

A Very Wide Bandwidth SIS Heterodyne
Receiver Design for Millimeter and
Submillimeter Astronomy

Thesis by
Frank Randolph Rice, III

In Partial Fulfillment of the Requirements for the degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2020

(Defended August 26, 2020)

© 2020

Frank R. Rice

All rights reserved.

ORCID: 0000-0002-9845-9762

*To my beautiful wife Fanny –
till death us do part.*

ACKNOWLEDGEMENTS

Several members of the research staff and graduate students in the submillimeter group at Caltech provided valuable guidance and assistance as I designed and built my receiver, and I gratefully thank them for their help. In particular, my two office mates, John Ward and Goutam Chattopadhyay, were invaluable collaborators for many of my efforts. John was not only instrumental to the development of the SuperMix library; he also took me to the CSO and taught me about observing with its heterodyne receivers. A tireless worker with an incredible work ethic, his contributions to this receiver development will be described throughout this text. Goutam is an expert in microwave and millimeter wave design, and he provided timely, expert assistance as I learned to use commercial modeling and development software. His work on the design of the JPL submillimeter frequency multipliers for the Hershel mission inspired my very successful waveguide probe design. Last but certainly not least, his SIS-based, dual polarization heterodyne receiver provided the first serious test of SuperMix as a modeling and design tool.

I thank and congratulate the JPL team led by Henry G. (Rick) LeDuc, who put in long hours perfecting the new techniques needed to make the superconducting mixer chips required by my design. Their tireless efforts to produce not only superior SIS junctions to meet my demanding specifications, but to also incorporate beam leads onto the very thin silicon substrate I required were really inspiring. In particular, I must mention Bruce Bumble, SIS fabricator par excellence, whose mixer chip made my receiver perform better than I ever would have expected.

I am grateful that I got to know the great Gerry (pronounced “Gary”) Neugebauer, whose office in Downs was directly across the hall from mine. Our many discussions of

not only physics and astronomy but also life in general were most important to me. As a leading member of the search committee who decided to bring me on at Caltech as a permanent hire, I also have him to thank for my long career here.

Professor Tom Phillips, brilliant physicist and engineer and the long-time director of the Caltech Submillimeter Observatory, provided unwavering support, encouragement, and outstanding advice during the development and maintenance of my instrument. Largely due to his patient, careful listening and timely hints, my receiver was able to evolve into a sensitive, useful instrument for astronomy. He once brought Arno Penzias down to the lab to introduce me and show him my receiver, which was an unforgettable experience for me.

Jonas Zmuidzinas. What can I say to express how important he has been to the success of this project and to my success at Caltech? Jonas is probably the smartest guy I know (and I have met some pretty smart ones around here). He is also pleasant, friendly, generous, and a good listener, all of which make for a very rare combination. He surprisingly took on this unknown, old guy who had been away from university for nearly twenty years and knew next to nothing about Caltech or the current state of physics research.



As a first-year graduate student, I absorbed his lectures given during an upper-division electricity and magnetism course (using, naturally, Jackson's text). I realized

then that he was exactly the guy from whom I wanted to learn science. The summer following my first year Jonas let me join his research group on a trial basis, and we started a collaboration that resulted in the highly-successful SuperMix software suite. I really enjoyed his style of developing me into a bonafide researcher, and I have adopted his regimen of weekly, one-on-one advisor-apprentice meetings as a model for the experimental physics courses I teach. Jonas kept me funded and paid for the many years I took to build and improve my receiver and also while I taught and worked on other projects for the CSO and submillimeter astrophysics groups. As a not-so-famous saying goes, his efforts “might not have kept the kids eating cookies, but kept the wolves from the door.”¹ He was also instrumental in Caltech’s decision to take me on permanently and provide a Lectureship appointment (which I still hold).

Finally, I gratefully acknowledge the support for this work generously provided in part by NASA/JPL and its Center for Space Microelectronics Technology, by NASA grants NAG5-9493, NAG5-4890, NAGW-107, and NAG2-1068, by the NASA/USRA SOFIA instrument development program, and by the Caltech Submillimeter Observatory (NSF grant AST-9980846). Invaluable support has also been provided by Caltech through funds received from royalties earned from sales of the *Feynman Lectures on Physics*.

Notes about the text:

The text was created and formatted using Microsoft® *Word 2010* using primarily Times New Roman and Calibri fonts. Program code was set in Lucida Console. Graphics were combined, formatted, and edited using Microsoft® *Office 2010* tools including *PowerPoint* and *Picture Manager*. Equations were created and typeset using *MathType 6.8*, © 2012 Design Science, Inc. Plots were generated using *Wolfram Mathematica*® versions 8 to 11. Solid models and mechanical drawings were generated using various versions of *Dassault Systèmes Solidworks*®. The final PDF was generated using Adobe *Acrobat X Pro*®. SuperMix code was compiled and executed using Cygwin64 and Cygwin/X running under 64-bit Microsoft® *Windows 7 Professional*.

¹ Told to me by Professor Dan Kleppner to describe the teaching assistantship he very generously offered me on behalf of MIT in 1996. Oh, how things would have turned out differently if I had accepted!

ABSTRACT

This text describes in some detail the design and operational history of an instrument used as the front-end receiver for a fast, broadband, high-resolution spectrometer for the 1.3 millimeter wavelength atmospheric window. Using only a single superconductor-insulator-superconductor (SIS) tunnel junction as its heterodyne detector, the receiver's novel design achieved then unprecedented RF and IF bandwidths and incorporated several innovations which have since been widely adopted within the millimeter and submillimeter wave research communities. Although intended as a relatively simple technology demonstrator and starting point for more refined and sophisticated designs, the receiver turned out to be a useful astronomical instrument in its own right, and it was deployed as a de facto facility instrument for several years at the Caltech Submillimeter Observatory. Also described are the author's contributions to another important aid to research and design efforts: the *SuperMix* software library for the analysis and optimization of high-frequency circuitry, especially developed to aid in the design of systems involving SIS and other superconducting elements. Finally, the text may serve as a useful introduction to the theory behind and methodology for modeling and design of SIS heterodyne mixers.

TABLE OF CONTENTS²

Acknowledgements	iii
Abstract	vi
List of Figures	x
Glossary	xiii
<i>Chapter 1 Introduction</i>	<i>1</i>
A history of the wide bandwidth receiver project	1
General description of the instrument	8
<i>Chapter 2 Heterodyne SIS mixer chip design</i>	<i>13</i>
Mixer chip operation and functional requirements	13
Technology innovations introduced by the design	16
Mixer chip general layout	18
SIS junction behavior, characteristics, and specifications	20
SIS RF embedding impedance requirements	31
SIS IF embedding impedance requirements	35
Mixer RF matching network circuit design	37
RF choke and IF matching circuit design	42
Mixer chip performance model results	49
<i>Chapter 3 Waveguide probe and mixer block</i>	<i>53</i>
Waveguide and waveguide probe	53
Mixer chip mounting channel	57
Feed horn waveguide interface	59
Other aspects of the mixer block design and mounting	61
Cryostat wiring considerations	63

² With any reasonably-competent PDF document reader, the entries in this table should act as hyperlinks to the relevant pages of this document. The same should be true of the citations and references to figures and equations throughout the body of the text. This functionality does not extend, unfortunately, to printed versions (at least, not at the time of this writing).

<i>Chapter 4 RF optics and receiver performance</i>	66
RF optics design	66
Receiver RF beam measurement	68
Optics noise temperature impact	73
<i>Chapter 5 SIS circuit modeling and SuperMix</i>	81
An example illustrating SuperMix capabilities	82
Circuit modeling and design optimization	87
Basic numerical data structures, linear algebra, and units	91
Representing a heterodyne receiver	95
<i>Chapter 6 Conclusion</i>	104
Astronomy and planetary science	104
SuperMix	105
Waveguide probe	108
Mixer design	109
Final comment	109
Appendices	110
<i>Appendix A : Mixer chip specifications and geometry</i>	111
<i>Appendix B : Mixer block and lna specifications</i>	116
Mixer block drawings	116
LNA performance specifications	125
<i>Appendix C : The scattering representation and noise</i>	126
The wave representation	126
The scattering matrix	128
Reflection by an impedance; Y and Z matrices	129
The Smith Chart	131
Modeling internal sources and noise	134
The excess quantum noise introduced by high-gain amplifiers	139
Y-factor and SSB vs. DSB noise in heterodyne receivers	141
<i>Appendix D : Receiver model predicted results</i>	146

Embedding impedance (reflection, Γ_{RF} or Γ_{IF}) study results	146
Mixer performance variation with SIS operating state	151
<i>Appendix E : Fast SIS harmonic balance algorithm</i>	<i>155</i>
Nonlinear calculation of the SIS harmonic currents	157
The physics behind the tunneling current calculations	159
Harmonic balance of the junctions and the linear network	160
Calculating the SIS Jacobian matrix	162
<i>Appendix F : Propagation in partially-filled waveguide</i>	<i>165</i>
Wave propagation in a lossless, cylindrical waveguide	165
Transverse and longitudinal fields	167
Propagation in partially-loaded, lossless, rectangular waveguide	169
Using a vacuum gap to increase the lowest cutoff frequency	174
Effect of the vacuum gap on the TE_{01} cutoff frequency	176
<i>Appendix G : Gaussian optics review</i>	<i>178</i>
The paraxial, fundamental Gaussian beam mode approximation	178
Edge taper, beam coupling, and aperture efficiency	180
The Gaussian beam parameter and beam transformations	183
<i>Appendix H : Cryogenic bias wiring optimization</i>	<i>186</i>
Temperature and heat transfer of an undersized (too thin) wire	189
Application to high-resistivity, cryogenic bias wiring materials	190
Optimizing a wire carrying an intermittent current	191
<i>Appendix I : SuperMix programs</i>	<i>193</i>
Building a circuit	193
Modeling superconducting microstrip lines	195
Using the optimizer	197
Calculating the SIS operating state C_k values	199
Embedding impedance modeling	205
Bibliography	211

LIST OF FIGURES

A March 2004 wide-bandwidth, low-resolution spectrum of the galaxy M82.	4
A 2007 high-resolution line survey of the Orion KL star forming region.	5
Various configurations of the receiver and its electronics at the CSO.	6
Panoramic, interior view of the CSO and telescope with the author's receiver (2010).	7
High-level block diagram of the receiver system, showing signal flow.	8
Photo of the 2004 mixer assembly mounted to the cryostat cold plate.	10
Drawings of the original (left) and improved (right) mixer assemblies.	10
Diagram and photo of the mixer chip showing its general features.	19
Diagrams illustrating mixer chip mounting relative to the RF waveguide.	20
DC I-V characteristic of a typical SIS device used in the prototype receiver.	22
Example of a pumped SIS DC I-V curve.	24
Complex SIS LO quantum admittance vs. DC bias.	26
Effect of RF embedding reflection (Γ_{RF}) on SIS mixer noise and IF stability.	33
Effect of RF embedding reflection on the SIS pumped I-V curve shape.	34
Effect of IF embedding reflection on mixer conversion gain and RF stability.	36
RF matching network design concept and circuit implementation.	38
Photo of the final RF matching network as manufactured by JPL.	39
Modeled SIS RF embedding impedance performance of the mixer chip.	41
Expected RF embedding impedance variation with SIS capacitance.	42
The RF choke and IF matching network structure.	43
Predicted coupling of RF signals to the IF output of the choke circuit.	44
Variations in the SIS RF embedding gamma caused by IF reflections.	45
Mixer chip bias tee and output wire bond connections.	46
Mixer chip bias tee photo and equivalent circuit model at IF frequencies.	47
Modeled SIS IF embedding impedance variation with IF frequency.	48
Modeled mixer chip + LNA DSB noise performance vs. SIS operating state.	50
Modeled mixer chip + LNA sideband performance vs. RF and IF frequency.	52
RF waveguide and waveguide probe design with dimensions shown in mils.	55
Smith chart showing the predicted waveguide probe performance.	56
Mixer chip channel cutoff frequency vs. channel height and final geometry.	58

Feed horn cylindrical to rectangular waveguide transition geometries.	59
Physical configuration of the mixer block, feed horn, and LNA assembly.	62
Receiver RF optical design with modeled Gaussian beam plots.	68
Receiver beam raw mapping data for the final cryostat optics configuration.	69
Final convolution model fit to the raw receiver beam map data of Figure 4-2.	71
Another view of the x vs. y data deviation asymmetry from the RF beam fit.	72
The deconvolved receiver RF beam as inferred from the beam map data.	72
Receiver RF optics transmission efficiency estimates vs. RF frequency.	75
RF optics and LO noise impacts on the modeled receiver performance.	78
Receiver DSB noise and Y-factors model vs. measured results.	80
G. Chattopadhyay's quasioptical, balanced mixer design using eight SIS junctions.	83
Fit of the mixer model to measured pumped I-V (left) and FTS (right) data.	85
Predicted and measured IF output power vs. RF source temperature and SIS bias.	86
The top-level circuit topology of a multi-SIS, heterodyne detector.	95
SuperMix representation of a nonlinear SIS as a heterodyne detector.	97
Representing a complete heterodyne system using class <code>mixer</code> .	99
Appendices:	
Waveguide probe and RF circuit detailed geometry.	112
RF choke interface detailed geometry (upper); choke arrangement (lower).	113
RF choke circuit and ground plane layer detailed geometries.	114
Ground plane and circuit connection bond pad geometries.	115
Assembled mixer block, feed horn, and horn adapter with dimensions.	116
Drawing views of the assembled mixer block.	117
Mixer block parts descriptions.	118
Mixer block top half drawing.	119
Chip channel and waveguide detailed dimensions.	120
Mixer block bottom half drawing.	121
Mixer block horn adapter.	122
Waveguide rectangular-cylindrical transition detailed dimensions (in mils).	123
Mixer block assembly mounting and telescope and LO optical path interface.	124
Measured performance of the 4-12 GHz LNA delivered in 2004.	125
The wave representation on a transmission line.	127

A 5-port structure with an incoming wave at port 1.	128
Reflection at a termination (1-port).	130
Impedance matching example.	132
Smith chart showing a plot of the variation of $\Gamma(\omega)$ for the circuit in Fig. C-4.	133
Effect of number of harmonics on the mixer analyses; LO 240 GHz, IF 8 GHz.	147
Mixer T_n contours and min T_n vs. IF frequency; LO 240 GHz, Upper Sideband.	148
Mixer T_n contours and min T_n vs. LO frequency; IF 8 GHz, Upper Sideband.	149
Mixer T_n contours and min T_n vs. LO frequency; IF 8 GHz, Lower Sideband.	150
Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 5.6.	152
Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 6.2.	153
Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 7.5.	154
SIS junctions attached to a linear network.	155
Example normalized SIS DC I-V curve and its Kramers-Kronig transform.	158
A cylindrical waveguide of arbitrary, but uniform, cross-sectional shape.	165
A rectangular waveguide filled (loaded) with two dielectrics.	169
Increasing silicon chip TE_{10} cutoff frequency by adding a vacuum gap.	175
Silicon dielectric TE_{01} cutoff frequency vs. dielectric filling fraction.	177
Aperture efficiency vs. edge taper.	182
Diagram illustrating the parameters used for the heat flow analysis of a wire.	186
A drawing of the RF circuitry wiring layer, identifying its components.	193
The SuperMix model of the SIS mixer chip RF circuitry wiring layer.	194

GLOSSARY

Some of the descriptions provided here may be rather “breezy” or incomplete, but should be sufficient to jog the reader’s memory or to get across the general idea (many others, alas, may prove too wordy and pedantic). More satisfying explanations may be provided in the body and the appendices, or may be relegated to the references.

Alpha, LO pumping (α): a measure of the amplitude of the local oscillator (LO) signal applied to an SIS mixer. A parameter used in John Tucker’s quantum heterodyne theory [1], it is given by $\alpha \equiv V_{\text{LO}}/V_{ph} = eV_{\text{LO}}/\hbar\omega_{\text{LO}}$, where V_{LO} is the LO signal voltage amplitude and ω_{LO} is the LO signal angular frequency. See *SIS DC I-V characteristic and photon-assisted tunneling* in Chapter 2.

Bolometer receiver: a *direct detection* receiver which uses a sensitive “thermometer” as its power detector. The incoming radiation is focused on a small, cold sensor (for instruments useful for astronomy, *very* cold). Absorbed energy causes a tiny temperature rise in the sensor, which in turn causes some electrical characteristic of the sensor (its resistance or, in several more recent designs, its resonant frequency) to change. Because a well-designed sensor can efficiently absorb energy from any polarization over a broad frequency range and is not subject to *quantum noise*, a bolometer-based instrument can be extremely sensitive to broadband radiation (such as the submillimeter thermal radiation emitted by “warm” (~ 30 K) dust surrounding many interesting astronomical sources).

CSO: the Caltech Submillimeter Observatory, located near the summit of Mauna Kea, Hawaii. Its domed, 10.4 meter diameter telescope fed a suite of both heterodyne spectrometers and bolometer cameras covering the submillimeter atmospheric windows with frequencies of 180–950 GHz (1.7 mm–320 microns). The observatory achieved “first light” in 1987 and ceased operation in 2015 [2]. During its too-short lifetime, the CSO was the site of many outstanding observations and the home to several major engineering developments in submillimeter astronomy instrumentation.

Direct detection: a direct detection receiver responds to incoming radiation by providing an output signal proportional to the average signal input power, where the average is over a time scale very much longer than the typical period of the incoming radiation's oscillation. As a result, any information regarding the instantaneous phase of the incoming signal is lost (contrast this with the action of a *heterodyne receiver*). Because a properly-designed direct detection receiver is insensitive to the phase of the incoming signal, it is not subject to the effects of *quantum noise*, unless the receiver system first coherently amplifies the signal prior to its input to the power detector.

Embedding impedance: the electrical impedance presented to an SIS junction by its surrounding circuitry. Specifically, the *RF embedding impedance* refers to the impedance presented to the SIS at RF frequencies (100's of GHz); the IF embedding impedance is that seen by the SIS at IF frequencies (~10 GHz). These impedances determine how efficiently signals are coupled into and out of the SIS, as well as determining the stability of the circuitry against unwanted oscillations. See Chapter 2.

Gunn diode: an inhomogeneously-doped (usually N-type GaAs) semiconductor. When a DC bias voltage is applied to the semiconductor, a very large electric field in its lightly-doped region may be generated. Conduction electrons accelerated by this field can undergo inter-band transitions out of the normal conduction band into a band where their effective mass is increased, reducing their conductivity. This Gunn effect (J. B. Gunn, 1962) gives the diode an unstable, negative resistance region of its I-V curve, resulting in very high frequency, somewhat chaotic oscillations of the diode current. Placing the device in a tunable resonant cavity can stabilize the oscillation frequency of the diode, resulting in an effective, powerful millimeter-wave source.

Harmonic balance: see *operating state*.

Heterodyne detection: process whereby a weak, high-frequency signal (the *RF source*) is combined with a much stronger, tunable, coherent signal (the *LO source*) and is then input to a nonlinear device (the detector, or *mixer*). The nonlinearity of the mixer generates additional signals at frequencies equal to sums and differences of all integer multiples of the RF and LO signal frequencies. For most applications, the output of interest is the simple difference of the LO and RF frequencies (called the *IF frequency*), and it is selected by low-pass filtering of the mixer output. If the LO signal is kept fixed in amplitude and frequency, then the IF output amplitude of a well-designed receiver is proportional to the RF signal amplitude, and its phase equals the difference in the phases of the LO and RF signals, making this a *phase coherent* process.

Heterodyne receiver: an astronomical receiver which uses *heterodyne detection*. The *phase coherence* between the RF signal input and the IF output of a heterodyne detector has three important consequences: first, the receiver can be designed to have very high frequency resolution, making it ideal for spectrometry; second, the receiver is unfortunately limited in sensitivity by *quantum noise*; third, a single IF output from the receiver can respond to only one component of polarization in the incoming radiation from a source, so that half the power from an unpolarized astronomical source will be undetectable. The polarization component (e.g., linear, circular, or elliptical) to which the IF output is sensitive is determined by the design of the telescope antenna and its feed system which couples the RF energy to the heterodyne detector.

IF: Intermediate Frequency. The frequency range into which the detector of a heterodyne receiver coherently converts incoming signal radiation. The detector described in this paper has a useful IF output frequency range of about 4–18 GHz. Subsequent amplification and signal processing by the IF instrument system (called the receiver's *back end* processing) may use only a subset of this range.

Instantaneous bandwidth: the frequency range over which a *heterodyne receiver* is sensitive while keeping its local oscillator (LO) frequency fixed. It is equal to the intermediate frequency (IF) bandwidth for a single-sideband receiver or twice the IF bandwidth for a double-sideband or sideband-separating instrument.

Large-signal response: see *operating state*.

LNA: low-noise amplifier. Until recently, the extremely high frequencies of the submillimeter wavelength bands have been generally beyond the capabilities of conventional electronic amplifier components. Consequently, the RF signal is first mixed with the local oscillator (LO) and converted to the much lower IF frequency by a heterodyne detector. The resulting IF signal is then amplified by the LNA, a cryogenically cooled microwave amplifier designed to be very linear and have low noise and high signal gain. The bandwidth and sensitivity of this amplifier (as represented by its noise temperature) will be one of the limiting factors to the overall performance of the heterodyne receiver system.

LO: Local Oscillator. A relatively high-power, stable, tunable, coherent signal source whose output is combined with the incoming RF signal radiation and input to a heterodyne detector to produce a much lower frequency IF output signal.

Loss tangent, dielectric: A measure of the power loss of a dielectric material. A material's dielectric constant will generally be complex-valued: $\varepsilon = \varepsilon_r - j\varepsilon_i$ (assuming a time dependence of $\exp(j\omega t)$, the electrical engineering phase convention). The imaginary part will characterize the lossy dielectric's conductive component. The *loss tangent* (usually written as $\tan \delta$) is the ratio of the dielectric constant's imaginary and real parts, so that $\varepsilon = \varepsilon_r(1 - j \tan \delta)$. As shown in Chapter 7 of Jackson [3], for materials whose loss is relatively small (nonconductors, for example), the fractional power loss per radian of phase along a signal's path would then be very nearly equal to $\tan \delta$, justifying its "loss tangent" moniker.

Noise temperature: a measure of the unwanted noise added to a signal by a transmission medium, amplifier, or other signal processing step. In this paper, "noise" is considered to result from random processes such as thermal excitations in a lossy element (*Johnson noise*), fluctuations in the flow of current caused by the discrete nature of electric charge (*shot noise*), or the fundamental fluctuations introduced by the quantum nature of the excitations of the (bosonic) normal mode frequencies of the inputs to the system (*quantum noise*). Noise temperature is a measure of the noise power spectral density expressed in Kelvin and referred back to the input:

$$T_n(\nu) \equiv \frac{1}{k_B} \left. \frac{dP_n}{d\nu} \right|_{\nu(\text{input-referred})} \quad (\text{see Appendix C}).$$

Operating state: for a nonlinear circuit element such as an SIS device used as a heterodyne detector, its *small-signal response* is modeled as its first-order variation away from its operating state: its quiescent, steady-state (no input signal) voltage and current waveforms. In the case of a nonlinear device used as a heterodyne detector, this quiescent state is determined by an applied DC bias voltage across the device along with the presence of a relatively high-power local oscillator signal at a fixed frequency. The device responds to these stimuli, producing what is called its *large-signal response*. To determine this response requires an iterative, numerical solution to a nonlinear system of equations, a procedure known as *harmonic balance*.

Photon-assisted quasiparticle tunneling: see *quasiparticle tunneling*.

Quantum noise: quantum noise provides a limit to the sensitivity of any phase-coherent amplifier such as a heterodyne receiver. It may be thought of as arising due to “spontaneous emission” at the output of a high-gain, coherent amplifier, the desired response being the “stimulated emission” of the amplifier output in response to the signal input. In the case of a *heterodyne detector*, it is the input RF frequency (not the much lower IF frequency) which determines the system’s *quantum noise limit*. The quantum noise limit corresponds to an equivalent *noise temperature* of 10 K at 208 GHz and is proportional to frequency. See *The excess quantum noise introduced by high-gain amplifiers* in Appendix C.

Quasioptical system: a system through which an electromagnetic signal propagates in a fairly well-defined direction and whose structure consists of components with transverse dimensions of only a few wavelengths. Many substructures of the RF signal feed systems for most microwave and millimeter wave receivers and transmitters may be categorized as quasioptical. See Chapter 3 and Goldsmith [4].

Quasiparticle: a quite powerful and subtle idea introduced in 1957–1959 by the great theorist Lev D. Landau. At very low temperatures electron-electron interactions in a metal can render invalid the simple but powerful approximation that an N -electron state may be modeled as a collection of N occupied single-electron states. Landau demonstrated, however, that for a large class of strongly interacting N -fermion systems, including the conduction electrons in most metals (but not including magnetically-ordered systems nor the superconducting ground state), an N -fermion state can nevertheless be adequately described as a collection of N occupied single-*something* states, the “somethings” behaving like fermions and dubbed “quasiparticles.” See, for example, Chapter 17 of [5] for a more thorough (and possibly satisfying) explanation.

Quasiparticle tunneling: a process by which the two electrons of a Cooper pair may separate, with one of the electrons tunneling through the insulating barrier of an SIS device. The result is a pair of now-occupied single-quasiparticle states, one in each of the two superconductors of the SIS, accompanied by the transfer of a single electron charge across the barrier. The tunneling may be either spontaneous or *photon-assisted*, depending on the energetics of the available single-quasiparticle states and the photon flux.

RF: Radio Frequency. In the context of this paper, the signal frequency range to which a telescope-receiver combination is sensitive, which is designed to be 200–300 GHz for the system described in this paper.

Scattering matrix: a numerical model of the behavior of a structure which responds linearly to sinusoidal, traveling-wave excitations entering into one or more *ports* of the structure. The structure’s scattering matrix \mathbf{S} describes how incident wave excitations generate outgoing wave responses at each of the structure’s ports. \mathbf{S} will generally have elements which are complex-valued and frequency-dependent. Scattering matrices are the representations of choice used by most high-frequency models of circuit behavior. See Appendix C.

Sideband: an RF frequency range converted by a *heterodyne detector* to its IF output. Since the IF frequency is given by the difference between the RF and LO frequencies, an RF signal with a frequency either greater than or less than the LO frequency could be converted to the same IF frequency. RF signals with frequencies given by the LO frequency plus the IF frequency range comprise the *upper sideband*; the corresponding range of RF frequencies below the LO frequency make up the *lower sideband*. A *double-sideband* detector, as is the subject of this paper, is the most simple, and it converts both upper and lower sideband RF signals. Unfortunately, it superimposes both sidebands in its single IF output, both doubling the background noise and making the correct interpretation of its output more difficult. The significantly more complicated *single-sideband* detector allows only one RF sideband to produce an IF output, and it is a key component in what are called *superheterodyne* receivers, mainstays of the commercial radio industry. Even more sophisticated heterodyne detectors provide two separate IF outputs, one for each sideband, and are therefore called *sideband-separating* mixers.

SIS: see *superconductor-insulator-superconductor*.

Small-signal response: see *operating state*.

S-matrix: see *scattering matrix*.

Smith Chart: a graph of the variation of the complex-valued *amplitude reflection coefficient* Γ (gamma) of an electromagnetic component or termination, usually parameterized by signal frequency. See *The Smith Chart* in Appendix C.

Superconductor-insulator-superconductor: a device constructed from two superconducting metals separated by a very thin (~ 1 nm) insulating layer. Under appropriate conditions, charge carriers originating in one of the superconductors may quantum-mechanically tunnel through the barrier and into the other, carrying an electrical current across the barrier. The charge carriers may be either Cooper pairs of electrons (Josephson tunneling) or single electrons (see *quasiparticle tunneling*). Josephson tunneling of Cooper pairs may be suppressed by applying an appropriate magnetic flux through the insulating layer and parallel to the superconducting layers' surfaces.

Tunnel junction: see *superconductor-insulator-superconductor*.

(page intentionally left blank)

Chapter 1

INTRODUCTION

This thesis focuses on the author's role in the design and development of what was meant to be a prototype of a new, high-performance instrument for millimeter and submillimeter astronomy. This instrument, variously known as "Z-rex," "Zrx," or, more familiarly, "Frank's receiver," was designed for fast, wide-bandwidth spectroscopy covering the 1.3 mm atmospheric window (180–300 GHz). It debuted on the Caltech Submillimeter Observatory's 10.4 meter telescope in mid-summer, 2003. Mainly because of the hard work and the timely, wise advice of several scientists, engineers, students, and technicians (many of whom saw its true potential), the instrument experienced a long, productive run at the observatory until finally being retired in 2012. Possibly of greater importance to submillimeter astronomy were the tools and methods used to design the instrument. These tools and methods saw wide application in the subsequent development of new, sophisticated instruments throughout those years at several sites around the world.

A history of the wide bandwidth receiver project

The author joined Caltech's Submillimeter Astrophysics Group in the summer of 1997 in order to begin the development of a new software modeling package for the design of arbitrarily complex, superconducting, high-frequency, heterodyne receivers using one or several *superconductor-insulator-superconductor* (SIS) tunnel junction devices as nonlinear mixing elements. Jonas Zmuidzinis, John Ward, and the author worked together over the next several months to write the first version of what has become the *SuperMix* C++ class library for the modeling and optimization of such circuitry. The author has continued to refine and expand this library during the intervening years; as of this writing the SuperMix package has nearly 48,000 lines of code and has been used successfully by various research groups around the world to design new submillimeter and microwave systems (SuperMix is discussed briefly in Chapter 5).

While SuperMix was being developed, analyses of SCUBA (Submillimeter Common-User Bolometer Array) observations were leading to a revolution in astronomers' understanding of the star formation history of the universe. Early success with SuperMix in 1999 increased the author's confidence that a heterodyne receiver could be designed having unprecedented bandwidth while maintaining sensitivity and stability. Such a receiver could potentially be used to search the 180–300 GHz atmospheric window for carbon monoxide (CO) line emissions from distant submillimeter sources; because the heterodyne receiver system would be designed to be a spectrometer, finding a CO line and measuring its observed frequency would result in an accurate redshift determination of its source. If sufficiently sensitive, this receiver could serve as a useful follow-up instrument because SCUBA was a two-color, imaging camera which could only provide very rough redshift estimates.

Several technical challenges would need to be overcome to make the new instrument feasible. The very wide bandwidth requirement would call for significantly new, thoroughly optimized designs of the SIS devices and the associated superconducting circuitry. An improved design for a waveguide probe (the element which couples the RF energy into the circuit) would be required, as would a new broadband, low-noise, microwave amplifier (LNA) to handle the output from the heterodyne detector. All of these elements required a properly-designed enclosure and cooling to liquid helium temperatures (4 Kelvin), as well as physical, optical, and electrical integration with the telescope and observatory. Finally, a cost-effective spectrometer system would be needed to process the large bandwidth available at the receiver's output. The LNA and the spectrometer system would be developed by others, but the author was responsible for the rest.

By early 2000 the author had finished studies which demonstrated the feasibility of a wideband receiver design using a single SIS device. Funding was secured in April 2000, in May the author passed his candidacy exam, and the project was officially underway. By July, the target electrical characteristics for the SIS device had been chosen, a candidate waveguide probe concept had been identified, and the basic topology of the superconducting micro-circuitry had been laid out. Concurrently with the author's efforts, design of new cryogenic microwave amplifiers was underway in Sander Weinreb's group

at Caltech [6], and the Wideband Analog Autocorrelation Spectrometer System (WASP II), under development by Andrew Harris of the University of Maryland, was identified as an appropriate spectrometer back-end for the instrument's initial testing [7]. An internal Preliminary Design Review of the receiver was completed in August 2000. The object was to design a prototype instrument for development and testing on the 10.4 meter telescope at the Caltech Submillimeter Observatory (CSO) on Mauna Kea.

After much effort, the waveguide probe design was finalized in early 2002. The mixer chip was to be manufactured on a 25 micron thick silicon wafer, an unusual choice for a waveguide-mounted SIS receiver, and the detailed mixer chip circuitry layout was underway. The initial chip design was completed in May 2002, and the first batch of mixer chips was delivered by the Jet Propulsion Laboratory (JPL) one year later. The initial mixer waveguide block manufacturing drawings were completed in July 2002, and a first version of the block was machined and delivered by Custom Microwave, Inc., several weeks later.

A custom active multiplier assembly for use as the heterodyne receiver's local oscillator (LO) was procured from Virginia Diodes, Inc., and was delivered in April 2002. This assembly would be the heart of a *synthesized LO chain*, a new way of generating the high-frequency LO signal. At that time it provided a modern, more flexible alternative to the proven, traditional *Gunn diode* LO source, and we wished to investigate the engineering challenges such a design would present. The lessons learned from the group's use of this LO system will be touched on only briefly in this document; they are addressed in much greater detail in Sumner [8].

Late in the spring of 2003 all the pieces of the receiver were finally available, although the SIS devices in the mixer chips then available did not meet all design goals (the JPL team also needed to develop new skills and techniques to successfully fabricate the author's challenging design, which they accomplished in stages). After several failed attempts and broken mixer chips, the author finally succeeded in mounting and wire bonding an operational SIS mixer chip to its mixer block by the end of July 2003. The assembled receiver was cooled and checked for the first time on 15 August 2003. The *noise temperature* of the receiver during that first test averaged a disappointing 400 K

(double-sideband: *DSB*), a noise level an order of magnitude higher than the design goal, but the observatory beckoned.

The complete receiver system was shipped to and installed on the telescope atop Mauna Kea less than two weeks later (25 August), and engineering testing continued. The receiver system was designed to mount on the CSO telescope's *Cassegrain relay optics* [9]. First light for the system, which included the active multiplier LO and three WASP spectrometers for a total IF output bandwidth of 11 GHz (22 GHz instantaneous RF bandwidth), was on 28 August. The first detected object was Mars at an LO frequency of 240 GHz, and the measured average receiver noise temperature was 290 K (DSB) (Mars happened to be at opposition and was a bright, unmistakable target). In spite of the very limited time available for testing and adjustments, this first engineering and observing run at CSO proved that the receiver design was essentially sound, and that a sensitive heterodyne spectrometer with very large instantaneous bandwidth was feasible — that with a bit more work, this prototype receiver could itself evolve into that instrument. The following March the receiver returned to CSO with a new mixer chip, a

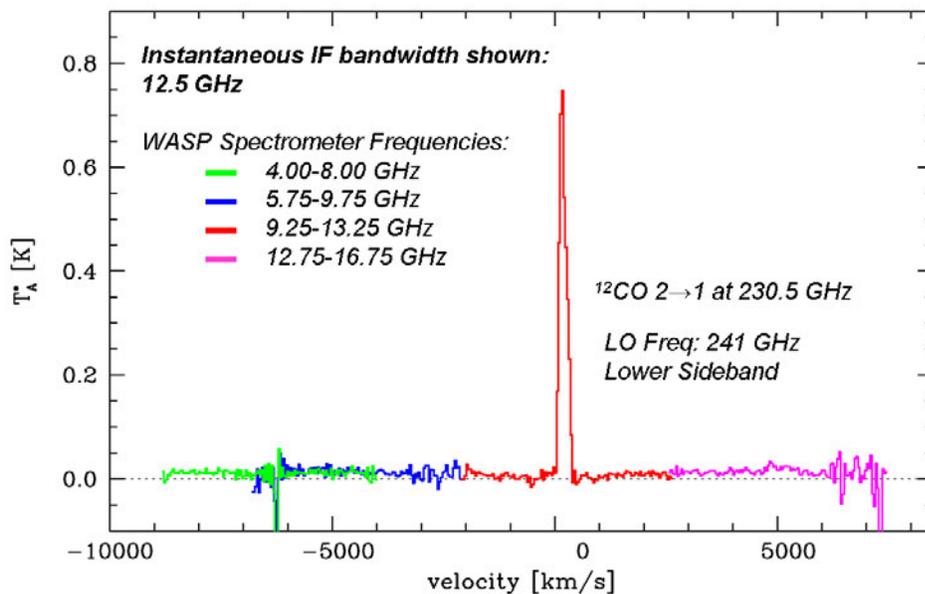


Figure 1-1: A March 2004 wide-bandwidth, low-resolution spectrum of the galaxy M82. The wide-band IF output of the receiver was fed to the Sumner-designed IF processor/down-converter [8]. The outputs of the down-converter were processed by an array of four WASP analog autocorrelation spectrometers designed by Harris [7]. This second observing run clearly demonstrated the new receiver's potential.

much quieter IF LNA, a better optical match to the telescope, and a very quiet Gunn diode LO borrowed from a CSO facility receiver. Noise temperatures were improved by a factor of three or four: now generally less than 100 K (DSB). The receiver was teamed with a full suite of four wide-bandwidth analog spectrometers (Harris and Zmuidzinas [7]) and was used to obtain wide-bandwidth spectra of several bright galaxies (Figure 1-1 on page 4).

After several more improvements,³ by mid-2005 the receiver's noise performance was another factor of two lower, and the instrument was ready to be used for serious astronomy. The 10.4 meter diameter of the CSO telescope was really too small for CO line searches of high-redshift sources, but the wide IF bandwidth made this new receiver ideal for rapid line surveys of galactic star-forming regions (Figure 1-2 below and Widicus Weaver et al. [10]).

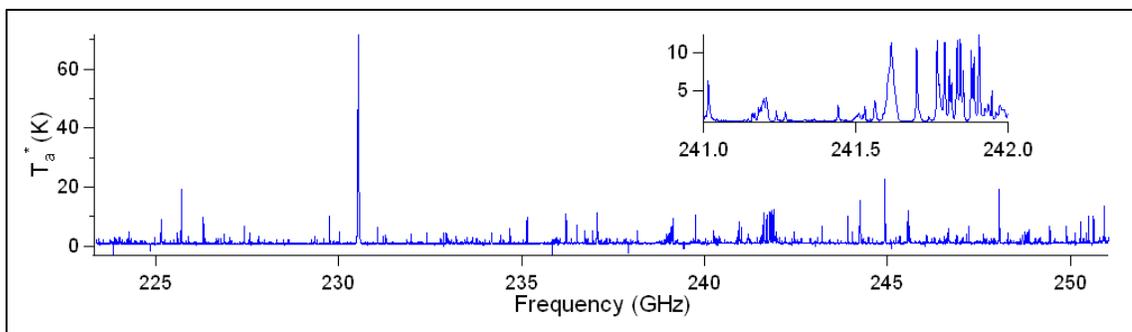


Figure 1-2: A 2007 high-resolution line survey of the Orion KL star forming region. Constructed from several hours of observation using the prototype receiver, this image is from a 2009 presentation by Radhuber et al. [11]. With a sensitivity more than an order of magnitude better than an earlier, similar CSO survey [12], the spectrum contains over 3850 lines with minimum intensities of 0.1K (antenna temperature). The background noise level was less than 30 mK; details of the observation are available in Sumner [8].

³ A spectacularly high-quality JPL rendition of the author's second-generation mixer block design (see Chapter 3), an outstanding JPL mixer chip which even included beam leads (see Chapter 2), and significant improvements in the receiver's electrical and optical interface to the telescope.

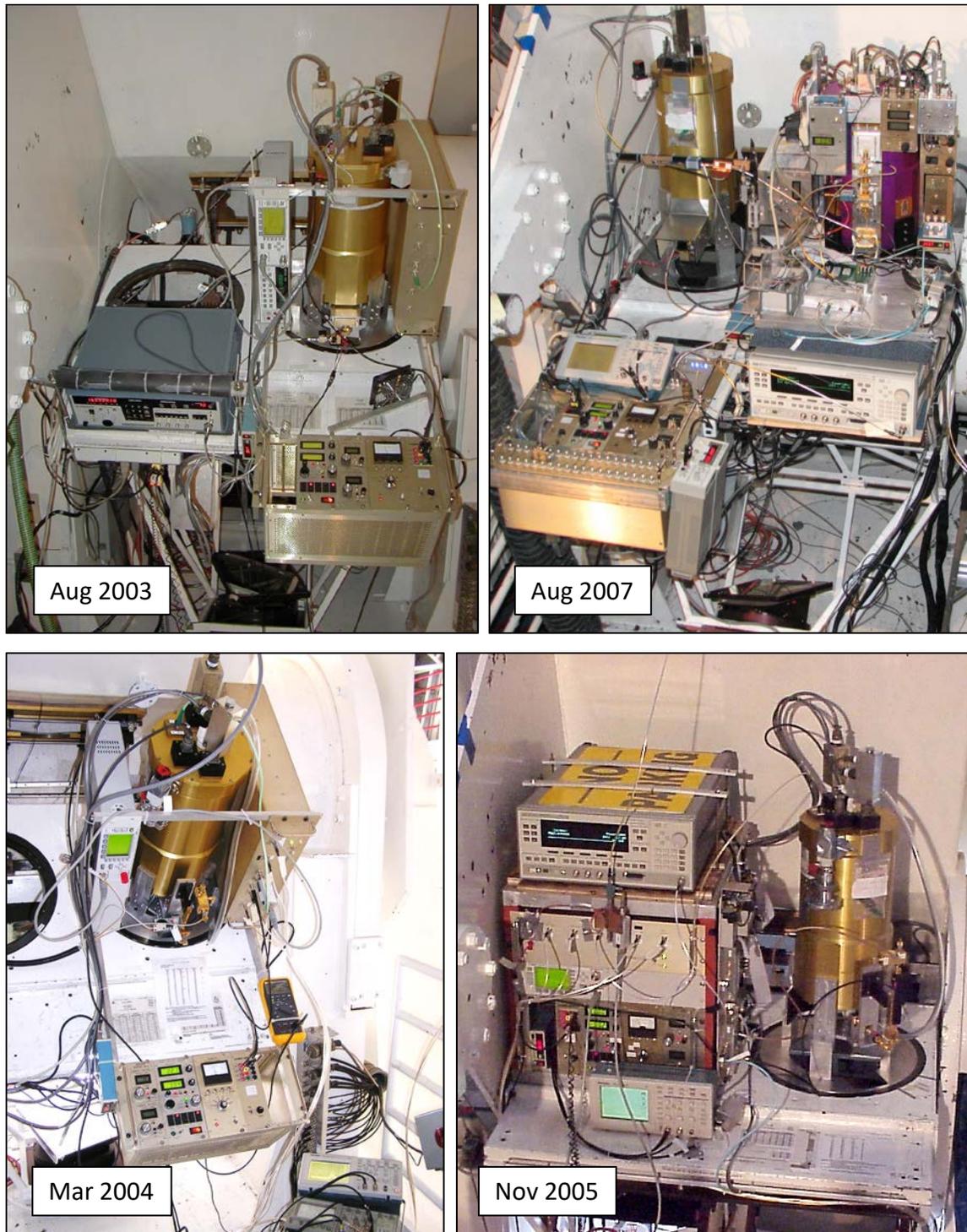


Figure 1-3: Various configurations of the receiver and its electronics at the CSO. The first observing run used a simple synthesized LO setup which was quite noisy. By 2007, it was more sophisticated and performed much better. Best sensitivities, however, required the Gunn LO (lower two photos). In the top right photo the receiver shares the telescope with “Barney,” Jacob Kooi’s 345 GHz prototype heterodyne receiver [13].

Except for a couple of shipments back to Caltech for troubleshooting and repairs, the receiver remained at CSO as a de facto facility instrument. It enjoyed DSB noise temperatures of less than 50 K for IF frequencies of 4–8 GHz, rising to no more than 75 K in the 8–15 GHz IF output range. During its tenure at the observatory the receiver underwent several reconfigurations of its external electronics as the observers and CSO staff gained more experience using and handling the instrument (Figure 1-3 on page 6 and Figure 1-4 below).

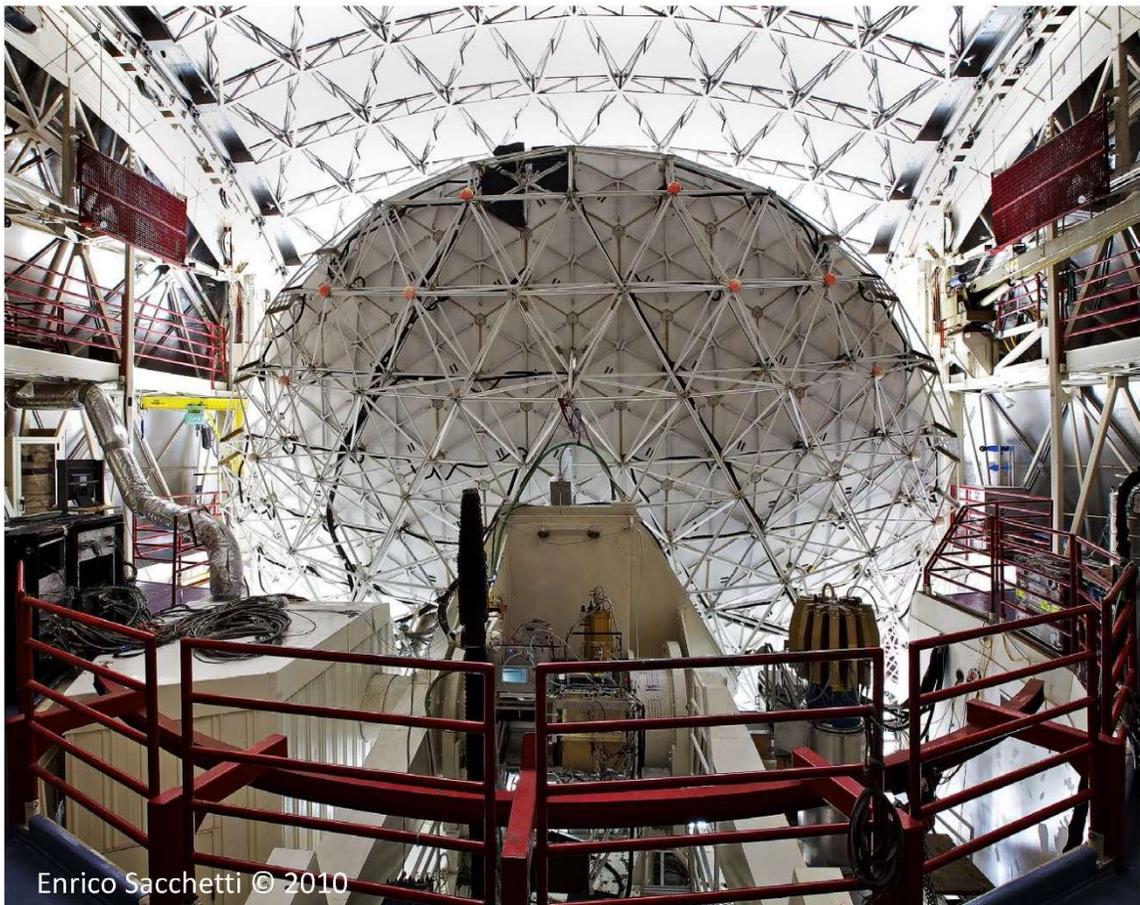


Figure 1-4: Panoramic, interior view of the CSO and telescope with the author's receiver (2010). This spectacular photo from the CSO mezzanine was taken by photographer Enrico Sacchetti (© October, 2010) [14]. It shows the receiver cryostat mounted on the telescope's Cassegrain Relay Optics assembly situated between its two elevation axis bearings [9].

The receiver was retired in 2012 and was replaced by a new, more sophisticated, and slightly quieter 230 GHz facility receiver. The first CSO heterodyne receiver to be

remotely operable, the follow-on instrument's design incorporated many of the lessons learned from this prototype effort. The author also designed the remotely-controlled bias and control electronics for this new facility receiver.

General description of the instrument

The instrument is a single-beam (single-pixel), double-sideband, heterodyne receiver designed to acquire frequency spectra of astronomical sources in the 200–300 GHz band (1.5–1.0 mm wavelength). Housed in a cryogen-cooled (LN₂ and LHe) dewar (cryostat), the heart of the instrument is a niobium-based, thin-film, superconducting receiver circuit on a small silicon chip. The single active element of the circuit is a small SIS (superconductor-insulator-superconductor) tunnel junction. The SIS junction is used in its single-electron (quasiparticle) tunneling mode; Cooper pair (Josephson) tunneling is suppressed by an adjustable magnetic field produced by a small superconducting electromagnet.

A tunable local oscillator (LO) signal is combined with the incoming source radiation (RF signal) by a beam combiner external to the cryostat. This combined signal is focused

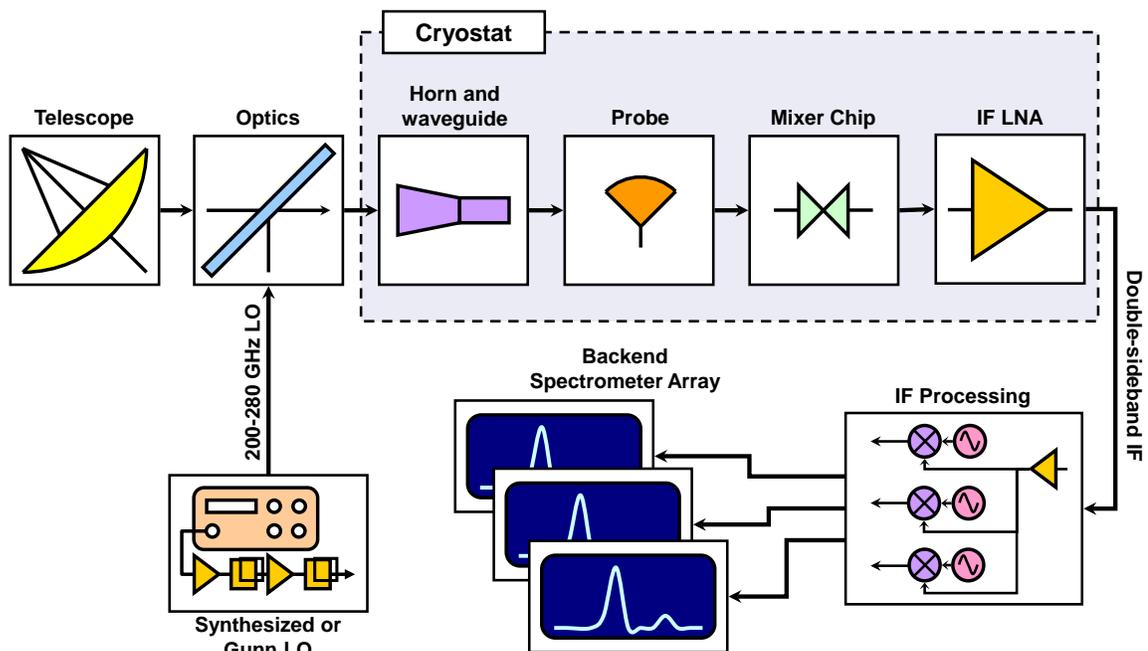


Figure 1-5: High-level block diagram of the receiver system, showing signal flow.

into a conical feed horn which couples it into a small rectangular waveguide. The silicon chip has an integrated *waveguide probe* (antenna) which then couples the waveguide signal into its heterodyne circuitry. The SIS device acts as a heterodyne mixer, producing a phase-coherent image of the source RF signal at its much lower frequency IF output (4–16 GHz). This IF signal is amplified by a cryogenically-cooled, low-noise microwave amplifier (LNA) and routed to the cryostat output.⁴ External to the cryostat, additional amplification and processing of the receiver IF signal is performed before routing it to any of several integrating microwave spectrum analyzers which produce and record the frequency spectrum of the astronomical source (see Figure 1-5). Rounding out the subsystems is an array of receiver bias and control electronics to operate the cryogenic system components.

As could be surmised by examining Figure 1-3 on page 6, the source signal from the telescope enters the receiver cryostat through a window on its bottom cover. Figure 1-6 and Figure 1-7 on page 10 show the configuration of the cryogenic components inside the dewar. The components are mounted to the “cold plate,” which is the bottom surface of a can holding the liquid helium. As shown in the figures, the optics and waveguide horn look downward to intercept the signal from the telescope (and local oscillator). Subsequent chapters of the text discuss the design of this assembly in some detail.

Two local oscillator sources have been used with the receiver: (1) a *synthesized LO chain*, a microwave signal generator (12–18 GHz) whose output undergoes filtering, power amplification, and frequency multiplication ($\times 15$) by a custom-built, millimeter-wave assembly; and (2) a *Gunn diode* millimeter-wave oscillator whose output is frequency multiplied ($\times 3$). The synthesized chain is more versatile and easily tuned, but can be noisy and susceptible to spurious outputs unless very carefully designed. It is also quite complicated and expensive. The Gunn LO is much simpler, potentially quieter, and powerful, but requires careful adjustment of tuning micrometers on the diode’s resonant cavity and is therefore not an agile frequency source.

⁴ The LNA eventually used with the receiver was a prototype 4–12 GHz unit primarily designed by Niklas Wadefalk while working with Sander Weinreb’s Microwave Research Group at Caltech. Its truly spectacular measured performance is specified in Appendix B.

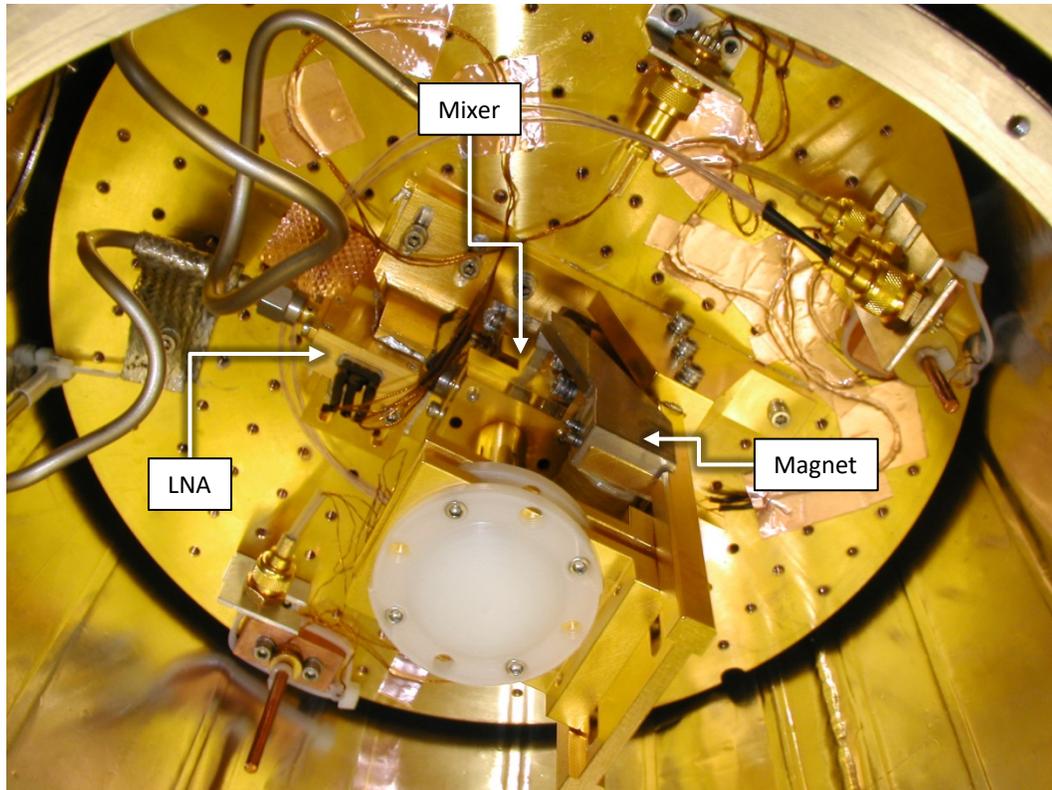
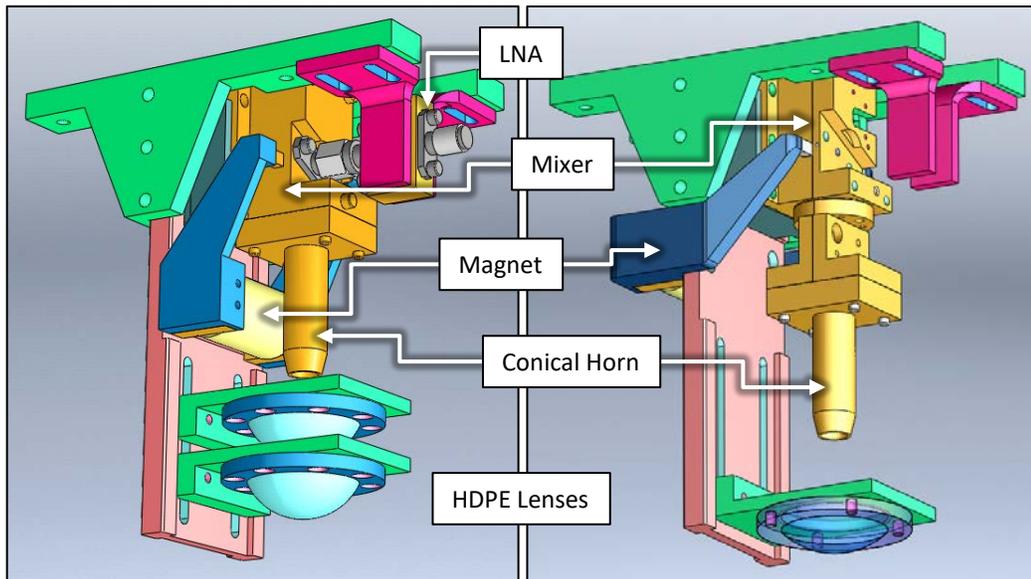


Figure 1-6: Photo of the 2004 mixer assembly mounted to the cryostat cold plate.
 Figure 1-7: Drawings of the original (left) and improved (right) mixer assemblies.



To stabilize its output frequency for high-resolution spectroscopy, the Gunn diode output must be phase-locked to a stable reference oscillator, greatly complicating the LO

system and potentially injecting spurious RF and IF signals into the receiver. In spite of these issues, the Gunn diode LO system represents a mature, stable technology widely used for radio and submillimeter-wave astronomy. On the other hand, the synthesized system represented a new, sophisticated, and flexible technology which has by now replaced Gunn diode LO systems in new designs.

Because the author's focus was on the mixer chip design, his initial goal in developing the mechanical assembly to house and cool the mixer and the optics to couple it to the telescope RF signal was to waste as little thought on these issues as was practical. Because his knowledge of cryostat design and of millimeter-wave optics was essentially nonexistent, he chose to copy as much of the design elements of other group-members' instruments as he could, and to even reuse their parts whenever possible.⁵ As the receiver's performance improved and it spent more time on the telescope, these elements of the design saw many improvements, several introduced by CSO facility personnel.

For example, the cryostat was sized and configured to be compatible with the CSO mounting adapters used for an instrument designed by John Ward for his thesis research [15], so that most of his components could either be reused or their designs copied. The dewar design was a slightly modified copy of that used for John's instrument, the modifications aimed at increasing the cryostat "hold time" (time available between fillings of the dewar cryogen cans to keep the receiver cold) and increasing the cryostat "headroom" (vertical space available for receiver components mounted to the cold plate).

The RF optics and waveguide horn consisted of "found" parts scrounged from the submillimeter instrument labs: a corrugated waveguide feed horn recovered from the first, decommissioned 230GHz CSO receiver and a plastic lens kit intended for use with quasioptical, twin-slot antenna SIS mixers such as those described in Bin [16]. The receiver bias and control electronics (to apply DC bias to the mixer chip, supply the current to the mixer magnet, and to supply power to the LNA) were repurposed components also originally designed for quasioptical instruments. The mixer electromagnet itself was another "found" part wound with superconducting niobium wire.

⁵ Imitation, of course, is the sincerest form of flattery – Charles Colton (1824).

Staying with this “junkyard” theme, the wiring connectors used on the cryostat cold plate as well as the cryostat external electrical connectors and vacuum feed-through components were all reused or spare parts from other projects. Even the cryostat RF window was a carefully-measured, slightly modified copy of windows used on previous projects.

As a result of the author’s lack of effort towards new design and procurement, both the development time and the receiver’s price tag were greatly reduced. Excluding the cost of producing the mixer chips at JPL, the total expenditures for the receiver components were well under \$150k; most of this cost was for the new synthesized LO amplifier-multiplier components, LNAs, and the IF output’s microwave amplifier-processor components. Excluding the cryogenic LNA and the mixer chip, less than \$50K in new expenditures were required for the cryostat and its associated optics, mechanical components, mixer waveguide block, bias electronics, and Gunn LO system.

*Chapter 2***HETERODYNE SIS MIXER CHIP DESIGN****Mixer chip operation and functional requirements**

The heart of a heterodyne receiver for submillimeter-wavelength astronomy is its mixer chip: in this case, a small silicon wafer supporting a superconducting circuit which includes a single superconductor-insulator-superconductor (SIS) tunnel junction as its active element. This chapter describes the design of this device in some detail—several aspects of which represented a clear departure from the conventional wisdom for SIS-based heterodyne mixer design at that time. The finely-tuned complexity of the circuit and the demands it placed on its fabrication required the introduction of new tools and techniques not previously used in submillimeter-wavelength astronomical instrument design. These new tools and techniques have proved very useful for the subsequent development of several other powerful astronomical instruments.

The noise of an SIS-based heterodyne receiver system is predominantly produced by six sources: (1) shot noise at the RF and IF frequencies caused by the SIS bias current, (2) noise produced by the cryogenic low noise amplifier (LNA) attached to the SIS IF output (mainly originating in shot noise from the LNA's bias currents), (3) thermal noise from the various RF sidebands input to the SIS and converted to the IF output, (4) thermal and shot noise generated in the receiver's local oscillator (LO) source, (5) ground-state quantum field fluctuations at each RF sideband which can be detected and amplified by the phase-coherent nature of a heterodyne receiver (see Appendix C), and, finally, (6) spurious signal interference coupled into the receiver system through various imperfections such as ground loops, improper grounding and isolation of the receiver electronics, spurious intermodulation and harmonic signals produced by signal generators associated with the LO and IF processing systems, or other sources of interference.

Of course, mitigating noise due to (6), spurious signal interference, requires careful electrical design of the entire receiver system and its interface to the telescope, which includes considerations well beyond the design of the mixer chip itself. Careful design of the LO and its coupling to the mixer can greatly reduce its contribution to the receiver

system noise. A much more complicated heterodyne detector design which combines multiple SIS mixers using RF and IF “hybrids,” signal combiners which mimic optical beam splitters and beam combiners, can also reduce the impacts of LO noise (4) and sideband thermal noise (3) on receiver sensitivity. These quite complicated receiver designs, however, are generally built up from simple, well-designed, SIS mixer building blocks, and the mixer chip which is the subject of this chapter is intended to be such a building block.

Proper SIS mixer chip design will ensure that (a) the SIS device itself is appropriately designed and fabricated to optimize its ability to efficiently detect incident RF radiation and convert it to the required IF output frequency; (b) the incident RF signal power is efficiently coupled to the SIS junction, minimizing reflective and dissipative losses; and (c) the resulting SIS IF signal output is efficiently coupled to the IF LNA. Successful execution of the design will then optimize the receiver’s sensitivity by minimizing the impacts of noise sources (1), (2), and (5) described above.

The receiver’s mixer chip is physically small (approximately 2 mm long by 1/4 mm wide), but it contains the great majority of the high-frequency electronic circuitry of the instrument. Incoming RF radiation combined with a much more powerful, coherent signal from the receiver’s LO are injected into the superconducting mixer circuitry by the mixer chip’s integrated *waveguide probe*. The chip’s RF circuit then couples the combined RF and LO signals to its SIS tunnel junction. When properly biased with a small, constant (DC) voltage, the highly nonlinear SIS device detects the tiny modulations of the input signal amplitude caused by the presence of the telescope RF signal. As a result, the SIS generates a microwave *intermediate frequency* (IF) signal corresponding to these input amplitude variations. Additional mixer chip circuitry filters and isolates the SIS IF output, which contains the information in the RF radiation collected by the telescope. The mixer chip IF output has a design bandwidth of over 12 GHz, and this signal is subsequently amplified and its power spectrum analyzed by additional receiver IF systems.

As outlined above, the mixer chip design thus integrates the following elements and functions:

- (1) An SIS tunnel junction device used to perform efficient heterodyne detection of the RF signal over the entire RF and IF operating bandwidths.
- (2) A waveguide probe to couple the RF radiation from the telescope and LO into the mixer circuitry. The probe must efficiently couple this radiation over the receiver's large RF design bandwidth of 200 GHz to 300 GHz. Its design is the subject of Chapter 3.
- (3) An *RF matching network* to efficiently couple RF power from the probe to the SIS device over this broad bandwidth, optimizing the receiver sensitivity without introducing instability and oscillations.
- (4) An *RF choke*: circuitry to block RF signals from propagating beyond the mixer chip into lower-frequency parts of the circuit, where they may be reflected back. This function is required to prevent the uncontrollable amplitude and phase of an unwanted reflection from adversely affecting the RF matching network performance.
- (5) An *IF matching network* to efficiently couple the SIS IF output signal to the processing electronics external to the mixer chip over its entire operating bandwidth of at least 12 GHz without introducing instability and oscillations.
- (6) A DC (zero frequency) coupling circuit to provide for constant-voltage biasing of the SIS device to optimize its response to the combined RF and LO signal.

Three additional, very important requirements (probably not clear from the above descriptions) must be imposed on the mixer chip design. The first requirement is for its circuitry to be efficiently and accurately referenced to a single, well-defined *ground potential* from DC up to the maximum RF frequency of the receiver system. This potential must also closely match that of the RF signal waveguide wall surrounding the waveguide probe. Unless this requirement is met, the receiver system can experience unacceptable frequency response variations and instability. The second requirement is that its IF output signal be a linear function of the telescope RF input signal over the expected total RF input power level range: the mixer chip (and the associated IF output

processing electronics) must have tolerable *output saturation* in the presence of large RF input power.

The final requirement, for this is after all a research project, is that the mixer circuit design should advance the “state of the art” in high-frequency, superconducting microcircuit design and fabrication technologies in such a way that it could lead to the practical realization of much more sophisticated designs and capabilities. It should push that development by setting a high bar (but not too high). To succeed, the mixer design must not require years-long, revolutionary fabrication technology development efforts, nor must it demand impractical levels of accuracy from the microcircuit fabrication process (both in circuit layer alignments and circuit feature dimensions). Regardless of its complexity, the design should be tolerant of expected fabrication process uncertainties without significantly affecting the mixer chip’s final performance.

Technology innovations introduced by the design

To successfully meet the requirements outlined in the previous section, the mixer chip development introduced several innovations involving design techniques, materials, and fabrication processes. Foremost among these was the development of the SuperMix software library, the powerful set of tools used to numerically model and optimize the design of the mixer chip circuitry. In fact, this effort was the first to use SuperMix as its primary tool for superconducting heterodyne SIS mixer design.

The use of SuperMix modeling and optimization code made it practical to develop a highly optimized mixer design which could efficiently operate over very wide RF and IF bandwidths without requiring adjustable tuning elements. Such a level of performance had not been available in previous receivers and was therefore a major design goal of the project. Achieving this goal resulted in two parallel innovations: firstly, the mixer’s RF bandwidth requirement was met by incorporating only a single SIS junction device. This achievement was notable because the very few contemporary SIS heterodyne mixers combining wide, fixed-tuned RF bandwidths with high-performance had accomplished this result only by using networks of up to as many as eight SIS junctions. Recent developments at Caltech and JPL, however, had demonstrated the fabrication of SIS

junction devices with unprecedented combinations of high-frequency response, heterodyne conversion efficiency, and noise level. Their work was motivated by the need for terahertz-frequency heterodyne receivers as part of the HIFI instrument (Heterodyne Instrument for the Far Infrared) in the joint European Space Agency – NASA *Herschel Space Observatory*, launched in 2009. These local SIS fabrication developments also proved to be particularly appropriate for this much lower RF frequency receiver effort (200–300 GHz vs. 1.0–1.3 THz for HIFI).

The second major design goal enabled by the use of SuperMix is the receiver's wide IF output frequency range. This goal has been accomplished by the incorporation of a unique circuit innovation on the mixer chip: combining both RF choke and IF matching network functions in a single small structure composed of a cascaded series of thin-film, superconducting transmission line sections. Although previous SIS mixer chip designs usually incorporated at least a part of the RF choke circuitry on chip, they then invariably relied on external, conventional microwave circuit elements to accomplish the required IF matching to the SIS devices. In particular, SIS IF matching has commonly been achieved by using a bulky *microwave isolator*, an element which effectively limits the system IF output bandwidth to a single octave. This new mixer IF design innovation completely eliminates the need for an isolator, allowing close physical coupling of the mixer chip to the IF system's first amplifier. The results are effective RF isolation along with very wide bandwidth IF performance; these results are accomplished using a circuit which occupies an area of only 0.04 mm^2 on the mixer chip ($0.8 \text{ mm} \times 0.05 \text{ mm}$).

A final innovation introduced by this effort has been the use of silicon as the waveguide probe mixer chip substrate. Previous submillimeter receivers almost universally used quartz as the substrate material for waveguide probe mixers. Silicon, however, offers a major advantage over quartz: it can be very accurately micro-machined using chemical etching processes unsuitable for use with quartz. These processes can take advantage of mature technologies developed for the commercial semiconductor industry, offering a relatively low-risk route to the incorporation of advanced mechanical and electrical features. In this design, the use of silicon has made practical the incorporation of gold metal *beam leads* whose dual purpose is to both mechanically support the mixer chip along its edges and to effectively couple its ground plane to the surrounding

waveguide structure (see Figure 2-1 on page 19). The use of beam leads with SIS waveguide mixers was very unusual at the time of this development, mainly because of the difficulties presented by their fabrication on a quartz substrate.

