Practical, high dynamic range parametric amplification with RF SQUID arrays

by

Ryan Kaufman

Bachelor of Science, University of Pittsburgh, 2019

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial

fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ryan Kaufman

It was defended on

October 29, 2024

and approved by

Michael Hatridge, Department of Physics and Astronomy

Roger Mong, Department of Physics and Astronomy

Thomas Purdy, Department of Physics and Astronomy

Evan Schneider, Department of Physics and Astronomy

Paul Leu, Industrial Engineering

Practical, high dynamic range parametric amplification with RF SQUID arrays

Ryan Kaufman, PhD

University of Pittsburgh, 2024

High-fidelity qubit readout, which is critical for error correction in large-scale quantum computers, requires a short and strong pulse transiting the qubit's measurement mode which is then processed by a necessarily high bandwidth, high saturation power, quantum-limited amplifier. Larger quantum processors additionally employ frequency multiplexing, expanding the required bandwidth by the number of channels to be amplified. Resonant Josephson parametric amplifiers (JPAs), weakly nonlinear harmonic oscillators that use Josephson junctions as nonlinear inductors, have been workhorses in enabling scaling to this point, but they suffer from limitations in dynamic range, bandwidth, and operation with commercial circulators or isolators that are straining their ability to further support scaling readout in ever-larger processors. This thesis shows several methods to combat these shortcomings in resonant JPAs. First, we show data from high saturation-power JPAs that utilize radio frequency superconducting quantum interference device (rf SQUID) arrays to carefully control the nonlinearity of the amplifier, increasing the saturation power while still maintaining their lauded quantum efficiency. With the success of this collaboration with scientists at the Advanced Microwave Photonics group at NIST Boulder, we discuss time-domain qubit measurement with these devices. Here we detail the nanosecond-to-nanosecond dynamics of a qubit measurement, exploring how discrete samples of a qubit readout signal processed by a parametric amplifier can best be weighted and integrated to extract the maximum amount of information possible, especially in the context of high-power qubit readout. Lastly, we address another limitation of reflection Josephson parametric amplifiers: their instantaneous bandwidth. Using similar rf SQUID designs, we show how to tune the equations of motion of a multi-pole device to create a broadband response, allowing for an increase in bandwidth from the 10 MHz scale to the 100 MHz or even 1 GHz scale. Additionally, we will discuss how the simulation methods used for the design of high saturation power devices can extend to these broadened devices. Following this design process, we will detail the fabrication of

more devices at NIST as well as measurements of a similar broadband Josephson parametric amplifier.

Table of Contents

1.0 Introduction and Basic Theory 1

1.1 Quantum-limited readout chains 1

1.1.1 A good experiment requires a good sensor 1

1.1.2 Amplifiers and Added Noise 2

1.1.3 Chains of amplifiers: Friis Equations and concatenated noise 5

1.1.4 Quantum limits to noise fluctuations 9

1.1.5 Thermalization of input signals 10

1.1.6 Parametric amplifiers and the quantum limit 12

1.2 Applications of quantum-limited preamplification 13

1.2.1 Dispersive qubit measurement 14

1.2.1.1 High efficiency for strong and weak measurements 14

1.3 Parametric processes, the Langevin equations, and input-output theory . . . 16

1.3.1 Quantization 16

1.3.2 Hamiltonian of a Parametric amplifier and the rotating-wave approximation 21

1.3.3 Langevin equations and scattering 24

1.4 Common properties of amplifiers 28

1.4.1 Instantaneous and Tunable Bandwidth 28

1.4.2 Frequency multiplexing 31

1.4.3 Saturation and Intermodulation 35

1.4.4 Pump coupling efficiency 40

1.4.4.1 One-port case: voltage division 40

1.4.4.2 Two-port case: ABCD parameters 43

2.0 High dynamic range RF-SQUID amplifiers 47

2.1 Introduction 47

2.2 Amplifier Circuit Design and Fabrication 48

2.3 Results . . . 54
2.4 Conclusion . . . 58
2.5 Periodic steady-state simulation . . . 59
2.6 Experimental details . . . 63
2.6.1 Fabrication . . . 63
2.6.2 Flux biasing: offsets . . . 63
2.6.3 Flux biasing: tunability . . . 64
2.6.4 Flux biasing: uniformity . . . 65
2.6.5 Saturation power calibration . . . 66
2.6.6 Saturation power variation over bias points . . . 67
2.6.7 High power measurement transitions . . . 68
3.0 Time-Domain qubit readout and weight functions . . . 70
3.1 Dispersive readout in the time domain . . . 70
3.1.1 Hardware setup . . . 70
3.1.2 Readout trajectories . . . 72
3.1.3 Cavity pulse-shaping . . . 74
3.2 Matched filters . . . 76
3.3 Majority vote pairwise filtering . . . 79
3.4 Applying Machine Learning: General Temporal Post Processor . . . 83
4.0 Broadband Reflection Parametric Amplifiers . . . 86
4.1 Networks of passively coupled resonators . . . 86
4.1.1 Over, under, and critical coupling . . . 86
4.1.2 Representing coupled electrical resonators in coupled-mode theory . . . 91
4.2 What is a filter? . . . 95
4.3 Impedance-matching a degenerate parametric amplifier . . . 97
4.3.1 Impedance of an active parametric amplifier . . . 98
4.3.2 Power insertion loss method and continued fraction expansion . . . 99
4.3.3 Prototype circuits and tradeoffs . . . 104
4.4 Design and Simulation workflow . . . 107
4.4.1 Types of Resonators . . . 107

4.4.2 Pumpistor Model vs. Parametric Inverter Model 110
4.4.3 Types of Inverters 111
4.4.3.1 Inverter Definition 112
4.4.3.2 Capacitive, Inductive, and Quarter-wave inverters 113
4.4.3.3 Minimizing Nonideality of Quarter-wave inverters 115
4.4.4 Group Delay, Quality Factor, and Fitting Multi-pole Networks 118
4.4.5 Simulating Parametric Amplifiers in ANSYS HFSS 122
4.4.5.1 Component Interpolation: Transmission Line Resonator 123
4.4.5.2 Compensating for negative capacitance inverter legs by shortening distributed resonator length 126
4.4.5.3 Layout assembly 127
4.4.6 Fitting simulation results to a circuit model 130
4.4.7 Pump port simulation 133
4.5 Results 134
4.5.1 Flux nonuniformity: justification for on-chip flux lines 135
4.5.2 Brief introduction to RF-SQUID SNAKE amplifiers 136
4.5.3 Broadband Amplifier Measurements 139
4.5.3.1 Lumped Element Snake Amplifier (LESA) Design 139
4.5.3.2 Gain and saturation power 141
4.5.3.3 LESA noise performance and readout efficiency 142
5.0 Conclusions and Perspectives 147
Appendix. : Code for polynomial long division 151
Bibliography 152

List of Tables

1.1 Table of two-tone mixing amplitudes for mixing products in a 4th-order nonlinear transfer function. 38

1.2 Table of amplitudes for mixing products in a 5th-order nonlinear transfer function. 39

2.1 Qubit coherence versus different states of the amplifier. 56

4.1 Table of component values for three different stages of 15% bandwidth amplifier design. 129

List of Figures

1.1 HEMT and RT amp noise performance vs. noise figure of RT amp, cold loss before the HEMT, and the gain of the HEMT. 7
1.2 Quantum versus classical noise plotted against physical temperature. . . 10
1.3 Analysis of input attenuation stack, 20dB at 4K, 20dB at 100mK, variable at base. 12
1.4 JPA-assisted readout added noise versus gain and cold loss. 13
1.5 Circuit diagram for a resonated rf-SQUID 17
1.6 Tunable bandwidth conceptual diagram. 28
1.7 Inductance modulation of an RF-SQUID with a shunting ratio (β) of 0.25 30
1.8 Common time-domain pulses and their power spectral densities 32
1.9 Channel widths in a frequency multiplexed readout scheme. 34
1.10 Definition of input saturation power, and schematic of intermodulation near saturation. 36
1.11 IP3 intersection plot 38
1.12 Division of a circuit into an environment impedance, coupler impedance, and system impedance for calculating PCE. 41
1.13 Capacitor PCE for a capacitively coupled JPA with Q = 10. 42
1.14 A summary of pump efficiencies of JPAs from other publications, reproduced from [62] and expanded. 43
1.15 The circuit for considering a second pump port in the system. 44
2.1 Configuration of an rf-SQUID array parametric amplifier at the base stage of a dilution refrigerator. 50
2.2 Simulating an rf SQUID array amplifier with periodic steady-state analysis. 51
2.3 Gain and saturation power of an rf SQUID array amplifier. 55
2.4 Weak measurement experiment to determine readout efficiency. 57

2.5 Circuit schematic for writing down the classical equations of motion of an rf SQUID array amplifier. 60
2.6 Simulation details: finding 20dB gain points of the amplifier. 62
2.7 Simulation details: Simulating pump power and pump efficiency required to operate the amplifier. 62
2.8 Bias flux sweeps for two different RF-SQUID array devices. 64
2.9 Flux biasing nonuniformity frequency components. 66
2.10 Photon number power calibration via qubit spectroscopy. 67
2.11 Saturation power variation over bias flux for a high-power RF-SQUID array 68
2.12 High power readout histograms for a transmon qubit 69
3.1 Full circuit diagram of the high-power rf-SQUID JPA experiment. . . . 71
3.2 Voltage response of a driven resonator detuned according to qubit state. 73
3.3 Readout trajectories for the first four qubit states at two different drive frequencies. 74
3.4 Shaped pulses used to decrease the ring-up and ring-down times of the intracavity population of the readout resonator. 75
3.5 Pairwise weight functions derived from Ref. [60] and their results with integrated mean values of trajectories. 78
3.6 Readout trajectories and weight functions on a real qubit measurement setup. 80
3.7 Readout histograms for 2-state data integrated with each respective weight function. 82
3.8 Classification performance of TPP versus FGDA for readout of real qubits. 84
3.9 Adaptation of TPP-learned filters with increasing measurement tone amplitude and evolving noise conditions. 85
4.1 Single resonator with internal loss 87
4.2 Overcoupling, Undercoupling, and Critical Coupling for a single resonator 88
4.3 Single LC resonator driven from two ports of differing impedance. . . . 89

4.4 Two-port, one parallel lossless resonator transmission in terms of port quality factors. 90
4.5 Two-port one-resonator system with internal loss. 91
4.6 Electrical circuit for impedance inverter-coupled resonators 92
4.7 Comparing accuracy of coupled-mode description of scattering vs. quality factor 95
4.8 Filter scattering generated by varying coupling strength. 96
4.9 Active amplifier impedance vs signal to idler coupling strength 98
4.10 3-pole matched amplifier circuit schematic. 103
4.11 20dB gain profile of multiple coupled-mode prototype circuits with 15% bandwidth, comparing required inductance modulation and quality factor. 105
4.12 Circuit representations of various useful impedance/admittance inverters. 113
4.13 Variation in the inverter values of common inverters vs frequency. . . . 114
4.14 Scattering parameters of a 3-pole Chebyshev prototype synthesized with ideal frequency-independent admittance inverters. 114
4.15 Capacitive inverters used to synthesize various fractional bandwidth filters. 115
4.16 Transmission line inverter variation over frequency. 115
4.17 Broadband amplifiers synthesized with quarter-wave inverters with respect to both bandwidth and the impedance of the resonators in the circuit. 116
4.18 Phase wraps and group delay for under and over-coupled resonators. . . 119
4.19 Phase wrap and group delay of 3-pole Chebyshev filter prototypes vs. fractional bandwidth of the designs. 121
4.20 Broadband parametric amplifier design flow. 122
4.21 Possible modes of propagation for electric fields in a CPWG transmission line architecture. 124
4.22 Resonator parameter interpolation from HFSS simulation. 126
4.23 Layout and group delay fit for a 15% bandwidth parametric amplifier with a pump port. 128

4.24 Fitting a multi-pole amplifier with exported scattering from HFSS in AWR Microwave Office. 130
4.25 Modelling the gain of a simulated broadened amplifier layout. 131
4.26 Fitting two-port phase coupling efficiency. 133
4.27 A circuit with a much more even pole-free PCE near the pump frequency. 134
4.28 Flux nonuniformity fitting to amplifier phase modulation profiles. . . . 135
4.29 Breakdown of rf-SQUID SNAKE design. 137
4.30 Schematic and layout of the Lumped-Element Snake Amplifier (LESA) 139
4.31 Gain and saturation power of eight different LESA devices. 141
4.32 Readout efficiency of LESA devices and noise performance near saturation compared to TWPA devices in Ref.[50]. 143
4.33 LESA IMD behavior at various inter-tone detunings. 144

Dedication

To my family. You gave me the spirit of invention, and the will to see this through.

Acknowledgements

In pursuing this degree, I owe a huge amount to my family, friends, and mentors who have helped me along the way. To start, I owe a huge thanks to my advisor Michael Hatridge for being a patient mentor and giving me both the freedom to pursue what I thought was interesting, and the guidance to make that effort useful. Additionally, I am very grateful that he gave me the opportunities, from visiting NIST to interning at Google, that set down the bedrock on which I will build my career.

Additionally, I want to thank my collaborators Joe Aumentado, Kat Cicak, and the entire NIST AMP group, who have been nothing short of instrumental in both inspiring me to think of new research paths and providing the environment and equipment to do so. Additionally, I want to thank Ofer Naaman and Ted White at Google. The discussions we had about amplifier research remain some of the work I am most interested in today, and I look forward to working with you soon!

My peers in the lab have also been immensely helpful, both for my work and my health. Even though I didn't always show it, the best part of my day was talking about science and life with all of you over lunch or coffee. Much of my work would have felt empty if I had no one to share it with, so thank you all for listening and helping whenever possible. I wish all of you the best in your careers!

I also owe a special thanks to Dr. Wargo. Your classroom showed me what real science could be, long before I would get the chance to do it professionally. The inspiration from that experience carried me through many of the difficulties in my education and still pushes me forward today.

To my friends Sam, Jordan, Nick, Zach, Katie, John, and Isaac in Pittsburgh, you are a godsend. Through both undergrad and grad school, I would not have made it through if not for your love and support. Thank you for putting up with all my BS, and giving me the backup that I needed to make it through.

Lastly, I owe all of this to my family. Mom and Dad, I owe you my work ethic and my ambition. You have always given me the freedom to pursue my passion, showed me the way

forward when I encountered difficulties, encouraged me to push myself to the limits of my abilities, and shown me what it means to help those in need in the family. The values that you instilled in me will provide a guiding light for the rest of my life, and for that I am forever thankful. The masonry work helped too, ten years of school is *nothing* compared to that.

1.0 Introduction and Basic Theory

1.1 Quantum-limited readout chains

1.1.1 A good experiment requires a good sensor

The design of any good scientific experiment requires a good sensor to make a precise and efficient measurement. In some ways, one could argue that the experiment itself is the *engineering* of a good sensor. In order to build ourselves a good experiment (a good sensor), we have to know the conditions around the observables that we would like to measure. In this respect, quantum computing is no different as an object of experimental study than any other piece of quantum theory. Therefore, I want to focus on the common factor of any quantum experiment: they are all very, very "cold". What I mean by this is that we need to take special care to perform the experiment at temperatures that place the energies of thermal fluctuations ($\approx k_b T$) well below quantum fluctuations of a harmonic oscillator at the relevant frequency of the experiment ($\hbar\omega/2$) in order to access quantum phenomena (coherence, entanglement, etc.) and measure them effectively. In this way, we can define a goal - a "quantum temperature" - that we strive to get as far below as possible: $T_Q = \frac{\hbar\omega}{2k_b T}$.

For transitions at 100 THz, $T_Q = 2400$K, meaning that even room temperature gives $T << T_Q$. However, for transitions at microwave frequencies relevant in this thesis and in many other quantum applications, the drive is closer to the 5 GHz level, which corresponds to $T_Q = 120$mK. Because of this choice, we require much more careful construction of the experimental apparatus that I will cover in this section. Not only do we need a dilution refrigerator (*not* covered in this thesis) in order to get the physical temperature of the system well below this mark, we also need to carefully send electrical signals from room-temperature equipment to the experiment and, critically, extract a delicate electrical signal at this cold temperature and deliver it to the comparatively "hot" room-temperature equipment. In this section, I will cover how we do this, what can go wrong, and how the amplifiers that I improved in this thesis are critical tools to accomplish this task. In the process of learning

this material for myself in the first few years of my PhD, I found it quite difficult to chain the references together in the proper order and bounce back and forth between various standards to gain an understanding of the core physics in addition to simultaneously learning about the latest developments in parametric amplifiers. In the next few sections, I hope to alleviate the struggle for future students by presenting a cohesive story from start to finish that covers not only the interesting quantum physics but also gives references to the best practices that we need to borrow from electrical engineering.

1.1.2 Amplifiers and Added Noise

To break into this topic, let's go over some basics of radio-frequency (rf) amplification in a classical context (which I am pulling directly from Ref. [47]), and then try to use those basics to work towards addressing our specific quantum problem. All amplifiers affect the Signal-to-Noise Ratio (SNR) of the signal that they are processing, which is calculated in power as $\mathrm{SNR} = S/N$, where S is the coherent signal power of a sinusoidal tone with frequency ω and the noise (N) is an integrated noise power within the bandwidth of interest. This effect is succinctly summarized by the *noise figure* given by[1]

$$F = \frac{\mathrm{SNR}_i}{\mathrm{SNR}_o} = \frac{S_i N_o}{S_o N_i}. \tag{1.1}$$

Here we arrive at the first quirk of (linear) amplifiers. Despite their ubiquitous usage in every single microwave transmitter and receiver, they *never* increase the *local* signal-to-noise ratio. I view the intuitive reason that this must be the case as the amplifier cannot tell the difference between what is signal and what is noise within some bandwidth. This fact that linear amplifiers must have $\mathrm{SNR}_o \leq \mathrm{SNR}_i$ means that every linear amplifier has to add noise, and we can describe the output of an amplifier with (again, linear) power gain G with $S_o = GS_i$, and $N_o = G(N_i) + N_{a,out}$, where $N_{a,out}$ is the added noise power (units

[1] I have to caution the reader already here that if you are just learning about all this, please stick to *just* one source. If you bounce back and forth between whichever site comes up on Google as the first result you'll get a lot of confusing definitions that are inconsistent across sources. For example, some sources call what I have defined here the noise *factor*, and reserve noise *figure* for the dB measure of this definition. That's not a huge discrepancy, but hopping back and forth requires careful attention to the units of every definition, just because microwave engineering has been around for quite some time now and there are a lot of independent sources that take liberties from the original textbook definitions.

of J*Hz) referred to the output of the amplifier. This definition of added noise has the most straightforward conceptual interpretation. Some noise was present initially, then went in and got amplified to the output. At some point, the extra noise was added inside the amplifier and the output noise includes this additive factor. Despite the conceptual tidiness, it is often more convenient in the analysis of large microwave networks to play a mental trick here, which is that adding some noise *inside* the amplifier has the same result as if the amplifier was *noiseless* and the extra noise was at the input of the amplifier, but divided by the gain, shown by

$$N_o = G(N_i) + N_{a,out} = G(N_i + N_a). \tag{1.2}$$

This new symbol N_a is the most common measure of added noise, called *input-referred* added noise. We will get to a more detailed discussion of the bounds on N_a shortly, but for now I want to focus on another useful idea, which is that the input noise can (classically) be assigned an *effective* noise temperature T_i within a certain bandwidth B given by

$$N_i = k_b T_i B. \tag{1.3}$$

With this definition, then, the input-referred added noise can be assigned an equivalent temperature similarly,

$$N_a = k_b T_a B \tag{1.4}$$

and the two can be related by the noise figure with

$$(F-1)\frac{N_i}{k_b B} = \frac{S_i N_o - S_o N_i}{S_o N_i}\frac{N_i}{k_b B} = (\frac{S_i}{S_o}N_o - N_i)\frac{1}{k_b B} = \\ (\frac{1}{G}(G(N_i + N_a)) - N_i)\frac{1}{k_b B} = \frac{N_a}{k_b B} := T_a. \tag{1.5}$$

Or, more succinctly,

$$\frac{T_a}{T_i} = F - 1. \tag{1.6}$$

Now, I want to caution that these temperatures are not physical temperatures. They are expressions of how much (classical, right now) thermal fluctuations would correspond to a particular added noise *power* within a certain bandwidth of a brick-wall filter whose

width is B. Fundamentally, they are proxies for noise power referred to the input of the amplifier. This is where standards established by microwave engineers cause some confusion. If you look at how the noise is actually measured, (Y-factor measurements, which I will not discuss here but Ref [47] does very well), one has to start with an input noise power (temperature) and increase it in a controlled way and look at how the measured output noise power (temperature) changes. Because it is easy to build a noise source at *room temperature*, the standard for quoting noise figure measurements is to assume an input noise temperature of 290K, then quote the noise figure with that input which will give (recall $F \geq 1$) an input-referred output total noise ($N_o/(k_b B)$) temperature above 290K, and the difference represents the added noise power ($N_a/(k_b B)$).

Please allow me to illustrate the problems with this standard when applied to our (admittedly very niche) context of cryogenic amplification. As an example, Low-Noise Factory (LNF) makes very good, incredibly low-noise cryogenic amplifiers made from High Electron-Mobility Transistors (HEMTs). For example, take the LNF LNC4_8F. They claim two things on their datasheets: a noise temperature of $T_a = 1.5$K and a noise figure of (in linear units) 1.005. Incredible! I want to emphasize, despite my incoming criticism, that this is an amazing feat of engineering. However, there is a problem with the communication. One would then naively expect that since these amplifiers operate at a *physical* cryogenic temperature of 4K, we could back up that noise figure by reversing Eq. 1.6 and calculating

$$F = 1 + \frac{1.5}{4} = 1.375. \tag{1.7}$$

Clearly, this does not reproduce their number. Why? It is precisely because despite the amplifier being *designed* to operate at 4K, and is actually operated there in our readout chains, they quote its noise figure in relation to 290K, because that is the standard. Using this, we of course recover their number which is

$$F = 1 + \frac{1.5}{290} = 1.005. \tag{1.8}$$

My claim is that this is misleading. The amplifier is *not* operated at 290K! The reason that 290K is the standard is because the *vast* majority of microwave electronics are used at

that temperature, and so it allows the easy application of noise figure to get back-of-the-envelope calculations of amplifier performance and it generally avoids a lengthy discussion about where the noise is referred to. However, this immediately breaks down when the input noise source is not at that 290K. This creates a silly proxy, because we have to first use the noise figure with the reference temperature to determine the added noise *power* that they're talking about, then convert that number into an effective temperature. In other words, you can't go from noise figure directly to noise temperature using the physical temperature of the device, noise power calculated from the reference temperature is a necessary middleman. To drive this home, I will walk through an example using the same physical temperature to reference temperature imbalance as LNF. Starting with a decidedly mediocre Minicircuits amplifier with an added noise temperature of 87K, I could then define a reference temperature of 16,820K so that I can quote a noise figure of $1 + 87/16,820 = 1.005$. It looks like a phenomenal amplifier! But in reality, the magnitude of noise figure as a metric is useless unless you are at the reference temperature. In this example, all I have to say about that is good luck amplifying on the sun. Because of this ambiguity in using noise figure across large differences in physical temperatures, I am going to avoid using noise figure whenever possible for the rest of this thesis, focusing instead on added noise temperature and added noise power.

1.1.3 Chains of amplifiers: Friis Equations and concatenated noise

Now that we have discussed added noise, noise temperature, and noise figure, let's discuss how these metrics change throughout a chain of amplifiers, which will be crucial to how we set up the readout chains in our quantum computers. The added noise behavior in a chain of amplifiers with gains $\{G_i\}$ and added noise temperatures $\{T_{e,i}\}$ is described by Friis' equation in Ref. [47]. I will show it below for the case of three concatenated amplifiers.

$$T_{chain} = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} \tag{1.9}$$

Again, to reiterate the same point that has come up multiple times now, this is the equivalent *added* noise temperature of the whole chain. The total input-referred noise tem-

perature is still $T_{total} = T_{in} + T_{chain}$. In any case, Friis' equation tells us the important fact that if the first amplifier in the chain has sufficiently high power gain, then it is mostly the only determining factor of the chain's total noise performance! This is good because it means we can separate amplifiers into two types that are good at different things. In the low-noise first amplification stage, we can build low-noise amplifiers (LNAs) with very good noise performance. Then, since the signal will already be amplified after the LNA stage, we can build power amplifiers to handle higher input power after the LNA, without worrying as much about added noise. This is how most microwave receiver setups are built, with the lowest noise amplifiers front and center, and additional amplifiers only serving to counteract losses in the chain and high added noise temperatures of room temperature instruments like spectrum analyzers. For some examples of this, see Fig. 1.1.

The last piece of classical information that we need to know for the purposes of this thesis is that even a good piece of room temperature equipment, say a spectrum analyzer, generally has a very high noise figure or very large effective added noise temperature. To give a ballpark number, it is generally around $T = 29$kK, because the first stages of a spectrum analyzer are an input attenuator and a mixer, each with a loss of around 10 dB. This gives a noise figure of at least 100, hence the high added noise temperature.

Finally, we have enough tools to walk this abstract discussion back to measuring quantum signals. Let's say that we have an input noise power at an equivalent temperature of $T_{in} = T_Q = 120$mK at $\omega_0 = 5$ GHz with an SNR of $\mathrm{SNR}_{in} = 10$. What happens when we connect that signal directly to a spectrum analyzer? What would we see? Ignoring the thermodynamics involved with directly connecting a 290K machine to a 120mK source, all we have to do is compute the total added noise power and compare that to the signal, which we have not amplified at all. The result when $T_o = 29$kK is

$$\mathrm{SNR}_o = \mathrm{SNR}_i \left(\frac{T_Q}{T_o + T_Q} \right) = 4.1 \times 10^{-5}. \tag{1.10}$$

Needless to say, this is exceedingly bad. In fact, if we calculate the added noise temperature, we find it to be entirely dominated by the added noise of the spectrum analyzer. Since this will turn into a discussion of quantum-limited amplifiers, let's compute the input-referred added noise in terms of *photon number* instead of temperature. With our 5 GHz input tone,

and the noise corresponding to a $\hbar\omega/2$ at 5 GHz, the input signal is 5 photons strong, but the output added noise in terms of photon number is a whopping 29 kK $k_b/(\hbar\omega) = 120000$ photons! It is entirely reasonable, then, that our signal is utterly swamped by this tsunami of noise.

How can we do better? I will skip a myriad of other examples of what *not* to do and skip to the best "off-the-shelf" case of the following: the LNF LNC4_8F HEMT amplifier I mentioned earlier, backed by a standard room temperature (RT) LNA with 20dB gain going directly into that same spectrum analyzer. Figure 1.1 analyzes this situation at a frequency of 5 GHz with respect to three key variables in order to determine which are the most important for the total added noise of the chain.

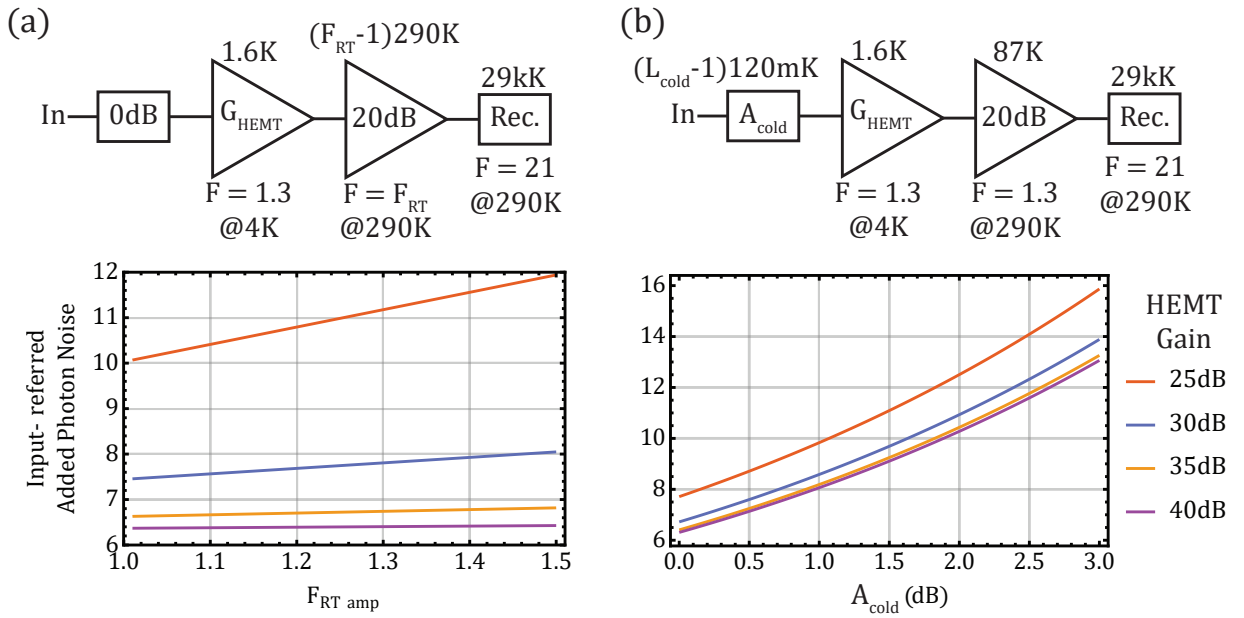


Figure 1.1: HEMT and RT amp noise performance vs noise figure of RT amp, cold loss before the HEMT, and the gain of the HEMT. Above each circuit element is its added noise temperature, below is its noise figure and physical temperature. Within each element is either its loss (rectangular attenuators) or its gain (triangular amplifiers). (a) An off-the-shelf readout chain for measuring quantum signals, viewed without any cold loss before the HEMT with respect to the noise figure of the amplifier at room temperature on the x-axis and the gain of the HEMT shown by the color of each line. (b) The same circuit as before, except now the noise figure of the RT amp is held constant at 1.3 corresponding to a reasonably economical amplifier.

The three variables I chose to analyze here are the following: the power gain of the HEMT, the noise figure of the room-temperature amplifier (F_{RT}), and the attenuation (A_{cold}) before the HEMT. It's reasonable to ask why I am analyzing the gain of the HEMT, since we do not have much control over what LNF will sell us. The reason I do this is that HEMTs usually

stay in use for quite some time, and (speaking from experience) they are highly susceptible to many graduate student-related sudden aging factors that may result in reduced gain, so it is useful to understand how important any gain variations are and how those losses in gain might ripple up through the chain and make themselves visible in careful calibrations.

In Figure 1.1 we can immediately see (by looking at the initial values of both subplots) that although this readout chain's input-referred added noise is much, *much* better than directly connecting to a spectrum analyzer, 6 photons is still not great. With this, we will still have an SNR of less than one with a 5-photon input. Why not increase the input power? I will discuss it in chapters 2 and 3, but the short version is that the quantum system that we are measuring simply breaks.

Next, Fig. 1.1(a) tells us how much the noise figure of the room-temperature amplifier matters. We can see that the answer depends on how much gain the HEMT is providing. If the HEMT is performing as it should by its specification (40 dB gain and a 1.5K noise temperature) then the added noise of the RT amp hardly matters, increasing the photon noise of the chain by much less than a photon in the range of the plot. However, if the HEMT is significantly under performing (as old HEMTs sometimes do) then the noise factor of the room temperature amplifier does matter, but in this case even a very good (read: expensive) room temperature amplifier will only make a bad readout chain slightly less bad. In this way, buying a fancy room temperature LNA with only a noise figure of 1.1 is not so much a universal boon as it is an insurance policy. If the equipment you are using is well-maintained and you know your HEMT is performing well you will not need it, but if you find out your HEMT was under performing over the last three months of your experiment then you will be glad you had it. All that's left for the reader to do is decide on the insurance premium. Would you like the right-hand side of the plot for $50, or the left-hand side for $5000? I'll leave the choice to you, but I chose the middle at about $120, and calibrated the HEMTs once per year.

But what if we are confident that the HEMT is working well? Figure 1.1(b) shows the effect of an unavoidable consequence of the physical size (≈meters) of the readout chain: there is inevitably some loss between the sample at the base stage of the dilution refrigerator and the HEMT amplifier, sitting multiple feet away at the 4K stage. This loss hits *much*

harder than the noise figure of the RT amplifier, and is often exacerbated by the inclusion of multiple circulators and isolators that lie between the device under test (DUT) and the HEMT. This loss can easily tally up to 3 dB, which halves the SNR, as you might expect.

The keen observer may point to Fig. 1.1 and ask "Wait, why is the first loss/attenuation stage listed with a physical temperature of 120 mK? I thought that dilution refrigerators routinely reached temperatures as low as 10 mK?" They do, and my label of 120 mK is a bit of a hack. I will explain the suspiciously high noise temperature in the next section.

1.1.4 Quantum limits to noise fluctuations

Let's give a little more detail on why the quantum temperature T_Q that I defined in the last section is important in the context of noise. Borrowing from Clerk et al. in Ref. [9], we see that we can define an input noise spectral density of a quantum circuit with real input impedance Z_s as

$$S_{v_{in},v_{in}} = Z_s \hbar\omega \coth\left(\frac{\hbar\omega}{2k_bT}\right). \tag{1.11}$$

Because this noise spectral density is *two-sided*, as in to calculate the noise power we have to integrate over both the negative and positive frequency bandwidth (effectively giving a factor of two) the high-temperature dependence of this expression does not match up with the equipartition theorem of one harmonic oscillator [58]. By applying the equipartition theorem to a model of a matched transmission line where each photon frequency occupies a harmonic oscillator with two degrees of freedom, we should have $S_{vv} = k_bT$ for $T >> T_Q$, but this function indicates twice that. Accordingly, I divide by two in all my analyses to make it back into a one-sided noise PSD[2]. This also brings this function in line with the high-temperature Rayleigh-Jeans approximation of the noise temperature discussed in Ref. [47], which also gives $S_{vv} = k_bT$ for $T >> T_Q$, again, for a one-sided noise PSD.

[2]This is a much abbreviated version of a much longer and more in depth conversation treated, at length, in Ref. [9] and Ref. [10]. This debate between whether and/or how to include negative versus positive frequencies shows up in many places, e.g. the design of single-sideband (and, generally, complex) mixers in Ref. [20] and of course the parametric amplifier coupled-mode formalism in Ref.[41]. For now I just want the reader to be aware of the terms "one-sided" vs. "two-sided" noise PSD.

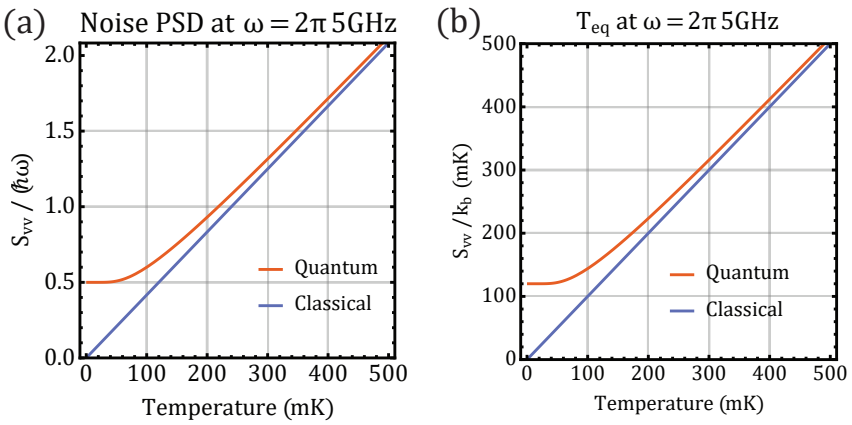


Figure 1.2: Quantum versus classical noise plotted against physical temperature. (a) Comparing the high-temperature classical approximation to the quantum expression in Eq. 1.11 (again, divided by two) in units of photon number. (b) Same as (a) but with equivalent noise temperature.

Figure 1.2 compares these two in both added photon number and effective noise temperature. The most important conclusion to draw is that below a certain physical temperature, the physical temperature no longer determines the noise power. Instead, quantum fluctuations take over, which only depend on frequency. The limit on the left-hand side of Fig. 1.2(b) is why I chose the 120 mK "temperature" for the attenuator in Fig. 1.1(b). This hack remedies the fact that the noise of cold (again, $T << T_Q$) attenuators is not determined by their physical temperature, and so we have to substitute the usual physical temperature that would be used for attenuators for the effective temperature of quantum fluctuations. The hyperbolic cotangent function in Eq. 1.11 effectively interpolates between the two. For $T << T_Q$, the proper way to avoid this hack is to consider how noise of concatenated elements add by using the math of quantum optics and beam splitters. To see how this works, let's analyze the other side of the signal chain: delivering input signals to the base stage of a dilution refrigerator. This next section will use the results of Ref. [27] and add some commentary that is relevant to the context of this thesis.

1.1.5 Thermalization of input signals

So far we have ignored the process of delivering a signal into the quantum system under test. This is the reverse conundrum to getting low power, low noise signals out of a dilution

refrigerator. How do we get high SNR signals *into* the dilution refrigerator? Once again, room temperature noise is a real problem. As a quick example, if we once again plugged a room temperature microwave generator directly into the DUT, then we would be bombarding it with hundreds of thousand of noise photons. Quantum systems do not remain quantum under these conditions.

To lower the amount and the temperature of the radiation, we run the signal through stages of attenuation at progressively lower and lower physical temperatures. Attenuators used in this way serve the reverse purpose of amplifiers. The idea, from Ref. [27], is that attenuators (and cold loss in general) act like a beam splitter, permitting a certain amount (1% in a 20 dB attenuator) of the noise from the last stage to pass, but re-emitting a thermal distribution of noise from its physical temperature. That thermal distribution is the Bose-Einstein distribution in our case, so we reach the following recursion equation for the total output noise $N_{o,i}$ of the ith attenuator in a chain from Ref. [27] which I have reproduced for ease of reference here as

$$N_{o,i} = \frac{N_{o,i-1}}{A_i} + \frac{A_i - 1}{A_i} \frac{1}{e^{\frac{\hbar\omega}{k_b T}} - 1} \tag{1.12}$$

where A_i is the attenuation in linear units (A = 100 for 20 dB). The nonlinearity of this recursion relation makes creating a linearized quantum version of Eq. 1.9 impossible (at least without some severe approximation), and so we are left with using this recursion relation in combination with Eq.1.11 in order to calculate the total quantum noises of cold attenuators and amplifiers. Figure 1.3(a) shows a typical attenuator stack that we commonly use in our dilution refrigerators.

Figure 1.3(b) shows the result for the input attenuation stack, which gives a thermal photon number occupation at the base stage versus the last-stage attenuation, in line with Ref. [27]. Why do we care about such small photon numbers? The tiny magnitude of the photon noise makes it seem like we could have just done the calculation classically and been done with it. Figure 1.3(c) goes the last step to explain why we need the fully quantum equations. Using Ref. [51], we can calculate the dephasing that occurs when these thermal photons interact with a readout resonator, leading to a cap on the T_2 of any dispersively coupled qubit that is *extremely* sensitive to ambient thermal photons. The temperature of

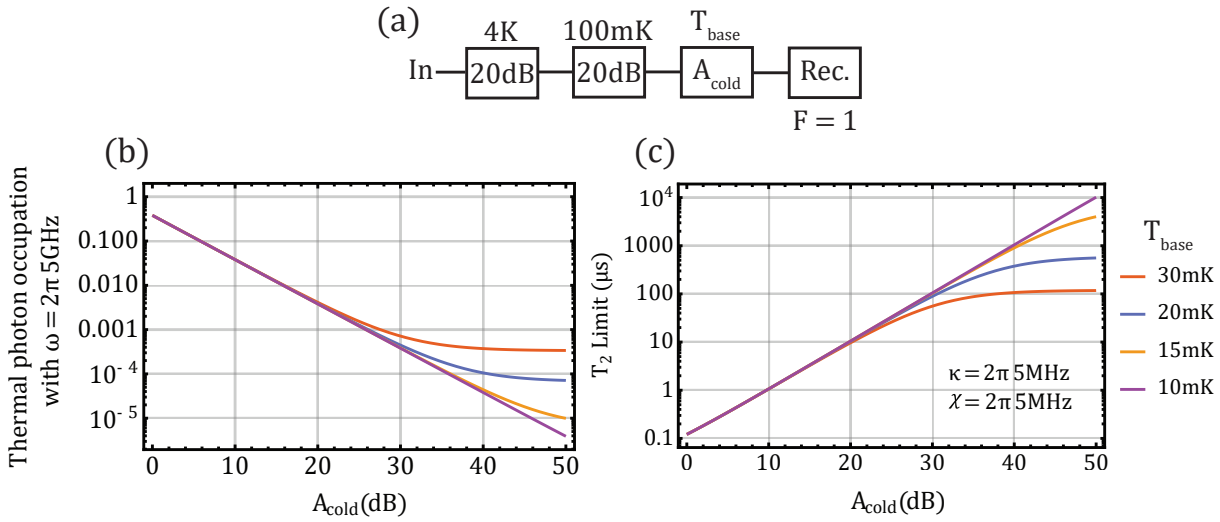


Figure 1.3: Analysis of a typical input attenuation stack, 20dB at 4K, 20dB at 100mK, with variable attenuation at the base stage. (a) The setup of attenuators with their physical temperatures and losses marked in dB.(b) Added photon noise number, including the minimum 1/2 lower bound for a phase-preserving amplifier [7] (c) T_2 upper bound using (b) as a thermal photon number from Ref. [51].

the last stage of attenuation also needs to be as cold as possible. Even temperatures of only 30 mK start to significantly contribute to decreases in T_2. Now that we have a way to describe the loss in cold attenuators, we can finally move on to a discussion of the full output chain with parametric amplifiers and loss, and how we can approach the quantum limit.

1.1.6 Parametric amplifiers and the quantum limit

One advantage of breaking up Friis equation into recursive steps is that each step need not be the same as the last, meaning that we can make the recursion relation in Eq. 1.12 take the quantum photon noise power spectral density (PSD) in Eq. 1.11 as an input. Figure 1.4 shows the result versus cold losses before the parametric amplifier and the parametric amplifier gain.

Comparing Fig. 1.4 to Fig.1.1(b) we can immediately see the sharp reduction to the added noise at high JPA gain, usually by over a factor of 10. This is the critical role of the JPA in a readout chain: it serves as the low-noise vanguard of the readout chain, protecting the SNR from the added noise of the HEMT stage. The catch is that we must be extremely careful to keep any losses before the amplifier to a minimum, else the added noise rises above

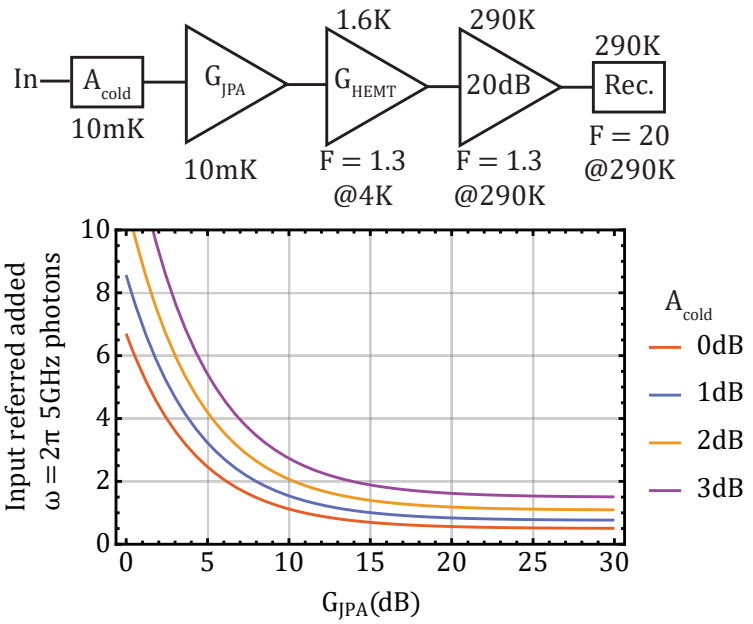


Figure 1.4: Quantum-limited JPA-assisted readout added noise versus gain and cold loss. The circuit diagram and all temperatures are shown above. Cold losses directly detract from the limits to added noise, making it common to refer to a system's efficiency in terms of a quantum-limited amplifier preceded by some cold loss.

the half-photon limit. Some losses from cabling and packaging are (for now) inevitable, and Fig. 1.4 shows that these losses directly add noise to the readout chain.

Now that we've covered the general steps of building a quantum-limited amplification chain, we can move on to why that might be of interest for some fundamental physical measurements on quantum systems.

1.2 Applications of quantum-limited preamplification

In this section, we will start by introducing the standard cQED dispersive qubit measurement scheme, the concept of strong and weak measurement, the important associated metrics, and their relationship to amplification. In the second part, we will introduce the state-of-the-art for directional parametric amplifiers. I would also like to acknowledge Florent Lecocq in the National Institute for Standards and Technology (NIST) for his assistance in writing this section.

1.2.1 Dispersive qubit measurement

A standard method for assessing the state of a superconducting qubit is to couple it to a readout resonator[5]. When strongly off-resonant, in the so-called dispersive regime, the frequency of the readout resonator depends on the qubit state, enabling a qubit measurement by probing the resonator's microwave response. This readout method is a quantum nondemolition (QND) measurement so long as the readout power is low enough, typically on the order of a few photons. Higher readout power increases the strength of the measurement, but eventually drives unwanted transitions in the qubit, destroying the state being measured. The measurement signal needs then to be amplified by over six orders of magnitude before reaching room temperature electronics, with minimal degradation of the signal-to-noise ratio. This is usually done in multiple stages of amplification that we discussed in the last section, starting from a parametric amplifier for the first stage [2]. The amplification chain can be characterized by a single numerical factor quantifying the SNR reduction, the system's *efficiency*. In the following, we will discuss the impact of this measurement efficiency and the limitation of traditional amplifier chains.

1.2.1.1 High efficiency for strong and weak measurements

The strength of a measurement is often divided in two categories, weak and strong measurements, both benefiting from a high measurement efficiency. Strong measurements are the canonical quantum measurements whereby the system being measured is fully projected into one of its eigenstates, e.g. 0 or 1 for a qubit. The fidelity of a strong measurement, defined as $1 - P(0|1) - P(1|0)$, the agreement between the measurement result and the qubit state, is critical for mid-circuit measurements acting as part of an error-correction scheme [28]. Secondly, the QND-ness of the measurement, defined in part by how often the measured qubit remains in the eigenstate detected by the measurement, is much more stringent, as a destructive measurement error destroys the utility of doing a mid-circuit measurement in the first place. Both of these constraints at once present a necessary constraint on maximizing the efficiency of the measurement, so that low photon numbers can extract the maximum fidelity measurement without activating destructive non-QND readout effects.

Weak measurements are not strong enough to fully project a qubit into an eigenstate of the measurement by the completion of the measurement. Instead, weak measurement gives the qubit a stochastic "kick" around the Bloch sphere, leaving the qubit in a mixed state, with its purity determined by the measurement efficiency [18]. In a perfectly efficient measurement, the qubit remains in a pure state throughout the process. Several publications have tracked the Bloch vector trajectory of the qubit, using it to extract the measurement efficiency [18, 21].

High measurement efficiency is crucial for analog feedback schemes used in state stabilization. Feeding any excess noise back into a qubit only increases its dephasing, resulting in the stabilization of a mixed state instead of a pure state. The loop delay of the feedback cycle also contributes to state impurity, competing with the qubit coherence time. However, previous work on measurement-based state stabilization has shown that amplifier efficiency is the primary limiting factor [67]. Typical amplification chains use a parametric amplifier as the first stage, thanks to noise performance fundamentally limited by Heisenberg's uncertainty principle [2]. However, both reflection parametric amplifiers and traveling-wave parametric amplifiers currently require external circulators or isolators to shield the qubit from excess back-action, with numerous drawbacks. First, since most multiplexing schemes have a maximum of ten readout resonators on a single amplifier, the increasing number of external magnetic circulators is becoming a major problem in large-scale quantum processors with thousands of qubits. Second, since circulators need strong magnetic fields to utilize the Faraday effect and create nonreciprocal scattering, they cannot be easily integrated on-chip with processors. Further, because they have to be outside of magnetic shielding, large numbers of circulators are taking up more space than the processors in the base-stage of dilution refrigerators. Finally, these circulators and associated wiring introduce loss that *decreases* the quantum efficiency of the readout chain, as Figure 1.4 shows in the last section.

Since the parametric processes that we take advantage of to create amplifiers and other devices that process quantum information, it is useful to understand how the circuit diagrams that we draw translate into quantum operators. This is the major bridge between electrical engineering and this part of quantum information processing. As a graduate student starting out, I found a lot of the connections between them (e.g. "How is impedance defined again?")

confusing, initially. Looking back, an accessible and excruciatingly detailed derivation of the quantization of an LC circuit would ave been useful to see early-on, so that I could link some ideas from electrical engineering to those in quantum optics. I found out later that this existed all along in Ref. [12], and I will reproduce the derivation in the next section at the level of a senior undergraduate who only knows electrical device equations and is vaguely familiar with quantum operators and commutation relations.

To a graduate student that has a class that covers this, the next few sections are likely to be duplications of what they have already seen. However, in my case, the classes that taught this material were few and far between and therefore practically inaccessible until I was a much older graduate student. I hope that the next few sections can be helpful to any students who may be in a similar predicament.

1.3 Parametric processes, the Langevin equations, and input-output theory

1.3.1 Quantization

In this section, we will give a brief introduction to how we model the dynamics of parametric amplifiers based largely on the work in [12]. First, we will briefly cover the quantization of a basic rf-SQUID Hamiltonian going to reach a raising/lowering operator description for a parametric gain process. Then, we will show the quantum Langevin equations for the system, calculate the commutator of an rf-SQUID Hamiltonian, and write the linearized equations for a single quantum mode under an external drive. We will move to a rotating wave approximation of these dynamics, and describe how applying a parametric pump at particular frequencies can activate various parametric processes. Once we have reached the rotating-wave description of a parametric gain process in a single-mode amplifier, we can describe the result in a matrix, and link to broader work that will be useful in later chapters. First, we begin the potential from an rf-SQUID resonator circuit shown in Fig.1.5 via the device equations.

The parallel connection allows us to connect the device equation for an inductor and

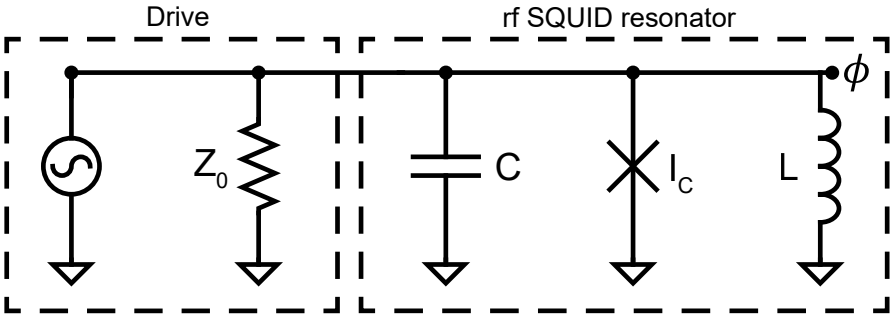


Figure 1.5: Circuit diagram for a resonated rf-SQUID

capacitor

$$V_L = L\frac{dI_L}{dt}, V_C = \frac{1}{C}\int I_c dt \tag{1.13}$$

to those of a Josephson junction below with reduced flux variable ϕ and reduced magnetic flux quantum ϕ_0 [8]

$$I_J = I_0\cos(\phi), V_J = \phi_0\frac{d\phi}{dt} \tag{1.14}$$

to allow us to write the energy of each circuit element below:

$$U_C = \frac{C}{2}V_C^2, U_L = \frac{L}{2}I_L^2, U_J = -E_J cos(\phi) \tag{1.15}$$

in terms of only a single variable ϕ and its derivatives, with $E_J := \phi_0 I_c$.

$$U_L = \frac{L}{2}\left(\frac{\phi_0\phi}{L}\right)^2 = \frac{\phi_0^2}{2L}\phi^2, U_C = \frac{C}{2}(\phi_0\dot{\phi})^2, U_J = -E_J\cos(\phi) \tag{1.16}$$

Notice in particular that because the voltage on the capacitor must be the same as the voltage on the junction, we have the following relationship between the charge and phase

$$V_J = \phi_0\frac{d\phi}{dt} = QC \implies Q = \frac{\phi_0}{C}\dot{\phi} \tag{1.17}$$

enabling an exact comparison to position and momentum: $p = m\dot{x}$ with $m = \frac{\phi_0}{C}$. Because the energy in the capacitor now appears analogous to kinetic energy, we are motivated to write the Lagrangian as

$$L = T - U = (U_C) - (U_L + U_J) = \frac{C\phi_0^2}{2}\dot{\phi}^2 - \frac{\phi_0^2}{2L}\phi^2 + E_J\cos(\phi) \tag{1.18}$$

For a classical simulation of this Lagrangian, it is valid to move forward by adding a driving voltage into the Lagrangian with external dissipation. We will do this in Ch. 2 to solve the driven dynamics of an rf-SQUID amplifier. For now, we will focus on how these energies are quantized, and describe the dynamics of the resulting operators.

