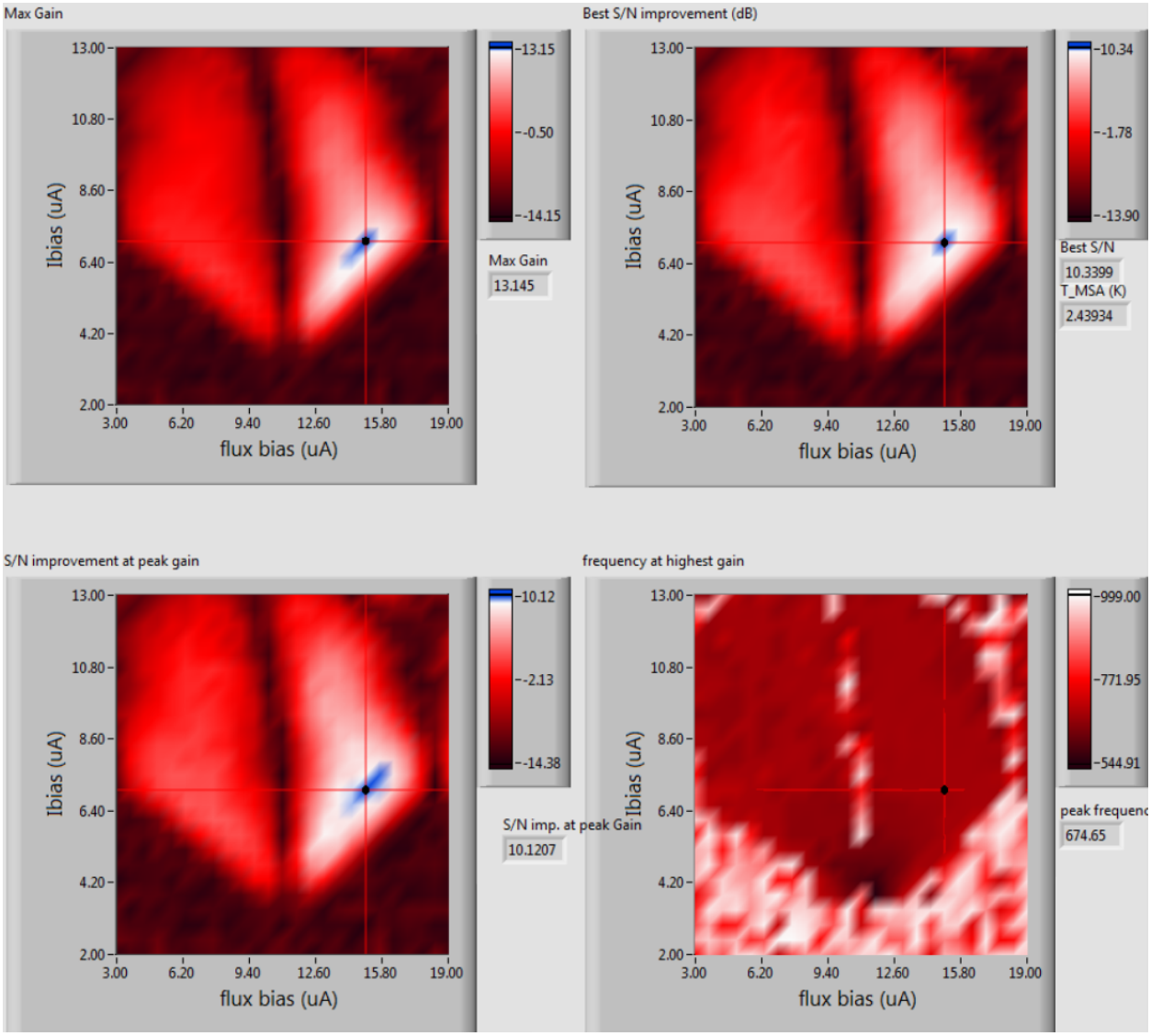

Figure 7.32: RF performance plots as a function of bias parameters, showing gain (top left), best S/N improvement at any frequency (top right), best S/N improvement at the peak gain frequency (bottom left), and frequency of peak gain (bottom right) as a function of the current and flux bias.

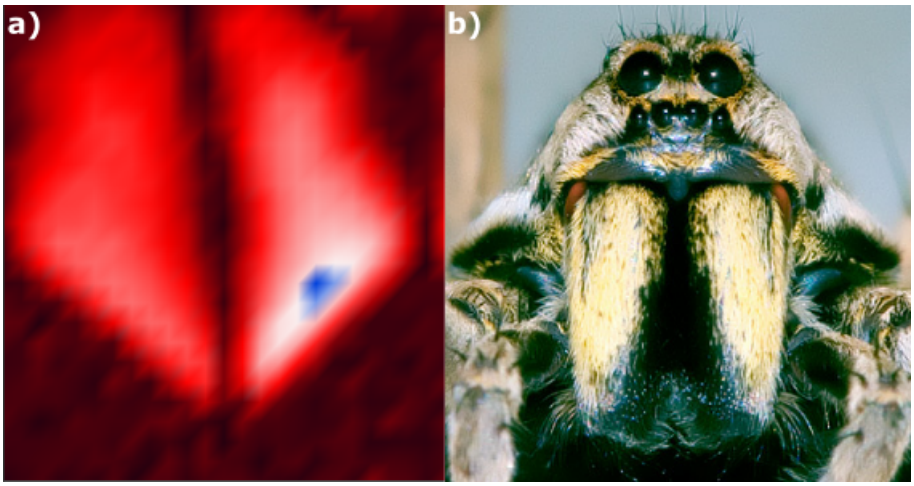

Figure 7.33: The spider-fang plot. (a) A plot of peak gain (color coordinate) of an MSA as a function of flux bias and current bias. The flux domain (x-coordinate) is close to ϕ_0, and the current domain (y-coordinate) ranges from about 2 to 13 μA. The dark regions are regions of nominal zero gain, for example because the SQUID is in the superconducting state (lower regions) or is flux biased to a minimum I_c (maximum V) so $V_\phi = 0$, for example at the central vertical ridge. Moderate gain is colored red, near-peak gain is colored white, and the very highest gain is colored blue. Greatest gain is observed where the $|V_\phi|$ is greatest, around $\phi_a = \phi_0/4$ or $3\phi_0/4$. (b) A common wolf spider, which has prominent fangs. This spider does not construct silk traps, but is an active hunter, stalking its prey and able to jump many times its body length. Truly a terrifying hunter to anyone living at the mm scale. Photo is approximately 50 times actual size. Photo courtesy of [118].

50 K, for an MSA T_N of 13.9 K, which is far short of expectations for an MSA at a bath temperature of 4.2 K. Even at the bias parameters with the best S/N improvement, the peak gain is 18.0 dB with a S/N improvement of 8.17 dB on a post-amplifier with a T_N of 50 K, for an MSA T_N of 7.26 K, which is still far short of expectations for an MSA in a liquid helium bath.

Comparing figures 7.31 and 7.35 we see that at higher frequencies the gain and Q are much higher, indicating that the feedback effects are much stronger. This is a general problem with tuning by changing the shape of the resonant standing wave–changing the shape of the standing wave also changes the degree or even sign of feedback, so the gain and noise temperature may vary in un-desired ways as the device is tuned. We can largely correct this effect by including a varactor at both the input coupling point *and* between the end of the input coil and ground.

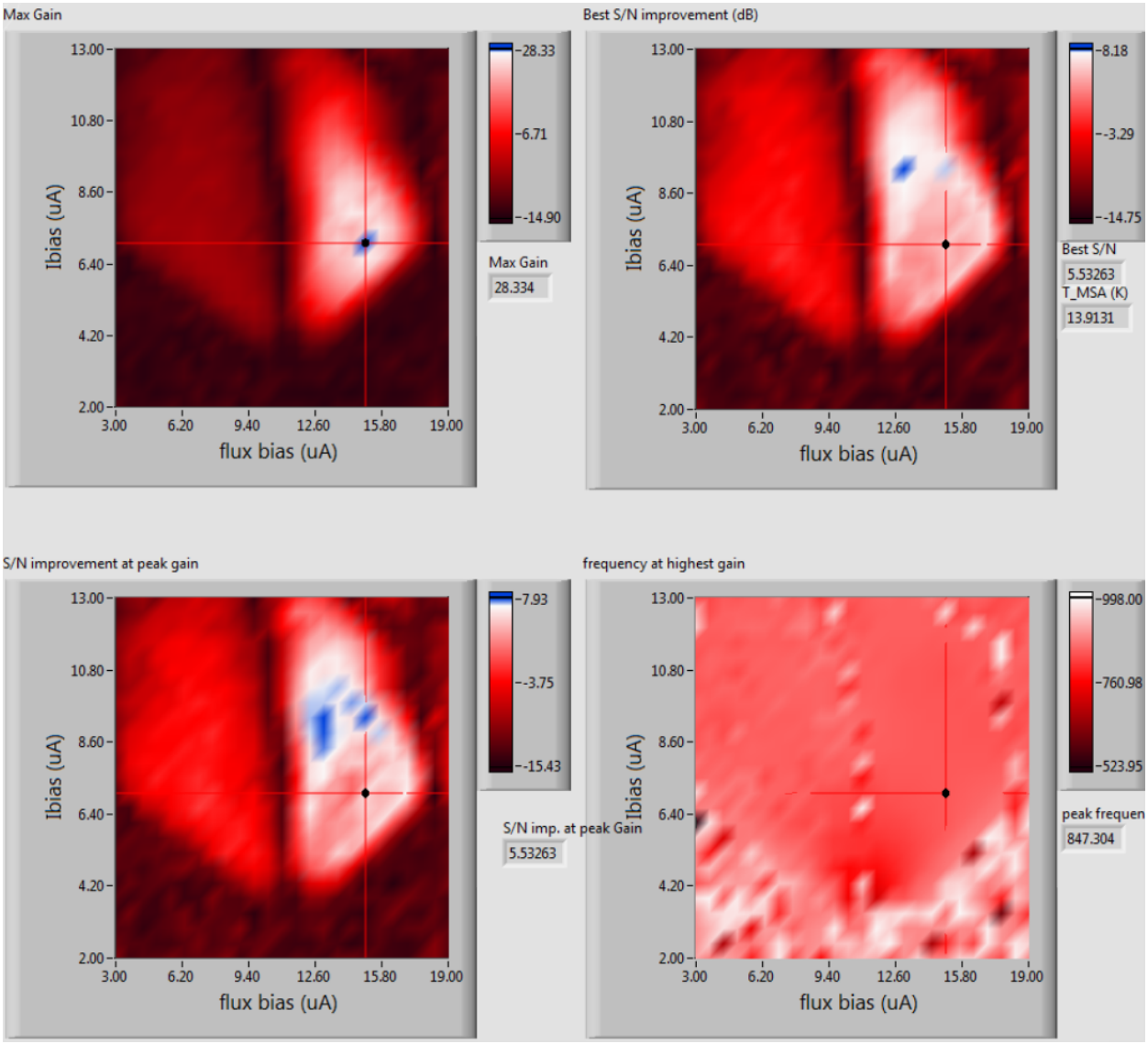

Figure 7.34: "Spider-fang" plot with the MSA tuned to maximum frequency. Note the enhanced asymmetry of the gain plot relative to low-frequency tuning, and "hollowing out" of the S/N improvement around the bias parameters for highest gain. The cursors (red orthogonal lines with a black dot at the intersection) indicate the bias with maximum gain, which no longer corresponds to the bias with best S/N improvement.

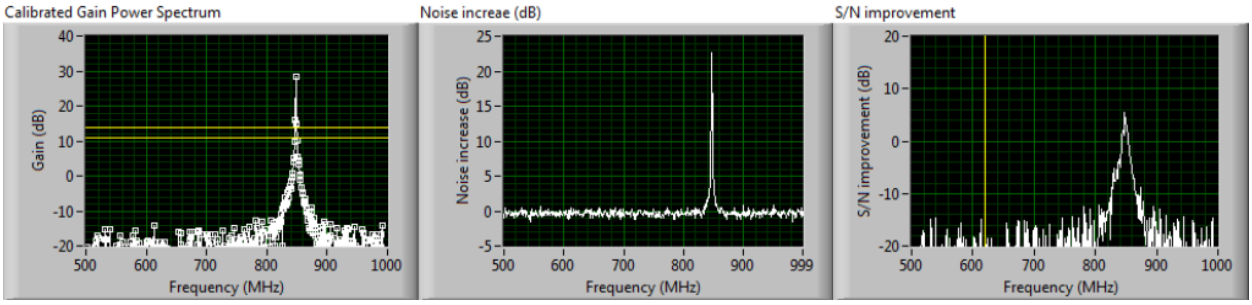

Figure 7.35: Gain, noise power, and S/N improvement spectra referenced at the bias parameters indicated by the cursors in figure 7.34.

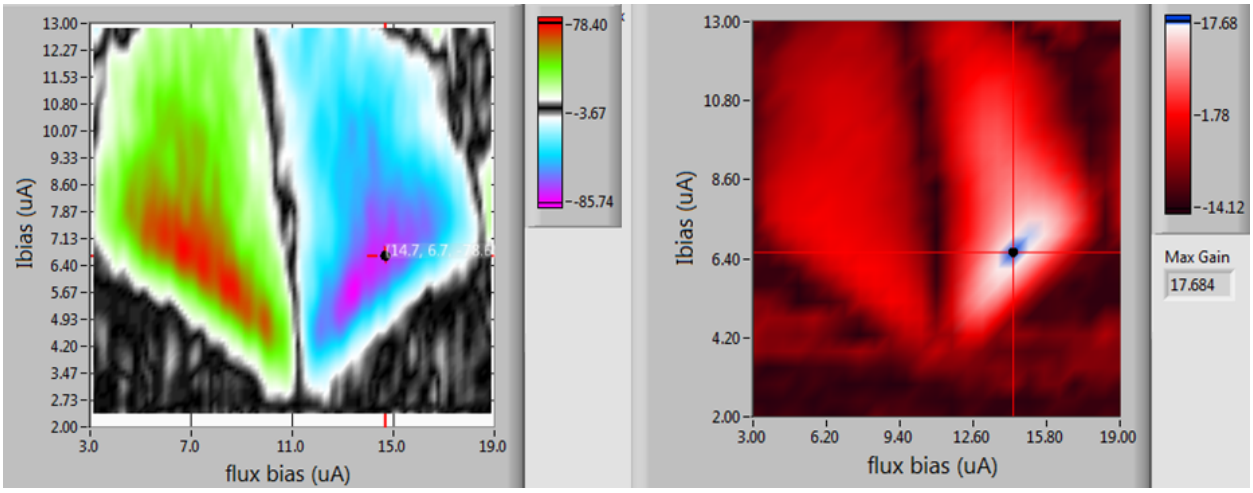

Figure 7.36: Anti-symmetry of the V_ϕ plot versus lack of symmetry in the gain spider-fang plot.

7.9 Two-Varactor Tuning

To understand the advantages of two-varactor tuning, we need to understand how tuning and bias parameters affect feedback. Consider figure 7.36 which shows the flux sensitivity (left) and gain (right) of a device as a function of bias parameters. The flux sensitivity is almost perfectly anti-symmetric, with a max V_ϕ of 78 μV/ϕ_0, and a min V_ϕ of -85 μV/ϕ_0. This suggests the gain should be about equal whether the MSA is biased around $\phi_0/4$ (red region) or $3\phi_0/4$ (violet region), albeit with a π-difference in the output phase between the two. Contrast this to the measured gain spider-fang plot, which shows a max gain of about 17 dB on the right lobe, but a max gain of only about 2 dB on the left lobe. The difference is due to the nature of the device feedback.

Figure 7.37 shows the device in question installed on the RF carrier board and a simplified schematic of the RF connections. This device has three varactors on the input (two

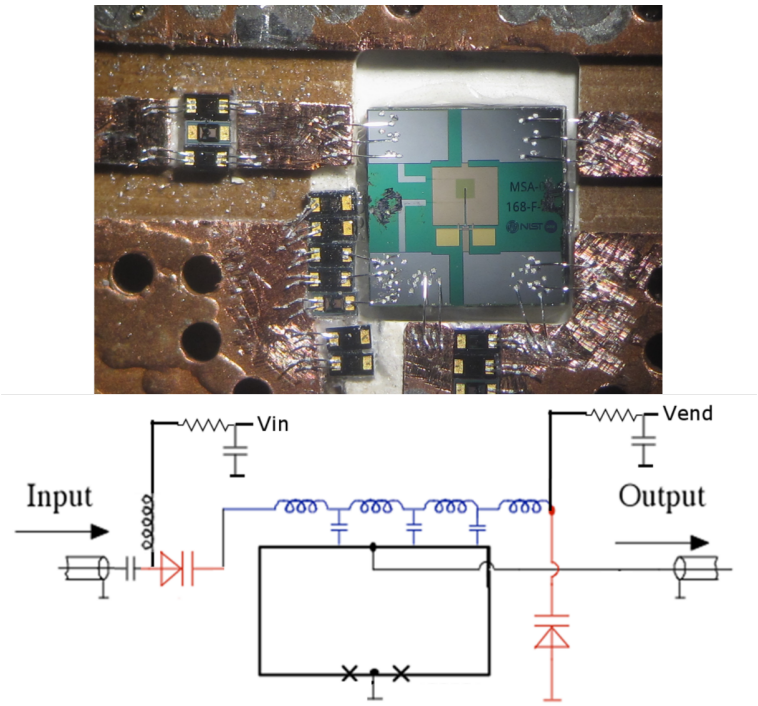

Figure 7.37: Device MSA2W2-168F-26B installed with tuning varactors at each end of the input coil. The simplified schematic shows the MSA input coil (blue) connected to the 50-Ω input line through a varactor, and the end of the input coil connected to ground through another varactor. Both varactors are independently controlled by the voltages V_{in} and V_{end}. The input coil is capacitively coupled to the "ground plane" of the SQUID washer, which is not ground at all but carries the amplified output signal.

connected) connecting the 50-Ω line to the MSA input coil and ten varactors connecting the end of the MSA input coil to ground. The SQUID voltage terminals are connected to ground and the output line. Critically, the SQUID washer is connected to the output line, so the amplified signal is capacitively coupled to the input coil. The effects this has on feedback are illustrated in figure 7.38, which shows a model of a standing wave on the MSA input coil, both in terms of the current amplitude and voltage amplitude.

Figure 7.38a shows a $\lambda/2$ standing wave, which is weakly coupled to the input (left) and has no connection at the end of the microstrip (right) so that the ends of the microstrip are current nodes, and the center is a current anti-node. The voltage standing wave (red) is also $\lambda/2$, but has anti-nodes at the ends and a node at the center. The microstrip lies over the "ground" plane of the SQUID washer but, as indicated in the diagram, the SQUID washer is not ground, but at the potential V_{out}, and assumed to be at a constant voltage across its surface at any moment in time. This voltage on the washer will couple to the voltage-component of the input coil standing wave. In figure 7.38a, the washer coupling will be positive on the left end of the input coil and negative on the right end (by the sign of the

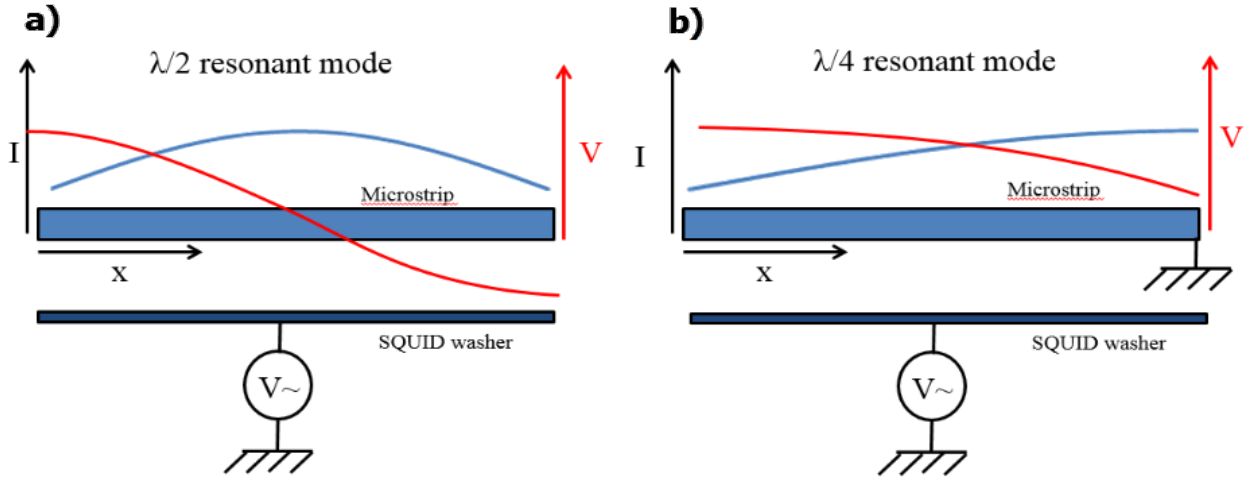

Figure 7.38: Model of a $\lambda/2$ and $\lambda/4$ standing wave on the MSA input coil. Here the spiral microstrip is "flattened out" for clarity and the entire length of the input coil is represented by a straight segment. The voltage on the SQUID washer (the amplified signal) couples to the voltage component of the microstrip standing wave. (a) With the end of the input coil floating, the voltage component of the $\lambda/2$standing wave has equal and opposite lobes ($\int_{in}^{end} V(x)dx = 0$) so the total capacitive coupling is zero. (b) With the end of the input coil grounded, the voltage component of the input coil is everywhere positive or zero, so capacitive feedback is positive.

voltage wave) for a net coupling of zero. Even for an active washer, there is no feedback to an ideal $\lambda/2$ standing wave. In contrast, consider the $\lambda/4$ standing wave of figure 7.38b–here the input is still weakly coupled, but the end is shorted to ground, for a current node and the input and current anti-node at the end. The voltage wave is an anti-node at the input, and a node at the end. Capacitive coupling to a signal on the SQUID washer is everywhere positive or zero, for a net positive coupling. With an active washer, there is positive feedback to an ideal $\lambda/4$ standing wave. It would seem that since tuning is between a $\lambda/2$ and $\lambda/4$ standing wave, the feedback must be zero or positive, but this is neglecting the reality that there must be finite coupling to the input line, so the left end of the standing wave is never truly a current node (voltage anti-node).

