Advancement of Heterodyne Focal Plane Arrays for Terahertz Astronomy

by

Caleb Henry Wheeler III

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

Approved October 2016 by the
Graduate Supervisory Committee:

Chris Groppi, Chair
Nathaniel Butler
Philip Christensen
Philip Mauskopf
Paul Scowen

ARIZONA STATE UNIVERSITY
December 2016
ABSTRACT

The Kilopixel Array Pathfinder Project (KAPPa) advances the number of coherent high-frequency terahertz (THz) receivers that could be packed into a single focal plane array on existing submm telescopes. The KAPPa receiver, at 655-695 GHz, is a high frequency heterodyne receiver that can achieve system temperatures of less than 200 K, the specification for ALMA band-9. The KAPPa receiver uses a novel design of a permanent magnet to suppress the noise generated by the DC Josephson effect. This is in stark contrast to the benchmark solution of an electromagnet that is both too expensive and too large for use in kilo-pixel arrays. I present a simple, robust design for a single receiver element that can be tessellated throughout a telescope’s focal plane to make a \( \sim 1000 \) pixel array, which is much larger than the current state-of-the-art array, SuperCam, at 64 pixels and \( \sim 345 \) GHz.

While the original goal to develop receiver technologies has been accomplished, the path to this accomplishment required a far more holistic approach than originally anticipated. The goal of the present work has expended exponentially from that of KAPPas promised technical achievements. In the present work, KAPPa and its extension, I present solutions ranging from 1) the creation of large scale astronomical maps, 2) metaheuristic algorithms that solve tasks too complex for humans, and 3) detailed technical assembly of microscopic circuit components. Each part is equally integral for the realization of a \( \sim 1000 \) pixel THz arrays.

Our automated tuning algorithm, Alice, uses differential evolution techniques and has been extremely successful in its implementation. Alice provides good results for characterizing the extremely complex tuning topology of THz receivers. More importantly, it has accomplished rapid optimization of an entire array without human intervention. In the age of big data astronomy, I have prepared THz heterodyne receiver arrays by making cutting edge community-oriented data analysis tools for
the future of large-scale discovery. I present a from-scratch reduction and analysis architecture developed for observations of 100s of square degree on-the-sky maps with SuperCam to address the gulf between observing with single dish antennas versus a truly integrated focal plane array.
DEDICATION

For my mother who passed away on September 27, 2016. Her strength and sacrifices gave me the freedom to pursue my dreams and showed me the will needed to survive and thrive in harsh world. I love you, Mom.
ACKNOWLEDGMENTS

I can connect a string of fortunate interactions with people that has led me down the path I walk today. I am lucky to have deep friendships, and I would like to acknowledge many of the pivotal moments here. These are people and times that helped me to achieve my doctorate. Most of these events are likely long forgotten by anyone but myself, but for me, these moments changed my world.

The names that are listed here do not include the stories that go with them for the sake of brevity in the current work. Darrel B. Carrol, Mr. Mahoney, John Eversmeyer, Mrs. Shelia Page, Paul P, Kristen P, Markesia B. Mr. (Coach) Steve Radomski, Joey Newton, Greg Stelzer, Tony Schilli, OJ Hodel, Joe Zinkl, Rick Muessig, Mike Stagg, Trevor Toland, Angela Speck, Mike Pagano, Todd Veach, Maureen Wheeler, Jordan Wheeler, and Caleb Wheeler.

Chris Groppi. I am not sure I could have finished my PhD with another advisor. I picked you to work with you because I liked your attitude and I wanted the responsibility of working with a new professor to start a new lab for a crazy project. I got what I bargained for; it turns out most of that stuff was really hard. But I wanted to work hard, not because I believed in our project, but because I believed in you. You treated graduate students like people, you worked hard all the time to get us what we needed, you worked hard for the students in your classes, you were excellent by every measure. When I saw this, I knew that your way could and should be the formula that succeeds, one that could displace the old way; this was something I could believe in. Before you, there were are a lot of people that gave me chance, but you were the first to have faith in what I was doing and to trust me to make the right call because I was the expert. I believe that is what I did, I didn’t make the receiver you asked for, I made the receiver system that was needed. That absolute creative freedom was exactly what I needed and allowed me
to grow unimpeded. I believe in you, Chris, thank you for your faith in me and your friendship.

Natalie Hinkel. You have the most complete view of who I am. You are my Rock. You accepted my obsession with my work, my strange way of perfection or chaos. Our honesty with each other has helped me to be honest with myself. We have always stood by each other. In the summer of 2016, when we moved to Nashville, I expected you to stand by me as everything in our life was changing: we left behind our old jobs and some of the longest friendships either of us have been lucky enough to have. I knew you would stand by me as I worked, mostly alone, writing up the last six years of my life. It was not the first time I had thrown myself into something completely and separated myself from the outside world. I knew it would be hard but that I could make it and survive with acceptable mental damage. I knew we could make it. When my mom died a month before my defense I was lost, and we were both broken, but together. I had 60 pages left to write, but stubborn like my mother, I decided to finish my dissertation; you pledged to support me. At the hardest moment of my life, when no one knew if I could finish, I held on to you, my Rock. Like the house built on the rock, I held onto you as the flood water came in; I grieved and worked and stay with my family. During the last week, I did not know if I could finish. I had lost all perspective on whether I was writing something that would be acceptable, my dissertation committee would not answer my emails, and the emotional toll from the loss of my mother had wounded my whole family. It felt like the flood water had gone above my head and it was dark, but I held my breath and held onto my Rock. The flood ended, the water receded, and you and I were still holding each other, gasping for air but still strong, still standing. I would never have asked you to do that for me, I would never ask so much from anyone, which is why it was such a beautiful gift. I love you, my Rock.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>xi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## CHAPTER

### 1 INTRODUCTION

1.1 The Frontier of the THz (Submm) Sky

1.1.1 Large Scale Mapping

1.1.2 Heterodyne receivers

1.2 Astronomical Observables

1.2.1 Rotational Line Physics

1.2.2 Giant Molecular Clouds

1.2.3 High Redshift Observations

1.3 The Theory for the KAPPa Detector

1.3.1 A Brief Overview

1.3.2 Mixing

1.3.3 Superconductivity

1.3.4 SIS Junctions as Mixers

### 2 KAPPA DESIGN AND FABRICATION

2.1 Cryogenic Engineering

2.1.1 Cryostats

2.1.2 Cryostat Renovation

2.2 KAPPa Pixel Design

2.2.1 SIS Device Design

2.2.2 Electrical Design

2.2.3 Low Power Dissipation LNA
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.4 Small Feature Design</td>
<td>46</td>
</tr>
<tr>
<td>2.2.5 Feedhorn Design</td>
<td>49</td>
</tr>
<tr>
<td>2.2.6 Optical Design</td>
<td>50</td>
</tr>
<tr>
<td>2.3 Magnet Design</td>
<td>51</td>
</tr>
<tr>
<td>2.3.1 Benchmark Electromagnet</td>
<td>53</td>
</tr>
<tr>
<td>2.3.2 Permanent Magnet Design</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3 Permanent Magnet Simulation</td>
<td>57</td>
</tr>
<tr>
<td>3 KAPPA PROCEDURES AND TESTING</td>
<td>61</td>
</tr>
<tr>
<td>3.1 SIS Device Testing</td>
<td>61</td>
</tr>
<tr>
<td>3.1.1 SIS Device Handling</td>
<td>61</td>
</tr>
<tr>
<td>3.1.2 SIS Device Testing and Selection</td>
<td>66</td>
</tr>
<tr>
<td>3.1.3 Device Delamination</td>
<td>73</td>
</tr>
<tr>
<td>3.1.4 KAPPa Pixel Assembly</td>
<td>74</td>
</tr>
<tr>
<td>3.2 Magnet Preparation</td>
<td>76</td>
</tr>
<tr>
<td>3.2.1 Measuring Magnetic Field Strength</td>
<td>76</td>
</tr>
<tr>
<td>3.2.2 Magnet Testing Procedure</td>
<td>80</td>
</tr>
<tr>
<td>3.2.3 Dialing the Required Magnetic Field Strength</td>
<td>82</td>
</tr>
<tr>
<td>3.2.4 Permanent Magnets For Future Arrays</td>
<td>83</td>
</tr>
<tr>
<td>3.3 The KAPPa Test-Bench and Characterization Database</td>
<td>85</td>
</tr>
<tr>
<td>3.3.1 Receiver Tuning Parameter Space</td>
<td>86</td>
</tr>
<tr>
<td>3.3.2 Automated Measuring System</td>
<td>88</td>
</tr>
<tr>
<td>3.3.3 Efficient Receiver Tuning Using Differential Evolution Strategies</td>
<td>90</td>
</tr>
<tr>
<td>3.3.4 Differential Evolution Algorithm Outline</td>
<td>92</td>
</tr>
</tbody>
</table>
### 5 SIMULATIONS TO DETERMINE THE EFFECTS COSMIC DENSITY VARIATIONS IN THE GLOBAL 21CM SIGNAL

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction to 21cm Astronomy</td>
</tr>
<tr>
<td>5.2 Calculation of the Global 21cm Signal</td>
</tr>
<tr>
<td>5.2.1 Consideration for Using 21cmFAST</td>
</tr>
<tr>
<td>5.2.2 Parameter Space Investigated</td>
</tr>
<tr>
<td>5.2.3 Simulations</td>
</tr>
<tr>
<td>5.3 Results of 21 cm Analysis</td>
</tr>
</tbody>
</table>

### 6 FUTURE WORK

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>REFERENCES</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Parameters for Simulations 1, 2, &amp; 3</td>
<td>151</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.1</td>
<td>Taurus Molecular Cloud in CO (2 → 1)</td>
</tr>
<tr>
<td>1.2</td>
<td>Cepheus Molecular Cloud in CO (3 → 2)</td>
</tr>
<tr>
<td>1.3</td>
<td>Gas Cycle in the ISM</td>
</tr>
<tr>
<td>1.4</td>
<td>SIS Junction Current-Voltage Curve</td>
</tr>
<tr>
<td>2.1</td>
<td>Form for 10 K Heat Strap</td>
</tr>
<tr>
<td>2.2</td>
<td>The 10 K Heat Strap</td>
</tr>
<tr>
<td>2.3</td>
<td>The Compact Assembled KAPPa Pixel</td>
</tr>
<tr>
<td>2.4</td>
<td>16 Pixel Receiver CAD, Whole Receiver</td>
</tr>
<tr>
<td>2.5</td>
<td>16 Pixel Receiver CAD, Detailed Pixel Cell</td>
</tr>
<tr>
<td>2.6</td>
<td>Detailed Electrical Layout of the KAPPa Receiver</td>
</tr>
<tr>
<td>2.7</td>
<td>Tool Design for the KAPPa Drilled Feedhorn</td>
</tr>
<tr>
<td>2.8</td>
<td>Electromagnet and Permanent Magnet CAD Models</td>
</tr>
<tr>
<td>2.9</td>
<td>Permanent Magnet System and Simulation</td>
</tr>
<tr>
<td>2.10</td>
<td>Permanent Magnet, Simulated Magnetic Field Across SIS Junction</td>
</tr>
<tr>
<td>3.1</td>
<td>Chip Carrier for Dipping SIS Devices</td>
</tr>
<tr>
<td>3.2</td>
<td>Current-Voltage Curves of Tested SIS Devices</td>
</tr>
<tr>
<td>3.3</td>
<td>Histograms of Tested SIS Device Properties</td>
</tr>
<tr>
<td>3.4</td>
<td>New Chip Carrier Circuit</td>
</tr>
<tr>
<td>3.5</td>
<td>Microscope Image of the Electromagnet and Gaussmeter</td>
</tr>
<tr>
<td>3.6</td>
<td>The Magnetic Flux Density for Three Permanent Magnets</td>
</tr>
<tr>
<td>3.7</td>
<td>Block Diagram of the KAPPa Receiver Characterization System</td>
</tr>
<tr>
<td>3.8</td>
<td>Investigation of the DE Mutation Parameter $F'_T$</td>
</tr>
<tr>
<td>3.9</td>
<td>Histogram of the IF Noise Temperature</td>
</tr>
<tr>
<td>3.10</td>
<td>Histogram of Mixer Conversion Loss</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>3.11 The Intersecting Lines Test for Optics Noise Contribution</td>
<td>112</td>
</tr>
<tr>
<td>3.12 SIS Device IV Sweep with Y-Factor</td>
<td>114</td>
</tr>
<tr>
<td>3.13 System Temperatures as a Function of IF Bandwidth</td>
<td>115</td>
</tr>
<tr>
<td>3.14 Best System Temperatures for KAPPa LO Frequencies</td>
<td>118</td>
</tr>
<tr>
<td>3.15 System Temperature as a Function of Electromagnet Current</td>
<td>120</td>
</tr>
<tr>
<td>3.16 System Temperature as a Function of SIS Current for All Permanent Magnet Tests</td>
<td>123</td>
</tr>
<tr>
<td>4.1 The sReduce Advanced Baselining Algorithm</td>
<td>132</td>
</tr>
<tr>
<td>4.2 A Reference Drawing for the Relationship Between Galactic Coordinates and Absolute Galactic Distances</td>
<td>137</td>
</tr>
<tr>
<td>4.3 The Ratio of Galactic Circular Velocity and Galactic Radius as a Function of Galactic Radius</td>
<td>138</td>
</tr>
<tr>
<td>4.4 Distance Solutions from an Emitter to the Sun</td>
<td>142</td>
</tr>
<tr>
<td>4.5 Measured CO (3 →2) Emission as a Function of Galactic Position</td>
<td>143</td>
</tr>
<tr>
<td>4.6 Average Spectrum for All Data in Subsample of the SuperCam Observations</td>
<td>144</td>
</tr>
<tr>
<td>5.1 Simulation 1, Fiducial Simulation</td>
<td>157</td>
</tr>
<tr>
<td>5.2 Simulation 2, Fastest Reionization</td>
<td>158</td>
</tr>
<tr>
<td>5.3 Simulation 3, Gradual Reionization</td>
<td>159</td>
</tr>
<tr>
<td>5.4 ( \langle S'<em>{21\text{cm}} \rangle - S</em>{21\text{cm}} ) as Function of Redshift for All Simulations</td>
<td>160</td>
</tr>
<tr>
<td>5.5 ( \langle S'<em>{21\text{cm}} \rangle - S</em>{21\text{cm}} ) as Function of Neutral Hydrogen for All Simulations</td>
<td>161</td>
</tr>
<tr>
<td>5.6 Z-Axis Averaged ( S_{21\text{cm}} ) Maps for Simulations 1, 2, &amp; 3 as Function of Redshift 6.5 - 9.0</td>
<td>162</td>
</tr>
</tbody>
</table>
5.7  Z-Axis Averaged $S_{21\text{cm}}$ Maps for Simulations 1, 2, & 3 as Function of Redshift 10.0 - 20.0 .................................................. 163

5.8  The Z-Axis Averaged $S_{21\text{cm}}$ Maps for Simulations 1, 2, & 3 as Function of Neutral Hydrogen Fraction .............................................. 164
Chapter 1

INTRODUCTION

1.1 The Frontier of the THz (Submm) Sky

The terahertz (THz) frequency range is a single decade in the electromagnetic spectrum defined from 0.3 THz to 3 THz. This same region is sometimes called submillimeter (submm) astronomy since it corresponds to a wavelength range of 1 mm to 0.1 mm. The THz is sandwiched between the low frequency end of the far infrared (FIR) from 20 THz to 0.3 THz (15µm to 1 mm) and is higher in frequency than the region known as the extremely high frequency (EHF) at 30 GHz to 300 GHz (10 mm to 1 mm) or millimeter band. In keeping with the naming of the radio bands of increasing frequency, radio instruments that operate in the THz could be called ludicrously high frequency (LHF) receivers. The THz region is the wild-west of detector development since THz photons challenge radio techniques with increasing the high frequency limits of detectors, while the far-IR techniques are starved with lower photon energies.

Most of our knowledge of dense, cold regions in our galaxy comes from FIR dust emission. However, dust makes up less than 1% of the total mass in cold dense clouds. The spectra of dust is quite complicated and depends on a number of parameters, for example: temperature, composition, size, and fractal dimension. Modeling of dust spectra tends to give ranges of compositions with large degeneracies and does not preserve kinematic information. Looking at dust also limits us to a single optical depth determined by the mean dust parameters.

The THz light emitted directly from gas and dust provides a unique window
to the landscape of our Universe. Within molecular clouds, gas alone constitutes \( \sim 99\% \) of the material by mass and is the driving force for the stellar physics and planetary formation within the cloud. Using coherent techniques that proved a precise measurement of light’s frequency, the kinematics for each individual species gas can be determined in addition to the gas’s density and position.

### 1.1.1 Large Scale Mapping

Cold (10-100K) gas lies in the space between stars known as the interstellar medium (ISM). This is the domain of large structures such as giant molecular clouds (GMC), whose spatial scales vary from hundreds of square degrees on the sky to smaller, individual protostellar stars. For observations of small scale structure, interferometers like the Atacama Large Millimeter Array (ALMA) are needed to obtain both high spatial resolution in addition to spectral resolution \( R \) required to fully resolve rotational lines, where \( R \) is define as

\[
\frac{\lambda}{\Delta \lambda} = R = \frac{\Delta \nu}{\nu},
\]

where \([\Delta \lambda, \Delta \nu]\) is the smallest change in [wavelength, frequency] that can be detected at [wavelength \( \lambda \), frequency \( \nu \)].

For the CO rotational line spectra observed in GMCs, thermal broadening from the Maxwellian distribution of molecular velocities dominates the line profile. The normalized line profile from the thermal component of an atom or molecule, \( \phi(\nu) \), is given as a function of observed frequency \( \nu \) as

\[
\phi(\nu) = \frac{c}{\nu_0} \sqrt{\frac{M}{2\pi k_b T}} \exp \left( -\frac{Mc^2}{2k_b T} \left( \frac{\nu - \nu_0}{\nu_0} \right)^2 \right),
\]

where \( M \) is the mass of the atom or molecule, \( T \) is the temperature, \( k_b \) is the Boltzmann constant, \( c \) is the speed of light, and \( \nu_0 \) is the rest frequency of the
emission line. Equation 1.2 take the familiar form of a Gaussian profile. From this we can determine the full width half maximum (FWHM) of the profile by solving the equation as:

\[ \frac{\phi(\nu_0)}{2} = \frac{c}{\nu_0} \sqrt{\frac{M}{2\pi k_b T}} \exp \left( - \frac{Mc^2}{2k_b T} \left( \frac{\Delta \nu/2}{\nu_0} \right)^2 \right) \]

\[ \iff 1 = \exp \left( - \frac{Mc^2}{8k_b T} \left( \frac{\Delta \nu}{2\nu_0} \right)^2 \right) \]

\[ \iff \ln \left( \frac{1}{2} \right) = - \frac{Mc^2}{8k_b T} \left( \frac{\Delta \nu}{\nu_0} \right)^2 \]

\[ \iff \ln(2) = \frac{Mc^2 \Delta \nu^2}{8k_b T \nu_0^2} \]

\[ \iff \frac{8 \ln(2) k_b T \nu_0^2}{Mc^2} = \Delta \nu^2 \]

\[ \iff \Delta \nu = \frac{\nu_0}{c} \sqrt{\frac{8 \ln(2) k_b T}{Mc^2}}. \]

From Equations 1.1 and 1.3 we can determine the spectral resolution, \( R \), required to resolve a thermally broadened emission line emission line profile \( \phi(\nu) \).

The $1.3 billion dollar ALMA cannot be used to efficiently map the 100 square degree structures within GMCs. In Figure 1.1 we see a 120 square degree image of the Taurus GMC made from measurements of the carbon monoxide (CO) J=1 \( \rightarrow \) 0 rotation transition. Making this map would take ALMA \( 6.2 \times 10^7 \) pointings; if it’s assumed that one pointing takes place per second, this map would require 2 years of continuous observation time. To observe large scale structure in the THz, we must be able to effectively map the sky while maintaining the spectral resolution to observe gas kinematics. This is problem that can be tackled by arrays of coherent THz receivers. For example, the Rayleigh Criterion is defined as:

\[ \sin(\theta_R) = 1.22 \frac{c}{\nu D}, \]

\[ (1.4) \]
where $\theta_R$ is the resolution angle, $D$ is the diameter of the circular telescopes primary optic, $\nu$ is the frequency of the observed light, and $c$ is the speed of light. Using Equation 1.4 we see increasing frequency improves the resolution as the diffraction limited beam angle $\theta_R$ decreases. This is a positive gain that improves the spatial resolution, but to maintain mapping speed for a doubling in frequency we must quadruple the number of array elements.

Figure 1.1: The [Pineda et al. (2010)] map shows carbon monoxide column density, where both axes are in degrees, for the Taurus molecular cloud. Note that the Taurus molecular cloud is about 120 square degrees on the sky.

Building detectors for the THz is a difficult task. Aside from being a technological frontier, there is very little technology injection from commercial sources. In addition, the water in the Earth’s atmosphere scatters, absorbs, and emits THz radiation, so ground based measurements take place in a few narrow windows of the electromagnetic spectrum. But when the odds are long the payoff is
big. We have little understanding of what the THz sky looks like or how well all
that gas is really coupled to the dust, so it is a worthy endeavor to better explore
this wavelength regime.

At this point a sane person might argue that building large arrays of high
spectral resolution detectors for the THz might be too ambitious for the nebulous
goal of figuring out what the gas is doing around us. Why invest resources when the
scientific return is unknown? We invest both time and money into the developing
the THz because we are at the beginning of a nonlinear increase in the data
collection rate of our instruments at the exact time when optical and radio
astronomy have already developed many of the techniques needed for making and
processing large scale maps. Collecting the techniques of other branches of
astronomy will require adaptation to the THz, but we already have a roadmap of
what worked when other parts of the spectrum were emerging from darkness, so to
speak. Astronomy has matured from observing only a few bright, nearby targets to
analyzing a variety of stars in the whole sky, which gives a statistical context for all
observations. The time of observing one object in order to write one paper will be
displaced by statistics derived from catalogs that automatically identify and process
all objects. The age of Big Data is upon astronomy, so the field is ready for the high
data rate of 1000-pixel THz array.

Every branch of astronomy is currently going through growing pains, adding
new axes to their old observations. For most branches of astronomy the axis is
frequency, time, or correlated detection; to make a comparable improvement for
high spectra resolution THz, we only need to add more pixels. Large scale surveys
of gas emission lines throughout the Milky Way would provide statistical tools to
address fundamental yet not well understood processes in astrophysics, such as star
formation, the life cycle of the ISM, and the chemical evolution of our solar
neighborhood. Additionally, large scale mapping tells us what is typical within the
galaxy and what is not.

1.1.2 Heterodyne receivers

A common example of a heterodyne receiver is an FM radio. An FM radio
signals are transmitted frequency range of 87.5 – 108.0 MHz, far above the audible
range of human hearing of 20 – 20,000 Hz. Through a process known as mixing
(discussed in detail in Section 1.3.2) radio stations can takes waves of electrons at
audible frequencies that might be formed from grooves on a vinyl record and up
converts them to the frequency of your favorite radio station. The high frequency
radio station signal is then down converted to audible frequencies using heterodyne
receivers that are so ubiquitous in our lives, that they are simply called radios.

To make the designation between heterodyne and direct detection receivers,
let use examine the etymology of the words. The stem of the word ‘hetero-’ is
derived from Greek and in this context means ‘different’. The suffix ‘-dyne’ comes
from the Greek word ‘dyna’ meaning power. This name is wholly appropriate
because a heterodyne receiver is able to distinguish between different photon powers,
unlike a direct detection receiver where each photon received can be treated as
having the same power. Direct detection receivers are generally made from detectors
that can be thought of as very sensitive thermometers: they can tell statistically
how much power is is absorbed, but only the average amplitude of a signal’s
waveform is measured. Having only the amplitude information makes it difficult to
distinguish the small difference in photon powers from light emitted at different
frequencies. Getting frequency information from a direct detection receiver requires
filtering or dispersive optics to limit what frequencies are entering the detector.

However, it is also possible to reunite the phase and amplitude information
in a direct detection receiver if you have knowledge of the phase of the light entering your detector. Unfortunately since we do not have control of the astronomical signals we observe, this trick cannot be used without first measuring the phase of the incoming signal, effectively making a heterodyne receiver. However knowing the phase information of the input signal can be used in conjunction with a direct detection receiver to determine the far-field beam pattern from near-field measurements [Davis et al., 2016].

For many astronomical targets, using a direct detection receiver is usually preferable because the instrument can be made more cheaply and reliably when hyper-spectral resolution is not needed. Such targets are usually broadband, or very faint, where every photon is needed for detection. Direct detection detectors have made great strides in detecting lower photon energies which require increasingly sensitive detectors. At the time of the present work, there is a fair bit of overlap in frequency between the optical-born techniques of direct detection instruments and the radio-born techniques of heterodyne instruments.

At the THz frequencies of $350 - 666$ GHz ($\sim 850 - 450$ microns), the cryogenic bolometer UK SCUBA2 instrument, with 10,240 pixel, is now employed at the 15m James Clerk Maxwell Telescope [Audley et al., 2004]. While SCUBA2 is by far the current state-of-the-art in direct detection detectors. SCUBA2, and all bolometer instruments, are insensitive to the the small differences in power between different frequencies of light. Bolometer arrays can be used as spectrometers if the light is filtered before incidence on the detector. Filters for a direct detection receiver typically define $\sim 10\%$ of the imaging bandwidths. Fabry-Pérot grating spectrometers can be constructed and placed in front of the incoherent arrays to disperse the incoming light, but offer modest spectral resolution ($R < 1000-10,000$) and cannot spatially and spectrally multiplex simultaneously. In the THz, direct
detection arrays are used to image broadband thermal emission from dust, or – with a careful filter – an array can constrain the total power emanating from a target at bandpass.

When an astronomical target requires very high spectral resolution ($R > \sim 10,000$, see Equation 1.1), a heterodyne receiver may be more appropriate than a direct detection receiver. Since the phase and amplitude are persevered and delivered directly to the analog-to-digital converter, the frequency and amplitude of the original astronomical signal can be reconstructed. The physical limit of spectral resolutions comes from the quantum uncertainty of a single photon being absorbed for a given spectral channel. Practically, however the noise temperature of the receiver and the sampling rate of the analog-to-digital converter will limit the receiver’s spectral resolution.

The coherent method is similar to the approach used for the infrared band, where the photoconductors are pushed to lower energies, namely $\sim 200$ microns (e.g. Herschel PACS per Poglitsch et al. 2008 and Spitzer MIPS via Rieke et al. 2004). The current state-of-the-art direct, coherent THz array is 64-pixel, for example SuperCam. While much improved from the single-pixel detectors, not even the scope of these larger arrays is adequate for wide field imaging of weak spectral lines. Both Goldsmith et al. (2009) and Kerr et al. (2009) highlighted the critical role of large focal plane arrays for the advancement of observational submm/THz astronomy. In particular, they emphasized that new, principal technologies would need to be developed in order to meet future scientific demands of facilities like the Cerro Chajnantor Atacama Telescope (CCAT) and the South Pole Telescope (SPT).

Heterodyne receivers, with their inherently high spectral resolution, can be used to measure molecular lines form a GMC accurately enough to determine the clouds relative velocities. In our own galaxy, this allows us to calculate the distance
of the GMC; for extragalactic sources, this means a very accurate determination of redshift. For astronomical and atmospheric applications, there is considerable interest in the behavior of gas species. To disentangle their motions, instruments with high spectral resolutions are required ($R > 10,000$). In molecular clouds, the gas makes up $\sim 99\%$ of the material by mass, and is the dominant player in the physics of star and planet formation. Kinematic information extracted from spectra collected using coherent techniques allows the detailed study of gas dynamics, in addition to the chemical information collected from the detection of a particular gas species.

1.2 Astronomical Observables

In this section we will discuss the underlying astrophysical applications for heterodyne receivers. We will explain the physics for detecting specific molecular lines within GMCs and high redshift targets.

1.2.1 Rotational Line Physics

Diatomic molecules can be separated into two classifications. The first is symmetric, non-polar molecules which have no permanent electric dipole moment, for example $\text{H}_2$. The second type of molecule are asymmetric, polar molecules such as $\text{CO}$. Because they are asymmetric, the permanent dipole moment of a polar molecule oscillates with respect to the line of sight about an equilibrium value as the molecule rotates and vibrates. The molecule then radiates at its rotational and/or vibrational frequency.

Because the research here is geared towards cooler gas within GMCs (see Section 1.2.2), we will only consider rotational molecular transitions. From quantum mechanics, we see that electronic, vibrational, and rotational energy states of a
A molecule must all be quantized. Thus, molecules are only allowed to transition to and from specific rotational energies, leading to emission and absorption of photons at discrete frequencies corresponding to quantized energy between states. The molecules can be approximated as a rigid motor such that the bond length is assumed to be fixed and the total angular momentum is

\[ L = \sqrt{J(J + 1)} \frac{\hbar}{2\pi}. \]  

(1.5)

Here, \( J = 0, 1, 2, \ldots \) is the quantum rotational state which changes in value when a rotational transition occurs, and \( \hbar \) is Planck’s constant. The rotational energy is defined as

\[ E = \frac{L^2}{2I} = \frac{J(J + 1) \hbar^2}{8\pi^2 I}, \]  

(1.6)

where \( I \) is the moment of inertia. During the event of a rotational transition from \( J \rightarrow J - 1 \), the change in rotational energy is

\[ \Delta E = \frac{\hbar^2}{8\pi^2 I} [J(J + 1) - (J - 1)J] = \frac{\hbar^2}{4\pi^2 I} J. \]  

(1.7)

Therefore, the transition frequency is defined as

\[ \nu = \frac{\Delta E}{\hbar} = \frac{\hbar J}{2\pi I} = \frac{\hbar J}{4\pi^2 mr_e^2} \text{ for } J = 1, 2, \ldots, \]  

(1.8)

where \( m \) is the reduced molecular mass and \( r_e \) is the equilibrium nuclear separation, per [Townes & Schawlow (1975)]. Therefore, when analyzing the successions of transitions for a particular molecule, the difference between those transitions are harmonics of the fundamental frequency as determined by the moment of inertia of that molecule.