By expanding the cosine term of the total energy, we can see how we might move forward in quantizing the conjugate phase and charge variables, as well as identify the Josephson junction as a nonlinear inductor.

$$U_{total} = U_C + U_L + U_J = \frac{Q^2}{2C} + \frac{\phi_0^2}{2L}\phi^2 - E_J\left(1 - \frac{\phi^2}{2} + \frac{\phi^4}{24} + ...\right)$$

Here the inductance-like term for the junction is $E_J\phi^2/2$, enabling $E_J = \phi_0^2/L_J$ to compare the Josephson junction to an inductor whose inductance $L_J = \frac{\phi_0}{I_0}$ with critical current I_0. Here you can also see why there's a negative sign in the Josephson energy, which is so that the quadratic term of the junction energy has the correct sign. It is also worth considering the same expansion when an external flux is allowed to thread the connection between the junction and the inductor, which will appear later in Ch. 2 when we discuss the classical equations of motion for the system. Moving forward, we have observed that charge and phase are conjugate variables in exactly the same way as position and momentum, so we can quantize these operators, changing variables by the transformations

$$\hat{\Phi} = \hat{\phi}\phi_0 = \phi_{zpf}(\hat{a} + \hat{a}^\dagger), \hat{Q} = Q_{zpf}(\hat{a} - \hat{a}^\dagger) \tag{1.19}$$

where the raising and lowering operators are unitless, inheriting all of their dimensions from the constants Q_{zpf} and ϕ_{zpf}. We can see what those constants must be by plugging them in, and comparing them to the known result for a classical harmonic oscillator combined with the constraint on the commutator

$$\left[\hat{a}, \hat{a}^\dagger\right] = \hat{a}\hat{a}^\dagger - \hat{a}^\dagger a = 1. \tag{1.20}$$

Doing so, we get

$$\hat{H} = \frac{Q_{zpf}^2}{2C}(\hat{a} - \hat{a}^\dagger)^2 + \frac{\phi_{zpf}^2}{2L_t}(\hat{a} + \hat{a}^\dagger)^2 \tag{1.21}$$

where $L_t = \frac{LL_J}{L+L_J}$. Follow by writing out the products

$$(a - a^\dagger)^2 = a^2 - aa^\dagger - a^\dagger a + a^{\dagger 2} = a^2 - (1 + a^\dagger a) - a^\dagger a + a^{\dagger 2} = -2a^\dagger a + a^2 + a^{\dagger 2} - 1 \tag{1.22}$$

and

$$(a + a^\dagger)^2 = a^2 + aa^\dagger + a^\dagger a + a^{\dagger 2} = 2a^\dagger a + a^2 + a^{\dagger 2} + 1. \tag{1.23}$$

Putting all this into the Hamiltonian, with the capacitive prefactor called α and the inductive prefactor called β we get that

$$\hat{H} = 2(\beta - \alpha)(a^\dagger a + \frac{1}{2}) + (\alpha + \beta)(a^2 + a^{\dagger 2}). \tag{1.24}$$

Moving on, we see that to match the Hamiltonian of a quantum harmonic oscillator we must have

$$\hbar\omega = \frac{\hbar}{\sqrt{L_t C}} = 2\,(\beta - \alpha) \tag{1.25}$$

with $Z = \sqrt{L_t/C} \implies L_t = Z^2 C$ we can write this as

$$2(\beta - \alpha) = \frac{2}{2C}\left(\frac{\phi_{zpf}^2}{Z^2} - Q_{zpf}^2\right) = \frac{\hbar}{ZC}. \tag{1.26}$$

Rearranging the right-hand equality we get

$$\left(\frac{\phi_{zpf}^2}{Z} - ZQ_{zpf}^2\right) = \hbar. \tag{1.27}$$

Here we use the commutation relation

$$\begin{aligned}[\hat{\Phi}, \hat{Q}] = \phi_{zpf}Q_{zpf}\left((a - a^\dagger)(a + a^\dagger) - (a + a^\dagger)(a - a^\dagger)\right) = \\ \phi_{zpf}Q_{zpf}(aa^\dagger - a^\dagger a - (-aa^\dagger + a^\dagger a)) = \\ 2\phi_{zpf}Q_{zpf}[a, a^\dagger] \implies \phi_{zpf}Q_{zpf} = \frac{i\hbar}{2}.\end{aligned} \tag{1.28}$$

Satisfying both Eq 1.27 and Eq. 1.28 requires that

$$Z = i\frac{\phi_{zpf}}{Q_{zpf}}.$$

This result says neatly that in this particular case, the impedance (a real number) describes the ratio between current and phase fluctuations, which for Z to be real must be two complex numbers separated by 90 degrees.

Lastly, we arrive at the final Hamiltonian, in which we also need to sort out the $\alpha + \beta$ term. We now know that

$$\frac{\beta}{\alpha} = \frac{\phi_{zpf}^2}{2L_t}\frac{2C}{Q_{zpf}^2} = -1$$

meaning that the $\beta + \alpha$ term is simply 0. This leads to a satisfying conclusion: the choice of impedance that satisfies the commutation relation also makes the Hamiltonian diagonal in the ladder-operator basis, with the final Hamiltonian [12]

$$\hat{H} = \hbar\omega(a^\dagger a + \frac{1}{2}).$$

Despite the simplicity of this example, the ideas are nearly universally applicable to large superconducting circuits, with one conceptual jump. The mode that we quantized here using the procedure from Ref. [12] was the only trivial eigenmode of the circuit, and it was easy to choose the charge across the capacitor and the flux as the canonically conjugate variables.

In circuits with many more nodes and branches on which to support many more voltages and currents, we still retain the idea of resonances, but they are more spread out and occupy multiple nodes and branches in the circuit. One of the first pieces of literature to deal with this specific issue is Ref. [45], which laid the foundation for this exact quantization treatment using finite-element EM simulation software. Work still continues in this field today, because without the assistance of software to determine the eigenmodes of the physical metal layout, it's generally quite difficult to choose the conjugate variables of a large multi-branch multi-node circuit. For very recent work on this topic, please see Ref. [46]. Now that we have an idea of the Hamiltonian of a harmonic oscillator, we can move on to more practical considerations with Josephson parametric amplifiers.

1.3.2 Hamiltonian of a Parametric amplifier and the rotating-wave approximation

In this section I will go through the approximations we need to use to go from a Hamiltonian to a scattering matrix under a parametric pump. Since parametric amplifiers operate with hundreds of photons in the amplifier, in amplifier design we mostly operate under the working approximation that we are in the classical correspondence limit, and we can describe everything except the noise performance by looking at the classical equations of motion without any quantum mechanics needed to describe the dynamics at all. Accordingly, the utility of this section is to remove the quantum mechanics entirely from the toolkit, but justify that removal with a series of approximations so that we can just talk about properties of matrices and scattering in the rest of the thesis and go back to check noise properties later.

First we start with three harmonic oscillators with lowering operators a_i and frequencies ω_i in a Hamiltonian with a single three-wave coupling term.

$$\frac{H}{\hbar} = \sum_j \omega_j a_j^\dagger a_j + g_3(a_1 + a_1^\dagger)(a_2 + a_3^\dagger)(a_3 + a_3^\dagger) \tag{1.29}$$

It's important to pause here and comment on the various situations that this Hamiltonian can represent. This Hamiltonian is written diagonalized, meaning that the three oscillators have no direct coupling to one another. This is a surprisingly weak constraint, meaning that it's a very flexible Hamiltonian.

For example, if these operators represent physical resonances of a circuit, this is something that will come out of an eigenmode EM simulation in Ansys HFSS, but not from a circuit synthesis of coupled harmonic oscillators. Alternatively, these raising and lowering operators do not *have* to represent physical resonances. They can also represent different driving frequencies of the same resonator since raising and lowering operators are orthogonal so long as they oscillate at different frequencies. In this case, we can think of the three-wave coupling term as just representing three different-frequency sources (that may even be delivered from the same physical port) of a flux drive that mix in a cubic nonlinearity.

In either case, the next step is a stiff pump approximation. We can pick whichever drive has the highest strength and ask what will happen if its magnitude is unaffected by the other

two. To keep the link to flux drives, I will stiffen the operator

$$(a_3 + a_3^\dagger) \to \eta \cos(\omega_p t), \tag{1.30}$$

where η is a complex, time-independent variable (phasor magnitude). The pump tone it scales oscillates at ω_p with a phase offset θ_p. This is equivalent to $a \to \frac{\eta}{2} e^{i\omega_p t}$.

Now we are left with a simple idea, that the term ηg_3 controls a time-varying coupling between a_1 and a_2, $g_{eff} = \eta g_3 \cos(\omega_p t)$ which gives

$$\frac{H}{\hbar} = \omega_1 a_1^\dagger a_1 + \omega_2 a_2^\dagger a_2 + \omega_p |\eta|^2 \cos^2(\omega_p t + \theta_p) + g_{eff}(a_1 + a_1^\dagger)(a_2 + a_2^\dagger) \tag{1.31}$$

With the time-dependence more clear, we can discuss the frequencies in the problem. The question is this: given the pump frequency and assuming the operators oscillate at their respective frequencies of ω_1 for a_1 and ω_2 for a_2, which terms in front of operators will not oscillate at all in this frame? The harmonic terms will clearly survive as their exponents cancel, and the pump term that used to be the harmonic term for a_3 will also survive, with amplitude $\frac{\omega_p |\eta|^2}{2}$ from expanding $\cos^2(x) = \frac{1+cos2x}{2}$. Most importantly, the two-body term $(g_{eff}(a_1 + a_1^\dagger)(a_2 + a_2^\dagger))$ may also survive, *depending on the pump frequency in relation to the signal frequency.*

To get a sense of the frequencies we can either do the transform formally (see [72]) or just stiffen all the modes and see which ones arithmetically cancel. I will favor the latter.

If all the modes are stiffened, we end up with a simple product of three cosines to consider.

$$H_{int} \propto \cos(\omega_1 t) \cos(\omega_2 t) \cos(\omega_p t) \tag{1.32}$$

The product of cosines has a nice intuitive picture, one that we will elaborate on much more in a later section on intermodulation. To see which pump frequencies will result in a surviving non-rotating component, all we have to do is simplify the product to

$$H_{int} \propto \left(\frac{\cos(\omega_1 + \omega_2 t) + \cos(\omega_1 - \omega_2 t)}{2} \right) \cos(\omega_p t). \tag{1.33}$$

Since the only way to arrive at a nonrotating term is to create a $\cos^2(x)$ term which will have a constant component, we see that we only have two choices of the pump frequency

that will lead to terms that survive the RWA, the sum $\omega_p = \omega_1 + \omega_2$ or the difference $\omega_p = \omega_1 - \omega_2$. This logic can be applied to any order interaction in any mixing Hamiltonian, it's a consequence of the correspondence principle and the fact that the same logic applies to classical mixers, frequency doublers, and more.

There are many caveats to applying the rotating-wave approximation. One is that it does not cope well with multiple narrowly detuned tones. The reason is that the RWA doesn't distinguish between "barely rotating" and "rotating fast", it discards them both the same way. In the case of narrowly detuned tones, we should consider keeping any terms that fall within the linewidth γ of one another. There is also another ambiguity, which is that the magnitude of "small" terms is up for debate. If the pump tone is very large, then even though the magnitudes of "fast-rotating" terms start small, they can be multiplied with quite large pump amplitudes, bringing them right back up to par with the harmonic terms and lower-order nonlinearities. This is not just a universal bane, it can be leveraged to do interesting things such as subharmonic gates on qubits [70]. The rotating-wave approximation and its various quirks comes up again and again. On the one hand the linearized analytical clarity of the result is appealing in theory. However, as certain tradeoffs (e.g. between decoherence and ease of driving gates) become unavoidable in the linear theory and strained in building larger quantum processors, nonlinear methods that take advantage of the terms we used to throw away are becoming more and more appealing.

While applications of rotating terms are interesting to investigate generally, here I want to focus just on the linear scattering theory, then deal with those higher-order terms only via experimentally measuring them later in chapter 4. In that vein, if we choose the pump frequency $\omega_p = \omega_1 + \omega_2$, ignore the pump energy term, and discard all remaining rotating terms, we see that the Hamiltonian looks like

$$\frac{H}{\hbar} = \omega_1 a_1^\dagger a_1 + \omega_2 a_2^\dagger a_2 + g_{eff} a_1 a_2 + g_{eff}^* a_1^\dagger a_2^\dagger. \tag{1.34}$$

This is the standard Hamiltonian for a nondegenerate parametric amplifier. Next, we'll use some of the tools of linear quantum optics to analyze how this Hamiltonian produces gain, and build to some useful expressions of the amplifier's gain and bandwidth.

1.3.3 Langevin equations and scattering

To describe the dynamics and see how Eq. 1.34 produces gain, we write the quantum Langevin equations for both a_1 and a_2 as

$$\dot{a_j} = \frac{i}{\hbar}[H, a_j] - \frac{\gamma_j}{2}a_j + \sqrt{\gamma_j}a_{j,in}. \tag{1.35}$$

Where we use input-output theory for each operator a_i, and so there are no drive components in the Hamiltonian H itself.

$$a_{i,out} + a_{i,in} = \sqrt{\gamma_i}a_i \tag{1.36}$$

Computing the RHS of Eq. 1.35 for a_1 and $a_2^\dagger$, we see that

$$\dot{a_1} = -(i\omega_1 + \frac{\gamma_1}{2})a_1 + ig_{eff}a_2^\dagger + \sqrt{\gamma_1}a_{1,in}, \text{ and} \tag{1.37}$$

$$\dot{a_2}^\dagger = -(-i\omega_2 + \frac{\gamma_2}{2})a_2^\dagger - ig_{eff}^* a_1 + \sqrt{\gamma_2}a_{2,in}^\dagger \tag{1.38}$$

By applying a pump at $\omega_p = \omega_1 + \omega_2$ we choose the sum term in Eq. 1.33, and create a positive feedback loop between a_1 and $a_2^\dagger$ with a characteristic rate g_{eff} which competes directly with the mode's decay rate γ_1. In combination with the conjugate equation for $a_2^\dagger$, we can see right away that the equations are coupled, which suggests a matrix approach with a complex exponential ansatz. Additionally, we choose to drop the hats off of the operators, and solve for just the operators' steady-state expectation values. There is a more detailed discussion of this step in Ref. [53] and Ref. [10]. The conceptual reduction of the argument there is that by dropping the operator nature of the equations, the amplifier is acting like a classical object that quantum fields are sent into, processed by, and sent out from [10]. Applying this idea, we can get $a_j = \langle a_j \rangle e^{s_j t}$, where now $\langle a \rangle$ is merely a complex number. Here $s = \sigma_j - i\omega_j^s$ with ω_j^s representing the frequency of an input signal to mode j. In this context, no one usually talks about σ, the exponential rise/fall component of the input. However, I will leave it in for the reason that it provides a crucial link to both stability analysis of coupled quantum Langevin equations and the optimization of broadband

directional parametric processes. We make this full switch to the complex domain with a Laplace transform, making the LHS into

$$\dot{a_j} \rightarrow s\langle a_j \rangle e^{st}. \tag{1.39}$$

We do the same with the RHS of both Eqs. 1.37 and 1.38 to get

$$s\langle a_1 \rangle = -(i\omega_1 + \frac{\gamma_1}{2})\langle a_1 \rangle + ig_{eff}\langle a_2 \rangle^* + \sqrt{\gamma_1}\langle a_{1,in} \rangle \tag{1.40}$$

$$s^*\langle a_2 \rangle^* = -(-i\omega_2 + \frac{\gamma_2}{2})\langle a_2 \rangle^* - ig^*_{eff}\langle a_1 \rangle + \sqrt{\gamma_2}\langle a_{2,in} \rangle^* \tag{1.41}$$

The last step is to move all the $\langle a_j \rangle$ terms to the left-hand side, and incorporate the output operators $a_{1,out}$ and $a_{2,out}$ with the input-output equations in the form $\langle a_{j,in} \rangle + \langle a_{j,out} \rangle = \sqrt{\gamma_j}\langle a \rangle$ to get a description of the amplitude scattering parameters $S_{jk} = \frac{\langle a_j \rangle}{\langle a_k \rangle}$. To make things a little easier, we can borrow a slightly modified definition from [41] with

$$\Delta_j = \frac{1}{\gamma_0}(is - \omega_j + i\frac{\gamma_k}{2}) \tag{1.42}$$

and

$$\beta_{jk} = \frac{c_{jk}}{2\gamma_0}. \tag{1.43}$$

Because the detunings of all signals from their modes can be kept the same, we can scan the response of the system with only one independent frequency parameter $\delta = \omega_j^s - \omega_j$, which lets us directly use the results of Refs. [49] and [41] to use only one variable s in the same way. Here, $\gamma_0 = \sqrt[n]{\gamma_1\gamma_2...\gamma_n}$ generalizes the resonant and coupling terms to multimode systems with different decay rates in each mode. In our case it is $\gamma_0 = \sqrt{\gamma_1\gamma_2}$. The mode cooperativity $c_{jk} = g_{eff}$ in this case. This lets us write Eqs.1.40 and 1.41 as

$$-i\gamma_0\Delta_1\langle a_1 \rangle = i\gamma_0\beta_{12}\langle a_2 \rangle^* + \sqrt{\gamma_1}\langle a_{1,in} \rangle. \tag{1.44}$$

$$i\gamma_0\Delta_2^*\langle a_2 \rangle^* = -i\gamma_0\beta_{12}^*\langle a_1 \rangle + \sqrt{\gamma_1}\langle a_{2,in} \rangle^* \tag{1.45}$$

If we eliminate the operators a_1 and a_2 with the input and output operators,

$$-i\gamma_0\Delta_1\frac{\langle a_{1,in}\rangle+\langle a_{1,out}\rangle}{\sqrt{\gamma_1}}=i\gamma_0\beta_{12}\frac{\langle a^*_{2,in}\rangle+\langle a^*_{2,out}\rangle}{\sqrt{\gamma_2}}+\sqrt{\gamma_1}\langle a_{1,in}\rangle. \tag{1.46}$$

$$i\gamma_0\Delta^*_2\frac{\langle a^*_{2,in}\rangle+\langle a^*_{2,out}\rangle}{\sqrt{\gamma_2}}=-i\gamma_0\beta^*_{12}\frac{\langle a_{1,in}\rangle+\langle a_{1,out}\rangle}{\sqrt{\gamma_1}}+\sqrt{\gamma_1}\langle a_{2,in}\rangle^* \tag{1.47}$$

From here it is clear we could eliminate variables and solve for whichever scattering parameter we need. However, there is a more convenient way to write this if we rearrange Eqs. 1.46 and 1.47 to read

$$-i\gamma_0\Delta_1\langle a_1\rangle-i\gamma_0\beta_{12}\langle a_2\rangle^*=\sqrt{\gamma_1}\langle a_{1,in}\rangle. \tag{1.48}$$

$$i\gamma_0\Delta^*_2\langle a_2\rangle^*+i\gamma_0\beta^*_{12}\langle a_1\rangle=\sqrt{\gamma_1}\langle a_{2,in}\rangle^*. \tag{1.49}$$

The last remaining step is to organize these equations into matrix form, where we can use the matrices M and K from Refs.[49] and [41]. Here I will write out the matrix form of the equations above and their conjugate equations in one go,

$$M=\begin{bmatrix}\Delta_1 & 0 & 0 & \beta_{12}\\ 0 & \Delta_2 & \beta_{21} & 0\\ 0 & -\beta^*_{21} & -\Delta^*_1 & 0\\ -\beta^*_{12} & 0 & 0 & -\Delta^*_2\end{bmatrix}, K=\begin{bmatrix}\sqrt{\gamma_1} & 0 & 0 & 0\\ 0 & \sqrt{\gamma_2} & 0 & 0\\ 0 & 0 & \sqrt{\gamma_1} & 0\\ 0 & 0 & 0 & \sqrt{\gamma_2}\end{bmatrix}, \tag{1.50}$$

which we can use to write a single matrix equation in place of these two coupled equations after defining the vectors

$$\vec{v}=\begin{bmatrix}\langle a_1\rangle\\ \langle a_2\rangle\\ \langle a^\dagger_1\rangle\\ \langle a^\dagger_2\rangle\end{bmatrix}, \vec{v}_{in}=\begin{bmatrix}\langle a_{1,in}\rangle\\ \langle a_{2,in}\rangle\\ \langle a^\dagger_{1,in}\rangle\\ \langle a^\dagger_{2,in}\rangle)\end{bmatrix}, \text{and } \vec{v}_{out}=\begin{bmatrix}\langle a_{1,out}\rangle\\ \langle a_{2,out}\rangle\\ \langle a^\dagger_{1,out}\rangle\\ \langle a^\dagger_{2,out}\rangle\end{bmatrix}. \tag{1.51}$$

This series of definitions allows the entire set of coupled quantum Langevin equations expressed simply as

$$-i\gamma_0 M\vec{v}=K\vec{v}_{in}. \tag{1.52}$$

We can even write input-output theory en masse in vector format with

$$\vec{v} = K^{-1}(\vec{v}_{in} + \vec{v}_{out}). \tag{1.53}$$

The goal of calculating all of the output averages in terms of the input averages is in sight, substituting Eq. 1.53 into Eq. 1.52 we arrive at

$$\begin{aligned} -i\gamma_0 M K^{-1}(\vec{v}_{in} + \vec{v}_{out}) &= K\vec{v}_{in} \\ (\vec{v}_{in} + \vec{v}_{out}) &= \frac{i}{\gamma_0} K M^{-1} K \vec{v}_{in} \\ \vec{v}_{out} = (\frac{i}{\gamma_0} K M^{-1} K - I)\vec{v}_{in} &= S\vec{v}_{in}, \\ S &= \frac{i}{\gamma_0} K M^{-1} K - I \end{aligned} \tag{1.54}$$

Now we can easily compute the reflection gain of the amplifier with S_{11} below as

$$S_{11} = \sqrt{G}(s) = \left(-1 - \frac{i\gamma_1 \Delta_2^*}{\gamma_0 \left(\beta_{12}\beta_{12}^* - \Delta_1 \Delta_2^*\right)}\right) \tag{1.55}$$

We can define $\rho = 2|\beta_{12}|/\gamma_0$ and look at an on-resonance drive to see that this simply yields

$$S_{11}|_{s=0} = -\frac{1+\rho^2}{1-\rho^2} \tag{1.56}$$

This lets us define a power gain of the amplifier, G, with the square of the value of the scattering parameter on-resonance,

$$G \equiv |S_{11}|^2 \tag{1.57}$$

.

Now that we have a convenient expression for the gain that we can trace all the way back through the rotating-wave approximation and the quantization of the oscillator Hamiltonian, we can move on to discuss the more relevant day-to-day "datasheet" properties of a parametric amplifier and how they interact with the engineering needs of a sizable quantum processor.

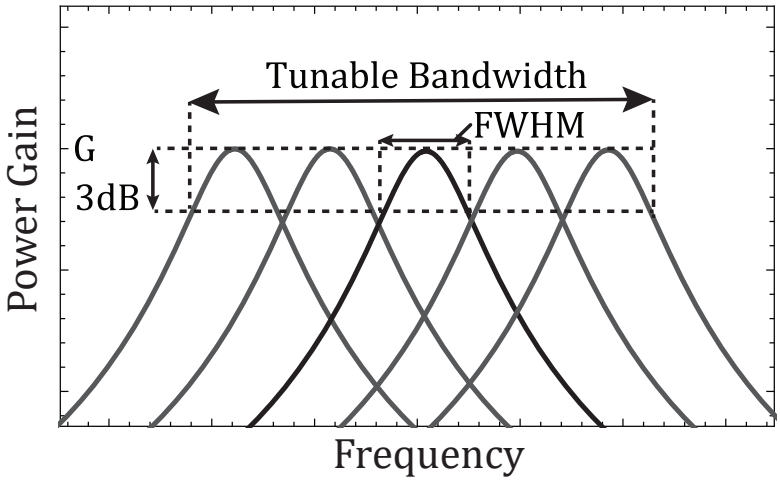


Figure 1.6: Tunable bandwidth conceptual diagram. Schematic of instantaneous gain of multiple gain curves, with instantaneous bandwidth labeled as "FWHM" and the tunable bandwidth over multiple bias points.

1.4 Common properties of amplifiers

In this section, we will go over many common properties of Josephson parametric amplifiers. First, we will discuss bandwidth and the distinction between instantaneous and tunable bandwidth. Next, we will discuss input and output saturation power, why they must be properties of any nonlinear system, and how to judge how much power a particular qubit will require to measure. To go with this, I will also discuss intermodulation as another consequence of nonlinear systems. Lastly, building off of the definitions of saturation power I will discuss pump efficiency from two different angles, and how we might more efficiently couple pump light into Josephson parametric amplifiers.

1.4.1 Instantaneous and Tunable Bandwidth

Instantaneous bandwidth describes the frequency window over which the power gain (or any scattering parameter) varies by a linear factor of two. It also goes by the names "3dB Bandwidth" and "Full-Width-Half-Maximum Bandwidth" (FWHM). In the case of a single-

mode degenerate parametric amplifier, which has a reflection scattering parameter [3] gain profile described by the complex function

$$S_{11}(\delta) = -\frac{4\delta^2 + \gamma^2(1+\rho^2)}{4\delta^2 + 4i\delta\gamma + \gamma^2(1-\rho^2)}. \tag{1.58}$$

Alternatively, expressed in absolute value in terms of the power gain G on-resonance,

$$|S_{11}(\delta)| = \frac{2\delta^2 + \sqrt{G}\left(\gamma^2 + 2\delta^2\right)}{\sqrt{\gamma^4 + 4\gamma^2\left(\delta^2 G + \delta^2\sqrt{G}\right) + 4\delta^4\left(\sqrt{G}+1\right)^2}} \tag{1.59}$$

By squaring this, then Taylor expanding this around small detuning, we can see

$$|S_{11}|^2(\delta) \approx G - \frac{4\delta^2\left(G^{3/2} + G^2 - G - \sqrt{G}\right)}{\kappa^2} + O\left(\delta^3\right) \tag{1.60}$$

For high gain, we can approximate the quadratic detuning term with just the G^2 term, and solve for when the gain becomes half the original:

$$|S_{11}|^2 = \frac{G}{2} \implies G - \frac{4\delta^2 G}{\gamma^2} = \frac{G}{2} \implies \delta = \pm\frac{\gamma}{2\sqrt{2G}} \tag{1.61}$$

So we see that the full-width half max is twice that value, $\mathrm{B} = \gamma/\sqrt{2G}$, making the gain-bandwidth product a constant

$$\sqrt{G}\mathrm{B} = \frac{\gamma}{\sqrt{2}}. \tag{1.62}$$

With a maximum external port coupling of $\gamma = 500$ MHz, we get a maximum bandwidth of $\mathrm{B} \approx 35$ MHz that is typical for single-mode reflection JPAs. However, for situations where we are only measuring a single resonator at a time, this bandwidth does not accurately capture the full utility of the amplifier. JPAs are (usually) widely tunable in resonant frequency, opening up the option to pump the amplifier at many different bias points to allow one device to cover a much larger *tunable* bandwidth, albeit not all at the same time and requiring a change in pump frequency. See, for example, the lightly shaded gain profiles in Fig. 1.6.

[3] Since the Langevin equations describe the relationships between ladder operators, the quantity we are discussing here is actually closer to a quadrature gain of an optical field, $a + a^\dagger$. Of course, this is directly related to the phase and current via the quantizations in Eq.1.19.

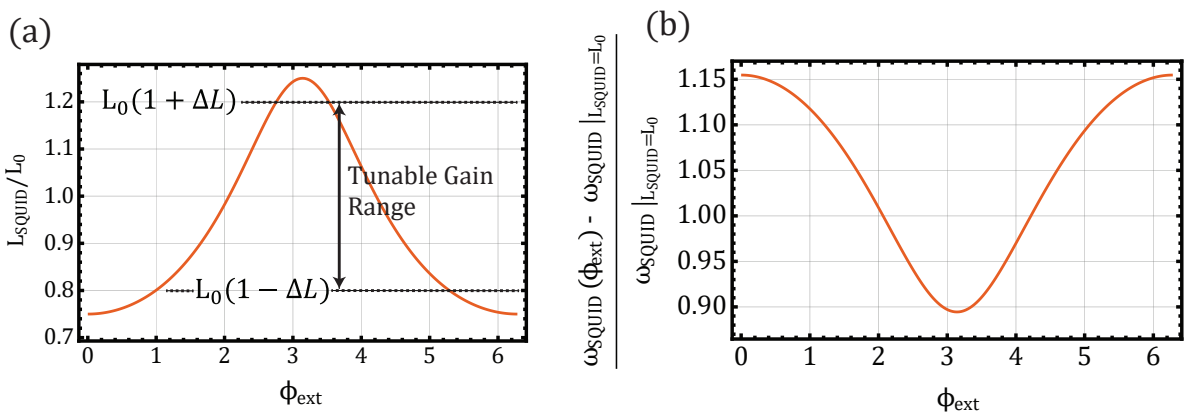


Figure 1.7: Inductance modulation of an RF-SQUID with a shunting ratio (β) of 0.25. (a) Sample inductance modulation of an RF-SQUID, showing that the value of interest here (only those inductances that correspond to gain) is usually somewhat less than the full inductance range. (b) The corresponding tunable resonance frequency range. This range, factoring in both (a) and pump detuning (which usually extends the tunable bandwidth by about the decay rate κ), forms the tunable bandwidth of the amplifier.

The reason that JPAs are tunable, and that they find so much utility (particularly in academic settings where readout resonator frequencies often have a large spread across many experiments) is that the resonant frequency of a JPA is explicitly dependent on a tunable inductor. In a parallel LC resonator, the resonant frequency is given by $\omega_0 = \frac{1}{\sqrt{LC}}$, and if we consider that the amplifier can get gain over a particular inductance range from $L_0(1-\Delta L)$ to $L_0(1+\Delta L)$ symmetrically around a midpoint then the resonant frequency can be tuned by

$$\frac{\omega_{max}^2}{\omega_{min}^2} = \frac{1-\Delta L}{1+\Delta L} \approx 1 - 2\Delta L + 2\Delta L^2. \tag{1.63}$$

Figure 1.7(a) shows the inductance modulation range calculated in this way, where we can see that the inductance modulation that corresponds to gain is slightly smaller than the full range because of the decrease in third-order nonlinearity near the extreme values of the inductance. Figure 1.7(b) shows how this corresponds to a tunning range in the resonant frequency, which when biased for gain creates a family of bias points shown in Fig. 1.6 as light gray curves.

The reason that I chose a symmetric inductance to call L_0 is because for the applications in this thesis, the amplifier needs 3-wave gain, which requires a third-order nonlinearity that is 0 without applying some external flux through a squid loop, and usually maximizes

somewhere near the middle of the flux modulation range (with some leeway for nearly-hysteretic devices and multi-loop devices where the true maximum can be considerably far from the middle of the tunable inductance range). Because of this requirement, a symmetric definition about the middle is more relevant than a definition at minimum inductance and dovetails nicely with the idea of fast ω_p inductance modulation δ_L around a bias point in chapter 4.

Rearranging this, we get an expression for the tunable bandwidth as a percentage of variation around the designed operating frequency $\frac{\mathrm{B}_t}{\omega_0} = \frac{\omega_{max}-\omega_{min}}{\omega_0}$ in terms of the inductance variation ΔL:

$$\frac{\mathrm{B}_t}{\omega_0} = \frac{\sqrt{1+\Delta \mathrm{L}} - \sqrt{1-\Delta \mathrm{L}}}{\sqrt{1-\Delta \mathrm{L}^2}} \approx \Delta \mathrm{L} + \frac{5\Delta \mathrm{L}^3}{8} + O\left(\Delta \mathrm{L}^4\right). \tag{1.64}$$

The tunable gain-capable inductance range ΔL cannot be made larger than one, and is usually comparable to β for rf-SQUID devices. Additionally, increasing the inductance of the resonator changes its impedance, meaning that for large tunable bandwidths the coupling rate to the external environment will increase with the inductance at low frequencies, making the instantaneous bandwidth a strong function of the resonant frequency. Additionally, making $\Delta L \to 1$ requires large nonlinearities in the Hamiltonian, creating an unfavorable tradeoff between tunable bandwidth and saturation power.

1.4.2 Frequency multiplexing

In order for us to evaluate JPAs as preamplifiers for large quantum processors, we have to go over some basics in signal processing. First, I will go over the goals of a multiplexed system, then I will show unavoidable artifacts in time-domain pulses, and lastly, I will discuss the aspects of these problems that JPAs can address with techniques we will show in later chapters.

Qubit readout is fundamentally time-domain. For reading out each single qubit, we need to generate a short pulse at the readout resonator frequency in the range of $5-10$ GHz, send it into the readout resonator, then amplify it back up to the room temperature equipment. We can imagine that the most basic extension of this method is to read one qubit at a time,

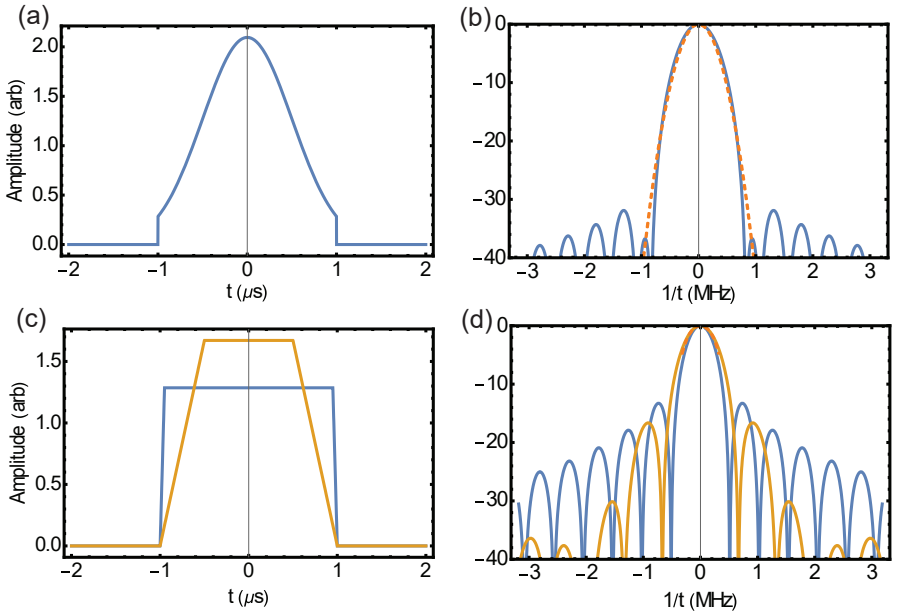


Figure 1.8: Common time-domain pulses and their power spectral densities. (a) Standard cut-off gaussian pulse, here drawn out to 2.5σ on either side. (b) Power spectral density (PSD) of the cut-off gaussian, with untruncated Gaussian approximation shown in orange dotted line. (c) Ramped-edge square pulses with different ramping slope. In blue, a rise time that is one-tenth the pulse duration, a decent approximation of most fast switches on a 100 ns pulse. In orange, a more conservative ramp time set equal to the constant pulse duration. All pulses enclose the same area. (d) PSD of the two ramped-edge square pulses.

switching the bias current and pump frequency of the amplifier to move the gain profile within its tunable bandwidth, stopping at one qubit at a time, reading out, then moving on. This is called *time-domain* multiplexing, and at small scale is certainly doable. Modern direct digital synthesis (DDS) is capable of synthesizing JPA pump frequencies allowing generators to change frequency very quickly (10 ns), and fast (few ns) DC current switching is already used to drive tunable-coupler quantum processors. The biggest problem with this method is how it scales. Common readout times of 100 ns per qubit would scale prohibitively beyond more than 10 qubits or so, because the last qubit would have to wait longer (decay more) before the amplifier could work its way to it, limiting its readout fidelity.

A far more scalable solution is frequency multiplexing, which takes advantage of the fact that multiplying tones together in the time domain adds frequency components to the signal. With this, one could multiply together multiple tones, say at 50, 100, and 150 MHz, upconvert them all up to readout resonator frequency (say at 5.05, 5.1, and 5.15 GHz), let

each tone interact with its own readout resonator, then reverse the process to demodulate each tone individually, acquiring an IQ value for each qubit *simultaneously*. We will primarily address frequency multiplexing here, because it is one of the most promising methods to scale. The expense for this expediency in time is that everything in the signal processing chain must process each of these tones, separated by some frequency spacing (channel width), in the same way. The bandwidth of the signal processing chain, which is always limited by the lowest bandwidth in the chain, sets the budget for deciding the channel width. Since the end goal is to process as much information as possible, we want as many high-fidelity channels as possible.

How do we decide the width of the channels? This is determined mostly by how long (and what shape) the pulse to each individual qubit is. To provide a few reference points, Fig. 1.8 provides a few examples of 2 µs long pulses to demonstrate how the shape affects the crosstalk we would expect in neighboring channels with no nonlinear mixing. The cut-off gaussian pulse in Fig. 1.8(a) is a logical choice when worried about minimizing channel width, however it is less common. Its power spectral density (PSD) in Fig. 1.8(b) shows an initially gaussian profile that gives way to a square pulse-like ripple. Figure 1.8(c) shows more common flat-top pulses, which are more easily extended in time, and Fig. 1.8(d) shows a particular trait of these flat-top readout pulses, although their FWHM PSD is determined solely by the length of the pulse (in this case about 500 KHz for a 2 µs pulse) the higher-order lobes of the pulse are hugely affected by the pulse shape. Without any ramping, the pulse carries a -20 dB component over 3 times the width of the primary pulse.

The level of acceptable crosstalk directly determines the channel width from these pulse spectra. For example, if we deem -10 dB peak crosstalk to be acceptable, then the gaussian pulse initially starts off worse than the others, with a channel width corresponding to significantly wider than the square pulse at 1 MHz vs. 750 KHz. However, if one has to maintain a more strict -30 dB peak crosstalk limit, the Gaussian pulse wins easily, requiring only a 2 MHz channel width compared to over 6 MHz for the square pulse and 2.5 MHz for the ramped-edge pulse. This process of delineating channels within an active bandwidth is shown in Fig. 1.9(a), where we mark channels based on the width of the pulse. Here we use the ramped-edge pulse as an example.

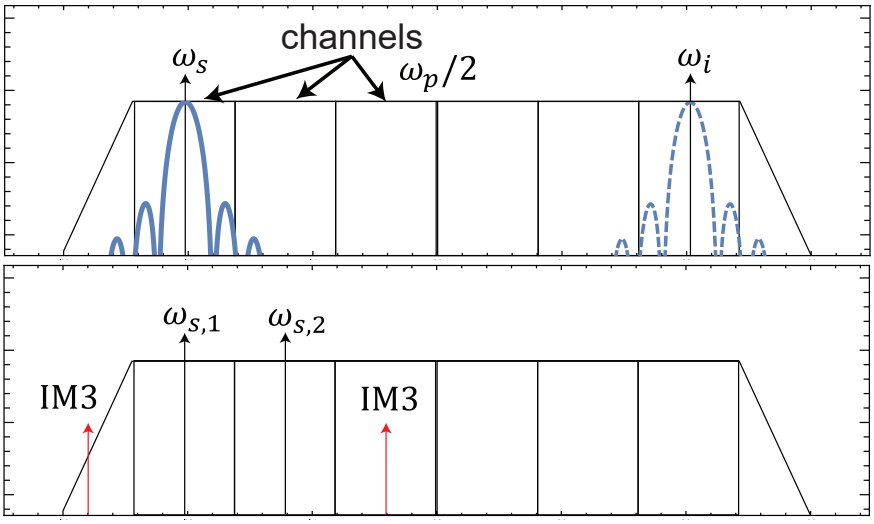


Figure 1.9: Channel widths in a frequency multiplexed readout scheme. (a) The process of deciding channel width from pulse width, here three channels are chosen on one side of the amplifier bandwidth. The idler tone, always generated at mirrored detuning around $\omega_p/2$ (shown in a dotted line), blocks the other three. (b) Intermodulation between two tones generates IM3 tones that cause crosstalk in adjacent channels.

In a degenerate(phase-sensitive) parametric amplifier, there is one extra wrinkle. Since the idler is generated in the same resonator and escapes via the same port, the bandwidth is cut in half immediately, because all of the channels on one side of $\omega_p/2$ are heavily blocked by the idlers generated by the signals on the other side of $\omega_p/2$. This same issue plagues traveling-wave parametric amplifiers. Nondegenerate(phase-preserving) parametric amplifiers can solve this issue by designing the idler network independently at a well-separated frequency. Another option if one wants to spread out the signals in frequency is to alternate signal with idler channels. In other words, dividing the gain profile into channels 1,2,3,4, 5, and 6 in Fig. 1.9 and applying this every-other strategy would mean that we could use channels 1, 3, and 5 to spread the incoming readout resonator pulses out. Although the pulses will have idler clones that run into the same frequency crowding issues up the readout chain, the qubit readout resonators themselves will not see these clones (to within the circulator isolation), and so the crosstalk-induced T_2 reduction will be lower.

Additionally, these spectra are calculated without factoring in the frequency response of other components in the readout chain. For example, the readout cavity linewidth κ will alter the spectrum of a reflected pulse, as it is often somewhere in the 1 to 50 MHz range,

not coincidentally comparable to the channel size. This resonant response will delay rising edges and prolong falling ones with an exponential time constant κ. We will analyze these effects in chapter 3.

The issue of idlers blocking half the bandwidth is a more specific case of a common problem in frequency multiplexing. The more general case (applicable even to modern cell, Bluetooth, and WiFi networks) is shown in Fig. 1.9(b). Here, nonlinearities in the amplifier mix together two adjacent tones into an undesirable red product, labeled "IM3" for intermodulation 3rd order, which we will discuss in detail in the next section. These inconvenient spurious tones generate crosstalk, and share a fundamental link to another important frequency multiplexing requirement: saturation power.

1.4.3 Saturation and Intermodulation

All amplifiers must eventually saturate, because the sources of energy which we have to amplify with are finite. In this section, we will consider the math behind saturation and one of its causes from the last section - intermodulation - with a simple transfer function outside of the context of devices. In later chapters, we will use these results to quantify the nonlinearity of rf-SQUID amplifiers. To start, a conceptual diagram of saturation in a degenerate parametric amplifier is shown in Fig. 1.10. The saturation curve in (a) gives an idea for saturation - simply the reduction of the gain at high signal input strength. The interesting part of saturation specifically in the context of parametric amplifiers is that it is inseparable from the idea of intermodulation. Nonlinearity causes *both* simultaneously. In this section, we will walk through intermodulation and saturation starting with general cases applicable in all nonlinear AC electronics, and gradually work up to more and more specific cases to parametric amplifiers.

The analysis of a nonlinear transfer function in an electrical circuit can usually be expressed as a taylor expansion in terms of the input. In regular electrical circuits, this is done in terms of voltage in Ref. [47]. For highly nonlinear functions (e.g. the sinusoidal device equation of the Josephson junction) there are some difficulties with this as the expansion is unable to capture dynamics that move between the cosine wells in the potential, except

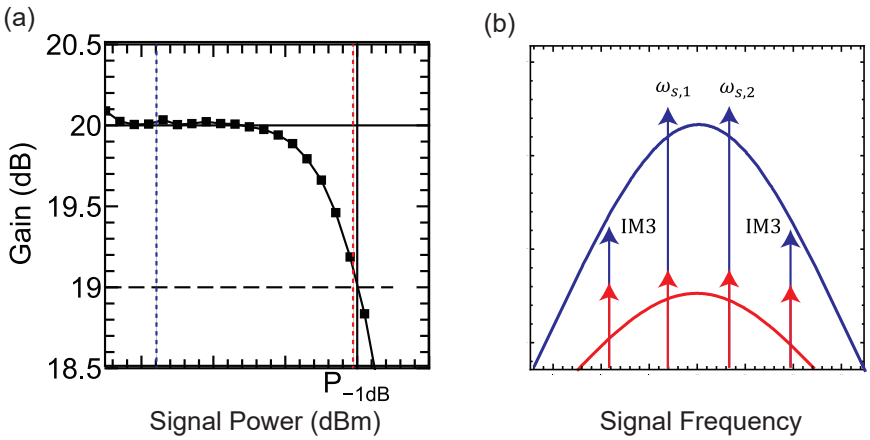


Figure 1.10: (a) Representation of the saturation of a single-tone input to an active degenerate parametric amplifier. As one turns up the input signal power, the gain decreases until a 1 dB drop is reached at input signal power P_{-1dB} reference powers marked with colored dotted lines. (b) Schematic of the behavior of intermodulation products as the amplifier saturates from the dotted reference points in (a). In blue, the behavior far from saturation where the gain is still linear, and the magnitudes of the intermodulation products are small compared to the signal inputs. In red, the behavior near saturation where the gain is compressed and the intermodulation products are considerable ($\approx 10\%$) compared to the signals.

in very high orders of the approximation. See Ref.[31]. Nonetheless, it is useful to see how different orders of nonlinearity manifest in the orders of the Taylor expansion, so we analyze the transfer function of a model amplifier below, with the flux transfer function

$$\phi_{out} = A_1\phi_{in} + A_2\phi_{in}^2 + A_3\phi_{in}^3 + A_4\phi_{in}^4. \tag{1.65}$$

There will be higher-order terms in any real transfer function, but we will find that we can use this simple function effectively to probe different orders of nonlinearity in the experiment. To see the importance of this expansion we consider the input with a near-resonant term, $\phi_{in} = \phi_s \cos((\omega + \delta)t)$. With this input, we know that there will be some mixing to higher order terms (multiples of $\omega + \delta$).

$$\frac{3}{4}A_3\phi_s^3 + A_1\phi_s. \tag{1.66}$$

In Eq. 1.66 we can see that the third-order term will influence the first-order linear gain at high enough signal strength. We can define the "input saturation power" as the point at which the linear gain, A_1, would be modified by 1dB (about 10% in linear units) from its

original value. In the case where A_1 and A_3 have opposite signs, this is the "input-referred 1dB compression point" noted as $\mathrm{IP}_{-1\mathrm{dB}}$. In the case where they have the same sign, an equivalent "input-referred 1dB **expansion** point" ($\mathrm{IP}_{+1\mathrm{dB}}$) may be defined. These input powers are

$$\mathrm{IP}_{-1\mathrm{dB}} \approx \phi^2_{s,-1dB} = -0.145\frac{A_1}{A_3}, \mathrm{IP}_{+1\mathrm{dB}} = \phi^2_{s,+1dB} \approx 0.167\frac{A_1}{A_3}. \tag{1.67}$$

Since the two terms scale differently with the input voltage and the magnitude of higher order terms $(A_2, A_3, ...)$ in Taylor expansions often start lower than the first order term, it makes sense to calculate an intercept between the two different orders within the original terms. This point is the "Third-order intercept" (TOI) or more commonly "intercept point - 3rd order" (IP3). When only the input part of the coordinate is referenced, it is referred to as "input-inferred intercept point - 3rd order" (IIP3), likewise (OIP3) when referred to the output power. Using Eq. 1.66, we see that the intercept is

$$\mathrm{IIP3} = \phi^2_{s,\mathrm{IIP3}} = -\frac{4A_1}{3A_3} \tag{1.68}$$

Interestingly, comparing these two allows for an easy test of how closely third order a system is. If the system is described mainly by fourth order and lower nonlinearities, then the ratio of the IIP3 and $\mathrm{IP}_{1\mathrm{dB}}$ is about 10dB.

Suppose the system is less than fifth-order nonlinear, and one can drive a sufficiently large input power to reach the third-order intercept directly. In that case, we can measure the magnitudes of A_1 and A_3 by the power-dependent behavior of the gain - simply by looking at the output tone at the input tone's frequency and observing the point where the output power reaches zero shown by the black line in Fig. 1.11. The major problem in practical situations with this method is that most devices are *not* nonlinear to less than fourth order, they contain higher order nonlinearities. This will eliminate the zero in the gain at the IP3 point and complicate the measurement.

However, we are not limited to measuring one tone in the system. If, instead, we put in two detuned tones in the system with inter-tone distance 2δ around ω, we can observe the

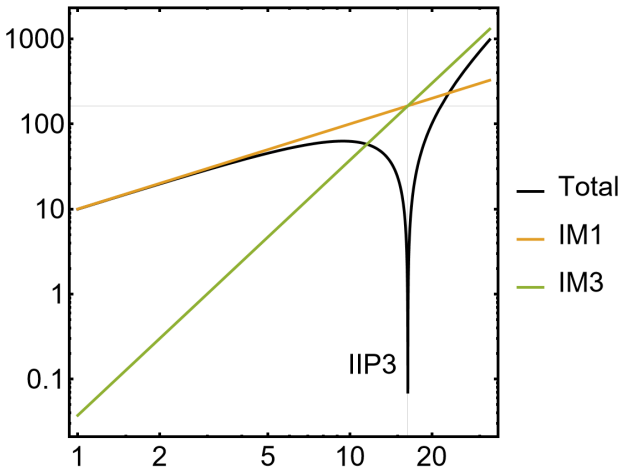


Figure 1.11: IP3 intersection plot. Log-log plot of the absolute value of each term in Eq. 1.66 for $A_1 = 10$ and $A_3 = -0.05$. The IM1, or "primary" tone has slope 1 on this plot, with the third-order term having slope 3. The x-axis value of their intercept is the IIP3 point. The combination of the two tones, shown as a black line, shows a decrease in gain as the tones become comparable, highlighting that the generation of intermodulation tones causes saturation.

power dependence of not only the primary tones, allowing us to observe gain compression in conjunction with the mixing products of the two inputs - the "intermodulation" tones. To see this, we consider a new input that includes both tones and calculate the Fourier components of the tones at $(\omega + \delta), (\omega - \delta), (\omega - 3\delta),$ and $(\omega + 3\delta)$ in Table 1.1.

Tone Order	Tone Frequencies	Tone Amplitude
1	$\omega \pm \delta$	$\frac{9}{4}A_3\phi_s^3 + A_1\phi_s$
3	$\omega \pm 3\delta$	$\frac{3}{4}A_3\phi_s^3$
5	$\omega \pm 5\delta$	0

Table 1.1: Table of one-tone mixing amplitudes for mixing products in a 4th-order nonlinear transfer function. Notably, the fifth-order term is zero. However, the third-order term provides a direct measure of the third-order nonlinearity.

Here we can see the utility of measuring with multiple tones. Firstly, in this transfer function truncated to fourth order, we see that the third-order intermodulation tones' amplitude is directly proportional to the third-order nonlinearity. While the first-order tone still contains the complicating higher-order cubic term, we can still acquire the linear part of the term by extrapolating its low-power dependence, which in the ideal case (low-enough powers used for extrapolation) will produce a line of slope one, similar to the orange line depicted in Fig. 1.11. By comparing the low-power dependence of the primary (order-1) terms

with the secondary intermodulation tones, we can extract the IP3 point without fitting a complicated power dependence of the primary. This method takes the nonlinearities and uses their higher-order nature to move their measurement to a different frequency, where the high sensitivity of a readout chain plus standard spectrum analyzer will easily be able to see them independently of the primary tones. Remarkably, the slopes of the third-order tone amplitude and the low-power dependence of the first-order tone amplitude have the exact same relationship as the original one-tone situation of Eq.1.66, allowing the two-tone measurement to extract the same IP3 as the one-tone measurement.