A model of a "$\lambda/2$" standing wave is shown in figure 7.39, now described with scare quotes because while the end may be reasonably described as a current node (as in an ideal $\lambda/2$ standing wave), the input has some finite coupling to the outside world so it is not quite a current node. As such, the voltage node is not quite at the center of the resonator but shifted somewhat closer to the input, and there is more "negative going" than "positive going" length of the voltage standing wave, or $\int_{in}^{end} V(x)dx < 0$, for net *negative* capacitive coupling, and negative feedback.

Figure 7.40 shows the feedback-with-tuning effect in practice. The left column shows

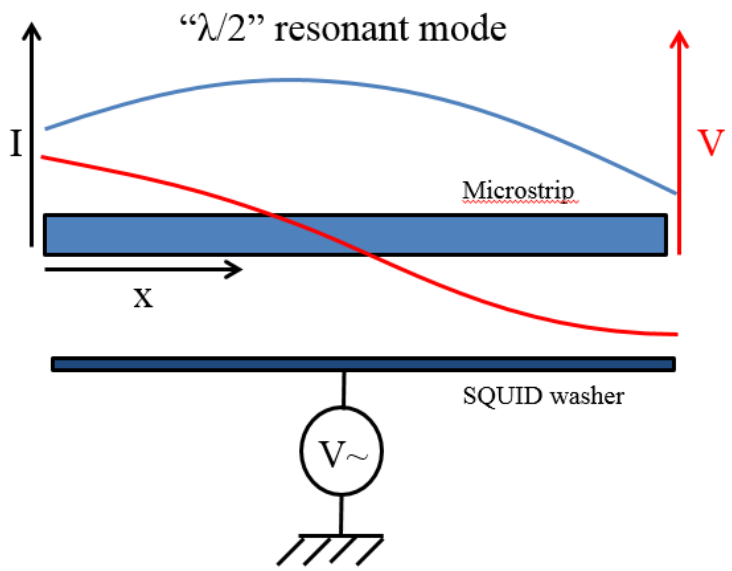

Figure 7.39: An almost $\lambda/2$ standing wave, corrected for finite input coupling. In this example, the input is not an exact node but the end is. This allows for a net *negative* capacitive feedback between the input coil and SQUID washer because $\int_{in}^{end} V(x)dx < 0$, in contrast to expectations that feedback must be zero to positive for a $\lambda/2$ to $\lambda/4$ standing wave.

model standing waves in the input coil calculated from the method described in section 5.1 based on the known coupling capacitances and external loading of the input coil at both ends, and the empirically determined input coil impedance Z_0. The center column shows the spider-fang gain plot, and the right column shows the gain spectrum measured at the bias parameters allowing the greatest gain (colored blue in the spider-fang plots).

In the first case 7.40a, the coupling capacitors are not equal, but chosen such that the phase on reflection coupled to the 50-Ω input line is nearly equal to the phase on reflection from ground. The standing wave is nearly symmetrical on the input coil, so like figure 7.38a there is virtually no capacitive feedback. The spider-fang gain plot of 7.40a is nearly symmetrical, consistent with expectations that gain correlates with $|V_\phi|$, irrespective of its sign. In the absence of feedback the gain is rather low, only about 6.4 dB, which leads to a system noise temperature that is rather high, about 10 K in this case, mostly due to a high contribution from the HEMT. In the second case 7.40b, the end of the input coil is connected to ground and the input is coupled by some moderate capacitance, for a standing wave on the input coil close to $\lambda/4$. This results in strong capacitive feedback, and a highly non-symmetrical spider-fang gain plot. The right-hand lobe is very bright due to strong positive feedback with a peak gain around 28 dB, but the left lobe is rather dim with a best gain of just a few dB because the opposite sign of V_ϕ turns strong positive feedback into strong negative feedback. With the gain so high and positive feedback so strong this device

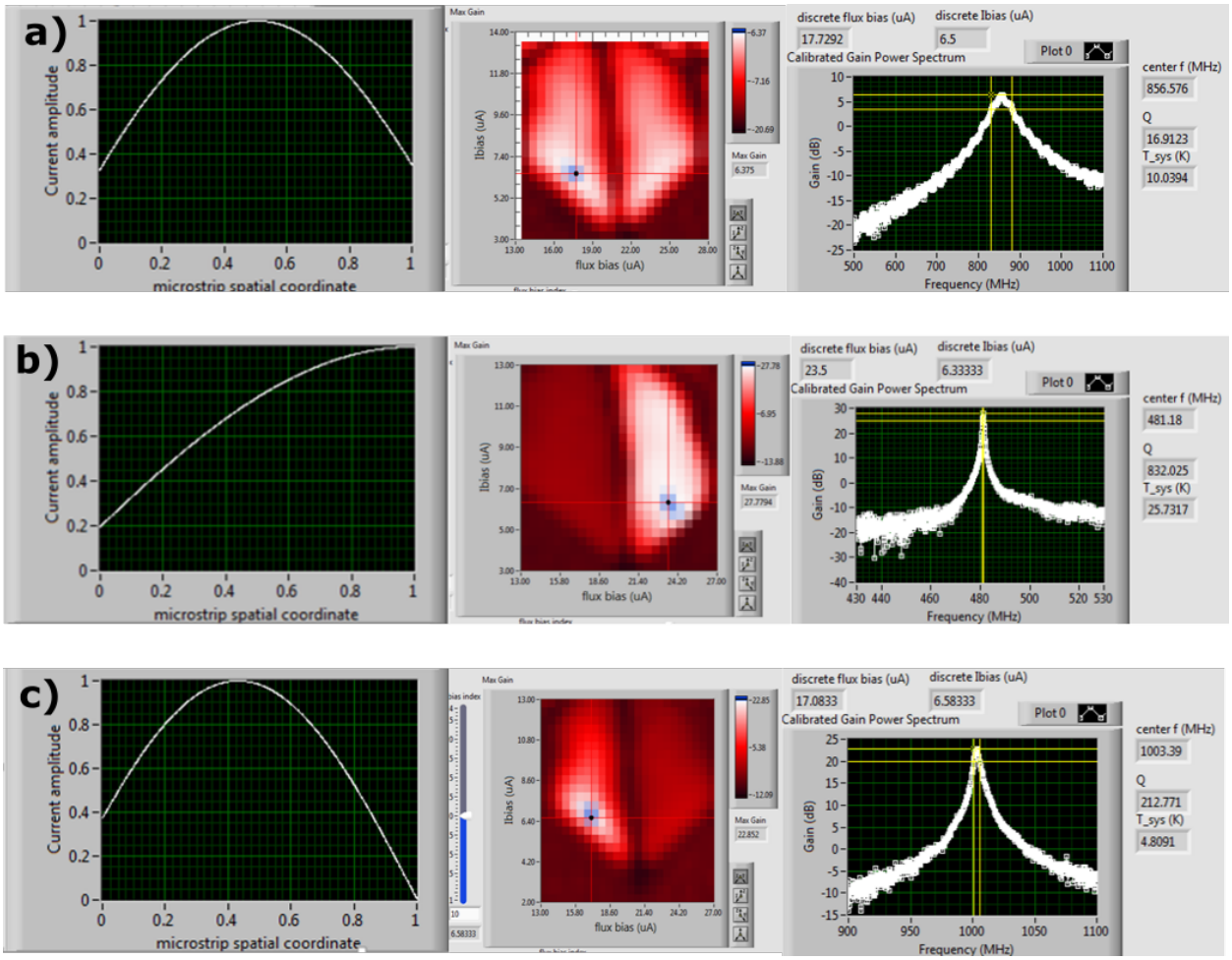

Figure 7.40: MSA input coil standing-wave shape, gain, and T_{sys}. (a) The standing wave shape is nearly symmetrical so capacitive feedback is negligible. The sipder-fang gain plot is nearly symmetrical because in the absence of feedback, the gain is ambivalaent to the sign of V_ϕ. The gain is rather low, only about 6 dB, resulting in a large T_{sys} of about 10 K. (b) The input coil end is shorted to ground, resulting in an almost $\lambda/4$ standing wave and strong capacitive feedback. The spider-fang plot is highly asymmetrical because the two lobes have opposite feedback signs. The gain is quite high (28 dB) but the system is almost self-oscillating and the noise temperature is also quite high, about 26 K. (c) The input coil end is open and coupled by a moderate capacitance to the input line. The standing wave is closer to $\lambda/2$, but shifted left as in figure 7.39. This reverses the sign of the feedback and the spider-fang gain plot exhibits greatest gain in the opposite lobe. The gain is now a more moderate 22 dB and the system noise temperature is the best of the three at 4.8 K.

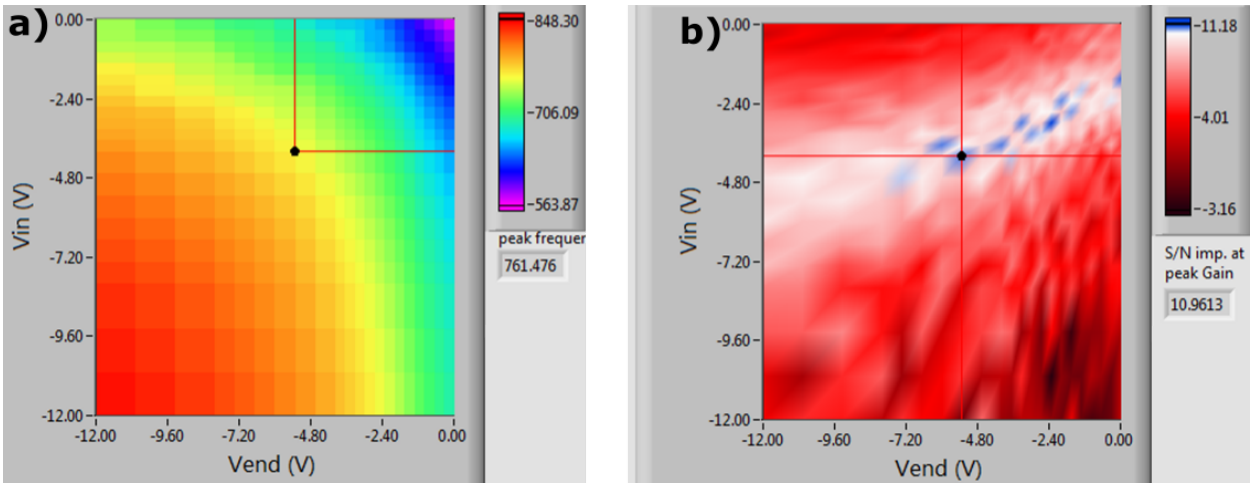

Figure 7.41: Two-varactor tuning and S/N improvement effects. (a) The tuning frequency ranges from 563 to 848 MHz (violet to red) across the available varactor tuning voltages. (b) The S/N improvement ratio as a function of the varactor tuning voltages. The "ridge" of best S/N improvement nearly spans the available frequency space.

is nearly self-oscillating and has a system noise temperature of 25 K. In the last case 7.40c, the end of the input coil is floating and the input is coupled somewhat more strongly than in 7.40a or b. The spider-fang gain plot is again asymmetrical, but the lobe with greatest gain is now the left one, because the shift of the standing wave left of center has changed the sign of the capacitive feedback. In this case of moderate feedback the gain is about 22 dB, and the system noise temperature is the best of the bunch at 4.8 K. (Lower temperatures are possible with careful choice of coupling, but the point is made.)

By controlling both the coupling capacitance between the 50-Ω input line and input coil and the input coil end and ground, we can choose the frequency of the standing wave *and* its shape, for optimal feedback and best T_{sys}. Informally, reducing the capacitance on both ends of the input coil "squeezes" the standing wave towards a $\lambda/2$ shape, increasing the frequency, and increasing the capacitance on each end reduces the frequency, while keeping the waveform centered. Adjusting the coupling capacitances in opposition shifts the standing wave right or left, while preserving the wavelength and frequency. In practice we can tune to the frequency we desire while positioning the standing wave to achieve optimal feedback and maintain a low T_{sys}. Figure 7.41 shows the tuned frequency and S/N improvement as a function of the two varactor control voltages.

On the tuning and S/N graphs of figure 7.41 the x-axis is the tuning voltage of the end varactor to ground and the y-axis is the tuning voltage of the coupling varactor to the 50-Ω input line. The greatest capacitance occurs at 0V (top and right) and the smallest capacitance occurs at -12 V (bottom and left). The tuning graph of 7.41a shows the lowest

voltage in violet at (0V, 0V) and the highest frequency in red[8] at (-12V, -12V). The device tunes from 564 to 848 MHz, and each constant-color arc represents a continuum of resonator shapes at a constant frequency. We can chose the most favorable shape at each frequency from the data of figure 7.41b, where we see that a white "ridge" of favorable S/N improvement crosses through almost the entire available frequency space. By controlling the input and end couplings independently, we can maintain optimal feedback and low T_{sys} while tuning throughout the frequency range.

7.10 Devices with Notable Performance

Here we present data from some select devices that demonstrated exceptional performance. It is important to explain that these are not "hero devices"–results presented here can be reproduced by nominally identical devices either from the same fabrication run or a later run. For example, device LFF-11A, LFF-11B, and MSA2W2-LFF-11A[9] will behave exactly the same if installed with the same coupling varactors, on the same RF carrier, in the same fridge setup. We would rarely test this device-modularity intentionally, but devices under development would sometimes not survive testing.[10] When a failed device was replaced with a nominally identical clone, we always found that measured performance was also nominally identical. For example if we show performance data for device LFF-11A[11] it may be understood that device LFF-11C (used in ADMX for the 560 to 640 MHz range) is capable of similar performance.

7.10.1 Device LFF-12H

Device LFF-12H demonstrated exceptional performance, and the related devices LFF-12A and LFF-12C were early deliveries to ADMX for operation in the 750-800 MHz 680-700 MHz range as un-tunable devices. Figure 7.42 shows a summary of the gain and T_N of the device as estimated from the S/N improvement, at a physical temperature of 60 mK. Each trace represents a different varactor tuning, and the current bias and flux bias may also differ between traces. The varactor voltages vary from -11 V (highest frequency, white) to

[8] Seems these colors should have been swapped, no?

[9] The original fabrication was a single wafer, and we did not use fabrication or wafer prefixes to identify devices. After the second fab run of two wafers, we instituted the prefix nomenclature ex: MSA2W2, meaning "second fabrication, second wafer".

[10] Almost without exception, devices would fail after a wire-bonding operation to the voltage terminals or other procedure that made electrical modifications to a point with low electrical resistance to the device. Presumably the failure was due to electrostatic discharge (ESD) across the junctions. Failure could be manifested as either a superconducting short or open circuit. Despite adhering to best ESD practices, each modification carried about a 2% chance of device failure. The only time a packaged device failed is when we applied about 1000 times too much energy to solenoid control lines of a cryogenic 6-way RF switch connected to the MSA while operating at mK temperatures. Although there is nominally no electrical connection between the control lines and the RF line, this was sufficent to knock out an MSA.

[11] May it Rest In Peace.

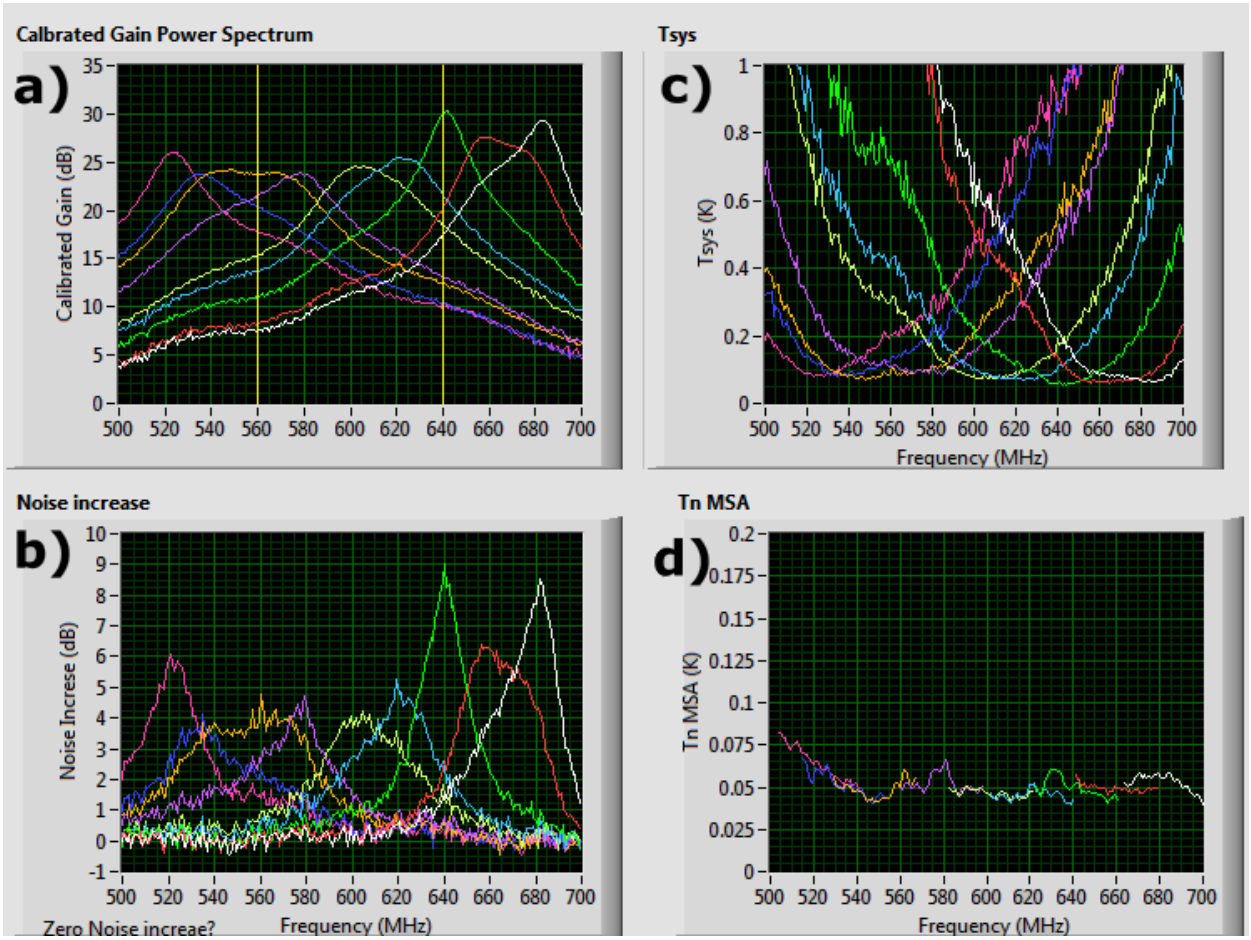

Figure 7.42: Gain, tunability, and T_N of device LFF-12H. This device has a respectable gain of 25 dB, and and even more respectable noise temperature of about 50 mK. The tuning range, from 520 to 680 MHz, is more than sufficient to cover the 560-640 MHz circulator range.

+1.2 V (lowest frequency, violet). At positive forward-biasing voltages, the varactors start to conduct and contribute to the noise noise. At a bias voltage of 0.7 or less, the noise is close to optimal.

Figure 7.42a shows the gain spectrum at 9 tuning/bias points. The gain ranges from 24 to 50 dB, and the peak frequency ranges from 520 to 680 MHZ. The yellow lines indicate the circulator frequency range we wish to match. Figure 7.42b shows the noise increase simultaneously measured with the gain. The noise increase spectra seem to follow the gain spectra, suggesting a consistent noise temperature. Figure 7.42c shows the system noise temperature calculated from the S/N improvement and estimated HEMT noise temperature. The minimum T_{sys} appears nearly constant. Figure 7.42d shows the estimated T_N of the MSA after subtracting the estimated HEMT contribution from the measured T_{sys}, with each

of the spectra truncated to the neighborhood near its minimum value for clarity. Except at the lowest frequencies where the varactors are starting to forward-conduct, the MSA T_N appears to be consistently around 50 mK. Because this is an estimate from S/N improvement and subject to systematic errors (like the HEMT noise temperature estimate) it cannot be taken as an authoritative measure of T_N, but it is sufficient to classify the performance as "really good."

The low S/N noise estimate motivates a heated load Y-factor measurement, with the results shown in figure 7.43. This test was performed at a physical temperature of 150 mK and a single set of tuning parameters for a peak gain around 625 MHz. The red trace shows the T_{sys} spectrum calculated from the Y-factor measurement, with a minimum of about 200 mK. The green trace is the estimated HEMT contribution, and the blue trace is the estimated MSA T_N calculated from the difference of measured T_{sys} and estimated HEMT contribution. The white line shows the standard quantum limit $T_Q = hf/k_B$ for the noise temperature of a phase-preserving amplifier. The difference between the red and green traces is rather small–in some places less than the statistical noise in either one, leading to crossover! When the HEMT contribution is a large part of the total noise, it is not really honest to claim a well-known MSA noise temperature, but we can claim without ambiguity that the MSA noise temperature is certainly less than 200 mK, and "by eye" appears consistent with a noise temperature of 50 mK.

7.10.2 Device LFF-11A

Device LFF-11A was designed for a slightly higher frequency range than LFF-12H, and the tuning range and noise temperature in a 4.2 K liquid helium bath is shown in figure 7.44. This device was coupled to the 50-Ω input line by a fixed 1.8 pF capacitor and tuned with two varactors coupling the end of the input coil to ground. The tuning is from 525 to 725 MHz, which is a good fit to the 644-714 MHz circulator range of ADMX. The gain is a little low at the higher frequencies, but this testing was done before we implemented 2-varactor tuning, so we suffered from frequency-dependent gain. The gain is also a little "wavy" with an period of about 25 MHz, indicating an impedance mismatch at the ends of a cable about 4-m long, consistent with the length of coax between the top of our 4-K probe and the MSA. With a noise temperature around 2 K (about half the bath temperature) and good tunability, we proceeded to testing in the dilution fridge.