By using multiple gas lines, we can make observations that would be difficult, if not impossible, via FIR radiation from dust emission. For example, we
can probe deep into dusty star forming regions, retrieve kinematic information about gas around newly formed protostars, and analyze the gas composition in the outflows of asymptotic giant branch (AGB) stars. Large scale surveys of these lines, both in the local universe and beyond, can be used to help answer long standing questions in astrophysics, such as the details of star formation, the life cycle of the ISM, and the chemical evolution of galaxies.

1.2.2 Giant Molecular Clouds

GMCs are dense concentrations of molecular gas, often ranging in mass from $10^3 - 10^7 \, M_\odot$ spanning ~5-200 pc in diameter with a mean density of $\sim 100 \, \text{cm}^{-3}$ (Murray, 2011). While perhaps not gravitationally bound (Dobbs et al., 2011), GMCs are the birthplace for the majority of stars (Murray, 2011). GMCs have highly irregular shapes, exhibiting clumps or filament structures (Scalo, 1990; Falgarone et al., 1991), see for example Figure 1.2. The range of structures results in a combination of hot ($\sim 100 \, \text{K}$) areas of active star formation, warm ($\sim 50 \, \text{K}$) regions, and cooler regions around 10 K.

However, star forming regions in GMCs can have much higher temperatures. Ionizing radiation from newly formed O-type stars can create compact HII regions (Osterbrock, 1989). These compact regions have large temperatures, $\sim 10,000 \, \text{K}$, that are much greater than the typical 10-100 K molecular regions. As higher energy Lyman lines in these compact HII regions are converted to Lyman alpha photons (and two lower energy continuum photons that make up the energy difference), there are an increasing number of optically thick Lyman alpha photons that are unable to escape the cloud. These compact HII regions increase in temperature until collisional excitations begin to play an increasing role in converting light to optically thin phonons that can transport energy out of the cloud.
(see Osterbrock (1989) section 4.2 for a detailed analysis).

Figure 1.2: The Cepheus OB3 molecular cloud showing both the CO (3→2) and CO (1→0) transitions from Sun et al. (2006).

The main theory for GMC formation is due to gravitational instability within the galaxy (Larson 1987, 1988, 1992; Elmegreen 1990b, a, 1991b, a). Gas can accumulate into a large masses that are in agreement with estimated cloud formation time, which is ∼40 Myr (Larson 1994). In addition, the occurrence of star formation in regions exhibiting a critical surface density of gas has been observed and agrees with the predictions by (Kennicutt 1989, 1990). It is also theorized that GMCs may be formed due to spiral density waves that shock compress the ISM. This would account for the presence of a large amount of both molecular gas and molecular gas within the spiral arms (Kennicutt 1989, 1990; Dobbs et al. 2006). Finally, the formation rate of GMCs may also be increased through random collisions of the clouds. However, it is doubtful that these events alone could
form current GMCs since the estimated building time is too slow by roughly a factor of two and because collisions may not always lead to coalescence \cite{Larson:1994}.

The lifetime of a GMC is relatively low, on the order of $\sim 10$ Myr years \cite{Blitz:1980, Larson:1981}, due to the motion of the gas and the sensitivity of radiation and other environmental conditions. Overly dense clumps of gas are formed when turbulence creates fluid flows which are shocked and then cooled. The clumps within the GMCs collide or accrete additional material and eventually lead to star formation \cite{Murray:2011}. Star formation is usually active during the majority of the lifetime of a GMC \cite{Mooney:1988}. Additionally, GMCs may experience a rapid burst of star formation as the clump accretion rate exponentially increases due to density perturbations from initial star formation. Star formation peaks at the end of the GMC’s lifespan before the clouds are disrupted by the radiation pressure of the newly formed stars \cite{Murray:2011}. The short lifetime, namely 10-20 Myr, of the GMCs is on par with the ages of the youngest stars associated with them and is similar to the dynamic time scales of the cloud’s gas \cite{Larson:1981}. Since there are more GMCs with star formation than not, and because the GMCs are disrupted by star formation, this implies that the period of star formation within the cloud is greater than half of the cloud’s 10-20 Myr lifetime \cite{Elmegreen:1991b}.

The irregular shapes and complexities of the structures within GMCs make them difficult to simulate. Fractal properties have been utilized by a variety of authors in order to better understand the clouds, namely \cite{Scalo:1990, Dickman:1990, Falgarone:1991, Falgarone:1992, Zimmermann:1992}. More specifically, the turbulent flows that create the surface boundaries of the clouds have been examined by \cite{Falgarone:1991, Sreenivasan:1991}. Other determinations fractal dimension with respect to the formation of shells, bubbles,
and clumps within the GMCs, and it particularly leads to star formation, was discussed by [Walch et al.] (2013).

**Figure 1.3:** The life cycle of the ISM from [Groppi et al.] (2010). SuperCam and KAPPa probe the bottom half of this diagram, studying the structure and evolution of molecular gas.

Star formation occurs within a GMC when a gaseous region of the cloud becomes overly dense ([Shu et al., 1987](#), see also Figure 1.3). This can happen through a variety of different ways. First, the accumulation of gas (and potentially other stars) causes a gravitational instability that eventually collapses ([Toomre, 1981](#)). Second, shock wave perturbations could occur throughout the cloud, potentially originating from the outer disk, bar, or nearby galaxies ([Roberts, 1969](#), [Grabelsky et al., 1987](#)). Third, shocks created by the rotation of a galactic bar and their subsequent negative torque inside their radii (or galactic bar potential-well)
creates an instability within the gas which accumulates behind the bar and forms stars (Matsuda & Nelson 1977).

The variety of different star formation environments makes clear that star formation within a GMC does not take place within the entirety of the GMC, but within smaller regions of the cloud (Mouschovias 1989; Shu 1995). While the temperatures of GMCs typically remain cold (∼20 K), areas of active star formation may become as hot as ∼100 K. These areas remain at temperatures at or below ∼100 K due to the presence of CO molecules which allow for cooling and are low enough for star formation (Daddi et al. 2010; Tacconi et al. 2010). Temperatures that are above ∼100 K, which occur in shock waves and boundaries of gravitational instability, are too hot for star formation. However, ionization fronts from high temperature HII regions formed from O-stars in GMCs can create shock waves in dense gas that can run ahead of the ionization front compressing the gas (Osterbrock 1989). This process may trigger more star formation in the GMC. Evidence for this is shown in Kraus et al. (2006) section 5.

Kilopixel Array Pathfinder Project (KAPPa) is optimized to examine line emission from ∼100 K molecular gas inside GMCs, as illustrated in Fig 1.3. KAPPa can be tuned from 600-700 GHz with up to 5 GHz instantaneous bandwidth to observe the J=6→5 rotational transition of both $^{12}$C$^{16}$O and $^{13}$C$^{16}$O in the Milky Way. The J=6→5 transition corresponds to hot (∼100 K) molecular gas; this is in contrast to the rest of the GMC which typically shows cold gas emission via the J=1→0 transition.

1.2.3 High Redshift Observations

In the submm sky, there are two dominant sources of light: thermal continuum radiation from dust and narrowband line emission. The continuum
radiation of reprocessed starlight by dust can provide an estimate of interstellar gas mass and a measurement of the main contributor to the bolometric luminosity (dust) of star forming galaxies. However, spectroscopic observations are needed to investigate the dynamics of the gas directly by measuring the narrowband line emission. Evidence indicates that peak star formation in our Universe occurred between $z \sim 1–3$ \cite{Shapley:2011}, therefore we must look at the cold gas of galaxies at these redshifts to better understand star formation \cite{Carilli:2013}.

Spectroscopy can be used to unambiguously separate and identify line emission and, in turn, determine redshift. A measurement of cold molecular gas, the raw material of star formation, starts with a measurement of the CO rotational lines (see Section 1.2.1). After multiple line observations, a gas temperature can be determined and the CO population can be calculated by use of the Boltzmann equation. Low J-number CO observations tell us how much molecular CO is present; using the CO/H ratio of our own galaxy, we can determine the minimum amount of available gas for star formation in the observed high redshift galaxy. Understanding galaxy assembly and star formation requires spectroscopic observations of molecular gas driving these processes.

Spectrometers are needed when broadband detections of light from galaxies of unknown redshift have multiple emission lines shifted into the detecting bandwidth. To use a THz receiver, such as KAPPa, in this manner would require an unprecedented tuning range of the local oscillator (LO) and the ability to use many LOs. Additionally KAPPa can deliver up to 5 GHz of instantaneous bandwidth, so after making the world’s best collection of LOs, the LOs would have to search through 100s of GHz of bandwidth going from a redshift of $z = 1$ to $z = 2$. This is the main impediment for using heterodyne receives like KAPPa to find galaxies of unknown redshift. A more appropriate tool might be a direct detection instrument
that used diffractive optics to create modest spectral resolution that would first identify the approximate redshift and location of high redshift galaxies. Knowing the approximate redshift would inform the tuning of efficient mapping instruments like KAPPa, or high spatial resolution instruments like ALMA.

1.3 The Theory for the KAPPa Detector

In this section, we will describe the goals of the KAPPa project as well as the electrical engineering techniques, such as mixing, superconductivity, and Superconducting-Insulating-Superconducting (SIS) devices, used to make the instrument functional. We give more overview here, as opposed to the previous introductory sections, so that the reader can better understand the background science required for the instrument.

1.3.1 A Brief Overview

The SIS device, see Section 1.3.3, inside the KAPPa pixel takes two input waves between 645-695 GHz and then outputs a single wave that is the difference in frequency of the inputs. For KAPPa, a difference frequency between 0.5-5 GHz is amplified and transmitted to an analog-to-digital converter or a total power meter. While a 5 GHz signal can be amplified by available electronics, a 645 GHz amplification requires a very small transistor gap; such amplifiers do not exist. The added noise from mixing is amplified along with the signal by the first amplifier, so it is preferable to amplify the signal before it is mixed because the process of mixing adds noise to the signal. Since we must mix the signal to lower frequency before it can be amplified, we choose an SIS device for mixing to minimize the added noise.

KAPPa is a compact pathfinder instrument with a 16-pixel two dimensional (2D) integrated focal plane at 660 GHz that was designed to be adapted in the
future to comprise a ∼1000 pixel THz heterodyne array. We developed the
techniques and the technology to needed to reduce cost and complexity of a THz
heterodyne receiver before attempting to construct the fully-featured science
instrument. With the KAPPa single pixel complete, it is now not only possible to
build a kilopixel heterodyne array out of modular elements, but we also developed
an automated characterization system for rapid tuning of the entire array.

1.3.2 Mixing

Mixing is a result of nonlinear transformation formation, where a linear
translation is like \( y = mx + b \). More formally, \( F \) is a linear transformation over the
set \( V \) if and only if both (1) \( F(\vec{v}_1 + \vec{v}_2) = F(\vec{v}_1) + F(\vec{v}_2) \) and (2) \( \alpha F(\vec{v}) = F(\alpha \vec{v}) \) are
true for any vector \( \vec{v} \) in \( V \) and for any scalar \( \alpha \). To represent a nonlinear
transformation we use some results from Taylor’s Theorem where any analytic
function \( F(x) \) can be represented over the open set \( S \) as an infinite sum of
polynomials centered on the point \( x_0 \) in \( S \) such that
\[
F(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots \quad \text{for every } x \text{ in } S.
\]

In circuit theory, ideal resistors, inductors, capacitors, and amplifiers are all
linear circuits. A simple nonlinear circuit is a diode. The ideal diode has very small
resistance for voltages greater than \( V_d \) and infinite resistance for voltages less than
\( V_d \). The ideal diode can be represented as a Taylor polynomial except for the single
point \( V_d \) where the derivative is not continuous and the function is therefore
non-analytic. If we look at the current of an ideal diode as a function of voltage, we
see a step function where current is zero until \( V_d \) and then is proportional to current
after \( V_d \). Consider the Taylor representation around zero volts \( V_0 = 0 \) for an ideal
diode as a function of current, \( I = +a_0 + a_1(V) + a_2(V)^2 + \ldots \).

Let us examine the first nonlinear term for the ideal diode \( I \) and define an
analytic function $I(V) = a_2(V)^2$. First, consider a single sine wave in voltage such that a received photon might make $V(t) = \sin(t)$, and let us set $a_2 = 1$:

$$I(V) = (V)^2$$
$$\iff I(t) = (\sin(t))^2$$
$$\iff I(t) = \sin^2(t)$$
$$\iff I(t) = \frac{1}{2} (1 - \cos(2t))$$
$$\iff I(t) = \frac{1}{2} - \frac{\cos(2t)}{2}.$$  \hfill (1.9)

This results show that the output of our nonlinear device will have a direct current component $\frac{1}{2}$ and a new sine wave, $\frac{\cos(2t)}{2}$, that is double the frequency of the input sine wave, $\sin(t)$, when $a_2 \neq 0$. The critical step is applying the double-angle and half-angle formulas to reduce the powers of the sine functions.

Looking at the next nonlinear term for the ideal diode, we define
\[ I(V) = a_3(V)^3 \] and set \( a_3 = 1 \):

\[ I(V) = (V)^3 \]
\[ \iff I(t) = (\sin(t))^3 \]
\[ \iff I(t) = \sin^3(t) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{3}{4} \sin(t) + \sin^3(t) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{3}{4} \left( \sin(t) - \frac{4}{3} \sin^3(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( 3 \sin(t) - 4 \sin^3(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( 2 \sin(t) - 2 \sin^3(t) + \sin(t) - \sin^3(t) - \sin^3(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( 2 \sin(t)(1 - \sin^2(t)) + \sin(t)(1 - \sin^2(t)) - \sin^3(t) \right) \]  
(1.10)
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( 2 \sin(t) \cos^2(t) + \sin(t) \cos^2(t) - \sin^3(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( (2 \sin(t) \cos(t)) \cos(t) + (\cos^2(t) - \sin^2(t)) \sin(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( (\sin(2t)) \cos(t) + (\cos^2(t) - \sin^2(t)) \sin(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( \sin(2t) \cos(t) + (\cos(2t)) \sin(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( \sin(2t) \cos(t) + \cos(2t) \sin(t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \left( \sin(2t + t) \right) \]
\[ \iff I(t) = \frac{3}{4} \sin(t) - \frac{1}{4} \sin(3t) . \]

The resulting output of this nonlinear term is the superposition of two waves, one that is the same frequency as the original, \( \frac{3}{4} \sin(t) \), and added to that is 3 times the frequency of the original wave, \( \frac{1}{4} \sin(3t) \). The tricky part of the math here used the sum of squares for sine and cosine, both the double angle formula for sine and cosine, the final step was the sine of a sum.

Now that we have seen that a nonlinear device can be used to multiply the frequency of an input signal, let us consider two waves created by the simultaneous
absorption of two photons each with frequency $t_1$ and $t_2$, such that the voltage on the wire can be expressed as $V = V(t_1) + V(t_2) = a_1(\sin(t_1) + a_2(\sin(t_2))$. Suppose this voltage is sent through a nonlinear device where $I(V) = a_2(V)^2$, then we have:

$$I(V) = (V)^2$$

$\iff I(t_1, t_2) = \left( a_1 \sin(t_1) + a_2 \sin(t_2) \right)^2$

$\iff I(t_1, t_2) = a_1^2 \sin^2(t_1) + 2a_1a_2 \sin(t_1) \sin(t_2) + a_2^2 \sin^2(t_2)$

$\iff I(t_1, t_2) = a_1^2 \left( \frac{1}{2} - \frac{\cos(2t_1)}{2} \right) + 2a_1a_2 \sin(t_1) \sin(t_2) + a_2^2 \left( \frac{1}{2} - \frac{\cos(2t_2)}{2} \right)$

$\iff I(t_1, t_2) = \frac{1}{2} \left( a_1^2 + a_2^2 - a_1^2 \cos(2t_1) - a_2^2 \cos(2t_2) \right) + 2a_1a_2 \sin(t_1) \sin(t_2)$.

The first two terms are the familiar single wave result, so we now turn our attention to the cross product in the final term and apply the product formula for sines, namely:

$$C(t_1, t_2) = 2a_1a_2 \sin(t_1) \sin(t_2)$$

$\iff C(t_1, t_2) = \cos(t_1 - t_2) - \cos(t_1 + t_2)$.

This simplified cross product is what makes a nonlinear circuit able to perform the mixing of two signals, which creates two new signals: one of difference frequency of $t_1 - t_2$ and one at the sum frequency $t_1 + t_2$. Let us now reunite the cross product with the other waveforms from Eq. 1.11

$$I(t_1, t_2) = \left( a_1 \sin(t_1) + a_2 \sin(t_2) \right)^2$$

$\iff I(t_1, t_2) = \frac{1}{2} \left( a_1^2 + a_2^2 - a_1^2 \cos(2t_1) - a_2^2 \cos(2t_2) \right) + \cos(t_1 - t_2) - \cos(t_1 + t_2)$.

By having a single controllable frequency, or the LO, a spectrum at one frequency range can be mapped to any equal bandwidth frequency range. The new signal that results from the mixing of the LO and radio frequency (RF) signal called the intermediate frequency (IF).
Mixing has several limitations, the first of which can be seen in the above calculations: some of the signal is lost in undesired mixer products. Even more undesired mixer products can arise from other terms of the Taylor representation. For an ideal mixer, one could specify the nonlinear response function that maximizes the desired output product. In the KAPPa SIS device after the junction, there are filters built into the transmission line to diminish unwanted mixer products before they reach the first amplifier. Down conversion is a process where a mixer receives a high RF signal, $f_{RF}$, with the high LO signal, $f_{LO}$, and outputs the low frequency difference product at a frequency: $f_{IF} = f_{RF} - f_{LO}$. Up conversion mixes a low frequency RF signal with an LO frequency (which could be much much greater than the RF signal) and outputs the sum product, $f_{IF} = f_{RF} + f_{LO}$.

Another issue that affects simple mixers like the KAPPa device is because the receiver is sensitive to both an upper-side and lower-side-bandwidths around the LO frequency. Having an upper- and lower-side-band, is what qualifies KAPPa as a dual side-band receiver. Consider a single IF frequency, $f_{IF}$, for a dual band receiver there are are two down converted RF frequencies such that $f_{RF1} = f_{LO} - f_{IF}$ and $f_{RF1} = f_{LO} + f_{IF}$. While these two RF signal will map to the same IF frequency during down conversion, the two RF signals are 180° apart so it is possible to isolate a single single band using image rejection mixers. See Chapter 12 of “Microwave Engineering” by David Pozar (2012) for a complimentary discussion of mixers.

For the KAPPa project, we use a dual side-band mixer, the principle advantage being that it is simpler to manufacture in the large quantity needed for a kilopixel array. The inescapable disadvantage is that, best case scenario, we receive noise from both the side-bands, doubling the receiver noise. Worst case is that there are astronomical lines in both side-bands mapping to the same IF which increases noise due to confusion. In practice, the astronomical lines are sparse and by tuning
the LO we can determine where side-band signal originates from and pick a tuning to maximize signal-to-noise, which is currently done by hand. However, in a kilopixel array, we are intent to use genetic algorithms and automated characterization routines to optimize the receiver in the lab as well as at the telescope. This work is for the KAPPa project is discussed in depth in Section 3.3.3.

1.3.3 Superconductivity

The mixing of the KAPPa detector takes place over an SIS junction. The SIS junction is between the superconducting antenna and the superconducting high frequency filter. All of these components are grown on silicon wafer and are delivered in a single package that is 0.3 mm × 1 mm × 0.005 mm, which is referred to throughout this paper as the SIS device. When we refer to the SIS device in terms of it electrical function inside the receiver, it is an SIS mixer.

When a THz frequency photon is absorbed, the Cooper pairs within superconductors are able to transmit THz frequency field changes. At a finite temperature, superconductors retain normal unpaired electrons. These unpaired electrons also respond to the field of a THz frequency photon, but unlike Cooper pairs, normal electrons will still scatter off the lattice of the superconductor. This process causes loss in the transmission of the THz signal which increases with frequency, like a normal wire. It is necessary to cool a superconducting THz transmission line sufficiently such that the losses from normal electrons is acceptable. The superconductors in the KAPPa SIS device are niobium, a type II superconductor with a critical temperature of 9.2 K. In KAPPa, the SIS device is operated at 4.2 K, or the bath temperature of liquid helium.

The SIS mixer, like all mixers, down-converts an RF signal to a lower frequency IF signal of the same bandwidth. We use an SIS device as the KAPPa
mixer for two advantages. First, the associated noise of an SIS mixer is much lower than that of a diode or field effect transistor. By bringing the noise level closer to the quantum limit, the original signal is less degraded than it would be by a ‘standard’ mixer. Additionally, the nonlinear current-voltage relationship, an the essential part of the frequency down conversion, occurs over a much smaller step in voltage, $\delta V$, which reduces the power requirement from the LO.

An SIS junction, a type of Josephson junction (JJ), is a sandwich of two superconductors separated by a thin insulating layer. By injecting a bias voltage on one of the superconducting layers, it becomes energetically favorable for a Cooper pair to tunnel through the insulator to the opposing superconductor. In 1973, Brian Josephson won the Nobel Prize in physics for his precise prediction of that Cooper pairs of a superconductor could tunnel through an non-superconducting barrier \( \text{[Josephson, 1962]} \). Named the DC Josephson effect, it is observed that Cooper pairs will travel unimpeded across a JJ in a loop indefinitely. The energy barrier of the gap in the JJ is too large for tunneling of normal electrons to be likely, however, below the critical temperature of the superconductor a super-short appears across the gap. This is allowed because, in the absence of applied currents and magnetic fields, all superconducting electrons can be represented by a single wave function. In this state, all Cooper pairs are in phase coherence. For a Cooper pair to tunnel across an SIS junction, the Cooper pair must have the same phase as that of the superconductors on the other side, or enough energy must be available to break the Cooper pair.

A way to visualize all superconducting electrons is to imagine them as dancers. At any moment, each dancer has a partner; this is the Cooper-pair. But the dance is really with all the superconducting electrons together, so the partners are constantly switching and moving around the dance floor as one single quantum
state. Dancers can switch seamlessly because each dancer is perfectly in time with all the others, which describes the same phase for all Cooper pairs. The switch is so fast that at any moment no dancer is ever alone; isn’t that nice?

A quasi particle describes a single member of a Cooper pair, but this is more of a convenient construct because all the quasi particles in a superconductor act together and are described by a single wave function $\Psi_{\vec{P}}$:

$$\Psi_{\vec{P}} = \Psi \exp(i\varphi(\vec{r})),$$

where $\vec{P}$ is the momentum vector of all the quasi particles and $\varphi(\vec{r}) = \vec{P} \cdot \vec{r}/\hbar$. Each superconducting material has a property known as the coherence length $\xi$, which can be thought of as a distance of influence of the superconductor. Different superconducting materials have different coherence lengths due to the variation in phonon interactions for a superconducting material: the coherence length is the distance over which the phonons of a superconductor can communicate. When two superconductors are separated by a distance much less than the coherence length, Cooper pairs can tunnel without destruction (the DC Josephson effect), communicating between the separate superconductors and aligning their phase. By placing two superconductors at a distance near to or greater than the coherence distance, it is possible to have the superconductors with partial-to-no phase coherence. Under the assumption of weak coupling across the junction, the current can be approximated as:

$$I = I_c \sin(\gamma) = I_c \sin(\varphi_2 - \varphi_1),$$

where $\varphi_1$ and $\varphi_2$ are the phases of the superconductors on each side of the junction and $\gamma$ is the phase difference. In the presence of no applied currents and magnetic fields the phase difference, $\gamma = 0$. Measuring the current of an SIS junction with a
small or zero voltage will result in current measurements, \( I \), that range from \( I_c \) to \(-I_c\). Since this function is bounded by the critical current \( I_c \), if we are able to reduce the critical current, we can suppress the DC Josephson effect.

### 1.3.4 SIS Junctions as Mixers

The JJs are a rich subject and are used in vast array of different applications, however, it is the nonlinear voltage current relationship of the JJ that makes it possible to use as a mixer. But to make the JJ behave as a diode, we must first suppress the ‘noise’ from the DC Josephson effect. In the KAPPa receiver, we accomplish this by creating an applied magnetic field in the region between the insulating region between the superconductors. With considerable work, it can be shown that the critical current in a magnetic field \( I_c(\vec{B}) \) for a rectangular SIS junction (like the KAPPa SIS junction):

\[
I_c(\vec{B}) = I_c(0) \left| \frac{\sin(\frac{\pi \Phi}{\Phi_0})}{\frac{\pi \Phi}{\Phi_0}} \right| ,
\]

\[
\Phi = L(2\lambda_L + d)(\hat{n} \times \vec{B}) ,
\]

\[
\Phi_0 = \frac{h}{2e} \approx 2.0678 \times 10^{-15} \text{ T} \cdot \text{m}^2 ,
\]

where \( \Phi_0 \), the fundamental quanta of magnetic flux, \( \gamma_L \) is the London penetration depth for the surface of each superconductor, \( d \) is the thickness of the non-superconducting gap, and \( L \) is the side length of the region between the superconductors perpendicular to the magnetic field, and \( (\hat{n} \times \vec{B}) \) is the magnetic field strength passing through the gap region. For an SIS junction with a circular contact region (circular boundary conditions) the equation replaces the sine function (rectangular boundaries) with a low-order Bessel function. This equation is in the familiar \( |\sin(x)/x| \) notation and produces a minimum of \( I_c = 0 \) for all non-zero integer flux quanta \( \Phi_0 \).
By applying a magnetic field that places integer flux quanta in the gap of an SIS junction, the critical current can be reduced to zero, which effectively turns off the DC Josephson effect. We have measured data during this scenario, shown in Figure 1.4. In the figure, you can see that there is almost no current flow across the SIS junction (dashed black line 594.6 Ω) until the bias voltage reaches 1.5 mV, which is enough to coax regular electrons across the gap. At a 2.5 mV there is enough power to break apart Cooper pairs and heat the SIS junction above the critical temperature (dotted line at 34.2 Ω), which demonstrates normal conduction properties of a resistor. Figure 1.4 shows real data that is a near-approximation of an ideal diode: very low current below 2.0 mV and then a sharp increase in current. For SIS devices, the quality improves as the first linear region from 0 - 1.5 mV (black dash line) increases resistance (becomes more flat), and the near vertical region from 1.6-2.2 mV (small black line) decreases resistance (becomes more vertical).

Injection of energy by an incoming photon allows the quasi particle to travel across the insulator, causing a measurable change in voltage on the receiving superconductor (Kooi, 2008). By injecting many photons from an LO, we can produce a current of quasi particles through photon-assisted tunneling. The amount of LO energy injected exceeds that added energy from 300 K blackbody photons found in our experiment, so loading from thermal THz sources does not increase quasi particle tunneling. In Section 1.3.2 we demonstrated that down conversion is possible with two signal $V_1$ and $V_2$ passing though a nonlinear circuit. Here we see that $V_1$ is produced by absorbing LO photons and $V_2$ is the ‘signal’ photons from the lab or any other source. By biasing to the voltage across the SIS junction, so that it takes exactly one photon to excite a quasi particle across the junction, we can optimize the mixing conversion to maximize the cross terms that down converts frequency and gives the system the best sensitivity.
Figure 1.4: The current-voltage sweep of the SIS junction. This figure shows the characteristic shape of the an SIS device with the DC Josephson effect suppress by magnetic flux in the insulating region. For this data, three flux quanta $3\Phi_0$ are in the the gap and no LO signal is applied. Three resistance of three linear regions are marked with values given in the legend.
KAPPA DESIGN AND FABRICATION

2.1 Cryogenic Engineering

Instrument building is the physical realization of a new idea. While great breakthroughs and new technology can provide advantages, most of the effort in building any new instrument is repeating known experiments. New features of an instrument are meaningless until tested in the context that makes them comparable to previous instruments. Instruments are a long chain of technologies that must work together to test a single new link in the chain. The environment of astronomy rewards those who output papers, but most of the work done by instrument builders will never be published. However, the unpublishable work to build a chain of technologies is the minimum bar to test one new idea that might end up in a paper. For much of astronomy instrumentation, but specifically for KAPPa, the first technology in the chain is cryogenics.

While it is tempting to skip any part of the technology chain, the most commonly skipped step is the consideration of cryogenic environment in which you experiment will be housed. From a researcher’s perspective, careful cryogenic engineering may seem like a chef being asked to build a special oven for their pies. However, cryogenics is not as reliable a technology as an oven. Every cryostat comes with some variation or defect that must be addressed before the experiment can begin. Often, months will be needed during any experiment to get the cryostat to perform tolerably. If you are on a small project with a limited budget, you are the cryogenic engineer. In the following section we will present the unpublishable and
poorly circulated knowledge needed to form the first link in a chain, from one conscripted cryogenic engineer to another.

Our cryostat was designed to house an array of 16 pixels. During the cryogenic fabrication, over 100 direct current (DC) condition lines were installed from the 300 K bias electronics to the 4 K mounting. However, our cryogenic heat load was dominated by: 1) the resistive dissipation from the electromagnet current before reaching the superconducting wire on the 4 K stage; 2) the first stage low noise amplifier (LNA); and 3) the ambient heat load while the experiment was off due to contributions of thermal radiation and heat conduction of wires and structural materials. However, this seemingly pedestrian accomplishment is the entry price to even consider working on heterodyne array of SIS receivers of $\sim 1000$ pixels.

Our design used $\sim 30$ cm of cryowire as a thermal break between thermal stages. We use Goretex GR as the material for our infrared filters at the 50 K and 10 K stage as optical windows. Since Goretex GR is known to have a slight polarization, each filter is aligned $90^\circ$ with respect to the other. Our vacuum window was a piece of 1 mil thick sheet of Mylar. The window for the KAPPa cryostat was design to accommodate a 16 pixel array, but adding thickness to the Mylar window would decrease system performance. All materials were fastened in a way so as to maintain contact pressure during the differential contraction of metals from room temperature to 4 K. The cryogenic properties of the IF signal cable designed for KAPPa uses a pure thermally conductive copper, but only a small cross section is necessary so it here load is comparable individually run coax cables with stainless steel thermal breaks.
2.1.1 Cryostats

The fundamental detector of the KAPPa project is a niobium SIS mixer (see Section 1.3.4). Temperature changes on an SIS device will change the inductance of the charge carriers (see a detailed discussion in Section 1.3.3). Thus it is desirable to have temperature stable cryogenic systems that are as far below the critical temperature as possible. We satisfy this requirement by using a liquid helium bath, which reaches a temperature of 4.2 K at one atmosphere of pressure. However, such low temperatures can pose many problems. For example, suppose we have a 4.2 K object that we need to insulate. The greatest heat source for our hypothetical 4.2 K object is the air. For large differences in temperature, the air becomes convective which speeds the conduction of heat. Also, the gaseous air components become liquid before reaching 4.2 K. For example, the primary constituent of air is nitrogen, or N₂, which becomes a liquid at 77 K and solid at 63 K. The solution to these problems is to put our 4.2 K object in a vacuum to at least 10⁻⁵ torr when warm, since pressure is a function of temperature (note: lowering the pressure of a space by cooling the space is referred to as cryopumping). One of the difficulties associated with vacuum systems is that they take time to reach the required pressure by means of pumping. Additionally, these vacuum systems sometimes leak, which can allow water vapor to expedite the corrosion of sensitive metals.

