Advancing the State-of-the-Art of Microwave Astronomy:

Novel FPGA-Based Firmware Algorithms for the Next Generation of Observational

Radio and Sub-millimeter Wave Detection

by

Jonathan Hoh

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2023 by the
Graduate Supervisory Committee:

Christopher Groppi, Chair
Tracee Jamison-Hooks
Edward Buie
George Trichopolus
Evan Scannapieco

ARIZONA STATE UNIVERSITY

May 2024

ABSTRACT

This dissertation presents a comprehensive study on the advancement of astrophysical radio, microwave, and terahertz instrumentation/simulations with three pivotal components. First, theoretical simulations of high metallicity galaxies are conducted using the supercomputing resources of Purdue University and NASA. These simulations model the evolution of a gaseous cloud akin to a nascent galaxy, incorporating variables such as kinetic energy, mass, radiation fields, magnetic fields, and turbulence. The objective is to scrutinize the spatial distribution of various isotopic elements in galaxies with unusually high metallicities and measure the effects of magnetic fields on their structural distribution.

Next, I proceed with an investigation of the technology used for reading out Microwave Kinetic Inductance Detectors (MKIDs) and their dynamic range limitations tied to the current method of FPGA-based readout firmware. In response, I introduce an innovative algorithm that employs PID controllers and phase-locked loops for tracking the natural frequencies of resonator pixels, thereby eliminating the need for costly mid-observation frequency recalibrations which currently hinder the widespread use of MKID arrays.

Finally, I unveil the novel Spectroscopic Lock-in Firmware (SpLiF) algorithm designed to address the pernicious low-frequency noise plaguing emergent quantum-limited detection technologies. The SpLiF algorithm harmonizes the mathematical principles of lock-in amplification with the capabilities of a Fast Fourier Transform to protect spectral information from pink noise and other low-frequency noise contributors inherent to most detection systems. The efficacy of the SpLiF algorithm is substantiated through rigorous mathematical formulation, software simulations, firmware simulations, and benchtop lab results.

ACKNOWLEDGMENTS

I would like to acknowledge the fact that the first draft of this dissertation is due in two hours and I dont have time to think of everybody I need to thank. This will be updated over the next few days.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Astronomy, Scale, and the Quest for the Quantum Limit

Across all categories of RF and mm/sub-mm astronomy, the universal goal of a detection mission is to capture electromagnetic fields from cosmic sources and extract meaningful information. What qualifies as "meaningful" is wholly dependent on the mission at hand; something considered noise by one observation team may be the primary target of another. The vast majority of instruments developed for space detection observe either from Earth, low-Earth orbit (LEO), or a gravitationally stable Lagrange point between the Earth and Sun. To humans, a space-observation satellite in LEO and a satellite at L2 (the second Sun-Earth Lagrange point) 1.5×10^6 km from Earth are surveying the cosmos from vastly different observation points. [Giovinazzi]. By no means is this objectively false, and compared to the distances encountered in daily life, it is completely rational to consider the space between these two satellites an unfathomable stretch of space.

Nevertheless, this human point of view is limited by the scale of what is encountered on Earth. Astronomy and cosmology dive into distance scales that are far outside the comprehension of the human mind. While the distance between an LEO and an L2 satellite is enormous to a human, zooming out to the scale of the Solar System provides a perspective where the same separation is quite humble. Zooming out once again enters the interstellar scale, where the distance between the Sun and its closest neighbor Proxima Centauri is over 4.2 light years, or $\approx 4 \times 10^{13}$ km [Bean]. Using this distance of 4.2 light years as the smallest distance seen on the interstellar scale, the distance between the two observation satellites (which moments ago seemed nearly infinite to humans) is ten million times smaller. At

the point where the scope reaches galactic scales, the separation between the LEO and L2 satellites is absolutely trivial; the satellites may as well be sitting on top of one another.

The massive scales encountered in cosmology, especially with respect to the instruments

Figure 1.1: The classic "Earthrise" photo shows how perspective changes scale. From the moon, everything that has ever existed in human history is found on the very surface of a single blue sphere. *Credit: Bill Anders, 1968*

used for observation, have profound effects on the nature of information in experimental astrophysics. Detection of information from deep space cannot be aided by simply placing a detection system in a "better" location. These signals are produced in environments that size and distance completely eclipse any concept of three-dimensional space that can feasibly be utilized by astronomers. The electromagnetic perturbations produced by distant galaxies, sought by astronomers, and measured by telescopes will be uniformly distributed regardless of the placement of the telescope.

This is not a claim that the readings from a telescope placed randomly on Earth will be identical to the readings of the same telescope placed on the moon or built into a satellite. In the case of ground-based telescopes, a significant amount of electromagnetic radiation from distant sources gets absorbed by the Earth's thick atmosphere, which ranges in opacity across the electromagnetic spectrum(Rees, 2012).

1.2   Drawing a Physical Picture: Molecular Clouds

The overwhelming scale of galactic and extragalactic sources clearly does not lead to uniform detection on smaller scales. Instead, this uniformity applies to the field distributions created by the sources. For a more salient example, consider a massive cosmic structure such as a molecular cloud. Molecular clouds are cold and dark regions of space composed mainly of molecular hydrogen (H_2) with other molecules scattered throughout. Although this environment is cold and diffuse compared to more mature galaxies, it also provides the perfect environment for the growth of protostars. Due to this property, molecular clouds are often referred to as "stellar nurseries" and understanding their evolutionary mechanisms is essential for demytifying star formation. Observing molecular clouds can be a tricky task due to their low virial temperatures, with regions ranging 100K all the way down to ¡10K. Fortunately, due to the abundance of H_2, there is a considerable presence of water vapor within the molecular clouds. Water vapor has a 556 GHz rotational emission produced by the spin-flip of the ortho-ground state rotational transition $1_{10} \rightarrow 1_{01}$ of the 2 hydrogen atoms that comprise H_2O, and this transition conveniently only needs around 27 Kelvin to trigger.

Without diving too deep into the abyss of statistical mechanics, assume that there is enough external radiation to excite the H_2O molecules within a molecular cloud. An equation can then be formed to express how often the 556 GHz rotational transition will occur. Since this value is a number over time, the equation which governs this process is aptly called the "rate equation", which takes form as:

$$R_{ij} = n_i B_{ij} \rho(\nu) + n_j A_{ji} \qquad (1.1)$$

Here, n_i and n_j are the number densities of molecules in states i and j, $\rho(\nu)$ is the energy density of the radiation field at frequency ν, and the constants A_{ji} and B_{ji} are the Einstein coefficients which account for all possible transitions between the two states. Very briefly, the Einstein coefficients can be described as follows:

1. (A_{ji}) is the spontaneous rate coefficient. This is the probability per unit time that a transition will occur spontaneously from state i to state j, with units of s^{-1} (per second). It represents the innate tendency of a system to transition from a higher-energy state to a lower one, emitting a photon in the process.

2. (B_{ji}) is the stimulated emission and absorption coefficient. This is a more complex term than A_{ji} and represents the probability per unit time, per unit energy density, that a transition will occur due to the presence of an external radiation field. Unlike the simple dimension of the spontaneous rate coefficient, its units are $m^3 \cdot J^{-1} \cdot s^{-1}$.

Within these Einstein coefficients, numerous elements in the physical environment of the molecular cloud are encoded. Those physical factors that are not encoded in these coefficients are either represented by the external radiation field $\rho(\nu)$ or are woven into the number densities n_i and n_j. Although these number densities are simply the physical number density of atoms in states i and j, calculating them requires the use of the Boltzmann distribution. Within this distribution, parameters such as the temperature and overall density of water vapor in the molecular cloud come into play. Expressed in full, the equation for determining the number density of atoms in a given state is:

$$n_i = n \frac{g_i}{Q} e^{-E_i/kT} \tag{1.2}$$

where n is the total number density, g_i is the degeneracy of state i, Q is the partition function, E_i is the energy of state i, k is the Boltzmann constant, and T is the temperature.

Leaving the world of statistical mechanics and re-entering the realm of astronomy, this rate equation based on the parameters of the molecular cloud is directly tied to the intensity of 556 GHz photons that can be seen. The intensity I of the radiation emitted in a given frequency band is directly related to the number of photons emitted per unit of time, per unit area, and per unit of solid angle. Mathematically, this can be expressed as

$$I = \frac{dN}{dt dA d\Omega} \tag{1.3}$$

where dN is the number of photons, dt is the observation time, dA is the area, and $d\Omega$ is the solid angle. In the case of a molecular cloud, the rate equation R_{ij} gives us the net rate of transitions from state i to state j per unit volume. Each of these transitions results in the emission of a photon with a frequency around 556 GHz. Therefore, R_{ij} can be thought of as the number of such photons emitted per unit time, per unit volume. Relating the rate R_{ij} to the observed intensity I, requires only integrating R_{ij} over the volume V of the molecular cloud that contributes to the observed emission:

$$I \propto \int_V R_{ij} dV \tag{1.4}$$

Using a proportionality constant here allows for the inclusion of factors like the energy of each emitted photon, the efficiency of the emission process, and any attenuation factors between the molecular cloud and the observer (Agarwal, 1974).

With that, the discussion once again returns to size scales. The reason that observations from space-based satellites contain so much more information is not due to their location, but rather the fact that they do not have to compete with the attenuation caused by the Earth's atmosphere.(Matsushita *et al.*, 1999). The relationship in eq.(3.4) shows that the intensity I of the observed emission is a measure of the energy received per unit area, per unit time, per unit solid angle. This is an intrinsic property of the molecular cloud and is independent of the observer's location (provided, of course, there is no intervening medium

5

to attenuate the signal such as an atmosphere). Molecular clouds are literally astronomical in scale, often spanning tens to hundreds of parsecs. Given their enormous size and the vast distances at which they are located (often hundreds or thousands of parsecs away), the angular size they subtend in the sky is relatively stable. Whether an observing satellite is in low-Earth orbit or a Lagrange point is irrelevant, as the angular size of the cloud remains essentially the same.

1.3   Front-End to Back-End Chains: No Tolerance for Weak Links

For those designing astronomy instrumentation, the uniform and isotropic nature of fields from distant sources means that building a system sensitive enough to collect power from a targeted galaxy will also be sensitive to all other sources that manage to reach the detectors. Infinite resources can be dumped into an ultrasensitive telescope capable of accurately measuring the Cosmic Microwave Background, but no matter where that telescope is placed, it will have to contend with the noise caused by the very same 556 GHz spectral line used for demystifying stellar evolution. The crux of the matter lies in the details of the instrument design, which includes not just the telescope optics but also the detectors, filters, and data processing algorithms. In many respects, the challenges posed by modern RF and millimeter-wave astronomy reach comical levels of complexity and difficulty. The reality of attempting to capture unique spectroscopic images from space is similar to sitting in front of an orchestra and attempting to determine whether the second-chair violin has a string out of tune. While each instrument may be playing an intricate set of notes, each note brimming with countless harmonics of meaningful information, it is all considered noise to those only interested in that second-chair violin. No matter where a listener sits, there is truly no seat in the concert hall that will make the job easier, only seats that amplify the already Sisyphean difficulty of the task.

Before buying too much into this metaphor, the fundamental differences between sound

produced by orchestral instruments and light produced by energy-state transitions in atoms must be kept clear, with the prior being a matter wave and the latter being electromagnetic. However, this does not change the comparability of the goals and challenges in isolating specific sources. Given the uniformity of signatures from celestial sources across the cosmos, any receiving instrument will necessarily also pick up radiation from unwanted field sources across a wide range of frequencies. Extraction of the target signal requires exceptional precision and accuracy at all stages of the detection system.

Consider a radio telescope designed to detect spectral lines from a distant molecular cloud. The telescope must collect as many photons as possible from this faint target, while excluding contaminating radiation from the myriad other sources shining across the radio spectrum. This necessitates an extremely sensitive front-end receiver and optics. However, even with a flawless telescope and noise-free amplifier, the back-end signal processing can undermine everything gained up front. Any noise added after light collection will be indistinguishable from the astronomical signal itself. This is because the electromagnetic waves from all field sources arrive identically at the aperture. Once collected, the only way to isolate the target photons is through spectroscopic filtering and noise reduction in the receiver electronics. Any spurious signals introduced along the detection chain cannot be separated from the 'clean' photons emerging from the telescope. This is why state-of-the-art instruments push the limits of low-noise engineering at every stage. The front-end optics and receivers are designed to efficiently gather target photons while minimizing contaminating stray light and thermal noise. The back-end signal chain then preserves this "noiseless" condition through extensive electromagnetic shielding, vibration damping, and thermal stability. Digitization, filtering, and data processing are performed with ultra-low-noise electronics and algorithms (Longo *et al.*, 2014).

Every aspect of a deep-space detection system is engineered to add as little noise as possible. Even the smallest amounts of added noise can override gains made earlier in the

system. The intrinsic quantum noise of the astrophysical signal itself is the only true limiter on the noise floor, and detection systems which achieve this incredible performance are referred to as "quantum limited". But absent a perfect detection system at all stages, this limit remains out of reach for most technologies available to astronomers. Thus, a sort of "cooperative arms race" materializes where improvements to any piece of a detection system force all the following elements to keep pace, lest they become the weakest part of the signal chain and negate any enhancements made in the overall signal-to-noise ratio.

1.4   The Endless War on Noise

1.4.1   Front-End Development

To the astronomical instrument designer, signal-to-noise (SNR) is the single most important metric of performance in a system. SNR is the simple but vital measurement of incoming fields that expresses how much of their total power is desired information. If there is more noise power in the bandwidth of a measurement than signal power, the signal is undecipherable. Only when signal power dominates the total power of a measured field will there be any chance to extract its contents. Due to the natural attenuation of electromagnetic fields as they propagate through space, signals get dimmer as their source distances lengthen. This unavoidable truth means that in order to successfully obtain information from distant galaxies, noise must be minimized to whatever extent possible by whatever means possible. For the front-end of a system, many noise sources originate from the natural world. Atmospheric effects play a significant role in introducing noise in RF and millimeter/submillimeter wave astronomy. The atmosphere of the Earth can cause phase noise due to an inhomogeneous and time-variable distribution of water vapor (Battat *et al.*, 2004). Water vapor in the atmosphere also creates a non-uniform tranmission win-

dow across frequencies, one of the most opaque being the very same 556 GHz line used for detecting molecular clouds (Matsushita and Matsuo, 2003) (Matsuo *et al.*, 1998a) (Matsuo *et al.*, 1998b). These sources of noise degrade distant, faint signals from space before they even interact with an instrument. Naturally, detectors can also introduce noise that affects the sensitivity and performance of the system. A common target of microwave astronomy is the CMB, an echo from the last moments of the Big Bang that exists in all parts of the universe at an incredibly low temperature of 2.7 kelvin. Even though these CMB photons are universally present in all space, they are not uniformly isotropic, and measuring the subtle differences in their angular distribution requires exceptionally precise measurement (Hajian and Souradeep, 2003). Pursuits like measuring the anisotropy of CMB photons are what create technologies such as kinetic inductance detectors: a resonating pixel that changes fundamental frequency with incident power due to interactions with superconducting electrons, and can reach single-photon levels of sensitivity. Since MKIDs rely on superconductivity to detect light, they require whole cryogenic coolers to lower them to operating temperatures of nearly absolute zero. The lack of resistivity and thermal noise in cryogenically cooled superconducting detectors has made them explode in popularity over the last decade and have set a new standard in the noise floor of astronomy instruments (C. Yates *et al.*, 2017) (Endo, 2015). At microwave and terahertz frequencies, the difficulty of extracting meaningful information from cosmic sources comes not only from the required SNR, but also from the oscillation frequencies themselves. Most analog-to-digital converters (ADCs) on the market today cannot exceed a few Gsps without introducing crippling quantization noise. A common but clever alternative to direct microwave detection is a heterodyne receiver. With the use of a local oscillator (LO), detectors in this category receive high-frequency signals from space and downmix them to RF frequencies that are much easier to process. This aids in avoiding the noise common in high-frequency microwave electronics, but replaces it with another dominating noise source. Heterodyne

mixing requires a device that can be switched on and off a billion times per second without losing fidelity of the signal, and as a result, most modern heterodyne systems use super-conducting mixers such as superconductor-insulator-superconductor (SIS) mixers, which as their name suggests, also require superconductivity to function. Though SIS mixers offer low noise performance, their conversion loss and noise performance can be affected by factors such as the quality of the superconducting materials used and the design of the mixer (Uzawa *et al.*, 1998) (Karpov, 1998) (Li *et al.*, 2008).

1.4.2   *Consequences for Back-End readout*

All of the detectors and mixers mentioned above are vulnerable to a litany of noise-producing conditions and are constantly being improved by groups around the world. The common goal that unites and motivates all advancements in these technologies revolves around lowering the intrinsic noise floor for microwave detection. But even in an ideal world with flawlessly fabricated receivers and mixers, detection of faint RF signals from space would not magically become simple. Advanced hardware components offer im-proved sensitivity, bandwidth, and multiplexing capabilities, but they also require sophis-ticated readout techniques to extract and process the acquired data efficiently and without ruining the signal information. MKIDs, for example, have the ability to be tuned to specific resonant frequencies, and by tuning each element of an array to a slightly different reso-nance, thousands of these incredibly sensitive detectors can be used like pixels in a camera. This is an enormous leap in sensitivity compared to past pixel technologies, but attempts to read out each detector individually would waste so much time, power, and effort that there would be no net benefit to using them. More sensitivity means that more informa-tion is collected from space, and that information must be converted to bits. Furthermore, increasing spectral bandwidth in spectroscopy requires a linear increase in sampling speed

of the detected information (Rantwijk *et al.*, 2016). Computers are not equipped to handle input data rates higher than what is required for consumer interfaces, and therefore can only process small-bandwidth spectra. To effectively use computers to read out an array of MKIDs (usually numbering in the thousands of individual detectors), there would need to be a computer for every single pixel. Since grants typically do not provide funds to purchase 2000-4000 individual computers, a more efficient solution is required. Field-programmable-gate-arrays (FPGAs) are arrays of CPU logic cells in the form of NAND gates that can be arranged to mimic virtually any physical circuit in existence. Due to their flexibility and capacity to process data at astronomically high rates, FPGAs have become the primary choice of hardware platform to tailor specific computational tasks in RF astronomy, including the readout of MKIDs.