On this dilution fridge run we suffered from a DC wiring failure. The voltage tuning of the varactors was entirely un-responsive, and the current and flux bias lines were connected by some finite conductivity, making a plot of DC characteristics as a function of independently controlled flux and current bias impossible. Nonetheless, we were able to bias this MSA to a favorable point and measure the gain and S/N improvement. With very little else to do, we measured the performance of this device at several fridge temperatures ranging from 50 mK to 4.2 K. Our measurements from this "heating study" are shown in figure 7.45. The gain remains fairly constant around 20 dB, but shifts from about 660 MHz cold to to 575 MHz warm, with most of the tuning happening between 200 mK and 2 K, and most spectra

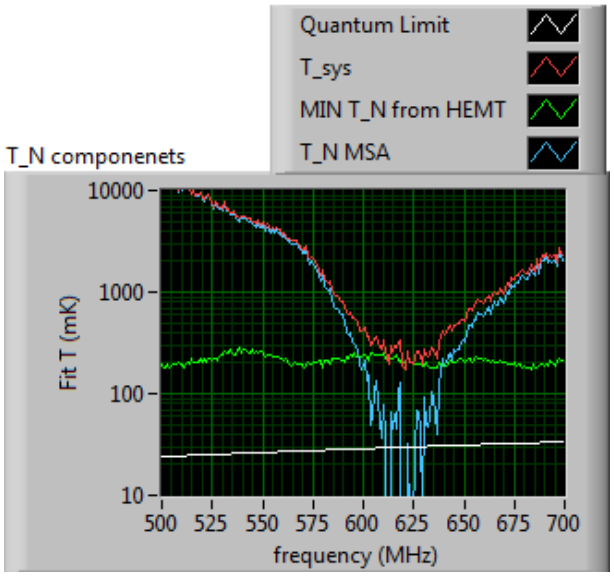

Figure 7.43: Y-factor heated load T_{sys} measurement of device LFF-12H. T_{sys} (red line) has a mimnumum around 200 mK, and is dominated by the HEMT contribution (green line). The white line indicates the standard quantum limit $T_Q = h \times f/k_B$. The blue line shows an attempt to extract T_N of the MSA by differencing the red and green lines, but because the statistical noise is in some places greater than the difference, the results are somewhat dubious. We can certainly claim that T_N of the MSA is less than 200 mK, and 50 mK appears to be a reasonable best guess.

clustering into two modes. This temperature-tuning was mysterious at the time, but is now understood to be due to the varactor temperature dependence described in section 7.1. The noise calculated from the S/N improvement rises with temperature, but remains relatively low. Figure 7.46 shows a summary of the noise temperature versus physical temperature.

The noise temperature remained always below the physical temperature, and was typically less than half the physical temperature. The trend of tracking the physical temperature seem to break only when the noise temperature approaches the quantum limit–at a physical temperature of 50 mK, T_N is 46 mK, and the quantum limit at this frequency of operation (665 MHz) is 32 mK.

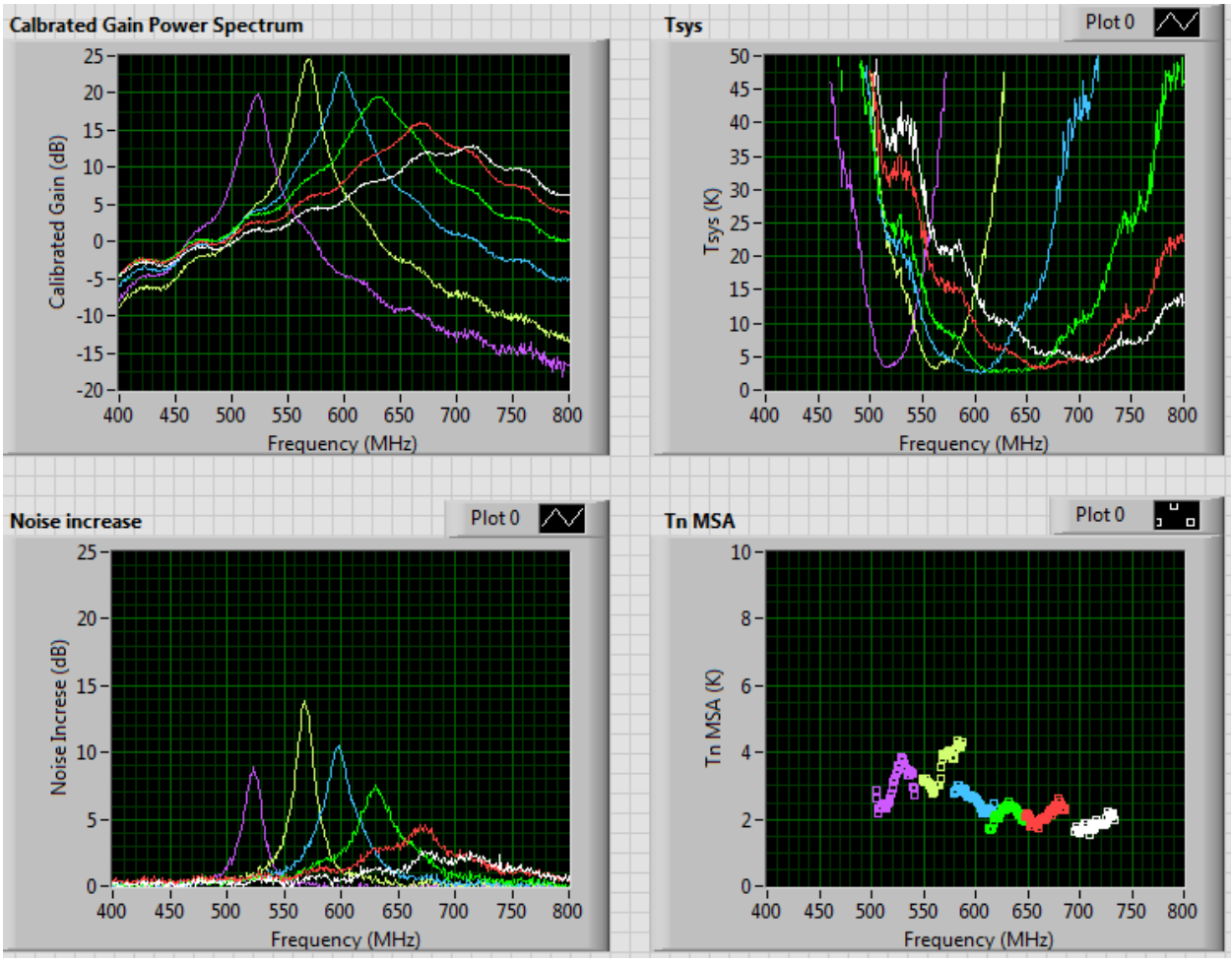

Figure 7.44: Performance of device LFF-11A at 4 K. The tunability of this device is from 525 to 725 MHz, a good match for the 2nd circulator range in ADMX (644-714 MHz). The noise temperature is around 2 K in a liquid helium bath, so with these performance parameters, it should be tested at milikelvin temperatures.

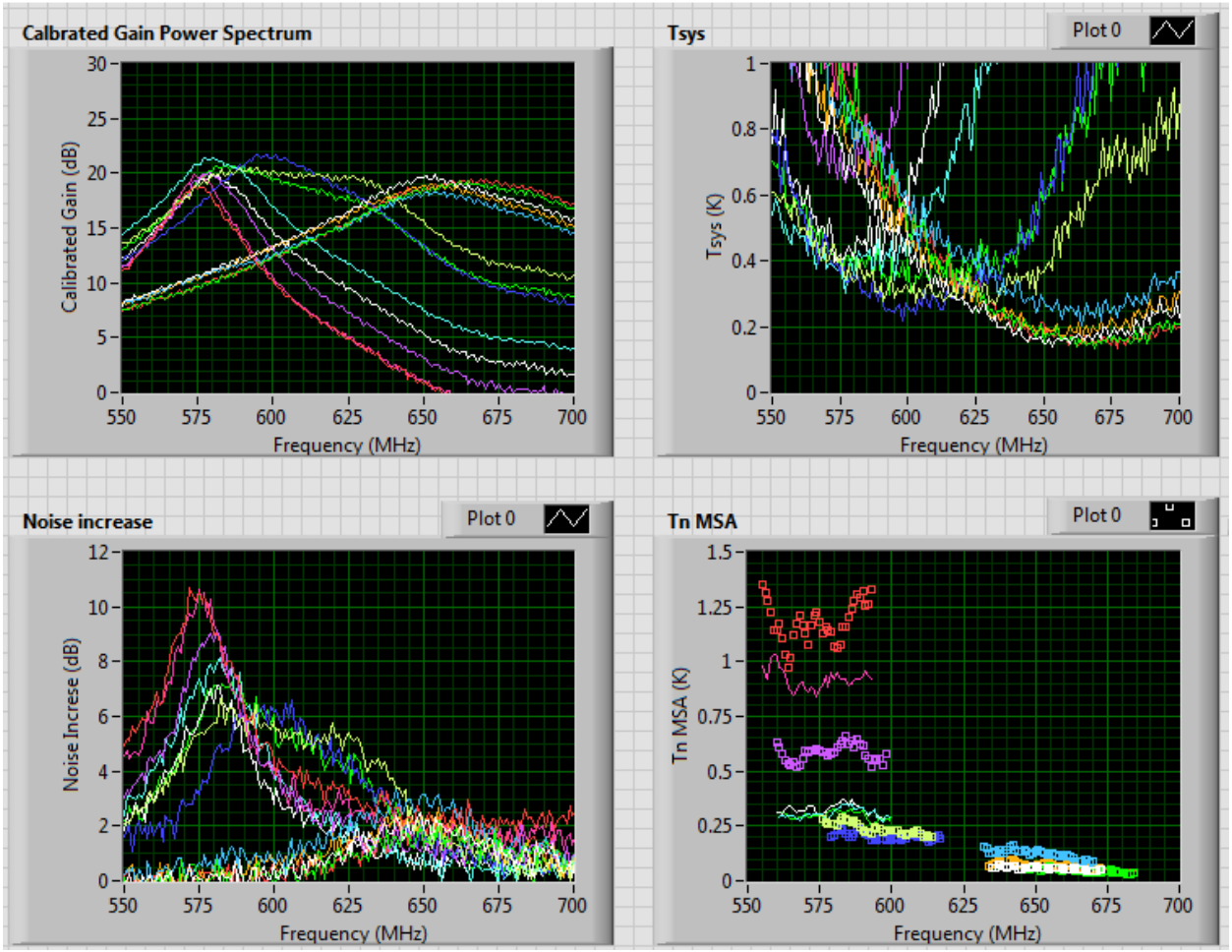

Figure 7.45: Performance of device LFF-11A at temperatures from 50 mK to 4.2 K. We could not tune this device and the varactor was floating, presumably at 0-V bias. The bias parameters were chosen for best S/N improvement at each temperature and the results collected here. At higher temperatures, the frequency was notably lower. This trend was quite a mystery until we delineated the varactor temperature dependence as described in section 7.1. The noise temperature scales with temperature, approaching the quantum limit at 50 mK and only about 1.25 K at a bath temperature of 4.2 K.

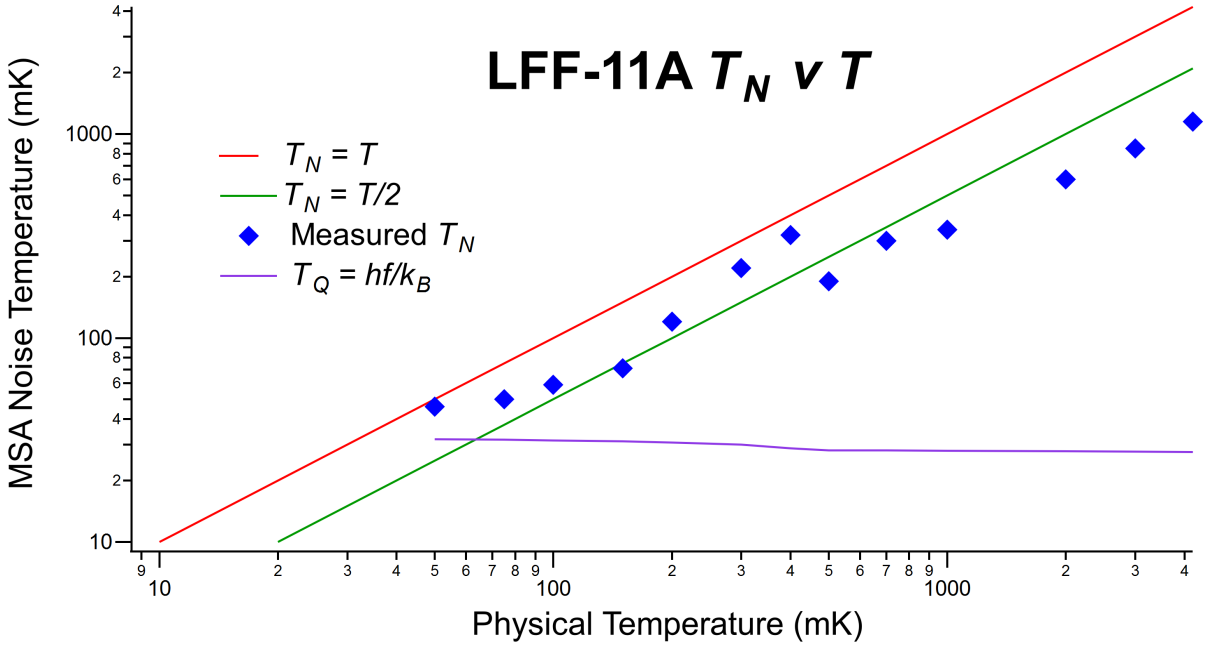

Figure 7.46: Summary of device LFF-11A T_N as a function of T_{bath}. The blue points are the MSA noise temperature measured by S/N improvement. The green and red lines are visual references showing the functions $T_N = T_{bath}$ and $T_N = T_{bath}/2$, respectively. The purple line shows the standard quantum limit $T_Q = hf/k_B$. The purple line rises somewhat at lower temperature because the device tuned to higher frequencies at lower temperatures. The T_N of this device was always below T_{bath} and typically below $T_{bath}/2$, approaching 1.44 times the quantum limit.

7.10.3 Device HFF-20A

MSA design HFF-20 is a low-frequency (more input coil turns) design from the high-frequency "HF" family of MSAs from the first fabrication. (Recall tables 5.5 and 5.6.) This device was designed for an unloaded $\lambda/2$ frequency of 1,994 MHz and in practice demonstrated excellent performance from about 830 to 1,010 MHz. All data presented here are with the device coupled to the input line via a 1.5 pF capacitor and tuned with one bank of ten parallel varactor diodes from the end of the input coil to ground.

The DC characteristics of this device are shown in figure 7.47. The DC landscape is rather smooth, and the shape is reminiscent of the "textbook" SQUID modulation curve. This suggests the device is almost entirely free of resonances or Shapiro steps. The one exception is a subtle "waver" to the landscape on the positive V_ϕ side of the "voltage hill", visually manifest by two parallel red stripes on the V_ϕ plot of figure 7.47. We explored the RF performance only in the region with negative V_ϕ (positive feedback) and very smooth flux modulation. The output impedance at our chosen bias point was about 16 Ω, suggesting

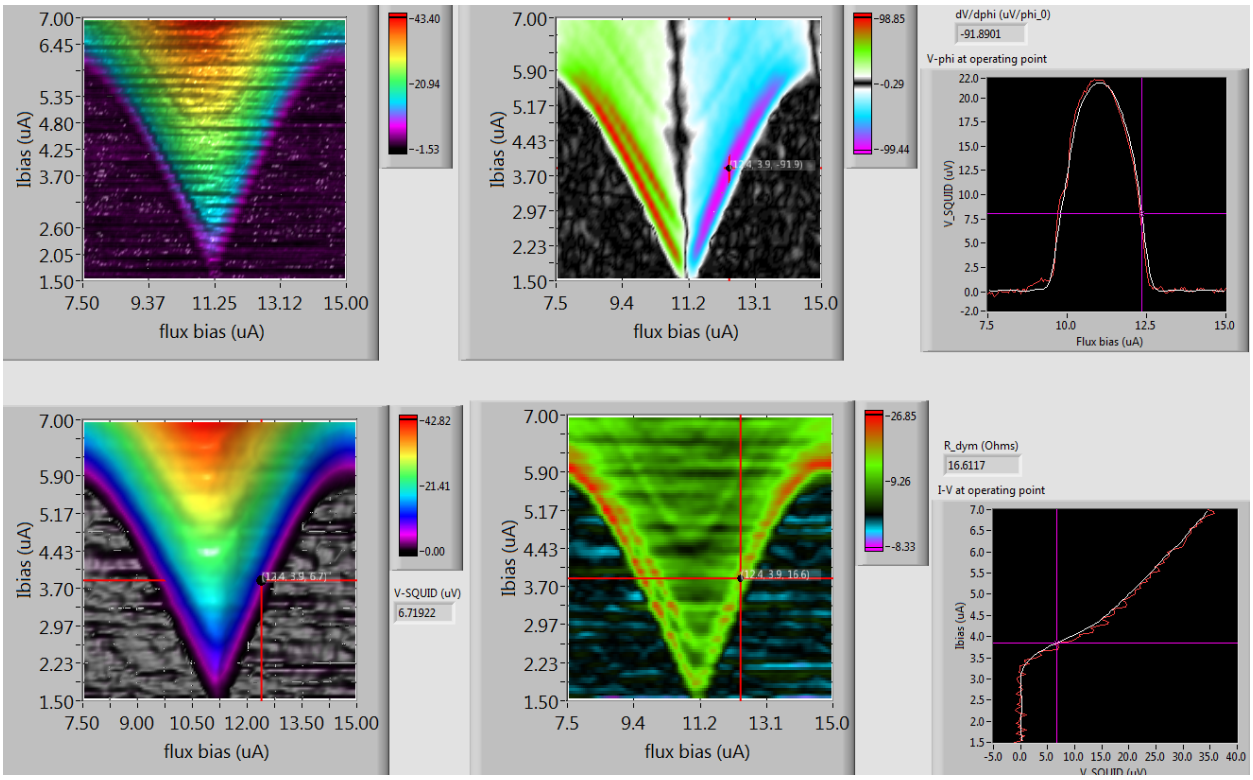

Figure 7.47: DC landscape of device HFF-20A. The four color landscapes correspond to one ϕ_0 of flux in X and 1.5 to 7 μA current bias in Y. Top left is the voltage as a function of current and flux bias. Horizontal "streaking" is due to DC amplifier drift. Bottom left is the voltage landscape after smoothing. The voltage on the left plots ranges from 0 to 43 μV. Top center is the flux derivative of the DC landscape, showing a V_ϕ sensitivity ranging from -99 to +98 $\mu V/\phi_0$. Center bottom is the dynamic impedance, ranging from 0 to 26 Ω. The red-line cursors on the color plots indicate a promising bias point, and cuts are taken through this point to generate the rightmost plots. Top right is a V-ϕ cut, and bottom right is an I-V cut. Purple cursors on the rightmost plots indicate the selected bias point. At the selected bias point, V_ϕ is -92 $\mu V/\phi_0$ and the dynamic impedance is 16 Ω. Note how "boring" this DC plot is compared to those shown in figures 7.25 through 7.28. Boring is *good*, and this device showed very promising performance.

it coupled only about half its amplified signal energy to the 50-Ω output line.

Despite an impedance mismatch at the output, this device demonstrated a gain as high as 30 dB, though the bias parameters yielding the most favorable noise temperature yielded a gain of 18 dB. Figure 7.48 shows a spider-fang plot of RF performance with the device tuned to its highest frequency, showing data for only the right "fang". The RF performance is relatively forgiving of bias parameters. The spider-fang plot uses the usual color map of black-red-white-blue, with blue marking only the very extrema. Note the relatively wide

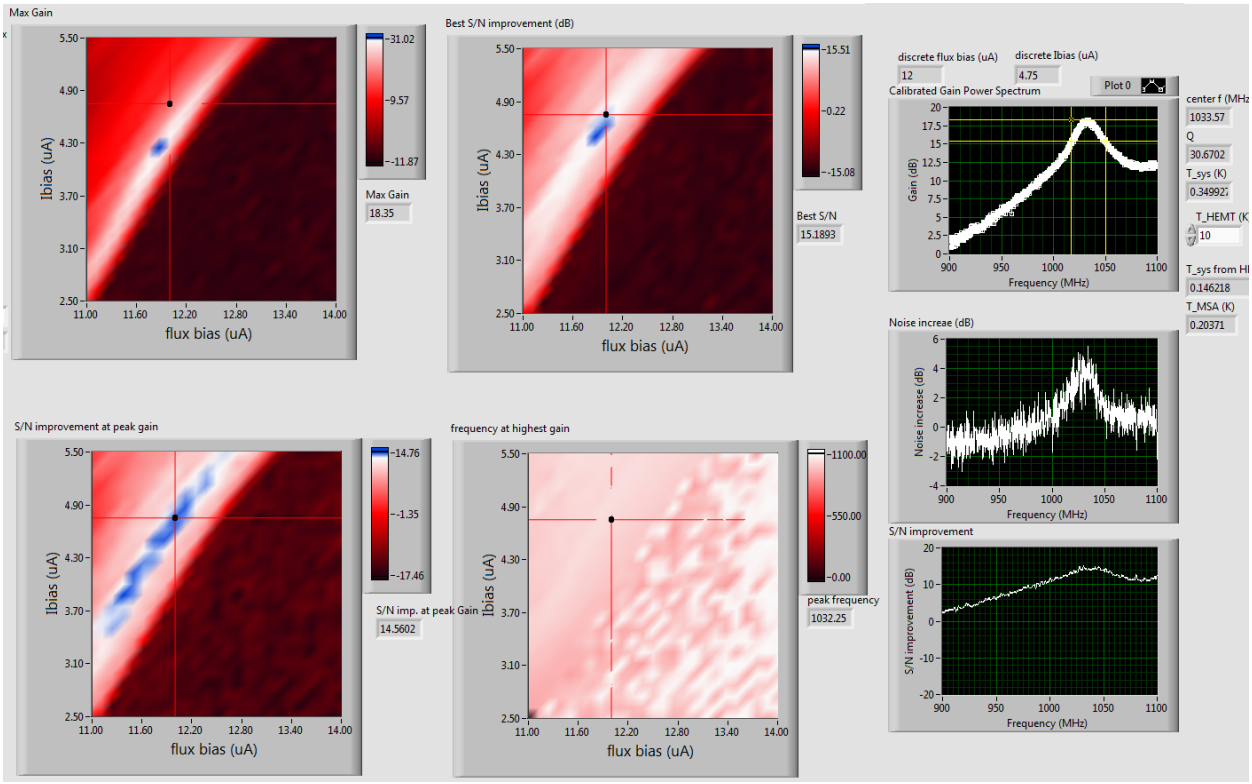

Figure 7.48: Spider-fang plot of HFF-20A, highest frequency tuning. This shows a spider-fang RF performance plot of device HFF-12A with the domain restricted to the right-hand fang. Top left shows gain, top center shows best S/N, bottom left shows best S/N at peak gain frequency, and center bottom shows frequency of peak gain. The cursors are positioned at a point of best S/N at the peak gain frequency, referencing the three right-hand spectra. The spectra are, top to bottom, gain, noise increase, and S/N improvement. The peak gain is about 18 dB at 1,032 MHz, with a S/N improvement of 14.5 dB, for an estimated $T_{sys} = 350$ mK and MSA $T_N = 203$ mK.

swath of white-to-blue on the lower left plot of figure 7.48. The gain and noise power spectra are relatively broad, with a Q of about 30. The gain spectrum has a strange asymmetry, suggesting an nearby gain peak around 1,200 MHz, but our instrumentation did not permit investigating this. The noise increase at the gain peak is only about 4 dB, suggesting a T_{sys} of about 350 mK, and an MSA T_N of about 200 mK.