Tone Order	Tone Frequency	Tone Amplitude
1	$\omega \pm \delta$	$\frac{25}{4}A_5\phi_s^5 + \frac{9}{4}A_3\phi_s^3 + A_1\phi_s$
3	$\omega \pm 3\delta$	$\frac{1}{8}\phi_s^3\left(25A_5\phi_s^2 + 6A_3\right)$
5	$\omega \pm 5\delta$	$\frac{5}{8}A_5\phi_s^5$

Table 1.2: Table of amplitudes for two-tone mixing products in a 5th-order nonlinear transfer function.

In conclusion, the power dependence of two-tone intermodulation product measurements provide a means to directly access the nonlinearity of a given transfer function. In essence, the experimentalist is able to reconstruct the local power dependence of the Taylor series expansion of the operating point simply by looking at the magnitudes of frequency-separated tones. Importantly, this test can be done *no matter the state* of the system. In particular, intermodulation measurmenets are able to be done when the device is actively powered. In Refs. [15, 62], this proved instrumental in showing that the nonlinearity of a device under pump is *not* dominated by the static (or unpumped) nonlinearity that other methods, such as a Duffing test, are able to access. In this section we considered the ideal two-tone measurement in a nonlinear system with native linear gain. In a later chapter, we will investigate how this enables a measurement of device parameters in pumped JPAs.

1.4.4 Pump coupling efficiency

1.4.4.1 One-port case: voltage division

The following section is adapted from an appendix of Ref. [21]. The advantage of utilizing a 3-wave JPA over a 4-wave JPA is the ability to use a single pump at twice the signal frequency, generally reducing the pump leakage into the signal band of the readout chain and making it easier to implement filters to further reduce pump leakage. However, this advantage comes at a cost, which is that the behavior of the circuit near the pumping frequencies has to be carefully considered independently of the behavior at the signal frequencies. For example, it is possible that far away from the signal band it is challenging to get the pump tone to interact with the parametric core of the system that you are driving, significantly raising the pump power requirements of the amplifier and possibly heating the dilution refrigerator it resides in. Therefore, it is important to consider how effectively a certain input pump power drives parametric processes.

We have already simulated η_{PAE} (shown in Eq. 2.5) for the core of the amplifier without the capacitive coupler, shown in Fig. 2.7(b), and so we will define another quantity that captures the difference between the input coupling at the signal frequency, captured effectively by the decay rate γ, and the pump frequency. This difference is not modeled in the periodic steady-state simulation, which uses the same decay rate for both signal and pump. To model the real circuit, which uses a coupling capacitor to alter the real part of the impedance on resonance, we will modify the "core efficiency" shown in Fig. 2.7(b) by another factor which we will calculate as a function of any coupling impedance.

As detailed in Appendix 2.5, the driving term we are concerned with is the circulating phase at the core node of the amplifier, which is the pump frequency component of the phase ϕ_{in} across the rf-SQUID array element. To model the effect of any arbitrary coupling series impedance, we consider the circuit in Fig. 1.12.

For the particular case of a series coupling impedance, the result is easily attainable by voltage division, allowing us to find the power dissipating in the real impedance outside the amplifier, given by $P_a = |V_{a,p}|^2/(2Z_0)$, as a function of the impedances and the required circulating phase

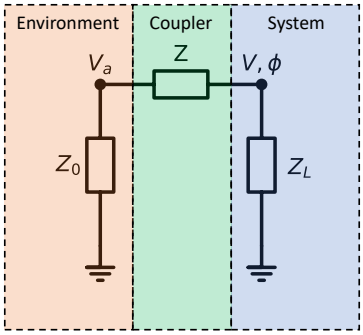


Figure 1.12: Division of a circuit into an environment impedance, coupler impedance, and system impedance for calculating PCE, dissipating in a real impedance Z_0, versus the circulating phase ϕ across the load impedance Z_L as a function of the coupling impedance Z.

$$P_a = \frac{1}{2Z_0} Re[|\frac{\phi_p \Phi_0}{2\pi} i\omega(1 + \frac{Z}{Z_L})|^2],$$

where ϕ_p and V_a are complex amplitudes for $\phi(t) = \phi_p e^{-i\omega t}$, $V_a = V_{a,p} e^{-i\omega t}$ respectively. This can be further simplified by assuming all-reactive loads $Z_L = iX_L$ and couplers $Z = iX$, as is the case for a JPA coupled via a capacitor as long as the pump is not dissipating in the amplifier,

$$P_a = \frac{(|\phi_p|\Phi_0\omega)^2}{2Z_0(2\pi)^2} \left(1 + \frac{X}{X_L}\right)^2.$$

We can further group the pre-factors with $\alpha = \frac{(\Phi_0)^2}{2Z_0(2\pi)^2}$ and rearrange the equality to form an efficiency metric

$$\eta_{PCE} := \frac{\alpha\omega^2|\phi_p|^2}{P_a} = \left(\frac{X_L}{X + X_L}\right)^2.$$

The efficiency that the core utilizes the pump which makes it to the core node is the topic of dedicated research of its own [19], so we define this simpler metric as phase coupling efficiency (PCE). This quantity allows the circuit designer to accurately anticipate the amount of pump power that will make it to the core of the amplifier under some assumptions, and serve as useful energy to the amplification process as opposed to being reflected off of the coupler. The phase coupling efficiency, pump efficiency, and targeted output saturation power combine to determine the total power requirement of the amplifier, the relevant quantity when considering heating problems in a dilution refrigerator.

$$\eta_{total} = \eta_{PCE} \cdot \eta_{PAE} \tag{1.69}$$

There are some limitations to this simple calculation, the chief among them is that there must not be a zero in the denominator, or else the efficiency diverges positively at that particular frequency. At first, this appears to be a good thing. However, a positive divergence necessarily implies that the pump power to induce *any* phase across the load becomes arbitrarily small. This consequence is incompatible with a stiff pump, and so it appears that the PCE can only be evaluated far away from such poles.

This can be calculated for a coupling capacitor C_c with reactance $X_{Cc} = -1/\omega C_c$ and is most useful in the approximation of driving far above resonance of the mode, where the load reactance of a parallel LC resonator is dominated by the capacitor C. This gives a high-frequency limit to the PCE and graphed by the dotted line in Fig. 1.13 of

$$\eta_{PCE} = \left(\frac{1}{1 + C_c/C} \right)^2 .$$

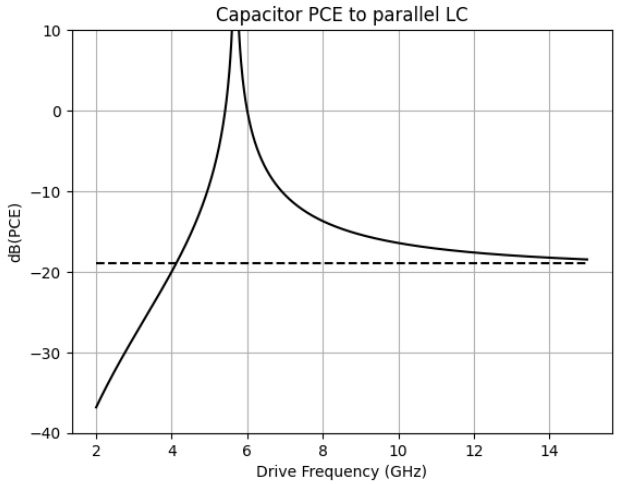


Figure 1.13: Capacitor PCE for a capacitively coupled JPA with Q = 10. The dotted line indicates the high-frequency limit determined by the ratio of shunt and coupling capacitance. Besides the inherent efficiency of the core, the PCE directly modifies how much external pump power is needed to drive gain for a three-wave JPA, increasing power requirements for 20dB gain by a factor of 10-20dB for the large coupling capacitor configuration a $Q \approx 10$ parametric amplifier.

Power-added efficiency is not usually the dedicated focus of JPA's fabricated in academic settings, with the primary focus usually being on quantum efficiency, bandwidth, and saturation power in that order of priority. Nonetheless, we find it interesting to review the

power-added efficiencies of many JPA's in the field. We found the first such review in [61] very helpful for perspective on the topic, and so have edited it here with some additions. While we do not make similar distinctions for total versus core power added efficiencies for all devices, due to the many differences in couplers and three versus four-wave nonlinearities employed, we still find it helpful to gauge our device's pump efficiency in reference to other devices fabricated in the field. With this in mind, we conclude that while the core efficiency is competitive at the $\eta_{PAE} = 1\%$ level, the choice of pump coupler places the η_{total} of our device below the norm.

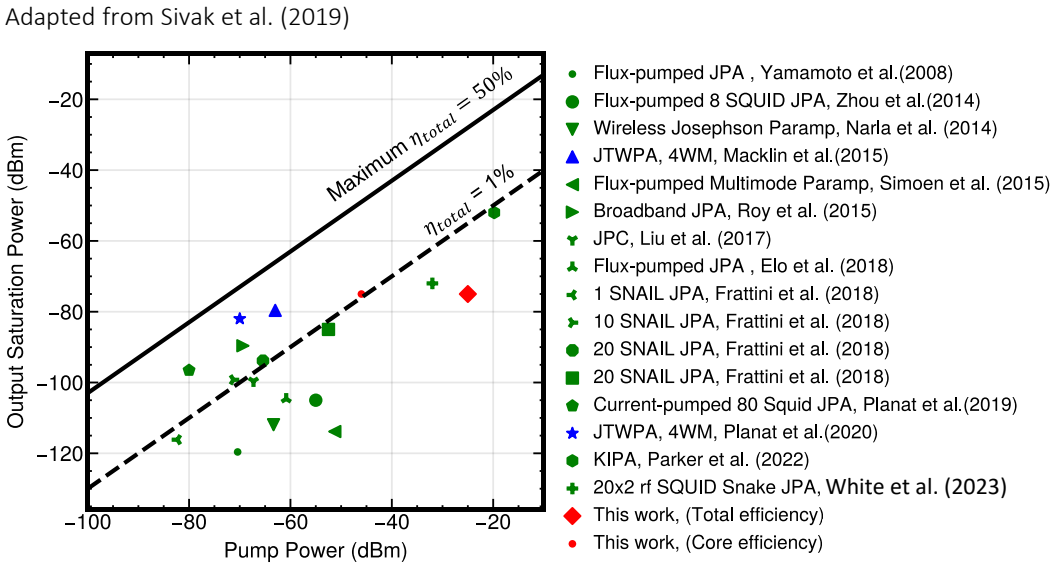


Figure 1.14: A summary of pump efficiencies of JPAs from other publications, reproduced from [62] and expanded. Our devices (red cross) lie well below the trend for other devices as well as our theoretical prediction (red diamond) because of reflections off of the coupling element.

The PCE and the PAE determine the total power requirement of the amplifier, and so future work will include coupler optimization to more effectively bring down the power requirement without creating too soft of a pump. Accounting for PCE will hopefully allow future devices to operate without the use of a hot pump line and diplexer.

1.4.4.2 Two-port case: ABCD parameters

One-port phase coupling efficiency is easy to calculate but it has little utility to solve the most pressing problem for pumping broadband amplifiers: delivery of the pump through a separate port. For a broadband amplifier, we have to deliver the pump through a separate

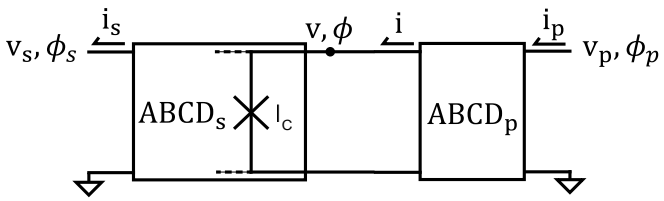


Figure 1.15: The circuit for considering a pump port in the system. Here we wrap the junction (and its inductance) into the signal-side ABCD matrix. This will be more convenient for the synthesis procedure in Ch. 4.

port because the signal port is structured like a bandpass filter around a signal frequency ω. As a simple example, the scattering through a filter falls off at about 20 dB per decade in power, per pole of the filter. This is equivalent to 12 dB per octave, where each octave is defined as a factor of two in the frequency. The pump for a 3-wave JPA is always at twice the center frequency of the gain, so we can estimate that the voltage or current attenuation at the pump frequency for a three-pole filter will be at least 36 dB.

For a high-power JPA, a 36 dB drop in efficiency is untenable. Assuming a standard core efficiency of about -40 dB and a desired output saturation power of -90 dBm, then this raises the power requirement to -10 dBm just outside the amplifier. This is, generously, a scorching amount of power. If the maximum output power of the room-temperature generator supplying the pump is 20 dBm, then this allows only a total of 30 dB of attenuation to the base stage of the refrigerator, which at the pump frequency already matches the line attenuation for stainless-steel SMA cables. In other words, there is no headroom for thermalizing the pump lines. Without thermalization, the superconductors at the chip level will be bombarded with high-temperature (super-gap) radiation. At the very least, this will correspond to an increase in internal loss and a loss of quantum efficiency.

One way to fix this is to introduce a second port that can be engineered independently of the first. Signal-side parametric amplifier matching networks can have many forms, as we'll discuss in Ch.4. Despite the variety, we can calculate the figures of merit from the "transmission matrix" commonly referred to as the ABCD parameters [47] of the system. The ABCD parameters describe the voltages and currents of the network shown in Fig. 1.15 as

$$\begin{bmatrix} v_p \\ i_p \end{bmatrix} = \begin{bmatrix} A_p & B_p \\ C_p & D_p \end{bmatrix} \begin{bmatrix} v \\ i \end{bmatrix}, \text{and } \begin{bmatrix} v \\ i \end{bmatrix} = \begin{bmatrix} A_s & B_s \\ C_s & D_s \end{bmatrix} \begin{bmatrix} v_s \\ i_s \end{bmatrix}. \tag{1.70}$$

Here I have to add a caveat that for historical reasons I do not completely understand, not all ABCD parameters have the same sign convention for the output current. Some simulation software takes the negative of the output currents convention mentioned here (e.g. AWR Microwave Office). It's important to note that the sign change of the current will take $B \to -B$ and $D \to -D$. Here though, I take the convention in Ref. [47], which allows for easy concatenation of matrices, uses no negative sign, and reflects the current directions in Fig. 1.15.

Each of these matrix definitions gives a system of equations that we would like to solve for v/v_p at a particular frequency, which is the equivalent of the voltage division case discussed in the previous section. This follows by eliminating the current variable between the two sections, and the result is

$$\frac{v}{v_p} = \frac{A_s}{A_p A_s - B_p C_s}. \tag{1.71}$$

This is already enough to calculate the efficiency, but to break this down a little more we can assume that (i) the signal side of the network can be reduced to a single complex admittance Y_s, and (ii) that the pump port can be reduced to a single series impedance with both these equivalent immittances at the *pump* frequency. With this translation, the efficiency becomes

$$\frac{v}{v_p} = \frac{1}{1 - Z_p \left(Y_r + Y_s\right)}. \tag{1.72}$$

This form of the equation tells us that to design a pump port, we could consider placing zeros in the impedance of the pump network near twice the pump frequency to guarantee effective pump coupling. Poles in the pump impedance, on the other hand, pose the same issue with divergence that we saw before in the single-port case and should be avoided. However, this may not guarantee low pump leakage. To see why, consider the ratio of the voltage at the pump vs. at the signal terminal. Again, the signal-side ABCD matrix includes

the 50 Ω termination here, so this voltage directly indicates the power dissipating at the port via v_s/Z_0^2.

$$\frac{v_s}{v_p} = A_p A_s + B_p C_s = 1 + Z_p Y_s. \tag{1.73}$$

This tells us why simply putting zeroes in the pump impedance doesn't guarantee low pump leakage. In fact the opposite, because it also necessarily implies that the pump voltage leaks directly through to the signal port unless there are also zeros in the signal-side *admittance* to compete.

2.0 High dynamic range RF-SQUID amplifiers

2.1 Introduction

Superconducting quantum computers are rapidly scaling to thousands of physical qubits [1]; more, high-fidelity qubit readout is a requirement for both Noisy Intermediate Scale (NISQ) [48] and error-corrected operation of these machines. To avoid further inflation of machine size via the use of ancilla qubits for error correction and other algorithms involving mid-circuit measurements, we require the readout to be Quantum Non-Demolition (QND), which leaves the qubit in the state indicated by the measurement outcome [43]. Moreover, dispersive superconducting qubit readout is limited by as yet imperfectly understood non-QND effects for strong readout drives [56, 59, 25, 11], and so one cannot increase the power of the measurement indefinitely to increase measurement fidelity. Further, due to the qubit's finite T_1, one must measure for a time much smaller than T_1 in order to attain high fidelity. Finally, we need a faithfully linear amplifier to amplify the short, few-photon quantum signal to be unaffected by the added noise of standard 4K commercial amplifiers in the upper stage of the readout chain while adding the minimum amount of noise allowed by quantum mechanics [7, 10].

Generally, we require these high-fidelity, QND measurements for each element of a quantum computer. Frequency multiplexing is a natural solution, allowing many qubit readout resonators to be measured from a single feed line as long as their resonant frequencies are sufficiently spaced. However, this spacing has to be larger than a few linewidths of the resonators to avoid crosstalk, necessitating a large bandwidth [39, 54, 69, 23] of the first amplifier in the chain. The presence of many signals at the same time also requires that the first amplifier be highly linear, both avoiding compression and excess intermodulation distortion [69, 23].

The Josephson Parametric Amplifier (JPA) is one type of first-stage amplifier that shows promise for its ease of manufacture. The JPA's megahertz-scale bandwidth typically falls well short of Josephson Travelling Wave Parametric Amplifiers (JTWPAs). However, the

difficulty of manufacturing the thousands of junctions required for JTWPAs, their intrinsic loss, and noise-rise near saturation [50], motivates the study of how we can improve the JPA to create a small, easy-to-fabricate, quantum-limited amplifier. The primary shortcoming of simple JPAs with a single junction or loop is that they have very limited saturation power, typically less than -110 dBm [71, 39, 73, 54, 15]. Depending on the choice of qubit parameters, this power may not be enough to amplify more than one or two qubits without distortion from saturation effects within the amplifier [69]. Diluting the nonlinearity of single junction/loop devices using a chain of DC SQUIDs, rf SQUIDs, and SNAILs has been explored both theoretically and experimentally, yielding substantially higher saturation power [26, 15, 61, 69, 23].

In this work, we design, fabricate, and characterize an extremely quantum-efficient and high-saturation power amplifier using a simple, single-mode device. The key component of our design process is numerical simulations of the un-truncated device Hamiltonian to fully understand and optimize the device within the constraints of fabrication. The key tuning parameters that we optimized are the device quality factor/impedance and the rf SQUID shunting ratio. The resulting simple design consists of merely two capacitors and an rf SQUID array using 25 junctions. We experimentally realized this device in a modified NIST tri-layer process. We demonstrate input referred saturation powers averaging -94.2 dBm ($\pm$ 1.4 dBm) with 20 dB of power gain with extremes approaching -91 dBm input saturation power and 28 MHz instantaneous bandwidth using only 25 5.5 µA junctions. Additionally, we show that our amplifiers have high quantum efficiency (62% referred to the plane of the qubit), allowing fidelity measurements that are only limited by the qubit's ability to tolerate photons in the readout resonator, not the amplifier's saturation power.

2.2 Amplifier Circuit Design and Fabrication

In this section, we present the device simulation and design within the bounds of fabrication. We begin by discussing the shortcomings of previous simulation efforts that truncate the nonlinearity in the device Hamiltonian. Following this, we discuss why we choose rf

SQUIDs as our nonlinear elements. We also elaborate on our simulation method, which does not require truncation, and optimize the design parameters within our circuit. Lastly, we translate these optimization results into circuit elements with an added impedance transformation to control device quality factor.

The problem with single-junction or single-loop JPA's is that even if their bandwidth is broadened to amplify the readout signals of multiple resonators, they struggle to reach the input saturation powers required to simultaneously amplify multiple readout signals at the 20 dB gain required to overwhelm the added noise at the next stage of amplification at 4K [39, 54]. Measurements of this effect are shown directly in [69] in readout on a superconducting processor. This problem of JPAs lacking power handling can be traced to the current through and phase difference across the Josephson junction at large input signals, which other works have focused on in some other similar amplifier designs [26, 15, 61].

These past works depend crucially on expansions of the junction nonlinearity around a particular bias point to find an analytical expression to guide design. While helpful to guide the designer's intuition at small signal strengths, truncations to the junction Hamiltonian have been shown to obfuscate the high signal power behavior of Josephson junction-based devices [31, 15, 61], sometimes requiring extensions up to seven or eight orders to capture the nonlinear behavior of the junctions near amplifier saturation [31]. Because one of our chief goals is to increase amplifier saturation power, we instead simulate the device at large signal powers by finding a periodic steady-state solution of the device's classical equation of motion in the time domain. This method is not without its drawbacks. Notably, it is more computationally intensive than methods such as harmonic balance and does not give any analytical understanding of the design. However, crucially, time domain simulations also do not require any simplification or truncation of the sinusoidal nonlinearity of the junction, and so are a far more accurate guide for the design of high-saturation power amplifiers.

To guide our design, we start with the same strategy as other work with DC SQUIDs [26] and Superconducting Nonlinear Asymmetric Inductive eLements (SNAILs) [16, 15]. If the oscillating phase of the signal increases to nearly 2π, higher-order nonlinearities in the sinusoidal potential of the junction begin to become comparable to the third-order nonlinearity required for gain and disrupt amplification. This has been shown previously in similar stud-

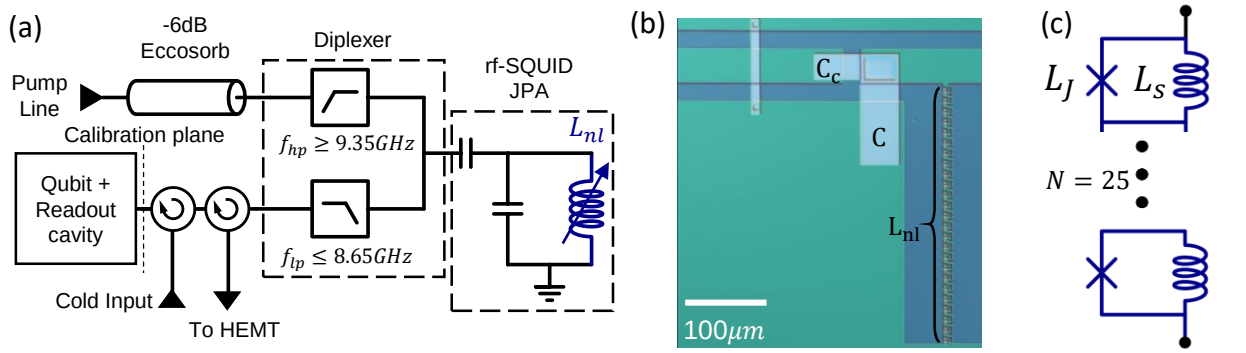


Figure 2.1: (a) The configuration of the measurement circuit in the mixing chamber of a dilution refrigerator. A full diagram is given in Appendix 2.6. The input power and quantum efficiency of the device are both calibrated from the plane shown at the readout resonator. (b) Device micrograph, showing the coupling capacitor C_c and the main capacitor C that are used to control the quality factor and resonant frequency, respectively. The 25 rf-SQUID long array is shown at right, with approximately a 250 µm span to the ground plane (c) Schematic of the nonlinear inductive element (labeled L_{nl} in (a)) of the amplifier.

ies of the same type of behavior in Josephson ring modulators [33, 31] as well as Josephson bifurcation amplifiers [26]. It has also been shown with fourth-order Kerr nonlinearity in SNAIL parametric amplifiers [16, 15, 61, 62]. All of these devices utilize a nonlinear shunt across the junction, formed by Josephson junctions of equal or higher critical currents. For high-inductance versions of these loops, an array of JJs is desirable for producing rather linear arrays with small geometric size. However, for large critical currents/small inductances, such as we use in amplifiers, it becomes quite feasible and desirable to use the geometric inductance of a short superconducting lead as the shunt, which has the additional advantage of eliminating concerns about array modes of the shunt itself [37, 15].

This device we refer to as an rf SQUID, though we acknowledge that is method of operation and biasing is rather different than the historical devices given this name [8]. We then carefully consider how best to optimize the circuit to maximize saturation power. To complement the dilution of the current through the junction by the linear inductive shunt, the phase is distributed by even division over an N-element series array similar to past work in Refs [26] and [15]. The full circuit with all external circuit elements is shown in Fig. 2.1(a). The result is similar to designs proposed in [40] but simpler, with only one array instead of two in parallel, albeit without an easy means of delivering flux on-chip over the entire array.

To simulate the device, we begin with the circuit diagram in Fig. 2.5(a), with a single

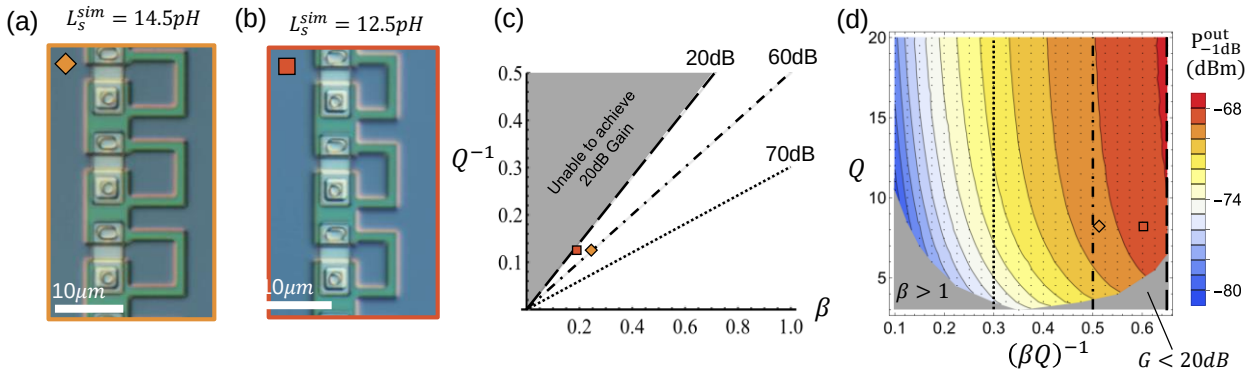


Figure 2.2: (a) Micrograph of three out of the 25 rf-SQUID loops in device A. (b) Micrograph of three out of the 25 rf-SQUID loops in device B. (c) The design space of a single-mode degenerate JPA as a function of shunt inductance (for a given junction inductance) and quality factor Q of the JPA mode in steady state at ω when pumped at 2ω. The junction inductance $L_J \approx 60$ pH was chosen for convenience given by a preexisting reliable fabrication process. When combined with the choice of a reasonably low quality factor Q, this determines the choice of β, shown altogether as an orange diamond in the design space. An additional device B with a lower simulated shunt inductance is shown as a red square. (d) The predicted output P_{1dB} of devices in a subsection of the design space. The dependence on the slopes in (c) is emphasized by dotted/dashed lines, which in this coordinate space are vertical.

nonlinear mode formed from an array of N rf SQUIDs of shunt inductance L_s in parallel with a Josephson junction with inductance L_J resonated by a capacitor C, and coupled to the external environment of characteristic impedance Z_0 around resonance. Assuming that the external magnetic fluxes through the rf SQUID,noted as ϕ_{ext}, are the same, the equation of motion (EOM) can be expressed as

$$\partial_t^2\phi(t) + \gamma\partial_t\phi(t) + \omega_L^2\phi(t) + \omega_J^2 \sin\left[\frac{\phi(t)+\phi_{\text{ext}}}{N}\right] = 2\gamma\partial_t\phi_{in}(t), \tag{2.1}$$

where the phase variable is defined by $\phi(t) = \int V(t)dt$ (see Fig. 2.5a), $\gamma = (CZ_0)^{-1}$ is the external decay rate, $\omega_L^2 = 1/(CL_s \cdot N)$ and $\omega_J^2 = 1/(CL_J)$ are two frequency constants.

Using this simulation method, we find that controlling the dimensionless ratio of shunt to Josephson inductance L_J, $\beta = L_s/L_J$ of each rf SQUID in the array and the quality factor Q of the resonator is critical to maximizing the saturation power of the rf SQUID while maintaining its ability to be pumped to produce gain. We find that the device should have approximately the same maximum achievable gain for constant $(\beta Q)^{-1}$, shown in Fig. 2.2(c)

by different dashed/dash-dot/dotted lines through the design space. This is qualitatively similar to discussions about the pQ product in SNAIL and JPC parametric amplifiers, with Q having the same meaning as here, and p being a ratio of SNAIL inductance to the total inductance of the mode. Much like in our rf SQUID case, there a higher pQ, analogous to a lower $(\beta Q)^{-1}$, is "safer" in the sense that it ensures the amplifiers will achieve 20dB of gain [4, 15].

From our numerical calculations, we find that the design parameters Q and β roughly outline a region where the amplifier will fail to achieve 20 dB gain if the rf SQUID is shunted with too small an inductor or the quality factor is too low, as shown by the dashed line determined by a slope $(\beta Q)^{-1} \approx 0.65$. For every point in the design space, we choose a flux bias point that maximizes the third-order nonlinearity in the rf-SQUID. However, we also find that variations in the saturation power with respect to flux bias are small in theory (see Fig. 2.6(b)). Our principal result is that we also find that the maximum saturation powers of the device occur on the boundary outlined by the same dashed line through the design space in Fig. 2.2 (c) and (d), and can increase the saturation power in a 25-element long array by over 10 dB compared to the lowest values. A higher inductive shunt produces a more nonlinear device with lower saturation power for a given Q, but more tunability in its inductance with respect to flux bias. For mode impedances that are controllably low (such that stray parallel capacitance and series inductances are small compared to the LC resonator that makes up the mode) at 5 GHz, this sensitivity corresponds to requiring single micron-scale control over the exact dimensions of the inductive shunt of each rf SQUID in the array. Fig. 2.2(d) shows the design space of the amplifier assuming an $N = 25$ rf-SQUID array in an embedding resonator with $\omega_0/2\pi \approx 6$ GHz.

Based on this understanding, we next scale the circuit by arraying N rf SQUIDs in series. A larger array lowers the nonlinearity by diluting the phase across each rf SQUID, raising the amplifier's saturation power, but also making it harder to pump. Scaling the array by holding total array inductance constant and decreasing both the individual junction and shunt inductances as the array number increases would increase the saturation flux linearly. This yields quadratic saturation power scaling if one could maintain a constant mode impedance by scaling the junction critical current up and the shunt inductance down

indefinitely. However, scaling the junction size up decreases its inductance, and since the designer has to maintain the same shunting ratio, the shunt inductance becomes difficult to control with optical lithography below about 10 pH if the junction size is a few microns. Additionally, the junction side of the rf-SQUID has a stray inductance of a few pH, which can become comparable to the shunt as the scaling is increased. In combination, these effects put a lower bound on the achievable shunt inductance, and therefore an upper bound on the critical current of the junction. With this restriction in mind, we choose to use 25 rf SQUIDs that only require a comfortably low inductance $12 - 15$ pH shunt inductor and a 5.5 µA junction as the basic ingredient for our array.

Our fabricated devices, shown in Fig. 2.2(a) and (b), are fabricated in a modified niobium trilayer process at the National Institute for Standards and Technology's Boulder Microfabrication Facility (layer specifications given in Appendix 2.6.1), where the high yield and consistency of the junctions allows for excellent device yield in even very long junction arrays [44]. Even so, spatial nonuniformity in the flux biasing in the array (discussed in Appendix 2.6.4 and 2.6.2) presents an issue with verifying device parameters in the experiment. Accordingly, we discuss only the nominal designed values here, with the understanding that these parameters can reasonably vary by 10 percent or so in the real device. We use a shunting ratio of $\beta = 0.25$ for device A, 0.21 for device B, and a junction inductance of 60 pH.

In total, the rf SQUID array forms about 290 pH of total effective inductance at zero DC-bias, and between 340 to 390 pH at the flux bias points which we expect to achieve 20 dB gain. Provided this inductance is paired with a resonant capacitor that produces $\omega_0/2\pi \approx 6$ GHz we arrive at a capacitor value of 2 pF. If this mode were directly connected to a 50Ω transmission line, we would expect a quality factor $Q \approx 4$. However, as mentioned above, we have to alter the quality factor due to the increased inductance of the array compared to the single rf-SQUID. To increase the quality factor well into the region where we expect to achieve 20 dB of gain (see Fig. 2.2(d)), we add a coupling capacitor $C_c \approx 0.26$ pF to increase the effective environmental impedance on resonance to $\approx 100\Omega$ as seen by the JPA resonator, and increase the quality factor to approximately $Q = 10$, with some added variance over bias flux because of the coupling capacitor's frequency-dependent impedance.

The effect of this variation with respect to flux and its consequences for the tunability of the amplifier's gain is discussed in section 2.6.3.

We measure the packaged JPA chip using a typical qubit readout amplification chain, with a few additions, such as eccosorb filters, a diplexer, and a modified pump line, that are detailed in the beginning of chapter 3.

2.3 Results

Cooled to 15 mK in a dilution refrigerator, and powered with a strong pump at 2ω the 25-long rf-SQUID array shown in Fig. 2.2(a) and in the larger device in Fig. 2.1(b), is capable of generating a three-wave gain of more than 20 dB over a range of bias points, shown in aggregate in Fig. 2.3(a) with more detail in Fig. 2.11(b). To be sure we account for any loss introduced by the amplifier, we take the maximum return loss over all bias fluxes at each frequency as the baseline for gain. After using this pessimistic estimate to tune to 20 dB of gain at each bias point, we measure input signal compression power up to a maximum of -91.5 dBm, which is among the highest values reported in the literature for a Josephson junction-based resonant parametric amplifier. This value is in good company with the extreme values of [15, 61, 62], and the average values in other rf-SQUID based devices [69, 23]. Notably, devices of this type achieve similar saturation powers even to JTWPAs that contain almost one hundred times more junctions [34].

Power is calibrated at 5.7 GHz by measuring resonator photon number-induced shifts of the qubit's transition frequency ($\omega_{01}/2\pi$) combined with separate measurements of the qubit-cavity dispersive shift, shown in Appendix 2.6.5 along the lines of [57].

Average input saturation power values are (-94.2 ± 1.4) dBm, with an average instantaneous bandwidth of (20 ± 6) MHz. As shown in Fig. 2.3(b) with more detail in Fig. 2.11, the gain varies somewhat in profile with respect to pump detuning at any given bias point, which we attribute to ripples in the line impedance. Further, because the pessimistic baseline we have chosen for gain is not necessarily flat with respect to frequency and is combined from all bias points, some of the gain profiles are asymmetric with respect to $\omega_p/2$. If we take a less

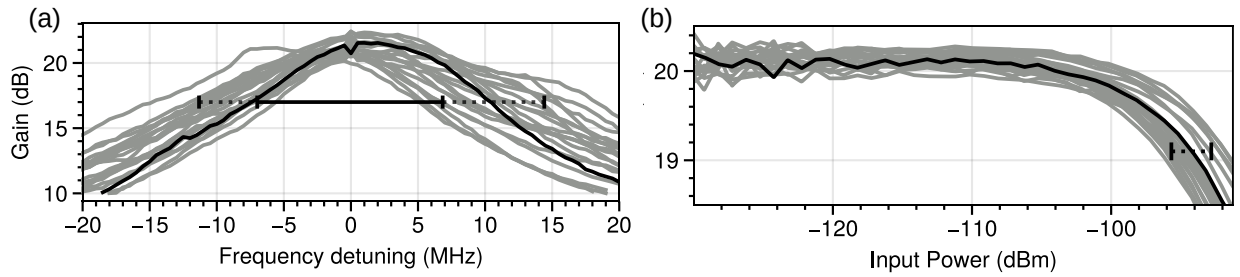


Figure 2.3: (a) Gain profiles with varying pump-detuning, with bandwidth deviations shown. (b) Saturation power of the $N = 25$, 14.5 pH shunted rf-SQUID array amplifier at multiple bias points. In both plots, the black lines are particular bias points that have similar values to the mean. The means and standard deviations are calculated from the full dataset shown in Fig. 2.11.

pessimistic baseline, only normalizing to one flux bias point at a time, they appear symmetric as expected. Some variation in the shape of the loss is expected with the device's low quality factor because the device responds to changes in the external environment over its linewidth of approximately $300 - 500$ MHz, with variations over flux bias. This significantly impacts the pumped operation of the device, because the characteristic ripples in the external environment occur on the same 100 MHz scale. Consequently, the gain profile should take on a modified profile depending on the local slope of the external environment at that particular signal frequency. In addition to the shapes shown, it is possible to see double-peaked gain profiles similar to [54]. We fall short of the expected $\sqrt{GB}$ gain-bandwidth product for many bias points partly because of this issue of variable port impedance. There is also significant uncertainty in the fit for the device linewidth due to the same impedance ripples, so it is possible the device linewidth may be lower than designed.

The input-referred saturation power of the amplifier depends slightly on bias settings and certain parameters that vary on the 1 dB level in the external environment, such as the loss between the resonator cavity and the amplifier chip. Within these variations, the data shows reasonable agreement with the theoretically predicted value from periodic-steady-state analysis of the device equation of motion, shown in Fig. 2.6. The high saturation power of the rf-SQUID JPA that we demonstrate is critical for enabling many-qubit readout experiments. However, despite the diplexer's rejection, the pump power (approximately -25 to -30dBm, based on room temperature loss calibrations) required and the lack of attenuation on the pump line might raise concerns over the thermal photon occupation in the readout resonator

and the corresponding limits imposed on T_2 from thermal photons in the readout resonator dephasing the qubit.

Amplifier State	T_1 (μs)	$T_{2,R}$ (μs)	$T_{2,E}$ (μs)
Replaced with short	81	8	20
Off	90	13	26
Detuned -500MHz	93	11	18
On (25dB Gain)	91	6.4	9.1

Table 2.1: Qubit coherence versus different states of the amplifier. The run in which the amplifier was disconnected was measured separately from the other three. The T_2 coherence differences with respect to the amplifier's state suggest that the reverse isolation between the amplifier and the qubit may not be sufficient.

To address this concern, we measure qubit coherence with varying pump conditions in Table 2.1. We speculate that the presence of the high-power pump line is not meaningfully contributing to decoherence at the 8 µs level, otherwise we would have expected that the qubit coherence would have improved in the absence of the amplifier and diplexer combination in row 1. It must be noted, however, that this measurement was taken on a different cooldown than the others, and qubits can have significant variation in coherence from cooldown to cooldown. Avoiding dephasing is crucial to the design of high-saturation JPAs, allowing the pump inefficiency to be tolerable in terms of qubit coherence albeit inconvenient for fridge design, requiring specially engineered lines with lower attenuation in order to deliver high power pumps without excess heating. However, we observe the T_{2R} time of the qubit dip distinctly when the amplifier is turned on, indicating that the amplifier introduces an extra source of dephasing onto the qubit from the gain alone. This effect is often observed in high-efficiency measurement setups [30, 29], and can be the result of insufficient isolation between the JPA and the readout resonator. To combat this, one can use additional circulators for isolation at the cost of a modest loss in quantum efficiency [23].

In order to calibrate the efficiency of the amplification chain from the plane of the qubit, we can measure the back-action that the amplifier introduces on the qubit. We turn the amplifier on at 25 dB gain, prepare the qubit in the $|+X\rangle$ state, and apply two measurements in sequence according to the pulse diagram in Fig. 2.4(a) [18].

The back-action of the weak measurement results in a sinusoidal pattern in the $\langle X \rangle_c$

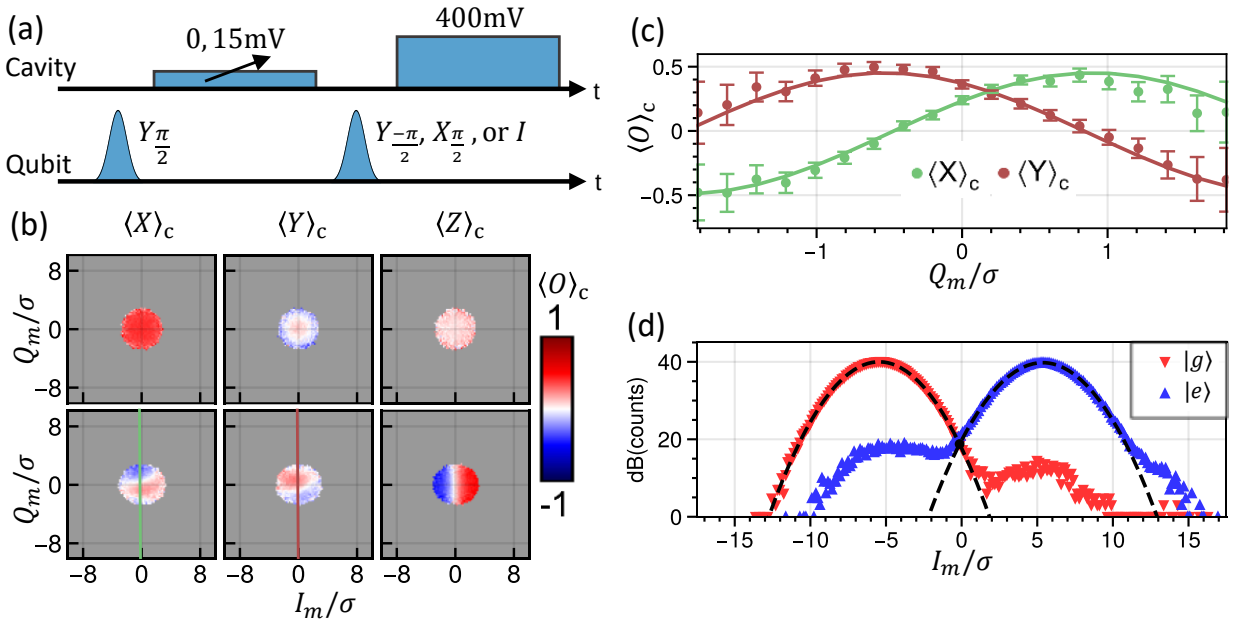


Figure 2.4: (a) Tomography of variable-strength measurement of a transmon qubit prepared in the $|+X\rangle$ state, with the amplifier biased to 25 dB of phase-preserving gain. (b) Conditional measurement back-action in the X and Y axes with respect to I and Q, increasing in frequency with measurement strength, is a measure of efficiency, showing sinusoidal oscillation in the perpendicular axis to measurement in the X and Y back-action of the qubit. As measurement strength increases, the contrast decreases as the qubit state collapses to the eigenstates of the measurement more quickly, but the oscillation frequency increases. (c) A line cut at $I_m = 0$ is fit to yield $(62.4 \pm 1.4)\%$ quantum efficiency, normalized to 1. The data shown are the averages from a Gaussian distribution of samples that varies over Q_m, with the variation in samples reflected in the error bars of the fit. (d) Gaussian fitting to an integrated 800 ns readout pulse to determine 99.3% separation fidelity, integrated from an IQ histogram containing over 10^6 counts. Red hue indicates counts in the qubit state $|g\rangle$, while blue indicates counts in $|e\rangle$.

and $\langle Y\rangle_c$ expectation values in Fig. 2.4(b), with linecuts in Fig. 2.4(c) with respect to the Q_m result of the weak measurement. This sinusoidal pattern has frequency set by the measurement strength $\overline{I_m}/\sigma$, but the amplitude is diminished by an exponential dependent on the measurement strength scaled by the quantum efficiency [18]. Because measurement strength is independently set by the distance between coherent pointer states in the IQ plane, we can fit these oscillations to extract the quantum efficiency. Because the IQ values are dependent on the entire readout chain, the quantum efficiency η that results from the fit represents the efficiency of the entire readout chain, including losses and added noise from the higher-stage amplifiers at 4 K and room temperature. Therefore, it acts as a safe lower bound for the efficiency of the rf-SQUID JPA itself.

With this method, we find a measurement efficiency of (0.62 ± 0.014) on a scale of 0 to 1, without compensating for state preparation and measurement error. This is among the highest reported quantum efficiency values for phase-preserving amplifiers [68, 66, 69, 23], however still not as efficient as phase-sensitive nonreciprocal amplification [29], serving as motivation to pursue integrating amplification closer to the readout resonator itself. The high quantum efficiency of our device also serves as evidence that the high-power nature of the rf-SQUID JPA does not detract from its efficiency.

The projective state readout fidelity of the qubit suffered from a below optimal $2\chi/\kappa = 0.348$, where $2\chi = 0.348$ MHz is the change in readout resonator frequency between $|g\rangle$ and $|e\rangle$ states of the qubit and $\kappa = 1$ MHz is the readout resonator linewidth. So, despite high readout efficiencies, the T_1 decay of the qubit limits the phase-preserving readout fidelity to 99.3% with a 0.8 µs pulse length, as shown in Fig. 2.4(d). As a result, despite the high saturation power of the amplifier, we were unable to increase the readout power and shorten the pulse time due to readout-induced state transitions in the qubit. For examples of high-power readout histograms, see Fig. 2.12.

Higher state transitions in qubit-cavity systems at high cavity occupation is an oft-observed phenomenon [56, 25, 59, 11]. However, it is not entirely without counter. One can design linear filters that take into account quantum noise and T_1 decay [17, 24]. Even with such filters the behavior of transmon qubits at very high readout resonator occupations is generally detrimental to readout performance, and remains a barrier to fast, high-fidelity, QND measurement. Future work to improve the state of the art, including the study of transmon dynamics at high power, will hopefully be enabled by a robust measurement chain that can easily tolerate the measurement powers that drive transmons into behavior resembling chaos [11].

2.4 Conclusion

To conclude, we have experimentally demonstrated the effectiveness of periodic steady-state simulations for capturing JPA dynamics. Using these tools we have successfully de-

signed, built, and fabricated a high saturation power JPA built on rf-SQUID arrays with enough power handling capability for hundreds of channels of simultaneous measurements, albeit without the bandwidth to demonstrate this capability. Additionally, we have shown that despite the device's need for a high-power pump, it does not reduce qubit coherence beyond what is expected for a single stage of isolation, which can be further circumvented by pulsing the amplifier. In addition, the device maintains quantum efficiency on par with some of the best examples of phase-preserving amplification in the literature.

Future work includes engineering a robust matching network to broaden the gain response of this amplifier with techniques similar to [41, 23]. However, we note that while these techniques outline an elegant solution that traces roots back to filter synthesis, they must also be unified with the saturation power calculations demonstrated in this work which account for the full nonlinearity of the Josephson junction to further assist the designer in anticipating saturation effects. More, we should implement a dedicated pump coupling port that can avoid the inefficiency of the coupling capacitor as a means of delivering the AC driving phase to the junction at the pump frequency. This could effectively integrate the external diplexer on-chip, which can be feasibly implemented with the existing fabrication process, which will greatly simplify the external circuitry required to operate the amplifier and conceivably reduce the loss between the amplifier and qubit to further increase the efficiency of measurements. Additional research is also being done on how to increase the amplifier's power-added efficiency once the pump reaches the resonator node [19].

In all, we demonstrate a simple and highly linear amplifier that can serve as a stepping stone for these innovations and as a workhorse in near-term quantum information experiments, including the study of the QNDnesss of strong qubit readout.

2.5 Periodic steady-state simulation

We analyze the dynamics of the rf-SQUID amplifier using periodic steady-state simulation similar to the methods used to analyze Josephson Ring Modulators in [31], starting Eq. (2.1) in the main text. Note that the phase variable $\phi(t)$ has DC components, noted as

$\phi_{\rm DC}$, and hence $\phi(t) = \varphi(t) + \phi_{\rm DC}$. The DC component can be determined by the current relation in each rf-SQUID,

$$\frac{\phi_{\rm DC}}{L_s \cdot N} + \frac{1}{L_J} \sin\left(\frac{\Delta\phi}{N}\right) = 0, \tag{2.2}$$

where we define $\Delta\phi = \phi_{\rm DC} + \phi_{\rm ext}$.

By separating the DC and AC components, we rewrite the EOM in Eq. (2.1) as

$$\partial_t^2\varphi(t) + \gamma\partial_t^2\varphi(t) + \omega_0^2\varphi(t) + \omega_J^2\left[\sin\left(\frac{\varphi(t)+\Delta\phi}{N}\right) - \sin\left(\frac{\Delta\phi}{N}\right) - \frac{\varphi}{N}\cos\left(\frac{\Delta\phi}{N}\right)\phi(t)\right] = 2\gamma\partial_t\phi_{in}(t), \tag{2.3}$$

where the mode resonant frequency is given by

$$\omega_0^2 = \omega_L^2 + \frac{\omega_J^2}{N}\cos\left(\frac{\Delta\phi}{N}\right) = \frac{1}{NC}\left[\frac{1}{L_s} + \frac{1}{L_J}\cos\left(\frac{\Delta\phi}{N}\right)\right]. \tag{2.4}$$

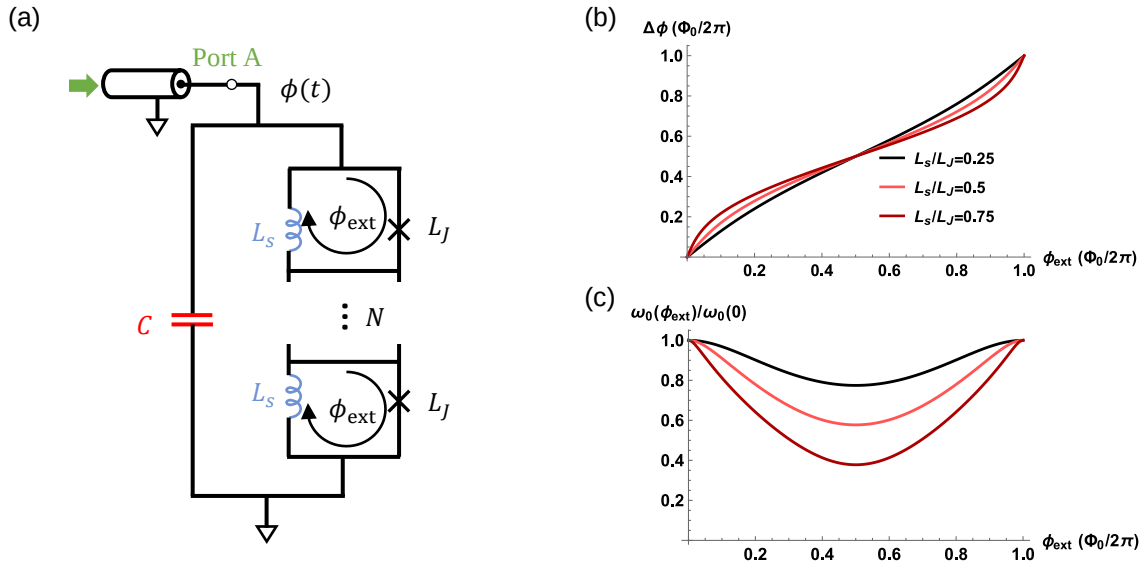


Figure 2.5: (a) The design with labeled node phases and circuit parameters under consideration for periodic steady-state solutions of the equation of motion. (b) Illustration of the relation between external flux bias to the total internal DC flux through the rf-SQUID ring. Amplifiers that have a higher $\beta = L_s/L_j$ ratio are more nonlinear. (c) The nonlinearity manifests in the tunability of the amplifier, with higher shunting ratios showing more flux modulation.