1.5   FPGA-based Firmware for Detector Readout

By every technical measurement, FPGAs are far superior to CPUs as a solution for modern astronomy readout. Much like software, firmware has quietly become part of every day life for most people whether they know it or not. Any interaction that occurs between the analog and digital worlds requires some level of firmware to function; from computer keyboards, to touchscreens, to phone cameras, the technology of the digital age is only possible due to the rules of interaction established by firmware. Software benefits greatly from its commercial visibility, and thanks to the explosion of software-based applications entering every corner of life, the concepts and underpinnings of what software is and how it is developed have become common knowledge. Even the most staunch Luddites have been forced to accept that software is intimately tied to fundamental elements of society, such as banks, healthcare, and commerce. Hardware holds an even more ubiquitous understanding in society as it benefits from being corporeal and physically observable. Given how present both hardware and software are in the cultural Zeitgeist, it would seem obvious that

11

firmware, the crucial link between the two, would be equally as understood. Yet, despite its equally widespread presence, the firmware is a veiled mystery to virtually everyone, except those who choose to consciously understand it.

One reason for the lack of knowledge about firmware is its inherent complexity. Firmware is typically written in low-level programming languages and is specific to the hardware it controls. Understanding and working with firmware requires expertise in hardware architecture, embedded systems, and low-level programming, all of which present brutal learning curves to beginners (Zaddach *et al.*, 2014). The complexity of firmware analysis and reverse engineering can be further compounded by the lack of access to source code or hardware documentation. Due to the intrinsic nature of firmware connecting hardware and software systems, it is often developed in some amount of secrecy as to not expose any vulnerabilities in consumer products. A lack of educational programs teaching firmware in a manner equivalent to those teaching software or hardware leaves a frightening absence of online resources for self-teaching. Even if one decides to learn from existing examples, reverse engineering firmware often involves manual effort and specialized tools to extract and analyze the code (Chen *et al.*, 2018) (Fowze *et al.*, 2019). Furthermore, firmware is not directly visible or accessible to end-users, unlike software applications for which graphical user interfaces have been standard fare for decades. Firmware does not benefit from being the end-product for users, operating at a lower level with almost all firmware interactions being between it and the hardware components of a device rather than with the device operator. (Lehniger *et al.*, 2022). If one were to imagine a person as a digital machine, the body representing hardware and the learned behaviors of the brain being software, then firmware is represented by the nervous system. Much like the human nervous system, firmware is the vehicle for properly transmitting and interpreting information between the outside analog world and the internal digital world.

1.6   Strengthening the Weak Link

As the technology for hardware and software in modern astronomy continues to grow exponentially, the unfortunate reality is that unless firmware grows alongside them, no meaningful change can progress. Throughout the next three chapters, I will unravel some of the knots in firmware holding state-of-the-art detection technologies from reaching their true potential. In doing so, I will not only illuminate the problems faced by the slow pace of firmware development, but also pose and defend my solutions to these difficulties. Before doing so however, it is crucial to contextualize the purpose of the instrumentation in question. Detection systems in astronomy maintain their place at the forefront of data sciences due to the absurd measurements that are required for solidifying theoretical estimates and models for the structure of the universe. Just as a scientist need understand the instrument that collected their data for a proper interpretation, instrument engineers do not have a full grasp on what is "meaningful" information until they themselves experience how their data is used for the larger picture of astrophysics knowledge. Thus, as a prelude to my proposed advancements in astronomy readout firmware, I will spend the next chapter introducing and elaborating on the results of supercomputer simulations based on the very galaxy formation process from which my firmware advancements hope to collect empirical readings.

Chapter 2

SUPERCOMPUTER SIMULATIONS OF HIGH METALLICITY GALACTIC FORMATION

2.1 Galactic Evolution and the Circumgalactic Medium

Galactic formation is a process that takes place over a vast range of time scales, spanning anywhere from hundred of millions to billions of years. The exact amount of time that passes during the life cycle of a galaxy can vary drastically due to cosmological parameters such as the nature of dark matter and the rate of expansion of the universe, as well as the complicated and improbable nature of the physics of star formation and gas cooling. Regardless of the specific conditions that create a galaxy, its formation is sure to be extraordinarily long and unique in its own way. The entire existence of humanity is but a mere pinpoint compared to the evolution of galaxies such as the Milky Way, and thus for most of history, astronomers did not even attempt to gain an understanding of the dynamics and details about formation processes. However, in the century that has passed since Einstein first penned his theory of general relativity laying out the rules that govern space and time themselves, constant breakthroughs in detection and simulation technology have allowed humans to peer into the past and piece together the puzzle of cosmic origins. In particular, the origins of how galaxies form from the nearly empty vacuum of space has puzzled astronomers for centuries. Starting around the turn of the 20th century, advancements to optical telescopes allowed for glimpses of spiral arms filled with gas around galaxies similar to the Milky Way. Interest in these exterior gasses only continued to ramp up as large radio telescopes became common and showed consistent noise in these boundaries between galaxies and the void beyond. Originally deemed the "Galactic Corona", the chaotic yet

near-invisible borders of galaxies were observed relentlessly with little to no illuminating findings.

In modern times, it is known that these "Galactic Coronas" are in fact an ionized halo of gas present around galaxies known as the circumgalactic medium (CGM). The difficulty experienced in observing details of the CGM was not the fault of astronomers in the past, but rather the incredibly diffuse nature of this gaseous medium. Little emission can be seen from the CGM and the method for understanding its makeup relies on the light from other sources. In particular, astronomers now use active galactic nuclei (AGN), one of the brightest objects in the universe and also known as Quasars, as background illumination. Through spectroscopic observations of CGM located conveniently between Earth and distant Quasars, astronomers can measure absorption lines rather than emission lines to determine the chemical makeup of these previously beffudling regions.

The invention and widespread use of absorption spectroscopy for elucidating details of the CGM has become one of the most important discoveries in the quest to understand galaxy formation processes (Tumlinson *et al.*, 2017). The interactions of the CGM consist of dynamic processes that allow for the transfer of matter and energy into their central galaxy such as inflows and outflows driven by various feedback processes. Acting as a fuzzy border between a galaxy and the surrounding intergalactic medium (IGM), it comes as no surprise that the dynamic activity in the CGM determines many fundamental characteristics that define galactic makeup. Through constant interaction, the CGM is fed rare elements and isotopes (known as metals) by galactic outflows from its inner galaxy along with accretion of intergalactic gases. This unceasing activity on both sides leaves the CGM as a major resovoir of galactic metals, indicating its important role in the heavy metal enrichment of galaxies (Tumlinson *et al.*, 2011). Simulations have shown that smaller, low-mass galaxies grow directly from the inflow of gasses from the CGM, along with accretion directly from the IGM. Results such as these highlight the essential nature of the CGM in the growth of

early galaxies (Zhang *et al.*, 2023).

A result of the constant feedback processes that occur in the CGM is an environment that does not comfortably sit in any given state of chemical equilibrium. Non-equilibrium chemistry within CGM refers to the chemical reactions that occur out of thermal equilibrium, where the reaction rates are not solely determined by temperature but also by other factors such as ionization rates, radiation fields, and the presence of shocks. The existence of non-equilibrium in the CGM has important implications for interpretation of galaxy formation and therefore has been the focus of countless studies, both in observation and simulations (Richings and Schaye, 2016). Observations have provided some of the most compelling results, showing that the presence of unexpected molecules in the inner winds of asymptotic giant branch stars is a result of non-equilibrium chemistry due to shock propagation (Cherchneff, 2006). Additionally, non-equilibrium effects have been found to impact the cooling functions of cooling gas in the diffuse interstellar medium (Richings *et al.*, 2014). This change in cooling effects is intrinsically linked to the metallicity within the CGM, therefore it is unsurprising that non-equilibrium chemistry in CGM is found to have profound effects on both the metallicity and star formation rates inside galaxies (Walch *et al.*, 2011).

2.2   A Brief History of CGM Understanding

2.2.1   Breakthroughs in Cosmology

Alongside the rapid advancement in RF technology that came as an unintended consequence of the post-WW1 technology boom, the concept of using machines as a method of computation also entered the consciousness of the scientific community. However, the advances made in astronomy during and after WW1 is barely a sneeze compared to the hurricane of detection and computation technology developed as a result of WWII. Putting aside

the ethical and moral conundrums of wartime science booms, this period of competitive research and global cooperation ushered in the mastery of radio frequency communications and the realization of machine computation in the form of the Turing machine. Once the war officially ended, entire new sectors of science and mathematics that were trail-blazed in secret by wartime research became available to the academic public at large, and from there the true digital renaissance was born. From there, Bardeen, Brattain, and Shockley at the legendary Bell Laboratories made the next major breakthrough for computation in 1947 with the invention of the first functioning transistor capable of amplifying electrical signals effectively. The transistor quickly rose to prominence as a means to store binary states of information which could be electrically manipulated to calculate complex mathematics. During this same period, the basic model of the RF telescope developed by Grote Rever in 1937 had been iterated upon by astronomy groups across the world, allowing for the first time the large scale analysis of the cosmos in wavelengths other than those seen by the human eye (Field, 2003). While RF telescopes were a groundbreaking invention, the lack of storage options for electromagnetic signals picked up by RF receivers severely limited the capacity to observe for meaningfully long periods. Fortunately transistor technology was shrinking in size at an exponential rate. Aggressive competition for developing digital storage methods meant that by the 1960s, transistors had become the building blocks of integrated circuits (ICs) which were able to pack thousands of transistors onto a single chip. These ICs provided a vector for astronomers to not only store, but also process the information from RF telescopes.

The Cold War between the United States and the Soviet Union meant that wartime funding was being dumped into all emerging areas of science and technology, especially research in digital signal processing. Decades of intensive research culminated in 1965 with James Cooley and John Tukey publishing a paper titled "An Algorithm for the Machine Calculation of Complex Fourier Series". This publication quickly gained traction as it allowed

17

for an incredibly efficient Fourier Transform estimation, today known as the Fast Fourier Transform (FFT). FFTs integrated into the back-end electronics of radio equipment meant that RF telescopes were not only able to take images outside of visible light, but also break it down into its fundamental spectral components. Spectral analysis immediately became an integral part of observational cosmology and its continuous application only improved its capabilities over the following decades (Wilson, 2011).

As with most data-centered technologies, the experimentation and discoveries made by RF astronomers using the FFT slowly trickled into consumer products. The optimization of ICs allowed the concepts of Turing machines to be condensed into incredible calculation machines known today simply as "computers". The ongoing demands of pushing the envelope in experimental cosmology continued to drive breakthrough after breakthrough until computers the size of classrooms such as ENIAC I and II shrunk down to sizes that could fit comfortably inside offices, and eventually on office desks. Theoretical astrophysicists were more than willing to accept larger amounts of computing power at lowering costs to help compute numerical solutions to problems without analytical answers. Although the contributions of computers to theoretical research was unequivocally game-changing for the latter half of the twentieth century, the concept of fully rendered, three-dimensional simulations was still far out of reach.

Modern Simulations of the CGM The last decade has seen significant progress in the ability to simulate and understand the CGM. Early on, studies focused on the overall structural processes of the CGM such as density, temperature, and metallicity distributions(Tumlinson *et al.*, 2017). As computational power and techniques advanced, simulations began to incorporate more detailed physics, including radiative cooling, feedback from star formation and active galactic nuclei, and the effects of galactic outflows (Hafen *et al.*, 2019). These large-scale simulations have been instrumental in understanding the origins and evolution of the CGM. They have provided insights into the processes driving

gas accretion to galaxies, the enrichment of CGM with metals, and the impact of feedback mechanisms on CGM properties (Hafen *et al.*, 2019) (Hafen *et al.*, 2020). Simulations have also been used to study the interaction between the CGM and galaxies, including the effects of galaxy mergers and the role of the CGM in the regulation of star formation (Davies *et al.*, 2020).

Comparisons of data collected through observation against simulation has been crucial for validating and refining hypothesized models of galaxy formation. In doing so, simulations have become accurate to the point of being able to reproduce observed properties of the CGM, such as the mass-metallicity relation, the distribution of ionized gas, and the kinematics of the CGM (Peeples *et al.*, 2019) (Sanchez *et al.*, 2019). Not only can the results of simulations be used to confirm observational discoveries, but their reliability has reached such heights that observational missions are carried out based on predicted simulation targets(Robert *et al.*, 2018). The ongoing advancements in computational capabilities and modeling techniques continue to provide more capabilities for simulation. High-resolution simulations with improved physical models are being developed to capture the complex processes occurring within the CGM, including the effects of magnetic fields, cosmic rays, and the interplay between different gas phases (Liang and Remming, 2019).

2.3   MAIHEM Models of the CGM

Emergent supercomputer technologies have vastly expanded the ability of researchers to borrow computational power from clusters across the United States for studies and experiments that are far too costly in time and power to be reasonably run on commercial CPUs. For those attempting to simulate CGM evolution, the last decade has ushered in a revolutionary number of resources to test and iterate through possible models. During this time, a code known as "Models of Agitated and Illuminated Hindering and Emitting Media" (MAIHEM) has been consistently improved upon with increasingly complex parameters

19

to simulate the dynamics of the CGM. As a cooling and chemistry package built using the FLASH hydrodynamics code, MAIHEM models hydrodynamics in three dimensions (3D). Importantly, it includes a non-equilibrium chemistry network of 65 ions, such as hydrogen, helium, carbon, nitrogen, oxygen, neon, sodium, magnesium, silicon, sulfur, calcium, iron, and electrons. It solves for various physical processes, including dielectric and radiative recombinations, collisional ionizations with electrons, charge transfer reactions, and photoionizations by a UV background. The code is used to simulate the multiphase nature of the magnetized CGM and its observable properties, such as absorption line tracers (Buie II *et al.*, 2022).

In his seminal work "The Influence of Turbulence and Magnetic Fields on the Non-Equilibrium Chemistry Evolution in the Halos Surrounding Milky Way-like Galaxies", Professor Edward Buie dives deep into the mechanics of CGM simulations using MAIHEM. The three main chapters of this dissertation are based off of three publications he wrote, each adding more complexity to the model. All three models notably contain a novel addition of turbulence which previously was not included in the MAIHEM simulation. Given the feedback mechanisms present in physical CGM, the inclusion of turbulence to the evolution of a non-equilibrium CGM model has substantial impact. From his results, turbulence can clearly be seen promoting the mixing of cold gas and hot ambient medium throughout the simulation. As a result, the intermediate snapshots of the simulation agree better with the CGM observations than any previous models. While the effects of including turbulence are clearly significant, only simulations of galaxies matching the parameters of the Milky Way were examined.

2.3.1   Solar Metallicity CGM Modeling

In this section, I will present the results of simulations from the turbulent MAIHEM model which explore galaxies which have metallicities much greater than that of the Milky Way. Analysis of observation from the COS-MAGIC survey show the existence of galaxies with metallicities equivalent to that of the Sun, three times higher than that of the Milky Way. Typically, galaxies follow a relationship known as the mass-metallicity relationship and typically do not deviate very far from this curve. However, at high and low stellar masses, the scatter of the relation is significantly higher, suggesting that merging events and/or stochastic accretion and star formation may drive galaxies outside the relation (Bothwell *et al.*, 2013). Rapidly, the galaxy becomes massive enough to produce and retain metals, and it evolves along the shallow branch toward higher metallicities (Agertz and Gieles, 2016). Furthermore, active starburst regions are capable of propelling shocks that affect the typical accretion inflows of CGM. In these cases, the additional influx of metals from the tumultuous starburst zones will force an unusually high amount of metals into the dynamic process. Realistically, a multitude of drivers can be attributed to the existence of Solar metallicity galaxies based purely on the number that have been observed in recent surveys.

To understand the intermediate mechanics of what occurs in these rare-ion-rich environments, the MAIHEM model was tuned to simulate the evolution of a three-dimensional volume equivalent to the size of the ionized halo around the Milky Way galaxy and broken down into thousands of smaller 3-d cells to aid in simulation. Virial parameters were used to initialize total mass of the galaxy and the general proportion of 12 percent baryonic material versus 88 percent dark matter determines the total mass of gas placed in the simulation. Temperature was also derived based on Virial relationships, in this case, the potential energy of the gravity from the dark matter halo of the galaxy is related through

21

the Virial theorem to a kinetic energy, and by a factor of the Boltzmann constant, used to calculate the starting temperature.

Column densities for Si ii, Si iv, N v, and O vi were the main focus of the chemical distribution in this high metallicity model. These ions were chosen because they are commonly observed in the halos of low-redshift galaxies and are representative of different ionization states. More specifically, Si ii and Si iv sample low and intermediate ionization state material, respectively. Inversely, O vi and N v sample higher ionization state material. These ions and their ratios were chosen in the larger context of understanding galaxy evolution because they provide insights into the physical processes and conditions in high-metallicity CGM. Previous runs of the turbulent MAIHEM simulation have shown that the presence of magnetic fields facilitates the development of an overall hotter CGM and the formation of dense structures within it, where magnetic pressure dominates. The magnetic fields also allow for the diffusion of angular momentum throughout the extended disk and gas cooling onto the disk, contributing to the maintenance of the disk at late times. Therefore, the presence of magnetic fields influences the overall structure and dynamics of the galaxy, particularly in the CGM and the disk. Given the complex relationship between magnetic fields, cooling rates, turbulence, and metallicity, runs of the high-metallicity were executed in and outside the presence of a magnetic field and the overall structure of these results was compared.