Imagine that the KAPPa receiver at 4.2 K resides in a vacuum chamber. The chamber itself is thermally linked to the air in the lab at 300 K. To keep the chamber from heating our 4.2 K device, we physically separate our device from the walls of the vacuum chamber by using G10 fiber glass. The equational heating rate, \( \frac{\partial Q_{thm}}{\partial t} \), of thermal conduction can be seen in Eq. 2.1:

\[
\frac{\partial Q_{thm}}{\partial t} = -k(T)A \frac{\partial T}{\partial x},
\]

(2.1)
where \( k(T) \) is the thermal conductivity of a material at temperature \( T \), \( A \) is the cross sectional area normal to the thermal gradient, and \( \partial T/\partial x \) is the magnitude of the thermal gradient. To minimize heat flow for a given temperature differential, Eq. \( \text{2.1} \) implies that we should choose long thin rod or strips of low \( k(T) \) material to physically separate hot and cold bodies. It is worth noting that \( k(T) \) is rarely constant from 300 to 4.2 K, due to the thermodynamic properties of a given material.

Once our 4.2 K device is thermally separated from the lab, we must also consider heat transfer by radiation. For a blackbody, the radiative heat transfer \( \partial Q_{\text{rad}}/\partial t \) is given as:

\[
\frac{\partial Q_{\text{rad}}}{\partial t} = \epsilon \sigma S (T_h^4 - T_c^4),
\]

(2.2)

where \( \epsilon \) is the emissivity, \( \sigma \) is the Stefan-Boltzmann constant, \( S \) is the surface area of the enclosed device, \( T_h \) is the temperature of the enclosure, and \( T_c \) is the temperature of the enclosed devise. Equation \( \text{2.2} \) demonstrates how radiative heating is particularly effective when the enclosed body has \( T_c \) close to zero. To slow the radiative heat transfer we block the radiation with successive layers of low emissivity radiative shielding. In practice, these shielding layers are made from rigid sheet aluminum or flexible Mylar which are thermally separated from the vacuum chamber. The outer layers of shielding can be cooled by putting them in thermal contact with cheap cryogens, like liquid N\(_2\), or by using a cryocooler. A cryocooler is similar to a refrigerator in that it uses the expansion of compressed helium as a working fluid, in place of the freon used in personal refrigerators. The innermost shield can be thermally coupled to a liquid helium bath to keep our 4.2 K device in thermal equilibrium with its surroundings, making \( \partial Q_{\text{rad}}/\partial t = 0 \). The innermost shield in our cryostat is thermally coupled to a 10 K stage, which is cooled by our cryocooler. The setup that we have hypothetically constructed is called a cryostat.
2.1.2 Cryostat Renovation

In order to build THz instruments, we require an environment in which we can test those instruments. For THz astronomy, this means building a cryostat. The discussion of basic principles of cryostat construction in §2.1.1 leaves us poised to examine the renovation that was completed on the science cryostat within Chris Groppi’s lab. Prior to this project, our science cryostat was sitting in a storage facility in Tucson, AZ for more than 1 year with the inner vacuum chamber exposed to the atmosphere. The cryocooler that was designed for this particular cryostat was salvaged by another project. Unfortunately, this salvage operation left a large hole in the vacuum casing, which allowed fine dust to enter and coat the inside. Therefore, the first step to building our cryostat was to get the cryostat to hold a vacuum. Missing port plates were redesigned and machined and new o-rings were purchased and installed.

After a few weeks of cleaning and installing new parts, we attempted our first vacuum test. The test was not successful: we were able to maintain a vacuum of $10^{-2}$ torr while the cryostat was on the pump, but pressure slowly regained when removed from the pump. At this point in the cryostat renovation, these symptoms could mean either: 1) the inside of the cryostat was still outgassing or 2) there was a tiny leak in the vacuum chamber. Outgassing would most likely to be due to water vapor, among other absorbed gases, within micro fissures in the aluminum vacuum casing. With the flow rate we observed, a leak would be difficult to find, such as a hair across an o-ring or something more sinister.

To eliminate the possibility of outgassing we constructed a thermal jacket made from aluminized bubble wrap to insulate a heater, originally designed to heat a 5 gallon bucket. We zip-tied the heater to the vacuum casing and covered the
cryostat with two layers of insulation. While this setup allowed us to bake out the
cryostat, we were not able to pump below $10^{-2}$ torr. The leak that had evaded our
detection turned out to be on a He fill-line that went from the top of the cryostat to
a tank that held liquid helium and cooled the 4.2 K stage. This leak could only be
detected by taking the cryostat apart and exposing the normally evacuated area
around the pipe while at the same time putting a vacuum on the He fill line that
was typically exposed to atmosphere. By using a special pump that has a helium
leak detector, we found the leak by spraying small amounts of helium from a
compressed helium cylinder onto the He fill tube and waiting for changes in the
helium levels read by the mass spectrometer. After fixing the leaking He fill tube,
the cryostat was able to reach $10^{-5}$ torr.

The repair to the leaking He fill tube involved resoldering the joint that had
developed a fracture. To facilitate this repair we needed to remove all the thermal
sinking to the fill lines. A common problem with cryostats that use a helium bath is
the sudden increase in the thermal gradient of the He fill tube as it approaches the
helium bath. To reduce thermal coupling of the helium bath to the atmosphere, the
He fill pipe is made from stainless steel, which has very low thermal conductivity,
see Eq. 2.1. This sudden thermal gradient can cause stressing of the metal,
particularly near solder joints, which contain two metals with different thermal
expansion rates. In order to reduce the stressing of the metal, the He fill tube can
be thermally sunk to the other cold stages of the cryostat. This makes the
temperature change more gradual, while also reducing the heat load on the helium
bath introduced by the He fill tube. Since the cryocooler draws heat from the
cryostat before helium bath is added, the He fill tube is also pre-cooled. Pre-cooling
helps to reduce sudden stresses on the metals of the He fill tube while the helium
bath is being filled for the first time.

34
Having fixed the vacuum leak we turned our attention to the thermal issues of the cryostat. A new cryocooler was ordered and needed to be installed. Two tasks needed to be completed to preparation for the dewar to be cooled: 1) the thermal sinking of the He fill tube needed to be reinstalled and 2) a thermal link between the 50 K stage of the cryocooler and the 50 K plate of the cryostat needed to be established. We would also have to make a simultaneous link between the 10 K stage and the 10 K plate.

We began by installing the thermal links, or heat strapping, to the He fill tube. A decision was made not to reinstall the original heat strapping since it was in poor condition and of questionable function. The original heat strapping was braided tinned-copper straps. While highly flexible, tinned-copper is not highly thermally conductive at low temperatures. New heat straps were constructed from 99.99% pure oxygen free copper, which offers excellent thermal conductivity even at low temperatures. Copper that has been bent or otherwise mechanically stressed has reduced thermal conductivity. One end of the heat strapping would be connected to a flat aluminum plate and the other end would be threaded along the curved surface of a cylinder on the He fill tube. To reduce mechanical stressing, and since the installation space was small, we needed to make forms to shape the heat strapping before it could be installed. In between the strap and the He fill tube, as well as the strap and the aluminum plate, a thin layer of indium was placed. Indium is a soft conductive metal that can be pressed between more rigid metals to conform to the surfaces of the rigid metals. This property of indium allows for the maximum surface area contact between the rigid metals providing excellent thermal contact. Instead of using the surface area of the head of a single screw as a thermal contact, as was done with the original heat straps, we use thick metal blocks in between the bottom of the screw head and the heat strap to apply even pressure over a greater
area. Blocks with one curved edge were designed for fastening to the curved He fill tube. The screws that were used to fasten the block were fitted with conical spring washers to keep constant tension on the blocks as the metal of the screw cooled and expanded. All the metals involved in the heat conduction of the strap (indium, copper, and aluminum) all oxidize and an oxidization layer is far less thermally conductive then the pure metal. For copper and aluminum, we sanded the contact surfaces with sandpaper until the surfaces were lustrous and then installed them quickly with gloves. For indium, this meant gently rubbing the contact surfaces with an isopropyl wipe and then quickly installing with gloves. The copper is still lustrous after many months since the cryostat is kept under constant vacuum, even when not in use.

Installing the thermal link between the cold stages of the cryostat and the cryocooler was more difficult than the installation of the heat strapping to the He fill tubes, but the process and materials used were very similar. The design requirements of the thermal coupling of the 10 K stage are more elaborate. To begin, we want cross-sectionally thick pieces of copper to conduct heat from the cryostat’s 10 K to the cryocooler. However, the copper we use can only be bent a single time to reduce the stressing of the copper. The copper strapping must be put through a 5 cm slot and, once inside the G10 fiber glass structure, the straps must fit perfectly while taking into account a high offset and two different bolt patterns. Once in place, the copper straps need to have indium installed between contact surfaces. The copper strap must also be installed using low profile techniques due to the limited overhead space. Finally, since all these metals will oxidize, the work must be done in a single work session. In contrast to the 50 K stage, heat strapping was greatly simplified with one caveat. Due to the geometry of the cryocooler, the 50 K strapping needed to be installed at the same time as the 10 K strapping. After
successive attempts and the creation of a practice form to simulate the size and offset of the 10 K stages, we had a heat strap design that could satisfy all of the above design requirements. This design and mold can be seen in Fig. 2.1. The installed 10 K heat strapping can be seen in Fig. 2.2.

![Figure 2.1](image)

**Figure 2.1:** We created a form specifically for the 10 K heat strap because the features were so shallow and due to the fact that the metal can only be bent once without work hardening, to maximize conduction.

After the thermal linking, the only remaining task was the installation of thermometers. Four thermometers were installed: one on the 50 K stage, one on the 10 K stage, and two in the 4.2 K experimental area. Installing wires in a low noise cryogenic environment can be tricky. To reduce the electric and magnetic fields generated by wires, wires with opposite currents are twisted together. Crossing of wires is not preferable but when wire must cross they should do so at $90^\circ$ angles to minimize inductance. The wire should also be thermally sinked with the
environment, to avoid external heating. Thermally sinking of wires involves taping them to the cold stage with thermally conductive tape over a distance of at least 10 cm.

After all of the above work, we were eventually able to cool the cryostat, fill it with liquid helium, and conduct a test of the hold time. A typical test of the hold time for a cryostat such as ours involves filling the helium tank as the 10 K stage reaches 80 K, then waiting for all the stages to reach their equilibrium temperature and finally filling the helium tank again as it runs out. The second addition of helium gives a better understanding of the hold time since there is no latent heat left in the system. We are proud to report that a month after the first liquid helium fill, we were still waiting for the liquid helium to evaporate from the first fill. At this point we allowed the cryostat to warm up, since hold time was far longer than we were willing to wait.
2.2 KAPP\textalpha{} Pixel Design

Toward our goal of creating a kilopixel array, it is desirable to have the entire footprint of all the pixel components fit under the aperture of the feedhorn (see Figure 2.3). This design constraint will allow large numbers of pixel cells to be ‘tiled’ in a 2D array without mechanical interference between adjacent pixels. One of the technologic hurdles that KAPP\textalpha{} addresses is to replace the electromagnet found inside of SIS receivers with a permanent magnet. To enable the new receiver architecture, we have created a versatile single pixel receiver capable of housing either a permanent magnet or an electromagnet without having to take apart the detector block or unmount the receiver from the cryostat. This feature will enable the unambiguous comparison of the single pixel performance while using either the electromagnet or the permanent magnet (see Section 3.5.4).

Our array design easily fits the SIS device, a matching network, a low power dissipation LNA, a permanent magnet, and an IF transmission line beneath a drilled conical feedhorn that is 6mm in diameter. Figure 2.3 gives a pictorial map of the pixel components within the KAPP\textalpha{} footprint. The permanent magnet is in the form of a ‘U’, with the SIS junction centered between the two poles or legs of the ‘U’. In simulations, this provided a very uniform magnet field across the SIS junction, which allows for mechanical tolerances so precision placement will not be required during the permanent magnet installation.

Unfortunately the KAPP\textalpha{} project was never able fabricate the 16-pixel array. One by one every impediment for creating the 16 pixel array was removed. The KAPP\textalpha{} team did maintain a working design of the 16-pixel array that was updated during various advancements as seen in Figures 2.4 and 2.5. The 16-pixel array would have been an irrefutable technology demonstration of the KAPP\textalpha{}}
Figure 2.3: The KAPPA single pixel detector with all components installed and wired in-house along with the high precision machining of the copper block housing the various components. This slide is a combination of work by Caleb Wheeler, Hamdi Mani, and Matt Underhill. This versatile single pixel prototype can have both an electromagnet and permanent magnet installed from the back of the pixel without separating the horn ad the detector block so as to unambiguously compare the suppression of Josephson noise across the SIS by both the permanent magnet and the electromagnet.

project, but instead we choose to invest our efforts in the less cosmetic, but greatest unseen impediment for creating arrays of 1000 pixels: the need for automated characterization and tuning (described in Section 3.3). Time and grant funding will determine if we made the prudent choice.

2.2.1 SIS Device Design

The SIS device was designed by Arthur Lichtenberger of the University of Virginia and Jacob Kooi at the California Institute of Technology. The design was adapted and improved from KAPPa’s predecessor SuperCam (first introduced in
Figure 2.4: The 16-pixel KAPPa array design. A more detailed rendering of an individual cell is given in Fig. 2.5

Section 1.1.2). The device were grown at the University of Virginia by Dr. Lichtenberger using a technique called silicon on insulator (SOI) over a period of months for 20 layers of depositions up to a thickness of 5µm. This includes the superconductor niobium as well as the gold beam leads. The installed device the was used to collect the almost all the data for the KAPPa experiment can be seen in Figure 2.3

The KAPPa Device is optimized for mixing at 670 GHz (CO $J = 6 \rightarrow 5$ transition) while SuperCam’s device was designed to operate at 345 GHz (CO $J = 3 \rightarrow 2$ transition). The KAPPa SIS device is physically smaller than the SuperCam device to meet KAPPa’s 6mm × 6mm footprint size. The SIS junction of KAPPa must be made smaller for optimization at an operating frequency double that of SuperCam’s. The KAPPa SIS device is small at 1450 µm × 546 µm × 5 µm
while the SIS junction that the KAPPa device houses is only $1\mu m \times 1\mu m$. This necessitates that the depositions must be extremely accurate. The KAPPa project benefited from device yields above 50%, a fortunate boon, since device testing and verification prior to installation was only successful for less then one device in 10 (see Section 3.1.1).

2.2.2 Electrical Design

For an array of 1000-pixels, it is necessary to have as many pixels as possible on a single, densely packed block. As the array size grows, it becomes increasingly important to condense the electrical routing. The single pixel KAPPa receiver required 4 DC lines to operate the WBA23 LNA and to bias the SIS device while
monitoring its current. To design electrical routing of the 16-pixels KAPPa array would require a multilayered circuit board to deliver individual voltages to each pixel. At minimum, one edge of the array would need to be dedicated to high-density connectors. Unfortunately, it seems that groups of amplifier and pixels cannot be biased in parallel. Aside from performance fluctuations from the manufacturing process, the same SIS device can delaminate over time changing the optimal bias conditions. Additional we saw the optimal bias conditions change as a function of LO power and frequency. When the LO was performing best and the receiver was operating under ideal magnetic field conditions, the optimal bias and current would shift. For a very large array of 1000-pixels, electrical routing of 4000 conductors will be difficult: any unobstructed surface on the side of the array would need to be used for high density conductors. One solution may be to put a small computer directly onto the array block and to adjust and maintain the bias voltages for each SIS device. This simplifies the issue of connecting 4000 conductions to the edge of the array, but it adds a source of heat to the receiver’s cold stage. Having a bias computer inside the cryostat also has the important advantage reducing the number of wires traveling from 300K to the 4K receiver stage, reducing cost and heat load. Transmission of the IF signals out of the cryogenic environment is also a challenge for heterodyne arrays. The single pixel KAPPa design uses a simple stripline transmission line to take the IF signal from the LNA output to a coax connector. The stainless steel coax cable of the KAPPa IF was the least reliable part of the entire KAPPa project. During every receiver cool-down, the solder joint would crack from the differential contraction of the two metals and required frequent replacement. It comes as no surprise that for the design of the KAPPa array it was desirable to create our own cable for multiple IF lines. After the IF signal leaves the LNA in the KAPPa array, the signal travels a short distance on a
top plane multilayer circuit board to a hot via, then out the rear of the detector block. The rear of the detector block is also where the permanent magnet is installed. So after the installation of the permanent magnet, we plan to connect the IF of the array to a flexible circuit board with G3PO connectors.

**Figure 2.6**: A detailed electrical layout of the KAPPa receiver.

### 2.2.3 Low Power Dissipation LNA

As outlined in Section 1.3.1, the first amplification of the signal is after the SIS mixer. Having a Low Noise Amplifier (LNA) directly after the SIS device is essential to preserving the signal to noise ratio. To scale to array of 1000 SIS pixels a careful trade study must be performed between amplifier gain, noise temperature,
bandwidth, and additionally there must be very low power dissipation and reactive matching to the SIS device. While noise temperature should be low and gain should be high, the economics of array building allow us to select an amplifier that adds some noise with the trade of having very low power dissipation. Additionally we needed a physically small device that we could mount very near to the IF output of the SIS device.

The low power LNA, WBA23 (Russell & Weinreb, 2012), was the ‘Goldilocks’ amplifier we selected to be used in the KAPPa experiment, having very small power dissipation while adding a small but acceptable amount of noise to the receiver (see Section 3.4.1). Additionally, this amplifier was used for an IF bandwidth of 0.5-5 GHz. This LNA was developed and manufactured by Damon Russell and Sander Weinreb at California Institute of Technology. This packaged device is a compact 800µm by 500µm and is easily integrated into the KAPPa footprint (see Figure 2.3). The impedance matching of the WBA23 to the SIS device was done by Damon Russell and Jacob Kooi. The matching network required only a single 2.5nH inductor wire bonded between the SIS device and the WBA23. The initial single pixel design called for a matching network that included a parallel matching capacitor in addition to the series matching inductor. However, it was determined prior to the any measurements with the receiver that the capacitor was not needed and remained disconnected for all experiments (this vestigial capacitor can be seen disconnected in Figure 2.3). Thus, only the matching inductor was included in the 16 pixel design.

Figure 2.6 shows that the WBA23 and SIS device were intended for integration, as the WBA23 includes a bias-T and resistors that allowed the KAPPa team to measure the current and voltage across the SIS junction with only 3 DC conductors reaching the the SIS device via the WBA23 (see the schematic in Figure
2.6 and compared to the installed device in Figure 2.3). Grounding the WBA25 required wire bonding it to many grounding pads (see Figure 2.3), to the surrounding detector block. For the single pixel design we used the minimum amount of space for grounding to the copper detector block, but since we hit our minimum footprint size with room to spare it was our intention to relax this requirement for the 16-pixel design. This would allow for a simpler and faster pixel assembly.

2.2.4 Small Feature Design

The KAPPa project benefited from being literally tangential to THz group’s world-class micro-milling facility at Arizona State University. The principal advantage to working in the micro-milling facility was being on the forefront of the constant stream of innovations, as well as the frank advice we received about the early versions of KAPPa. Here we wrestled with the KAPPa design until we found a version that could truly be scaled to ∼1000 pixels.

The SIS derive requires a precise pocket design for electromagnetic reasons. The mechanical design of the detector block starts with an oval waveguide $410 \mu m \times 125 \mu m$ with a $60 \mu m$ radius of curvature at the edges of the device. With a depth of $114 \mu m$ (a quarter wave length for light at 660 GHz), this feature is an example of the 3-1 rule, where we try not to cut to depths that are three times the diameter of the cutting tool. At large aspect ratios of tool length to tool diameter, the bending of the tool due to the cutting forces while engaging the metal causes $\mu m$ level deflections in the tool. These deflections can be in excess of the precision needed for guided light at THz frequencies. This is why we use an oval waveguide (see Figure 2.3) as opposed to the more common rectangular design that would have a lower cross-pol. An electromagnet backshort and the ground plane under the SIS device were required to be machined in a way consistent with the simulation results.
The single pixel KAPPa receiver was aligned in the usual way with two precision ground 1/16” stainless steel pins in two apposing corners of the array. Every time the pixel was opened, separating the horn and detector block, one would imagine all the fragile epoxied components shredding under the force of opening. Fortunately it was the opposite that occurred, the pixel survived for over 20 openings, including once when it was dropped it on the floor from a height of 1 meter during an explanation of the experiment. It was an amazingly robust piece of technology in contrast to the stainless steel IF coax cable outside the pixel would fail regularly. The DC Omnetics connector and electromagnet survived around 40 connections. This robustness will be needed if a kilopixel array is to make it up a rickety road to high altitude telescopes or to survive a parachute landing as a part of a balloon payload.

The most difficult feature to design was the feed-through for both the electromagnetic and the permanent magnet that were installed from the rear of the detector block. While the KAPPa pixel requires only a permanent magnet, it was important that the single pixel be able to use an electromagnet so that a comparison of performance could be made. The aspect ratio of the permanent magnet feed-through was much greater than 3-1 rule, however the feed-through required far less precision than the other features of the KAPPa pixel footprint. The permanent magnet feed through was designed to be drilled, but this created a square-peg-round-hole problem as the permanent magnet used field concentrators that were rectangular after their initial construction. It was our intention to precision mill each field concentrator from a rectangular shape too a circular one to match the feed through, however this proved to be unnecessary as the oversized features for the electromagnet could easily accommodate the permanent magnet field concentrators without additional milling. For future array the same approach
will be used of creating an oversized hole in the detector block and having a corresponding feature machined into the Horn block to receive the field concentrators. This allows for tight mechanical tolerances of the permanent magnet will be maintained from the pocket on the rear of the detector block that holds the permanent magnet and concentrators in place with respect to the SIS device. This will keep the machine of the detector block relatively simple and removes the need for milling of the field concentrators after they have been initially manufactured with a wire by electro discharge machining (wire EDM) as seen in Figure 2.5.

After the SIS junction, the SIS device has a microstrip with large and small sections that filter out high frequency mixer products, with some effort this can be seen on the SIS device in Figure 2.3. Like all microstrip transmission lines, planes of the detector block must be at a set distance achieve the microstrip designed electrical impedance. This requires that the SIS device sealed in copper with machined features both on the horn block in addition to the detector block. Good conduction between the detector block and the horn block for the features directly around the SIS device is important as small gaps can allow for unwanted single to enter the SIS device and additionally the desire astronomical signals can escape. Conduction is done by electrons and electron repulsion is the force keeping the surfaces of two tangent metals from passing right through one another. So it is no surprise that conduction between to adjacent metals is directly proportional to the pressure between the metals. This is why the KAPPa horn block has minimal corresponding feature that that surrounds the SIS device but is is relieved elsewhere. The depth of the relief of the KAPPa horn-block was the distance from the circuit board that would have the least interference for the IF frequencies of 1-5 GHz for the signals that were outbound on the wire bonds and microstrip.
2.2.5 Feedhorn Design

Feedhorns are horn shaped antenna that transition an electromagnetic wave from the impedance of free space to the impedance of waveguide. Feedhorns are usually characterized in the context of their beam pattern, the probability of transmission of a photon from a given direction. For the KAPPa array we were looking to have a beam pattern that was compatible with the Sub-millimeter Telescope on Mt. Graham Observatory in Arizona. Additionally we wanted to make the mouth of the feedhorn, at the transition to free space, to be a small as possible so that the pixels could be packed as densely as possible while also being within 10% of a Gaussian beam pattern.

What complicates the matter is that KAPPa’s 2-D integrated design necessitates that the feedhorns be machined from a single copper block that can be precisely aligned with the detector block below. The requirement of precision alignment and scalability prevent the construction of many individual split-block feedhorns that would later need to be chained together to form the whole array. Our solution was to use our precision milling machine to drill all feedhorns into a single block.

With simplicity and scalability as our allies we worked with our collaborator, Jamie Leech of the University of Oxford, to design a feedhorn that could match the beam pattern of the feeding telescope to the single mode waveguide of our detector, be made from drillable cascade of cones, and have a small as possible aperture to free space. This design can be seen in Figure 2.7. Working with tool makers we constrained our feedhorn solutions to be series of conical features where the opening angle of the feature could only decrease as the feedhorn when from the oval waveguide to the 6 mm aperture to free space. As the first two profiles of the feed
horn were nearly the same angle and beyond the tolerance of the tool maker, these surfaces were combined into a single surface. Working with our tool designer, we came up with a system to use two tool single flute tools to cut two profiles of the feedhorn as seen the CAD drawing of out tool design in Figure 2.7. The first tool cuts the primary profile at the mouth an the next profile before the KAPPA circular waveguide. The second tool cuts transition to the oval waveguide. After cutting the the feedhorn and waveguide transition surfaces, the horn block must be flipped over and the profile of the oval waveguide is precision milled form the other side.

While the feedhorn accepts all polarizations of light, the oval waveguide and the probe on the SIS device have only been optimize for one polarization. The transition from feed horn to waveguide stares with a sub-wavelength 150µm circular waveguide section the same diameter as the oval waveguide’s largest diameter an then transitions into the oval waveguide with a 30° taper transitioning to the waveguide’s minor axis.

2.2.6 Optical Design

For simplicity, we have chosen an SIS RF detector that is only sensitive to a single polarization. To support a single polarization of radiation with minimal loss inside a conductor, we can construct a rectangular cavity (or waveguide) inside a a conductor that is slightly bigger than λ/4 × λ/2, where λ = 0.461 mm is the wavelength of radiation in air corresponding to 0.660 THz. We use such a waveguide for our RF detector. The RF detector, a fan shaped device that is drawn in red Figure 2.5, must be placed a distance of λ/4 from the fifth (terminating) wall of the waveguide to assure maximum absorption efficiency. Our detector must then be housed in a solid block of metal, hereafter referred to as simply ‘the block.’ We chose to make the block out of copper for its high electrical and thermal conductivity.
2.3 Magnet Design

Using a technique known as mixing, SIS devices in THz heterodyne receivers allow for the down-conversion of high RF signals to lower IR frequencies. The IF can then be transmitted and analyzed by commercially available components. For an SIS device to have efficient down-conversion at high frequencies, an applied magnetic field is required to suppress Josephson Noise (Leridon, 1994), or Cooper pairs tunneling across the SIS junction. Down conversion is most efficient when the applied magnetic field creates an integer number of magnetic flux quanta in the superconducting regions of the SIS junction. However, there is a considerable degradation of performance for non-integer flux quanta applied magnetic fields. Further, the KAPPa receiver has very poor response when no applied field is present. The historic choice for magnetic field generation within an SIS receiver has
been electromagnets because they allow for precise tuning of the magnetic field. However, electromagnets are too large and consume too much power to be used in future high pixel-count SIS array receivers. Our strategy to solve the issue of power consumption and packing density for future THz arrays is to replace a tunable electromagnet with an off-the-shelf fixed permanent magnet systems for magnetic field generation.

KAPPa makes the technological advancements needed to create arrays of THz heterodyne SIS receivers that are on the order of ~1000 pixels. For comparison, SuperCam is currently a state-of-the-art SIS receiver with a 64-pixel array (Groppi et al., 2010). Many technological improvements have been made by the KAPPa team to realize the future of large heterodyne arrays (Groppi et al., 2012; Wheeler et al., 2014). To make an array with larger numbers of pixels, we are required to make each pixel smaller, simpler, and cheaper. The fulcrum of this work has been whether the magnetic field generation required for SIS receivers can be simplified without substantial degradation of performance that would negate the benefits of large array construction.

The KAPPa receiver was designed to house either an electromagnet or a permanent magnet to compare the performance with different modes of magnetic field generation. The flexibility of the KAPPa receiver allows for an electromagnet and permanent magnet system to be installed interchangeably from the backside of the receive’s detector block. Installation from the backside negates the need to open the detector block and disturb the SIS device or supporting electronics. This feature provides a direct comparison of the same system’s performance and an unambiguous test-bed for different magnetic field generation techniques. While the permanent magnets apply a fixed magnetic field, there are many ways to adjust the field strength of a magnet prior to installation in the receiver.
Figure 2.8: Solid models of the electromagnet and permanent magnet systems. The electromagnet footprint, 20 mm × 5 mm, viewed from the front (panel A) and the back (panel B) and the permanent magnet footprint 3.175 mm × 3.175 mm, viewed from the front (panel C) and the pack (panel D) of the detector block. The detector block houses the SIS junction and supporting electronics. In the front view, the feedhorn block has been made translucent, however the outline of 6 mm diameter feedhorn aperture can still be seen. In panel D, The permanent magnet holder has been made translucent. The permanent magnet holder is needed to reuse the same mounting scheme required by the electromagnet.

2.3.1 Benchmark Electromagnet

The electromagnet used as a benchmark in the KAPPa project was originally designed for SuperCam 64-pixel heterodyne focal-plane array. The SuperCam team, through considerable effort, minimized the manufacturing cost and complexity of these electromagnets. Our goal has been to shrink the footprint of the single pixel to fit in a square 6 mm × 6 mm to fit under a 6 mm diameter feedhorn aperture. However, the 20 mm × 5 mm footprint of a single electromagnet is almost 3 times
the goal footprint for the entire KAPPa pixel, see Figure 2.8. Despite the many disadvantages, the electromagnet’s field can be tuned at any time during the operation of the receiver. As discussed in Section 3.5.3 by tuning the receiver’s magnetic field we can make a map of receiver performance as a function of magnetic field strength.