Because of their commercial importance, very high quality silicon wafers are readily available. Unlike crystalline quartz, silicon also has highly advantageous, isotropic dielectric characteristics. At submillimeter wave frequencies and when cooled to the operating temperature required for the niobium superconducting circuitry (typically 4 K), silicon is a very nearly ideal, lossless dielectric. The choice of silicon has one important disadvantage, however, when used for a waveguide mixer substrate. Silicon's dielectric constant of 11.9 is significantly larger than that of quartz (5.6). When used to support a waveguide probe, the dielectric substrate must be positioned in a channel extending away from a waveguide wall (Figure 2-2 on page 20). This channel can allow electromagnetic radiation to escape from the waveguide and propagate within it. If the channel supports wave propagation for radiation within the design bandwidth of the receiver, then its presence can have a crippling effect on the receiver's frequency response. Silicon's large dielectric constant will lower the cutoff frequencies of the channel propagating modes, so a careful selection of the geometries of the mixer chip substrate and the channel supporting it is critical. The solution is to use a thin substrate and to mount it in a channel which provides a properly-sized vacuum gap above it; Chapter 3 and Appendix F hold the details. In this design, the silicon substrate is very thin: 25 microns. Consequently, its fabrication by JPL demanded the development of special processes and techniques.

Mixer chip general layout

The final prototype mixer chip design has the general layout and features shown in Figure 2-1 on page 19. Its design will be discussed in detail in subsequent sections of this chapter. Precise dimensions of the various mixer chip features are provided in Appendix A starting on page 111. Thin-film, niobium circuitry is deposited in a few layers on one surface of the chip's small, thin ($1990 \times 230 \times 25$ micron) silicon wafer. A niobium metal ground plane covers approximately one half of the wafer's surface. Deposited above this ground plane is a silicon oxide (SiO) insulating layer followed by another niobium layer which defines the mixer circuitry and its integrated waveguide RF probe. When cooled to

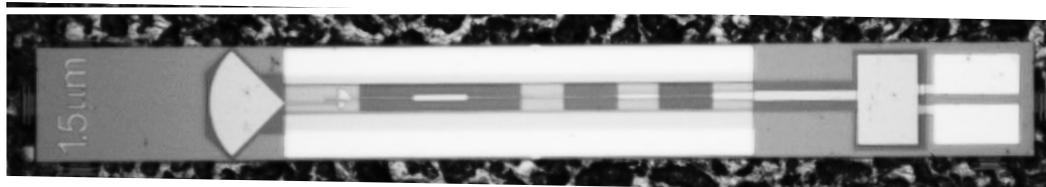
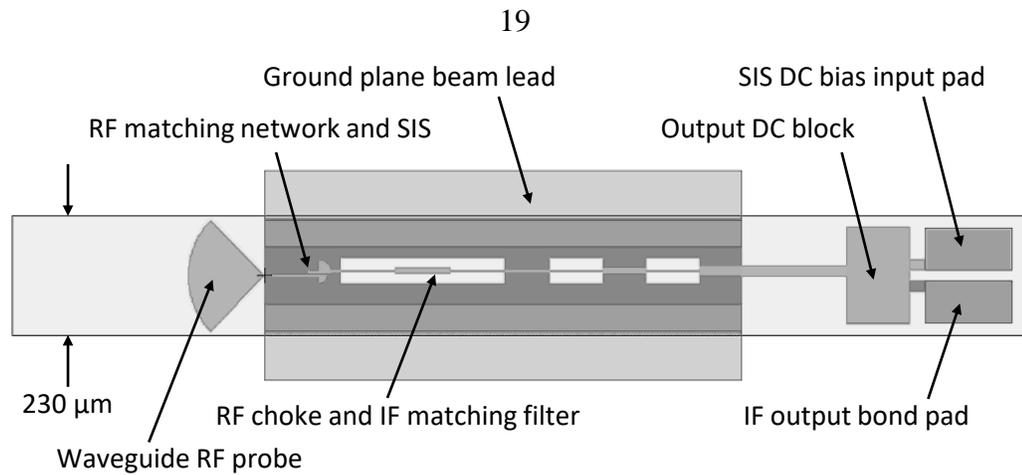


Figure 2-1: Diagram and photo of the mixer chip showing its general features.

The dimensions of the silicon substrate are $1990 \times 230 \times 25$ microns. The circuit ground plane and wiring are superconducting niobium; the beam leads and bond pads are gold. The photo is of a mixer chip from the first generation manufacturing run at JPL in 2003, which did not include beam leads. The “1.5 μm ” in the photo shows the size of the SIS device on that particular chip. Several sizes were included to accommodate manufacturing tolerances; the writing did not affect the chip behavior because it did not protrude into the mixer waveguide (see Figure 2-2 on page 20).

its operating temperature of approximately 4 K, the mixer chip’s niobium circuitry becomes superconducting.

Gold plating was layered onto the bond pads and over the outer edges of the ground plane to support wire bonding to these surfaces. The first chip fabrication runs did not include the gold beam leads, and the first chip used for initial receiver testing did not have them. Later chip runs did include beam leads, once their fabrication was perfected. The final version of the receiver did incorporate one of these later chips.

The single active element of the circuitry is a tiny, niobium SIS tunnel junction (1.3×1.3 micron) employing an aluminum nitride (AlN_x) insulating barrier on the order

of one nanometer thick.⁶ The SIS junction was formed from several layers stacked between the niobium circuit and ground layers. The chip was fabricated by a JPL Microwave Experiment Systems and Technology Section team led by Dr. Henry G. (Rick) LeDuc; their development of the several processes required to successfully produce the design demanded many months of hard work, as detailed in Kaul et al. [17].

How the mixer chip is mounted relative to the receiver RF waveguide is illustrated in Figure 2-2. The apex of the mixer chip's pie-shaped waveguide probe is aligned flush with the waveguide wall, and the mixer chip substrate is then long enough to extend completely across the waveguide as shown in that figure. Chapter 3 addresses the details of the waveguide and the mixer chip mounting structure designs.

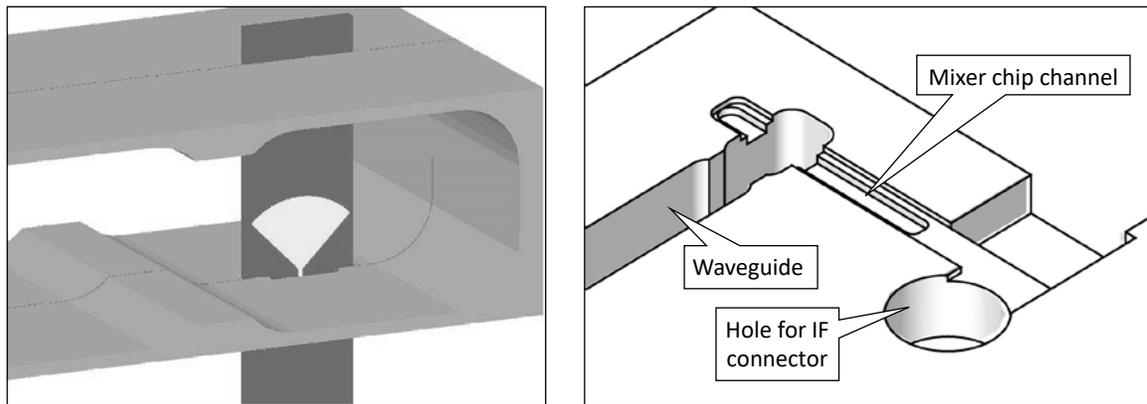


Figure 2-2: Diagrams illustrating mixer chip mounting relative to the RF waveguide.

SIS junction behavior, characteristics, and specifications

Here we review some of the relevant physics of an SIS device in the presence of RF frequency radiation needed for the mixer design effort. An SIS junction is a two-terminal electronic device with a highly nonlinear voltage-current relationship (discussed later).

⁶ Actually, several choices of SIS dimensions were included in the mask layout for a chip fabrication run from a single blank silicon wafer. SIS sizes ranged from 1.2×1.2 micron to 1.6×1.6 micron. In this way fabrication uncertainties could be accommodated by choosing that SIS size which gave results closest to the specified design.

Because the scale of the SIS nonlinearity can be quite dramatic over an energy range of a small fraction of an meV (10^{-3} electron volt), oscillating signals with frequencies of 100 GHz or higher have photon energies comparable to or even much larger than this SIS nonlinearity energy scale (at 300 GHz, $h\nu = 1.24$ meV). Consequently, at millimeter wave frequencies and beyond a classical analysis of the behavior of this device is completely inadequate. When the SIS is properly biased with a small, constant (DC) voltage, an oscillating LO signal induces a corresponding current through the junction because of a purely quantum-electrodynamic effect called *photon-assisted, quasiparticle tunneling*, first observed by Dayem and Martin in 1961–1962 [18]. This behavior of the SIS can also make it a remarkably sensitive device for *heterodyne detection* of a weak RF signal: the slight modulation of the amplitude of a strong LO combined with a weak RF signal causes a corresponding modulation in the SIS current at the difference of the LO and RF frequencies, thus generating a broadband, *intermediate frequency* (IF) signal. A successful, comprehensive quantum theory of heterodyne detection by SIS quasiparticle tunneling was developed by John Tucker in 1979 [1]. This quantum theory demonstrated that the output IF power of the SIS can exceed that of the input RF signal: the SIS junction can act not only as a detector, but also as an amplifier of the resultant IF signal, a behavior not possible according to purely classical analysis of a nonlinear circuit element (the additional power, of course, is supplied by the strong LO signal and by the DC bias on the SIS). In the following sections we investigate the consequences of this quantum behavior of the SIS with an eye toward how to best make use of its capabilities as a heterodyne detector.

SIS DC I-V characteristic and photon-assisted tunneling

A typical SIS DC I-V characteristic is shown in Figure 2-3 on page 22. The SIS I-V shows an extremely sharp nonlinearity near voltages of $\pm V_{\text{gap}}$, the niobium superconductor's 4.2 K *gap voltage* of 2.796 mV (its *gap energy* is therefore 2.796 meV)

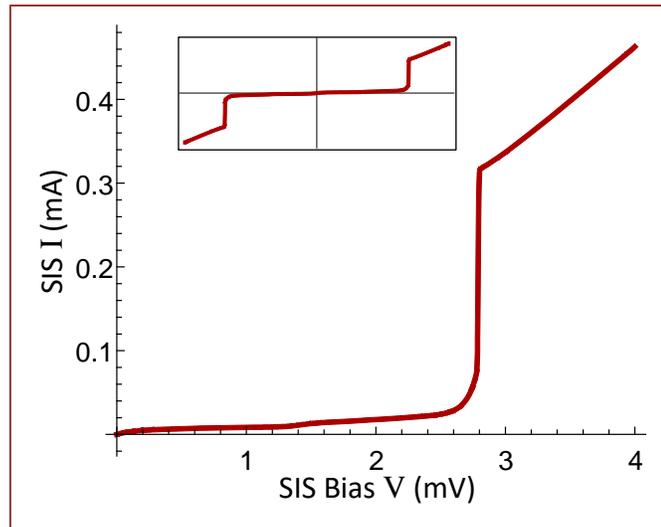


Figure 2-3: DC I-V characteristic of a typical SIS device used in the prototype receiver. This is the measured I-V curve at 4.2 K of the mixer chip SIS installed in the receiver in 2005. The SIS gap voltage (V_{gap}) is 2.796 mV (the steepest part of the curve), and its normal resistance (R_n) is 7.67 ohm (given by the slope well above V_{gap}). The inset shows the I-V behavior for negative bias voltages. This SIS device was fabricated at JPL [17].

[19] [20].⁷ As shown in the inset of Figure 2-3, the I-V is an odd function of the bias voltage. Some characteristics of the nonlinearity in this I-V curve: the SIS current increases by more than a factor of 11 between 2.6 and 2.9 mV, and its maximum slope is over 22 mA/mV, representing a dynamic resistance of only 0.6% of the 7.67 ohm SIS normal resistance (R_n).

The SIS junction's very thin insulating layer offers a potential barrier to conduction electrons on either side of it, so electron state wave functions on either side of the barrier have evanescent characters within the barrier. If the barrier is thin enough, then wave

⁷ The identification of the superconductor gap energy with the SIS junction gap voltage was made by Ivar Giaever in 1960 [19] [20]. His experiments unequivocally confirmed the existence of the superconducting energy gap formed around the normal-metal Fermi energy predicted by BCS theory, earning him the 1973 Nobel Prize in physics. Giaever also invented the “semiconductor picture” of the superconducting state, now widely used to provide an intuitive picture of quasiparticle (single electron) tunneling in SIS devices.

functions from the two sides of the barrier may have a significant overlap integral, leading to the possibility for electron tunneling across the barrier. In the superconducting state, however, few of these single-electron states are occupied; the superconductor's normal-state band gap at the Fermi level represents the minimum energy required to break a superconducting Cooper pair and generate a pair of occupied single-electron states, one above and one below the Fermi energy.⁸

With a small applied DC bias voltage across the SIS junction, the difference in energy between the highest unoccupied quasiparticle state below the Fermi level on one side of the barrier and the lowest quasiparticle state above the Fermi level on the other side is $\Delta E = eV_{\text{gap}} - eV_{\text{bias}}$. A slow, adiabatic increase in the electric potential across the insulating barrier (applied by increasing V_{bias}) has little effect until the potential difference reaches the superconductor's gap voltage V_{gap} . At that point, however, a Cooper pair on the high-potential side of the barrier can spontaneously break, with one quasiparticle state becoming occupied below the Fermi level on the high-potential side, as the other electron of the pair tunnels through the barrier to occupy a state above the Fermi level on the low-potential side of the barrier. This behavior gives rise to the dramatic increase in tunneling current at the superconductor's gap voltage shown in Figure 2-3. Continued increases in the DC bias voltage beyond V_{gap} simply increase the number of empty single-particle states with equal energies available on the two sides of the barrier, increasing the tunneling current very nearly linearly with DC bias: this slope defines what is called the SIS junction's *normal resistance*, R_n .

A careful examination of the SIS I-V shown in Figure 2-3 will reveal that the tunneling current just above V_{gap} is less than V_{gap}/R_n , as one might otherwise naively expect. In fact, for an ideal SIS junction at 0 K (no leakage current below V_{gap}), a microscopic theory of the tunneling Hamiltonian in terms of the quasiparticle densities of states near the Fermi level [21] shows that $I_{\text{dc}}(V_{\text{gap}}^+) \approx (\pi/4) V_{\text{gap}}/R_n$. For a junction at finite temperature, one should add the leakage current expected just below V_{gap} . For the

⁸ Actually, the idea of single-electron states in a cold metal is misleading, because of the no longer negligible electron-electron interactions. Instead, one talks of Landau's single-“quasiparticle” states; see the Glossary.

SIS in Figure 2-3, $(\pi/4)V_{\text{gap}}/R_n = 0.28 \text{ mA}$; adding the leakage current (0.03 mA) results in a just-above-gap current very close to the measured 0.32 mA.

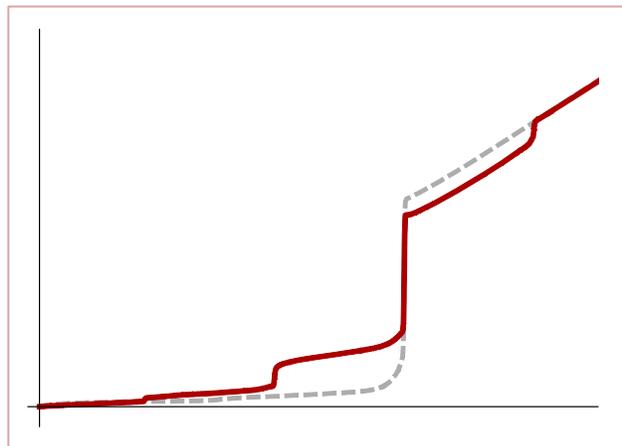
The presence of a high-frequency oscillation in the electric field across the barrier can result in *photon-assisted tunneling* of single electrons across the barrier. Conceptually, if the oscillation's photon energy $h\nu > eV_{\text{gap}} - eV_{\text{bias}}$, then an absorbed photon can supply the additional energy needed to break a Cooper pair and tunnel one of its electrons across the barrier. This very low energy analog of Einstein's photoelectric effect leads to the formation of the *photon steps* observed by Dayem and Martin in what is now called a "pumped" SIS DC I-V curve, as illustrated in Figure 2-4.

The photon step in this simulated I-V curve appearing just below V_{gap} is generated by photon-assisted tunneling due to absorption of a single photon per quasiparticle pair creation; it has a width of $V_{ph} = h\nu/e = 1 \text{ mV}$ for the 240 GHz simulated signal. Another, smaller step is visible at bias voltages within a millivolt below the first step. This second step signifies two-photon absorption events. At bias voltages above V_{gap} , photon-assisted tunneling of electrons back across the barrier can regenerate Cooper pairs, reducing the net DC tunneling current, as shown.

The equations which describe Withington and Kollberg's multi-harmonic extension of Tucker's quantum theory of the SIS junction are given in Appendix E: *Fast SIS harmonic balance algorithm* starting on page 155 [22] [1]. Also in that appendix is a very brief description of the origin of these equations from the time evolution of the relative phases of the quasiparticle states on either side of the insulating barrier and its effect on

Figure 2-4: Example of a pumped SIS DC I-V curve.

SuperMix was used to simulate this DC I-V curve (solid) using the SIS I-V characteristic from Figure 2-3 (dashed). Tucker's theory was used to model the DC I-V curve in the presence of a 240 GHz signal with 0.9 mV peak amplitude across the SIS junction. The photon steps are clearly visible.



the electron tunneling Hamiltonian. Consider the lowest-order nontrivial relationship derivable from those equations: assume that the voltage across the SIS barrier is $v(t) = V_0 + V_{\text{LO}} \cos(\omega_{\text{LO}} t)$, and that the resulting current is $i(t) = i_0 + \text{Re}[i_1 \exp(j\omega_{\text{LO}} t)]$, where i_1 is the complex current phasor of the oscillating signal across the SIS junction, and ω_{LO} is its angular frequency (anticipating that this will become the local oscillator frequency). Ignoring higher harmonics of ω_{LO} and using the notation from Appendix E, then the lowest-order nonzero coefficients $A_{(m),n}$ defined by equations (E.6) and (E.7) are $A_{0,1} = J_0(\alpha)$ and $A_{\pm 1,1} = J_1(\alpha)$, where $\alpha \equiv V_{\text{LO}} / V_{ph} = e V_{\text{LO}} / \hbar \omega_{\text{LO}}$ specifies the LO signal amplitude and the J_i are Bessel functions. From (E.5) the nonzero convolution coefficients $C_{(k)}$ are then $C_0 = A_{0,1}$ and $C_{\pm 1} = A_{\pm 1,1}$. Finally, using (E.2) and (E.3), the SIS pumped I-V current is found to be:

$$\begin{aligned}
 i_0(V_0) &= J_0(\alpha)^2 I_{\text{dc}}(V_0) + J_1(\alpha)^2 \left[I_{\text{dc}}(V_0 - V_{ph}) + I_{\text{dc}}(V_0 + V_{ph}) \right] \\
 &\approx I_{\text{dc}}(V_0) + \frac{1}{4} V_{\text{LO}}^2 \left[\frac{I_{\text{dc}}(V_0 - V_{ph}) - 2I_{\text{dc}}(V_0) + I_{\text{dc}}(V_0 + V_{ph})}{V_{ph}^2} \right] \quad (2.1) \\
 &\sim I_{\text{dc}}(V_0) + \frac{1}{4} \alpha^2 (V_0 + V_{ph}) / R_n; \quad I_{\text{dc}}(V > V_{\text{gap}}) \gg I_{\text{dc}}(V < V_{\text{gap}}).
 \end{aligned}$$

In this expression V_0 is the SIS DC bias voltage and V_{LO} the applied signal voltage. $I_{\text{dc}}(V)$ is the quiescent SIS DC current at voltage V from its I-V characteristic, and $i_0(V_0)$ is the resulting pumped I-V current at the DC bias voltage. Expanding the first equation above to second order in α results in the approximation shown: the SIS DC current acts as a *square-law detector* of the applied signal voltage (V_{LO}), but instead of using the second derivative of the I-V curve at the DC bias point, the quantum theory requires the *second finite difference* in the signal's photon voltage V_{ph} around the DC bias point. The final expression gives an order of magnitude estimate for a DC bias point on the first photon step below the SIS gap voltage. In the example shown in Figure 2-4, $\alpha \approx 1$, and the pumped DC bias current of the photon step is indeed roughly $1/4$ of the quiescent SIS I-V characteristic curve current just above the gap.

Expressions to a similar order of approximation for the oscillating component of the tunneling current are (with $\alpha \approx 1$):

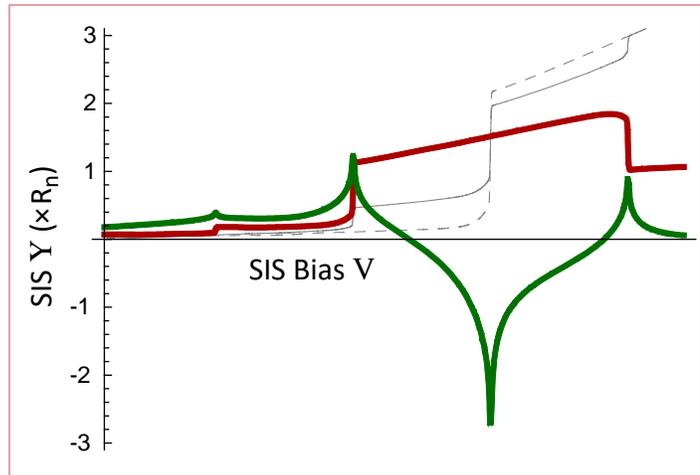
$$\begin{aligned}
 i_1(V_{LO}) &= J_0(\alpha)J_1(\alpha) \left[I_{dc}(V_0 + V_{ph}) - I_{dc}(V_0 - V_{ph}) \right] + \\
 &\quad j J_0(\alpha)J_1(\alpha) \left[I_{kk}(V_0 - V_{ph}) - 2I_{kk}(V_0) + I_{kk}(V_0 + V_{ph}) \right] \\
 &\approx 0.3V_{LO} \left[\frac{I_{dc}(V_0 + V_{ph}) - I_{dc}(V_0 - V_{ph})}{V_{ph}} \right] + \\
 &\quad j 0.3V_{LO} \left[\frac{I_{kk}(V_0 - V_{ph}) - 2I_{kk}(V_0) + I_{kk}(V_0 + V_{ph})}{V_{ph}} \right].
 \end{aligned} \tag{2.2}$$

The oscillating tunneling current at ω_{LO} has a dissipative term (real part) with a resistance $\sim R_n$.

Evidently, the oscillating tunneling current given by (2.2) also has a non-dissipative, reactive term (actually a *susceptance*) with no classical counterpart (we have not included the SIS physical capacitance in this analysis). The function $I_{kk}(V)$ in (2.2) is the *Kramers-Kronig transform* of the SIS DC I-V characteristic $I_{dc}(V)$. It is defined in Appendix E, equation (E.4) and is illustrated in Fig. E-2 on page 158. The term $(-j I_{kk})$ provides an analytic extension of I_{dc} into the complex plane in a way that enforces the causal relationship between $v(t)$ and $i(t)$. From an examination of $I_{kk}(V)$ in Fig. E-2, it should be clear that this susceptance, which involves a second difference of I_{kk} around the DC bias point, will be capacitive ($\text{Im}[Y] > 0$) for DC bias voltages over the lower portion of the

Figure 2-5: Complex SIS LO quantum admittance vs. DC bias.

Superposed on the unpumped and pumped DC I-V curves (dashed and solid gray) are the real (red) and imaginary (green) parts of the SIS admittance for an LO of 240GHz with α of 0.9 (same as in Figure 2-4). The units of the admittance are $1/R_n$. This is a SuperMix-generated result.



first photon step, but it will become inductive approaching V_{gap} (see Figure 2-5). This imaginary part of the SIS susceptance has its origin in virtual transitions of tunneling electrons between quasiparticle states differing by energy $\hbar \omega_{\text{LO}}$, causing a phase shift in $i_1(V_{\text{LO}})$ in some ways analogous to the quantum origin of the index of refraction of visible light in transparent, dielectric materials. The sharper the current jump at the SIS gap, the higher the peak in I_{kk} there, and thus the larger the reactive terms at the edges of the photon steps. A much more thorough description of the effects discussed here is provided by Mei Bin's thesis [16], which is suggested as further reading before tackling Tucker's paper. This variation in quantum susceptance with DC bias voltage can be useful for making fine adjustments to the impedance match between the SIS and the RF circuit.

Appendix E, equation (E.13) gives the formula for the small signal admittance matrix describing the heterodyne performance of an SIS junction.⁹ According to that expression the coupling of the upper and lower sidebands to the IF output, in the limit that $\omega_{\text{IF}} \ll \omega_{\text{LO}}$, is

$$Y_{0,\pm 1} = \pm \frac{1}{V_{ph}} \sum_{k=-\infty}^{\infty} C_{(k)} C_{(k \mp 1)}^* \left[I_{\text{dc}}(V_0 + kV_{ph}) - I_{\text{dc}}(V_0 + (k \mp 1)V_{ph}) \right]. \quad (2.3)$$

Using the same lowest-level approximation as before, wherein the nonzero $C_{(k)}$ become $C_0 = J_0(\alpha)$ and $C_{\pm 1} = \pm J_1(\alpha)$,

$$Y_{0,\pm 1} = J_0(\alpha) J_1(\alpha) \left[\frac{I_{\text{dc}}(V_0 + V_{ph}) - 2I_{\text{dc}}(V_0) + I_{\text{dc}}(V_0 - V_{ph})}{V_{ph}} \right]. \quad (2.4)$$

Again, a second finite difference of the SIS I-V curve around the DC bias point determines the heterodyne responsivity of the device. $Y_{0,\pm 1}$ is maximized (at this level of approximation) for $\alpha = 1.08$ and $J_0(\alpha) J_1(\alpha) = 0.34$. This condition therefore provides a typical LO pumping target amplitude, explaining the choice of $\alpha \approx 1$ in the previous

⁹ The *admittance matrix* \mathbf{Y} expresses the output currents of a device in terms of its input voltages. In this case the element indices designate frequencies: index 0 is the IF, and ± 1 designate the upper and lower RF sidebands. See: *Representing a heterodyne receiver* in Chapter 5 and also Appendix C.

examples. For $V_{ph} \approx 1$ mV, appropriate for a 230 GHz receiver, then $Y_{0,\pm 1} \sim 1/R_n$, a not-unexpected result.

At this lowest order of approximation $Y_{0,0}$, the SIS output admittance at the IF frequency, is given by the slope of the pumped I-V curve at the DC bias voltage, V_0 . In general, this slope at the bias point can be expected to be somewhat less than $1/R_n$, as is the case in Figure 2-4 on page 24. Because the SIS input admittance at the RF frequency can be expected to be $\sim 1/R_n$, as is $Y_{0,\pm 1}$, the RF input current and resulting IF output current amplitudes are about the same. The SIS IF output current will induce a voltage across the impedance produced by the parallel combination of $Y_{0,0}$ and the IF load admittance, an equivalent impedance which can exceed R_n . Thus the mixer IF output voltage amplitude can exceed the RF input voltage amplitude. This RF-to-IF voltage gain results in a similar power gain because the RF and IF currents are approximately equal. Thus it is possible to achieve power gain from an SIS when used as a heterodyne detector, performance not predictable from a purely classical theory.

SIS capacitance and receiver RF bandwidth

Of course, the capacitance of the two closely spaced superconducting surfaces of the SIS will add another reactive term to the SIS current. This capacitance can dominate the SIS conductance at high frequencies, greatly reducing the RF voltage across the insulating barrier and therefore its heterodyne detection capabilities. A major function of the linear mixer circuitry matching the RF source to the SIS detector will be to “tune out” this capacitive shunt by providing a compensating inductance over the design RF bandwidth of the receiver. An order of magnitude estimate of the maximum RF bandwidth attainable for the SIS as a heterodyne detector is given by the device’s $R_n C_{SIS}$ -limited frequency: $f_{BW} \sim (2\pi R_n C_{SIS})^{-1}$. If this frequency is too small, then multiple SIS devices will be required to meet the receiver RF bandwidth and sensitivity requirements. Making an SIS device with a thinner insulating barrier exponentially increases the electron tunneling probability, thereby exponentially decreasing its R_n . Its capacitance, however, grows only inversely with barrier thickness reductions, so a thinner barrier can increase the SIS f_{BW} . For JPL-produced SIS junctions with an aluminum nitride insulating barrier on the order of 1nm thick, the specific capacitance C/A between the two superconducting terminals had been determined to be approximately 85 fF/micron².

Thus for a target bandwidth of 120 GHz, the required specific normal resistance $R_n A$ would need to be about 15–16 Ω micron². This specific resistance is related to the DC Josephson critical current density (at an infinitesimally small bias voltage) by the formula $I_C/A \approx (\pi/4) V_{\text{gap}}/(R_n A)$ [16]. SIS fabricators typically use I_C/A as a target parameter. The required receiver design RF bandwidth therefore demands a target I_C/A of 14–16 kA/cm², somewhat higher than that of devices used in earlier 230 GHz receivers.

SIS saturation and response nonlinearity

The usefulness of a heterodyne receiver for astronomy will be judged in large part by its ability to acquire reliable, accurate spectroscopic data over a large dynamic range of source and background brightness levels. Dark, transparent skies can offer a background brightness temperature of less than 30 Kelvin, whereas a telescope’s “hot load” used for instrument calibration approaches 300 Kelvin. To be most useful, the receiver’s IF output power is required to respond quite linearly to this range of RF source power levels. Unfortunately, without careful design, the SIS heterodyne mixer’s response linearity can be compromised in the face of such large variations in RF input power, especially in the case of a system with large RF and IF bandwidths.

Problems with SIS mixer linearity, also known as mixer *saturation*, are primarily the results of two effects: excessive SIS RF input RMS voltages (input saturation) and excessive SIS IF output RMS voltages (output saturation). Each of these sources of nonlinearity can be mitigated by limiting the corresponding instantaneous bandwidth (RF or IF) to which the SIS junction is exposed or by reducing the SIS normal resistance R_n so that a given RF or IF power level results in smaller SIS voltage excursions. Because this receiver is designed to handle large RF and IF bandwidths using a single SIS junction, the first option is unavailable to us. In this section we explore how a small SIS R_n may reduce saturation to tolerable levels.

First, consider input saturation due to excessive RF input levels. The heterodyne theory assumes that the RF input provides a very small perturbation to the LO signal incident on the SIS mixer. The resulting small signal admittance matrix describing the heterodyne performance then results from a first-order expansion about the quiescent SIS operating state established by the SIS DC bias and the LO. The actual RF power incident on the SIS mixer is determined by the source spectral power density integrated across the

receiver's instantaneous RF bandwidth. For a source brightness temperature of 300 Kelvin and an equivalent RF bandwidth of 120 GHz, the resulting total RF input power is $P_{\text{RF}} = k_B T \nu_{\text{BW}} \approx 0.5 \text{ nW}$. For optimum performance, the LO voltage amplitude will be given by $1 \approx \alpha = eV_{\text{LO}} / \hbar\omega_{\text{LO}}$, and the required LO power to achieve this amplitude will be $P_{\text{LO}} \approx V_{\text{LO}}^2 / 2R_n$. At 200 GHz with $R_n \approx 8 \Omega$, this gives $P_{\text{LO}} \approx 43 \text{ nW}$. Thus the incident RF power level is only just over 1% of the optimum LO power level. Since the ratio of the voltage amplitudes is given by the square root of the power ratio, the effective SIS pumping α in the presence of a 300 Kelvin RF source will increase by only about 0.6%, changing the mixer conversion efficiency, equation (2.4), by at most a few times 10^{-5} , a negligible effect. If insufficient LO power is available to achieve optimum pumping α , however, the situation can change dramatically. For example, with only half the required LO power available, incident RF power at 300 Kelvin can now change the mixer conversion efficiency by 0.5%, a potentially unacceptable level of nonlinearity. Thus to avoid mixer input saturation it will be important to ensure that sufficient LO power is available to optimally pump the SIS.

Now consider the potential problem of SIS mixer IF output saturation. Again assume a 300 Kelvin RF source. The simple, double-sideband SIS heterodyne mixer design will convert both the upper and lower sidebands containing this RF signal into the IF output (we ignore the contributions from higher frequency sidebands). With an equivalent SIS mixer IF bandwidth of approximately 15 GHz (not all of which can be handled by the LNA and subsequent electronics), a total of 30 GHz of RF signal could be down-converted to the SIS IF output. Using a calculation similar to that above, the total, integrated RF source power then converted to the IF by the SIS could be $\approx 0.1 \text{ nW}$. The SIS RF input impedance is $\sim R_n$, 8Ω , whereas the design impedance of the SIS and LNA combination seen by the IF output signal at the SIS is $\sim 3R_n$ (this design target is addressed in a subsequent section of this chapter). Assume that the total integrated SIS IF output power could be nearly the same as that present in the RF (i.e., that the mixer achieves an average conversion gain of unity). The RMS IF output voltage in this case would be $V_{\text{IF}} \approx \sqrt{3R_n P_{\text{IF}}} \approx 0.05 \text{ mV}$, or 5% of the 1 mV wide pumped I-V photon step for a 230 GHz LO.

To lowest order, the 230 GHz receiver's mixer conversion gain will be at a maximum for an SIS DC bias voltage V_0 near the center of the first photon step, and it then decreases quadratically to zero approaching the edges of the photon step. With a 300 Kelvin RF source, the SIS IF output will vary as Gaussian noise with the RMS voltage V_{IF} calculated above. Because the IF output frequency will be no higher than about 20 GHz, this IF output will affect the SIS mixer performance just as if the SIS operating state were adiabatically varied by continually adjusting its DC bias voltage away from V_0 in the same random manner. A Gaussian distribution of DC bias voltages with mean V_0 and standard deviation $V_{IF} \approx 5\%$ of the photon step width will produce a mean mixer conversion gain of only 99% of its value at V_0 . This reduction in conversion gain is potentially problematic. It is possible, however, that the assumptions going into this calculation are quite optimistic, so it can be hoped that output saturation when faced with a 300 Kelvin RF source will not actually reach this level. As will be seen at the end of this chapter (Figure 2-20 on page 52), the actual mixer conversion gain should be generally less than 50%, so the expected IF voltage excursions should again keep saturation at a tolerable level.

SIS RF embedding impedance requirements

The sensitivity of the SIS heterodyne receiver at any particular RF frequency ν (in Hz, not angular frequency) is characterized by its noise temperature $T_n(\nu)$:

$$T_n(\nu) \equiv \frac{1}{k_B} \left. \frac{dP_n}{d\nu} \right|_{\nu(\text{input-referred})} \quad (2.5)$$

where $dP_n/d\nu$ is the system's noise power spectral density (in, say, watts/Hz). The *definition* of noise temperature T_n as a measure of the input-referred $dP_n/d\nu$ is valid for any source of noise, and not only noise of thermal origin. By "input-referred," we mean that the total noise signal at the output of the system is interpreted as an added noise signal at the input of a noise-free (but otherwise equivalent) system which would reproduce the observed output noise. What this means in practice is that the system's output noise power spectral density is divided by the system's power gain at each frequency. See *Modeling internal sources and noise* in Appendix C.

For the SIS mixer followed by the IF LNA, the *RF embedding impedance*¹⁰ seen by the SIS device determines its contribution to the heterodyne detector's noise temperature, which, as mentioned at the beginning of this chapter, is due primarily to shot noise from the SIS DC bias current (when being pumped by the LO source) along with the thermal noise and ground state quantum fluctuations from the RF sideband channels. How much noise the LNA and following components then contribute to the total, input-referred noise temperature is determined by the SIS detector's heterodyne *conversion efficiency*, which is in turn determined by its RF responsivity $Y_{0,\pm 1}$ as well as the *IF embedding impedance* it sees. Here we address the SIS noise temperature contribution as a function of the RF embedding impedance; in the next section we consider the effects of the IF embedding impedance on its conversion efficiency.

The RF embedding impedance, in combination with the SIS DC bias voltage and LO pumping level, also determines the *output admittance* of the SIS that the IF circuit sees, or, equivalently, the SIS output reflection coefficient Γ_{SIS} in the IF frequency range. It is quite possible for the SIS mixer output admittance $Y_{0,0}$ at the IF frequency to acquire a negative real part, in which case $|\Gamma_{\text{SIS}}| > 1$, and oscillations can be induced in the IF and DC bias circuitry. If this happens, then instead of doing its job and amplifying the RF input signal from the telescope, the receiver will choose to “whistle a happy tune,” which will also, unfortunately, make the observer trying to use the instrument most unhappy! Clearly, the RF circuit design must avoid this impedance regime, which then sets strict limits on the range of RF embedding impedance which can be tolerated by the SIS device.

Using the receiver's SIS I-V curve shown in Figure 2-3, SuperMix was used to model the effects of varying RF embedding impedance on the SIS mixer's noise performance. A selection of typical results is provided by Figure 2-6 on page 33, showing mixer noise contours and IF output instability regions for three different LO frequencies. These

¹⁰ The expressions “embedding impedance,” “embedding admittance,” and “embedding reflection” will be used essentially interchangeably in this and the following sections. Although they have differing units, they all represent the same physical concept: the electrical characteristics of the mixer circuitry to which the SIS device is connected.

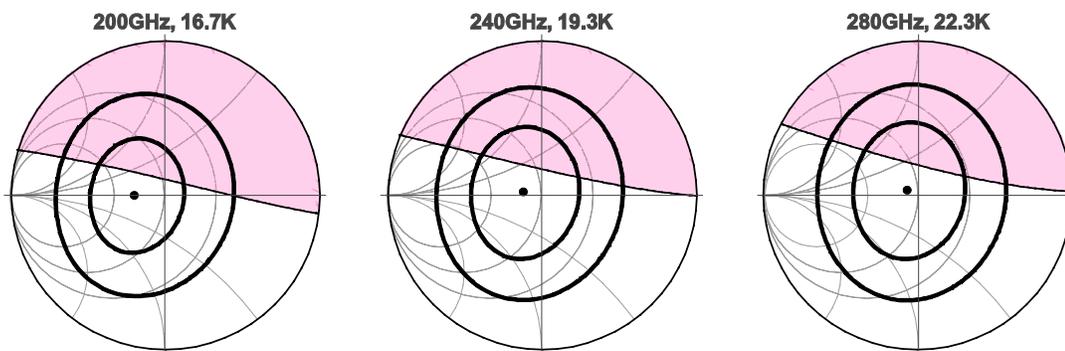


Figure 2-6: Effect of RF embedding reflection (Γ_{RF}) on SIS mixer noise and IF stability. Using the measured I-V curve shown in Figure 2-3, a SuperMix simulation assumed that the SIS was pumped at the frequencies shown with LO amplitude $\alpha = 1.0$ and that the SIS DC bias was at the center of the primary LO photon step below V_{gap} . The Smith charts' normalizing impedance is the SIS R_n , and they cover the full range of RF embedding $|\Gamma_{\text{RF}}| < 1$ and show calculated mixer noise temperature contours at 2 and 3 times $h\nu_{\text{LO}}/k_B$ (black ovals). The black dot is at the RF Γ calculated for minimum mixer T_n , whose calculated value is shown following the LO frequency. The shaded areas designate regions of IF instability, i.e. SIS IF output $|\Gamma_{\text{SIS}}| > 1$. These results are for the RF upper sideband at an 8 GHz IF.

simulation results are plotted on *Smith charts*, described in Appendix C. The Smith charts' noise and IF stability contours are plotted in the complex plane of RF embedding reflection coefficient as seen by the SIS, $\Gamma_{\text{RF}} = (1 - R_n Y_{\text{RF}})/(1 + R_n Y_{\text{RF}})$. The SIS normal resistance R_n serves as each chart's normalizing impedance, and it is represented by the origin at the center of each plot ($\Gamma_{\text{RF}} = 0$). The thin gray arcs in the charts show a few lines of constant real and constant imaginary parts of the RF embedding admittance, Y_{RF} , and illustrate the conformal map of the complex admittance plane onto the complex Γ plane. These model results are for the upper RF sideband (USB) and an IF frequency of 8 GHz. Results are quite similar for the LSB and for IF frequencies within an octave of this value (4–16 GHz). See Appendix D for a much more comprehensive set of model results (the SuperMix source program used to generate them is presented in Appendix I).

Note that Γ_{RF} for these charts is defined to include the SIS physical capacitance *as part of the RF circuit*; this parallel capacitive admittance adds to the admittance of the rest of the RF circuitry when calculating Y_{RF} and its associated Γ_{RF} . As can be seen from the charts, minimum mixer noise temperature is realized at a nearly real-valued RF

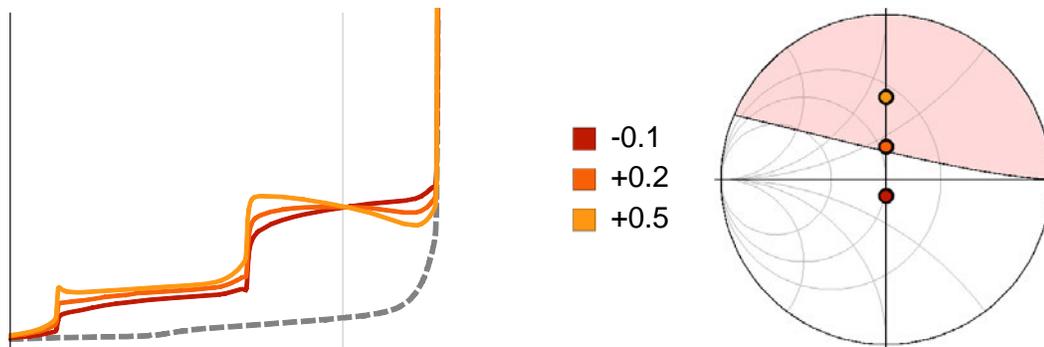


Figure 2-7: Effect of RF embedding reflection on the SIS pumped I-V curve shape. SIS pumped with a 240 GHz LO such that pumping amplitude $\alpha = 1.0$ when DC bias is at the center of the primary LO photon step below V_{gap} (vertical grid line in pumped I-V plots). The three choices of RF embedding reflection have $\text{Re}(\Gamma_{\text{RF}}) = 1$ and $\text{Im}(\Gamma_{\text{RF}})$ as shown, taking the SIS IF output impedance from stable to unstable (right-hand Smith chart). The slope of the pumped I-V curve at the DC bias point is negative for Γ_{RF} in the unstable region. These are modeled results using SuperMix.

embedding impedance slightly less than the SIS normal resistance. At 240 GHz, for example, the middle chart shows that the minimum mixer T_n is 19.3 K, 1.7 times the quantum limit, attained with $\Gamma_{\text{RF}} = -0.12$, or about $0.8R_n$. The slight imaginary part of the optimum Γ_{RF} evident in the chart compensates for the quantum susceptance of the SIS at the selected DC bias point (see Figure 2-5 on page 26).

The onset of IF output instability can be better appreciated by comparing SIS pumped I-V curves for RF embedding impedances outside and inside the stability region. Consider Figure 2-7, which shows the SIS pumped I-V behavior for a 240 GHz LO with power sufficient to achieve a pumping amplitude $\alpha = 1.0$ when the SIS DC bias is at the center of the primary LO photon step. At IF frequencies the real part of the SIS output admittance is well-approximated by the slope of the pumped I-V characteristic curve at the SIS DC bias voltage. Three choices for Γ_{RF} are plotted in the Smith chart in Figure 2-7, and their associated pumped I-V curves are plotted as well. Clearly, the unstable IF choice (in the shaded region) has a negatively-sloped pumped I-V curve at the DC bias point, whereas the stable choices for Γ_{RF} result in a positive I-V curve slope.

The origin of these various pumped I-V curve slopes can be understood by again considering the imaginary part of the SIS junction's quantum admittance plotted in

Figure 2-5. The imaginary part goes from capacitive (positive) at the low-voltage end of the photon step to inductive (negative) as the bias approaches V_{gap} . For RF embedding admittances with a capacitive reactance, this implies that the SIS junction's match to the LO source is worse where the SIS is also capacitive and better where it is inductive. The better the match, the more LO power absorbed and the larger the SIS pumped I-V current. Thus the SIS pumped I-V current increases as the DC bias point moves toward the SIS gap, resulting in a stable, positively-sloped I-V curve. The converse is true if the RF source has an inductive reactance, leading to the unstable, negatively-sloped I-V curve.

SIS IF embedding impedance requirements

In the previous section it was mentioned in passing that the SIS heterodyne detector's contribution to the mixer noise temperature T_n is not affected by the IF load impedance seen by the SIS. This is generally the case: the signal/noise ratio at the output of a circuit gain stage is unaffected by the input impedance of a subsequent stage, because the fraction of power coupled into the subsequent stage (determined by the stages' impedance match at the interface) is the same for both the signal and noise. The SIS T_n characterizes the SIS detector's input-referred noise power spectral density, and it determines the SIS detector's signal/noise ratio when compared to some "standard" RF input signal power spectral density. Since that ratio will be unaffected by the IF load impedance the SIS sees, the SIS T_n is likewise unaffected. The IF load impedance, however, determines the fraction of the SIS IF signal power coupled into the subsequent IF output stage, which then affects the conversion efficiency (conversion gain or loss) of the RF into the IF output stage. Noise added by the IF amplifiers will therefore have a larger impact on the receiver system signal/noise ratio if this conversion efficiency is poorer than it could be.

Another important consideration when designing the IF output matching circuitry is its effect on the input impedance the SIS+IF circuitry presents to the RF source. As was the case with the RF circuit's embedding impedance, a poor choice for the IF embedding impedance (the IF load impedance seen by the SIS) could result in unacceptable RF

reflection by the mixer, causing circuit oscillations at RF frequencies and establishing large amplitude, unstable standing waves in the telescope optics.

Again using SuperMix along with the receiver's SIS I-V curve shown in Figure 2-3, the effects of varying IF embedding reflection coefficient Γ_{IF} on the SIS mixer's conversion efficiency and RF stability may be investigated. A selection of typical results is shown in Figure 2-8, displaying mixer conversion efficiency contours and RF instability regions vs. LO and IF frequencies. The model assumed an RF embedding impedance of $0.8R_n$, near the impedance for optimum mixer noise temperature found in the previous section. The IF embedding Smith charts' normalizing impedance is chosen to be $3R_n$, and from the charts it is clear that this choice for the IF embedding impedance Z_{IF} should result in slight mixer conversion gain over most of the receiver's design RF

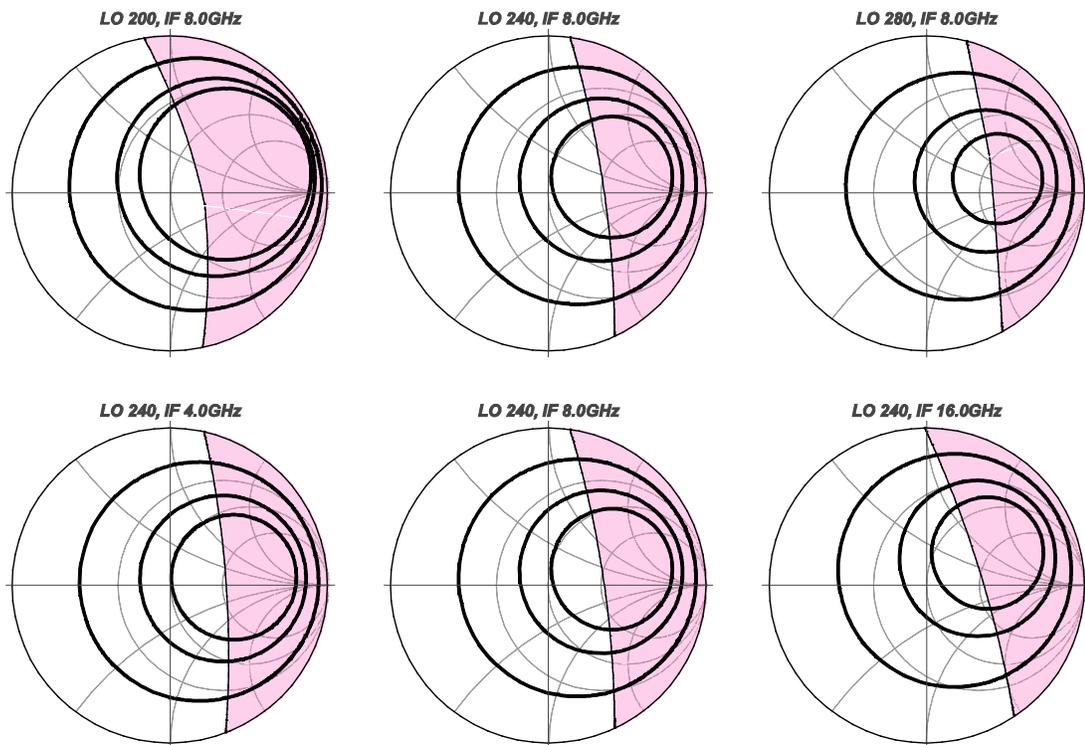


Figure 2-8: Effect of IF embedding reflection on mixer conversion gain and RF stability. SuperMix models assumed an LO amplitude of $\alpha = 1.0$ and DC bias at the center of the primary LO photon step below V_{gap} , and the assumed RF embedding impedance was $0.8R_n$. The Smith charts' normalizing impedance is $3R_n$, and they cover the full range of IF embedding $|\Gamma_{IF}| < 1$. Mixer RF USB gain contours of -3 , 0 , and $+1$ dB are shown. The shaded areas designate regions of potential RF instability, i.e. SIS mixer RF input $|\Gamma_{SIS}| > 1$ at either the upper or lower sideband.

and IF frequency ranges (the center contour is for 0 dB, or unity gain). Potential RF reflection gain (and instability) sets in at an IF embedding impedance only slightly above this value at the lowest LO frequencies, so the mixer chip designer must take this into account. Another consideration is the introduction of mixer output saturation as the mixer conversion efficiency is increased, as discussed in an earlier section of this chapter.

Low noise, cryogenic microwave amplifiers (LNAs) are nearly universally designed to have input impedances of $50\ \Omega$ within their design bandwidths; consequently they would provide a good match to an SIS with $R_n = 17\ \Omega$. The large bandwidth required by the receiver design, however, demands an SIS R_n of only about half this value. Moreover, microwave LNAs often have input impedances which diverge quite dramatically away from their nominal values at frequencies beyond their design bandwidths, usually becoming very nearly purely reactive. Directly connecting such an amplifier to the SIS can result in quite unpredictable and therefore unacceptable mixer behavior. The same is true for the DC biasing circuitry required for the SIS. The mixer chip design must therefore guard against possible negative effects of IF LNA and DC bias circuit impedance variations for all frequencies at which the SIS could respond, both IF and RF. It must also transform the LNA input impedance presented to the SIS at IF frequencies to an acceptable value ($\lesssim 3R_n$) to avoid RF instability or excessive conversion loss.

Mixer RF matching network circuit design

As discussed in an earlier section, the receiver RF bandwidth requirement drives the SIS device's $R_n C_s$ product to a fairly low value, even though its physical capacitance C_s then becomes rather large compared to more conventional receivers in this frequency range: $R_n \approx 8.5\ \Omega$ and $C_s \approx 140\ \text{fF}$ were the initial SIS design specifications (Appendix A on page 111 lists these specifications). Referring again to the Smith charts in Figure 2-6 on page 33, the RF circuit must tune out C_s and present a resulting resistance of slightly less than R_n over the receiver's RF design bandwidth of 200–300 GHz. The circuit design must also ensure that the embedding impedance it presents to the SIS mixer is kept under tight control in order to minimize mixer noise temperature while avoiding the IF instability region covering nearly all of the Γ_{RF} upper half-plane.

As will be discussed in Chapter 3, the waveguide probe feeding the RF and LO signals to the RF matching circuit has a nominal impedance of $Z_{\text{probe}} = 36.8 \Omega$ (see especially Figure 3-2 on page 56). The RF circuitry must efficiently transform this impedance in order to match the SIS requirements while also remaining compact and simple in order to reduce its capacitive load on the IF output circuit and its losses at RF frequencies.

Importantly, the RF design must also be tolerant of typical UV contact lithography alignment and etching errors as well as manufacturing tolerances around the SIS target current density. The mixer chip's thin film microstrip circuitry design was limited to a minimum width of 2.7 microns so that lithography errors would have minimal effect on the circuit's performance. With superconducting niobium wiring and ground plane and SiO dielectric, this specification limited the characteristic impedances of the microstrip transmission lines to a maximum of about 18Ω (design film thicknesses are also specified in Appendix A).

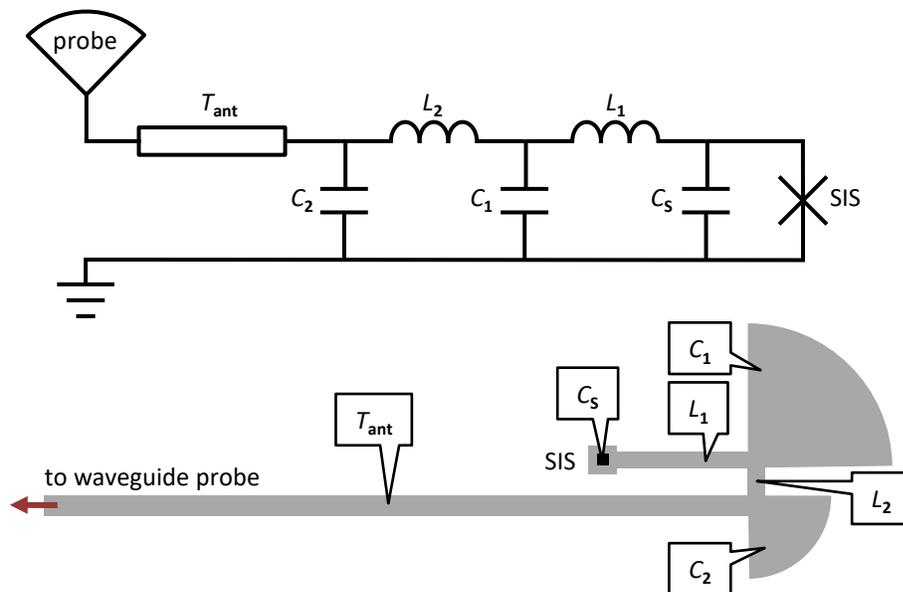


Figure 2-9: RF matching network design concept and circuit implementation.

The schematic of the RF conceptual design circuit topology was implemented on the mixer chip with a thoroughly optimized, thin-film, superconducting microstrip circuit layout. The compact design ($130 \text{ microns} \times 40 \text{ microns}$) effectively tunes out the SIS physical capacitance (C_s) while matching the probe impedance to the required SIS embedding impedance. The IF matching network connection to C_1 is not shown.

Straightforward analysis of the SIS RF embedding impedance requirements demonstrates that a simple parallel or series inductance tuning element to resonate out the SIS C_s is inadequate to achieve the broadband matching performance required. On the other hand, a properly-designed, lumped-element, *C-L-C-L-C ladder network* could indeed meet the bandwidth goal (the ladder's final C would be the SIS C_s). An additional quarter-wave transmission line impedance transformer stage would then be added to lower the probe impedance to that required at the input to the ladder network (Figure 2-9 on page 38).

Although conceptually the RF circuit joining the waveguide probe to the SIS device started as a transmission line transformer followed by a lumped-element ladder, the analyses and optimizations were performed incrementally using a series of thin film, high frequency circuit models implemented with the SuperMix software package (Chapter 5 and Appendix I). Lumped-element capacitors are replaced in these circuits by 90 degree

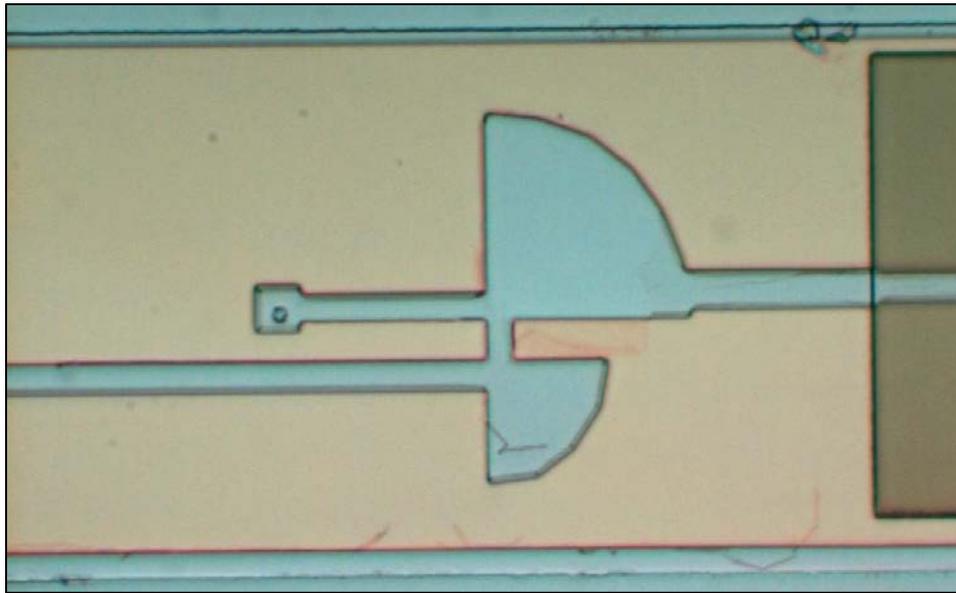


Figure 2-10: Photo of the final RF matching network as manufactured by JPL.

This false-color photo is of a mixer chip from the first generation manufacturing run showing the RF matching network wiring and SIS. Also shown is how the SIS IF output is extracted using the microstrip extending to the right from the larger radial stub. The first transition to CPW transmission line in the IF network is visible at the far right in the photo and is discussed in the next section. The SIS device is also visible as the small oval inside and below the center of the square pad connecting it to the microstrip wiring layer.

radial stub elements whose radii are small compared to a signal wavelength. Similarly, series inductance elements may be emulated with short, high impedance transmission line segments. Figure 2-9 on page 38 illustrates the relationship between the initial conceptual design and the microstrip implementation of the RF circuit; Figure 2-10 on page 39 then shows a photo of the resulting mixer chip RF network and SIS as manufactured by JPL.

The final circuit design optimization used a complete, detailed model of the thin-film, superconducting microstrip circuitry and SIS device. The effects of stray capacitance and inductance at the microstrip circuit junctions and corners were analyzed using *Ansoft* (now ANSYS) HFSS 3-D finite-element electromagnetic simulations and were included as elements in the SuperMix model. Parasitic perturbations introduced by the IF matching circuitry (discussed in the next section) were also included in the final optimizations. Design dimensions and detailed layout of the RF circuit elements are provided in Appendix A.

As explained above, a conservative design for the matching network ensured that it would remain stable and work reasonably well even in the face of possible SIS and mixer chip manufacturing tolerances. For example, the actual mixer chip selected for the receiver had a measured SIS R_n of 7.67 ohm rather than the 8.5 ohm design target, and slight mask alignment errors are also indicated by the offset position of the SIS in the chip photo (Figure 2-10). The predicted performance of the RF matching network design is illustrated in Figure 2-11 and Figure 2-12. As should be clear from a comparison of these figures to the embedding impedance study charts in Figure 2-6 on page 33 and in Appendix D, the conservative matching network design ensured that the SIS IF output would remain stable over a wide range of SIS and other mixer chip parameter variations.

In particular, the design turned out to be also reasonably tolerant of errors in the assumed value of the SiO dielectric constant used to characterize the microstrip and radial stub electrical properties. As the author was designing the mixer chip circuitry, a preliminary analysis of an experiment meant to measure the electrical properties of thin film, superconducting microstrip lines at millimeter wave frequencies gave a value of 7.5 for the SiO dielectric constant, significantly different from its conventionally accepted value of 5.6. Perhaps overly trusting of these early findings, the author used the 7.5 value when optimizing his mixer chip circuit design. A more thorough analysis of that

experiment's results revised the dielectric constant value down to 6.2, unfortunately too late for the author's design, which had already been submitted to JPL for manufacture. Even more troubling, perhaps, was the SiO value found from a SuperMix model analysis of another, quite complicated SIS mixer design (described in the first section of Chapter 5, SIS circuit modeling and SuperMix). This independent analysis yielded a value of 5.7, very close to the conventional value of 5.6. This result was obtained in mid-1999 [23], but was unfortunately forgotten while designing the mixer chip a year or two later.

Figure 2-11 shows that even though an erroneously high choice of 7.5 for the assumed SiO dielectric constant was used in the design, the actual chip behavior even for an SiO dielectric constant of 5.6 and using the actually delivered SIS with its smaller R_n remains quite close to the design prediction, except perhaps at the lowest RF frequencies (below 190 GHz). Figure 2-12 on page 42 shows that the tolerance of the RF design to errors in the assumed value of the SIS physical capacitance is also quite good, again except for the lowest RF frequencies in the case of a quite small value assumed for the SIS capacitance.

An unfortunate but unavoidable consequence of the conservative RF matching design is an increase in the mixer noise temperature compared to the best obtainable (according to the embedding impedance study). In fact, the model predicts that with this design the

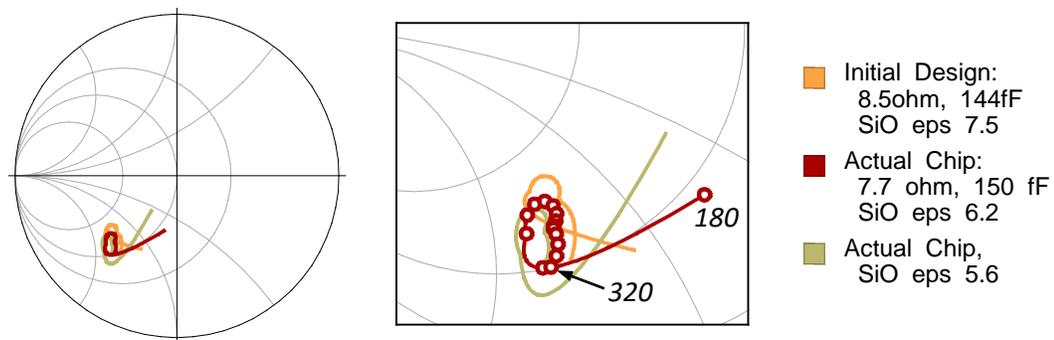


Figure 2-11: Modeled SIS RF embedding impedance performance of the mixer chip. A SuperMix model of the final mixer chip design (including the waveguide probe and IF matching network) predicted the embedding impedance presented to the SIS. The Smith chart normalizing impedance is the SIS R_n , as were the charts in Figure 2-6 on page 33. The models shown are for the design SIS specification and for the actual mixer chip installed in the receiver (whose DC I-V characteristic is shown in Figure 2-3 on page 22). The circles indicate RF frequencies from 180 GHz to 320 GHz in 10 GHz increments.

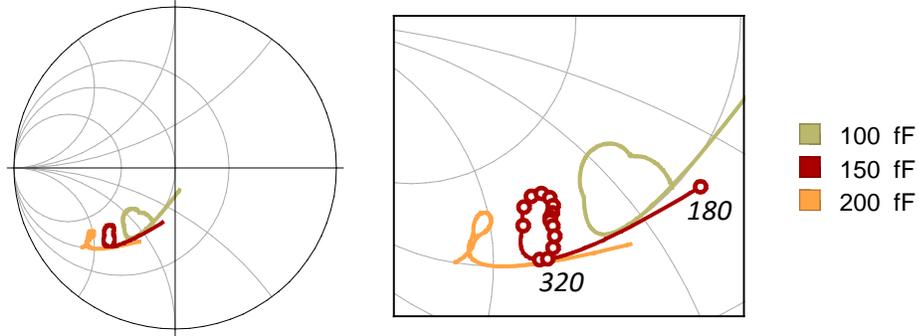


Figure 2-12: Expected RF embedding impedance variation with SIS capacitance. RF embedding impedance as used here includes the SIS device physical capacitance C_S as part of the embedding network. Shown are SuperMix predicted RF embedding impedances presented to the SIS as the SIS physical capacitance is varied away from its expected value. The Smith chart normalizing impedance is the measured 7.7 ohm SIS R_n , and the model assumed a SiO dielectric constant of 6.2, to allow for a direct comparison to the red curve in Figure 2-11. The circles on the nominal model result (150fF) indicate RF frequencies from 180 GHz to 320 GHz in 10 GHz increments.

expected noise temperature contribution of the SIS mixer should generally be between 2 and 3 times the quantum noise limit over the receiver's bandwidth, an increase of 10 to 15 K over the minimum attainable according to the RF embedding study (the quantum noise temperature limit is 10 K at 210 GHz).

RF choke and IF matching circuit design

The center area of the mixer chip is dominated by a relatively large circuit structure consisting of a series of alternating co-planar waveguide (CPW) and thin-film microstrip transmission lines (refer again to Figure 2-1 on page 19). This circuit serves two purposes: (1) at RF frequencies it forms a series of alternating high and low impedance, $\frac{1}{4}$ -wave transmission line segments which isolates incoming RF and LO signals from the external IF and SIS biasing circuitry; and (2) at the receiver's relatively low IF frequencies, within a range of 2–18 GHz, it acts as a multi-stage L-C cascade which transforms the mixer's external IF load (the LNA input) to the proper embedding impedance needed to optimize the SIS performance. In short, this structure serves as both the mixer's *RF choke* and its *IF matching network*. As mentioned earlier in this chapter,

combining these two functions into a single, thin-film network integrated with the SIS mixer chip has been an important innovation introduced with this design.

The circuit's RF choke function will be addressed first. Consider a $\lambda/4$ length of transmission line terminated by a load impedance Z_L . The total, round-trip phase length of a signal traversing the line to and from the termination Z_L would then be 180° , so a load reflection Γ would transform to a reflection of $(-\Gamma)$ when exiting the other end of the line. If the transmission line's characteristic impedance is Z_C , then from equation (C.4) on page 130 in Appendix C, the impedance Z presented by the combination of the $\lambda/4$ -wave transmission line and its termination would be given by $Z/Z_C = Z_C/Z_L$. Consequently, a series of two $\lambda/4$ -wave sections with characteristic impedances Z_1 , and Z_2 (Z_1 closest to the load Z_L) will then transform the load impedance to $Z = (Z_2/Z_1)^2 Z_L$.

The RF choke circuit, Figure 2-13, is a series of four such transmission line pairs, in each case with $Z_2 > Z_1$. The four identical high impedance (Z_2) sections share a common characteristic impedance of 103Ω , the largest practical impedance attainable using CPW transmission line on a silicon substrate given the mixer chip's dimensions and the chip manufacturing tolerances (exact dimensions of the design are available in Appendix A). The design characteristic impedances of the low impedance sections are shown in the figure; all were chosen to optimize the structure's IF matching performance, as discussed later in this section. To simplify the design process, the lengths of all eight sections are such that each has a 90° phase length at a single, common RF frequency of 287 GHz, a

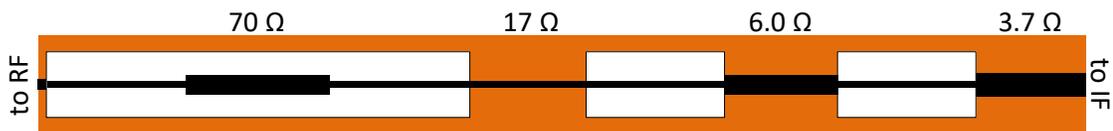


Figure 2-13: The RF choke and IF matching network structure.

The circuit consists of a total of eight $\lambda/4$ transmission line sections with alternating high and low characteristic impedances. The high impedance, co-planar waveguide (CPW) sections have a common characteristic impedance of 103Ω , whereas the low impedance sections (all but one are thin-film microstrip lines) have the characteristic impedances shown in the figure. The horizontal black strip represents the signal wiring conductor, which varies in width from section to section. The ground plane is shown by the orange area, which extends under the microstrip wiring. The white areas are breaks in the ground plane which form the CPW sections. The design frequency determining the lengths of the sections is 287 GHz.

frequency selected during design optimization of the circuit.

At its design frequency, this cascade of $\frac{1}{4}$ -wave sections will multiply the IF load impedance attached at its right end by a factor of over 18 million for RF signals present at its left end, whereas it is attached to the RF circuitry at a particularly low-impedance point: the radial stub C_1 (Figure 2-9 and Figure 2-10). The real test of the choke's effectiveness is, however, how well it isolates RF and LO signals from the mixer's IF output. Varying, unpredictable reflections of such high frequency signals from the IF and DC bias components could wreak havoc with the effectiveness of the RF matching to the SIS (recall, however, that the optimized choke circuit itself was also included when calculating the mixer circuit's RF embedding impedance in the previous section). Given this criterion, the RF choke design should be quite effective, as implied not only by the previous section's Figure 2-11, but also as shown explicitly in Figure 2-14. The effects of RF and LO signal reflections from the IF circuitry downstream of the choke structure should be completely negligible throughout the receiver's design RF bandwidth.

This observation is reinforced by considering the effect of a perfect reflection by the IF circuitry on the RF embedding impedance seen by the SIS. Figure 2-15 on page 45 shows the magnitudes of the Γ errors (in dB) which would be introduced to the SIS RF embedding impedance (Figure 2-11) by a 100% RF signal reflection at the RF choke

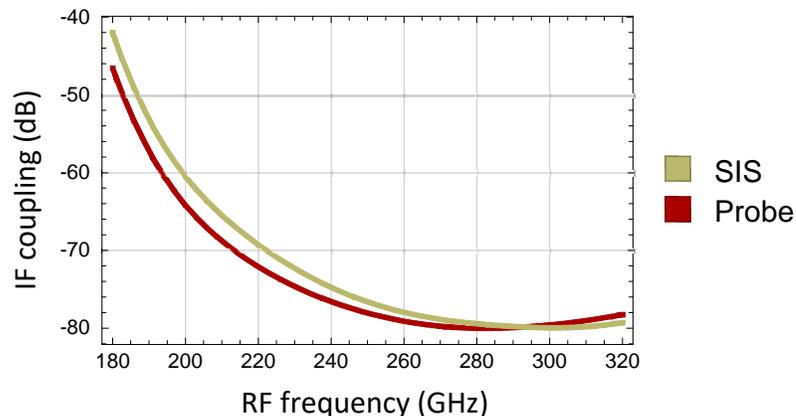


Figure 2-14: Predicted coupling of RF signals to the IF output of the choke circuit. SuperMix model results showing the predicted coupling attenuation of RF signals from the waveguide probe and the SIS upon reaching the IF output of the RF choke structure. The model assumed the measured 7.7 ohm SIS R_n , the modeled antenna probe impedance, SiO dielectric constant of 6.2, and a 50 ohm IF load.