To calculate the pump power that produces 20 dB gain, we apply a procedure similar to the tuneup of the real amplifier. At a particular flux value, we can turn up the pump power and analyze the gain of a weak signal. Repeating this process over a half-quanta of flux values produces a plot shown in Fig. 2.6(a), showing a contour of gain points that are

all viable bias points. We have divided this contour into two pieces, representing the lower pump power 20 dB gain points in green, and the higher pump power points in red. Despite also looking like 20 dB gain points, points along the red contour are not guaranteed stability and so we neglect them in the analysis, continuing with only the bias points along the green curve. To investigate the variation over the flux bias of the simulated amplifier, we move along the green contour, testing both the saturation power $P_{-1\mathrm{dB}}$ and the core power added efficiency

$$\eta_{PAE} = \frac{P_{\mathrm{out,1dB}}(\omega_s) - P_{\mathrm{in}}(\omega_s)}{P_{\mathrm{pump}}(\omega_p)}, \tag{2.5}$$

(which reduces to approximately $P_{\mathrm{out},-1\mathrm{dB}}/P_{\mathrm{pump}}$ at 20 dB gain and near amplifier saturation) at each bias point of the amplifier, with the results shown in Fig. 2.6(b)-(c). We find only weak variation in saturation power, but more pronounced variation in the pump efficiency at saturation.

The marginal variation is seen along the bias flux for saturation power, reinforcing notions put forward in [61] and [31] that the static Kerr of the amplifier mode does not act as the primary control of the saturation power, however the significant variation in pump power without a change in saturation power also complicates attributing saturation solely to pump depletion. It may be possible, for example, that the effects of Kerr and pump depletion are opposed to one another in how they affect saturation, with the low (static) Kerr bias points toward $\phi_{\mathrm{ext}} \approx 0.25\phi_0$ having more saturation power relative to the pump, but less pump power to draw from at 20 dB gain. Continued study of this equation of motion is necessary to explain these effects in greater detail [19].

Lastly, we sweep through a significant subset of all possible families of amplifiers of this circuit type to investigate how both the pump power added efficiency and the pump power required to achieve 20 dB gain varies. In Fig.2.7(a) we see that within a given gain family of amplifiers (a vertical line cut through either (a) or (b)), the required pump power increases with respect to the quality factor, and in general that the higher saturation power families of amplifiers determined by the x-axis coordinate require disproportionately higher pump powers at higher quality factors.

We can see this disproportionality directly in Fig.2.7(b), which shows that the pump

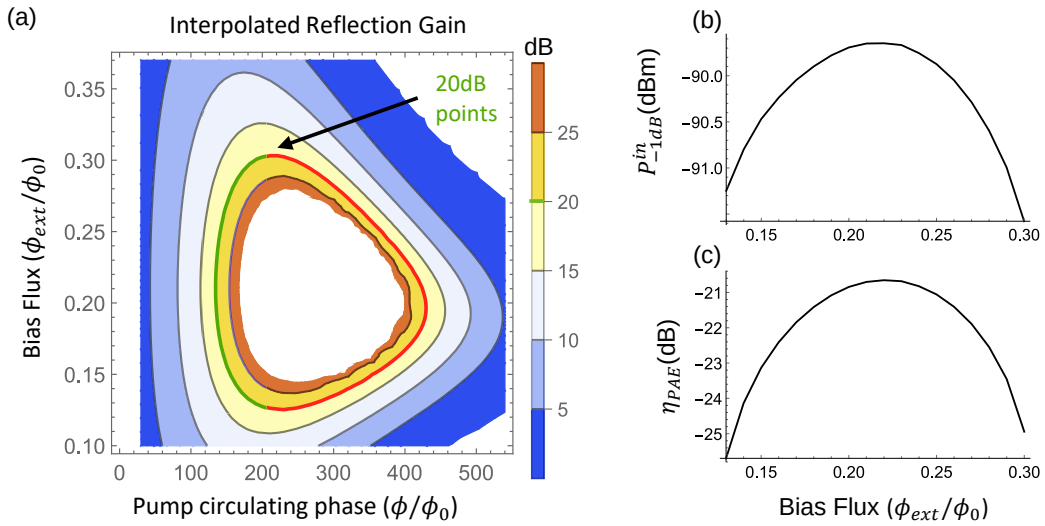


Figure 2.6: (a) The biasing space of a given parametric amplifier, in this case matching the simulation parameters of the amplifier in the main text with a shunt inductance of 15 pH, a junction inductance of 60 pH, and a quality factor of 8.1, with 25 rf SQUIDs arrayed in series. The green highlighted contour indicates the powers that are selected for the rest of the analysis in (b) and (c). (b) Extracted P_{1dB} along the green contour shown in (a). (c) Variation in pump efficiency with respect to the bias flux.

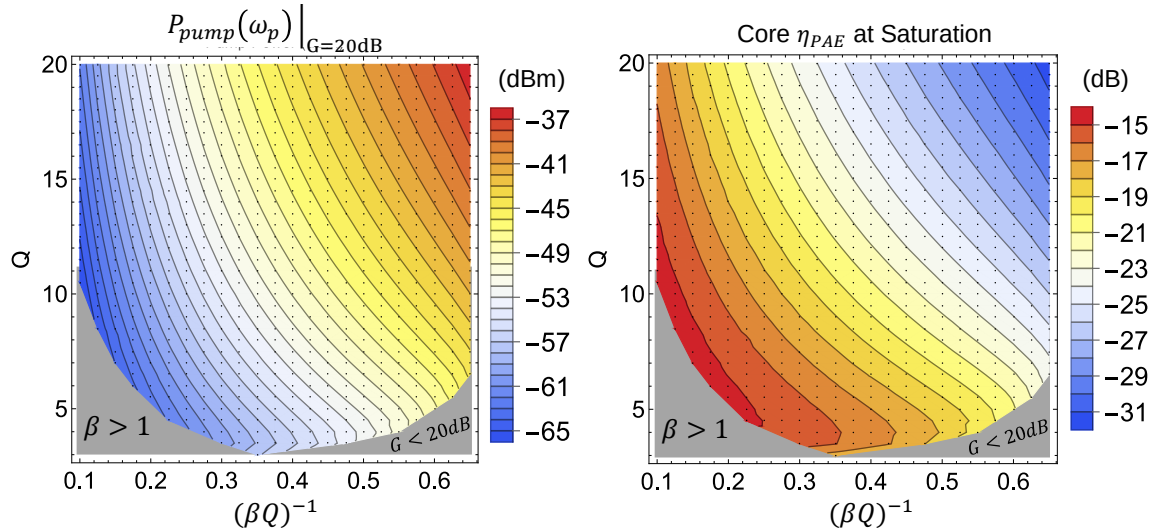


Figure 2.7: (a) The simulated required pump power of the design space of the JPA's in this work. (b) A similar diagram to the previous one for pump efficiency.

efficiency drops significantly with higher quality factor and decreases yet more as one designs for larger and larger (higher and higher saturation power) $(\beta Q)^{-1}$. We note that despite the simulated pump efficiency of approximately -20 dB for the parameters of our devices, the amplifier in this work performs worse in total efficiency, only achieving approximately -40 dB

pump efficiency because of the pump reflecting off of the coupling capacitor, discussed in Appendix 1.4.4.1.

2.6 Experimental details

2.6.1 Fabrication

Our devices were fabricated at Boulder Microfabrication Facility (BMF) at National Institute of Standards and Technology using a Nb/Al-AlOx/Nb trilayer process with low-loss amorphous silicon (a-Si) as an inter-wiring dielectric as described in [25]. From the substrate up the nominal layer thicknesses are as follows: 381- µm-thick intrinsic Si wafer (20kΩ cm resistivity) with 200 nm Nb, 8 nm Al-AlOx, and 110 nm Nb forming a trilayer that is patterned top-down on the substrate to define Josephson junctions and the bottom Nb wiring layer. These structures are blanketed by 310-nm-thick a-Si dielectric layer. Vias patterned through the dielectric form contacts between the junctions, or bottom wiring, and the 300-nm-thick Nb deposited as a top wiring layer over a-Si. The top and bottom wiring layers together with the low-loss a-Si between them form thin-film parallel-plate capacitors. The fabrication was further modified from the description in [30] to insert a thin bi-layer ($<$ 20 nm) of Nb/Al between a-Si and the top Nb wiring layer to facilitate improved patterning of layers and increase device yields. The full details of the fabrication will be provided elsewhere. The top Nb wiring is bonded onto an external PCB with 1% Al-Si wirebonds to carry signals off-chip into the external environment.

2.6.2 Flux biasing: offsets

This amplifier is sensitive to external flux offsets caused by the earth's magnetic field as well as other stray magnetic fields such as nearby ferrite circulators and isolators. In order to shield the device from such stray fields, the amplifier package is enclosed first by a thick 6061 aluminum superconducting shell and then again by a high permeability cryoperm shield made from Amumetal 4K. Despite these precautions, we are able to see distortions in some

setups, such as flux offsets even when no bias current is running through the superconducting electromagnet used to bias the amplifier, seen in Fig. 2.8(b). However, freshly annealed cryoperm cans are able to reject these external flux offsets effectively, shown for a different device in Fig. 2.8(a) by the high degree of symmetry around zero bias current.

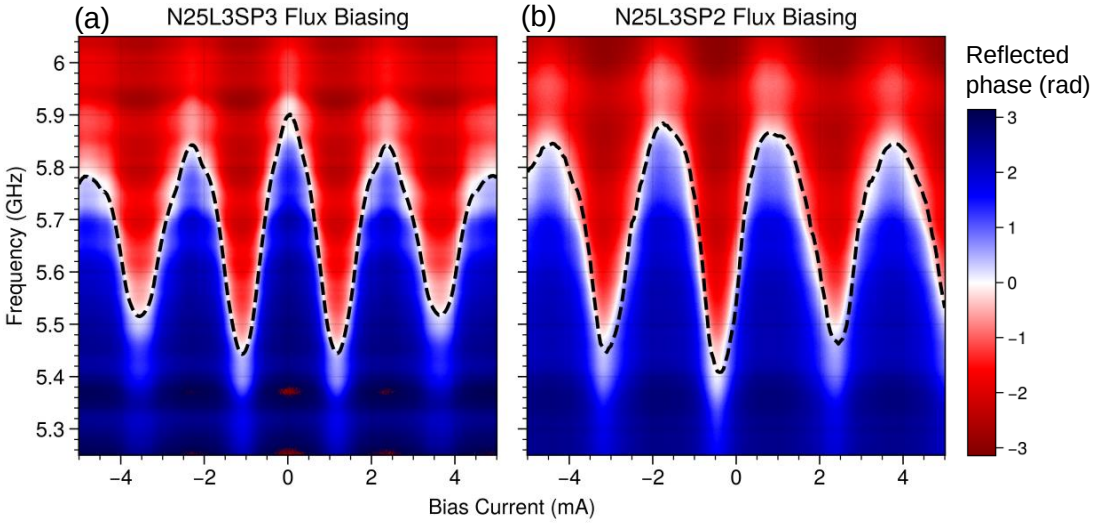


Figure 2.8: Bias flux sweeps for two different RF-SQUID array devices. (a) The flux sweep of one of the two tested devices that make up the data in the main text. The first flux sweep behaves as we expect an rf-SQUID should, with the maximum frequency at zero external bias field. The oscillation in frequency damps at higher fluxes, which suggests that the array is coupling to the external field in an asymmetric way and/or that the field is not homogeneous over the length of the array. (b) The flux sweep of the device tested for quantum efficiency in this work. We find that in this particular device, the zero bias flux is not at the highest frequency of the device consistently over multiple cooldowns, likely due to damage on an older cryoperm can that was used for shielding.

2.6.3 Flux biasing: tunability

One item that weighs on the choice of the quality factor in the design of the amplifier is the tunable frequency range over which the amplifier can achieve 20 dB gain. Because of the low instantaneous bandwidth of a typical reflection JPA, their chief practical strength lies in their tunability to whichever frequency the readout cavity needs to be driven for measurement. Because the readout cavity frequency of a typical transmon qubit is fixed and designs vary over a few GHz for modern transmons, it is necessary to be able to tune the gain center frequency of the amplifier on the order of a few hundred megahertz or to broaden the amplifier's gain response using an impedance matching circuit to minimize the number of separate amplifiers required to cover the entire readout band. With this in mind,

we observe that the amplifiers in the family that barely achieve 20 dB gain shown on the right-hand side of Fig. 2 only achieve that gain in a narrow range of bias flux. Despite having the best saturation power, these devices would effectively only have a tunable bandwidth comparable to their instantaneous bandwidth, motivating increasing the shunt inductance slightly to favor a more practical tunable bandwidth. Our devices are simulated to achieve gain for an array inductance of approximately 320 to 420 pH, which gives a nominal tunable range of approximately 700 MHz measured from the maximum frequency.

In practice, however, the device achieves a much narrower tunable bandwidth than this, only about 150 MHz, which can be seen in Fig.2.11(b). One would expect that perhaps the reduction in tunable bandwidth could arise from flux nonuniformity in the array, and so we analyze this possibility next.

2.6.4 Flux biasing: uniformity

Our high-power JPA is biased by the use of a superconducting electromagnet set into a cutout underneath the copper amplifier package. With a mm-scale distance from the plane of the array to the edge of the electromagnet solenoid and the finite radius of the coil of approximately 1 cm, we anticipate that there should be very little flux nonuniformity in the array from the nonuniformity in the magnet field alone.

To investigate this, we sweep the current through the magnet over the maximum possible values without causing heating of the magnet loom connections in the refrigerator, then analyze a fast Fourier transform (FFT) of the resonant frequency modulation in Fig. 2.9(b). We can then identify peaks in the spectrum in Fig. 2.9(b). The transform correctly identifies the primary modulation frequency highlighted by a green vertical line about the peak value.

The origin of the second-highest peak is unknown. One possible explanation is that large superconducting planes at the boundaries of the array are shielding the first and last few rf SQUIDs from receiving the full external flux through their loops. This local shielding would be very similar on both the ground and resonator node boundaries of the array. In effort to remedy this, future efforts will include tapers to the array, diminishing the local effect of

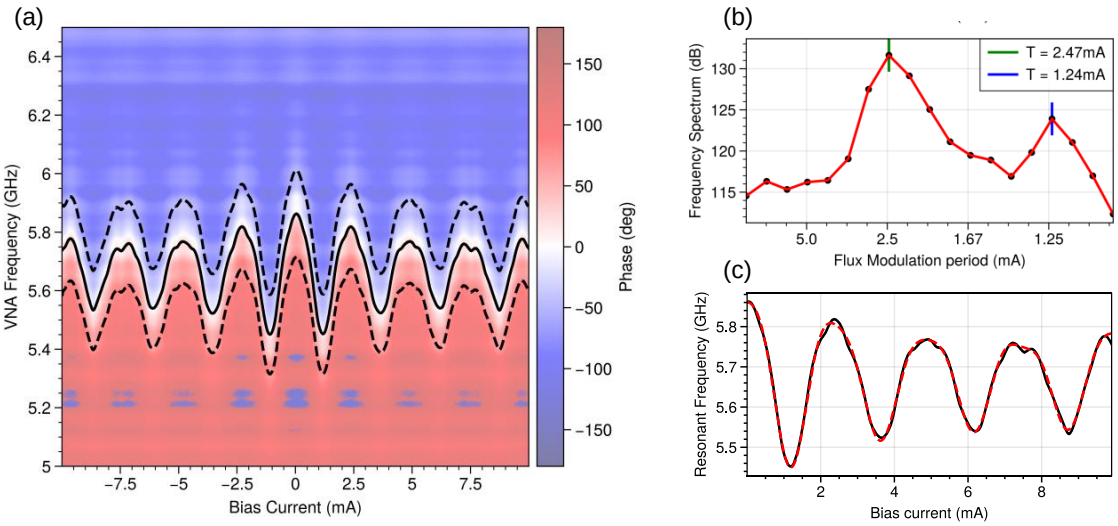


Figure 2.9: Flux biasing nonuniformity frequency components. (a) A wider fluxsweep of the device shown in Fig. 2.8(a) that is suitable for the multiple periods required for accurate analysis in the Fourier domain. Fitted resonant frequency is shown by the solid black line, and the fitted quality factor at each flux bias is indicated by the bracketing dotted lines. (b) A discrete cosine transform of the fitted resonant frequency, truncated approximately 20 dB below the highest peak and convolved with a 3-sample wide Blackman window, shown on a log scale. The modulation period of the highest peak is the closest analog of a single flux quanta threading each loop in the array. However, it is clear that not all of the rf SQUIDs are in sync with one another. (c) The resonant frequencies are compared directly with an inverse discrete cosine transform of (b), serving to verify that the smoothing and truncation of the full transform have not appreciably altered the data.

any adjacent superconductors.

2.6.5 Saturation power calibration

Power calibrations are often difficult to do in a straightforward way for devices at the base of a dilution refrigerator. Measurements of the input attenuation up to the plane of the amplifier at room temperature are easy, but they are often inaccurate, as all of the losses are susceptible to changes in temperature. The most straightforward way to calculate power for an amplifier serving to read out a qubit is to use the qubit as a power meter [57]. Independent measurements of the qubit can extract the χ frequency shift of the qubit per photon in the readout cavity. This allows qubit spectroscopy at varying readout powers to directly measure steady-state power in the readout resonator by watching the qubit shift in frequency. As we show in Fig. 2.10(a), after establishing this baseline with the qubit,

the only tasks remaining are to link the DAC drive voltage to a voltage on a spectrum analyzer. This is simple to do by simply removing the drive from the fridge and attaching it to a spectrum analyzer, a necessary step in eliminating mixer LO leakage and tuning single-sideband upconversion in any case. The results of these calibrations as well as similar procedures for the VNA are shown in the remaining subplots in Fig. 2.10. In our case, the conversion loss of the upconversion mixer and the cable losses are roughly balanced by the additional amplification in the chain.

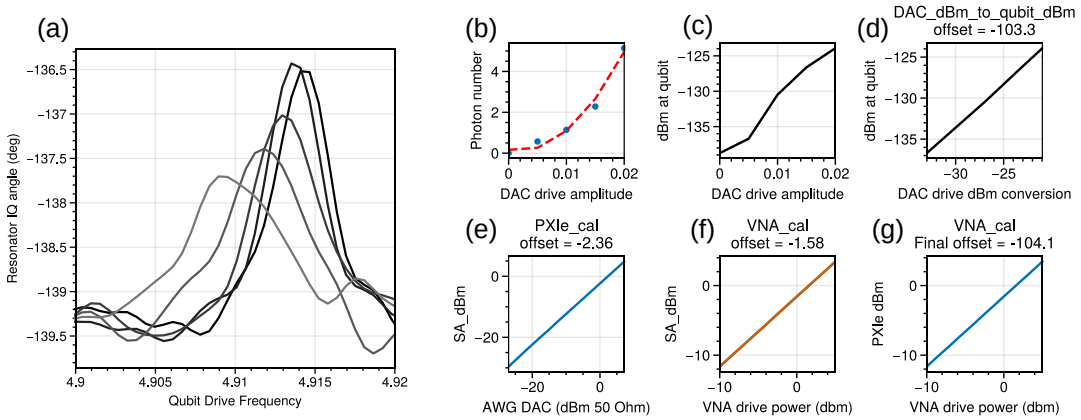


Figure 2.10: Photon number power calibration via qubit spectroscopy. (a) Qubit spectroscopy with varying readout resonator power. The χ and κ of the qubit are independently calibrated using standard reflectometry and measurement post-selection. (b) Extraction of steady-state resonator photon number versus DAC voltage drive, exhibiting the quadratic variation expected. (c) conversion of the previous plot to dBm (d) Conversion of the x-axis of the previous plot to dBm. This forms the principle part of the calibration (e)-(g) Calibration the rest of the measurement chain against the spectrum analyzer using the same cable that they would connect into the dilution refrigerator with, and subtracting them to amount to a small change to the final offset. Powers are swept over a wide range to check the linearity of the measurement chain.

2.6.6 Saturation power variation over bias points

Many amplification schemes focus on the available bandwidth of the amplifier. Considering resonant JPA's without matching networks typically lack instantaneous bandwidth, they need to be flux tuned to different operating points, and the pump moved in frequency, in order to achieve gain at a variety of frequencies. A figure of merit of such a device is the tunable bandwidth - the range where the amplifier can operate when allowing for any possible tuning. Because changing the flux bias and pump tuning of the amplifier likely

affects the saturation power, we measure multiple bias points distributed over the rf-SQUID JPA's tunable bandwidth of approximately 150 MHz.

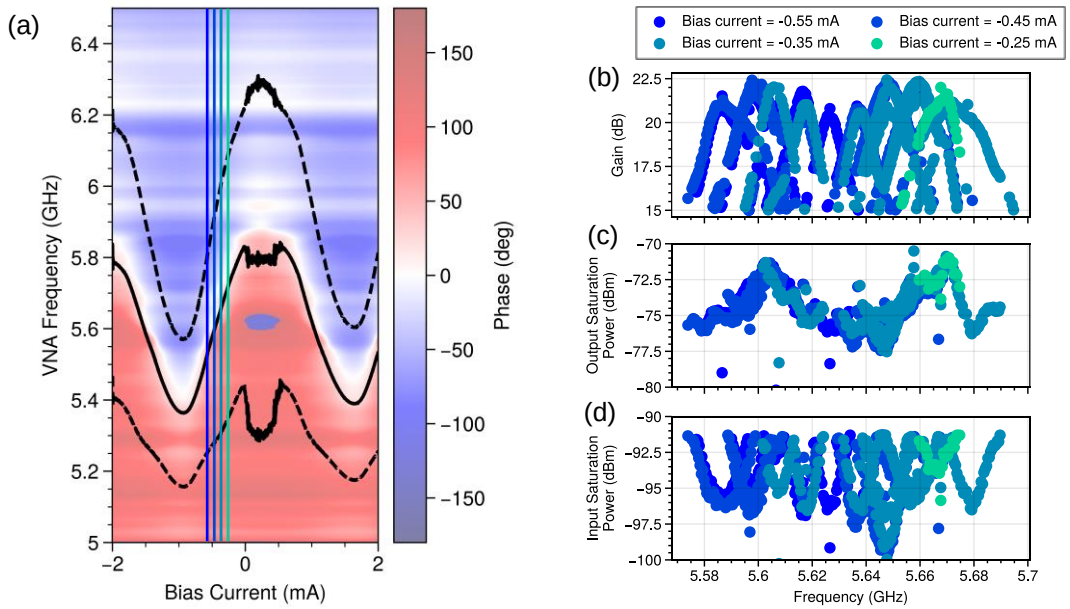


Figure 2.11: Saturation power variation over bias flux for a high-power RF-SQUID array. (a) A flux sweep of the device used for saturation power data in this work, with vertical lines marking particular bias points of interest for producing gain in the amplifier. (b)-(d) Various gain bias points of the amplifier used for saturation data after accounting for the maximum possible loss at the given frequency over any flux bias of the amplifier, shown with their measured input and output saturation powers. The colors of the points correspond to their flux bias points.

We find some few-dB variation in the saturation power over the bias points of the amplifier, with most bias points achieving close to the average value reported in the main text. This is supported by the simulations, and similar to many previously reported results for array amplifiers [15, 61, 62].

2.6.7 High power measurement transitions

A natural next step in qubit measurement with a high-power JPA is to increase the power of the measurement to attain a higher integrated SNR with a shorter measurement time. While the amplifier is more than capable of handling the additional power, up to several hundred photons at this readout resonator frequency, the readout resonator plus qubit system is certainly not, as discussed in the main text. Here we show an example of how this process can go awry by showing integrated readout histograms.

In Fig. 2.12(c) and (d) we give just one example of the measurement results one might expect to see when measuring at powers where the QND nature of dispersive measurement breaks down because of measurement-induced transitions in the qubit state. Transitions occurring faster than the integration time of the readout (here approximately 800ns, not including the ring-down time of the resonator) are visible as the "tracks" between states, increase in frequency at higher powers, as noted in multiple other publications [56, 59, 25, 11]. In particular, we even see transitions that are supposed to be forbidden, as well as many multi-photon transitions. In this case, the time resolution to view these transitions is limited by the ring-up and ring-down time of our 1 MHz linewidth resonator, and so we leave the study of these effects in further detail to future work.

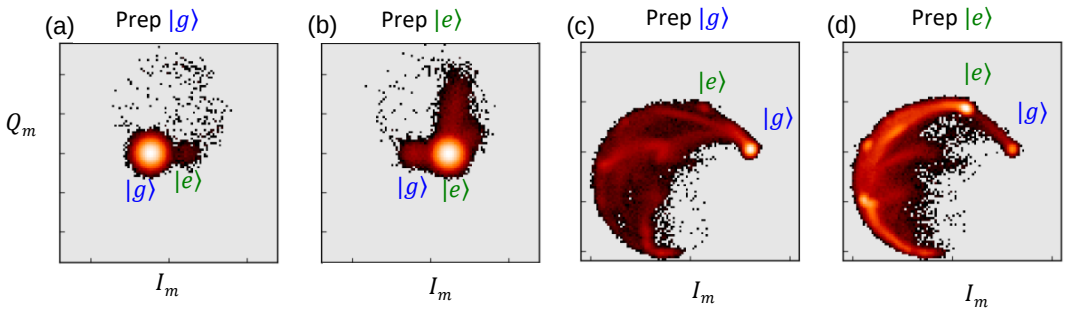


Figure 2.12: High power readout histograms for a transmon qubit. (a) and (b) measurement histograms of the transmon shown in the main text with the amplifier on, counts shown in logarithmic color scale. (c) and (d) High power measurements of the transmon in the main text. Because of the high saturation power of the rf-SQUID JPA, we can probe the extremely high power response of the transmon. Multiple higher states and transitions between states are easily separable, enabling future studies of transition rates and effective qubit temperature as a function of readout power.

3.0 Time-Domain qubit readout and weight functions

Readout is a fundamentally time-domain task. In this section, we will focus on the design of the readout stack by walking through the transmitting and receiving chain of the readout electronics. This design clears the way for a high-frequency carrier tone to deliver an IQ-modulated pulse to the readout cavity via an upconversion mixer. Once there, we will discuss the dynamics of the readout cavity, and how the design of the cavity dispersive shift, χ, and its external power decay rate, κ, influences the trajectory of the voltage response that is analyzed to determine the qubit state. Lastly, we will discuss how these trajectories can best be weighted, and give an overview of how we can build weight functions for reading out more than two states of the qubit from our publication in Ref. [24], which opens some interesting options for implementing qutrits and generally qudits using the same transmon hardware with high efficiency, high fidelity readout. Qutrits, in particular, are of interest to the error correction community for their potential to leverage error detection and correction protocols with much less physical circuit complexity than would otherwise be needed in a qubit-centric system.

3.1 Dispersive readout in the time domain

3.1.1 Hardware setup

In order to understand how a readout cavity is driven from the equipment available at room temperature, we should first discuss how the circuit is connected. To keep the discussion focused while remaining sufficiently general, I will target the setup we used in Ref. [21], shown in Fig. 3.1. From a single input port, the device is driven through two different lines via a frequency diplexer with a low-pass 3 dB cutoff at 8.65 GHz, a high-pass 3 dB cutoff at 9.35 GHz, and over 80 dB rejection from the high-pass port to the common port at the frequencies in this data. The high-pass line provides the pump through a very lightly

attenuated (20 dB attenuator at 4 K) line, and the low-pass port receives the reflected signal from the cavity via the circulator and two Keycom 15 cm superconducting NbTi semi-rigid coax cables. The diplexer enables rejection of the thermal noise of the high-power pump line without losing pump power at the pump frequency at the cost of additional cabling and insertion loss of approximately 1 dB from the diplexer, both of which must reduce the quantum efficiency of the measurement.

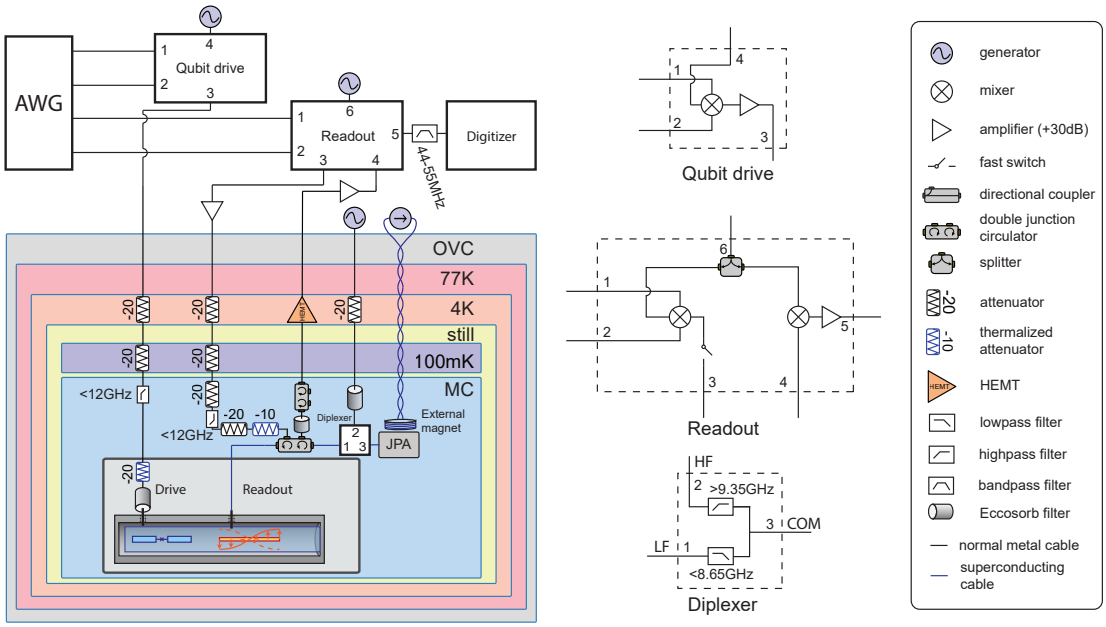


Figure 3.1: The full circuit diagram of the high-power rf-SQUID JPA experiment, which shows a typical setup for driving a readout resonator in reflection for qubit measurement.

In Fig. 3.1, we run the transceiver with an IF frequency of 50 MHz from the AWG, digitally demodulated at that same IF frequency in the ADC, both of which have an analog bandwidth of about 1 GHz, placing their first Nyquist boundary at 500 MHz. This is all to say that our IF frequency is well within their sampling range. Specifying this IF frequency lets us categorize this receiver most succinctly as a superheterodyne transceiver. Operating the receiver has a few distinct advantages. One, it largely eliminates the need for a fast rf switch at the input, although we use one anyway. Two, it allows us to easily and quickly change the frequency of the drive, which makes it much quicker to find optimal readout and qubit drive frequencies. Lastly, it's much easier to filter the receiving side with a single-sideband mixer, allowing us to eliminate the idler tone generated by the parametric amplifier as it falls in the rejection band of the mixer [20].

Using this setup, we can create any envelope of pulse shape we want to send to the readout cavity via IQ heterodyne upconversion, transmit down the lines of the fridge, where attenuation along the way thermalizes the signal, then amplify it all the way back up, starting with the high power rf SQUID JPA from the last section and ending with a final rf low noise amplifier (LNA) before the downconversion single-sideband (SSB) mixer, followed by a Stanford Research Systems gain block to amplify further the 50 MHz signal after downconversion just prior to entering the digitizer. We have shown [18, 21] this setup to be highly quantum efficient when the rf SQUID JPA is on, meaning that the final demodulated IQ signal-to-noise ratio (SNR) is close to the SNR at the interface of the readout cavity within the noise limits of phase preserving amplification [7].

3.1.2 Readout trajectories

Building on the previous section's hardware setup, let's look at the major characteristics of what we expect for the time-domain voltages coming out of the digitizer in Fig. 3.1 when the readout resonator is driven with a simple square pulse and we measure the reflected wave. While the full response has to be treated most properly by solving the full stochastic master equation, for example in Ref. [60], here I will treat the different qubit states simply as different resonant frequencies for a resonator, each with linewidth κ and separated in frequency by the dispersive shift χ.

Both χ and κ are marked in Fig. 3.2(a), where I show the steady-state phase response of an overcoupled $Q = 10$ resonator (see chapter 4 for a definition of couplings for resonators) as an example. The drive frequency in Fig. 3.2 is marked by a vertical dotted line and set exactly between the resonant frequency of the resonator when the qubit is in the $|g\rangle$ state shown in orange and the resonant frequency of the resonator when the qubit is in the $|e\rangle$ state, shown in blue. Reference [5], among many others, have shown this to be the optimal driving point for extracting two-state readout information from a dispersively coupled transmon qubit. In Fig. 3.2(b) I summarize the response of this optimal $\chi/\kappa = 1$ resonator to a readout pulse. The input is a simple square wave of time duration $10/\kappa$, shown in yellow. Because the resonator is driven in reflection, the IQ magnitude of the response at $t = 0$ right at the start

of the pulse is not zero. However, it is also not immediately indicative of the qubit state, as *all* of the state-dependent responses start at -1 in the quadrature-phase component. The intracavity amplitude, which I will show later on, is what holds qubit state information and builds up with the characersitic rate κ in the resonator. Over time, this light also leaks out of the resonator from the same port that drove it there, and interferes with the initial reflection to generate the profiles in the in-phase (I) and quadrature-phase (Q) components we see in Fig. 3.2. There are several characteristics of the I and Q quadrature responses worth discussing, as they provide information about the detuning of the drive from the resonator as well as the resonator's κ. Since the $|g\rangle$ readout resonator frequency, which we can call $\omega_{r,|g\rangle}$, is detuned positively from the drive, the resonator does not respond smoothly, instead yielding some oscillation in the envelope response to the drive. The characteristic frequency of this oscillation is directly indicative of the ratio of the detuning to the linewidth of the resonator, and the exponential rise and decay rate corresponds to the linewidth itself. This is equally true in Fig. 3.2(c), but with the opposite detuning and therefore negative quadrature-phase component.

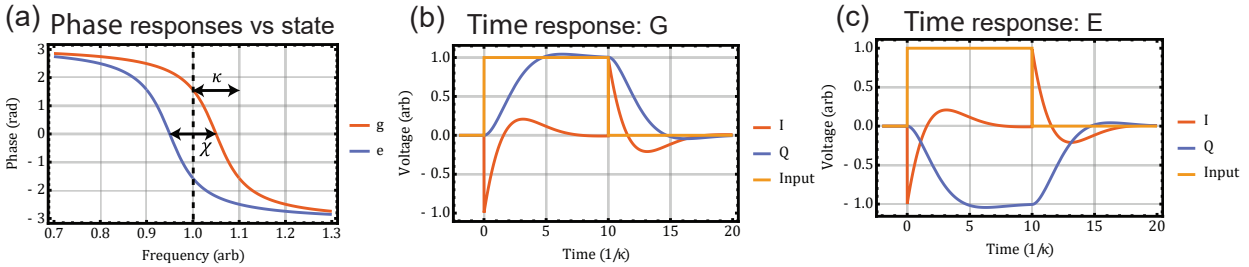


Figure 3.2: Voltage Response of a driven resonator detuned according to qubit state. (a) Steady-state phase responses of the readout resonator shown driven at the frequency indicated by the black arrow. (b) Time-domain response of the resonance corresponding to the ground state in (a) shown in the rotating frame of the drive. (c) Time-domain response of the resonance corresponding to the excited state in (a) shown in the rotating frame of the drive.

If we generalize slightly, it becomes onerous to analyze every single state in time-dependent traces as in Fig. 3.2. Instead, we can elect to make the time dependence implicit, and view the trajectories as parametric curves in the IQ-plane. Figure 3.3 does exactly this for two different measurement drive frequencies, highlighting the ability to optimally read out different pairs of qubit states by changing the drive frequency. Here I will only show the ring-up for clarity between the initial point and the steady state IQ value.

In Fig. 3.3 we can compare the parametrized curves to the steady-state phase values

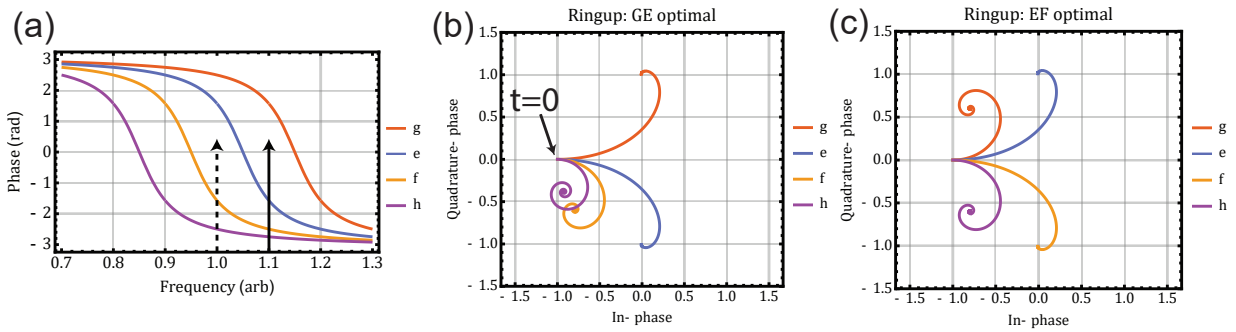


Figure 3.3: Readout trajectories for the first four qubit states at two different drive frequencies. (a) Steady state phase response of the readout resonator frequencies of the first four qubit states on the assumption that all of their χ's from adjacent states of the resonator are identical, with the GE optimal drive indicated by the solid arrow and the EF drive indicated as a dotted arrow. (b) Readout trajectories corresponding to the drive frequency $(\omega_{r,g} - \omega_{r,e})/2$ indicated by the solid arrow in (a). (c) Readout trajectories corresponding to the drive frequency $(\omega_{r,e} - \omega_{r,f})/2$ indicated by the dotted arrow in (a)

to find a satisfying intuitive explanation to why a $\chi/\kappa = 1$ is the optimal value for maximum SNR. In Fig. 3.3(b), we can see that our first choice of drive frequency is optimal for distinguishing $|g\rangle$ from $|e\rangle$ because it gives them the maximum possible separation in the IQ plane for a given drive amplitude, 180°. We can also see, however, that this choice of drive frequency only maximizes the $|g\rangle, |e\rangle$ separation, and that any other pair of states, say $|e\rangle, |f\rangle$, is not maximally separated. We can choose a different frequency for this task, as shown in Fig. 3.3(c) for EF readout indicated by the dotted arrow in Fig. 3.3(a). These trajectories are exactly what will be integrated to determine a final IQ point that will serve as the measurement. These IQ positions, in turn, will be used to make a decision on the final qubit state in the real experiment. Before we get into how we integrate, I want to comment on one big degree of freedom in the measurement that sees a lot of common use: pulse shaping.

3.1.3 Cavity pulse-shaping

The square-wave input pulse shapes shown in Fig. 3.2(b) and (c) are very simple. One inconvenient cost of that simplicity is that we have to wait for quite some time for the light to populate the resonator: the characteristic ring-up time $1/\kappa$ of the resonator. Is this a cost we have to pay? Luckily, the answer is no. At least, we don't have to pay *that* cost. Prior

work [38] has shown that because we can alter the input pulse very quickly, we can shape it to get more light into the resonator without over-populating the cavity. Fig. 3.4(a) shows an example of the routine (CLEAR) shown in Ref. [38].

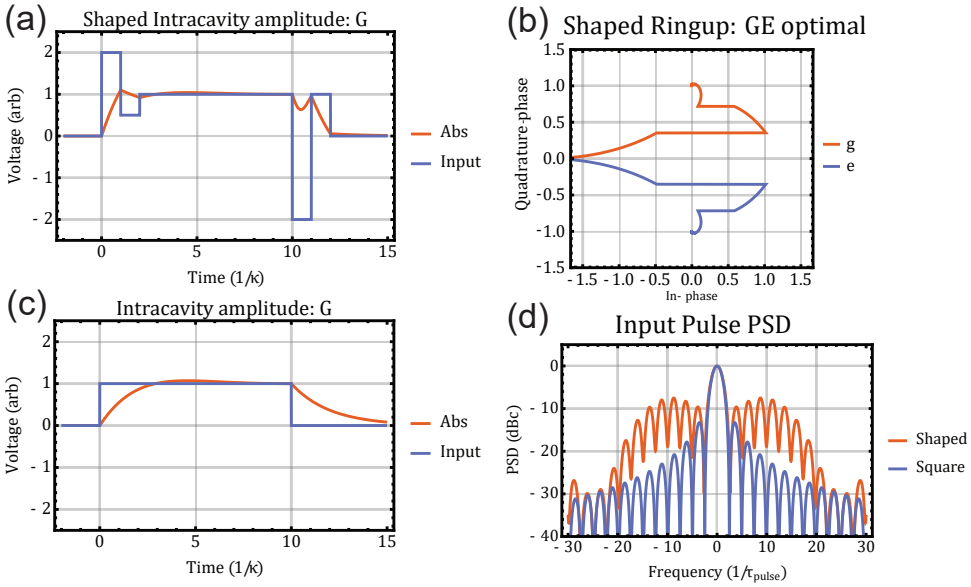


Figure 3.4: Shaped pulses used to decrease the ring-up and ring-down times of the intracavity population of the readout resonator. (a) A modified implementation of the routine in [38], but here the parametrization of the system allows it to be applied for any $\chi/\kappa = 1$ system. (b) New ring-up trajectory of the system with the input pulse given in (a). Sharp discontinuities indicate sudden changes in the input. (c) Un-shaped response of the intracavity population with a square wave input for comparison with (a). (d) Pulse PSD for the shaped (orange) vs un-shaped (blue) inputs, highlighting that adding fast time dependence to shape the pulse dramatically increases the sidelobes of the pulse

In Fig. 3.4(a), I show an adaptation of the routine in Ref. [38] that is similar to the shape that I used in Fig. 2.4(d) to produce higher readout fidelity than would otherwise have been possible. To see that this is the case, we need to talk about the intracavity population, given via input-output theory as $(V_{in} + V_{out})/\sqrt{\kappa}$, where both V_{in} and V_{out} are complex magnitudes. Plotted in Fig.3.4(a) and Fig. 3.4(c), we can see that the intracavity amplitude would usually decay relatively slowly compared to how fast the input turns off, but by manually over and undershooting the beginning and ends of the input pulse, we can bring the intracavity population into rough alignment with the duration of the input pulse itself. The cost of this improvement is shown in Fig. 3.4(d), where we can see that the power spectral density (PSD) of the pulse in frequency space is drastically altered by adding this seemingly-helpful time-domain shape, generating sidebands that scatter a signficant portion

of the input drive over 10x farther than it would otherwise reach in an unshaped square wave, which we have already discussed in Fig. 1.8.

While pulse shaping is undoubtedly beneficial for single-resonator readout, decreasing the effective reset time by over a factor of 5 in Fig. 3.4(a) vs Fig. 3.4(c), its place in frequency-multiplexed readout circuits is debatable. The much-increased crosstalk between adjacent channels poses a major constraint on resonator spacing, and therefore further increases already-difficult bandwidth requirements for parametric amplifiers in the readout chain.

3.2 Matched filters

Now that we have established the time-domain dynamics of the readout resonator in the last section, we have to ask how to extract information about the qubit state most efficiently. There are, to put it lightly, a couple of rules of thumb to abide by here, because there is a long and vibrant history in physics of asking how effectively an experimental apparatus can measure any particular variable. So I will start this section with a brief discussion of a few references that put the (very large) conversation of effective and precise physical measurements into this specific, narrow context of measuring state-dependent resonator responses in transmon qubits.

Clerk *et al* in Ref.[10] give a comprehensive introduction to amplification and quantum efficient measurements of transmon qubits. There, they quantify two different rates of information extraction in the system. The first is the dephasing rate Γ_ϕ, which describes how quickly the photon number noise outside of the readout resonator extracts information about the qubit state (and throws it away). The second is the Γ_{meas}, which describes how quickly a *coherent* measurement can extract information about the qubit state. Remarkably, as Ref.[10] shows very clearly via two methods, the measurement rate can only ever equal the dephasing rate, it cannot exceed it. The ratio of the two determines the quantum efficiency by $\eta = \Gamma_{meas}/\Gamma_\phi$. These two rates are often measured independently and directly compared to each other to determine quantum efficiency of a particular measurement, as in Ref. [29]

for example. Measuring Γ_ϕ is easy, it can be directly related to the dephasing rate T_2 via an equivalent external temperature that gives a corresponding photon number noise near the resonator frequency [57]. Measuring Γ_{meas}, however, is highly susceptible to a long sequence of decisions about how to process the information coming from a dispersive measurement. It is the biggest and most often ignored step - calibrating weight functions - that I intend to focus on here. This is also one of the most critical steps in Ref. [6], which gives a general way of extracting quantum efficiency to supplement the methods we use in chapter 2 from Ref. [18].

I view weight functions as a sequence of decisions in the time-domain. At every moment in time, which so far I have treated as a continuum but in reality is a discrete set of samples, we have to decide how relevant the information we have at that time is to determining which state the qubit is in. To state directly how important the weight functions are, it is *very easy* to make the wrong decision. For example, it would be foolish to weight the $t \in [0, 1]$ information in Fig.3.4(c) in the same way as the $t \in [5, 6]$ information, because the intracavity photon number in the former is much lower than the photon number in the latter. In other words, the measurement *rate* is much faster in the $t \in [5, 6]$ case, so that information is more important. This is the foundation of the argument in Ref. [60], which defines in Eq.(12) the ideal complex weight function for dispersive measurement in terms of the mean intracavity fields in the readout resonator for the excited and ground states of the qubit: α_e and α_e, respectively. I have adapted it below for a reflection resonator (whose "transmission coefficient" from Ref. [60] is now 1, because there is only a single port) as

$$S(t) = \sqrt{\kappa}(\alpha_e(t) - \alpha_g(t)). \tag{3.1}$$

I have also dropped the superscripts from Ref. [60], because we are only measuring a single qubit in our case. This has a tidy conceptual foundation, which is that *any* difference in *either* quadrature of the intracavity fields for the two different qubit states results in measurement induced dephasing Γ_{meas} and therefore must be factored into an ideal weight function. If we fail to do this properly, the recorded measurement efficiency Γ_{meas} will fail to accurately reflect the maximum possible efficiency set by the losses and other hardware in the system. In other words, we only get the highest efficiency if we make the right decisions

about which information matters. This is the same spirit of and is consistent with the ideas in Refs. [55, 35, 17]. However, one notable thing that this kind of reasoning leaves out is the decay of the qubit over time, which Ref. [17] covers in more detail. The short version of that discussion is that since the qubit is more likely to have decayed as time progresses in the measurement (especially since recent study has concluded that the decay rates for each transmon state under strong drive are *much* different than idle decay for various reasons [56, 25, 11, 65, 13]) the information later in the trajectory is less valuable than information earlier in the trajectory, and so the weight functions should reflect this.

This logic readily expands to any given pair of qubit states. In Fig. 3.5(a)-(c) I show examples of what these weight functions look like in the time-domain for pairs of different qubit measurement states. To acquire the integrated result of any particular measurement, all we have to do is multiply these complex weight functions with any particular measurement trajectory, then integrate it over the time of the measurement. In Fig. 3.5(d)-(f), I show the results, which are optimal as-defined in Ref. [60], produce different IQ values for the measurement depending on the filter used.

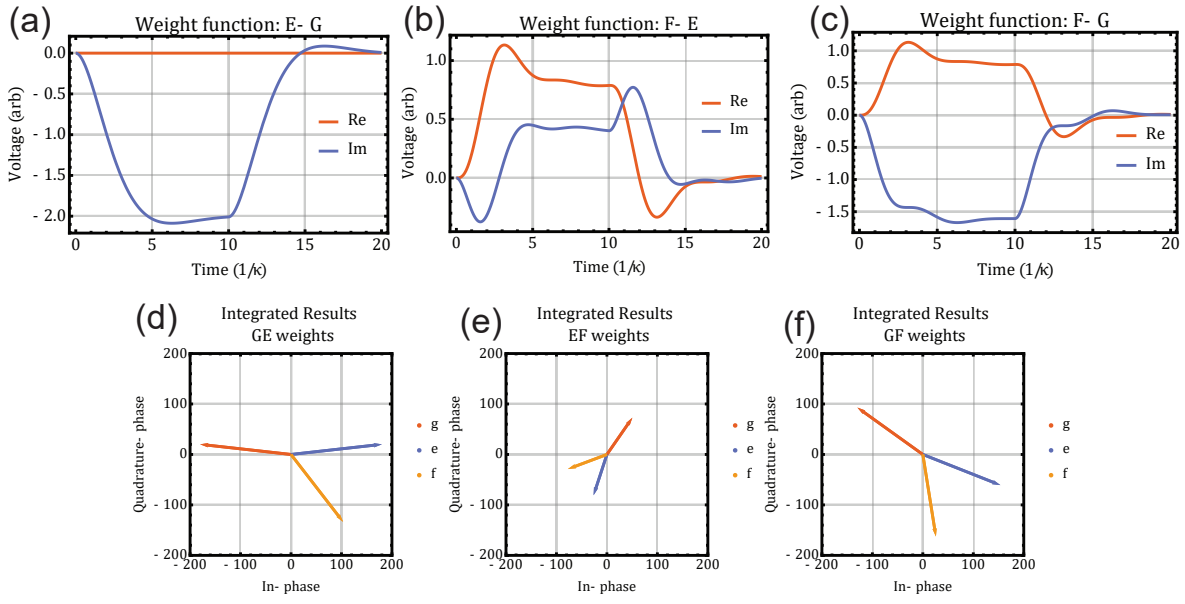


Figure 3.5: Pairwise weight functions derived from Ref. [60] and their results with integrated mean values of trajectories. (a)-(c) the real and imaginary components of the weight functions for (a)E-G readout, (b)F-E readout, and (c)F-G readout given the un-shaped square wave input pulse parameters in Fig. 3.2(b) and (c). (d)-(f) Integrated measurement IQ results for each qubit state readout trajectory in Fig. 3.3(a) weighted by the functions in (a)-(c), showing that the measurement IQ result will be very different depending on which weight function is used.

3.3 Majority vote pairwise filtering

In the three-state readout case, the matched filtering approach shown in Fig. 3.5 has a problem. If we can only generate the matched filters in a pairwise fashion, how can we incorporate the information from all three matched filters to optimally distinguish all three qubit states from one another? This was one of the primary motivating questions in our work with Saeed Khan in Hakan Türeci's group at Princeton from 2021-2023. In the end, the correct answer was a complete and self-consistent way of calculating the matched filters for N different qubit states [24], called the "temporal post-processor" (TPP). The TPP behaves more like traditional matched filtering in electrical engineering, where the answer is mapped from 0 to 1 in an N-dimensional normalized vector space - each state represented by a unit vector. The N-dimensional answer vector will have its largest component in a given direction, and the state corresponding to that direction is the answer. This approach is very general and easily trainable - even analytically calculable from the measured mean trajectories, much like the filter Silveri *et. al.* give in Ref. [60]. Because the method is based solely on matrix multiplication, it's also easily integrated into an FPGA, although this hasn't been shown experimentally yet.

Before I discuss a few major results from that publication, I would like to comment on a predecessor to this result - majority-vote filtering - that we ended up omitting from the final publication because its capabilities were eclipsed by the more general TPP result in Ref.[24]. However, the framework that we built with it did live on in the publication, where we compared the performance of the TPP to each matched filter, and broke down the machine learning component of the TPP in terms of the weight functions that it learned versus the matched filters in the white-noise case.

The idea behind the majority vote filter is simple: in a three-state case, why don't we process each individual filter *in parallel*, then at the end let them each vote for what they think the state is between the two states that they were designed to measure? The GE-matched filter decides whether the state is G or E, the GF-matched filter decides between G and F, and the EF-matched filter decides between E and F. Then the votes are tallied, and if two filters agree on a state, that is the final decision of the total filter. The catch

is, of course, that they might *all* disagree in multiple ways. For example, GE could decide G, GF could decide F, and EF could decide E. In this case, no decision is rendered - a null result. Although it seems strange, this is not necessarily a false physical situation, the qubit could have accidentally been excited outside of the three-state manifold. One benefit of this method is that it is easy to implement in an FPGA setup that already has matched filter capability - you just have to duplicate the setup and run it in parallel for the other two filters, provided you have enough space in the fabric. So in this sense, it has less of an initial programming overhead than implementing the full TPP in Ref. [24].

To show how the majority vote filter works and to lend some data to the previous section, I will walk through a particular implementation of a majority-vote filter for three-state readout on the same amplifier and qubit+cavity combination as chapter 2.