The magnetic field of the electromagnet is generated by a wrapped solenoid of superconducting wire. A DC current on superconducting wire does not produce a thermal load on the KAPPa cryostat’s 4.2 K stage. However, the low thermal conductivity cryowires proceeding the superconducting magnet wire produce a larger heat load than all other receiver components combined, including the first stage low noise amplifier (LNA). Smaller currents can be used if the solenoid has additional wraps, but this increases the cost, complexity, and size of the electromagnet. Due to their physically smaller SIS junctions, the KAPPa SIS devices require a higher magnetic field strength than their SuperCam counterparts. The SuperCam electromagnets were machined using high magnetic permeability metals. After machining the electromagnet components are wrapped with superconducting wire. This process is expensive and time consuming and the result is an electromagnet that is several times larger than all the other pixel components combined. To make an array of \(~1000\) SIS pixels, with densely packed pixels, we must shrink the size of the magnet system.

2.3.2 Permanent Magnet Design

Using the feedhorn aperture as the maximum footprint size of a pixel, it was required that all of the pixel components fit into a \(6 \text{ mm} \times 6 \text{ mm}\) cell. Our original design called for the electromagnets to be placed in the same plane as the SIS device and the LNA. Even with strong rare earth neodymium magnets, the field strength
Figure 2.9: The permanent magnet system. The left side shows a solid model of the permanent magnet system. The permanent magnet is cylindrical commercially available magnet shown in blue. The magnetic field concentrators, (ping-pong paddles) are shown in orange. The narrow 'handles' of the field concentrators increase the magnetic field over the SIS junction. The right side shows the magnetic field vectors for the permanent magnet system simulated in CST.

and orientation required for the SIS device called for magnets that were too large to fit within the 6 mm × 6 mm cell. This issue was mitigated by using a ‘U’ shaped permanent magnet system (see Figure 2.9) that is installed from the rear of the pixel, the side of the detector block that was opposite of the feedhorn, as shown in Figure 2.8. With this improvement, the plane of the SIS device only needs to have two small holes that the legs of the ‘U’ shaped permanent magnet system pass through. We were able to make the legs of the permanent magnet system thin, allowing the legs to be placed closer to SIS device decreasing the required strength of the permanent magnet. This design has three advantages: 1) The desired decrease in pixel footprint size; 2) There was more room for different permanent
magnet geometries, allowing us to select from a wider variety of magnet sizes and strengths to provide the required field strength; and 3) The effects of a pixel’s magnet are minimized on neighboring pixels since alternating the magnetic field direction for neighboring pixels can further minimize the effects of neighboring magnets has on a given pixel.

Magnets are difficult to machine because they are brittle and the magnetized chips that form during the machining process will be attracted to any ferrous metal used in the machining process. Therefore, it is preferable for both cost and simplicity of design to utilize unmodified commercially available magnets that are a standard. To reshape the magnetic field for the KAPPa SIS device, we machine steel concentrators to make the legs of the ‘U’ shaped permanent magnet system, see Figure 2.9. The 1010 steel has a high magnetic permeability and can have its magnetic properties simulated in finite element codes. The final permanent magnet system design utilized a cylindrical permanent magnet and two steel field concentrators shaped like ping-pong paddles. The large, flat ‘paddles’ of the concentrators sandwiches the permanent magnet, while the two thin handles extend into the plane of the SIS device channeling the magnetic field to reconnect across the SIS device (see Panel C in Figure 2.8).

All of the iron yokes of the permanent magnets were created in a single batch. The small size of the paddles makes them difficult to machine using a traditional milling process because they cannot easily be held. To solve this problem we employed a machining technique known as wire electric discharge machining (EDM). After machining the profile of the permanent magnet field concentrators into a long single piece of iron, a high voltage wire was used to slice the profiled iron into the final concentrator shape. After slicing, a finished concentrator simply needs to be cleaned and deburr before being fitted with a permanent magnet and installed.
into a pixel.

This mechanical design allowed for us to shrink the footprint of the magnets in an SIS pixel but the cost of the permanent magnet system is considerably less than the SuperCam electromagnet counterpart. The machining cost of one SuperCam electromagnet is $300, the cost of one neodymium magnet is $0.15, and the machining of the entire batch of 50 permanent magnet yokes was comparable to the cost of one SuperCam electromagnet.

Using a combination of simulation and results from the electromagnet pixel tests we were able to determine the field strength need to suppress Josephson noise within the SIS device. However because the change of material properties, such as magnetic permeability, as the permanent magnet is cooled from 300 K to 4.2 K we decided to hedge our bets and design a magnet pocket in the rear of the detector that was over sized. In the event that more magnetic field strength was need than had been budgeted, we would simple design a new set of yokes for a larger stronger permanent magnet. This caused us to design a carriage for the permanent that would allow for larger magnet to be installed at a later time. Fortunately this issue did not arise during this experiment.

2.3.3 Permanent Magnet Simulation

To suppress the noise associated with the DC Josephson effect that is inherent to SIS devices, we provide a magnetic bias field that is 45° from a wall of the waveguide, in plane with the mixer, centered on the center of the waveguide. This maximizes area for the Magnetic Flux to pass through insulating layer in the SIS Junction. Typically this is done with a small adjustable electromagnet. Once an SIS detector is cooled and operating, the electromagnet can be tuned to determine the needed magnetic field to produce optimal performance. Unfortunately,
electromagnets are too large to be used with a pixel spacing of 6 mm. For our single pixel prototype we are capable of producing the magnetic bias field with a tunable electromagnet as well as with a permanent magnet. It is our goal to replace the electromagnets of our SIS detector with cheap permanent magnets smaller than the cell size of our pixels. A solid model of the single pixel with both the electromagnet and the permanent magnet installed can be seen in Figure 2.8.

Keeping the goal of a ∼1000 pixel array in mind, we chose commercially available magnets that can be interchanged to produce different strengths of field. For our field, we require a magnetic field of ∼600 Gauss. To occupy less space in the already crowded back block of the detector, we placed the permanent magnet on the side of the back block, opposite the SIS mixer. To get the field to the SIS detector, we used a small (0.5 mm diameter) steel yoke. The yoke and the permanent magnet make a ‘U’ shape where the permanent magnet is in the bend of the ‘U,’ the yokes are the straight legs of the ‘U,’ and the SIS detector is at the top of the ‘U’ in the reconnecting magnetic field, as shown in right panel of Figure 2.9.

In the setup described above, we can change the strength of the magnetic field by changing the separation distance of the yoke towers, by changing the strength of the permanent magnets, or using a permanent magnet with a different thickness or radius. From Fig. 2.10, we can see the magnetic field created by the simulation seen in Fig. 2.9 evaluated on the line that crosses the SIS receiver. Looking at the plot, we notice that the magnetic field is fairly flat around the SIS device in the center of the curve. This flatness makes the yoke and permanent magnet difficult to misalign. In turn, this makes the magnetic bias easy to install for a ∼1000 pixel array. Placing the magnet in a different area than the SIS mixer and LNA also helps to free up space in a 6 mm × 6 mm cell.

Microassembly of the components in a heterodyne pixel is a tedious process,
thus the alignment of ∼1000 permanent magnets in a future heterodyne arrays should be as simple as possible. With this in mind, the permanent magnet system should produce a magnetic field that change by less than 10% for any possible misalignment of the SIS device or the permanent magnet system. We used Computer Simulation Technology (CST) EM Studio to evaluate various geometries of permanent magnet system with the goal of designing a magnetic field is as flat as possible in the region of the SIS device.

![Figure 2.10: The magnetic field for the permanent magnet system across the SIS device. The right image shows the geometry of the permanent magnet system. The magnetic field strength is measured across the path of the blue line that connects the left and right field concentrators. The center of this path is the location of the SIS device at 0.85 mm from each concentrator. The plot on the left side shows the magnetic field strength along the path of the blue line in the right image.](image)

During the CST evaluations, our goal was to having a uniform magnetic field in a region around the SIS device that was 1.5 times that of the maximum mechanical tolerances for both the alignment of the SIS device as well as the permanent magnet system. The maximum mechanical tolerances for the microassembly of the SIS device and permanent magnet system were less than 0.05
mm. Figure 2.10 shows the field strength between two concentrators on a path that goes through the SIS device. We were able to achieve desired field flatness for a length scale greater than that of the mechanical tolerances by extending the yokes above the plane of the SIS device and LNA by 1.0 mm. The field flatness could have been improved by making the area of the field concentrators leg larger for the side facing the SIS device, but the conclusion of our trade study indicated that it was more desirable to have a thin yoke leg that could be brought as close to SIS junction as possible.

Our simulation results informed us of the gradient and the approximate field strength that we could expect based on the data sheets from our commercially available magnets. Simulation helped to choose an effective geometry that would be tolerant of misalignment. After finding an effective geometry using CST evaluations we would then determine what could be mechanically designed. Mechanical design and magnetic field design required many iterations before all design criteria were achieved. Small pieces and tight tolerances are good for the magnetic field design but can be expensive to machine. A future design might benefit from much larger pass-through for concentrators in the detector block which lead to smaller, shallower features in the horn block. This would make the detector block features easier to machine while still maintaining the tight mechanical tolerances by capturing the tops of the magnet concentrators in the horn block. Simulation was done under the expectation that real magnets would exhibit larger than advertised field strengths that varied from magnet to magnet.
Chapter 3

KAPPA PROCEDURES AND TESTING

3.1 SIS Device Testing

The technologies in the THz regime progress slowly: our detectors do not benefit from the commercial electronics revolution like other frequencies in astronomy. In THz astronomy, the application drives detector development. In the future, a single THz pixel will be completely integrated much like charge-coupled devices (CCDs) are today. Eventually, transistors will be able to operate at THz frequencies and THz engineers will no longer be required to down-convert the signal before amplifying. Hopefully, by this point in our career, we will have moved to higher and more difficult frequencies.

The mission of KAPPa is to develop technologies needed to make an array of $\sim 1000$ THz pixels. A large part of that mission has been to further integrate the SIS device with other components of the pixel. The highest frequency receivers for the future of astronomy will likely have their components grown directly on to the detector where they will be used. KAPPa has progressed THz technology in that direction, but full integration is still out of reach. In this section we present our hard-earned lessons from working with SIS devices.

3.1.1 SIS Device Handling

One impediment to realizing coherent $\sim 1000$ pixel array receivers is the selection of the best performing SIS devices. Before we could select best SIS devices, the KAPPa team was required to develop a procedure to handle SIS devices with
minimal loss. This proved to be extremely difficult. The SIS devices are physically small and the silicon parts of the device can crack and destroy the device. The sensitivity of the KAPPa devices to electrostatic discharge (ESD) has proven to be far greater than that of the SuperCam devices previously handled by our group.

The ESD sensitivity was a particular problem for our lab located in Tempe, Arizona. The arid desert air, with low water content, is desirable for transmitting signals through the atmosphere. However, in low humidity regions, static charges can build up on surfaces. In our lab, these charges will quickly exceed safe levels for the SIS device. To combat the problems of ESD, we took the normal approach of simply grounding every surface ESD mats, wrist and leg straps, and protective clothing and shoes. Unfortunately this had almost no effect, only one in five devices would withstand the testing procedure on a particularly lucky day when no devices were no lost during mounting or removal. While complaining about the difficulty of ESD sensitivity, Dr. Phil Christensen, a geology professor at Arizona State University, recommended that we try testing our devices in a humidity controlled room. Specifically, he recommended that humidity be above 40% or when polar water molecules in the atmosphere collide with materials with static charge and strip the excess electrons away before the charges build to unsafe levels. This turned out to be the key to handling the KAPPa devices safely in Arizona. This point is belabored because the SuperCam SIS devices were assembled in the same lab but did not have ESD sensitivity of the high-frequency-smaller-SIS-junction KAPPa Devices, and it is likely that all future devices at frequencies of KAPPa or higher will encounter the same sensitivity. After switching to a humidity control room for the KAPPa device assembly, not a single device was lost.

Once the SIS devices were fabricated and delivered, they are easily freed from the wafer but can be subsequently very difficult to handle. The devices are
surprisingly flexible and durable but are still easily destroyed during handling and mounting. The forces needed to safely handle the devices are far below the forces normally exerted by human hands. For the size scale of the KAPPa SIS devices, electrostatic forces dominate and gravity only adds a small perturbation. ASU assembled all of the pixels for SuperCam. We learned to assemble devices from the person who pioneered the technique for SuperCam, Hamdi Mani. The technique for handling the SuperCam devices was to float the devices in isopropyl alcohol and to use a grounded paintbrush with most of the bristles removed to pick up the devices. The handler of the paintbrush, through the mastery of electrostatic forces, would then try to scoop the device from the alcohol, allow surface tension to hold it to the paintbrush, and then deposit in a working space that contains more isopropyl alcohol. As the isopropyl evaporates from the working space, the device can be moved; at this stage, micron level alignment could be done. This work is extremely time consuming and tedious, and is unlikely for a randomly selected graduate student to excel at this procedure. At ASU, only two people learned how to do this procedure, making it one of the least scalable aspects of KAPPa.

Mr. Mani also mastered the technique of picking up the devices with a small piece of gel (Gelpack #0). This gel was a packing material that the small lump circuits components are shipped in, a relatively abundant resource in our lab. Using the gel on the end of a grounded needle, one could pick up the components and place them without the need for isopropyl alcohol. In a high pressure performance, Mr. Mani could pick up the devices with the gel and place the directly into electrically conductive epoxy, saving about 15 minutes of work as opposed to the paintbrush placement. We found these techniques too risky for final placement but very effective low risk movements. Because of the stickiness of the gel, it can be very hard to remove the devices and control the orientation by which they stick to
the gel. In its current configuration, the gel is a tool for the advanced packager.

Fortunately for our successor, there are several ways in which the handling of SIS devices can be improved. The major flaw of the paintbrush is that it is not easy to grab or let go of the SIS device, increasing the time needed to work on the installation. Ideally, someone might want tweezers to handle the device, however, humans lack the dexterity and control to hold an SIS device without breaking the silicon substrate. With some limited success, we were able to construct a pair of paintbrush-tweezers. However, the bristles stuck out at too great an angle to properly pick up devices. The density of the bristles was low enough that the bristles pass each other with very few touching. To pick up a device, lay the bristles parallel to the surface then squeeze the tweezers to scoop up the devices, while keeping in mind that many times the bristles will flick the device out of the view of the microscope. This version of the paintbrush-tweezers works fine when the bristles are wetted with alcohol and saves time picking up the SIS devices, constituting a significant increase in the easy of handling when compared to a regular paintbrush. Longer bristles might help this problem, but longer bristles may also be more likely to store static charge and would be harder to control.

Again we employed the effective strategy of complaining to geologists about the difficulty of our work. During a private communication with geologist Nathaniel Borneman, he recognized the similarity between handling SIS devices and picking up and sifting through small grains of rocks and minerals. Every first year student in his lab learns to do this process, as it is essential chemically testing the individual components of a complex conglomerate. Aside from the usual array of tweezers, students in this lab can also use a vacuum tool to pick up and place small grains. This vacuum tool is control via a foot pedal leaving the user’s hands free to operate the microscope and vacuum. Previously this idea has been rejected for SIS device
handling because the air flows through the nozzle of the vacuum can build up charges on the SIS device from ions in the air. This is a reasonable fear, but the vacuum may be the best option for 1000 pixel array. To mitigate charge build up, we propose to modify the vacuum nozzle to be grounded so that excess charge can be dissipated. Additionally, this work should be completed in humidity control room where the water in the air is helping to dissipate electrical charge.

The final improvement came at the cost of detector yield. The reason that SIS devices cannot be picked up with tweezers is because the silicon substrate is fragile compared to the forces of human hands. However, the SIS device can be picked up by its gold beam leads, which is what we did on when we were trying to remove a device we successfully tested (see Section 3.1.2). Unfortunately, this would often mangle the beam leads making them unsuitable for installation in the KAPPa pixel. We often used parts of the beam leads to make small adjustments to the position of the SIS device during installation. It is our hope that a future batch of SIS devices could be constructed with extra gold tabs connected to the beam leads by a thin strip of gold. These gold tabs could be used to handle the SIS device, pick it up directly with tweezers, place it, fasten the device in place using a combination of epoxy and/or Crystal Bond, and then the extra tab could simply be torn away leaving a perfectly mounted device. The principal disadvantage to this technique is that the handling tab will take up area on the substrate where the SIS devices are grown, decreasing the maximum number of devices per substrate. However the maximum number devices is unimportant if they cannot be handled in an inexpensive and efficient manner. Going to larger scale, simplicity is tantamount.
3.1.2 SIS Device Testing and Selection

To select the best SIS devices before mounting them within the mixer block, we perform a DC sweep of the device at ambient temperatures and at the working temperature of 4.2 K. To test the devices at 4.2 K, we employ the use of a dipstick which consists of three major elements: the chip carrier in which the SIS devices are mounted, the junction box that provides ESD protection when switching between devices, and a stainless steel tube that houses the wiring as well as provides a thermal break between the junction box and the chip carrier (see Figure 3.1). The SIS devices are mounted to the chip carrier using electrically conductive epoxy. The junction box contains considerable capacitive filtering, resistive chokes, and channel grounding when devices are not being tested. Chip carriers would typically start with all devices at nominal DC resistance at room temperature but, during the transition to 4.2 K, many devices become electrically open. The diminishing yield of devices when using dipstick measurement technique suggests that this method would not be effective for finding large numbers of high performing devices for a large array.

Selection of ~1000 SIS devices by dip testing is impractical. Future array developers should consider a method to test the SIS devices in-situ before they are ever removed from the wafer. In theory, this can be done using a cryogenic probe station, however cooling the entire SIS wafer to the working temperature of 4 K will require innovative probes or advanced baffling to shield the wafer from 300 K light. Within the KAPPa, experiment it was noticed that many of the best preforming device where colocated on the device wafer. An advanced dipping procedure that is outlined in the current section could be employed to test a representative sampling of devices. The neighbors of the best performing SIS devices could then be
Figure 3.1: In the left panel we see a picture through a microscope of a successfully mounted SIS device for testing with the dipstick. On the right we see the same chip carrier mounted on the end of the dipstick prior to being submerged in liquid helium. A successfully tested SIS device must be cut free from the electrical conductive epoxy using a scalpel and very steady hands.

preferentially installed directly into the kilopixel array without ever being tested. This will cause a reduction in array performance, but this may be an acceptable trade when considering the price of a cryogenic probe station.

When the SIS devices survived the dipping process, they are removed from the chip carrier by cutting away a small portion of the device’s beam leads with a scalpel under a microscope and then stored. An SIS device that is ready to be removed can be seen in Figure 3.1. The removal process further reduced the yield of usable and measured devices. We place devices with alcohol using a grounded paintbrush (as discussed in detail in Section 3.1.1). Once a device has been roughly placed and the alcohol has evaporated, precision placement is then done with a
single bristle of a paintbrush. The results presented in the present work are obtained using a device that was tested on the dipstick and successfully remounted in this manner.

Almost all of the SIS devices tested were either electrically open or shorted, this is likely due to the lack of a humidly controlled room at the time of testing, as discussed in Section 3.1.1. However 3 months of testing did result in 19 devices that exhibited the characteristic of an SIS junction during the DC sweep of current and voltage. The current-voltage sweep of all 19 devices can all be seen in Figure 3.2. For simplicity, each device in Figure 3.2 has the same current and voltage scale. Devices were first rejected if the bias voltage needed to cause normal electron tunneling was below the expected range of 2-3 mV, denoted by a diamond symbol in Figure 3.2, see for example ND41 (top-left corner panel) and OF28 (middle panel).

For each device in Figure 3.2, we calculated the device resistance in 3 regions which is displayed for each device in the legend of each panel. The exact region we used to calculate the resistance is donated by a black line that overlays the colored line of each devices IV curve. Region 1 uses a dash-dot line and is the resistance of the SIS device before the tunneling voltage. Region 2 is the smallest and is denoted with a solid line. Region 2 is also referred to as the step resistance and is between that tunneling bias voltage and critical current. Finally, Region 3 is the non-superconducting resistance of the device for currents higher than the critical current.

Ideally, a quality factor for each SIS device can be determined from the ratio of impedance of Region 1 to that of Region 2. Histograms of the impedance in Regions 1, 2, and 3, the quality factor, as well as the tunneling current, critical current, and voltages for the 19 tested devices in Figure 3.3. Region 1 is ideally infinite resistance, or a horizontal line in Figure 3.2, and Region 2 is zero resistance,
or a vertical line in Figure 3.2. However we do not see the expected behavior for an SIS device in these tested curves, with the sole exception of device NB27. In the absence of a magnetic field, we expect the SIS devices to exhibit the DC Josephson effect (see Section 1.3.3): the devices conduct current regardless voltage. We observed the DC Josephson effect for all devices when installing them into the KAPPa receiver, but only device NB27 exhibited during the dip tests. Armed with a greater knowledge of superconduction theory, it seems that our devices were being current-pumped by a RF source inside the liquid helium tank. This source was likely the tank itself with the warm parts of the dipstick. All of the device were exposed to the background THz radiation as the device were tested uncovered, exactly as displayed in Figure 3.1. The tank hypothesis is further validated as many of the device (NF25, ZZ70, OB46) showing Shapiro steps, indicative of pumping RF radiation. It is possible that this RF signal was picked up on the wire of the dipstick, the original design only had RF choke on the switching box and not near the chip charier.

Regardless of the reason, the resistance of Region 1 was not a valid for use in determining a quality factor. This can be seen in the large spread in device quality factors presented in Figure 3.3 as compared to the tighter grouping of Region 2 impedances. Devices were instead selected based on having the lowest resistance in Region 2 in Figure 3.3, where the lowest Region 2 impedances were below 10 Ω. Later in the KAPPa experiments, we noticed an increased contact resistance of the SIS device due to delimitation if the electrically conductive epoxy (discussed in Section 3.1.3). This was discovered well after device dip testing, so we suspect that the additional issue may have added to the resistance of Region 2 in some cases.

Our current dipping process does not include the use of a magnetic field to suppress noise from the DC Josephson effect. We select the best devices by...
Figure 3.2: Above are all the successfully tested SIS devices, measured at 4.1 K with ambient magnetic field. Current and voltage scales are the same for all devices. Each device’s tunneling voltage and critical current are denoted per device with a diamond and circle, respectively. The resistance of three regions is displayed for each device where Region 1 is the resistance before the tunneling voltage, Region 2 is the step resistance between that tunneling bias voltage and critical current, and Region 3 is the non-superconducting resistance of the device for currents higher than the critical current.

measuring the resistance of the step region of the IV curve and choosing devices with the lowest step resistance, Region 2 in Figure 3.2. Typically, devices are chosen using measurement of the quality factor (the ratio of the subgap resistance to the normal resistance), but the lack of magnetic field in our dipping setup prevented accurate measurement of the subgap resistance. The steepness of the IV step was
chosen as a reasonable proxy.

A magnet could have been installed to suppress the DC Josephson effect (see Section 1.3.3). As it later became apparent from working with the SIS device as a receiver, a large magnetic field can suppress the DC Josephson effect without an integer magnetic flux quanta, as is needed for low magnetic fields. Operating at a high magnetic field makes a noisy receiver, but it works well to measure the DC properties of the SIS device without noise. If testing devices, we could place the magnet on the chip carrier and within a few tries we could find a range of field that should work for the small variation in SIS junction size.

**Figure 3.3:** Histograms of the measured properties seen in Figure 3.2.
Our chip carries were made from pieces of microstrip that have been cut up with scalpels (see Figure 3.1). This allowed us to test, but it was extremely difficult to mount and remove the devices. For example, the microstrip would be destroyed after about 5 cycles and need to be replaced. When preparing to select 100s of devices for a 16 pixel array, a printed circuit board (PCB) was designed to make device mounting and removal an easy process. The new chip carrier PCB can be seen in Figure 3.4 and it solves several of the flaws of the previous system. The new PCB is designed to maintain the 4-wire measurements of the original chip carriers while having on-board surface mounted resistors (blue packaged item in Figure 3.4) as protection against large voltages and ESD as well as a space for an optional capacitor or diode in parallel with the SIS device.

This new chip carrier design eliminates the need for wire bonding as the gold vias are designed to be soldered to the available posts of the old chip carrier, reducing the technical skill needed to test SIS devices. Advanced users of epoxy may notice a waxy film that develops on curved epoxies, this was the case with the electrically conductive epoxy used in KAPPa experiments. The waxy film only affected the removal of SIS devices from the the old chip carriers. The waxy film would seep from flexible gold beam lead to under the inflexible silicon body of the SIS device. Then, the fragile body of the SIS device was effectively glued to the microstrip substrate (reference Figure 3.1 to see the silicon body of the SIS device over the white micro strip substrate). The devices could be removed after solvents were used to desolve the wax, a time consuming process that resulted in the loss of several tested and well performing SIS devices. The new chip carrier design eliminates this issue by having the body of the SIS device fully suspended above the PCB’s dielectric substrate. Because the new chip carrier PCB’s traces were manufactured with a ‘2 oz gold’, a standard that deposits a 2.8 mils (71.1 µm) think
layer of gold, the 5 µm thick SIS device can be easily suspended above the PCB (see panel A of Figure 3.4).

Figure 3.4: A solid model showing the a new PCB designed specifically to test the KAPPA SIS devices. Panel A is the most zoomed-in image showing the SIS resting comfortably on the traces of the PCB, while panel B is panned out to show the fully assembled 4-wire test circuit for a single SIS device. Panel C show the entire PCB board without the SIS device, blue protection resistors, and yellow protection diodes to show the PCB as it would be delivered from the circuit manufacturer.

3.1.3 Device Delamination

The ground beam leads of the SIS device (denoted with a white box in Figure 2.3) are both electrically and mechanically attached to the detector block using electrically conductive epoxy. Here we may have benefitted from using a larger and rougher surface to attach the ground leads of the SIS derive. The electrically conductive epoxy we used slowly delaminates as the device was thermally cycled from it operating temperature at 4.2 K to the ambient temperature of the lab.
Physically, we could measure and track this process as a function of the number of cool downs by measuring an increasing series resistance across the SIS device. This of course led to degraded performance. During the first installation the SIS device, we operated under the incorrect hypothesis that applying the smallest amount of electrically conductive epoxy would combat thermal differential expansion by having relatively small linear distances for expansion. In practice, we found that it required depositing large amounts electrically conductive epoxy that totally filled the ground lead feature to stop the delamination of the SIS device. The future SIS device and pocket design may include a 100 $\mu$m by 20 $\mu$m crush zone where the electrical the conductive epoxy and the ground beam lead are tightly sandwiched between the detector and horn block to prevent delamination.

3.1.4 KAPPa Pixel Assembly

Assembling a KAPPa pixel requires glueing a number of components with electrically conductive epoxy. Aside from the SIS device, the minimum components for a single KAPPa pixel includes the WBA23 LNA, an ion beam milled inductor, three filtering capacitors, and one capacitor used as a landing pad for the SIS device’s IF beam lead. It is very difficult to get the epoxy in the bottom of the 250 $\mu$m by 250 $\mu$m by 250 $\mu$m trench (and not on the side of the trench which can short the capacitor). As a result of tests with the early version of the detector block, the KAPPa pixel components are mounted on open steps so that each component can be installed without collisions with trench walls (see Figures 2.3 and 2.5). Both the 16 pixel design and single pixel components were designed to be installed by a right-handed person under a microscope.

By keeping in mind the effort it would take to build a 1000 pixel version of our array, we focus on improving our assembly techniques. Here we compare two
assembly techniques developed at ASU. We begin with a capacitor already mounted, where the IF beam lead of the SIS mixer is already in place. The epoxy that holds the capacitor must be cured before the SIS mixer can be mounted.

Mounting the SIS mixer requires the use of alcohol to float the fragile device. It cannot be picked up with tweezers and must be handled with a paint brush using the surface tension alcohol (see Section 3.1.1).

The KAPPa team developed a new innovative technique for installing the SIS device, which was developed to eliminate the use of Crystalbond, used in the SuperCam experiments. Once the capacitor is mounted and the IF beam lead of the SIS mixer is in place and cured, we use metallic epoxy to mechanically attach the alignment tabs to the block. In addition, we also use epoxy to electrically connect the capacitor and the IF beam lead. In this way, we have great mechanical and electrical conductivity. We learned to align the SIS mixer by cutting a piece of gel from Gelpack #0 and putting it on a probe tip, as was discussed in Section 3.1.1. Gelpack #0 is a sticky substrate that is used for the transportation of small electronics that would otherwise be damaged or lost during transportation. Using a razor to cut a corner from the Gelpack, we coax a sliver onto a probe tip, which is like a very fine needle. Once the gel is on the probe tip, we can pick up the SIS mixer without damaging it and align it in as many dimensions as we can move the probe tip. The epoxy pulls the SIS mixer from the gel-coated probe tip. Since we have eliminated the use of Crystalbond, we can use heat at the wire bonding station and heat epoxy to reduce the cure time. This, and subsequent improvements in assembly techniques, will be integral for the eventual assembly of a ∼1000 pixel array.
3.2 Magnet Preparation

In this section we go into the detailed procedure of measuring the flux density of the magnets and testing them magnets prior for use in KAPPa.

3.2.1 Measuring Magnetic Field Strength

To measure the magnetic field strength of the permanent and electromagnets, we used a commercially available gaussmeter, namely the 5180 Gauss/Tesla Meter by F. W. Bell. The entire device consisted of a handheld meter with digital output display and a selection of detachable probes. The field strength was measured using the transverse Hall probe, which allows us to take measurements of the magnetic flux density between magnetic poles. Correct orientation of the encased Hall Effect sensor within the transverse probe was necessary to achieve flux measurements near and in-between the field concentrators of the permanent magnets. The probe itself was a thin plastic stem only 1.14 ± 0.1 mm thick housing the small sensor which is located 0.85 ± 0.1 mm from the tip of the probe. The diameter of the active sensor area for the transverse probe is 0.38 mm as given by the manufacturer. This active area of flux being measured is significantly larger than the true area of flux passing over the 1µm × 1µm SIS junction.

The measurements of magnetic field strength required keep the Hall element within the gaussmeter normal to the incoming flux lines. It was important to move the probe as little as possible during measurement to minimize contamination from ambient magnetic fields. An aluminum precision motion stage was required to keep the stage from being magnetized. It proved extremely challenging to measure the magnetic fields of the small magnets used in the KAPPa receiver and the region where the SIS junction is located is even smaller. The regions that we were
Figure 3.5: Microscope image of the electromagnet and gaussmeter. A sweep of the magnetic field is performed while the electromagnet is stepped horizontally on a precision platform compared to a fixed transverse sensor probe of the gaussmeter. The direction of probe sweep is indicated by a black arrow, and the engraved lines on the moving platform mark the distance from concentrators at 1 mm intervals.

attempting to measure were much smaller than the characteristic size of the probe used for measurement. The only way to get around this scale issue was to make many measurements at very small steps and then repeat this process a statistically significant number of times.