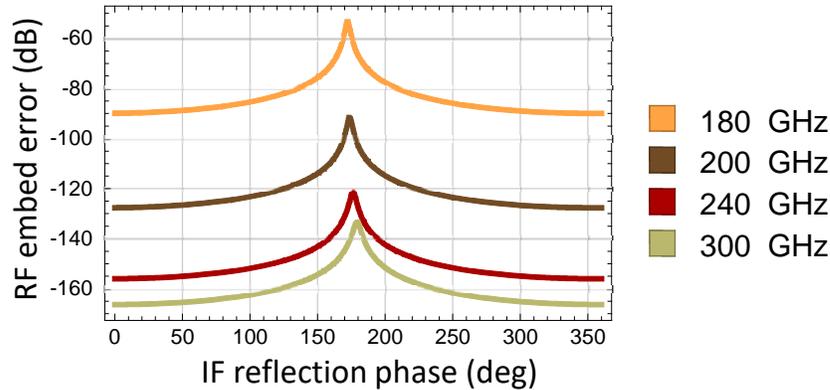


Figure 2-15: Variations in the SIS RF embedding gamma caused by IF reflections. Plots of SuperMix model results showing the magnitude of the predicted errors in the SIS RF embedding gamma (cf. Figure 2-11 on page 41) introduced by a perfect reflector at the RF choke's IF output as a function of the phase of the reflection and for various RF frequencies. The model assumed the measured 7.7 ohm SIS R_n , the modeled antenna probe impedance, SiO dielectric constant of 6.2, and it used a 50 ohm IF load as the reference for the error magnitudes introduced by 100% reflection at the IF output.

output. The plots show the error as a function of the reflection phase for various RF frequencies within the receiver's design RF bandwidth. Even for the worst case, an IF load which at 180 GHz is very nearly a short circuit (172° phase on reflection), the effect on the SIS RF embedding gamma is a change of only -53 dB, or about 0.002 in error amplitude; Figure 2-15 shows that the error drops rapidly as the reflected phase moves away from that worst-case angle. For RF frequencies of 200 GHz and above, the worst-case reflection introduces maximum errors which are more than 10 times smaller.

Now consider the second critical design function of the transmission line structure shown in Figure 2-13: IF output load matching to the SIS. It must transform the input impedance of the IF circuitry connected at its output in such a way that the SIS device sees an IF embedding impedance which provides adequate mixer conversion efficiency while keeping the SIS mixer stable at RF frequencies (refer again to the Smith charts in Figure 2-8 on page 36). At these frequencies (less than 20 GHz), the lengths of the transmission line sections in the IF matching + RF choke network are very much shorter than $\lambda/4$; consequently, they behave somewhat like lumped elements, the CPW (high-impedance) lines like series inductors and the microstrip (low-impedance) lines like

parallel capacitors.¹¹ Including the SIS physical capacitance and the effective capacitance of the RF matching network at IF frequencies, this arrangement then emulates a C - L - C - L - C ladder circuit which can transform the nominally 50 ohm IF circuit load impedance to the roughly 2 to $3R_n$ (15 – 24 ohm) range needed for optimal SIS matching.

In addition, although the original design optimizations of the IF matching circuit did not include its effects, the mixer chip also contains a “bias tee” circuit with wire bond pads for its DC bias and IF output connections, as shown in Figure 2-16 and the photo in Figure 2-17 on page 47. These structures can be accurately modeled as lumped elements because they are quite small compared to the IF wavelengths ($\lambda = 20$ mm at 15 GHz).

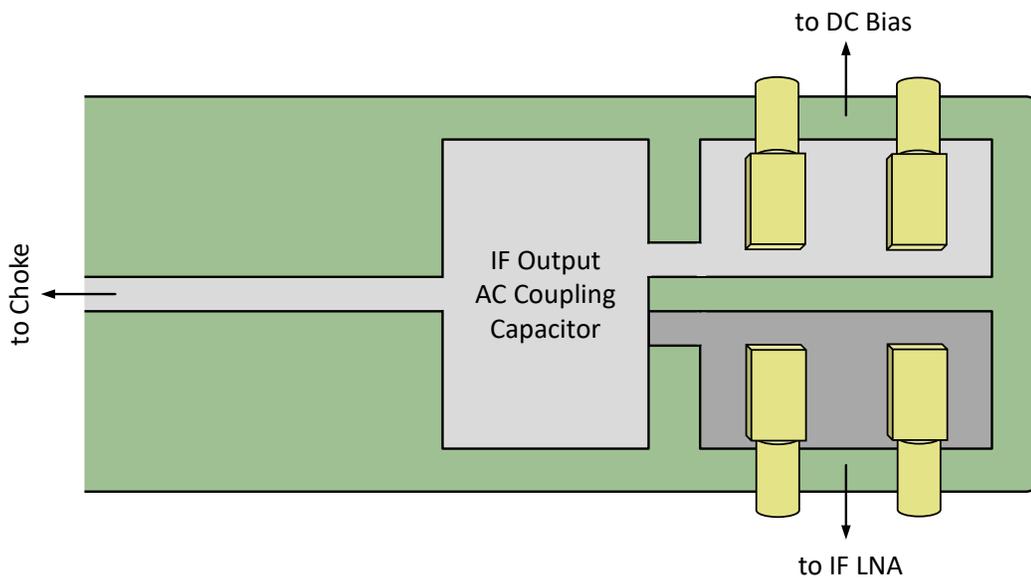


Figure 2-16: Mixer chip bias tee and output wire bond connections.

Lumped-element models of the components in this structure (Figure 2-17 on page 47) are included in the mixer IF matching calculations shown later (Figure 2-18 on page 48). Also included are models of the 1 mil diameter bond wires (shown here in gold) and the bond pad on the mixer DC bias interface board (not shown). The 25 micron thick silicon substrate (shown in green) rests on the mixer block surface, which acts as the bias tee circuit ground. The substrate width (vertical here) is 0.23 mm. A $50\ \Omega$ microstrip line connects the RF choke to the bias tee.

¹¹ A quick estimate of the equivalent inductance or capacitance of a transmission line with characteristic impedance Z_0 and which happens to be $\frac{1}{4}$ -wavelength long at frequency ν (in Hz) is $L \approx Z_0/(4\nu)$, $C \approx 1/(4Z_0\nu)$.

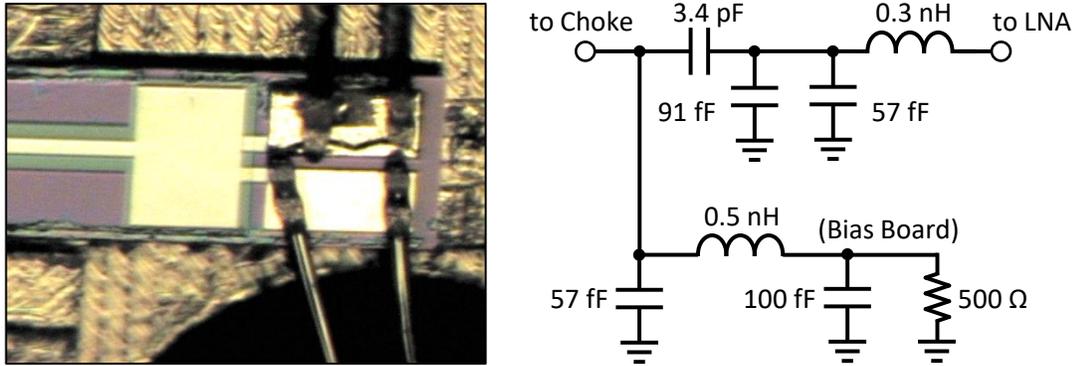


Figure 2-17: Mixer chip bias tee photo and equivalent circuit model at IF frequencies. The right-hand graphic shows the lumped-element SuperMix model of the bias tee used for calculations involving the IF output circuitry. The IF output AC coupling capacitor (large yellow rectangle in the photo) uses the mixer chip's very thin SiO film as its dielectric and therefore has a relatively large capacitance value: 3.4 pF (assuming the SiO dielectric constant is 6.2). Its bottom conductor (connected to the IF output bond pad, 57 fF) also forms an additional 91 fF capacitor with the mixer block surface (at ground potential) through the chip's silicon dielectric. The inductances of the wire bond connections and the bias board capacitance (not shown in the photo) were estimated from their geometries. A 50 Ω microstrip connects the RF choke output to the coupling capacitor.

Figure 2-17 shows the resulting circuit. The capacitor values were calculated from the component dimensions and the dielectric constants using a simple, approximate formula: $C = \epsilon A/d$, where ϵ is the dielectric's permittivity, A is the plate area, and d is the dielectric thickness. Pairs of 1 mil gold wire bonds were used for both the LNA and DC bias connections to reduce their combined inductance. The resulting inductances used were estimated to be equivalent to 0.4 nH per millimeter of bond length, which corresponds to a characteristic impedance of approximately 120 Ω for the connections.

With the entire mixer chip circuitry model available, the SIS IF embedding impedance as a function of IF frequency may be calculated and compared to the results of the embedding impedance study described previously: the model results are presented in Figure 2-18 on page 48. As mentioned earlier, the original design calculations and optimizations unfortunately neglected to include models of the bias tee circuit and wire bonds. Note, however, that their inclusion makes significant changes to the model results

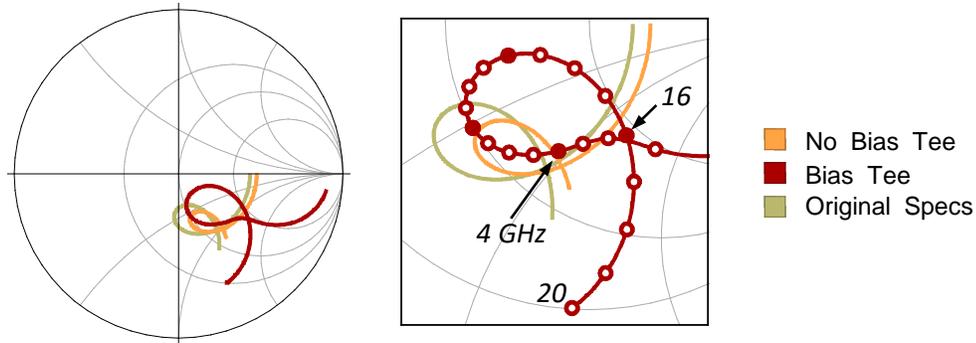


Figure 2-18: Modeled SIS IF embedding impedance variation with IF frequency. The curves show the SIS embedding gamma for IF frequencies from 0.1 GHz to 20 GHz. The Smith chart normalizing impedance is 3 times the SIS R_n , as in Figure 2-8 on page 36. The orange vs. red plots show the effect of the bias tee circuitry on the embedding impedance. The green plot shows the original design analysis and optimization result, which did not include the bias tee circuitry and used different physical parameters for the SIS and for the assumed SiO dielectric constant (see Figure 2-11 on page 41 for details). The red circles on the close-up figure detail are at 1 GHz intervals from 1 GHz to 20 GHz. The filled circles are at 4, 8, 12, and 16 GHz.

at frequencies below a gigahertz. This effect is caused primarily by the bias tee's AC coupling capacitor, which becomes an open circuit at low frequencies.¹²

A more serious problem may be introduced by the bias tee: a series LC resonance resulting from the combination of the DC bias bond wires and the DC bias wire bond pads. These components act as a short circuit for signals near their resonant frequency (refer back to the circuit diagram in Figure 2-17). It is therefore important for this resonant frequency to be outside the range of the required IF frequency band. Unfortunately, in the first generation receiver taken to the CSO this resonance was close to 8.5 GHz, very much within the specified IF frequency range (ouch!). Once this cause of the IF frequency response problem was correctly identified, the DC bias board bond pad area was reduced and the bond wires doubled up and shortened, which moved the resonance out of the IF band.

¹² The corner frequency of the high-pass filter formed from the 3.4 pF AC coupling capacitor and the LNA's assumed 50 Ω input impedance is $(2\pi RC)^{-1} = 0.94$ GHz.

In any event, comparing these IF embedding results to the analysis presented in Figure 2-8 on page 36 seems to indicate that the circuit might provide SIS mixer conversion efficiencies approaching 0 dB (minimal conversion loss). We will find in the next section, however, that this estimate will turn out to be optimistic. These results also indicate that the IF embedding impedance (especially for the full model including the bias tee) also seems to flirt with the RF instability region of the Smith chart near the IF bandwidth limits, and it flagrantly violates that boundary for frequencies well below the specified IF band (below 1 or 2 GHz). SIS RF reflection gain at frequencies very near the LO frequency did not appear to have any noticeable effects on the instrument's performance on the telescope, however, so bullet dodged, possibly because of losses in the receiver's RF optics, the topic of Chapter 4.

Mixer chip performance model results

The predicted performance of the mixer chip design presented in this chapter is summarized in the plots presented here. These results were generated using a SuperMix model of the complete mixer chip design along with the measured noise performance of the receiver's cryogenic LNA graphed in Fig. B-10 on page 125. The modeled noise and mixer conversion performance calculations refer to the RF and LO signals present at the entrance of the mixer block's rectangular waveguide (just upstream of the waveguide tuning step shown in Figure 3-1 on page 55) and do not include estimates of any waveguide losses. Additional losses introduced by the RF optics as well as the noise added by the LO source are also left out of these calculations. These far from insignificant contributors to the receiver's noise level are addressed in Chapter 4.

SuperMix was first used to determine the SIS DC bias voltage and LO power level required to minimize the modeled circuit's *double-sideband noise temperature* at a single, representative IF frequency (see *Y-factor and SSB vs. DSB noise in heterodyne receivers* on page 141 of Appendix C). The chosen IF frequency for these calculations was 6 GHz, the center of the 4–8 GHz IF band commonly used for CSO heterodyne astronomical observations. A representative selection of the results of this analysis is shown in Figure 2-19 on page 50. This figure's array of contour plots demonstrates how the mixer's noise performance is expected to vary with operating state over a range of LO

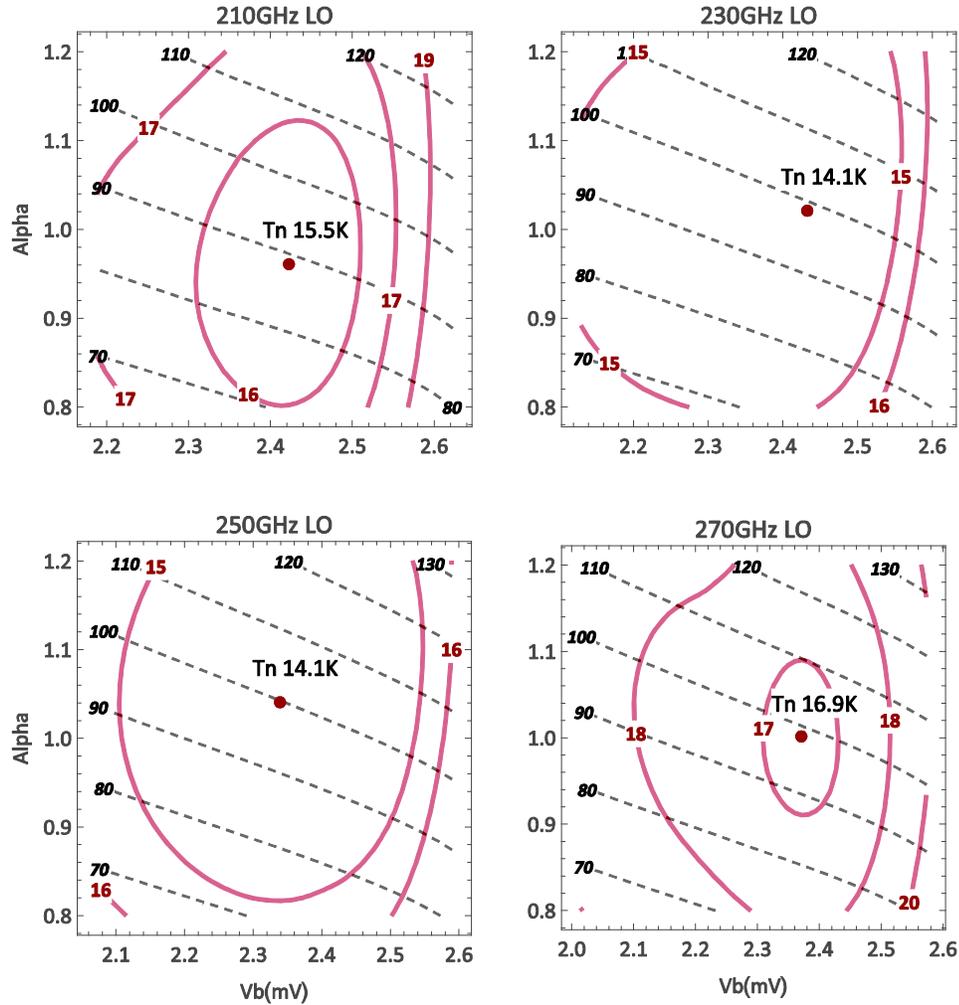


Figure 2-19: Modeled mixer chip + LNA DSB noise performance vs. SIS operating state. These graphs are contour plots (red curves) of predicted double-sideband noise temperature as a function of SIS DC bias voltage (x -axes) and LO pumping α (y -axes) for various LO frequencies (see the *Glossary* for a definition of α). Noise temperatures were calculated at an IF frequency of 6 GHz. The optimum SIS operating state is indicated by a red dot, and its corresponding T_n is indicated. The dashed lines denote various SIS LO-pumped DC bias current levels corresponding to the SIS operating states (in microamps). The assumed SiO dielectric constant is 5.6.

frequencies. The DSB quantum noise limit at, for example, 230 GHz is 5.5 K (one half of the SSB quantum noise limit $h\nu/k_B$), so the mixer + LNA are predicted to have noise levels of generally about 3 times the quantum limit (5.5 K DSB at 230 GHz).

The model results presented here assume a mixer chip SiO dielectric constant with a more conventional value of 5.6, rather than the design value of 7.5 or even the 6.2 value

presented earlier in this chapter. Predicted performance graphs using values of 6.2 and 7.5 are available in Appendix D. As expected, a choice of $\epsilon_{\text{SiO}}=5.6$ results in worse predicted noise performance at the lower LO frequencies (by 2 K or more DSB compared to using 6.2). In retrospect, however, $\epsilon_{\text{SiO}}=5.6$ is probably the most realistic value to use.

Note from the graphs that the optimal SIS operating states are predicted to be close to a DC bias voltage of 2.4 mV and with LO signal amplitudes at the SIS that are approximately equal to their photon voltages: $\alpha=eV_{\text{LO}}/h\nu_{\text{LO}}\approx 1$, as was expected from the application of Tucker's theory discussed early in this chapter. From the graphs, these LO power levels will result in SIS DC bias currents of approximately 100 μA (refer back to Figure 2-4 on page 24 for a typical SIS pumped DC I-V curve). Note that the optimized SIS DC bias voltages average only about 40% of an LO photon step width away from the SIS V_{gap} of 2.8 mV, rather close to the SIS gap voltage. In any event, the graphs show that the SIS operating state optima are rather broad, so, according to the models, setting the SIS bias to 2.4 mV while providing sufficient LO power to achieve a 100 μA SIS bias current should be nearly optimal at all LO frequencies. This was indeed found to be the case when operating the completed receiver, where optimal bias voltages were between 2.3 and 2.4 mV, and optimal pumped I-V currents were generally close to 100 μA .

The final plots for this chapter, shown in Figure 2-20 on page 52, are of the expected mixer + LNA upper and lower sideband noise temperatures and expected conversion efficiencies for these optimized SIS conditions, again assuming a SiO ϵ of 5.6. The predicted single-sideband noise temperatures are less than three times the quantum limit except at the lowest RF frequencies (below about 215 GHz for the lower RF sideband). The mixer conversion efficiency predicted by the model is lower than one would like, generally between -3 and -5 dB. As a consequence, the LNA's contribution to the predicted SSB noise temperature is 10–20 K, about equal that of the SIS. Finally, note that the model indicates that for most of the RF frequency range (i.e. below 260 GHz) observations should use the lower RF sideband. Chapter 4 will add the expected RF optical performance and LO source noise estimates to the model and will then compare the complete receiver model predictions to the available laboratory and observatory measurements.

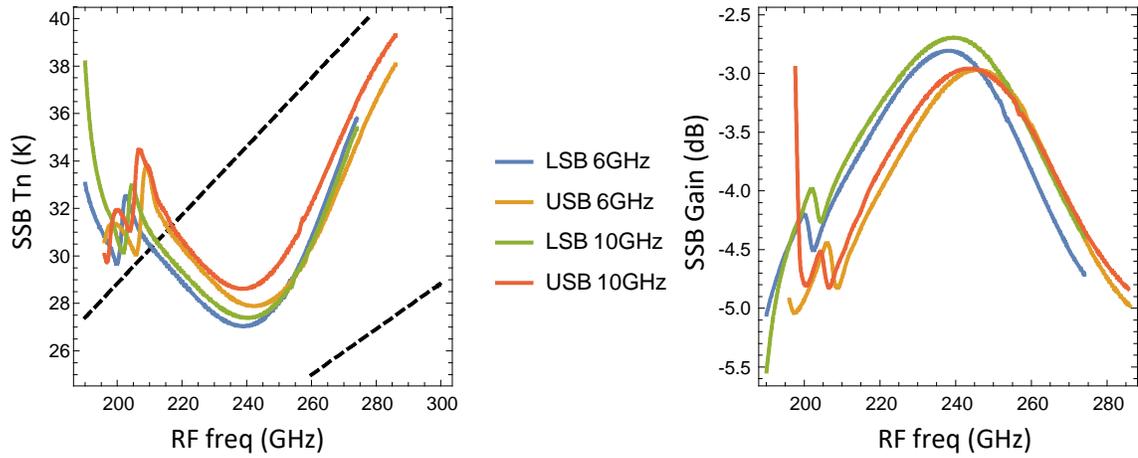


Figure 2-20: Modeled mixer chip + LNA sideband performance vs. RF and IF frequency. SuperMix-generated model results for an assumed SiO dielectric constant of 5.6 and at near-optimal SIS DC biasing and LO power. Upper and lower sideband results are plotted for IF frequencies of 6 and 10 GHz. The RF frequencies are the actual sideband frequencies, $\nu_{LO} \pm \nu_{IF}$. The dashed lines in the noise performance graph show $2\times$ and $3\times$ the quantum limit noise temperature (which is 10 K at 208 GHz).

Chapter 3

WAVEGUIDE PROBE AND MIXER BLOCK

Waveguide and waveguide probe

Through initial development and testing of the SuperMix software suite (Chapter 5) the author had, by mid-1998, become quite familiar with Caltech's quasi-optical, twin-slot designs for submillimeter SIS heterodyne mixers. In the 180–300 GHz range, however, such mixers would not offer bandwidth or noise performance comparable to that of a more traditional, horn-fed waveguide approach. The then-current SIS receivers at CSO used rectangular waveguide mixers, albeit with adjustable waveguide elements which required manual tuning whenever the desired RF observing frequency was changed by more than a few percent. An important requirement for this new receiver was to eliminate the need for such adjustments: it was to be a *fixed-tuned* design over that frequency range, a feature which is now, two decades later, commonplace.

To achieve a 5:3 frequency range in a fixed-tuned design would require a new approach to coupling the RF signal from the waveguide to the SIS mixer: not only would the SIS mixer chip need much more sophisticated, optimized RF circuitry (Chapter 2), that circuitry would also need to be very efficiently coupled to the incident RF signal from the waveguide. To accomplish this latter requirement both the rectangular waveguide and the mixer's waveguide probe geometries would need careful, coordinated design. A hint as to where to start was provided in a paragraph in a 1999 conference paper by Withington et al. [24]: in it they briefly described the remarkable performance they had measured of a waveguide probe constructed with a 90° radial-sector geometry and radius of about 0.4 of its waveguide height.

Design work on the waveguide and probe design then began in mid-2000 aided by using *Ansoft* (since acquired by ANSYS®) *HFSS*, a 3-D, finite-element electromagnetic field (EM) modeling software suite. Starting from a geometry scaled from the Withington et al. waveguide probe structure, the author constructed an HFSS model of a rectangular waveguide containing a radial waveguide probe on a silicon substrate (the plane of the radial probe was coincident with the waveguide centerline plane formed from the RF

signal propagation direction and the waveguide height, as will be shown in Figure 3-1 on page 55). Using the desktop computers available at that time, accurate HFSS EM field calculations of a complicated structure unfortunately could take several hours. At that time HFSS also did not include a capability to optimize EM structure designs, and its ability to adaptively mesh a complicated 3-D spatial model was fairly primitive. After several weeks, the author finally discovered how to manually divide the HFSS model structural elements so that it could more quickly generate reliable, accurate results. Unfortunately, an optimized EM structure geometry for the waveguide + probe would still need to be found manually or estimated using some other program such as SuperMix.

The author modeled the radial probe as located in the center of a tee-shaped structure consisting of a long, rectangular RF waveguide section whose midpoint was joined to one end of a long channel extending from the waveguide wall. This channel then held a thin-film microstrip on a silicon substrate connected to the waveguide probe. The lengths of these sections were chosen to ensure that any evanescent fields emanating from the waveguide-probe interface would have very small amplitudes at the RF ports of the structure. The resulting RF waveguide and microstrip propagation characteristics and coupling of various EM modes by this basic probe geometry were calculated using HFSS and its scattering matrix was “de-embedded” to an equivalent matrix connecting the modes at the location of the probe. This result was then used to create an equivalent SuperMix circuit model.

After a few weeks of exploring various effects on the probe performance introduced by changes in its radius, the waveguide height, waveguide back-short positioning, and other geometry changes, the author realized that a simple probe configuration probably could not achieve the performance required by the receiver’s design goals. Goutam Chattopadhyay, his office mate and a SuperMix collaborator, happened to be working on designs for Schottky-diode terahertz frequency multipliers for the Herschel Space Observatory. Discussions with Goutam led the author to realize that adding a fixed tuning element to the waveguide just upstream of the probe (toward the signal source) might improve its RF coupling performance. Initial HFSS model trials demonstrated that this turned out to be the case: adding a small, easily implemented reduction in the height of

the RF waveguide over a short distance just upstream of the probe could result in a dramatic improvement in the waveguide probe performance.

Following this lead, an additional set of SuperMix models was constructed from HFSS calculations for other waveguide geometry features such as steps in its height, fillets (rounded corners), and the waveguide's terminating back-short. These element models were then combined in a SuperMix model which could optimize the relative positioning of these features to estimate a candidate optimized waveguide and probe geometry. An HFSS model evaluation of this full-featured geometry was then used to refine the SuperMix model to improve its optimization. After a couple of iterations of this procedure, a complete HFSS model of the final waveguide and probe structure was built and used to generate final performance results for use with the mixer chip design discussed in the previous chapter.

The resulting waveguide probe structure finally adopted for the receiver in January 2002 is shown in Figure 3-1, and its HFSS-modeled performance is shown in Figure 3-2 on page 56. That figure provides a Smith chart of the RF waveguide probe's reflection coefficient presented to the mixer chip RF circuitry. An ideal probe would have zero

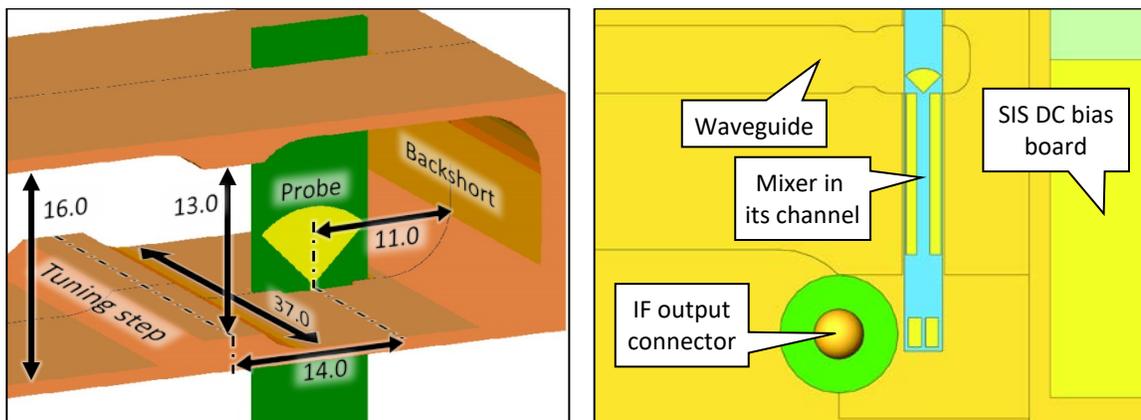


Figure 3-1: RF waveguide and waveguide probe design with dimensions shown in mils. The 150 micron radius waveguide probe is formed from superconducting niobium film integrated onto the SIS mixer chip. The silicon chip substrate is 230 microns wide by 25 microns thick. The chip is mounted in a channel machined into one half of the two-piece mixer block as shown at right. The rectangular RF waveguide machined into the mixer block halves has the dimensions (in mils) shown in the left figure. All rounded fillet radii in the waveguide structure are 5.0 mils, suitable for precision end-mill machining.

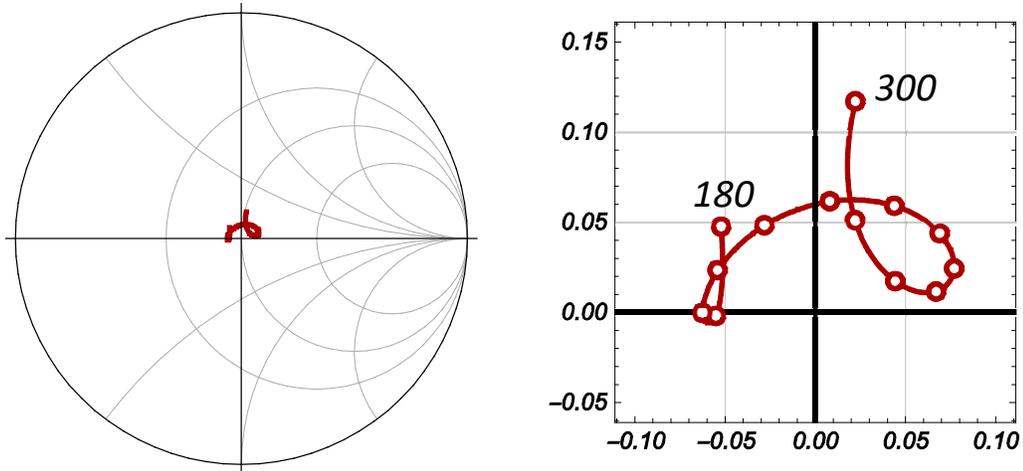


Figure 3-2: Smith chart showing the predicted waveguide probe performance. Modeled using *Ansoft* (now ANSYS) HFSS, the plots show the expected probe reflection coefficient as seen by the RF matching network. The circles in the right-hand plot are spaced at 10 GHz intervals covering 180–300 GHz. The chart normalizing impedance is 36.8Ω , the real part of the nominal probe impedance over the RF design bandwidth. The imaginary part of the probe's nominal impedance is predicted to be $+3.2 \Omega$. The probe's impedance match to its nominal impedance is better than -22 dB over the RF bandwidth.

reflection over the RF band, indicating perfect coupling of the RF energy between the mixer circuitry and the waveguide. The figure illustrates that for an appropriate choice of the RF circuitry's input impedance, 36.8Ω , the RF signal coupling from the waveguide will be very nearly ideal, with an average efficiency of better than 99% over the RF band and worse-case coupling exceeding 98.5% (at 300 GHz). If the RF mixer circuitry were designed to present an impedance with a small capacitive component, then its worst-case coupling efficiency could exceed 99% across the receiver's entire RF band.

To facilitate fabrication by an American machine shop, dimensions of the waveguide structure were chosen to be defined using *mils* (0.001 inch, 25.4 microns). The waveguide is, of course, designed to be used in its primary, TE_{10} mode, and has cross-sectional dimensions of 37.0×16.0 mils, slightly less than full-height (which would have a height of $37.0/2 = 18.5$ mils). The design cutoff frequency of the TE_{10} mode of the waveguide is thus $c/(2 \times 37.0 \text{ mils}) = 159.5$ GHz, and the cutoff frequencies of the next propagating waveguide modes are TE_{20} at 319 GHz and TE_{01} at 369 GHz. This choice of waveguide dimensions is therefore suitable for the receiver RF bandwidth of 180–300 GHz. The

waveguide is designed to be machined in two halves, the split being the plane normal to the center of the long axis of its rectangular cross section. The two halves of the mixer block can then be machined using standard end milling and are described in a later section. To support successful, precise machining of the waveguide structure, its corners were designed to be rounded (filleted) with a common radius of 5 mils. These fillets were, of course, incorporated in the HFSS EM models.

Concurrently and in coordination with the author's efforts, Jacob Kooi, a research engineer and instrument maker in our group at Caltech, spent a lot of time and effort further investigating this probe design. He built scale models of the waveguide probe and ran many HFSS simulations to more fully explore the effects of design geometry variations on the probe coupling performance. Our joint findings were incorporated into the designs of Hershel instrumentation and his designs of a new suite of CSO heterodyne receivers deployed several years later (in 2012) [25]. Because of Jacob's efforts to get the word out as early as 2003 [26], however, this radial waveguide probe design has found wide application in submillimeter RF circuitry, as mentioned in Chapter 6.

Mixer chip mounting channel

As has been previously described, the mixer chip is mounted in a channel which extends away from the RF waveguide wall. Proper design demands that this channel, which includes the dielectric mixer substrate, must not have waveguide modes which support propagation of RF radiation within it. It is important, therefore, to ensure that the cutoff frequencies of propagating waveguide modes in the channel holding the silicon chip are higher than the highest design RF frequency of the receiver. Unfortunately, the silicon mixer chip must be wide enough to support its integrated waveguide probe, which requires nearly 1/4 mm. Combined with silicon's large dielectric constant of 11.9, the existence of mixer channel propagating modes with cutoff frequencies within the receiver's 180–300 GHz operating range is a distinct possibility: $c / (2 \times 0.25 \text{ mm} \times \sqrt{11.9}) = 174 \text{ GHz}$.

Because the minimum width of the silicon chip is constrained by the waveguide probe design, steps must be taken to raise the lowest cutoff frequency of these channel waveguide modes. As is well known, this may be accomplished by using a very thin

silicon chip and adding a vacuum-filled volume adjacent to the chip which can serve to lower the effective dielectric constant of the silicon + vacuum filling the channel. A study was therefore undertaken to understand the dependence of the mode cutoff frequencies on the geometry of a rectangular waveguide partially filled with dielectric in order to optimally design the silicon mixer chip support channel. Completed in mid-2000, the results of that analysis are detailed in Appendix F. The final mixer chip channel design cross section is illustrated in Figure 3-3; also shown in the figure is a graph of the variation in channel cutoff frequency with the size of the vacuum gap. It was calculated using the results derived in Appendix F for the TE_{10} cutoff frequency of a partially-dielectric-filled rectangular waveguide.

Note from the diagram in Figure 3-3 that the ground plane (including the chip beam leads) deposited on the top surface of the mixer chip effectively divides the chip channel into two independent waveguides: a completely vacuum-filled one above the mixer circuitry and another which includes the chip substrate underneath the mixer circuitry. It is this latter waveguide whose cutoff frequency concerns us. Note from the graph that even a small vacuum gap can significantly raise the TE_{10} cutoff frequency: a 0.5 mil gap under the mixer chip as used on either side of the central, larger gap would be sufficient

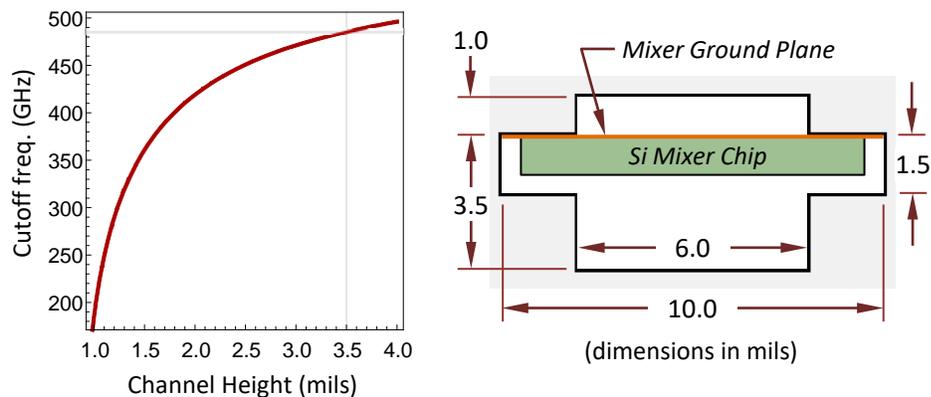


Figure 3-3: Mixer chip channel cutoff frequency vs. channel height and final geometry. The graph shows the calculated TE_{10} cutoff frequencies for a mixer chip 10 mils wide and 1 mil high (slightly larger than the actual chip) and including uniform vacuum gaps creating the total channel heights shown. The frequencies were found by solving the transcendental system of equations (F.29) – (F.31) derived in Appendix F. For a total channel height of 3.5 mils, the calculated cutoff frequency is just over 485 GHz. The actual chip channel geometry is shown in the right-hand diagram.

to increase the cutoff frequency to over 360 GHz, well outside the design bandwidth of the receiver. Because the 3.5 mil section of the channel is not as wide as the mixer chip, the actual cutoff frequency will be somewhat lower than that predicted in the chart. Even if the resulting effective gap height were only half of its 3.5 mils, however, the expected cutoff frequency would be over 400 GHz. This channel geometry was included in the HFSS model when calculating the predicted probe performance shown in Figure 3-2, showing that its design is completely satisfactory.

Feed horn waveguide interface

The RF radiation from the telescope optics is coupled into the mixer waveguide using a corrugated, conical feed horn. The output signal from the horn is coupled into its integrated, circular waveguide (with a diameter d of 45.3 mils). The circular waveguide's

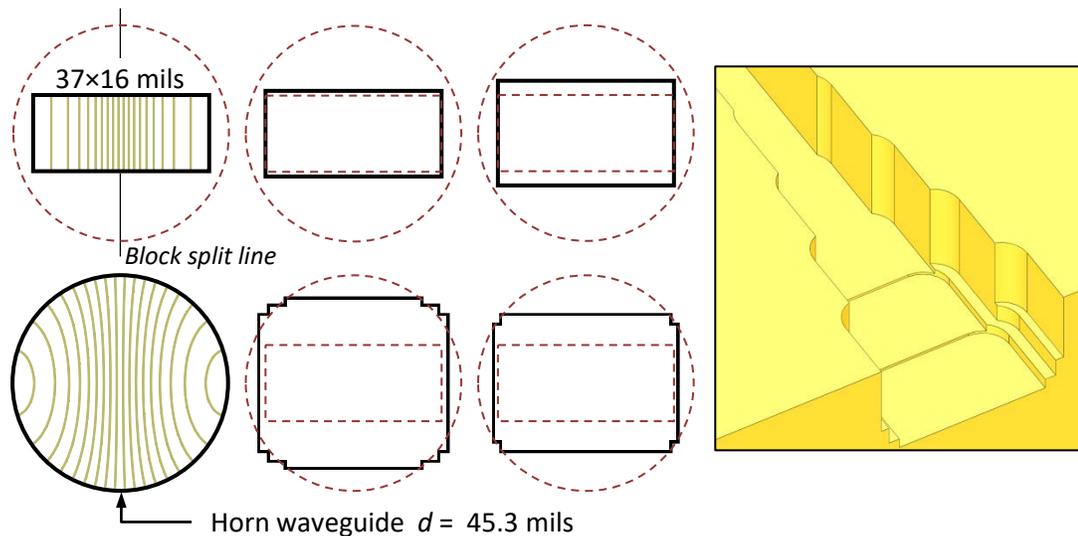


Figure 3-4: Feed horn cylindrical to rectangular waveguide transition geometries.

A four-section, $\frac{1}{4}$ -wave transformer serves to transition the corrugated feed horn output waveguide to the mixer rectangular waveguide. Cross section geometries for the individual waveguide sections are illustrated in the diagrams starting at the upper left corner and proceeding clockwise. As with the waveguide probe interface shown in Figure 3-1, the transition is designed for end-mill machining into a split-block configuration with a 5 mil fillet radius on all cuts. The right-hand image shows an oblique view drawing of the designed waveguide transition machined into one half of the split mixer block. Detailed dimensions are provided in Fig. B-8 on page 123.

dominant, TE_{11} mode (described later) carries the RF signal which must then be coupled efficiently into the mixer waveguide's rectangular TE_{10} mode. A four-section transformer integrated into the end of the mixer block waveguide is designed to accomplish this task. The end result of the transformer design process is shown in Figure 3-4 on page 59, with detailed dimensions available in Appendix B, Fig. B-8.

The left-most waveguide cross sections in Figure 3-4 are of the mixer block's rectangular waveguide and the horn's circular waveguide. Shown in green are electric field line plots for the two respective dominant waveguide modes (the densities of the displayed field lines are proportional to the electric field magnitudes). The circular waveguide's modes are naturally expressed in terms of solutions using cylindrical coordinates, which are products of Bessel and trigonometric functions: $\propto r^{-1}J_n(k_m r) \cos(n\theta)$ for the radial TE component and $\propto J'_n(k_m r) \sin(n\theta)$ for the angular component. For TE_{nm} modes the radial wavenumber k_m is chosen so that at the wall radius a , $k_m a$ is the m^{th} zero of J'_n , the first derivative of J_n (see for example [27]). The radial wave number k_m turns out to be the cutoff wavenumber of the circular waveguide for that mode, thus determining its cutoff frequency. For its dominant TE_{11} mode $k_1 a = 1.84118378\dots$, and with $a = 45.3/2$ mils, $f_c = 152.8$ GHz. Contrast this value with the 37 mil rectangular waveguide's dominant TE_{10} mode cutoff frequency of 159.5 GHz. This difference in cutoff frequencies slightly complicated the design of the waveguide transformer, but use of tools such as HFSS and SuperMix made the task straightforward.

It should be clear from the field line plots in Figure 3-4 that the circular TE_{11} and the rectangular TE_{10} modes are naturally related by a continuous transformation between the two waveguide shapes. The waveguide transformer approximates this transformation using a finite number of discrete steps. A straightforward approach to designing a stepped waveguide transformer is available when the various sections are designed to have a common cutoff frequency, because then the characteristic impedance ratios of the sections are independent of wavelength. In this case, long-established methods for the design of transmission line matching transformers can be used as described by Bathker [28]. In our case, however, the waveguides to be matched have different cutoff frequencies, and each one's characteristic impedance will vary with frequency as

$(1 - f_c^2/f^2)^{-1/2}$. Over the receiver's broad RF bandwidth of 180–300 GHz, the circular and rectangular waveguides' impedance ratio will therefore vary by about 12%.

Given the respective cutoff frequencies and frequency-dependent characteristic impedances of the two waveguides, SuperMix was used to calculate the impedances, cutoff frequencies, and lengths of the four transformer segments required to optimize the matching over the design RF bandwidth. Following Bathker, the two transformer segments closest to the mixer waveguide are rectangular, sharing its 37 mil width and therefore cutoff frequency. The other two segments approximate appropriately truncated cylindrical waveguide cross sections. These cross sections were actually stepped as shown in Figure 3-4 to accommodate end-mill machining of the waveguide sections. All of the required machining cuts to form each half of the split-block waveguide transformer end with 5 mil fillets. HFSS was used to model the resulting structure to determine the transformer's matching performance. This process was iterated a few times to ensure that the actual stepped geometry performed adequately. The final design was completed in March 2002, and final HFSS modeling predicted an RF signal reflection at the interface approaching –30 dB over much of the receiver bandwidth.

Other aspects of the mixer block design and mounting

The receiver's wide IF bandwidth design requires that the mixer chip output be physically very closely coupled to its accompanying cryogenic, low-noise IF amplifier (LNA). The reason for this requirement is that residual impedance mismatches are to be expected between the mixer chip output, the LNA input, and the interconnecting transmission line (see, for example, Figure 2-18 on page 48). By keeping the phase length of the interconnection between the mixer chip and the LNA short, impedance mismatches then may not excite large standing waves (cavity resonances) in this interconnection at IF frequencies.

Additional design requirements on the mixer block assembly include: (1) the waveguide connecting the receiver's RF feed horn to the mixer chip waveguide probe should have smooth walls and should be kept short to minimize RF signal losses; (2) an electromagnet must be included to supply the magnetic field required to suppress

Josephson tunneling current oscillations through the SIS junction; and (3) SIS DC bias current must be supplied to the mixer chip, and the DC bias voltage across the SIS junction must be monitored. Finally, the mixer block and its mounting must provide effective thermal coupling between the superconducting mixer chip and the cryostat cold plate (cooled to 4.2K by a liquid helium reservoir) while also keeping the RF feed horn aperture accurately aligned with the optics coupling it to the external telescope RF signal. The resulting mixer block design is shown in Figure 3-5 below and is also detailed in the various diagrams found in Appendix B. The drawings in that appendix and in this section are of the “second generation” design completed in 2004 and fabricated at JPL in 2005.

The connection from the mixer chip IF output to the LNA circuitry is only about 20 mm long and is limited by the length of the high-frequency microwave connector joining the LNA and mixer block. As was discussed earlier in this chapter, the connection from the mixer chip to the mixer block’s IF connector is accomplished using a pair of parallel wire bonds each only about 1/3 mm long (see Figure 2-16 on page 46 and Figure 3-1 on page 55). Although the total resulting length of the mixer chip to LNA connection

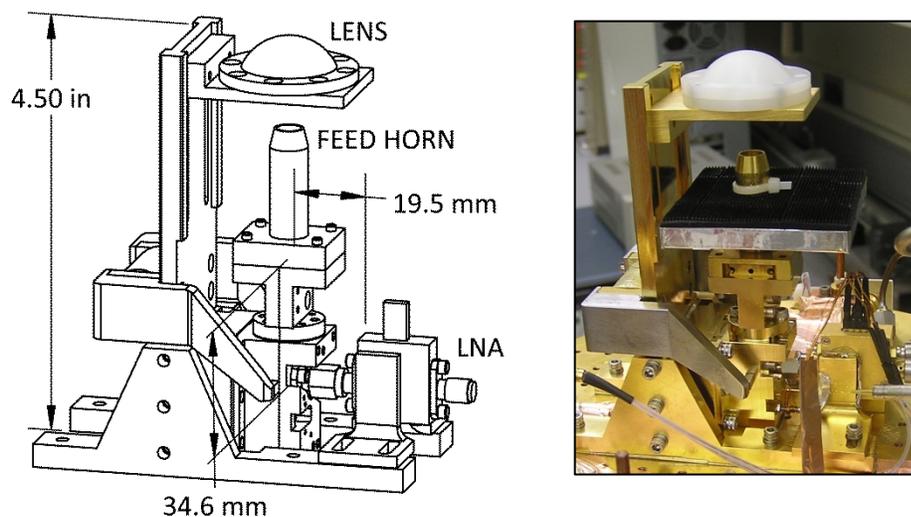


Figure 3-5: Physical configuration of the mixer block, feed horn, and LNA assembly. The photo shows the complete assembly mounted to the cryostat 4.2K cold plate. An interface assembly joins the feed horn to the mixer block in this second-generation version installed in 2005. Dimensions in millimeters show the approximate physical lengths of the RF and IF signal paths from the SIS mixer chip to the feed horn and LNA. Compare to Figure 1-6 and Figure 1-7 on page 10 and to the drawings in Appendix B.

is 1 wavelength at 15GHz, it could be further reduced only by integrating the LNA circuitry into the mixer block. It also should be noted that this close, direct coupling of the mixer block IF output connector to the SIS mixer chip is potentially dangerous, because the mixer chip is very susceptible to damage from electrostatic discharge (ESD). Once the LNA is attached to the mixer block, however, it serves to isolate the mixer chip from ESD introduced at the cryostat IF output connector.

The cylindrical–rectangular waveguide transition described in the previous section was incorporated into the first-generation mixer block RF waveguide, and the RF feed horn was connected directly to that mixer block. In the second-generation design this transition was moved into a separate interface assembly, and the mixer block RF waveguide was terminated with a standard waveguide interface connector (see Appendix B, Fig. B-2 and Fig. B-7). This was done so that the mixer block assembly would be compatible with future, more complicated receiver designs planned by JPL. Although this design change added an inch to the RF waveguide length, JPL's fabrication incorporated very high-quality, very smooth waveguide machining, greatly mitigating the losses introduced by this additional length. Luckily, the author's cryostat design included just enough vertical space above its 4.2 K cold-plate to accommodate this length increase, as shown in Appendix B, Fig. B-9 on page 124.

The photo in Figure 3-5 shows the complete support structure attaching the mixer assembly to the cryostat cold plate. This structure, fabricated from high thermal conductivity copper, effectively keeps the mixer and LNA temperatures stable at 4.2 K. The niobium wiring of the mixer's electromagnet is also kept superconducting by this same thermal connection. Lastly, this structure's rigidity also accurately maintains the receiver's optical alignment with the telescope, discussed further in the next chapter.

Cryostat wiring considerations

Finally, a few notes concerning heat burden introduced by the electrical signal and control connections to the cryostat's cold receiver components. The cryogen "hold time," the time it takes for either the cryostat LN2 or LHE reservoir to empty, should be long enough that an undue servicing burden is not placed on the observatory staff when the

receiver must be kept continuously operational for several days. In the case of the author's receiver, the design target was a hold time of at least 26 hours, so that cryogen servicing was only required once daily (with some scheduling flexibility). Proper sizing of the cryogen reservoirs is, of course, an important design task to accomplish this goal, but so is proper design of the cryostat electrical wiring.

The mixer chip, LNA, and the Josephson current suppression magnet each require DC electrical bias and control connections, as do the sensing diodes used to monitor the temperatures of the 4.2 K and 77 K cryostat stages. Of course, the LNA IF output signal also must be routed out of the cryostat. Each of these connections also acts as a heat conduit which increases the load on the cooled stages and consequently the rate of cryogen evaporation. Many of these connections carry a negligible electrical current at very low frequencies (essentially DC), so long, thin, low-conductivity wires can be used to reduce their heat loading on the cooled stages. A few, however, such as the LNA and magnet bias connections, are required to carry up to a few dozen milliamps of current when the receiver is in active operation (typically 40–60 mA for the receiver's magnet bias wiring). If these wires have excessive electrical resistance, then ohmic heating within them can dominate the heat load they introduce to the cooled stages. Proper sizing and choice of materials for these wires are then important considerations when designing the cryostat.

The author began building up the receiver cryostat internal wiring in mid-2000.¹³ A literature search at that time provided little useful information on the sizing of wiring carrying low, but nonnegligible, currents across large temperature gradients. A short theoretical study was undertaken to guide the wiring design effort, the results of which are provided in Appendix H beginning on page 186. This information was informally made public on Jacob Kooi's CSO instrumentation website a few months later. As should

¹³ Assisting the author in this effort was Jonathan Simon, then a Caltech physics undergraduate student. Currently Professor at the University of Chicago, Jon leads a large, active group of researchers in the experimental study of many important topics at the interface of condensed matter, quantum optics, and quantum information.

have been expected, these results turned out to be a duplication of another investigator's earlier work: a 1963 paper by Wladimir Mercouroff [29].

Chapter 4

RF OPTICS AND RECEIVER PERFORMANCE

RF optics design

This chapter discusses relevant features of the design of the remaining component of the “cold” receiver subsystem: the RF optics. As illustrated by the images of Figure 1-3 on page 6 and Figure 1-4 on page 7, the receiver cryostat was designed to be mounted on the CSO telescope’s Cassegrain Relay Optics (RO) assembly. The optical and mechanical interfaces of receiver systems to this assembly are described in papers by Caltech’s Gene Serabyn [9] [30]. According to [9], the RO produces a (geometric, ray optics) $f/4.48$ telescope focus located 7.60 inches (193 mm) above its mounting surface at either of two instrument locations (see Figure 1-3). The optical axis at each location is normal to the RO mounting surface and is coincident with the axis of the RO instrument mounting plate’s rotation bearing.

Because the author’s receiver was designed to be a “single pixel” heterodyne spectrometer and not an imaging system, its interface to the telescope RO optics was considerably simplified. The receiver RF optics have a single axis which is coincident with the long axis of symmetry of its cylindrical cryostat, so the cryostat need only be mounted with this long axis coincident with the RO optical axis and at the correct height above the RO mounting surface. The receiver optics must then properly match the mixer RF feed horn to the RO $f/4.48$ signal. In this section we describe how the receiver’s simple optical design accomplishes this task. A full, three-dimensional model of the RF input signal electromagnetic field is far too cumbersome for this optical system design task, so a much simpler approximation of the signal fields was used: a *fundamental Gaussian beam mode* solution of the *paraxial* approximation of the Helmholtz wave equation, an approach explained in more detail in Appendix G: *Gaussian optics review* starting on page 178. The now-standard reference for Gaussian beam analysis is the text by P. F. Goldsmith [4], and it was used extensively by the author as he refined the receiver’s RF optical design.

As previously described, the telescope RF signal is coupled into the mixer block waveguide using a cylindrical, corrugated feed horn. The feed horn was graciously given to the author for use with his receiver by Jacob Kooi, a Caltech submillimeter group colleague. Although the detailed provenance of the horn is unclear, it was probably used in a 230 GHz CSO prototype receiver built around 1990 [31]. The feed horn's detailed design and performance were not available to the author, but, based on the results presented in [31], it was probably designed for 200–290 GHz operation. Careful measurements of the horn indicated that its circular aperture has a diameter of 6.5 mm and a slant length (along the horn's conical waveguide surface) of 40.1 mm. These dimensions were used to estimate the feed horn's Gaussian beam pattern shape using the methods described in [4]. The horn's dimensions indicated that this fundamental Gaussian mode would have a waist radius that would be only very weakly dependent on the RF frequency, so its asymptotic beam f /ratio would vary with RF frequency, from about $f/1.71$ at 180 GHz to $f/2.75$ at 300 GHz (assuming a 10 dB *edge taper*, as explained in Appendix G).

The receiver RF optics design is required to transform the feed horn's frequency-dependent beam to the telescope's $f/4.48$ signal beam over its RF frequency range. After much effort "reinventing the wheel," the author rediscovered a fact that had at that time been known for over a decade: a single lens, properly chosen and positioned, can match a corrugated feed horn's frequency-dependent Gaussian beam pattern to a fixed f /ratio, waist position, and edge taper (within fairly broad limits). An initial design used in the 2003 version of the receiver resulted in disappointing performance. Analyzing the results of a map of that design's beam pattern led to an improvement in the model parameters used to numerically optimize the design. The final design used a plano-convex HDPE (high-density polyethylene) lens with a surface radius of 0.55 inch (18 mm) positioned 0.54 inch from the feed horn aperture. Figure 4-1 on page 68 shows a schematic of this design along with the modeled beam calculations for the limits of the receiver design RF bandwidth. The model predicted that at all design RF frequencies the receiver's Gaussian beam would have a very nearly constant beam waist position and an asymptotic $f/4.48$ pattern with a 10.2 dB edge taper.

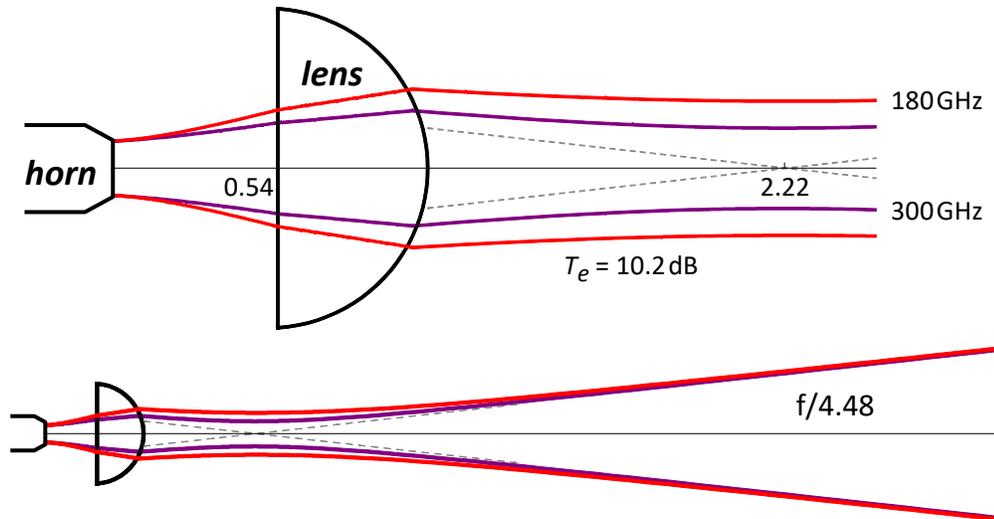


Figure 4-1: Receiver RF optical design with modeled Gaussian beam plots. The positions of the lens planar surface and the final beam waist (0.54 in and 2.22 in, respectively) are measured from the feed horn aperture. Also shown are 180GHz (red) and 300GHz (violet) Gaussian beam contours at a common edge taper of 10.2dB. The spherical surface of the lens has a radius of curvature of 0.551 inch (18 mm).

The final design configuration of the receiver mixer block, feed horn, optics and cryostat is detailed in Appendix B, Fig. B-9 on page 124. The receiver beam's resulting *confocal distance* (see Appendix G) provides an order of magnitude estimate of the “depth of field” around its beam waist position. In this case it is just over 1 inch at 300 GHz, increasing to 1.7 inch at 180 GHz. This implies that the relative position of the beam waist and the telescope focus (along the beam axis) is not all that critical and need only be matched to within a few tenths of an inch.

Receiver RF beam measurement

By the summer of 2005 it was clear that the receiver's optical coupling to the telescope needed to be improved. Initial laboratory receiver RF beam measurements demonstrated that the beam waist position and edge taper were both in error. An analysis of this data resulted in an improved model of the optical characteristics of the HDPE lenses when cooled to LHe temperatures, and the optics were then redesigned. This updated design has been described in the previous section. Concurrently, the cryostat

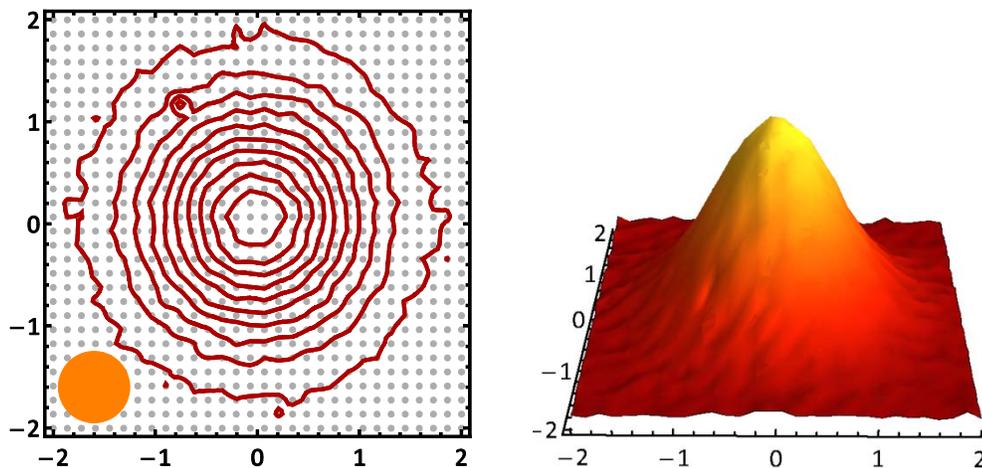


Figure 4-2: Receiver beam raw mapping data for the final cryostat optics configuration. The contour plot of the SIS response also shows the beam mapper source raster positions as gray dots, and the orange circle shows the physical size of the source (0.7 inch diameter). The scales show (x, y) positions normal to the beam axis in inches. The optical axis path length from the mapper to the cryostat RF window was 14.25 ± 0.2 inches.

mounting was modified to raise it another 3.8 inches above the relay optics mounting surface to bring the redesigned receiver beam waist to the telescope focus. Following these modifications the receiver beam was again measured to check the effectiveness of the final design, and the results of this measurement are presented in this section.

The Caltech submillimeter group's receiver "beam mapper" consisted of a small, hot source superimposed on a room-temperature background. The source thermal radiation output was then modulated by a room-temperature shutter, and its position was x - y scanned across the receiver RF input beam under computer control. The receiver's SIS mixer was used in its *direct-detection* mode: with no local oscillator power applied, the SIS bias current at its DC bias voltage (2.3 mV) was very slightly modulated by the presence of the modulated hot source in the receiver's RF beam. A lock-in amplifier was used to measure the amplitude of this bias current modulation as the source position was varied. Because the local oscillator and receiver IF output signals were not used, the beam mapping data provided a measurement of the receiver beam profile averaged over its full RF bandwidth.

The beam mapping was performed in early September 2005 with the invaluable aid of the author's colleague Tasos Vayonakis. The mapping data is presented in Figure 4-2.

This measured beam pattern data is actually that of the receiver beam shape convolved with the intensity pattern of the beam mapper's hot source, so we must deconvolve the measured data to recover the receiver beam. If the hot source intensity pattern were a circular Gaussian, then its convolution with a circular Gaussian receiver beam would remain Gaussian with a variance given by the sum of the component Gaussian variances. This is an adequate working assumption for an initial fit to the mapping data.

A clearly bad data point at $x = -0.76$ inch, $y = 1.17$ inch was replaced by the mean of its neighboring point values, and a circular Gaussian function was then fit to the data using chi-squared minimization. First, the data uncertainty was estimated by calculating the variance of a set of 16 data values, 4 taken from each of the 4 corners of the full data set. The standard deviation of the fluctuations in this data subset was 39 dB below the full data set's maximum measured value and nearly 26 dB below its mean value. The final data fitting function parameters included not only the Gaussian amplitude and width, but also x and y offsets for the beam center as well as a constant background value. The resulting reduced chi-squared of the fit was 1.6, showing that the convolved beam shape is unlikely to be exactly Gaussian. Even so, the fit residuals averaged only 25% larger than the assigned data uncertainty and appeared to be dominated by random fluctuations. This model provided good initial results for the beam matching calculations, especially for the beam center offsets.

A subsequent, final fit assumed a circular Gaussian receiver beam convolved with a circular hot source disk of uniform intensity and known radius. The circular symmetries of these component patterns imply that their convolution will be a function of radius only, greatly simplifying the numerical convolution calculations. Assuming that the initial fit correctly determined the x and y offset values, chi-squared minimization of the convolved beam pattern function to the data was performed to determine the receiver beam width and amplitude and the background level best-fit parameter values. The results of this final fit are shown in Figure 4-3. The fit reduced chi-squared value was 1.18, differing enough from 1 (almost 4 sigma, given 900 data point values) to indicate that a slight deviation from the assumed circular Gaussian beam shape was still present.

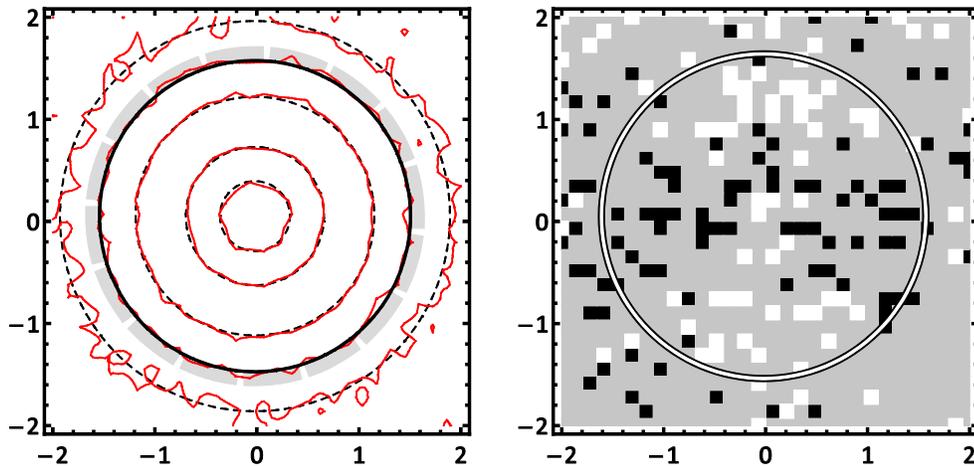


Figure 4-3: Final convolution model fit to the raw receiver beam map data of Figure 4-2. The left plot shows fit (black) vs. data (red) contours at 0.5, 2, 6, 10.2 (solid line), and 16 dB below the fit central peak value. The right plot shows the pattern in the final fit residuals (data point – function values). The black and white squares in this plot denote measured values which were noticeably below and above the fit results, respectively. The thick, gray, dashed circle in the left plot and the thin white circle in the right plot represent the calculated edge radius of an $f/4.48$ beam from the receiver at the location of the beam mapper. The circle's thickness in the left plot represents the uncertainty in this calculation. Scales are inches in the plane normal to the beam axis.

At least one aspect of this deviation is evident in the residuals plot in Figure 4-3. The gray area in that plot denotes data points whose residuals were no more than 1.4 times the estimated data uncertainty, whereas the black and white squares denote data points with residuals exceeding those limits. A clear pattern is evident which indicates that there is a very slightly elliptical component to the data with axes nearly aligned with the x and y directions in the plot. This effect was most likely caused by some x - y asymmetry of the beam mapper itself, such as it not being aligned precisely orthogonally to the receiver beam axis. It could also indicate, however, that the actual receiver beam was slightly elliptical, possibly because of the 90° reflection in the beam path between the cryostat and the beam mapper or because of some misalignment between the RF lens and feed horn axes. In any event, these deviations are small: only about 6% of the data residuals (57 out of 900) exceeded twice the assigned data uncertainty, and the worst offender had a residual amounting to 4.4% of the on-axis fit intensity, or 3.8 times the assigned uncertainty. Figure 4-4 illustrates the relative sizes of these deviations from the model fit.

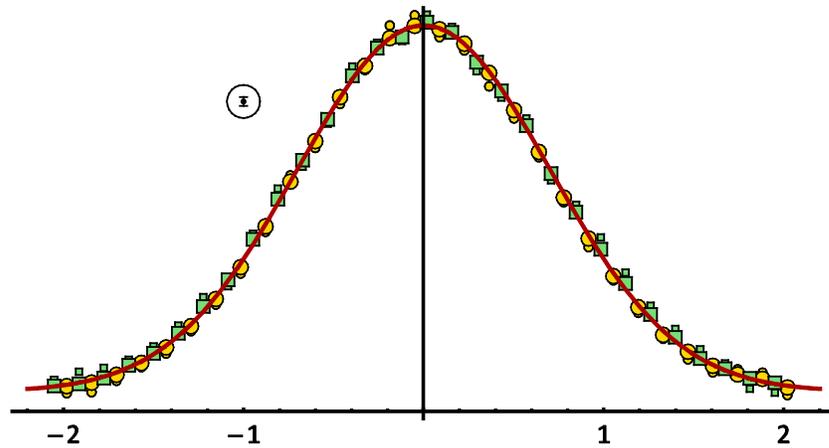


Figure 4-4: Another view of the x vs. y data deviation asymmetry from the RF beam fit. This plot illustrates the relative sizes of the fit residuals for data slices within ± 0.5 in of the x -axis (yellow circles) and y -axis (green squares). The larger symbols show the mean residual values and the smaller symbols show the maximum individual residual values within those slices. In each case the residual values have been added to the model fit result along the respective axis (red curve). Both axis origins have been translated to the beam center for this plot. Inside the black circle is a point with the assigned data uncertainty error bars to use for comparison. Clearly, the data were precise enough to uncover what are actually quite small deviations in the beam shape.

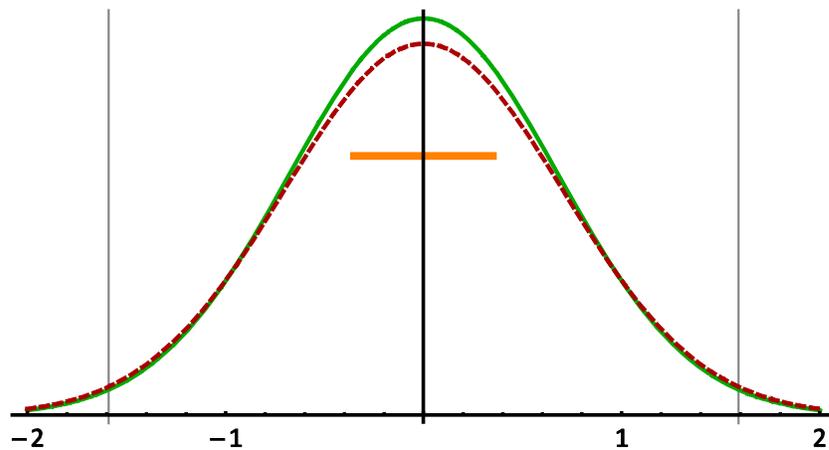


Figure 4-5: The deconvolved receiver RF beam as inferred from the beam map data. The green curve shows the radial variation of the receiver RF Gaussian beam model. Subsequent convolution with the beam mapper hot source circular profile (source diameter indicated by the orange line) results in the dashed red detection response profile, which is the function directly sampled by the data measurements. The gray vertical lines mark the edge of the telescope's $f/4.48$ beam: their intersections with the green profile give a resulting receiver edge taper of 11.9 dB.

The deconvolved, circular receiver beam model profile is shown in Figure 4-5 on page 72. This profile represents the receiver's power sensitivity, so that its value varies with radius as $\exp(-2r^2/w^2)$. The receiver's Gaussian beam width at the beam mapper location is $w=1.36$ inch, and therefore its edge taper is 11.9 dB at the radius corresponding to the telescope's $f/4.48$ geometrical optics limit, a bit greater than the design target value of 10.2 dB. This change would reduce the expected receiver's antenna illumination aperture efficiency by only about 2% (see Appendix G).