2.4   High-Metallicity Results
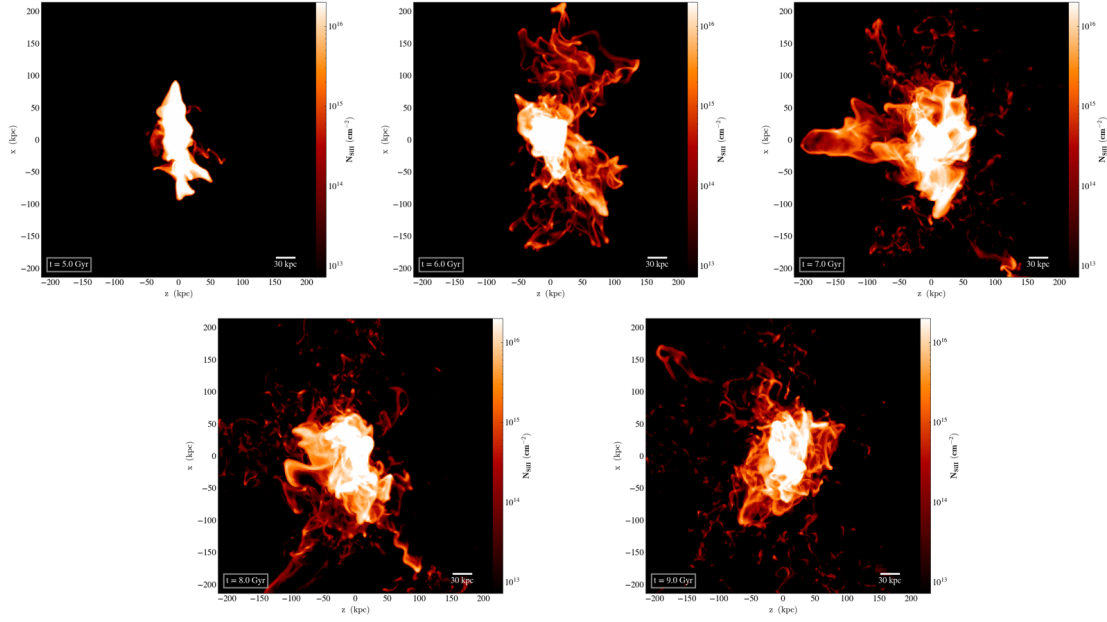
Figure 2.1: Plots showing the column densities of Si ii across 5Gyrs starting around 5Gyrs into formation and ending at almost 10 Gyrs.
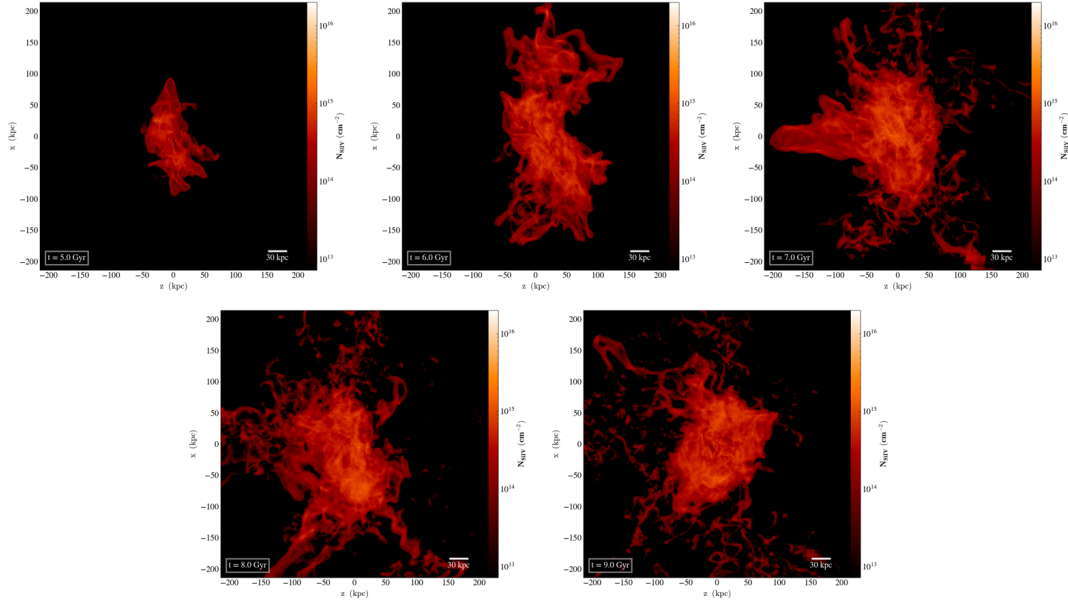
Figure 2.2: Plots showing the column densities of Si iv across 5Gyrs starting around 5Gyrs into formation and ending at almost 10 Gyrs.
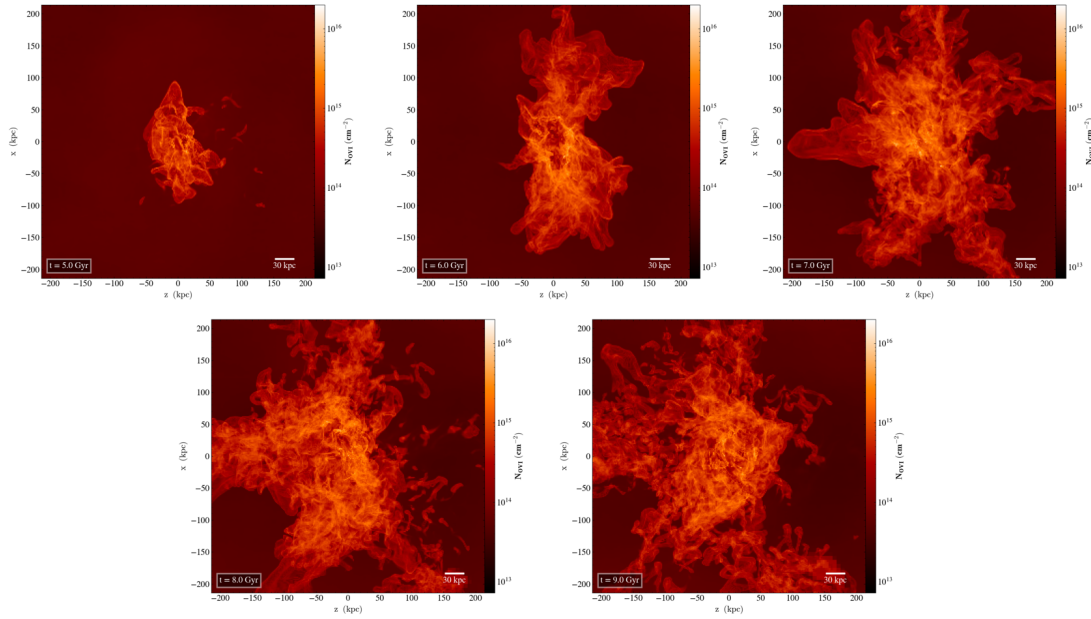
Figure 2.3: Plots showing the column densities of O vi across 5Gyrs starting around 5Gyrs into formation and ending at almost 10 Gyrs.
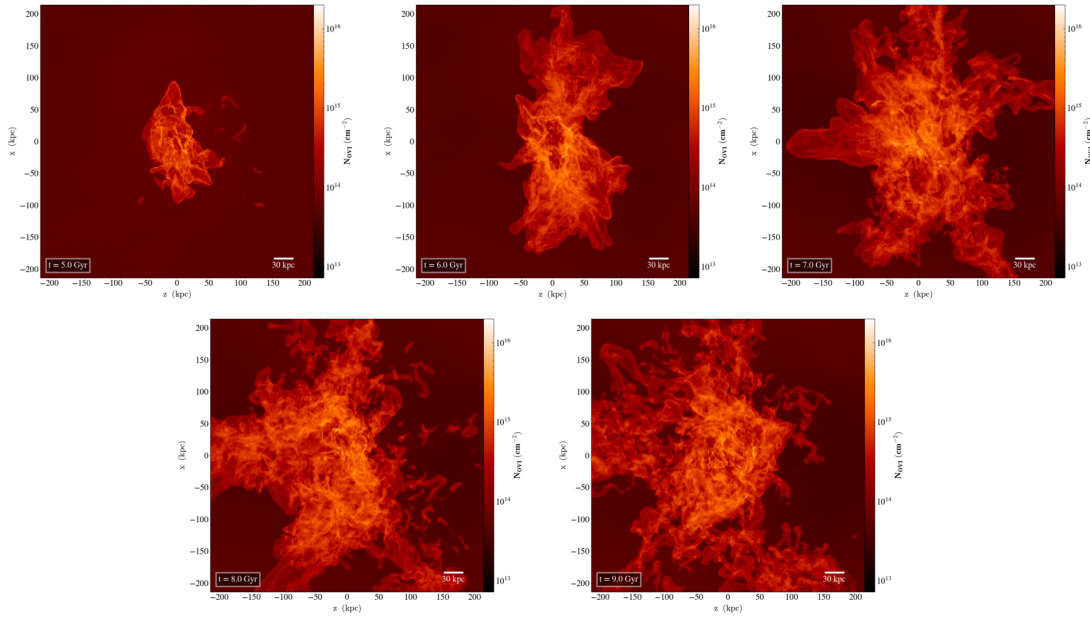
Figure 2.4: Plots showing the column densities of O vi across 5Gyrs starting around 5Gyrs into formation and ending at almost 10 Gyrs.

Chapter 3

TONE-TRACKING ALGORITHM DESIGN

3.1   Introduction

Over the past 25 years, low temperature detectors have made considerable waves in the astronomy community. With operating temperatures dipping below a single Kelvin, these cryogenically cooled devices allow for the detection of extremely faint photons with minimal noise and have become the near-ubiquitous standard for millimeter-wave and sub-millimeter wave astronomical instruments on single-dish telescopes. In general, super-conducting heterodyne arrays are used for spectroscopic surveys while lower resolution spectroscopy and photometric imaging are achieved with direct detection arrays. A common building block of these direct detection arrays are bolometers with superconducting thermometers known as Transition Edge Sensors (TES) which are then monitored and read out in a multiplexed fashion with current amplifiers called superconducting quantum interference devices (SQUIDS) 30. While TES bolometers have retained their position as a mainstay in millimeter and sub-millimeter wave astronomy, the past decade has seen the rise of a competing superconducting technology which no longer requires the need for SQUIDS. This cutting-edge detector dubbed the "kinetic inductance detector" (KID) has recently been flown on missions such as OLIMPO and BLAST-TNG and shows great promise for advancing ultra-high sensitivity observational astronomy. Unfortunately, the growing use of KIDs for astronomy has also illuminated certain problems with their implementation. Current readout techniques for KIDs greatly reduce their theoretical dynamic range; an issue that needs to be remedied if KIDs are to continue their widespread use.

3.1.1   What are KIDs?

As their name would imply, kinetic inductance detectors are highly-sensitive detectors which take advantage of the kinetic inductance properties of superconducting films. Understanding the mechanics of KIDs requires a cursory knowledge of inductance, resonators, superconductivity, and Cooper pairs. At the most simplistic level, a KID is comprised of a superconducting resonator that modulates its resonant frequency and quality factor when incident power in the form of photons is absorbed into the superconducting material. Readout of the KIDs is achieved by coupling the resonators to a microwave transmission line. By designing different KID pixels to have unique resonant frequencies, microwave frequency multiplexing is possible with relatively low effort compared to other superconducting detectors. 46

Kinetic Inductance

In kinematics, mass has the property known as inertia which is defined as the resistance to a change in motion. The electromagnetic analogue to inertia is inductance, which instead of resisting change in motion, resists change in current to a system as $v = L\frac{di}{dt}$.

In a system such as the superconducting films used for KIDS, there are two types of inductance; magnetic (geometric) and kinetic. Magnetic inductance is due to energy stored in the magnetic field which opposes change in current and is completely determined by the geometry of the conductor. Kinetic inductance on the other hand is due to the kinetic energy of charge carriers in a conductor. In a standard metal conductor, electrons cannot move completely freely due to vibrational modes of the conductor's lattice. Therefore, kinetic inductance is typically negligible compared to the magnetic inductance. In superconductors however, electron mobility is unmatched and kinetic inductance can become a dominating factor in the overall inductance.

27

Circuit Properties of a KID

A fundamental characteristic of KIDs is their design as resonant RLC circuits. Fig. 3.1 shows a lumped element model for a KID capacitively coupled to a microwave transmission line of impedance Z_0. This model consists of an RLC resonator whose inductance is
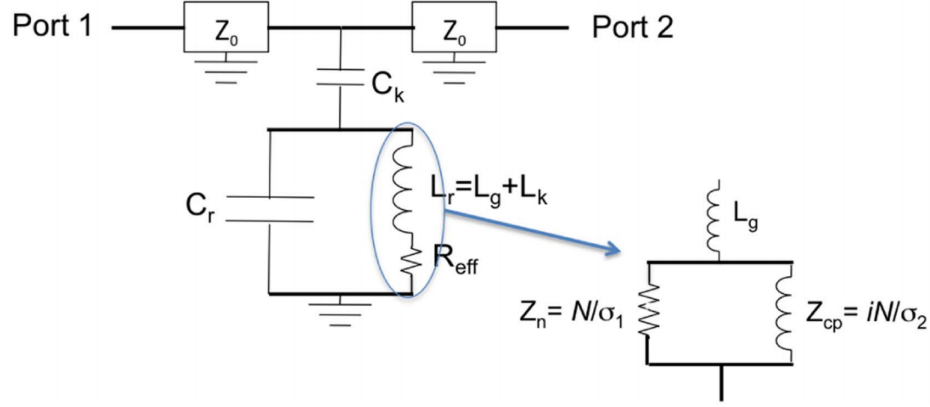
Figure 3.1: The lumped circuit model of a single KID resonator coupled to a transmission line. Reprinted from "Transition Edge Sensors and Kinetic Inductance Detectors in Astronomical Instruments," by P. D. Mauskopf, 2018, Publications of the Astronomical Society of the Pacific, 130:082001 (28pp).

the summation of the geometric inductance (magnetic field inductance) of the resonator's material and the kinetic inductance due to quasiparticles. The resonators capacitance is represented by C_r and its effective resistance is given as R_{eff}. The resonator is coupled to a microwave transmission line by a coupling capacitor with capacitance C_k. The overall resonant frequency of the lumped circuit is therefore described in Eq. 3.1 where L_T is the combined magnetic and kinetic inductance 30.

$$\omega_0 = 2\pi f_0 = 1/\sqrt{L_T(C_c + C_r)} \tag{3.1}$$

As stated before, KIDs are superconductors and their total inductance will be dominated by the kinetic component. This formula for resonant frequency (Eq. 3.1) is little more than

the fundamental property of LC resonators which states that resonant frequency $\omega_0 \propto \frac{1}{\sqrt{LC}}$.
Viewing the detector as an RLC resonator coupled to a microwave transmission line, it is straightforward that KIDs must also share the transmission parameters of a resonant tank circuit. Fig. 3.2 shows both the S_{21} and phase delay of a signal ringing down a transmission line capacitively coupled to a KID as seen in Fig. 3.1. Like an LC tank circuit, a frequency sweep passes through unaffected until it reaches the resonant frequency of the KID at which point the signal is grounded and goes mostly unseen at port 2. Furthermore, the phase plot shows a logarithmic phase shift that approaches zero as the signal moves towards resonant frequency 31.

Figure 3.2: The transmission parameters of an ideal KID resonator centered about a resonant frequency f_0. The top plot shows the S_{21} of a signal traveling through the coupled transmission line and the bottom shows the phase shift. Reprinted from "Microwave Kinetic Inductance Detectors," by Benjamin A. Mazin, 2004, thesis defense, California Institute of Technology Pasadena, California.

Superconductivity

The last essential piece to understanding the function of KIDs is how superconductivity is used of for the purposes of detecting incoming photons. Superconductivity in relation to resonators is an astoundingly complex topic for which the physics are given a detailed

treatment in a review by Zmuidzinas 46. Thankfully for the purposes of understanding the fundamentals of KIDs, a much more elementary understanding will suffice. In a standard conductor, a lattice of positively charged nuclei form according to the properties of the material. This lattice is analogous to a series of balls on springs and therefore undergoes vibrational modes from energy in the system known as "phonons".

Typically, a material above its critical temperature will have significant vibrations in its lattice but if it is a conductor, electrons are mostly free to move about. When a voltage is applied, electrons will attempt to flow in the direction of the electric field. However, since the lattice is vibrating, the electrons will bounce off of the positive charges and will not have a straight path from one side to the other. They will also transfer energy into the lattice during collisions, heating the metal and establishing the property known as "resistance." In a superconductor however, the lattice vibrations are so small that electrons can move with the electric field and largely avoid any collisions, hence why there is zero DC resistance.

While the electrons in a superconductor may not "collide" with the positive charges, they do attract them as they pass by which removes energy from the electron and creates a phonon in the lattice. The presence of a phonon causes a region of positive charge as the nuclei are more densely packed then their equilibrium state. Since electrons are able to freely move within the lattice, the over-density of positive ions attracts a second electron which becomes bound to the first through this electron-phonon interaction and the resulting pair of electrons is known as a "Cooper Pair". In terms of energy levels, the electrons in a Cooper Pair are located in the superconducting region where a band-gap of energy determined by the properties of the cooled material separate them from the resistive conduction region.

The diagram in Fig. 3.3 shows the energy diagram of a Cooper pair of electrons and their interaction with incident photons. Under isolated conditions, the energy band-gap prevents the electrons from exiting the superconducting region. However, in the case where an in-
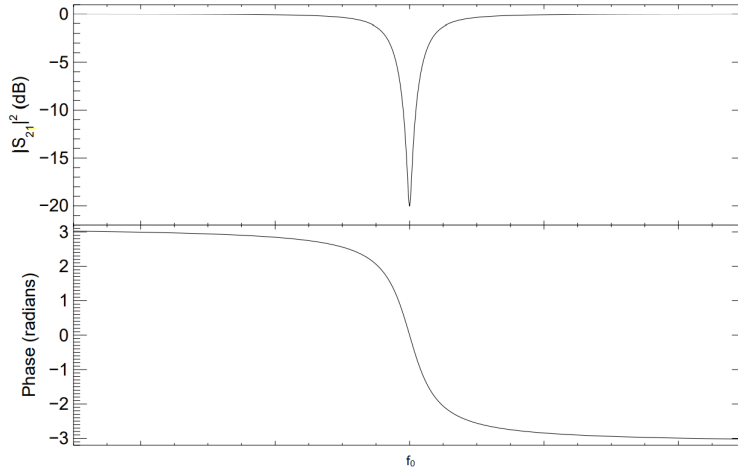
30
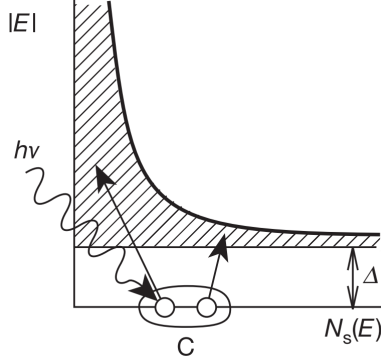
Figure 3.3: The transmission parameters of an ideal KID resonator centered about a resonant frequency f_0. The top plot shows the S_{21} of a signal traveling through the coupled transmission line and the bottom shows the phase shift. Reprinted from "Microwave Kinetic Inductance Detectors," by Benjamin A. Mazin, 2004, thesis defense, California Institute of Technology Pasadena, California.

cident photon with sufficient energy interacts with a critically-cooled KID, the Cooper pair can be broken and the electrons jump the band-gap into the resistive conduction region. The effective electric mass of electrons below the band-gap are considerably smaller than those above and thus the process of breaking Cooper pairs has the effect of increasing the kinetic inductance of the KID. Since KIDs are resonant circuits with inductance dominated by the kinetic component, the result is that the resonant frequency of the KID is modulated downward. In short, incident power on a cooled KID causes a shift in resonant frequency and this complicated process is what allows KIDs to be used as extremely sensitive photo-detectors. The effect of incident power on the S_{21} of a KID coupled to a microwave transmission line can be seen in Fig. 3.4 where the blue curve shows the original resonator and the shifted orange line shows the transmission after modulation.