Using a microscope and the engraved scale on the platform, the probe and concentrators were brought to a specific distance, which was often less than or equal to 1 mm apart. When the field concentrators are attached to our cylindrical neodymium magnets, there is only $1.7 \pm 0.1$ mm of separation between the concentrators. This presents a challenge since the sensor probe thickness, compared to the spacing of the concentrators, leaves little room for movement. Very few data
points can be gathered in this space because the probe can collide with the magnet, potentially causing damage the sensor probe, or simply changing the orientation of the probe with the magnet. To give an estimate of values for the magnetic field strength between concentrators, a single data point was measured for several permanent magnets by centering the probe at the location between the poles where the field would be redirected over the SIS junction in the final pixel. However, this would not be enough to give us confidence in the permanent magnet. A second set of measurements was taken at the tip of the iron concentrators. Here the probe would scan laterally from one concentrator to the next to determine the relative field strength and uniformity, which should exhibit similar behavior as the field lines between concentrators as it does a known distance away. Once the full distance was traveled, the magnet was sent back to starting position and the sweep was repeated several times. This was done for consistency and to verify reproducibility. Four identical permanent magnets were measured in this fashion. Figure 3.5 shows a microscope image of the probe sweeping procedure for the electromagnet.

The electromagnets were tested in a similar fashion. The base of the electromagnet was bonded to the platform and the leads were connected to a variable power supply. The power supply let us measure the magnets for a range of currents starting from 45mA and leading up to 80mA. The typical operating current of the electromagnet, when installed in the KAPPa receiver, was 65mA. The chosen spread in current allowed us to observe the electromagnet at range of field strengths that could be expected in the receiver. The concentrators on the electromagnet were spaced further apart than the ones in the permanent magnet. The electromagnet had one concentrator that was much shorter than the other (seen in Figure 3.5) creating an asymmetric magnetic field strength (seen in Figure 3.6). Because of the asymmetry, the field strength at the center of the concentrators varied more than the
Figure 3.6: The average flux density of a set of three identical permanent magnets (data points with filled circles) over a single sweep between concentrators. The average flux density of an electromagnet with current between 45mA – 80mA is shown with filled squares. The distance is measured from the center of one concentrator to the next, while the separation of the sensor to the concentrators is 0.85 ± 0.1 mm.

permanent magnet systems adding to the difficulty of comparison. The asymmetry of the electromagnet affects the flux density at the ends of its concentrators, which makes it difficult to compare the uniformity of the permanent magnet systems. Figure 3.6 shows that the permanent magnet system exhibits a symmetric behavior with a uniform central field, not unlike simulation. Because the permanent magnet demonstrated generally higher flux densities then required, we attempted to lower the magnetic field of the permanent magnets as discussed in Section 3.2.3.
3.2.2 Magnet Testing Procedure

Before each measurement of the permanent magnet, the alignment of the probe magnet system was checked to make sure nothing has been moved out of place. The transverse probe was tilted vertically within its holder clamp, preventing any rotation or horizontal translation of the probe, then the probe was fit into the zero flux chamber. The zeroed gaussmeter would be able to ignore any ambient fields and display only readings from our magnets. As per manufacturer specifications, a 15 minute warm up time was observed before measurements to allow for maximum accuracy. During this time a calibration sweep was conducted to verify that nothing would obstruct the movement of the magnet and that the sensor was indeed reading the flux lines. One full sweep consists of moving the platform 2.5 ± 0.1 mm. This is the distance between the centers of one concentrator to the next and results in the equivalent motion of moving the sensor probe in a straight line from one concentrator to the other. Once the full distance was traveled the magnet was sent back to starting position and the sweep was taken again several times. This was done for consistency and to verify repeatability. Four identical permanent magnets were measured in this fashion.

The moving platform of the magnet stand was interchangeable, so a second one was used for the electromagnet. The electromagnets were attached to the platform using the same crystal bond method as in the permanent case. The base of the electromagnet was bonded to the platform and the leads were connected to a variable power supply. This power supply let us measure the magnets for a range of currents starting from 45mA and leading up to 80mA. The typical operating current of the magnet would be 65mA so the chosen spread in current allowed us to observe the electromagnet not only at the best case scenario for operation but with the
potential for fluctuations in power. The concentrators on the electromagnet were spaced further apart than the ones in the permanent magnet system causing a sweep to displace the platform further. The geometries differed slightly as well and one concentrator was shorter than the other, meaning the field strength at the center of the concentrators varied more so than in the permanent magnet case.

We attempted to measure the electromagnet using the same procedure as for the permanent magnets, but the gaussmeter probe and the electromagnet were incompatible with the permanent magnet procedure due to the geometry. This issue highlights the need for a smaller gaussmeter probe for a future project. Despite difficulties, a relative comparison was still possible. After determining the electromagnet current needed for the ideal magnetic field inside the KAPPa device by experiment, we removed the electromagnet and provided it with the same current to measure the strength of the ideal magnetic field.

The scanning measurements were carried out by keeping the sensor probe in a fixed position while moving the magnets laterally with respect to the sensor. Initial attempts consisted of moving the magnets by hand over equally spaced intervals, which proved wildly inaccurate. The later attempts involved machining a precision magnet stand out of aluminum, which allowed us to move the magnet in smooth uniform motions. One aluminum block at the center of the machined stand served as a mobile platform that could be moved by the turn of single screw. Each full revolution of the screw would displace the platform by $0.45 \pm 0.1$ mm. The magnet was attached to the moving platform via the same adapter housing which would be used to fit the permanent magnet to the pixel block. Since the actual neodymium magnet is much smaller than the electromagnet an adapter for the permanent magnet was made to give it similar dimensions as the electromagnet. This engineering allowed for the interchange of the different magnets. Crystal bond
adhesive was used to place the adaptor in the moving platform.

Once a permanent magnet was inserted into the adapter, the sensor probe was aligned with either the magnetic field concentrator or the neodymium magnet such that the center of the probe was in the same plane as the center of the concentrator. Using a microscope and the engraved scale on the platform, the probe and concentrators were brought to a specific distance, which was often less than or equal to 1 mm apart. After alignment was achieved the only intentional motion was that of the lateral displacement of the magnet platform and the vertical raising and lowering of the probe stand which was necessary only when replacing magnets into the adapter or zeroing the magnetic field detected by the probe.

3.2.3 Dialing the Required Magnetic Field Strength

Permanent magnets can have their magnetic field reduced by heating them to their Curie temperature ($T_c$) for small amounts of time. In this way we are able to purchase magnets that have a stronger field than required and heat them until they have exactly the desired field. The technique was also used to demagnetize any of the iron field concentrators that may have been in too strong of an applied field. The $T_c$ for our cylindrical neodymium magnets was 310$^\circ$ C. An ordinary hotplate was used for the procedure of weakening the magnets. Once the hotplate was warmed to $T_c$, a single magnet was placed at the center. With constant heating for only 5 seconds at $T_c$, the flux density of the heated magnet was reduced to 60% of its original value. Following the first successful KAPPa test using a permanent magnet a second permanent magnet was heated in this same way and brought to 63% of the strength of first magnet.

Prior to the measurement of the electromagnet field strength, several sets of permanent magnet systems were constructed. Each system has a different magnetic
field strength corresponding to a different round of heating at the magnet’s $T_c$. We then compared that field strength of the set of permanent magnet systems to the measured electromagnet field under typical operating currents. Based on our observations from sweeping the electromagnet (discussed in Section 3.5.3), we know that there is a zone of infinite system temperature that is seen for weak magnet fields. Therefore, it was prudent to select a permanent magnet system with a stronger field than that what had been determined to be ideal at the lab bench.

### 3.2.4 Permanent Magnets For Future Arrays

The gaussmeter used to measure the flux density of the magnets needs to have a probe that is much smaller. The probe calculated the average magnetic field through the area of the Hall element. However, the area of the probe was much greater than the change in magnetic field that we were attempting to measure. A future SIS focal plane array might consider a custom gaussmeter for the purposes of determining magnetic field strength. The manufacturer also carries an ultra-thin version of the transverse probe that is half as thick which may allow for more effective readings. Inexpensive alternatives include making a simple circuit with a commercially available linear Hall Effect sensor for a homemade gaussmeter that is less robust. These Hall sensors cost on average less than $5.00 for an individual component and can be driven with 5V from a power supply. The potential drawback of using them is that the range of flux density they can measure is significantly lower than the current gaussmeter used, however the measurements for the magnet concentrator system thus far are within their range. The Hall element may also be deeper within the sensor making it difficult to compare measurements to the existing gaussmeter because the location of the flux being measured will not be the same for either probe. Benefits include that these Hall sensors come in
different geometries that may be more effective at measuring flux density directly between the concentrators. Hall effect sensors like the model MLX90215 with Hall elements 0.2 mm in diameter have smaller active areas for measuring the flux density. The small size of these Hall effect sensors and their relative low cost would also allow future teams to make multiple gaussmeters and experiment with different configurations for positioning the probe relative to the magnets. A similar moving platform like the one used to shift the magnets laterally could be used to move the probe closer to the magnets and set a more accurate separation distance between the probe and magnets. This could be vital to pinning down exactly what flux density is passing over the junction.

A future THz array of 1000 pixel would benefit from first making a single pixel prototype for reasons of permanent magnet calibration. Despite how well our permanent magnets worked, this is due to the extremely conservative path we and our plan to use greater than 3 integer flux quanta for suppression of the DC Josephson effect. With a gaussmeter capable of measuring smaller changes in magnetic field and a firmer understanding of how the magnetic field changes magnetic permeability at 4K, the low noise minima could be assessable for future array builder. This may seem contradictory, however, given the equipment the would need to be purchased for such a large scale project, additional carful measurement and simulation could allow for a 30 K decrease in the system noise by using the first minima from a single flux quanta. If there is a fear that the system could work in a single pixel but not an array, than the array can simply be build so that a more powerful magnet can be installed in some or all pixels to achieve the 3rd or 4th flux quanta for noise minima. We would certainly be willing to try.
3.3 The KAPPa Test-Bench and Characterization Database

At the beginning of the KAPPa experiment, we identified many technological advances we would need to develop to facilitate the creation of a ∼1000 pixel array. However, at the end of the KAPPa experiment, rapid and automated receiver characterization is one of the signature technologies in which we invested. Because the KAPPa receiver was the first major project of the THz group at ASU, the normal route of comparing the KAPPa receiver to an established working receiver was not available.

The characterization procedure for KAPPa’s predecessor, SuperCam, involved waiving a paddle of Eccosorb cooled with liquid nitrogen in front of the cryostat, reading a value from a powermeter, and then changing the receivers settings. This was all that was needed when working with a well understood receiver. The KAPPa team did not set out to create a fundamentally new receiver test bench and characterization database. However, once it was constructed and implemented, it wasn’t long before the receiver was able to outperform the graduate students that were used to determine the characterization. It hardly seems visionary considering the push for greater automation in all sectors of human life, but the KAPPa instrument required an entirely new data infrastructure in order to be fully and automatically characterized.

Looking back through at the first data taken with the KAPPa receiver, the problem with our LO was visible and could have been diagnosed. However, the LO was a purchased piece of equipment and less likely to be a problem then a fundamentally new and untested receiver. So the KAPPa team was forced to take the long path to victory, and prove that receiver was working correctly. Human characterization by the waving-the-cold-paddle method utterly failed to identify the
complex parameter topology where the reliever was operating ideally. It seemed
that the receiver was working at random but repeatable settings. It was clear a new
strategy would need to be implemented to provide data that could diagnosis this
complex issue.

KAPPa characterization flow is simple, repetitive tasks such that receptive
decision making are removed from the user’s control. The database is fundamentally
flexible, making it easy to add new measurement types or experimental equipment.
When queuing the database of all KAPPa data ever taken for a specific analytic test,
each test knows what data it needs and ignores incomplete data sets. When adding
test equipment, or a new analysis technique, the KAPPa experiment requires no
additional integration to work within the existing system. The user of the KAPPa
receiver is now solely responsible for maintaining and constructing test equipment
and procedures while the KAPPa receiver can automatically characterize and tune
itself during round-the-clock operation. Reports were sent from the instrument to
the user by email. By creating this characterization, we have drastically reduced the
number of human hours spent characterizing a receiver, allowing for more complex
receiver operation. It is this improvement, more than any other, that will make it
possible to build ~1000 pixel array. It sets a powerful standard new standard in the
THz by removing humans from repetitive tasks, no matter the complexity.

3.3.1 Receiver Tuning Parameter Space

Full characterization of the single pixel system requires an investigation of a
large number of parameters. To enable correct interpretation of the single pixel
results with the permanent magnet, we first must fully characterize the pixel with
an electromagnet installed. The block diagram in Figure 3.7 shows the measurement
equipment and control systems used to explore the parameter spaces of magnetic
field strength, LO power, LO frequency, IF bandwidth as well as the IF amplification chain and bias electronics. However, before characterizing the pixel we must first ensure signal integrity on the IF chain.

Radio frequency interference (RFI) in the IF band of 0.5 to 5.0 GHz decreases the sensitivity of our receiver. To combat the problem we deploy a chain of filtering capacitor on all of the direct-current (DC) bias lines entering the cryostat. These filtering capacitors were placed in three key areas: immediately after the first IF amplifier (see Figure 2.3), at the wiring harness on the 4 K stage, and on the outside of the hermetic military connector taking the DC lines into the cryostat. While all of these measures improved performance, the greatest single improvement was made by enclosing the hermetic military connector on the outside of the cryostat in a small aluminum box, with in-line filtering capacitors for the DC pass through-in to the box. Currently, we have no detectable RFI remaining within the IF band.

Ultimately, we need our receiver to be optimized for astronomical observations, so it is to our benefit to characterize the KAPPa receivers in a manner consistent with observations. The bias system used to set the receiver’s SIS bias voltage allows for feedback in order to maintain a constant voltage as IF current fluctuates from changing RF inputs. We use this system to sweep bias voltage while recording the SIS current and IF total power in a 60 MHz wide IF band centered at 1.4 GHz. For each SIS bias position, we expose the pixel to both an ambient temperature blackbody source (∼295K), and blackbody source cooled by liquid nitrogen (∼77K). From the ratio of IF power measured these two different temperatures we calculate a Y-factor for each SIS bias position.
Figure 3.7: Block diagram of the KAPPa receiver characterization system. The Y-factor measurements can be made in an automated and highly repeatable fashion by controlling the optical chopper, SIS bias, magnet field strength, LO power, LO input frequency, and IF bandwidth from the same machine that records bias conditions and total power output.

3.3.2 Automated Measuring System

In order to characterize a large number of pixels, we have developed an automated measurement system. Each component of the system is controlled by a single computer that also stores and processes raw data, allowing for instant feedback and fine adjustment. The block diagram in Figure 3.7 outlines how each system is connected to the control computer.

The optical chopper enables us to make Y-factor measurements in an
automated and highly repeatable manner. The chopper’s stepper motor is controlled through a serial port, allowing us to switch between an ambient blackbody source and a source cooled by liquid nitrogen at any time interval. The design and specifications of the optical chopper can be found in (Kuenzi et al., 2014). The liquid nitrogen bath of the cool source is capable of lasting as long as ten hours on a single fill. Thus far, measurements have occurred for as long as 30 continuous hours. However there is no limit on continuous operation. Since there can be 10 hours between liquid nitrogen fills, continuous measurements can be conducted by a single person who can then retain a standard human sleeping schedule.

Another critical component of the automated measuring system is the SIS bias and electromagnet monitoring and control system (bias computer). The bias computer was designed and fabricated by our collaborators at the University of Arizona. With the bias computer, we are able to set the electromagnet current to within $10^{-6}$ of the system’s dynamic range. Because of hysteresis effects in the superconducting magnet coil, the control computer first sets the magnet to the maximum current and then slowly steps down until the current has reach the desired value. The bias computer also has its own internal feedback system for setting the SIS bias voltage, which requires that bias sweeps be much longer then the time constant of the bias computer’s feedback system. In addition, we use the bias computers SIS monitoring system in a feedback loop with the voltage controlled attenuator to set the LO power. With the electromagnet set in a reference position, we set the SIS bias voltage to 1.8 mV, then the control computer automatically adjusts the LO’s attenuator until the desired SIS current is maintained. The dynamic range of the SIS current setting is 2 – 40 µA, but a typical value are between 5 – 30 µA.

The last two controlled systems, LO frequency and IF bandwidth, are set
without the use of feedback, greatly simplifying their implementation. The LO frequency is adjusted through a general purpose interface bus (or GPIB) port on the signal generator which provides the input tone for the LO. The input frequency, 13.6 – 14.4 GHz, is multiplied 48 times in frequency (see Figure 3.7) to a range of 650 – 692 GHz. Adjustment of the LO frequency is done prior to the determination of LO power discussed above. The tunable bandwidth filter that sub-selects within the range of the IF bandwidth is set by a single input voltage. A lookup table is used to map the input voltage to center frequency.

3.3.3 Efficient Receiver Tuning Using Differential Evolution Strategies

Differential evolution (DE) is a population-based class of search algorithms that searches for a solution iteratively by improving upon previous solutions, called candidate solutions or agents (Storn & Price, 1997). Metaheuristic search algorithms such as DE do not guarantee that a global maximum will be found for a given function. However, DE is robust against functions with complicated, nondifferentiable topologies, as well as simple-uniform topologies, because it does not require a continuous function for a given range of parameters. Upon completion of testing, DE requires the response topology of every parameter-set to be ranked according to fitness criteria. The differential nature of the algorithm means it is ideal for operating over continuous ranges of input parameters. Because of this, and the ability to search a function of unknown topology, DE algorithms make an excellent tool for end-to-end characterization of experimental THz heterodyne receivers.

DE has many useful applications, for example, in the field of digital signal processing. Finite impulse response and infinite impulse response filters are utilized to condition digitally sampled data. To make the most of limited computational resources, DE is used to optimize the desired filter response against parameters such
as stopband rejection and passband ripple. This is particularly important when operating in challenging environments, such as space, where designs are required to be extremely efficient in size, weight, and power. DE also provides the capability to match a digital filter to any arbitrary response allowing equalization of channels with fast fading characteristics. Used in conjunction with mechanical modeling software, DE has been used to optimize designs against environmental stress factors such as vibration (Necmettin, 2014). Similarly to how it is employed with the KAPPa instrument, DE can be used to find optimal settings in systems with high configurability, such as optimization of water pump systems (Babu & Angira, 2003). The breadth of DE optimization applications for engineering and design demonstrates the versatility of this technique.

As with other fields of science, THz astronomy is moving to larger format arrays for survey missions. With larger arrays, tasks once completed judiciously by humans require automated optimization. The KAPPa receiver is a demonstration of technological advancements required to build arrays of thousands of heterodyne receivers that tessellate the KAPPa design. Minimum characterization of the single KAPPa test pixel requires tuning a magnetic field, bias voltage, and LO output power as a function of the KAPPa LO’s 40 GHz tuning range. Because of the difficulty in manufacturing and installing SIS devices that are the heart of the KAPPa pixel, the tuning required will vary from pixel-to-pixel. Other portions of the KAPPa IF chain, such as the LNA, can also be tuned and optimized using DE. For a single pixel the tuning is difficult; for an array of thousands of SIS pixels, optimization cannot be done using by-hand methods.

During early KAPPa experiments requiring quick optimization, the LO output contained two tones instead of the desired single tone over 99% of the 650-690 GHz tuning range. Analyzing DE topology of KAPPa isolated the LO as
the issue, exonerating the rest of the system. This spurred the realization of DE as a powerful diagnostic tool allowing us to map the optimal receiver response topology with system changes. After making improvements to the LO (see Section 3.3.5), the complexity of the receiver response topology decreased. At this point, the role of the KAPPa DE algorithm changed to verify that any LO tuning could be optimized in a consistent manner.

The KAPPa DE algorithm is hereafter referred to as Alice; named for the Red Queen character in Lewis Carroll’s famous novels. In the novel “Through the Looking Glass”, the Red Queen says to Alice, “[In Wonderland] it takes all the running you can do, to stay in the same place” (Carroll, 1871). The Red Queen Effect, an evolutionary hypothesis stating that organisms must constantly evolve simply to survive in competition with other constantly evolving organisms (Van Valen, 1973), derives its name from this same line of Carroll’s novel. We designed the DE algorithm, Alice, to quickly optimize any pixel, however it also allows for a given parameter space to be efficiently mapped.

3.3.4 Differential Evolution Algorithm Outline

DE algorithms search parameter spaces for better solutions by mutating the population members with each generation. Once the members of a generation have been measured and ranked according to a cost function, a DE strategy can be used to generate a new population of parameters to test. Alice was developed from the pseudo code in (Ali et al., 2009). Mutation is the property that sets DE apart from other evolutionary algorithms. It is customary to denote the mutation operator as $F$. For simplicity of illustration, let us assume there is only one continuous parameter under test, and that the vectors $X_1$, $X_2$, $X_3$, $X_4$ are all random order lists containing the parameter value for each member of the current population. We can
produce a new generation of test parameters, \( Y \), to test by using the DE strategy:

\[
Y = X_1 + F(X_2 - X_3).
\]  
(3.1)

For this strategy, we can say \( Y \) is the mutation of \( X_1 \). The magnitude of the mutation is controlled by \( F \) and the direction of mutation is determined by the differential operation of \( (X_2 - X_3) \). Here, \( F \) can be any real number, but negative numbers are equivalent to their positive counterparts since the genes in \( X_1, X_2, X_3 \) are in random orders. For values of \( F \) that are much larger than one, it is more likely that the mutants in population \( Y \) will have genes that are far from \( X_1 \). The function \( F \) should therefore be tailored to the gradients of the parameter space under test. A second perturbation can be added using what is known as the crossover. The crossover rate, \( c \), is in the set \([0,1]\). Let \( R \) be a list of random numbers with each element in that list contained in the set \([0,1]\), we make a mask, \( M \), such that for each element \( r_j \in R \) there exists \( m_j \in M \) such that:

\[
m_j = \begin{cases} 
1, & r_j < c \\
0, & r_j \geq c.
\end{cases}
\]  
(3.2)

This mask is used to choose between the genes in \( Y \) and the set of crossover genes in \( X_4 \) to create the final test population \( Z \). If for each element \( m_j \in M \) we determine \( z_j \in Z \) as:

\[
z_j = \begin{cases} 
if m_j = 1, & x_j \\
if m_j = 0, & y_j,
\end{cases}
\]  
(3.3)

where \( z_j \in Z, y_j \in Y, x_j \in X_4 \), and if \( m_j = 1 \) then \( z_j = x_j \) and if \( m_j = 0 \) then \( z_j = y_j \). From this, it can be seen that the boundary case \( c = 0 \) causes \( Z = Y \) and the case \( c = 1 \) causes \( Z = X_4 \).
Before the new population, \( Z \), can be tested and evaluated, some amount of clean up needs to be done to ensure that the perturbed genes in \( Z \) fall in the range of parameters given, since it is possible to perturb a gene outside of the boundaries specified for that gene. After testing, the cost function can be applied to \( Z \).

Consider any \( z_j \in Z \), \( x_j \in X_1 \), and let \( CF(x) \) denote the cost function. Then if \( CF(z_j) < CF(x_j) \), the child \( z_j \) replaces the parent population member in \( X_1 \) otherwise the more successful parent remains in the mating population to create the next generation of children to test.

While we only show the simplest DE strategy, many more strategies are accessible in *Alice*. Additional DE strategies and ideas for choosing initial conditions can be found in (Ali et al., 2009). For example, if quick convergence is desired, many mutations can be made using only the best population member. Again, consider only one gene under test. The vectors \( X_1 \) and \( X_2 \) are unsorted lists of all genes in the current population. Now we create a list that is the same length as \( X_1 \), \( X_2 \) but contains the gene(s) of best population member according to the cost function called \( X_b \). We can now implement a strategy that is many mutations of the best member such as \( Y = X_b + F(X_1 - X_2) \). After improving the KAPPa experiment’s LO, quick convergence for verification became more important than thoroughly searching a parameter space for all solutions. The final phase of KAPPa primarily employed strategies like the one above achieving convergence after two or three generations.

### 3.3.5 The Mutation Parameter ‘\( F \)’

For the majority of the KAPPa experiment, *Alice* continued to use the default parameters set during the algorithm’s initial run. Only after the KAPPa LO had been improved and after the most aggressive testing had been successfully completed did we revisit *Alice’s* settings. During the final stage of the KAPPa
experiment we tested the mutation parameter, $F$. For the test of the mutation parameter, we used simplest DE strategy outlined in Section 3.3.4 and a crossover probability of 0.6. We used a Y-factor, $Y_m$, based cost function $CF(Y_m)$. For $Y_m \geq 1.75$ we set the cost $CF(Y_m) = 0$, for $Y_m < 1.75$ we set the cost $CF(Y_m) = (1.75 - Y_m)^2$. We chose a goal Y-factor of less than the theoretical maximum performance of $Y_m \approx 1.8$ for the receiver so that anomalously high Y-factors would not skew the results away from a true optima. We disabled the algorithm from exiting after reaching its goal, so each test of the mutation parameter had 7 generations, 1 initial population of 50 members made from randomly generated parameters, and 6 generations with 40 members produced through DE. This provided us with a total of 290 Y-factor measurements. The free parameters for Alice search included: SIS junction bias voltage, LO power, IF bandwidth, and LO frequency. The electromagnet had been replaced with a permanent magnet at this epoch in the KAPPa experiment.

The strength of the magnetic field across the SIS junction can greatly affect the Y-factor measurement. Magnetic flux quanta trapped in the superconducting layers of the SIS device suppress the DC Josephson Effect. By using an electromagnet comprised of superconducting solenoid while choosing a specific current, we can dial the strength of the magnetic field and determine the precise relationship between magnet current and Y-factor. We control LO power using a voltage control attenuator provided on our Virginia Diodes LO. However, LO power scales non-linearly with input voltage and is highly dependent on optical alignment, as well as LO frequency. To eliminate this uncertainty we interpret SIS current at a fixed SIS bias voltage of 1.8 mV to scale as linearly with LO power. This calibration is performed before any bias sweeps. Early tests indicate the best Y-factor results lie in between 10 - 15 $\mu$A (Y-factor $\sim 2$) but good Y-factors ($> 1.7$) can be found.
from 8 - 25 µA. Full receiver characterization utilizes two additional parameters: IF bandwidth and LO frequency. While the automated measurement system (Section 3.3.2) is fully capable of controlling these two parameters, they have been kept constant for Figure 3.8. We use an LO frequency of 672 GHz and are measuring the IF band at 1.42 GHz. While we expect the Y-factor to be effected by these parameters, we do not anticipate the parameters of magnet field and LO power to be correlated with either LO frequency or IF bandwidth.

The KAPPa receiver performed well and had a simple response topology while testing the DE parameter, $F$. In this case, the metric that we judge the performance of Alice is total time to convergence to the optimal solution. This is slightly different than the least number of steps. The receiver can become very unstable for parameters far from the optima and this requires the control code to switch the range of the power meter for less sensitivity and larger dynamic range. So the very best settings are those that converge quickly with the least number range switches for the power meter.

In Figure 3.8 we see the results of the mutation parameter $F$ investigation. Namely, $F = 0.1, 0.2$ had the fastest convergence with $F = 0.1$ also finishing the fastest by having the least number of power meter range resets. The extreme mutation parameters $F = 0.01, 1.0$ had the worst convergence: $F = 0.01$ responded too slowly and $F = 1.0$ mutated out of the optimal parameter range. In some cases, the subsequent generation did more poorly overall than the prior generation; this is likely due to the influence of the crossover probability that is most pronounced in the low mutation parameters. Other than this test, the mutation parameter for all other application of the Alice algorithm was set to 0.6, a low convergence mutation parameter on this graph. However, convergence is not the only metric for which DE algorithms can be judged. Thoroughly searching a complex topology would require
a different selection of DE parameters and different analysis techniques.

![Figure 3.8](image)

**Figure 3.8:** Investigation of the DE mutation parameter $F$. Using differential evolution strategy outlined in Section 3.3.4 and a crossover rate $c = 0.6$, we looked at the convergence of *Alice* using 7 different mutation parameters $F = 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0$, each with a unique symbol and color. Each generation is placed at the average time of measurement for that generation and the average Y-factor for that generation, the error bars are the standard deviation for the Y-factors in that generation.

### 3.3.6 The Algorithm Controlled Receiver Characterization Test Bench

*Alice* interacts with the rest of the KAPPa control code to set the adjustable input parameters of the KAPPa receiver. The KAPPa control code then measures and analyzes the receiver’s performance. Based on the measured performance, new input parameters are created for the subsequent generation of performance measurements. This process occurs iteratively until the algorithm’s goals have been met. *Alice* requires only parameter ranges and the fitness ranking for tested
parameters. Testing a black-box system requires a highly robust and automated measurement setup. Creating a robust automated test bench may be the single largest barrier to implementing a DE algorithm for tuning a device.

Vectorization of *Alice* allows for one to *n*-dimensional parameter spaces to be tested. However, it follows that population size (the number of parameter sets tested per generation) must increase as the number of parameters being search increase. Thus, it can be more efficient to hold some parameters constant to characterize the parameter space before global optimization. In evolutionary terms, the input parameters correlate to the genes of organisms. When operating *Alice*, a random initial population (bounded within specified ranges) can be generated, or best results of a previous run can be used as the initial seed.

Unlike purely computational DE processes, each member of the population (single set of parameters) must be tested in series. This can lead to results that are not purely a function of the input parameters. For example, after 4 hours of operation, an amplifier may go into oscillation, thus changing the noise performance of the system. Evaporation of liquid nitrogen, used to cool a cold-noise source, is another common problem that leads to varying system performance. One can imagine all manner of problems leading to non-ideal measurement, which is where *Alice* deviates from purely computational DE processes. To combat anomalously high measured performance, *Alice* implements checks to test if the best parameter sets give reproducible results. If results are not reproducible, the anomalously high population member (set of input parameters) is removed from the population and no longer contributes to subsequent generations. After a specified number of generations without progressing toward the goal, *Alice* will abort and send email notification. After completion, a full characterization is completed using the best population members to verify the receiver configuration performs as expected. While
DE is a quick and powerful tool, the main limitation of Alice's current application is the speed at which measurements using each population member can be made.