A low T_N estimated via S/N improvement justifies a heated load Y-factor measurement, and the results of this measurement are shown in figure 7.49. The Y-factor method gives a system noise temperature at 1,027 MHz of $T_{sys} = 327$ mK and estimated MSA noise temperature $T_N = 180$ mK, just a little lower than estimated by S/N improvement. We also performed a Y-factor measurement with the device tuned to a somewhat lower frequency, near 906 MHz. The gain and feedback change with tuning (a limitation of single-varactor

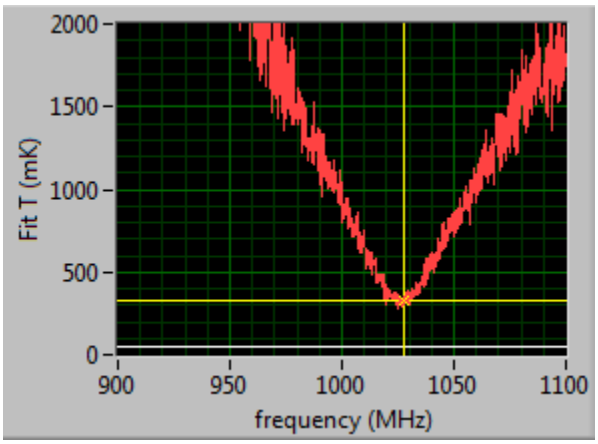

Figure 7.49: System noise temperature with HFF-20A tuned to maximum frequency, measured with a heated load. The promising noise temperatures suggested from the S/N improvement motivated a Y-factor heated load test. The system noise temperature measured at 1,032 MHz is $T_{sys} = 327$ mK, for an estimated MSA noise temperature of $T_N = 180$ mK after subtracting the estimated HEMT contribution.

tuning), so that in this case the gain is somewhat lower at a lower frequency. Figure 7.50 shows the Y-factor measurement at a lower tuning, with a system noise temperature at 906 MHz of $T_{sys} = 497$ mK and estimated MSA noise temperature $T_N = 181$ mK, this time with the estimated HEMT contribution and estimated MSA T_N shown explicitly. The scatter in T_N is a sizable fraction of the estimated T_N. A more accurate measurement would require a HEMT with a lower noise temperature.

This device showed great promise, but the varactor temperature dependence that emerges at temperatures lower than 2 K pushed this device to frequencies outside of the immediate range of ADMX, so development was halted. Figure 7.51 shows accessible frequencies as a function of temperature. At 4.2 K, the MSA tunes over the frequency band of the 715-850 MHz circulator, but at mK temperatures the tuning range is restricted to around 830 to 1,010 MHz, largely outside of the immediately applicable range. This device was delivered to the LLNL Rare Event Detection Group, which collaborates with ADMX on the detector design.

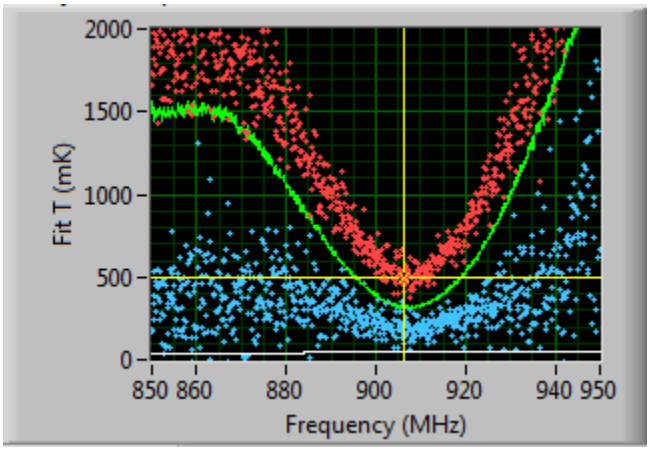

Figure 7.50: System noise temperature with HFF-20A tuned near its center frequency, measured with a heated load. This device was tested only with single-varactor tuning, so gain and T_N vary with tuning. Here we see the results of a Y-factor noise temperature measurement with the device tuned near the center of its range. Red points are T_{sys}. Green line is the estimated HEMT contribution to T_{sys}. Blue points are the estimated MSA T_N. The system noise temperature measured at 906 MHz is $T_{sys} = 497$ mK, for an estimated MSA noise temperature of $T_N = 181$ mK after subtracting the estimated HEMT contribution.

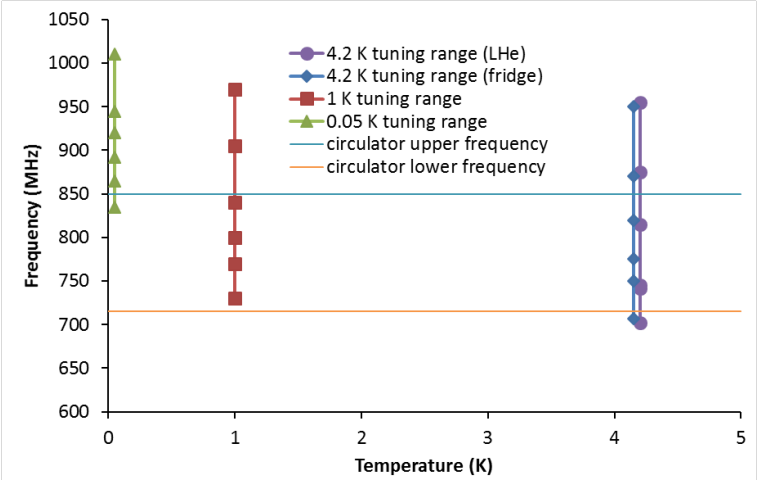

Figure 7.51: HFF-12A Tuning range with temperature. Tests at 4 K suggested this device was well matched to the 715-800 MHz range defined by ADMX circulators. Purple trace shows the tuning range in a 4.2 K liquid helium bath, and blue shows the tuning range at 4.2 K installed in the dilution fridge. Red shows the tuning range at 1 K in the fridge, and green shows the tuning range at 50 mK. Since this unexpected temperature dependency proved the device would not be compatible with immediate operations in ADMX, we curtailed further development with this device.

7.10.4 Device MSA2W2-45F-24A

MSA design MSA2W2-45F-24 has the highest design-frequency of any device designed for this project, with a nominal $\lambda/2$ standing wave frequency of 3,066 MHz. All data presented here are from a liquid He bath at 4.2 K with the device coupled to the input line via a 1.6 pF capacitor and the end of the coil left floating (no tunability). Our data collection and thus presentation of this device are rather abbreviated, but it has some interesting results nonetheless. Figure 7.52 shows the DC landscape. The landscape is relatively smooth, though at this low resolution it is not possible to resove very fine structure. At a promising-looking bias point marked by the cursor, the flux sensitivity $V_\phi = -31\ \mu\text{V}/\phi_0$ and the dynamic impedance is 8 Ω. The low dynamic impedance suggests only about 27% of the available amplified signal is transmitted to the 50-Ω output line.

The RF performance spider-fang plot is shown in figure 7.53, with cursors indicating the particular bias parameters for the displayed spectra. The gain and estimated MSA T_N are quite good at 21 dB and 1.7 K (physical temperature is 4.2 K), but the operating frequency presents a puzzle: the nominal design frequency is 3,066 MHz, but the operating frequency is 1,161 MHz. Part of the frequency shift is due to loading from the input: a 3,066-MHz resonator with $Z_0 = 159\ \Omega$, coupled to a 50-Ω line through a 1.6 pF capacitor and the other end floating would have a loaded resonant frequency of 1,875 MHz. Although the bonding pad at the end of the input coil is not explicitly wire-bonded to external components, it is not truly floating–it will have some capacitance to the surrounding ground planes. Figure 7.54 shows the MSA installed in RF carrier board. The un-bonded end-of-coil bonding pad (bottom left pad of the device) is about one mm^2, and represents a non-negligible capacitance. There is no ground plane behind the MSA chip, so all stray coupling is to co-planar conductors. Explicit calculation of a co-planar capacitance is non-trivial, but we can make a practical estimation[119] that the stray capacitance is about 0.1 pF. Adding 0.1 pF to the the end of the input coil reduces the expected frequency to 1,733 MHz. Supposing we have badly mis-calculated the stray capacitance of the bonding pads so that the actual capacitances are 2.1 and 0.5 pF on the input and end respectively, the expected frequency is 1,275 MHz, still higher than the observed frequency. Supposing the microstrip impedance is other than designed, our high-estimate capacitive loading would tune the device to the observed 1,161 MHz if $Z_0 = 199\ \Omega$, 25% *higher* than the design impedance.

A lower than designed capacitance between the input coil and SQUID washer would result in a lower bare frequency and higher Z_0 (more tunable by external capacitors), consistent with the observed low operating frequency. Supposing the coil-to washer capacitance were 1/2 the designed value (perhaps due to a doubly thick SiO_x layer), the bare frequency would be $\sqrt{2}$ lower (2,167 MHz) and Z_0 would be $\sqrt{2}$ greater (225 Ω). Assuming input and end capacitances of 1.7 and 0.1 pF respectively, this gives us an operating frequency of 1,205 MHz, still not quite as low as observed, but close. We find it unlikely that the fabrication process could be off by a factor of two or more, though it may be a small contributing factor. Absent further testing, the cause of the low operating frequency remains unknown.

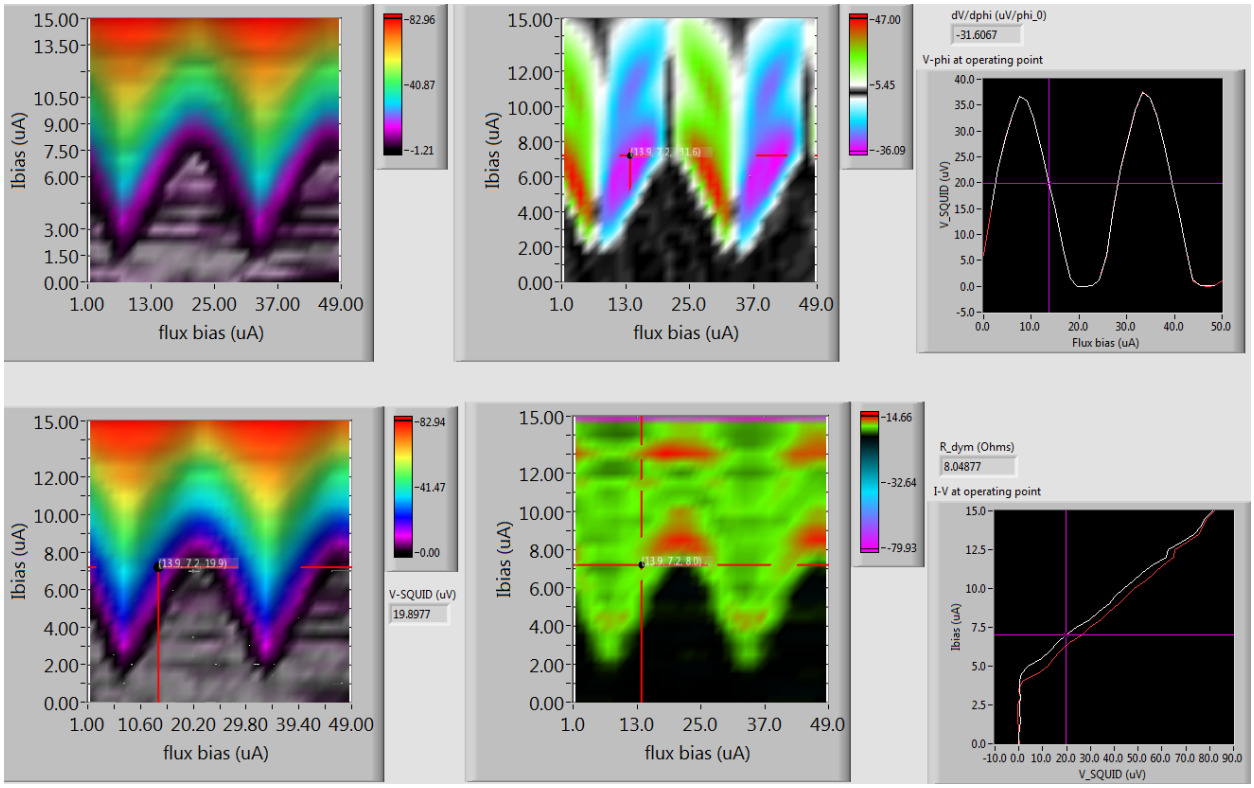

Figure 7.52: DC landscape of device MSA2W2-45F-24A. The four color landscapes correspond to slightly less than 2 ϕ_0 of flux in X and 0 to 15 μA current bias in Y. Top left is the voltage as a function of current and flux bias. Bottom left is the voltage landscape after smoothing. The voltage on the left plots ranges from 0 to 83 μV. Top center is the flux derivative of the DC landscape, showing a V_ϕ sensitivity ranging from -36 to +47 $\mu V/\phi_0$. Center bottom is the dynamic impedance, ranging from 0 to 15 Ω. The red-line cursors on the color plots indicate a promising bias point, and cuts are taken through this point to generate the rightmost plots. Top right is a V-ϕ cut, and bottom right is an I-V cut. Purple cursors on the rightmost plots indicate the selected bias point. At the selected bias point, V_ϕ is -31 $\mu V/\phi_0$ and the dynamic impedance is 8 Ω. This data was collected at rather low resolution so is is difficult to assess its smoothness, but it shows no indication of being "interesting" like figure 7.26.

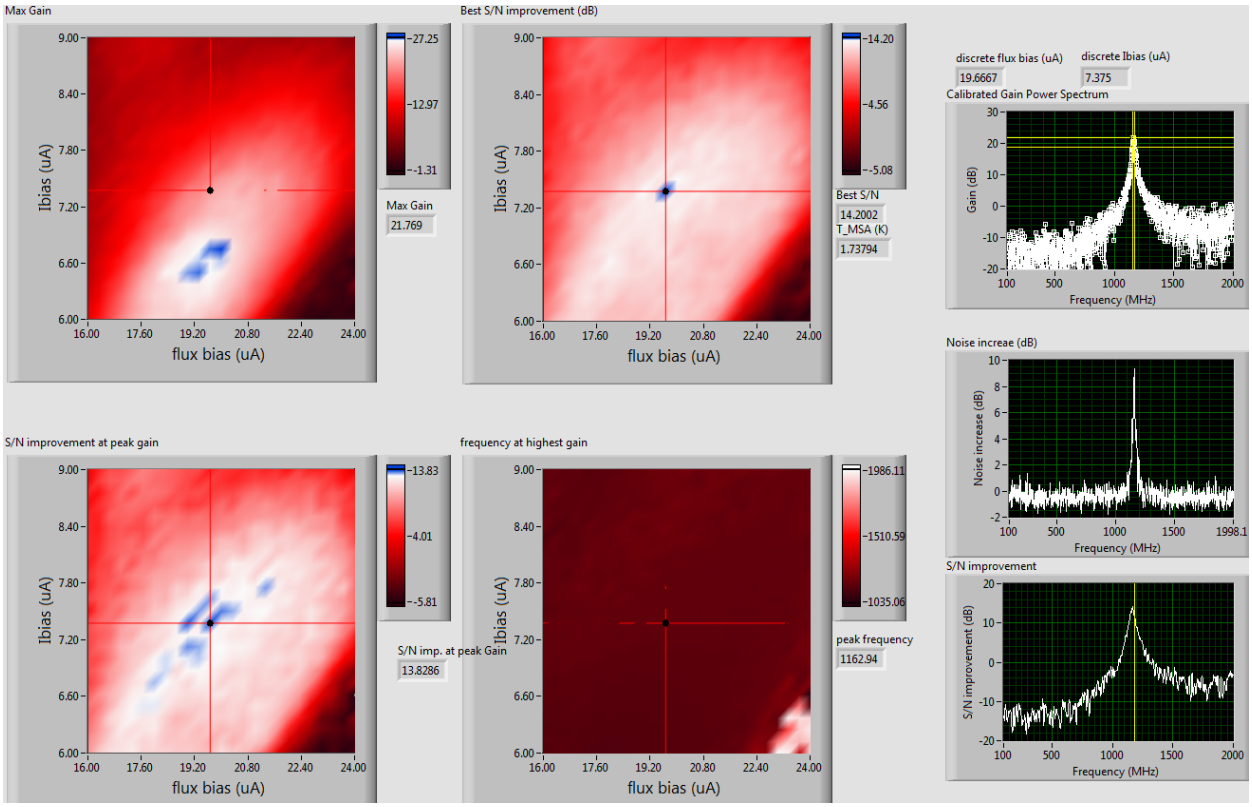

Figure 7.53: Spider-fang RF performance plot for device MSA2W2-45F-24A. The peak gain is 27dB, but the gain at the bias point with best S/N improvement (indicated by the cursors) is 22 dB. At this bias point, the frequency is 1,161 MHz, the estimated T_{sys} is 2.07 K, and the estimated MSA T_N is 1.7 K. This estimated noise temperature is very good for a device at a physical temperature of 4.2 K. The operating frequency is far lower than the nominal design frequency of 3,066 MHz.

7.10.5 Device MSA2W2-454F-24A

MSA design MSA2W2-454F-24 has nearly the lowest design frequency of any device designed for this project, with a nominal $\lambda/2$ standing wave frequency of 540 MHz and Z_0 of 282 Ω. All data shown here are from tests in a 4.2-K liquid helium bath with the input coupled via a 1.6 pH capacitor and the input coil end floating (no tuning). Our measured RF performance, shown in figure 7.55 demonstrates that the RF properties are remarkably insensitive to the bias parameters. At a typical bias point (marked by the cursor) the gain is 23 dB at 354 MHz with a Q of 77 and an estimated T_N of 4.9 K. A T_N greater than the physical temperature is not at all impressive, but no optimization was attempted here. As designed and coupled, we would expect a frequency of 382 MHz, about 8% higher than observed. We take this to mean that the design principles we used in designing the first fabrication batch hold true in the second batch as well, at least for low frequency, high-inductance "square hole" SQUID

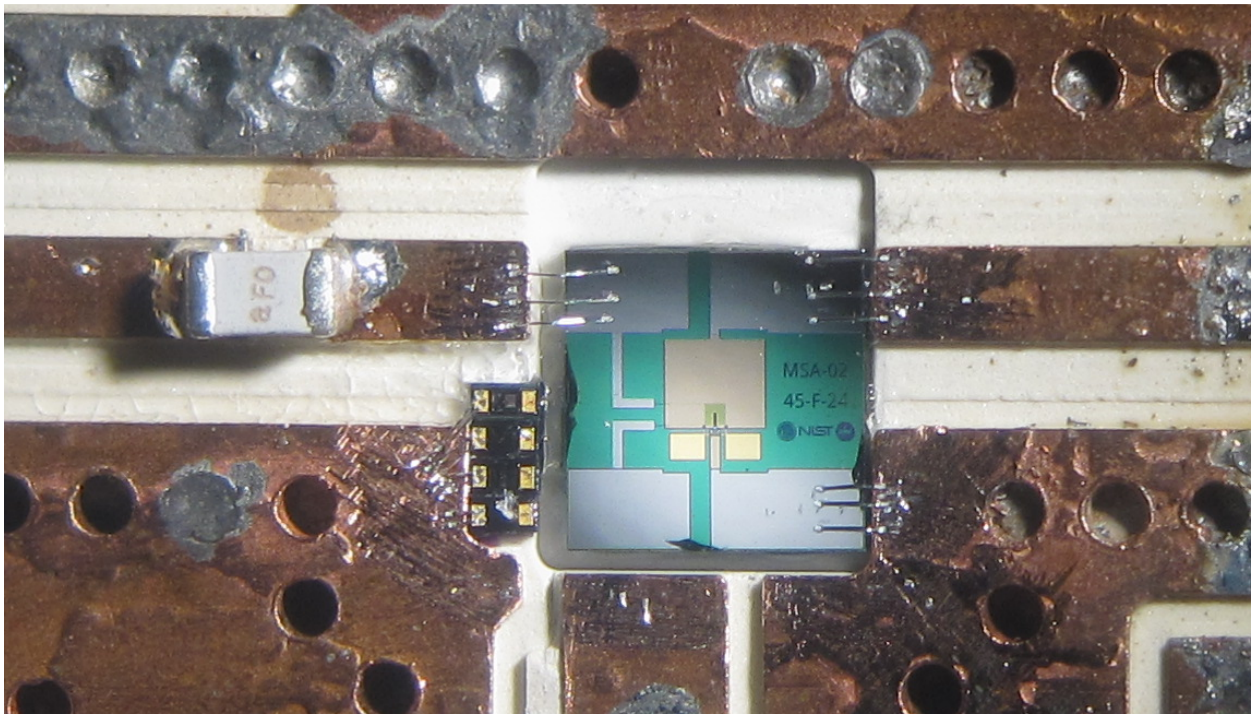

Figure 7.54: Device MSA2W2-45F-24A installed on its RF carrier board. The input (top left bonding pad) is coupled to the 50-Ω input line via a 1.6 pF capacitor, and the end of the input coil (bottom left pad) is unconnected. The unconnected bonding pads will have some stray capacitance to ground, which may account for part of this device's surprisingly low operating frequency.

designs similar to this one.