Optics noise temperature impact

Next we evaluate how signal losses in the RF optics would be expected to impact the receiver system noise temperature. The RF signal from the telescope must pass through four optical elements on the way to the feed horn: the LO beam combiner, the cryostat window, the heat shield IR filter, and the RF focusing lens. The elements' assumed material characteristics over the receiver RF bandwidth are summarized in Table 4-1 on page 74. This information will be used to generate an estimate of the RF signal power loss caused by these components. Because the dielectric loss tangents are small, their power losses will be dominated by specular reflections at material interfaces.¹⁴

Except for the LO beam combiner, the surfaces of the various elements are approximately normal to the RF signal direction of propagation, and each element has a diameter which is much larger than the RF beam width. For the purpose of estimating losses, we may treat the signal propagation through these elements as 1-dimensional and can use the scattering matrix formalism outlined in Appendix C. In particular, every material interface represents a change in the medium's wave impedance, resulting in a partial reflection as described by equation (C.3) on page 130. A medium's wave impedance varies as $Z = \sqrt{\mu/\epsilon}$, whereas its refractive index varies as $n = \sqrt{\mu\epsilon}$. Because the dielectric materials in Table 4-1 are nonmagnetic, we can use (C.3) for the reflection

¹⁴ For a relatively low-loss dielectric, the material's loss tangent describes a signal's fractional power loss per radian of phase along its direction of propagation (see also Chapter 7 of Jackson [3]).

Table 4-1: RF optics material properties

Characteristics derived mainly from data in Lamb [32], but also from [33] and [4].

Material	Temp (K)	n	$\tan \delta (\times 10^{-4})^e$	Element
HDPE ^a	290	1.52	4	Window Lens
	4	1.57	4	
Microporous PTFE ^b	290	1.22	3	AR coating
Fluorogold ^c	77	1.61	69	IR filter
BoPET ^d	290	1.83	200	LO beam combiner

a High-density polyethylene.

b Polytetrafluoroethylene, trade named *Teflon* ® The Chemours Company (a DuPont spin-off). Microporous PTFE was developed to serve as a chemical filter material. It has the trade name *Zitex G* ® Saint-Gobain. The material used had a 50% PTFE filling factor.

c A solution of tiny, elongated glass particles embedded in PTFE. Developed as a gasket material, it is an effective scatterer of IR, but not submillimeter radiation. This product name was ® Fluorocarbon, Inc., and is not to be confused with a fluorescent dye of the same name, currently ® Fluorochrome, LLC.

d Biaxially-oriented polyethylene terephthalate polyester, originally trade named *Mylar* ® DuPont (among others).

e The material's *loss tangent*, the ratio of its dielectric constant's imaginary and real parts (see the *Glossary*).

coefficient Γ if we substitute $Z_0 = 1/n_i$ and $Z = 1/n_f$ when crossing from dielectric n_i into dielectric n_f . The fractional power transmitted through the interface is then $t = 1 - |\Gamma|^2 = 4n/(n+1)^2$, where $n = n_f/n_i$.¹⁵

The various rays of the RF signal passing through the volume of the lens, as well as those passing through the vacuum spaces between elements, have a variety of different phase lengths. For these cases we can therefore ignore the effects of cavity resonance modes between the interfaces. For the interiors of the IR filter and the cryostat vacuum

¹⁵ This expression assumes that n is real-valued, thus ignoring the materials' loss tangent contributions at the interface. More generally, $t = 4\text{Re}(n)/|n+1|^2$. Even in the case of Mylar® this correction would be completely negligible.

window, however, the effects of reflections between their surfaces must be considered. Short transmission lines with characteristic impedances and wave propagation constants determined by their refractive indices may be used to represent frequency-dependent transmission models of these elements. In the case of the HDPE vacuum window, a thin Zitex layer was bonded to its exterior surface to serve as an antireflection coating, again represented by an appropriate transmission line. The resultant modeled fractional power transmissions through the optical elements are shown in Figure 4-6. Clearly, surface reflections from the HDPE lens will noticeably reduce the receiver efficiency at all frequencies, and internal reflections in the vacuum window and IR filter will cause the low-frequency performance to especially suffer.

The beam combiner is treated slightly differently. Its 1 mil thick Mylar film is oriented 45° to the RF beam axis with the RF linear polarization to which the receiver is sensitive in its plane of incidence. The resulting complete element is called a “4-port directional coupler,” but the only matrix element of interest here is that for RF transmission through the element. The transmission (s_{21}) and reflection (s_{11}) coefficients at a dielectric interface as a function of angle of incidence and polarization can be found in almost any reference describing electromagnetic wave propagation. Here are, for example, the formulas as presented in the classic text by Born and Wolf [34], again

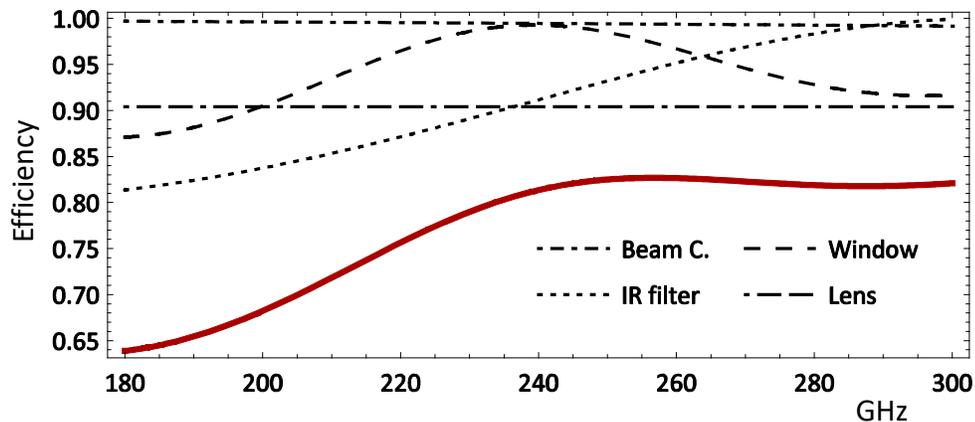


Figure 4-6: Receiver RF optics transmission efficiency estimates vs. RF frequency. The calculated signal power transmission efficiency estimates for the various elements in the RF optics: beam combiner, cryostat window, heat shield IR filter, and focusing lens. The red curve shows the predicted overall optical efficiency, which is the product of the individual element efficiencies.

assuming nonmagnetic materials, going through an interface from index n_1 into n_2 , with electric field polarized in the plane of incidence, and for angles of incidence θ_1 and refraction θ_2 :

$$s_{21} = \frac{2n_2 \cos \theta_1}{n_2 \cos \theta_1 + n_1 \cos \theta_2}; \quad s_{11} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2}. \quad (4.1)$$

The other two coefficients we need for the interface are $s_{22} = -s_{11}$ and $s_{12} = s_{21}(n_1 \cos \theta_2) / (n_2 \cos \theta_1)$. With these formulas for the behavior of the RF signal at the interface, the resulting transmission coefficient through a plane-parallel dielectric plate of thickness d is then:

$$\mathcal{T} = |t|^2; \quad t = \frac{s_{12}s_{21} \exp(2i\beta)}{1 + s_{11}s_{22} \exp(2i\beta)}; \quad \beta = (2\pi/\lambda_{vac})n_2 d \cos \theta_2. \quad (4.2)$$

As described in [34], this is one of *Airy's formulæ* (G. B. Airy, 1833). \mathcal{T} is the fraction of an incident plane wave's power transmitted through the plate, where β is the component of the signal's wave vector along the thickness of the plate times the plate thickness – in other words, the plane wave's phase change directly across the plate's thickness. The formula is for a lossless dielectric, but because the beam combiner has $d = 1 \text{ mil} = 25 \text{ micron} \ll \lambda_{vac}$ and has $\tan \delta \ll 1$ a lossless calculation will be adequate for our application. The resultant power transmission as a function of RF frequency is also shown in Figure 4-6: it varies nearly linearly from 99.7% at 180 GHz to 99.2% at 300 GHz.

The beam combiner's reflection will be the difference between 100% and these transmission values, giving the fractional power from the LO source coupled into the rest of the optics, which will then further attenuate its power as shown in Figure 4-6. The choice of polarization orientation and thickness of the beam combiner greatly reduces the noise temperature contribution from the LO source, but had the unfortunate byproduct of also limiting the LO power available to the mixer, especially at the low end of the RF frequency range. This left the mixer "LO starved" and resulted in disappointing noise performance at frequencies below about 220 GHz.

Because we are interested only in the overall RF signal transmission through the system and are ignoring cavity resonances between the elements, we can model the

elements as a series of impedance-matched attenuators with frequency-dependent power attenuations given by Figure 4-6. It must be remembered, however, that for our receiver assembly the loss of signal through each optics element is completely dominated by reflection, not absorption (as in a matched attenuator). The lost RF signal is not replaced by thermal radiation directly from the element, as would be the case for a true absorptive medium, but rather by reflection of thermal radiation emitted by the surrounding receiver structure.

The structure's surfaces are predominantly composed of low-loss, highly-reflective, metallic materials (gold and aluminum) with reasonably low thermal emissivities at the receiver's RF frequencies. The thermal radiation scattered by an optical element into the RF signal path will then be some weighted average of the physical temperatures of the cryostat structure surfaces. We therefore incorporate this scattering into our model by setting the temperature of the attenuator used to model an element to the expected radiation temperature that element scatters toward the receiver feed horn. In the case of the RF lens and IR filter, this radiation would likely be dominated by the 77 K heat shield surrounding the cold plate. The vacuum window, on the other hand, would predominantly reflect radiation from the heat shield, which in turn is reflecting the ambient temperature radiation from the cryostat body. Finally, as mentioned in the previous paragraph, reflection of noise (thermal and otherwise) from the local oscillator source would comprise the contribution of the LO beam combiner.

The basic, traditional way one might use to calculate the noise temperature impact of each optical element would be to proceed as follows: (1) the element's noise contribution introduced by reflection is given by the reflected source noise temperature multiplied by the element's power reflection coefficient, or, equivalently for our attenuator model, by setting the attenuator's temperature to that of the reflected noise source; (2) this noise is then added to the calculated noise temperature of the down-stream receiver system, and (3) this combined noise is divided by the element's power transmission efficiency to give the noise temperature of the receiver system now referred to that element's input. This procedure is iterated through each additional element situated upstream from the system, finally arriving at the input-referred noise temperature of the receiver including the RF optics.

This procedure would be fine for a two-port amplifier but not for the heterodyne mixer, because it neglects the additional, frequency-dependent noise introduced into the mixer's other RF sidebands. A more correct procedure would be to use SuperMix to model each optical element as a frequency-dependent attenuator with a temperature that is determined by estimating the background radiation temperature coupled by the element through reflection into the RF path. A circuit created by cascading these elements is inserted before the heterodyne mixer's RF waveguide probe input. SuperMix can then properly calculate the noise impacts of the elements in every RF sideband.

The SuperMix model results are presented in Figure 4-7, which one should compare to the mixer chip + LNA predictions in Figure 2-20 on page 52. Among the assumptions that went into the model were that the noise scattered into the receiver beam by the IR filter and by the lens could be represented by 70 K black-body radiation, by a 290 K black-body for the noise scattered by the vacuum window, and by a 1000 K black-body for the noise accompanying the LO signal to the beam combiner. Changing any of these assumptions would, of course, change the receiver noise temperature predictions in Figure 4-7. At mid to high RF frequencies the model predicts that the optics impact would be to slightly more than double the receiver's SSB noise temperatures, but that the

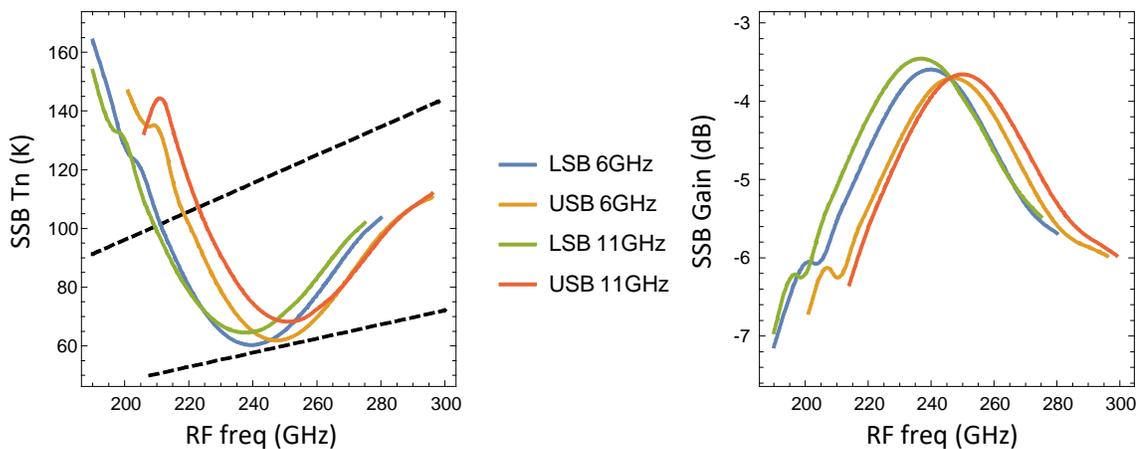


Figure 4-7: RF optics and LO noise impacts on the modeled receiver performance. SuperMix-generated model results for the mixer upper and lower sidebands are plotted for IF frequencies of 6 and 11 GHz. The RF frequencies are the actual sideband frequencies, $\nu_{LO} \pm \nu_{IF}$. The dashed lines in the noise performance graph show $5\times$ and $10\times$ the quantum limit noise temperature (which is 10 K at 208 GHz). Compare with the basic mixer chip + LNA results in Figure 2-20 on page 52.

optics efficiency losses near the lower end of the RF band will have a much greater impact. The original receiver design target was to keep the receiver noise temperature below 100 K SSB, which the model predicts to be met for RF frequencies above 210 GHz. More work would clearly be needed to meet that noise target for RF frequencies below that threshold.

Comparison with laboratory measurements

Throughout the receiver's development and operation, limited sets of Y-factor laboratory measurements of its noise performance were conducted. The most comprehensive set of measurements of the final receiver configuration was obtained in February, 2007, while the receiver was at Caltech for repairs. The measurements were conducted using a Gunn diode LO source by the author and Melanie Leong, a CSO staff senior electronics engineer. Comparisons of the full, RF optics + mixer + LNA model's Y-factor predictions with this measured data set are shown in Figure 4-8 on page 80. The Y-factor measurements were of the total receiver IF output powers over 4 GHz bandwidths centered at 6 GHz and 11.25 GHz. The figure compares these measurements to model predicted Y-factors for IF output frequencies corresponding to these band center frequencies. The model DSB noise temperatures were calculated by first calculating the IF output noise power spectral density when the RF sideband inputs are terminated by matched, 0K sources. That result was then divided by the sum of the calculated upper and lower sideband RF to LO power gains. Model Y-factor predictions were calculated by determining the IF output power level ratio when the RF input source was set to two RF test temperatures: in this case, 295 K and 80 K. As explained in Appendix C, one must be careful when converting a measured Y-factor to its corresponding DSB noise temperature, because at these source test temperatures the ground state quantum fluctuation contribution to the receiver output noise level all but disappears.

Clearly, and not too surprisingly, the model predicted better performance than was observed in the lab. The difference was as high as about 13K (DSB) for the 6 GHz IF, an increase of almost 50% over the model prediction. The higher frequency IF output band showed a nearly constant reduction of about 0.25 in its measured Y-factors versus those of the lower IF output band, a noise increase of approximately another 20 K (DSB). A possible reason for this discrepancy is that the higher IF output band was first down-

converted to 4–8 GHz by heterodyne mixing with a local, fixed oscillator signal at 16.5 GHz. This process could have injected additional noise (possibly in the form of narrow-frequency, spurious signals in addition to the oscillator’s output noise) which accounted for the excess noise observed in the laboratory measurements. The down-conversion was performed by an IF processor designed and assembled by Matthew Sumner, and it is described in his thesis [8]. This processor assembly offered a particularly difficult microwave design problem, and it did work well at handling such a challenging task.

These results also clearly illustrate an inherent, general problem with models which attempt to predict a system’s noise performance: such models rarely include every relevant noise source. Consequently, noise models are best interpreted as establishing lower bounds on the noise levels of an actual system.

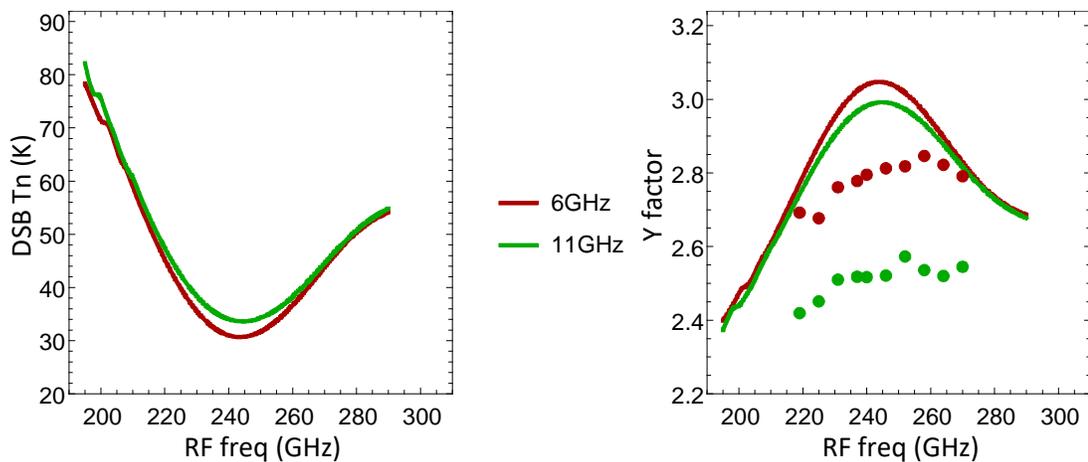


Figure 4-8: Receiver DSB noise and Y-factors model vs. measured results.

SuperMix-generated double-sideband (DSB) noise temperature predictions are plotted for IF frequencies of 6 and 11 GHz. The corresponding Y-factor predictions are plotted versus measured values from February, 2007, using a Gunn diode LO and for IF output frequency bands of 4–8 GHz and 9.25–13.25 GHz (the latter first down-converted to 4–8 GHz).

Chapter 5

SIS CIRCUIT MODELING AND SUPERMIX

This chapter and its associated appendices discuss the representations and algorithms used to model and optimize the design of the prototype receiver's SIS mixer chip and its associated waveguide structure as well as its interfaces to the IF output and DC bias circuitry. The most important design tool used for this effort was the SuperMix C++ class and function library, developed primarily by John Ward and the author, with important contributions from Goutam Chattopadhyay, Jonas Zmuidzinas, and, later, Anastasios Vayonakis, our associates in the Caltech submillimeter astrophysics instrumentation group [35]. Under Jonas's vision and guidance, the initial development of SuperMix began in 1996, and we saw its first public release in April 2000. Following the 2001 version 1.3 release, the author has been the SuperMix library's sole maintainer (19 years and counting). As of this writing, the publicly available version of the library, version 1.6, was released in 2009 [36] [37]. It contains over 47,500 lines of C++ code (including the 6800 line regression test suite).

The SuperMix software was intended to be a practical, useable tool available for any research group developing new high-frequency, superconducting designs. It needed to be fast, efficient, and compatible with a wide range of computing platforms without requiring proprietary, commercial software to support it. We decided that the most straightforward way to achieve this goal was to create a library of routines which could be compiled to optimize execution speed. This choice eliminated interpreted software environments such as Python or Java, and also eliminated solutions such as application packages written for commercial environments such as Mathematica® or MatLab®. Consequently, we chose C++ which could be compiled using the widely-available, free GNU compiler *gcc* running under a Linux-type operating system.¹⁶ At the time of the original library development in 1997–1998, the *gcc* compiler implemented only a subset of the full C++ language specification, and, particularly importantly, it was missing a

¹⁶ GNU and Linux are copyright © the Free Software Foundation, a nonprofit [55].

complete and bug-free implementation of the C++ *Standard Template Library* (STL). Consequently, the SuperMix library incorporates several independently developed data structures (such as a complex data type) implementing functionality that was later incorporated into *gcc*. In most cases, the original SuperMix implementation has been retained, mainly because it works.¹⁷

Designed from the beginning to model and optimize high-frequency, superconducting circuitry and, in particular, heterodyne mixers using SIS devices, SuperMix owes much of its power to a few key numerical algorithms originally developed by others and which have been modified, often expanded and generalized, and then incorporated into the library. In particular, we must mention some of the most important features which set SuperMix apart from other design tools: (1) those algorithms used to model the electrical characteristics of thin film microstrip transmission lines constructed from superconducting metal layers (Caltech’s Mei Bin [16] provides a thorough description); (2) the connection algorithms used to combine the scattering and noise correlation matrices of individual components and sub-circuits into representations of larger, more complicated circuits (Caltech’s Scott Wedge [38]); (3) the SIS mixer small-signal and harmonic balance algorithms by John Tucker [1] and Withington and Kollberg [22].

An example illustrating SuperMix capabilities

Before jumping into detailed descriptions of the SuperMix code and its design, we provide an example of its modeling capabilities and numerical accuracy. ’Tis said, “The proof of the pudding is in the eating,” so let’s take a bite, shall we? As the first complete version of the library was receiving its finishing touches, Goutam Chattopadhyay, a SuperMix contributor and the author’s office mate at the time, was completing the testing of his primary doctoral thesis project, a 530 GHz, dual-polarization, quasioptical, heterodyne receiver containing eight SIS junctions [23]. After he described his design to the author, the author then remarked that Goutam could convert the receiver into a single-polarization, balanced design by properly orienting the linear polarizations of the RF and

¹⁷ “If it ain’t broke, don’t fix it.” – Anonymous.

LO signals and by using a 180° *hybrid* circuit to subtract the two polarizations' IF output phasors. This he did, thus constructing and testing the very first balanced heterodyne receiver operating at a submillimeter wavelength [39]. Given the complexity of its design and the thoroughness of Goutam's laboratory measurements, this instrument was destined to become the ideal test case for SuperMix.

Goutam's balanced mixer chip is diagrammed in Figure 5-1. Slot antennas in the ground-plane receive radiation incident from a direction normal to the plane of the figure. The E-field component of incident radiation which is perpendicular to the long axis of a slot induces a voltage across the slot. The top layer circuitry is designed to develop a "virtual short" to the outer edge of a slot at RF frequencies, so that the slot voltage induced by the sum of the appropriate E-field components of the RF and LO signals appears between the inner ground-plane and the wiring layer at the slot. The eight SIS junctions mix these RF- and LO-induced voltage signals, producing the four IF output

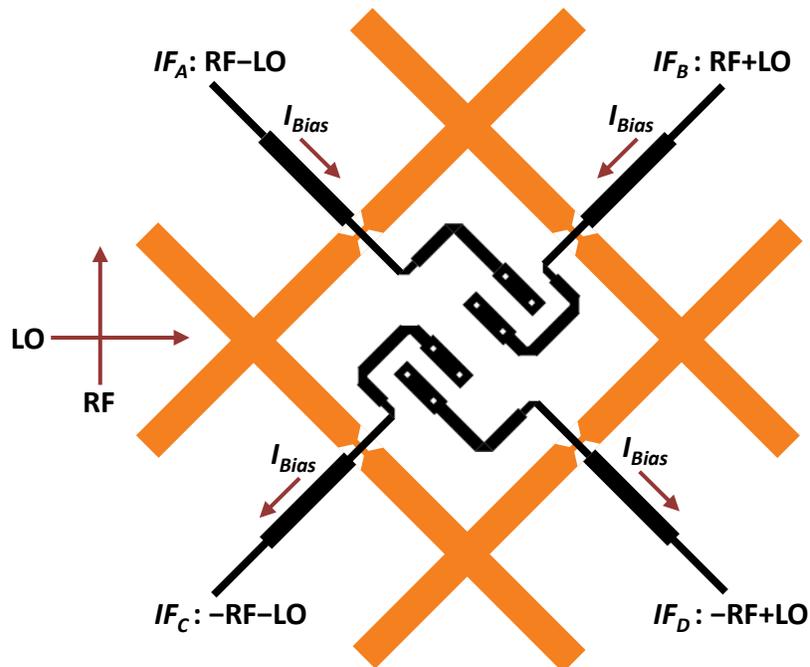


Figure 5-1: G. Chattopadhyay's quasioptical, balanced mixer design using eight SIS junctions. The orange area represents the four-slot antenna configuration. The top wiring layer is diagrammed in black. The small white squares in the wiring layer are the SIS junctions; the white background is the circuit ground-plane. The LO and RF signal phasor polarizations are shown as arrows, as are the SIS DC bias currents through the wiring layer. Deposited on a silicon substrate, the slot antenna lengths are each 0.2 mm for this 530 GHz design.

signals shown. The IF output wiring also serves at the SIS DC bias source and sink, the positive DC terminal applied to IFs A and B, and the negative terminal to IFs C and D. The inner ground plane area between the slots completes the DC circuit through the eight SIS junctions.

With the linear polarizations of the RF and LO signals as shown and the DC bias of the junctions as depicted in Figure 5-1, the four IF outputs have phasors related to the RF and LO modulations shown. Clearly, the LO modulation can be removed from the IF output by combining the four IF signals: $IF_{out} = (IF_A + IF_B) - (IF_C + IF_D) \propto RF$. This is the essence of a balanced mixer design: the final IF output has no contribution from any LO modulation (noise in the LO channel) [39]. To match the SIS junction impedances to the slot antennas, each slot antenna RF circuit uses two quarter-wave transmission line impedance matching sections and a small transmission line section between each pair of SIS junctions to tune out their capacitances. In addition, each IF output has a multi-section RF block circuit (not shown), and a combination of two transmission line segments and two capacitors make up the 180° hybrid which combines the four IF outputs into a single, balanced IF signal.

A SuperMix model of the entire superconducting RF-IF mixer circuit shown in Figure 5-1 (including the eight nonlinear SIS junctions), the RF blocks, IF hybrid, and optical components of the receiver was implemented by the author with Goutam's assistance; his thesis provides a copy of the rather long C++ source code listings of that model [23]. Several of the physical parameters of a thin-film, superconducting circuit cannot be precisely controlled during fabrication: the SiO insulating layer thickness and its dielectric constant are among these, at least that was the case in 1997–1998. Additionally, the receiver LO signal was generated by a Gunn diode source with no phase lock to a precision frequency reference and no way to accurately measure its output power. All of these uncertainties contributed potential inaccuracies to the SuperMix model.

The measured SIS DC I-V characteristic of the mixer chip could provide accurate determinations of the SIS gap voltage and normal resistance, however, along with its heterodyne mixing performance, so that was a start. A simple SuperMix program read in measured unpumped (no LO) SIS DC I-V characteristic data of the combined, eight-junction array. It accurately determined the average SIS V_{gap} (2.684 mV) and R_n

(12.55 Ω) as well as the small measurement offsets in the data; it also built the two normalized I-V files (*iv.dat* and *ikk.dat*) needed by the nonlinear mixer models.

To determine several other important physical characteristics of the materials used in the mixer chip, Goutam's measured *Fourier transform spectrometer* (FTS) data were analyzed. The FTS instrument measures the frequency response of the receiver by monitoring the changes in the SIS current produced by a wide-bandwidth, modulated-temperature source passed through a Michelson interferometer. Its theory and use are described in Mei Bin's thesis [16]. The program simultaneously analyzed a measured SIS pumped I-V (LO applied) data set. The full model of the balanced receiver was used to generate predicted FTS and pumped I-V responses, and a few key uncertain physical parameters of the model were adjusted to best match the model predictions to the measured data (in addition to accurately determining the pumped I-V LO frequency and power). The results are shown in Figure 5-2.

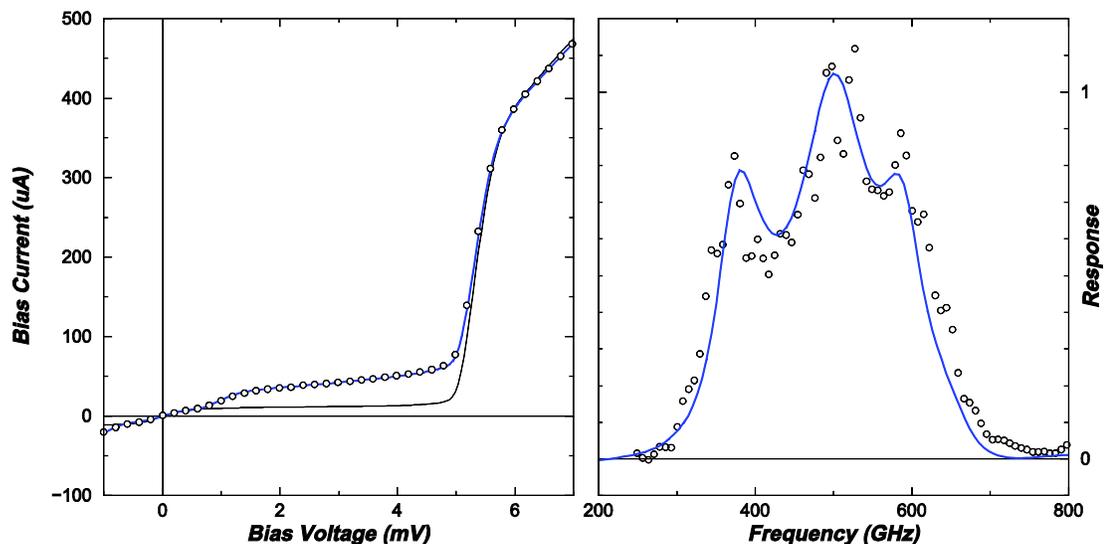


Figure 5-2: Fit of the mixer model to measured pumped I-V (left) and FTS (right) data. The blue curves are the model predictions; the data points are the black circles. The unpumped I-V model is also shown on the left graph (black curve). By adjusting a few key physical characteristics of the materials used in the mixer chip, the SuperMix model of the eight-junction, balanced receiver was able to quite accurately predict its measured performance. The vertical scale of the FTS response is arbitrary.

Of note are the optimized values of the SuperMix circuit model parameters:

SIS junction capacitance: 153.9 fF	SiO dielectric thickness: 242.5 nm
Nb film resistivity: 5.127 $\mu\Omega$ cm	SiO dielectric constant: 5.682
LO power: 231 nW	LO frequency: 525.8 GHz

This was the first demonstration that a full, nonlinear computer model could accurately determine several uncertain material characteristics of a complicated, superconducting, thin-film mixer chip with so few laboratory measurements. To prove that the resulting model was accurate, model predictions of the receiver IF output power generated by hot and cold RF broadband thermal sources (as a function of SIS DC bias voltage) were compared to actual measurements. The results shown in Figure 5-3 clearly demonstrate that SuperMix is an accurate tool for SIS mixer modeling and optimization. It compares measured data to that predicted by the SuperMix receiver model. Note that the match is really very good, except for SIS bias voltages between about 2.0 and 2.5 mV. In this range the measured data shows additional IF output power generated by AC *Josephson tunneling* of Cooper pairs of electrons generating a spurious RF signal within

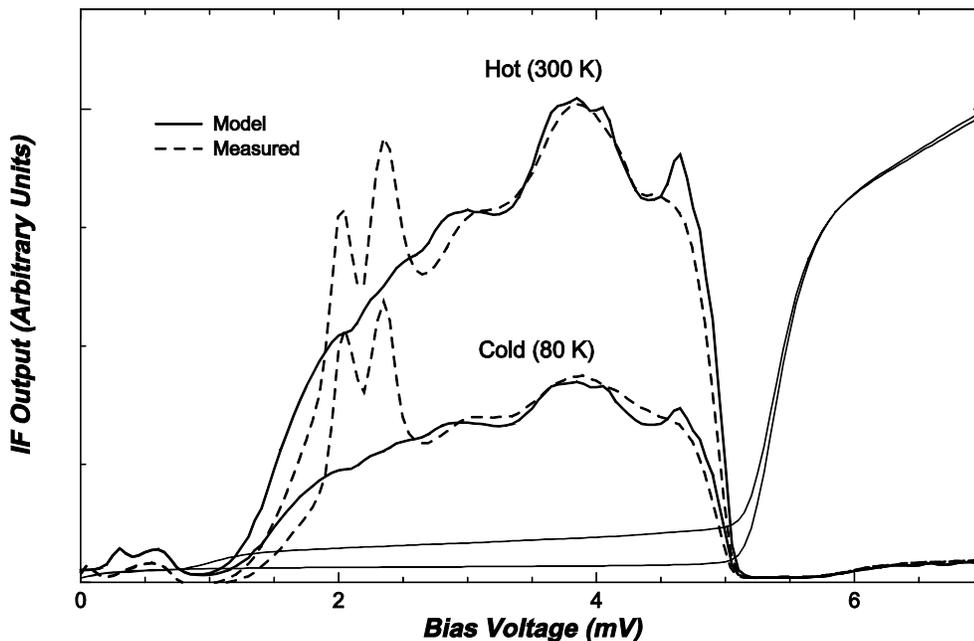


Figure 5-3: Predicted and measured IF output power vs. RF source temperature and SIS bias. The results speak for themselves: SuperMix works.

the receiver's RF bandwidth. This spurious signal can be suppressed by a properly applied magnetic field within the SIS junctions' insulating barrier; the SuperMix simulation of Tucker's quasiparticle tunneling theory does not include this effect.

Circuit modeling and design optimization

Now to delve into some details of how SuperMix does what it does and how to use it. The rest of this chapter and its associated appendices concentrate on a selection of the author's particular contributions to the library, but first priority must be given to the overall philosophy and structure of the SuperMix data structures and algorithms used for circuit modeling, nearly all of which are solely the results of the creativeness, intelligence, and hard work of John Ward [35] [15]. This section briefly describes his creations, along with the circuit optimization implementation created by Jonas Zmuidzinas.

Calculating and retrieving a circuit's response

A fundamental mathematical model of a circuit's input-output behavior in response to electromagnetic signals is to use a *wave representation* of the signals and a *scattering matrix representation* of the circuit. This is by far the most elegant and commonly-used approach for modeling high frequency circuits and structures, and it was the clear choice for SuperMix. The basic mathematics of this representation are described in Appendix B starting on page 116. Electromagnetic waves enter and leave a circuit at all of its several *ports*. The representation describes the numerical relationships between these incident and outgoing waves. A Fourier-space model is used, and each harmonic wave's amplitude and phase are captured using a single complex-valued *phasor*. The data structure of the representation of a circuit with n ports (n a positive integer) requires three objects: its $n \times n$ *scattering matrix* \mathbf{S} , its $n \times n$ *noise correlation matrix* \mathbf{C} , and its n -element *source vector* \mathbf{b}_s ; all are functions of frequency.

Any circuit or component modeled by SuperMix is stored in a C++ object whose type has inherited the capability to calculate and return a numerical data structure containing these arrays describing its behavior. Executing a circuit object's `get_data()` member function tells the object to recalculate its response using the settings of the global

parameters and then return that response in a data structure of type `sdata`. This structure contains member matrices `S` and `C` and a source vector `B`. The implementation of these complex-valued vectors and matrices is described in a later section. The following C++ code fragment shows the basic method (as a rule, reading code can be tedious and unproductive, but give it a go anyway):¹⁸

```
// Model will be an object of type circuit (discussed later)
circuit Model;
    // more code goes here to describe what the circuit contains

// These global parameters hold the conditions for the calculation
// units are also discussed in a later section
device: : f = 6.0*GHz;    // the frequency
device: : T = 4.2*Kelvin; // the temperature
device: : Z0 = 50*Ohm;   // the normalizing impedance

sdata Response = Model.get_data(); // the calculated behavior

// output the complex-valued element S21 of the circuit's resulting S matrix
cout << Response.S[2][1]; // C-style syntax for indexing an element of S
```

Building circuits from sub-circuits

Once the abstract class types had been defined to implement components which could calculate and report their individual scattering representation responses, John Ward had to add a scheme which would enable a SuperMix user to build up complicated circuit structures iteratively from sub-circuits or individual components. The solution he developed is very flexible, capable, and computationally efficient: the SuperMix `circuit` class.¹⁹ An appendix of John's thesis [15] describes the inner workings of this class, but suffice it to say that when its `get_data()` is called, a `circuit` object efficiently applies Wedge's connection algorithms [38] to an internal, hierarchical data

¹⁸ C++ (or C) keywords and code snippets will be highlighted by using a `console` font. The “//” starts a comment which continues to the end of the current line. Comments are in a `colored` font to offset them from the code.

¹⁹ Actually, his was an iterative approach. As John evolved and refined his implementation of the SuperMix `circuit` class, he adapted the basic n -port classes and their implementation to enhance their overall efficiency and usability. His efforts paid off: following the SuperMix version 1.0 release, his implementation of the n -port and circuit classes has remained unchanged. Well done (or, as we old sailors would say, *Bravo Zulu*).

structure of its constituent components and sub-circuits. The object first builds that data structure as its `connect()` and `add_port()` member functions are called to assemble the circuit.

Use of the circuit class is best illustrated by an example: see the section *Building a circuit* on page 193 of Appendix I. That section provides example code to model the SIS mixer chip RF matching network. The following section of that appendix then shows how to use SuperMix to construct models of the superconducting transmission lines used as elements in the mixer chip circuitry. Jonas Zmuidzinas implemented much of the sophisticated numerical modeling for physical structures such as the conductivity calculations for superconductors and the propagation properties of microstrip transmission lines.

Parameters and optimization

Successfully modeling a high-frequency circuit is only the first step. The real power of SuperMix lies in its ability to optimize the circuit design by adjusting the values of parameters describing the circuit's components, such as the length and width of a microstrip transmission line component. To accomplish this, the optimizer routines must have read and write access to these numerical parameters. SuperMix accomplishes this through its `parameter` classes. Most of the numerical values needed to specify an element's electrical behavior, such as a resistor's resistance value or a transmission line's length, are defined using member variables of class `parameter`. In normal use, these variables are interchangeable with and behave just like the C++ type `double`. They have greatly-enhanced capabilities over a simple `double` variable, however. For example, they can perform optional bounds checking and limiting of their values to physically realistic ranges. A `parameter` object can also be assigned to refer to another abstract, parameter-like variable, so that its value mimics that of the referenced object. Both of these capabilities are particularly useful when optimizing a design.

Given the `parameter` class as the means for adjusting circuit parameters, the SuperMix optimization scheme (created by Jonas Zmuidzinis with additions by John Ward) is implemented using the following sequence of steps (in a user's code):

1. Identify or define `parameter` variables for each of the various circuit parameter values to be optimized. The optimizer will use these variables to control the circuit behavior (length and width of a microstrip, for example).
2. Declare error term objects, each of which returns the square of some aspect of the circuit's behavior away from a target value (such as noise temperature, gain, input match, etc.). Define swept operating ranges over which these error terms will be averaged (usually frequency ranges).
3. Declare an object of class `error_func`, an error function which can also set the values of circuit parameters and calculate a weighted average of the resulting error term values over their swept ranges.
4. Declare an object derived from class `minimizer` and pass it the error function. SuperMix provides two predefined minimizer classes: `powell`, which is an efficient, multi-dimensional, local minimum finder; and `montecarlo`, a fairly capable random global minimizer. Or you can write your own.
5. Call the minimizer's member function `minimize()`. It will communicate various choices for the parameter values to the error function and then examine the error function's result. It will iterate until it optimizes the parameter values.

SuperMix provides a large set of predefined error terms so that a user can quickly define quite complicated error functions. Defining custom error terms is also fairly straightforward. An example of a SuperMix optimization using its predefined classes is given in the section *Using the optimizer* in Appendix I on page 197. This example offers a complete program which can be compiled and run.

Basic numerical data structures, linear algebra, and units

John Ward’s implementation of component and circuit modeling required efficient and flexible implementations of complex-valued vectors and matrices. The author developed these structures concurrently with John’s efforts.

Complex numbers

Amplitudes and phases of harmonic signals are best represented using complex numbers. The SuperMix implementation of complex-valued quantities uses a pair of C++ type `double` variables packaged into the class `Complex`. The operator “overloading” capabilities of C++ were fully employed to implement standard mathematical operations on this complex number representation, so for complex-valued variables `x`, `y`, and `z`, and real-valued (`double`) variables `a` and `b`, expressions such as

```
Complex z, x(5.3, 0), I(0, 1); // initialize to (real, imag), else 0
z = x + y;   z = x * y;   z = cos(x);   z = pow(x, y);   z = conj(x);
z = a + I*b;   a = real(x);   b = arg(x);   a = abs(x);   // etc.
```

work as expected (the SuperMix `Complex` constant `I` is actually predefined). Also as expected, the code will automatically cast a `double` or an `int` to type `Complex` as required. Finally, there are rather complete sets of formatted input and output routines that the user can configure as needed to delimit the two parts of a complex value, which can be input or output in either Cartesian or polar representation (with phase in radians or degrees).

Vectors and Matrices

As the SuperMix complex number representation was being completed, the author implemented the representations of complex-valued vectors and matrices required for circuit calculations. These calculations require flexible data structures—in particular, vectors and matrices of potentially unlimited capacities, as well as lookup tables and traditional C-style arrays. Consider first the representation of vectors of numerical values. Because the nonlinear device calculations will often involve (usually infinite) Fourier sums (series) of complex-valued coefficients, the index range of a vector may be required to extend to positive or negative infinity. Storing all of these element values would require a large chunk of computer memory! The (maybe obvious) solution chosen was to allocate enough memory to hold nonzero values for elements whose index was within a

specified range of 0. Outside of that specified range, the vector was assumed to consist of elements whose values equaled 0. This maximum index range (for nonzero element values) could be pre-allocated when a vector object was created, but could then be changed dynamically under program control during execution. Because of the range of index values required for various programming needs, SuperMix vector classes support three different indexing modes: `Index_C`, the traditional “C-style” indexing mode from 0 to $n-1$ for n elements; `Index_1` for 1 to n indexing; and `Index_S` for $-n$ to $+n$ indexing of $2n+1$ elements. By implementing controlled element access, any indexing operation which falls outside of a vector’s allocated range returns an element value of 0, and not an error. By default, SuperMix creates an empty vector object, which is not very useful until reallocated. Also by default, vectors have `Index_1` indexing mode. A vector’s index mode as well as its size may be changed dynamically during program execution.

SuperMix defines class `Vector` to hold complex-valued elements and class `real_vector` for elements of type `double`. Thus the declarations

```
Vector A, V(10), Y(20, Index_S);
real_vector y(16, Index_C);
real_vector x(y);
```

define three new complex-valued vector variables: `A` with mode `Index_1` and the default allocation size 0, `V` also with mode `Index_1` and allocated to hold 10 nonzero elements, and `Y` which can hold 41 elements (valid index range from -20 to $+20$). The real-valued vector variable `y` is initially allocated to hold 16 nonzero elements. Finally, `x` is created and initialized to be an independent, identical copy of `y`. A rich suite of basic operations on vector objects are included in the SuperMix library. Here are a very few examples (`v` is of type `Vector`, `s` is a `double` or `Complex` value, `i` and `j` are `int` index values, the function `f()` returns a `double` or `Complex` value):

```
v.get(i) = s; // set the single indexed element of v to s
v = s;      // set all allocated elements of v to s
v /= s;     // scale the allocated elements of v
s = v.get(i); s = v.read(i); // set s to the indexed v value
v.swap(i,j); // exchange the values of these indexed elements
v.unit(i);  // a unit vector: set v[i] to 1, all others to 0
v.apply(f); // for a function f(), set v[i] = f(v[i]) for all i
            // or, if f() is so defined, v[i] = f(v[i],i) for all i
```

The above “unary” vector operations can only change a vector’s allocated elements.

Conceptually, SuperMix vectors have an infinite number of elements, all of which vanish outside the allocated index range. Thus, for example, a binary vector operation such as vector addition will always be valid, regardless of the vectors' allocations. Binary vector operations return a result which is appropriately sized to include all resulting nonzero elements. This is a general rule regarding how the various vector and matrix operations are defined. Vector assignment will, if necessary, reallocate the left-hand side vector so that it holds the complete result of an operation (**u**, **v**, and **w** are of type **Vector**, **s** is a **double** or **Complex** value, **i** and **j** are **int** index values):

```
v = w; w = u + v; v += u; w = real(v); v = conj(u);
u = scalmult(v, w); // u[i] = v[i]*w[i] for each index value i
s = dot(v, u); // an inner product: the sum of conj(v[i])*u[i]
s = v * u; // the same as s = dot(v, u)
s = norm(v); // the same as s = dot(v, v)
```

Of course, it is also possible to loop over just the allocated vector elements:

```
for( int j = w.minindex(); j <= w.maxindex(); ++j) w[j] += 2*j; .
```

Note the traditional, C-style indexing of the vector elements. If no memory has been allocated for **w**, then it is guaranteed that **w.maxindex()** will be less than **w.minindex()**, so such a looping construct will not access any unallocated elements.

SuperMix matrix classes are straightforward extensions of the vector classes. Class **Matrix** holds complex-valued elements. The two matrix indices (row and column) can have different index modes. Matrices may also be resized during program execution, and, as with vectors, matrix operands are padded with zero-valued elements whenever required to make operations valid. For example, given complex-valued matrices **A**, **B**, and **C** (not necessarily square), complex-valued vectors **x** and **b**, real-valued scalar **s**, and integer **n**:

```
A = identity_matrix(n, Index_S); // diagonal elements = 1, others = 0
C = A * B; // matrix multiplication, which will always work
A += s; A = real(B); A = transpose(B); A = dagger(B);
A = Inverse(B); // so that A*B == B*A == identity matrix
x = row(n, A); B = columnmatrix(x); A = rowmatrix(b);
b = A * x; b = x * B; // matrix-vector multiplication (dot products)
B = solve(A, C); // so that A * B == C
```

If, for example, an inverse of a matrix does not exist, then the result of the operation has no allocated elements (is empty). As one might expect, other similar operations are defined to provide a basic, but reasonably complete, linear algebra capability. Special,

fast, optimized routines are also available to speed up calculations needed for SIS harmonic balance determination, etc.

Physical units and constants

The numerical values stored in the elements of SuperMix scalars, vectors, and matrices usually represent physical quantities with associated units such as meters, coulombs, volts, and hertz. The author added a very simple but convenient implementation to SuperMix in order to accommodate units and conversions between compatible units. The internal representations of physical quantities are scaled to be appropriate for high-frequency structures: gigahertz (GHz), microns (μm), etc. Units are then implemented using constant, real-valued numbers which can be used to properly adjust input or output quantities to these internal scales. For example, various physical quantities may be defined as

```
double Vbias = 2.3*Milli*Volt, rho = 5.0*Micro*Ohm*Centi*Meter;
Complex Z = (3.2 + I*4.5)*Ohm;
double g = 9.8*Meter/(Second*Second);
double phase1 = -92.7*Degree, phase2 = Pi/4;
```

Of course, input values may be scaled in the same way: multiply the input value by the units. Useful physical constants are also predefined by the SuperMix library:

```
// electrostatic potential of an electron at distance r:
double phi(double r){
    return -eCharge/(4.0*Pi*epsVacuum * r);
}
```

To convert calculated numerical values into appropriate units for output, one simply divides by the desired units instead of multiplying (continuing the above example):

```
double v = phi(2.3*Nano*Meter); // potential 2.3nm from an electron
cout << v/Volt; // output the calculated potential in volts
```

Unfortunately, there is no unit compatibility checking in SuperMix—that responsibility is left to the user. As long as all defined physical values, input data, and output results include units as described above, then numerical results should be correctly scaled to the desired physical units.

Representing a heterodyne receiver

A fully nonlinear model of a heterodyne mixer circuit using SIS junctions is quite complicated with lots of parts which must mesh well for its calculations to be efficient and accurate. This section describes the author's implementation of the SuperMix `mixer` class and of its SIS model elements. SuperMix models a heterodyne receiver by using a special C++ class object to interface two very different types of elements: (1) standard linear circuits with scattering matrix representations which are functions of frequency, and (2) two-terminal, nonlinear elements (SIS junctions) which require operating state calculations and which can mix signals at different frequencies (e.g. RF and LO inputs) to produce new signals at frequencies which are linear combinations of their input frequencies (e.g. IF outputs). The SuperMix library class `mixer` manages the interface between these distinct types. It first performs large-signal *operating state* calculations for the nonlinear elements and then uses these results to generate the small-signal, heterodyne mixing performance of the linear + nonlinear receiver system.

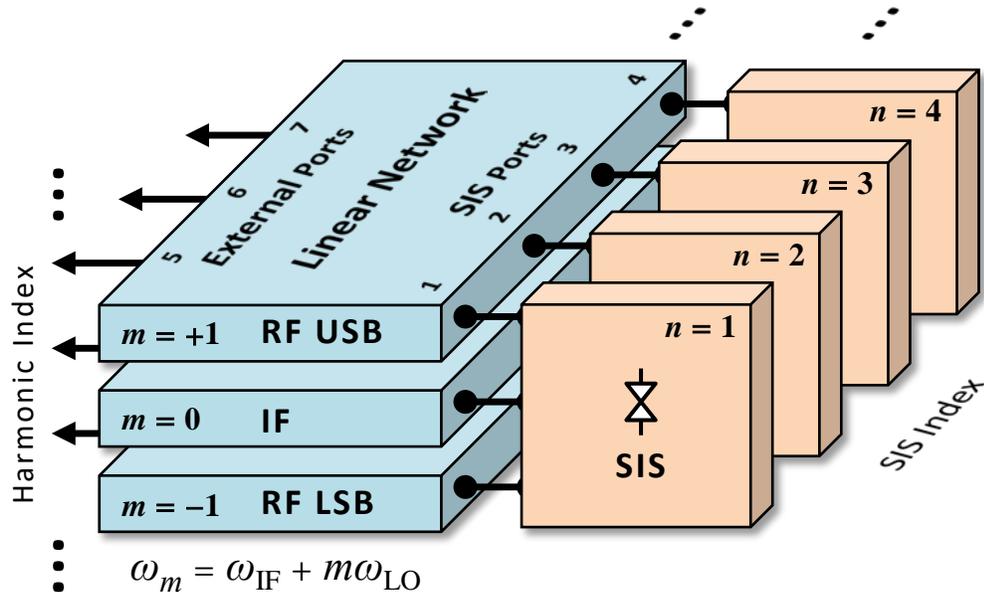


Figure 5-4: The top-level circuit topology of a multi-SIS, heterodyne detector. The linear circuitry can only connect signals at the same frequency. The linear circuit response (**S** matrix and **C** matrix) must be evaluated at each frequency ω_m . The SIS junctions' small-signal admittance matrix elements $Y_{m,m'}$ connect the signals at different frequencies, each at one port of the linear circuit.

For a given output IF frequency ω_{IF} and LO frequency ω_{LO} , the signal present at the IF output of a heterodyne system is a linear combination of the signals present at a potentially large set of input frequencies: DC ($\omega = 0$) and $\omega_m = m\omega_{\text{LO}} + \omega_{\text{IF}}$ where m is any integer. Of course, $m = 0$ corresponds to the IF; $m = \pm 1$ represent signal frequencies in the *upper and lower RF sidebands* of the LO frequency (*USB* and *LSB*), and these are usually the primary RF signals of interest. The linear circuitry of the heterodyne receiver will respond to signals at each of these frequencies independently of the presence of signals at other frequencies—only the nonlinear (SIS) elements can transfer power between the various signal frequencies. The top-level, conceptual topology of the heterodyne receiver circuitry (ignoring DC) is shown in Figure 5-4. The SuperMix `mixer` class calculates the various frequencies and manages the port and frequency connections in order to determine the modeled receiver's response and its expected noise. It is designed to handle any number of independent SIS junctions and any number of harmonics m of the LO frequency.

Modeling a single SIS device

Before discussing the `mixer` class in detail, we first describe the model of a single SIS junction. A SuperMix SIS model is encapsulated in its `sis_device` class, which cannot be used as a circuit element in the same way as linear elements like resistors or transmission lines. It is intended to only be interfaced to the rest of the heterodyne receiver circuitry using the `mixer` class. To accurately calculate the SIS behavior when used as a heterodyne detector, its *operating state* must first be determined. An SIS device model has an operating state determined by a set of harmonically related voltages: a complex-valued vector of RMS voltage phasors at nonnegative, integer multiples of a single local oscillator (LO) frequency: $m\omega_{\text{LO}}$ (a real-valued SIS DC bias voltage is also included as the element with $m = 0$). When used as part of a receiver circuit, a mixer object will supply an operating state vector as an argument to each `sis_device` object's `large_signal()` member function, along with the LO frequency and the number of harmonics to use in its analysis. This member function then initiates an SIS operating state calculation to determine the corresponding vector of harmonic current phasors through the junction.

As pointed out in Chapter 2, the dramatic nonlinearity of the SIS device I-V characteristic at scales of a fraction of a millivolt requires that a quantum theory be used to describe its electrical behavior (Figure 2-3 on page 22). Withington and Kollberg's extension of John Tucker's quantum mixing theory provides the solution [22] [1]. The nonlinear system of equations which determines the SIS operating state is encoded in `sis_device` and its associated classes and is described in detail in this text's Appendix E on page 155 and sections of Appendix I starting on page 199.

The SIS device I-V characteristic and its analytic extension (using a Kramers-Kronig transform as described in Appendix E) must be input to a SuperMix `sis_device` object as a pair of data files. The SuperMix class `ivcurve` uses these files to build a spline interpolation of their data for use during `sis_device` calculations. The input I-V data files are usually normalized using the SIS device *gap voltage* and *normal resistance*, parameters which set the voltage and current scales of the SIS I-V. For the niobium-based

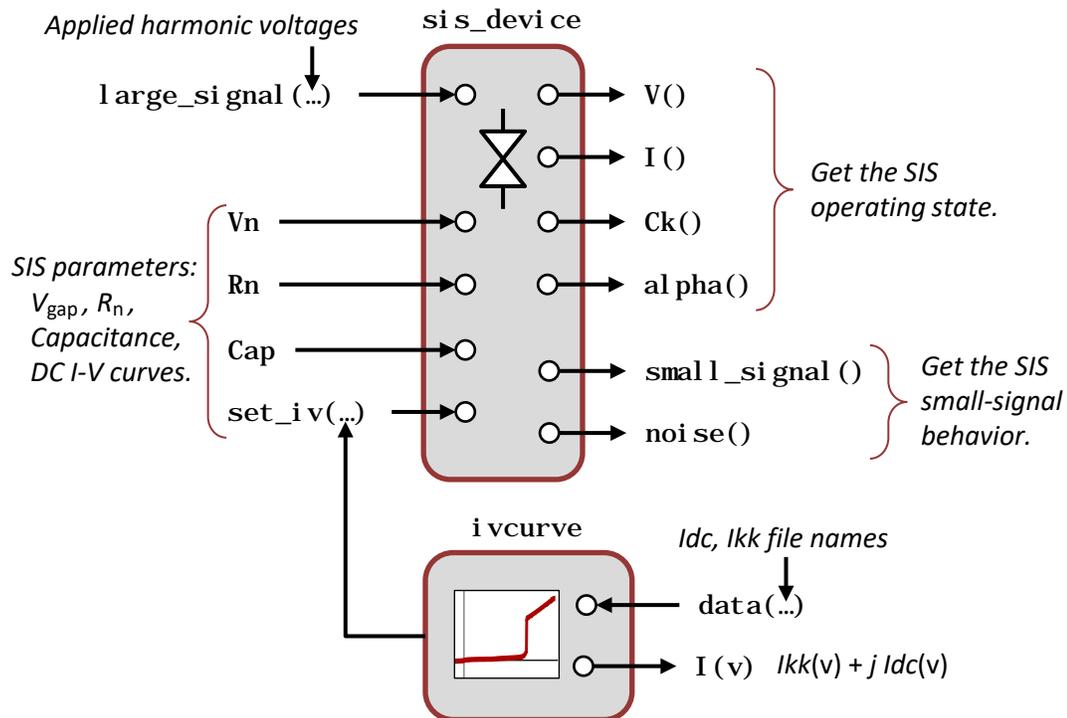


Figure 5-5: SuperMix representation of a nonlinear SIS as a heterodyne detector. The `sis_device` class encapsulates the numerical routines required to apply Tucker's quantum mixer theory to the determination of the SIS operating state and its resulting heterodyne mixing small-signal admittance matrix.

SIS device last used with the subject receiver, these parameter values were 2.79 mV and $7.7\ \Omega$, respectively. In addition, the closely-spaced superconducting terminals of the SIS tunnel junction form a capacitor whose displacement current provides a parallel conduction path through the device. This capacitance, on the order of 0.15 pF for the receiver's SIS, is therefore another important characteristic of the device which must be included in the `sis_device` model. Once all of these parameters have been specified and the SIS operating state has been determined, `sis_device` can then calculate its heterodyne detector *small signal admittance matrix* and *noise current correlation matrix* using the quantum mixing theory. Figure 5-5 provides a graphical summary of the `sis_device` external interface (parameters and member functions) through which these capabilities are accessed.

Some code to show how to declare and set up a single SIS junction element:

```
// Declaring and defining an SIS device model:
sis_device SIS;
ivcurve IV;
IV.data("iv.dat", "ikk.dat"); // read in the named IV data files
SIS.set_iv(IV);
SIS.Vn = 2.79*mV; SIS.Rn = 7.7*ohm; SIS.Cap = 150*Femto*Farad;
```

Building a heterodyne mixer model

A submillimeter-wave heterodyne receiver system's linear circuitry is required to handle a very broad range of frequencies: the RF and LO are both at hundreds of GHz, whereas the IF frequency is nearer to 10 GHz. In addition, active components require DC power to bias them and determine their operating states. The complicated receiver circuitry will usually exhibit very different behaviors in these three disparate frequency regimes. Consequently, the receiver circuitry may be best represented by different models, one applicable to each frequency range. The `mixer` class expects that three different linear circuit objects will be used to represent this circuitry. Figure 5-6 on page 99 diagrams the `mixer` class interface and its primary functions.

The `mixer` class member function `set_rf()` identifies the model to be used for the linear circuit at the LO and the various RF frequencies; `set_if()` identifies the model for IF frequencies, and `set_bias()` identifies the DC bias model of the circuitry. The user first builds the three linear circuit models from individual elements using class `circuit` as described earlier in this chapter. The names of these circuit objects are then

passed as arguments to the corresponding `mixer` function calls. Each of these linear circuit models must have ports dedicated to the SIS devices (Figure 5-6). The RF circuit must also have at least one more port dedicated to receiving the RF signal input, and the IF circuit needs at least one port for the IF signal output.

Once `sis_device` objects are created for the receiver's SIS junctions, they are each identified to the mixer using its `add_junction()` member function. Functions `set_LO()` and `harmonics()` assign the LO frequency and set the number of harmonics to be used in its harmonic balance and small-signal analyses. The receiver IF frequency

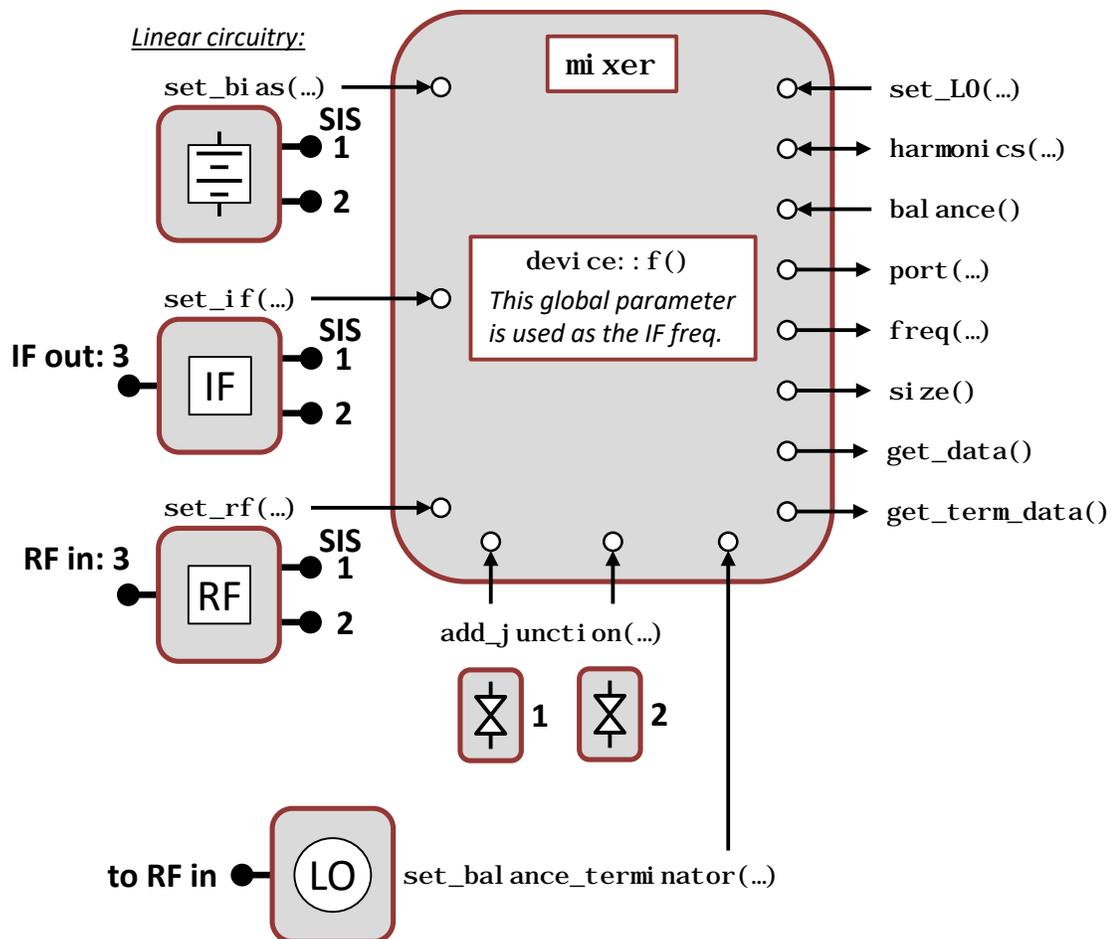


Figure 5-6: Representing a complete heterodyne system using class `mixer`. A `mixer` class object manages the interface between the system's linear circuitry and its nonlinear SIS junctions. It is used to perform both harmonic balance to set the SIS operating states and subsequent small-signal heterodyne receiver performance analyses.

will then be set using the value of the SuperMix global parameter `device : f`.

Here is an example of code to build a simple, single-SIS mixer model:

```
// Declare and build the RF, IF, and DC bias circuits:
circuit RF, IF, Bias;
    // code goes here to build the circuits...

// the mixer will use the SIS device defined in the previous example
mixer Mix;
Mix.add_junction(SIS); // Mix will attach the SIS to port 1 of each circuit
Mix.set_RF(RF).set_IF(IF).set_bias(Bias); // the calls can be cascaded
Mix.harmonics(3); // use 3 harmonics for the simulation
```

The first task of the mixer model is to set the operating states of the SIS junctions using its harmonic balance routines: `initialize_operating_state()` followed by `balance()`. The IF circuit model is not used for this calculation; only the RF circuit, which should also supply the LO source signal, and the DC bias circuit, which must supply the individual SIS DC bias voltages, are used. For harmonic balance calculations, the RF circuit's signal input ports are assumed to be terminated by passive terminations each with the SuperMix global characteristic impedance `device : Z0`; if this behavior is inappropriate, then individual RF port terminations may be specified using `mixer`'s `set_balance_terminator()` member function. For a simple, double-sideband receiver, the LO signal is often combined with the RF using optics external to the mixer RF circuitry. In this case the easiest way to incorporate the LO signal into the mixer model is to use a LO power source for the RF balance terminator (as indicated in the example of Figure 5-6). This was the method used when modeling and optimizing the wide bandwidth mixer chip for the subject receiver. More complicated receiver circuits such as balanced or sideband separating designs should incorporate the LO power source into the RF circuit model.

The mixer member function `initialize_operating_state()` is not shown in Figure 5-6, but it serves an important function nonetheless. Its default behavior is to treat the SIS junctions as open circuits and calculate the DC and LO voltages applied to each one (in other words, calculate the \mathbf{Z} matrices and source voltage vectors \mathbf{S}_v of the RF and DC bias circuits with the embedded LO and DC voltage sources included). These voltages are applied to the SIS junctions using their `large_signal()` member functions (Figure 5-5 on page 97). The applied voltages and the SIS-calculated currents

are then used as initial values for the mixer's `balance()` algorithm. The harmonic balance mathematics and algorithm used by `balance()` are described in Appendix E.

Continuing our example, define the mixer LO and perform a harmonic balance:

```
// We'll use a 1-port source for the LO and assign its frequency and
// signal power:
parameter fL0 = 230.0*GHz, pL0 = 100.0*Nano*Watt;
generator L0; // the 1-port LO power source
L0.source_f = &fL0; L0.source_power = &pL0; // now changes to fL0 or pL0 will
// be passed on to the LO source
Mixer.set_balance_terminator(L0).set_LO(&fL0); //again cascading function calls
Mixer.initialize_operating_state().balance(); // that's all there is to it
```

Once the harmonic balance has been completed, the mixer small-signal analysis can proceed. The mixer member function `get_data()` performs the tasks required to analyze the heterodyne circuit topology diagrammed in Figure 5-4 on page 95 and return the response scattering matrix **S** and noise correlation matrix **C** connecting the various RF sidebands and the IF output(s). The mixer retrieves each SIS junction's small-signal admittance matrix **Y** and current noise correlation matrix **C_I** by calling their individual `small_signal()` and `noise()` member functions. The RF circuit's response (**S** and **C** matrices) are calculated at each of the sideband frequencies $\omega_m = m\omega_{LO} + \omega_{IF}$, and the IF circuit response is calculated at ω_{IF} . All of these results are then connected into the sideband circuit topology of Figure 5-4 using an algorithm equivalent to John Ward's `circuit` implementation, but optimized for the mixer's unique sideband topology. The resulting **S** and **C** matrices are then returned by the `mixer` object's `get_data()` member function.

Determining the port index of the mixer response matrix for one of the RF sideband inputs or the IF output of the full mixer model is no trivial matter. Consequently, the `mixer` class provides the member functions `port()` and `freq()` to help a user find a correct port index for its calculated **S** and **C** matrices and the port's corresponding signal frequency. For example, refer back to Figure 5-6 on page 99. The RF input is at port 3 of the RF circuit model; the upper sideband (USB) input corresponds to harmonic index $m = +1$. The IF circuit model's output happens to be at its port 3 also, and the IF harmonic index is $m = 0$ ($\omega_m = m\omega_{LO} + \omega_{IF}$). These port indices are those of the RF and IF sub-circuits of the mixer, not the port indices of the full mixer circuit. The `mixer` function call `port(sub-circuit port #, m)` returns the corresponding port index of the full model.

For example, to get the USB to IF conversion gain from the mixer model's small-signal \mathbf{S} matrix, we can use the following code:

```
// Calculate mixer RF USB to IF gain (continuing previous examples).
sdata response = Mix.get_data(); // perform the small-signal analysis

int USBin = Mix.port(3, 1); // RF circuit port 3 at harmonic +1 (USB)
int IFout = Mix.port(3, 0); // IF circuit port 3 at harmonic 0 (IF out)

double gain = zmagsq(response.S[IFout][USBin]); // magnitude squared of S
double freq = Mix.freq(USBin); // the USB frequency given the LO and IF freqs
```

Including RF sideband quantum noise

As described in the final sections of Appendix C, a simple, double-sideband heterodyne mixer can amplify ground-state quantum fluctuations present at its various RF sideband inputs, adding at least an extra $\hbar\omega_{\text{LO}}/k_B$ of noise to the receiver output (USB and LSB). This amounts to an added noise temperature of 11 K at 230 GHz. Because this noise originates in quantum fluctuations in the electromagnetic fields of the RF sideband input channels, SuperMix does not include it in the heterodyne mixer noise calculations triggered by a `get_data()` call. As described in Appendix C, SuperMix includes this input-associated quantum noise by using the Callen-Welton passive noise formula (C.11) for one-port, dissipative terminations such as a matched, `device : Z0` terminator on a circuit input (the Callen-Welton formula is also applied to any passive, dissipative elements in the mixer circuitry).

The most straightforward way to get SuperMix to include the ground-state quantum noise in a mixer noise calculation is to assign a 0 K, matched (`device : Z0`) terminator to each open RF input port using the `set_balance_terminator()` member function; a 1-port `sterm` (found in SuperMix *elements.h*) could serve. If the LO source used for harmonic balance calculations has already been assigned as a balance termination, then one simply sets its temperature to 0 K. Now calling `get_term_data()` will return \mathbf{S} and \mathbf{C} matrices with only open IF ports; all RF sideband ports will be terminated with 0 Kelvin loads, and each will contribute noise representing ground-state quantum fluctuations to the IF outputs' noise. To then refer the IF output noise back to the appropriate RF sideband input in order to calculate the mixer noise temperature, simply divide by the sideband to IF power gain found using a mixer `get_data()` call.

Again continuing our example code:

```
// Calculate mixer RF USB to IF noise temp (continuing previous examples).
// The LO source has been defined as the terminator for the RF input port.
LO.Temp = 0.0; // now just ground-state quantum noise from the RF input ports

// Fetch the IF output noise from the get_term_data() noise matrix
double Nout = Mix.get_term_data().C[IF][IF].real; // just the real part
// divide by the mixer gain to get the input-referred noise temperature
double Tn = Nout/gain; // gain was calculated in previous example
```

An analogous method may be used to generate predicted Y-factor or hot/cold load responses by setting the RF termination's temperature to a desired value and returning the resulting IF output noise power using the mixer's `get_term_data()` member function:

```
// Calculate mixer IF responses to RF hot and cold source temps.
// The LO source is the terminator for the RF input port.
parameter T_hot(295*Kelvin), T_cold(80*Kelvin);

// Fetch the IF output noise resulting from each RF termination temperature
LO.Temp = T_hot;
double P_hot = Mix.get_term_data().C[IF][IF].real;
LO.Temp = T_cold;
double P_cold = Mix.get_term_data().C[IF][IF].real;

// The Y-factor is the ratio of these two output noise powers
double Y = P_hot/P_cold;
```

Chapter 6

CONCLUSION

By any measure, it is clear that the wide-band heterodyne receiver project has been a success. Intended to be an otherwise simple, bare-bones project to develop and demonstrate the tools and techniques required to design and fabricate advanced, optimized, superconducting mixers for millimeter and submillimeter astronomy, the receiver was capable enough to have a long, successful tenure as a CSO scientific instrument. The techniques and tools used to develop its SIS mixer have seen wide application throughout the terahertz community. The following sections provide citations of a variety of papers not otherwise used as references for the material in this document. Hopefully they will indicate the lasting scientific and technological impacts of these efforts.