There are two limiting factors for the speed of operation for the KAPPa system. First is relaxation time of the electronics after changing parameter states. Second is the speed of the optical chopper as it switches from a hot (300 K) load to a cool (77 K) load, see Section 3.3.2. The fastest rate during testing, one measurement per second, proved to be too violent for the optical chopper. A rate of one measurement per three seconds, used in most testing, minimized chopper repair time. Minimizing the rate of measurement may be an insurmountable impediment for implementing DE to optimize future systems. For this reason, DE is not a catch-all, but should be judiciously applied to systems with complex or vast tuning topology.

While a generation of population members is tested, Alice is idle. After measurements are completed and the data is processed, the measured results are fed into Alice's cost function. The cost function ranks the measured results according to a defined metric. A simple cost function, comparing the ratio of IF power between a hot and cold load, is currently utilized. This ratio is an estimate of receiver Y-factor. When the Y-factor is high, the receiver is performing well and the set of genes of the population members are highly ranked. In practice, the cost function requires artful handling. For example, all ratios above a threshold are marked equally so as not to give credence to anomalously good results.

For the KAPPa receiver, there are three zones in the SIS bias voltage parameter space that give high Y-factors. There is only one zone, however, where the receiver is stable enough to be used for the longer integrations required during astronomical observations. For this reason, we explored adding additional components to the cost function to highly rate receiver stability. The current test
for receiver stability, measurement of Allan time, takes greater than 1000 seconds to complete one measurement. The long measurement time makes the current setup to measure Allan time impractical for use with Alice.

### 3.3.7 Optimization with Alice

Using Alice to optimize settings led us to some interesting results. We had observed for some time that the SIS device at the heart of the KAPPa receiver was slowly delaminating over time. This could be observed as an increase in contact resistance across the device. This affected the optimal bias voltage and LO pump power, something that might not have been observed had there not been an automated optimization procedure. Also the optimal LO pump power changed from 13 to 10 $\mu$A after the seed tone of the LO was filtered. We hypothesize the 3 $\mu$A difference was from unwanted tones that were subsequently filtered out. The SIS device in the KAPPa receiver requires an applied magnetic field to suppress Josephson Noise. When starting the KAPPa experiment, we expected receiver performance to be a strong function of applied magnetic field from the electromagnet. However, Alice confirmed that the function is only sharp near a magnetic field strength of zero.

By the end of the KAPPa experiment, it was realized that we could run every user controlled component with Alice. Instead of painstakingly confirming the bias current of various LNA’s in the KAPPa IF chain by individual measurements, we could have simply hooked them to a programmable power supply and optimized their currents with the rest of the receiver’s free parameters. The overall goal of the KAPPa project was to develop technology to enable the construction of THz heterodyne receiver arrays that have an order of magnitude more pixels than the current state-of-the-art 64-pixel array. Tuning and optimization of a 1000-pixel
array based in SIS devices is a task that should not be attempted by-hand. Due to unavoidable manufacturing and installation difficulties, SIS devices can require different bias currents. LO’s can have a non-uniform power delivery across the focal plane of the array. But by judicious selection of cost functions one can optimized over the entire array at one time or make individual adjustments to each pixel.

3.3.8 Application and Extension of Differential Evolution

The present work is the beginning of possible improvements to receiver tuning and optimization using DE. The natural scalability of DE, requires little augmentation from its current state to be applied to arrays greater than 1000 heterodyne pixels, makes it the right choice for KAPPa. The required augmentation resolves the difference between parameters that can be set for individual pixels and those that must be optimized for all pixels globally. Some parameters, such as LO frequency, result in global change over the array, and other parameters, such as bias voltage, can be tuned for individual pixels. Succeeding rounds of DE need to be applied to first-time local properties of each pixel before continuing to a global tuning. This requires the design of a cost function that can factor in the performance of a specific pixel under test and how the varying parameters of that single pixel affect overall performance of the array. As mentioned in Section 3.3.3, DE has vast application and is not restricted in functionality to heterodyne instruments.

DE algorithms created to test and calibrate receivers in the lab can be extended to provide active receiver calibration as a function of environmental or system changes over time. When moving a receiver to a new location, all manner of environmental differences can affect receiver performance. For example, components can delaminate from vibration or thermal cycling which changes their resistance and
thus the required bias voltages (per Section 3.1.3), a faulty microwave oven at the new location can cause unwanted tones in the IF signal, and components can simply overheat in the rarefied atmospheres at high altitude telescopes. With everything that can go wrong, having DE available to control and test your system can be a valuable tool, guaranteeing you have ideal optimization for any distinct installation or time. Additionally, DE is a powerful diagnostic tool for testing theories concerning the cause of non-ideal performance. For the present work, application of the default setting for the crossover rate and DE strategy mutation parameters is powerful enough to yield an optimal solution. The task of optimizing the DE mutation parameters has not been completed; however, fitness can easily be determined by measuring convergence time of the algorithm. Thus, it is possible to wrap one DE algorithm, such as Alice, with another DE algorithm that selects the mutation parameters for Alice. Optimization of convergence in this way is helpful for systems that might need to be tuned often over their lifetime.

Possible future advancements of Alice need to resolve two standing issues for the user. First, the SIS bias voltage has three regions with repeatable high Y-factor. Only one of these regions, however, can be used in practice because the Y-factor variation is too high to be readout optimally, resulting in higher error. Because Allan time measurements for the current KAPPa configuration take hours, as compared to seconds for Y-factor, receiver stability cannot be accounted for with the current cost function. Future improvements can be made to the algorithm to first optimize the receiver for Y-factor, and then test and optimize receiver stability for those solutions with high Y-factor. The resultant ranking of solution sets shows which mutation parameters are best for astronomical observation. The second possible enhancement for Alice centers on the ability to search for multiple solutions that meet the fitness criteria. In many cases, it is desirable to not only find one
solution, but also all possible solutions in parameter space. When a sufficient solution is found during the search, subsequent tests can be bounded to avoid a region around the known optimal solution. This forces the algorithm to continue searching for additional solutions without repeating permutations that have already been measured and confirmed.

As the field of astronomy moves to larger arrays and data sets, the need for standardization and automated optimization becomes increasingly unavoidable. Luckily, many disciplines currently benefit and contribute to the advancement of metaheuristic optimization algorithms. Astronomy will continue to benefit greatly from interdisciplinary crossovers.

3.4 Receiver Performance

Let us discuss some of the standard tests developed for assessing the performance of the KAPPa receiver and its constituent components. All data presented in this section results from querying the KAPPa database. The database is vast and will be an invaluable tool for future instrument builders within the THz group. We present typical data, but much of it comes from the last several months of testing after performance issues from the LO were adequately resolved.

To begin: a review of the basic electoral engineering terms that will be ubiquitous in the rest of the section. All quantitative fields of study prefer to work in convenient number systems and RF electrical engineers are no exception. Power is measured in logarithmic units called decibels (dB) and relates to power $P$ in watts (W) in the following:

$$L_p = 10 \log_{10} \left( \frac{P}{1W} \right) \text{dB.}$$

For small signals, as are typical for astronomy instruments, we often use a 1 mW as
the reference power and milli-decibels (dBm) as

$$L_p = 10 \log_{10} \left( \frac{P}{1 \text{ mW}} \right) \text{ dBm}. \quad (3.5)$$

When referring to the difference between a reference power $P_0$ and a measured power, we can refer to the difference in dB by applying the familiar setup:

$$L_p = 10 \log_{10} \left( \frac{P}{P_0} \right) \text{ dBm}. \quad (3.6)$$

Finally, we point out a few mental math tricks to make working in this system simpler: Multiplication or division by a factor of 10 in watts means addition or subtraction, respectively, by 10 in dB and dBm. Multiplication or division by a factor of 2 in watts means addition or subtraction, respectively, by 3 in dB and dBm.

For increasing signal power ($P_{out} > P_{in}$), such as that seen as a signal passes through an active circuit such as an amplifier, we refer to linear gain $G$ as

$$G = \frac{P_{out}}{P_{in}}, \quad (3.7)$$

and can be converted to dB (as it often is) using Equation 3.6. For decreasing signal power ($P_{out} < P_{in}$), such as that seen as a signal passes through as passive circuit (such as a resistor), we refer to linear loss $L$ as

$$L = \frac{P_{in}}{P_{out}}. \quad (3.8)$$

From this it is clear that inverse of $G$ is $L$. If these are chosen in the way suggested above and expressed in dB, only positive values will be produced values, which is standard.

We define signal to noise ratio, $SNR$, as,

$$SNR = \frac{P_{signal}}{P_{noise}}, \quad (3.9)$$
where $P_{\text{signal}}$ is the power from the signal and $P_{\text{noise}}$ is the power from the noise. We are now equipped to define the noise factor, $F$, of a circuit component as,

$$F = \frac{SNR_{\text{in}}}{SNR_{\text{out}}},$$  \hspace{1cm} (3.10)

which takes into account the SNR in and out. Here we can see that a noiseless circuit component, one that adds no additional noise, has $F = 1$. All physical circuit components, with finite temperatures, will have $F > 1$, since $F$ is a unitless ratio. This is distinguished from noise figure, $NF$, as

$$NF = 10 \log_{10}(F),$$  \hspace{1cm} (3.11)

which is quoted in units of dB. Finally we define noise temperature $T_N$ in K as

$$T_N = (F - 1) \, 290 \, \text{K}. \hspace{1cm} (3.12)$$

The noise temperature, $T_N$, is the preferred unit for radio instrument builder as it can be used to to make direct comparison to objects with a finite temperature to calibration sources. As many of the figures and discussion in this section use noise temperature, we will also define it directly as

$$T_N = \frac{P k_B}{B}, \hspace{1cm} (3.13)$$

where $P$ is power in W, $B$ bandwidth in Hz over which $P$ is measured, and $k_B$ is the Boltzmann constant. Because of it relationship to physical units, $T_N$ is the preferred term for describing noise calculated from fundamental measurements, while $F$ and $NF$ are most convenient for working with cascaded systems of components with known noises.

We commonly measure the $T_N$ for the KAPPa receiver and evaluate its performance using the Y-factor method. The Y-factor measures the system noise by
exposing a circuit element to two different known temperature loads and measuring
the component’s output power in response to each load. The value \( Y \) is defined as

\[
Y = \frac{P_{\text{hot}}}{P_{\text{cold}}},
\]

(3.14)

where \( P_{\text{hot}} \) and \( P_{\text{cold}} \) are the output power of the component under a hot and cold
load, respectively. We see that a \( Y = 1 \) corresponds a component that is insensitive
(blind) to the temperature changes of the load. We can use a measured \( Y \) to
calculate the \( T_N \) of the component under test and we can apply the following as

\[
T_N = \frac{T_{\text{hot}} - YT_{\text{cold}}}{Y - 1},
\]

(3.15)

where \( T_{\text{hot}} \) and \( T_{\text{cold}} \) are the hot and cold load temperatures in K, respectively. For
all KAPPa receiver tests, \( T_{\text{hot}} = 293 \) K, the ambient temperature of the laboratory,
and \( T_{\text{cold}} = 77 \) K, the boiling temperature of liquid nitrogen. In subsequent
calculations we used the slightly conservative \( T_{\text{hot}} = 290 \) K and \( T_{\text{cold}} = 80 \) K in an
allowance for imperfect emissivity of out blackbody sources. For KAPPa this was
several layers of Eccosorb, a common carbon imbedded RF absorber. Since we
measured the entire system using this method, we can specifically define the system
temperature for our setup as

\[
T_{\text{sys}} = \frac{290 \text{ K} - Y \times 80 \text{ K}}{Y - 1},
\]

(3.16)

where \( T_N \) from Equation 3.15 is replaced with system temperature, \( T_{\text{sys}} \).

### 3.4.1 Shot Noise Tests

The first test we conducted looked at the KAPPa receiver performance for
everything after the SIS device. By biasing the SIS device with a current in excess
of the critical current, we render the device useless as an RF mixer. Instead it
becomes a source of shot noise since it is now behaving as a normally conducting resistor. Using the device in this way, we can check the entire IF amplification chain and determine the noise that it contributes to the final system temperature. The majority of the noise that will be contributed in the IF chain is from the first IF amplifier. This is best understood in terms of the cascaded amplifier (Friis) equation as

\[ F_{total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots, \]  
(3.17)

where \( F_n \) and \( G_n \) are the noise factor and gain of the \( n \)th amplifier in a cascade as defined in equations 3.10 and 3.7, respectively. If we differ from convention and allow \( G_N < 1 \), we can model a lossy component such as transmission lines and capacitors. Based on this equation, we can see why it is so important to have some amplification (gain) before the signal is attenuated over a long transmission line, because all additions of noise will be reduced by a factor of the 1st amplifier’s gain \( G_1 \). This can be see in both Figures 2.3 and 2.5 as a driving design requirement for the pixels footprint.

Fundamentally, shot noise, or thermal noise, arises from the random thermal motion of electrons based on their temperature. In the classic paper by Nyquist (1928), we see that the square of average expectation voltage, \( \langle v^2 \rangle \), for the measurements across a noisy resistor with resistance, \( R \), at physical temperature \( T \) is given by

\[ \langle v^2 \rangle = 4 R B k_b T. \]  
(3.18)

From this we can define the thermal noise, \( P_{shot} \), as

\[ P_{shot} = V I = \frac{V^2}{R} = \frac{\langle i^2 \rangle}{R} = 4 B k_b T B = \langle i^2 \rangle R, \]  
(3.19)

where \( \langle i^2 \rangle \) is the square of the expectation current from thermal noise. The square
expectation current \( \langle i^2 \rangle \) across an SIS device is given by

\[
\langle i^2 \rangle = 2eBI,
\]

(3.20)

where \( I \) is the direct current measured across the SIS device, and \( e \) is the fundamental charge of one electron [Dubash et al., 1995]. Here it is good to note the while one side of the SIS device is conducting normally with resistance \( R \) the other side is still superconducting under the applied bias voltage, thus charges must tunnel across the SIS junction as Cooper pairs that are subsequently thermalized and broken upon reaching the normally conducting side of the SIS junction.

From equations 3.19 and 3.20 we are now able to calculate the physical temperature of the SIS junction with measurable qualities as

\[
T_{\text{SIS}} = \frac{2eBI R}{4Bk_b} = \frac{eIR}{2k_b} = \frac{eV}{2k_b},
\]

(3.21)

where \( V \) is the bias voltage across the of the SIS device. We can use Equation 3.15 to calculate the noise of the entire IF system but first we must collect the the relevant information. To review, the KAPPa receiver test code needs to measure the voltage and current across the SIS device and record the output power for some portion of the IF bandwidth, let us call these two measurements \((V_1, I_1, P_1)\) and \((V_2, I_2, P_2)\) where \(V_1, V_2\) are much greater then the critical voltage. The resistance \( R_{\text{SIS}} \) of the SIS device in the ‘nominally conducting region’ is given as,

\[
R_{\text{SIS}} = \frac{V_2 - V_1}{I_2 - I_1}.
\]

(3.22)

Therefore, we calculate the IF system noise \( T_{\text{IF}} \) using Equation 3.15 as,

\[
T_{\text{IF}} = \frac{T_{\text{hot}} - Y T_{\text{cold}}}{Y - 1} = \frac{\frac{eV_2}{2k_b} - \frac{P_2}{P_1} \frac{eV_1}{2k_b}}{\frac{P_2}{P_1} - 1}.
\]

(3.23)

In practice the KAPPa receiver code takes the average of many measurement in the ‘normally conduction region’ of the SIS device to drive down measurement
error and ensure that the SIS device is fully in the normally conducting state. For a future receiver, this test should be conducted to optimize the LNA bias voltage and other components that can be optimized with computer feedback, as discussed in Section 3.3.7. Unfortunately the KAPPa project was already using all computer controllable power supplies available at the time.

In Figure 3.9 we see all the shot noise measurements for where $T_{IF}$ was calculable. This is in agreement with the expected noise temperature of the KAPPa first stage amplifier, the WBA25, where $T_{WBA25} < 15$ K (Russell & Weinreb 2012) for the 0.5-4 GHz IF bandwidth for which the KAPPa was measured. For a detailed discussion of the WBA25 amplifier see Section III of Russell & Weinreb (2012).

![Figure 3.9: A histogram of the IF noise temperature, $T_{IF}$, for all KAPPa measurements that meet the criteria for Equation 3.23. The first stage KAPPa LNA was set to a nominal low power dissipation current of 1.5 mA for all measurements.](image-url)
3.4.2 Mixer Conversion Loss

When measuring the thermal noise of the IF system, we can measure the mixer conversion loss if we additionally do a Y-factor measurement of the entire system for the load temperatures $T_{\text{hot}} = 290$ K and $T_{\text{cold}} = 80$, as described in Equation 3.16. This must occur while collecting the measurements of the receiver output power, $P_{\text{hot load}}$ and $P_{\text{cold load}}$. From the shot noise analysis, we collect the $(P_{\text{shot1}}, T_{\text{shot1}})$ and $(P_{\text{shot2}}, T_{\text{shot2}})$ as described in Section 3.4.1 and for Equation 3.22. We calculate the mixer conversion loss as

$$L = \frac{P_{\text{shot2}} - P_{\text{shot1}}}{P_{\text{hot load}} - P_{\text{cold load}}} \frac{T_{\text{hot}} - T_{\text{cold}}}{T_{\text{shot2}} - T_{\text{shot1}}}. \quad (3.24)$$

The unitless mixer conversion loss can be transformed to the dB scale through Equation 3.6. Figure 3.10 shows the conversion loss as a histogram for all KAPPa receiver measurements that meet the requirements above. This is slightly more restrictive set than Figure 3.9 but still encompasses most measurements of the KAPPa system.

3.4.3 Intersecting Lines Test

The intersecting line technique is a Y-factor based test that allows for the calculation of the optics noise temperature, $T_{\text{opt}}$. Here we follow the method from Ke & Feldman (1994). Graphically, the intersecting line technique is one of the easiest of the KAPPa receiver tests. Several KAPPa Y-factors are taken in accordance with Equation 3.16 at several LO signal strengths. We measure relative LO signal strengths at a specific SIS bias voltage, $V_{\text{bias}}$, in addition to a specific magnetic field by observing the current across the SIS junction, $I_{\text{SIS}}$. In the absence of the DC Josephson effect, the current is caused by ‘pumping’ from photon assisted tunneling across the SIS junction. While the optimal $I_{\text{SIS}}$ is dependent on the
receiver tuning at the time of the measurements, take for example Figure 3.11 where the optimal $I_{\text{SIS}} = 14\mu\text{A}$. While the receiver is well biased, we can collect a series of measurements for receiver output power $P_{\text{hot load}}$ and $P_{\text{cold load}}$ for decreasing LO pump current, $I_{\text{SIS}}$. The lines in Figure 3.11 are formed from each measurement pair using only two data points, one at 77K and one at 300K with each line corresponds to a different LO pump power, where 14 $\mu\text{A}$ of pump is optimal. The temperature at which these points intersect gives us the noise from the optics in front of the mixer. The red vertical line is the conservative optics temperature corresponding to 44 K. This test is tricky in practice, because it lacks the robust number of measurements to determine the each line. Also, the instrument operator must perform these measurements quickly to avoid issues from receiver gain instability inherent in SIS receivers.
Figure 3.11: The intersecting lines test for determining the noise contribution of the optics. The y-axis scales as power in mW, but it is simply the raw voltage output of a variable range powermeter. However, the absolute scale is unimportant for this test.

Analytically we follow the example of Ke & Feldman (1994) and write the total output of the KAPPa receiver $P_{\text{Tot}}$ as,

$$P_{\text{Tot}} = \left( (T + T_{\text{opt}}) * G_{\text{opt}} G_{\text{SIS}} + T_{\text{SIS}} G_{\text{SIS}} + T_{\text{IF}} \right) * k_b B G_{\text{IF}},$$

(3.25)

where the gains and noise temperature pairs $(G_{\text{opt}}, T_{\text{opt}})$, $(G_{\text{SIS}}, T_{\text{SIS}})$, $(G_{\text{IF}}, T_{\text{IF}})$ are for the optics, SIS mixer, IF system – respectively, $T$ is the load temperature (either 290 or 80 K), $B$ is the IF bandwidth of measurement, and $k_b$ is the Boltzmann constant. For the KAPPa SIS device the $T_{\text{SIS}} G_{\text{SIS}}$ is independent of LO signal power (this is true in low phase noise regions of the LO output as a function of frequency). With the ability to hold all other values in Equation 3.25 constant and varying the load temperature $T$, $P_{\text{tot}}(T)$ is a straight line where is $T_{\text{opt}}$ is simply the
negative x-intercept (this format of a linear equation is called point-slope). By decreasing the optimal LO pump power we can verify our assumptions and measure $T_{\text{opt}}$ robustly for each LO power as in shown in Figure 3.11.

3.5 KAPPa Receiver Performance as a Function of $T_{\text{sys}}$

The ultimate measure of a receiver’s performance is the determination of end to end noise temperature $T_{\text{sys}}$ (see Equation 3.16). For the KAPPa receiver we physically measure Y-Factor (Equation 3.14) before determining system temperature. When the receiver is not working well we calculate Y-factor as a function of bias voltage, as seen in Figure 3.12 on the right Y axis, and we expect to see a Y-factor around 1.7 for a range of typical bias voltages. As the tuning topology of SIS receiver is difficult to navigate, we deploy a suite of diagnostic tests. Using the current-voltage curve of the SIS device, we determine the resistance of different linear regions, the black lines over plotted on the current-voltage curve with resistances reported in the legend of Figure 3.12. These regions are consistent with the regions of Figure 3.2. The absolute scale of total power measurements, light and dark blue marker in Figure 3.12 corresponding to the left axis, can also be useful as the best Y-factors occur under the low total power conditions where dark current from the DC Josephson effect is fully suppressed is fully suppressed. From this data more complex tests can be preformed (such as those discussed in sections 3.4.1 and 3.4.3) if the diagnosis is unclear. In Figure 3.12, the KAPPa receiver appears to have a slightly wrong magnet field bias as indicated from the unusually high total power readings.

The current-voltage and total power-voltage curves seen in Figure 3.12 are generated from sampling the output of a total power detector after passing through an IF filter (see Figure 3.7). Total power stability over longer integration times
Figure 3.12: This is a typical diagnostic Plot for the KAPPa receiver. The typical shape of the SIS device can be seen on the purple line, this line is pumped by LO despite the legend caption. The right axis scales to the dark and light blue total power measurements from which Y-factor, the green line and right axis, are determined.

(minutes) will determine the effectiveness of setting the receiver to any set of parameters that produce a high Y-factor.

3.5.1 System Temperature vs. IF frequency

To measure the IF bandwidth response of the KAPPa receiver, we use a tunable filter with resonate yttrium-iron-garnet (YIG) sphere. The YIG sphere changes resonance frequency in the presence of a magnetic field. Our YIG filter for the KAPPa IF system has a 3 dB bandwidth of 15 MHz, a functional tuning range
measured from 0.5 GHz at 0 voltage on the filters electromagnetic solenoid to 4 GHz at a solenoid voltage of 10 V. The increase in system temperature, $T_{sys}$, for the middle voltages of 5 – 6 V (as seen in Figure 3.13) is consistent with simulations from KAPPa team member Jacob Kooi before the SIS device was constructed.

**Figure 3.13:** In this plot we show the well-performing data points per voltage bin, which is marked with an average line of $T_{sys}$ as a function of IF bandwidth – measured with the electromagnet installed. Note that the middle of the IF band at 5 – 6 V has an increased $T_{sys}$ consistent with simulations of the SIS device before it was constructed.
3.5.2 System Temperature vs. LO frequency

The KAPPa DE algorithm, *Alice*, (discussed in sections 3.3.4) originally developed as a tool to identify the source of poor signal-to-noise performance as a function of LO frequency. At the time, KAPPa receiver tests performed at the expected system temperature of 200 K but only at the LO’s center frequency of 672 GHz and some other select frequencies. While we first analyzed performance in 1 GHz steps, testing discovered that using random test frequencies is beneficial to finding optimal LO frequencies. Identification of all LO frequencies meeting performance requirements can isolate the problem to the LO chain. Unfortunately, each of the LO frequencies meeting performance require independent optimization of LO power and SIS bias voltage. We quickly realized that a human cannot perform the decision making process for the hundreds of hours required to optimize for each working LO frequency point. Furthermore, the unresolved uncertainty that the issue stems purely from the LO leads human optimization to be even less practical.

Utilization of *Alice* to characterize the receiver isolated performance issues to the LO chain. In lieu of sending the LO back to the manufacturer, we decided to try to improve the 14 GHz seed tone that enters the LO and subsequently is multiplied in frequency 48 times before continuing to the KAPPa receiver. We designed a new signal source from a tunable yttrium-iron-garnet (YIG) oscillator and added additional filtering from a tunable YIG filter over the same range. While the YIG oscillator generated tone had fewer and smaller spurious tones, the filtering of LO seed tone provided the needed improvement. Once the LO input signal had additional filtering, the parameter space for which the receiver performed well became a weaker function of LO frequency. At this point in the KAPPa experiment, optimization with *Alice* provided only a marginal improvement as compared to the
default value for each parameter.

In Figure 3.14, we see that the best system temperatures for the lifetime of the entire KAPPa experiment as function of LO frequency. The data sets plotted show the improvements in system performance as a function of LO frequency. Before the additional filtering was added to the LO seed frequency, Alice searched for the best LO frequencies within a 1 GHz bin. Optimization with Alice greatly improved system performance in the 1 GHz bins compared to evenly spaces searches and searches of random LO frequencies. This can be seen in Figure 3.14 as the difference between the dotted line with triangle markers, showing the response when testing on the integer values of LO frequency, and the dashed line with diamond markers where Alice searched for the best frequencies within that 1 GHz bin. This worked extremely well for the center frequencies of 658-675 GHz and gave us the evidence needed to start investigation into the LO chain as the source of the unwanted noise. Since the IF bandwidth is spans greater than 4 GHz changing the LO frequency within a 1 GHz bin would not have effected the receivers ability to detect a particular spectral line from an astronomical source. While losing access to some LO frequencies is not a desired property of a receiver, it did highlight that DE could be used to find tunings that could make an otherwise useless receiver perform as well as possible.

The 2nd permanent magnet system greatly improved the KAPPa receiver’s ease of use compared to the previous electromagnet and permanent magnet. It became challenging to find settings that were non-optimal. During this final configuration of the KAPPa receiver, there were multi-hour periods of time where the receiver operation at system temperatures that were lower than what had ever been achieved with the electromagnet (Figure 3.14 diamond markers from 675 to 690 GHz). During the KAPPa experiment, measurements of the current across the
**Figure 3.14:** Best system temperatures as a function of LO frequency. The solid line with circle markers is the best of all data for the entire KAPPA experiments. The dot-dashed line with star markers is the best of all electromagnet data, the electromagnet allows for precise tuning of the magnetic field. Permanent magnet 1 and 2 are the triangle dotted and diamond dashed lines, respectively. Each data point is calculated from the mean of the 5 best data points in a 1 GHz bin, the error bars represent the standard deviation. All points are filtered to be within the expected SIS bias voltage and current ranges. Additional filtering removed single measurements with system temperature standard deviations greater than 10%. Only permanent magnet 2 data was taken at integer values of LO frequency, other data has components that were created by a search algorithm that was allowed to have any rational LO frequency. Electromagnet data was taken using a thinner beam splitter. The thinner beam splitter accounts for high system temperatures at the extremes of the LO tuning range due to insufficient LO power.

SIS device were used to tune the LO power. For the 2nd permanent magnet, we found that the optimal current across the SIS device was lower then it had been for electromagnet and 1st permanent magnet. We hypothesis that this is the result of better suppression of Josephson noise by the field of the 2nd permanent magnet, which also highlighted the need to reapply our optimization algorithm after any
3.5.3 *Electromagnet Results*

In this section, magnetic field strength (|B|) is discussed in terms of current through the solenoid (a component of the electromagnet system), in lieu of absolute calibrated field strength. While the absolute field through the SIS junction is calculable, it relies on measurements spatial changes in the magnetic field that are smaller than the probe of our gaussmeter. In addition, we are only able to make 300 K measurements of the magnetic field on a lab bench, which may not fully correspond to in situ measurements at 4.2 K where the magnet permeability of materials increases. Thus we present solenoid current (IS) as proxy for |B|.

Figure 3.15 shows the best receiver system temperatures (T$_{sys}$) as a function of the electromagnet solenoid current (IS). As seen in Figure 3.15, there are several minima for T$_{sys}$ as a function of IS. As IS increases, the range of IS per T$_{sys}$ minima increases and T$_{sys}$ between the minima decreases. In practice, the first minima for T$_{sys}$ is not useful for setting the receiver. This is due to the hysteretic effects in IS when setting magnetic fields that are larger than range of currents that fall in the 1$^{st}$ minima. Setting the electromagnet at the first minima required slowly stepping the electromagnet current and halting the stepping when optimal T$_{sys}$ had been achieved and requires two sweeps if the optimal T$_{sys}$ is not known. The other minima were more tolerant of variation in the magnetic field strength. For ease of use, we commonly set the electromagnet to the fourth minima in Figure 3.15, corresponding to IS greater than 50 mA. Because T$_{sys}$ between minima decreases as IS increases, the receiver is more likely to have measurable response for non-optimal fields near the fourth minima. This was an important consideration when tuning the permanent magnet field strength.
Figure 3.15: System temperature as a function of electromagnet current. The error bars show the spread of the 5 points per 1 mA bins. These points were taken using the center LO frequency of 670 GHz and at a 15 MHz IF bandwidth centered at 2 GHz. We can see several minima in system temperature, at 14 mA, 28 mA, 37 mA, and 54 mA. As the electromagnet current increases, the minima can be found over larger ranges of magnet current. Using this data, larger magnet field strengths for the permanent magnets are more likely to display low system temperatures even in the case of non-optimal field strength.

3.5.4 Permanent Magnet Results

During the first permanent magnet test on the KAPPa receiver, we could not predict how the magnet field strength might change as the permanent magnet system was cooled to 4.2 K. To maximize the likelihood of a positive result, we attempted to calibrate our permanent magnet to be in the fourth minima for $T_{sys}$ (the 50-62 mA range in Figure 3.15). For the fiducial parameters of LO frequency of 670 and 675 GHz at a 15 MHz IF bandwidth centered at 2 GHz, the first permanent magnet yielded a system temperatures of near 200 K (see Figure 3.14). Figure 3.15
shows the best system temperatures achieved with the electromagnet in the 170–180 K range. We concluded that we had likely over-shot the magnetic field strength for the 1st permanent magnet system. The 2nd permanent magnet system was designed and measured to have a 63% reduction in field strength compared to the 1st permanent magnet system.