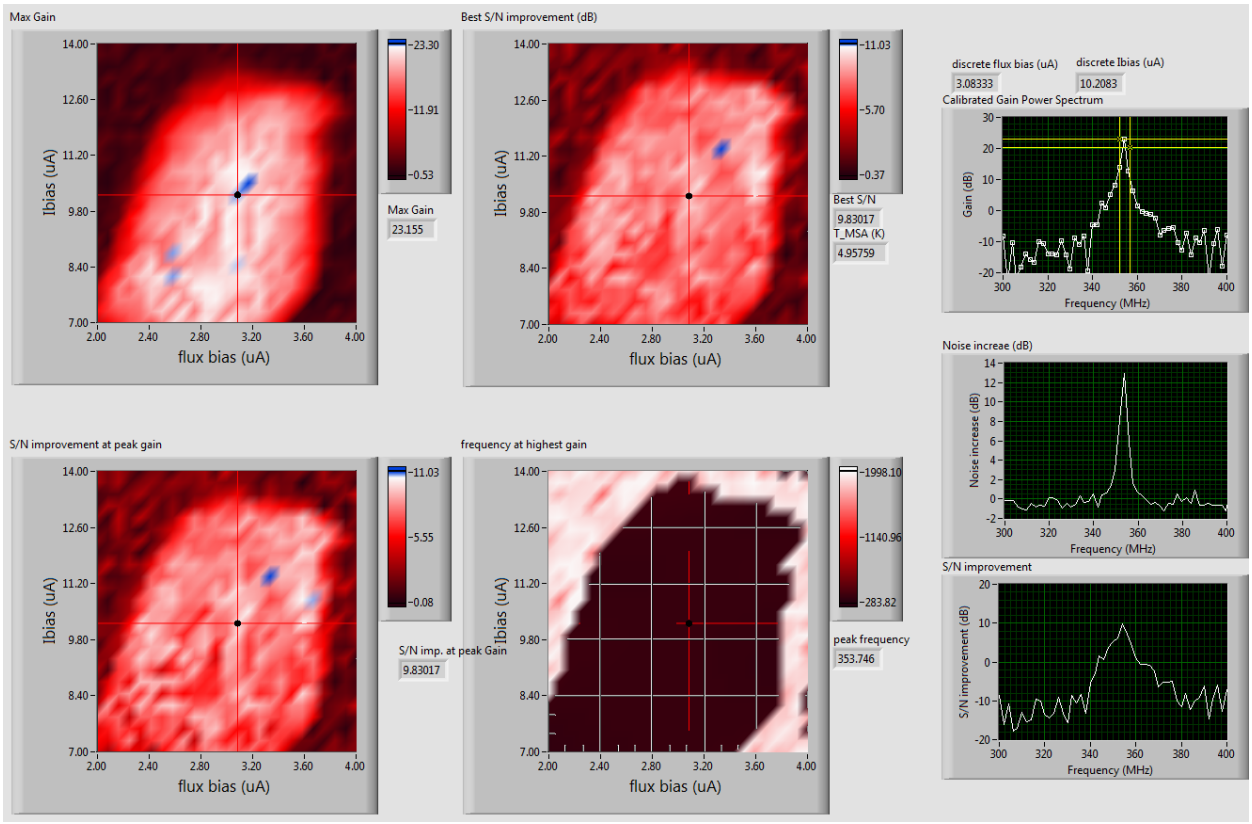

Figure 7.55: Spider-fang RF performance plot for device MSA2W2-454F-24A. Only the ''right fang" is shown here. The gain, S/N improvement, and frequency are all remarkably insensitive to the bias parameters. At a typical bias point marked by the cursors the gain is 23 dB at 354 MHz with a Q of 77, and the estimated T_N is 4.9 K. The resonant frequency is only 8% higher than we'd expect from this MSA with loading from the 50-Ω input line. This demonstrates the general soundness of our designs at lower frequencies, but tunability and T_N optimization tests were not performed.

Chapter 8

Concluding Remarks

8.1 Summary of Results

We have provided an overview of two of the outstanding mysteries in physics today: the dark matter problem and the strong CP problem. We hope our treatment of these topics is simple enough to be useful to a person approaching these topics for the first time, and thorough enough to whet the appetite and enable further investigation by a motivated reader. We presented the axion as a well-motivated solution to both the dark matter and strong CP problems, and gave an overview of ADMX, the longest running and most sensitive axion search to date. We hope to have provided some historical breadth to our discussion, such that someone new to the the field might appreciate ADMX in the larger context of astronomy, cosmology, and physics as they stand today.

With the need for a low-noise microwave amplifier in ADMX established, we moved to a description of the superconducting concepts most essential for understanding the operation of the MSA, and presented a summary of relevant results for MSAs produced prior to this work.

We discussed the specific considerations of our own MSA designs, with detail and clarity sufficient, we hope, to prepare the reader to duplicate or improve on our designs. We also included a discussion of "nuts and bolts" considerations in constructing a practical cryogenic RF experiment. Although our discussion of practical considerations is the part of this work least likely to be remembered as historically significant[1], we suspect it is the part most likely to be useful to what readership beyond the faculty committee[2] may read this dissertation.

We presented our measured performance of MSAs, with a diversion to the surprising temperature dependence of the varactors we used to tune the MSAs. We included reports of the performance parameters critical to operation in ADMX (ex: gain, T_N), and provided an accounting of how those measurements were calibrated and verified. We also included parameters more useful to the MSA designer (for example, Z_0, dynamic range) who may be

[1] A shockingly low bar, we admit.

[2] A number we hope to be as large as $2^{11} - (23 \times 89)$. Mersenne primes, anyone? No? ...nevermind.

motivated to compare their own measurements to that of a known functional device. We introduced new ways of visualizing MSA performance (the DC voltage landscape and spider-fang plot) that (we think) help to clarify understanding of particular device performance.

8.2 Future Directions

The project of ADMX is primarily one of detection–that is, gathering evidence of the presence or absence of axions in the galactic halo. The MSA is a phase-preserving amplifier that will by its nature add an equal amount of quantum noise to each quadrature of measurement, so it is subject to the standard quantum limit of noise temperature $T_Q = hf/k_B$, about 34 mK at our frequencies of interest and rising with frequency at about 48 mK/GHz. The minimum quantum noise added by any amplifier is fundamentally constrained by quantum mechanics, but the noise need not be distributed equally to all quadratures. An amplifier that can "squeeze" its quantum noise to one quadrature, for example the JPA can beat the standard quantum limit by radiating the resonator with squeezed vacuum and amplifying along the axis with reduced quantum uncertainty. In principle the "leftover quantum noise" of such a system is limited only by the degree of squeezing and amount of extra (non-quantum) noise added by the amplifier. Neither is limited by physical principle, but limited only by the engineering challenges of achieving low "noise over quantum" and a high degree of squeezing.[3] The JPA-squeezed state approach is currently under development in experiments like HAYSTAC, and beating the standard quantum limit becomes more important at higher frequencies.

The ideal amplifier for a detection search is a particle counter–an amplifier that has "infinite" uncertainty in the phase of the signal and arbitrarily small uncertainty in particle number. The phase qubit (or any qubit that allows for high-fidelity readout of the occupation number) could be used as a particle counter with arbitrarily low background noise at the cost of revealing no phase information. If[4] an axion detection is achieved, the resonator design is simplified because wide frequency tunability is no longer needed, and with existence proved, phase-measurements become scientifically important and the phase-preserving amplifier becomes a valuable asset.

If axions are the primary constituent of mark matter on our galaxy they likely exist as a Bose-Einstein condensate with a correlation length of 100 m or more. One can imagine an "ADMX network" of detectors spaced by distances ranging from 10's of meters to, ideally, the entire globe[5], each with its own phase-preserving amplifier. By integrating the phase data between sites, we could identify the flow direction, or multiple simultaneous flow directions of the axion field. This would be a boon to astronomy, revealing a detailed local map of

[3]The engineering challenges are non-trivial, but if given a choice between overcoming physical law or barriers of engineering, we unreservedly recommend choosing the engineering challenge.

[4]Or when?

[5]Lets put some in orbit at the Earth-Sun Lagrange 3 and Lagrange 4 points as well. Why think small?

dark matter[6] caustics, telling us about the formation history of the Milky Way, and a boon to science as a whole giving us a direct window into a whole new sector of particle physics.

To further develop the MSA, we would like to develop a detailed superconducting model of the coupled input coil and SQUID washer at the level of a time-resolved finite-element simulation, which may shed light on the details of capacitive feedback and the surprisingly low inductance of the input coil. We would also like to model the effects of higher harmonics, both in terms of feedback and junction dynamics. We see hints of harmonic feedback effects in the Shapiro step-like features of the DC voltage landscapes, and have observed markedly different behavior between the output connected to a 50-Ω attenuator and a circulator matched to the signal band, presumably due to strong out-of-band reflections from the circulator. Perhaps, issues with out-of-band reflections could be avoided with the microwave equivalent of dichroic glass–a coupler that transmits in-band frequencies to port A and dumps out-of-band frequencies to port B, which could be dumped to a thermalized resistive load, if desired.

MSAs can do more than chase axions. There are not currently many options for a researcher who desires a near quantum-limited amplifier that operates at frequencies significantly below 1 GHz. The best commercially available amplifiers (for example the LNF-LNC0.2-3A by Low Noise Factory) can operate as low as 200 MHz with a T_N of 1.7 K, about 170 times higher than the quantum limit. JPAs can operate near the quantum limit at frequencies of several GHz, and have been made to operate as low as 700 MHz or so, but are unlikely to be useful at much lower frequencies barring a breakthrough in low-loss, high-value capacitor design. We have designed and tested MSAs that operate as low as 350 MHz (before tuning lower with varactors) and MSAs operating as low as 90 MHz with a T_N of 60 mK have been reported in prior work.[120] We have delivered MSAs developed from this project to the Gabrielse lab at Northwestern University for detecting axial oscillation of a single electron in a Penning trap, the Tobar Lab at the University of Western Australia for reading out micro-mechanical oscillators, and the Center for Axion and Precision Physics Research (CAPP) in Daejeon, Korea.

We are not aware of any phase-preserving amplifiers that have a T_N near the quantum limit that operate well above 1 GHz, though the Josephson traveling-wave parametric amplifier (JTWPA) shows promise, particularly for applications requiring a wide instantaneous bandwidth. MSAs in this work have been demonstrated operating at a little above 1 GHz, and MSAs with higher-harmonic input coils have been demonstrated operating as high as 2.5 GHz. The MSA remains the lowest T_N amplifier available in the 100 MHz to 1 GHz range, and may yet remain a competitive choice for a phase-sensitive amplifier at frequencies up to several GHz. We hope this technology continues to develop, as the need for lower-frequency quantum-limited amplifiers will only grow as the study of macroscopic quantum systems expands and advances.

[6] Would we still call it dark matter at this point?

Bibliography

[1] DMX. *Cradle 2 the Grave*. Bloodline Records and Def Jam Recording. Audio CD. Track 1: X Gon' Give It to Ya. 2003.

[2] H. Kamerlingh Onnes. "Further experiments with liquid helium. C. On the change of electric resistance of pure metals at very low temperatures etc. IV. The resistance of pure mercury at helium temperatures," in: *KAWA, Proceedings* (1910-1911). URL: http://www.dwc.knaw.nl/DL/publications/PU00013358.pdf.

[3] H. Kamerlingh Onnes. "Further experiments with Liquid Helium. G. On the Electrical Resistance of Pure Metals, etc. VI. On the Sudden CHange in the Rate at which the Resistance of Mercury Disappears." In: *KAWA, Proceedings* 13.124c (1911), pp. 799–802.

[4] H. Kamerlingh Onnes. "Further experiments with liquid helium J. The imitation of an Ampre molecular current or a permanent magnet by means of a supra-conductor." In: *KAWA, Proceedings* 17.104c (1914), pp. 278–283. URL: http://www.dwc.knaw.nl/DL/publications/PU00012675.pdf.

[5] Broom R. F. "An Upper limit for the Resistivity of a Superconducting Film". In: *Nature* 190 (4780 1961), pp. 992–993. DOI: 10.1038/190992b0. URL: http://dx.doi.org/10.1038/190992b0.

[6] D. J. Quinn III and W. B. Ittner III. "Resistance in a Superconductor". In: *Journal of Applied Physics* 33.2 (1962), pp. 748–749. DOI: 10.1063/1.1702504.

[7] D. J. Quinn and W. B. Ittner. "Resistance in a Superconductor". In: *Journal of Applied Physics* 33.2 (1962), pp. 748–749. DOI: 10.1063/1.1702504. eprint: https://doi.org/10.1063/1.1702504. URL: https://doi.org/10.1063/1.1702504.

[8] Joel Primack. *Dark Matter: History and Cosmology*. Accessed: 2019-02-21. June 2016. URL: http://physics.ucsc.edu/~joel/Primack-DM&Cosmology-Lamat.pdf.

[9] V. M. Slipher. "Spectrographic Observations of Nebulae". In: *Popular Astronomy* 23 (Jan. 1915), pp. 21–24.

[10] C. Wirtz. "Über die Bewegungen der Nebelflecke". In: *Astronomische Nachrichten* 206.13 (), pp. 109–116. DOI: 10.1002/asna.19182061302. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asna.19182061302. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asna.19182061302.

[11] Edwin Hubble. "A relation between distance and radial velocity among extra-galactic nebulae". In: *Proceedings of the National Academy of Sciences* 15.3 (1929), pp. 168–173. ISSN: 0027-8424. DOI: 10.1073/pnas.15.3.168. eprint: http://www.pnas.org/content/15/3/168.full.pdf. URL: http://www.pnas.org/content/15/3/168.

[12] Galina Weinstein. "George Gamow and Albert Einstein: Did Einstein say the cosmological constant was the "biggest blunder" he ever made in his life?" 2013. URL: https://arxiv.org/abs/1310.1033.

[13] Wayne Hu and Martin White. "The Cosmic Symphony". In: *Scientific American* (Feb. 2004), pp. 44–53. URL: https://doi.org/10.1051/0004-6361/201321591.

[14] Robert P. Kirshner. "Hubble′s diagram and cosmic expansion". In: *Proceedings of the National Academy of Sciences* 101.1 (2004), pp. 8–13. ISSN: 0027-8424. DOI: 10.1073/pnas.2536799100. eprint: http://www.pnas.org/content/101/1/8.full.pdf. URL: http://www.pnas.org/content/101/1/8.

[15] Allan R. Sandage. "Cosmology: A search for two numbers". In: *Physics Today* 23.2 (1970), pp. 34–41. DOI: 10.1063/1.3021960. URL: https://physicstoday.scitation.org/doi/10.1063/1.3021960.

[16] N. Suzuki et al. "The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above z $>$ 1 and Building an Early-type-hosted Supernova Sample". In: *The Astrophysical Journal* 746.1 (2012), p. 85. URL: http://stacks.iop.org/0004-637X/746/i=1/a=85.

[17] URL: https://www.nasa.gov/images/content/735683main_pia16873-full_full.jpg.

[18] Planck Collaboration. "Planck 2013 results. XVI. Cosmological parameters". In: *Astronomy and Astrophysics* 571 (2014), A16. DOI: 10.1051/0004-6361/201321591. URL: https://doi.org/10.1051/0004-6361/201321591.

[19] Wayne Hu. *Ringing in the New Cosmology, Intermediate Guide to the Acoustic Peaks and Polarization*. Feb. 2001. URL: http://background.uchicago.edu/~whu/intermediate/intermediate.html.

[20] S. Perlmutter et al. "Cosmology from Type Ia Supernovae". In: (1998). eprint: arXiv:astro-ph/9812473.

[21] Volker Springel et al. "Simulations of the formation, evolution and clustering of galaxies and quasars". In: *Nature* 435 (June 2005). Article, pp. 629–636. URL: http://dx.doi.org/10.1038/nature03597.

[22] Daniel J. Eisenstein et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *The Astrophysical Journal* 633.2 (2005), p. 560. URL: http://stacks.iop.org/0004-637X/633/i=2/a=560.

[23] J. H. Oort. "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems". In: *Bulletin of the Astronomical Institutes of the Netherlands* 6 (Aug. 1932), p. 249.

[24] F. Zwicky. "On the Masses of Nebulae and of Clusters of Nebulae". In: *Astrophysical Journal* 86 (Oct. 1937), p. 217. DOI: 10.1086/143864.

[25] Y. Sofue et al. "Central Rotation Curves of Spiral Galaxies". In: *The Astrophysical Journal* 523.1 (1999), p. 136. URL: http://stacks.iop.org/0004-637X/523/i=1/a=136.

[26] Joel Primack. *Rio Lectures 1: A Brief History of Dark Matter*. Oct. 2008. URL: http://physics.ucsc.edu/~joel/Primack08RioLectures/.

[27] F.D. Kahn and L. Woltjer. "Intergalactic Matter and the Galaxy." In: *Astrophysical Journal* 130 (Nov. 1959), p. 705. DOI: 10.1086/146762.

[28] Sir F. W. Dyson F. R. S., A. S. Eddington F. R. S., and Mr. C. Davidson. "IX. A determination of the deflection of light by the sun′s gravitational field, from observations made at the total eclipse of May 29, 1919". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 220.571-581 (1920), pp. 291–333. ISSN: 0264-3952. DOI: 10.1098/rsta.1920.0009. eprint: http://rsta.royalsocietypublishing.org/content/220/571-581/291.full.pdf. URL: http://rsta.royalsocietypublishing.org/content/220/571-581/291.

[29] R.A. Alpher. "Origin and Relative Abundance of the Chemical Elements." PhD thesis. The George Washington University, Washington, D.C, 1948.

[30] NASA/WMAP Science Team. *Nucleosynthesis in the Early Universe*. Dec. 2010. URL: https://map.gsfc.nasa.gov/universe/bb_tests_ele.html.

[31] Cristian Armendariz-Picon and Jayanth T. Neelakanta. "How cold is cold dark matter?" In: *Journal of Cosmology and Astroparticle Physics* 2014.03 (2014), p. 049. URL: http://stacks.iop.org/1475-7516/2014/i=03/a=049.

[32] Randall Monroe. *Dark Matter Candidates*. Aug. 2018. URL: https://xkcd.com/2035/.

[33] J R Primack, D Seckel, and B Sadoulet. "Detection of Cosmic Dark Matter". In: *Annual Review of Nuclear and Particle Science* 38.1 (1988), pp. 751–807. DOI: 10.1146/annurev.ns.38.120188.003535. eprint: https://doi.org/10.1146/annurev.ns.38.120188.003535. URL: https://doi.org/10.1146/annurev.ns.38.120188.003535.

[34] Gerard Jungman, Marc Kamionkowski, and Kim Griest. "Supersymmetric dark matter". In: *Physics Reports* 267.5 (1996), pp. 195–373. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(95)00058-5. URL: http://www.sciencedirect.com/science/article/pii/0370157395000585.

[35] C. Alcock et al. "The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations". In: *The Astrophysical Journal* 542.1 (2000), p. 281. URL: http://stacks.iop.org/0004-637X/542/i=1/a=281.

[36] P. G. Harris et al. "New Experimental Limit on the Electric Dipole Moment of the Neutron". In: *Phys. Rev. Lett.* 82 (5 Feb. 1999), pp. 904–907. DOI: 10.1103/PhysRevLett.82.904.

[37] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[38] Stephen L. Adler. "Axial-Vector Vertex in Spinor Electrodynamics". In: *Phys. Rev.* 177 (5 Jan. 1969), pp. 2426–2438. DOI: 10.1103/PhysRev.177.2426. URL: https://link.aps.org/doi/10.1103/PhysRev.177.2426.

[39] J. S. Bell and R. Jackiw. "A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model". In: *Il Nuovo Cimento A (1965-1970)* 60.1 (Mar. 1969), pp. 47–61. ISSN: 1826-9869. DOI: 10.1007/BF02823296.

[40] E. Shuryak. *The QCD vacuum, hadrons and superdense matter*. World Scientific, Jan. 1988, p. 137. ISBN: 978-9971978327.

[41] Varouzhan Baluni. "CP-nonconserving effects in quantum chromodynamics". In: *Phys. Rev. D* 19 (7 Apr. 1979), pp. 2227–2230. DOI: 10.1103/PhysRevD.19.2227. URL: https://link.aps.org/doi/10.1103/PhysRevD.19.2227.

[42] P. Di Vecchia. "The Dynamics of the Pseudoscalar Mesons at Arbitrary Θ in Large N Quantum Chromodynamics". In: *Acta Phys. Austriaca Suppl.* 22 (1980), pp. 341–381.

[43] R.J. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten. "Chiral estimate of the electric dipole moment of the neutron in quantum chromodynamics". In: *Physics Letters B* 88.1 (1979), pp. 123–127. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(79)90128-X. URL: http://www.sciencedirect.com/science/article/pii/037026937990128X.

[44] B. Lee Roberts and William J. Marciano. *Lepton dipole moments (Advanced series on directions in high energy physics ; Vol. 20)*. World Scientific, 2010.