Astronomy and planetary science

The majority of the CSO science using this instrument was performed by Dr. Matthew Sumner [8] (then a Caltech Ph.D. student, advisor Prof. Zmuidzinas; now with Sandia National Laboratories) and Prof. Susanna Widicus Weaver (first as a Caltech Ph.D. student, advisor Prof. Geoffrey Blake; now at Emory University) and members of her research groups [10] [11]. Although the original motivation for the instrument's design was to perform extragalactic carbon monoxide rotational emission line searches for redshift determinations [40], its astrophysical science impact has been mostly confined to galactic line surveys of protostellar regions (see again Figure 1-2 on page 5). Here is a list of pertinent papers:

Luyao Zou, Brian M. Hays, and Susanna L. Widicus Weaver, "Weakly bound clusters in astrochemistry? Millimeter and submillimeter spectroscopy of trans-HO₃ and comparison to astronomical observations," *J. Phys. Chem. A*, vol. 120, no. 5, pp. 657–667, 2016. <https://doi.org/10.1021/acs.jpca.5b09624>.

Brett A. McGuire, P. Brandon Carroll, James L. Sanders, III, Susanna L. Widicus Weaver, Geoffrey A. Blake, Anthony J. Remijan, "A CSO search for l-C₃H⁺: detection

in the Orion Bar PDR,” *Monthly Notices of the Royal Astronomical Society*, vol. 442, no. 4, pp. 2901–2908, August 2014. <https://doi.org/10.1093/mnras/stu1047>.

McGuire, Brett Andrew, *Time-Domain TeraHertz Spectroscopy and Observational Probes of Prebiotic Interstellar Gas and Ice Chemistry*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.)), 2015, doi: 10.7907/Z9B27S79. <https://resolver.caltech.edu/CaltechTHESIS:07232014-152145064>.

M. C. Sumner, G. A. Blake, A. I. Harris, M. Leong, T. G. Phillips, F. Rice, S. L. Widicus Weaver, H. Yoshida, and J. Zmuidzinas, “Millimeter line surveys of class-0 protostars and targeted searches for complex organics in high-mass star-forming regions,” *Protostars and Planets V*, 8550, 2005. <https://www.lpi.usra.edu/meetings/ppv2005/pdf/8550.pdf>.

Susanna L. Widicus Weaver, Matthew C. Sumner, Frank Rice, Jonas Zmuidzinas, Geoffrey A. Blake, “High-sensitivity broadband spectral line surveys of star forming regions with the CSO,” *International Symposium On Molecular Spectroscopy, 64th Meeting*, Ohio State University, June 2009, abstract id.RG01. http://molspect.chemistry.ohio-state.edu/symposium_64/symposium.

Michał Drahus, David Jewitt, Aurélie Guilbert-Lepoutre, Waław Waniak, James Hoge, Dariusz C. Lis, Hiroshige Yoshida, Ruisheng Peng, and Albrecht Sievers, “Rotation state of comet 103P/Hartley 2 from radio spectroscopy at 1 mm,” *ApJ. Letters*, vol. 734, no. 1, L4, May 2011. <http://dx.doi.org/10.1088/2041-8205/734/1/L4>.

Nadine Wehres, Susanna Widicus Weaver, Eric Herbst, Darek Lis, et al., “Detection, identification and correlation of complex organic molecules in 32 interstellar clouds using submm observations,” *Proceedings of the International Symposium on Molecular Spectroscopy*, Urbana, IL, June 2014. <https://doi.org/10.15278/isms.2014.RF10>.

SuperMix

The SuperMix circuit modeling and design software described in Chapter 5 was extensively used for the design and analysis of the author’s mixer chip. Subsequently, it has proved to be an important tool used during the design and optimization of nearly every Caltech and JPL instrumentation effort employing superconducting circuitry and

operating in the submillimeter and terahertz bands. Among the instruments whose designs involved SuperMix are the *SAMBA* bolometer array, the heterodyne instruments for the *Herschel Space Observatory*, transition edge sensor bolometer arrays for the *SPIDER* CMB polarimetry experiment, the upgraded CSO facility heterodyne instruments, and the *SuperSpec* millimeter wave grating spectrometer:

Alexey Goldin, James J. Bock, Cynthia L. Hunt, Andrew E. Lange, Henry G. LeDuc, Anastasios Vayonakis, and Jonas Zmuidzinas, “Design of broadband filters and antennas for SAMBA,” *Proc. SPIE 4855, Millimeter and Submillimeter Detectors for Astronomy*, February 2003. <https://doi.org/10.1117/12.459677>.

A. Karpov, D. Miller, F. Rice, J. A. Stern, B. Bumble, H. G. LeDuc, and J. Zmuidzinas, “Low noise 1 THz–1.4 THz mixers using Nb/Al-AlN/NbTiN SIS junctions,” *IEEE Trans. Applied Superconductivity*, vol. 17, no. 2, pp. 343–346, June 2007. <https://doi.org/10.1109/TASC.2007.898277>.

C. L. Kuo, J. J. Bock, et al., “Antenna-coupled TES bolometer arrays for CMB polarimetry,” *Proc. SPIE 7020, Millimeter and Submillimeter Detectors and Instrumentation for Astronomy IV*, 70201I, July 2008. <https://doi.org/10.1117/12.788588>.

J. W. Kooi, R. A. Chamberlin, R. Monje, B. Force, D. Miller and T. G. Phillips, “Balanced receiver technology development for the Caltech Submillimeter Observatory,” *IEEE Transactions on Terahertz Science and Technology*, vol. 2, no. 1, pp. 71–82, January 2012. <https://doi.org/10.1109/TTHZ.2011.2177726>.

Attila Kovács, Peter S. Barry, Charles M. Bradford, et al., “SuperSpec: design concept and circuit simulations,” *Proc. SPIE 8452, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VI*, 84522G, September 2012. <https://doi.org/10.1117/12.927160>.

SuperMix has also been adopted by many research groups around the world for their design efforts, and, as of this writing, decades after its initial release, SuperMix continues to be widely used for modeling and optimization. Here are included a few citations describing recent efforts for which it served as a design tool:

Paul Grimes, Ghassan Yassin, Phichet Kittarat and Stafford Withington, “Analysis of subharmonic SIS mixers using SuperMix,” *Proceedings 17th International Symposium on Space Terahertz Technology*, Paris, France, pp. 94–97, 2006.

https://www.researchgate.net/publication/252900217_Analysis_of_subharmonic_SIS_mixers_using_SuperMix.

(Oxford, UK; Mahidol University, Bangkok, Thailand; Cambridge, UK)

B. I. Ivanov, M. Trgala, M. Grajcar, E. Il'ichev, and H.-G. Meyer, “Cryogenic ultra-low-noise SiGe transistor amplifier,” *Rev. Sci. Instrum.*, vol. 82, 104705, October 2011.

<https://doi.org/10.1063/1.3655448>.

(Institute of Photonic Technology, Jena, Germany; Novosibirsk State Technical University, Russia; Comenius University, Slovakia)

B. Tan, G. Yassin, P. Grimes, J. Leech, K. Jacobs and C. Groppi, “A 650 GHz unilateral finline SIS mixer fed by a multiple flare-angle smooth-walled horn,” *IEEE Transactions on Terahertz Science and Technology*, vol. 2, no. 1, pp. 40–49, January 2012. <https://doi.org/10.1109/TTHZ.2011.2177736>.²⁰

(Oxford, UK; Harvard-Smithsonian Center for Astrophysics; I. Physikalisches Institut, University of Cologne, Germany; Arizona State University)

(Oxford, UK; Harvard-Smithsonian Center for Astrophysics; I. Physikalisches Institut, University of Cologne, Germany; Arizona State University)

J. D. Garrett, B. K. Tan, F. Boussaha, C. Chaumont and G. Yassin, “Simulating the Behavior of a 230-GHz SIS Mixer Using Multitone Spectral Domain Analysis,” *IEEE Transactions on Terahertz Science and Technology*, vol. 9, no. 6, pp. 540–548,

November 2019. <https://doi.org/10.1109/TTHZ.2019.2938993>.²¹

(Oxford, UK; GEPI, Observatoire de Paris–PSL, Paris, France)

²⁰ Several of the authors listed here have published quite a few other papers describing mixers designed using SuperMix. This citation provides a relatively recent example.

²¹ In this quite recent paper the authors describe an SIS nonlinear simulation package called QMix, which is designed to handle multiple large-signal inputs, rather than a single LO signal used by SuperMix. They spend several sentences contrasting their software and SuperMix. Finally, something new!

Waveguide probe

The waveguide probe design described in Chapter 3 has found wide application throughout many of the world’s submillimeter and terahertz technology communities. The many citations of Jacob Kooi et al.’s 2003 announcement and description of the design [26] may be useful to gauge this interest—as of this writing *Web of Science* lists 54 citations and *Google Scholar* lists 99, the latest in 2020 and a dozen of which were published in 2017 or later. Among the more interesting or representative papers wherein a waveguide probe based on our design was described:

H. Rashid, S. Krause, D. Meledin, V. Desmaris, A. Pavolotsky and V. Belitsky, “Frequency multiplier based on distributed superconducting tunnel junctions: theory, design, and characterization,” *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 5, pp. 724–36, September 2016. <https://doi.org/10.1109/TTHZ.2016.2583201>.

J. Cheron and E. Grossman, “High gain 220GHz power amplifier MMICs with minimal footprint,” *2016 IEEE MTT-S International Microwave Symposium (IMS)*, San Francisco, CA, pp. 1–3, 2016. <https://doi.org/10.1109/MWSYM.2016.7540038>.

Andree-Labsch, S., Jacobs, K., Stutzki, J. et al, “Near quantum limited Nb-Al-AIO_x-Nb mixers on 9 μm thick silicon substrates around 350 GHz,” *J Infrared Milli Terahz Waves* 35, pp. 300–317, 2014. <https://doi.org/10.1007/s10762-014-0052-5>.

Interestingly, this last paper describes a mixer chip with technologies remarkably similar to the author’s, including not only its radial waveguide probe, but also its thin silicon substrate, beam leads for ground connections, integrated IF matching network design, and 4–12 GHz IF frequency range. It did use a pair of SIS devices rather than the author’s single-SIS approach.

Mixer design

This final citation is of a paper describing a sideband separating receiver constructed using the author's mixer chip design:

J. S. Ward, K. A. Lee, J. Kawamura, G. Chattopadhyay, and P. Stek, "Sensitive broadband SIS receivers for microwave limb sounding," *2008 33rd International Conference on Infrared, Millimeter and Terahertz Waves*, Pasadena, CA, pp. 1–2, 2008. <https://doi.org/10.1109/ICIMW.2008.4665436>.

Final comment

Needless to say, during the course of this project the author was not only introduced to an enormous range of technologies and laboratory techniques, but he also had the privilege to make friends with many outstanding scientists, engineers, and other professionals throughout the terahertz community. Projects come and go, and nearly everyone directly involved with this one has moved on to other things, the author included. The skills and friends, however, will remain relevant throughout my career, no matter where it leads.

APPENDICES

Appendix A: MIXER CHIP SPECIFICATIONS AND GEOMETRY

Table A-I: Mixer chip material initial design specifications

<i>Silicon substrate:</i>		
Dimensions: 1990×230×25 μm		
<i>Niobium circuitry:</i>		
V_{gap} : 2.9 mV	T_{C} : 9.2 K	ρ (normal): 5.0 μΩ cm
Ground plane thickness: 200 nm	Circuitry thickness: 400 nm	
<i>SiO insulator:</i>		
Dielectric const. (ϵ): 7.5*	Loss tan ($\tan \delta$): 0.0012	Thickness: 350 nm
<i>SIS:</i>		
Materials: niobium with aluminum nitride barrier (Nb-Al-AlN _x -Nb)		
V_{gap} : 2.9 mV	Critical current density (J_{C}): 16 kA/cm ²	
$R_{\text{N}A}$: 14.3 Ω (μm) ²	Specific capacitance: 85 fF/(μm) ²	
Area: 1.69 (μm) ²	R_{N} : 8.5 Ω	C : 140 fF f_{RC} : >130 GHz

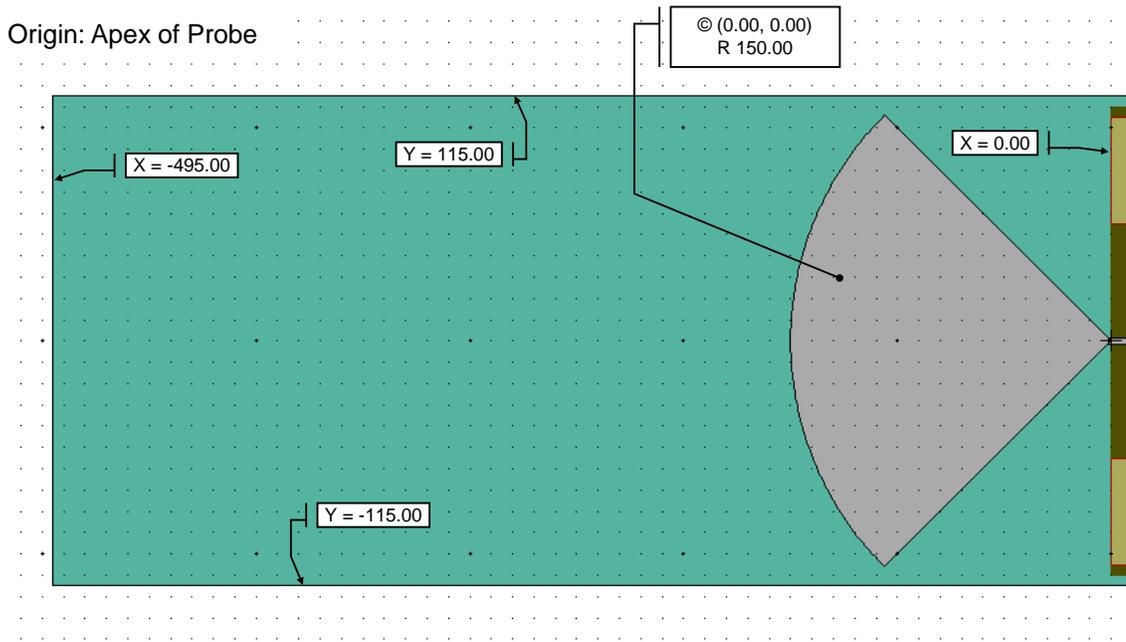
* This value for ϵ_{SiO} and its loss tangent were based on an analysis of experimental measurements performed by A. Vayonakis at Caltech in 2000 – 2002. A later reevaluation revised the value to 6.2, but too late to update the design. Note also, however, that analysis of the behavior of the balanced mixer chip described in Chapter 5 resulted in a best-fit value for ϵ_{SiO} of 5.7, very close to the conventionally accepted value of 5.6.

The overall layout of the mixer chip is shown in Figure 2-1 on page 19. The following figures provide detailed specifications of the circuit geometry used to fabricate the first-generation prototype mixer chips. In these diagrams the silicon substrate is shown in blue-green ■, the niobium circuitry is gray ■, the niobium ground plane ■, and the gold bond pad layer is ■. The blue-green substrate is hidden, however, in the detailed RF and choke circuitry diagrams.

Dimensions in microns

Grid: 10x10 microns

Origin: Apex of Probe



Dimensions in microns

Grid: 2x2 microns

Origin: Apex of Probe
(shown as a cross)

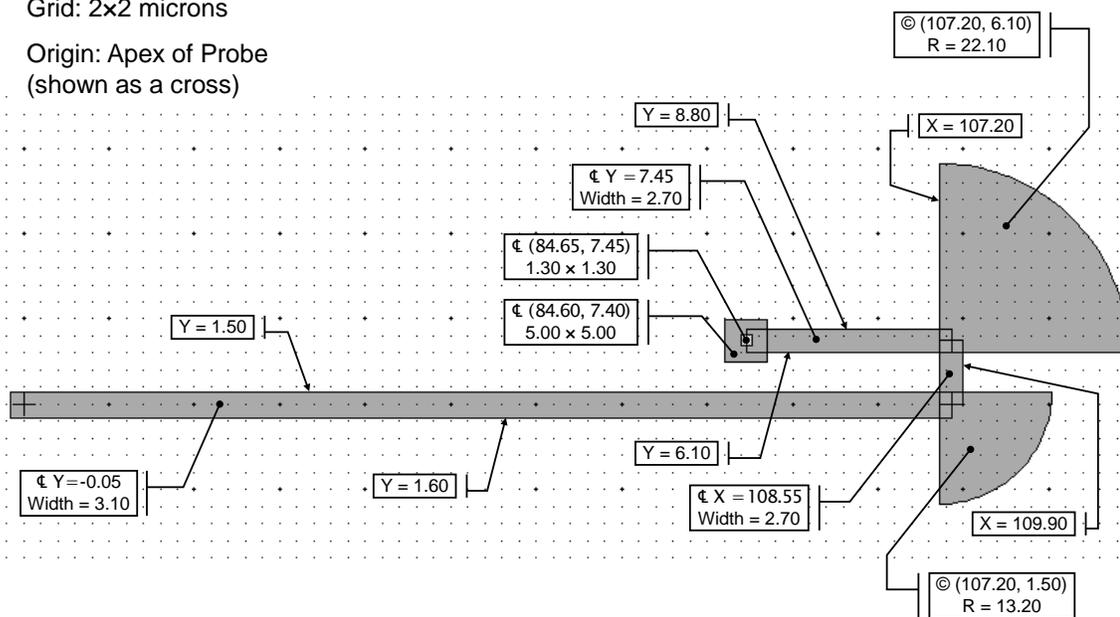


Fig. A-1: Waveguide probe and RF circuit detailed geometry.

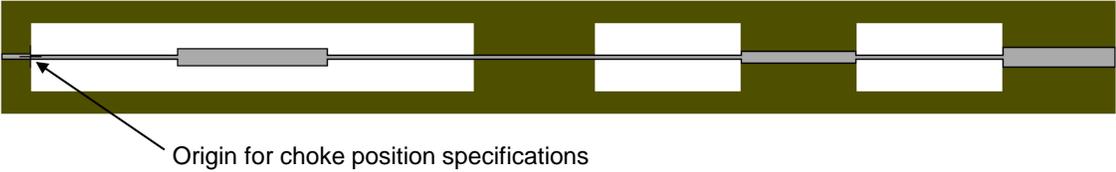
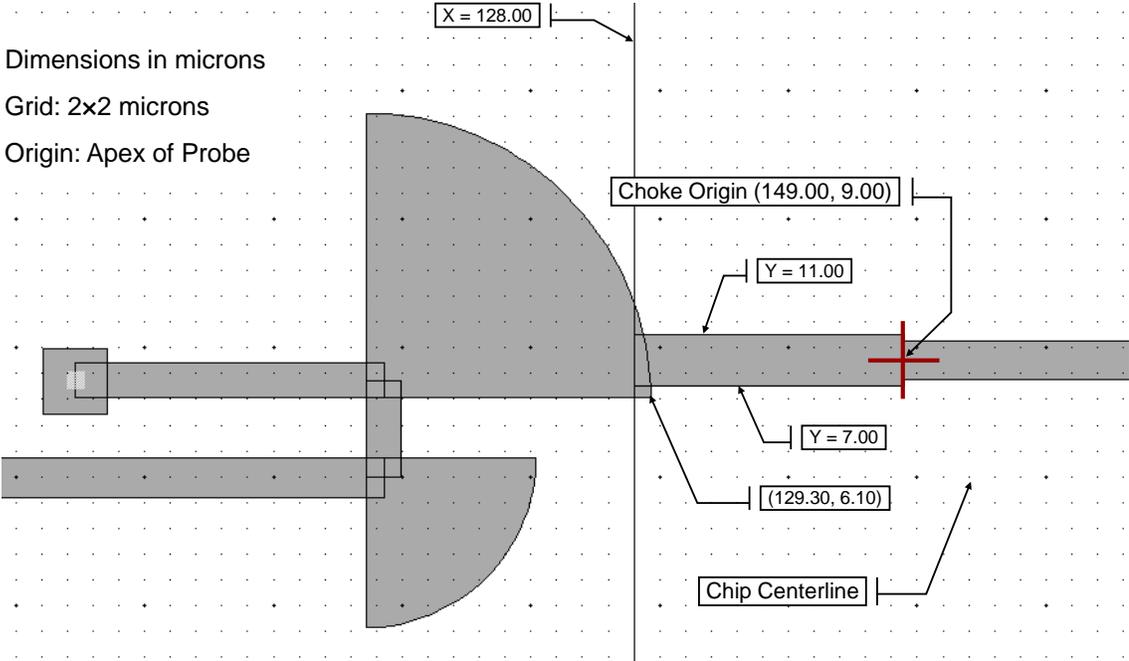


Fig. A-2: RF choke interface detailed geometry (upper); choke arrangement (lower).

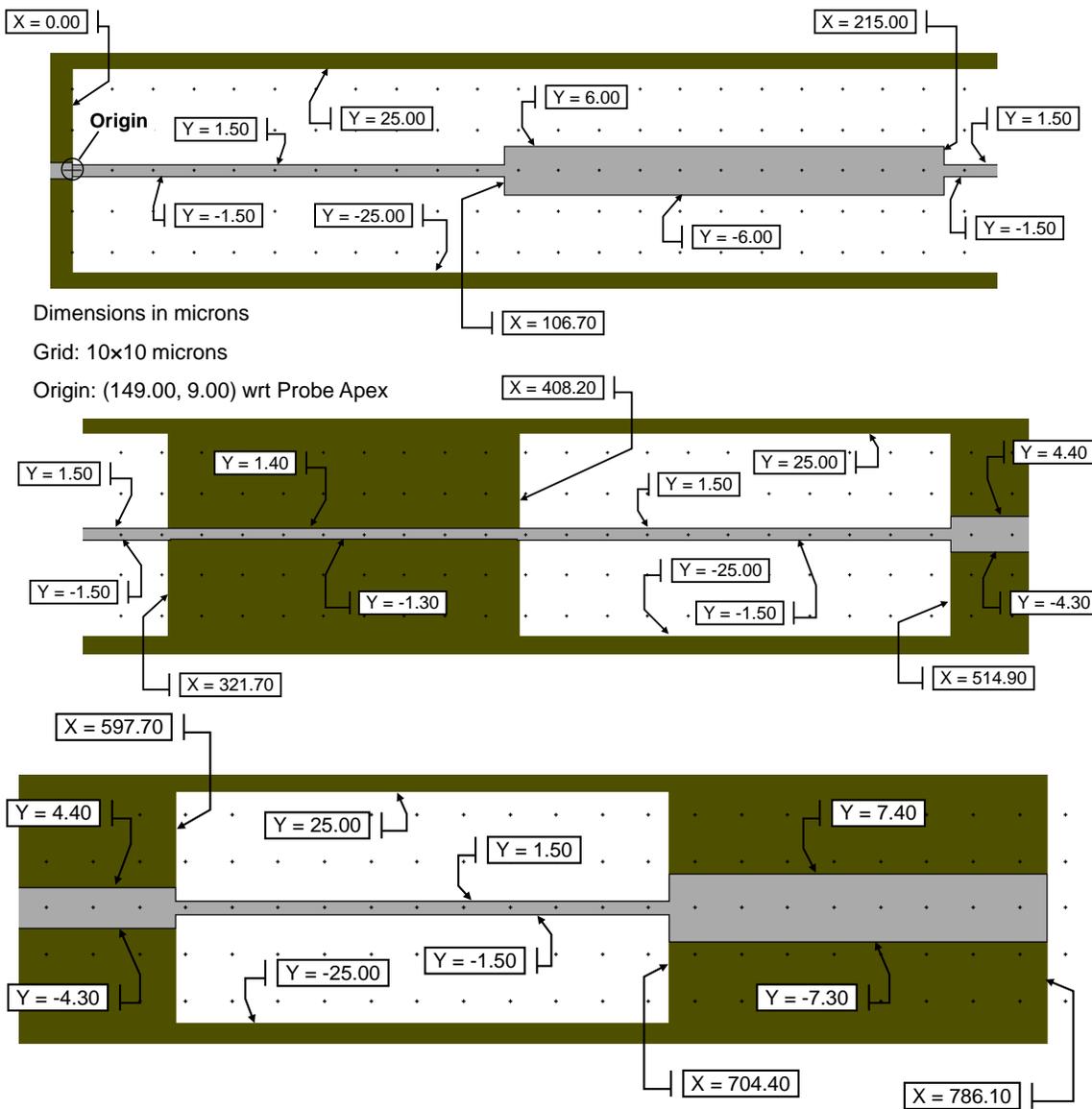


Fig. A-3: RF choke circuit and ground plane layer detailed geometries.

Dimensions in microns

Grids: 10x10 microns

Origin: Apex of Probe

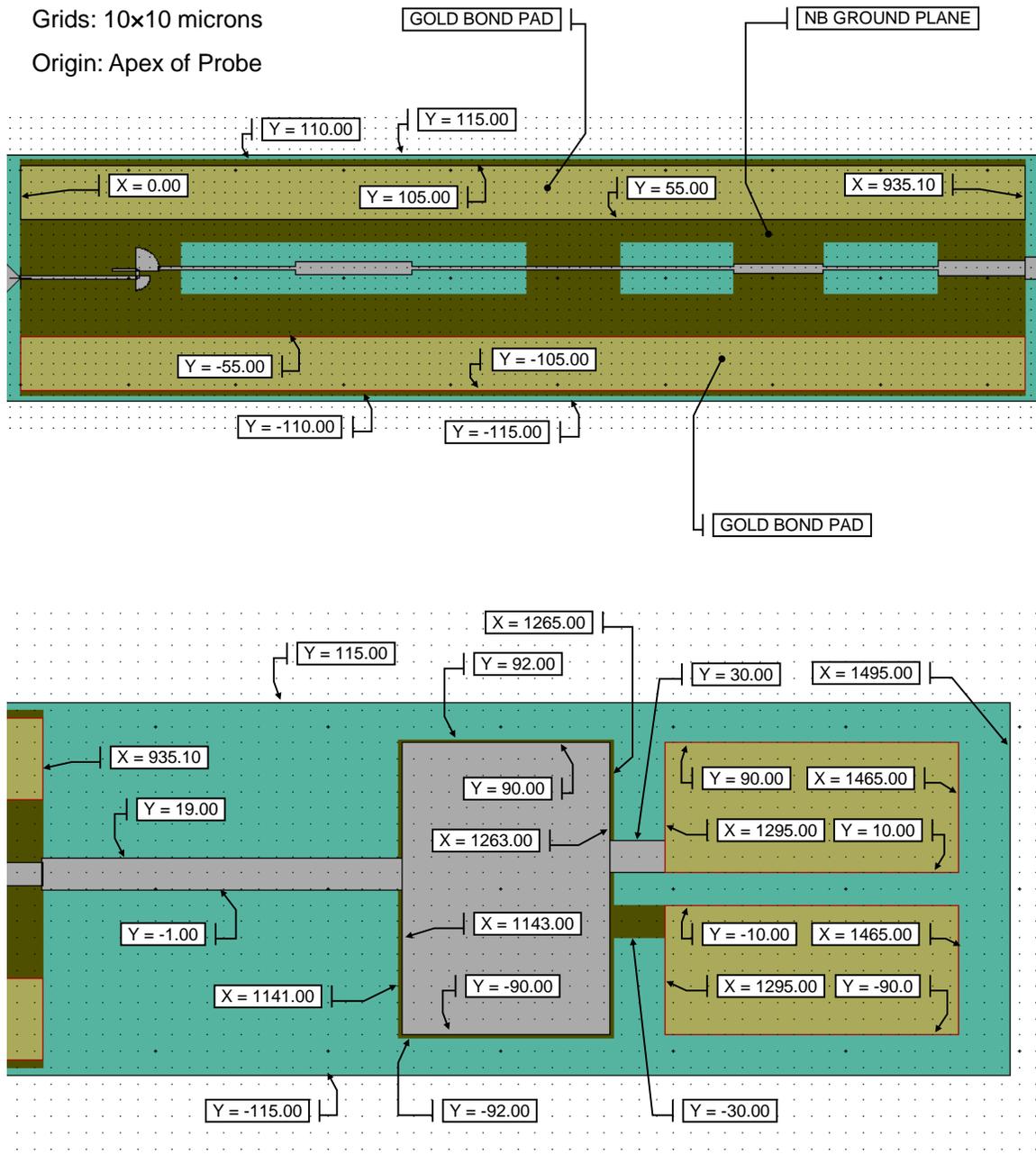


Fig. A-4: Ground plane and circuit connection bond pad geometries.

Appendix B : MIXER BLOCK AND LNA SPECIFICATIONS

Mixer block drawings

Fig. B-1: Assembled mixer block, feed horn, and horn adapter with dimensions.

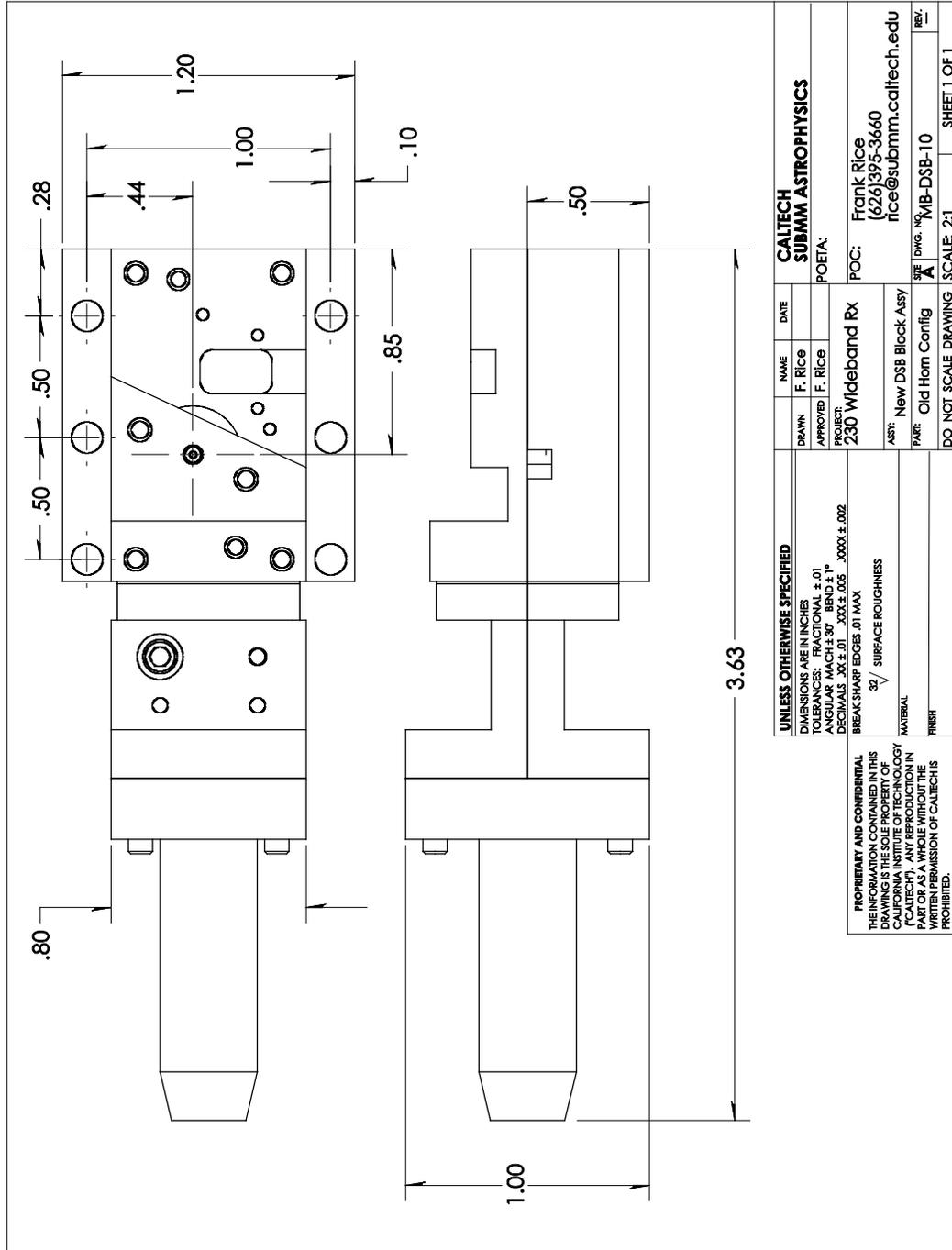


Fig. B-2: Drawing views of the assembled mixer block.

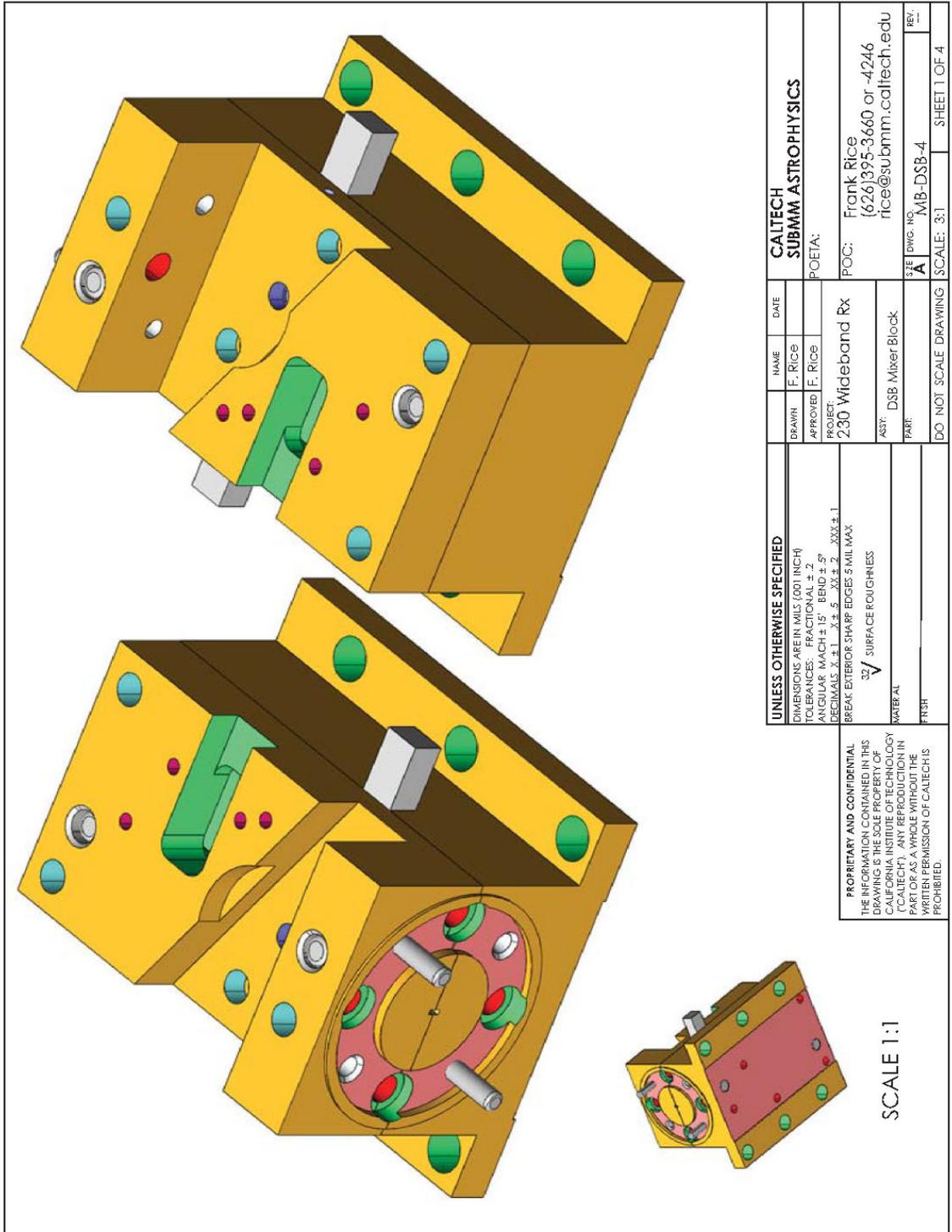


Fig. B-3: Mixer block parts descriptions.

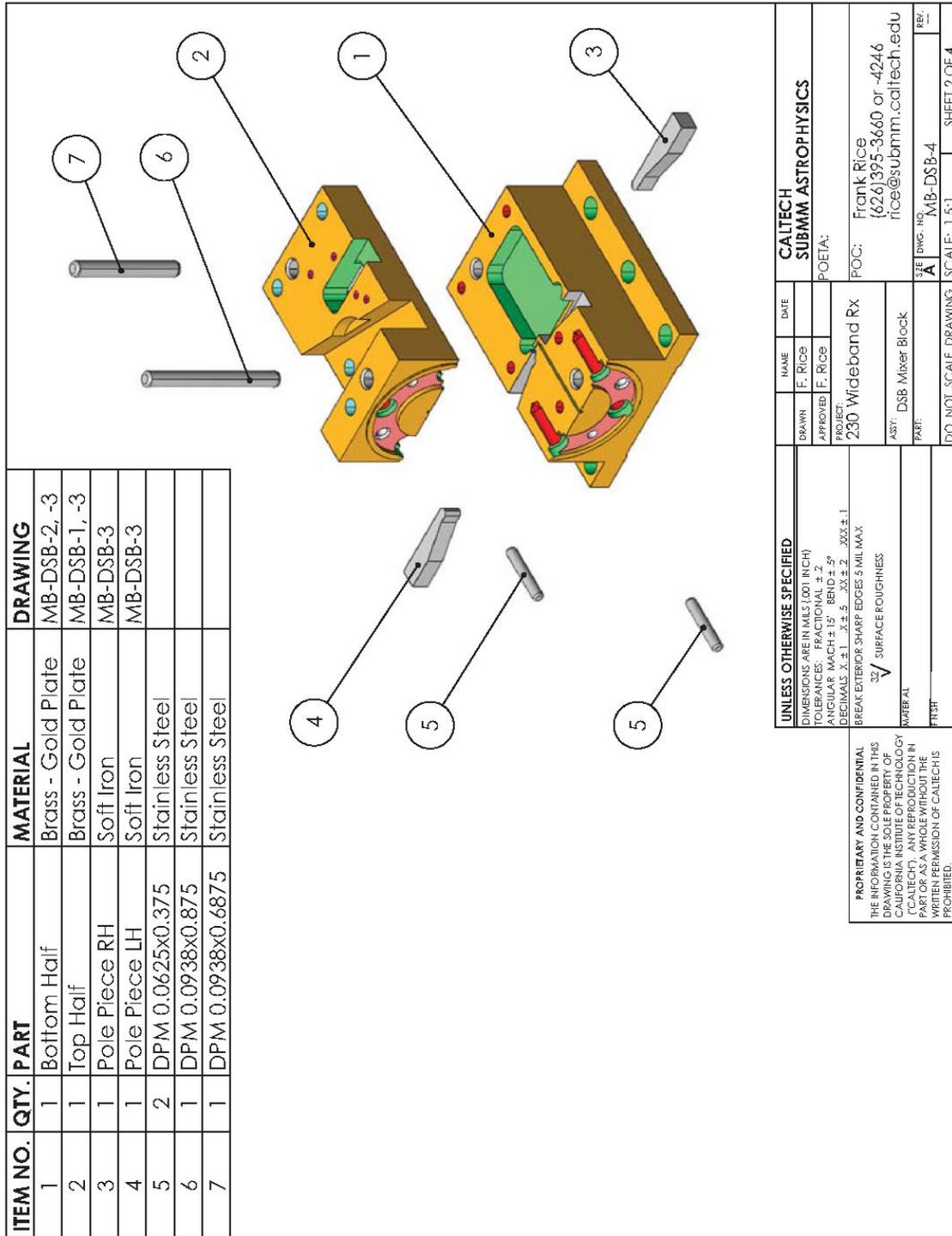


Fig. B-4: Mixer block top half drawing.

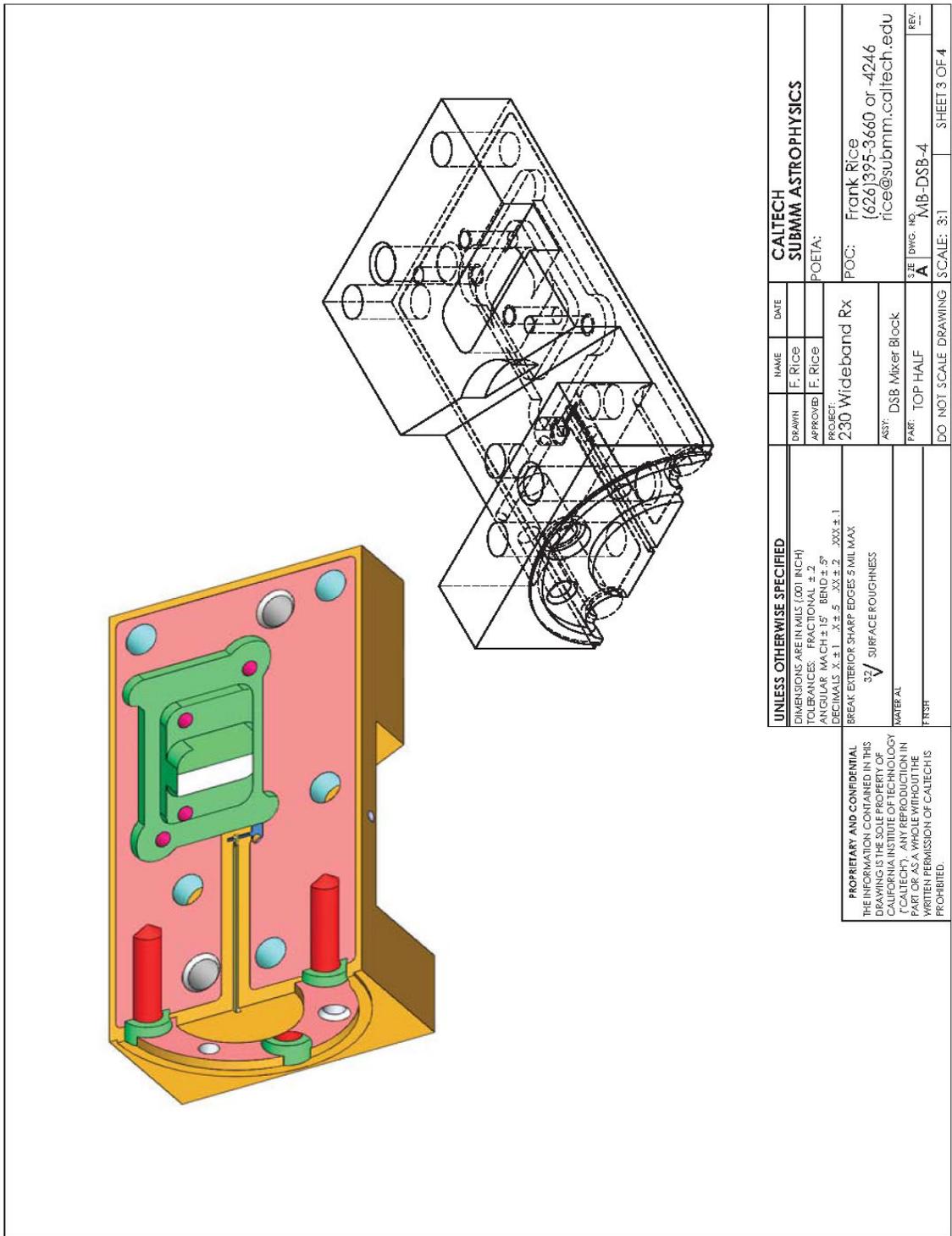


Fig. B-5: Chip channel and waveguide detailed dimensions.
 This is an excerpt from the mixer block top half drawing (dimensions in mils).

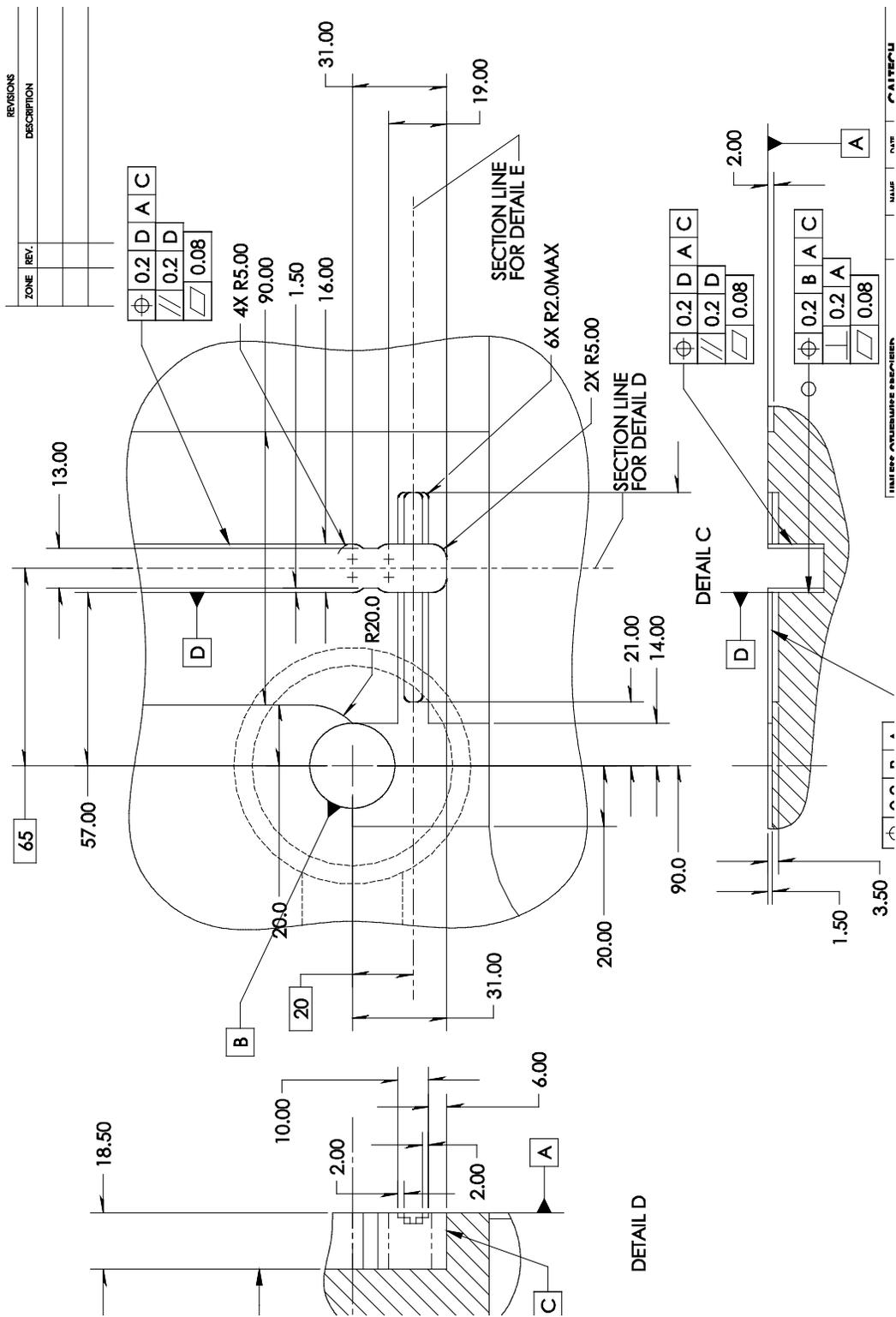


Fig. B-6: Mixer block bottom half drawing.

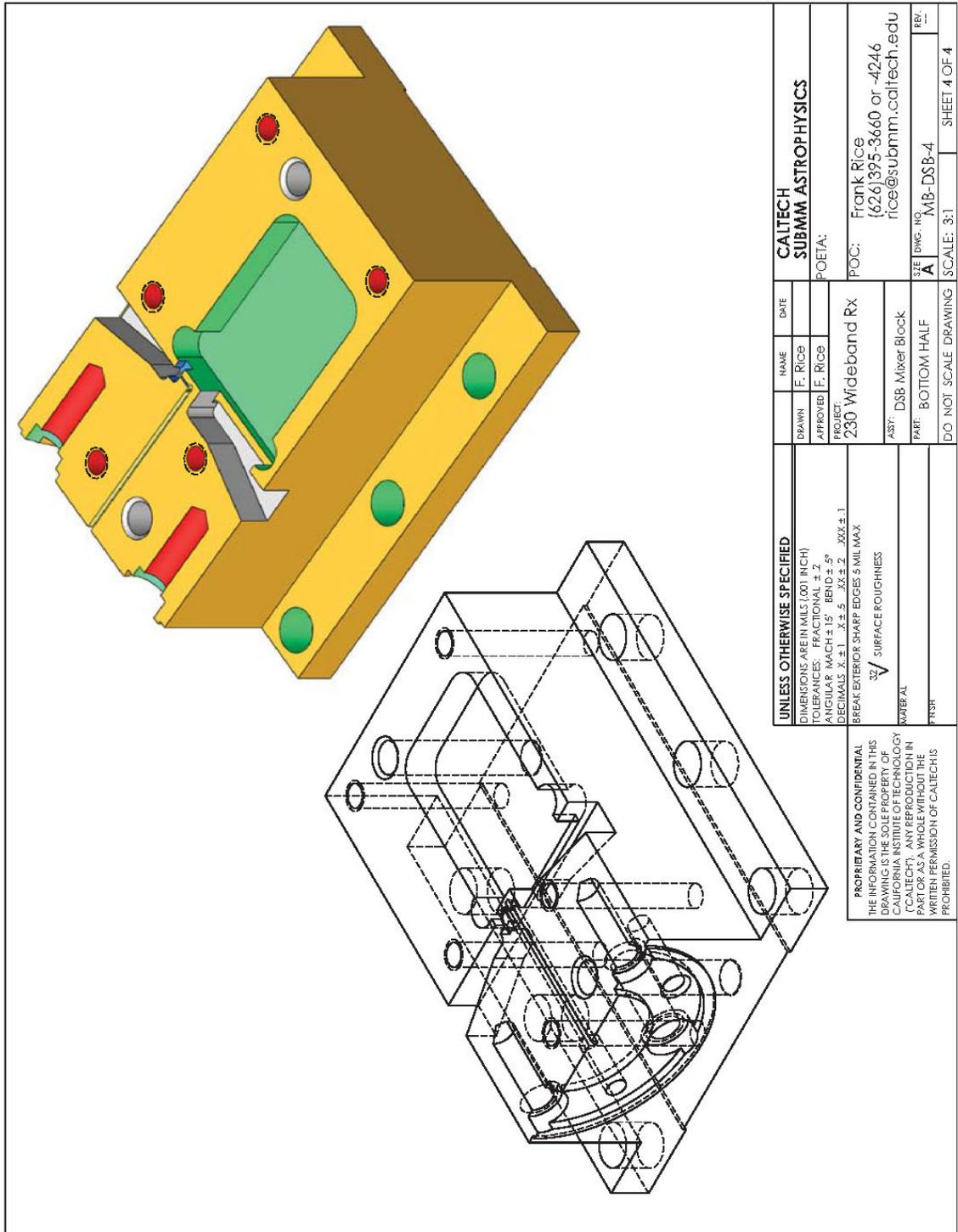


Fig. B-7: Mixer block horn adapter.

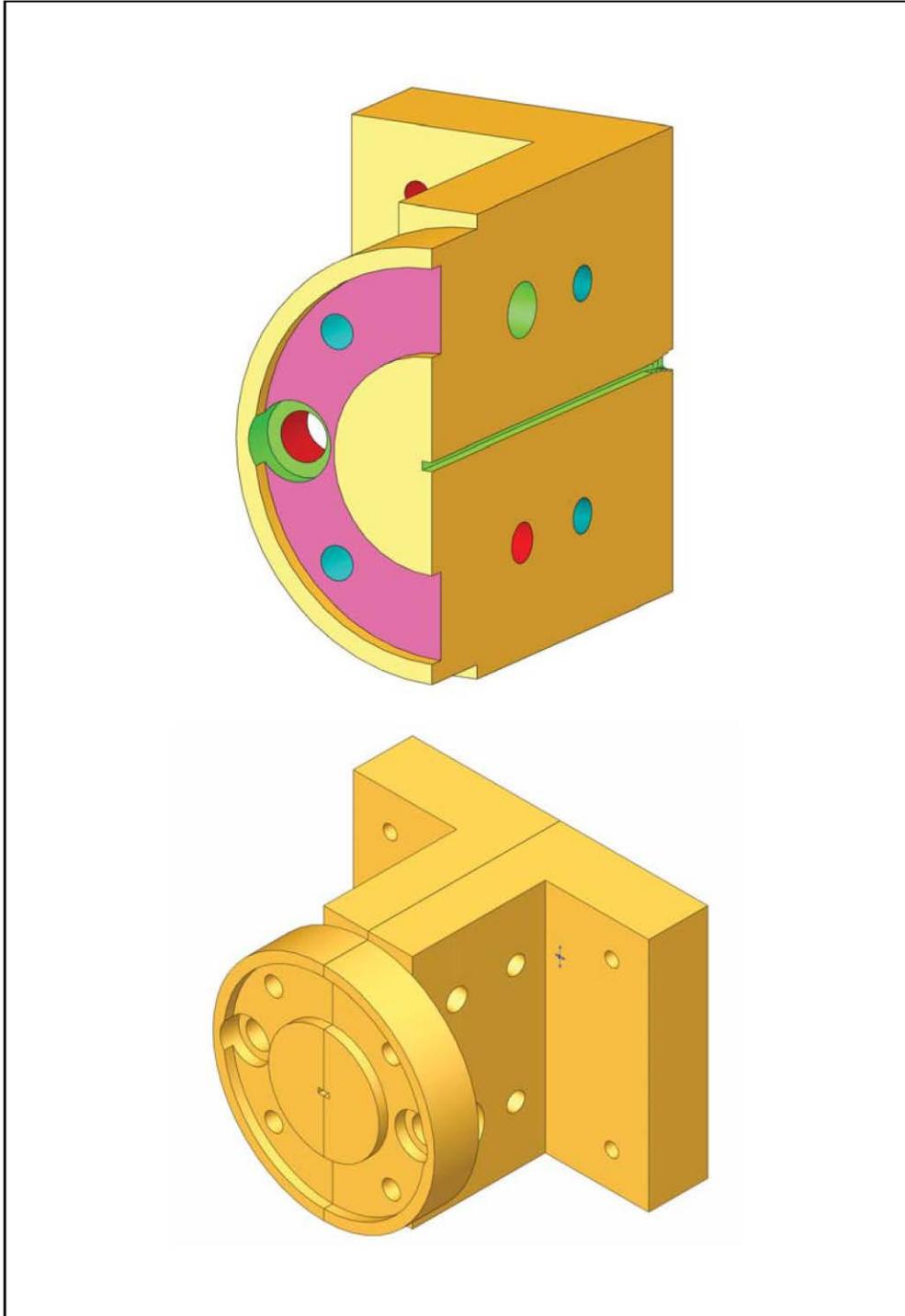


Fig. B-8: Waveguide rectangular-cylindrical transition detailed dimensions (in mils). This structure transforms the feed horn cylindrical waveguide to the mixer block rectangular waveguide. Shown are details from the mixer block horn adapter drawings.

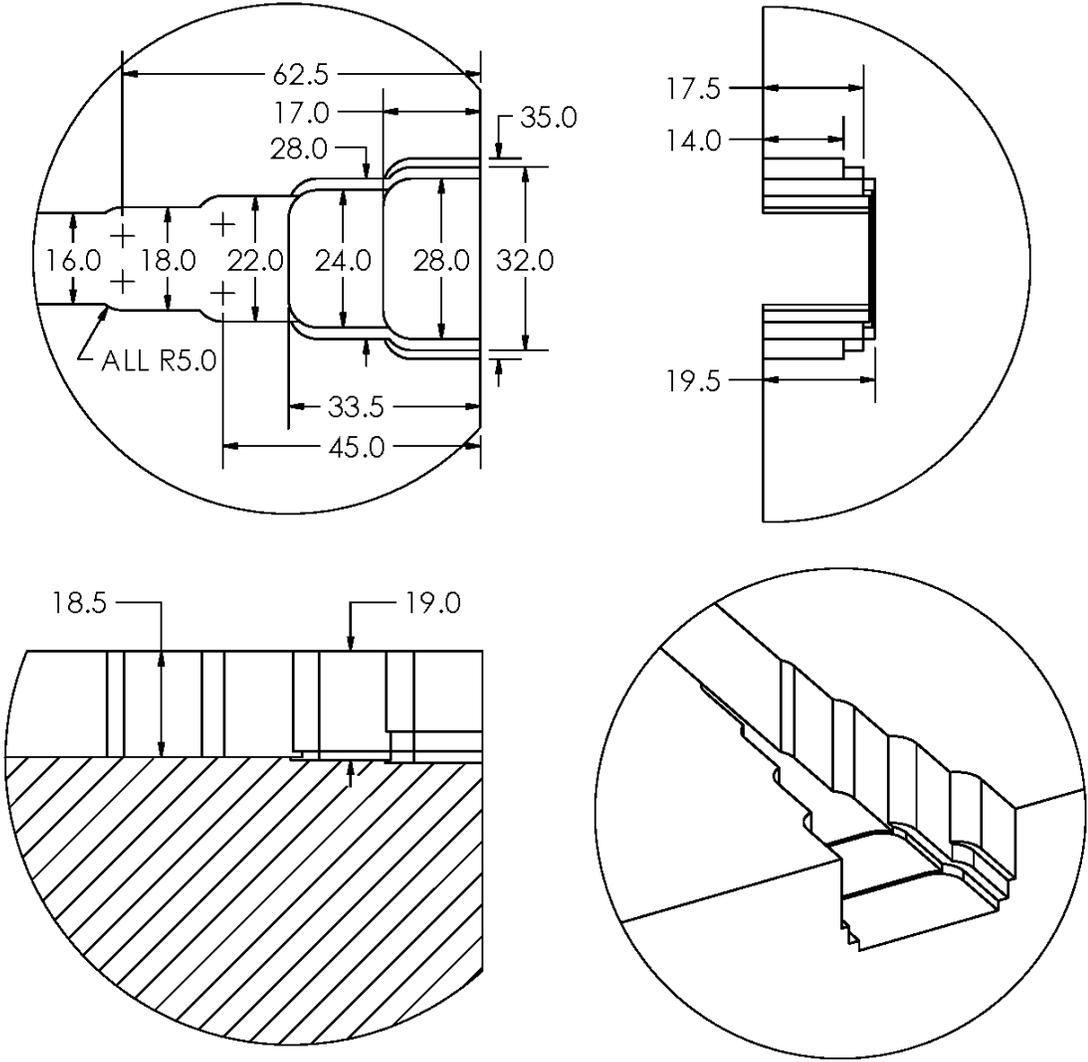
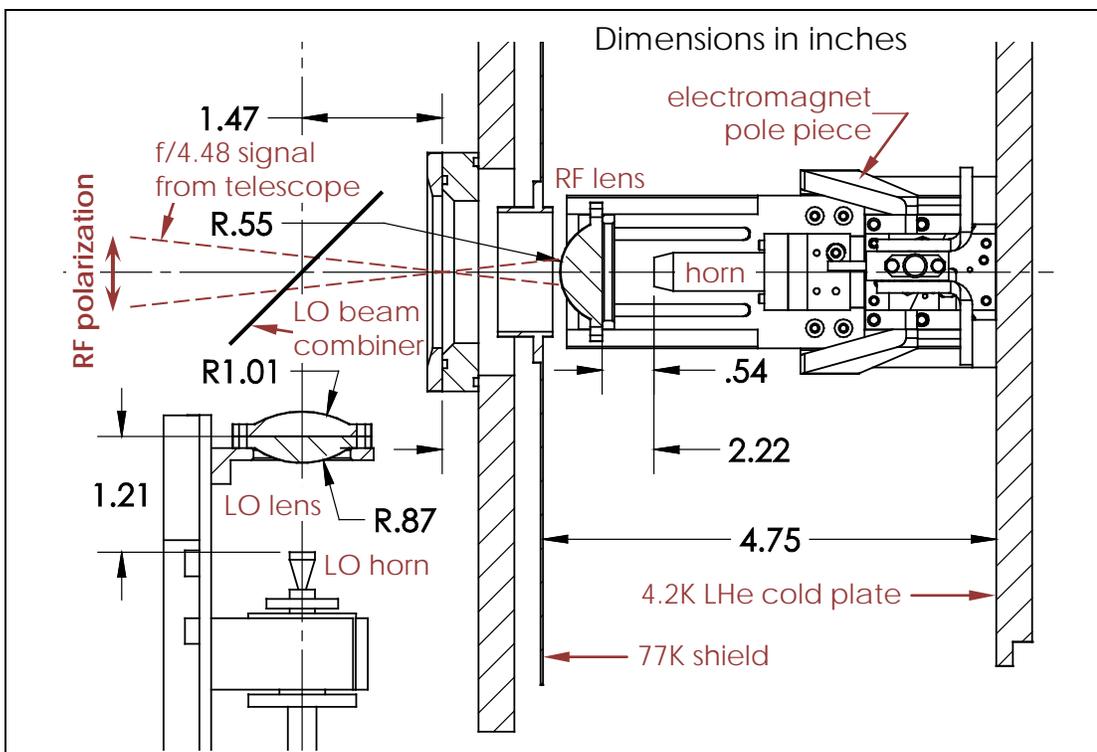


Fig. B-9: Mixer block assembly mounting and telescope and LO optical path interface. This September 2005 configuration shows the relative positions and orientations of the major components in the optical signal path from the telescope, including the feed horn and its focusing lens, the LO source and its focusing lens, and the LO signal beam combiner. The beam waist from the telescope optics is located very near the surface of the cryostat's RF signal entrance window, which is in turn located 2.22 inches from the mixer feed horn aperture. The lenses are HDPE. The beam combiner is 1 mil thick Mylar. The cryostat vacuum window is 1mm HDPE with a Zitex antireflection coating on its outer surface. The 77 K heat shield window near the RF lens (not shown), constructed from 0.5 mm thick Fluorogold, serves as an IR filter.



LNA performance specifications

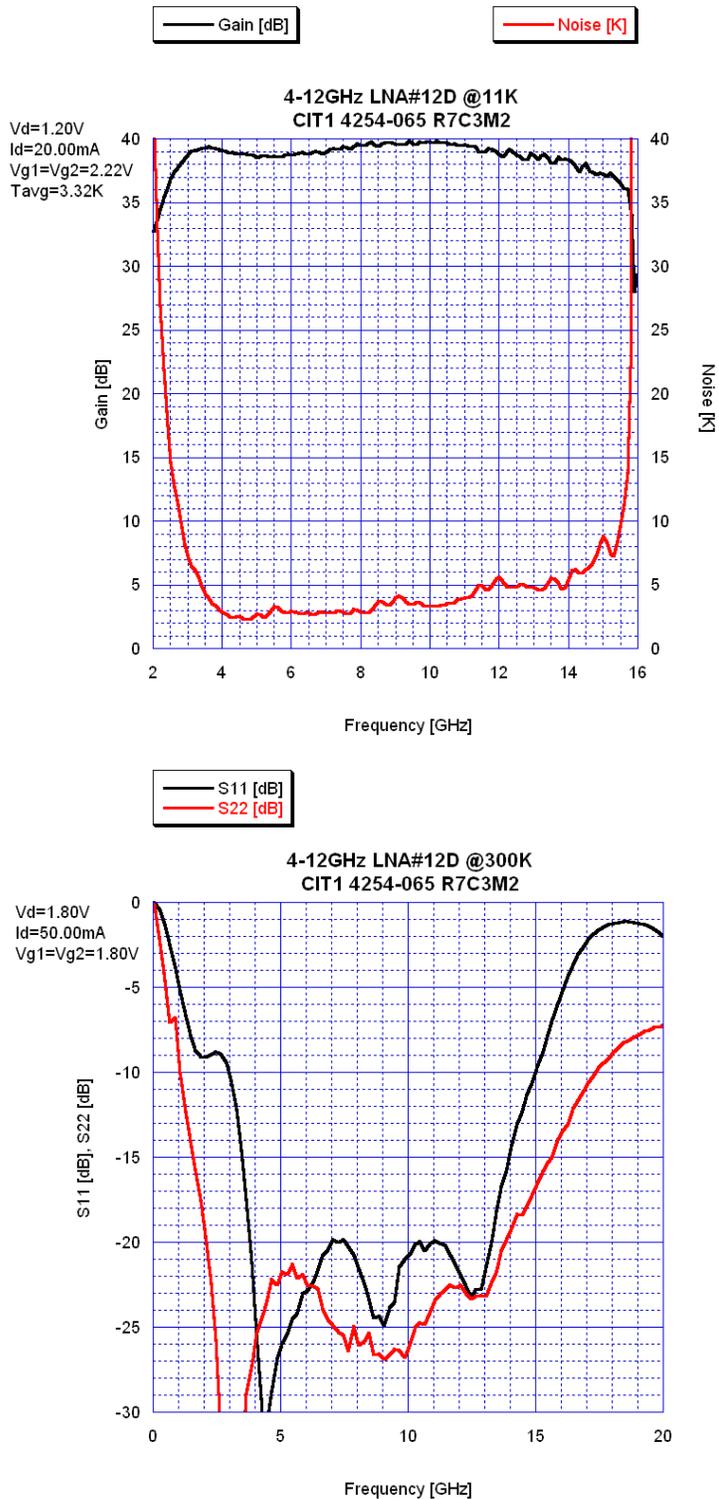


Fig. B-10: Measured performance of the 4-12 GHz LNA delivered in 2004. A commercial version of this amplifier is available. [41]

Appendix C : THE SCATTERING REPRESENTATION AND NOISE

This appendix covers a variety of basic topics in order to provide a somewhat brief review of the representations of circuit elements used by SuperMix and referred to throughout the body of the text. For a more satisfying coverage of the topics herein, refer to any reasonably complete text on radio-frequency or microwave circuit analysis. Also of note are the dissertations of Wedge [38] and Ward [15]. Both of these papers provide detailed descriptions of how the representation of a complicated circuit (including noise) may be constructed from the representations of its component elements. In particular, [15] describes the very clever data structures and algorithms John Ward created for SuperMix.

The wave representation

High-frequency circuit and structure modeling almost universally relies on a *wave representation* of electromagnetic signal flow and an associated *scattering matrix representation* of each circuit element's characteristics, and SuperMix is no exception. This representation is especially convenient because many elements of high-frequency circuits may be modeled as transmission lines, waveguides, or similar structures supporting signal wave propagation. Such structures support one or more *propagation modes*, each of which is characterized by two generally complex-valued, frequency-dependent parameters: the mode's *wave number*, $k(\omega)$, also called its *propagation constant*, and its *characteristic impedance*, $Z_c(\omega)$. For a harmonic disturbance with frequency ω incident on an ideal, lossless transmission medium, $k(\omega)$ and $Z_c(\omega)$ are both either purely real (and the disturbance propagates) or purely imaginary (and only an evanescent wave is present).

The SIS mixer chip circuitry consists of a variety of two-conductor transmission lines along with a waveguide probe and a single SIS device. At any point on a two-conductor transmission line, an electromagnetic wave propagating along it induces a voltage and current. Refer to the left-hand graphic in Fig. C-1 on page 127, where it is assumed that a single, sinusoidal wave with frequency ω propagates from right to left along a transmission line which has characteristic impedance $Z_c(\omega)$. The physical significance of

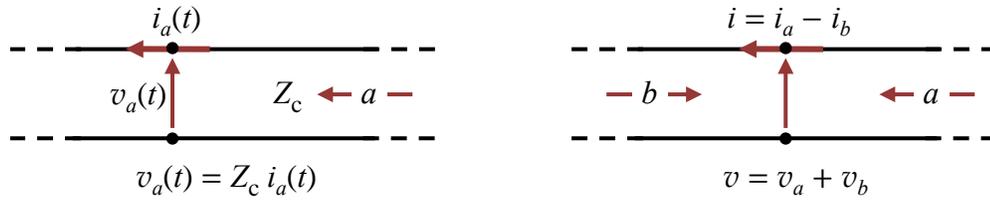


Fig. C-1: The wave representation on a transmission line.

Left: A single wave a propagating along a transmission line with characteristic impedance Z_c showing the relationship between the voltage and current induced by a at some point on the line.

Right: A transmission line with harmonic waves a and b propagating along it in opposite directions. The total current and voltage at any point on the line is the superposition of the two wave currents and voltages, as shown.

Z_c is that it specifies the ratio of the voltage and current phasors (complex amplitudes) induced at any position on the line by the wave. The direction of propagation of the wave is given by the direction of the current flow $i(t)$ in the more positive of the two conductors when $v(t)$ is at a maximum. Thus the wave's propagation direction naturally corresponds to the direction of the average flow of power it induces along the transmission line.

In SuperMix, the complex-valued voltage and current phasor amplitudes $v(\omega)$ and $i(\omega)$ are chosen to represent RMS (root mean square) voltages and currents, so the average power flow through some point along the transmission line induced by the wave becomes

$$P = \text{Re}[vi^*] = (\text{Re}[Z_c]/|Z_c|) |v|^2/|Z_c|. \quad (\text{C.1})$$

All of the quantities in the above equation may be functions of frequency. Z_c is real and positive for a propagating, lossless transmission line. In this case, the voltage and current phasors are in phase, and the power in the wave is just given by the product of their RMS magnitudes.