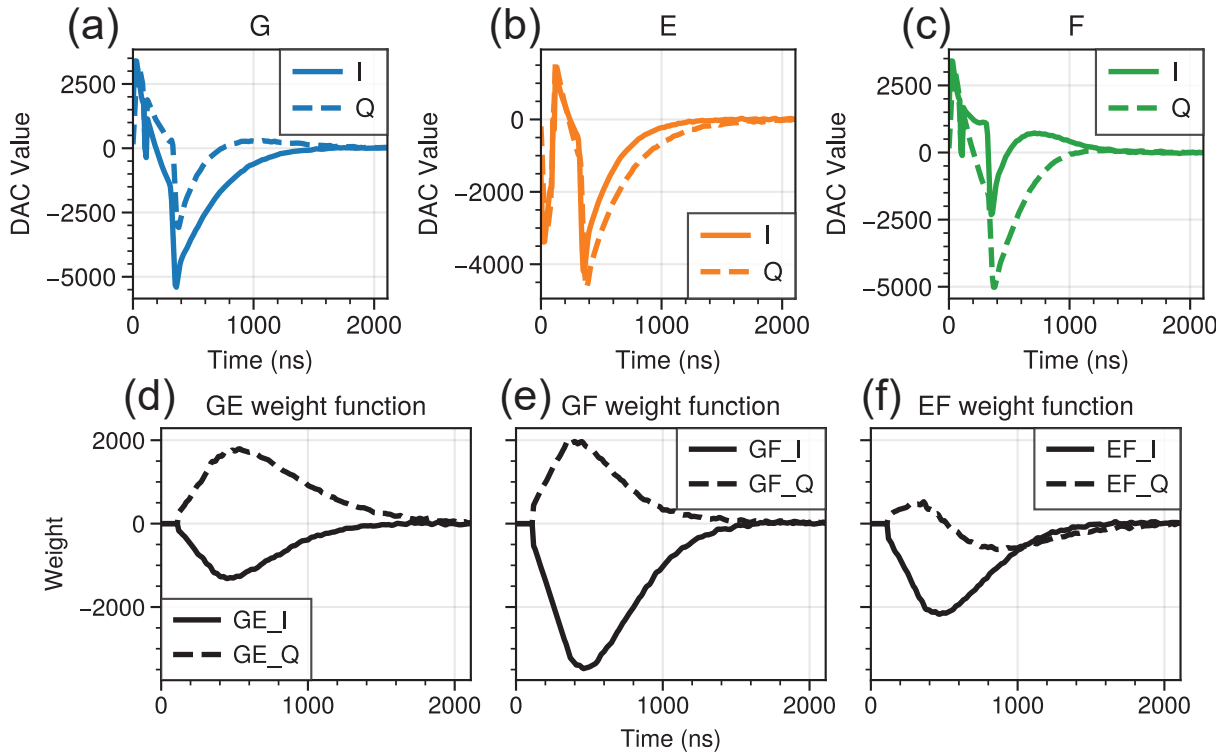


Figure 3.6: Readout trajectories and weight functions on a real qubit measurement setup. (a)-(c) Real and imaginary components of the reflection resonator response for the qubit in the (a)$|g\rangle$ (b)$|e\rangle$, and (c)$|f\rangle$ states when tuned for readout roughly between the $|g\rangle$ and $|e\rangle$ states, averaged over approximately 30000 records per qubit state. The beginning of the time trace approximately corresponds to t = 0 in Fig. 3.2. (d)-(e) The weight functions generated for the (d)GE, (e)GF, and (f)EF pairs of states.

In Fig. 3.6 I show the readout trajectories of the readout resonator discussed in the readout experiment in chapter 2 when driven with a simple square wave input after down-

conversion from the drive on a single-sideband mixer to an IF frequency of 50 MHz, and software demodulation at an IF frequency of 50 MHz down to baseband. As we can observe via the relatively long ring-down time of this resonator, the external power decay rate is quite slow. As an exercise in quickly estimating the decay from the voltage profile, the $1/e$ decay time of the average voltage profile is an estimate for $2/\kappa \approx 360$ ns, which equates to a $\kappa/(2\pi) \approx 900$ KHz, which is fairly close to a more careful fit to $\kappa/2\pi = 1$ MHz which includes complex spiraling from the detuning as well. This slowness is the primary barrier to a higher fidelity in this scenario without a shaped pulse as in Ref. [38] and Fig. 3.4(a). Since in this particular instance this data was taken to verify that the TPP could learn "regular" quantum trajectories, this pulse is *not* shaped as it was to produce the (much higher) readout fidelity in Fig. 2.4(d). It was also a barrier that in general the pulse shaping that optimizes intracavity population becomes highly dependent on the pulse time when the cavity does not reach steady-state during the measurement, creating a difficult optimization problem when sweeping both readout pulse time and amplitude. From the average complex trajectories in Fig. 3.6(a)-(c), we then calculate the weight functions in Fig. 3.6(d)-(f) according to Eq. 3.1, substituting the states for each pair accordingly.

From these weight functions, we can integrate the results into histograms where we can get our first good look at the noise distributions of these readout states. Figure 3.7(a) through (c) shows histograms of the trajectories in Fig. 3.6(a) through (c) integrated against the appropriate filter from Fig. 3.6(d) through (f). The result for each measurement record is a set of three IQ values, one for each filter. From here, we have to decide how to use this information to generate the "vote" for each filter. Generally, we fit a 2D Gaussian probability distribution to each visible "blob" in Fig. 3.7(a) through (c) and then compare the distributions for each state by subtracting one from the other. The results are shown in Fig. 3.7(d)-(f) by the color, which indicates the "decision" of the filter for that point in IQ space, and the shading which is indicative of the strength of that prediction at that point in IQ space. A couple interesting features are worth discussing. The first and most important feature is the separatrix for each filter, emphasized by the dotted lines in Fig. 3.7(d)-(f) along the boundary of the change in hue. The separatrix is the contour defined in IQ space where the two fitted probability distributions are equal, meaning that any measurement

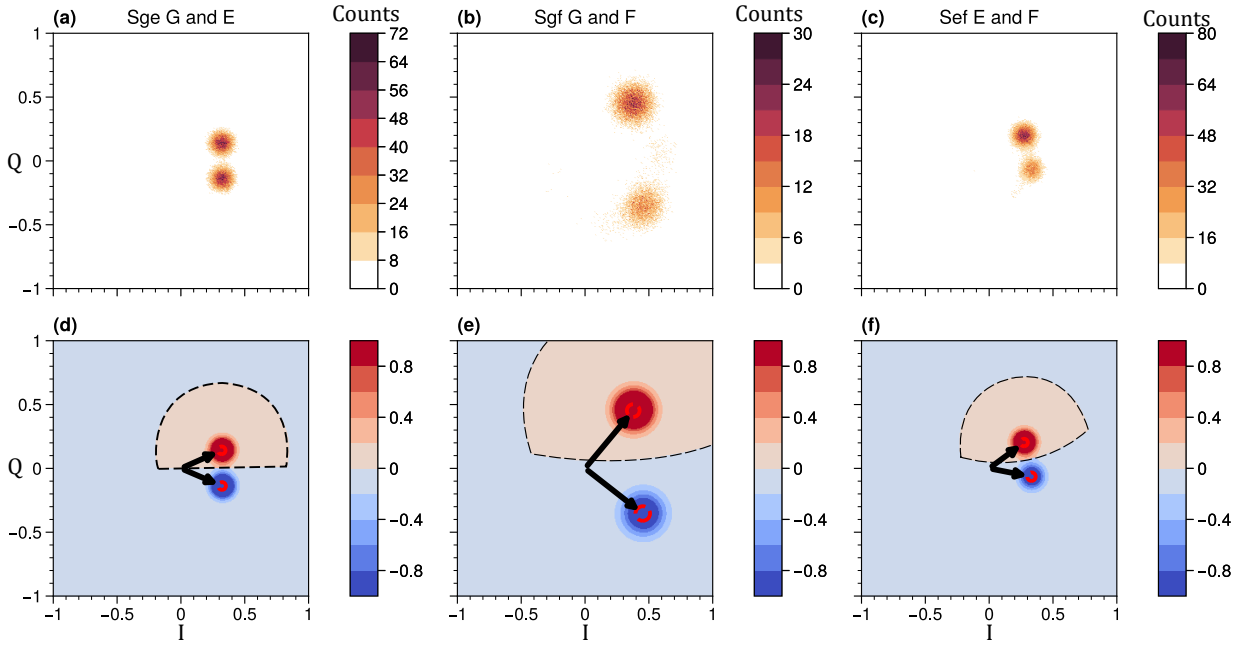


Figure 3.7: Readout histograms for 2-state data integrated with each respective weight function. (a)-(c) Integrated counts for pairs of states that correspond to the associated matched filter from Fig. 3.6(d)-(f). Importantly, the noise is symmetrically Gaussian, showing little distortion along any preferred axis. (d)-(f) Fitted Gaussian distributions for the state combinations in the subplots above, where regions of IQ space are colored to indicate both the state prediction (hue) and strength of the prediction (shading) on a normalized scale of -1 to 1. Separatrices where the state prediction changes in IQ space are indicated by black dotted lines.

record that lands on this contour is equally likely to herald one state or the other. In many cases, the distributions of the two states have different standard deviations, which generates the curvature we see in the separatrices. A simple version of this separatrix (the one that is normally calculated in FPGA implementations of these filters) is defined by the perpendicular bisector of the line segment defined by the centers of the distributions. This line is simple to calculate, but actually does not capture the curvature over the whole IQ space and so will result in small errors. This is one reason why the TPP is generally favorable over the majority vote filter, as it can capture this curvature more effectively in the weight function transformation.

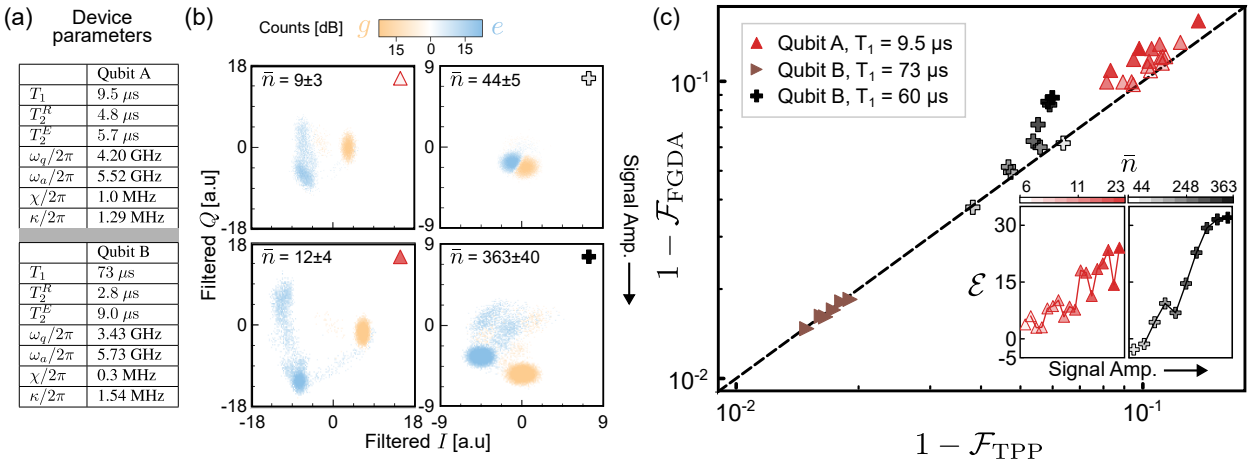
The last step of the majority-vote filter is to take a vote for each record and record both the individual matched filter fidelities and the total classification fidelity, including any null results. For the data shown in Figs. 3.6 and 3.7, the result reads as [G fidelity, E fidelity,

F fidelity, total fidelity] = [0.948, 0.918, 0.820, 0.895], with four null results over all 90k records. In general, we can see that this choice of readout parameters (a detuning that lands between G and E) favors G vs. E fidelity. Admittedly, these numbers aren't particularly good. This is likely due in part to the prohibition on pulse shaping, meaning that the T_1 limit for fidelity solely from integration time is somewhere around $(T_1 - \tau_{int})/T_1 \approx 97\%$ in this case. Additionally, this is just one subset from a fairly wide scan in both readout power and readout voltage which is displayed more completely in Ref. [24].

3.4 Applying Machine Learning: General Temporal Post Processor

To conclude this chapter, I would like to discuss a few key observations we made with the Türeci group in Ref. [24] and a few ideas on the future of quantum filtering. First, let's take a look at the performance of the machine-learned weight functions of the TPP in Ref. [24] versus the analytical matched filter (FGDA) also calculated in Ref. [24] for two-state discrimination via Fig.3.8. For context, the data I showed in Figs. 3.6 and 3.7 are members of the black triangle dataset from Qubit B. Portions of this section, Figures 3.8 and 3.9 particularly, are adapted from Ref. [24].

The first thing I would like to point out in Fig. 3.8 is that the gains in fidelity of the machine-learned weights in Ref. [24] compared to the optimal matched filter are fairly marginal, especially at high fidelity (where it is actually slightly worse). This is not necessarily a fair comparison, though, because the TPP is completely general to any noise condition and here it is being fed a data type that is mostly Gaussian noise and highly mean-separated. In any case, the gains are notably larger in the cases for large readout amplitude, as shown in the inset of Fig. 3.8(c), which suggests that high power readout may be introducing correlated noise sources that the TPP is picking up on and the matched (gaussian noise) FGDA filter is missing. This highlights the greatest benefit of the TPP, which is that it is capable of learning arbitrary noise conditions and adapting accordingly, which means that it can compensate for highly nontrivial changes in physical parameters (such as the coherence times of higher transmon states) during the sweep of an external parameter like readout power or



	Qubit A
T_1	9.5 μs
T_2^R	4.8 μs
T_2^E	5.7 μs
$\omega_q/2\pi$	4.20 GHz
$\omega_a/2\pi$	5.52 GHz
$\chi/2\pi$	1.0 MHz
$\kappa/2\pi$	1.29 MHz

	Qubit B
T_1	73 μs
T_2^R	2.8 μs
T_2^E	9.0 μs
$\omega_q/2\pi$	3.43 GHz
$\omega_a/2\pi$	5.73 GHz
$\chi/2\pi$	0.3 MHz
$\kappa/2\pi$	1.54 MHz

Figure 3.8: **Classification performance of TPP versus FGDA for readout of real qubits.** (a) Parameters of various dispersive qubit-cavity systems used for gathering readout data. Coherence measurements are subject to 10% variation over time. (b) Representative qubit readout histograms under boxcar filtering as a function of measurement signal amplitude. (c) Readout data for three dispersive qubit-cavity systems is analyzed and the resulting classification infidelities for binary ($C = 2$) state classification are plotted against each other. The dashed line marks $1-\mathcal{F}_{\text{FGDA}} = 1-\mathcal{F}_{\text{TPP}}$. For datasets with variable shading of markers (red and black), more opaque markers indicate stronger measurement tone amplitudes, with corresponding resonator photon number $\bar{n}$ indicated via colorbars. Inset: Percentage fewer errors $\mathcal{E}$ computed for indicated datasets with increasing input signal amplitude.

readout duration.

To emphasize that last point, we can see in Fig. 3.9(a) that at low readout power, even *across datasets with completely different qubits and readout equipment*, the learned weights match up well with the matched filter. Only at high readout power do they diverge in interesting ways. There is a lot of future work to be done on this subject, perhaps an entire processor could be characterized in this way to put these filters through their paces. Further, we could imagine utilizing the TPP-learned filters to discriminate states with zero mean difference, opening up interesting avenues to extract information from highly non-classical systems such as interferometric readout [32] and readout systems with in-situ parametric amplification. In general, our explorations into weight functions and optimal readout resulted in a much more general method of building weight functions that could prove extremely valuable in boosting readout fidelities above a fault-tolerant threshold in a highly non-classical system with correlated noise sources.

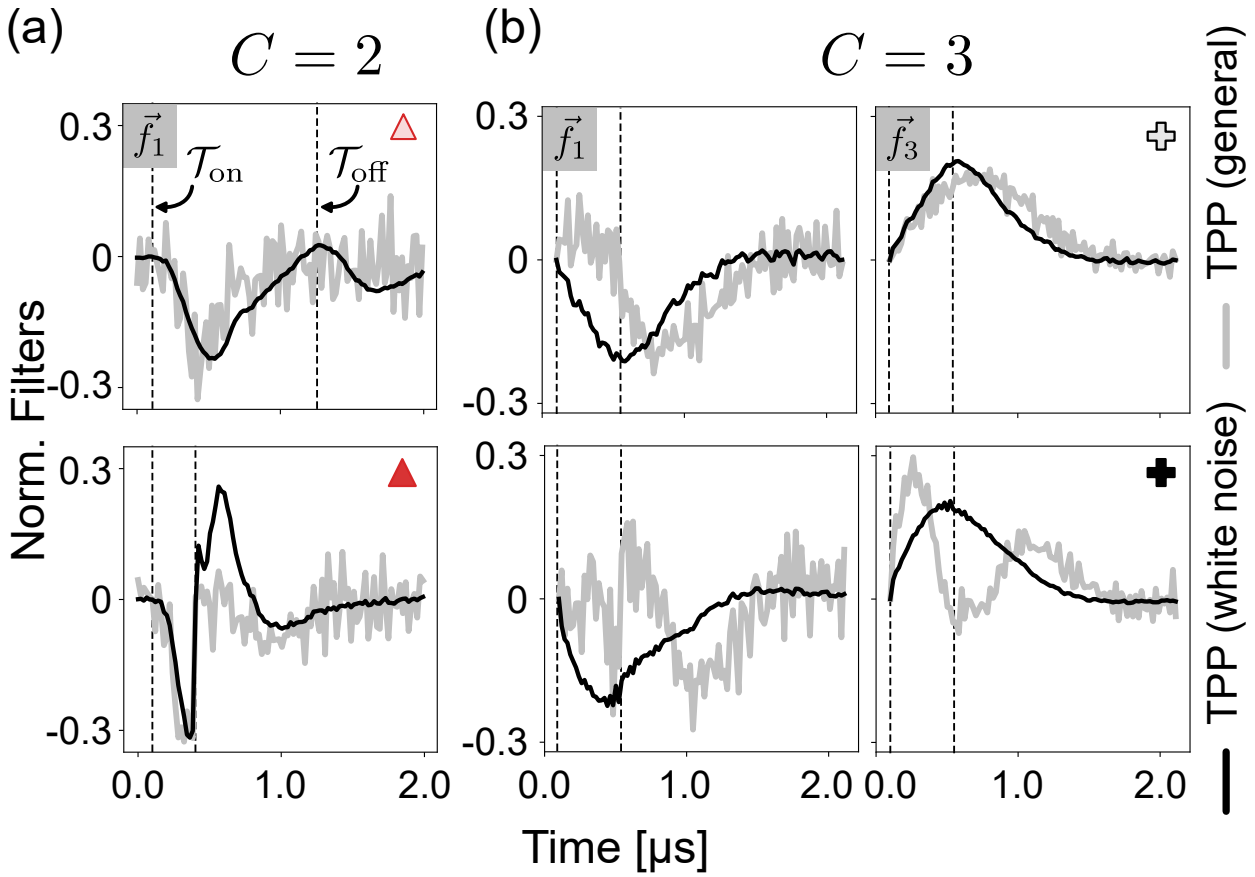


Figure 3.9: **Adaptation of TPP-learned filters with increasing measurement tone amplitude and evolving noise conditions.** Black curves are normalized TPP filters under the white noise assumption; for binary state classification, these are identical to standard matched filters. Gray curves are general TPP filters with no assumptions on noise statistics. (a) Filter $\vec{f}_1$ for binary ($C = 2$) classification, and (b) filters $\vec{f}_{1,3}$ for $C = 3$ state classification. In both cases, at weaker amplitudes the general TPP filter temporal profile closely matches that of the TPP filter assuming white noise. However, for stronger measurement amplitudes, a marked difference between the white noise TPP filter and the general TPP filter is observed.

4.0 Broadband Reflection Parametric Amplifiers

In this chapter, we will walk through the theory, design, and utility of coupled resonator networks in the context of parametric amplification. The end result will be a discussion of the simulation and design workflow for designing broadband parametric amplifiers, a summary of a design in progress for broadband parametric amplifiers with fractional bandwidths in excess of 15%, and finish with a discussion of results from a functioning parametric amplifier with 10% bandwidth.

My hope is to lead up to these topics in order of increasing complexity. First, we will discuss over, under, and critical coupling in only a single resonator to understand how light couples into and out of a resonant system. Building on this, we will dip into multi-coupling, multi-port systems and discuss extensions of the ideas of over and under-coupling, with a few rules of thumb on most concisely characterizing these systems. Next, we will bridge some gaps in vocabulary between these results and coupled-mode theory from chapter 1, which will prove a powerful tool in synthesizing broadband amplifiers using techniques from filter synthesis.

4.1 Networks of passively coupled resonators

4.1.1 Over, under, and critical coupling

The quantization example in chapter 1 gives a decent idea of why so much emphasis is placed on *resonant* circuits in cQED - they are easy to quantize and think about in terms of quantum optics. Because of their ubiquity, resonators and resonant modes in general have a nearly universal grip on most of the circuit design in superconducting quantum computing, and as a result it's extremely important to understand how they can couple to external feedlines.

In a circuit model of the mode, there is always a single number, the total quality factor

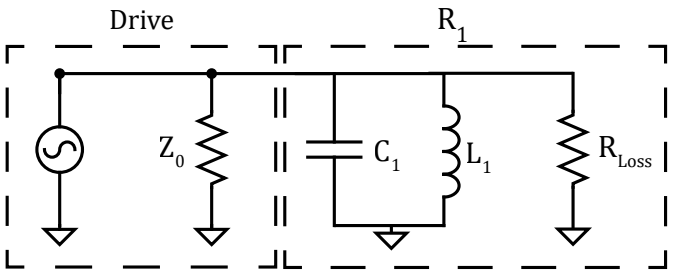


Figure 4.1: Single parallel LC resonator with internal loss R_{Loss}.

Q, that tells us the rate of *power* decay of the light in the resonator (the voltage decay is half as fast) via $Q_{total} = \frac{\omega_0}{\kappa_{total}}$. While this total number is useful, modes usually have multiple physically separate sources of loss that appear in parallel or in series electrically. This section will recap the parametrization of circuits in terms of quality factors $Q_k = \omega_0 R_k$ that can succinctly describe these losses attributed to individual loss sources represented by the resistors R_k or the external transmission line impedance Z_0, as well as develop an intuition behind how these ideas extend to larger coupled resonator circuits.

The canonical example of lossy resonance (in this chapter, an R without subscript represents a loss from an aribtrary non-port resistance in a circuit, while R_i represents the ith resonator in a circuit) is the parallel RLC circuit with a source impedance of the drive Z_0 shown in Fig. 4.1.[1]

The scattering parameter (just the reflection coefficient, because this is a one-port circuit) of the circuit in terms of the circuit elements is given by

$$S_{11} = -1 + \frac{2\omega\omega_1 Z_1 R_{loss}}{\omega\omega_1 Z_1 R_{loss} + i\left(\omega - \omega_1\right)\left(\omega + \omega_1\right) Z_0 R_{loss} + \omega\omega_1 Z_0 Z_1}. \tag{4.1}$$

We can use the characteristic impedance of the resonator $Z_1 = \sqrt{\frac{L_1}{C_1}}$ to describe the circuit, and I have already applied the transformation $L_1 \to \frac{Z_1}{\omega_1}$, and $C_1 \to \frac{1}{\omega_1 Z_1}$ to be able to discuss everything in terms of the impedance of the resonator. To remove a discussion of the specific values of the impedances and use their ratios, we use the quality factors defined by

[1] Although series RLC circuits are just as universal, they are described directly in Pozar [47], so I am not going to reproduce it here. The same ideas apply, though with an inverted impedance.

$$Z_0 \to Q_e Z_1, \text{ and } R_{loss} \to Q_i Z_1 \text{ to reach}$$
$$S_{11} = -1 - \frac{2i\omega\omega_1 Q_i}{\omega^2 Q_e Q_i - i\omega_1\omega\left(Q_e + Q_i\right) - \omega_1^2 Q_e Q_i}. \tag{4.2}$$

Here we can plot a few key results. On resonance, with $\omega = \omega_1$, we see that the scattering will generally be less than unit magnitude, with

$$|S_{11}(\omega_1)| = \left|1 - \frac{2Q_i}{Q_e + Q_i}\right|. \tag{4.3}$$

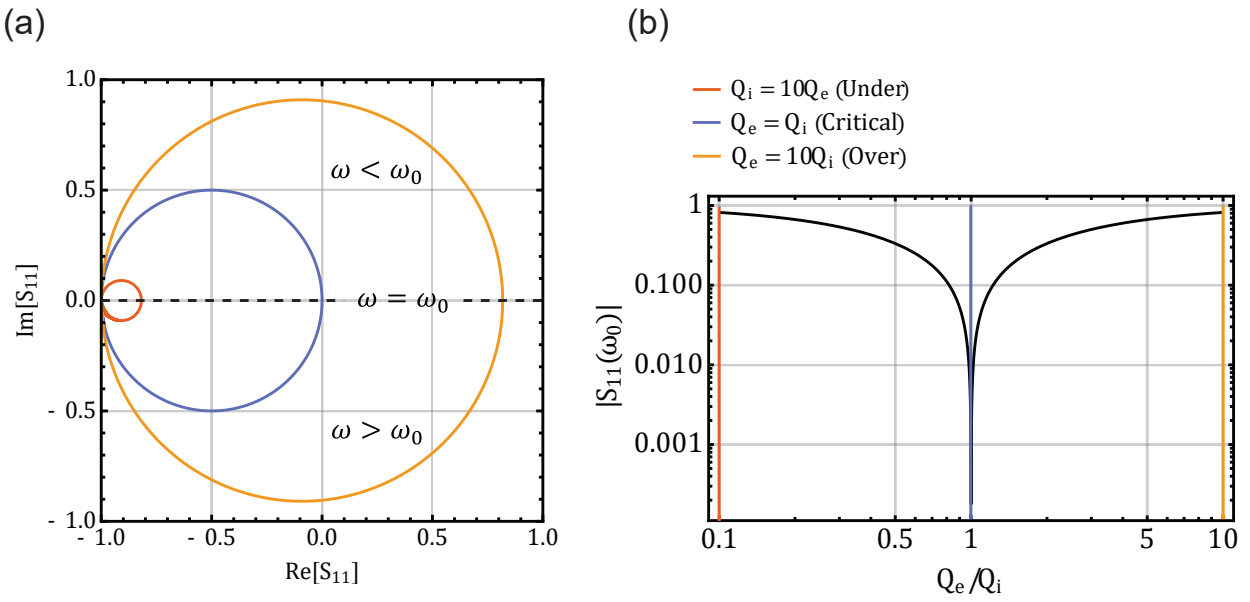


Figure 4.2: Depiction of the coupling regimes for the single parallel RLC resonator in Fig. 4.1. (a) Polar plot of the scattering parameters in Eq. 4.2 over all frequencies, with regions marked. On resonance, all the scattering lies on the real component, with magnitude determined by the balance of quality factors. (b) On-resonance scattering magnitude in Eq. 4.3 vs. quality factor ratio on a log-log scale, with line markers that correspond to the appropriate case in (a)

The ratio of quality factors determines the size of this dip. Illustrated by Fig. 4.2 (a), the scattering traces out a circle in the complex plane, with a real axis intersection on-resonance. These cases are clear-cut for a single resonator. If the quality factors are equal, there is no reflection on resonance. However, the case is initially less clear if we switch vocabulary. What if the internal resistance R represents not the internal loss mechanisms of the circuit, but instead another port? Consider the scattering of the circuit depicted in Fig. 4.3.

Now we have three more scattering parameters, but they are easy to compute. The reflection coefficient is exactly the same as before, with the caveat that we have now relabeled

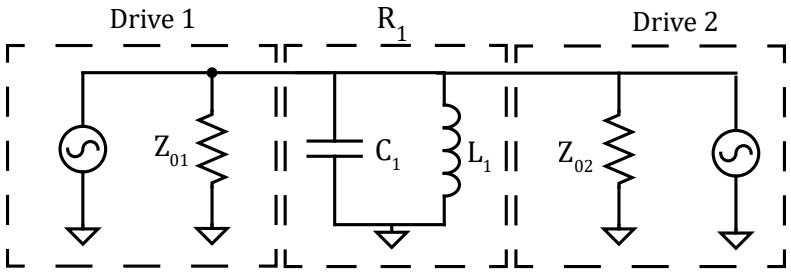


Figure 4.3: Single LC resonator driven from two ports of differing impedance. Z_{0i} represents the ith port's impedance.

$Z_0 \to Z_{01}$ and $R_{loss} \to Z_{02}$. This gives the exact same idea of quality factors, but now rather than discriminating between internal and external resistance, we can label the quality factors $Q_1 = Z_{01}/Z_1$ and $Q_2 = Z_{02}/Z_1$ where now the subscript of Q_i represents the port. To calculate the other parameters, We can apply the same results as before, but now the scattering on-resonance at $\omega = \omega_1$ is the matrix

$$S = \begin{bmatrix} \sqrt{\left(1 - \frac{2Q_2}{Q_1+Q_2}\right)^2} & \frac{2\sqrt{Q_1Q_2}}{Q_1+Q_2} \\ \frac{2\sqrt{Q_1Q_2}}{Q_1+Q_2} & \sqrt{\left(1 - \frac{2Q_1}{Q_2+Q_1}\right)^2} \end{bmatrix}. \tag{4.4}$$

Here I have calculated the scattering based on the ABCD matrix method, which is covered in more detail in section 4.1.2. One important note is that this method is exactly equivalent to using source and sink impedances of the same magnitude Z_0, but applying impedance inverters K_1 and K_2 on either side of the resonator such that

$$Z_{0,i} = \frac{K_i^2}{Z_0} \implies Q_i = \frac{K_i^2}{Z_0 Z_1} \implies \frac{Q_i}{Q_j} = \frac{K_i^2}{K_j^2}. \tag{4.5}$$

This makes it easy to calculate coupling ratios based on component values, and vice-versa. For example, take a capacitive impedance inverter (the most common in our field) on either side, with inverter constant on-resonance of $K_{c,i} = \frac{1}{\omega_1 C_i}$, then the coupling ratio is simply

$$\frac{Q_1}{Q_2} = \frac{C_{c2}^2}{C_{c1}^2}. \tag{4.6}$$

Considering Eq. 4.4, the diagonal scattering parameters are exactly analogous to Fig. 4.2(a), where the internal loss has simply been replaced by the other port. However, we can also take a look at the off-diagonal scattering in Fig. 4.4. In Fig. 4.4(a) we can see a similar pattern as in the reflection case in Fig. 4.2(a), where the scattering traces out a circle in the complex plane. The major difference here is that the center ($S_{11} = 0$) is where the divergences lie ($\omega \to 0$ from above or ∞ from below), vs. $S_{11} = -1$ in the reflection case.

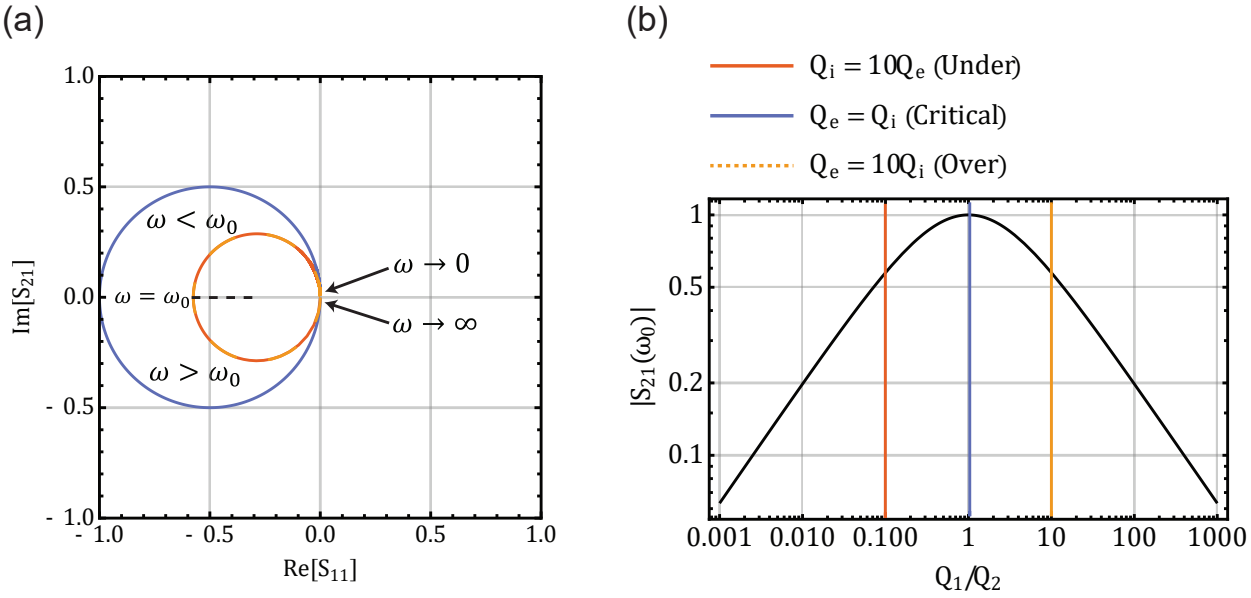


Figure 4.4: Two-port, one parallel resonator transmission in terms of port quality factors. (a) Transmission in terms of port quality factor ratios, extending the idea of over and under coupling to multi-port systems. Unlike reflection, transmission knows no difference between over and under-coupling. (b) Transmission on-resonance as a function of quality factor ratio, with lines marking the accompanying traces in (a).

To make this language of quality factors universal, we have to combine the previous two exercises in Fig. 4.1 and Fig. 4.3 to include two ports with an additional source of internal "loss," which can equally well be other terminated (but undriven/unmonitored) ports. With an equivalent parallel loss resistance R in parallel (shown in Fig. 4.5) with the resonator that translates to $Q_i = R_{loss}/Z_1$, we get the scattering matrix

$$\begin{bmatrix} 1+\frac{2iQ_2\omega\omega_1Q_i}{Q_1Q_2\omega^2Q_i-i\omega_1\omega(Q_2Q_i+Q_1(Q_i+Q_2))-Q_1Q_2\omega_1^2Q_i} & \frac{2i\sqrt{Q_1Q_2}\omega\omega_1Q_i}{Q_1Q_2\omega^2Q_i-i\omega_1\omega(Q_2Q_i+Q_1(Q_i+Q_2))-Q_1Q_2\omega_1^2Q_i} \\ \frac{2i\sqrt{Q_1Q_2}\omega\omega_1Q_i}{Q_1Q_2\omega^2Q_i-i\omega_1\omega(Q_2Q_i+Q_1(Q_i+Q_2))-Q_1Q_2\omega_1^2Q_i} & 1+\frac{2iQ_1\omega\omega_1Q_i}{Q_1Q_2\omega^2Q_i-i\omega_1\omega(Q_2Q_i+Q_1(Q_i+Q_2))-Q_1Q_2\omega_1^2Q_i} \end{bmatrix}. \tag{4.7}$$

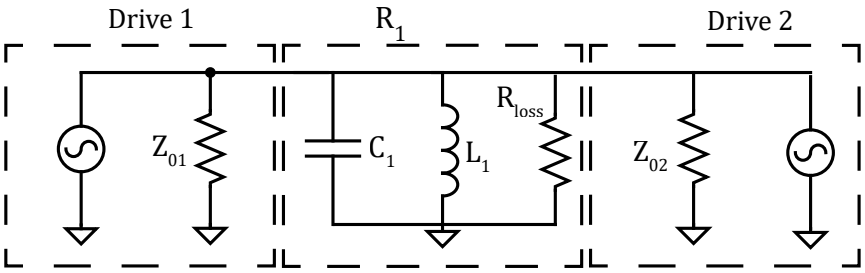


Figure 4.5: Two-port one-resonator system with internal loss R_{loss}.

This scattering matrix, while unwieldy, gives the ability to answer the question of "Where on earth is my light going?" by using pairwise measurements of a multi-port system. If we measure two ports and they *both* seem undercoupled, then the light must be *overcoupled* to another source of loss which we could model as another resistance R_{loss} in the system. If you have no more ports to check, these equations allow you to quantify the magnitude of the internal loss. However, if you do have more ports to check, you can continue this process. In the case of one extra third port, it can be connected and measured instead of the second. If you find that the light is overcoupled to that port, then you can measure $|S_{11}|^2 + |S_{22}|^2 + |S_{33}|^2$ to quantify the internal loss, and if that quantity is near one, rest easy knowing that the overcoupled port will be where most of the light goes at that frequency, no matter which other port is driven. This allows for easy, algorithmic characterization of high Q, multi-port resonator systems.

Concluding this section, port couplings are easily quantifiable in terms of quality factors, and able to completely characterize the system. Next we move to describing highly coupled resonator circuits using the matrix formalism developed in Ch.1 and in Ref. [41].

4.1.2 Representing coupled electrical resonators in coupled-mode theory

Since coupled-mode theory describes the scattering between resonators in the rotating frame, we would hope that there would be a simple correspondence between its description of scattering and the description one would get from calculating the scattering matrix just from electrical network theory in Ref.[47]. Let's see if this works out. Consider the network shown in Fig. 4.6 which shows two resonators linked by an impedance inverter K. The goal

of this section is to show a direct link between these circuit values and the values in the M-matrix in Eq. 1.54

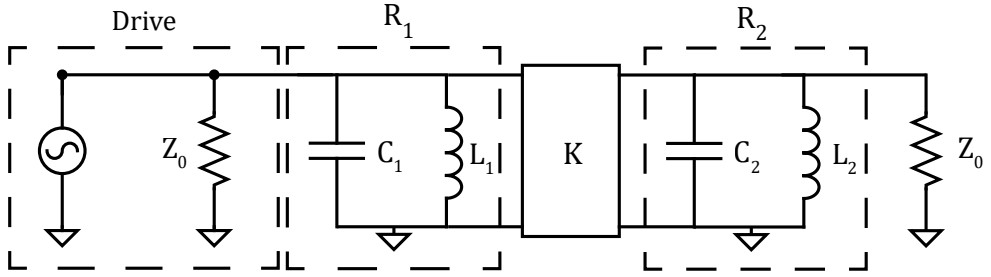


Figure 4.6: Diagram of two resonators coupled by an impedance inverter with inverter value K.

$$M_{cpl} = \begin{bmatrix} \Delta_1 & \beta_{12} \\ \beta_{12} & \Delta_2 \end{bmatrix}. \tag{4.8}$$

via the scattering matrix calculated via Eq. 1.54.

After reaching this link, we will utilize the relationship between the two to discuss generalizations of over, under, and critical coupling in the coupled-mode theory formalism. We will accomplish this by calculating the scattering matrix through the network in Fig. 4.6 by first concatenating ABCD matrices, and then transforming the result to the scattering matrix. To begin, we can use the ABCD representation of a general shunt impedance and the ABCD representation of an inverter given by [47]

$$T_{shunt} = \begin{bmatrix} 1 & 0 \\ \frac{1}{Z} & 1 \end{bmatrix}, \text{ and } T_{inv} = \begin{bmatrix} 1 & iK \\ \frac{i}{K} & 1 \end{bmatrix}. \tag{4.9}$$

After calculating the ith resonator ABCD matrices T_{Ri} from the parallel impedances of L_i and C_i, we can calculate the total ABCD matrix result simply as $T_{total} = T_{R1}T_{inv}T_{R2}$. Following [47] we can convert this matrix to a scattering matrix by specifying the impedance

of the port Z_0. I have reproduced the transform here for ease of reference.

$$\begin{aligned}
\Delta &= A + B/Z_0 + CZ_0 + D \\
S_{11} &= \frac{A + B/Z_0 - CZ_0 - D}{\Delta} \\
S_{12} &= \frac{2(AD - BC)}{\Delta} \\
S_{21} &= \frac{2}{\Delta} \\
S_{22} &= \frac{-A + B/Z_0 - CZ_0 + D}{\Delta}
\end{aligned} \tag{4.10}$$

In order to directly compare the two equations we use the variable changes to the characteristic impedance of the resonators Z_j, followed by a change to their decay rates and frequencies γ_j and ω_j.

$$L_j = \frac{Z_j}{\omega_j}, \text{ and } Z_j = Z_0 \frac{\gamma_j}{\omega_j} \tag{4.11}$$

We can focus on any scattering parameters because the entire matrix must be equal component-wise. Since S_{21} has a slightly simpler form, we'll start there. Computing the results from the ABCD parameters and Eq. 4.10 and using the substitutions from Eq. 4.11 we arrive at

$$S_{21} = \frac{2i\gamma_1\gamma_2 k Z_0}{\gamma_1\gamma_2 Z_0^2 - \frac{k^2}{\omega^2}\left(-i\gamma_1\omega + \omega^2 - \omega_1^2\right)\left(-i\gamma_2\omega + \omega^2 - \omega_2^2\right)}. \tag{4.12}$$

To compare the matrix coupled-mode calculation with a real β_{12} yields

$$S_{21} = \frac{4i\beta_{12}\gamma_1\gamma_2}{4\beta_{12}^2\gamma_1\gamma_2 + \left(\gamma_1 - 2i\left(\omega - \omega_1\right)\right)\left(\gamma_2 - 2i\left(\omega - \omega_2\right)\right)}. \tag{4.13}$$

To relate between the two, we can start by moving to an on-resonance degenerate case, where $\omega = \omega_1 = \omega_2$. This gives the equality

$$\frac{4i\beta_{12}}{4\beta_{12}^2 + 1} = \frac{2iKZ_0}{K^2 + Z_0^2}, \text{ requiring } \beta_{12} = \frac{Z_0}{2K}. \tag{4.14}$$

Eq. 4.14 gives a satisfying direct link between the coupling rate (which we can see experimentally with g_2 splitting) and the inverter value which we will make extensive use of in filter synthesis. Plugging this result into Eq. 4.13, we see that we are not quite out of

the woods yet, there is still an important functional difference between the two scattering equations.

$$S_{21} = \frac{2i\gamma_1\gamma_2 k Z_0}{\gamma_1\gamma_2 Z_0^2 + k^2\left(\gamma_1 - 2i\left(\omega - \omega_1\right)\right)\left(\gamma_2 - 2i\left(\omega - \omega_2\right)\right)} \tag{4.15}$$

Comparing Eq. 4.15 to Eq. 4.12, and again making the equations degenerate with $\omega_1 = \omega_2 = \omega_0$ we reach an interesting impasse in the equality

$$\left(\gamma_1 - 2i\left(\omega_{cm} - \omega_0\right)\right)\left(\gamma_2 - 2i\left(\omega_{cm} - \omega_0\right)\right) = \frac{1}{\omega^2}\left(-i\gamma_0\omega + \omega^2 - \omega_0^2\right)\left(-i\gamma_2\omega + \omega^2 - \omega_0^2\right). \tag{4.16}$$

As I was working through this, I anticipated that this would be an easy identity, maybe with just a rescaling of the frequency, but this is not the case! Without relabeling $\omega \to \omega_{cm}$ on the LHS of Eq. 4.16, there would be no way to force an equality with a frequency-independent β_{12}. Fig. 4.7 shows a graphical representation of the problem, where the symmetry point is completely different between the coupled-mode and the electrical scattering for low quality factors. While the difference between the coupled-mode and electrical response is largest at low quality factor, the real circuit is always symmetric around DC, or 0 frequency. This is a necessity for any electrical circuit, and follows from the requirement that the time-domain voltage response must be real-valued [47]. Another consequence is that the real component of the input impedance must be even, while the imaginary component must be odd.

Luckily, there is a way to rectify the issue, which is to move the symmetry point of the coupled-mode response to DC. This is accomplished by the transform

$$\omega_{cm} \to \omega_0 - \frac{\omega_0}{2}\left(\frac{\omega_0}{\omega} - \frac{\omega}{\omega_0}\right) \tag{4.17}$$

The first term of Eq. 4.17 moves the response to the frame of the resonators (symmetry around $\omega = 0$), then applies a low-pass to bandpass transformation to move the response back to the correct frequency (ω_0) but while maintaining the symmetry around DC. This moves the scattering parameters into exact equivalence with the electrical circuit, emphasized by the dashed overlay in Fig. 4.7. The transform also readily extends to longer coupled resonator and coupled-mode networks. In the process of editing this thesis, I also discovered this same

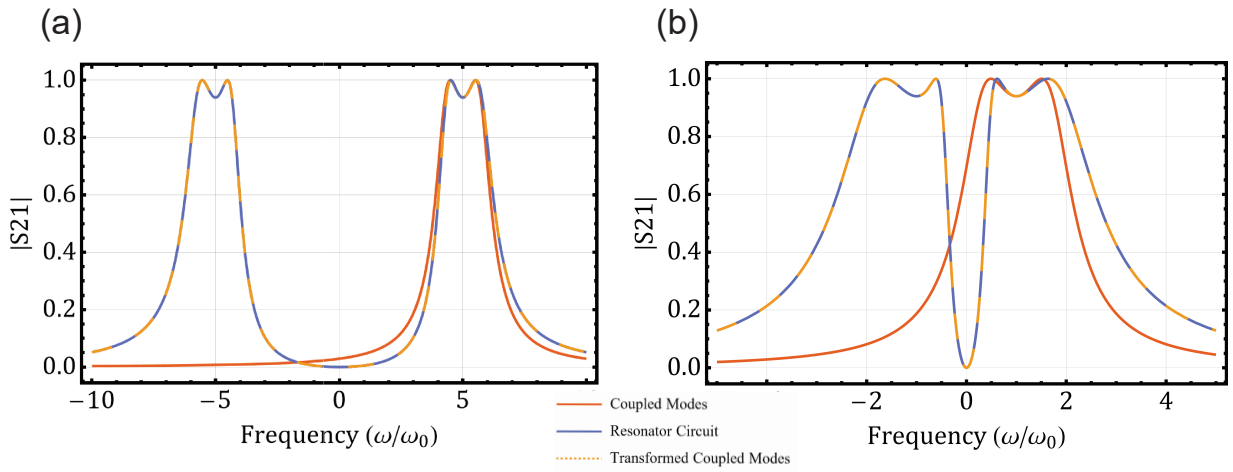


Figure 4.7: Comparing accuracy of coupled-mode description of scattering vs. quality factor in the degenerate ($\omega_1 = \omega_2 = \omega_0$) case. In both the high and low Q cases, the coupled-mode description is always symmetric about ω_0. (a) At reasonable quality factor ($Q = 5$ in this case) the coupled mode approach in orange is a fairly good approximation at zero detuning. (b) For very low quality factor ($Q = 1$ in this case), the differences between the resonator scattering and the coupled-mode description become more drastic, and the approximation only holds on-resonance.

transformed result present in Ref. [53]. Interestingly, they state that this result comes from leaving out the RWA in the process of simplifying the coupled QLEs.

In practice, this transformation is only needed to fit to very low Q experimental data. The design process of a coupled-resonator network is the same regardless of the transform in Eq. 4.17, and in fact the transform preserves all of the needed results of gain, bandwidth, etc. of a parametric amplifier designed with the methods in Ref. [41]. Accordingly, we will work in the untransformed coupled-mode theory whenever possible, focusing instead on the result in Eq. 4.14 to discuss the couplings within a multi-resonator circuit and how they translate to various types of filters.

4.2 What is a filter?

Let's analyze a particular case of coupled resonators with β_{12} near 0.5, equal decay rates $\gamma_0 = \gamma_1 = \gamma_2$ and no internal loss. Equally valid would be to say that $K = Z_0$, and $Z_1 = Z_2$, with both ports having the same source impedance Z_0, or $Q_1 = Q_2$ with $K = Z_0$. All of

these various forms of vocabulary say the same thing for the scattering. The S_{21} and S_{11} of this network versus a normalized detuning are shown in Fig. 4.8.

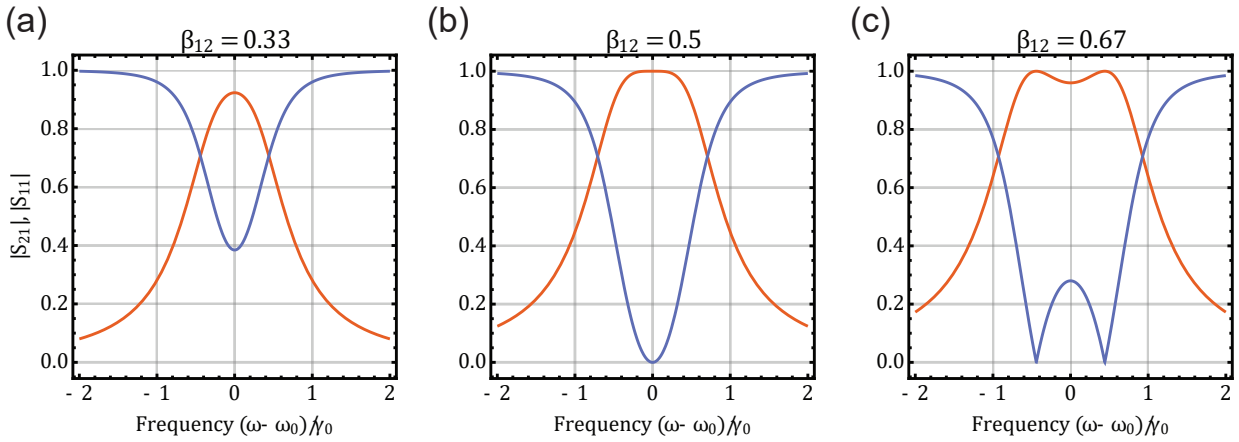


Figure 4.8: Filter scattering generated by a varying coupling strength. (a) With an under-tuned coupling between the resonators, the light is not efficiently coupled from one port to the other, creating a structure that is not a traditional filter, but shares more in common with the overcoupled case in the single-port resonator. (b) A critically tuned coupling between the resonators, allowing for a maximally flat reflection curve near resonance. (c) An over-tuned coupling between the resonators, which allows them to exchange light *faster* than each resonator's decay rate to its respective port.

Figure 4.8 shows a range of couplings, which is instructive to analyze in the two resonator case, as we don't yet need the sledgehammer of mathematics and filter design in Ref. [41] to understand the scattering. In the case of Fig. 4.8(a), the coupling rate between the resonators is slower than half of both the decay rates to the resonators' respective ports. Conceptually, there is no splitting in the response because the light either decays through the port it came from, or transfers to the other resonator, where it decays much faster than it would have transferred back to the the resonator it came from. In other words, there is no time for the light to cycle back to one resonator before it decays. This isn't so useful as a filter, because it reflects significantly, and so doesn't effectively transfer energy from one port to another.

In case (b), we have an interesting scenario. Here, the coupling rate is just as fast as both the decay rates, allowing the resonators to not only perfectly match the light at $\delta = 0$, but to also cooperate for a significant bandwidth $(1/\sqrt{2\gamma_1\gamma_2})$ around $\delta = 0$. In other words, the splitting is half the decay rate, and so where one resonator begins to fail, the other picks up the light and passes it along like a bucket brigade. They cooperate in such a way as to ensure that the leading term of the reflection around $\delta = 0$ is proportional to δ^2, the first-order slope has been *canceled*. This is the definition of a Butterworth filter, one of the

canonical examples of filter design. In other words, critical coupling between two degenerate resonators is synonymous with a 2nd-order Butterworth filter!

Case (c) rounds out the utility of this intuition, which is that as you increase the coupling, the resonators split, and the transmission response becomes multi-pole. This splitting is visible in the response because the light has time to bounce back and forth *between* the resonators before it must decay *out* of them. Because of this bouncing, the matching isn't perfect at the origin. However, it often doesn't *need* to be perfect, and the FWHM bandwidth of the filter clearly benefits from this more-than-critical coupling. While I will not show it here in mathematical detail, the maximum ripple in the response defines the same ripple as another canonical filter: the Chebyshev filter.