[45] Steve K. Lamoreaux and Robert Golub. "The Neutron Electric Dipole Moment: Yesterday, Today and Tomorrow". In: *Lepton Dipole Moments*. WORLD SCIENTIFIC, 2011. Chap. 15, pp. 583–634. DOI: 10.1142/9789814271844_0015.

[46] Ann Nelson. "Naturally weak CP violation". In: *Physics Letters B* 136.5 (1984), pp. 387–391. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)92025-2. URL: http://www.sciencedirect.com/science/article/pii/0370269384920252.

[47] Stephen M. Barr and A. Zee. "Solution of the Strong CP Problem by Color Exchange". In: *Phys. Rev. Lett.* 55 (21 Nov. 1985), pp. 2253–2256. DOI: 10.1103/PhysRevLett.55.2253. URL: https://link.aps.org/doi/10.1103/PhysRevLett.55.2253.

[48] R. D. Peccei and Helen R. Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.1440.

[49] R. D. Peccei and Helen R. Quinn. "Constraints imposed by CP conservation in the presence of pseudoparticles". In: *Phys. Rev. D* 16 (6 Sept. 1977), pp. 1791–1797. DOI: 10.1103/PhysRevD.16.1791. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1791.

[50] Wim de Boer. "The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology". In: *Time and Matter 2013 (TAM2013) Venice, Italy*. 2013. arXiv: 1309.0721 [hep-ph]. URL: http://inspirehep.net/record/1252561/files/arXiv:1309.0721.pdf.

[51] Apodaca Representaciones. *Promotional photo for Pedro Fernandez*. 717 E Tidwell Rd Houston, TX 77022. URL: http://ww4.hdnux.com/photos/40/63/73/8601339/3/1024x1024.jpg.

[52] Thomas Lin, Olena Shmahalo, Lucy Reading-Ikkanda, and Ryan Schude. "Roberto Peccei and Helen Quinn, Driving Around Stanford in a Clunky Jeep". In: *Qunta magazine* (2017). URL: https://www.quantamagazine.org/roberto-peccei-and-helen-quinn-driving-around-stanford-in-a-clunky-jeep-20170615/.

[53] Steven Weinberg. "A New Light Boson?" In: *Phys. Rev. Lett.* 40 (4 Jan. 1978), pp. 223–226. DOI: 10.1103/PhysRevLett.40.223. URL: https://link.aps.org/doi/10.1103/PhysRevLett.40.223.

[54] F. Wilczek. "Problem of Strong P and T Invariance in the Presence of Instantons". In: *Phys. Rev. Lett.* 40 (5 Jan. 1978), pp. 279–282. DOI: 10.1103/PhysRevLett.40.279. URL: https://link.aps.org/doi/10.1103/PhysRevLett.40.279.

[55] William A. Bardeen and S.-H.H. Tye. "Current algebra applied to properties of the light Higgs boson". In: *Physics Letters B* 74.3 (1978), pp. 229–232. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(78)90560-9. URL: http://www.sciencedirect.com/science/article/pii/0370269378905609.

[56] C. Edwards et al. "Upper Limit for $\frac{J}{\Psi} \to \gamma + Axion$". In: *Phys. Rev. Lett.* 48 (14 Apr. 1982), pp. 903–906. DOI: 10.1103/PhysRevLett.48.903. URL: https://link.aps.org/doi/10.1103/PhysRevLett.48.903.

[57] M. Sivertz et al. "Upper limit for axion production in radiative Υ decay". In: *Phys. Rev. D* 26 (3 Aug. 1982), pp. 717–719. DOI: 10.1103/PhysRevD.26.717. URL: https://link.aps.org/doi/10.1103/PhysRevD.26.717.

[58] Christian Boutan. "A Piezoelectrically Tuned RF-Cavity Search for Dark Matter". PhD thesis. University of Washington, 2017.

[59] Jihn E. Kim. "Weak-Interaction Singlet and Strong CP Invariance". In: *Phys. Rev. Lett.* 43 (2 July 1979), pp. 103–107. DOI: 10.1103/PhysRevLett.43.103. URL: https://link.aps.org/doi/10.1103/PhysRevLett.43.103.

[60] M.A. Shifman, A.I. Vainshtein, and V.I. Zakharov. "Can confinement ensure natural CP invariance of strong interactions?" In: *Nuclear Physics B* 166.3 (1980), pp. 493–506. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(80)90209-6. URL: http://www.sciencedirect.com/science/article/pii/0550321380902096.

[61] Michael Dine, Willy Fischler, and Mark Srednicki. "A simple solution to the strong CP problem with a harmless axion". In: *Physics Letters B* 104.3 (1981), pp. 199–202. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(81)90590-6. URL: http://www.sciencedirect.com/science/article/pii/0370269381905906.

[62] A. R. Zhitnitsky. "On Possible Suppression of the Axion Hadron Interactions. (In Russian)". In: *Sov. J. Nucl. Phys.* 31 (1980). [Yad. Fiz.31,497(1980)], p. 260.

[63] R.L. Davis. "Cosmic axions from cosmic strings". In: *Physics Letters B* 180.3 (1986), pp. 225–230. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(86)90300-X. URL: http://www.sciencedirect.com/science/article/pii/037026938690300X.

[64] R.L. Davis and E.P.S. Shellard. "Do axions need inflation?" In: *Nuclear Physics B* 324 (Sept. 1989), pp. 167–186. DOI: 10.1016/0550-3213(89)90187-9. URL: https://www.sciencedirect.com/science/article/pii/0550321389901879.

[65] Diego Harari and P. Sikivie. "On the evolution of global strings in the early universe". In: *Physics Letters B* 195.3 (1987), pp. 361–365. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)90032-3. URL: http://www.sciencedirect.com/science/article/pii/0370269387900323.

[66] David J. Gross, Robert D. Pisarski, and Laurence G. Yaffe. "QCD and instantons at finite temperature". In: *Rev. Mod. Phys.* 53 (1 Jan. 1981), pp. 43–80. DOI: 10.1103/RevModPhys.53.43. URL: https://link.aps.org/doi/10.1103/RevModPhys.53.43.

[67] Michael S. Turner. "Cosmic and local mass density of "invisible" axions". In: *Phys. Rev. D* 33 (4 Feb. 1986), pp. 889–896. DOI: 10.1103/PhysRevD.33.889. URL: https://link.aps.org/doi/10.1103/PhysRevD.33.889.

[68] R. M. Bionta et al. "Observation of a neutrino burst in coincidence with supernova 1987A in the Large Magellanic Cloud". In: *Phys. Rev. Lett.* 58 (14 Apr. 1987), pp. 1494–1496. DOI: 10.1103/PhysRevLett.58.1494. URL: https://link.aps.org/doi/10.1103/PhysRevLett.58.1494.

[69] K. Hirata et al. "Observation of a neutrino burst from the supernova SN1987A". In: *Phys. Rev. Lett.* 58 (14 Apr. 1987), pp. 1490–1493. DOI: 10.1103/PhysRevLett.58.1490. URL: https://link.aps.org/doi/10.1103/PhysRevLett.58.1490.

[70] Edward W. Kolb and Michael Stanley. Turner. *The Early Universe*. Addison-Wesley, 1990.

[71] Subrahmanyan Chandrasekhar. *An Introduction to the Study of Stellar Structure*. Chicago: Dover Publications, 2015.

[72] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[73] Peter W. Graham, Igor G. Irastorza, Steven K. Lamoreaux, Axel Lindner, and Karl A. van Bibber. "Experimental Searches for the Axion and Axion-Like Particles". In: *Ann. Rev. Nucl. Part. Sci.* 65 (), pp. 485–514. DOI: 10.1146/annurev-nucl-102014-022120. arXiv: 1602.00039 [hep-ex].

[74] P. Sikivie. "Experimental Tests of the "Invisible" Axion". In: *Phys. Rev. Lett.* 51 (16 Oct. 1983), pp. 1415–1417. DOI: 10.1103/PhysRevLett.51.1415. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1415.

[75] N. Du et al. "Search for Invisible Axion Dark Matter with the Axion Dark Matter Experiment". In: *Phys. Rev. Lett.* 120 (15 Apr. 2018), p. 151301. DOI: 10.1103/PhysRevLett.120.151301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.151301.

[76] S. Asztalos et al. "Large-scale microwave cavity search for dark-matter axions". In: *Phys. Rev. D* 64 (9 Oct. 2001), p. 092003. DOI: 10.1103/PhysRevD.64.092003. URL: https://link.aps.org/doi/10.1103/PhysRevD.64.092003.

[77] S. J. Asztalos et al. "Experimental Constraints on the Axion Dark Matter Halo Density". In: *The Astrophysical Journal Letters* 571.1 (2002), p. L27. URL: http://stacks.iop.org/1538-4357/571/i=1/a=L27.

[78] S. J. Asztalos et al. "Improved rf cavity search for halo axions". In: *Phys. Rev. D* 69 (1 Jan. 2004), p. 011101. DOI: 10.1103/PhysRevD.69.011101. URL: https://link.aps.org/doi/10.1103/PhysRevD.69.011101.

[79] S. J. Asztalos et al. "SQUID-Based Microwave Cavity Search for Dark-Matter Axions". In: *Phys. Rev. Lett.* 104 (4 Jan. 2010), p. 041301. DOI: 10.1103/PhysRevLett.104.041301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.041301.

[80] J. Hoskins et al. "Modulation sensitive search for nonvirialized dark-matter axions". In: *Phys. Rev. D* 94 (8 Oct. 2016), p. 082001. DOI: 10.1103/PhysRevD.94.082001. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.082001.

[81] J.V. Sloan et al. "Limits on axion-photon coupling or on local axion density: Dependence on models of the Milky Way's dark halo". In: *Physics of the Dark Universe* 14 (2016), pp. 95–102. ISSN: 2212-6864. DOI: https://doi.org/10.1016/j.dark.2016.09.003. URL: http://www.sciencedirect.com/science/article/pii/S2212686416300504.

[82] S. Al Kenany et al. "Design and operational experience of a microwave cavity axion detector for the 20-100 μeV range". In: *Nucl. Instrum. Meth.* A854 (2017), pp. 11–24. DOI: 10.1016/j.nima.2017.02.012. arXiv: 1611.07123 [physics.ins-det].

[83] L. Zhong et al. "Results from phase 1 of the HAYSTAC microwave cavity axion experiment". In: *Phys. Rev.* D97.9 (2018), p. 092001. DOI: 10.1103/PhysRevD.97.092001. arXiv: 1803.03690 [hep-ex].

[84] Gray Rybka et al. "Search for dark matter axions with the Orpheus experiment". In: *Phys. Rev.* D91.1 (2015), p. 011701. DOI: 10.1103/PhysRevD.91.011701. arXiv: 1403.3121 [physics.ins-det].

[85] M. Hatridge, R. Vijay, D. H. Slichter, John Clarke, and I. Siddiqi. "Dispersive magnetometry with a quantum limited SQUID parametric amplifier". In: 83.13, 134501 (Apr. 2011), p. 134501. DOI: 10.1103/PhysRevB.83.134501. arXiv: 1003.2466 [cond-mat.mes-hall].

[86] V. E. Manucharyan et al. "Microwave bifurcation of a Josephson junction: Embedding-circuit requirements". In: *Phys. Rev. B* 76 (1 July 2007), p. 014524. DOI: 10.1103/PhysRevB.76.014524. URL: https://link.aps.org/doi/10.1103/PhysRevB.76.014524.

[87] B.D. Josephson. "Possible new effects in superconductive tunnelling". In: *Physics Letters* 1.7 (1962), pp. 251–253. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(62)91369-0. URL: http://www.sciencedirect.com/science/article/pii/0031916362913690.

[88] B.D. Josephson. "Supercurrents through barriers". In: *Advances in Physics* 14.56 (1965), pp. 419–451. DOI: 10.1080/00018736500101091. eprint: https://doi.org/10.1080/00018736500101091. URL: https://doi.org/10.1080/00018736500101091.

[89] P. W. Anderson and J. M. Rowell. "Probable Observation of the Josephson Superconducting Tunneling Effect". In: *Phys. Rev. Lett.* 10 (6 Mar. 1963), pp. 230–232. DOI: 10.1103/PhysRevLett.10.230. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.230.

[90] W. C. Stewart. "CURRENT VOLTAGE CHARACTERISTICS OF JOSEPHSON JUNCTIONS". In: *Applied Physics Letters* 12.8 (1968), pp. 277–280. DOI: 10.1063/1.1651991. eprint: https://doi.org/10.1063/1.1651991. URL: https://doi.org/10.1063/1.1651991.

[91] D. E. McCumber. "Effect of ac Impedance on dc Voltage Current Characteristics of Superconductor Weak Link Junctions". In: *Journal of Applied Physics* 39.7 (1968), pp. 3113–3118. DOI: 10.1063/1.1656743. eprint: https://doi.org/10.1063/1.1656743. URL: https://doi.org/10.1063/1.1656743.

[92] J. Clarke and A. I. Braginski. *The SQUID handbook*. Wiley-VCH, 2004.

[93] Claudia D. Tesche and John Clarke. "dc SQUID: Noise and optimization". In: *Journal of Low Temperature Physics* 29.3 (Nov. 1977), pp. 301–331. ISSN: 1573-7357. DOI: 10.1007/BF00655097. URL: https://doi.org/10.1007/BF00655097.

[94] Claude Hilbert and John Clarke. "Measurements of the dynamic input impedance of a dc SQUID". In: *Journal of Low Temperature Physics* 61.3 (Nov. 1985), pp. 237–262. ISSN: 1573-7357. DOI: 10.1007/BF00681634. URL: https://doi.org/10.1007/BF00681634.

[95] Michael Mück, Marc-Olivier André, John Clarke, Jost Gail, and Christoph Heiden. "Radio-frequency amplifier based on a niobium dc superconducting quantum interference device with microstrip input coupling". In: *Applied Physics Letters* 72.22 (1998), pp. 2885–2887. DOI: 10.1063/1.121490. eprint: https://doi.org/10.1063/1.121490. URL: https://doi.org/10.1063/1.121490.

[96] Michael Mück, Marc-Olivier André, John Clarke, Jost Gail, and Christoph Heiden. "Microstrip superconducting quantum interference device radio-frequency amplifier: Tuning and cascading". In: *Applied Physics Letters* 75.22 (1999), pp. 3545–3547. DOI: 10.1063/1.125383.

[97] R.F. Bradley. "Cryogenic, low-noise, balanced amplifiers for the 300-1200 MHz band using heterostructure field-effect transistors". In: *Nuclear Physics B - Proceedings Supplements* 72 (1999). Proceedings of the 5th IFT Workshop on Axions, pp. 137–144. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(98)00516-7. URL: http://www.sciencedirect.com/science/article/pii/S0920563298005167.

[98] Michael Mück, J. B. Kycia, and John Clarke. "Superconducting quantum interference device as a near-quantum-limited amplifier at 0.5 GHz". In: *Applied Physics Letters* 78.7 (2001), pp. 967–969. DOI: 10.1063/1.1347384. eprint: https://doi.org/10.1063/1.1347384. URL: https://doi.org/10.1063/1.1347384.

[99] Michael Mück, Christian Welzel, and John Clarke. "Superconducting quantum interference device amplifiers at gigahertz frequencies". In: *Applied Physics Letters* 82.19 (2003), pp. 3266–3268. DOI: 10.1063/1.1572970. eprint: https://doi.org/10.1063/1.1572970. URL: https://doi.org/10.1063/1.1572970.

[100] Michael Mück, Bernd Schmidt, and John Clarke. "Microstrip superconducting quantum interference device amplifier: Operation in higher-order modes". In: *Applied Physics Letters* 111.4 (2017), p. 042604. DOI: 10.1063/1.4985384.

[101] D. Kinion and John Clarke. "Microstrip superconducting quantum interference device radio-frequency amplifier: Scattering parameters and input coupling". In: *Applied Physics Letters* 92.17 (2008), p. 172503. DOI: 10.1063/1.2902173. eprint: https://doi.org/10.1063/1.2902173. URL: https://doi.org/10.1063/1.2902173.

[102] D. Kinion and John Clarke. "Microstrip superconducting quantum interference device amplifier: Conditional stability". In: *Applied Physics Letters* 96.17 (2010), p. 172501. DOI: 10.1063/1.3377898. eprint: https://doi.org/10.1063/1.3377898. URL: https://doi.org/10.1063/1.3377898.

[103] D. Kinion and John Clarke. "Superconducting quantum interference device as a near-quantum-limited amplifier for the axion dark-matter experiment". In: *Applied Physics Letters* 98.20 (2011), p. 202503. DOI: 10.1063/1.3583380. eprint: https://doi.org/10.1063/1.3583380. URL: https://doi.org/10.1063/1.3583380.

[104] National Communications System Technology & Standards Division. "Federal Standard 1037C". In: *General Services Administration Information Technology Service* (1996). URL: https://www.its.bldrdoc.gov/fs-1037/fs-1037c.htm.

[105] Sean R. O'Kelley. *srisyadasti/MSA-development: Supplementary Data and Software*. Jan. 2018. DOI: 10.5281/zenodo.1163747. URL: https://doi.org/10.5281/zenodo.1163747.

[106] J. Jaycox and M. Ketchen. "Planar coupling scheme for ultra low noise DC SQUIDs". In: *IEEE Transactions on Magnetics* 17.1 (Jan. 1981), pp. 400–403. ISSN: 0018-9464. DOI: 10.1109/TMAG.1981.1060902.

[107] Keith T. Regner et al. "Broadband phonon mean free path contributions to thermal conductivity measured using frequency domain thermoreflectance". In: *Nature Communications* 4 (Mar. 2013). Article, p. 1640. URL: http://dx.doi.org/10.1038/ncomms2630.

[108] V.B Efimov and L.P Mezhov-Deglin. "Phonon scattering in diamond films". In: *Physica B: Condensed Matter* 263-264 (1999), pp. 745–748. ISSN: 0921-4526. DOI: https://doi.org/10.1016/S0921-4526(98)01280-0. URL: http://www.sciencedirect.com/science/article/pii/S0921452698012800.

[109] Ray Radebaugh. *NIST Material Measurement Lab, Cryogenics Technologies Group*. Feb. 2018. URL: http://cryogenics.nist.gov/MPropsMAY/materialproperties.htm.

[110] K. Flachbart, A. Feher, Š. Jánoš, Z. Málek, and A. Ryska. "Thermal conductivity of NbTi alloy in the low-temperature range". In: *physica status solidi (b)* 85.2 (1978), pp. 545–551. ISSN: 1521-3951. DOI: 10.1002/pssb.2220850217. URL: http://dx.doi.org/10.1002/pssb.2220850217.

[111] R Berman, P R W Hudson, and M Martinez. "Nitrogen in diamond: evidence from thermal conductivity". In: *Journal of Physics C: Solid State Physics* 8.21 (1975), p. L430. URL: http://stacks.iop.org/0022-3719/8/i=21/a=003.

[112] P.G. Klemens. "Thermal Conductivity and Lattice Vibrational Modes". In: ed. by FREDERICK SEITZ and DAVID TURNBULL. Vol. 7. Solid State Physics. Academic Press, 1958, pp. 1–98. DOI: https://doi.org/10.1016/S0081-1947(08)60551-2. URL: http://www.sciencedirect.com/science/article/pii/S0081194708605512.

[113] S. M. Wasim and N. H. Zebouni. "Thermal Conductivity of Superconducting Niobium". In: *Phys. Rev.* 187 (2 Nov. 1969), pp. 539–548. DOI: 10.1103/PhysRev.187.539. URL: https://link.aps.org/doi/10.1103/PhysRev.187.539.

[114] David C. Larsen, ed. *Thermal Conductivity 16*. Boston, MA: Springer US, 1983, p. 162. ISBN: 978-1-4684-4265-6. DOI: 10.1007/978-1-4684-4265-6.

[115] T. Pynchon. *Gravity's Rainbow*. Penguin Publishing Group, 2012, p. 711. ISBN: 9781101594650. URL: https://books.google.com/books?id=GGPm4I3BbxAC.

[116] Francesco D'Angelo, Zoltán Mics, Mischa Bonn, and Dmitry Turchinovich. "Ultra-broadband THz time-domain spectroscopy of common polymers using THz air photonics". In: *Optics Express* 22(10) (2014), pp. 12475–12485. DOI: 10.1364/OE.22.012475. URL: https://www.osapublishing.org/oe/viewmedia.cfm?uri=oe-22-10-12475&seq=0.

[117] C. C. Grimes and Sidney Shapiro. "Millimeter-Wave Mixing with Josephson Junctions". In: *Phys. Rev.* 169 (2 May 1968), pp. 397–406. DOI: 10.1103/PhysRev.169.397. URL: https://link.aps.org/doi/10.1103/PhysRev.169.397.

[118] Roz Batten. Point of View challenge on dcpchallenge.com. Accessed: 2018-11-05. Oct. 2007. URL: https://www.dpchallenge.com/image.php?IMAGE_ID=600869.

[119] John M. Martinis, Rami Barends, and Alexander N. Korotkov. "Calculation of Coupling Capacitance in Planar Electrodes". In: *arXiv e-prints*, arXiv:1410.3458 (Oct. 2014), arXiv:1410.3458. arXiv: 1410.3458 [physics.class-ph].

[120] Marc-Olivier André, Michael Mück, John Clarke, Jost Gail, and Christoph Heiden. "Radio-frequency amplifier with tenth-kelvin noise temperature based on a microstrip direct current superconducting quantum interference device". In: *Applied Physics Letters* 75.5 (1999), pp. 698–700. DOI: 10.1063/1.124486. eprint: https://doi.org/10.1063/1.124486. URL: https://doi.org/10.1063/1.124486.

Appendices

Appendix A

Varactor Parameter Extraction

Generating a graph of varactor parameters as a function of temperature and applied voltage, (as in figure 7.7) is not entirely straightforward. The raw S_{11} data contains the needed signal, but it also contains a large amount of spurious reflected power. The reflected signal is small because several 10's of dB of attenuation is explicitly added between room temperature and the mK stage to avoid illuminating the sample with 300 K microwave radiation. This attenuation diminishes both the illuminating power and reflected power. The spurious signal is large because a very high quality SMA connector on a microwave line may have a reflection coefficient of -30 dB, but a value of -20 dB or greater is more typical, and there are necessarily many couplings between the room temperature and mK stages of the dilution fridge. Even with great care, the reflected signal is small relative to the spurious reflected power. Fortunately, the spurious reflections are static while the signal of interest varies with varactor voltage, allowing us to see motion against a "bright, static" background.