The wave itself (Fig. C-1, left) may be represented at any point along the line by a phasor a , which in SuperMix is assumed to have the same phase as its induced voltage. For an ideal, propagating transmission line the wave phasor amplitude is chosen to represent the power in the wave, so that $P = |a|^2$. More generally,

$$P = |a|^2 (\text{Re}[Z_c] / |Z_c|) \Rightarrow a(\omega) = V(\omega) / |Z_c|^{1/2}. \quad (\text{C.2})$$

Now consider the case of two waves a and b with the same frequency but opposite directions of propagation (Fig. C-1, right). Assuming that the transmission medium is linear, the voltages and currents induced on the transmission line by the two waves will be superposed. The waves' voltage phasors will add, as shown in the figure, but their current phasors will subtract, since the waves propagate in opposite directions. The net average power flow at any point on the line will still be given by $P = \text{Re}[vi^*]$, where v and i are now the superposed voltages and currents. In terms of the two wave amplitudes and assuming a real-valued characteristic impedance, then $P = |a|^2 - |b|^2$, where positive power flow would be in the direction of wave a .

The scattering matrix

The wave representation provides a very general and useful tool to describe signal propagation even along transmission media such as waveguides which are not easily representable using the two-conductor model of Fig. C-1 (with its well-defined voltages and currents). Consider a structure with several *ports*, each connected to an independent wave mode of one of several transmission lines. We pick some numbering scheme whereby each port gets a unique integer index, as with the 5-port structure shown in Fig. C-2. In this example, the ports are imagined to be connected to semi-infinite, lossless transmission lines. For now, assume that only the line connected to port 1 of the structure has an incident wave. This incoming wave's complex-valued amplitude phasor is

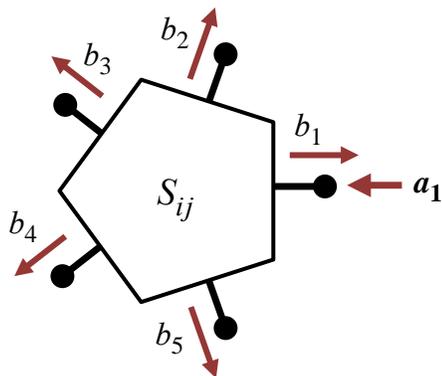


Fig. C-2: A 5-port structure with an incoming wave at port 1.

The structure's scattering matrix \mathbf{S} describes the relationship between the incident and emitted waves.

designated a_1 . In response to this stimulus, the structure emits waves with phasors b_i from all of its ports, including a reflected wave at port 1, b_1 . Assuming that the structure's response is linear, it can be described by a 5×5 complex-valued, frequency-dependent *scattering matrix* $\mathbf{S}(\omega)$ with elements S_{ij} . The outgoing phasors in Fig. C-2 are then $b_i = S_{i1} a_1$. With waves incident at all ports, and forming column vectors \mathbf{b} and \mathbf{a} from the wave phasors b_i and a_j , the relationships between the incident and outgoing phasors are given by the linear matrix relationship $\mathbf{b} = \mathbf{S} \cdot \mathbf{a}$.

So far we have said nothing about the characteristic impedances of the transmission media connected to the ports. In general, each port could have a unique, frequency-dependent, characteristic impedance for its associated transmission mode. This would, for example, be the appropriate approach for a scattering matrix connecting different propagation modes in a waveguide structure. Such a situation leads to what are called *generalized* scattering matrices. The model used to define a scattering matrix \mathbf{S} in SuperMix, on the other hand, is a more conventional one wherein all ports are assumed to interface to identical, lossless transmission lines with a common, real-valued, frequency-independent characteristic impedance called the matrix's *normalizing impedance*, Z_0 . SuperMix by default uses 50Ω , but that value can be changed to something more appropriate. Mixer model programs, for example, might set Z_0 to the SIS device's normal resistance.

Reflection by an impedance; Y and Z matrices

In addition to assuming a common normalizing impedance Z_0 for all ports of an \mathbf{S} matrix representation, SuperMix makes a useful assumption about the nature of the transmission lines its elements connect: it assumes that each port of a circuit element can still be represented as an interface to a two-conductor transmission line with a voltage-current relationship determined by Z_0 as described in Fig. C-1 on page 127. Given our previous comments about the general utility of the wave representation, this assumption may seem unnecessarily restrictive. It turns out that this is not really the case—in fact, it makes implementation of the general concept of “impedance” using SuperMix much more straightforward, because impedances may then be modeled as ratios of voltage and current phasors.

Consider the simple 1-port terminating a transmission line with characteristic impedance Z_0 shown in Fig. C-3. If the termination's impedance is Z , then what is the value of its single \mathbf{S} matrix element, S_{11} ? At the port, we know that $v = v_a + v_b$ and $i = i_a - i_b$ (with the sign conventions shown by the arrows in the figure), where $Z_0 = v_a/i_a = v_b/i_b$. But it must also be the case that $Z = v/i$. Thus $S_{11} = b/a = v_b/v_a = (Z - Z_0)/(Z + Z_0)$.

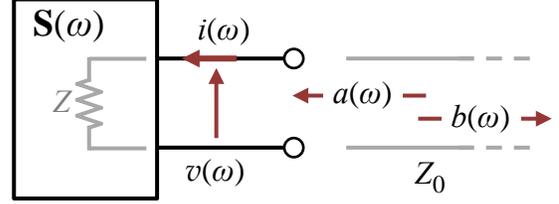


Fig. C-3: Reflection at a termination (1-port).

So we are led to the familiar result describing the reflection of a wave at an impedance interface. The complex-valued reflection coefficient $\Gamma(\omega)$, our 1-port's S_{11} , is given by

$$\Gamma(\omega) = \frac{Z(\omega) - Z_0}{Z(\omega) + Z_0}. \quad (\text{C.3})$$

Conversely, given a reflection coefficient $\Gamma(\omega)$ and the normalizing impedance Z_0 , the termination's impedance $Z(\omega)$ must be

$$Z(\omega) = Z_0 \frac{1 + \Gamma(\omega)}{1 - \Gamma(\omega)}. \quad (\text{C.4})$$

The power absorbed by the termination (or continuing on in a transmitted wave if Z is the characteristic impedance of another transmission line represented by the "termination") is $P = |a|^2 - |b|^2 = |a|^2(1 - |\Gamma|^2)$. This factor of $(1 - |\Gamma|^2)$ or $(1 - |S_{11}|^2)$ is a common sight when analyzing circuits using the scattering matrix representation.

The idea of the 1-port terminating impedance with its relation $v(\omega) = Z(\omega)i(\omega)$ extends in a straightforward way to the *impedance representation* (\mathbf{Z} matrix) of a multi-port, linear circuit, wherein the independent variables are the ports' input current phasors, and the voltages are calculated using the matrix equation $\mathbf{v} = \mathbf{Z} \cdot \mathbf{i}$. A multi-port's \mathbf{S} and \mathbf{Z} matrix representations are related by the matrix versions of (C.3) and (C.4):

$$\begin{aligned} \mathbf{Z} &= Z_0 (\mathbf{I} + \mathbf{S}) \cdot (\mathbf{I} - \mathbf{S})^{-1} \\ \mathbf{S} &= (\mathbf{Z} - Z_0 \mathbf{I}) \cdot (\mathbf{Z} + Z_0 \mathbf{I})^{-1}. \end{aligned} \quad (\text{C.5})$$

In the expressions (C.5) the symbol \mathbb{I} is the identity matrix. The order of the matrix multiplications does not matter because the factors commute. Similarly, the *admittance representation* (\mathbf{Y} matrix), wherein $\mathbf{i} = \mathbf{Y} \cdot \mathbf{v}$, is also useful. Clearly, $\mathbf{Y} = \mathbf{Z}^{-1}$, so:

$$\begin{aligned} Z_0 \mathbf{Y} &= (\mathbb{I} - \mathbf{S}) \cdot (\mathbb{I} + \mathbf{S})^{-1} \\ \mathbf{S} &= (\mathbb{I} - Z_0 \mathbf{Y}) \cdot (\mathbb{I} + Z_0 \mathbf{Y})^{-1}. \end{aligned} \tag{C.6}$$

The advantage of the wave representation and its scattering matrix \mathbf{S} is that for nearly any useful circuit configuration, the scattering matrix \mathbf{S} exists and is nonsingular. Unfortunately, this is not the case for the \mathbf{Y} and \mathbf{Z} matrices of some common circuit configurations (ports internally connected in series and or in parallel, for example).

The \mathbf{S} matrix of an ideal, lossless network is *unitary*: $\mathbf{S}\mathbf{S}^\dagger = \mathbf{S}^\dagger\mathbf{S} = \mathbb{I}$, where \mathbf{S}^\dagger is the conjugate transpose of \mathbf{S} . Furthermore, if a network is passive and does not contain, for example, ferromagnetic materials, then it is probably *reciprocal* (equivalent to time-reversal symmetry), and \mathbf{S} then equals its transpose: $\mathbf{S} = \mathbf{S}^T$.

The Smith Chart

Much high-frequency design effort is devoted to efficiently coupling a signal from one subsystem to the next. An incident signal wave $a(\omega)$ arriving at the input port of a circuit may suffer some reflection $b(\omega)$, and the net power transfer into that port of the circuit is $P = |a|^2 - |b|^2$, with $b/a = \Gamma$. If the circuit is passive, then $|\Gamma(\omega)| \leq 1$, so the reflection coefficient Γ lies within the unit circle in the complex plane. A parametric plot in the complex plane of the variation of $\Gamma(\omega)$ over a design's desired frequency range is called a *Smith chart*,²² and it has for many years served as a useful tool for visualizing a circuit's performance.

Consider this example: A 200–300 MHz signal originates from a transmission line source with a characteristic impedance of 30Ω . The circuit input is a pure resistance of 8Ω , so, from equation (C.3), the reflection coefficient $\Gamma = -0.6$, independent of

²² After Bell Laboratories engineer Phillip Smith (1905–1987).

frequency. Nearly 40% of the incident signal power will be reflected by the input to the circuit. The designer adds a series inductance L and parallel capacitance C to the circuit to attempt to reduce the magnitude of $\Gamma(\omega)$ over the signal's frequency range. The modified input to the circuit is shown in Fig. C-4.

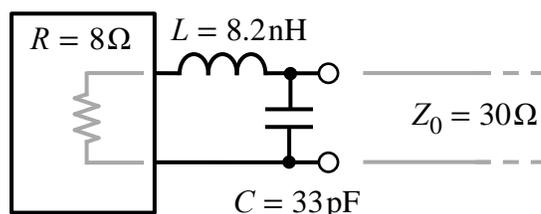


Fig. C-4: Impedance matching example. By adding the L and C as shown, less incoming power will be reflected over the design frequency range.

The engineer's objective is to use the added L and C to transform the circuit's 8Ω input resistance to a value closer to the 30Ω line impedance. The input impedance of the modified circuit is the series impedance of R and L in parallel with that of C , thus changing the reflection coefficient $\Gamma(\omega)$. The result is plotted in the Smith chart on page 133.

The Smith chart provides a parametric plot in the complex plane of the $\Gamma(\omega)$ variation with frequency. The center of the plot is at complex 0, and the unit circle is shown in bold black. As can be seen from the red curve denoting $\Gamma(\omega)$, the designer's simple modification has resulted in a significant improvement in the reflection power loss: within most of the design frequency range $|\Gamma(\omega)| \ll 0.25$. Because the added components are purely reactive, they dissipate no power. Thus the unreflected power does indeed go into the original circuit's 8Ω input, as desired. The SuperMix optimizer capability can be used to determine these component values: see *Using the optimizer* in Appendix I.

The Smith chart also generally includes coordinate arcs for various values of the input impedance Z which one can use to quickly estimate how a change in Z can affect the reflection coefficient Γ . Because the complex function $\Gamma(Z)$ defined by (C.3) is analytic over the right half-plane of complex-valued Z , it defines a conformal map $Z \rightarrow \Gamma$. Thus the vertical and horizontal lines in the Z -plane of constant real part and constant imaginary part map to orthogonal circles in the Γ -plane, as shown by the few loci plotted in light-gray in Fig. C-5. The entire right half of the Z -plane is mapped to the interior of the unit circle in the Γ -plane.

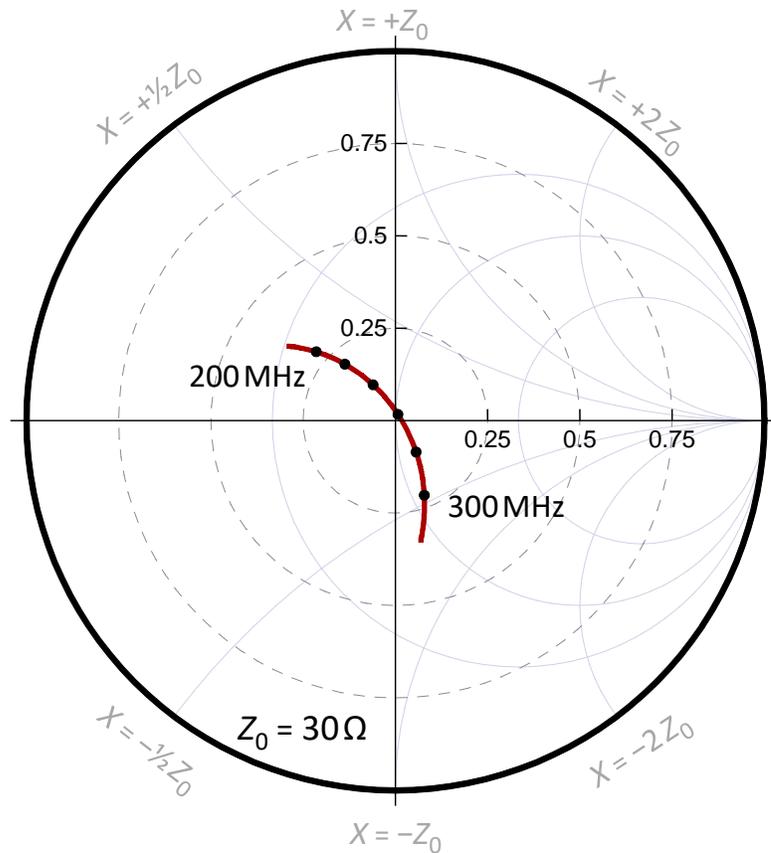


Fig. C-5: Smith chart showing a plot of the variation of $\Gamma(\omega)$ for the circuit in Fig. C-4. The black dots show $\Gamma(\omega)$ at 20 MHz intervals over the range 200–300 MHz. The dashed circles show fixed values of the magnitude of Γ . The faint solid circles with centers on the real axis are lines of constant Real $[Z(\omega)]$ (series resistance) equal to $\frac{1}{2}Z_0$, Z_0 , and $2Z_0$. The arcs of circles tangent to the real axis are lines of constant Imaginary $[Z(\omega)]$ (series reactance) equal to the values shown.

A variation of this scheme is to conformally map the circuit's complex admittance Y -plane onto the Γ -plane. Again, the right half-plane of Y is mapped to the interior of the unit circle. The loci of constant conductance (real part) and susceptance (imaginary part) are again circles. These Y -coordinate circles take the form of a reflection about the imaginary- Γ axis of the circles for Z shown on the Smith chart in Fig. C-5. Such Y -based Smith charts are used extensively in the sections of this text describing the optimal SIS embedding impedance investigations of Chapter 2.

Modeling internal sources and noise

Not all circuits are passive. *Active circuits* have internal power sources and emit coherent waves \mathbf{b} from their ports even if no incoming waves \mathbf{a} are present. In addition, most passive circuits emit random, broadband *noise* from their ports which originates in their resistive elements. In this section we address how these effects are incorporated into the wave representations of our circuits.

A circuit with one or more internal power sources will output a frequency-dependent vector of wave phasors $\mathbf{b}_S(\omega)$ which then adds to the response of the circuit to incident waves, so that $\mathbf{b} = \mathbf{S} \cdot \mathbf{a} + \mathbf{b}_S$. This model assumes that the internal source is *coherent*, in the sense that at any frequency for which $\mathbf{b}_S(\omega) \neq \mathbf{0}$ the relative phases and amplitudes of the port output waves produced by the source are well-defined and given by the ratios of the various component phasors of \mathbf{b}_S . Generally, this will not be the case for internal sources of noise, so they must be modeled differently.

In high-frequency circuits nearly all randomly-generated sources of noise fall into three basic categories: *thermal noise*, generated by the statistics of the microscopic degrees of freedom of a system at a finite temperature; *quantum noise*, a fundamental, limiting level of background noise present at the outputs of phase-preserving (coherent), high-gain systems such as amplifiers and mixers; and *shot noise*, generated in some types of devices when carrying a finite power flow and arising due to the discrete nature of charge (electrons) or quanta of radiation (photons). Other sources of random noise, such as *flicker* or *1/f noise* are usually relatively unimportant when analyzing the high-frequency behavior of a circuit or structure.²³ Note that both *quantum noise* and *shot noise* arise from the particle nature of “fields”: both originate from the fundamentally “quantum nature of nature.”

We start by considering the thermal noise generated by a passive circuit, using a simple, 1-port termination of a transmission line to illustrate the idea (Fig. C-3 on page

²³ Actually, $1/f$ noise does play a major role when considering the gain stability of very high-gain systems such as astronomical instruments: the gain of a system often displays $1/f$ -like variations at time scales longer than a few seconds, and steps must be taken to mitigate that effect.

130). Assume that no coherent wave signals a or b are present. When the transmission line is in thermal equilibrium at temperature T , it will have a Bose-Einstein distribution of energies in thermally-generated photons on the line: photons traveling in both directions and at all frequencies and with random phases: *black-body* radiation.²⁴ If the termination is perfectly reflective, $|\Gamma(\omega)| = 1$, then radiation from the line entering the termination will reflect back with the same power regardless of frequency, and thermal equilibrium of the radiation on the line will be maintained.

If, however, the 1-port termination absorbs some thermal radiation from the line because $|\Gamma(\omega)| < 1$, then not all of the power in the incoming thermal radiation from the line will reflect. Thus the equilibrium distribution of radiation on the line will be disturbed by the termination's absorption. When in thermal equilibrium with the transmission line, the termination must therefore emit thermal radiation with power just sufficient to replace that absorbed from the line: power proportional to $1 - |\Gamma(\omega)|^2$. We desire a formula to describe this extra thermal radiation emitted by the 1-port when it is in thermal equilibrium at temperature T . Before we do this, however, we need to digress for a bit in order to understand the nature of the sources of random fluctuations present in passive channels and dissipative terminations.

The random fluctuations induced by a Planck black-body distribution of incident thermal wave power from a transmission line (only one boson mode at each frequency: those waves traveling right-to-left in Fig. C-3) have a power density of (note the use of the frequency in hertz, $\nu = \omega/2\pi$):

$$\frac{dP_n}{d\nu} = \frac{h\nu}{\exp(h\nu/k_B T) - 1} \quad (\text{C.7})$$

where k_B is, of course, Boltzmann's constant. For a receiver system which detects an incoming RF source by first generating an output proportional to the long-term average RF power absorbed by its detector (for example, a direct-detection receiver using a bolometer as its first active element; see the Glossary), the black-body thermal noise

²⁴ Bose-Einstein statistics arise from the quantum nature of the electromagnetic field, as well. Thus all of the noise sources we are considering are of a fundamentally quantum origin.

distribution along with any shot noise generated in the detector itself is pretty much the whole story.

On the other hand, if the receiver processes an incoming RF signal using some sort of active gain device which responds to the input channel's instantaneous amplitude (for example, a linear, high-gain RF amplifier or, as in our case, an SIS mixer) the formula (C.7) for the system's *equivalent input noise power* must be modified. For example, consider the spectral density of the voltage fluctuations at the input of an amplifier which are generated by an impedance-matched source with resistance $R(\nu)$. In this case, $\langle V^2 \rangle = 4R\langle P_n \rangle$, and the black-body power density (C.7) would generate mean squared voltage fluctuations with spectral density given by

$$\langle V^2(\nu) \rangle = 4R(\nu) \frac{dP_n}{d\nu} = 4R(\nu) \frac{h\nu}{\exp(h\nu/k_B T) - 1}.$$

It turns out that this formula is incomplete, as first pointed out in 1951 by Callen and Welton [42]. Even at zero temperature each harmonic wave mode will exhibit ground-state amplitude fluctuations completely analogous to that of a quantum harmonic oscillator with the same characteristic frequency. Thus the correct voltage fluctuation formula should add the mode ground-state energy $h\nu/2$ to the boson distribution term. Phase-coherent, linear amplification provided by the electrical circuitry we wish to model with SuperMix works by amplifying the instantaneous amplitude of the source. The black-body formula SuperMix uses for such a circuit's *equivalent input noise power* is therefore augmented by including this ground-state *quantum "noise"* and it becomes

$$\left. \frac{dP_n}{d\nu} \right|_{\text{input-referred}} = \left(\frac{h\nu}{\exp(h\nu/k_B T) - 1} + \frac{h\nu}{2} \right) = \frac{h\nu}{2} \coth\left(\frac{h\nu}{2k_B T}\right). \quad (\text{C.8})$$

We must be careful how we interpret this Callen-Welton formula. The input-referred equivalent noise power expression (C.8) does not mean to imply that there is actually power flow from the source channel into our detector originating from these ground-state amplitude fluctuations. No device can have the net effect of transferring energy out from a mode already in its ground state, because, of course, that mode has no lower energy state into which it can transition. Rather the only net effects on our phase-coherent system of an input mode's ground-state amplitude fluctuations are to stimulate transitions

of one or more of the system's internal modes. This situation is analogous to how the spontaneous transition of an excited atom to a lower energy level can be analyzed as emission stimulated by the electromagnetic field's ground-state fluctuations. On the other hand, ground-state fluctuations of the field cannot result in the spontaneous, permanent excitation of an otherwise-closed system to a higher energy level.²⁵

Keeping this important caveat in mind, then, at temperatures where $h\nu \ll k_B T$, expression (C.8) takes the Rayleigh-Jeans limiting form $dP_n/d\nu = k_B T$. This form motivates the conventional expression of any system's input-referred noise power spectral density $dP_n/d\nu$ as its *noise temperature* T_n at that frequency, as first mentioned in the Glossary and in Chapter 2:

$$T_n(\nu) \equiv \frac{1}{k_B} \left. \frac{dP_n}{d\nu} \right|_{\nu(\text{input-referred})} . \quad (\text{C.9})$$

The *definition* of noise temperature T_n as a measure of the input-referred $dP_n/d\nu$ is valid for any source of noise, and not only noise of thermal origin. By "input-referred," we mean that the total noise signal at the output of the system is interpreted as generated by an added noise signal at the input of an equivalent but noise-free system which would reproduce the observed output noise.²⁶ What this means in practice is that a system's input-referred noise power spectral density is calculated by dividing its output noise power spectral density by its signal power gain at each frequency.

Returning to our 1-port terminator of Fig. C-3, at each frequency ν it has absorbed a fraction $1 - |\Gamma(\nu)|^2$ of the incident noise, where that noise was distributed with temperature T according to (C.7). For the passive terminator to remain in thermal

²⁵ Ground state fluctuations do play a role in short, *virtual* transitions of a system to a higher energy which are then quickly followed by transitions to an even lower energy state, so that the electromagnetic field gets a net energy transfer *from* the system.

²⁶ To reiterate, the *total noise signal* from a linear, phase-coherent system's output also includes the possibly amplified ground-state amplitude fluctuations of the bosonic modes at the system's inputs, equation (C.8). Thus, for our purposes, we include the extra Callen-Welton $h\nu/2$ of ground-state input noise fluctuations as part of the total system noise when calculating its noise temperature. Not everyone agrees with this convention: see, for example, Kerr [45].

equilibrium with the input radiation, it must emit thermal noise which will exactly replace the amount it has absorbed. Therefore, its output noise power density $dP_n/d\nu$ must be

$$\left. \frac{1}{k_B} \frac{dP_n}{d\nu} \right|_{\nu(\text{output})} = \left[1 - |\Gamma(\nu)|^2 \right] \left. \frac{1}{k_B} \frac{dP_n}{d\nu} \right|_{\text{black-body}}. \quad (\text{C.10})$$

Note how $dP_n/d\nu$ is expressed in temperature units by dividing it by k_B . Expression (C.10) may be generalized in a straightforward fashion to any passive n-port with scattering matrix \mathbf{S} and at temperature T , resulting in the passive device's *noise correlation matrix* \mathbf{C} :

$$\mathbf{C}(\nu) = \left[\mathbb{I} - \mathbf{S}\mathbf{S}^\dagger \right] \left. \frac{1}{k_B} \frac{dP_n}{d\nu} \right|_{\text{black-body}}.$$

Clearly, in this formula \mathbf{C} is expressed in temperature units. If a network or circuit absorbs no power, then its \mathbf{S} matrix must be unitary, so $\mathbf{S}\mathbf{S}^\dagger = \mathbb{I}$, and its noise correlation matrix elements all vanish. Otherwise, the diagonal elements of $\mathbf{C}(\nu)$ provide the circuit's added noise power spectral density at each of its ports.²⁷ The off-diagonal elements give the complex noise phasor correlation coefficients: $C_{ij} = \langle c_i c_j^* \rangle$, again in temperature units.

Since this paper is concerned with a receiver system incorporating a phase-coherent RF detector (its SIS mixer), the noise correlation matrices of our passive RF elements should incorporate the input-referred, Callen-Welton ground-state noise fluctuations of (C.8):

$$\mathbf{C}(\nu) = \frac{h\nu}{2k_B} \coth\left(\frac{h\nu}{2k_B T}\right) \left[\mathbb{I} - \mathbf{S}\mathbf{S}^\dagger \right]. \quad (\text{C.11})$$

²⁷ Unlike the formula (C.9), the diagonal elements of \mathbf{C} , although expressed in temperature units, *are not* noise temperatures $T_n(\nu)$ because they are not “input referred.” A multiport \mathbf{S} matrix, in and of itself, doesn't differentiate between “input” and “output” ports: in general, all ports have incident waves and emit outgoing waves.

Equation (C.11) provides an adequate noise model for passive, linear network elements which do not amplify incident waves, and this noise calculation is the default provided by SuperMix to determine the noise of a sub-circuit when given its \mathbf{S} matrix. It assumes that the frequencies are high (no $1/f$ noise), and that the currents present in the circuit are small and generate negligible levels of shot noise. In this case, passive thermal noise calculated using (C.11) is appropriate. In the unlikely event that the output of such a network of passive circuit elements never reaches an active, high-gain, phase-preserving amplifier of some sort, then ground-state quantum fluctuations should not be included in noise calculations. The unembellished, Bose-Einstein black-body expression of (C.7) should instead be used, which SuperMix will not do. This is unimportant for low-frequency circuitry, because in that case $h\nu \ll k_B T$ and the expression (C.11) will approach the Rayleigh-Jeans result.

The excess quantum noise introduced by high-gain amplifiers

For a phase-coherent (phase-preserving) system such as an SIS mixer or any other high-gain, linear amplifier, even the Callen-Welton equation (C.11) with its ground-state fluctuation-induced noise is still not the whole story: we need to include an additional element of quantum noise which all high-gain, phase-coherent systems must generate. An excellent, relatively easily followed theoretical analysis of the quantum noise limits of high-gain, phase-coherent systems is the subject of a 1982 paper by Caltech's Carl Caves [43]. In this paper, he demonstrates that: (1) there is an uncertainty principle relationship between the real and imaginary parts of a coherent wave's complex amplitude phasor at the system's output,²⁸ and (2) this fact implies that for systems which greatly amplify both the real and imaginary parts of an input's phasor (phase-insensitive, coherent amplification), there must be additional noise at the amplifier's output corresponding to an added system noise temperature of at least another $h\nu/2k_B$. Along with the input channel's inherent, independent quantum-fluctuation-induced noise floor of $h\nu/2k_B$, this

²⁸ Or, equivalently, the amplitudes of the amplifier output's in-phase ($\cos\omega t$) and quadrature-phase ($\sin\omega t$) components.

implies that the quantum noise temperature limit of any coherent, phase-insensitive, high-gain amplifier is $h\nu/k_B$, where the frequency ν corresponds to that of the input signal, not the output signal (e.g. the RF frequency of a mixer, not its IF frequency). Caves's analysis has successfully withstood the attacks of very clever skeptics through the years since its publication, and stands today as a remarkable, nearly universally accepted conclusion.

The ports of a linear, coherent amplifier exchange incident and outgoing bosonic excitations (photons, in our case). One port (the input) has an incident photon rate representing the signal to be amplified (plus, at a minimum, its associated quantum noise amplitude fluctuations); another (the output, which may be the same port as the input) generates outgoing photons which are phase-coherent with those of the signal input along with some number of incoherent, noise photons. Caves defines the amplifier's gain G as the ratio of the number of the signal input-stimulated bosons at the output to those signal bosons which were absorbed by the input (to simplify the following formula we assume that G is nonnegative real; if the output frequency is the same as that of the input, then G is the amplifier's power gain). He then shows that, for any particular input frequency, the uncertainty relation between the amplifier output phasor's real and imaginary parts can be satisfied only if some other, independent source of quantum noise is also present, and its noise leads to a total input-referred noise temperature of at least:

$$T_n(\nu) \geq \underbrace{\frac{h\nu}{2k_B}}_{\text{input noise}} + \underbrace{\frac{h\nu}{2k_B} \left| 1 - \frac{1}{G} \right|}_{\text{minimum possible noise added by amplifier}} = \frac{h\nu}{k_B} \underbrace{\left| 1 - \frac{1}{2G} \right|}_{\text{valid for } G \geq 1}. \quad (\text{C.12})$$

Equality in (C.12) could only be achieved by an ideal, quantum-limited device at 0 Kelvin. Also note again that $h\nu/2k_B$ of the amplifier's noise temperature comes from its response to the input source channel's quantum noise amplitude fluctuations and is independent of the amplifier's internal workings. Multiplying (C.12) by G would give the circuit's output noise power; upon a little reflection, the reader should be able to see that this result is indeed consistent with the purely passive noise calculations of (C.10) and (C.11).

For the case of an ideal, 2-port,²⁹ high-gain amplifier, the possible internal sources of the other $h\nu/2k_B$ implied by (C.12) seemed too obscure to be acceptable to some critics. On the other hand, it should be clear that an amplifier which outputs a much higher power and therefore a much greater photon rate than that provided by the input must contain some source of those photons. Because the amplifier's output wave is phase-coherent with the input, this photon source must somehow involve one or more internal amplifier modes with transition frequencies comparable to the input frequency (these internal modes and transitions may be virtual). Consequently the amplifier's internal modes would then also serve as independent sources of quantum ground-state amplitude fluctuations, stimulating the generation of more noise terms of at least $h\nu/2k_B$. In the case of a simple, double-sideband mixer such as the SIS detector used in the author's heterodyne receiver, the input fluctuations present at each of the two RF sidebands (upper and lower) excite noise contributions of at least $h\nu/2k_B$ to the single IF output, so the inequality (C.12) is certainly satisfied. Actually, all RF sidebands will contribute various fractions of their quantum noise to the total mixer noise, along with thermal fluctuations determined by the sidebands' termination temperatures. The SIS device itself contributes additional thermal noise, and the nonzero DC current through the SIS (due to leakage currents and the LO photon step in its DC I-V curve) contributes shot noise. A detailed quantum theory of SIS mixer noise, consistent with Caves's result, is derived in Michael Wengler's thesis [44]. All of these additional noise terms make near-ideal, quantum-limited performance a real challenge.

Y-factor and SSB vs. DSB noise in heterodyne receivers

How is the noise performance of a heterodyne receiver based on a simple, single-detector mixer characterized and measured? As briefly discussed at the end of the last section, such a receiver qualifies as a phase-insensitive, coherent photon amplifier for signals present in one of its RF sidebands. At any particular IF frequency, the receiver's total output noise includes contributions generated by the quantum and thermal noise

²⁹ One port serves as the input, the other is the output port.

present in the receiver's RF sidebands along with thermal and shot noise contributions generated by the SIS and the IF circuitry. The relative intensity of the desired IF output signal to that of the IF output noise defines the receiver's signal to noise ratio (S/N). If the RF input signal intensity to the system is known, then the receiver's IF output S/N could be used to determine the receiver's equivalent input noise temperature using (C.9).

Performing this noise temperature calculation, however, requires one to know the receiver system's RF→IF power gain, a potentially elusive quantity requiring careful calibration. The *Y-factor* method is an alternative commonly used to estimate the system's noise temperature without requiring a full gain calibration. If the receiver's RF→IF power relationship is linear and *the system's noise temperature is independent of the RF input signal power*, then measuring the IF output powers for two known RF signal input intensities would be sufficient to determine both the receiver system's power gain and, more importantly, its equivalent input noise temperature. This is the essence of the Y-factor method. With two RF input intensities (in temperature units) T_h and T_c ("hot" and "cold," respectively), the Y-factor is defined as the corresponding IF output power ratio P_h/P_c . If the system is linear and its equivalent input noise temperature is T_n , independent of the RF intensities T_h and T_c , then

$$Y \equiv \frac{P_h}{P_c} = \frac{T_h + T_n}{T_c + T_n},$$

$$\text{and } T_n = \frac{T_h - YT_c}{Y - 1}. \quad (\text{C.13})$$

In practice, the two calibrated RF signals are most often supplied by efficient, broadband, black-body radiators at different physical temperatures: one usually at room temperature, typically near 297 K, and one first cooled by immersing in liquid nitrogen, thus typically near 80 K. For example, assume that during lab testing holding each of these two radiators in turn near the RF signal input window of the 180–300 GHz heterodyne receiver system resulted in IF output power levels of 327 and 117 μW . Assume further that the effective IF pass band for these power measurements was 4–8 GHz and the LO frequency was 240 GHz. At these frequencies and temperatures, black-body radiation from an RF source is well into the Rayleigh-Jeans regime so that the signal intensity each then injects into the receiver's RF input channel is proportional to its

physical temperature and is independent of RF frequency. The measured $Y=327/117=2.80$. Using hot and cold temperatures of 297 and 80K in (C.13), the resulting estimated average receiver noise temperature over this frequency range would be $T_n = 41$ K, but see the important caveat explained at the end of this section.

The above example gives an experimental determination of the simple heterodyne receiver's *double-sideband* noise temperature $T_{n(\text{DSB})}$. The RF sources used were broadband, black-body radiators which injected signal power into both the upper and lower RF sidebands of the heterodyne receiver (as well as into all of its higher order RF sidebands). A simple heterodyne mixer formed from a single SIS device can convert signals from the two primary RF sidebands into the IF output with roughly equal efficiencies, so that the total IF output power produced by a broadband RF source is approximately twice that produced by an equal intensity source confined to only a single sideband. Consequently, the observed IF output S/N is twice what would be expected from an equal intensity source confined to a single RF sideband. The *single-sideband* receiver noise temperature will therefore be approximately twice that determined using (C.13): $T_{n(\text{SSB})} \approx 2 T_{n(\text{DSB})}$.³⁰ The above example would then result in a single-sideband noise temperature estimate of 82K, the relevant value when determining the sensitivity of the receiver for obtaining line spectra of astronomical sources (as in Figure 1-2 on page 5). Note that this noise value is approximately 7 times the quantum limit at 240 GHz, namely 11.5 K. The literature nearly always quotes values for $T_{n(\text{DSB})}$ obtained by measuring Y-factors rather than converting them to their single-sideband equivalents (probably because they look better!). A sideband separating or image rejecting heterodyne receiver would generate an IF output dominated by the RF signal from one

³⁰ Actually, doubling $T_{n(\text{DSB})}$ provides only an average of the upper and lower sideband $T_{n(\text{SSB})}$ noise temperatures. Because their RF frequencies differ by twice the IF output frequency, the sideband noise contributions differ slightly, and the SIS RF embedding impedances of the two sidebands will differ as well, further affecting their conversion gains and noise temperatures. It also ignores the broadband RF signals introduced to the higher-order sidebands and their contributions to the measured IF output S/N . These effects may be quite small in most cases, but they exist.

sideband only, and Y-factor measurements would then provide much better direct estimates of $T_{n(SSB)}$.

One final and very important caveat concerning this technique: when the physical temperatures of both the hot and cold sources are large compared to the quantum noise limit $h\nu/k_B$, then the bosonic, quantum contributions to their RF amplitude fluctuations described by the Callen-Welton formula (C.8) on page 136 have all but disappeared. The ratio of their squared RF amplitude fluctuations is very nearly equal to the ratio of their physical temperatures, and the extra quantum noise contributions of the sources described in the previous sections are much less than $h\nu/k_B$. Return now to our 240 GHz example with hot and cold source physical temperatures of 297 K and 80 K. The quantum limit noise temperature is $h\nu/k_B = 11.5$ K, but calculations using (C.8) show that the equivalent noise temperatures of the sources are only increased to 297.04 K and 80.14 K, respectively, and not increased by the full quantum limit value! Thus, at these source temperatures, Y-factor estimates of a receiver's T_n will include only an insignificant fraction of the quantum noise limit.

On the other hand, a Y-factor estimation of the atmosphere-telescope-receiver total system noise temperature while observing at a facility such as the CSO is accomplished by: (1) inserting an ambient temperature, black-body absorber to inject a T_h source into the RF beam (usually between the telescope's Cassegrain mirror and the following optics), and (2) then looking at blank sky, providing an assumed $T_c \approx 0$ K. The standard Y-factor formula becomes $T_n = T_h / (Y - 1)$, and this estimate then not only includes the extra thermal noise and losses from the atmosphere and telescope, but also includes the full quantum fluctuation contribution in the measured Y . Using $T_c = 0$ K in the formula (C.13), although its actual equivalent noise temperature would be, from (C.8), $h\nu/2k_B$ in each sideband, thus reduces the measured Y value, effectively adding this equivalent noise temperature to the receiver's calculated T_n .³¹ These problems regarding the proper interpretation of Y-factors and other frustrations introduced by the nonlinearity in the

³¹ In fact, this method slightly over estimates the total system noise, including quantum noise, as one can easily show. The error is typically on the order of a few percent. Assuming a value for T_c which is a few degrees above 0 K can correct this error.

Callen-Welton formula were raised in 1999 by Kerr [45]. Over two decades later, a consensus within the submillimeter wave community on how to consistently characterize coherent amplifier and detector noise is yet to be reached.

Appendix D: RECEIVER MODEL PREDICTED RESULTS

This appendix provides more complete presentations of modeled mixer performance predictions than those presented in the main body of the text.

Embedding impedance (reflection, Γ_{RF} or Γ_{IF}) study results

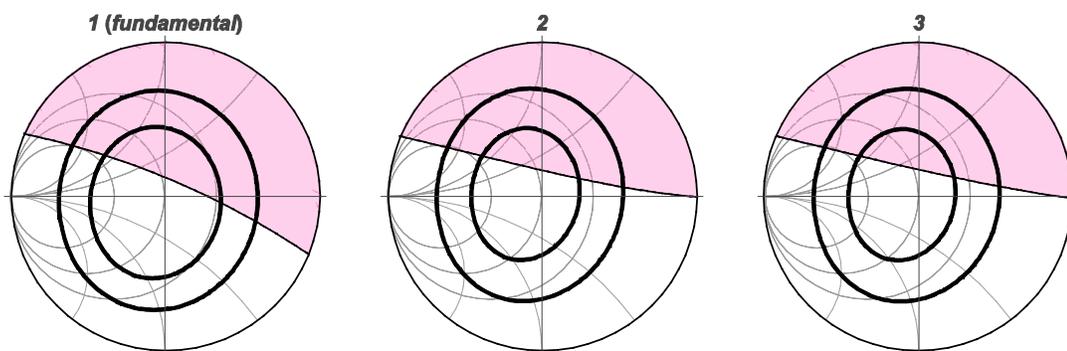
This section provides a more extensive collection of model results for the SIS embedding impedance studies discussed in Chapter 2. Smith Y-type charts with a normalizing impedance of the SIS normal resistance R_n are used for the RF embedding impedance model results; Z-type charts with a $3R_n$ normalizing impedance are used for the IF embedding results. Unless otherwise noted, the measured mixer SIS I-V curve of Figure 2-3 on page 22 was used, along with 4 K SIS temperature, $V_{\text{gap}} = 2.79$ mV, and $R_n = 7.7$ ohm. The physical SIS capacitance is included as part of the RF linear embedding circuitry, and not as part of the SIS device model. Mixer noise temperature T_n calculations assumed that the RF input source temperature was 0 K, and the resulting ground-state fluctuation densities of $h\nu/2k_B$ at all RF sidebands are included in the mixer T_n model. The mixer “quantum limit” T_n was then calculated using the LO frequency: $h\nu_{\text{LO}}/k_B$.

Results begin on the next page.

Number of harmonic sidebands used in the analyses

The mixer analyses used 2 LO harmonics (4 sideband frequencies), although the SIS operating state was calculated assuming a sinusoidal excitation (no higher LO harmonics). The RF embedding impedance was assumed to be the same at all of these sideband frequencies, since the model does not assume a particular RF circuit configuration. Shown below are the effects on the RF and IF embedding model results for various numbers of harmonics. Note that going to 2 harmonics resulted in a noticeable change, whereas adding even more harmonics had little further influence.

RF embedding results:



IF embedding results:

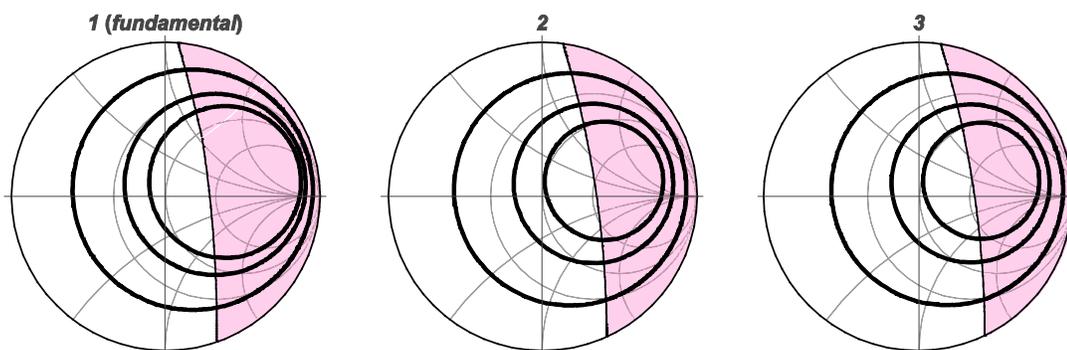


Fig. D-1: Effect of number of harmonics on the mixer analyses; LO 240 GHz, IF 8 GHz. The RF results show mixer noise contours at 2 and 3 times the quantum limit along with the IF instability regions. The IF results show mixer gain contours of -3 , 0 , and $+1$ dB along with the RF instability regions. Analyses using only the fundamental RF sidebands (1 harmonic) predict slightly lower noise temperatures and slightly higher conversion gains than analyses using more harmonics. The predicted instability regions are also slightly affected. Models are for RF USB and assumed SIS DC bias at the center of the primary photon step with LO pumping $\alpha = 1$.

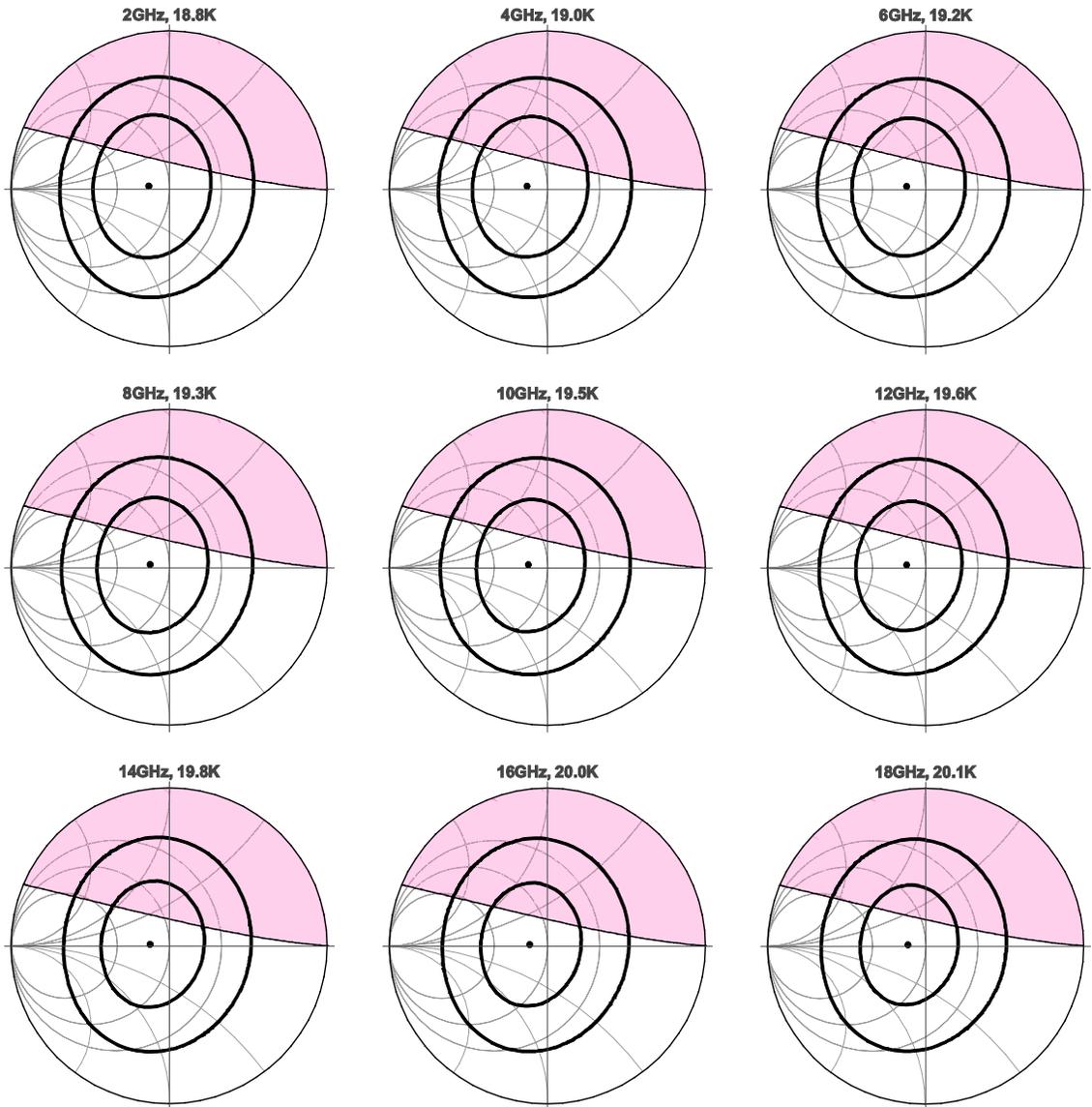
RF embedding model results

Fig. D-2: Mixer T_n contours and min T_n vs. IF frequency; LO 240 GHz, Upper Sideband. SIS DC bias at the center of the primary photon step with LO pumping $\alpha = 1$. Shown by the point inside the noise contours, the RF embedding impedance for minimum T_n is nearly independent of IF frequency, changing by less than 3% over the IF frequency range shown.

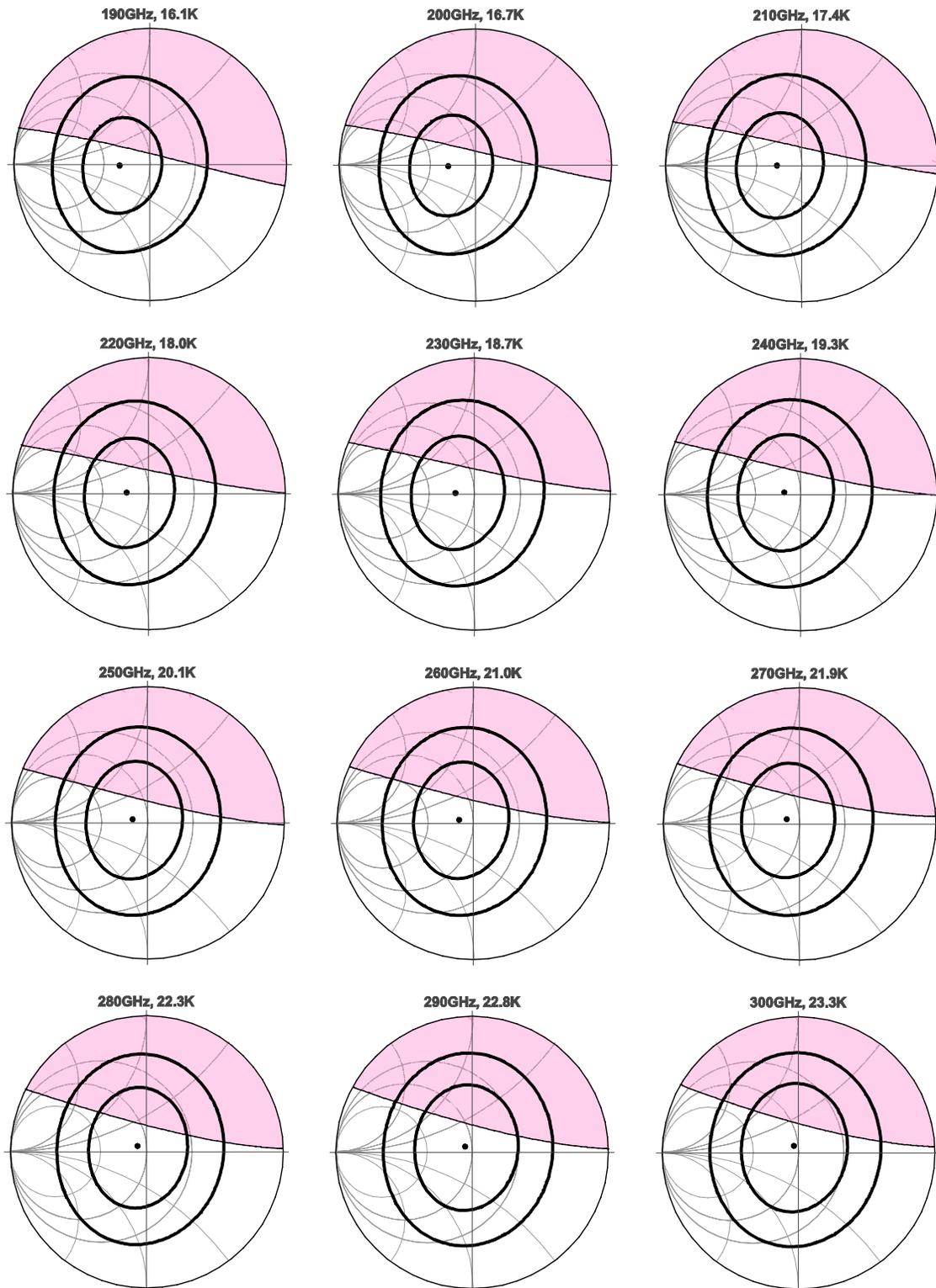


Fig. D-3: Mixer T_n contours and min T_n vs. LO frequency; IF 8 GHz, Upper Sideband. SIS DC bias at the center of the primary photon step with LO pumping $\alpha = 1$. The Z_{RF} for min T_n is very nearly real and rises from $0.64 R_n$ at 198 GHz to very close to R_n at 308 GHz.

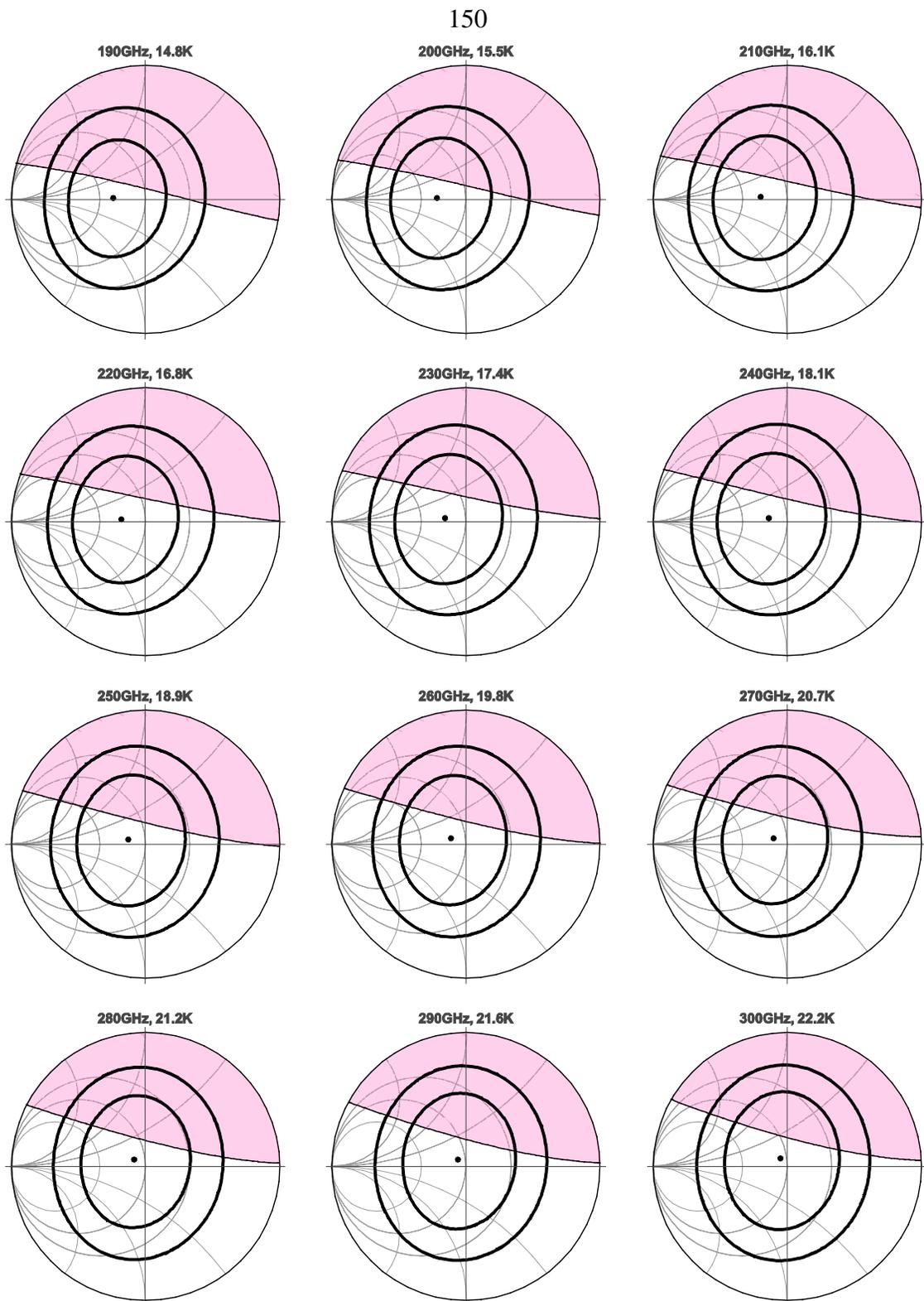


Fig. D-4: Mixer T_n contours and min T_n vs. LO frequency; IF 8 GHz, Lower Sideband. SIS DC bias at the center of the primary photon step with LO pumping $\alpha = 1$. The Z_{RF} for min T_n is very nearly real and rises from $0.61 R_n$ at 182 GHz to just below R_n at 292 GHz.

Mixer performance variation with SIS operating state

This section provides a more extensive collection of graphs showing the SuperMix predicted variations of mixer + LNA double-sideband (DSB), input-referred noise temperature versus SIS operating state (DC bias voltage and applied LO power). Results are presented as a function of LO frequency and for different choices for the mixer chip SiO dielectric constant. See Chapter 2 for a discussion of the details. DSB, input-referred noise temperatures are obtained by dividing the calculated system output noise power by the sum of the two calculated sideband power gains. Recall that the mixer chip design assumed a SiO dielectric constant value of 7.5. As expected, then, the modeled minimum noise temperatures for this dielectric constant value are lower than those found when the assumed value of the dielectric constant is changed.

Results begin on the next page.

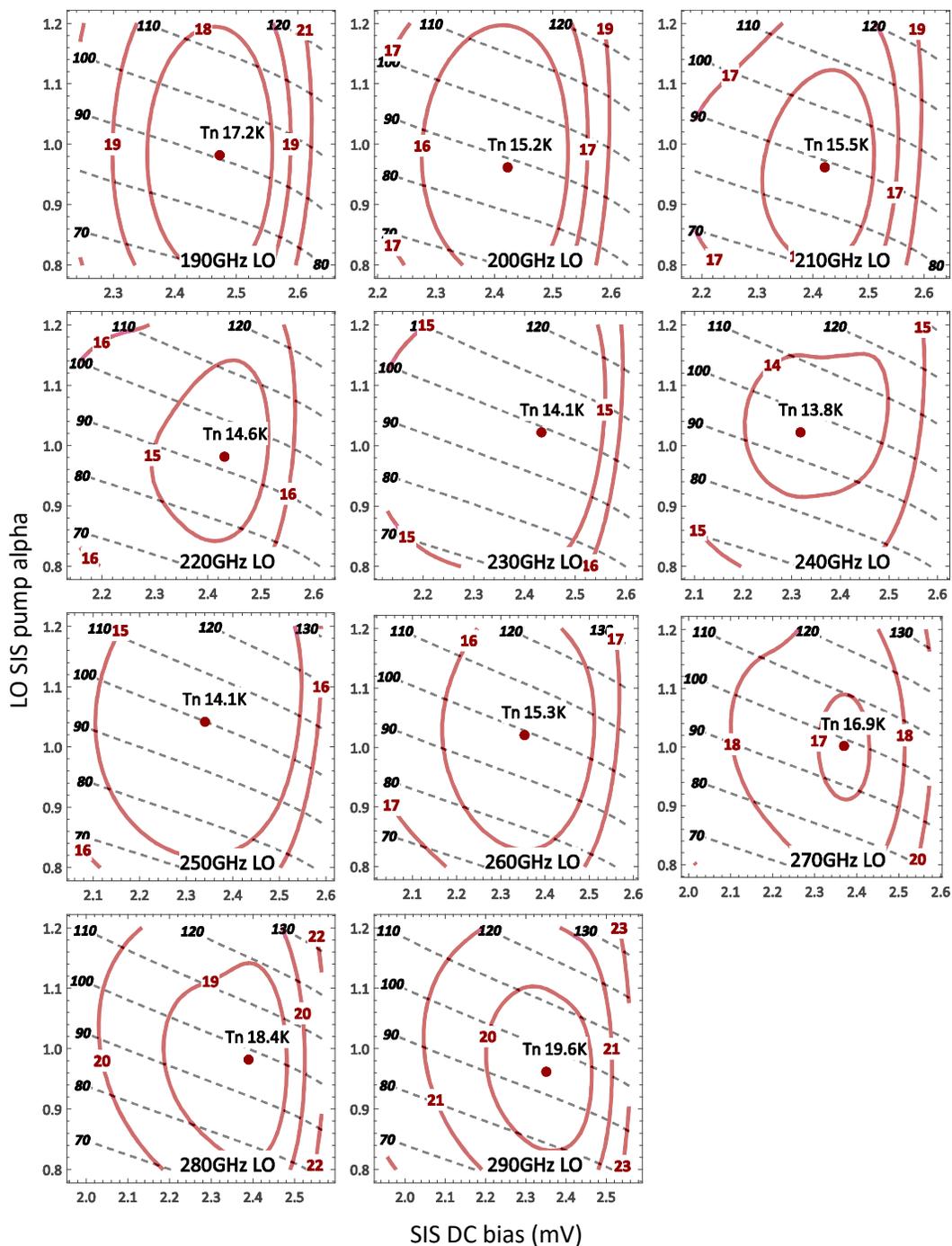


Fig. D-5: Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 5.6. Predicted double-sideband noise temperature as a function of SIS DC bias voltage (x-axes) and LO pumping α (y-axes) for various LO frequencies. Noise temperatures were calculated at an IF frequency of 6 GHz. The optimum SIS operating state is indicated by a red dot, and the corresponding minimum DSB T_n is indicated. The dashed lines denote SIS DC bias current levels (in microamps) corresponding to the SIS operating states.

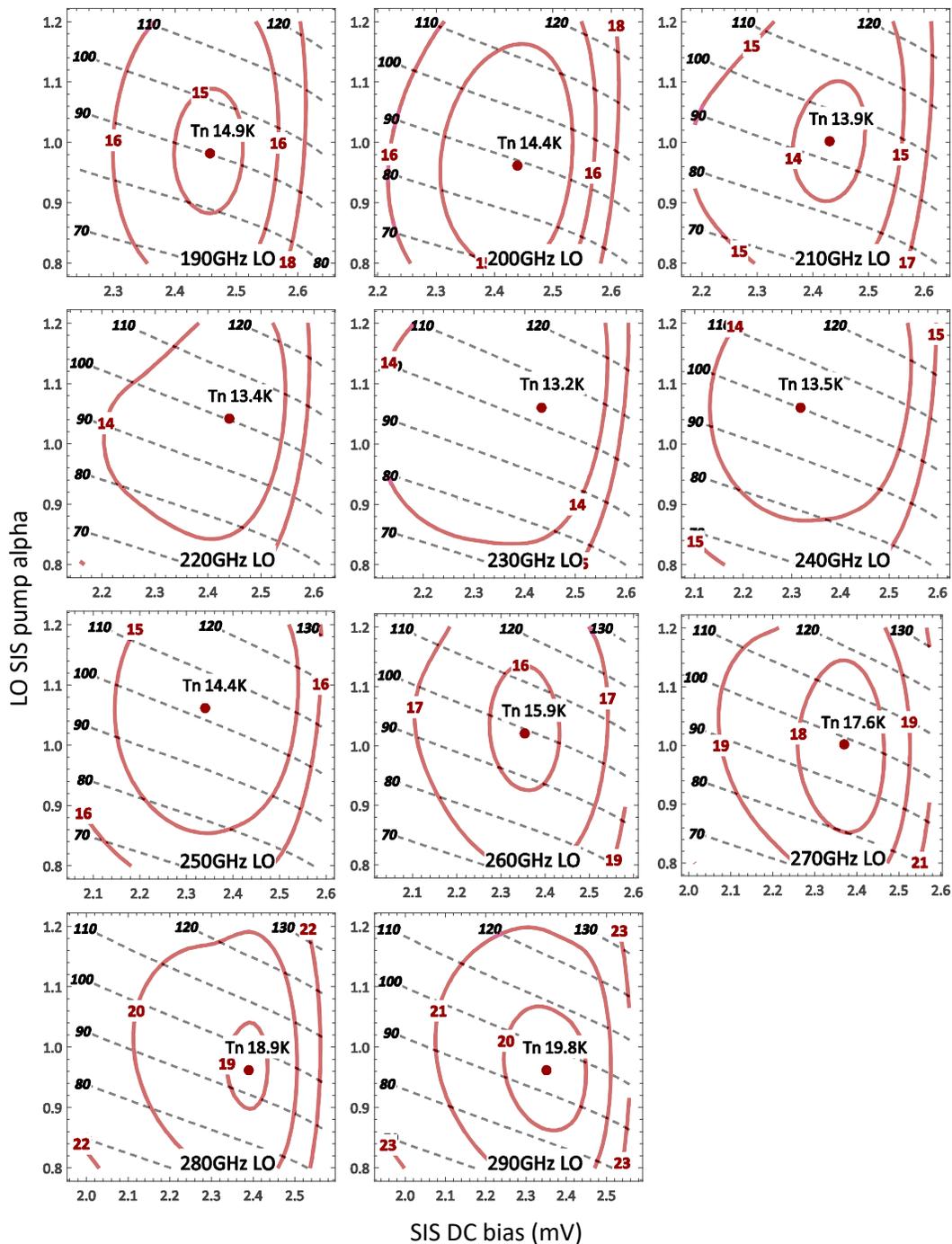


Fig. D-6: Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 6.2.

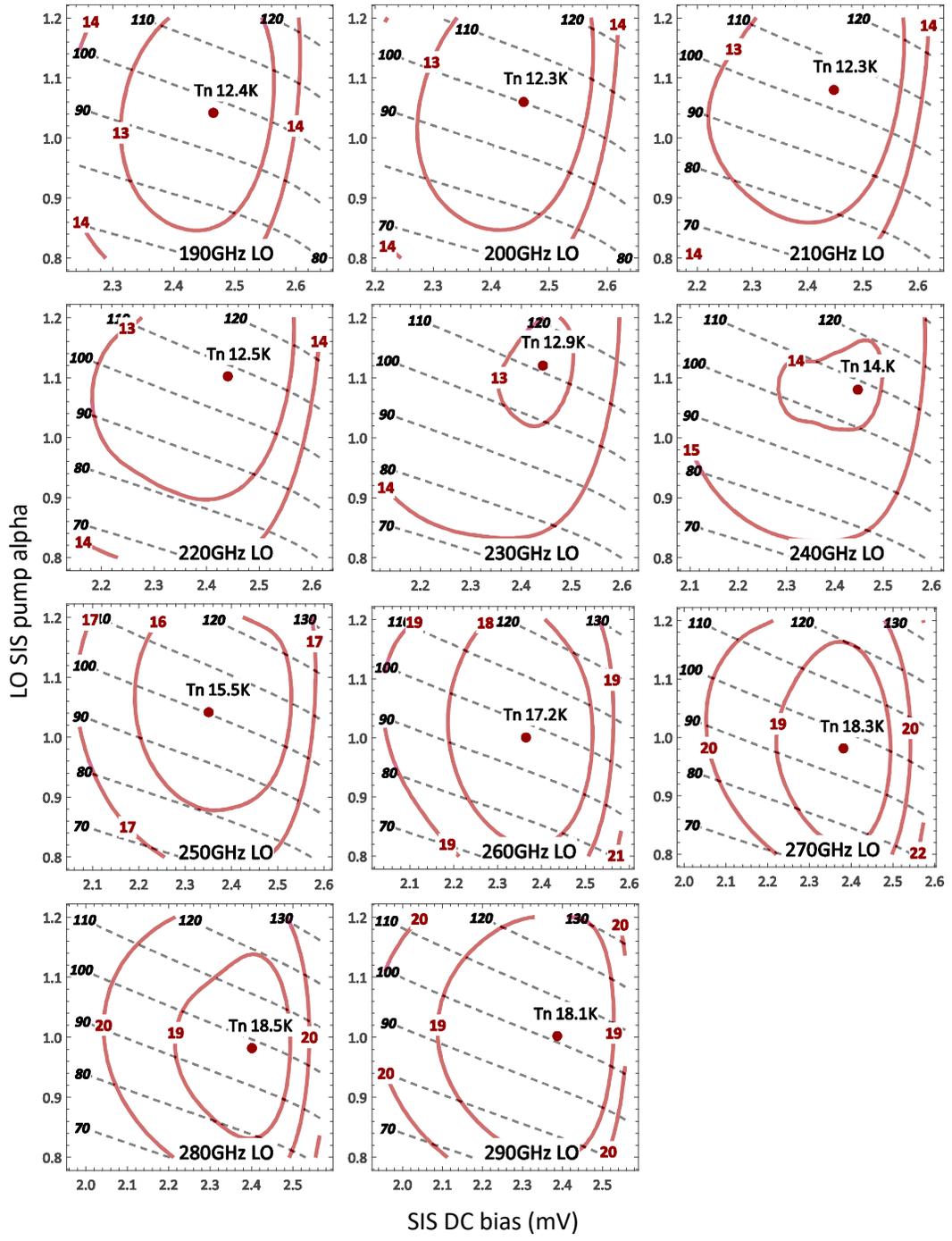


Fig. D-7: Mixer + LNA DSB T_n vs. SIS operating state for SiO dielectric constant = 7.5.

Appendix E : FAST SIS HARMONIC BALANCE ALGORITHM³²

The small signal response of an electronic circuit containing nonlinear elements, such as a mixer, is calculated by performing a perturbation analysis of its quiescent operating state. The operating state is defined by the large signal voltage and current waveforms across its nonlinear elements. Clearly it is essential to accurately determine these waveforms in order to proceed with the small signal analysis. *Harmonic balance* is one well-established method for accomplishing this task.

For the purposes of harmonic balance, a circuit containing two-terminal, nonlinear devices is typically modeled as shown in Fig. E-1, wherein each nonlinear device is connected to its own port of an otherwise linear embedding network. This embedding network includes the large-signal DC and AC sources which establish the circuit's overall operating state. The goal of harmonic balance is to determine the voltage and current waveforms $v(t)$ and $i(t)$ at each nonlinear device, thereby fixing its individual operating state and its resultant small signal behavior.

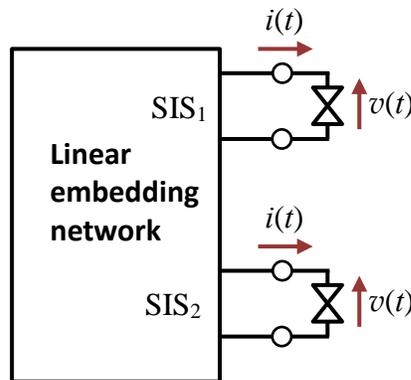


Fig. E-1: SIS junctions attached to a linear network.

Traditionally, harmonic balance proceeds by analyzing the linear network in the frequency domain, constructing its Thevenin or Norton equivalent at each of the

³² The derivation in this Appendix closely follows that first presented in a 1999 paper by Rice, Ward, et al. [50].

harmonic frequencies present in the network. Concurrently, a time domain calculation of the response of each nonlinear element is conducted starting with some suitable initial driving function (current or voltage) sourced by the network. The results of the two calculations are compared following suitable transformation; differences lead to adjustment of the harmonic signals present at the ports of the linear network and the corresponding time domain waveforms used for the driving functions of the nonlinear elements. The process is iterated until reaching convergence of the solutions. Hicks and Khan [46] provide a detailed analysis of this technique and its convergence properties.