In Figure 3.14, we compare the electromagnet results to those of the two permanent magnet systems for system temperature and a function of LO frequency. The constituent data for the electromagnet and 1st permanent magnet were created using a searching algorithm that allowed the LO frequency to be any rational number, while the 2nd permanent magnet data is a simple sweep of integer values of LO frequency (Wheeler & Toland, 2016). Since the LO does not have uniform amplitude modulated noise performance as a function of frequency, there were times when the searching algorithm could find lower system temperatures for LO frequencies near, but not on, the integer value of frequency. In this way, better system temperatures for the less optimized field of the 1st permanent magnet could be achieved over the more optimized field of the 2nd magnet.

Overall the permanent magnet performance was within 50 K of the electromagnet performance (Figure 3.14) and improvements in the permanent magnet performance are still possible. The ease of operation for the KAPPa system was greatly increased with the installation of permanent magnets. Because of hysteretic effects, the electromagnet was the most difficult input parameter of the KAPPa system to tune. With the dominant heat load from the electromagnet current gone, the KAPPa cryostat doubled helium hold time. The installation of the permanent magnet was a major boon for the receiver operations team. The operation of the receiver was simplified without the need to tune the electromagnet. A full characterization of the receiver dropped from weeks of testing to only 3 days,
and the hold time of the receiver’s cryostat was doubled without the added heat load from the electromagnet.

The permanent magnet, was the last run of tests for the KAPPa system and thereby benefited from the work of the DE algorithm *Alice* (see Section 3.3.6). So instead of attempting to set the voltage and current using feedback loops, we would simply set a parameter space to investigate and walked away. Prior to using *Alice* we would set the LO signal power by running a feedback loop to attenuate the LO until a desired SIS ‘pump’ current was achieved, often between 12-14 $\mu$A. Only after the receiver was warm and we were analyzing the data did we notice that the best performance points (pink circles in Figure 3.16) were at a lower than expected SIS currents. This further impressed the notion that the complex receiver tuning tasks are best left to algorithms with no prior assumptions.
Figure 3.16: All permanent magnet tests with $T_{sys}$ plotted as a function of SIS current. This data was measured using a differential evolution algorithm programmed to sweep a variety of parameters while optimizing others. The swarm-like behavior is a result of the optimization of a continuous parameter space. Each point is determined from a pair of hot and cold load measurements, the number of points demonstrates the robustness of the KAPPA receiver and its test system through the long duration of operation needed to take these measurements.
Chapter 4

ON SKY OBSERVATIONS

4.1 SuperCam

4.1.1 SMT installation

The SuperCam instrument is the predecessor to KAPPa, and as such, the work here builds on the hard-learned lessons from the SuperCam team. Working with SuperCam validated many of the decisions that we made as a part of the KAPPa team. For example, we chose to design and create a single, flexible, and compact circuit board that snaps in place to carry the IF signals from the receiver to the outside world for the KAPPa experiments. This decision never seemed more important then while installing 256 SMA connections under a 400 lb SuperCam receiver at the Sub-millimeter Telescope on Mt. Graham Observatory in Arizona. During the SuperCam installation, we saw some amplifiers and equipment overheat due to the difference in air density in the lab where the equipment was tested and density at the telescope. Having a $\sim 70\%$ air mass means fan-based heat exchangers are less effective. Most problems could be solved with the installation of large fans purchased from any local store which were zip-tied to the mounting structure of the receiver.

4.2 sReduce: A New Heterodyne Receiver Data Reduction Pipeline

An array of heterodyne receivers like KAPPa can be thought of as system of many individual radio receivers. Many current radio telescopes have a single dish, acting as a large antenna, that feeds a single radio receiver. Telescopes of this type
have forged a trail for future receivers of increasingly high-frequency by solving many of the observational and data reduction processes needed for heterodyne receiver observations. However, for both current heterodyne receivers, like SuperCam, and future kilopixel receivers, like KAPPa, we must re-examine the techniques of single dish data processing.

Receivers like KAPPa and SuperCam are built to rapidly map many square degrees of the sky. Because many pixels will look at the same region of the sky, a single pixel in an array receiver will not dwell at a single point for as long a time as a single dish receiver would. Therefore, instead of thoroughly characterizing a point on the sky by single instrument with known response over a relatively short period of time, many pixels with different gains can contribute shorter integrations over a variety of timescales. One can treat each pixel in a heterodyne receiver array as a single dish telescope and apply the single dish telescope methodology for data reduction and later add together the results for all pixels. This technique has two drawbacks: 1) the low signal to noise level for each measurement of a single pixel makes it very difficult to correct the gain and provide a consistent signal response from pixel to pixel, and 2) it is impossible to spatially correlate data across all pixels. For example, if most pixels see no signal at a given point in the sky and a small minority of pixels see a signal, it can be flagged for further investigation or simply rejected as the result of some transient process. Correlating dating from individual pixels could show crosstalk between pixels or oscillation on the input of the spectrometer. By understanding correlated behavior, we can both improve receiver performance and diagnose potential problems.

For many reasons, we have created a new data reduction architecture, called sReduce, that is capable of fully accessing the range of correlations available across multiple heterodyne pixels, further enabling current and future heterodyne array
receivers. Because of the close relationship with the KAPPa and SuperCam teams, we are able to complete this final KAPPa advancement using SuperCam data obtained on telescope for the 310° galactic longitude region of the sky. Since we are building a new data architecture, we can take advantage of modern computing processing power and memory. Additionally we have chosen to write this architecture in Python 2.7, consistent with the KAPPa receiver test code. Aside from being easily exported for a future kilopixel array, coding in Python gives the added advantage of easily importing existing computational tools from the astronomy community and beyond. Additionally, our code is designed to be installed cross-platform and will be made freely available to the astronomy community, anyone who is simply interested, or is assigned to do data reduction as a part of the class project. Please contact the author for source code.

4.2.1 An Introduction To sReduce

The data reduction pipeline for KAPPa, that is currently operating on on-the-sky SuperCam data, is called sReduce, which stands for spectral reduction code. While currently in use for SuperCam, the architecture of the code is general enough to easily handle any type of data that is channelized, including single channel data.

What makes sReduce so flexible is the ability to delete data without deleting an entire data axis. The data structure for SuperCam in sReduce is set up such that each single measurement contains 64 spectra with each spectra containing 900 channels. To take full advantage of any type of correlation, not only time or sky coordinates, we must be able to delete unwanted spectra once they have been flagged for removal. Since the data structure of sReduce does not depend on rectangular arrays, this is a simple process and simply shortens the runtime to
process one spectrum. Channel data is not deleted in this fashion but can be
masked or replaced for individual channels or groups of channels.

The sReduce code is set up as a scripting language. Once data has been
loaded, it has methods that operate on all the data, individual measurements of 64
spectra, or single spectra. These methods can be called in any order to process data
straight from the command line for simple functions such as baselining, co-adding
spectra (per pixel), finding channels of low standard deviation across the spectra
(per pixel), or deleting data. Many of the processes have been collected into a script
that cleans and prepares the data prior to mapping to a rectangular grid for
post-processing analysis. Once the power and flexibility of sReduce is demonstrated,
it is our hope that future contributions will improve sReduce in order to make it a
‘living tool,’ free to all in the astronomy community and the public. Currently the
code is not publicly available because there is only a single read-in class for
SuperCam-specific data, most of which is currently unpublished. We hope that
sReduce will aid in publishing SuperCam data, such that it may become a tool for
rapidly analyzing future SuperCam observations on any users personal computer.

sReduce is not only the data reduction code for SuperCam, it is the future
data pipeline for a kilopixel array receiver. Such future receivers may make maps of
the whole sky. With this in mind, sReduce was designed to reduce data in sections,
or work with entire data sets. sReduce can also run batches of data so that runs
that exceed the random access memory of a computer can be broken into smaller
chunks. Because sReduce has no need for specially shaped arrays, it creates batches
by starting at the extremes of an observed region and finds all nearby points
growing in the next circle from the starting point. Once the user specified file
number limit has been reached, sReduce starts at the other end of the observable
region and grows circles encompassing the remaining group points. This process
eliminates the need to rescale border regions between different scans, since each grid point flags all spectra (the file name and pixel number) within a certain radius to be processed by the regridder. As a result, spectra in overlapping regions are flagged together regardless of whether they were part of the same run or not. Each grid point is processed a single time in a manner consistent with the processing of all other grid points.

4.2.2 Sky Data Acquisition

To take astronomy measurements at a telescope using SuperCam or other similar receivers, we must take additional steps as compared to optical detectors. First, we must integrate for a long time to see faint targets, a common practice to reduce the effect of uncorrelated noise for all astronomy. However, heterodyne receivers are famously unstable, with gain stability changing on the order of seconds. To illustrate this issue, suppose we take two measurements separated by a time much larger than the stability timescale, say 100 seconds, from the same receiver looking at exactly the same target. Intuitively, we would expect the results to be the same within the error of the systematics. However, over that time, the gain – or output power for each unit of input power – of a heterodyne receiver changes by increasing or decreasing, so for each measurement we must apply a unique gain correction. If we were to take one long measurement, we would be unable to apply all the gain corrections and thus lose information. Instead we read out the receiver rapidly, applying the individual gain collections based on calibration targets of known temperature that. In this way, data from a variety of gain responses can be co-added and we can continue to reduce uncorrelated noise. While observing, THz receivers must strike a balance between having as much time as possible on the sky and having a good understanding of the receiver gain from calibration using known
In optical astronomy, sky measurements are often taken while tracking the sky, moving the telescope and rotating the imaging device in such away that the image on the camera stays fixed. Optical astronomers worry about star tracking, optical alignment, and error from a telescope's fine motion controllers. Copying this technique for the THz would be extremely difficult because of our large beam sizes and lack of bright, well-characterized sources to track from. Pointing corrections are often made from planets or other bright radio sources, which requires slewing the telescope far from the intended field of observation. One way to map a large target and minimize systematic errors of the telescope slewing is to stop the telescope at a point on the sky and allow the rotation of the Earth to slew the receiver over the targeted field. Once the rotation of the Earth brings the telescope to the end of the targeted field, we need only move the telescope once to begin the other extreme the target area. This is a form of on-the-fly (OTF) mapping, where images are taken as the telescopes slews from one position to another.

OTF mapping allows for sub-beam spatial resolution and pointing accuracy, although it does create some problems from a data processing perspective. Each pixels beam must now be thought of as an Airy function convolved over the direction of sky rotation. To properly weight the spectrums measurement for a given point near the region that was measured, we must know not only the angle difference from measurement location to the new point, but we must also know the direction and speed of the slew. The slew rate will be determined by the declination of the telescope at the time of measurement, as declinations near the celestial equator will have the fastest slew rate. Currently, the sReduce code only takes the angle into account and uses a simple Gaussian convolution. However we are aware of that this tactic may be overly simplified and we have plans to improve the
sReduce convolution once the beam profiles for each pixel are determined. To make this results faster within the sReduce architecture, we will use look up table for each beam to determine the Airy function and then apply a convolution representing the telescopes slew.

Because the pixels of SuperCam are in a regular grid, the mapping efficiency for points on the sky changes as a function of the grids orientation to the direction of slew during OTF mapping. The sky is more efficiently mapped if the direction of slew is different then the repetition axis of the SuperCam. Put simply, some regions of the sky are poorly measured because they slipped between the SuperCam beams during OTF mapping. As a result, maps made from SuperCam data can very in the amount of noise contained two regions of the map separated by a few beam sizes. For this reason, sReduce can flag regions with a low density of measured spectra, although the exact use of this information, in order to inform analysis on SuperCam data, is still being discussed.

4.2.3 Advanced Baselining Algorithm

Baselining in data reduction is a process in which the noise floor of the spectrum is set to zero for all channels so that any signal in the spectral channels can be measured in a consistent way. This process is an additional cleaning that is done after calibration. As mentioned in Section 4.2.2, the gain of the receiver changes at a characteristic time scale, unfortunately that time scale can be different across each pixel. Some calibrations may not occur at close enough intervals to remove all manner of instability effects from each pixel. Aside from the changing gain, there are many waveforms that can be imposed on a signal spectrum while in transmission as continuous waves on a coax. One problem is standing waves, which can occur in two orthogonal polarizations as the result of reflections from a
mismatched impedances. Additionally there is the possibility of cross-talk from signals on other cables, or even something as simple as a noisy power supply on the IF processing computer. For KAPPA, discussed in at length in Chapters §1 §2 §3 we were concerned with the ‘why’ and ‘how,’ but for SuperCam we must work to clean the data that we have.

For most of the SuperCam data, a very simple polynomial fit is all that is needed to clean the spectra. Using sReduce, we can do a computationally inexpensive linear fit using regions of each pixel’s spectra where there is the lowest amount of variation (the points for weighting the baseline are represented by the blue dots in Figure 4.1), where we suppose that is low variation is cased by lack of signal. We can also subtract larger polynomials, a user control option in sReduce, but we are hesitant to do such a subtraction as there is no physical basis for which we can justify our spectra having high order polynomial behavior. There is also the concern that we would start removing the signal in some pixels when trying to find the appropriate baseline for all functions. Realizing that polynomials are not the only basis set one can use to represent the function of a spectrum, we set out to find a more appropriate basis. The periodic structure of the spectra that were difficult to baseline (see the red spectrum in Figure 4.1) made it appear that something like a sine wave would be appropriate. However the periodic structure was far too jagged to be properly fit by sine waves. We considered fitting with a number of sine waves, but this introduced the fear of creating or destroying signal from constructive interference from multiple wave of different frequency.

We then realized that the shape of the spectra often looked like the radius-time equations of a Keplerian orbit. This was the key idea that allowed us to see the the baselines as the superposition of two orthogonal sine functions being projected onto a single spectrum after the data acquisition. As the IF signal is
Figure 4.1: The advanced baselining algorithm uses a bounded least squares to fit elliptical functions to the SuperCam data spectra. Only spectra that are flagged as poorly fitted by the algorithm are fitted by this aggressive baseliner. The gently curving lines dark green, light green, and black lines are the tier 1, 2, and 3 baseline fits, respectively. The red spectrum is the original spectrum, but does not include points flagged by masking operations. The purple spectrum is the final subtracted spectrum and includes the regions that were masked.

transmitted on a wire, two modes of standing waves can form. Moreover, each electrical transformation of the signal can be thought of as a complex transformation. Classes of functions that are periodic in both complex dimensions, and contain simple poles are called elliptical functions in the study of complex analysis. As an instrument builder, we proceeded conservatively and have decided to use the simple elliptical function using only two orthogonal sine waves that have the same period, which is similar to a Keplerian orbit.

sReduce’s advanced baselining algorithm takes a tiered approach to fitting and subsequently removing a baseline. The structure is not specific to the functions described below and instead uses PYTHON’S scipy.optimize package for a bounded least squares fit (i.e. least_squares). All spectra are initially compared to a simple
linear fit using the parameters of slope, $m$, and y-intercept $b$ such as,

$$y = mx + b$$

(4.1)

where $m$ and $b$ are unbounded free parameters (this first fit is the dark green line in Figure 4.1). For most spectra, the cost of this fit falls under the user specified value, and the baseline is simply subtracted.

When the cost is high, a second tier of fitting is triggered where the free parameters are $m$, $b$, phase $\theta_0$, amplitude $A$, and period $T$ demonstrated as,

$$y = mx \left( A \cos \left( \frac{2\pi x}{T} + \theta_0 \right) \right) + b .$$

(4.2)

Here, $A$ and $\theta_0$ are unbounded and $T$ is bound between 0.3 and 4, where we restrict the number of periods to be less than four so that the signal of CO emission is not confused for a sine peak. In practice, $T$ will be within one harmonic of the initial value used, since the least squares solver uses a gradient method to determine the path to a solution. In the future, a more complex optimization could be applied that takes into account harmonics of solutions. This solution is represented by a light green line in Figure 4.1. At this point, the algorithm accesses if there was a significant benefit to the tier-two fitting over the tier-one linear fit. If there is a 10% improvement in cost, a user specified variable, the algorithm selects the tier two solution for baselining and moves on to tier three.

In tier three, we introduce the eccentricity parameter, $e$. In astronomy and math, this variable has its own definition, but it is more appropriate to define it in this context as the ratio between the amplitudes of two orthogonal sines waves of the same period. The tier three equation is:

$$y = mx \left( A \cos \left( \frac{2\pi x}{T} + \theta_0 \right) \right) \frac{1 - e^2}{1 + e \cos\left( \frac{2\pi x}{T} \right)} + b ,$$

(4.3)
where free parameters are $m$, $b$, phase $\theta_0$, $A$, $T$ and where $e$ is bounded between 0 and less than 1. For an astronomer reading this and looking for an analogue, this is the distance of an object from an observer as a function of time in a Keplerian orbit. For an unbound object, simply moved closer to the observer, we achieve Equation 4.1. For an object moving in a circular orbit around another body but still linearly moving away from the observer, it is represented in Equation 4.2. And finally we add the additional perturbation in with respect to an eccentric orbit to order to determine Equation 4.3. The tier three baseline fit is selected if there is again a 10% improvement in the cost of the fit compared to the sine wave fit. In Figure 4.1 we can see that the tier 3 fit was selected because it has been plotted with the blue markers, indicating it as the baselined function to be subtracted from the red (original) spectra resulting in the purple (baselined) spectra.

For an electrical engineer, there is no analogue to an indefinitely receding object inside the circuit. In this case, it could be more satisfying to consider the linear part of the equation as a very long period sine wave for which we are in the regime where a linear approximation is acceptable. However, there is a far more practical reason for including this feature: the linear part allows a simple baselining approach for all spectra until such time as the Advanced Baseline algorithm is needed at a later step, then the algorithm can easily undo a poor linear fit from the earlier baseline. This makes the sReduce code more modular and better to use as scripting language.

4.2.4 Automatic Rejection Using Spatially Correlated Data

As seen in Section 3.3.3 the KAPPA team is quite fond of data processing techniques developed for advance radar applications. Another simple but novel adaptation for THz astronomy is to match-filter spectra at the regridder. While
most of sReduce is designed for flexibility, the regriddrer has been designed for clarity and speed. For a single new grid point, the regriddrer finds all spectra within the user specified radius. However, despite the proceeding algorithms best efforts there are still many spectra that will add more noise than signal to this new grid point. We realized that the spectra could be easily flagged by eye because they are very different from the typical spectra for that grid point. By simply adding all the spectra together, we can make an average spectrum and compare each constituent spectra to this average. Spectra above a user-specified threshold can be rejected before polluting the final grid point.

Once the initial match-filter was in place, it’s been a simple process to upgrade it. Currently, the message filter takes into account the spatial weighting, giving spectra nearest to the beam center a higher priority. Additionally the user can select only a fraction of the available spectra with the lowest $T_{sys}$ or lowest residuals after baselining. New filters and can implemented quickly, as the regriddrer is the most modular and well documented part of the sReduce code. The future match-filter may use library files to improve computational speed but this comes with the danger of biasing the data with expected results.

4.3 Creating Galactic Maps

We can use SuperCam data to make three dimensional (3D) maps of the $J=3 \rightarrow 2$ rotational transition of CO throughout the galaxy (see Section 1.2.1). Schematically, this is done by attributing the frequency shift of the $J=3 \rightarrow 2$ rotational transition from the rest frame to the difference in velocity from our galaxy’s rotation.

We begin this process by determining the line of sight velocity for each frequency in the SuperCam IF band. The shifted frequency, $f$, is given by the
doppler formula,

\[ f = \left(1 + \frac{\Delta v}{c}\right)f_0, \quad (4.4) \]

where \( f_0 \) is the rest frequency of the line emission, \( c \) is the speed of light, and \( \Delta v \) is the line of sight (LOS) velocity difference between the emitter and the observer.

The LOS velocity, in \( km/s \), is the preferred axis, as opposed to channel number or frequency, for working with SuperCam data and can be seen in the Figure 4.1.

Next we must take into account the Sun’s relative motion throughout the galaxy so that we can find the velocity relative to the rest frame of the Milky Way, hereby referred to as the local standard of rest (LSR). Using Equation 1 from \textit{Bhattacharjee et al. (2014)} we see that the LSR velocity \( v_{\text{LSR}} \) can be calculated from the LOS velocity \( v_{\text{LOS}} \) as

\[ v_{\text{LSR}} = v_{\text{LOS}} + U_\odot \cos b \cos l + V_\odot \cos b \sin l + W_\odot \sin b, \quad (4.5) \]

where \( l \) is the galactic coordinate for longitude, \( b \) is the galactic coordinate for latitude, and \((U_\odot, V_\odot, W_\odot)\) are the relative velocities of the Sun compared to the LSR within a Cartesian coordinate system (see Figure 4.2). We take \((U_\odot, V_\odot, W_\odot) = (11.1, 12.24, 7.25) km/s\), consistent with \textit{Bhattacharjee et al. (2014)} from \textit{Schönrich et al. (2010)}.

Using the results from \textit{Binney & Merrifield (1998)}, we write the circular velocity \( V_c(R) \) of an object in orbit of the Milky Way. We define \( V_c(R) \) as a function of distance from the galactic center \( R \) at an observed LSR velocity \( v_{\text{LSR}} \), namely:

\[ V_c(R) = \frac{R}{R_0} \left[ \frac{v_{\text{LSR}}}{\sin l \cos b} + V_0 \right]. \quad (4.6) \]

Here, \( R \) is the projection of the observed object’s distance from the galactic center, \( r \), onto the equatorial plane as seen in Figure 4.2. We know that \( R \) can easily be calculated from the law of cosines with the knowledge that \( \cos(b) = \frac{A}{r_h} \) (see Figure 136).
Figure 4.2: A reference drawing showing the relationship between galactic coordinates $l$ and $b$, the relative velocity of the Sun $(U_\odot, V_\odot, W_\odot)$ and galactic radius $R$, distance of source to the galactic center $r$, distance of the Sun to the source $r_h$, circulate orbital velocity of the sun $V_0$, and distance of the Sun to the galactic Center $R_0$.

From Figure 4.2 and Equations 4.6 and 4.7, $r_h$ is the distance from our Sun to the observed object, $R_0$ is the distance from our Sun to the galactic center, and $V_0$ is the Sun’s circular rotation speed. Staying consistent with Bhattacharjee et al. (2014), we use values of $(R_0, V_0) = (8.3 \, kpc, \, 244 \, km/s)$ based on observations of masers and stars around Sagittarius A (Bovy et al. 2009; Gillessen et al. 2009). We can rearrange Equation 4.6 as

$$\frac{V_c}{R} = \frac{1}{R_0} \left[ \frac{u_{LSR}}{\sin l \cos b} + V_0 \right], \quad (4.8)$$
a ratio which can be calculated from SuperCam observations and Equations 4.4, 4.5, and 4.6. A \( V_c/R \) ratio can be calculated for each spectral channel and each measurement of SuperCam data, but the results are highly dependent on the values used for \((U_\odot, V_\odot, W_\odot)\) and \((R_0, V_0)\).

Using the values for the galactic rotation curve from Bhattacharjee et al. (2014), we can plot \( V_c/R \) as a function of \( R \) as seen in Figure 4.3. Unfortunately, we do not live in a galaxy where \( V_c(R)/R \) is a one-to-one function and thus the inverse function to calculate \( R \) is ambiguous. To fully determine the distance to the emitters of the CO J=3 \( \rightarrow \) 2 line emission, we require additional information. One possibility is explored in Schlafly et al. (2014) where dust redding along the LOS is used to infer the distance of GMCs. From Figure 4.3 we can see that ambiguity in galactic radius occurs at 15\( kpc \) and 18\( kpc \). Additionally, \( V_c(R)/R > 25 \ kpc \) begins to flatten out, making measurements of precise galactic radius more difficult for larger galactic radii.

![Figure 4.3: The ratio of galactic circular velocity \( V_cR \) and galactic radius \( R \) as a function of galactic radius \( R \). Data is selected from the \( V_c(R) \) table values in Bhattacharjee et al. (2014).](image)
Despite these uncertainties, we can continue toward our map of 3D CO emission for observed GMC in the Milky Way by using Equation 4 from Bhattacharjee et al. (2014). Here we use the Cartesian coordinates for an observed source of line emission as

\[
x = r_h \cos(b) \sin(l)
\]

\[
y = R_0 - r_h \cos(b) \cos(l)
\]

\[
z = r_h \sin(b),
\]

where \( l \) and \( b \) are the galactic coordinates.

The remaining task is to solve for Equation 4.7 for \( r_h \). We will go through the algebraic steps below and arrive at a solution that can be easily implemented and is more intuitive than the standard representation. Let us begin with Equation 4.7, where

\[
R = \sqrt{R_0^2 + r_h^2 \cos^2(b) - 2R_0r_h \cos(b) \cos(l)}
\]

\[
\implies R^2 = R_0^2 + r_h^2 \cos^2(b) - 2R_0r_h \cos(b) \cos(l)
\]

\[
\iff 0 = \cos^2(b)r_h^2 - 2R_0 \cos(b) \cos(l)r_h + R_0^2 - R^2.
\]
In this format, we can apply the quadratic formula and solve for \( r_h \) as

\[
\begin{align*}
    r_h &= \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \\
    r_h &= \frac{2R_0 \cos(b) \cos(l) \pm \sqrt{4R_0^2 \cos^2(b) \cos^2(l) - 4 \cos^2(b)(R_0^2 - R^2)}}{2 \cos^2(b)} \\
    r_h &= \frac{2R_0 \cos(b) \cos(l) \pm 2 \cos(b) \sqrt{R_0^2 \cos^2(l) - R_0^2 + R^2}}{2 \cos^2(b)} \\
    r_h &= \frac{R_0 \cos(l) \pm \sqrt{R_0^2 \cos^2(l) - 1} + R^2}{\cos(b)} \\
    r_h &= \frac{R_0 \cos(l) \pm \sqrt{R^2 - R_0^2 (1 - \cos^2(l))}}{\cos(b)} \\
    r_h &= \frac{R_0 \cos(l) \pm R_0 \sqrt{\left(\frac{R}{R_0}\right)^2 - \sin^2(l)}}{\cos(b)} \\
    r_h &= \frac{R_0 \left(\cos(l) \pm \sqrt{\left(\frac{R}{R_0}\right)^2 - \sin^2(l)}\right)}{\cos(b)} \\
    r_h &= \frac{R_0 \cos(l)}{\cos(b)}.
\end{align*}
\]

This result is the generalization of Equation 2 in \textit{Roman-Duval et al.} (2009) for galactic latitude, \( b \neq 0 \), and is most recognizable from the 3rd to last step in Equation [4.11]. The resulting \( r_h \) is valid for all of \( R \), but there are several cases to consider before implementing the method.

First, let us consider the case where \( \left(\frac{R}{R_0}\right)^2 - \sin^2(l) = 0 \), where emitters are moving tangentially to LOS. In this instance, we have a unique solution solution for \( r_h \) namely,

\[
    r_h = \frac{R_0 \cos(l)}{\cos(b)}.
\]

Sources of line emission that satisfy this solution are at the point tangent to the LOS on their circular path around the galaxy, such that there is a right angle between the galactic center and the Sun at the emitter. Since the \( V_c(R) \) can be uniquely determined at points that \( R^2 - R_0^2 (1 - \cos^2(l)) = 0 \), this is often used to determine...
the rotational velocity of objects within our galaxy (Binney & Merrifield 1998).

The result of Equation 4.11 is valid for $R < R_0$ and $R_0 \leq R$ but gives two solutions for distance for most cases. For $R < R_0$, there is yet another source of ambiguity and it is why maps of galactic structure require additional constraints to determine a unique distance. The two set of solutions can be seen in Figure 4.4. While $R_0 < R$ gives two solutions as well, one solution is in front of the observer on the LOS and one solution is behind the observer eliminating the ambiguity. The sReduce code is designed to return far solutions by default but near solutions can be selected as in Figure 4.4.

For the case of $R^2 - R_0^2 (1 - \cos^2(l)) < 0$, the solutions for $r_h$ become imaginary. This could be the result of the particular velocities of a GMC deviating from the assumptions above. Equally, possible is that this is a noisy spectral channel that appeared to be CO emission. sReduce is programed to reject such solutions before creating a map of the galaxy.

Based on this we can turn spectra of CO emission in to a spatially correlated map of the Galaxy. A small fraction of the available SuperCam data has been reduced in this manner and can be seen in Figure 4.5. The columns of structure correspond to the spectral lines of CO J:3 $\rightarrow$ 2 transition. The average spectra for this map can be seen in Figure 4.6. Both figures are made from three separate SuperCam observation runs spanning over $-0.35^\circ < b < 0.25^\circ$ in galactic latitude and $309.9^\circ < l < 310.4^\circ$ in galactic longitude. This narrow region gives a tube like appearance, to show more more structure the z-axis of Figure 4.5 has been stretched giving a more oval appearance when viewed from the Sun’s location of $(X,Y,Z) = (0,8.3,0)$ kpc. Since this is within $90^\circ$ of the galactic center at $l = 0^\circ$, we can are unable to restrict our solutions to be within or outside the Sun’s orbit of the galaxy, Figure 4.5 shows the ‘far’ solutions (as labeled with the green line in
Figure 4.4: Using the sReduce analysis package, we can calculate all velocities with emission in an observed SuperCam spectrum in a single command. The output of this command is shown here and demonstrates the cases for \( r_h \) solutions. This plot is generated using the \( V_c(R) \) table data from Bhattacharjee et al. (2014), the parameters data for \((U_\odot, V_\odot, W_\odot)\) from Schönrich et al. (2010) and \((R_0, V_0)\) from Bovy et al. (2009); Gillessen et al. (2009) as mentioned in the text.

For perspectives within \(90.0^\circ < l < 270^\circ\), the usual ambiguity of solutions is avoided.
Figure 4.5: The units of scale are in kpc. The Z axis has been stretched to show more structure, thus the metric is in in inflated in z. The color scale is in Kelvin above the background radiation of the sky. This coordinate system is consistent with Bhattacharjee et al. (2014).
Figure 4.6: This is the average spectrum of the data in Figure 4.5, the Y-axis is in Kelvin above the sky temperature, and the x axis is CO J:3 → 2 LOS velocity in km/s.
Chapter 5

SIMULATIONS TO DETERMINE THE EFFECTS COSMIC DENSITY
VARIATIONS IN THE GLOBAL 21CM SIGNAL

The 21cm emission is often seen from the same cold, dense gas clouds that produce the CO emission. GMCs (see Section 1.2.2) can be investigated at different optical depths using both CO and \( \text{H}_2 \), giving a view of gas clouds that are opaque at higher frequencies.