3.1.2 The KID Dynamic Range Conundrum

In order to understand the dynamic range issue surrounding KIDs, one must first understand KID readout. Starting from a single KID coupled to a microwave transmission line as shown in Fig. 3.1, a signal sent through the line will have transmission parameters
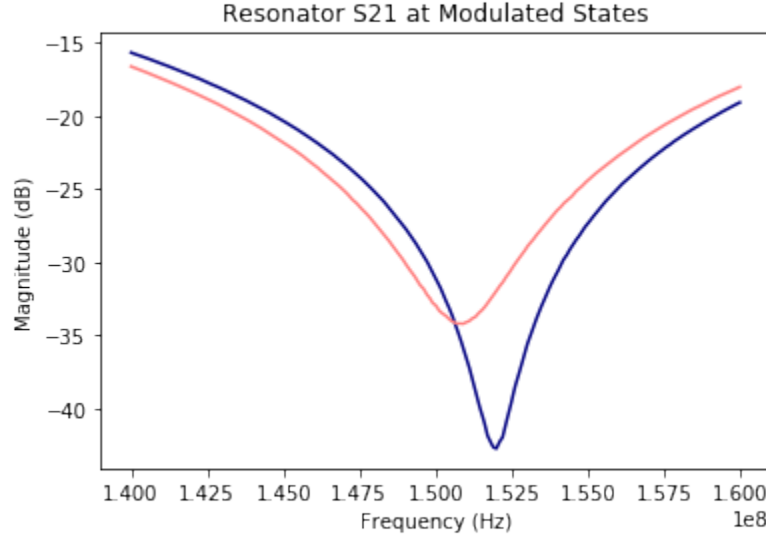
Figure 3.4: S_{21} of a 152.5 MHz resonator in an unmodulated state (blue) and modulated state (orange) due to a change in incident power. This effect is the basis for how KIDs detect faint photons.

similar to Fig. 3.2. In order to determine the incident power on the detector, a probe tone at the original resonance of the KID is inserted at port 1 and monitored at port 2. As photons interact with the KID, the transmission parameters will modulate and the amplitude seen at port 2 will change (shown in Fig. 3.5) . Assuming the relationship between incident power and change in transmission is known, the amount of power on any KID can be determined. A similar effect is seen in phase space (Fig. 3.6) and can likewise be used to relate modulation to incident power. While the example shown is of a single KID, tuning an array of KIDs to different frequencies allows for frequency multiplexing as long as a probe tone for each resonance is sent down the coupled transmission line and demodulated upon finishing.

For small amounts of incident power, monitoring a probe tone in both amplitude and phase is a perfectly acceptable means of reading out KIDs. When there are large changes to the incident power on a KID, problems begin to arise. As the intensity of light interacting with a KID increases, the change in amplitude seen at port 2 begins to shrink rapidly. Once a KID is saturated with enough power, the change in transmission is small enough that it is
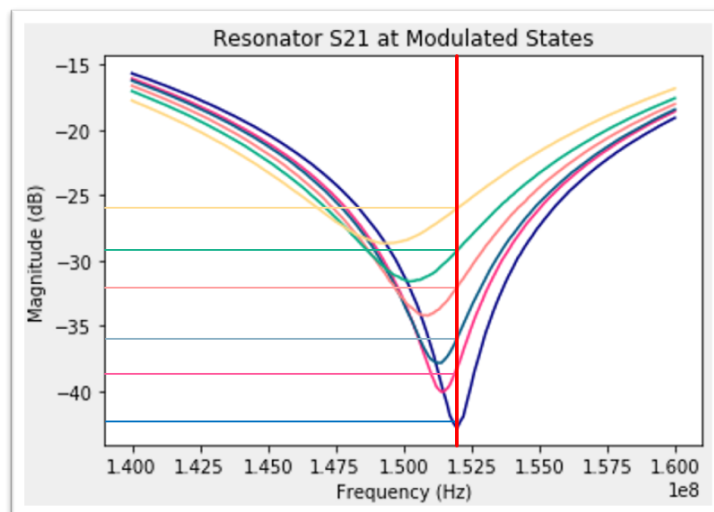
Figure 3.5: A single resonator in multiple states of modulation probed with a tone at the rest resonance. As incident power modulates a KID, the change can be seen though the variation in probe tone transmission.
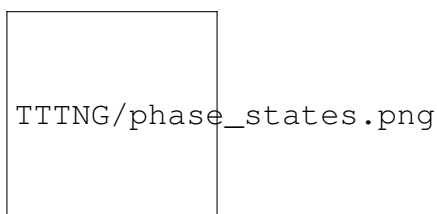
Figure 3.6: The phase delay of a KID isolated from incident power (blue) and modulated by incoming photons (orange) with a probe tone from port 1.

lost in system noise, mainly due to amplifiers. Fig. 3.7 shows an exaggerated example of this difference in visibility. Attempting to read out useful information from a KID which is dominated by amplifier noise is a fruitless effort and the detector is dead weight until the issue is corrected. At this time, the only remedy for this problem is a time-consuming VNA sweep which occurs mid-flight and wastes precious observation time. Losing amplitude information to amplifier noise is not uncommon during flights and on the recent BLAST-TNG mission this problem occurred every few degrees when scanning the sky.

The next NASA mission planning to use MKIDs as a means of deep-space detection is the Terahertz Intensity Mapper (TIM) balloon-borne telescope, lead by a group at the
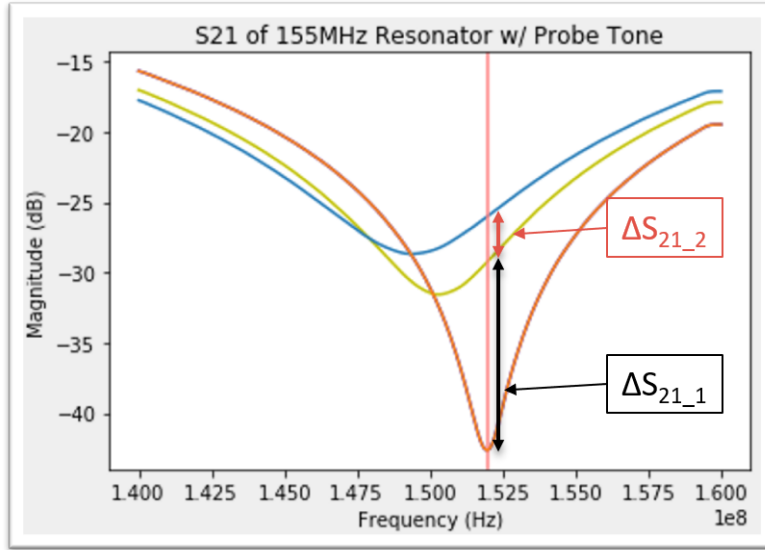
Figure 3.7: An exaggerated example depicting the diminishing returns of incident power on a KID. Increasing power returns smaller and smaller changes in amplitude until the difference between two modulated states is dominated by amplifier noise. As of now, the only correction possible is a time-intensive mid-flight VNA sweep.

University of Illinois in collaboration with a number of other research institutions. TIM aims to survey CII, OI, and OIII in 100 galaxies to demystify many aspects of star formation while simultaneously producing deep tomographic maps of the 3-D structure of the universe. The architecture, spectrometer, and readout system of TIM is modeled from the BLAST-TNG mission that recently flew at McMurdo Base in 2019. Among the lessons learned from BLAST-TNG was the need for a more efficient method to overcome the KID dynamic range issue 3.

3.2 The Tone-Tracking Algorithm

3.2.1 Firmware development

The solution to the restricted dynamic range of KIDs encountered in previous missions is deceptively simple; design a firmware to constantly track the resonance of the KIDs as they modulate. By always knowing where each of the resonant frequencies in a KID array

34

is located, the need for mid-flight VNA sweeps is eliminated without sacrificing valuable phase and transmission information. Rather than having a static probe tone at a starting resonance that is reset once information is lost to noise, this "tone-tracking" firmware will constantly update the probe tone to the current resonance of each KID in an array and use feedback information to calculate incident power as before. This puzzle presents itself like many other control system problems and is tackled using a number of digital control techniques.

The first and most important element to a control system is finding an "invariant" or a system characteristic that does not change with respect to the independent variable. In this case, the independent variable is the incident power on the KIDs. Invariant properties allow for a reference which is essential for any control system. Resonators have such a property due to their electrical length at resonance being zero. As seen in the phase plot of Fig. 3.2, resonators have zero phase shift for signals directly on resonance. This means that regardless of incident power on a KID, the phase of a tone at resonance will never change (ignoring linear phase shifts from signal lines connected to the system). The plot in Fig. 3.8 shows the invariance of resonant phase w.r.t. different incident power levels. The black line which stays constant regardless of resonant frequency shows the resonant phase for each of the modulations.

Determining an invariant for a tone-tracking system is a vital first step, but the goal of the algorithm is to move a probe tone in frequency space, not phase. Fortunately there is a linear regime in the relationship between phase and frequency of a resonator that allows for transformations between phase and frequency information. By calculating the slope of this linear portion, shifts in phase detected by the control system can be converted to shifts in frequency of the probe tone output. This important area is highlighted in Fig. 3.9.

The plots in Fig. 3.9 show the linear regime of the same resonator in a modulated and unmodulated state. Two concerns naturally arise from this figure. First, different modula-

35
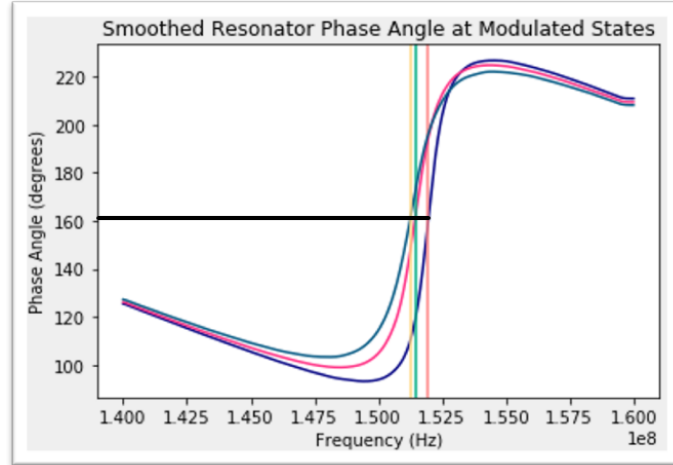
Figure 3.8: A phase plot of a resonator in different states of modulation superimposed on resonant probe tones for each state. As seen by the black line, the phase shift does not change between modulations (note that this phase plot is only of the resonator itself and no cabling which would cause a linear phase change). Due to this consistency w.r.t. incident power, resonant phase acts as an effective invariant for a tone-tracking control system.
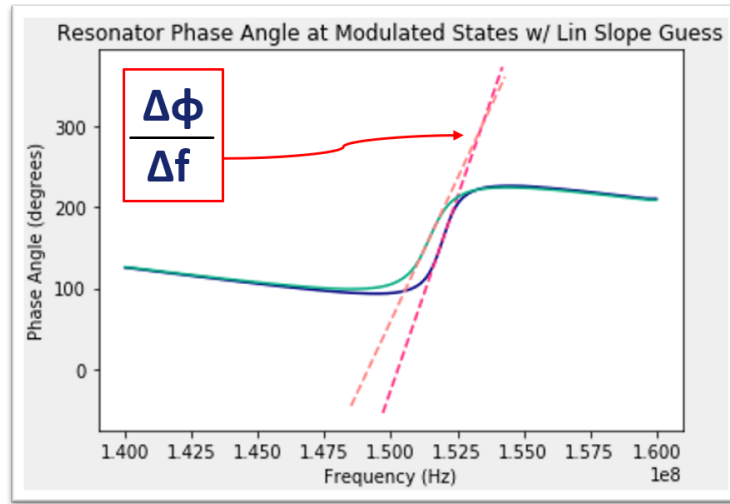
Figure 3.9: A plot highlighting the linear regime in the phase/frequency relationship of a resonator in both the rest state and a modulated state. Previous applications of KIDs have confirmed that resonant frequency will not stray from this linear portion. While modulation of the KID causes changes in slope, the tone tracking system is designed in a way where these differences can be ignored.

tions have differing slopes. To account for this, the tone-tracking firmware uses incremental changes so that as long as the slope used is a good general estimate, the variable nature of

the linear regime can be ignored. This feature will be examined in the full breakdown of the algorithm. Second, using a linear relationship to relate phase and frequency relies on resonance never moving outside of the linear regime. Due to the low incident power on KIDs during balloon-borne missions, the resonant frequency is never expected to shift more than a line width as confirmed in the BLAST-TNG mission. Therefore, it is safe to use the slope of the phase/frequency linear regime as a means of converting from phase space to frequency space. If more accuracy is needed a non-linear fit to the resonance can be used.

With the tools of an invariant and a conversion factor in hand, it is now possible to design an algorithm to track the resonant tone of a KID. Before observing, one can determine the original resonant frequency of the KID (f_{res0}), the resonant phase (ϕ_0) and the linear regime slope ($\frac{\Delta\phi}{\Delta f}$). Once power falls on a KID, there will be a change in phase of ($\Delta\phi$). Putting these together, it only takes a simple set of equations to reach our final goal of the new resonant phase (f_{resnew}). These equations are shown in Eq. 3.2 and Eq. 3.3.

$$\frac{\Delta f}{\Delta\phi} * \Delta\phi = \Delta f \tag{3.2}$$

$$f_{res0} + \Delta f = f_{resnew} \tag{3.3}$$

Fig. 3.10 provides a graphical representation of how the tone tracking algorithm uses known data and relationships to shift the probe tone to the modulated resonance of the KID. In digital control terms, the tone-tracking algorithm is a combination of a phase-locked loop (PLL) feedback-controlled with a proportional/integrative (PI) controller. The next section will break down these aspects into more detail.
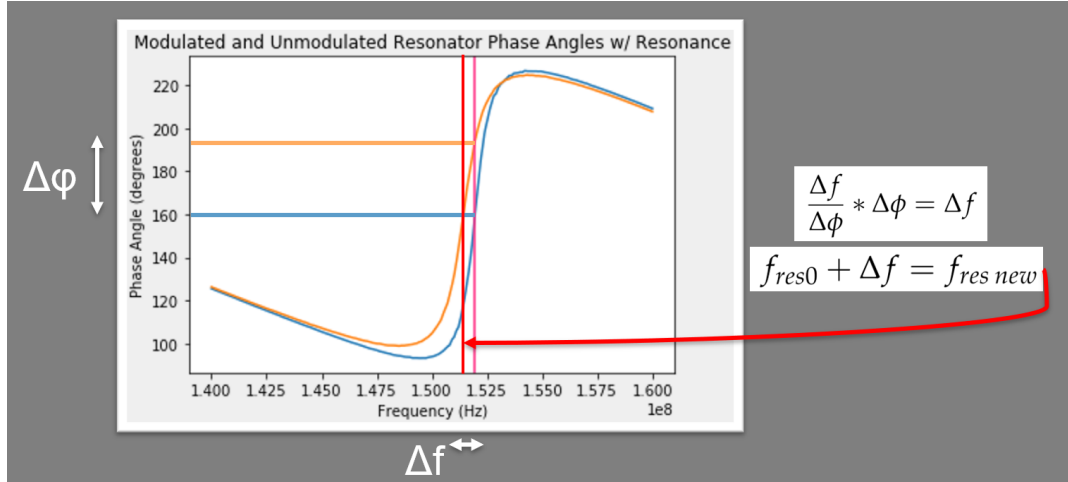
Figure 3.10: A graphical representation of the tone tracking algorithm.

3.3 Simulations and Results

Creating a digital control system to tone-track resonant frequencies of KIDs is an exciting step in maximizing the potential dynamic range of detectors but means little unless it is executed on hardware. The upcoming TIM mission aims to use an array of 2048 MKID (microwave kinetic inductance detector) pixels simultaneously through frequency multiplexing. Due to the enormous data rate requirements involved with such a high number of detectors, a field programmable gate array (FPGA) is the correct choice of hardware. FPGAs are logical processors which are preferred to CPUs for astronomy-related readout due to their highly re-configurable logic and unmatched bandwidth capacity. The downside to FPGAs compared to standard computers is that they require firmware modules to run which are significantly more difficult to develop than computer programs.

Fig. 3.11 shows the block diagram of the tone tracking algorithm for a single KID according to the firmware design. Starting with a probe tone at the KID's original resonant frequency, the signal is outputted by the digital to analog converter (DAC) and travels through the transmission line coupled to the KID. The affected tone then re-enters the FPGA by a digitizer (ADC) in quadrature and is multiplied with a complex conjugate of
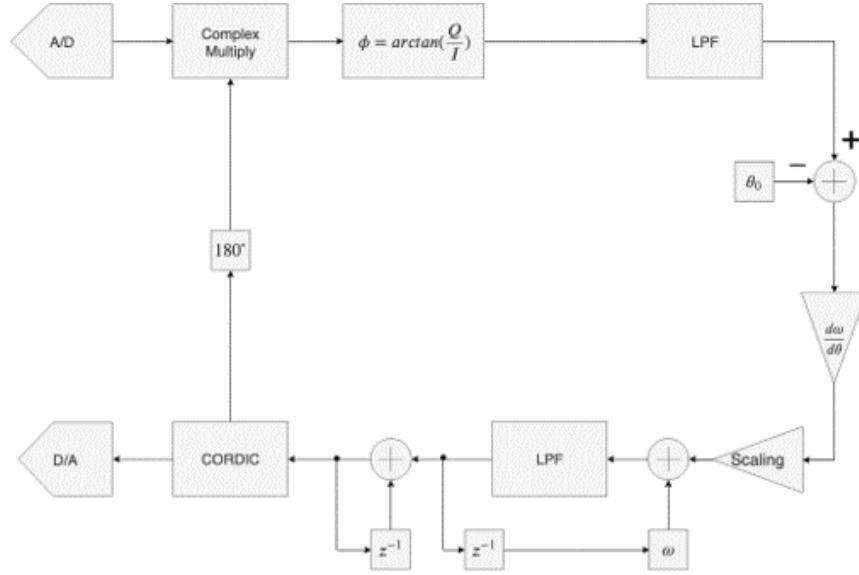
Figure 3.11: A block diagram of the single-channel tone-tracking algorithm firmware. The diagram can be interpreted as follows: 1. Probe tone leaves DAC and goes through observing KID 2. Resulting tone returns to the ADC and is mixed to DC 3. Phase of input is calculated and integrated 4. Differential is taken to find $\Delta\phi$ 5. $\Delta\phi$ is scaled by estimated linear slope ratio 6. Differential frequency is added to original probe frequency 7. Probe tone is made with a CORDIC design and enters DAC 8. Return to step one and repeat

the probe tone to mix down to DC. The phase of the DC signal is calculated and fed through a filter in order to remove small fluctuations. A differential is taken between the new phase and the resonant phase to find the keystone value of $\Delta\phi$. This differential is multiplied by the inverse slope of the estimated phase/frequency ratio to convert it to a differential frequency. Since there is lag time due to the length of the tone-tracker as well as uncertainty in the accuracy of the linear regime slope, this frequency difference (Δf) is scaled down considerably as to not "overshoot" the goal. The scaled down (Δf) is then added incrementally over multiple loops through the firmware to the original probe tone and inserted into a CORDIC wave generator to make a new probe tone closer to resonance. Finally, the probe tone exits the DAC and repeats the cycle until resonance is found and a phase differential of zero is achieved.