This appendix details the software and data analysis tools used to remove the spurious background signal and extract varactor parameters from the cleaned signal. Section 7.1.3 takes the signal analysis as given and presents the temperature dependence of the varactor parameters. Throughout this section, references are made to specific file formats, datasets, and custom software. All such references refer to electronic material available at [105]. While it is possible to follow this section without reference to this supplementary material, the motivated reader is able to fully audit the software, methods, and data.

Raw data from the S_{11} sweeps is saved both in a human-readable file format (*.xls) for guaranteed future readability and in a binary file format (*.dat) more amenable to immediate computer analysis. The human-readable format generates two files for each tuning voltage, one for amplitude and one for phase. The temperature, frequency range and tuning voltage are embedded in the file names. The human-readable files consist of two columns of text; one for frequency and the other for amplitude or phase. The binary files group all data at a given temperature in a single file, with a structure described in figure A.1. Storing the data in human-readable form requires about twice the storage capacity as the binary form, but it is well worth it. Keeping data in a human-readable format ensures future compatibility and readability, and makes possible outside review by independent researchers. Sample data

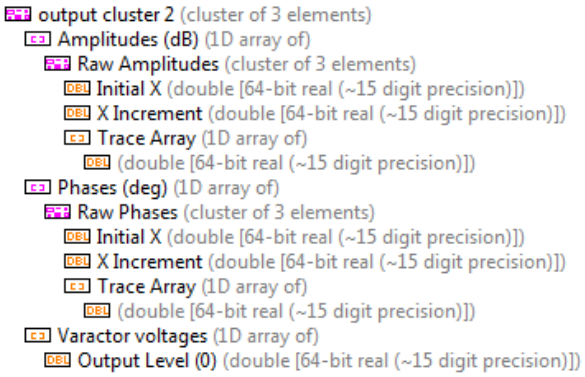

Figure A.1: Graphical type definition of the saved voltage sweep data. This is useful for someone using the software available at [105].

Figure A.2: File path to *.dat data file. This is useful for someone using the software available at [105].

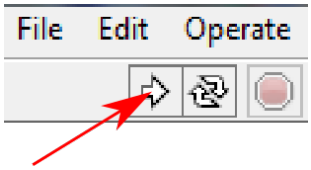

Figure A.3: The "run once" button. This is useful for someone using the software available at [105].

taken at 300 K and 100 mK are available in both "human-readable" (text) and "machine-readable" (binary) formats on github [105] along with the LabVIEW typedef necessary to interpret the binary file.

The program "linear resonator analysis_2.vi" was created for the visualization and analysis of the binary S11 data, and its functioning is described here. The reader is encouraged to find the software and sample data at [105] and follow along. The binary data (*.dat file) in "\Data\Linear Resonator for Varactor characterization\300K evenly spaced V voltage sweep" is used in the examples here, but any *.dat file in the "\Data\Linear Resonator for Varactor characterization\" directory may be used. If following with the software, first enter the path to the *.dat file as shown in figure A.2 (click the folder icon to browse to it), then hit the "run once" arrow in the top left, as shown in figure A.3.

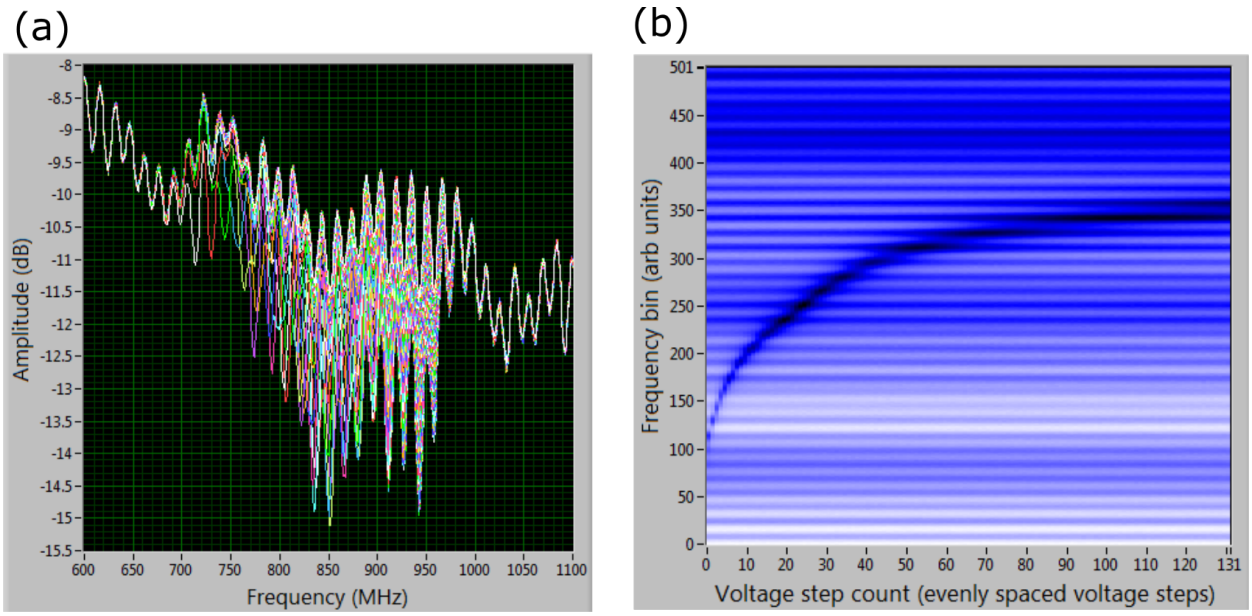

Figure A.4: Raw S_{11} amplitude spectra. Tuning is apparent from about 750 to 975 MHz, but there is a strong systematic signal common to all traces. (a) Overlaid S_{11} amplitude spectra. Each color corresponds to a different varactor voltage. (b) Map of all S_{11} amplitude spectra. Systematic signal is apparent in the horizontal bands. Tuning trend is apparent in the dark "square root" shaped band.

A.1 Removing Contamination from Raw Data

The spectra are loaded and displayed in their raw form, shown in figures A.4 and A.5. The data we need are visible (log-like curves in A.4b and A.5b) but a large amount of systematic contamination precludes curve-fitting in this state.

Figure A.4a (A.5a) shows all of 130 amplitude (phase) vs. frequency spectra overlaid on the same graph, and figure A.4b (A.5b) shows the same amplitude (phase) data on an intensity chart with uncalibrated scales in frequency and tuning voltage as the vertical and horizontal axis respectively, with color coding for the amplitude (phase). Whereas the spectra would ideally be Lorentzian functions centered at various frequencies, they are in practice contaminated by a strong ripple, with dominant periodicity of about 14 MHz and also other slower variations. The frequency of 14 MHz corresponds to a 1/2 wavelength of about 7 m given a propagation velocity of $v = 0.67c$, typical of most 50-Ω cabling. This matches the length of cable between the room-temperature VNA and mK sample box. The contamination must be removed before any data can be extracted.

Among the amplitude spectra, the contamination is common to each spectrum but the spectra differ substantially among themselves, and the "true function" is strictly non-positive, so the contamination can be removed by the following simple operation: at each

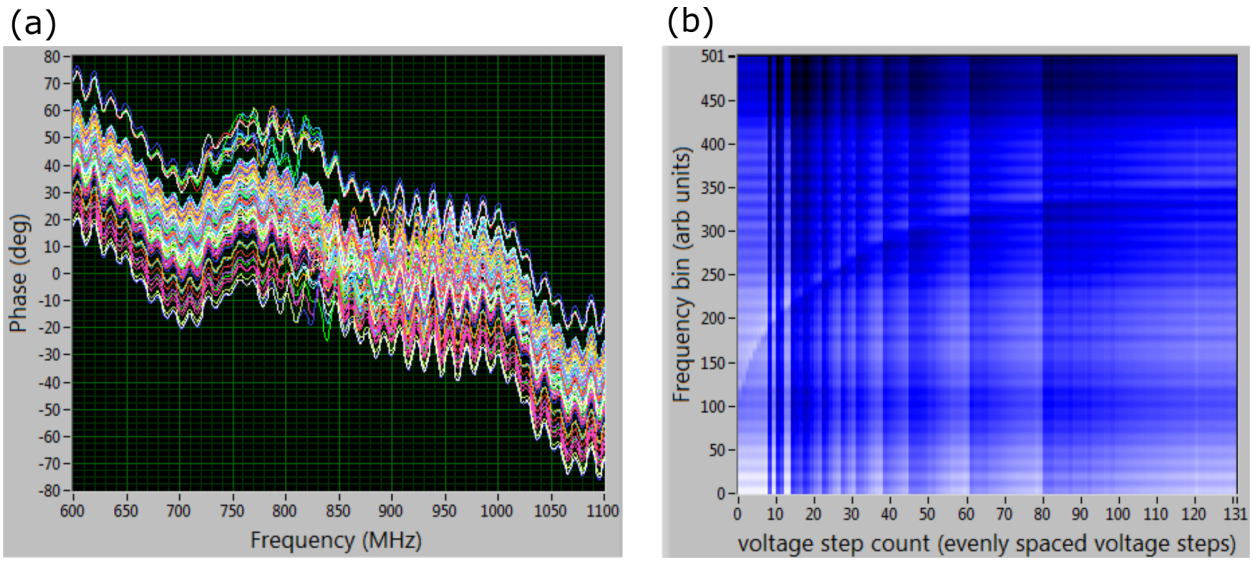

Figure A.5: Raw phase S_{11} spectra. (a) Overlaid phase spectra. Each color corresponds to a different varactor voltage. A systematic ripple with a period of about 14 MHz period is clear, but tuning trends are obscured by various phase offsets. (b) Map of all phase spectra. The 14 MHz systematic function appears as horizontal bands, and occasional phase offsets appear as vertical bands. The real tuning signal is faintly visible as a "square root"-like contrast edge, rising left to right.

frequency the largest amplitude recorded across *any* spectrum is noted and attributed to the contaminating ripple. This "amplitude de-ripple" function is shown in figure A.6.

The "amplitude de-ripple" function is subtracted from each amplitude function and the results are shown in figure A.7. This leaves about 0.1 dB offset error equal to about 1/2 the peak-to-peak noise remaining, but it is inconsequential and does not affect the extraction of important parameters from the spectra. Compare figures A.7 and A.4 to see the effectiveness of this amplitude "de-ripple" operation.

Correcting the phase data shown in figure A.5 is less straightforward because the "true" function has both positive and negative excursions, and the contamination has multiple systematic components. First of these is an arbitrary phase offset for each spectrum created as an artifact of the VNA phase calculation. Second is a 14 MHz ripple corresponding to the 7 m of cabling between the sample and VNA. Third is an overall trend in phase v. frequency due to an imprecise signal delay setting in the VNA. To correct the first contaminant, each spectrum is shifted by a constant phase such that the phase at the lowest frequency is zero, with the results shown in figure A.8. Compare figures A.5 and A.8 to see the effect of this initial correction. The vertical bands have been removed but the horizontal bands remain.

To remove the 14 MHz ripple and other remaining trends, a "background contamination" function is constructed from the data represented in figure A.8. The high-frequency half of

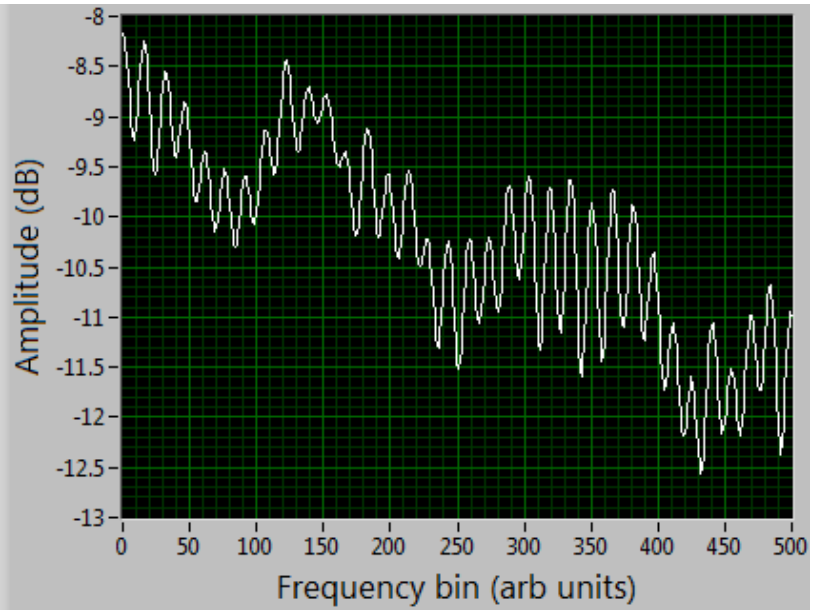

Figure A.6: The "amplitude de-ripple" function. At each frequency, the greatest amplitude value of all 131 tunings is selected to produce this "de-ripple" function. The difference of the raw data and this function gives functions clean enough for automatic fitting.

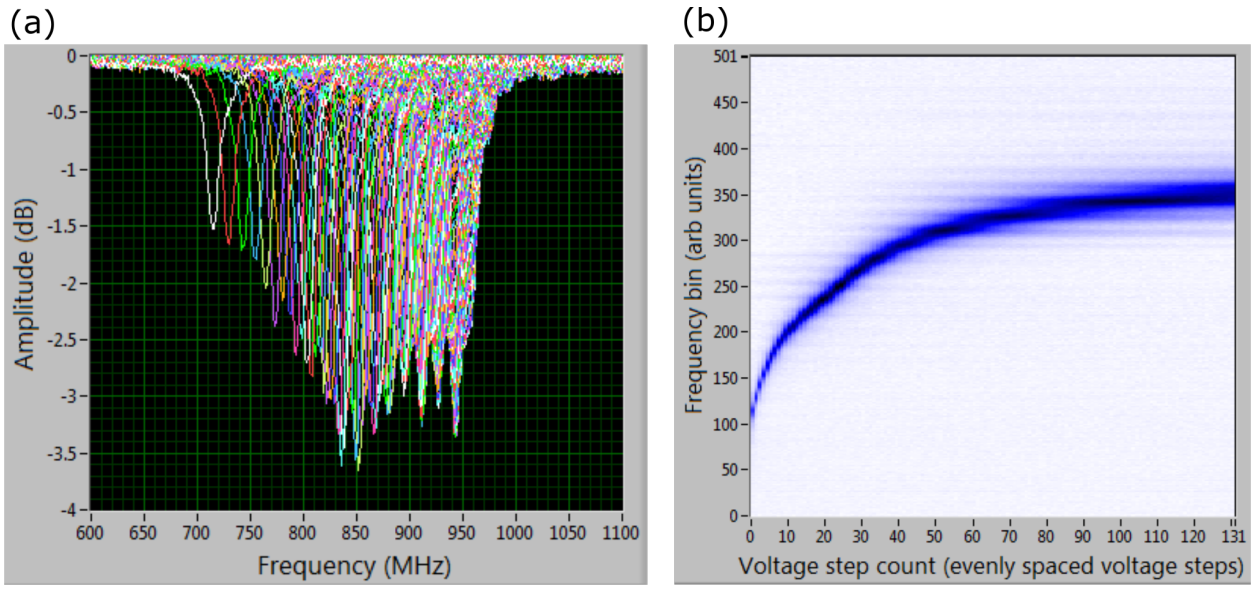

Figure A.7: Amplitude spectra with contamination removed. (a) Overlaid amplitude spectra. Each spectra is a readily recognizable Lorentzian, and no systematic trend is obvious. (b) Map of all amplitude spectra. The tuning trend is now very clear, with some small systematic contamination remaining, visible as faint horizontal bands.

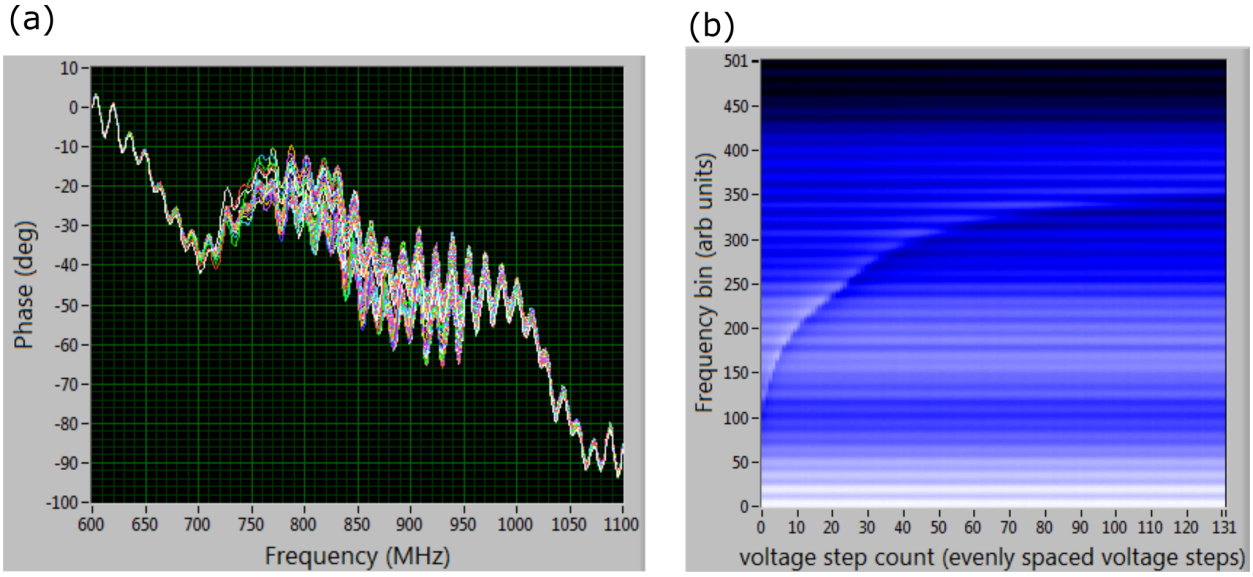

Figure A.8: Phase spectrum after offset. Each phase spectra has been offset such that $\theta(600\text{ MHz}) = 0$. (a) Overlaid phase spectra. There appears to be a tuning trend, but systematic effects still dominate. (b) Map of all phase spectra. Systematic trends appear as horizontal bands, and the tuning trend appears at a "square root"-like contrast edge (dark below, light above) rising from left to right.

the lowest-tuned frequency spectrum is taken to be all "contamination" (i.e. assume the actual phase = 0° here). Similarly, the low-frequency half of the highest-tuned spectrum is identified with the other half of the "contamination function". These two half-spectra will compose the whole "phase contamination" spectrum, but are discontinuous at the center frequency where they meet. To make a continuous "phase contamination" function, each candidate half-function is added to a 1/4-wave sine function such that they meet smoothly at the center frequency without discontinuity. (There is nothing special about the 1/4 sine function. It is just a convenient function with zero amplitude on one end and zero slope at the other.)

Figure A.9 demonstrates the construction of the phase "background contamination" function. The highest-tuned phase spectrum is in red, the lowest-tuned phase spectrum is in white, and the stitched-together background phase in green. An uncontaminated S_{11} phase spectrum will be equal to zero far from the resonance, negative just below resonance, positive just above the resonance, and equal to zero on resonance. Casual inspection of figure A.9 shows that the difference between the red and green traces has the correct form for an S_{11} phase signal. The difference between the white and green traces also has the correct form, though the, though that difference is more subtle in this case. In both cases, the phase contamination is of greater magnitude than the signal, so it is not yet clear if the phase can

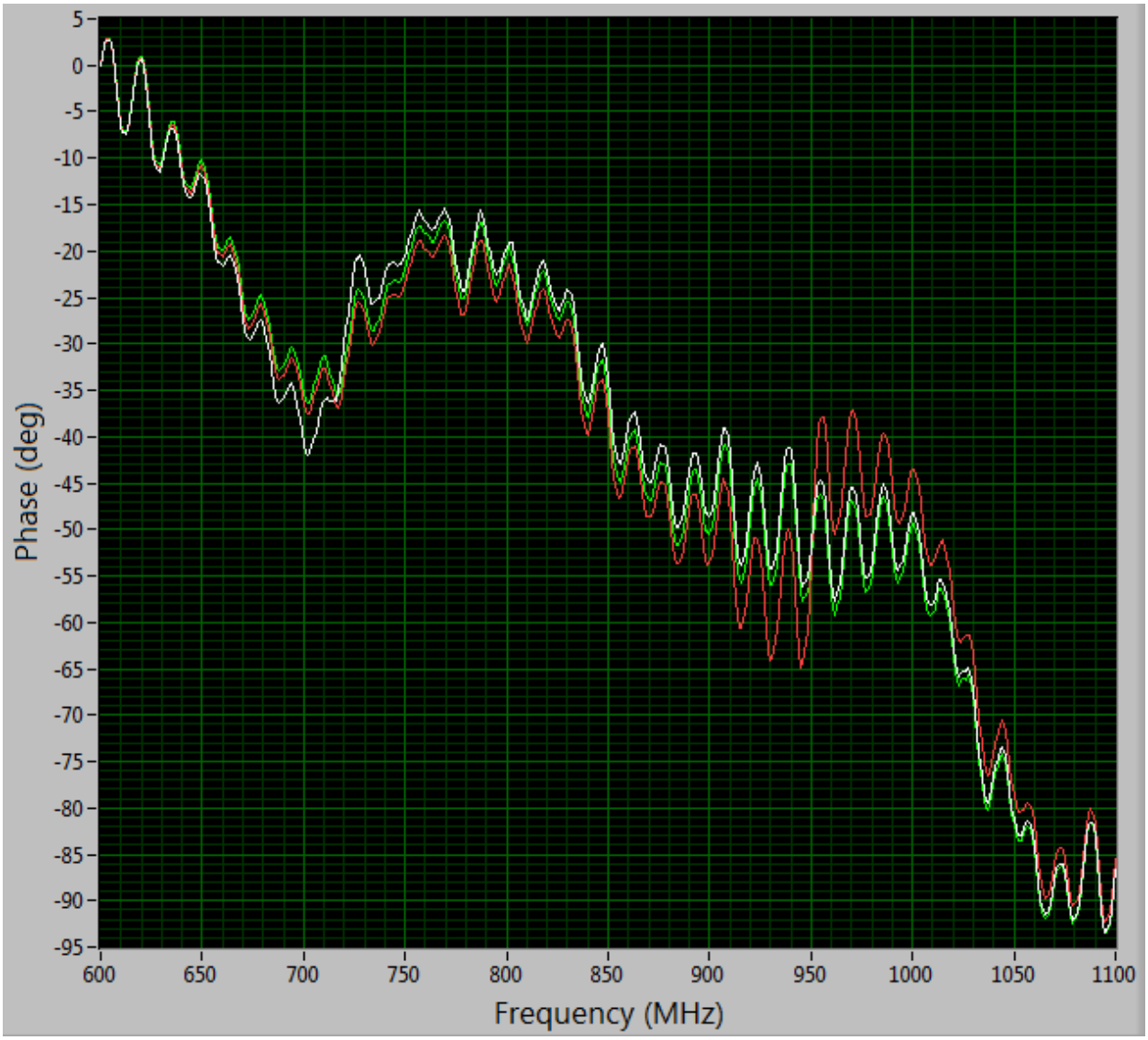

Figure A.9: Stitched-together phase background. This function is the attempt to remove the common systematic contamination of the phase data. White is the lowest tuned spectrum, and red is the highest tuned spectrum. Green is a composite function made from the left half of the red function and right half of the white function, with a 1/4 sine shape added to each side as needed to ensure a smooth transition at the center.