5.1 Introduction to 21cm Astronomy

The ionization state of hydrogen marks three phases in the evolution of the Universe. The first transition is from a universe filled with ionized hydrogen (HII) to one with neutral hydrogen (HI), which is observed from measurements of the cosmic microwave background, or CMB, occurring at \( z \sim 1090 \) \(^{(\text{Komatsu et al., 2009})}\). When hydrogen became neutral and the energy density of the Universe continued to decrease, the first baryonic structures formed. These structures continued to collapse, and thus, the first stars and eventually galaxies were formed. Such early structures emitted radiation capable of ionizing HI in the universe. Prior to \( z \sim 6 \), observations have shown that the Universe had been almost completely reionized \(^{(\text{Fan et al., 2006})}\). The transition period during which hydrogen went from predominately neutral to the current state of predominately ionized the epoch of reionization (EoR).

Unlike the transition that created the CMB, EoR remains a frontier in the astronomy. Current observations can place some important constraints on the timing of EoR and the sources of the ionizing photons. Direct observation of the 21
cm line of HI (21cm signal) during EoR would provide astronomers a powerful tool for investigating the structure formation in the early universe and bridge a gap in our understanding of the cosmos. Astronomers are rapidly approaching the ability to make observations of the redshifted 21cm signal, namely the EDGES project (Mozdzen et al., 2016) as well as DARE (Burns et al., 2012).

These initial observations of the redshifted 21cm signal will be averages over many comoving Mpc. As we investigate these global averages of the 21cm signal, we keep in mind that the averaging forces an additional degeneracy at a single observed redshift, namely we are unable to distinguish between a few bright sources versus a uniform signal over the observed region. Different physical processes would drive these two cases, but would produce the same signal power. To discuss this degeneracy further, we must first investigate the 21cm signal from a region that can be described by a single set of physical parameters.

An analysis of the 21cm signal is given in Pritchard & Loeb (2010) as a differential brightness temperature compared to the CMB at a given redshift, and is reproduced in Eq. 5.1 by:

$$T_b(z, \vec{\theta}) = 27 x_H \left( \frac{T_s - T_{\text{CMB}} (1 + z)}{T_s} \right) \left( \frac{1 + z}{10} \right)^{1/2} \times (1 + \delta_b) (1 + \delta_x) \left( \frac{\partial_r v_r}{(1 + z) H(z)} \right)^{-1} \text{mK},$$

where $x_H$ is the neutral fraction of hydrogen, $\delta_x$ is the fractional variation in the neutral fraction of hydrogen, $\delta_b$ is the density in baryons, $T_s$ is the 21cm spin temperature, $T_{\text{CMB}} = 2.728$K is the CMB temperature at $z = 0$, $H(z)$ is the Hubble parameter, and $\partial_r v_r$ is the time derivative of the velocity of an emitter along our line of sight. The baryon density, $\delta_b$, is defined in terms of the average density, $\rho_{\text{ave}}$, and the density of a region, $\rho$, via $\delta_b = (\rho - \rho_{\text{ave}})/\rho_{\text{ave}}$. Overall, Eq. 5.1 expresses the
strength of the 21cm signal as a differential brightness temperature, compared to the CMB. A brightness temperature of 0 means that the 21cm signal is indistinguishable from the CMB. Negative brightness temperature corresponds to a 21cm absorption feature in the CMB signal, while positive brightness temperatures shows emission. In our present work, we consider that global average of the brightness temperature which is defined as $T_b(z) = \langle T_b(z, \hat{\theta}) \rangle_{\hat{\theta}}$. It is the global average, $T_b(z)$, that will be measured by the first generation of instruments.

Analytic and semi-numerical calculations have approximated have calculated $T_b(z)$ with increasing sophistication. Because of the large number of unconstrained parameters during the EoR, many different models are needed to properly explore possible histories of the EoR. The work in Furlanetto (2006) and Sethi (2005) provide models that can be used to predict the power of the 21cm signal as a function of redshift. Both papers analyze the various physics and processes that effect the 21cm signal, Furlanetto (2006) provides a discussion of possible ionizing sources. A more detailed analysis is carried out in Pritchard & Loeb (2008) where the impact on the global 21cm signal by small scale fluctuations.

The first observations of the 21cm signal will be from regions on the order of 100 Mpcs. Simulations capable of modeling these scales are required to capture smaller scale fluctuations while simulating a region much greater than the Jean’s length. A variety of simulation tools exist for modeling such scales. We refer the reader to Zahn et al. (2011) for a comparison of four such simulations.

At the onset of this work, it was not clear weather global avenges of $x_H$ and $\delta_b$ are truly representative of the many regions that contribute to that average. Work done by Aubert & Teyssier (2010), using their radiative transfer code to analyze regions as large as 100 Mpc with $1024^3$ cells, shows that the average neutral fraction lies in between two characteristic regions of gas and is not representative of
During the EoR, the dominate terms in Eq. 5.1 are the neutral fraction of hydrogen, $x_H$, and the density in baryons term, $(1 + \delta_b)$. For this analysis, we focus on these two terms, the product of which we define as the unitless value $S_{21cm}$:

$$T_b \propto S_{21cm} = x_H (1 + \delta_b).$$

(5.2)

Here we use the 21cmFAST model (Mesinger et al., 2011) to create simulations that calculate the expected global average of the 21cm signal $\langle S_{21cm}(z) \rangle$, or the value of the 21cm signal as a function of redshift averaged over the whole sky. Historically, this value has been estimated from analytic treatment that required the assumption that $\langle S_{21cm}(z) \rangle = \langle x_H \rangle (1 + \langle \delta_b \rangle)$. Since matter is conserved, $\langle \delta_b \rangle = 0$, the estimation becomes: $\langle S_{21cm}(z) \rangle = \langle x_H \rangle$.

However, at the EoR, this assumption requires further examination. The source of ionizing photons may come from Population II or III stars, quasars, or some combination thereof. These ionizing sources are correlated with baryon density, $(1 + \delta_b)$. In contrast, we do not expect to see any ionizing source in under-dense regions where $(1 + \delta_b) < 1$. As a result, ionization will first occur in the most dense regions, making the most rarefied regions the last to be ionized.

If we consider the universe during EoR when $x_H = 0.5$, half of the total hydrogen in the universe is ionized, but the ionized regions make up the most dense parts of the universe. The average density of the neutral hydrogen gas, which is capable of emitting 21cm radiation, is less than $(1 + \langle \delta_b \rangle) = 1$. Therefore, the $\langle S_{21cm}(z) \rangle = \langle x_H \rangle$ estimate over-approximates the 21cm signal. Our anecdote falls short when self shielding of ionizing sources is considered, but still outlines our motivation for investigation.

The simulations we present here are based on a ΛCDM cosmology with
\[ \sigma_8 = 0.82, \, h = 0.7, \, \Omega_m = 0.28, \, \Omega_\Lambda = 0.72, \, \Omega_b = 0.046, \, \Omega_0 = 1, \, Y_{\text{He}} = 0.24, \] and a power law index of \( n_S = 0.96 \). These have been chosen to be consistent with the constraints from WMAP (Komatsu et al., 2009).

### 5.2 Calculation of the Global 21cm Signal

To estimate \( \langle S_{21\text{cm}(z)} \rangle \) during EoR we need to create a simulation with sufficiently large volumes that evolves as a function of redshift. The volume must be much larger than the density fluctuations as well as the Jean’s length but must contain cells small enough to resolve the density fluctuations. To accomplish this task, we create a cubic simulation space within 21cmFAST (Mesinger et al., 2011) from which we are able to output data cubes of the density in matter, \((1 + \delta_m)\), and the neutral fraction of hydrogen, \(x_H\). Note that this code treats dark matter and baryons together as a single entity when calculating the density matter density field, so we switch from \((1 + \delta_b)\), as seen in Eq. 5.1, to \((1 + \delta_m)\) to more accurately reflect the quantity we are using. We expect these to be highly correlated on the scale of our cell size. From our data cubes, we determine the signal from each \(i^{th}\) cell in our simulation via \( S_{21\text{cm}}^i = x_H^i (1 + \delta_m^i) \). We calculate the global average 21cm signal as the mean value of \( S_{21\text{cm}}^i \) over all cells in our simulation, which we denote with a bold ‘S’ or \( S_{21\text{cm}} \), as shown in Eq. 5.3:

\[
S_{21\text{cm}} = \left( \frac{\sum_{i=1}^{N} x_H^i (1 + \delta_m^i)}{N} \right).
\] (5.3)

We will compare this to the analytic result, denoted with a prime, ′ or \( \langle S'_{21\text{cm}} \rangle \), as shown in Eq. 5.4:

\[
\langle S'_{21\text{cm}(z)} \rangle = \langle x_H \rangle (1 + \langle \delta_m \rangle) = \langle x_H \rangle.
\] (5.4)
5.2.1 Consideration for Using 21cmFAST

The number of cells in our 21cmFAST simulation is limited by the available random access memory (RAM) of the machine on which the code is run. The recommended resolution for cell size is smaller than 1 Mpc on a side. With these considerations, the simulations we run have $1350^3$ cells within a simulation volume of 200 Mpc on a side for the initial density field. The code then interpolate the initial cells to produce $450^3$ cells (these new cells are three times as long on a side as the cells from the initial conditions and therefore represent 27 times the volume of the original cell) but with the same box length, for subsequent numerical calculations. The simulation spaces have periodic boundary conditions which serves two functions: 1) to close the system such that many thermodynamic arguments are valid, and 2) to help make the simulation valid at the edges of our simulation space.

A hindrance to large-scale numerical cosmological modeling is the huge range of relevant scales. Analysis of the global 21cm signal requires simulation boxes on the order 100 Mpc. However, reionization is driven by ionizing photons that come from early generations of stars; star formation occurs in regions that span $\sim 50$pc. The smallest scales involved in this process is the cross section of a single hydrogen of a single atom by an ionizing photon.

Some approximations are required to unite the physics of the large-scale behavior of small-scale physical processes. The 21cmFAST simulation approximates physical processes in the scale of a unit cell. It also numerically computes the evolution of the cells with reference to the parameters of the individual cell and the parameters of its neighbors. Additionally, due to the semi numerical nature of the code, the cells in our simulation boxes must undergo Gaussian smoothing to alleviate cell effects from the analytic treatment within cells (Mesinger et al. 2011).
Table 5.1: Parameters for Simulations 1, 2, & 3

<table>
<thead>
<tr>
<th></th>
<th>$T_{\text{vir,min}}$ (K)</th>
<th>$\zeta$</th>
<th>$R_{\text{max}}$ (Mpc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation 1</td>
<td>$1 \cdot 10^4$</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>Simulation 2</td>
<td>$5 \cdot 10^5$</td>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td>Simulation 3</td>
<td>$1 \cdot 10^4$</td>
<td>30</td>
<td>3</td>
</tr>
</tbody>
</table>

Matter and dark matter are treated as a single entity, as mentioned in Section 5.2. However, matter has an associated pressure that resists collapse, while dark matter does not. However, dark matter is the dominate form of matter and determines the gravitational potential. In the resolution of this simulation, baryonic matter experiences pressure on scales smaller than the simulated cells. In interest of computational speed, dark matter and baryonic matter can be treated together within this code.

5.2.2 Parameter Space Investigated

Many cosmological parameters, such as the Hubble constant, critical energy densities of matter radiation, the cosmological constant, and the fraction of helium to hydrogen created in the Big Bang, are now known to such accuracy that they can be treated as fixed parameters for our simulation. However, the source of ionizing radiation is not well known; it is currently thought to be from the first or second generation of stars. There is also uncertainty in the number of ionizing photons per stellar baryon. We address these uncertainties by doing three simulations, manipulating the three variables described below. The values we use for each of our three simulations can be seen in Table 5.1.

We first analyzed $T_{\text{vir,min}}$, which sets the minimum virial temperature, in
Kelvin, for a region to be counted as contributing to the production of ionizing photons. Recall from the virial theorem $2\langle T \rangle = \langle U \rangle$, or that twice the kinetic energy is equal to the potential energy. For very large systems of particles, we replace the kinetic energy of all particles with a characteristic temperature, using $T_{\text{vir}, \text{min}}$ to describe the state of collapse of a system via the virial theorem. Regions that have a virial temperature less than $T_{\text{vir}, \text{min}}$ are thought to be too loosely bound to survive an external ionizing field that is present during the EoR. Work by Dijkstra et al. (2004) suggests that $T_{\text{vir}, \text{min}} < 10^4$ K during the EoR.

Another free parameter is the ionization efficiency factor, $\zeta = m_{\text{ion}}/m_b$, where $m_{\text{ion}}$ is the ion mass and $m_b$ is the mass in baryons. The efficiency factor can also be calculated as $\zeta = f_{\text{esc}} f_s N_{\gamma/b} (1 + n_{\text{rec}})^{-1}$, where $f_{\text{esc}}$ is the escape fraction of ionizing photons, $f_s$ is the star formation efficiency, $N_{\gamma/b}$ is the number fraction of ionizing photons per stellar baryon, and $n_{\text{rec}}$ is the typical number of times a hydrogen atom will recombine (Furlanetto et al., 2004). In our code, we fix the value of $\zeta$, since the decomposition only offers more unconstrained variables. From Eq. 2 found in Furlanetto et al. (2004), a region can be considered fully ionized if the collapse fraction $f_{\text{coll}} \geq \zeta^{-1}$. Combining this and the definition of the collapse fraction, we determine Eq. 5.5:

$$\zeta^{-1} \leq f_{\text{coll}} = \text{erfc} \left[ \frac{\delta_c(z) - \delta m}{\sqrt{\sigma^2_{\text{min}} - \sigma^2(m)}} \right]$$

(5.5)

where $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) \, dt$, where $\delta_c(z)$ is the critical density for collapse, $\delta_m$ is the mean linear overdensity in the region in question, $\sigma(m)$ is the variance of density on scale m, and $\sigma^2_{\text{min}} = \sigma^2(m_{\text{min}})$.

Finally, the parameter $R_{\text{max}}$ sets the maximum size of an ionizing bubble of radiation, in Mpc. In practice, the maximum size of ionized bubble is determined to
be the mean free path of ionizing photons. This parameter simply stops the growth of ionization bubbles at a maximum radius. We can set this value to a low number to simulate a universe that is has a large optical depth for ionizing photons, for example a dusty universe.

5.2.3 Simulations

We do three characteristic simulations, the parameters of which can be seen in Table 5.1. Simulation 1 is our fiducial model, using the same parameters as outlined in §3.3 of Mesinger et al. (2011). For Simulation 2, we set parameters to allow for large mean free paths of ionizing photons, with very rare, $\zeta = 200$, but massive ionization bubbles, $T_{\text{vir,min}} = 5 \cdot 10^5$ K. The parameters of this simulation were set to produce rapid ionization of the universe by a few ionizing sources. Simulation 3 is the other extreme: by restricting the size of the ionization bubbles, but keeping $T_{\text{vir,min}} = 10^4$ K and $\zeta = 30$ as in our fiducial model, we expect to see a slow reionization made by many small bubbles.

All of our simulations have been seeded with a density field of $1350^3$ cells over a distance of $200^3$ Mpc$^3$ at a redshift of $z = 300$. Our fiducial model, Simulation 1, has a smaller number of cells, $450^3$, due to interpolation from the initial conditions, but covers the same volume of $200^3$ Mpc$^3$ starting at a redshift of $z = 120$. In addition to a density map, at this redshift we also calculate simulation volumes that tabulate the acceleration due to fluctuations in the density field in three orthogonal axes, namely $\partial v/\partial x$, $\partial v/\partial y$, and $\partial v/\partial z$. From the density and velocity volume, a temperature volume is created. From the temperature, we calculate the neutral fraction of hydrogen, $x_H$. From a redshift $z = 120$ down to $z = 1$, the simulation recalculates the density from the previous redshift’s density and acceleration in $\Delta z = 0.25$ steps. For each redshift step, we extract the density
field, the three acceleration components, and $x_H$. For Simulation 1, this means $5$ arrays $\times 477$ redshift steps $\times 450^3$ double precision cells per box.

The difference in the parameters from Simulation 1 and subsequent Simulations 2 & 3 only effect the models during EoR. With this in mind, we started Simulations 2 & 3 at $z = 60$, well before any hints of reionization. We start all three simulations with the same initial Gaussian field, which determines the initial density fluctuations. As a consequence, the three simulations are identical, within numerical error, until reionization. This can be seen in the bottom three panels of Fig. 5.7 that show $S_{21\text{cm}}$ at $z = 20$. For Simulations 2 & 3 we save the same 5 parameters as Simulation 1 in $\Delta z = 0.25$ step for $z \in [60, 1]$. It should be noted that the simulations are only valid from redshifts greater than $z = 3$, but reionization for all simulations ends prior to this redshift at $z \sim 5$.

5.3 Results of 21 cm Analysis

Our simulations resulted in over one terabyte of data, which we have culled into a series of plots. At every redshift step, $\Delta z = 0.25$, we find the average of the neutral fraction, $\langle x_H \rangle$, and the density, $\langle \delta_m \rangle = 0$ per cell to calculate $S'_{21\text{cm}} = \langle x_H \rangle \cdot (1 + \langle \delta_m \rangle) = \langle x_H \rangle$, or the current analytical approximation as outlined in Section 5.2. At the same time, we do cell by cell multiplication of $x_H$ and $(1 + \delta_m)$. We then calculate the average of the product, or $S_{21\text{cm}}$ from Eq. 5.3. We present these plots for Simulations 1, 2, & 3 in Figs. 5.1, 5.2, & 5.3, respectively. In all three simulations, $S_{21\text{cm}}$ is below $S'_{21\text{cm}}$ during EoR confirming our hypothesis that the analytic estimations of 21cm signal over-estimate the brightness temperature $T_b$.

In Fig. 5.4 we show the difference $S'_{21\text{cm}} - S_{21\text{cm}}$ for all three simulations. We find that the redshift space for which the 21cm signal deviates from the $S'_{21\text{cm}}$
varies for each model. The amount of deviation is model-dependent. However, each model only deviates from the estimation by a maximum of 10% at some redshift. Measuring this 10% deviation in brightness temperature of a few mK is beyond the ability of current instruments. But this correction will be of vital importance for understanding the first observations of the redshifted 21cm signal and the physics that produced it.

To compare these measurements in a more quantitative way, we express $S'_{21\text{cm}} - S_{21\text{cm}}$ as function of $x_H$, shown in Fig. 5.5. From this plot we see that all three simulations start with the same slope when the neutral fraction is 1, however there are three different slopes when each simulation reaches $x_H = 0$. The interpretation of these slopes is that ionization in all three simulation starts in the same way, but as the simulations evolve the differences in their respective parameters begin to be asserted. The simulations, in order of largest to smallest final slope, are: Simulation 3, 1, & 2.

In Fig. 5.5 the maxima of all three curves are at three different values of $x_H$. In order of smallest to greatest neutral fraction for the peak value of $S'_{21\text{cm}} - S_{21\text{cm}}$, we have Simulation 3, 1, & 2. It should also be noted that the only intersection for any of the simulations in Fig. 5.5 occurs at the $x_H = 0.0$ and 1.0. As a result, Simulation 3 is strictly greater than Simulation 1, which in turn is strictly greater than Simulation 2. The trends in this plot suggest that parameters intermediate to Simulations 2 & 3 would yield a curve that was intermediate of the curves created by Simulations 2 & 3 in Fig. 5.5, thus the curves of Fig. 5.5 representative a family of curves.

While Figs. 5.1 – 5.5 demonstrate the phenomena we set out to study, they fail to fully capture the differences in the evolution in each of the three simulations. In Figs. 5.6 and 5.7 we show the z-axis averaged $S_{21\text{cm}}$ for each of the three
simulations as a function of redshift. Since all three simulations start from the same density field, their z-axis averaged $S_{21\text{cm}}$ maps are identical prior to $z = 20$. From these figures we can see that Simulation 2 takes the longest time to begin reionization at $z = 10$. Since Simulation 2 requires a high virial temperature to start ionization, it must wait to be sufficiently collapsed before reionization can begin.

Simulations 1 & 3 are nearly identical until $z = 10$, when the ionization bubbles in Simulation 1 grow greater than those in Simulations 3, with a limiting value of 3 Mpc. Simulation 1 finished ionization first prior to $z = 8$, while Simulations 2 & 3 both finished reionization around $z = 5$. Simulation 2 had the fastest reionization and Simulation 3 had the most gradual.

To aid in structural comparisons for each of the three simulations we present z-axis averaged $S_{21\text{cm}}$ maps of constant $x_H$ for all three simulations in Fig. 5.8. While the simulations shown in Fig. 5.8 are all of constant $x_H$, Simulation 2 has the greatest $S_{21\text{cm}}$. This simulation is capable of having dense regions that are not dense enough such that the virial temperature becomes greater than $5 \cdot 10^5$ K and thus are not counted as contributing ionizing photons, see Section 5.2.2 for details. Most striking in Fig. 5.8 is the differences in dynamic range of the $S_{21\text{cm}}$ in each of the three simulations. Simulation 2 has the greatest dynamic range at all three values of $x_H$, while Simulation 3 is relatively homogenous over the range of $x_H$. The homogeneity of Simulation 3 arises from reionization being a product of many small ionization bubbles. Note that Simulation 2 & 3 both have $x_H = 0.25$ at around $z = 7$, but they have a totally different spatial dependence of $S_{21\text{cm}}$.  

156
Figure 5.1: Fiducial simulation, Simulation 1.
Figure 5.2: Simulation 2, which has the fastest reionization. It was ionized by large ionization bubble that are seeded only in the densest regions.
Figure 5.3: Simulation 3, with the most gradual reionization. It was ionized by small, 3 Mpc, ionization bubbles. This simulation had the great deviation from $S'_{21\text{cm}}$. 

\[ \chi_H (1 + \delta_m) \]

\[ \langle S'_{21\text{cm}} \rangle \]

\[ S_{21\text{cm}} \]
Figure 5.4: $\langle S'_{21\text{cm}} \rangle - S_{21\text{cm}}$ as function of redshift $z$ for all three simulations. Simulation 1 (or Sim 1) is demarcated as the solid cyan line, Sim 2 is the dashed pink line, and Sim 3 is the wider spaced dashed green line.
Figure 5.5: The $\langle S'_{21\text{cm}} \rangle - S_{21\text{cm}}$ as function of neutral hydrogen $x_H$ for all three simulations. The color coding is similar to Fig. 5.4.
Figure 5.6: The z-axis averaged $S_{21\text{cm}}$ maps for Simulations 1, 2, & 3 as function of $z$. The columns from, left to right, are Simulations 1, 2, & 3, respectively. All scales are the same, where dark red is the largest values $S_{21\text{cm}} > 1$, and dark blue is the smallest values $S_{21\text{cm}} = 0$. Rows are constant in redshift $z$. The neutral fraction $x_H$ is also given in each panel.
Figure 5.7: The z-axis averaged $S_{21\text{cm}}$ maps for Simulations 1, 2, & 3 as function of $z$. The columns from left to right, are Simulations 1, 2, & 3. All scales are the same, dark red is the largest values $S_{21\text{cm}} > 1$, and dark blue is the smallest values $S_{21\text{cm}} = 0$. Rows are constant in redshift $z$. The neutral fraction $x_H$ is also given in each panel.
Figure 5.8: The $z$-axis averaged $S_{21\text{cm}}$ maps for Simulations 1, 2, & 3 as function of $x_H$. The columns from, left to right, are Simulations 1, 2, & 3, respectively. All scales are the same, dark red is the largest values $S_{21\text{cm}} > 1$, and dark blue is the smallest values $S_{21\text{cm}} = 0$. Rows are constant in neutral fraction $x_H$. The redshift $z$ is also given in each panel.
Chapter 6

FUTURE WORK

The KAPPa project, and its various extensions, has successfully advanced the field of large THz heterodyne array receivers. However, builders of the next kilopixel array will still face some considerable obstacles. THz technologies progress extremely slowly compared to other fields without the added benefit from large-scale commercial applications for ultra-low noise, passive techniques used in astronomy. Here we will discuss a few of the less obvious tasks that remain for future THz instrument builders.

An important innovation of the KAPPa project, not discussed in the present work but developed by other KAPPa team members, is the flexible IF transmission line that snaps onto the back of the array receiver. This flexible circuit board transmits the IF signals on microstrip lines from the 4 K array receiver to external 300 K spectrometers. It is a marvelous piece of engineering and it also solves a serious problem of manufacturing thousands of individual coax cables. We believe that this is the correct approach for both simplicity and scalability. However, we are concerned about feeding this flexible circuit board through the vacuum wall of the arrays cryostat. The potential for catastrophic failure exists, but a larger concern is that small vacuum leaks will develop over time from thermal differential expansion between the vacuum walls and the flexible circuit. It is our hope that the next builder of a kilopixel array will solve this problem in a forward-looking and sustainable way for long-term use in facility instruments.

Automation is the key to progress in the future of astronomy. It is our sincerest hope that the automation, optimization, and data analysis tools developed
for KAPPa are not shelved because of their own complexity to make short-term gains. Short term progress, and personal prestige, are the enemies of revolutionary innovation. If the reader is interested, please find our team, we can build something together.

Data reduction for large THz receiver arrays, will also benefit from optimization techniques of differential evolution, although this may be excessive for some applications. It is our vision that code, for use on a single laptop computer, will automatically calibrate a kilopixel array receiver both in the lab and on the telescope, informing data acquisition on the sky. The approach of taking lots of data and working hard to reduce it all later is unnecessary. This is why all code for the KAPPa project and SuperCam data reduction project have been written in a single language, Python 2.7, an astronomy community standard. THz instrument builders have worked in a vacuum for too long, no pun intended, and it is time to join the rest of the astronomy community to develop resources together. A fully integrated and networked system would make it immediately apparent when calibration scans were needed more or less often. Moreover, a feedback loop with single code, networked computers could automatically make adjustments more rapidly the human ever could and, more importantly, improve the overall quality of data. This will be increasingly important for future space and balloon missions using THz arrays where human interaction will be limited, if available at all.

The largest blind spot in building kilopixel array receivers is in the field of digital signal processing, namely building spectrometers to read the arrays IF output. Fortunately, this is one of the few areas of THz instrument building where there is considerable crossover with commercial application. If the price of commercial spectrometers continues to drop, we may be able enough to purchase these as a part of an overall kilopixel array. However, if such spectrometers are out
of the price range, cheaper student projects should start developing low-budget spectrometers in advance of a kilopixel THz array. This may even be a lucrative side business for an ambitious professor, as spectrometers have many applications in astronomy and beyond.

Since Jordan Wheeler (younger brother to Caleb Wheeler) is a direct detection THz instrument building astronomer, we are excited to imagine how direct detection and heterodyne THz instruments can work together to accomplish their goals. Wheeler, the lesser, works on a direct detection instrument called SuperSpec that sacrifices one spatial axis for moderate spectral resolution, with 100 GHz of instantaneous bandwidth. SuperSpec is optimized to look for line emission from a 200 – 300 GHz band, with a spectral resolution of $R \approx 500$. For galaxies at $z > 1$ there will be two available CO rotational lines in SuperSpecs band.

Nearby star forming galaxies ($z \approx 0$) are shown to have both warm and cool gas components (Rangwala et al., 2015). Similarly, Kamenetzky et al. (2014, and references therein) used archival data from Herschels SPIRE that contained observations of the $J = 4 \rightarrow 3$ through $J = 13 \rightarrow 12$ $^{12}$CO rotational transitions for several unresolved sources. Higher temperature rotational transitions are particularly important since they enable a probe for the hotter gas component of galaxies. The populations of these high-J rotation lines of CO require a more massive warm-gas component than the ground based observations of $J = 0 \rightarrow 1$ lines indicate.

ALMA’s large spectral resolution also provides more dynamical information for the galactic target, whether the gas is rotating in a disk, contained in an outflow, or coupled to an active galactic nucleus. An interesting problem involves modeling the hot and cool gas components of nearby star forming galaxies using follow-up observations from ALMA and extending the developed formalism of Kamenetzky.
et al. (2014), that uses a photon escape probability, to model the molecular gas in galaxies. The mechanism by which the warm component of gas is heated is not well constrained in high redshift sources. It is unknown if the source of heating is stellar winds, supernovae remnants, protostellar outflows, or a combination. However, understanding the warm component will be needed to add context to the direct detection observations, such as those made using SuperSpec, of unresolved sources of high-J transitions during the Universe's peak star formation epoch.

By using an efficient mapping heterodyne mapping instrument like KAPPa, we would be able to characterize a larger number of high redshift galaxies at a spectral resolution that would resolve modular lines, but each galaxy would be unresolved spatially. Such an instrument would be very useful when tuned to redshifts where the galaxies are undergoing rapid transitions. We expect that as investigations with instrument’s like SuperSpec, start making their first maps, interesting redshifts will be identified. Shortly after these maps are made, we hope to be standing by with a heterodyne instrument tuned to these interesting redshifts. Aside from collecting precise redshift information, we hope to automatically characterize outflows and gas dynamic of these galaxies, thus allowing us to statistically characterize a large swath of sky during a transitional redshift. The large-scale statistics determination would be an exciting look in galaxy evolution.
REFERENCES


Aubert, D., & Teyssier, R. 2010, The Astrophysical Journal, 724, 244


Carroll, L. 1871, Through the Looking-Glass, and What Alice Found There (Macmillan Publishers, United Kingdom)


169


Josephson, B. 1962, Physics Letters, 7, 251


170

Nyquist, H. 1928, Transactions of the American Institute of Electrical Engineers, Volume 47, Issue 2, pp. 617-624, 47, 617


Pozar, D. 2012, Microwave Engineering (Wiley, New Jersey)

—. 2010, Physical Review D, 82, 23006, (c) 2010: The American Physical Society


Shu, F. H. 1995, in Molecular Clouds and Star Formation, ed. Y. Chi & J. You, 97


172
Townes, C. H., & Schawlow, A. L. 1975, Microwave spectroscopy. (Dover Publications)
Van Valen, L. 1973, Evolutionary Theory, 1, 1