The System Generator function of Matlab's Simulink is the stepping stone between the

39
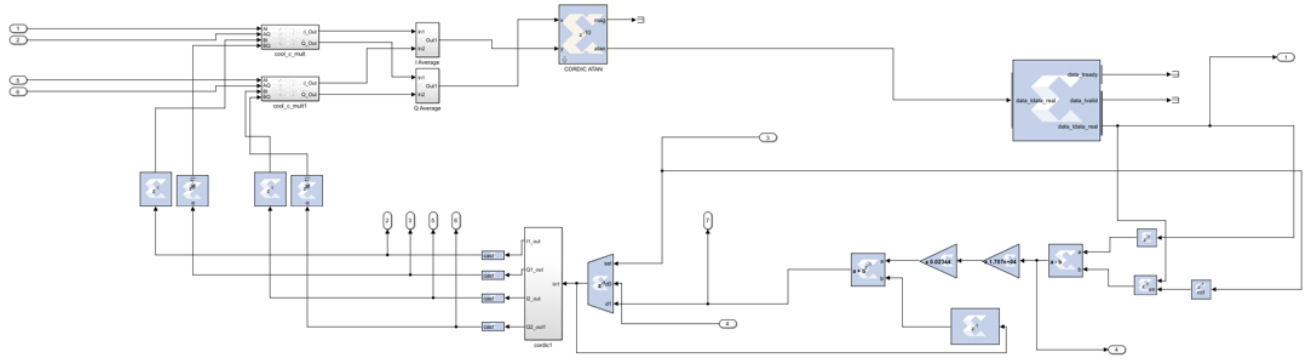
Figure 3.12: Simulink design of the tone-tracking algorithm. Since there is a factor of two difference between the FPGA fabric clock and the digitizer, a parallel design is required as two samples will be digitized for every clock cycle on the FPGA. Similar to the previous block diagram, the schematic can be read as follows: 1. Parallel probe tones leave the CORDIC block and are modified by the simulated resonator 2. The I and Q of the resulting pair of signals are averaged together 3. Phase of input is calculated and filtered 4. Differential is taken between current phase and resonant phase to find $\Delta\phi$ 5. $\Delta\phi$ is multiplied by estimated linear slope ratio and scaled down 6. Differential frequency is added to original probe frequency 7. New probe tone is generated by CORDIC block 8. Return to step one and repeat

block diagram and usable firmware as it allows for graphical DSP design and converts the designs into a hardware defined language (HDL) compatible with FPGAs. Fig. 3.12 Shows the design for a single-channel tone-tracking firmware that can be converted into HDL and uploaded to an FPGA. While similar to the block diagram, the digitizer operates at twice the clock speed of the FPGA fabric and therefore two samples will be fed for every clock cycle on the FPGA. To account for this, a parallel design was created so that the tone-tracker can handle two samples at once.

3.3.1   Simulation Results

With the tone-tracker constructed in the world of Simulink, simulations can finally be underway to test the validity of the algorithm. To accomplish this, a sufficiently analogous "digital resonator" was created and fed into the ports of the tone-tracker. The first test was

40

simply giving the resonator a modulation and watching how long it took for the algorithm to reach zero phase differential as seen in Fig. 3.13. Two modulations were done; an unrealistically large one for fun and a smaller more realistic blip. In the larger case, the algorithm found resonance on the order of 10 microsecond with a small overshoot. The second resonance shift was located in less than half the time with no oscillations.

For a more realistic scenario, white noise was added to the system and another set of tests was run. Fig. 3.14 shows two simulated resonance shifts with an added white noise component. The tone tracking algorithm located the resonance in both cases on millisecond timescales with a negative-exponential-shaped curve.

The final simulation result of note is the waveform generated during a resonance shift. The timestream in Fig. 3.14 is the output of the CORDIC wave generator inside the tone-tracking firmware. Unlike lookup tables (LUTs) typically used for MKID readout which create waves by reading values from registers, the CORDIC design uses trigonometric identities to move around the unit circle and produce a sine wave. This difference allows the CORDIC to save considerable memory as it only needs a phase input to output a waveform. Due to the unknown frequencies needed for tone-tracking, the CORDIC design working successfully will greatly reduce the resource requirement of the firmware.

3.4 Discussion

3.4.1 Examining Simulation Results

The results from tone-tracking firmware simulations show that the algorithm is a valid means of relocating KID resonances mid-observation. The ideal noiseless tests shown in Fig. 3.13 verify that the control system is capable of rapidly reducing phase differential to zero. The large resonance shift that occurs in the first half of the noiseless simulation does show a slight oscillatory behavior due to overshooting the target frequency. However,

41
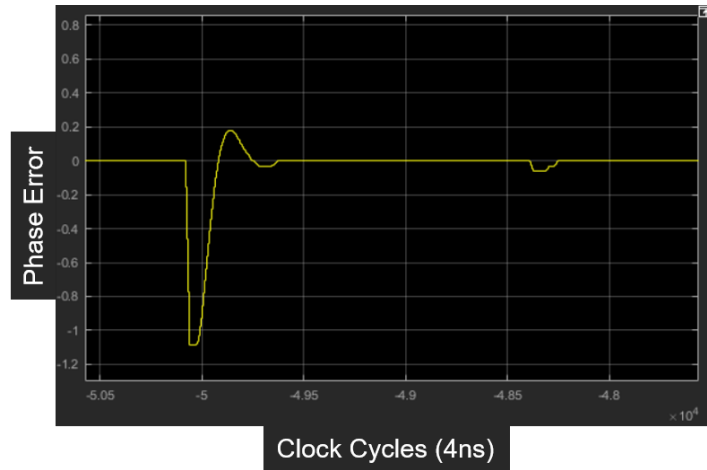
Figure 3.13: Noiseless simulation results of tone-tracking algorithm measuring the difference between the real-time phase and resonant phase. Zero means the tone is tracked. The first spike represents a large phase shift and a dampened oscillation is visible. The second bump is a much smaller and more realistic phase shift.
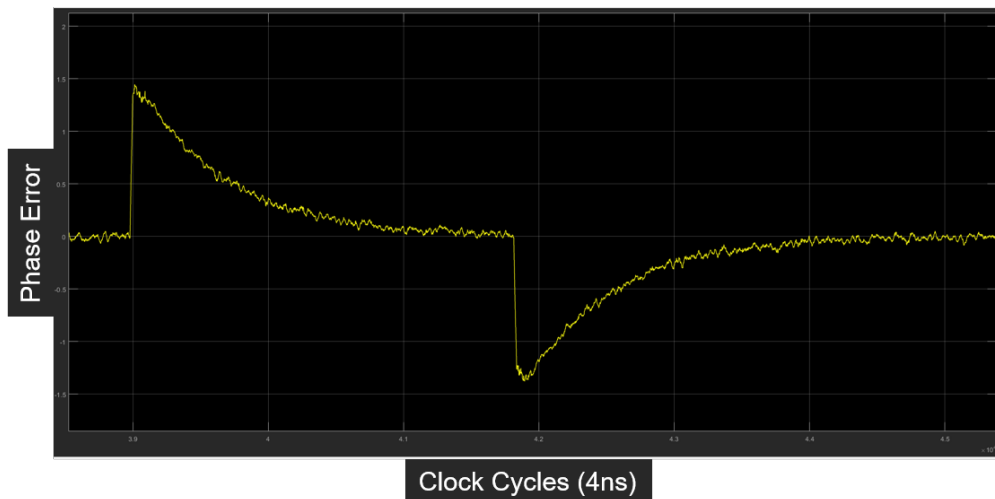
Figure 3.14: Noisy simulation showing two simulated resonance shifts both finding the resonance on millisecond timescales.
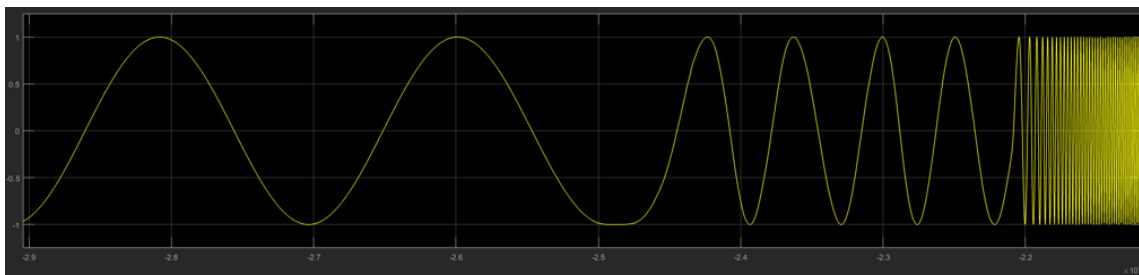
Figure 3.15: A timestream of the waveform exiting the CORDIC wave generator as a resonance shift increases the frequency.

this feature is normal for PI controllers. Tweaking the magnitude of the integrative portion of the algorithm could help reduce this effect, but a phase shift of such large magnitude will not be seen during TIM observations and the smaller more realistic shift does not experience this characteristic.

The simulation with white noise in Fig. 3.14 is much closer to the behavior one would expect to see during deployment. Though the resonance shifts used in this test are also more pronounced than what will be seen on TIM, all of the phase error plots seen throughout the white-noise tests share a similar shape and time-scale regardless of modulation magnitude. The negative exponential shape of the phase differential implies that noise will significantly increase the timescale of relocating resonances. However, the shape of the plot also shows that the integrative portion of the controller is effective in eliminating the overshooting seen in the noiseless simulations. In its current state, the noisy tone-tracker was able to relocate resonances on millisecond timescales. Detector sampling rates from BLAST-TNG and other projects are on a similar millisecond timescale, confirming that the current tone-tracking algorithm is just fast enough to be effective in similar observation scenarios. The algorithm uses a scaling reduction to slow down the modulation of the output probe tone in order to avoid overshooting or runaway instabilities. If future tests show that the timescale to relocate resonance is too slow for TIM, an effective fix would be to tweak the scaling reduction factor and increase the allowable frequency shift per loop of the algorithm.

43

Finally, the output of the CORDIC wave generator in Fig. 3.15 shows great promise for reducing resource requirements in comparison to the LUTs used for BLAST-TNG. One of the more difficult requirements of a viable MKID tone-tracker is the need for sub-kilohertz resolution on the output of the wave generator. When designing with LUTs, frequency resolution can only be improved by increasing the length of the tables, in turn requiring more memory usage for the wave generator. CORDIC generators on the other hand have a frequency resolution determined by the number of bits in their accumulators. This resolution, which is equal to the minimum producible frequency from the CORDIC, is mathematically defined by Eq. 3.4 where N is the number of bits in the CORDIC accumulator.

$$f_{min} = f_{resolution} = f_{clk}/(2^{N-1}) \tag{3.4}$$

Therefore, in order to meet the goal of a sub-kilohertz f_{min}, the CORDIC generator requires an accumulator with N=20 bits, producing a frequency resolution of 477Hz. This means that while the CORDIC design is effective at reducing resource consumption as compared to LUTs, the bit growth from a 20-bit accumulator will need to be counteracted if the design is to be compatible with surrounding readout firmware. As usual, there is no free lunch. Furthermore, in order to assure that the readout noise is not dominated by the CORDIC, a frequency resolution as low as a single Hertz is required. Once tone-tracking is tested and deemed ready for deployment on missions, the CORDIC accumulator will likely need to be raised to N=30 bits to achieve Hz resolution. Since this requirement only comes into play once tone-tracking is placed on TIM, it is likely that increasing the CORDIC to this size will not be attempted until late in the development process.

3.4.2 Hardware Woes

While a great deal of time has been spent developing, designing, and simulating this tone-tracking algorithm, it is apparent from the results that there is still a significant amount of work to be done before the firmware is ready for deployment. The largest hurdle so far has been an inability to use proper hardware. The tone-tracking firmware is quite resource intensive even for a single-channel design and scaling up to 2048 resonating pixels for TIM will require a cutting-edge FPGA. Fortunately, we have recently come into possession of the new Xilinx ZCU111 RF System-on-Chip (RFSoC) FPGA which has the power to support the tone-tracking firmware. The RFSoC contains 8x 12bit-DACs and 14bit-ADCs running at 4GSPS integrated directly on the chip as well as an arm microprocessor capable of running Linux and Python and an Ethernet connection that can handle up to 1Gb/s bandwidth.

The Xilinx RFSoC is not only desirable for tone-tracking firmware, but for the entire MKID readout system as well. Readout for BLAST-TNG was handled by a stack of four ROACH2 boards which have been used as the standard astronomical FPGA for over a decade now. Due to the increased speed and bandwidth, a single Xilinx RFSoC can handle the workload of all four ROACH2 boards used in BLAST-TNG. Therefore, a goal of the TIM mission is to create a general architecture for the Xilinx RFSoC which will allow us to port over the BLAST-TNG firmware with the aim of making TIM the first balloon-borne mission to fly with the new board.

As to be expected, the impressive power of the Xilinx RFSoC comes with an equally impressive price tag. For the past year, our lab was lucky enough to be in possession of a single Xilinx RFSoC for developing general readout architecture, porting the BLAST-TNG firmware, and designing the tone-tracking firmware. The recent COVID-19 pandemic has made shared use of the board a challenge and forced a considerable slow-down in the

realization of tone-tracking firmware. Fortunately over the last month, an increase in supply of Xilinx RFSoC boards has allowed for the acquisition of more units and tone-tracking development has once again resumed. With renewed access to necessary hardware, there is now a clear path to the deployment of a tone-tracking firmware.

3.4.3   Hurdles to Overcome

Before the COVID-19 pandemic halted development, a number of challenges arose that must now be solved before lab testing of tone-tracking can begin. The first of these is estimating the resource allocation of the tone-tracking firmware. In a perfect world, tone-tracking would be uploaded to the same Xilinx RFSoC as the rest of the readout firmware needed for TIM. Resource monitoring for the single channel tone-tracking design suggests that this ideal scenario is likely impossible due to the resource requirements of the current tone-tracking algorithm. Whether or not tone-tracking capabilities will need a dedicated RFSoC board must be determined well before deployment as it will impact power, weight, and monetary requirements.

A more straightforward complication lies in the fact that the tone-tracking algorithm is designed around resonant phase invariance with respect to incident power. This assumption is true when looking at KIDs as isolated resonators, but in practice, signals will be traveling through cabling and transmission lines as they enter and exit the MKID array. These paths outside of the resonators will have frequency dependent phase delays that must be accounted for inside the tone-tracking firmware. While the phase effects of the transmission are linear and wholly predictable, the lengths of all cables and transmission lines must be known precisely in order to be corrected. Accounting for this effect will be a top priority once the tone-tracking firmware is successfully uploaded and operating on a Xilinx RFSoC.

The final and most devious roadblock so far in tone-tracking development is the manipulation of bit-widths throughout the firmware. Due to the number of complex multipliers

and arithmetic functions found in the algorithm, bit growth is rampant and when unrestricted can produce CORDIC outputs with bit-widths nearing the hundreds. Since the tone-tracker will be used in tandem with the rest of the general readout architecture being developed, input and output bit-widths must be compatible. The task of suppressing bit growth is not simple; too many bits will lead to resource gluttony and incompatibilities while too few bits will lose precision and greatly increase bit-noise. Once the RFSoC general architecture is complete, great care will be taken to balance all of these competing requirements.

3.5   Conclusion and Next Steps

The ability to constantly track the resonant frequency of MKIDs in-situ has been desired for quite awhile and the complexity of the problem speaks to why it still does not exist. Significant progress has been made over the past year in developing an FPGA-based tone-tracking system, but there is a great deal more to accomplish before it can be used in future MKID observations. The first step is the completion of the general readout architecture for the Xilinx RFSoC. While this architecture has already interfaced successfully with the ADC and DAC units built into the board, there remains work to be done on accessing and communicating with the on-board DDR4 memory and direct-memory-access blocks, both vital components needed for tone-tracking. Completing the general architecture will also finalize the input and output bit-width requirements for the tone-tracker and will greatly simplify installing other firmwares onto the Xilinx RFSoC.

Since bit growth seems to be a considerable issue with the tone-tracking firmware, a Python simulation imitating the algorithm is under development that will focus on where bit-widths can be reduced. Developing a testing environment in python will greatly accelerate the tweaking process as making even small changes to firmware modules is an arduous and time-consuming testing procedure.

47

Furthermore, the development of firmware which can read and live-update the tone-tracker with the slope of the linear phase/frequency regime for each KID would help reduce the time it takes for the algorithm to locate resonances. Although using a general estimate for the slope as done in the current firmware effectively completes the goal of tone-tracking, knowing a precise conversion value will allow for a less aggressive scaling reduction and in turn allow the firmware to locate resonances in fewer cycles.

Once these steps are taken, a single-channel version of the tone-tracking algorithm can be uploaded to the Xilinx RFSoC and tested first using room temperature resonators and eventually cryogenically cooled MKIDs. A successful single-channel test will provide information on resource utilization that will direct how to best scale to a 2048-channel, frequency multiplexed design for use on TIM and other future MKID-based observation missions. Finally, all designs, software, and firmware modules will be uploaded to GitHub and made open-source for others to use and modify so that MKIDs can continue their rising popularity in the astronomy community.

Chapter 4

DEVELOPMENT OF A SPECTROSCOPIC LOCK-IN FIRMWARE (SPLIF) FOR THE READOUT OF QUANTUM-LIMITED SENSITIVITY DETECTORS

4.1 Detection Noise: The Pernicious Pink Power Problem

1/f noise, also known as flicker noise or pink noise, is a ubiquitous phenomenon in various physical systems, including electronic circuits, biological systems, and astrophysical detectors. Its unique frequency-dependent power spectral density $S(f) \propto 1/f^\alpha$ poses significant challenges for detection systems. This section aims to elucidate why 1/f noise is particularly difficult to remove compared to other types of noise, such as white noise or thermal noise.