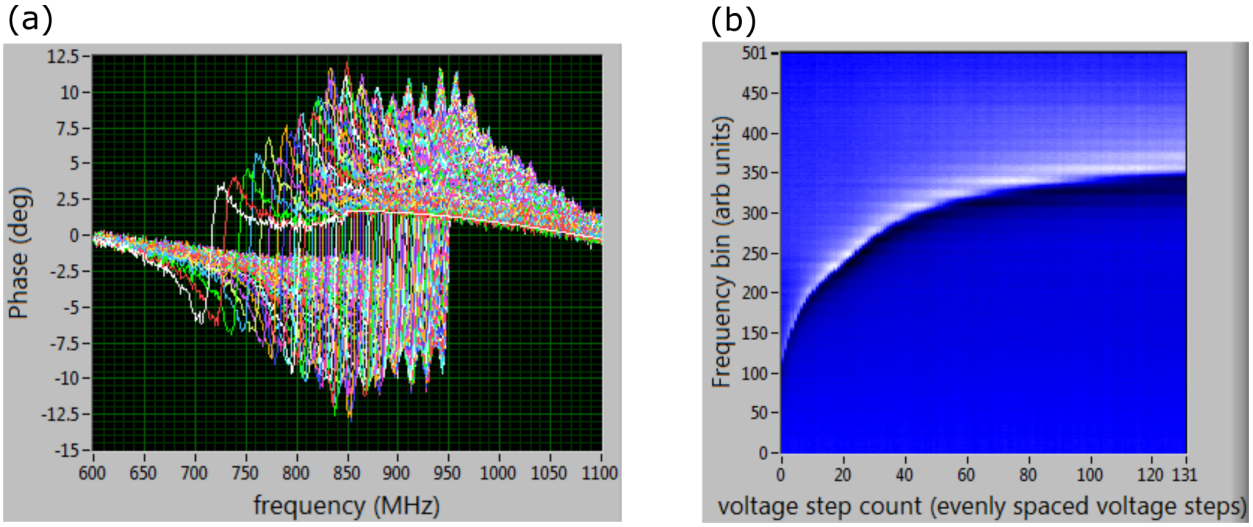

Figure A.10: Fully corrected phase spectra. (a) Overlaid phase spectra. Both the "down-up" phase shape and tuning trends are now visible. (b) Map of all phase spectra. The tuning trend appears at a clear "square root"-like contrast edge (dark below, light above) rising from left to right. A small amount of systematic contamination remains as faint horizontal and vertical banding.

be successfully de-contaminated.

The "phase contamination" background phase shown as a green trace in figure A.9 is subtracted to all phase spectra, with the result shown in figure A.10a. Note that every spectra now has the correct general form and a tuning trend from about 720 MHz to is 950 MHz is clear. The tuning trend is now the dominant feature of the map in figure A.10b as the "square root" contrast edge, with negative values (black) just below resonance and positive values (white) at frequencies just above resonance. A small amount of contamination remains: visible cross-hatching in figure A.10b, and the 1/4 sine offset to the phase is clear in the lowest-tuned phase spectrum visible in figure A.10a, but the results are now sufficiently clean to reliably extract device parameters via an automated routine. Compare figures A.10 and A.5 to see the effectiveness of the complete phase "de-ripple" operation.

A.2 Extracting Device Parameters

The "contamination removal" operation described in section A.1 is completed more or less instantly by the program "linear resonator analysis_2.vi", which then turns to the job of extracting device parameters from the cleaned data. This starts by fitting a Lorentzian function to each amplitude spectrum, which can be monitored in the window "Data with fit" shown in figure A.11. In addition, amplitude and phase data are plotted as real and

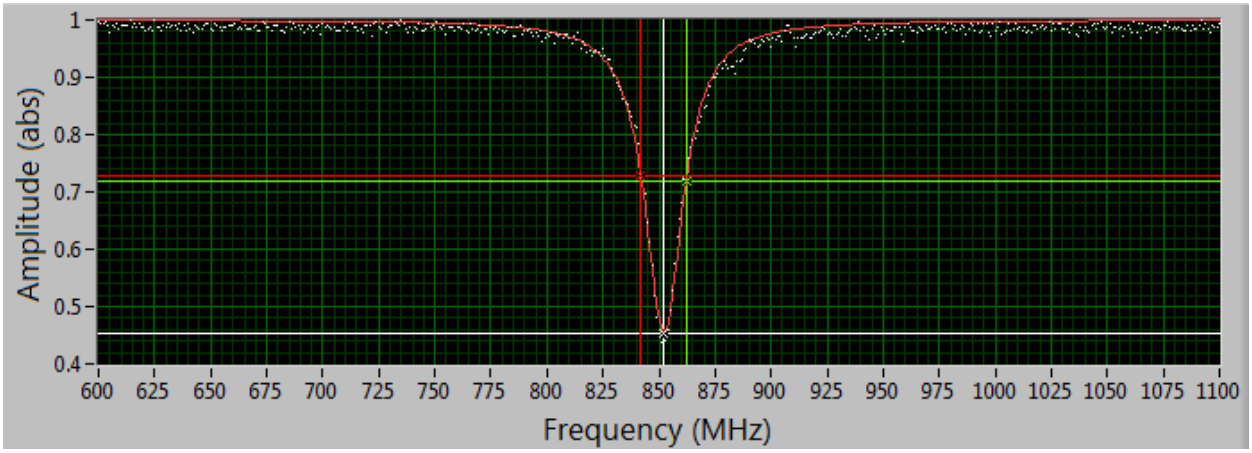

Figure A.11: Sample fit to amplitude spectrum. White points are data and red line is the fit. After signal conditioning to remove the systematic contamination the fit to data is quite tight. The fit is not perfect: notably, the signal conditioning method results in data that are strictly < 1, while the measurement noise should result in occasional values > 1. Despite this, the extracted parameters of f and Q are reliable. White cursors mark center frequency and amplitude minimum, extracted from Lorentzian fit. Red and green cursors mark the -3dB points.

Table A.1: Arguments for the numerical model used to extract C_{end} from spectral data.

R_{in}	C_{in}	Z_0	f_0	f	R_{end}	C_{end}
50	0.3	50	1005	[from data]	0	–
Ω	(pF)	Ω	(MHz)	(MHz)	Ω	(pF)

imaginary components of the S_{11} signal in figure A.12. Figure A.12a shows data from a single spectra as white boxes and a fit to an ideal resonator shown as a red line. Figure A.12b shows an overlay of 131 spectra as colored lines and the center point of each fit circle as a white box.

These "monitoring portals" allow for the human oversight often necessary to ensure the automated fitting algorithms behave reasonably. In figure A.11 the white cursor marks the extracted center frequency, and the red and green cursors mark the -3dB points used for calculating Q. The measurements are extracted from the fit, rather than the raw data, so contamination from the remaining ripple is minimized.

To extract the varactor value C_{end}, the resonant frequency is input to the numerical model described in section 5.1 with arguments listed in Table A.1.

The results of the "value extraction" process at a single temperature are shown in figure A.13. For varactor biases from +1 V to -12 V, figure A.13a shows tuning from about 710 MHz to 950 MHz, figure A.13b shows a rather complex variation of Q between about 30 and 44, figure A.13c shows a calculated capacitance of about 2.5 to 0.1 pF per varactor,

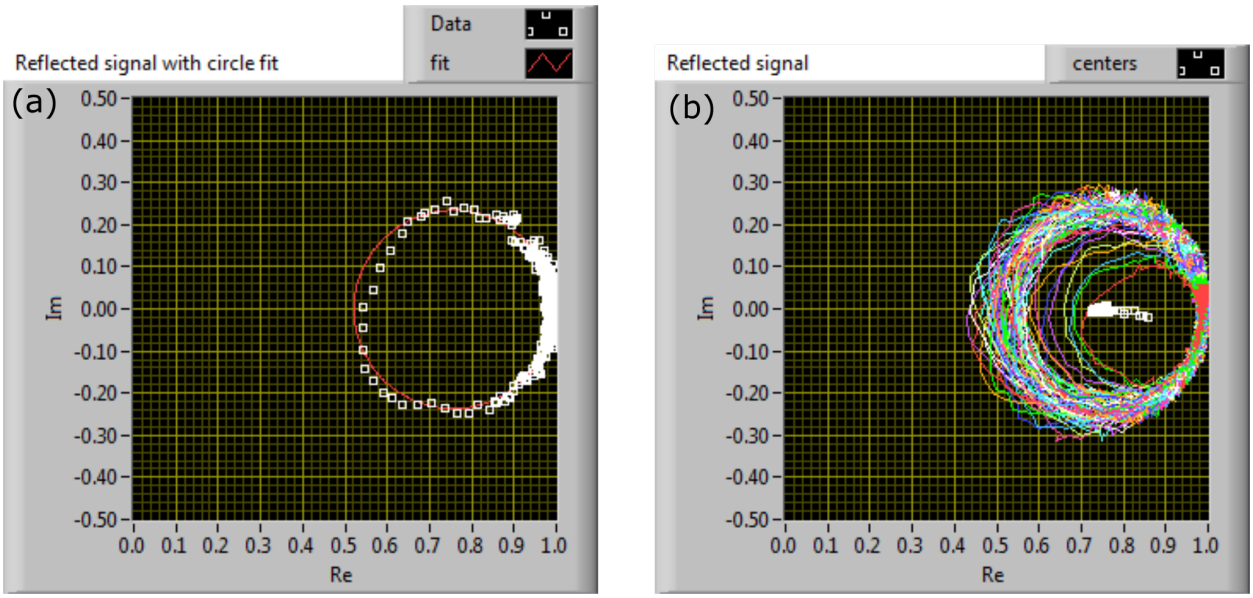

Figure A.12: Complex plots of reflected signals. Ideal ideal reflection data would from a perfect circle. Data here demonstrates that the remaining systematic errors are largely eliminated by the fit to an ideal reflection model. (a) Single-tuning I-Q plot with data and fit. (b) Multiple-tuning I-Q plot with data only.

and figure A.13d shows the Q data against frequency rather than voltage, which reveals the "fast" Q fluctuations to be roughly periodic in f, suggesting the origin is from the testbed and not the varactors themselves.

The data in figure A.13 were taken at even voltage intervals, but the capacitance and thus tuning vary most rapidly at voltages near 0 V and we would prefer data spread evenly across frequencies. With the data represented in figure A.13a as a guide, voltage steps are selected for a second data collection run with approximately even frequency intervals. Figure A.14 demonstrates the results, with a nearly linear function of frequency vs. data point. Compare to figures A.14 and A.7b to assess the effectiveness of choosing uneven voltage intervals to achieve even frequency intervals.

Figure A.15 shows the dependence of frequency, Q, and C as a function of varactor bias in the same way as figure A.13, but with even frequency steps.

The data are collected and analyzed as described here at 24 temperatures ranging from 300 K to 60 mK. The extracted values for f, Q, and C versus bias voltage are available at [105] in the folder "\Data\Linear Resonator for Varactor characterization\60mK to 300K extracted f, Q, and C vs V". These data are in the form of comma separated values (*.csv), so it is easily readable by any software, but the reader is encouraged to use the provided program "linear resonator plotter.vi" (either as LabVIEW program or Wndows compatible standalone) to review these data easily in a graphical format.

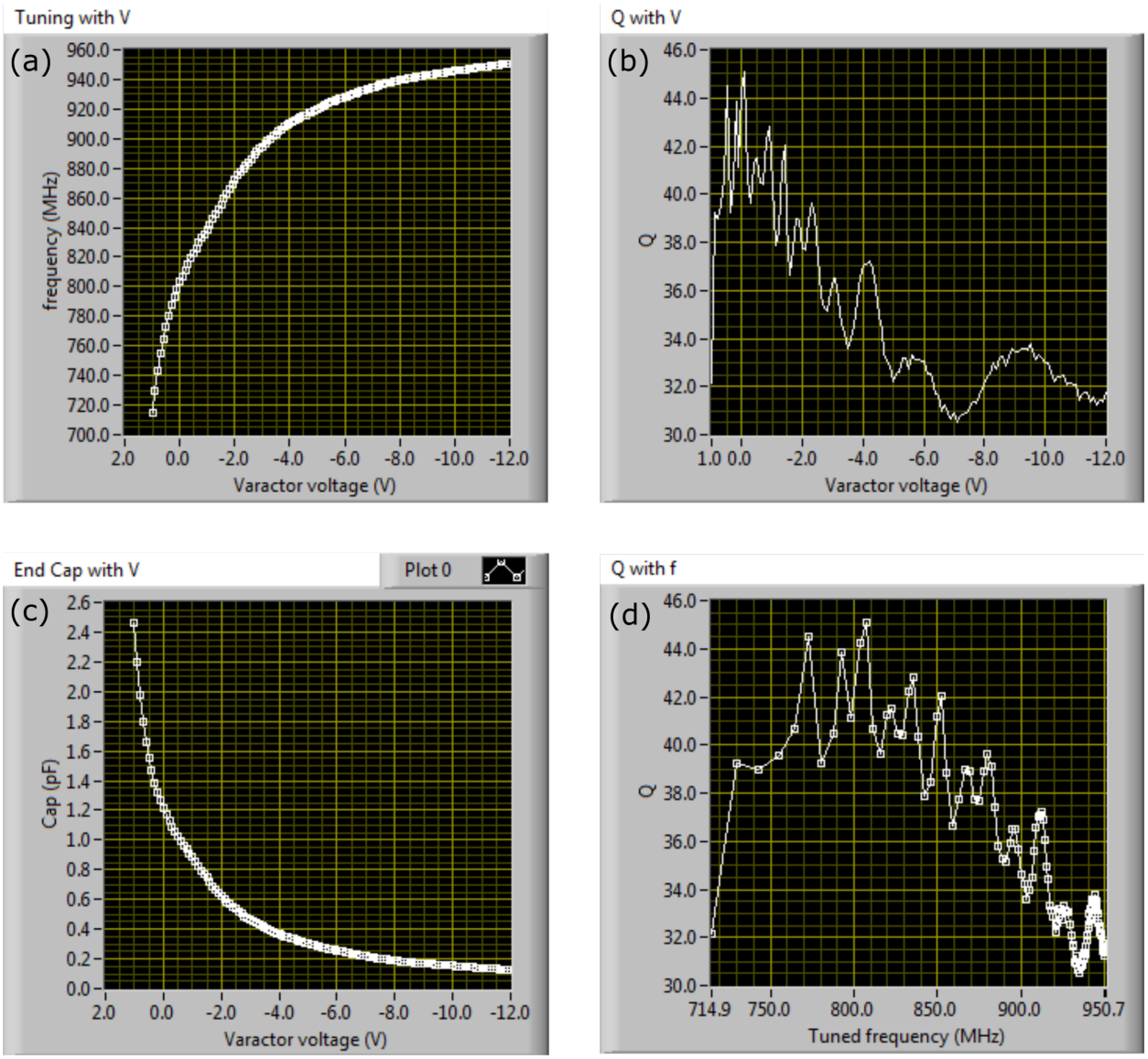

Figure A.13: Parameters extracted from resonator spectra, T = 300 K, evenly spaced voltages. (a) Resonator frequency vs. varactor voltage. (b) Resonator Q vs. varactor voltage. (c) Varactor capacitance vs. voltage. (d) Resonator Q vs. frequency.

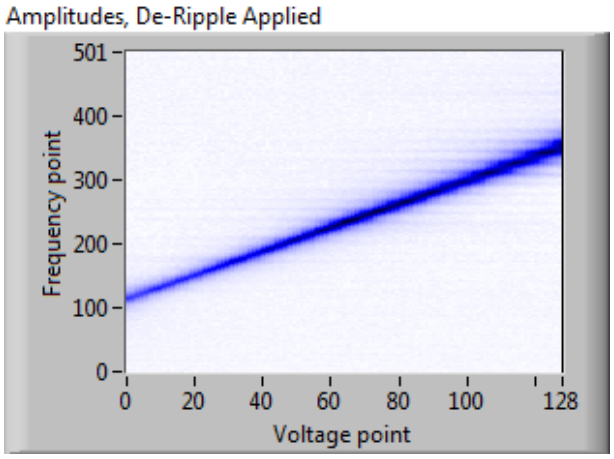

Figure A.14: Map of all amplitude spectra, tuning steps chosen for even frequency steps.

This analysis reveals a surprising temperature dependence of the varactor capacitance, explored further in section 7.1.3.

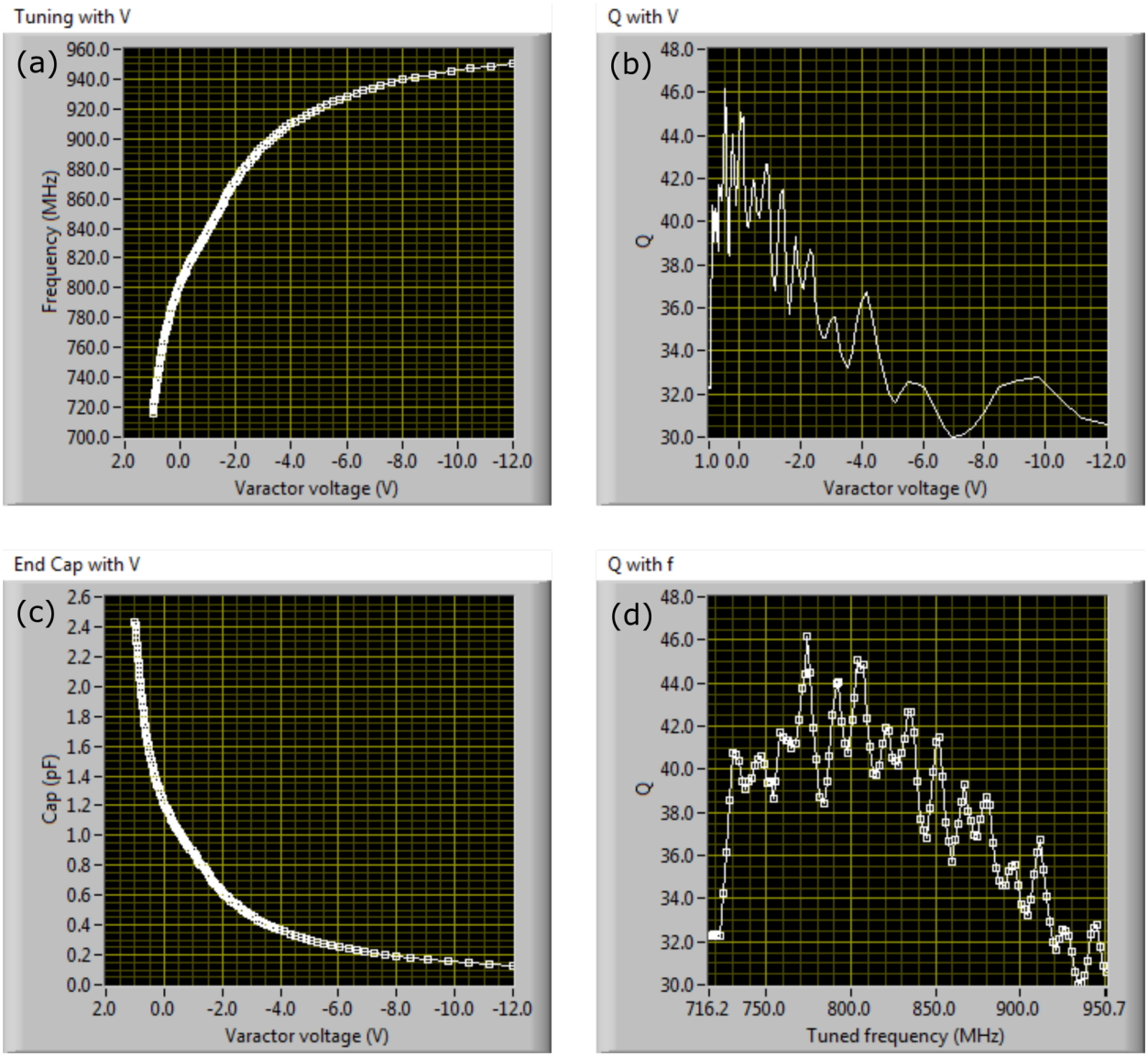

Figure A.15: Parameters extracted from resonator spectra, T = 300 K, evenly spaced frequencies. (a) Resonator frequency vs. varactor voltage. (b) Resonator Q vs. varactor voltage.(c) Varactor capacitance vs. voltage. (d) Resonator Q vs. frequency.