This game of relative coupling rates versus decay rates gives an intuitive picture of what is happening inside of a filter. In a Butterworth filter, each element cooperates perfectly so as to cancel higher and higher orders of the Taylor expansion of the reflection at the origin. In this sense, the design space is a point. Only one set of device parameters will produce this ideal maximally flat response at the origin. In a Chebyshev filter, the light can slosh around, bouncing off of multiple resonators before ultimately decaying. This sloshing generates ripples and can generate wider bandwidth at the cost of some ripple in the multi-pole response.

Now that we have developed an intuition for how filters work using coupled mode theory, we can lean on the well-developed math of filter synthesis to make any filter that we want. The difficult part of this work was done in Ref. [41], and so I will focus on a few illustrative examples in the next section.

4.3 Impedance-matching a degenerate parametric amplifier

The standard use case of filters discussed in the previous section permits light to pass at some frequencies, but reflect or absorb (a case not covered here) it at all others. In this thesis, however, we will take advantage of another use case: impedance transformation. To do this, we'll first discuss the amplifier impedance we need to transform and the benefits of

doing so.

4.3.1 Impedance of an active parametric amplifier

The gain profile of a parametric amplifier derived in Eq. 1.58 can be described in a simple way by comparison with the standard form of the reflection coefficient of a one-port system given by

$$\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0}. \tag{4.18}$$

Matching these two up, we can imagine that in order to reach a large reflection gain ($|\Gamma| >> 1$) we need ian in input impedance Z_in that approaches $-Z_0$. Within this picture, we get an expression for the input impedance of a parametric amplifier that is actively pumped (also given in Ref. [41]), plotted in Fig. 4.9.

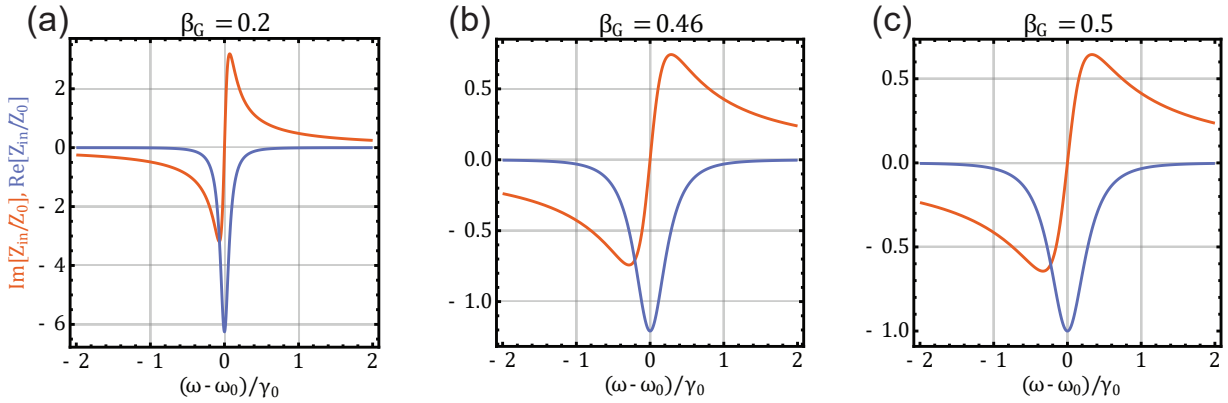


Figure 4.9: Active amplifier impedance vs signal to idler coupling strength. (a) Below critical gain, the imaginary component of the impedance slope is very sharp, and the real component is large and negative. (b) Very near critical gain, here pictured at about 20 dB, the imaginary impedance takes on a shallower slope, and the real component is near $-Z_0$). (c) At critical gain, the imaginary component of the slope is nearly the same, and the real component is exactly $-Z_0$, causing a divergence in Eq. 4.18.

Using the results from Eq. 1.55, we can calculate the input impedance via inverting Eq. 4.18. Carrying this out in Fig. 4.9, we can see that the slope of the imaginary impedance at zero detuning opens a "window" into the real (negative) input impedance of the circuit to enable a large gain. This gives a clear avenue towards broadening the gain response - widening the window.

Our task is now set, we have to engineer a network that cancels the slope of the impedance, shown in Fig. 4.9(b), for 20dB gain over a certain bandwidth. This will create a structure that ensures the reflection gain maintains a certain value over a specified fractional bandwidth.

4.3.2 Power insertion loss method and continued fraction expansion

As stated in the last section, we want to create a linear network with roughly zero imaginary impedance over a wide window that will define our fractional bandwidth. This process is well-developed in electrical engineering for filter synthesis, and there is a formal link between this method and coupled mode theory established first in Ref. [41]. This subsection will walk through the power insertion loss method used in that publication for calculating filter prototypes, to determine the coupling rates between resonators in a filter and explain the motivation behind each step to build the tools needed to compare different types of filters in the next subsection. Portions of this analysis are pulled from Ref.[22], where I used it to derive the filter prototype coefficients for a matched parametric amplifier that uses Legendre polynomials. That publication, in turn, uses the methods outlined in Ref [41].

First I will give a very brief introduction to the mathematical ideas we will need. As physicists, we are very familiar with how to approximate a function around a point with a Taylor series. A more broad problem that we need to borrow some math from for this task is the approximation of a function in a *window* because we need the imaginary component of the input impedance to be nearly zero over the entire bandwidth! To do this we need an orthogonal set of basis functions with some nice properties, similar to how we would break down something like a square-well problem.

How might we do this? Let's begin with normalizing the bandwidth with a new frequency variable $s = 2i\frac{\omega-\omega_0}{\Delta\omega}$, where $\Delta\omega = \omega_c - \omega_0$ defines a cutoff frequency ω_c, the point after which we don't need our approximation to hold any longer. For further simplicity, let's parameterize by ω/ω_c so that the cutoff occurs at 1, and so that we reduce our question to approximating a function in the domain $[-1, 1]$.

Luckily, there are a plethora of helpful functions at a variety of maximum orders n in ω

that we can use to accomplish this task. Among them are the binomial functions, $(\frac{\omega}{\omega_c})^{2n}$, the Chebyshev polynomials $T_N(\frac{\omega}{\omega_c})$, the Legendre Polynomials $L_n(\frac{\omega}{\omega_c})$, and so on. While my contribution to this technique focuses on the Legendre polynomials, and the provided example mathematica notebooks as well as [22] focused on their advantages, I will mostly use the Chebyshev polynomials as an example in any plots and diagrams in the walkthrough of this method.

Ideally, we would create a function that perfectly cancels the imaginary impedance in Fig. 4.9. Perfect cancellation over the entire target bandwidth is impossible with only reactive elements. However, we can get close with many orders of approximation within the window. However, many orders in the approximation would require many circuit elements to synthesize, and that is usually something we would like to avoid because it is generally quite hard to perfectly simulate, layout, and fabricate many (10's) of circuit elements without significant compounding error.

While we might be able to imagine creating an amplifier with 20 dB gain in a brick-wall style shape with a thirty resonator network, could we instead do it with only three within some tolerance, say 1 dB? The power insertion loss method for filter design allows us to analyze this task by using a particular one of these well-suited functions for approximating a function in a window to model the ideal behavior within some degree of tolerance, while only requiring n resonant elements to do so. In other words, the power insertion loss method allows the designer to quantify the tradeoffs in the filter circuit while minimizing the overall number of circuit elements.

Let's walk through an example with some commentary to see how filter synthesis math helps us generate a customized gain profile of an amplifier. The power-insertion-loss function itself is given by

$$\mathrm{PIL} = A(1 + k^2 \mathrm{F}(n, \omega/\omega_c)) \tag{4.19}$$

where $\mathrm{F}(n, \omega/\omega_c)$ can be any order n of any one of the families of the helpful functions mentioned previously, and we have introduced two degrees of freedom k and A. This function relates to the reflection coefficient with

$$\Gamma^2\left(\frac{\omega}{\omega_c}\right) = \frac{\text{PIL}-1}{\text{PIL}}. \tag{4.20}$$

Now that we have this relationship to the reflection, we can implement the constraints that will eliminate our degrees of freedom. In the case of an amplifier, we specify that the reflection in the middle of the band is related to the power gain G by

$$\Gamma^2(0) = \frac{1}{4G-2}. \tag{4.21}$$

Here we need to address a few immediate questions. First, why is the designed reflection inversely proportional to G? The reason is that we are building the network to match to a certain *magnitude* of real impedance. A positive impedance of the same value as a negative will flip the sign of the gain, making it look like a match instead. Traditional filter synthesis easily deals with positive resistances, not so much negative, so we favor positive resistance for now. Second, why is the expression $4G-2$ instead of simply G? The answer is given in Ref [41]: this is the phase-sensitive gain where the signal and idler of the amplifier are correlated. The 4 is the usual 6 dB increase for phase-sensitive power gain vs phase-preserving power gain.

With those addressed, we move to the second constraint, which is that we define a bandwidth by constraining the value of Γ^2 at $\omega = \omega_c$ with

$$\Gamma^2(1) = \frac{1}{4G/\xi - 2} \tag{4.22}$$

Where I have introduced the ripple parameter ξ, which is usually given in some small number of dB. These two constraints fully solve the problem and determine A along with k. Generally, the solution is given by

$$A = \frac{4G-2\xi}{\left(F(1)^2k^2+1\right)(4G-3\xi)}, \text{ and } k = 2\sqrt{\frac{G\left(\frac{1}{\xi}-1\right)}{F(0)^2(4G-3)\left(\frac{4G}{\xi}-2\right) - \frac{2F(1)^2(2G-1)(4G-3\xi)}{\xi}}}. \tag{4.23}$$

Specific cases are easy to obtain by switching out the polynomials for our matching method of choice, as we have already done for Legendre polynomials in Ref. [22]. After

one substitutes in a particular order of a particular family of polynomial, we arrive at a completely numerical function of two polynomials for the reflection coefficient, and we will make the switch to a complex variable $\omega/\omega_c \to is$ for the next step with

$$\Gamma^2 = \frac{r_n s^{2n} + r_{n-1} s^{2(n-1)} + ... + r_0}{d_n s^{2n} + d_{n-1} s^{2(n-1)} + ... + d_0} \tag{4.24}$$

Where I am emphasizing the fact that the polynomials are fully determined at this point by using the coefficients $\{r_n\}$ and $\{d_n\}$ to represent the positive coefficients of the numerator and denominator polynomials, respectively. All these coefficients must be positive, and this forms a special kind of polynomial known as a *Hurwitz* polynomial. These polynomials are guaranteed to have roots that are either real-valued, or appear in conjugate pairs.

Now the procedure shifts towards the end goal: synthesizing the *impedance* that can realize this reflection. Here we have an issue. The impedance is related to Γ, not Γ^2. Since Hurwitz polynomials always have roots in conjugate pairs of the form $\pm\sigma_i \pm i\omega_i$, we can choose to construct Γ by choosing the poles and zeros for which $\sigma_i < 0$. The main reason we do this is to ensure that the Laplace transform that we took in chapter 1 converges along the imaginary axis of the complex plane, or in other words that there is a valid Fourier transform for the system. The convergence of that transform also guarantees a type of stability known as bounded-input bounded-output (BIBO) stability [47].

Now that we have constructed Γ by selecting the poles p_i and zeros z_i with $\sigma_i < 0$ from the pole-zero decomposition of Γ^2, we can construct the impedance of the network with

$$\Gamma = \frac{\prod_i (s - z_i)}{\prod_i (s - p_i)} := \frac{R}{D} \implies Z = \frac{R + D}{D - R} \tag{4.25}$$

The last remaining step is to translate this impedance into a physically realizable network. For our simple topology, the heavy lifting is done by Cauer ladder synthesis, which expresses the input admittance of arbitrary-sized chains of inverter-coupled (represented by J_{jk}) resonators (represented by their characteristic impedances $Z_j = \sqrt{L_j/C_j}$ with the continued fraction expansion

$$\frac{Y_{in}}{Y_0} = \left(g_0 + \frac{1}{\frac{1}{g_1 s} + \frac{1}{\frac{1}{g_2 s} + \frac{1}{\frac{1}{g_3 s} + \ldots}}} \right). \tag{4.26}$$

This is handled in Ref. [41] with the equations reproduced here for ease of reference as

$$\begin{cases} J_{01} = \sqrt{\frac{\omega_0}{\Delta\omega} \frac{1}{Z_0 Z_1 g_0 g_1}} \\ J_{i,i+1} = \frac{\omega_0}{\Delta\omega} \sqrt{\frac{1}{Z_i Z_{i+1} g_i g_{i+1}}} \text{ for } 0 < i < N-1 \\ J_{N,N+1} = \sqrt{\frac{\omega_0}{\Delta\omega} \frac{1}{Z_N Z_{N+1} g_N g_{N+1}}} \end{cases} \tag{4.27}$$

With this expansion of the impedance/admittance we can make a connection to the physical circuit resonator impedances and inverters, we just have to get the expression in Eq. 4.25 into this form. This is doable by polynomial long division, which the numpy python package is particularly adept at. An updated version of the code in Ref. [41] with example usage for a 3rd order 1 dB ripple Legendre filter with 15% bandwidth is given below.[2]

This will directly produce the g_n that we need to continue the synthesis from the poles and zeroes of the impedance function. Now that we have the tools from Ref [41] with a little bit of extra explanation, here I will give a few examples of different prototype circuits, and why we might choose one over the other. A general diagram of a 3-pole matched parametric amplifier is given in Fig. 4.10 as it would appear in a physical circuit.

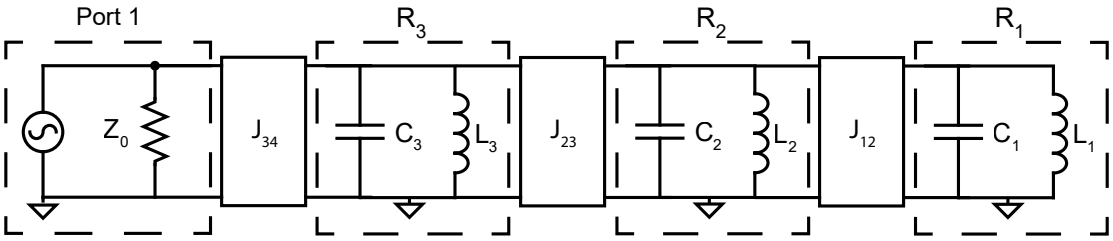


Figure 4.10: 3-pole matched amplifier circuit schematic, as it would appear physically.

The structure is very similar to what we have discussed earlier in the chapter with coupled resonators, except we have slightly more freedom in the three-resonator system. The inverters are calculated in a similar way to Eq. 4.14, but there are a few extra steps to reach it from Eq.65 in Ref. [41], reproduced here using only impedances in

[2]The old code doesn't work anymore since python went from 3.10 to 3.11, and numpy changed polynomial methods in 2.0. I have provided a replacement in Appendix A.

$$\beta_{jk} = \frac{\omega_0}{2\gamma_0}\frac{\sqrt{Z_j Z_k}}{K_{jk}}, \tag{4.28}$$

where I have replaced the admittance inverter J_{jk} with the equivalent impedance inverter $1/K_{jk}$.

Finally, for actual synthesis the inverter values $K_{jk} = 1/J_{jk}$ are calculated from Eq. 4.27. With the special case of only two resonators, we recover Eq. 4.14. The major conceptual reason why you would expect there to be a difference is that here we have multiple resonators that do *not* have their own port couplings! Instead, they have to inherit their losses from other resonators. We would already have needed this before if we omitted the second port in the previous section.

4.3.3 Prototype circuits and tradeoffs

Now that we have the ability to calculate any Chebyshev, Legendre, or Butterworth prototype that we want from the procedure given in the last section, let's compare these networks head-to-head in Fig. 4.11.

Figure 4.11(a) compares the gain profile of the three networks, where we can clearly see that the Butterworth network is maximally flat, as designed, but the Chebyshev and Legendre prototypes have some characteristic ripple. Off the bat, we have to acknowledge that the Legendre filters have a slightly different definition to the ripple parameter, because their value at $\omega = \omega_0$ changes slightly with the filter order. Practically speaking it's not of much relevance here, so I leave the detail in Ref. [22] but suffice it to say we have to be careful to compare the Legendre filters graphically with the Chebyshev filters to ensure their ripples are actually comparable, hence Fig. 4.11(a). With a rough equivalence established, we can move to compare the prototypes' characteristics.

There are a few possible figures of merit we might care about. One is of course the ripple in the gain, but the difference between 19 dB and 20 dB power gain is minimal in terms of overwhelming HEMT added noise, so it does not take priority. More important is the required pump power to drive an amplifier with a certain bandwidth. This requires factoring in a nonlinear inductive element (L_1 in Fig. 4.10), which we did in detail using

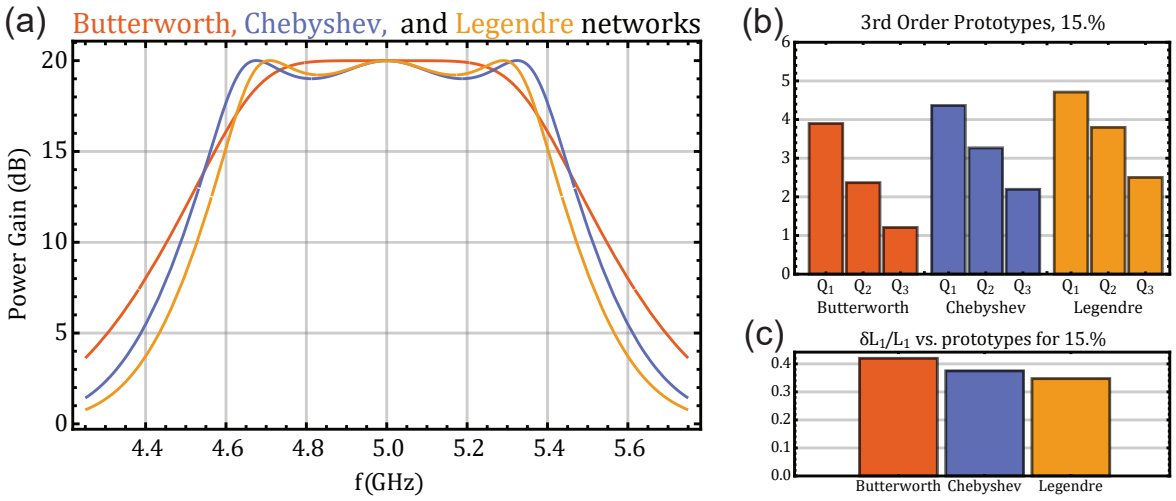


Figure 4.11: Twenty decibel gain profile of multiple coupled-mode prototype circuits with 15% bandwidth, comparing required inductance modulation and quality factor. (a) Gain profile of multiple different filter prototypes. For Legendre and Chebyshev prototypes, the ripple parameter is set to 1 dB. (b) Comparing the quality factors of each resonator as inherited from the adjacent resonators across prototype circuits, showing that the Chebyshev and Legendre circuits have a distinct edge over the Butterworth prototype. (c) Comparing the required inductance modulation for each design, showing the importance of choosing a high-Q prototype filter. A ratio of $\delta L/L = 0.5$ is the upper bound.

periodic steady-state analysis in Chapter 2. There, the conclusion for a single-mode circuit is clear: for a high saturation power rf-SQUID array, it will not reach 20dB power gain unless it feels a quality factor of at least $Q = 3.5$ at its phase-sensitive point. Since the purpose of the impedance matching network is to open a wider window around the phase-sensitive point to broaden the gain, it stands to reason that we should check to be sure that broadening the bandwidth of the amplifier does not decrease the quality factor of the "core" mode (R_1 in Fig. 4.10) below this threshold.

With this in mind we move to Fig. 4.11(b), to analyze in particular the quality factor Q_1, which is exactly the Q we are concerned with. Here, all three prototypes are above the threshold, but it's clear that there is a noticeable difference between the three different network types, with the Butterworth filter requiring the most inductance modulation, followed by the Chebyshev filter. The Legendre filter takes the gold in this example, although there is a small difference in the actual bandwidth in Fig. 4.11(a) that may make up the difference. The takeaway is that tolerating a larger ripple can create a larger effective quality factor,

and this larger quality factor may enable a designer to design a larger bandwidth filter for the same inductance modulation. To see this, we can borrow the "decrement relation" from Ref. [41] below with

$$Q = \frac{\omega_0}{\Delta\omega} g_0 g_1. \tag{4.29}$$

Here we can see the origin of the larger effective Q in Legendre filters and Chebyshev filters. The networks are synthesized by less tightly coupled resonators, resulting in higher effective Q, and a larger g_1 coefficient. We can calculate the required $\delta L_1/L_1$ to create the 20 dB gain in the prototype from Ref. [41] via

$$\frac{\delta L_1}{L_1} = 2\sqrt{\alpha} = 2\sqrt{\frac{\Delta\omega^2\left(\sqrt{G}-1\right)}{g_1^2\left(\sqrt{G}+1\right)\omega_0^2}} \times \begin{cases} \frac{1}{g_{N+1}} \text{ if N is even} \\ g_{N+1} \text{ if N is odd} \end{cases} . \tag{4.30}$$

Figure 4.11(c) drives the benefit of higher Q home by computing the inductance modulation using Eq. 4.30. Here we can see that the required modulation for a 15% bandwidth amplifier can be reduced from 43% with a Butterworth network to just under 35% with a Legendre network, which is a small but meaningful difference that could enable some 20% further increase in the engineerable bandwidth of the amplifier while requiring no changes to the nonlinear inductive element.

One additinoal element that I will not cover in detail here is that the actual eigenmodes of the system are no longer bound to just a single resonator. In fact, the "poles" of the amplfiier response come from modes that live spread out across each individual resonator in the system, much the same way hybridized resonators have eigenmodes that take symmetric and antisymmetric forms in the two resonators, forming a higher frequency "common" mode and a lower frequency "differential" mode. The poles here operate similarly in a conceptual way, except the filter design process carefully distributes the eigenmodes in frequency to generate teh correct impedance profile. Future work can hopefully identify these eigenmodes in greater detail, and give some insight into how much they each participate in the nonlinear element of what we think of as the "core" resonator, perhaps utilizing the work in Ref. [46].

Now that we have discussed the benefits of choosing networks with ripple over a Butterworth prototype, we can move to a more grounded discussion of the choice of circuit elements

that can be used to create both the resonators and the admittance inverters in Fig. 4.10, how they deviate at the edges of the band, and how we can rank them by ideality.

4.4 Design and Simulation workflow

4.4.1 Types of Resonators

After using the power insertion loss method to calculate the g_n filter prototype coefficients from the chosen filter prototype based on the tradeoffs in the last section, we can begin constructing a real circuit. From Eq. 4.27, we can see that we need to choose the impedances of the resonators first before we can calculate inverters. This is usually done based on what is available in the fabrication process, so I will provide a few brief examples here based on cases I have encountered.

Lumped Element Resonators

Lumped element resonators with impedances in the $5 - 20\ \Omega$ range are easy if one has access to multiple metal layer lumped capacitors and has a trusted, well-verified coiled inductor model calibrated to a real superconducting sample. Higher impedances are possible for lumped-element resonators, but in the absence of a space constraint it is better to avoid large (> 0.5 nH) coiled inductors that often carry a non-negligible parasitic capacitance. With this as a practical upper bound, then at 7 GHz we are limited to a maximum lumped-element resonator impedance of $Z_{lumped} = \omega_0 L_{max} \approx 22\ \Omega$.

Distributed resonators - $\lambda/4$ and $\lambda/2$

In a single-layer case, $40-100\ \Omega$ resonators are easy to make using $\lambda/4$ or $\lambda/2$ resonators. Fundamentally, this comes from the characteristic impedance of a transmission line being easily varied over a similar range. Not exactly a coincidence, considering that's one of the reasons why $50\ \Omega$ was chosen as the standard characteristic impedance in the first place [47]. However, the effective impedance of a $\lambda/4$ or $\lambda/2$ resonator is *not* simply the characteristic

impedance of the line it is composed of. This is also mentioned in Chapter 8 of Ref. [47]. Because of the transcendental nature of their impedance functions, there are two prefactors that I will calculate below. For a shorted $\lambda/4$ resonator with characteristic impedance Z_c connected to a port impedance Z_0 with resonant frequency ω_0 the full admittance with variable frequency ω is

$$Y_{\lambda/4} = \frac{1}{Z_0} + \frac{1}{iZ_c \tan \frac{\pi}{2}\frac{\omega}{\omega_0}}. \tag{4.31}$$

With this in hand, we can use

$$Q = \frac{\omega_0}{2}\frac{\mathrm{Im}[Y'(\omega_0)]}{\mathrm{Re}[Y(\omega_0)]} \tag{4.32}$$

for any admitance function Y that has a zero at ω_0. The denominator is easy here $(1/Z_0)$, but the numerator requires a little analysis. For our $\lambda/4$ short, we can calculate the derivative of Eq. 4.31 and Taylor expand it around $\omega = \omega_0$ to reach

$$\mathrm{Im}[Y'[\omega_0]] = \frac{\pi}{2\omega_0 Z_c} + \frac{\pi^3 (\omega - \omega_0)^2}{8\omega_0^3 Z_c}. \tag{4.33}$$

Taking the first term and using Eq. 4.32 we can see that the quality factor is

$$Q_{\lambda/4,\mathrm{short}} = \frac{\pi}{4}\frac{Z_0}{Z_c}. \tag{4.34}$$

A similar process can be used for an open-terminated $\lambda/2$ resonator, which will yield

$$Q_{\lambda/2,\mathrm{open}} = \frac{\pi}{2}\frac{Z_0}{Z_c}. \tag{4.35}$$

Interestingly, the difference between these two prefactors for the $\lambda/4$ vs. $\lambda/2$ cases gives more flexibility to distributed lines. If your process skews impedances high and you would like lower resonator impedances, then use $\lambda/4$ shorts. If not, or perhaps your process skews toward low impedances (say because you have crossovers to heal ground planes) then consider using open $\lambda/2$ resonators if space permits. Arguing for bounds on the impedance of distributed resonators after applying the right prefactor boils down to an argument of what characteristic impedance of transmisison line is "easy" to make. Using an EM simulator like HFSS or a waveguide handbook helps translate geometry restrictions to impedances. I'll

provide a small glimpse here via Cadence's TXLine for coplanar waveguide lines (CPW), which are the most common type that we use on-chip.

Let's start with high-impedance lines. If you are trying to make a high impedance coplanar waveguide line, it is natural to make the center trace as thin as possible, however it is decidedly *unwise* (I am speaking from unfortunate experience here) to make it any thinner than 1.5 µm, because the resolution of optical lithography (whether maskless like the MLA 100/150 or masked like a low NA ASML stepper) is not reliable below this length. Even if the laser itself is good, there can still be variations in resist height and resist development that can cause significant variation in such thin features. With a 400 µm high $\epsilon_r = 10$ substrate, this restricts the remaining design parameter to the gap, which must not be smaller than 1.5 µm for the same reason the center trace can't be thinner, and cannot be larger than 1/10 the substrate height, else we risk accidentally exciting microstrip waveguide modes with the ground plane underneath the substrate instead of the CPW modes we are after. With these parameters, a high-impedance choice of the center trace can obtain a range of $[Z_{low}, Z_{high}] = [60\ \Omega, 135\ \Omega]$. If instead we choose a low impedance transmission line center trace width, again following the same rule as the gap maximum by setting the width no larger than 1/10 the substrate height, we achieve a range of $[Z_{low}, Z_{high}] = [26\ \Omega, 62\ \Omega]$.

This is convenient! Distributed element CPW resonators have a massive effective dynamic range in their resonator impedances, and on top of that the lower bound picks up more or less right where the upper bound for lumped-element resonators leaves off. Unfortunately, the most prohibitive factor of distributed element resonators is the space requirement. At 7 GHz a $\lambda/2$ resonator will take up a colossal 9 mm of meandered line. On a 4 mm x 4 mm chip, this is an expensive ask, and limits the design. Because of this, $\lambda/2$ elements are not favorable, and so we often take the hit from the $\pi/4$ prefactor in Eq. 4.34 instead and build $\lambda/4$ shorts, with a modified total effective impedance range of $[Z_{\mathrm{low},\lambda/4}, Z_{\mathrm{high},\lambda/4}] = [20\ \Omega, 106\ \Omega]$

On the bright side, building distributed-element resonators generally guarantees you will hit the resonant frequency very close to the designed value, because the main factor in the resonance is the easily measured and easily designed length of the transmission line, within some small variation from how the line is terminated and the ideality of the ground plane along the line and at the termination that still warrants using an EM simulator to verify the

design.

With the resonators accounted for, we can choose the impedances in Eq. 4.27, and so now with demands set on the inverter J values from both the prototype and the resonator impedances, we can turn to realizing the inverter circuit elements themselves.

4.4.2 Pumpistor Model vs. Parametric Inverter Model

Before we go further, we should discuss two methods that we can use to calculate the scattering parameters of a parametric amplifier circuit. The first is the "pumpistor" model, explored in superconducting circuits in Ref. [64] and implemented soon after in Ref. [39]. Reference [64] has several versions of the pumpistor model, but the main takeaway is that for the degenerate case $\omega_s = \omega_i = \omega_p/2$ one can model the gain of the parametric amplifier as a negative resistor in parallel with the shunt LC resonator that forms the amplifier. For a single-mode, un-broadened amplifier this model lines up with the impedances presented in Fig. 4.9. One has to caution, however, that a negative frequency-independent resistor *only* works *exactly* on-resonance.

I must admit that for a long time I thought this was the *only* pumpistor model, and so I applied it wantonly, producing scattering that was asymmetric around $\omega_p/2$ that I will show briefly later in Fig. 4.15(b). If you stop and think abut this (as is so often the case as a stressed grad student, I didn't at the time) this doesn't make sense. How can the signal and idler have different gain? The device is a *mixer* that cannot distinguish between signal and idler. You can interchange one for the other in the real experiment and get the same result even in a nondegenerate gain experiment. As you might guess, this application of a negative resistance model is incorrect. In fact, Ref. [64] also gives a nondegenerate pumpistor model that can describe broadband impedance matched amplifiers, even using some of the same language as Ref. [41], describing the pumped nonlinear element as a means of reflecting the impedance of the idler onto the signal circuit. This intuition paved the way for the more rigorous synthesis procedure developed in Ref. [41]. Additionally, Ref. [64] also develops a saturation mechanism with higher-order expansions of the DC-SQUID device equations and incorporating the pump. It is this saturation mechanism that is often referred to as the

reason devices matched with impedance converting tapers like those in Ref. [39] have higher saturation *flux* than single-mode devices.

Despite this utility of the older model, in this thesis I will focus only on the model presented in Ref. [41], mostly because it is free of the parameters of any given nonlinear inductor, and uses the same inverters to describe both nondegenerate and degenerate amplifiers. The cost is that the model in Ref. [41] has no way of describing saturation. The technique that rectifies this is to design the circuit using a linearized circuit inductance and inductance modulation as in Ref. [41], and then once that design is done move to a full nonlinear simulator of your choice. There are many options for this with various tradeoffs.

In our work with our theory collaborators Chenxiu Liu and David Pekker, we take the brute force approach of writing down the equations of motion, then integrating them in the time domain. We wait for the system to reach a periodic steady state, which means that the system has fallen into a cycle that is periodic in time, with no changes in the envelope. From here, we can truncate the simulation, remove the initial transient, then do a Fourier transform of the steady-state data [31, 21].

For example, WRSPICE has native Josephson junction support, but requires you to perform the frequency domain analysis by transforming time-domain results yourself. More modern simulators like Keysight ADS and AWR/Cadence Microwave office have easier graphical interfaces, but require careful junction models like those shown in Ref. [42].

Generally, these programs have the final say over whether the design will work, and so the full nonlinear simulations are where the linear component values are adjusted and the design is finalized. That said, even the linear elements are imperfect inverters across the band, and so in the next section we will discuss these linear nonidealities and how to compensate for them.

4.4.3 Types of Inverters

In this subsection, I will go over the design tradeoffs of different types of impedance inverters, similar to the discussion of resonators in the previous section. Designing an admittance inverter with a specified value $J = 1/K$ *at a particular frequency* ω_0 is well-covered

in Pozar and other microwave engineering literature. However, in the former's case, there is not much discussion of how the impedance inverter values vary over frequency and might skew an otherwise ideal filter prototype. This is what I intend to cover here, with some additional details on which inverters are best suited for high-impedance or low-impedance filters over a large or small bandwidth around our typical center frequency of $\omega_0 = 7$ GHz.

4.4.3.1 Inverter Definition

An impedance inverter is defined by how it transforms the input impedance of a load Z_L. The input impedance is usually defined as the effective impedance "seen" from a certain point in the circuit, but a more formal definition is available from using ABCD matrices. If an ABCD matrix gives the relationship between input and output voltage and currents as

$$\begin{bmatrix} v_{in} \\ i_{in} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_{out} \\ i_{out} \end{bmatrix}, \tag{4.36}$$

then we can describe the input impedance of a circuit with that ABCD matrix representation by terminating it in an open circuit, or requiring that the output current i_{out} be zero. Then we can see that the ABCD matrix equations reduce to

$$v_{in} = Av_{out}, i_{in} = Cv_{out} \implies \frac{v_{in}}{i_{in}} = \frac{A}{C} = Z_{in}. \tag{4.37}$$

This makes it easy and unambiguous to discuss input impedances. For an ideal impedance inverter, the input impedance of a load Z_l (whose previous input impedance is just Z_l) is transformed to $Z_{in} = -K^2/Z_l$. This is accomplished by the ABCD matrix given before as

$$T_{inv} = \begin{bmatrix} 1 & iK \\ \frac{i}{K} & 1 \end{bmatrix}. \tag{4.38}$$

At first this doesn't seem like the simplest possible answer, why is it necessary for B and C to be complex? The answer is that for real values of K, impedance and their equivalent admittance inverters are passive, and so passive elements must be reciprocal, requiring that $AD - BC = 1$. The i prefactor on a real K or $J = 1/K$ ensures $BC < 0$ and that we do not violate reciprocity.

4.4.3.2 Capacitive, Inductive, and Quarter-wave inverters

Now that we know the fundamentals of impedance inverters, let's briefly tabulate the types of inverters we can use. Fig. 8.38 from Ref. [47](Pozar) is a decent starting point. I have reproduced the parts of it that are relevant here for ease of reference in Fig. 4.12, with some additional inverters I would also like to discuss.

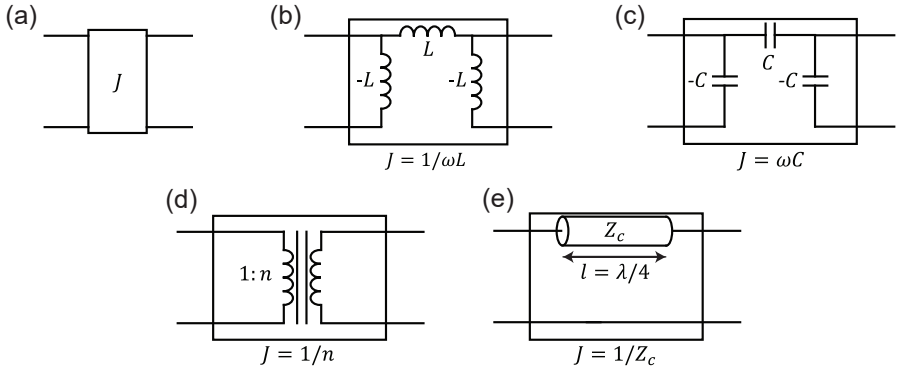


Figure 4.12: Circuit representations of various useful impedance/admittance inverters. (a) Ideal admittance inverter, defined by Eq. 4.38 with $K \to 1/J$. (b) Inductive π admittance inverter (c) Capacitive π admittance inverter (d) $1 : n$ transformer, which also functions as an admittance inverter. (e) $\lambda/4$ segment of a transmission line, one of the canonical examples of an admittance inverter used as a matching network.

The inverters pictured in Fig. 4.12 all have their listed J values *at a particular frequency* ω_0. This is an important caveat, especially because the ideal inverter is frequency-independent. Inevitably, this discrepancy leads to distortions in the gain profile of the real circuit in comparison to the ideal inverter circuit. Fig. 4.13 shows the variation in the inverter values graphically to supplement the equations in Fig. 4.12 for the capacitive, inductive, and transformer inverters.

In the capacitive and inductive cases in Fig. 4.13, we can see that their inverter values vary roughly linearly around ten percent of the resonant frequency in either direction. This variation distorts the gain profile of real networks synthesized with these methods compared to the ideal inverter, which is more or less just a mathematical object. If we calculate the inverter values for a 3-pole Chebyshev circuit with varying bandwidth and calculate the scattering using the inverter model in Ref. [41] we arrive at Fig. 4.14, which shows the target that we're hoping to hit with real circuit inverters later in the section.

For capacitive networks, I show the effect of this in Fig. 4.15 vs. the fractional bandwidth

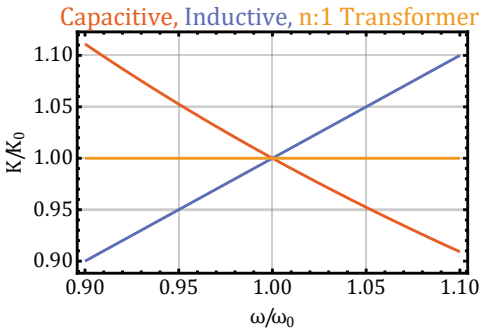


Figure 4.13: Variation in the inverter values of common inverters vs frequency. Only the transformer remains frequency-independent.

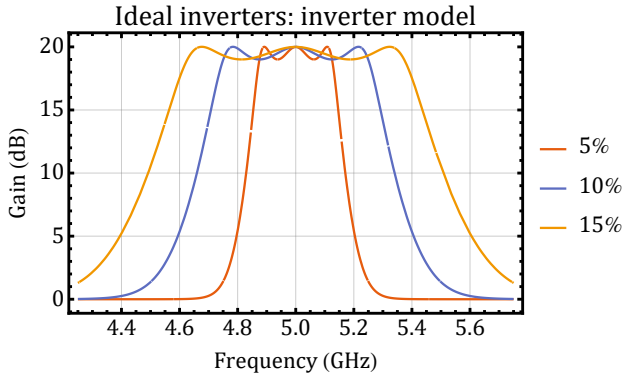


Figure 4.14: Scattering parameters of a 3-pole Chebyshev prototype synthesized with ideal frequency-independent admittance inverters. The scattering is the same no matter the choice of resonator impedances.

of the filter. Although each design works at exactly the center frequency, the inverters are less and less effective towards the edge of the band as we demand more and more fractional bandwidth out of the design. The distortion from the nonidealities in the inverters even overwhelms the Chebyshev ripple programmed into the prototype circuit.

Are quarter-wave inverters any better? In Fig. 4.16(a) we can see that transmission line inverters suffer from a similar problem in variation around their center frequency. However, they have an additional quirk, how well they maintain their inversion constant depends on the impedance of the quarter wave line compared to the impedance of the inverted load. If the ratio of the impedance of the quarter-wave inverter is small, then the variation of the inversion across frequency will also be small. This gives an additional quirk to the design: the designer is incentivized to keep the resonator and inverter impedances as close to the reference impedance Z_0 as possible for a broadband design. However, even with using this

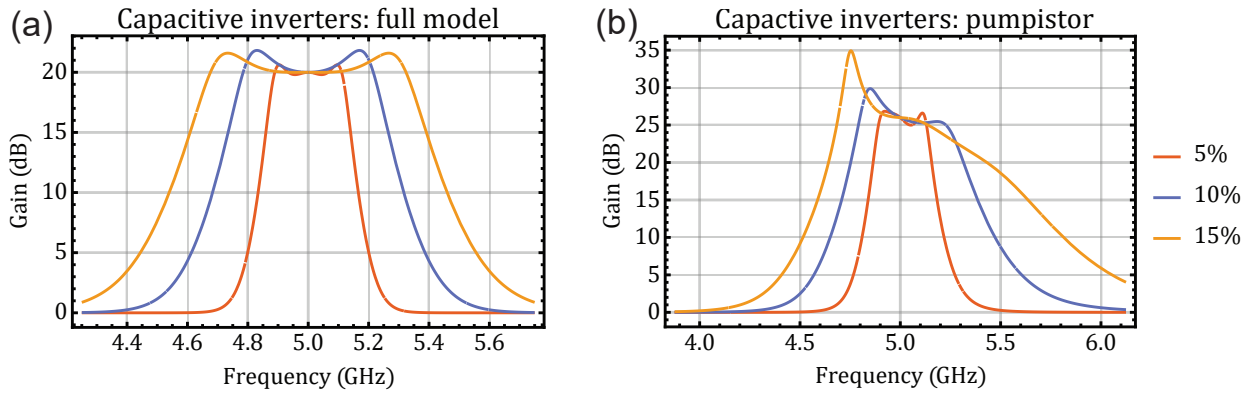


Figure 4.15: Capacitive inverters used to synthesize various fractional bandwidth filters. (a) Using the inverter model as in Ref [41] to represent the inverters in both the idler and signal side of the circuit at their appropriate frequencies. (b) The same prototype, but calculated with INCORRECT application of the pumpistor model, literally employing a negative resistor for all frequencies, not just the only valid point, which I only provide as a means of providing a red-flag to anyone who reads this that negative resistor scattering away from degenerate signals and idlers is not correct.

guideline, transmission line inverters suffer from the additional downside that they also add a non-negligible imaginary component to the inverter constant, shown in Fig. 4.16(b), that will complicate the filter profile in any case. This imaginary component also gets worse with increasing impedance contrast in the circuit.

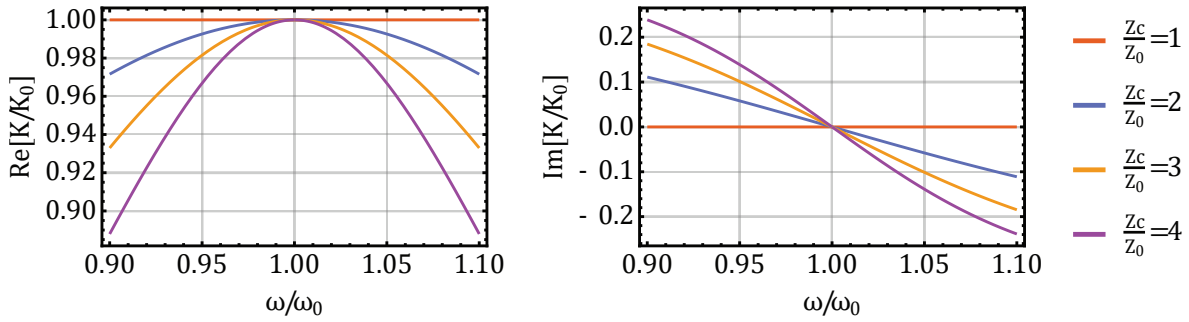


Figure 4.16: Transmission line inverter variation over frequency. (a) Variation in the real component of the inverter value, similar to the components in Fig. 4.13 but with even symmetry around resonance as opposed to odd symmetry. (b) Imaginary components of the inverter value, unique to the quarter-wave inverter, further complicate their broadband behavior.

4.4.3.3 Minimizing Nonideality of Quarter-wave inverters

Using transmission line inverters to synthesize the same broadened amplifier prototype as in Fig. 4.15, we reach the scattering shown in Fig. 4.17.

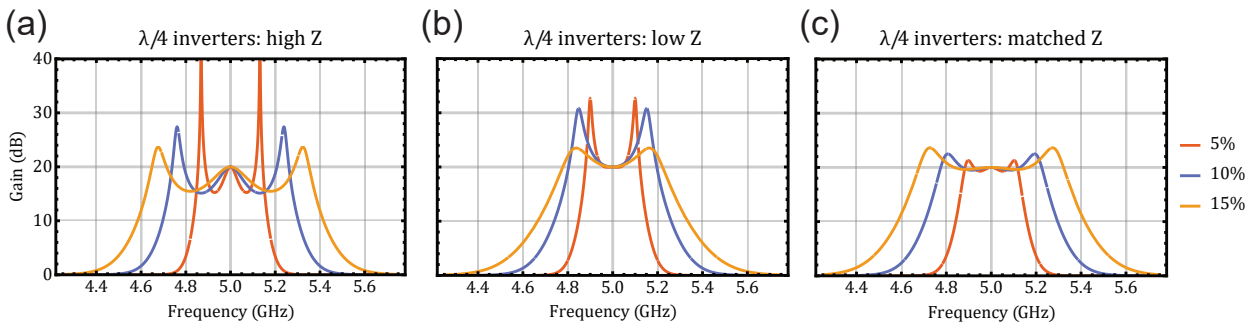


Figure 4.17: Broadband amplifiers synthesized with quarter-wave inverters with respect to both bandwidth and the impedance of the resonators in the circuit. (a) Same prototype, higher impedance $Z_1 = 150\,\Omega, Z_2 = 120\,\Omega, Z_3 = 120\,\Omega, Z_0 = 50\,\Omega$ network. The nonideality of the quarter-wave inverters manifests as *increased* bandwidth, but also more drastic ripple. (b) A transmission line inverter with lower impedance $Z_1 = 5\,\Omega, Z_2 = 20\,\Omega, Z_3 = 20\,\Omega, Z_0 = 50\,\Omega$ network. The nonideality of the quarter-wave inverters manifests as reduced bandwidth. (c) A transmission line inverter with the most "ideal" circumstances, where the impedances of the resonators are all set such that the inverters have a value of $J_{jk} = 1/Z_0$. The nonideality of the inverter shows nonetheless, but much less severely than the other cases.

In contrast with the capacitive inverter circuits in Fig. 4.15, the quarter wave inverter circuits in Fig. 4.17 have a variation in their performance that depends on the individual inverter deviations away from $J_{jk} = 1/Z_0$ for each inverter. I have divided these into two distinct types. In Fig. 4.17(a) we have a case where the resonator impedances are *above* Z_0, the reference impedance by a large margin, and so the inverters are also high impedance. In this case the gain is majorly distorted, and it is pretty clear that the filter we end up with is nothing like the one that we originally designed in Fig. 4.14. In the process of writing this, I decided to investigate a little more and generated a few other cases.

The low-impedance case shown in Fig. 4.17(b) also suffers from distortion, but instead of ripple, we see reduced bandwidth instead. However, there is a way to minimize the distortion by using transmission line inverters. With the guidance from Fig. 4.16, we could guess that we should set the transmission line inverter values to as close to the reference impedance as possible. In order to accomplish this, we can look at Eqs. 4.27. Here we can work inwards from the reference impedance Z_0, setting the impedances of the *resonators* to minimize what we are asking of the *inverters*. The result for a 3-pole circuit that sets all inverter values to $1/Z_0$ is given in Eq. 4.39, although in general, this can work for any size network.

$$Z_1 \rightarrow \frac{g_2 Z_2}{g_1 g_4^2}, Z_2 \rightarrow \frac{g_3 g_4^2 Z_3}{g_2}, Z_3 \rightarrow \frac{\Delta\omega Z_0}{g_3 g_4 \omega_0} \tag{4.39}$$

This final case is shown in Fig. 4.17(c), and indeed the scattering is much better than the other two cases! This is what I would suggest for developing transmission line inverter-based broadened parametric amplifiers. It also begs the question: what are the inverters actually *doing* if they all have inverter values of, effectively, one? The outer inverter, the one that links resonator three with the real external reference impedance, is doing absolutely nothing. We can totally omit it from the design, as looking out from the amplifier into a purely real reference impedance through a quarter wave line of equal impedance to the reference is completely equivalent to de-embedding the port by a quarter wavelength. This is a trick we already employed in the designs from Ref. [22]. The inner inverters, however, are not the same as this and cannot be omitted. As a small proof by contradiction, if you *did* omit them, the resonator elements could be combined in parallel and the entire circuit would collapse into a single pole. There is no intuitive justification to try this in any case, since the loads on either side are no longer real impedances, and so the simple calculation used to generate Fig. 4.16 by comparing it to a real reference impedance is no longer valid. A more careful analysis would have to generalize to resonant loads, and analyze the effect on the slope of the imaginary component of the admittance around resonance.

Even with tricks such as adjusting the inverter impedances to the reference impedance, in designing a broadband amplifier with any type of frequency-variable inverter, we will inevitably have to "fudge" the design a bit. We can do this by trimming the component values by hand until the gain profile looks more similar to the one in Fig. 4.14, or asking an optimizer to do this same thing but *much* faster. Most circuit design environments like AWR/Cadence Microwave Office or Keysight ADS are more than capable of this type of optimization. Because most circuit simulators, either on the finite-element level such as ANSYS HFSS or on the circuit design level such as in AWR Microwave Office, do not allow the designer to easily do frequency mixing to simulate the real gain that would occur in the circuit, we need a proxy variable that is not the gain as an optimization value.

It is a valid option to simulate the circuit using any load you want at the core of the amplifier. For example, setting a real termination of the value $R_{pa} = Z_1 \frac{\omega_0 g_0 g_1}{\Delta\omega}$ would produce

scattering in a linear simulator with inverse scattering to Fig. 4.15(b) in the capacitive case, which would look somewhat similar to a filter that these programs are well-suited to optimize to. While this is a valid option, it lacks a nice physical intuition, requiring the entire structure of filter synthesis to reach the reason why the negative resistance is that particular value demanded by g_0, g_1 and the bandwidth. Building on the arguments of Chapter 2, in the next section I will justify turning to a parameter that has much more physical significance to the idea of a one-port parametric amplifier and can easily be verified experimentally: the group delay.

4.4.4 Group Delay, Quality Factor, and Fitting Multi-pole Networks

In this subsection, I will introduce group delay, its relation to quality factor of resonant circuits, and how to use it to fit multi-pole single-port resonant circuits. First let's start with the definition of group delay from any scattering parameter S as

$$\tau_g = -\frac{d\phi}{d\omega}, \text{ where } \phi = \arctan\frac{\mathrm{Im}(S)}{\mathrm{Re}(S)}. \tag{4.40}$$

The group delay (not to be confused with phase delay, which is ϕ/ω) describes how the peak of an amplitude-modulated signal is delayed at the output of a circuit with respect to the input. Quality factor has a similar interpretation, a higher Q can be similarly viewed as more time that the light spends "bouncing around" inside a resonator before it leaves the circuit. Around resonance, then, we should expect the two to be related. To show exactly how they are related for a single-mode circuit, let's start with the reflection scattering parameter of a single-mode overcoupled circuit, which is given in Eq. 4.2. Using Eq. 4.40 we calculate the group delay expressed in terms of the quality factors Q_i and Q_e and resonant frequency ω_0 as

$$\tau_g(\omega) = \frac{2\omega_0\left(\omega^2+\omega_0^2\right)Q_eQ_i^2\left(Q_e^2\left(\left(\omega^2-\omega_0^2\right)^2Q_i^2-\omega^2\omega_0^2\right)+\omega^2\omega_0^2Q_i^2\right)}{Q_e^4\left(\left(\omega^2-\omega_0^2\right)^2Q_i^2+\omega^2\omega_0^2\right)^2+2Q_e^2\left(\omega^2\omega_0^2\left(\omega^2-\omega_0^2\right)^2Q_i^4-\omega^4\omega_0^4Q_i^2\right)+\omega^4\omega_0^4Q_i^4}. \tag{4.41}$$

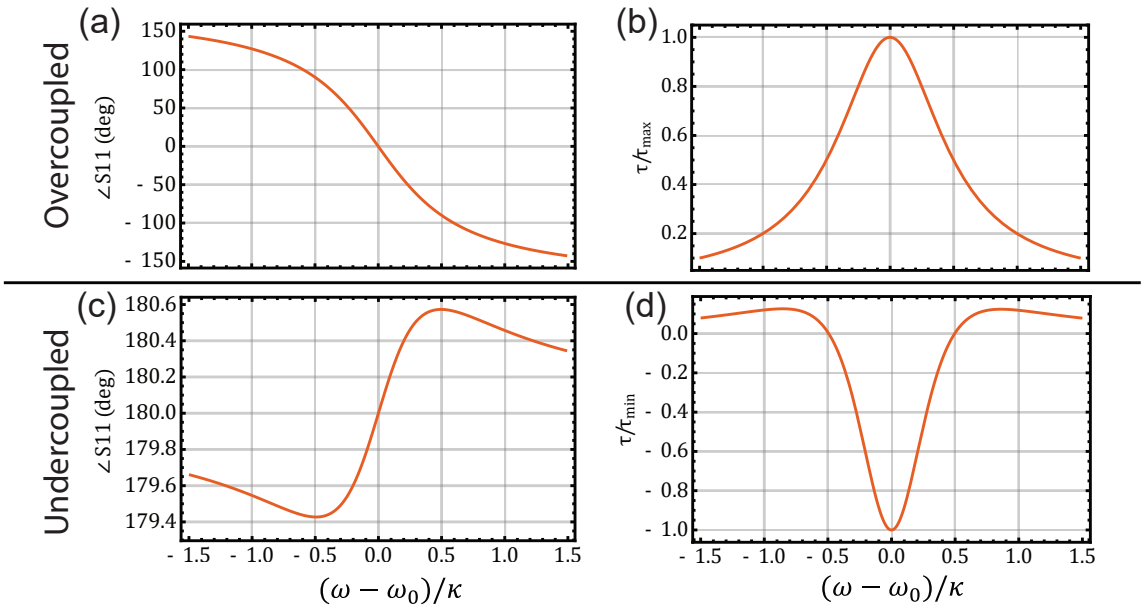


Figure 4.18: Phase wraps and group delay for under and over-coupled resonators. (a) Phase wrap for a $Q_i = 100Q_e$ over-coupled resonator, showing full 360° wrap. (b) Group delay of the same resonator as (a), showing a peak aligning with the τ_{max} value defined in Eq. 4.42. (c) Phase wrap for $Q_e = 100Q_i$ undercoupled resonator. (d) Group delay for the resonator in (c).