An effective method of measuring the differences in noise sources is by looking at their spectra; in partucular, the squared spectral characteristics. By squaring the values that comprise the spectra, much like statistical variance versus deviation, there is no longer a need to consider any negative values, as they will all be converted to positives. In astronomy and wave physics, the squared Fourier transform of a given signal is known as the power spectral density (PSD). When discussing noise, the PSD provides the greatest insight into the unique nuances between different noise sources, as well as the techniques that can be used to mitigate such noise. The PSD of 1/f noise is given by:

$$S(f) = \frac{A}{f^\alpha} \tag{4.1}$$

where A is a constant and α is usually close to 1. This PSD indicates that the noise power is inversely proportional to the frequency, making it problematic at low frequencies

(Kaulakys and Meskauskas, 1998).

Mathematically explored, pink noise shows a unique characteristic to other common noise types that encompasses this concept of low-frequency domination. Interestingly, the mathematical analysis of pink noise PSD analysis shows that this signal dominance actually occurs on both ends of the frequency spectrum. Characterizing different colors of noise by their behavior at spectral extremes helps elucidate how and why this seemingly paradoxical relationship exists. The following examination of pink noise and other common types of intrinsic noise through a lens on PSD mathematics will give a more cogent explanation of why pink noise is such a pest to astronomers.

1) **1/f Noise (Pink Noise)**

The PSD of 1/f noise is given by:

$$S(f) = \frac{A}{f^\alpha} \tag{4.2}$$

To check the finiteness of the integral at both ends of the spectrum, we evaluate:

1a) Low-frequency end:

$$\int_0^{f_1} \frac{A}{f^\alpha} \, df \tag{4.3}$$

This integral is finite for $\alpha < 1$.

1b) High-frequency end:

$$\int_{f_2}^{\infty} \frac{A}{f^\alpha} \, df \tag{4.4}$$

This integral is finite for $\alpha > 1$.

For typical 1/f noise, α is close to 1, making the integral finite at both ends.

2) **White Noise**

The PSD of white noise is constant:

$$S(f) = A \tag{4.5}$$

2a) Low-frequency end:

$$\int_0^{f_1} A\, df = A f_1 \tag{4.6}$$

This is finite.

2b) High-frequency end:

$$\int_{f_2}^{\infty} A\, df = \infty \tag{4.7}$$

This is infinite.

3) **Brownian Noise (Brown Noise)**

The PSD of Brownian noise is:

$$S(f) = \frac{A}{f^2} \tag{4.8}$$

3a) Low-frequency end:

$$\int_0^{f_1} \frac{A}{f^2}\, df = \infty \tag{4.9}$$

This is infinite.

3b) High-frequency end:

$$\int_{f_2}^{\infty} \frac{A}{f^2}\, df = \frac{A}{f_2} \tag{4.10}$$

This is finite.

4.1.1 Implications for Filtering

The finiteness of the PSD integral at both ends for 1/f noise implies that neither high-pass nor low-pass filtering can completely remove this noise. In contrast, white noise can be mitigated using low-pass filters due to its infinite high-frequency component, and Brownian noise can be mitigated using high-pass filters due to its infinite low-frequency component.

In the context of lowpass filtering through accumulation (or integration), the finite low-frequency end of 1/f noise becomes particularly problematic. Accumulation acts as a low-pass filter, emphasizing low-frequency components. For 1/f noise, this means that the noise remains in the accumulated signal, affecting the quality of the filtered output. Furthermore, pink noise is non-stationary, meaning its statistical properties change over time. Stationary interference such as white noise or Brown noise allows for processing techniques such as the Fourier Transform to separate signal from noise based on the statistics of the output. Since pink noise changes with time, attempting to apply standard noise-removal techniques becomes a vastly more difficult challenge(Li *et al.*, 2021). In short, once pink noise has tainted a signal, there is no general solution for its removal.

These unique mathematical properties in combination with the persistent nature of 1/f noise make it a formidable challenge for asrtonomers. Its impact on both heterodyne and direct detection systems makes it a ubiquitous hurdle that often imposes hard limits on senstivity. In the context of a quantum-limited sensitivity detector like the MKID, achievements such as single-photon detection become almost entirely based on a system's ability to counteract pink noise(Mazin, 1970).

4.2   Protecting Signals from Noise with Modulation

One method of mitigating low-frequency noise is through the use of amplitude modulation. While this will not remove any noise already present in the signal, any noise introduced at a frequency lower than the modulation frequency will be removed during the demodulation and downmixing stages. This, conceptually, is at the root of the SpLiF algorithm. If a signal from space is modulated directly after a reciever, any low frequency noise that would otherwise be added by detection, mixing, or signal processing is erased by demodulating with an identical frequency to the modulator. In the case of the SpLiF, the modulated signal is passed through an FFT and each of the FFT bins is individually

demodulated and filtered. By applying this method, one can essentially perform lock-in amplification on an arbitrary number of spectral channels given a high enough clock rate. The essential nature of modulation protecting a signal from low-frequency noise cannot be taken for granted. In order to explicitly express the mathematics of this claim, a proof is required.

4.2.1   Definitions for Proof

Let $s(t)$ be the original high-frequency signal, and $m(t)$ be the modulation signal with frequency f_m. Also, let $n(t)$ be the noise with a frequency f_n such that $f_n < f_m$. First, the signal is modulated. The modulated signal $s_m(t)$ can be represented as:

$$s_m(t) = s(t) \times m(t) \tag{4.11}$$

Next, add the noise $n(t)$ to the modulated signal, yielding:

$$y(t) = s_m(t) + n(t) \tag{4.12}$$

Quadrature demodulation involves multiplying $y(t)$ with $\cos(2\pi f_m t)$ and $\sin(2\pi f_m t)$ and then low-pass filtering.

$$y_I(t) = \text{LPF}\left[y(t)\cos(2\pi f_m t)\right] \tag{4.13}$$

$$y_Q(t) = \text{LPF}\left[y(t)\sin(2\pi f_m t)\right] \tag{4.14}$$

Here, LPF denotes a low-pass filter. After low-pass filtering, components of $y(t)$ that are at frequencies much higher or much lower than f_m are eliminated. The quadrature components $y_I(t)$ and $y_Q(t)$ can be combined to reconstruct $s(t)$;

$$\hat{s}(t) = y_I(t)\cos(2\pi f_m t) + y_Q(t)\sin(2\pi f_m t) \tag{4.15}$$

4.2.2 Proof of Noise Elimination

Proving this feature of filtering will require that $\hat{s}(t)$ is equal to $s(t)$, confirming the noise $n(t)$ has been removed. The demodulated signal without low-pass filtering is:

$$y(t)\cos(2\pi f_m t) = (s_m(t) + n(t))\cos(2\pi f_m t) = s_m(t)\cos(2\pi f_m t) + n(t)\cos(2\pi f_m t)$$
(4.16)

Since f_m is much higher than f_n, the term $n(t)\cos(2\pi f_m t)$ will contain components at frequencies much higher and much lower than f_m. The low-pass filter will remove these components, leaving only $s_m(t)\cos(2\pi f_m t)$.

Similarly for $y_Q(t)$, the reconstruction $\hat{s}(t)$ would thus only contain $s(t)$, as any noise components would have been eliminated by the low-pass filter. Going into more depth, consider $n(t)$ as a sum of sinusoidal components, all with frequencies less than f_m:

$$n(t) = \sum_{k=1}^{N} A_k \cos(2\pi f_k t + \phi_k)$$
(4.17)

where $f_k < f_m$ and A_k and ϕ_k are the amplitude and phase of each noise component. Adding this noise to the signal outputs $y(t) = s_m(t) + n(t)$. Upon quadrature demodulation, multiply $y(t)$ by $\cos(2\pi f_m t)$ and $\sin(2\pi f_m t)$ and apply a low-pass filter. Focusing on $y_I(t)$:

$$y_I(t) = \text{LPF}\left[y(t)\cos(2\pi f_m t)\right]$$
(4.18)

Expanding $y(t)$:

$$y_I(t) = \text{LPF}\left[(s_m(t) + n(t))\cos(2\pi f_m t)\right]$$
(4.19)

This Becomes:

$$y_I(t) = \text{LPF}\left[s_m(t)\cos(2\pi f_m t) + \sum_{k=1}^{N} A_k \cos(2\pi f_k t + \phi_k)\cos(2\pi f_m t)\right]$$
(4.20)

The mixed term:

$$\cos(2\pi f_k t + \phi_k)\cos(2\pi f_m t) \tag{4.21}$$

can be expanded using the trigonometric identity

$$\cos(A)\cos(B) = \frac{1}{2}[\cos(A+B) + \cos(A-B)] \tag{4.22}$$

$$\text{to}: \tag{4.23}$$

$$\frac{1}{2}\left[\cos(2\pi(f_k + f_m)t + \phi_k) + \cos(2\pi(f_k - f_m)t + \phi_k)\right] \tag{4.24}$$

Finally, the lowpass filter comes to clean up any unwanted signal. The components $\cos(2\pi(f_k + f_m)t + \phi_k)$ and $\cos(2\pi(f_k - f_m)t + \phi_k)$ will have frequencies that are either much higher than f_m or much lower. Low-pass filtering will thus attenuate these components, leaving only $s_m(t)\cos(2\pi f_m t)$.

In a similar vein, the $y_Q(t)$ component will also only leave $s_m(t)\sin(2\pi f_m t)$ after low-pass filtering.

4.3   The SpLiF Algorithm

Here I present the general overview for an FPGA-based, algorithmic solution for low-frequency and pink noise so commonly found in high-sensitivity detection systems. The spectroscopic lock-in firmware (SPLIF) processes a modulated input signal, specifically an amplitude-modulated signal. The FPGA-based firmware runs on the Roach 2 board with the Vertex 6 FPGA.

The SPLIF module when run on the ROACH2 uses a fabric clock speed of 256MHz, though it can be slowed down if ever desired. In order to create the modulation/demodulation signals (both of these processes are driven by the same digital waveforms), the SPLIF module relies on two lookup tables of the same depth and user input, with the only difference being the initial value vectors; one being a sine wave and the other being a cosine wave. These

waves are perfectly 90 degrees out of phase, allowing these lookup tables to act as a quadrature pair waveform generator (or IQ waveform generator).

The fundamental or minimum frequency of the I/Q pair produced by such a set of LUTs is determined by the "depth" of the LUT's dataframe. In other words, the number of "steps" in a LUT waveform generator is equivalent to $2^{bitdepth}$ since this is equivalent to the number of samples required to create one wavelength worth of data. In our case we are working with a 20-bit depth, and thus the LUTs are filled with 2^{20} values representing a full 2π radians around the unit circle. Since the current SPLIF implementation is set to run on an FPGA with a fabric clock of 256Mhz, the fundamental frequency is actually a precise 244.140625 Hz (found by simply taking the quotient of $\frac{256MHz}{2^{20}}$. Since this is a rational number that can be expressed as an 14-bit value (11110100.001001 to be exact), we can represent it fully without any quantization noise when it is used to create a quadrature reference later in the design. There, the exact representation of the modulation frequency will need to be known with zero error in order to calculate how many delays are in a quarter clock-cycle of a wave at this fundamental frequency.

The waveform generator can be controlled by the user to produce different carrier frequencies. The modulation tone is obtained by slicing off the signed bit of the sine wave and sending it to the sync out port. The sync out port outputs a square wave oscillating between 3.3 volts and 0 volts, replicating the internally generated wave.

The modulated signal, containing the information, is mixed with the square wave and then fed into a digitizer. The digitizer samples the signal at a rate of 512 MHz, providing real and imaginary parts of each sample.

The sampled data is then passed through a polyphase filter bank fast Fourier transform (PFB-FFT), which outputs the spectral information of the data. Each spectral bin represents a frequency range, and the FFT operates in a time-domain multiplexed fashion. The output consists of complex values.

To perform lock-in amplification, the real and imaginary parts of the FFT output are squared, added together, and then square root is taken to obtain the magnitude spectrum in the time domain.

Each spectral bin oscillates at the modulation frequency, reflecting the power level of the information signal. The modulation frequency is determined by the original waveform generators. To maintain phase invariance, a quadrature portion is added to the FFT magnitude. A variable delay is introduced to achieve the 90-degree phase shift. This ensures that the demodulation scheme is phase-immune.

To demodulate, a digital complex multiplier multiplies the demodulation waveform (the in-phase and quadrature components) with the FFT magnitude. This results in a demodulated version of the original signal. However, higher-frequency artifacts are produced due to the demodulation process.

A low-pass filter is necessary to remove these artifacts. Instead of individually filtering each spectral bin, all the data is accumulated and stored in BRAM (Block RAM). This simplifies the process and avoids the need to separate and filter each bin individually.

4.4 Block Diagram of the SpLiF Algorithm
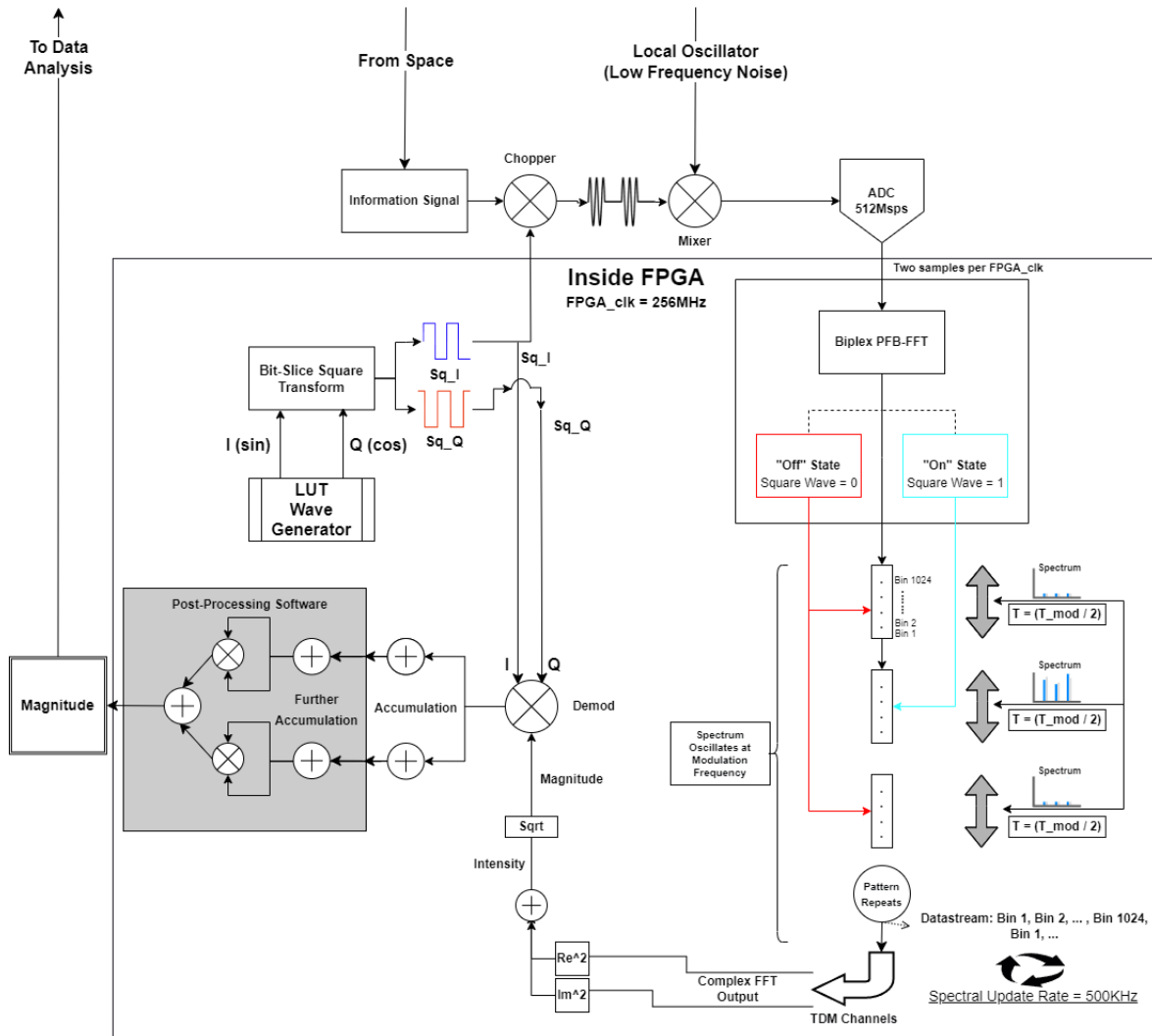
Figure 4.1: The block diagram of the SpLiF, explained below

4.5   Block Diagram Explained

In theory, the SPLiF algorithm can be effectively applied to direct-detection systems as long as the incoming signal can be precisely modulated at a given frequency provided by an FPGA. The block diagram for the SPLiF, however, is specifically designed for heterodyne detection systems, in particular those which detect terahertz frequency signals from space. While the terahertz regime is a small slice of the electromagnetic spectrum, 98 percent of the photons in the universe fall into this range (mainly due to CMB radiation). Along with these CMB photons, there are a number of crucial spectral lines located in the terahertz regime that act as important tracers for galaxy evolution, a classic example being the rotational transition line of water vapor located 556GHz. At such high frequencies, direct digitization of information is beyond the current capabilities of astronomy hardware and therefore must be handled through means of heterodyne mixing.

Mixing down signals like the 556GHz spectral line from water vapor requires some local oscillator capable of creating a steady signal of near-identical frequency. In the diagram of the SPLiF algorithm, the local oscillator shown is responsible for the creation of the heterodyne downmixing signal somewhere in the desired teraherz regime. Creating a stable sinusoid at terahertz frequencies is no simple task and requires the use of yet another mixing product from two optical laser sources. Two lasers are tuned so that their outputs are separated by the desired LO frequency, and by lowpass filtering the product from mixing the lasers with one another, a suitable local oscillator signal can be created anywhere in the terahertz range. While this advanced technique for terahertz signal creation has proved to be an effective method, it is not without its faults. The most pernicious issue created by the LO production is a low-frequency drift noise that dominates most information signals when integrated for any useful amount of time. The amount of low-frequency noise created

by the LO system is so overwhelming that a system had to be created in order to overcome it. This, in fact, is what the SPLiF algorithm was created to solve. By combining the principles of lock-in amplification with the modern standards of spectroscopic astronomy, I aim to show an FPGA-based firmware algorithm capable of reading out these noisy heterodyne detectors while avoiding the low-frequency noise which limits their capabilities.