In the case of mixers utilizing SIS devices as the nonlinear elements, the traditional harmonic balance approach outlined above has a very serious limitation: the time domain analysis must proceed for many cycles in order to determine the steady-state waveforms, greatly increasing the computation time required per iteration of the procedure. This limitation was overcome by Withington and Kollberg [22], who developed an algorithm for performing the nonlinear analysis of such devices purely in the frequency domain, a technique they refer to as spectral domain analysis. Starting from the mixer theory of John Tucker [1], they extended his results to include the effects on the SIS operating state of additional SIS excitation voltages at all harmonics of the LO frequency, thus developing the equations needed to analyze the frequency domain current response to the set of large-amplitude voltages induced by the DC bias and local oscillator (LO) source applied to a single SIS junction. Iteratively applying these induced harmonic currents to the SIS junction's linear embedding network results in improved estimates of the applied SIS harmonic voltages. Thus, harmonic balance may be achieved using calculations purely in the frequency domain.

This appendix extends those results to networks containing multiple SIS devices. The large signal solution is found using a multidimensional Newton-Raphson technique to achieve very rapid convergence, even when several harmonics and junctions are included in the analysis. This technique requires the determination of each SIS device's Jacobian matrix with respect to changes in the large signal voltages applied to it. As will be shown, this matrix can be easily generated from the small signal RF conversion matrix derived from Tucker's theory. A final improvement is to conduct the linear circuit calculations

using the scattering matrix representation rather than an impedance or admittance representation (which might not exist for certain circuit configurations).

Nonlinear calculation of the SIS harmonic currents

To perform the spectral domain analysis of the SIS operating state, the voltages and currents across each junction are represented as:

$$v(t) = V_0 + \operatorname{Re} \left[\sum_{m=1}^{\infty} v_m e^{jm\omega_{\text{LO}} t} \right]; \quad i(t) = I_0 + \operatorname{Re} \left[\sum_{m=1}^{\infty} i_m e^{jm\omega_{\text{LO}} t} \right] \quad (\text{E.1})$$

where V_0 and I_0 are the constant (DC) components, ω_{LO} is the local oscillator frequency, and the v_m and i_m are complex phasors. In the case where the circuit contains a single junction, the Fourier coefficients in (E.1) can be represented by the single, complex valued vectors \mathbf{v} and \mathbf{i} ; if there are multiple junctions, these vectors will still be used, but the individual components will be denoted by v_m^n and i_m^n , where m is the harmonic number ($m=0$ for the DC component) and n is the SIS junction index, ranging from 1 to the total number of junctions in the circuit. Withington and Kollberg's extension of Tucker's mixing theory provides an "admittance" representation of an SIS device: the junction currents \mathbf{i} are calculated from the junction voltages \mathbf{v} , $\mathbf{i} = \mathbf{i}(\mathbf{v})$.

The currents through an SIS junction may be calculated as follows [22]:

$$i_0 = \operatorname{Re} [i_{(0)}]; \quad i_m = i_{(m)} + i_{(-m)}^*, \quad m > 0 \quad (\text{E.2})$$

$$\text{where:} \quad i_{(m)} = -j \sum_{k=-\infty}^{\infty} C_{(k)} C_{(k+m)}^* \mathbf{I}(V_0 + kV_{ph}), \quad -\infty < m < \infty. \quad (\text{E.3})$$

In this equation V_0 is the junction DC bias voltage, $V_{ph} = (\hbar/e)\omega_{\text{LO}}$ is the *photon voltage* of the LO frequency, and the $C_{(k)}$ are complex coefficients defined below. The parentheses around the subscripts in (E.2) and (E.3) are used to distinguish harmonic indices of Fourier coefficients which range over negative as well as positive values; harmonic indices without the parentheses may take on nonnegative values only. This notation will be used extensively throughout the rest of this appendix.

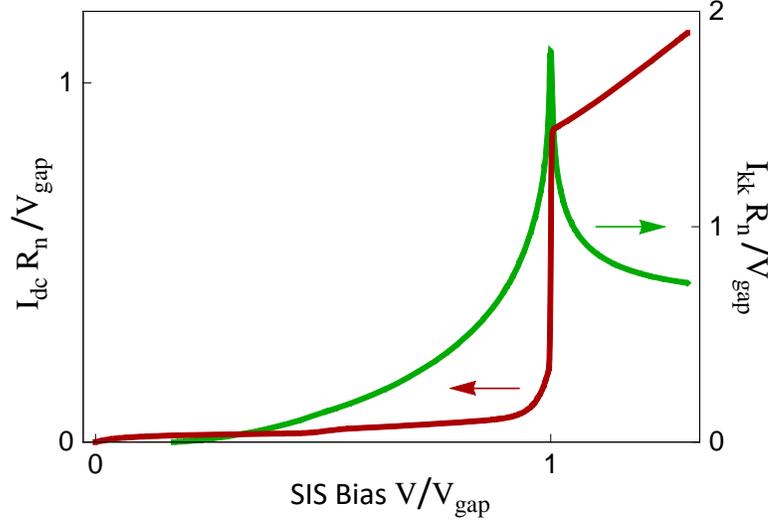


Fig. E-2: Example normalized SIS DC I-V curve and its Kramers-Kronig transform. The voltages and currents are normalized using the SIS gap voltage and normal resistance as shown. The DC I-V is shown in red, using the left-hand vertical scale; the right-hand scale is for the Kramers-Kronig transform (green). These curves were generated from the receiver's measured I-V data (Figure 2-3 on page 22). Note the sharp peak in I_{kk} near the SIS gap voltage.

In (E.3), $-jI(V)$ is a complex-valued, analytic extension of the SIS DC I-V characteristic of the SIS junction: $I(V) \equiv I_{kk}(V) + jI_{dc}(V)$. $I_{dc}(V)$ is the measured DC I-V characteristic curve of the junction, and $I_{kk}(V)$ is its Kramers-Kronig transform. A typical $I_{dc}(V)$, $I_{kk}(V)$ pair is shown in Fig. E-2, where both functions have been normalized so that the junction's gap voltage and normal resistance are both unity. The version of the Kramers-Kronig transform $I_{kk}(V)$ shown in Fig. E-2 is that due to Tucker, which formally subtracts off a conventional K-K transform of $I_{dc}(0)$ as shown in (E.4). Tucker's version ensures that $I_{kk}(0) \equiv 0$ while allowing a logarithmic divergence in I_{kk} at large V for many typical SIS I-V characteristic curves $I_{dc}(V)$. The resulting expression for $I_{kk}(V)$ is thus:

$$I_{kk}(V) = \frac{1}{\pi} P \int_{-\infty}^{\infty} I_{dc}(v') \left(\frac{1}{v' - V} - \frac{1}{v'} \right) dv'. \quad (\text{E.4})$$

In (E.4), " P " denotes the *Cauchy principle value* of the integral. $I_{dc}(V)$ is an odd function and $I_{kk}(V)$ an even function of the voltage V , so that $I(-V) = I^*(V)$.

The complex coefficients $C_{(k)}$ are defined by:

$$C_{(k)} = \lim_{n \rightarrow \infty} C_{(k)}^n; \quad C_{(k)}^0 = \delta_{0,k}; \quad C_{(k)}^{n>0} = \sum_{m=-\infty}^{\infty} C_{(k-nm)}^{n-1} A_{(m),n} \quad (\text{E.5})$$

where $\delta_{0,k}$ is Kronecker's delta and the $A_{(m),n}$ are given by:

$$A_{(m),n} = J_{(m)}(\alpha_n) e^{-jm\phi_n}. \quad (\text{E.6})$$

α_n and ϕ_n are the magnitude and phase of v_n , the n^{th} Fourier component of the SIS voltage, where its magnitude has been normalized by dividing by the photon voltage of the LO's n^{th} harmonic frequency, nV_{ph} :

$$\alpha_n e^{j\phi_n} \equiv v_n / (nV_{ph}) = ev_n / (n\hbar\omega_{LO}). \quad (\text{E.7})$$

Expressions (E.2) through (E.7) define the SIS harmonic currents i_n in terms of its harmonic voltages v_n , thereby defining the vector function $\mathbf{i}(\mathbf{v})$. Note, however, that the functions $A_{(m),n}$ defined by (E.6) and (E.7) are not analytic in the v_n ; this will be an important consideration when constructing the Jacobian matrix used by the Newton-Raphson algorithm.

The physics behind the tunneling current calculations

To understand the origin of the quite complicated expressions (E.2) through (E.7), consider the physics of single electron (actually, a single quasiparticle) tunneling through the insulating barrier of an SIS. The interaction between the many-particle states on either side of the barrier can be represented using a transfer Hamiltonian which can be quantified by considering the surface integral of the overlap of the states within the barrier (Cohen, Falicov, and Phillips [47]). The transfer Hamiltonian's effect on the SIS is then manifest in the observed tunneling current through the insulating barrier. The tunneling current will vary with the potential difference across the barrier because of the superconducting energy gap in the density of quasiparticle states (Bardeen [48]), giving rise to the observed SIS DC I-V characteristic (Fig. E-2).

Now introduce an oscillating potential across the barrier: $v(t) = V_0 + V_{LO} \cos(\omega_{LO}t)$. Assume that the frequency of the oscillation is so low that (in the absence of tunneling)

the superconductors on either side of the barrier would maintain their thermal equilibrium density matrices. The presence of the oscillating potential will cause the relative phase of the corresponding quasiparticle wave functions on the two sides of the barrier to pick up an extra $\omega_{\text{LO}}t$ term. Consequently, so will the phases of the tunneling Hamiltonian and its resulting tunneling current operator. The tunneling current matrix elements will therefore each have an additional factor of

$$\begin{aligned} \exp\left[-j\left(\frac{e}{\hbar}\right)\int V_{\text{LO}}\cos(\omega_{\text{LO}}t)dt\right] &= \exp\left[-j\left(\frac{eV_{\text{LO}}}{\hbar\omega_{\text{LO}}}\right)\sin(\omega_{\text{LO}}t)\right] \\ &= \sum_{m=-\infty}^{\infty} J_m(eV_{\text{LO}}/\hbar\omega_{\text{LO}}) e^{-jm\omega_{\text{LO}}t}. \end{aligned}$$

Clearly, this effect introduces tunneling currents at all harmonics of the LO frequency; this is the gist of Tucker's theory [1].

These harmonic currents, however, flow through the attached linear embedding network, inducing additional voltages at all harmonics across the SIS. These voltages in turn generate additional phase terms in the tunneling operator matrix elements, leading to more factors like the sum in the above expression. Expanding the product of all of these factors and gathering like terms results in a sum of terms like

$$J_m(eV_{\text{LO}}/\hbar k\omega_{\text{LO}})J_n(eV_{\text{LO}}/\hbar k'\omega_{\text{LO}})e^{-j(mk\omega_{\text{LO}}+nk'\omega_{\text{LO}})t}$$

for integers m , n , k , and k' . The Fourier transform of this product becomes a convolution, thus leading to the sequence of sums (E.7) to (E.3), which is Withington and Kollberg's extension to Tucker's theory [22].

Harmonic balance of the junctions and the linear network

The linear embedding network in Fig. E-1 has a single port for each junction. Its behavior will be represented by a frequency-dependent scattering matrix $\mathbf{S}(\omega)$ and an embedded wave source vector $\mathbf{b}_s(\omega)$, which will include the SIS junctions' DC bias voltages and the local oscillator source at ω_{LO} (we ignore its noise correlation matrix for the purposes of harmonic balance). Also indicated in Fig. E-1 are the conventional voltages and currents at the ports. As detailed in Appendix C, at any particular frequency

the scattering matrix representation of the linear embedding network imposes the following conditions on the vectors representing the various waves, voltages, and currents at the SIS ports (note that \mathbf{i} is the vector of currents *out* of the network's ports):

$$\begin{aligned}\mathbf{b} &= \mathbf{b}_s + \mathbf{S} \cdot \mathbf{a} \\ \mathbf{v} &= (\mathbf{b} + \mathbf{a})\sqrt{Z_0} \\ \mathbf{i} &= (\mathbf{b} - \mathbf{a})/\sqrt{Z_0}\end{aligned}\tag{E.8}$$

where Z_0 is the normalizing impedance of the scattering matrix representation of the network. A major advantage of this representation is that it nearly always exists, while some common network topologies lack an impedance or admittance representation. Note also that the voltages and currents in (E.8) are RMS, whereas the voltage and current phasors in (E.1) to (E.7) are peak amplitudes. This means that we must be careful about factors of $\sqrt{2}$ for the AC coefficients when solving the harmonic balance equations.

Each SIS junction will be excited by a set of harmonic voltages and a DC bias voltage, and each junction will respond by passing a set of harmonic currents and a DC bias current as described by the equations in the previous section: $\mathbf{i}=\mathbf{i}(\mathbf{v})$. With this nonlinear relationship and the linear equations (E.8), the operating states of the SIS junctions must satisfy the nonlinear vector equation (\mathbb{I} is the identity matrix):

$$Z_0(\mathbb{I} + \mathbf{S}) \cdot \mathbf{i}(\mathbf{v}) + (\mathbb{I} - \mathbf{S}) \cdot \mathbf{v} - 2\sqrt{Z_0} \mathbf{b}_s = \mathbf{0}.\tag{E.9}$$

This is the equation that must be solved for \mathbf{v} (and thus \mathbf{i}) by the harmonic balance routine, determining the junction operating states.

SuperMix uses a *Newton-Raphson* iterative root-finding algorithm to solve (E.9): given a vector function $\mathbf{y}(\mathbf{x})$ for which you want to find an \mathbf{x} such that $\mathbf{y}(\mathbf{x})=\mathbf{0}$, and the i^{th} guess $\mathbf{y}(\mathbf{x}_i)=\mathbf{y}_i$, generate a (hopefully improved) guess \mathbf{x}_{i+1} by extrapolating linearly to $\mathbf{y}=\mathbf{0}$, e.g. $\|d\mathbf{y}/d\mathbf{x}\|_{\mathbf{x}_i} \cdot (\mathbf{x}_{i+1} - \mathbf{x}_i) = -\mathbf{y}_i$. Using the Jacobian matrix $\|d\mathbf{y}/d\mathbf{x}\|$, one then solves this *linear* system for the \mathbf{x} vector increment $\Delta\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$. Taking the Jacobean matrix of the left-hand side of (E.9), a Newton-Raphson iteration requires the solution of:

$$\begin{aligned}2\sqrt{Z_0} \mathbf{b}_s - Z_0(\mathbb{I} + \mathbf{S}) \cdot \mathbf{i}(\mathbf{v}) - (\mathbb{I} - \mathbf{S}) \cdot \mathbf{v} \\ = [Z_0(\mathbb{I} + \mathbf{S}) \cdot \|d\mathbf{i}/d\mathbf{v}\| - (\mathbb{I} - \mathbf{S})] \cdot \Delta\mathbf{v}.\end{aligned}\tag{E.10}$$

The harmonic balance routine will seek a solution to the nonlinear system (E.9) by first calculating the linear network characteristics \mathbf{S} and \mathbf{b}_S at the relevant harmonic frequencies $\omega_m = m\omega_{LO}$. After assuming an initial trial SIS voltage vector \mathbf{v} , the routine must update \mathbf{v} by finding the SIS junctions' Jacobian matrix $\mathbf{Y} \equiv \|d\mathbf{i}/d\mathbf{v}\|$ and then solving (E.10) for the correction $\Delta\mathbf{v}$. The process is repeated until the corrections become sufficiently small. The actual implementation of a satisfactory algorithm is somewhat more complicated: it must continually monitor the convergence behavior of the iterations and be prepared to adjust $\Delta\mathbf{v}$. Potential difficulties include the algorithm's behavior near saddle points or local extrema of the left-hand side of (E.10). Based on a *Numerical Recipes* algorithm [49], the SuperMix Newton-Raphson solver (class `newton`) was implemented by Jonas Zmuidzinis in 1997.

Before proceeding with a derivation of the SIS junctions' Jacobian matrix $\|d\mathbf{i}/d\mathbf{v}\|$, let us again consider the elements of the vectors and matrices in (E.9) and (E.10), and how those elements should be indexed. The set of harmonic frequencies will be given by $\omega_m = m\omega_{LO}$ for nonnegative integers m ($m=0$ represents the DC bias). Members of the set of SIS junctions will be identified by the positive integer n , which will also designate the port numbering of the linear embedding network. As mentioned before, the elements of the voltage and current vectors in (E.9) and (E.10) will be identified by the notation v_m^n and i_m^n . The elements of \mathbf{S} connecting the different ports are evaluated at each frequency ω_m , and we can define $S_{m,m'}^{n,n'} = \delta_{m,m'} S^{n,n'}(\omega_m)$ (because the network is linear, frequencies do not mix within it). The nonlinear SIS relationship $\mathbf{i} = \mathbf{i}(\mathbf{v})$ will vary from junction to junction, so for junction n , $\mathbf{i}^n(\mathbf{v}^n) = \delta_{n,n'} \mathbf{i}^n(\mathbf{v}^n)$. This latter idea will also apply to the Jacobian matrix in (E.10): its nonzero elements will be $Y_{m,m'}^{n,n'} = \delta_{n,n'} (\partial i_m^n / \partial v_{m'}^n)$, again because harmonic currents in one SIS do not directly depend on voltages across a different SIS. In many cases, the elements of the embedded source vector, $(\mathbf{b}_S)_m^n$, will be nonzero only for the SIS DC bias voltages ($m=0$) and the local oscillator frequency ($m=1$).

Calculating the SIS Jacobian matrix

The Jacobian matrix $\mathbf{Y} \equiv \|d\mathbf{i}/d\mathbf{v}\|$ in equation (E.10) relates each individual SIS's harmonic currents to its applied voltages. Thus it takes the form of an admittance matrix,

Y. However, the individual complex-valued harmonic current phasors i_m^n are not analytic functions of the complex-valued harmonic voltages $v_{m'}^n$. This means that well-defined, unique complex numbers for the partial derivatives $\partial i_m^n / \partial v_{m'}^n$ are unavailable. We must instead work with pairs of complex-valued partial derivatives $\partial i_m^n / \partial \text{Re}[v_{m'}^n]$ and $\partial i_m^n / \partial \text{Im}[v_{m'}^n]$. We will find that these derivatives are derivable from the SIS device's heterodyne detector small-signal admittance matrix with complex-valued elements $Y_{(m,m')}$, the integers m and m' representing RF sidebands $\omega_m = m\omega_{\text{LO}}$ for all $-\infty < m < \infty$ (for harmonic balance calculations, we let $\omega_{\text{IF}} \rightarrow 0$, which then becomes a form of *homodyne* detection). The elements $Y_{(m,m')} = \partial i_{(m)} / \partial v_{(m')}$ are well-defined, complex-valued partial derivatives. There are enough elements in this matrix to form linearly-independent combinations which yield the $\partial i_m^n / \partial \text{Re}[v_{m'}^n]$ and $\partial i_m^n / \partial \text{Im}[v_{m'}^n]$ needed to use the Newton-Raphson iteration (E.10). In fact, as shown in Rice et al. [50], for positive integers m and m' :

$$\begin{aligned} \frac{\partial i_m}{\partial \text{Re}[v_{m'}]} &= Y_{(m,m')} + Y_{(m,-m')} \\ \frac{\partial i_m}{\partial \text{Im}[v_{m'}]} &= j [Y_{(m,m')} - Y_{(m,-m')}]. \end{aligned} \quad (\text{E.11})$$

If the i_m^n were analytic in the $v_{m'}^n$, then $\partial i_m^n / \partial \text{Im}[v_{m'}^n] = j \partial i_m^n / \partial \text{Re}[v_{m'}^n]$, which is clearly not the case.

The 0th components of \mathbf{i} and \mathbf{v} are DC voltages and currents, which we take to be real by definition (not to be confused with the analytic extension I_{kk} of the SIS DC I-V curve). These components also correspond to peak rather than RMS amplitudes. Taking all this into consideration, along with the limit that, as $\omega_{\text{IF}} \rightarrow 0$, then $Y_{(m,m')}^* = Y_{(-m,-m')}$:

$$\begin{aligned} \frac{\partial i_0}{\partial v_0} &= Y_{(0,0)}; \quad \frac{\partial i_{m \neq 0}}{\partial v_0} = 2Y_{(m,0)} \\ \frac{\partial i_0}{\partial \text{Re}[v_{m' \neq 0}]} &= \text{Re } Y_{(0,m')}; \quad \frac{\partial i_0}{\partial \text{Im}[v_{m' \neq 0}]} = \text{Im } Y_{(0,-m')}. \end{aligned} \quad (\text{E.12})$$

These are the relations we need to form the Jacobean matrix $\|d\mathbf{i}/d\mathbf{v}\|$ in (E.10) from the elements of the individual SIS small-signal admittance matrices $Y_{(m,m')}$.

The calculation of the small-signal admittance matrix (in the limit that $\hbar\omega_{\text{IF}} \ll$ than the energy scale of the SIS DC I-V nonlinearity near the gap energy) is derived in [22]; here is an alternative version more suitable for computation:

$$Y_{(m,m')} = \frac{-j}{2(m'V_{ph} + V_{if})} \sum_{k=-\infty}^{\infty} C_{(k)} C_{(k+m-m')}^* \begin{bmatrix} \mathbf{I}(V_0 + kV_{ph}) - \\ \mathbf{I}(V_0 + kV_{ph} - m'V_{ph} - V_{if}) + \\ \mathbf{I}^*(V_0 + (k+m-m')V_{ph}) - \\ \mathbf{I}^*(V_0 + (k+m-m')V_{ph} + m'V_{ph} + V_{if}) \end{bmatrix}. \quad (\text{E.13})$$

Again, $\mathbf{I}(V) \equiv \mathbf{I}_{\text{kk}}(V) + j\mathbf{I}_{\text{dc}}(V)$ is the complex-valued SIS DC I-V curve, $V_{ph} = (\hbar/e)\omega_{\text{LO}}$, and $V_{if} = (\hbar/e)\omega_{\text{IF}}$. The coefficients $C_{(k)}$ were defined in (E.5) to (E.7). These equations are coded into SuperMix to perform the small-signal SIS heterodyne mixer analysis. To perform the large-signal harmonic balance, we need the limit of (E.13) as $\omega_{\text{IF}} \rightarrow 0$. Substituting $V_{if} = 0$ works for $m' \neq 0$. When $m' = 0$, the summation terms in (E.13) become derivatives of $\mathbf{I}(V)$:

$$Y_{(m,m' \neq 0)} = \frac{-j}{2m'V_{ph}} \sum_{k=-\infty}^{\infty} C_{(k)} C_{(k+m-m')}^* \begin{bmatrix} \mathbf{I}(V_0 + kV_{ph}) - \mathbf{I}(V_0 + (k-m')V_{ph}) \\ + \mathbf{I}^*(V_0 + (k+m-m')V_{ph}) \\ - \mathbf{I}^*(V_0 + (k+m)V_{ph}) \end{bmatrix}, \quad (\text{E.14})$$

$$Y_{(m,0)} = -\frac{j}{2} \sum_{k=-\infty}^{\infty} C_{(k)} C_{(k+m)}^* \left[\mathbf{I}'(V_0 + kV_{ph}) - \mathbf{I}'^*(V_0 + (k+m)V_{ph}) \right].$$

Equations (E.11), (E.12), and (E.14) provide the calculations necessary to determine the elements of the Jacobian matrix $\|d\mathbf{i}/d\mathbf{v}\|$ used in the harmonic balance Newton-Raphson iterations (E.10). That expression includes another potentially time-consuming matrix calculation: $(\mathbb{I} + \mathbf{S}) \cdot \|d\mathbf{i}/d\mathbf{v}\|$. Because of the block-diagonal natures of the two matrices, however, the product elements are

$$\left[(\mathbb{I} + \mathbf{S}) \cdot \left\| \frac{d\mathbf{i}}{d\mathbf{v}} \right\| \right]_{m,m'}^{n,n'} = \left[\delta_{n,n'} + S^{n,n'}(\omega_m) \right] \left[\frac{\partial i_m}{\partial v_{m'}} \right]^{n'}$$

and no summing over an index is required to calculate an element of the product.

Appendix F : PROPAGATION IN PARTIALLY-FILLED WAVEGUIDE

The first few sections of this appendix present mathematical derivations of general equations for the propagation of waves in a waveguide filled with various dielectrics. In particular, the cutoff frequency expressions needed for the mixer chip mounting channel design are derived.³³

Wave propagation in a lossless, cylindrical waveguide

Consider a long waveguide of uniform cross-section as in Fig. F-1. The waveguide is filled with an arbitrary number of various homogeneous, isotropic dielectrics, each of which has some permeability μ and permittivity ε which are uniform, constant, real-valued scalars. The waveguide has ideal, lossless conducting walls and has no geometry or material property variations along the direction of propagation in the guide, which is assumed to be aligned with the z -axis, as shown in the figure.

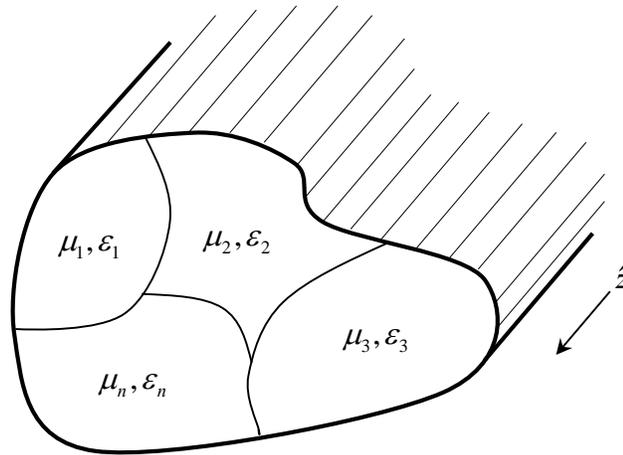


Fig. F-1: A cylindrical waveguide of arbitrary, but uniform, cross-sectional shape. The interior of the waveguide is filled with some number of different dielectric materials.

³³ This study of the cutoff frequencies of partially-loaded rectangular waveguide was substantially completed in August 2000. A portion of the author's handwritten notes from that effort have been made available online by Jacob Kooi [56].

Within each individual dielectric region of the waveguide there are no free charges or currents. Since μ and ε are uniform, constant scalars within each region, the wave equations for the electric and magnetic fields within the i^{th} region may be written as

$$\left(\nabla^2 - \frac{\mu_i \varepsilon_i}{c^2} \frac{\partial^2}{\partial t^2} \right) \begin{Bmatrix} \mathbf{E}_i \\ \mathbf{H}_i \end{Bmatrix} = 0 \quad (\text{F.1})$$

(using Gaussian units). Assuming harmonic solutions for waves propagating in the $+\hat{z}$ direction, we may assume solutions of the form

$$\mathbf{V}(x, y, z, t) = \mathbf{V}(x, y) \exp[i(k_z z - \omega t)]. \quad (\text{F.2})$$

Wave solutions of equation (F.2) propagating in the \hat{z} direction will, of course, have real wave numbers $k_z > 0$; evanescent waves have imaginary k_z with $ik_z < 0$. Using the form (F.2) for the solutions to (F.1), the result is the familiar Helmholtz equations:

$$\left(\nabla_{\text{T}}^2 + \frac{\mu_i \varepsilon_i}{c^2} \omega^2 - k_z^2 \right) \begin{Bmatrix} \mathbf{E}_i(x, y) \\ \mathbf{H}_i(x, y) \end{Bmatrix} = 0, \quad (\text{F.3})$$

$$\nabla_{\text{T}}^2 \equiv \hat{x} \frac{\partial}{\partial x} + \hat{y} \frac{\partial}{\partial y}.$$

Note that the vector amplitudes $\mathbf{E}_i(x, y)$ and $\mathbf{H}_i(x, y)$, although functions of only the transverse coordinates (x, y) , are 3-dimensional vectors which include components in the \hat{z} direction. Since the remainder of this section will consider solutions to (F.3), functions of (x, y) only, these function arguments will be dropped in the equations to follow.

The boundary conditions on the fields \mathbf{E}_i and \mathbf{H}_i must be considered next. At a conducting wall of the waveguide (\hat{n} is the unit normal to the wall surface):

$$\begin{aligned} \hat{n} \times \mathbf{E}_i &= 0, & \hat{n} \cdot \mathbf{H}_i &= 0, \\ (\hat{n} \cdot \nabla)(\hat{n} \times \mathbf{H}_i) &= 0. \end{aligned} \quad (\text{F.4})$$

The final condition in (F.4) may not be obvious, but can be derived by considering Maxwell's equation for $\nabla \times \mathbf{H}_i$ together with the other two boundary conditions. At an interface joining dielectrics numbered j and k with surface normal \hat{n} ,

$$\begin{aligned} \hat{n} \times (\mathbf{E}_j - \mathbf{E}_k) &= 0, & \hat{n} \cdot (\varepsilon_j \mathbf{E}_j - \varepsilon_k \mathbf{E}_k) &= 0, \\ \hat{n} \times (\mathbf{H}_j - \mathbf{H}_k) &= 0, & \hat{n} \cdot (\mu_j \mathbf{H}_j - \mu_k \mathbf{H}_k) &= 0. \end{aligned} \quad (\text{F.5})$$

The differential equations (F.3) with boundary conditions (F.4) and (F.5) form a typical eigenvalue problem for which solutions may be found only if the

$$\gamma_i^2 \equiv \frac{\mu_i \varepsilon_i}{c^2} \omega^2 - k_z^2 \quad (\text{F.6})$$

take on certain discrete values. The eigenvalues γ_i^2 define the various propagating modes, giving the dispersion relation $k_z(\omega)$ for each mode. Note that the γ_i^2 are just the squared magnitudes of the transverse components of the wave vectors in their respective regions, and that $k_z^2 + \gamma_i^2$ is just the squared magnitude of the wave vector of an unguided plane wave of frequency ω propagating in a medium with material properties μ_i and ε_i . Note further that for a given mode, the longitudinal wave number k_z must be the same in all dielectric regions. This is required by phase continuity along the length of the guide, so that if the boundary conditions (F.5) are satisfied at one value of the coordinate z in the guide, then they are identically satisfied everywhere within the guide.

Transverse and longitudinal fields

Because of the uniformity of the waveguide along the \hat{z} direction, it is useful to decompose all field vectors and vector operators into transverse and longitudinal components as was done with the gradient operator in (F.3). With $\partial/\partial z = ik_z$ for the harmonic solutions (F.2), we get the following forms for a general vector \mathbf{V} and the vector differential operations on it:

$$\begin{aligned} \mathbf{V} &= \mathbf{V}_T + \hat{z}V_z ; \text{ where } \mathbf{V}_T \cdot \hat{z} \equiv 0 \\ \nabla &= \nabla_T + ik_z \hat{z}, \end{aligned} \quad (\text{F.7})$$

$$\nabla \cdot \mathbf{V} = \nabla_T \cdot \mathbf{V}_T + ik_z V_z, \quad (\text{F.8})$$

$$\begin{aligned} (\nabla \times \mathbf{V})_T &= \hat{z}(ik_z \mathbf{V}_T - \nabla_T V_z) \\ \hat{z} \cdot (\nabla \times \mathbf{V}) &= \nabla_T \times \mathbf{V}_T. \end{aligned} \quad (\text{F.9})$$

Maxwell's equations for the harmonic solutions within any particular dielectric region then become:

$$\nabla_{\mathbf{T}} \cdot \begin{Bmatrix} \mathbf{E}_{\mathbf{T}} \\ \mathbf{H}_{\mathbf{T}} \end{Bmatrix} = -ik_z \begin{Bmatrix} E_z \\ H_z \end{Bmatrix} \quad (\text{F.10})$$

$$\nabla_{\mathbf{T}} \times \mathbf{E}_{\mathbf{T}} = i\mu \frac{\omega}{c} H_z \hat{z} \quad ; \quad \nabla_{\mathbf{T}} \times \mathbf{H}_{\mathbf{T}} = -i\varepsilon \frac{\omega}{c} E_z \hat{z} \quad (\text{F.11})$$

and the transverse components of Maxwell's curl equations, solved for $\mathbf{E}_{\mathbf{T}}$ and $\mathbf{H}_{\mathbf{T}}$,

$$\begin{aligned} \gamma^2 \mathbf{E}_{\mathbf{T}} &= i \left[k_z \nabla_{\mathbf{T}} E_z - \mu \frac{\omega}{c} (\hat{z} \times \nabla_{\mathbf{T}} H_z) \right] \\ \gamma^2 \mathbf{H}_{\mathbf{T}} &= i \left[k_z \nabla_{\mathbf{T}} H_z + \varepsilon \frac{\omega}{c} (\hat{z} \times \nabla_{\mathbf{T}} E_z) \right] \end{aligned} \quad (\text{F.12})$$

with γ^2 given by (F.6).³⁴ These equations obviously depend on the material properties of the dielectrics, and so will differ from region to region within the waveguide.

If E_z and H_z are functions which solve the Helmholtz equations (F.3) in a particular region, then $\mathbf{E}_{\mathbf{T}}$ and $\mathbf{H}_{\mathbf{T}}$ defined by (F.12) identically satisfy the other Maxwell equations (F.10) and (F.11), as well as (F.3). Therefore, the boundary value problem for the fields only requires one to solve (F.3) for the scalar functions E_z and H_z in each region of the waveguide. Since E_z and H_z are parallel to all dielectric and conductor boundaries in the waveguide, the boundary conditions (F.4) and (F.5) for them become

at conductor surfaces:

$$E_z|_{\text{wall}} = 0 \quad \hat{n} \cdot \nabla_{\mathbf{T}} H_z|_{\text{wall}} = 0, \quad (\text{F.13})$$

at dielectric region boundaries:

$$E_z|_{\text{region } j} = E_z|_{\text{region } k}; \quad H_z|_{\text{region } j} = H_z|_{\text{region } k}. \quad (\text{F.14})$$

Note that if E_z and H_z satisfy these boundary conditions, then $\mathbf{E}_{\mathbf{T}}$ and $\mathbf{H}_{\mathbf{T}}$ consistent with (F.12) will satisfy the conductor boundary conditions (F.4). The matching conditions (F.5) at the dielectric boundaries, however, when applied to $\mathbf{E}_{\mathbf{T}}$ and $\mathbf{H}_{\mathbf{T}}$, will (in general)

³⁴ Equations (F.10) – (F.12) are equivalent to equations (8.23) – (8.26b) in Jackson [3].

provide additional conditions on E_z and H_z through (F.12), and thus determine the allowed sets of values for the various eigenvalues γ_i in the dielectric regions.

Propagation in partially-loaded, lossless, rectangular waveguide

The previous discussion will now be specialized to the case of a rectangular waveguide filled with two lossless dielectrics as shown in Fig. F-2 below. In terms of the original problem of investigating the wave propagation properties of the channel containing the silicon mixer chip, a and h become the width and thickness of the silicon chip. In the final analysis relevant to design of the mixer chip mounting channel, the region with properties μ_2, ϵ_2 becomes an added vacuum gap of height b . The goal then will be to determine the effect of this gap on the cutoff frequencies of the various propagating modes in the channel.

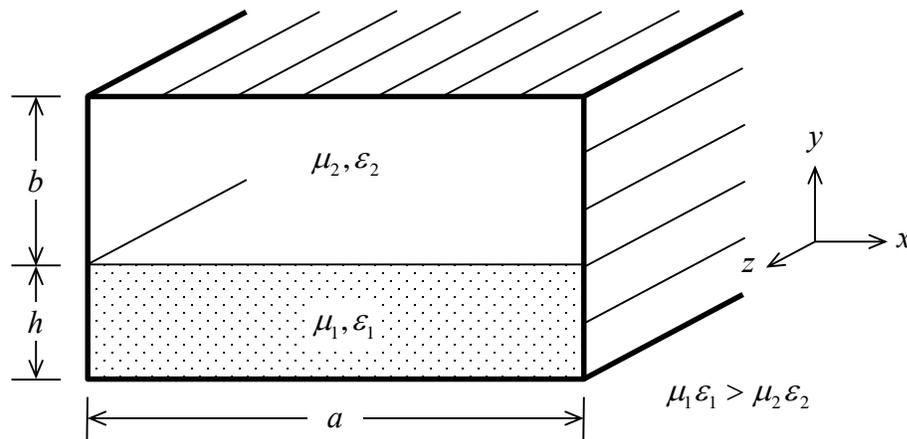


Fig. F-2: A rectangular waveguide filled (loaded) with two dielectrics.

Consider first the transverse variations of the longitudinal field components, E_z and H_z . The boundary value problem for each of these is given by (F.3), (F.13), and (F.14). Clearly, both E_z and the normal derivative of H_z must vanish at each conductor wall. Solutions to the transverse Helmholtz equations (F.3) in each region which identically satisfy these conditions are given in (F.15) and (F.16):

$$H_{z1} = A_1 \cos\left(\frac{m_H \pi x}{a}\right) \cos(k_{H1} y); \quad E_{z1} = B_1 \sin\left(\frac{m_E \pi x}{a}\right) \sin(k_{E1} y) \quad (\text{F.15})$$

$$\gamma_1^2 = \frac{\mu_1 \varepsilon_1}{c^2} \omega^2 - k_z^2 = \left(\frac{m_H \pi}{a}\right)^2 + k_{H1}^2 = \left(\frac{m_E \pi}{a}\right)^2 + k_{E1}^2$$

$$H_{z2} = A_2 \cos\left(\frac{m_H \pi x}{a}\right) \cos(k_{H2} (b + h - y))$$

$$E_{z2} = B_2 \sin\left(\frac{m_E \pi x}{a}\right) \sin(k_{E2} (b + h - y)) \quad (\text{F.16})$$

$$\gamma_2^2 = \frac{\mu_2 \varepsilon_2}{c^2} \omega^2 - k_z^2 = \left(\frac{m_H \pi}{a}\right)^2 + k_{H2}^2 = \left(\frac{m_E \pi}{a}\right)^2 + k_{E2}^2.$$

In the equations, the horizontal transverse mode indices m_H and m_E are nonnegative integers. If $m_E = 0$, then $E_{z1} = E_{z2} = 0$ and the mode is *TE* (*Transverse Electric*). Matching the fields at the dielectric boundary $y = h$ requires that

$$A_1 \cos(k_{H1} h) = A_2 \cos(k_{H2} b); \quad B_1 \sin(k_{E1} h) = B_2 \sin(k_{E2} b) \quad (\text{F.17})$$

where, in accordance with (F.15) and (F.16), A and B represent the H_z and E_z amplitudes, respectively, in each dielectric region.

The final requirement is to match the corresponding transverse fields \mathbf{E}_T and \mathbf{H}_T , given by (F.12), at the dielectric interface using (F.5) (recall that these fields will automatically satisfy the boundary conditions at the conductor walls, since E_z and H_z do). The matching conditions for the boundary normal (\hat{y}) components of the transverse fields demand that

$$\varepsilon_1 E_{y1}(y = h) = \varepsilon_2 E_{y2}(y = h); \quad \mu_1 H_{y1}(y = h) = \mu_2 H_{y2}(y = h). \quad (\text{F.18})$$

Substituting into (F.12), the condition on E_{y1} and E_{y2} requires that

$$\frac{\omega}{c} \frac{m_H \pi}{a} \left[A_2 \frac{\mu_2 \varepsilon_2}{\gamma_2^2} \cos(k_{H2} b) - A_1 \frac{\mu_1 \varepsilon_1}{\gamma_1^2} \cos(k_{H1} h) \right] \sin\left(\frac{m_H \pi x}{a}\right)$$

$$= k_z \left[B_2 \frac{\varepsilon_2 k_{E2}}{\gamma_2^2} \cos(k_{E2} b) + B_1 \frac{\varepsilon_1 k_{E1}}{\gamma_1^2} \cos(k_{E1} h) \right] \sin\left(\frac{m_E \pi x}{a}\right). \quad (\text{F.19})$$

Equation (F.19) and the corresponding equation resulting from the condition on H_{y1} and H_{y2} have some important implications:

- (1) If $k_z > 0$ (a propagating wave), then phase matching in the x direction along the dielectric boundary requires that the horizontal mode indices $m_H = m_E \equiv m$, where $m \in \{0, 1, 2, \dots\}$.
- (2) For $k_z > 0$, then if $m \neq 0$, the regions' H_z amplitudes B_1 and B_2 cannot both vanish unless both E_z amplitudes A_1 and A_2 do as well. Similarly, considering the equation for the H_y 's, then $A_1 = A_2 = 0 \rightarrow B_1 = B_2 = 0$. *There cannot be a pure TE mode solution unless $m = 0$, nor can there be a pure TM mode for any horizontal mode index value m .*
- (3) Since $\gamma_1^2 = (m\pi/a)^2 + k_{H1}^2 = (m\pi/a)^2 + k_{E1}^2$, and similarly for γ_2^2 , then we must have the vertical wave numbers $k_{H1} = k_{E1} \equiv k_1$ and $k_{H2} = k_{E2} \equiv k_2$. Therefore,

$$\boxed{\begin{aligned} \gamma_1^2 &= k_1^2 + \left(\frac{m\pi}{a}\right)^2; & \gamma_2^2 &= k_2^2 + \left(\frac{m\pi}{a}\right)^2; & m &\in \{0, 1, 2, \dots\} \\ k_z^2 &= \frac{\mu_1 \varepsilon_1}{c^2} \omega^2 - \gamma_1^2 = \frac{\mu_2 \varepsilon_2}{c^2} \omega^2 - \gamma_2^2. \end{aligned}} \quad (\text{F.20})$$

Equations (F.20) provide the eigenvalues and, thus, the dispersion relations $k_z(\omega)$ for the various waveguide modes specified by (m, k_1, k_2) . Note that (as one might expect), the first line of (F.20) just resolves the squared transverse wave vector magnitudes in the two dielectric regions of the waveguide into vertical and horizontal component magnitudes. The vertical components k_1 and k_2 must be related by:

$$\boxed{k_1^2 - k_2^2 = \gamma_1^2 - \gamma_2^2 = (\mu_1 \varepsilon_1 - \mu_2 \varepsilon_2) \omega^2 / c^2 > 0} \quad (\text{F.21})$$

$\therefore k_1^2 > k_2^2$, since $\mu_1 \varepsilon_1 > \mu_2 \varepsilon_2$ (by assumption).

Clearly, (F.21) demands that $k_1^2 > 0$. On the other hand, valid solutions exist for $k_2^2 < 0$, as will be shown below. The other boundary conditions (F.17) – (F.19) become:

$$\boxed{\begin{aligned} A_1 \cos(k_1 h) &= A_2 \cos(k_2 b) \\ B_1 \sin(k_1 h) &= B_2 \sin(k_2 b) \end{aligned}} \quad (\text{F.22})$$

$$\left[B_1 k_1 \varepsilon_1 \gamma_2^2 - A_1 k_z (\mu_1 \varepsilon_1 - \mu_2 \varepsilon_2) \frac{\omega m \pi}{c a} \right] \cos(k_1 h) + B_2 k_2 \varepsilon_2 \gamma_1^2 \cos(k_2 b) = 0 \quad (\text{F.23})$$

$$\left[A_1 k_1 \mu_1 \gamma_2^2 + B_1 k_z (\mu_1 \varepsilon_1 - \mu_2 \varepsilon_2) \frac{\omega m \pi}{c a} \right] \sin(k_1 h) + A_2 k_2 \mu_2 \gamma_1^2 \sin(k_2 b) = 0. \quad (\text{F.24})$$

Equations (F.20) – (F.24) provide 6 independent equations for the 7 unknowns ($k_z, k_1, k_2, A_1, A_2, B_1, B_2$) in terms of the dimensions and material properties of the waveguide, the frequency ω , and the horizontal transverse mode index m . The overall amplitude of the wave is arbitrary, so an under-determined system of equations is to be expected. In fact, it must be the case that the 4 equations involving the H_z and E_z wave amplitudes, (F.22) – (F.24), form a singular system so that the determinant of the amplitudes' coefficients must vanish. This condition along with (F.20) provides 3 independent equations for (k_z, k_1, k_2), so that the propagation modes and their dispersion relations may be found for any given frequency ω and horizontal mode index m . Because the vertical transverse wave numbers k_1 and k_2 appear as arguments of sine and cosine functions, there may be multiple solutions for a given ω and m which define the various vertical modes of the guide, analogous to the vertical mode index in the case of a rectangular waveguide filled with a single dielectric material.

In the case of a rectangular waveguide filled with a single dielectric material, the equations determining the amplitudes of the longitudinal fields H_z and E_z decouple, and the pure *TM* modes (wherein $H_z \equiv 0$) and pure *TE* modes ($E_z \equiv 0$) form a complete set of linearly independent mode solutions. In the case of partially-loaded waveguide, this is generally not the case. Equations (F.23) and (F.24) couple the H_z and E_z amplitudes through terms $\propto k_z (\mu_1 \varepsilon_1 - \mu_2 \varepsilon_2) (m\pi/a)$; this coupling coefficient will vanish only if:

- (1) Both dielectrics have the same refractive index.
- (2) $m=0$, giving the vertical-only *TE* modes.
- (3) $k_z=0$, the cutoff frequency for a mode.

To find the cutoff frequencies of the various modes, we must find the various possible combinations (m, k_1, k_2) which make the determinant of the linear system (F.22) – (F.24) vanish for the particular case of $k_z=0$. As pointed out above, each waveguide mode

becomes either pure *TE* or pure *TM* at cutoff because the terms coupling (F.23) and (F.24) vanish. The equations for the *TE* modes ($B_1 = B_2 = 0$) become

$$\begin{aligned} A_1 \cos(k_1 h) - A_2 \cos(k_2 b) &= 0 \\ A_1 \frac{k_1}{\varepsilon_1} \sin(k_1 h) + A_2 \frac{k_2}{\varepsilon_2} \sin(k_2 b) &= 0. \end{aligned} \quad (\text{F.25})$$

Setting the determinant of the coefficients of A_1 and A_2 in (F.25) to zero gives

<i>TE</i> modes at cutoff:	
$\frac{k_2}{\varepsilon_2} \cos(k_1 h) \sin(k_2 b) + \frac{k_1}{\varepsilon_1} \sin(k_1 h) \cos(k_2 b) = 0.$	(F.26)

Similar considerations lead to an equation for the *TM* modes ($A_1 = A_2 = 0$):

<i>TM</i> modes at cutoff:	
$\frac{k_2}{\mu_2} \sin(k_1 h) \cos(k_2 b) + \frac{k_1}{\mu_1} \cos(k_1 h) \sin(k_2 b) = 0.$	(F.27)

Equations (F.20) with $k_z = 0$ provide a second expression relating k_1 and k_2 :

at cutoff:	
$\frac{1}{\mu_1 \varepsilon_1} \left[k_1^2 + \left(\frac{m\pi}{a} \right)^2 \right] = \frac{1}{\mu_2 \varepsilon_2} \left[k_2^2 + \left(\frac{m\pi}{a} \right)^2 \right].$	(F.28)

For each integer value of the horizontal mode index m , the various (k_1, k_2) pairs which jointly solve (F.26) and (F.28) or (F.27) and (F.28) fully specify the various *TE* and *TM* modes at cutoff. The cutoff frequencies for each mode are then given by k_1 and k_2 using (F.21) or (F.20). In the literature equations corresponding to (F.26) and (F.27) are often derived in a quite different fashion and are expressed in terms of tangents, but those functions may have undefined values at perfectly acceptable combinations of the vertical wave numbers.³⁵

³⁵ See, for example, Chapter 6 of Collin [27], “Inhomogeneously filled waveguides.”

Using a vacuum gap to increase the lowest cutoff frequency

The problem at hand is to maximize the lowest cutoff frequency of the channel holding the mixer chip by adding a vacuum gap. The mixer chip substrate is a nonmagnetic slab with width a and height (thickness) h , and with $h < a$. It has the material properties $\mu = 1$, $\varepsilon > 1$. The vacuum gap has height b (with $\mu = \varepsilon = 1$, cf. Fig. F-2 on page 169). Without a gap, the lowest cutoff wave number and frequency will be the substrate's TE_{10} mode with $k_0 = \pi/a$ and $f_0 = \omega_0/2\pi = c/(2a\sqrt{\varepsilon})$, and for this case $k_1 = 0$. As will be shown, the TE_{10} mode's cutoff frequency will increase as a gap is added. First consider the effect on the TE_{10} mode cutoff frequency as the gap height b is increased from 0. Using (F.20) with, of course, the longitudinal wave number $k_z = 0$, and the definition of the no-air-gap TE_{10} mode cutoff frequency f_0 , the relation between the cutoff frequency f_{10} and k_1 is

$$\left(\frac{f_{10}}{f_0}\right)^2 = 1 + \left(\frac{k_1 a}{\pi}\right)^2 > 1 \quad \text{if } b > 0. \quad (\text{F.29})$$

Thus this cutoff frequency rises as an air gap is added. The relation between k_2 and f_{10} is

$$\left(\frac{f_{10}}{f_0}\right)^2 = \varepsilon \left[1 + \left(\frac{k_2 a}{\pi}\right)^2 \right] < \varepsilon \quad \text{if } h > 0. \quad (\text{F.30})$$

The inequality obtains because only if the waveguide were completely vacuum-filled could the f_{10} cutoff frequency increase by a factor of $\sqrt{\varepsilon}$ above f_0 . As shall be shown, even in the case of an infinite gap a nonzero value for the substrate thickness h will keep the TE_{10} mode cutoff frequency below that of a purely vacuum-filled waveguide. Thus it must be the case that the gap vertical wave number $k_2^2 < 0$ for this mode. Defining $\kappa_2^2 \equiv -k_2^2$, equation (F.26) becomes

$$k_1 \sin(k_1 h) \cosh(\kappa_2 b) = \varepsilon \kappa_2 \cos(k_1 h) \sinh(\kappa_2 b). \quad (\text{F.31})$$

Finding numerical solutions of the transcendental system of equations (F.29) – (F.31) is straightforward, and the solution with the smallest $k_1 > 0$ gives f_{10} . Typical results for silicon ($\varepsilon = 11.9$) are shown in Fig. F-3 on page 175, which shows the increase in the TE_{10} mode cutoff frequency as a vacuum gap is added. Note from the left-hand graph that adding even a small gap to a thin chip results in a dramatic increase in the cutoff

frequency. The achievable frequency increase (f_{10}/f_0) for any particular chip thickness/width ratio approaches a limit which is less than that of an empty waveguide of the same dimensions ($f_{\text{vacuum}}/f_{\text{silicon}} = \sqrt{\epsilon} = 3.45$). This limiting frequency is very nearly reached if the waveguide is square: its total height equals its width, as is clearly shown in the right hand plot of Fig. F-3.

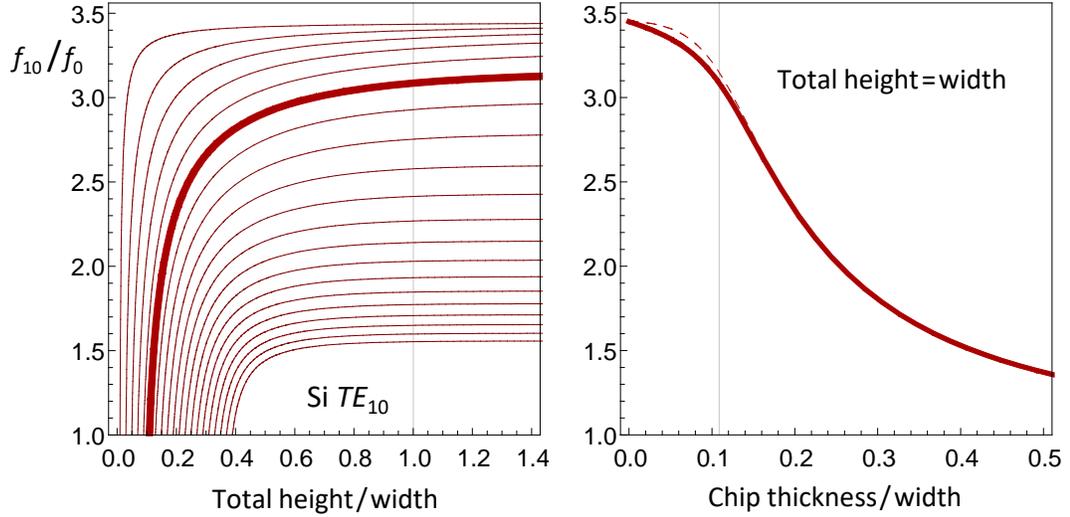


Fig. F-3: Increasing silicon chip TE_{10} cutoff frequency by adding a vacuum gap.

Left graph: family of curves showing TE_{10} cutoff frequency as a function of total waveguide height for chip thickness/width ratios ranging from 0.01 to 0.39. The curve for the actual mixer chip ratio (25/230) is highlighted.

Right graph: using a fixed waveguide height equal to the chip width, it shows the variation in cutoff frequency with chip thickness/width ratio. The vertical line is at the actual mixer chip thickness ratio. The maximum achievable cutoff frequency increase is shown as the thin dashed line (for an infinite vacuum gap).

To calculate the maximum achievable frequency increase, note that as the gap size $b \rightarrow \infty$, $\cosh(\kappa_2 b)/\sinh(\kappa_2 b) \rightarrow 1$, and equation (F.31) becomes

$$k_1 \tan(k_1 h) = \epsilon \kappa_2 . \quad (\text{F.32})$$

Solving this equation with (F.29) and (F.30) provides the maximum achievable cutoff frequency increase for a given chip thickness ratio. As mentioned previously, this limiting frequency is very nearly reached if the vacuum gap is sufficient to make the height of the waveguide supporting the dielectric substrate equal its width. This

observation turns out to be a good rule of thumb for common substrate materials regardless of dielectric constant, although it is probably overkill for thin chips, for which smaller vacuum gap heights will usually suffice.

Effect of the vacuum gap on the TE_{01} cutoff frequency

Adding a vacuum gap increases the total height of the waveguide. As the height b of the gap becomes significant, the waveguide mode with the next lowest cutoff frequency will become TE_{01} .³⁶ Its cutoff frequency will depend on the total height of the guide, $h + b$, and will decrease with increasing b . For a mixer chip substrate with given width a , thickness h , and dielectric constant ε , the goal is to choose the gap height b which maximizes the lowest cutoff frequency, which will occur when the two cutoff frequencies are equal: $f_{10} = f_{01}$. Increasing b further will (undesirably) lower f_{01} below f_{10} .

With $m = 0$, equations (F.26) and (F.28) simplify considerably: (F.28) states that $k_1 = \sqrt{\varepsilon} k_2$, and thus, from (F.26),

$$\sqrt{\varepsilon} \cos(k_1 h) \sin\left(\frac{k_1}{\sqrt{\varepsilon}} b\right) + \sin(k_1 h) \cos\left(\frac{k_1}{\sqrt{\varepsilon}} b\right) = 0. \quad (\text{F.33})$$

For TE_{01} , the smallest $k_1 > 0$ satisfying (F.33) should be chosen. The cutoff frequency is then (with k_0 and f_0 as previously defined):

$$\frac{f_{01}}{f_0} = \frac{k_1}{k_0} = \frac{k_1 a}{\pi}. \quad (\text{F.34})$$

For example, assume that the waveguide is square ($a = h + b$). If completely filled with dielectric ($a = h$ and $b = 0$), then obviously the TE_{01} and TE_{10} cutoff frequencies both equal f_0 . If the guide is empty (vacuum-filled), then this cutoff frequency goes up by a factor of $\sqrt{\varepsilon}$. A useful function is the variation in f_{01}/f_0 as the dielectric filling fraction

³⁶ For thin substrates and very small b , such that the chip width $a \gg h + b$, other horizontal modes such as TE_{20} will have lower cutoff frequencies than TE_{01} . As b increases, however, these mode cutoff frequencies also increase along with f_{10} . With b sufficiently large, f_{01} will be lower than any of these other modes' cutoff frequencies.

$h/(h+b)$ varies from 0 to 1. This function for the case of a silicon substrate is plotted in Fig. F-4.

The usefulness of this result is that a frequency ratio f_{01}/f_0 for a given filling fraction found from Fig. F-4 may be multiplied by the actual waveguide aspect ratio $a/(h+b)$ to determine the TE_{01} cutoff frequency in terms of f_0 for any waveguide partially loaded with dielectric by that same filling fraction. For example: a rectangular waveguide with a width equal to twice its total height and partially loaded with 20% silicon would have $f_{01}/f_0 = 2$ (actual aspect ratio) \times 2.4 (from the plot) = 4.8.

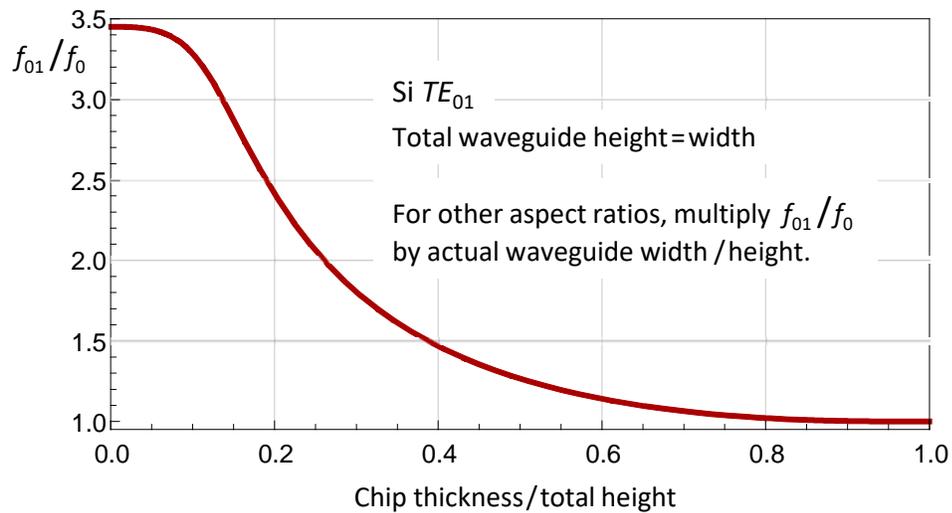


Fig. F-4: Silicon dielectric TE_{01} cutoff frequency vs. dielectric filling fraction.

Appendix G: GAUSSIAN OPTICS REVIEW

This appendix reviews some of the physics and mathematics of the quasioptical, fundamental Gaussian beam mode propagation model used to design the receiver RF optical system described in Chapter 3. This appendix provides the barest overview of the subject; for further details the interested reader should consult P. F. Goldsmith's now standard, 1998 text on *quasioptical systems* [4]. In it he defines a quasioptical system as one wherein an electromagnetic signal propagates in a fairly well-defined direction through an optical structure consisting of components whose transverse dimensions are only a few wavelengths long. In the case of the author's receiver, RF wavelengths are on the order of a millimeter and the RF structure has dimensions of no more than an inch or so (~30 mm), so Goldsmith's approach is appropriate.

The paraxial, fundamental Gaussian beam mode approximation

In many cases of microwave to submillimeter wave system design, a simple approximation to a full description of the propagating signal fields proves to be very useful: *Gaussian beam mode* solutions of the *paraxial wave equation*. This wave equation provides an approximation to the full, time-independent Helmholtz harmonic wave equation, $(\nabla^2 + k^2)\psi = 0$, in which $\psi(x, y, z)$ is any electromagnetic field component's complex-valued phasor and $k(x, y, z)$ is the harmonic signal's wave number.³⁷ Its derivation and solutions are detailed in Goldsmith [4].

Assume that the wave's direction of propagation is along \hat{z} . To arrive at the paraxial approximation to the Helmholtz equation, define the complex-valued function $u(x, y, z)$ so that $\psi(x, y, z) = u(x, y, z) e^{-jkz}$. Next assume that the axial (along \hat{z}) variation of u is small over distances of order a wavelength and also small compared to its transverse

³⁷ It is assumed here that the full, time dependent, harmonic solution for each field component would then be given by $\psi(x, y, z) e^{j\omega t}$, so that in this section we use the *electrical engineering phase convention* for the harmonic wave solutions, as does Goldsmith [4]. Of course, by using the Helmholtz equation, we have also assumed that the wave medium is linear.

variations. With these assumptions, the Helmholtz equation reduces to the paraxial wave equation for $u(x, y, z)$, Goldsmith's equation (2.5) and our equation (G.1) below:

$$\nabla_{\text{T}}^2 u - 2jk \frac{\partial u}{\partial z} = 0. \quad (\text{G.1})$$

The symbol ∇_{T}^2 denotes the *transverse* Laplacian operator. Because of the assumed cylindrical (azimuthal) symmetry of the receiver's RF optical system, cylindrical coordinates are appropriate, and $u(x, y, z) = u(r, z)$, so that $\nabla_{\text{T}}^2 = \partial^2/\partial r^2 + r^{-1} \partial/\partial r$.

With this assumed azimuthal symmetry, one solution to (G.1) is called the *fundamental Gaussian beam mode*, wherein the magnitude of $u(r, z)$ varies as

$$|u(r, z)| = \frac{\sqrt{2/\pi}}{w(z)} \exp\left[-\frac{r^2}{w(z)^2}\right] \quad (\text{G.2})$$

$$\text{with } w(z) = w_0 \sqrt{1 + (z/z_c)^2} \quad \text{and} \quad z_c = kw_0^2/2 = \pi w_0^2/\lambda. \quad (\text{G.3})$$

Clearly, for any axial position z the magnitude of u varies with radial distance r as a Gaussian, e -folding at $r = w(z)$, called the *beam radius*. The real, positive constant w_0 denotes the Gaussian beam's *waist radius*, and the real, positive constant z_c is called the beam's *confocal distance*. We have arbitrarily chosen the z -coordinate origin in (G.2) and (G.3) so that $w(0) = w_0$, and we have normalized u in (G.2) so that its squared magnitude integrated over any plane of constant z is unity. The waist radius w_0 completely determines the beam characteristics shown in (G.2) and (G.3) and is a free parameter of the solution to the second-order partial differential equation (G.1); the solution's other arbitrary constant was set by u 's normalization and by choosing $u(0, 0)$ to be positive real, determining both the solution's magnitude and phase at the origin. In particular, this choice will make $u(r, 0)$ positive real for all r (at the position of the beam waist).

The Gaussian beam's radius $w(z)$ varies hyperbolically with a minimum value of w_0 . For $|z| \gg z_c$, the radius asymptotically approaches $w(z) = (z/z_c)w_0$, corresponding to an f /ratio of $z_c/2w_0$. The beam's surfaces of constant phase in this region are very nearly spherical, and the beam's behavior becomes well described using geometric (ray) optics. On the other hand, near the beam waist ($|z| \lesssim z_c$) the beam radius is fairly constant, and

the Gaussian wave in this region is said to be “collimated,” with approximately planar surfaces of constant phase (the beam is perfectly planar at the beam waist) [4]. Interestingly, starting from a position $|z| \gg z_c$ and passing through the beam waist to another distant position z' , the Gaussian beam signal phase shifts by an extra 180° relative to the expected $k(z-z')$, a result consistent with the signal phase reversal expected upon passing through a geometric optics focal point. Of this additional phase shift, 90° occurs between the two confocal distance points $\pm z_c$ to either side of the beam waist.

Edge taper, beam coupling, and aperture efficiency

The fundamental Gaussian beam expression (G.2) for $u(r, z)$ gives the beam amplitude as a function of position, and therefore its squared magnitude will be proportional to the beam *flux density* (power/area) at position (r, z) , independent of azimuth angle ϕ .³⁸ For a given axial position z , the exponential $\exp(-r^2/w^2)$ gives the beam’s amplitude relative to that on the beam axis ($r=0$), and its square, $\exp(-2r^2/w^2)$, gives the relative beam flux density. Expressed in decibels, this relative reduction in flux density at r is called the beam’s *edge taper*, T_e , a function of the relative distance (r/w) from the beam axis:

$$T_e(r) = -20 \log\left(e^{-r^2/w^2}\right) \text{ dB} \approx 8.686 r^2/w^2 \text{ (dB)}. \quad (\text{G.4})$$

Thus the *edge taper* (in dB) is conventionally given a nonnegative value, even though it represents a reduction in flux density relative to that on the beam axis.

There exist, of course, many additional solutions of the paraxial wave equation (G.1). A countably infinite set of higher-order, Gaussian-like, normalized modal solutions for $u(r, \phi, z)$, all mutually orthogonal as well as orthogonal to the lowest order, fundamental Gaussian solution specified by equations (G.2) and (G.3), allows for the expansion of an arbitrary wave solution into a linear combination of this complete, closed set of modes. For example, given an arbitrary, but normalized and axially-aligned (common z -axis),

³⁸ Actually, this would be the case only for a monochromatic signal. Otherwise we should refer to the signal’s *spectral flux density*, $\partial^2 \text{power} / (\partial \text{area} \partial \text{frequency})$, at radius r and frequency ω .

generally complex-valued, paraxial wave solution $a(r, \phi, z)$, the *field coupling coefficient* c_a of this beam to the fundamental Gaussian beam $u(r, z)$ in some chosen plane with constant z would be given by the integral

$$(c_a)_z \equiv \langle u|a \rangle_z = \iint_{r, \phi} u(r, z)^* a(r, \phi, z) r dr d\phi, \quad (\text{G.5})$$

with normalization $\langle a|a \rangle_z = 1 = \iint_{r, \phi} |a(r, \phi, z)|^2 r dr d\phi.$

The fraction of the total flux in a (in the paraxial limit) coupled to u would then be given by $|c_a|^2$.

Now consider this coupling expression in the case of a very simple model of the CSO telescope + receiver detecting a monochromatic signal from a distant point source. The flux density of the source is uniform across the main telescope dish with a total incident power equal to this flux density times the projected area of the dish, $A_a = \pi r_a^2$, where r_a is the circular dish's radius. The signal's normalized beam function as intercepted by the telescope dish would then be $a(r \leq r_a) = 1/\sqrt{A_a}$, and $a(r > r_a) = 0$.³⁹ If the fundamental Gaussian beam mode associated with the receiver's optical coupling to the telescope has a beam radius of w at the position z of the main telescope dish, then if properly focused its reflection from the dish will have a planar wavefront (phase independent of r) with beam radius w . Using (G.5), the magnitude of the coupling of the signal flux to the receiver would then be $\alpha \equiv (r_a/w)^2 = 0.115 T_e(r_a)/\text{dB}$

$$|c_a|^2 = \frac{1}{A_a} \left[2\pi \int_0^{r_a} u(r, z) r dr \right]^2 = \frac{2}{\alpha} \left(1 - e^{-\alpha} \right)^2 \equiv \eta_{a, \text{illum}} \quad (\text{G.6})$$

where: $\alpha \equiv (r_a/w)^2 = 0.115 T_e(r_a)/\text{dB}$ (as defined in [4]).

³⁹ Another assumption: besides being monochromatic, the RF signal is assumed to have a single, well-defined polarization which matches the receiver beam's polarization at the telescope dish. If the source radiation is unpolarized (a good assumption for most astronomical sources), then only half of its total flux will match the receiver's polarization, and the rest of the source's incident power will remain undetected.

The result (G.6) estimates the telescope's *illumination aperture efficiency*, $\eta_{a,\text{illum}}$, which must have a value in the range $0 < \eta_{a,\text{illum}} < 1$.⁴⁰ It provides an example of the estimation of a single factor contributing to the telescope's overall aperture efficiency, η_a : the fraction of the incident signal power on the telescope's collecting area (main dish, in the case of the CSO) emitted by a distant point source which will then be coupled into the instrument's detector (receiver). Fig. G-1 plots $\eta_{a,\text{illum}}$ as a function of the Gaussian beam edge taper $T_e(r_a)$ at the edge radius of the telescope main dish. It shows that an edge taper of just over 10.9 dB results in an optimum $\eta_{a,\text{illum}}$ of about 82% when an astronomical point source signal must be coupled into a receiver's fundamental Gaussian beam mode.

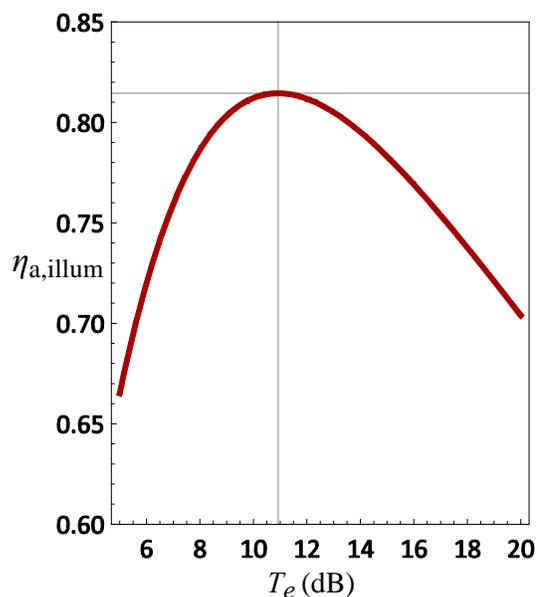


Fig. G-1: Aperture efficiency vs. edge taper. Maximum illumination efficiency is 82% for an edge taper of 10.9 dB. This result assumes ideal focusing and alignment as well as no secondary mirror obstruction.

This simple, nearly ideal example calculation, illustrative of the effect of coupling a source to a Gaussian beam, yields a result which can be quite far from the actually measured, final aperture efficiency η_a of a telescope such as the CSO. It ignores such factors as the surface accuracies, alignment, and reflectivities of the various mirrors in the telescope optical train (5 mirrors for the signal to reach the receiver mounted to the CSO relay optics), as well as the signal blockage and diffraction caused by the CSO telescope's Cassegrain secondary mirror and its mounting supports (0.48 m mirror diameter compared to the main mirror's 10.23 m). The effects of these additional considerations on the aperture efficiency are outlined by Goldsmith [4]. In the case of the CSO and its relay optics, the maximum aperture efficiency at 230 GHz is estimated to be ~55 – 65% for an edge taper of 10 – 11 dB [51] [52].

⁴⁰ An example of the *Schwarz inequality*.

The Gaussian beam parameter and beam transformations

Let us now return to the task of solving the paraxial wave equation (G.1). As shown in [4], a solution may be found by assuming that it has the following form:

$$u(r, z) = A(z) \exp\left[-\frac{jk r^2}{2q(z)}\right]. \quad (\text{G.7})$$

The unknown, complex-valued parameters in the proposed solution are $A(z)$ and $q(z)$, and we expect that they may be expressed in terms of the beam's waist radius w_0 and confocal distance z_c , as was done in (G.2). $A(z)$ will, of course, be chosen to achieve the desired normalization of u , whereas $q(z)$, called the solution's *Gaussian beam parameter*, will determine the variation of the beam's radius $w(z)$ and phase shift $\Phi(r, z)$ with position.