This is clearly not so useful, but I wanted to provide it for thoroughness considering how thoroughly we'll cut it down next. To see how we should interpret this function, let's evaluate it on resonance in two distinct cases below.

$$\tau_g(\omega = \omega_0) = -\frac{4\left(Q_e Q_i^2\right)}{\omega_0\left(Q_e^2 - Q_i^2\right)} \approx \begin{cases} \frac{4Q_e}{\omega_0} := \tau_{max}, \text{ if } Q_i >> Q_e \text{ (Overcoupling)} \\ -\frac{4Q_i^2}{Q_e\omega_0} := -\tau_{min}, \text{ if } Q_e >> Q_i \text{ (Undercoupling)} \end{cases} \tag{4.42}$$

This answer is quite useful! It's very clear that the group delay (an easily measurable variable that you can read right off of a VNA!) is directly related to the quality factor on-resonance. To be clear, this is not always foolproof, because very small group delays are hard to measure in a very long measurement line. However, in the cases of highly overcoupled resonators that have $Q_e > 10$ it is very convenient to measure the quality factor by reading off the delay this way. Fig. 4.18 shows the shapes of both the resonator phase wraps and the group delay in both the under and overcoupling case.

While group delay is useful in the overcoupling case as a direct reflection of quality factor, it is admittedly not useful, even confusing, in the undercoupling case. Because of the presence of both Q_e and Q_i in the undercoupled expression in Eq. 4.42, we can't use it to measure either Q directly without a fit, and further the value is *negative*. What does a negative group delay mean? There is quite a lot of discussion about this sort of thing [3], but in the end it is just an artifact of a highly resonant circuit. When pulsed *only within the linewidth of the resonance*, the circuit distorts the incoming pulse such that the output peak is earlier in the waveform than it was before in the input in a sort of "predictive" fashion. This appears as a negative group delay, but is just a quirk of the filter response being very catered to the input. Imagine a lock fit to a key. When you use the key on the lock and the door opens, would we say that the lock was able to predict the structure of the key? No, the information in the key is already embedded into the lock because you could construct the key by disassembling the lock and looking at the pins. There is nothing transmitted, just a switch activating a sequence of *predetermined* events. In the same way, there is no speed of light barrier being broken in the case of an undercoupled resonator with a negative group delay. It appears as an advance only because that particular pulse within the band is the key required to unlock a response that appears that way. Lastly, I should also address the third critical coupling case of the one-resonator system. In this case, $Q_e = Q_i$ and the group delay becomes infinite. This makes a strange sort of sense though, since in that case the light is perfectly transferred away from the port, and no reflection is measured. If you never get the light back, infinite delay is a somewhat reasonable interpretation.

The link between quality factor and group delay, particularly in the overcoupled case, opens an opportunity for interpreting large single-port filter networks for parametric amplifiers. Since they do not have a second port to access transmission scattering in a real circuit, and their overcoupled nature prevents the interpretation of an amplitude response in the scattering, the phase response and the associated group delay provide the only source of information to fit the response of the linear response of the real circuit, and extract any differences between the real circuit and the design.

If we analyze the group delay response of the ideal 3-pole scattering shown in Fig. 4.14, we see the result in Fig. 4.19. Looking at the phase wraps, we see what we might expect

with increasing fractional bandwidth designs, which is that the phase wrap gets spread over a wider and wider window in frequency. This causes the delay to become lower and lower, also spread over wider and wider windows in frequency.

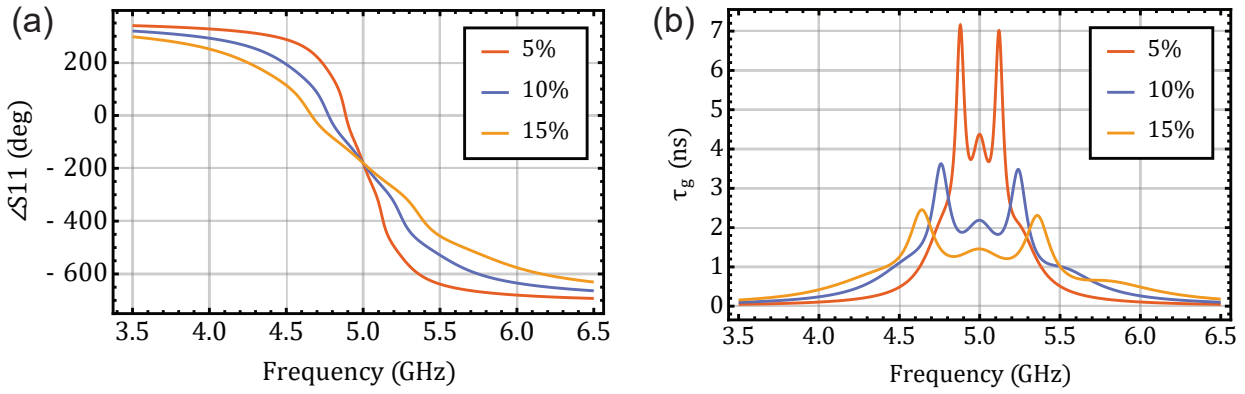


Figure 4.19: Phase wrap and group delay of 3-pole Chebyshev filter prototypes. (a) Phase wraps of various fractional bandwidth 3-pole Chebyshev prototypes with ideal inverter elements. (b) Group delay calculated form the phase wraps, shown in nanoseconds.

The objective in calculating the group delay shown in Fig. 4.19 is two-fold. The first is just to provide a practical means of experimentally verifying that the linear scattering of the amplifier, once fabricated, is close to the circuit prototype. The second is to try to make a bridge back to the relationship between the quality factor and the saturation power that we established in Chapter 2, and predict whether the amplifier will be pumpable by this metric in addition to the metrics I have already introduced in Fig. 4.11. A first order check is to make sure that the quality factor that we can calculate by applying Eq. 4.42 to the group delay in Fig. 4.19 matches up with the quality factor of the core resonator for the 15% design shown in Fig. 4.11(b). I'll show the practical objective and its effectiveness in the next subsection, but we can address the quality factor link now. If we calculate the quality factor of the 5 GHz center frequency 15% prototype by applying the overcoupling version of Eq. 4.42 to the delay in Fig. 4.19 we get a value of about $Q_e = 1.46\ \text{ns} * (2\pi\ 5\ \text{GHz})/4 \approx 11.5$. Because this doesn't align very well with the quality factor from Fig. 4.11(b), we are at an impass. However, we can still use the group delay as a benchmark target for simulating and measuring the single-port scattering of a matched prototype, which will be critical in the next section to bring the scattering of the physical circuit into agreement with the ideal design.

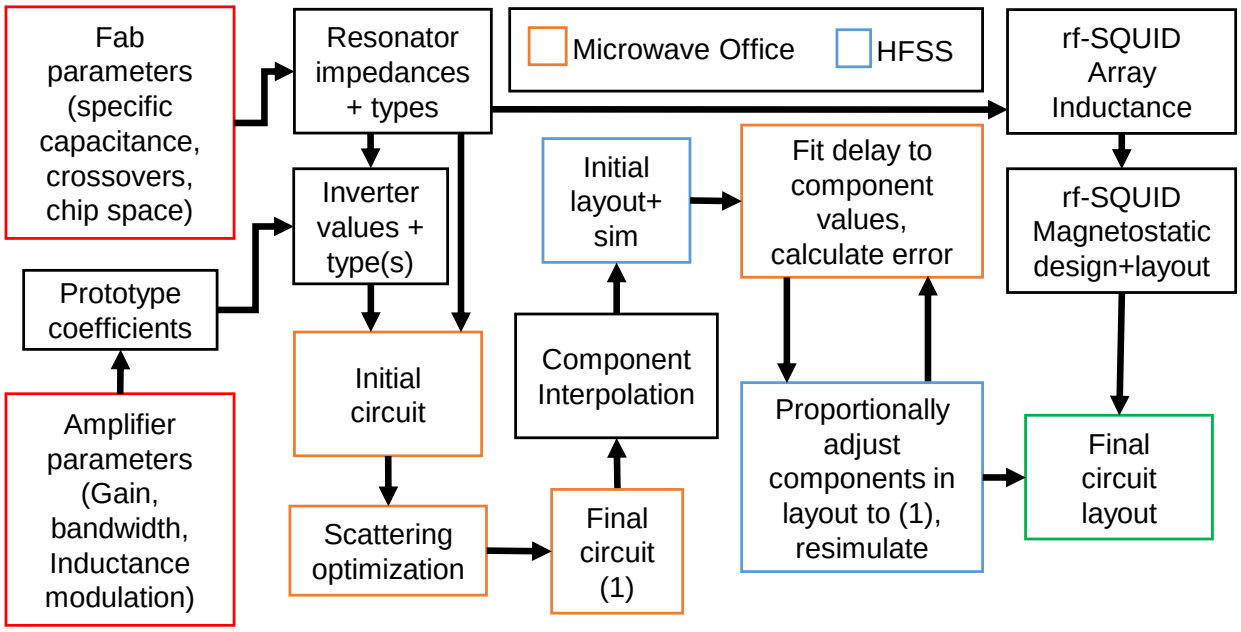


Figure 4.20: Broadband parametric amplifier design flow. Starting from the red squares, the process can be described by an initial design stage to determine the circuit followed by parameter interpolation to inform initial layout, then lastly a feedback loop to bring the group delay of the layout into agreement with the design. Lastly, the magnetostatic design of the rf-SQUID covered in chapter 2 completes the design of an rf-SQUID array with a particular inductance.

4.4.5 Simulating Parametric Amplifiers in ANSYS HFSS

In this subsection, we will walk through the simulation steps that build up to the final layout of a parametric amplifier. For a general reference to the design flow of this process, see Fig. 4.20. The previous subsection covered everything up until the final circuit design in the bottom middle of Fig. 4.20, and established the group delay as a viable metric to match the linear cirrcuit parameters. In this section, I will cover the latter half that translates circuit component values into physical dimensions in the layout.

I take a bottom-up approach here, because in large circuits where multiple resonators can interact with one-another via either intentional or unintentional parts of the layout it can often be difficult to extract which part of the circuit is causing a particular problem. Therefore, the goal is to establish confidence in each individual component, and only then combine them into the final design. This also sets up the backbone of eventually producing a Process Design Kit (PDK), a critical component of any production chain for radio frequency circuits. I will begin with the resonators, showing the layout and design parameters, show

the individual layouts in HFSS, and then give the results of simulations that span the entire design space, ending with justifying the use of an interpolator to design any particular resonator within the bounds of the design space.

4.4.5.1 Component Interpolation: Transmission Line Resonator

For a resonator to behave as we anticipate in a larger circuit we have to accurately know two things about it: the resonant frequency ω_0 and its impedance Z_r. A reasonable question to ask is: why do we need a simulator at all? We know the transmission line's length and the dimensions of the line, can't we use analytical expressions to calculate the resonator frequency and impedance? The answer is that of course we can, and this is exactly how we will set rough bounds on the simulation design space, but unfortunately there are many nonidealities in the layout that complicate this first-order approach, and if this is the only validation tool we use, the layout will simply be wrong. The first and biggest problem to look out for is grounding. Because we use coplanar waveguide lines for the resonators, the entire length of the resonator divides the ground plane, and although through-silicon-vias are gaining ground in some foundries [36], we do not yet have the luxury of being able to create vias down to a second "healing" ground plane. If we leave this division in the ground plane, we may accidentally generate/excite "stripline" modes between the two ground planes, hopping over the center trace entirely. Figure 4.21(a) shows the field pattern schematically of this stripline mode of propagation, as well as the desired CPW mode and another parasitic mode that goes directly through the substrate as a microstrip mode.

Since the top layer of the chip is only truly grounded at the sides of the chip via wirebonds to the package, the ground plane is able to locally support a different potential on either side of the transmission line. These parasitic modes are harmful to the scattering of the circuit and to the design as a whole because the resonances that they can create can couple to other resonators in the circuit, which moves their frequencies and adjusts their impedances. In short, we want to eliminate any modes of propagation that are not CPW modes. In Fig. 4.21(b) we see how vias to the lower ground plane accomplish the goal of removing stripline modes by essentially leaning on a low-inductance connection to the below-substrate

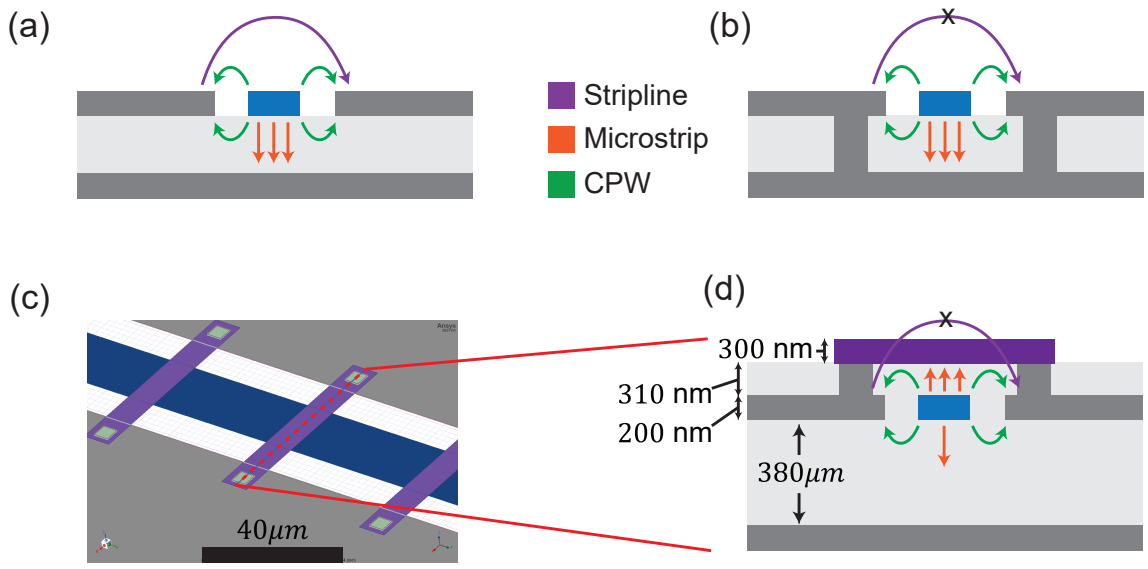


Figure 4.21: Possible modes of propagation for electric fields in a CPWG transmission line architecture. (a) Electric field line schematics that show the rough field directions of all the modes in a grounded coplanar waveguide (CPWG) transmission line. The electric field in green corresponds to the desired TEM mode, all others are parasitic. Without any mitigation, any asymmetry in the left and right upper ground planes will allow the stripline mode in purple, while a wide enough center trace will permit an orange microstrip mode to travel as well. (b) Grounding vias eliminate the stripline mode, as the two sides are held at the same potential to within the impedance of the via. (c) Diagram of actual CPWG layout in a component of an HFSS simulation, with red dotted cross-section. (d) Dotted cross-section from (c), showing the lack of vias through the substrate, instead using vias up to through a deposited dielectric layer to a top-metal crossover instead.

ground plane to eliminate any potential difference between the left and right ground planes on the surface layer. Since we do not have this ability in practice but still need to remove stripline modes, we are left to Fig. 4.21(c) and (d) where I show a crossover in ANSYS HFSS simulation of a CPW and a schematic of the red dotted cross-section in Fig. 4.21(d). Here we can see that instead of going under the line, through 380 µm of substrate, we go over with another metal layer on the surface. This is very effective in healing the ground plane, but the cost is that it heavily loads the line capacitively, with the capacitor microstrip-like upward-pointing orange fields in 4.21(d). On top of this, this is no longer a common format of transmission line. Periodic loading of crossovers is not an analytically tractable transmission line problem. This is yet another reason why we have to simulate resonators made from these types of lines in finite-element electromagnetic simulators like ANSYS high frequency

structure simulator (HFSS).

So how do we simulate the line? One could imagine that we could set a source and load impedance and try to drive a wave from one to the other over a very long section of line. If the transmission line impedance matches the (equal) source and load impedances, we would expect the reflection to be minimized. This works and is something we have tried before, however, the definitions of wave ports and lumped ports in HFSS complicate this simple idea of what exactly the port impedance is. For example, you may set up two equally sized wave ports at either end of the line and see zero reflection. Great, right? But the problem is, buried in the menus of HFSS is a calculated port impedance for every wave port, and you may find on inspecting this value that the port impedance is calculated as 40 Ω for both ports. This means that this carefully calculated scattering is excellently matched to 40 Ω, but when you cut that line out and use it in a real circuit where the external impedance is 50 Ω, it will *not* be well-matched. All this is to say that if you do use this method, **check the port impedance**.

Additionally, the size of the simulation can hypothetically be infinite, so how would we decide how to truncate it? Practically speaking, eigenmode simulation of a particular resonator is easier than all of this and gets at the goal (a resonator of specified impedance and resonant frequency) more directly. Pick the format of the distributed element resonator, and build the model with a particular crossover density and dimension, then simulate the quality factor and resonant frequency via eigenmode simulation in HFSS. I show an example of the layout of this simulation in Fig. 4.22(a) with three parameters, the CPW gap, CPW width, and the length of the downturned segment, highlighted in red. In Fig. 4.22(b) and (c), we see that these three parameters are swept over a large range to try to capture as much of the design space as possible. Because the design space is three-dimensional, we cannot pick out the best parameters by eye, only looking at 2D slices. Instead, we interpolate the entire parameter space, and then use gradient descent to find a minimum-cost point in the space for a set of goal impedance and resonant frequencies.

Interpolating through the design space allows us as designers to decide on the architecture, do one overnight sequence of 200 or so eigenmode simulations, and then be able to pick whatever resonator we want within the bounds of the space with a reasonable confidence

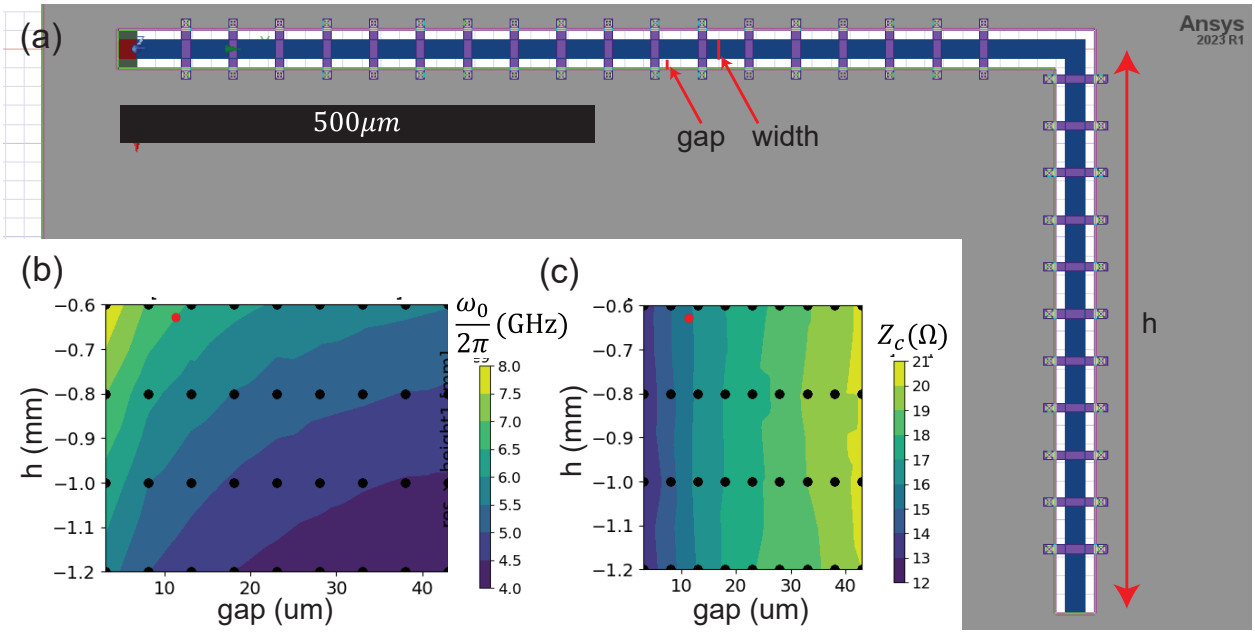


Figure 4.22: Resonator parameter interpolation from HFSS simulations. (a) Layout of the$\lambda/4$ resonator in HFSS, bent at 90 degrees to reduce the footprint of the resonator. The center trace of the CPW resonator is highlighted in blue and the crossovers in violet. The port is a 5000 Ω (The reason that we use 5000 kΩ here is that eigenmode is bad at picking up low-Q modes, so we artificially make it a high-Q mode by using a very large external impedance.) lumped resistance in the top left, shown in red. CPW gap, width and the length of the segment labeled "h" are all swept parameters. (b) Color plot of the resonator impedance (related to the quality factor of the resonator by Eq. 4.34 and the port resistance) with respect to the length of the adjusted segment "h" and the CPW gap. The width, while it is also swept, is a constrained parameter at 20.3 µm in both color plots to make 2D visualizations possible. The black dots in the color plot are the HFSS simulation values,with the colors representing two-dimensional slices through a first order interpolation of the entire three-dimensional parameter space. The red dot in both (b) and (c) indicates the optimum point found via gradient descent.

that the result will be close. This drastically lowers the overall time needed to simulate a large, multi-resonator design. It reduces it to piecing together previously calibrated LEGO-like pieces, and then optimizing the full design with small perturbations on the calibrated components.

4.4.5.2 Compensating for negative capacitance inverter legs by shortening distributed resonator length

As a brief aside, I want to address why the optimization target in Fig. 4.22(c) is *not* the center frequency of the amplifier, at 6.5 GHz, and equivalently why the electrical lengths of

the resonators in Fig. 4.1 are not targeted at 90° at 6.5 GHz. The reason is that the ideal capacitive inverter pictured in Fig. 4.12(c) has negative capacitance legs in addition to the intuitive series capacitance. This negative capacitance has to be absorbed by the neighboring resonators (the capacitance has to be altered if there is no resonator on one side to absorb it, the exact formula is given in Ref. [41] in Appendix D). Pozar gives the formula for how to absorb a negative capacitance in a distributed resonator, specifically the $\lambda/4$ resonators we use here. To compensate for a negative capacitance C_c, a $\lambda/4$ resonator at frequency ω_0 of characteristic impedance Z_c can be shortened by an electrical length θ_r given in radians by

$$\theta_r = Z_c \omega_0 C_c. \tag{4.43}$$

This gives the final length of the $\lambda/4$ resonator in radians as

$$\theta = \frac{\pi}{2} - \theta_r = \frac{\pi}{2} - Z_c \omega_0 C_c. \tag{4.44}$$

This allows us to build a capacitively coupled filter network with distributed element resonators, which can be easier to calibrate than lumped-element resonators in some circumstances, and also have a much higher dynamic range of impedance values as discussed earlier in the chapter. An important caveat is that this DOES NOT WORK past a certain point in θ_r. The obvious limit is that the line length cannot be negative, but the more practical limit is about $\theta_r/\theta \approx 0.7$. The intuitive reason is that the line stops acting capacitive beyond that point because the electrical field distribution starts to fall off. The mathematical reason is that the capacitance of the line is calculated by an expansion of the line's reactance, and that expansion is not very accurate beyond that point.

4.4.5.3 Layout assembly

Once we have calibrated each component in the design, say for a 15% bandwidth amplifier centered around 6.5 GHz, we can assemble them into the final layout. Fig. 4.23(a) gives a 3-pole Chebyshev response with 1 dB of ripple in the gain response. The tuned filter values labeled in Fig. 4.23(b) are given in Table 4.1.

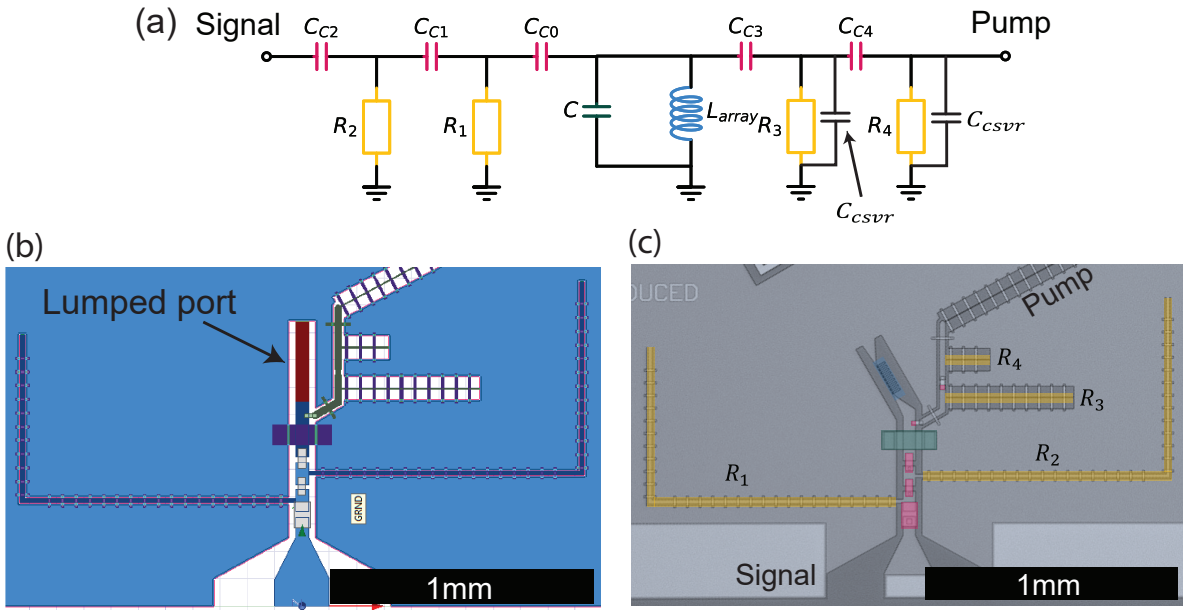


Figure 4.23: Layout and group delay fit for a 15% bandwidth parametric amplifier with a pump port. (a) Circuit diagram for a 3-pole amplifier with a built-in pump port. Values are given in Table 4.1. (b) HFSS model of the circuit showing that the lumped port must be the same length as the real array. (c) Real picture of the fabricated device. A tilt was added to the array (in haste, I admit) to attempt to better the grounding along the pump line delivery circuit.

The values in Table 4.1 represent each stage of the design process. The filter synthesis procedure given in detail earlier in the chapter and in Ref. [41] give the first column after a choice of resonator impedances. We tune these synthesis values to level out the inverter nonidealities shown in Figures 4.15 and 4.17. Mostly, these tuning values give only small variations to each circuit element. Following the flow diagram in Fig. 4.20, we move to calibrating the resonators with the procedure detailed in Fig. 4.22. We calculate the coupling capacitors more naively, with a specific capacitance that is given by the fabrication process, and the layer thicknesses shown in Fig. 4.21(d). This yields the layout shown in Fig. 4.23(b), and eventually the real device in Fig. 4.23(c).

Following this, we use a driven modal HFSS simulation to extract the scattering parameters of this layout without the core inductance of the array. Because we can't model flux quantization in the HFSS simulation, the array itself is replaced by a lumped 50 Ω port **of the same length as the array**, we then extract the scattering parameters to this fictional "port," export a touchstone file (which just describes the scattering parameters) into AWR

Component name	From synthesis	Tuned circuit	Fit to HFSS
C_{c0}	0.234 pF	0.3 pF	0.31 pF
C_{c1}	0.246 pF	0.246 pF	0.27 pF
C_{c2}	0.694 pF	0.599 pF	0.61 pF
C_{c3}	N/A	0.08 pF	0.074 pF
C_{c4}	N/A	0.116 pF	0.157 pF
L_{array}	400 pH	393 pH	388 pH
C_{array}	1.265 pF	1.199 pF	1.119 pF
R_1	15 Ω, 73.16°	17.7 Ω, 73.8°	15.2 Ω, 73.3°
R_2	15 Ω, 73.29°	15.8 Ω, 74.9°	15.2 Ω, 73.3°
R_3	N/A	40 Ω, 21.4°	40.2 Ω, 21.9°
R_4	N/A	40 Ω, 7.4°	40.2 Ω, 7.4°
C_{csvr}	N/A	0.1 pF	0.044 pF
Pump feed line	N/A	N/A	50 Ω, 171°@13 GHz

Table 4.1: Table of component values for three different stages of 15% bandwidth, 3-pole, 1 dB ripple Chebyshev, 6.5 GHz amplifier design.

microwave office. It's worth noting that in principle we could put a lumped inductor of the correct value into HFSS of the same dimensions of the port that I show in Fig. 4.23(b). There is a good reason to do this, which is that HFSS's meshing algorithm favors meshing areas where electric fields are concentrated. By placing the lumped inductor at the port, we would move the core mode resonance to the amplifier's center frequency, therefore ensuring it was meshed properly. However, the problem is that this method requires a separate simulation for each core mode inductance. This means that if you wanted to verify that the core inductor participates heavily in the amplifier by varying its inductance (we do) then you need to run *dozens* of simulations. If it's meshed well, each simulation takes at least an hour. Waiting over 12+ hours for each specific layout change is simply too expensive. This is why I favor using the lumped port approach, and instead of letting HFSS determine the meshing, I determine it myself by setting a surface distance mesh of about 10 µm on each of the core mode components.

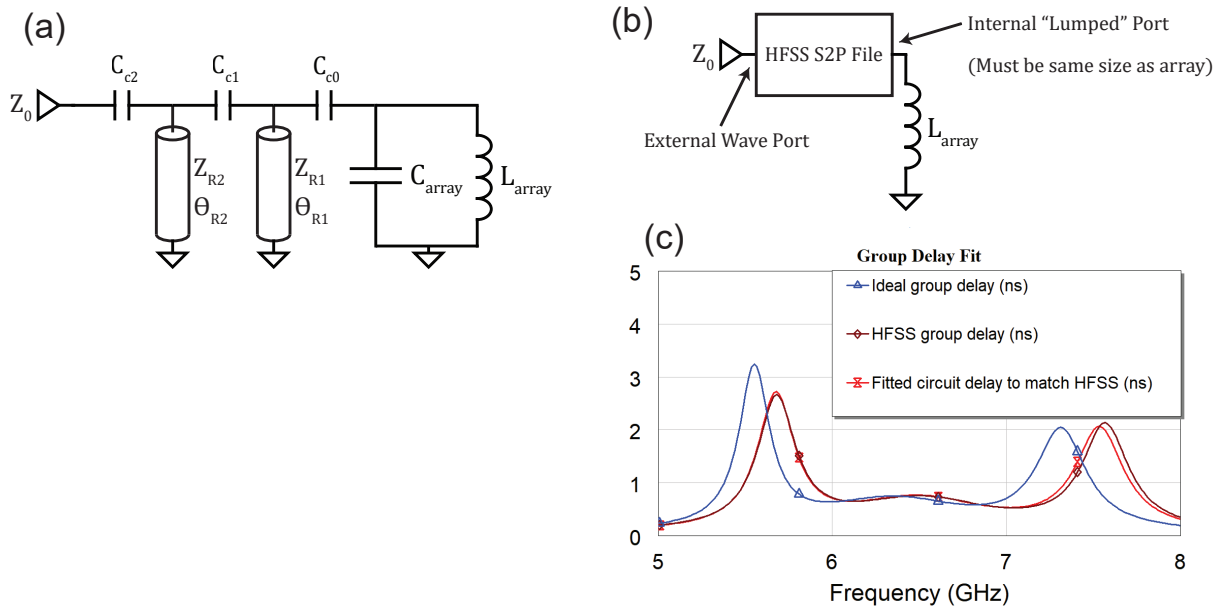


Figure 4.24: Fitting a multi-pole amplifier with exported scattering from HFSS in AWR Microwave Office. (a) The model we use to fit the circuit, with initial guesses based on the tuned circuit values. (b) The HFSS model imported from a touchstone file, loaded by the array inductance. (c) Fitting the group delay from the model (which includes the loaded inductor) to the HFSS simulation shown in (b), and comparing them both to the ideal tuned circuit delay. This fitting process determines the circuit model we will use for modelling the gain.

4.4.6 Fitting simulation results to a circuit model

In the next stage, we compare a model shown in Fig. 4.24(a) to the HFSS simulation of the layout shown in Fig. 4.23(b) via the exported scattering parameters and a load inductor representing the biased array inductance shown in Fig. 4.24(b). The result is a good fit to the group delay in Fig. 4.24(c) with the values in the third column of Table 4.1. We need to fit this model for the last and most important validating step of this process, which is a check to ensure that the amplifier circuit layout that is simulated in HFSS can get enough gain.

The goal of the following procedure is to determine the small signal gain of the amplifier while pumped. The most straightforward way to do this in regular microwave engineering, say to simulate a mixer, is to put all of the external drives into either a nonlinear (periodic steady-state simulation and/or a harmonic balance simulation, then analyze the output components in Fourier space to determine the scattering parameters. The problem with this

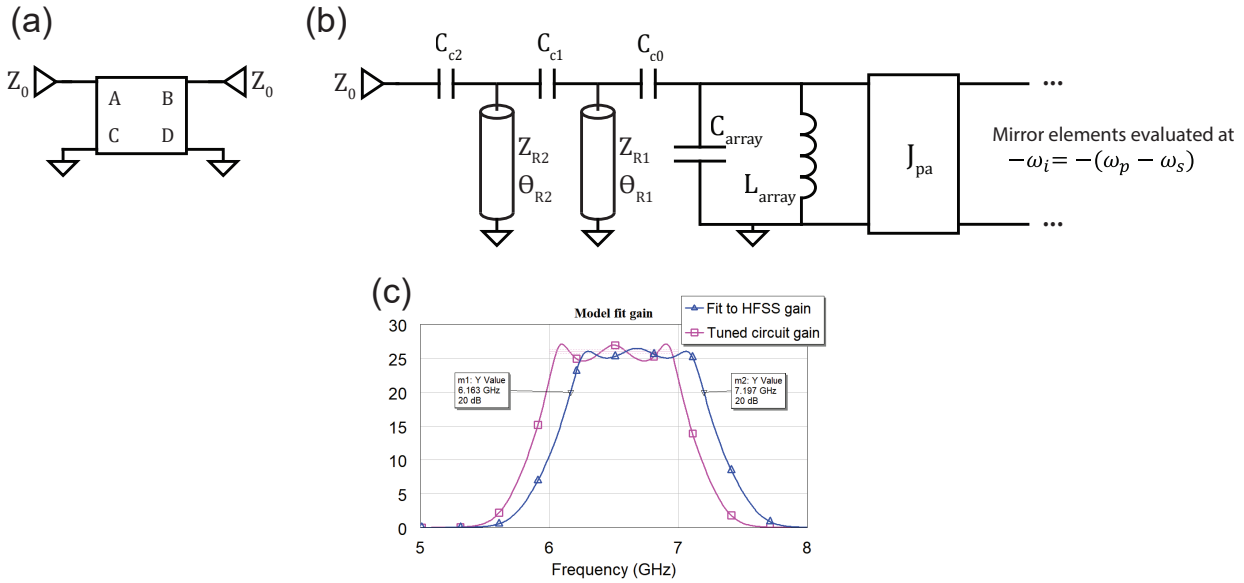


Figure 4.25: Modelling the gain of a simulated broadened amplifier layout. (a) An example of an "idler frequency" circuit element defined by a custom ABCD matrix in AWR Microwave office. (b) An example of a full model of an actively pumped broadened JPA, with the entire idler half of the circuit described by the appropriate circuit element mirrors, all made in the fashion of (a). (c) The result of tuning the center admittance inversion constant (labeled "J_fit" here in the circuit diagram) and the pump frequency (which enters into the model via $\omega_i = \omega_p - \omega_s$ in (a)) to produce gain.

approach in Josephson junction-based circuits is that, until very recently [42], there was not an easily accessible junction model based on real junction parameters available to easily plug into a harmonic balance simulation. Additionally, writing down the equations of motion for each of the three nodes of the system and solving them in a fashion similar to Chapter 2 was not practical in the time available. So with that preface, here is the procedure I thought up to calculate the small signal gain of an HFSS-simulated amplifier layout without access to a nonlinear simulation.

In Fig. 4.25(a), I show how to make the first essential ingredient in modeling an "idler frequency" side of an amplifier in an AWR Microwave office circuit. Because we essentially need to fool Microwave Office into doing a low-order harmonic balance simulation, we can't use the global frequency option that we normally would. In this case, I define the signal frequencies as an array in a FREQSWP object as "omega_s." Using this as a starting point (and the x-axis in all the plots) I define the idler frequencies as "-(omega_p-omega_s)"

in line with Ref. [41], then evaluate all the idler frequency elements at that frequency. In Fig. 4.25(b) I show how these are connected in AWR. Because they are defined as ABCD matrices, we have to choose an order for the circuit. We follow Ref. [41] here, but any two parallel elements (except the core inductor and its neighboring capacitor) can be freely exchanged. The last step in this process mimics the tuneup of the real amplifier. We have to decide on the pump frequency and the core inverter value J_{pa} that links the signal and idler side of the network, which is a proxy for the pump flux and inductance modulation that we need to use in the real device. By looking at Ref. [41] and doing some algebra, we can see that this inverter value is related to the inductance modulation in Eq. 4.30 by

$$J_{pa}Z_{core} = \frac{\sqrt{\alpha}}{1-\alpha} = \frac{\delta L}{2L_{array}\left(1-\left(\frac{\delta_L}{2L_{array}}\right)^2\right)}. \tag{4.45}$$

At most, $\delta L/L_{array}$ can reach one, so $\alpha = \delta L/(2L)$ is capped at 0.25. This means that for the amplifier to be physically realizable, the product $J_{pa}Z_{core}$ must be lower than 2/3 for the amplifier to be physically realizable. In the case of Fig. 4.25 the core resonator impedance is $\omega_0 * L_{array} = 16\ \Omega$, and the fitted $J_{fit} = 0.016$, so the product is $J_{pa}Z_{core} = 0.253$, well within the bound. If we work out the required inductance modulation δ_L/L_{array}, it works out to a quadratic equation with two roots, $\alpha = 0.057$ or $\alpha = 17.5$. The second one is not physical, but the first is and we can use it to work out $\delta L/L_{array} = 2\sqrt{\alpha} = 0.46$, about the middle of the range of $\delta L/L_{array}$. What this means is that if the operating inductance L_{array} is in the middle of the rf-SQUID array inductance modulation range (a fair assumption) then you have to be sure that the inductance can modulate by roughly 50% in either direction, which means the resonant frequency has to modulate by the square root of that, about 70% in either direction. This is a tall order, but an rf-SQUID with a $\beta = 0.75$ can achieve this. This emphasizes that having a minimal external series inductance is as critical as a large β because an appreciable amount of external series inductance will cut off the top of the inductance modulation range and prevent the resonant frequency from modulating over a wide enough range. Once the model is simulated in this way to produce gain over an acceptable bandwidth with an achievable inductance modulation, the signal-side engineering of the layout is done.

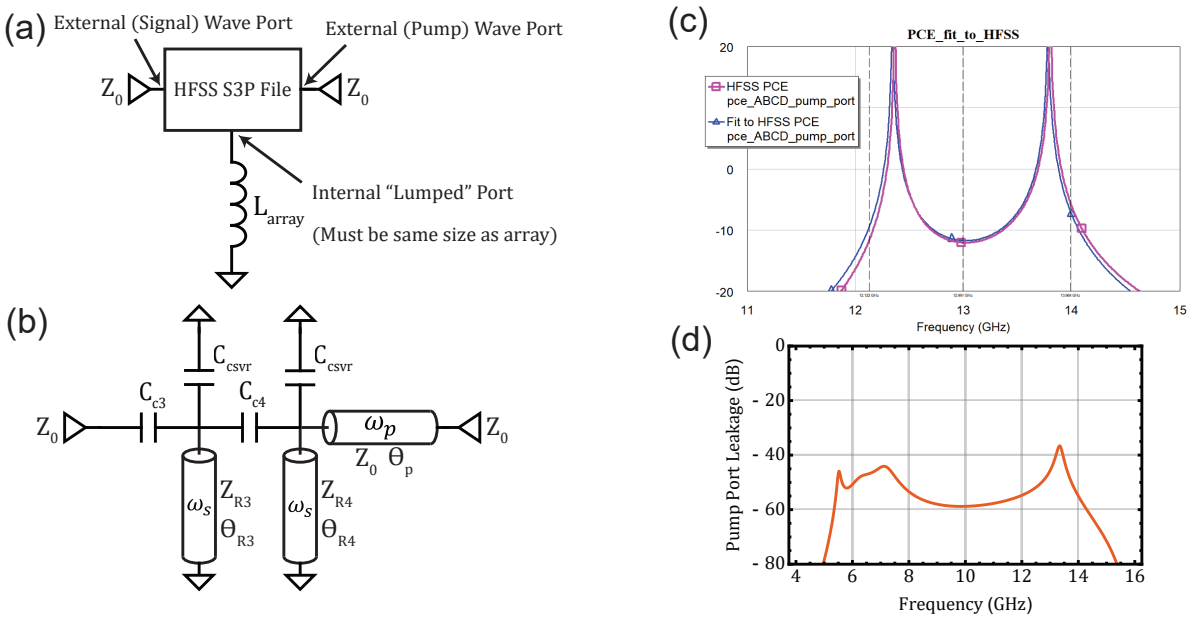


Figure 4.26: Fitting two-port phase coupling efficiency. (a) The setup in AWR Microwave Office for extracting the phase coupling efficiency directly from voltage ratios of an exported HFSS simulation. (b) The fitted circuit to the voltage ratio extracted via (a), with circuit values in Table 4.1. (c) The fit line in blue using Eq. 1.71 vs. the voltage ratio extracted from (a). (d) the pump leakage calculated using Eq. 4.12.

4.4.7 Pump port simulation

In order to calculate the pump port scattering, we can design a circuit that branches to the pump port from the core resonator, and validate the PCE from chapter 1 Eq. 1.71 via Microwave Office in a similar way. Ideally, this pump port will present such a large impedance at the signal frequency that it would not influence the signal-side matching network, so I will begin by considering only the signal-side network's effects on the PCE, and not vice versa. As a starting point, I used ANSYS NuHertz to generate a 2-pole Chebyshev high-pass filter with a cutoff frequency of slightly less than the pump frequency to generate the filter shown in Fig. 4.26(a). Using Eq. 1.71 on this circuit in addition to the terminated signal-side network's ABCD parameters, we can fit the blue line in Fig. 4.26(c) to the pink line that represents the simulation data.

Finding that this PCE is roughly 10 dB better than the coupling capacitor in chapter 2 gives reasonable confidence that the pump will be sufficiently coupled to the core mode to drive parametric gain. The fitted circuit includes a pump feed line which was essential to fit

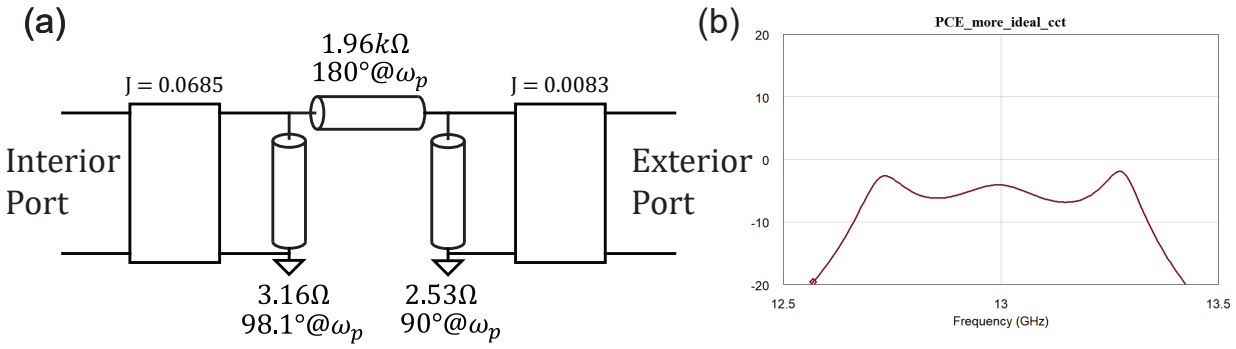


Figure 4.27: A circuit with a much more even pole-free PCE near the pump frequency. (a) A three-pole circuit with a Chebyshev-like PCE. (b) Plot of the PCE for the circuit in (a) with the same setup as the real pump coupling circuit in Fig. 4.26(b).

to the simulation data from HFSS, but was initially absent in the filter design. Additionally, future work must avoid the poles in the PCE response that we see in Fig. 4.26(c). As a proof of concept, I supply an example of such a circuit in Fig. 4.27(a) with a plotted pole-free PCE in Fig. 4.27(b). The impedance of the middle resonator (here shown as a series $\lambda/2$ element instead of a shunt $\lambda/4$) is impossibly high. Still, it can be easily changed with the addition of additional admittance inverters on either side.

4.5 Results

In this section, I will detail the initial results for the broadband circuit in Fig. 4.23 and the problems that the circuit encountered that we did not anticipate, with a discussion of its two biggest problems: flux nonuniformity and parasitic series inductance. Then I will discuss measurements that I took on a similar broadband amplifier during a student research internship at Google Quantum AI in Fall 2022. This device uses a slightly different version of an rf SQUID array that we talked about in detail in chapter 2, so I will first introduce it briefly and make a direct link to that previous work, and then discuss measurements of the broadband gain and saturation profiles, noise performance, and intermodulation measurements that tie back into the groundwork in chapter 1.

4.5.1 Flux nonuniformity: justification for on-chip flux lines

The simulation and design of rf SQUID array amplifiers depend on a fairly even distribution of the bias flux across each of the individual rf SQUIDs in the array. How do we tell if this is an accurate assumption in the real experiment? We got a small sample of this in Fig.2.9, which is that when scanning the flux modulation of the device there will be distortions that can be attributed either to flux offsets or flux nonuniformity. Flux offsets are the easier of the two: the peak of the flux modulation will be moved away from the origin, as in Fig. 2.8(a). Flux nonuniformity is more subtle, and manifests as distortions in the flux modulation with an overall envelope to the modulation controlled by the magnitude of the nonuniformity of the external flux. To quantify exactly how uniform, we can simply consider a sum of all of the inductances at each flux, with some SQUIDs getting a factor more bias flux than the average, and other SQUIDs less.

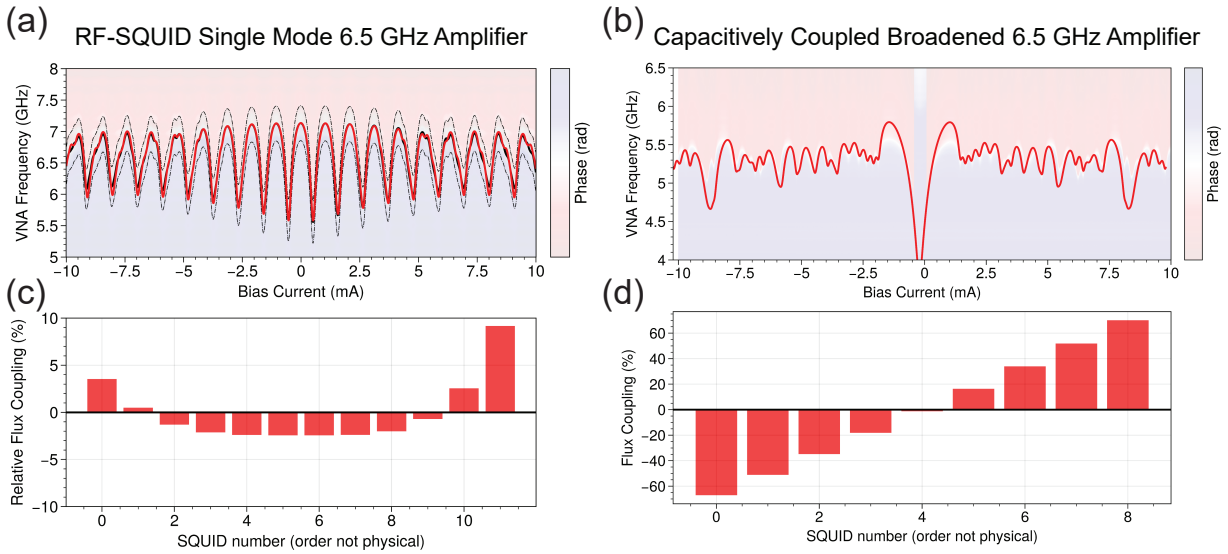


Figure 4.28: Flux nonuniformity fitting to amplifier phase modulation profiles. (a) The flux nonuniformity of a typical resonant, single-pole JPA. (b) The flux modulation profile of the broadened circuit shown in Fig. 4.23(b). (c) The plot of the fitted nonuniformity that generates the red line in (a) indicates a reasonable $< 10\%$ variation across the array. (d) The fit to the nonuniformity that generates the red line in (b), where we see that the much larger variation on order of 50% nearly completely destroys the flux modulation profile of the device, with less than a half quanta remaining accessible before the nonuniformities interfere and curtail the flux modulation.

Figure 4.28 shows that the changes in flux nonuniformity can be drastic in a broadband design even without the layout near the array changing that much. This suggests that

merely controlling the dimensions of the spindle magnet delivering the flux and the local geometry around the array is not enough to guarantee flux uniformity across the array, and that large screening currents may be flowing around the perimeters of the cutouts in the ground plane of the chip. This motivates the creation of on-chip flux lines that run parallel to the array and terminate in the ground plane. If done carefully, such as being designed with either a very high or very low characteristic impedance, these flux lines would couple minimally at the signal and pump frequency while opening a DC mutual flux coupling to the array SQUIDs. Modified prototypes that incorporate on-chip lines are currently being fabricated for testing, and future work will determine if that method is more effective. In the meantime, Fig. 4.28(b) captures everything that crippled these devices. Not only did the flux nonuniformity prohibit the large inductance modulation required to generate the broadband gain, the center of the available inductance modulation range (where it would need to be biased for maximum inductance modulation) is a full 1.5 GHz away from the center of the matching network response at 6.5 GHz. These difficulties are currently being remedied in a series of follow-up designs that include fast flux lines, and varying core inductances.

4.5.2 Brief introduction to RF-SQUID SNAKE amplifiers

In Fall 2022, I participated in a student research internship where I measured the performance of multiple broadband, high saturation power amplifiers based on rf SQUID arrays. These arrays were first introduced in Ref. [40] as elements of high saturation power cryogenic mixers, and they have a few advantages over the basic rf SQUID arrays I introduced in chapter 2. In this section, I will briefly introduce these arrays, discuss their various engineering advantages, and make a connection back to the more basic rf-SQUID arrays I discussed in chapter 2.