For the purpose of analyzing the SPLiF algorithm and its eventual software simulation development, a simple model for the LO seen in the diagram is most prudent. The phase noise of the local oscillator can be ignored if the LO signal is separated into two quadrature components. This simplification is possible due to the mathematics of wave mixing. When combining two all-real signals, relative phase between the two is essential for insuring predictable mixing products. This is explained in the mathematical formulation seen later on.//

The local oscillator has a variety of safeguards in place to minimize any mechanical motion. Even with this in place, imperceptible movements can still impact the effectiveness of the laser system in charge of producing and holding the LO signal. Due to a slew of small effects added together with a seemingly unavoidable low-frequency drift in amplitude, 1/f or "flicker noise" runs rampant in the LO signal. The LO does not contribute to noise reduction, in fact it itself is the noise that the SPLiF aims to overcome by upconverting the information signal before it reaches the LO. Mixing with a THz LO is not a choice, but rather a harsh necessity.

With this in mind, the front-end which exists outside the diagram becomes less of a mystery. Say for example the SPLiF is to be used on a telescope of some kind which is capable of sensing THz photons from space. More specifically, the goal of this observation mission is to map the intensity of the 556 GHz water vapor signature in an area of space

that is believed to be a stellar nursery. The science signal, therefore, would be the information collected by the telescope. Using some sort of pixel array, be it MKIDs or any other means of high-sensitivity detectors set up in a frequency-domain multiplexed fashion, the telescope is able to capture the light from the area of space in question. This array of pixels is all connected to a single output transmission line which transfers the output information of the telescope to the back end of the system; this being marked on the diagram as the "From Space" signal.

Before moving to the chopping stage, the instrumentation used for implementing the SPLiF must be understood. In its current state, the SPLiF is suitable for placement on the ROACH2 FPGA; the gold standard FPGA used by astronomers for over a decade. The ROACH2 is synchronized with a clock which drives its fabric rate to 256MHz. Furthermore, the ZDOK (clarification required) adapters on the front of the ROACH2 allow for the attachment of a combined ADC/DAC knows as the MUSIC ADC which samples incoming analog data at a rate of 512Msps. Furthermore, the ROACH2 includes an SMA output port called "syncout" which is capable of producing analog, square clock signals which oscillate between 0 and 3.3V. The syncout feature is what makes the ROACH2 capable of executing the SPLiF algorithm since it is able to create a steady modulation signal without including a digital to analog converter in the firmware compilation.

Inside the FPGA, the diagram shows a quadrature waveform generator which uses lookup tables (LUTs) to create sinusoidal waveforms at a given frequency in complex exponential form. Physically expressed, this complex exponential is represented by two waveforms at precisely the same frequency, one of which is delayed by 90 degrees with respect to the other. This pair of signals can be represented mathematically as $\cos(\omega t) + i\sin(\omega t)$. The purpose of creating a complex quadrature representation of this modulation frequency

61

will become clear at the demodulation step, but for now, only the in-phase component of the generated waveform is required. The LUTs are designed so that digital samples are expressed as signed-2s compliment values which have a least-significant bit (LSB) equal to 1 when the sample is negative, and 0 when the sample is positive. By slicing off only the LSB of both the generated waveforms, the waveforms which were previously sinusoidal are now square waves oscillating between 0 and 1, and 180 degrees out of phase with the original signs of the waveforms. The real (in-phase) component of the generated square wave is branched off in two directions, the first of which heads for the syncout port. Measured in the analog world at the output of the syncout port, a square wave with the exact frequency of the waveform generated internally to the FPGA will be seen oscillating between 0 and 3.3V.

4.6   Mathematical Formulation of the SpLiF Algorithm

Mixing down signals such as the 556GHz spectral line from gaseous H_2O requires a reference signal from the local oscillator capable of creating a steady sinusoidal waveform of nearly identical frequency. In the diagram of the SpLiF algorithm, the local oscillator shown is responsible for the creation of the heterodyne downmixing signal somewhere in the desired Teraherz regime.

Creating a stable, sinusoidal waveform at terahertz frequencies is no simple task and requires the use of yet another mixing product from two optical laser sources. Two lasers are tuned so that their outputs are separated by the desired LO frequency, and by low-pass filtering the product from mixing the lasers with one another, a suitable local oscillator signal can be created anywhere in the terahertz range. Although this advanced technique for creating terahertz signals has proven to be an effective method, it is not without faults. The most pernicious issue created by LO production is low-frequency drift noise. The amount of low-frequency noise created by the LO system is so overwhelming that a system had to

be created to overcome it. This, in fact, is what the SpLiF algorithm was created to solve. For the purpose of analyzing the SpLiF algorithm and its eventual software simulation development, a simple model for the LO seen in the diagram is most prudent. The phase noise of the local oscillator can be ignored if the LO signal is separated into two quadrature components. This simplification is possible due to the mathematics of wave mixing. When combining two all-real signals, relative phase between the two is essential for insuring predictable mixing products. However, by representing waveforms as complex exponentials expressed on the real and imaginary planes, the mixing process can be executed without any dependence on relative phase. The reason behind this subtle yet potent difference is a result of how phase is encoded into a waveform.

4.6.1   Fully Constraining Waveforms with Phase Quadrature

Phase is a fundamental concept in waveform analysis and signal processing, playing a crucial role in a variety of applications, from communication systems to radar technology. This dissertation chapter delves into the concept of phase in waveforms and explores the unique properties and abilities offered by signals in phase quadrature. In the context of waveforms, phase refers to the position of a point within a wave cycle and can be measured in degrees or radians. For a simple sinusoidal wave, the phase indicates the state of oscillation at a certain point in time. It is an important parameter that, along with amplitude and frequency, defines a wave. Changes in phase can lead to constructive or destructive interference, resulting in signal amplification or attenuation, respectively.

Phase quadrature refers to the relationship between two waves that are 90 degrees out of phase with each other. In other words, the peak of one wave coincides with the zero-crossing of the other. This relationship is particularly significant in the realm of signal processing and communications.

Signals in phase quadrature enable the representation of a signal in two dimensions - often referred to as the I (In-phase) and Q (Quadrature) components. This forms the basis of Quadrature Amplitude Modulation (QAM), a key technique in many digital communication systems, including Wi-Fi and 4G/5G cellular networks. QAM allows for the transmission of two different signals simultaneously on the same carrier frequency, effectively doubling the capacity of the communication link.

4.6.2 Analyzing the effects of Quadrature Demodulation Versus All-Real Demodulation

In order to show the difference between demodulating the digital timestreams of the FFT channels using a quadrature pair of signals compared to just the in-phase portion, the clearest proof is through using a simplified scenario. As in the SpLiF algorithm, the downmixing stage occurs after taking continuous FFTs of the modulated digital information signal. To simplify the mathematics as much as possible, some realistic assumptions must be made.

1. The original information signal is comprised of a single, pure sinusoid that lies within the bandwidth of the FPGAs ADC

2. The original information signal we wish to recover does not change in amplitude

3. No noise sources need to be considered (although their presence does not change the final conclusions on phase-dependence as long as they are not dominant compared to the original information signal)

4. The ADC samples 0 Volts when the modulation signal is at zero

5. The frequency of the sinusoid is much greater than that of the modulation signal
   - If these two values are too close to one another, there will not be enough samples to take a full spectrum in either the "on" of "off" state of the modulated input

64

6. The duty cycle of the amplitude modulation is precisely 50%

   - Based on the architecture of the modulation/demodulation waveform generators, a 50% duty cycle is quite accurate. Refer to the section on LUT waveform generation for proof.

Given these assumptions, the picture of the post-FFT samples should become much clearer. Each spectral channel measured over time should oscillate between zero and a constant value that is equal to the channel power, if there were no modulations before sampling. Without any modulation before sampling, we would expect to see a post-FFT-spectra that has virtually no amplitude in any channel except for the channel that encompasses the frequency of the input sinusoid (referred to from here onward as the "bin of interest"). For the sake of clarity, it will be useful to also define a term for the amplitude expected in the bin of interest in the case of no modulation at all. This will be referred to as the "unmodulated amplitude". Extending this context to the SpLiF model, the timestream of the bin of interest should be a near-perfect square wave. It now becomes clear that the goal of the digital downmixing is the retrieval of a DC signal with an amplitude equal to one half of the unmodulated amplitude. The following mathematic examination of downmixing under both the quadrature and all-real demodulation cases will prove that achieving the aforementioned goal without precise phase calibrations is only possible in the first case.

Enumerating Fundamental Elements

Mathematically expressing the predicted differences between complex and all-real demodulation, even with the assumptions made above, is no small feat. In order to effectively show the nuance that separates the two cases, a rigorous formulation is required. There are a number of functions and variables that will act as the main characters in this analysis, and properly defining these is essential before starting any mathematical evaluations. Thus, as

a preamble to the full formulation, definitions are required for the following:

1. External analog signals (i.e. information-carrying signal of interest, modulation signal)

2. Internally generated demodulation waveforms

3. Demodulated, pre-filter signal for both cases

4. Demodulated, post-filter signal for both cases

Clarification: Using Sinusoidal Modulation vs. Square Modulation

Properly defining the waveforms in the preamble (particularly their phase components) will be essential for an accurate analysis of the subtle differences between the two demodulation scenarios. To make the mathematics as simple as possible, it is enormously beneficial to consider only sinusoidal waveforms in this test scenario, as opposed to the practical implementation of the SPLiF which uses a mix of sinusoidal and square waveforms.

To demonstrate the mathematical equivalence between using square waves and sinusoidal waves for modulation and demodulation, an intuitive start is with the spectral makeup of square waveforms through Fourier decomposition, like so:

$$sq(t) = \frac{4}{\pi} \sum_{n=1,3,5,...}^{\infty} \frac{1}{n} \sin(n\omega t) \tag{4.25}$$

This representation shows that a square wave is, in essence, a sum of odd harmonics of sinusoidal waves. By showing that the mathematical concepts of quadrature demodulation are phase-independent for sinusoidal waves, the same can then be said for square waves, as they can be decomposed into a sum of sinusoids. Indeed, virtually any waveform with a valid Fourier decomposition could be used as the modulation waveform as long as the demodulation waveforms are identical in shape and contain quadrature representations in

sample space. In lieu of proving this with a rigorous mathematical proof, studies of super-conducting qubits under periodic latching modulations have empirically shown this extrapolation to be true in the case of modulating and demodulating signals [Silveri et al., 2014]. Were the SPLiF algorithm nonlinear, an argument could be made that the higher harmonic frequencies of the square waves would not have a predictable system response. However, as explained in the walkthrough of the SPLIF module block diagram, the algorithm consists only of linear operations. Thus, the harmonic components of the square waves can be treated as a summation of sinusoids that all undergo linear transformations. In short, given the linearity of the SPLiF, solving for the phase dependence of different digital downmixing scenarios with sinusoidal modulation/demodulation frequencies will provide the same conclusions as using square waveforms.

Mathematical Definitions

Now that any confusion about waveform simplifications has been addressed, formal mathematical assignments for the fundamental elements enumerated above can begin in good faith. Special attention must be given to the phases ϕ and ϕ_d as these represent the phase due to input transmission lines and the phase due to digital latency, respectively. Starting with defining the necessary signals:

$$\text{Original sinusoidal signal: } A(t) = A_{\text{DC}} + A\cos(2\pi f_s t + \phi)) \tag{4.26}$$

$$\text{Modulated signal: } R(t) = A(t)\cos(2\pi f_m t) \tag{4.27}$$

$$\text{Demodulation signals: } S_{\text{complex}}(t) = \exp(-j2\pi f_m t + \phi_d) \tag{4.28}$$

$$S_{\text{real}}(t) = \cos(2\pi f_m t + \phi_d) \tag{4.29}$$

Next, a syntax must be established for the quadrature demodulation signals. Demodulated signals $Z_{\text{complex}}(t)$ and $Z_{\text{real}}(t)$ can be expressed as follows [Error Analysis].

Complex demodulation:

$$Z_{\text{complex}}(t) = R(t) \times S_{\text{complex}}(t) \tag{4.30}$$

All-real demodulation:

$$Z_{\text{real}}(t) = R(t) \times S_{\text{real}}(t) \tag{4.31}$$

Finally, define the lowpass-filtered outputs of the demodulated signals:

$$Z_{\text{complex, LPF}}(t) = \text{LPF}\{Z_{\text{complex}}(t)\} \tag{4.32}$$

$$Z_{\text{real, LPF}}(t) = \text{LPF}\{Z_{\text{real}}(t)\} \tag{4.33}$$

Where the operator $\text{LPF}\{X(t)\}$ represents passing a signal $X(t)$ through a lowpass filter. In practice, this can be done analytically by removing any time-varying components which average to zero on time scales lower than a tenth of a second. This "cutoff" frequency of 10Hz is a general approximation based on the minimum integration time of the SPLiF module in its current form, and the derivation for this timescale will be discussed in the section on hardware implementation of the algorithm.

Hypotheses

With the fundamental elements of this analysis established, the desired hypothesis for the upcoming proof must be formally elaborated. Evaluating equations (3.12-3.21) in the context of modulation/demodulation, this proof aims to prove the following hypotheses:

1. $Z_{\text{complex, LPF}}(t)$ is phase-independent and maximal regardless of relative phases.

2. $Z_{\text{real, LPF}}(t)$ is phase-dependent and reaches its maximal value only when $\phi = \phi_d$.

Starting with the expansion of $Z_{\text{complex}}(t)$ and $Z_{\text{real}}(t)$, further simplification and evaluation will require the use of multiple trigonometric identities. The product-to-sum identities

(also known as Prosthaphaeresis formulas) are as follows:

$$\cos(A)\cos(B) = \frac{1}{2}[\cos(A-B) + \cos(A+B)] \tag{4.34}$$

$$\sin(A)\sin(B) = \frac{1}{2}[\cos(A-B) - \cos(A+B)] \tag{4.35}$$

Applying the first of these identities to equation [3.12]:

$$R(t) = \frac{A}{2}[\cos(2\pi(f_s + f_m)t + \phi) + \cos(2\pi(f_s - f_m)t + \phi)] \tag{4.36}$$

Proof for Divergent Demodulation Techniques

At this point, the proof must finally split into its two scenarios of complex demodulation and real demodulation, starting from the equations:

$$Z_{\text{complex}}(t) = R(t)\exp(-j2\pi f_m t + \phi_d) \tag{4.37}$$

$$Z_{\text{real}}(t) = R(t)\cos(2\pi f_m t + \phi_d) \tag{4.38}$$

Immediately, both $Z_{\text{complex}}(t)$ and $Z_{\text{real}}(t)$ show to be nasty looking equations. Simplifying any further demands even more trigonometric identities, as well as Euler's exceptionally powerful identity to bridge complex exponents and trigonometric functions.

Starting with eq. (3.24) for $Z_{\text{complex}}(t)$, substitute $R(t)$ from eq. (3.24):

$$Z_{\text{complex}}(t) = \frac{A}{2}[\cos(2\pi(f_s + f_m)t + \phi) + \cos(2\pi(f_s - f_m)t + \phi)]\exp(-j2\pi f_m t + \phi_d) \tag{4.39}$$

Applying Euler's identity:

$$e^{(-j\theta)} = \cos(\theta) - j\sin(\theta) \tag{4.40}$$

Eq. (3.27) becomes:

$$Z_{\text{complex}}(t) = \frac{A}{2}[\cos(2\pi(f_s + f_m)t + \phi) + \cos(2\pi(f_s - f_m)t + \phi)][\cos(2\pi f_m t - \phi_d) - j\sin(2\pi f_m t - \phi_d)] \tag{4.41}$$

69

Repeating the same process of substituting $R(t)$ with eq.(3.24) for $Z_{\text{real}}(t)$:

$$Z_{\text{real}}(t) = \frac{A}{2}\left[\cos(2\pi(f_s + f_m)t + \phi) + \cos(2\pi(f_s - f_m)t + \phi)\right]\cos(2\pi f_m t + \phi_d) \quad (4.42)$$

Once again, the prosthaphaeresis formulas are needed. In this case, referring to equations (3.22-3.23), A and B are $(2\pi f_s t + \phi)$ and $(2\pi f_m t - \phi_d)$ respectively. Starting with the complex demodulation case:

$$Z_{\text{complex}}(t) = \frac{A}{2}\left[\cos(2\pi f_s t + \phi)\cos(2\pi f_m t - \phi_d) - \sin(2\pi f_s t + \phi)\sin(2\pi f_m t - \phi_d)\right]$$

$$(4.43)$$

Further simplification requires yet another trigonometric identity for the difference of products:

$$\cos(A)\cos(B) - \sin(A)\sin(B) = \cos(A + B) \quad (4.44)$$

Applying this identity to equation [3.31] leads to:

$$Z_{\text{complex}}(t) = \frac{A}{2}\cos(2\pi f_s t + \phi - \phi_d) \quad (4.45)$$

Repeat the same process to finish the real-valued downmixing. After rearranging $Z_{\text{real}}(t)$:

$$Z_{\text{real}}(t) = \frac{A}{2}\left[\cos(2\pi f_s t + \phi)\cos(2\pi f_m t + \phi_d)\right] \quad (4.46)$$

Finally, substituting the product of cosines identity:

$$Z_{\text{real}}(t) = \frac{A}{4}\left[\cos(2\pi(f_s + f_m)t + \phi + \phi_d) + \cos(2\pi(f_s - f_m)t + \phi - \phi_d)\right] \quad (4.47)$$

Lowpass Filtering

With the preceding exercise in trigonometry complete, the signals for both quadrature and real-valued demodulation are in-hand and ready for the final stage of lock-in amplification: lowpass filtering.

Starting with the complex demodulation technique, the only portion of the demodulated signal found at low frequencies is a constant related to the original amplitude (A):

$$Z_{\text{complex, LPF}}(t) = \frac{A}{2} \tag{4.48}$$

This is maximal and phase-independent, as hypothesized.

Next, for the situation where demodulation is implemented with all-real signals:

$$Z_{\text{real, LPF}}(t) = \frac{A}{4} \cos(\phi - \phi_d) \tag{4.49}$$

This, unlike the quadrature demodulation case, is phase-dependent and reaches its maximal value $\frac{A}{4}$ only when $\phi = \phi_d$.

In summary, by simplifying the assumptions of the different scenarios possible for digital downmixing in the SpLiF algorithm, the conclusions reached are:

1. $Z_{\text{complex, LPF}}(t)$ is maximal and phase-independent.