Again assume that the beam waist is located at $z=0$. The solution for $q(z)$ is then found in Goldsmith to be

$$q(z) = z + jz_c \quad \text{and} \quad \frac{1}{q} = \frac{1}{R(z)} - j \frac{w_0^2}{z_c w(z)^2} \quad (\text{G.8})$$

$$\text{where } R(z) = z \left[1 + (z_c/z)^2\right] \text{ with } R(z \rightarrow 0) = \infty.$$

$R(z)$ in (G.8) is the local radius of curvature of the beam's spherical surface of constant phase passing through $(0, z)$. The constant phase surface at the beam waist is a plane, as was mentioned earlier. As for $A(z)$, its determination is now straightforward. With our chosen normalization and taking $u(0, r)$ to be positive real, then from (G.2) we must have $A(0) = \sqrt{2/\pi} / w_0$. Completing the solution for all z :

$$A(z) = \frac{\sqrt{2/\pi}}{w(z)} \exp[j\Phi_0(z)], \quad \text{with } \Phi_0(z) = \arctan(z/z_c). \quad (\text{G.9})$$

As can be seen from this expression, $A(z)$ not only incorporates the beam's normalization factor, it also includes the phase reversal across the beam waist mentioned earlier.

The complex-valued beam parameter $q(z)$ provides a straightforward way to analyze the transformations that various optical components will have on the characteristics of a Gaussian beam. As explained in Goldsmith [4], the *ray matrix* or, alternatively, the “*ABCD*” *matrix* approach used to analyze systems in the geometric optics limit can be

equally effective in the Gaussian optics regime. In geometric optics, an individual ray enters an optical component located at z with some radial displacement r and with some slope dr/dz . The ray then follows some path through the component, exiting at z' with the different radial position r' and slope dr'/dz . The component's $ABCD$ matrix describes this ray transformation:

$$\begin{pmatrix} r' \\ dr'/dz \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r \\ dr/dz \end{pmatrix}. \quad (\text{G.10})$$

The effect of a succession of optical components may then be modeled as a series of matrix multiplications. For rays converging to a focus on the z -axis at $z=0$, their radial positions and slopes are related by the radius of curvature of the wavefront at z : in the paraxial limit, $r \ll R$, and $(dr/dz) = r/R = r/z$, where $z < 0$ implies that the focal point is downstream from the current location. Thus a slight modification of (G.10) provides an expression for the transformation of the rays' local radii of curvature $R \rightarrow R'$, which in this case is also the transformation of the distances from the element to the entering and exiting rays' respective focal points:

$$z' = \frac{Az + B}{Cz + D}.$$

The Gaussian optics counterpart of this geometric optics relationship between the entering and exiting wavefronts' radii of curvature (or distances from their respective foci) is the transformation by an optical element of the complex-valued Gaussian beam parameter $q(z)$. Replacing z in the above expression by $q(z)$ gives a complex-valued version which happens to correctly describe the effect of the optical component on a Gaussian beam:

$$q'(z') = \frac{Aq(z) + B}{Cq(z) + D}. \quad (\text{G.11})$$

For this reason, $q(z)$ is also often referred to as a Gaussian beam's *complex radius of curvature* (recall that $q = z + jz_c$, and that for large z , $R = z$). The real parts of the input q and output q' values give directed distances z and z' to the respective beam waists, where $z > 0$ implies that the waist is upstream (toward the beam source) from the current

position. The imaginary parts provide the respective confocal distances z_c and z_c' , which in turn determine the beam waist radii using (G.3) on page 179.

For the heterodyne receiver optical design, $ABCD$ matrices for the relevant optical “elements” are given in Table G-I:

Table G-I: $ABCD$ matrices

Propagation for distance ζ	$\begin{pmatrix} 1 & \zeta \\ 0 & 1 \end{pmatrix}$
Flat interface into relative index n	$\begin{pmatrix} 1 & 0 \\ 0 & 1/n \end{pmatrix}$
Curved interface into relative index n $R > 0$ if concave toward input beam	$\begin{pmatrix} 1 & 0 \\ \frac{n-1}{nR} & \frac{1}{n} \end{pmatrix}$

Appendix H: CRYOGENIC BIAS WIRING OPTIMIZATION⁴¹

Several receiver subsystems are mounted to the cryostat's 4 Kelvin cold-plate, and each of these subsystems requires electrical biasing and monitoring. DC wiring connects these devices to the 290 Kelvin bias electronics external to the cryostat. This wiring can conduct a significant amount of heat to the cold-plate if it is not carefully designed, unacceptably reducing receiver hold-time. If a wire carries an insignificant current (as would a wire to a high-impedance voltage monitor, for example), then ohmic heating may be safely ignored, and the wire should be thin, long, and made from a low thermal conductivity alloy. Other wires may need to carry several milliamps of current, however (magnet and LNA bias supplies, for example). In this case, it is important to optimize the wire dimensions to minimize the total heat load on the cold-plate, including that due to ohmic heating generated by the electrical current through the wire.

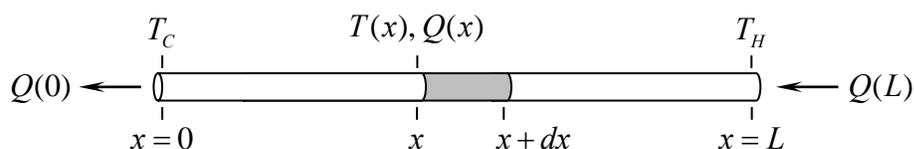


Fig. H-1: Diagram illustrating the parameters used for the heat flow analysis of a wire. A wire of length L and cross-sectional area A connects objects at temperatures T_C and T_H . Position x along the wire is defined as shown, along with temperature $T(x)$ and heat flow $Q(x)$. The wire's thermal conductivity and electrical resistivity are both functions of temperature ($\kappa(T)$ and $\rho(T)$, respectively), and will therefore vary with position along the wire.

Consider first the simple case wherein a conductor carries such a small current that electrical power dissipation is very much less than the heat conducted along it, so that ohmic heating may be ignored. The conductor connects the cold-plate (at temperature T_C) to a hotter thermal reservoir (temperature T_H), as diagrammed in Fig. H-1. When in

⁴¹ This study of cryogenic bias wiring optimization was substantially completed in August 2000. A portion of the author's handwritten notes from that effort have been made available online by Jacob Kooi [57]. The results herein duplicate those of a much earlier work by Mercouff [29].

steady-state, the temperature $T(x)$ at any point x along the wire is time-independent, so the heat flow everywhere along the wire must be the same: $Q(0) = Q(L) = Q(x)$ for this case. The calculation of the heat flow is straightforward, and the formula may be found in many standard texts. The derivation starts with the definition of the thermal conductivity, thusly:

$$\frac{Q}{A} = \kappa(T) \frac{dT}{dx} \quad (\text{H.1})$$

$$\therefore (Q/A) dx = \kappa(T) dT$$

$$\boxed{Q = \frac{A}{L} \int_{T_C}^{T_H} \kappa(T) dT.} \quad (\text{H.2})$$

The integral (H.2) is known as the *thermal conductivity integral* and is tabulated or graphed for many different materials used in cryogenic applications by various vendors such as, for example, Lake Shore Cryotronics, Inc. [53].

If the wire carries a nonnegligible current I , then ohmic heating of the wire must be included in the heat transfer calculation. Returning to Fig. H-1, the heat transfer $Q(x)$ is no longer constant along the wire, because the power dissipation in each segment dx adds a term to the heat transfer, as shown in (H.3).

$$\rho I^2 \frac{dx}{A} = Q(x) - Q(x+dx) = -dQ \quad (\text{H.3})$$

$$\therefore \frac{dQ}{dx} = -\frac{\rho I^2}{A}.$$

With (H.1),

$$\frac{dQ}{dT} = \frac{dQ/dx}{dT/dx} = -\frac{I^2}{Q} \rho \kappa$$

$$\therefore \boxed{Q^2(0) - Q^2(L) = 2I^2 \int_{T_C}^{T_H} \rho(T) \kappa(T) dT.} \quad (\text{H.4})$$

Evaluation of equation (H.4) requires another thermal integral, this time of the product of the electrical resistivity and the thermal conductivity as functions of the temperature, and it has units of (volts²). Unfortunately, this integral does not have a conventional name and is not tabulated by the standard references. It is apparent from (H.4) that to minimize heat transfer to the cold-plate, $Q(0)$, one should choose wire

dimensions such that $Q(L)=0$: no heat is transferred at the high-temperature end of the wire, and therefore all heat received at the cold-plate is due to ohmic heating of the wire. The minimum heat transfer to the cold-plate is then

$$\boxed{Q(0)_{\min} = I \sqrt{2 \int_{T_C}^{T_H} \rho(T) \kappa(T) dT}} . \quad (\text{H.5})$$

The square root term in this equation evaluates to an *optimal voltage drop* across the wire which will minimize the heat load on the cold plate when a current is flowing, and this voltage depends only on the material properties of the wire and its end temperatures and not on the current I . The designer's job is to choose the correct L/A for the wire so that this optimal voltage drop obtains when the design current I flows through it.

To determine how the wire geometry L/A affects the heat flow, start again with equations (H.1) and (H.3) and integrate twice to find the relationship between L/A and the heat transfer at the hotter end, $Q(L)$, resulting in (H.6).

$$\begin{aligned} Q(x) &= \sqrt{Q^2(L) + 2I^2 \int_T^{T_H} \rho(T') \kappa(T') dT'} = \kappa A \frac{dT}{dx} \\ \frac{dx}{A} &= \frac{\kappa(T) dT}{\sqrt{Q^2(L) + 2I^2 \int_T^{T_H} \rho(T') \kappa(T') dT'}} \\ \therefore \frac{L}{A} &= \int_{T_C}^{T_H} \frac{\kappa(T) dT}{\sqrt{Q^2(L) + 2I^2 \int_T^{T_H} \rho(T') \kappa(T') dT'}} . \end{aligned} \quad (\text{H.6})$$

Equation (H.6) implicitly defines the heat transfer $Q(L)$ at the hotter end of the wire in terms of the wire dimensions L/A . It is important to note that this derivation is only valid if $Q(x) \geq 0$ everywhere along the wire, since the positive square root was chosen in the first line above (H.6). This will not be correct unless the temperature monotonically increases along the wire from T_C to T_H . Nevertheless, for optimal wire sizing this condition is satisfied, so for $Q(L)=0$ the optimal wire dimensions to carry current I are given by equation (H.7) on page 189.

$$\boxed{\left[\frac{L}{A}\right]_{\text{optimum}} = \frac{1}{I} \int_{T_C}^{T_H} \frac{\kappa(T) dT}{\sqrt{2 \int_T^{T_H} \rho(T') \kappa(T') dT'}}} \quad (\text{H.7})$$

Equation (H.7) provides the optimal wire dimensions L/A for a wire carrying the steady current I , and (H.5) gives the resulting minimized heat transfer $Q(0)_{\min}$ at the cold-plate.

Note that the optimal wire design ensures that dT/dx vanishes just at the hot end of the wire. At all other points on the wire $dT/dx > 0$, and the expression just above (H.4) on page 187 is defined everywhere along the wire. If the wire is too thick or too short (L/A is smaller than the optimal value), then $Q(L) > 0$, and its value may be found by numerically solving the integral equation (H.6) for it in terms of the specified L/A . In this case, $dT/dx > 0$ everywhere along the wire, including the hotter end at L . Once $Q(L)$ is determined, (H.4) may be used to determine the heat load on the cold-plate, $Q(0)$.

Temperature and heat transfer of an undersized (too thin) wire

If L/A is larger than the optimal value given by (H.7), then the wire is too thin for the current it is required to carry, and ohmic heating will warm it to such an extent that along some portion of its length $T(x) > T_H$. Therefore $dT/dx < 0$ between the wire's maximum temperature point and the end at L . In this case, heat is also transferred from the wire to the "hot" reservoir at L . Clearly, dT/dx will vanish at the point of maximum wire temperature, and the expression above (H.4) is invalid there. Let the wire's maximum temperature be T_{Max} and let x_{Max} be its position on the wire. Because $T_{Max} > T_H$, heat transfer along the wire is everywhere away from x_{Max} . Since dT/dx vanishes at x_{Max} , each wire segment on either side of x_{Max} has optimal dimensions to carry current I along its temperature differential: T_{Max} to T_C and T_{Max} to T_H . Each wire segment then satisfies equations (H.7) and (H.5) for its respective length/area. Therefore, to determine T_{Max} , x_{Max} , and the resulting heat load on the cold-plate:

$$\frac{x_{Max}}{A} = \frac{1}{I} \int_{T_C}^{T_{Max}} \frac{\kappa(T) dT}{\sqrt{2 \int_T^{T_{Max}} \rho(T') \kappa(T') dT'}} ; \quad \frac{L - x_{Max}}{A} = \frac{1}{I} \int_{T_H}^{T_{Max}} \frac{\kappa(T) dT}{\sqrt{2 \int_T^{T_{Max}} \rho(T') \kappa(T') dT'}}$$

$$\begin{aligned} \therefore \frac{L}{A} - \left[\frac{L}{A} \right]_{\text{optimum}} &= \frac{2}{I} \int_{T_H}^{T_{Max}} \frac{\kappa(T) dT}{\sqrt{2 \int_T^{T_{Max}} \rho(T') \kappa(T') dT'}} \\ Q(0) &= I \sqrt{2 \int_{T_C}^{T_{Max}} \rho(T) \kappa(T) dT}. \end{aligned} \quad (\text{H.8})$$

The first of equations (H.8) defines an implicit relation which determines T_{Max} from I and L/A (along with the optimal L/A for current I calculated from (H.7)). Once T_{Max} has been calculated, the second equation of (H.8) determines the heat load on the cold-plate. It should also be noted that even though the wire's temperature variation is not monotonic, the integral expression (H.4) on page 187 remains correct.

Application to high-resistivity, cryogenic bias wiring materials

Approximate solutions to the heat transfer integral equations are straightforward for wires made from relatively high-resistivity materials commonly used for cryogenic applications. Common cryogenic wire materials such as Manganin® and phosphor bronze have nearly constant resistivity over the 4 K to 77 K temperature range: according to the graphs found on the Lake Shore Cryotronics website, Manganin's resistivity averages $45 \mu\Omega \text{ cm}$ and varies by only 6% over that range; phosphor bronze (with 5% tin) averages $11 \mu\Omega \text{ cm}$ and varies by just over 3% [54]. A resistivity ρ independent of temperature simplifies the integrals of (H.4) – (H.10) to standard thermal conductivity integrals: $\approx 520 \text{ mW/mm}$ for Manganin and $\approx 1200 \text{ mW/mm}$ for phosphor bronze between 4 K and 77 K, according to the same source [53].

Consider equations (H.5) and (H.7) for minimum heat transfer of a current-carrying wire. Equation (H.5) shows that there is an optimal voltage drop across the wire which determines the wire's optimal resistance when carrying a specified current I . This resistance in turn determines the wire's dimensions if ρ is assumed constant,

$$Q(0)_{\min} = I^2 R_{\text{opt}} = I^2 \left(\frac{\rho L}{A} \right) = I \sqrt{2 \rho \int_{T_C}^{T_H} \kappa(T) dT}$$

$$\boxed{\left[\frac{L}{A}\right]_{\text{optimum}} = \frac{1}{I} \sqrt{\frac{2}{\rho} \int_{T_c}^{T_H} \kappa(T) dT}} \quad (\text{H.9})$$

(constant I , and ρ independent of T).

As mentioned before, the minimum heat load on the cold-plate is equal to the entire ohmic heating of the wire, which, as is easily derived from (H.9) and (H.2), is equal to twice the heat transfer on the same wire when $I=0$.

For example, if a phosphor bronze wire connects a 4 K cold-plate to a 77 K stage, then by using its characteristics listed above, the optimal wire voltage drop is found to be 16 mV, and the optimal $L/A \approx 520 \text{ amp mm}^{-1}/I$; for Manganin the optimal voltage drop would be 23 mV, and the optimal $L/A \approx 48 \text{ amp mm}^{-1}/I$. Assume that a bundle of 32 AWG Manganin wires must carry a 20 mA current continuously to and from a superconducting magnet (40 mA total). If the length of the bundle must be approximately 40 cm, then how many wires should make up the bundle? The cross sectional area of a single 32 AWG wire is 0.0324 mm^2 . The optimal current for a single wire would then be $4 \text{ mA} \approx (48 \text{ amp mm}^{-1}) \times (0.0324 \text{ mm}^2) / (40 \text{ cm})$. Thus the bundle should contain 10 wires, 5 for each direction of the 20 mA current. The total heat load on the 4 K stage produced by this optimized, current carrying wire bundle would be its total ohmic heating: $(40 \text{ ma}) \times (23 \text{ mV}) \approx 0.9 \text{ mW}$.

Optimizing a wire carrying an intermittent current

In many cases, a wire to the cold-plate will not carry a large current I continuously, but only while the cooled instrument is active. When the instrument is idle, the current carried by the wire may be zero. Now the goal is to minimize the average heat transfer to the cold-plate over a complete cycle of activity and inactivity of the instrument. The wire dimensions must be a compromise: if the wire is sized to minimize heat transfer while the current I flows, then its dimensions will cause excessive heat transfer when the instrument is idle. Consequently, the wire should be made longer and thinner than a wire carrying a continuous current I . Let f be the fraction of time that the current I is applied; the rest of the time the current vanishes. Assume also that the time required for the

system to reach steady state as the current is switched on or off is a negligible fraction of an instrument cycle. The average heat load on the cold-plate would then be given by:

$$\begin{aligned} \overline{Q(0)} &= f[Q(0)]_{\text{on}} + (1-f)[Q(0)]_{\text{off}} \\ &= f I \sqrt{2 \int_{T_c}^{T_{Max}} \rho(T) \kappa(T) dT} + (1-f) \frac{A}{L} \int_{T_c}^{T_H} \kappa(T) dT. \end{aligned} \quad (\text{H.10})$$

This function must be minimized by choosing the correct wire dimensions, L/A , where, of course, T_{Max} is considered an implicit function of L/A by using (H.8) with (H.7) as described in an earlier section.

Appendix I : SUPERMIX PROGRAMS

The following pages provide a few simple program fragments which demonstrate some features of the SuperMix library as well as excerpts illustrating several key numerical routines developed by the author. As with the code fragments included in Chapter 5, comments are in a **colored font** to offset them from the code. The SuperMix distribution is available for download at [36] or [37]. It includes a suite of example programs illustrating many of the library's features. The library header files also contain extensive comments describing the library's features and syntax.

Building a circuit

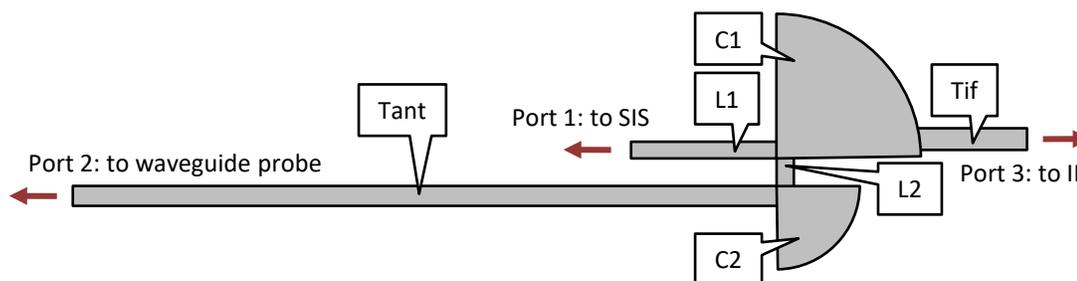


Fig. I-1: A drawing of the RF circuitry wiring layer, identifying its components.

The RF matching circuit of the receiver mixer chip, described in Chapter 2, was designed to properly match the waveguide probe impedance to the SIS junction's requirements. It also provides the IF output connection from the SIS to the IF matching network. The RF matching circuit is composed of several microstrip transmission lines and a pair of radial stubs, as shown in Fig. I-1 above. It has three ports: one each for the SIS, waveguide probe, and IF output. To model this circuit using SuperMix, one first defines circuit elements for each of the required components. To properly connect them and build the circuit model, a few more components are needed: two 3-port **branch** elements and a single **open_term**. The SuperMix model circuit is shown in Fig. I-2 on page 194. The code fragment required to build the circuit follows the figure (refer to the Fig. I-2 for the elements' names and their port numberings).

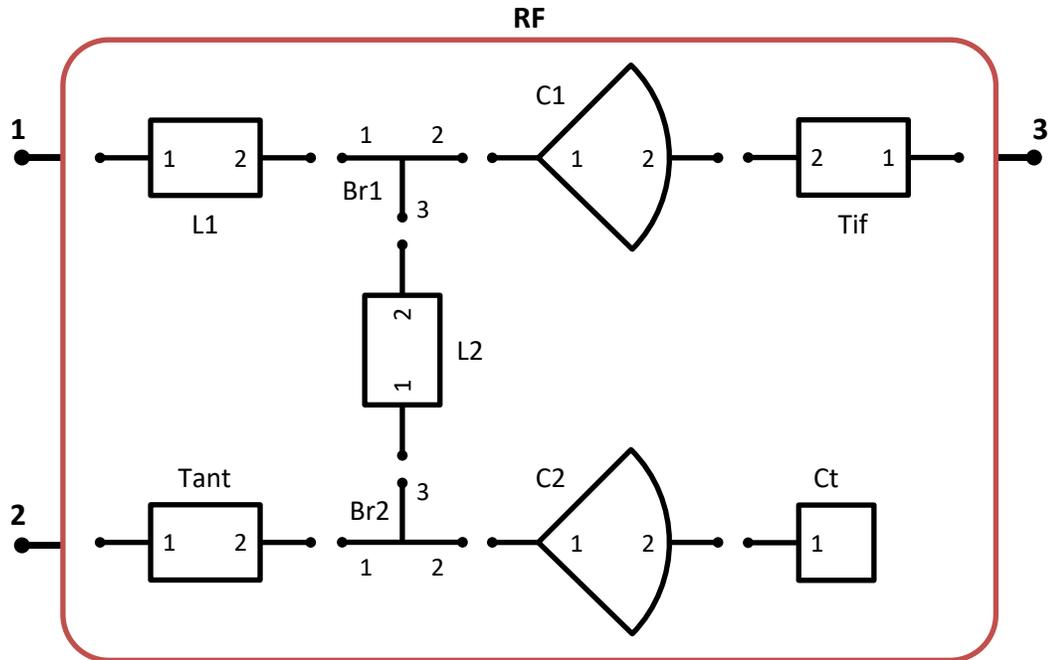


Fig. I-2: The SuperMix model of the SIS mixer chip RF circuitry wiring layer. The diagram shows the corresponding model elements with their port numberings. Two branch elements and an open-circuit termination must be added to complete the model.

Code example I-1: Using the circuit class to build the RF matching network model.

```

// declare the circuit and its elements
circuit RF;
microstrip Tant, L1, L2, Tif;
radial_stub C1, C2;
branch Br1(3), Br2(3); // declare 3-port branches
open_term Ct; // an open-circuit termination for C2

// build the circuit by defining element connections
RF.connect( Tant, 2, Br2, 1 );
RF.connect( Br2, 2, C2, 1 );
RF.connect( C2, 2, Ct, 1 );
RF.connect( Br2, 3, L2, 1 );
RF.connect( L2, 2, Br1, 3 );
RF.connect( Br1, 1, L1, 2 );
RF.connect( Br1, 2, C1, 1 );
RF.connect( C1, 2, Tif, 2 );

// Define the circuit's external port connections
RF.add_port( L1, 1 ); // ports must be added in order
RF.add_port( Tant, 1 ); // RF's port 2
RF.add_port( Tif, 1 ); // RF's port 3

```

As can be seen from the code, to add an element to a circuit model, simply include it as an argument in a `connect()` or `add_port()` call. Added elements can be sub-circuits also built using class `circuit`, so complicated structures can be built up in stages. An element or sub-circuit can also be used as an element by more than one independent circuit object, but the elements of an individual circuit must be separate objects. When building a circuit exactly one reference to each port of each element must be made using either `connect()` or `add_port()`.

Modeling superconducting microstrip lines

SuperMix includes predefined, accurate physical models for the superconducting microstrip and radial stub elements required by the SIS mixer model. In this section, we show how to use a selection of them. Much of the example code in this section is excerpted from the SuperMix distribution example program *microstrip.cc*. Modeling a physical transmission line such as superconducting microstrip involves three steps: (1) declare and assign physical properties to the materials which make up the structure; (2) declare the transmission line object and assign the materials to it; and (3) assign the object's remaining physical characteristics such as its length and width. First, the materials:

Code example I-2: the materials used to construct the microstrip transmission line

```
// superconducting niobium layer for both ground plane and top strip
super_film nb;
  nb.Vgap      = 2.9*Milli*Volt; // the 0 Kelvin gap voltage
  nb.Tc        = 9.2*Kelvin;
  nb.rho_normal = 5.0*Micro*Ohm*Centi*Meter;
  nb.Thick     = 3000*Angstrom;

// the dielectrics take epsilon and loss tangent at construction
const_dielectric SiO(5.6, .0001), vacuum(1.0, 0);
```

The conducting layers need a specified thickness so that their surface impedance may be calculated. The math is contained in the SuperMix library files `supcond.cc` and `surfaceZ.cc`. The superconducting film surface impedance is a function of frequency, and its calculation is nontrivial. Since optimizations usually involve sweeps over frequency, the `super_film` class by default automatically builds and maintains an interpolation of

the object's surface impedance as a function of frequency, greatly speeding up optimizations. For a calculation at a single frequency, however, or if the film temperature or thickness must be varied, then building the interpolation table wastes time, so it should be disabled using the object's member function `no_interpolate()`. Next, the microstrip:

Code example I-2 (continued): defining the microstrip properties

```
// superconducting niobium layer for both ground plane and top strip
microstrip line;
  line.top_strip(nb).ground_plane(nb);
  line.superstrate(vacuum).substrate(SiO);
  line.sub_thick = 2500*Angstrom; // dielectric substrate thickness

// the microstrip top strip width and line length often need to be optimized
// here are a pair of starting estimates, though
  line.width = 2.0*Mi cron;
  line.length = 100.0*Mi cron;
```

Now the length and width can be optimized along with other circuit parameters as shown in the next section. To determine good starting values for such an optimization, the simple procedure used in the library example program *microstrip.cc* could be used. The object of that program is to design a 1/4-wave microstrip transformer to match a 30 Ω source to a 10 Ω load at 230 GHz. The microstrip line should therefore have a characteristic impedance with real part equal to $(30 \times 10)^{1/2} = 17.3 \Omega$, and its length then determined by the line's propagation constant. Its member functions `Zchar()` and `Kprop()` provide the needed info.

Code example I-2 (continued): setting the desired microstrip characteristic impedance

```
// operating conditions for the calculation (usually set early in the code)
device: : T = 4.2*Kelvin; // operating temperature (superconductors!)
device: : f = 230*GHz; // this is the design frequency for the match
device: : Z0 = 30.0*0hm; // one of the impedances to match

// a simple iterative procedure to correct the line's width
double Ztarget = sqrt(30.0*10.0)*0hm; // the target line impedance
while( fabs(line.Zchar().real/Ztarget - 1.0) > 1.0e-6 )
  line.width *= line.Zchar().real/Ztarget; // increasing width lowers Zchar
```

Finally, use the propagation constant (imaginary part is the wave number) to adjust the length of the transmission line.

Code example I-2 (continued): using the line's calculated propagation constant

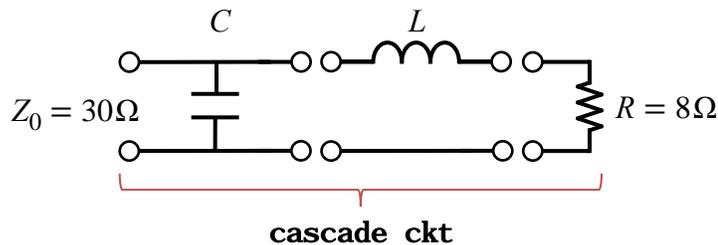
```
// set the line length to 1/4 wavelength (at device::f = 230*GHz)
double lambda = 2.0*Pi/line.Kprop().imaginary;
line.length = 0.25*lambda;

// print the line's width and length in microns
cout << line.width/Micron << " " << line.length/Micron << endl;
```

The calculated width and length (in microns) will be 2.25 and 122, respectively.

Using the optimizer

Here is a simple example showing how to optimize circuit parameters. The circuit to be optimized is the L-C matching network shown in Fig. C-4 on page 132. Its SuperMix model is built using class `cascade` to concatenate a sequence of 2-port elements, finally terminated by a 1-port (in this case), as shown below:



The code to build and find the L and C values which best optimize the power coupling between the 30 ohm source impedance (represented by the input transmission line with normalizing impedance $Z_0 = 30\Omega$) and the load (represented by a SuperMix `zterm` with $Z = R = 8\Omega$) over a frequency range of 200–300 MHz is shown in Code example I-3 on page 198.

Code example I-3: a simple application using the SuperMix optimizer.

```

#include "supermix.h"
int main()
{
    device::Z0 = 30.0*Ohm;    // the impedance of the line to match

    // the matching circuit:
    zterm    load(8.0*Ohm);    // the load to match
    capacitor C; C.parallel();
    inductor L; L.series();
    cascade ckt;                // cascade elements one-by-one
    ckt.add(C).add(L).add(load); // build the matching circuit

    // an error term to optimize the input match.
    // the target S11 is 0. The error is the squared difference.
    smag s11(ckt); // the magnitude of an S parameter
    s11.in(1).out(1).match(0.0);

    // frequency sweep over which to calculate the mean error
    // arguments to sweep(): minimum, maximum, increment, units
    sweeper band;
    band.sweep(device::f, 200.0, 300.0, 1.0, MHz);

    // the error function: give it control of L and C.
    // arguments to vary(): minimum, initial, maximum, units
    error_func ef;
    C.set(ef.vary(1.0, 10.0, 100.0, Pico*Farad));
    L.set(ef.vary(1.0, 10.0, 50.0, Nano*Henry));

    // add the single error term with its associated sweep
    ef.add_term(1.0, s11, band); // weight, error term, sweeper

    // the optimizer itself, using our error function
    powell opt(ef); // Powell's method local minimizer, using our ef
    opt.verbose(); // output the iterations as they occur
    // call the optimizer:
    double error_val = opt.minimize();

    // output the results
    cout << "Final error function value: " << error_val << endl;
    cout << "Final parameters are:" << endl;
    cout << "  C(pF): " << C.C/(Pico*Farad) << endl;
    cout << "  L(nH): " << L.L/(Nano*Henry) << endl;
}

```

The program's output is shown below. With the `verbose()` option chosen for the optimizer routine, it outputs the results from every intermediate iteration. The final error function value is the mean squared deviation of the magnitude of the circuit's reflection coefficient away from the target value of 0 (perfect match). This gives the mean fraction of the reflected power over the frequency sweep sample points.

Program output from Code example I-3.

```

Iteration 1
Parameters:
10 10
Function value: 0.271293

Iteration 2
Parameters:
33.4831 8.31831
Function value: 0.0244185

Iteration 3
Parameters:
34.2414 8.05832
Function value: 0.0243092

Final error function value: 0.0243088
Final parameters are:
C(pF): 34.0372
L(nH): 8.16922

```

Rounding these optimized values off to something reasonably available for purchase (33 pF and 8.2 nH), the resulting match is shown in the Smith chart, Fig. C-5 on page 133. The mean power reflected, $|\Gamma(\omega)|^2$, is less than 2.5% over the band. The worst case is 8% reflected at 200 GHz. Choosing 8.6 nH reduces the worst case reflected power to just over 7%, and the mean reflection is just over 2.5%, so that may be a better choice.

Calculating the SIS operating state C_k values

SuperMix uses Withington and Kollberg's extension of Tucker's quantum mixer theory to determine SIS operating states and their resulting small-signal admittance matrices using calculations purely in the frequency domain [22] [1]. The equations and operating state calculation algorithm are described in detail in Appendix E. In this section are selections of the SuperMix library code used to perform a key part of the numerical calculation: determination of the complex-valued vector of coefficients $C_{(k)}$ defined in equations (E.5) to (E.7) on page 159. This code is contained in the implementation of the `ckdata` class, found in the SuperMix library files *junction.h* and *ckdata.cc*, written by the author in 1998.

Arrays of Bessel function values

The $C_{(k)}$ calculation requires an array of Bessel function values $J_n(x)$ for a specified argument value in the general range $0 \leq x \leq 10$. The algorithm used is based on the traditional backward recursion [49]. The tricky part is twofold: (1) to determine how high an order n is needed to ensure that all Bessel functions whose magnitudes $|J_n(x)|$ might exceed ZEROTOL (10^{-6}) are included, and (2) to start the recursion at a large enough order m so that the calculated values of $J_0(x)$ to $J_n(x)$ have errors of less than ZEROTOL. The execution of the first trick is performed by `find_max_bessel_n`, which properly sets the size of the array `bessel_values`.

Code example I-4: calculating the maximum Bessel function order needed.

```

int ckdata::find_max_bessel_n(const double x)
{
    int nmax; // the maximum order of |Jn(x)| > ZEROTOL
    if ( x < ZEROTOL ) { // for x near zero, J1(x) ~ x/2.0
        nmax = 0;
    }
    else if ( x < 2*sqrt(ZEROTOL) ) { // in this range, J2(x) ~ x*x/8.0
        nmax = 1;
    }
    // the following formulas need modification if ZEROTOL != 1.0e-6
    else if ( x < 10.0 ) { // a quadratic fit in this range
        nmax = int( 2.165 + 4.678 * sqrt(x) + 0.5786 * x );
    }
    else { // for |x| > 10, a linear fit suffices
        nmax = int( 10.0 + 1.25 * x );
    }

    // now reallocate bessel_values if necessary, and return
    if ( nmax > bessel_values.Rmaxindex(nmax) )
        bessel_values.resize(1, nmax); // left index range should be {0,1}
    return bessel_values.Rmaxindex();
}

```

For an argument of 2, about the maximum for most practical applications of SIS mixers, `find_max_bessel_n(2.0)` returns 9 (`int(9.93789)`), so in most cases the number of Bessel function values to calculate will be quite reasonable.

The final trick is performed by `bessel()`, which then fills `bessel_values` with the Bessel functions $J_0(x)$ to $J_n(x)$ and their derivatives, calculated by backward recursion: $J_{i-1}(x) = (2i/x)J_i(x) - J_{i+1}(x)$ and $J_i'(x) = [J_{i-1}(x) - J_{i+1}(x)]/2$, along with the normalization condition: $1 = J_0(x) + 2 \sum J_{2j}(x)$.

Code example I-4 (continued): bessell function calculator.

```

int ckdata::bessel(const double x)
{
    int n,m;           // Jn is max returned order, Jm starts the recursion
    double *J, *Jprime; // will alias the rows in bessell_values

    n = find_max_bessel_n(x); // Jn, Jn' will be the highest in answers
    m = 2 * int(1 + (sqrt( double(n) ) + n)/2); // an even value for m

    J = bessell_values[0]; // could change with each call to
    Jprime = Bessell_values[1]; // find_max_bessel_n(), because of resize()

    if ( x == 0.0 ) {
        J[0] = 1.0; Jprime[0] = 0.0; // only Jo and Jo' will be returned
        return n;
    }

    double Jj, Jjm, Jjp; // terms in the recursion: Jj(x), Jj-1(x), Jj+1(x)
    int sum_flag = FALSE; // a boolean toggle for including even terms in sum
    const double Limit = 1.0e100; // don't let unnormalized results get
    const double Scale = 1.0e-100; // too big (unlikely). Scale = 1/Limit
    double mult = 2.0/x; // we know that x != 0.0

    // the backwards recursion loop
    int j;
    for ( j = m, Jj = 1.0, Jjp = 0.0; j > 0; --j ) { // remember, m is even
        Jjm = j*mult*Jj - Jjp; // the recursive calculation

        if ( j <= n ) { // Save results for Jj and Jj'
            J[j] = Jj; Jprime[j] = (Jjm - Jjp) / 2.0;
        }

        // Add the results for even j-1 into normalization sum:
        if ( sum_flag == TRUE )
            sum += Jjm;
        sum_flag = !sum_flag; // for next trip thru loop

        // Rescale values if results are getting too big:
        if ( fabs(Jjm) > Limit ) {
            Jjm *= Scale; Jj *= Scale; sum *= Scale;
            bessell_values *= Scale;
        }
        Jjp = Jj; Jj = Jjm; // step the terms down one
    } // end main recursion loop

    // now j = 0; and Jj = Jo(x), Jjp = J1(x) (unnormalized)
    J[0] = Jj;
    Jprime[0] = -Jjp; // Jo'(x) = - J1(x)

    sum = 2.0*sum - Jj; // now sum = normalization formula
    bessell_values /= sum; // normalize the answer
    return n;
}

```

In case you missed it, the trick is the calculation of integer **m** in the line following the call to `find_max_bessel_n()`. After `bessel(x)` executes, `bessel_values[0]` holds the vector of Bessel functions $J_0(x)$ to $J_n(x)$, and `bessel_values[1]` holds the values of their derivatives. `bessel(x)` returns the max order n in the arrays. Only the values for positive Bessel function indices are stored; the symmetries of the functions $J_{-m}(x) = (-1)^m J_m(x)$ will be applied when calculating the final $C_{(k)}$ values.

The $A_{m,n}$ and C_k calculations

The next step is to calculate the terms $A_{(m),n}$ using equations (E.6) and (E.7). The code for this is straightforward, calculating a complex-valued vector of results for nonnegative m at a specified n . The `ckdata` member function is `fillA()`, called with the complex-valued argument $\alpha_n e^{j\phi_n}$ defined in (E.7). The results are stored in `Amj_values` (using “ j ” for our “ n ”). `fillA()` returns the max m value calculated.

Code example I-5: calculating the $A_{m,n}$ terms.

```
// a is the complex version of alpha, ie: V/Vphoton
int ckdata::fillA(const Complex a)
{
    double alpha = abs(a);
    // unit = exp(-I*arg(a)), a unit vector
    complex unit = (alpha == 0.0) ? complex(1.0) : conj(a/alpha);

    int max = bessel(alpha); // calc the Bessel fcn array

    // now resize Amj_values if required
    if (max > Amj_values.maxindex(max)) // then Amj_values is too small
        Amj_values.resize(max);

    // lastly fill in the Amj_values
    max = Amj_values.maxindex(); // merely a safety precaution
    int m; Complex phase; // phase will be exp(-I*m*arg(a))
    double *J = bessel_values[0]; // the vector of Bessel functions

    for ( m = 0, phase = 1.0; m <= max; ++m, phase *= unit )
        Amj_values[m] = J[m] * phase;

    return max;
}
```

Only the $m \geq 0$ values are calculated by `fillA()`. The relation $A_{(-m),n} = (-1)^m A_{(m),n}^*$ is used to get the others.

Now we calculate the $C_{(k)}$ values using (E.5), which at first glance seems nightmarish. The gist of (E.5) is that the calculated $C_{(k)}$ values are iterated until the results do not change, with the n th iteration convolving in the calculated $A_{(m),n}$ values. The saving grace is that the user selects how many harmonics $h \geq 1$ to include in the calculation of the SIS operating state, so that all higher harmonic voltages $V_{n>h} \equiv 0$. Thus for $n > h$ only $A_{(0),n} = 1$ does not vanish, and therefore $C_{(k)}^{h+1} = C_{(k)}^h = C_{(k)}$, ending the sequence after h steps. The first step is easy, because from (E.5), $C_{(m)}^1 = A_{(m),1}$.

With these thoughts, we begin: first, the data structure used to hold the $C_{(k)}^n$ and the $C_{(k)}^{n-1}$ is a two-row, complex-valued, static SuperMix `Matrix` object inside the `ckdata` member function `calc()`. This function manages the calculation and storage of the sequence of convolution sums in (E.5) and lastly fills the `ckdata` member `Vector Ck` with the final result. Each step in the sequence of sums is actually calculated by the hidden, “helper” function `convC()`.

Code example I-5 (continued): function `convC()` for calculating the C_k .

```
static void
convC ( const int j,          // the harmonic number
        const int curr,     // the row of C which will hold the results
        Matrix & C,         // convC will write into the row C[curr]
        Vector & Amj        // the vector of Amj values for harmonic j
      )
{
  int k, kmax;              // loop index into C and its limit
  int m, amax;              // another loop index (into Amj) and limit
  int sign;                 // holds +/- 1 for the A(-m)j

  kmax = C.Rmaxindex();    // C[curr] row max index value (max k)
  amax = Amj.maxindex();
  for ( k = -kmax; k <= kmax; ++k ) {
    Complex *pCk = & C[curr][k]; // only do this indexing calculation once
    *pCk = C[!curr][k] * Amj[0]; // C[!curr] holds previous Ck iteration
    for ( m = 1, sign = -1; m <= amax; ++m, sign *= -1 )
      *pCk += C.read(!curr, k - j*m) * Amj[m] // term for +m
              + C.read(!curr, k + j*m) * sign * conj(Amj[m]); // term for -m
  }
}
```

The access of elements of the previous iteration using `C.read()` ensures that it returns 0 if beyond the valid range of `C`, which is $-kmax$ to $+kmax$. The code for `calc()` should now be clear.

Code example I-5 (continued): function calc() for calculating the C_k .

```

ckdata & ckdata::calc(
    const double fL0, // The Large-Signal (L0) frequency
    const Vector & V // The Large-Signal (L0) harmonic voltages
)
{
    const int INITSIZE = 20; // for temporary table alloc
    static Matrix C(2,INITSIZE, Index_C, Index_S); // hold intermediate results
    C.fillall(0.0);
    int curr = 0; // will be either 0 or 1: picks a row of C
    int cmax, pmax; // individual Rmaxindex() for current and previous C result
    int harms = V.maxindex(); // number of harmonics

    // initialize C[curr] for the first iteration
    C[curr][0] = 1.0;
    cmax = C.Rmaxindex(0); // elements of C beyond +/-cmax are all 0 here

    // loop through V values, convolving C with the previous result
    int amax;
    double scale = RmsToPeak*VoltToFreq / fL0; // convert Vj to alpha_j

    for ( int j = 1; j <= harms; ++j ) {
        // at this point, all C elements beyond +/-cmax == C.Rmaxindex() are 0
        // this will be a loop constant
        // if Vj is zero, the Ck don't change
        if (V[j] == 0.0) continue;

        // Vj is nonzero here
        curr = !curr; pmax = cmax; // step to other row of C; pmax is old cmax
        amax = fillA(scale * V[j] / j);
        // now adjust size of C
        cmax += (j * amax); // the new index range for nonzero Ck
        if(cmax > C.Rmaxindex(cmax)) // then the capacity of C is too small
            // grow capacity of C, plus some extra
            cmax = C.resize(1, cmax+INITSIZE).Rmaxindex(cmax); // max is still cmax

        // now the previous row, C[!curr], is zero for all elements beyond pmax
        // and the size of C is at least cmax, so we can use raw data access
        // and C.Rmaxindex() == cmax, so loop constant is still true.

        // do the convolution sum. This is written as a separate function to
        //support future expandability for the harmonic balance routine
        convC(j, curr, C, Amj_values);
    }

    // C[curr] has the results
    //this will reallocate Ck to be big enough
    Ck = row(curr, C).shrink(ZEROTOL);
    Tol = ZEROTOL;
    return *this;
}

```

The final result in the `ckdata` member vector `Ck` has its size adjusted so that only values greater than `ZEROTOL` (10^{-6}) are saved.

Embedding impedance modeling

This section contains an example program to generate the RF embedding impedance study results presented in Chapter 2 and Appendix D. It uses the SuperMix SIS device and mixer models described in Chapter 5. In addition, the code demonstrates some other convenient features of the SuperMix library, such as command line parameter handling, help prompts, and comment header block generation in a program output. The complete program is presented, which is a few pages long. It will be broken up into segments to make each of its separate tasks clearer. The first several lines of the program define the command line arguments required by the program. Here is the code:

Code example I-6: The RF embedding study program (part 1).

```
// rfembed.cc
// SuperMix source file for a study of RF embedding gamma on SIS
// mixer performance.

char * title =
    "rfembed:                                     \n"
    "Generate a table of SIS noise temperatures and IF output gammas as \n"
    "a function of RF embedding gamma for use in a study of the effect \n"
    "of RF embedding impedance on SIS mixer performance.           \n"
    "\n"
    "Normalizing impedance is the SIS Rn; this is also assumed to be the \n"
    "IF load impedance. Device temperature is assumed to be 4 Kelvin, \n"
    "but RF embedding source temperature is 0 Kelvin. Provide LO and IF \n"
    "freqs, SIS Vgap, Vbias, LO pumping alpha, and number of points in \n"
    "the -1 to +1 interval of gamma values. \n";

#include "supermix.h"
#include "extras/commentstream.h"
#include "extras/cmd_line.h"
#include <cstdlib>

commentstream comment(cout, "# "); // comments will be prefixed with "#"

// =====
// The command-line parameters
command_line_string
    IDC ("IV", "input IV data file.");
```

```

command_line_string
  IKK ("IKK", "input IKK data file.");
command_line_parameter
  Rn (Ohm, "Rn", "SIS normal resistance (ohm)");
command_line_parameter
  Vgap (mVolt, "Vgap", "SIS gap voltage (mV)");
command_line_parameter
  Vbias (mVolt, "Vbias", "SIS DC bias voltage (mV)");
command_line_parameter
  alpha (1.0, "alpha", "LO pumping alpha (eV/hfLO)");
command_line_int
  H ("harmonics", "number of harmonics in the mixer analysis", true, 1, 3);
command_line_int
  Np ("points", "number of points in each gamma dimension (re, im)", true, 3,
200);
command_line_parameter
  fLO (GHz, "fLO", "LO freq (GHz)");
command_line_parameter
  fIF (GHz, "fIF", "IF freq (GHz)");

```

The SuperMix header file *extras/cmd_line.h* contains the declarations for command line argument handling. It will also provide automatic help prompting if the input argument list is improperly formatted and can generate an output comment header block to document the parameter values used in a simulation. Individual command line arguments may be strings, parameters (real-valued numbers), or integers, and they are expected in the order declared in the program file, as illustrated above. Prompt strings are provided for user help, and units and limits may be included. Given the code above, a total of ten command line arguments are expected; if there are too many or too few arguments specified, then the code above will generate a help prompt, as will be described in further detail below.

The RF embedding study program (part 2): helper functions.

```

// convert a complex impedance to a gamma or vice versa
inline complex ZtoG(complex Z, double Zo) { return (Z-Zo)/(Z+Zo); }
inline complex GtoZ(complex G, double Zo) { return Zo*(1+G)/(1-G); }
// =====
// convert an alpha to an RMS LO voltage
inline double AtoV(double a, double fLO) { return
a*(fLO/VoltToFreq)/RmsToPeak; }
// =====
// calculate quantum noise temperature limit at freq f
inline double FtoN(double f) { return f*(hPlanck/BoltzK); }
// =====
// convert a bias and RMS LO voltage combination into a state vector
Vector state(double bias, double vLO)
{ Vector s(2, Index_C); s[0] = bias; s[1] = vLO; return s; }

```

The various physical constants used in the above functions are declared in *units.h*, which is included by *supermix.h*.

The C++ `main()` routine starts with `get_command_line()`, which reads the command line arguments and checks them against the command line parameters the program expects.

The RF embedding study program (part 3): check the command line arguments.

```
int main(int argc, char **argv)
{
    get_command_line(argc, argv, title);
```

If the arguments are too few, too many, or improperly formatted, a help prompt is automatically generated (user console input in bold):

```
$ ./rfembed
rfembed:
Generate a table of SIS noise temperatures and IF output gammas as
a function of RF embedding gamma for use in a study of the effect
of RF embedding impedance on SIS mixer performance.

Normalizing impedance is the SIS Rn; this is also assumed to be the
IF load impedance. Device temperature is assumed to be 4 Kelvin,
but RF embedding source temperature is 0 Kelvin. Provide L0 and IF
freqs, SIS Vgap, Vbias, L0 pumping alpha, and number of points in
the -1 to +1 interval of gamma values.

Usage: ./rfembed <IV> <IKK> <Rn> <Vgap> <Vbias> <alpha> <harmonics> <points>
<fL0> <fIF>
```

Where:

- * <IV>: input IV data file.
- * <IKK>: input IKK data file.
- * <Rn>: SIS normal resistance (ohm)
- * <Vgap>: SIS gap voltage (mV)
- * <Vbias>: SIS DC bias voltage (mV)
- * <alpha>: L0 pumping alpha (eV/hfL0)
- * <harmonics>: number of harmonics in the mixer analysis Must be between 1 and 3 inclusive.
- * <points>: number of points in each gamma dimension (re, im) Must be between 3 and 200 inclusive.
- * <fL0>: L0 freq (GHz)
- * <fIF>: IF freq (GHz)

Here is an example of a properly formatted command line with all required arguments:

```
$ ./rfembed.exe "../common/iv.dat" "../common/ikk.dat" 7.7 2.8 2.3 1.0 2 51
240 8
```

Continuing with `main()`, next it builds the very simple mixer model using the supplied command line arguments. No harmonic balance will be required, because the SIS is assumed to be pumped by a *sinusoidal* LO with the specified frequency (`fL0`) and amplitude given by the specified α (`alpha`). Thus the code simply sets the SIS operating state using its `large_signal()` member function. The RF circuit is especially simple: by using a `transformer`, the SIS RF embedding impedance is changed by modifying the transformer's impedance on the SIS side, which will be its port 1. The IF and DC circuits are even simpler: only a pass-through to the mixer IF output using a two-port `branch`, and a 0-impedance termination (`short_term`) for the DC bias circuit.

The RF embedding study program (part 3): define the model.

```

// -----
// global device settings
device::Z0 = & Rn;
device::T = 4.0*Kelvin;
device::f = fIF;
// -----
// set up SIS
sis_device sis;
sis.Rn = & Rn;
sis.Cap = 0.0; // no capacitance for this analysis
ivcurve iv(IDC.arg.c_str(), IKK.arg.c_str());
sis.set_iv(iv);
sis.Vn = Vgap;
sis.large_signal(state(Vbias, AtoV(alpha, fL0)), fL0, 1);
// -----
// other mixer components; build mixer
transformer RF; // RF.Z1 will be the embedding Z
zterm RF_source; // will terminate RF port for noise calcs
    RF_source.Z = & device::Z0;
    RF_source.Temp = 0.0;
branch IF(2); // 2-port just provides access to the mixer IF output
short_term BIAS; // operating state isn't set using harmonic balance
mixer mix;
    mix.harmonics(H).add_junction(sis).set_LO(&fL0);
    mix.set_bias(BIAS).set_if(IF).set_rf(RF);
    mix.set_balance_terminator(RF_source, 2);

int LSB = mix.port(2, -1);
int USB = mix.port(2, 1);

```

The last two lines above calculate and save integer values for the RF upper and lower sideband mixer ports; the single IF output will be the mixer's port 1.

Next the program generates a comment header with all of the parameter values used in the model. By sending output to the `comment` stream declared and initialized in part 1 of the code listing, each header line will be automatically prefixed with the specified comment delimiter string “# ”.

The RF embedding study program (part 4): output the comment header block.

```

// Output a header with the command-line parameters
char* delim = "-----\n";
state_display::command_line(comment, argc, argv, "rfembed:");
comment << delim;
state_display::info(comment);

// quantum noise temperature limit
comment << fixed << setprecision(1)
    << "Quantum noise limit (at LO freq, in Kelvin): "
    << FtoN(fLO) / Kelvin << endl;

// sis pumped bias current
comment << "SIS pumped DC bias current (uA): "
    << real(sis.I()[0]) / (Micro*Amp) << endl;
comment << delim;

// column headings
comment << "Gamma (Re Im)" << "\t" << "Tn(LSB)" << "\t" << "Tn(USB)"
    << "\t" << "|IF refl|" << endl;
comment << delim;
cout << fixed << setprecision(3);

```

The output generated by this code is shown further below.

Now for the code which calculates a table of the mixer model performance as a function of RF embedding reflection coefficient Γ . It will output upper and lower sideband noise temperature (including ground-state quantum noise) and the squared magnitude of the SIS IF output reflection coefficient, which will exceed unity if the SIS output impedance has a negative real part.

The RF embedding study program (part the last): output a table of model results.

```

// loop over RF gamma values and output table
double noiseL, noiseU, IFR;
int i, j;
double d = 2.0 / (Np - 1); // step size in RF gamma
Complex Grf(-1, -1); // start at RF gamma = -(1+I)
for ( i = 0; i < Np; ++i, Grf += Complex(d, 0) )
    for ( j = 0, Grf.imaginary = -1.0; j < Np; ++j, Grf += Complex(0, d) ) {
        RF.Z1 = GtoZ(Grf, device::Z0);
        noiseL = noiseU = mix.get_term_data().C[1][1].real;
    }

```

```

const sdata & s = mix.get_data();
// output RF gamma:
cout << real(Grf) << "\t" << imag(Grf) << "\t";
// output LSB and USB Tn:
noiseL /= (norm(s.S[1][LSB])*Kelvin);
if (noiseL > 1000.0) noiseL = 1000.0;
noiseU /= (norm(s.S[1][USB])*Kelvin);
if (noiseU > 1000.0) noiseU = 1000.0;
cout << noiseL << "\t" << noiseU << "\t";
// output IF reflection norm (<= 1.0 for stability)
IFr = norm(s.S[1][1]); if (IFr > 2.0) IFr = 2.0;
cout << IFr << endl;
}
}

```

The first several lines of a typical program output are shown below. Note the automatically-generated comment header information it includes.

```

# rfembed:
# Invoked with the command line:
# ./rfembed ../common/iv.dat ../common/ikk.dat 7.7 2.8 2.3 1.0 2 5 240 8
# -----
# State Data:
# input IV data file. : ../common/iv.dat
# input IKK data file. : ../common/ikk.dat
# SIS normal resistance (ohm) : 7.7
# SIS gap voltage (mV) : 2.8
# SIS DC bias voltage (mV) : 2.3
# L0 pumping alpha (eV/hfL0) : 1
# number of harmonics in the mixer analysis : 2
# number of points in each gamma dimension (re, im) : 5
# L0 freq (GHz) : 240
# IF freq (GHz) : 8
# Quantum noise limit (at L0 freq, in Kelvin): 11.5
# SIS pumped DC bias current (uA): 92.6
# -----
# Gamma (Re Im) Tn(LSB) Tn(USB) |IF refl|
# -----
-1.000 -1.000 1000.000 1000.000 0.053
-1.000 -0.500 1000.000 1000.000 0.095
-1.000 0.000 1000.000 1000.000 0.333
-1.000 0.500 1000.000 1000.000 1.173
-1.000 1.000 1000.000 1000.000 2.000
-0.500 -1.000 1000.000 1000.000 0.046
-0.500 -0.500 31.590 33.802 0.137
-0.500 0.000 22.339 24.227 0.491

```

This program, with its many command line arguments, is really meant to be called by a user interface program which constructs a properly-formatted command line and further processes the many lines of model results it generates. The author made extensive use of Wolfram *Mathematica*® for this purpose.

BIBLIOGRAPHY

- [1] John R. Tucker, "Quantum limited detection in tunnel junction mixers," *IEEE J. Quantum Electronics*, vol. QE-15, no. 11, pp. 1234-1258, Nov 1979.
<https://doi.org/10.1109/JQE.1979.1069931>
- [2] Walter R. Steiger. A Brief History of the Caltech Submillimeter Observatory.
[Online]. http://www.cso.caltech.edu/cso_history/CSO_History/
- [3] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. New York: John Wiley & Sons, Inc., 1999, ISBN: 978-0471309321.
- [4] Paul F. Goldsmith, *Quasioptical Systems: Gaussian beam quasioptical propagation and applications*. New York: IEEE Press, 1998, ISBN: 0-7803-3439-6.
- [5] Neil W. Ashcroft and N. David Mermin, *Solid State Physics*. Boston: Cengage Learning, Inc. (formerly Thomson Learning), 1976, ISBN: 978-0030839931.
- [6] Shu-I Hu, *Cryogenic wideband LNA design and characterization*, University of Michigan, Ann Arbor, Thesis (Dissertation (Ph.D.)) 2003.
<http://hdl.handle.net/2027.42/123607>
- [7] A. I. Harris and J. Zmuidzinas, "A wideband lag correlator for heterodyne spectroscopy of broad astronomical and atmospheric spectral lines," *Rev. Sci. Instr.*, vol. 72, no. 2, pp. 1531-1538, Feb 2001. <https://doi.org/10.1063/1.1334629>
- [8] Matthew Casey Sumner, *Optimizing end-to-end system performance for millimeter and submillimeter spectroscopy of protostars : wideband heterodyne receivers and sideband-deconvolution techniques for rapid molecular-line surveys*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.))
CaltechTHESIS:05302011-143941777, 2011.
<http://resolver.caltech.edu/CaltechTHESIS:05302011-143941777>

- [9] E. Serabyn, *Cassegrain relay optics for the CSO chopping secondary*, Caltech Submillimeter Observatory, Hilo, HI, CSO Technical Memo 4 (revised) 1995. http://cso.caltech.edu/docs/memos/optics_memo4.pdf
- [10] Susanna L. Widicus Weaver, Jacob C. Laas, Luyao Zou, Jay A. Kroll, Mary L. Rad et al., "Deep, broadband spectral line surveys of molecule-rich interstellar clouds," *Astrophys. J. Suppl.*, vol. 232, no. 1, Aug 2017. <https://doi.org/10.3847/1538-4365/aa8098>
- [11] Mary L. Radhuber, Jay A. Kroll, Jacob C. Laas, Thomas A. Anderson, Matthew C. Sumner et al., "Submillimeter spectral line survey of Orion," in *2009 CCU Workshop*, Green Bank, WV, 2009, Poster presentation. [http://www.virginia.edu/ccu/WorkshopInfo\(Post-workshop\).html](http://www.virginia.edu/ccu/WorkshopInfo(Post-workshop).html)
- [12] E. C. Sutton, Geoffrey A. Blake, C. R. Masson, and T. G. Phillips, "Molecular line survey of Orion A from 215 to 247 GHz," *Astrophys. J. Suppl.*, vol. 58, pp. 341-378, Jul 1985. <http://resolver.caltech.edu/CaltechAUTHORS:20120919-152843984>
- [13] Jacob W. Kooi, Attila Kovacs, Matthew C. Sumner, Goutam Chattopadhyay, Riley Ceria et al., "A 275-425 GHz tunerless waveguide receiver based on AlN-barrier SIS technology," *IEEE Tran. Microwave Theory Tech.*, vol. 55, no. 10, pp. 2086-2096, Oct 2007. <https://doi.org/10.1109/TMTT.2007.905503>
- [14] Enrico Sacchetti. (2010, Oct.) Flickr. [Online]. <https://www.flickr.com/photos/enricosacchetti/12421786263/>
- [15] John Strawn Ward, *Observations of carbon monoxide in the starburst galaxy M82 with a 690 GHz wide spectral bandwidth receiver*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.)) CaltechETD:etd-01112002-085240, 2002. <http://resolver.caltech.edu/CaltechETD:etd-01112002-085240>
- [16] Mei Bin, *Low-noise THz niobium SIS mixers*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.)) CaltechTHESIS:07172014-124951267,

1997. <http://resolver.caltech.edu/CaltechTHESIS:07172014-124951267>

- [17] Anupama B. Kaul, Bruce Bumble, Karen A. Lee, Henry G. LeDuc, Frank Rice et al., "Fabrication of wide-IF 200-300GHz superconductor–insulator–superconductor mixers with suspended metal beam leads formed on silicon-on-insulator," *J. Vacuum Sci. & Tech. B*, vol. 22, no. 5, pp. 2417-2422, 2004.
<https://doi.org/10.1116/1.1798831>
- [18] A. H. Dayem and R. J. Martin, "Quantum interaction of microwave radiation with tunneling between superconductors," *Phys. Rev. Lett.*, vol. 8, no. 6, pp. 246-248, Mar 1962. <https://doi.org/10.1103/PhysRevLett.8.246>
- [19] Ivar Giaever, "Electron tunneling between two superconductors," *Phys. Rev. Lett.*, vol. 5, no. 4, pp. 464-466, Nov 1960. <https://doi.org/10.1103/PhysRevLett.5.464>
- [20] Ivar Giaever, "Energy gap in superconductors measured by electron tunneling," *Phys. Rev. Lett.*, vol. 5, no. 4, pp. 147-148, Aug 1960.
<https://doi.org/10.1103/PhysRevLett.5.147>
- [21] John R. Tucker and Marc J. Feldman, "Quantum detection at millimeter wavelengths," *Rev. Mod. Phys.*, vol. 57, no. 4, pp. 1055-1113, Oct. 1985.
<https://doi.org/10.1103/RevModPhys.57.1055>
- [22] Stafford Withington and Erik L. Kollberg, "Spectral-domain analysis of harmonic effects in superconducting quasiparticle mixers," *IEEE Tran. Microwave Theory Tech.*, vol. 37, no. 1, pp. 231-238, Jan 1989. <https://doi.org/10.1109/22.20043>
- [23] Goutam Chattopadhyay, *Dual polarized and balanced receivers at millimeter and submillimeter wavelengths*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.)) CaltechETD:etd-02032004-163517, 2000.
<http://resolver.caltech.edu/CaltechETD:etd-02032004-163517>
- [24] S. Withington, G. Yassin, J. Leech, and K. G. Isaak, "An accurate expression for the input impedances of one-sided microstrip probes in waveguide," in *Proceedings*

Tenth International Symposium on Space Terahertz Technology, Charlottesville, VA, 1999, pp. 508-518.

- [25] J. W. Kooi, R. A. Chamberlin, R. Monje, A. Kovács, F. Rice et al., "Performance of the Caltech Submillimeter Observatory dual-color 180-720 GHz balanced SIS receivers," *IEEE Tran. Terahertz Sci. and Tech.*, vol. 4, no. 2, pp. 149-164, 2014.
<https://doi.org/10.1109/TTHZ.2013.2293117>
- [26] J. W. Kooi, G. Chattopadhyay, S. Withington, F. Rice, J. Zmuidzinas et al., "A full-height waveguide to thin-film microstrip transition with exceptional RF bandwidth and coupling efficiency," *Intl. J. IR & MM Waves*, vol. 24, no. 3, pp. 261-284, 2003.
<https://doi.org/10.1023/A:1021903132609>
- [27] R. E. Collin, *Field Theory of Guided Waves*. New York: McGraw-Hill, 1960, A second edition is available: IEEE Press, 1991.
- [28] D. A. Bathker, "A stepped mode transducer using homogeneous waveguides," *IEEE Tran. Microwave Theory Tech.*, vol. 15, no. 2, pp. 128-130, Feb 1967.
<https://doi.org/10.1109/TMTT.1967.1126392>
- [29] Wladimir Mercoureff, "Minimization of thermal losses due to electrical connections in cryostats," *Cryogenics*, vol. 3, no. 3, pp. 171-173, Sep. 1963.
[https://doi.org/10.1016/0011-2275\(63\)90009-2](https://doi.org/10.1016/0011-2275(63)90009-2)
- [30] E. Serabyn, "A wide-field relay optics system for the Caltech Submillimeter Observatory," *International Journal of Infrared and Millimeter Waves*, vol. 18, no. 2, pp. 273-284, 1997, DOI: <https://doi.org/10.1007/BF02677920>.
<https://link.springer.com/content/pdf/10.1007/BF02677920.pdf>
- [31] M. Chan, T. G. Phillips, B. Bumble, H. G. LeDuc J. W. Kooi, "A low noise 230 GHz heterodyne receiver employing $0.25\text{-}\mu\text{m}^2$ area Nb-AlO₂/Nb tunnel junctions," *IEEE Tran. Microwave Theory Tech.*, vol. 40, no. 5, pp. 812-815, May 1992. <https://doi.org/10.1109/22.137383>

- [32] James W. Lamb, "Miscellaneous data on materials for millimetre and submillimetre optics," *Intl. J. IR & MM Waves*, vol. 17, pp. 1997–2034, 1996.
<https://doi.org/10.1007/BF02069487>
- [33] Dominic J. Benford, Michael C. Gaidis, and Jacob W. Kooi, "Optical properties of Zitex in the infrared to submillimeter," *Applied Optics*, vol. 42, no. 25, pp. 5118-5122, Sep. 2003. <https://doi.org/10.1364/AO.42.005118>
- [34] Max Born and Emil Wolf, *Principles of Optics*, Seventh (expanded) ed. Cambridge, U.K.: Cambridge University Press, 1999, ISBN: 0 521 642221.
- [35] John Ward, Frank Rice, Goutam Chattopadhyay, and Jonas Zmuidzinas, "SuperMix: a flexible software library for high-frequency circuit simulation, including SIS mixers and superconducting elements," in *Proc. Tenth Intl. Symp. on Space Terahertz Technology*, Charlottesville, VA, 1999, pp. 269-281.
<https://www.nrao.edu/meetings/isstt/papers/1999/1999269281.pdf>
- [36] Frank Rice and John Ward. SuperMix. [Online].
<http://www.submm.caltech.edu/supermix/>
- [37] Frank Rice. (2018) SuperMix alternate download site. [Online].
<http://www.sophphx.caltech.edu/SuperMix/>
- [38] Scott William Wedge, *Computer-aided design of low noise microwave circuits*, California Institute of Technology, Pasadena, Thesis (Dissertation (Ph.D.)) CaltechETD:etd-09172003-101254, 1991.
<http://resolver.caltech.edu/CaltechETD:etd-09172003-101254>
- [39] Goutam Chattopadhyay, Frank Rice, David Miller, Henry G. LeDuc, and Jonas Zmuidzinas, "A 530-GHz balanced mixer," *IEEE Microwave and Guided Wave Lett.*, vol. 9, no. 11, pp. 467-469, Nov 1999. <https://doi.org/10.1109/75.808038>
- [40] Frank Rice, Matthew Sumner, Jonas Zmuidzinas, R. Hu, Henry G. Leduc et al., "SIS mixer design for a broadband millimeter spectrometer suitable for rapid line surveys

- and redshift determinations," in *Proc. SPIE 4855, Millimeter and Submillimeter Detectors for Astronomy*, February, 2003. <https://doi.org/10.1117/12.459710>
- [41] Cosmic Microwave Technology, Inc. (2019) Cryogenic Low Noise Amplifiers. [Online]. <https://www.cosmicmicrowavetechnology.com/cryogenic-low-noise-amplifiers>
- [42] Herbert B. Callen and Theodore A Welton, "Irreversibility and generalized noise," *Phys. Rev.*, vol. 83, no. 1, pp. 34-40, July 1951. <https://link.aps.org/doi/10.1103/PhysRev.83.34>
- [43] Carlton M. Caves, "Quantum limits on noise in linear amplifiers," *Phys. Rev. D*, vol. 26, no. 8, pp. 1817-1839, Oct 1982. <https://link.aps.org/doi/10.1103/PhysRevD.26.1817>
- [44] Michael James Wengler, *Heterodyne detection with superconducting tunnel diodes*, California Institute of Technology, Pasadena, CA, Thesis (Dissertation (Ph.D.)) CaltechETD:etd-02012007-084647, 1988. <http://resolver.caltech.edu/CaltechETD:etd-02012007-084647>
- [45] Anthony R. Kerr, "Suggestions for revised definitions of noise quantities, including quantum effects," *IEEE Tran. Microwave Theory Tech.*, vol. 47, no. 3, pp. 325-329, Mar 1999. <https://doi.org/10.1109/22.750234>
- [46] R. G. Hicks and P. J. Khan, "Numerical analysis of nonlinear solid-state device excitation in microwave circuits," *IEEE Tran. Microwave Theory Tech.*, vol. 30, no. 3, Mar 1982. <https://doi.org/10.1109/TMTT.1982.1131059>
- [47] M. H. Cohen, L., M. Falicov, and J. C. Phillips, "Superconductive tunneling," *Phys. Rev. Lett.*, vol. 8, no. 8, pp. 316-318, Apr 1962. <https://doi.org/10.1103/PhysRevLett.8.316>
- [48] J. Bardeen, "Tunnelling from a many-particle point of view," *Phys. Rev. Lett.*, vol. 6,

no. 2, pp. 57-59, Jan 1961. <https://doi.org/10.1103/PhysRevLett.6.57>

- [49] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York: Cambridge University Press, 2007, ISBN 978-0521880688.
- [50] Frank Rice, John Ward, Jonas Zmuidzinis, and Goutam Chattopadhyay, "Fast harmonic balance of SIS mixers with multiple junctions and superconducting circuits," in *Proc. Tenth Intl. Symp. on Space Terahertz Technology*, Charlottesville, VA, 1999, pp. 282-297.
<https://www.nrao.edu/meetings/issstt/papers/1999/1999282297.pdf>
- [51] Dr. Jonas X. Zmuidzinis, private conversation, 2000.
- [52] Dr. Thomas G. Phillips, private conversation, 2003.
- [53] Lake Shore Cryotronics, Inc. (2018) Lake Shore Cryotronics. [Online].
https://www.lakeshore.com/Documents/LakeShoreTC_1.pdf
- [54] Lake Shore Cryotronics, Inc. (2018) Lake Shore Cryotronics. [Online].
<https://www.lakeshore.com/products/Cryogenic-Accessories/wire>
- [55] Free Software Foundation. The Free Software Foundation (FSF). [Online].
<https://www.fsf.org/>
- [56] Frank Rice, "Propagation in waveguide III," in *CSO Heterodyne Instrumentation Homepage.*, 2000, <http://www.submm.caltech.edu/cso/receivers/>. [Online].
http://www.submm.caltech.edu/cso/receivers/thermal/par_dielectric_filledWG.pdf
- [57] Frank Rice, "Optimum wire dimensions," in *CSO Heterodyne Instrumentation Homepage.*, 2000, <http://www.submm.caltech.edu/cso/receivers/>. [Online].
http://www.submm.caltech.edu/cso/receivers/thermal/opt_wiredimensions.pdf