2. $Z_{\text{real, LPF}}(t)$ is phase-dependent and reaches its maximal value only when $\phi = \phi_d$.

This matches perfectly with the original hypothesis that quadrature demodulation is phase-independent while all-real demodulation is not. Note that in addition to phase independence, eq.(3.36) has a final filtered amplitude equal to $\frac{A}{2}$ while eq. (3.37) only reaches an amplitude of $\frac{A}{4}$ at its maximal condition of $(\phi = \phi_d)$. in order for demodulation to be phase-independent, only the demodulation signal need be expressed in quadrature. It is common for FPGA-based algorithms to use a double-ended signal on both inputs to a complex multiplier in order to maintain phase information about the signal. However, the situation in the SPLiF algorithm differs since the signal to be demodulated is an all-real magnitude from the FFT output, and by nature, has an imaginary component equal to zero. In the case where the real and imaginary parts of the complex output from the FFT are used as the in-phase and quadrature components of the information signal, an all-real demodulation

signal is a viable, phase-independent method of downmixing. However, a pivotal mechanic of the SPLiF algorithm relies on the fact that the magnitude of the FFT outputs are taken before reaching the demodulation stage, and therefore, this is not a feasible solution for phase-dependence.

4.6.3   Translation of Analog to Digital Proof

Yet another added complexity for deriving rigorous mathematical definitions of firmware is the interplay between the analog and the digital world. In the case of the SpLiF algorithm, the original information signal and the modulation signal are both analog and outside of the FPGA, while the rest of the signals are digitally-expressed invocations, created by digitizing the analog input and manipulating the resulting samples. While it may seem tempting to account for this discrepancy in the formulation, doing so will turn an already-dense proof into a truly abhorrent mess of notation. Fortunately, for the purpose of showing the expected differences between the two methods of demodulation, handling all signals as if they are continuous, analog, and time-dependent is perfectly acceptable (Khan *et al.*, 2021).

Lattice Functions in the Context of Signal Demodulation

In the realm of signal processing, particularly in the demodulation of complex and real signals, lattice functions play an indispensable role. These mathematical constructs serve as the backbone for the intricate algorithms that govern the transition from analog to digital signals.

Let $L(x)$ represent a lattice function. In the context of the quadrature demodulation proof, this function can be expressed as:

$$L(x) = \sum_{n=-\infty}^{\infty} s(n) \cdot \delta(x - nT) \tag{4.50}$$

where $s(n)$ is the signal sample at the n-th point, $\delta(x)$ is the Dirac delta function, and T is the sampling period.

Analog to Digital Transition

The transition from analog to digital signals is a pivotal aspect of modern signal processing. Mathematically, this transition can be represented as:

$$x[n] = x_a(nT) \tag{4.51}$$

where x[n] is the digital signal and $x_a(t)$ is the analog signal. The validity of this transition in the context of demodulation is supported by the Nyquist-Shannon sampling theorem, which ensures that the continuous signal can be perfectly reconstructed from its samples if it is band-limited and sampled above the Nyquist rate (Ishidoshiro *et al.*, 2012).

Validation for Digital Signals

The proof provided earlier elucidates the demodulation process for analog signals. However, it is imperative to understand that the same mathematical principles hold true for digital signals. The lattice functions serve as a bridge, ensuring that the demodulation process remains consistent across both domains.

In conclusion, lattice functions not only provide a mathematical framework for the demodulation of signals but also serve as a linchpin in the transition from analog to digital domains. Employing lattice functions ensures a seamless and mathematically sound transition from analog to digital signals, thereby maintaining the integrity and reliability of the data. [Lattice Functions for the Analysis].

Even though Lattice functions allow a connection between mathematics in the digital and analog spaces, it does not fully confine the combination of the two. In order to show that they can be used interchangeably and within the same proof, the same set of conditions must be proven in sample space alone. Once this is shown, the Lattuce functions proof assures that the mathematical result of mixing the two spaces will be the same as the results from either individual case.

Consider the original sinusoidal signal $A[n]$ in sample space, where n represents the sample index. The modulated signal R[n] can be represented as:

$$R[n] = A[n]\cos(2\pi f_m n) \tag{4.52}$$

Next define the complex demodulation signal $S_{\text{complex}}[n]$:

$$S_{\text{complex}}[n] = \exp(-j2\pi f_m n + \phi_d) \tag{4.53}$$

The demodulated signal $Z_{\text{complex}}[n]$ can be obtained by multiplying $R[n]$ and $S_{\text{complex}}[n]$:

$$Z_{\text{complex}}[n] = R[n] \times S_{\text{complex}}[n] \tag{4.54}$$

Substituting the expressions for $R[n]$ and $S_{\text{complex}}[n]$:

$$Z_{\text{complex}}[n] = A[n]\cos(2\pi f_m n) \times \exp(-j2\pi f_m n + \phi_d) Z_{\text{complex}}[n] = A[n]\cos(2\pi f_m n)\cos(\phi_d) + jA[n]\cos(2$$

$$\tag{4.55}$$

Real Demodulation in Sample Space

The modulated signal R[n] remains the same as before, $R[n] = A[n]\cos(2\pi f_m n)$

Next, define the real demodulation signal $S_{\text{real}}[n]$:

$$S_{\text{real}}[n] = \cos(2\pi f_m n + \phi_d) \tag{4.56}$$

The demodulated signal $Z_{\text{real}}[n]$ can be obtained by multiplying $R[n]$ and $S_{\text{real}}[n]$:

$$Z_{\text{real}}[n] = R[n] \times S_{\text{real}}[n] \tag{4.57}$$

Substituting the expressions for $R[n]$ and $S_{\text{real}}[n]$:

$$Z_{\text{real}}[n] = A[n]\cos(2\pi f_m n)\cos(2\pi f_m n + \phi_d) Z_{\text{real}}[n] = \frac{1}{2}A[n]\left(\cos(2\pi(2f_m)n + \phi + \phi_d) + \cos(\phi_d)\right) \tag{4.58}$$

Lowpass Filtering

The process of lowpass filtering the demodulations in sample space is identical to the process in the time domain. The lowpass-filtered version of $Z_{\text{complex}}[n]$, denoted as $Z_{\text{complex, LPF}}[n]$, can be obtained by removing the high-frequency components from $Z_{\text{complex}}[n]$. Similarly, the lowpass-filtered version of $Z_{\text{real}}[n]$, denoted as $Z_{\text{real, LPF}}[n]$, can be obtained by removing the high-frequency components from $Z_{\text{real}}[n]$. In this case, the low-frequency components are f_m and $-f_m$.

For $Z_{\text{complex}}[n]$, the lowpass-filtered version $Z_{\text{complex, LPF}}[n]$ will contain:

$$Z_{\text{complex, LPF}}[n] = A[n]\cos(\phi_d) \tag{4.59}$$

This is maximal and phase-independent, as hypothesized.

For $Z_{\text{real}}[n]$, the lowpass-filtered version $Z_{\text{real, LPF}}[n]$ will contain:

$$Z_{\text{real, LPF}}[n] = \frac{1}{2}A[n]\cos(\phi - \phi_d) \tag{4.60}$$

This is phase dependent and reaches its maximal value $\frac{1}{2}A[n]$ only when $\phi = \phi_d$.

To no surprise, conclusions align with the phase independence of quadrature mixing compared to all-real mixing for time-dependent waveforms.

4.7   Results From Tests

This is the final section of this dissertation and I messed something up. It will be fixed and the results will be added in the next few days.

REFERENCES

Agarwal, G. S., "Quantum statistical theories of spontaneous emission and their relation to other approaches", in "Quantum Optics", edited by G. Höhler, Springer Tracts in Modern Physics, pp. 1–128 (Springer, Berlin, Heidelberg, 1974), URL `https://doi.org/10.1007/BFb0042382`.

Agertz, O. and M. Gieles, "The origin of the milky way globular clusters", Monthly Notices of the Royal Astronomical Society **465**, 3622–3636 (2016).

Aguirre, J., "15-apra15-0081 nasa proposal", (2016).

Battat, J., R. Blundell, J. M. Moran and S. Paine, "Atmospheric phase correction using total power radiometry at the submillimeter array", The Astrophysical Journal (2004).

Bothwell, M., R. Maiolino, R. Kennicutt, G. Cresci, F. Mannucci, A. Marconi and C. Cicone, "A fundamental relation between the metallicity, gas content and stellar mass of local galaxies", Monthly Notices of the Royal Astronomical Society **433**, 1425–1435 (2013).

Buie II, E., E. Scannapieco and G. M. Voit, "Modeling Photoionized Turbulent Material in the Circumgalactic Medium III: Effects of Co-rotation and Magnetic Fields", The Astrophysical Journal **927**, 1, 30, URL `http://arxiv.org/abs/2201.09325`, arXiv:2201.09325 [astro-ph] (2022).

C. Yates, S. J., A. Baryshev, O. Yurduseven, J. Bueno, K. Davis, L. Ferrari, W. Jellema, N. Llombart, V. Murugesan, D. J. Thoen and J. J. A. Baselmans, "Surface wave control for large arrays of microwave kinetic inductance detectors", Ieee Transactions on Terahertz Science and Technology (2017).

Chen, J., W. Diao, Q. Zhao, C. Zuo, Z. Lin, X. Wang, W. Lau, M. Sun, R. Yang and K. Zhang, "Iotfuzzer: discovering memory corruptions in iot through app-based fuzzing", (2018).

Cherchneff, I., "A chemical study of the inner winds of asymptotic giant branch stars", Astronomy and Astrophysics **456**, 1001–1012 (2006).

Davies, J., R. Crain and A. Pontzen, "Quenching and morphological evolution due to circumgalactic gas expulsion in a simulated galaxy with a controlled assembly history", Monthly Notices of the Royal Astronomical Society **501**, 236–253 (2020).

Endo, A., "Superconducting on-chip spectrometery for millimeter-submillimeter wave astronomy", Ieice Transactions on Electronics (2015).

Field, A. J., "The Most Technologically Progressive Decade of the Century", American Economic Review **93**, 4, 1399–1413, URL `https://pubs.aeaweb.org/doi/10.1257/000282803769206377` (2003).

Fowze, F., D. Tian, G. Hernandez, K. Butler and T. Yavuz, "Proxray: protocol model learning and guided firmware analysis", Ieee Transactions on Software Engineering pp. 1–1 (2019).

Hafen, Z., C. Faucher-Giguère, D. Anglés-Alcázar, J. Stern, D. Kereš, C. Hummels, C. Esmerian, S. Garrison-Kimmel, K. El-Badry, A. Wetzel, T. Chan, P. Hopkins and N. Murray, "The origins of the circumgalactic medium in the fire simulations", Monthly Notices of the Royal Astronomical Society **488**, 1248–1272 (2019).

Hafen, Z., C. Faucher-Giguère, D. Anglés-Alcázar, J. Stern, D. Kereš, C. Hummels, C. Esmerian, S. Garrison-Kimmel, K. El-Badry, A. Wetzel, T. Chan, P. Hopkins and N. Murray, "The fates of the circumgalactic medium in the fire simulations", Monthly Notices of the Royal Astronomical Society **494**, 3581–3595 (2020).

Hajian, A. and T. Souradeep, "Measuring Statistical isotropy of the CMB anisotropy", The Astrophysical Journal **597**, 1, L5–L8, URL `http://arxiv.org/abs/astro-ph/0308001`, arXiv:astro-ph/0308001 (2003).

Ishidoshiro, K., Y. Chinone, M. Hasegawa, M. Hazumi, M. Nagai and O. Tajima, "Innovative Demodulation Scheme for Coherent Detectors in CMB Experiments", Review of Scientific Instruments **83**, 5, 056104, URL `http://arxiv.org/abs/1203.2407`, arXiv:1203.2407 [astro-ph, physics:physics] (2012).

Karpov, A., "<title>receivers for ground-based millimeter-wave radio telescopes</title>", (1998).

Kaulakys, B. and T. Meskauskas, "Modeling 1/f noise", Physical Review E **58**, 6, 7013–7019, URL `http://arxiv.org/abs/adap-org/9812003`, arXiv:adap-org/9812003 (1998).

Khan, S., Y. Zhu, M. Jawad, M. U. Safder, M. H. Jaffery and S. Javid, "Analysis of the satellite navigational data in the Baseband signal processing of Galileo E5 AltBOC signal", URL `http://arxiv.org/abs/2101.01381`, arXiv:2101.01381 [eess] version: 2 (2021).

Lehniger, K., A. Saad and P. Langendoerfer, "Finding gadgets in incremental code updates for return-oriented programming attacks on resource-constrained devices", Annals of Telecommunications - Annales Des Télécommunications **78**, 209–220 (2022).

Li, J., M. Takeda, Z. Wang, S.-C. Shi and J. Yang, "Low-noise 0.5thz all-nbn superconductor-insulator-superconductor mixer for submillimeter wave astronomy", Applied Physics Letters (2008).

Li, Y., M. G. Santos, K. Grainge, S. Harper and J. Wang, "HI intensity mapping with MeerKAT: 1/f noise analysis", Monthly Notices of the Royal Astronomical Society **501**, 3, 4344–4358, URL `http://arxiv.org/abs/2007.01767`, arXiv:2007.01767 [astro-ph] (2021).

Liang, C. and I. Remming, "On the model of the circumgalactic mist: the implications of cloud sizes in galactic winds and haloes", Monthly Notices of the Royal Astronomical Society **491**, 5056–5072 (2019).

Longo, G., M. Brescia, G. S. Djorgovski, S. Cavuoti and C. Donalek, "Data Driven Discovery in Astrophysics", (2014).

Matsuo, H., A. Sakamoto and S. Matsushita, "Fts measurements of submillimeter-wave atmospheric opacity at pampa la bola", Publications of the Astronomical Society of Japan (1998a).

Matsuo, H., A. Sakamoto and S. Matsushita, "<title>fts measurements of submillimeter-wave opacity at pampa la bola</title>", (1998b).

Matsushita, S. and H. Matsuo, "Fts measurements of submillimeter-wave atmospheric opacity at pampa la bola: Iii. water vapor, liquid water, and 183ghz water vapor line opacities", Publications of the Astronomical Society of Japan (2003).

Matsushita, S., H. Matsuo, J. R. Pardo and S. J. E. Radford, "Fts measurements of submillimeter-wave atmospheric opacity at pampa la bola ii : Supra-terahertz windows and model fitting", Publications of the Astronomical Society of Japan (1999).

Mauskopf, P. D., "Transition edge sensors and kinetic inductance detectors in astronomical instruments", Publications of the Astronomical Society of the Pacific **130**, 990, 082001, URL https://doi.org/10.1088%2F1538-3873%2Faabaf0 (2018).

Mazin, B. A., "Microwave kinetic inductance detectors", URL https://resolver.caltech.edu/CaltechETD:etd-10042004-120707 (1970).

Peeples, M., L. Corlies, J. Tumlinson, B. O'Shea, N. Lehner, J. O'Meara, J. Howk, B. Smith, J. Wise and C. Hummels, "Figuring out gas amp; galaxies in enzo (foggie). i. resolving simulated circumgalactic absorption at 2 z 2.5", The Astrophysical Journal **873**, 129 (2019).

Rantwijk, J., M. Grim, D. Loon, S. Yates, A. Baryshev and J. Baselmans, "Multiplexed readout for 1000-pixel arrays of microwave kinetic inductance detectors", Ieee Transactions on Microwave Theory and Techniques **64**, 1876–1883 (2016).

Rees, W. G., *Physical Principles of Remote Sensing* (Cambridge University Press, 2012), 3 edn., URL https://www.cambridge.org/core/product/identifier/9781139017411/type/book.

Richings, A. and J. Schaye, "The effects of metallicity, uv radiation and non-equilibrium chemistry in high-resolution simulations of galaxies", Monthly Notices of the Royal Astronomical Society **458**, 270–292 (2016).

Richings, A., J. Schaye and B. Oppenheimer, "Non-equilibrium chemistry and cooling in the diffuse interstellar medium - i. optically thin regime", Monthly Notices of the Royal Astronomical Society **440**, 3349–3369 (2014).

Robert, P., M. Murphy, J. O'Meara, N. Crighton and M. Fumagalli, "Exploring the origins of a new, apparently metal-free gas cloud atz = 4.4", Monthly Notices of the Royal Astronomical Society **483**, 2736–2747 (2018).

Sanchez, N., J. Werk, M. Tremmel, A. Pontzen, C. Christensen, T. Quinn and A. Cruz, "Not so heavy metals: black hole feedback enriches the circumgalactic medium", The Astrophysical Journal **882**, 8 (2019).

Tumlinson, J., M. Peeples and J. Werk, "The circumgalactic medium", Annual Review of Astronomy and Astrophysics **55**, 389–432 (2017).

Tumlinson, J., C. Thom, J. Werk, J. Prochaska, T. Tripp, D. Weinberg, M. Peeples, J. O'Meara, B. Oppenheimer, J. Meiring, N. Katz, R. Davé, A. Ford and K. Sembach, "The large, oxygen-rich halos of star-forming galaxies are a major reservoir of galactic metals", Science **334**, 948–952 (2011).

Uzawa, Y., Z. Wang and A. Kawakami, "Terahertz nbn/aln/nbn mixers with al/sio/nbn microstrip tuning circuits", Applied Physics Letters (1998).

Walch, S., R. Wünsch, A. Burkert, S. Glover and A. Whitworth, "The turbulent fragmentation of the interstellar medium: the impact of metallicity on global star formation", The Astrophysical Journal **733**, 47 (2011).

Wilson, T. L., "Techniques of Radio Astronomy", URL `http://arxiv.org/abs/1111.1183`, arXiv:1111.1183 [astro-ph] version: 1 (2011).

Zaddach, J., L. Bruno, A. Francillon and D. Balzarotti, "Avatar: a framework to support dynamic security analysis of embedded systems' firmwares", (2014).

Zhang, S., Z. Cai, D. Xu, R. Shimakawa, F. Battaia, J. Prochaska, R. Cen, Z. Zheng, Y. Wu, Q. Li, L. Dou, J. Wu, A. Zabludoff, X. Fan, Y. Ai, E. Golden-Marx, M. Li, Y. Lu, X. Ma, S. Wang, R. Wang and F. Yuan, "Inspiraling streams of enriched gas observed around a massive galaxy 11 billion years ago", Science **380**, 494–498 (2023).

Zmuidzinas, J., "Superconducting microresonators: Physics and applications", Annual Review of Condensed Matter Physics **3**, 1, 169–214, URL `https://doi.org/10.1146/annurev-conmatphys-020911-125022` (2012).