Figure 4.29(a) shows the circuit diagram of an rf SQUID "snake" amplifier from Refs. [69, 23]. We can see the similarity to the rf SQUID arrays from chapter 2 in Fig. 4.29(b), where I have split the bridging inductor between the SQUIDs into two parallel elements of twice the inductance. This lets us split up the individual SQUIDs to see with the substitution $\beta = 4L_1 + L_2$ that individually they share the hysteresis point of the rf SQUIDs in chapter 2,

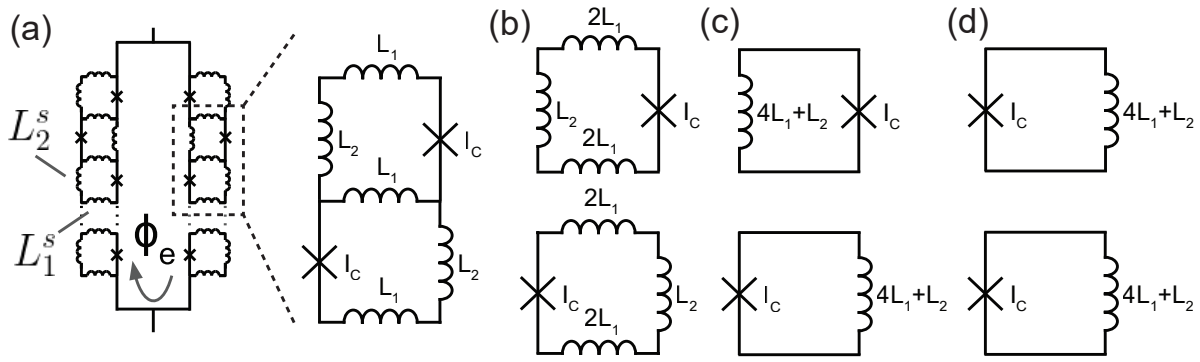


Figure 4.29: Breakdown of rf-SQUID SNAKE design.(a) Diagram of the rf SQUID SNAKE design reproduced from Refs.[40, 69, 23], with a focus on a single unit cell of the design. (b) Splitting the unit cell into two equal-inductance segments. (c) Combining the inductances in each segment to a single inductor to see the similarity to a single-segment rf-SQUID. (d) Rearranging the orientation to a ferromagnetic orientation of mutual inductance to make the setup identical to that of Fig. 2.1.

in other words the junction inductance competes with the total inductance in each loop as shown in Fig. 4.29(c). However, the analogy is not exact enough to determine the inductance with respect to the external flux bias as easily. The inductance of the snake array is given in Refs. [40, 69] as

$$L_{\text{snake}} = \frac{N\left(\left(L_1 + L_2\right) L_j + L_1 L_2 \cos\left(\delta_0\right)\right)}{2\left(L_j + \left(4L_1 + L_2\right)\cos\left(\delta_0\right)\right)} \tag{4.46}$$

If we parameterize this in terms of the total shunt inductance of each rf SQUID shown in Fig. 4.29(c) as $L_s = 4L_1 + L_2$ then we can see more clearly the comparison to the rf SQUID arrays that we had before with

$$L_{\text{snake}} = \frac{N}{2}\frac{(2L_1 + L_2)^2 + L_1 L_2(1 + \beta\cos(\delta_0))}{L_s(1 + \beta\cos(\delta_0))}. \tag{4.47}$$

Looking at this version of the inductance we can see the competition between three different current paths through the snake array. In one case, where $L_2 >> L_1$, we have the case that is truly identical to the rf-SQUID arrays in chapter 2 with a total shunt inductance $L_s = 4L_1 + L_2$ and the shunting ratio $\beta = L_s/L_j$ producing the inductance

$$L_{\text{snake, 1}} = \frac{N}{2}\frac{L_s}{1 + \beta\cos(\delta_0)}. \tag{4.48}$$

In a second case, we could imagine a reversal of the first, where $L_1 >> L_2$. In this case, we find that

$$L_{\text{snake, 2}} = \frac{N}{2}\frac{L_s}{4(1+\beta\cos(\delta_0))} \tag{4.49}$$

In the third and final case, one where β is very small, we have the situation where almost none of the current goes through the junctions, and the entire inductance of the array loses all flux dependence and approaches

$$L_{\text{snake, 3}} = \frac{N}{2}(L_s - 2L_1). \tag{4.50}$$

The major difference between the snake and a regular rf SQUID array is that the extra degree of freedom between L_1 and L_2 means that these three cases all have *different* amounts of total linear inductance in the low-β limit, and so the ratios between L_1, L_2, and L_j must all be provided to specify the current flow. Two of the paths (case one and two) favor the Josephson junction loops in some way, with case one favoring the JJs and L_2 and case two favoring the JJs and L_1. Case three corresponds to the remaining current path, where the junctions are avoided altogether and L_1 and L_2 are favored by the current.

In the devices from Refs. [69, 23] that I focus on in this section, the values of the inductances are $L_1 = 2.6$ pH, $L_2 = 8.0$ pH, and $L_j = 20$ pH, which means that of the three cases I present here, we can allow the intuitive explanation from case one to reign, and view the rf SQUID snake array as acting mostly like a regular rf-SQUID array discussed in chapter 4, but with a comparatively large $\beta = 0.9$.

The last major difference that I should comment on before moving to the data is that the arrangement of the orientation of the SQUID array between Fig. 4.29(c) and Fig. 4.29(d) makes an important practical difference. The arrangement in Fig. 4.29(c) is convenient for layout, but means that a global magnetic field will shift neighboring rf SQUIDs in *opposite* flux bias directions. This will create an un-compensatable flux difference between SQUIDs in the array that will reduce its flux modulation range, and therefore also reduce the maximum inductance modulation $\delta L/L$, which will cripple the device's ability to reach 20 dB gain. Therefore, magnetic shielding is critical for this anti-ferromagnetic arrangement of the rf SQUIDs, but somewhat less so for the ferromagnetic arrangement in Fig. 4.29(d).

Despite this initial cost, as a whole the snake array is an improvement over the rf SQUID arrays I presented earlier in chapter 2. The reason is that it easily facilitates a uniform flux bias via the flux transformer that delivers ϕ_e in Fig. 4.29(a). This flux transformer not only allows for easy, uniform DC flux biasing that solves the problems presented in Fig. 4.28(b), but also amplifies the flux at the pump frequency as well, increasing the PCE of the device.

4.5.3 Broadband Amplifier Measurements

Now that we have an idea of what the rf SQUID array accomplishes and its similarities to the rf SQUID arrays in chapter 2, we can investigate how it performs when integrated into a broadband amplifier design. This section will draw heavily from my publication in Ref. [23], but I will try to offer some new insights into the data , offer some perspective on how future work could improve on these results, and the various difficulties I ran into along the way.

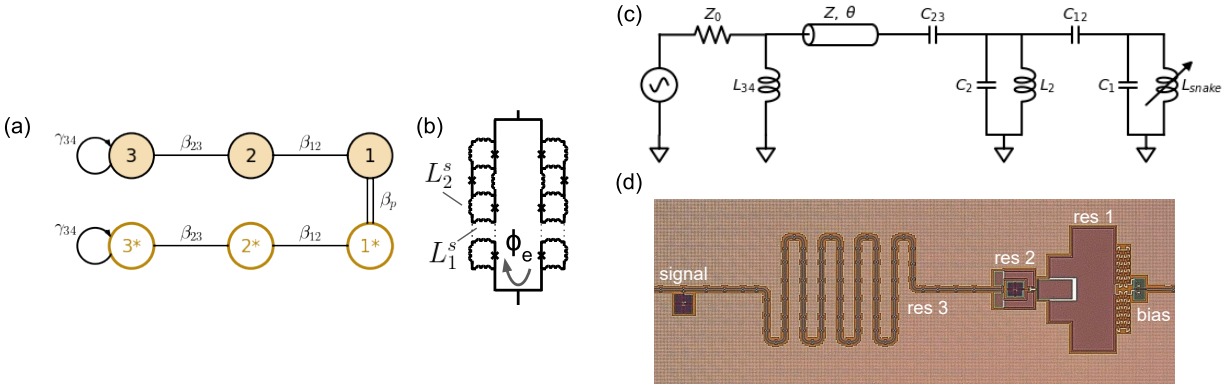


Figure 4.30: Schematic and layout of the Lumped-Element Snake Amplifier (LESA). (a) The coupled-mode diagram of the broadened circuit, showing both the signal and idler components of the network, which live in the same physical resonators. (b) Zoomed in schematic of the snake array, discussed in the last section. (c) Full circuit diagram of the LESA device, with resonator one being a series transmission line resonator, and the others being lumped. (d) Picture of the LESA device, where we can see the relative sizes of all the resonators. Since this is a multi-layer process, the resonator and transmission lines are all healed by crossovers.

4.5.3.1 Lumped Element Snake Amplifier (LESA) Design

Figure 4.30 shows the design and layout of the 3-pole Lumped-Element Snake Amplifier (LESA) device which is the star of this section. Although I did not personally design the

device, we can understand it within the framework that I laid out earlier in the section. The component values are all given in Ref. [23], but the short version is that this is a 13.5% bandwidth, 0.5 dB ripple Chebyshev designed around a center frequency of 4.9 GHz. The most interesting thing about the design is the choice of the first inverter and resonator. The idea behind using a shunt inductor as the first inverter rather than a series capacitor as I showed earlier is related to the wirebond inductance. With a series capacitor, the often quite large (≈ 0.5 nH) wirebond inductance will resonate with this capacitor, and as we can see in Table 4.1 the most external capacitor is often the *largest* of all the coupling capacitors. This is a recipe for disaster because the accidental resonance of that the wirebond inductance and the coupling capacitance is around 9 GHz, which is dangerously close to the amplifier resonance. By replacing this inverter with a shunt inductor, we can avoid this resonance entirely. Additionally, this means that the (again, rather large) wirebond inductance will only form a frequency-independent (!!) voltage divider with the shunt inductance. In the case of the LESA devices, this voltage division favors the LESA, with a shunt inductance of $L_{32} = 1.32$ nH giving a voltage ratio of about $L_{shunt}/(L_{wb} + L_{shunt}) \approx 0.72$ assuming a wirebond inductance of 0.5 nH.

The transmission line resonator that bridges the shunt inductance inverter to the rest of the circuit was a custom element worked out in Appendix C of Ref. [23] specifically to accommodate the inductive input coupling. I will leave the details there, but the reason why a transmission line resonator is convenient for this purpose is that it can be compensated both capacitively and inductively by trimming the length of the resonator by an amount that corresponds to its characteristic impedance. High-impedance transmission line resonators are easily able to compensate for series inductance with changes in length, and low-impedance resonators are well-suited to compensate for shunt capacitances by small changes in their length. By choosing an intermediate impedance (in the LESA design the impedance is 50 Ω), we can do both, allowing a transition from a capacitive series admittance inverter to an inductive shunt inverter at the cost of large nonidealities that arise from somewhat large changes in length. In the case of the LESA, an initial 90° line was cut all the way down to 38.6°, then had to be trimmed by about 20% to 32.6° to compensate for the various inverter nonidealities in the circuit.

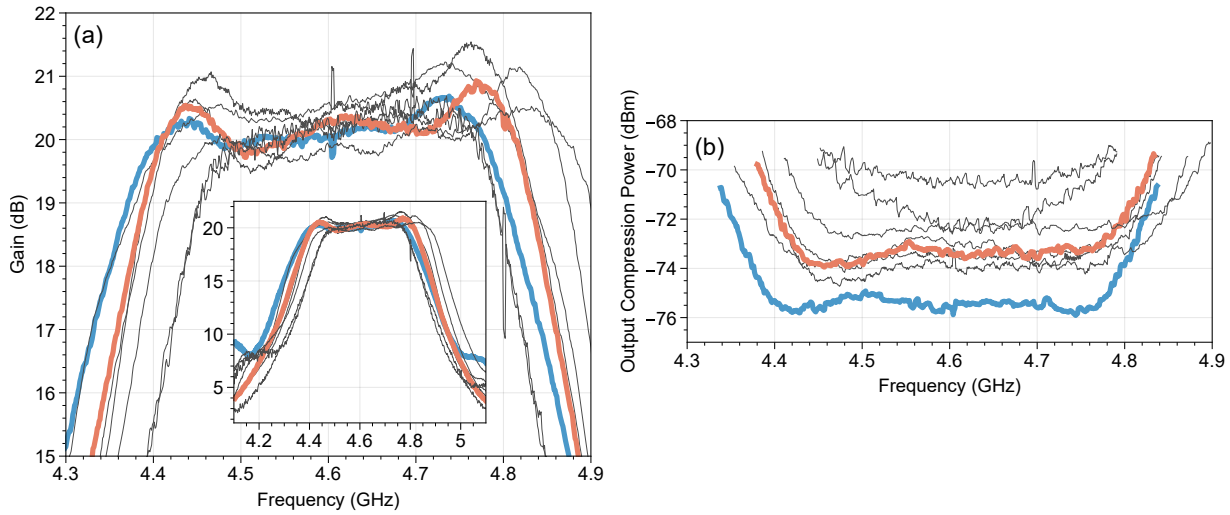


Figure 4.31: Gain and saturation power of eight different LESA devices. (a) The gain profiles across eight separate nominally identical LESA devices measured in a single cooldown, all connected to a 53-qubit Sycamore processor. Devices A and F, named for their output line names, are highlighted in blue and orange respectively. (b) Output-referred saturation power of the same devices, with LESA A and F highlighted.

4.5.3.2 Gain and saturation power

The LESA devices we measured have a roughly similar tuneup procedure to the rf SQUID parametric amplifiers I showed in chapter 2. First, the device is flux-tuned via the bias line and flux transformer in the right side of Fig. 4.30(d) so that the phase drop across each SQUID is about $0.35\phi_0$, which moves the core mode resonance to approximately the center frequency of the amplifier between $4.5 - 4.9$ GHz. Next, a pump is turned on near $2\omega_0$ and the gain profile is hand-tuned to achieve the level 20 dB gain response in Fig. 4.31(a). The optimization space of these amplifiers is not always so clear, there are often points where the gain profile is very similar, but the pump detuning and flux bias are adjusted to counter one-another to produce a slightly different profile at a slightly different gain. Generally, the profiles in Fig. 4.31 were adjusted to maximize bandwidth at 20 dB gain. We also checked these by scanning the parameter space of flux bias, pump detuning, and pump power and minimizing a cost function formed by comparing the device response to a flat 20dB response over a 400 MHz bandwidth and found that the hand-tuned profiles were very near the optimal points in a fairly broad plateau of reasonable bias points in the design space.

After tuning the particular bias points, the VNA input power, calibrated to photon number by a procedure identical to that in Fig. 2.10, was turned up until the gain deviated by 1 dB from the original value. The saturation power of the LESA devices is uniformly excellent, with even the lowest of them outperforming single-junction devices spectacularly, with similar saturation powers to the devices in chapter 2 over the entirety of their bandwidth. Interestingly, the LESAs uniformly experienced a very small 0.2 dB rise in gain across the entire bandwidth, before falling normally to 19 dB. This can be attributed to a slightly negative pump detuning and overall negative Kerr coefficient using logic from Ref. [33]. There is an interesting line of inquiry here about the eigenmode structure of a multi-pole amplifier. For example, is it only the Kerr coefficient of the *core* mode that matters? How are the eigenmodes distributed in frequency? Are the cross Kerrs between these modes all the same sign and/or similar magnitude? These questions will have to be left to future work.

4.5.3.3 LESA noise performance and readout efficiency

As with the devices in chapter 2, we have to measure the readout chain efficiency of the LESAs to understand if they are effective preamplifiers and to get an idea of the losses in the readout chain. Figure 4.32(a) benchmarks the readout efficiency of all eight LESA devices across the entire 53-qubit Sycamore processor they are collectively connected to. This efficiency calibration procedure is detailed in Ref. [69] and is similar to Ref. [6]. The essence of it is that after calibrating the χ and κ of the qubit readout resonator, we can record how quickly the SNR of the measurement increases with integration time, and from this rate, we can extract the quantum efficiency of the measurement.

One important aspect of operating a high saturation power amplifier over its full (half) bandwidth with multiple multiplexed channels is how the noise responds to gain saturation. One can imagine, for example, that the noise might increase independently of the gain in some way, and so it is important to record not only the gain compression but also the *noise* compression of the amplifier chain. In Fig. 4.32(b) we do exactly this for LESA A. An external blocking tone is used to saturate the amplifier, with a vertical dotted line indicating the input saturation power for ease of reference. During saturation, we measure both the ground and

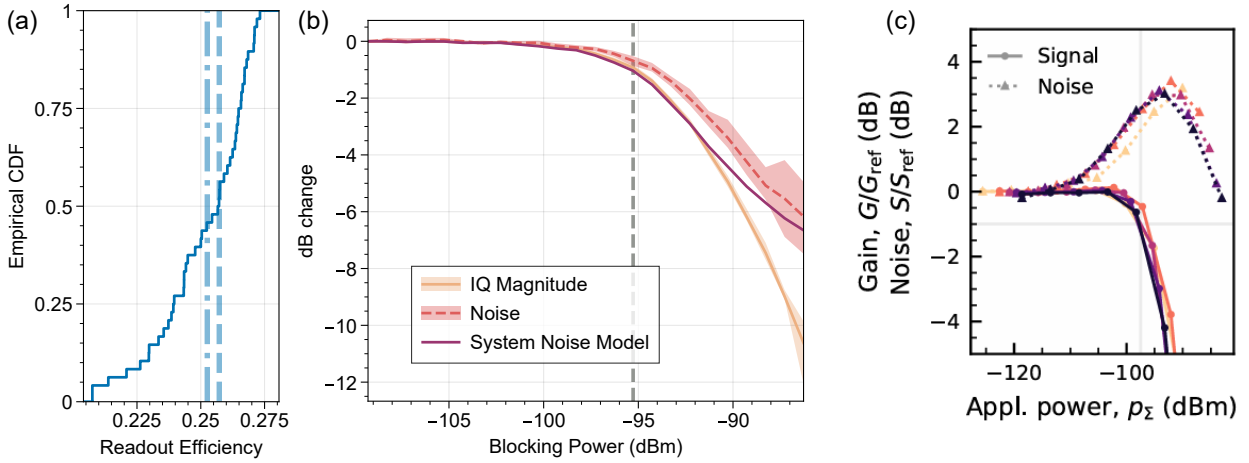


Figure 4.32: Readout efficiency of LESA devices and noise performance near saturation compared to TWPA devices in Ref.[50]. (a) Cumulative Distribution Function (CDF) of the readout efficiency across the entire 53-qubit Sycamore processor, with a median of 0.257 and mean of 0.253. (b) Measurement of the noise during the compression of the LESAs from an additional phase-locked blocking tone delivered at the edge of the idler side of the amplifier bandwidth. Solid lines represent the mean over the qubits, and the shading represents the range of the data. A system noise model that accounts for gain compression alone. is shown in purple. (c) Noise performance of a typical TWPA, reproduced from Ref. [50], showing a significant noise rise near saturation

excited state IQ histograms of the qubit readout resonators, optimized for optimal fidelity beforehand. We then record the decibel change of the means of the histograms average across all of the qubits in yellow (range shown in light shaded yellow) as well as the decibel change of their standard deviations as a representation of the noise (pink dotted line, pink shading). What we find is that the noise saturates *slower* than the gain, which is *not* explained by a noise model that only accounts for gain compression. This indicates that near and well into saturation there is another noise source not well accounted for in the model. Notably, although this seems bad, I point out in Fig. 4.32(c), reproduced from Ref. [50], that it is not *nearly* as bad as it is for TWPAs. In travelling-wave devices (which do not reject any parametric process other than those with one wave in the stopband) there is a large noise rise right at saturation. This means that their SNR saturates well before their gain. I view this as a chink in the TWPA's broadband armor. Because they permit *so many* parametric processes in their broadband architecture (again, shown by the multitude of intermodulation products in Ref. [50]), they allow many ways to introduce more noise back into the signal frequency. Since intermodulation products drastically increase near saturation, the noise

gets worse exactly where it hurts the most, right near gain compression.

One could level the exact same criticism at broadband reflection amplifiers, however I will point out they take the *exact* inverse strategy of TWPAs. Rather than permitting everything except a stop band, they reject everything outside of the filter bandwidth, so much so that the pump has to have a special route carved out to reach the core at all. This means that they permit a considerably smaller number of intermodulation products, and we can speculate that this could be why they add less noise near saturation. The Goldilocks resolution to this argument between bandwidth and noise could be something like a TWPA with a prescribed bandwidth, which combines both approaches.

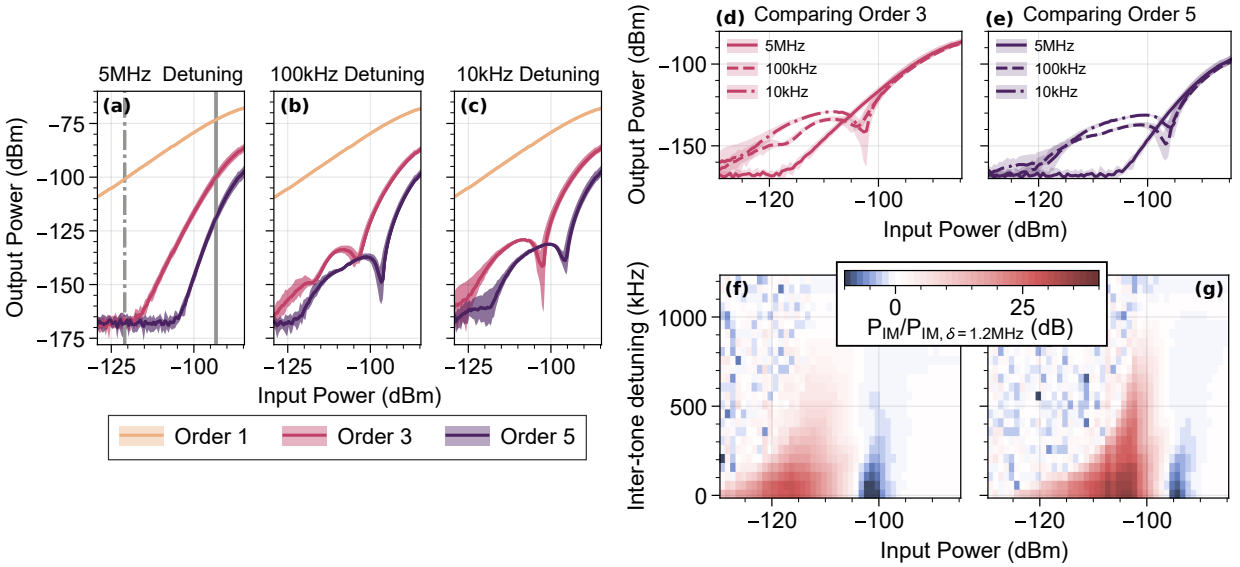


Figure 4.33: LESA IMD behavior at various inter-tone detunings, varying input power. (a)-(c) The fundamental frequencies (f_1, yellow), 3rd order IM3 ($2f_1 - f_2$, pink) and 5th order IM5 ($3f_1 - 2f_2$, purple) with setup that matches Fig. 1.9(b), with an inter-tone detuning of (a) 5 MHz, (b) 100 KHz, and (c) 10 KHz. The center frequency of the two-tone measurement is swept over the lower half-bandwidth of the amplifier on the opposite side of the readout resonators, with the shading representing the range of the data over those values and solid lines representing the averages over center frequency. (d) Comparison of 3rd order IM3 at various inter-tone detunings. (e) Comparison of 5th order IM3 at various inter-tone detunings. (f)-(g) Excess IMD product power, compared to large inter-tone detuning (1.2 MHz) as a function of input power and inter-tone detuning for (f) IM3 and (g) IM5 products. The color bar is somewhat skewed to highlight negative dips more than positive excess.

To try to understand this mechanism of added noise further, we decided to investigate the intermodulation products of the LESA amplifiers directly, and we found a fair few surprises that took some time to explain. In Fig. 4.33(a), we see the type of intermodulation power behavior that I outline in chapter 1, with the primary having a slope of one, IM3 with

slope 3, and IM5 with slope 5. This behavior is what would be of practical importance in measuring the readout resonators with these amplifiers, as the resonator detunings are of order 50 MHz in this setup. We can see from this that at standard readout powers shown by the gray slash-dot line, the total harmonic distortion (the sum of all of the intermodulation tones compared to the primary tone) is less than -70 dBc, which is excellent compared to TWPAs, which in Ref. [50] are shown to have a total harmonic distortion of at least -40 dBc, depending on how one adds up all the products.

Again, this seems like a consistent explanation, and we were satisfied with this result. However, there is a wrinkle in the intermodulation performance of these devices. Intermodulation distortion is only relevant when the inter-tone detuning is comparable the channel width, as we depict in Fig. 1.9(b). Since most telecommunication occurs over bandwidths that reflect the audio range of our ears ($\approx$ 10 KHz), intermodulation measurements are historically done at much lower detuning, in the 100 KHz to 10 KHz range. When we lowered the intertone detuning to this level, partly to check that our measurements were working correctly, we saw the behavior in Fig. 4.33(b) and (c), where not only was there much more distortion compared to the large-detuning level (as shown in Fig. 4.33(d) and (e)), the distortion had a much more intricate power dependence than just a different slope. This is interesting for a few reasons. For an intermodulation product to be changing in slope on a log-log plot, that means that it must have a power dependence that is not just a single power, but rather multiple different power dependencies that are changing. This suggests that two different physical sources might govern the power dependence. There is the original amplifier nonlinearity, anticipated in chapter 1 and shown in Fig. 4.33(a), but also another power-dependent effect. When the two are equal in magnitude, they produce a dip when they destructively interfere. Otherwise, they add and produce extra distortion. What could this be?

It was around this point that I ran out of time in my student research project and left Google. We had a few ideas about what this could be at the time. Was it a high-order intermodulation product interfering with the amplifier pump? Or maybe a nonlinearity further up in the readout chain? The problem with these explanations is that they were too linear, and none of them could produce more than one dip in the third order intermodulation

as shown in Fig. 4.33(b). We called Mark Dykman and his researchers as cavalry to help work through this issue. They worked out that the explanation that best fits the data is saturable nonlinearity in two-level system (TLS) defects in the device capacitors. Since a TLS bath will saturate at some level of input power, it makes sense that at some point, the amplifier intermods eventually have to win out and once again dominate the distortion at high input power. Since the saturable nonlinearity levels off, it is present in all orders of the intermodulation at varying strengths. Lastly, the coherence of the TLS bath offers an explanation for why there is no excess IMD effect at large inter-tone detuning. In our case, the most prominent dip in the TLS is explained by a TLS $T_2 = 2T_1 = 4$ µs, meaning that if the tones are detuned by more than $1/T_2 = 250$ KHz, then the loss becomes more and more time-independent, as the TLS bath can no longer track the modulation of $f_2 - f_1$. So, in short, the saturability of the TLS bath explains the input power dependence and their coherence explains the inter-tone detuning dependence.

5.0 Conclusions and Perspectives

In this thesis, we have demonstrated large improvements in both the saturation power and bandwidth of resonant Josephson parametric amplifiers. These improvements are vital in scaling the high fidelity, quantum efficient readout that JPAs are known for to match the needs of error correction codes in larger and larger quantum processors. In chapter 2, we provided a simple method to increase the saturation power of JPAs more than 100-fold over single-junction devices with only 25 rf SQUIDs, and demonstrated devices with -93 dBm input saturation power that can process *thousands* of photons without any loss in performance. This rf SQUID array structure is easily fabricated even in simple Dolan-bridge fabrication methods that are widely available in university nanofabs. These devices are robust to small flux nonuniformities and global flux offsets, making them ideal devices for biasing via external magnets in small-scale high-power experiments. Additionally, we show that a frequency diplexer can be used to separate the pump and signal side of the amplifier, allowing a high-power, lightly attenuated pump line to coexist in the same system as a low-power, quantum-limited readout chain. Lastly, we also showed that the high-power nature of these devices (and the hot pump line) do not detract from the lauded efficiency of the JPA by performing a weak measurement experiment and tracking the measurement backaction directly. We measure a quantum efficiency lower bound of 0.62, which is among the highest efficiencies measured in the field. We also show high-fidelity measurement limited primarily by readout-activated state transitions in the qubit at high powers, rather than any property of the amplifier itself. The non-QND nature of strong measurement tones in circuit QED remains an open problem for the field which motivates the pursuit of optimal filtering [24] and a greater understanding of the limitations of dispersive readout in circuit QED [59, 11, 13].

In chapter 3, we use these rf-SQUID amplifiers to turn up the power of qubit measurement drastically, reaching over 20 dB above traditional measurement powers to investigate optimal weight functions for qubit readout and observe high power measurement transitions with high efficiency. These transitions, previously invisible without a quantum-efficient high-power amplifier, are major barriers to increasing qubit readout fidelity beyond a fault-tolerant

threshold. Now, with high-power JPAs, they are easily visible, opening exciting avenues for future study of transmon readout dynamics. As one example, we show that we can effectively train and use ideal weight functions for multi-state, high-power transmon qubit readout that are able to reduce two-state readout errors by up to 20% [24].

With the utility of high-power amplifiers established in a single-qubit case, chapter 4 showed how we can use filter synthesis techniques to design the quantum Langevin equations of motion for a much more broadband gain response in resonant parametric amplifiers. These techniques give a pathway to increase the bandwidth by over 10x compared to single-pole devices from chapter 2. Building on this, we explore how we might use different filter prototypes to achieve the same bandwidth with lower inductance modulation, reducing the pump power required to operate the amplifier. We covered the careful simulation methods needed to design these devices in great depth, and detailed a design for a 15% bandwidth amplifier with a dedicated high-efficiency pump port.

Lastly, we showed how a modified rf SQUID array design, similar to the devices in chapter 2, can be integrated into a similar broadband design, and showed measurements from eight different devices which all successfully achieved over 400 MHz of instantaneous bandwidth with at least -95 dBm input saturation power over the same range. We also demonstrated the quantum efficiency of these devices and showed that their added noise does not experience the rise near saturation seen in competing traveling wave devices. Further, we showed intermodulation distortion measurements that correlate this idea to drastically lower harmonic distortion at common readout powers.

With the innovations detailed in this thesis, resonant parametric amplifiers offer an easy-to-fabricate, high-power, wide bandwidth method of bringing high quantum efficiency readout chains to bear on quantum processors. As quantum processors scale up, we have shown that with careful control and simulation of the device nonlinearity and the impedance profile around resonance, the input power handling and bandwidth of JPAs are able to keep up at current state-of-the-art efficiencies. Future work in the same vein will entail interesting applications of some of the same methods, such as increasing the efficiency of the energy transfer of the pump to the signal band [19]. Ultimately, we hope that the research in this thesis provides a thorough understanding of a reliable device that brings utility-scale

error-corrected quantum computing closer to reality.

While the results in this thesis provide near-term practicality, there is good reason to believe that some major changes in the technology will be needed to increase the quantum efficiency of measurements further. While there have been demonstrations of very high efficiency single-qubit readout (Refs. [14, 29]) these systems have not yet been scaled to larger processors, and even these large efficiencies fall short of the thresholds for analog feedback. An intriguing part of this issue is the topic of exactly how amplification should be integrated closer to the processor.

On one hand, using the methods from chapter 2 we could consider applying broadband synthesis methods from chapter 4 to previously demonstrated [49, 63, 30, 29] directional amplifiers in order to open a promising route to integration closer to the processor. This is one of the most accessible routes to approach industrially, and one that I intend to pursue after graduation. However, it falls victim to the qubit's readout resonator being a rather unreliable impedance environment. After all, its frequency explicitly changes depending on qubit state, so how might we design a system that both detects this change and achieves nonreciprocity in the signal response no matter the qubit state? Another option is integrating interferometric schemes like Ref. [32], however these heavily depend on the external off-the-shelf components needed to create the interferometer. Lastly and most radically, several efforts are underway, both by students in the Hatlab and the NIST AMP group among others, to dissolve the boundary between the amplifier and the processor altogether. This creates the opportunity for time-domain pulsed readout schemes such as those demonstrated in Ref. [52] as well as generating nonreciprocal gain processes where the qubit's readout resonator is actively participating in the amplifier.

Ultimately, not only are the different methods of amplification in a sort of arms race against one-another for long-term viability, the entire field is competing with the designers of transmon qubits and other dispersively coupled readout schemes. While we improve our amplifiers, they seek to invent readout methods that tolerate ever higher numbers of photons in the readout resonator to begin with. If this number can be pushed two orders of magnitude over the 20-50 photon added noise of a HEMT plus typical losses, the typical readout fidelities (and readout times) we see today may be achievable without a parametric amplifier at all.

Luckily, we amplifier designers need not worry. After all, parametric amplifiers with a high power tolerance would still be helpful even for these systems in reducing their readout time for the same total fidelity. In either case, high-power parametric amplifiers operating at the quantum limit of added noise would remain faithfully good sensors, paving the way for exciting new opportunities in the study of quantum information.

Appendix : Code for polynomial long division

```
import numpy as np
from numpy.polynomial import Polynomial as P
def gs_from_poly(num, den, n, debug = False):
    gs = [1]
    for order in range(n):
        quo, rem = divmod(num, den)
        num = den
        den = rem
        gs.append(quo.coef[-1].real)
    gs.append(1/quo.coef[0].real)
    if debug: print('Calculated gs: ', gs)
    return gs
#paste in your particular polynomial zeros and poles here:
zeros_R = [-0.479494, -0.239747 - 0.878883*1j, -0.239747 + 0.878883*1j]
poles_D = [-1.89473, -0.947363 + 1.81452*1j, -0.947363 - 1.81452*1j]
num_zeros = len(zeros_R)
num_poles = len(poles_D)
# construct R and d from power series polynomials
R = P.fromroots(zeros_R)
D = P.fromroots(poles_D)
z_factor_num = D + R
z_factor_den = D - R
num_deg = z_factor_num.degree()
den_deg = z_factor_den.degree()
print(gs_from_poly(z_factor_num, z_factor_den, 3))
```

Bibliography

[1] IBM quantum roadmap to build quantum-centric supercomputers | IBM research blog, 2024. `https://research.ibm.com/blog/ibm-quantum-roadmap-2025`.

[2] Jose Aumentado. Superconducting Parametric Amplifiers: The State of the Art in Josephson Parametric Amplifiers. *IEEE Microwave Magazine*, 21(8):45–59, August 2020. Publisher: Institute of Electrical and Electronics Engineers Inc.

[3] Andor Bariska. Time Machine, Anyone? - Andor Bariska.

[4] N. Bergeal, R. Vijay, V. E. Manucharyan, I. Siddiqi, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret. Analog information processing at the quantum limit with a josephson ring modulator. *Nature Physics 2010 6:4*, 6:296–302, 2 2010.

[5] Alexandre Blais, Arne L. Grimsmo, S. M. Girvin, and Andreas Wallraff. Circuit quantum electrodynamics. *Reviews of Modern Physics*, 93(2):025005, May 2021. Publisher: American Physical Society.

[6] C. C. Bultink, B. Tarasinski, N. Haandbæk, S. Poletto, N. Haider, D. J. Michalak, A. Bruno, and L. DiCarlo. General method for extracting the quantum efficiency of dispersive qubit readout in circuit QED. *Applied Physics Letters*, 112(9):092601, March 2018.

[7] Carlton M. Caves, Joshua Combes, Zhang Jiang, and Shashank Pandey. Quantum limits on phase-preserving linear amplifiers. *Physical Review A - Atomic, Molecular, and Optical Physics*, 86:063802, 12 2012.

[8] John Clarke and Alex I. Braginski. *The SQUID Handbook*, volume 1. Wiley Blackwell, 5 2005.

[9] A. Clerk. Quantum noise and quantum measurement. In Michel Devoret, Benjamin Huard, Robert Schoelkopf, and Leticia F. Cugliandolo, editors, *Quantum Machines: Measurement and Control of Engineered Quantum Systems*, pages 61–112. Oxford University PressOxford, 1 edition, June 2014.

[10] A. A. Clerk, M. H. Devoret, S. M. Girvin, Florian Marquardt, and R. J. Schoelkopf. Introduction to quantum noise, measurement, and amplification. *Reviews of Modern Physics*, 82:1155–1208, 4 2010.

[11] Joachim Cohen, Alexandru Petrescu, Ross Shillito, and Alexandre Blais. Reminiscence of classical chaos in driven transmons. *PRX Quantum*, 4:020312, 4 2023.

[12] M.H. Devoret. *Quantum fluctuations in electrical circuits.* Edition de Physique, France, 1997. INIS Reference Number: 29063476.

[13] Marie Frédérique Dumas, Benjamin Groleau-Paré, Alexander McDonald, Manuel H. Muñoz-Arias, Cristóbal Lledó, Benjamin D'Anjou, and Alexandre Blais. Measurement-Induced Transmon Ionization. *Physical Review X*, 14(4):041023, October 2024. Publisher: American Physical Society.

[14] A. Eddins, J. M. Kreikebaum, D. M. Toyli, E. M. Levenson-Falk, A. Dove, W. P. Livingston, B. A. Levitan, L. C. G. Govia, A. A. Clerk, and I. Siddiqi. High-Efficiency Measurement of an Artificial Atom Embedded in a Parametric Amplifier. *Physical Review X*, 9(1):011004, January 2019. Publisher: American Physical Society.

[15] N. E. Frattini, V. V. Sivak, A. Lingenfelter, S. Shankar, and M. H. Devoret. Optimizing the nonlinearity and dissipation of a SNAIL parametric amplifier for dynamic range. *Physical Review Applied*, 10:054020, 11 2018.

[16] N. E. Frattini, U. Vool, S. Shankar, A. Narla, K. M. Sliwa, and M. H. Devoret. 3-wave mixing Josephson dipole element. *Applied Physics Letters*, 110:222603, 5 2017.

[17] Jay Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf. Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement. *Physical Review A - Atomic, Molecular, and Optical Physics*, 76:012325, 7 2007.

[18] M. Hatridge, S. Shankar, M. Mirrahimi, F. Schackert, K. Geerlings, T. Brecht, K. M. Sliwa, B. Abdo, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret. Quantum back-action of an individual variable-strength measurement. *Science*, 339:178–181, 1 2013.

[19] Nicholas Mario Hougland, Zhuan Li, Ryan Kaufman, Boris Mesits, Roger Mong, Michael Hatridge, and David Pekker. *In Preparation*, 2024.

[20] Doug Jorgesson. IQ/IR/SSB Mixer Primer, May 2020.

[21] Ryan Kaufman, Chenxu Liu, Katarina Cicak, Boris Mesits, Mingkang Xia, Chao Zhou, Maria Nowicki, José Aumentado, David Pekker, and Michael Hatridge. Simple, high saturation power, quantum-limited, rf squid array-based josephson parametric amplifiers. 2 2024.

[22] Ryan Kaufman and Ofer Naaman. Parametric Amplifier Matching Using Legendre Prototypes, March 2023. arXiv:2303.00184 [physics, physics:quant-ph].

[23] Ryan Kaufman, Theodore White, Mark I. Dykman, Andrea Iorio, George Stirling, Sabrina Hong, Alex Opremcak, Andreas Bengtsson, Lara Faoro, Joseph C. Bardin, Tim Burger, Robert Gasca, and Ofer Naaman. Josephson parametric amplifier with chebyshev gain profile and high saturation. *Physical Review Applied*, 20:054058, 5 2023.

[24] Saeed A. Khan, Ryan Kaufman, Boris Mesits, Michael Hatridge, and Hakan E. Türeci. Practical Trainable Temporal Postprocessor for Multistate Quantum Measurement. *PRX Quantum*, 5(2):020364, June 2024. Publisher: American Physical Society.

[25] Mostafa Khezri, Alex Opremcak, Zijun Chen, Kevin C. Miao, Matt McEwen, Andreas Bengtsson, Theodore White, Ofer Naaman, Daniel Sank, Alexander N. Korotkov, Yu Chen, and Vadim Smelyanskiy. Measurement-induced state transitions in a superconducting qubit: Within the rotating-wave approximation. *Physical Review Applied*, 20:054008, 11 2023.

[26] Bogdan A. Kochetov and Arkady Fedorov. Higher-order nonlinear effects in a Josephson parametric amplifier. *Physical Review B - Condensed Matter and Materials Physics*, 92:224304, 12 2015.

[27] S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinsoo, R. Keller, J. Lütolf, C. Eichler, and A. Wallraff. Engineering cryogenic setups for 100-qubit scale superconducting circuit systems. *EPJ Quantum Technology*, 6(1):1–29, December 2019. Number: 1 Publisher: SpringerOpen.

[28] Sebastian Krinner, Nathan Lacroix, Ants Remm, Agustin Di Paolo, Elie Genois, Catherine Leroux, Christoph Hellings, Stefania Lazar, Francois Swiadek, Johannes Herrmann, Graham J. Norris, Christian Kraglund Andersen, Markus Müller, Alexandre Blais, Christopher Eichler, and Andreas Wallraff. Realizing repeated quantum error correction in a distance-three surface code. *Nature*, 605(7911):669–674, May 2022. Publisher: Nature Publishing Group.

[29] F. Lecocq, L. Ranzani, G. A. Peterson, K. Cicak, X. Y. Jin, R. W. Simmonds, J. D. Teufel, and J. Aumentado. Efficient qubit measurement with a nonreciprocal microwave amplifier. *Physical Review Letters*, 126:020502, 1 2021.

[30] F. Lecocq, L. Ranzani, G. A. Peterson, K. Cicak, R. W. Simmonds, J. D. Teufel, and J. Aumentado. Nonreciprocal microwave signal processing with a field-programmable Josephson amplifier. *Physical Review Applied*, 7:024028, 2 2017.

[31] Chenxu Liu, Tzu Chiao Chien, Michael Hatridge, and David Pekker. Optimizing Josephson-ring-modulator-based Josephson parametric amplifiers via full Hamiltonian control. *Physical Review A*, 101:042323, 4 2020.

[32] G. Liu, X. Cao, T.-C. Chien, C. Zhou, P. Lu, and M. Hatridge. Noise Reduction in Qubit Readout with a Two-Mode Squeezed Interferometer. *Physical Review Applied*, 18(6):064092, December 2022. Publisher: American Physical Society.

[33] G. Liu, T. C. Chien, X. Cao, O. Lanes, E. Alpern, D. Pekker, and M. Hatridge. Josephson parametric converter saturation and higher order effects. *Applied Physics Letters*, 111:20, 11 2017.

[34] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi. A nearâ€"quantum-limited Josephson traveling-wave parametric amplifier. *Science*, 350:307–310, 10 2015.

[35] Easwar Magesan, Jay M. Gambetta, A. D. Córcoles, and Jerry M. Chow. Machine Learning for Discriminating Quantum Measurement Trajectories and Improving Readout. *Physical Review Letters*, 114(20):200501, May 2015. Publisher: American Physical Society.

[36] Justin L. Mallek, Donna-Ruth W. Yost, Danna Rosenberg, Jonilyn L. Yoder, Gregory Calusine, Matt Cook, Rabindra Das, Alexandra Day, Evan Golden, David K. Kim, Jeffery Knecht, Bethany M. Niedzielski, Mollie Schwartz, Arjan Sevi, Corey Stull, Wayne Woods, Andrew J. Kerman, and William D. Oliver. Fabrication of superconducting through-silicon vias, March 2021. arXiv:2103.08536 [cond-mat, physics:physics, physics:quant-ph].

[37] Nicholas A. Masluk, Ioan M. Pop, Archana Kamal, Zlatko K. Minev, and Michel H. Devoret. Microwave characterization of josephson junction arrays: Implementing a low loss superinductance. *Physical Review Letters*, 109:137002, 9 2012.

[38] D. T. McClure, Hanhee Paik, L. S. Bishop, M. Steffen, Jerry M. Chow, and Jay M. Gambetta. Rapid Driven Reset of a Qubit Readout Resonator. *Physical Review Applied*, 5(1):011001, January 2016.

[39] J. Y. Mutus, T. C. White, R. Barends, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Kelly, A. Megrant, C. Neill, P. J.J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, K. M. Sundqvist, A. N. Cleland, and John M. Martinis. Strong environmental coupling in a Josephson parametric amplifier. *Applied Physics Letters*, 104:263513, 6 2014.

[40] O. Naaman, D. G. Ferguson, A. Marakov, M. Khalil, W. F. Koehl, and R. J. Epstein. High saturation power Josephson parametric amplifier with ghz bandwidth. *IEEE MTT-S International Microwave Symposium Digest*, 2019-June:259–262, 6 2019.

[41] Ofer Naaman and José Aumentado. Synthesis of parametrically coupled networks. *PRX Quantum*, 3:020201, 6 2022.

[42] Ofer Naaman, Ted White, Mohamed Awida Hassan, Derek Slater, Sean Mcilvane, Edwin Yeung, and Philip Krantz. Modeling flux-quantizing Josephson junction circuits in Keysight ADS, August 2024. arXiv:2408.07861 [cond-mat, physics:physics, physics:quant-ph].

[43] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 12 2010.

[44] J. Niemeyer, L. Grimm, W. Meier, J. H. Hinken, and E. Vollmer. Stable josephson reference voltages between 0.1 and 1.3 v for high-precision voltage standards. *Applied Physics Letters*, 47:1222–1223, 12 1985.

[45] Simon E. Nigg, Hanhee Paik, Brian Vlastakis, Gerhard Kirchmair, S. Shankar, Luigi Frunzio, M. H. Devoret, R. J. Schoelkopf, and S. M. Girvin. Black-box superconducting circuit quantization. *Physical Review Letters*, 108(24):240502, June 2012. arXiv: 1204.0587 Publisher: American Physical Society.

[46] Andrew Osborne, Trevyn Larson, Sarah Garcia Jones, Ray W. Simmonds, András Gyenis, and Andrew Lucas. Symplectic Geometry and Circuit Quantization. *PRX Quantum*, 5(2):020309, April 2024. Publisher: American Physical Society.

[47] David M Pozar. *Microwave engineering; 3rd ed.* Wiley, Hoboken, NJ, 2005.

[48] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 08 2018.

[49] Leonardo Ranzani and José Aumentado. Graph-based analysis of nonreciprocity in coupled-mode systems. *New Journal of Physics*, 17(2):023024, February 2015. Publisher: IOP Publishing.

[50] Ants Remm, Sebastian Krinner, Nathan Lacroix, Christoph Hellings, François Swiadek, Graham J. Norris, Christopher Eichler, and Andreas Wallraff. Intermodulation distortion in a Josephson traveling-wave parametric amplifier. *Physical Review Applied*, 20:034027, 9 2023.

[51] Chad Rigetti, Jay M. Gambetta, Stefano Poletto, B. L.T. Plourde, Jerry M. Chow, A. D. Córcoles, John A. Smolin, Seth T. Merkel, J. R. Rozen, George A. Keefe, Mary B. Rothwell, Mark B. Ketchen, and M. Steffen. Superconducting qubit in a waveguide cavity with a coherence time approaching 0.1 ms. *Physical Review B - Condensed Matter and Materials Physics*, 86:100506, 9 2012.

[52] Eric I. Rosenthal, Christian M. F. Schneider, Maxime Malnou, Ziyi Zhao, Felix Leditzky, Benjamin J. Chapman, Waltraut Wustmann, Xizheng Ma, Daniel A. Palken, Maximilian F. Zanner, Leila R. Vale, Gene C. Hilton, Jiansong Gao, Graeme Smith, Gerhard Kirchmair, and K. W. Lehnert. Efficient and Low-Backaction Quantum Measurement Using a Chip-Scale Detector. *Physical Review Letters*, 126(9):090503, March 2021. Publisher: American Physical Society.

[53] Ananda Roy and Michel Devoret. Introduction to parametric amplification of quantum signals with Josephson circuits. *Comptes Rendus. Physique*, 17(7):740–755, 2016.

[54] Tanay Roy, Suman Kundu, Madhavi Chand, A. M. Vadiraj, A. Ranadive, N. Nehra, Meghan P. Patankar, J. Aumentado, A. A. Clerk, and R. Vijay. Broadband parametric amplification with impedance engineering: Beyond the gain-bandwidth product. *Applied Physics Letters*, 107:262601, 12 2015.

[55] Colm A. Ryan, Blake R. Johnson, Jay M. Gambetta, Jerry M. Chow, Marcus P. da Silva, Oliver E. Dial, and Thomas A. Ohki. Tomography via correlation of noisy measurement records. *Physical Review A*, 91(2):022118, February 2015. Publisher: American Physical Society.

[56] Daniel Sank, Zijun Chen, Mostafa Khezri, J. Kelly, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J.J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher,

T. White, J. Wenner, Alexander N. Korotkov, and John M. Martinis. Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation. *Physical Review Letters*, 117:190503, 11 2016.

[57] David Isaac Schuster. *Circuit Quantum Electrodynamics*. PhD thesis, Applied Physics, 2007.

[58] R. Shankar. *Principles of Quantum Mechanics*. Springer US, New York, NY, 1994.

[59] Ross Shillito, Alexandru Petrescu, Joachim Cohen, Jackson Beall, Markus Hauru, Martin Ganahl, Adam G.M. Lewis, Guifre Vidal, and Alexandre Blais. Dynamics of transmon ionization. *Physical Review Applied*, 18:034031, 9 2022.

[60] Matti Silveri, Evan Zalys-Geller, Michael Hatridge, Zaki Leghtas, Michel H. Devoret, and S. M. Girvin. Theory of remote entanglement via quantum-limited phase-preserving amplification. *Physical Review A*, 93(6):062310, June 2016. Publisher: American Physical Society.

[61] V. V. Sivak, N. E. Frattini, V. R. Joshi, A. Lingenfelter, S. Shankar, and M. H. Devoret. Kerr-free three-wave mixing in superconducting quantum circuits. *Physical Review Applied*, 11:054060, 5 2019.

[62] V V Sivak, S Shankar, G Liu, J Aumentado, and M H Devoret. Josephson array-mode parametric amplifier. *Physical Review Applied*, 10:24014, 2020.

[63] K. M. Sliwa, M. Hatridge, A. Narla, S. Shankar, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret. Reconfigurable Josephson Circulator/Directional Amplifier. *Physical Review X*, 5(4):041020, November 2015. Publisher: American Physical Society.

[64] Kyle M. Sundqvist and Per Delsing. Negative-resistance models for parametrically flux-pumped superconducting quantum interference devices. *EPJ Quantum Technology*, 1(1):1–21, December 2014. Number: 1 Publisher: SpringerOpen.

[65] Ted Thorbeck, Zhihao Xiao, Archana Kamal, and Luke C. G. Govia. Readout-Induced Suppression and Enhancement of Superconducting Qubit Lifetimes. *Physical Review Letters*, 132(9):090602, February 2024. Publisher: American Physical Society.

[66] S Touzard, A Kou, N E Frattini, V V Sivak, S Puri, A Grimm, L Frunzio, S Shankar, and M H Devoret. Gated conditional displacement readout of superconducting qubits. *Physical Review Letters*, 122, 2019.

[67] R. Vijay, C. Macklin, D. H. Slichter, S. J. Weber, K. W. Murch, R. Naik, A. N. Korotkov, and I. Siddiqi. Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature*, 490(7418):77–80, October 2012. Publisher: Nature Publishing Group.

[68] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff. Rapid high-fidelity single-shot dispersive readout of superconducting qubits. *Physical Review Applied*, 7:054020, 5 2017.

[69] Theodore White, Alex Opremcak, George Sterling, Alexander Korotkov, Daniel Sank, Rajeev Acharya, Markus Ansmann, Frank Arute, Kunal Arya, Joseph C. Bardin, Andreas Bengtsson, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Zijun Chen, Ben Chiaro, Josh Cogan, Roberto Collins, Alexander L. Crook, Ben Curtin, Sean Demura, Andrew Dunsworth, Catherine Erickson, Reza Fatemi, Leslie Flores Burgos, Ebrahim Forati, Brooks Foxen, William Giang, Marissa Giustina, Alejandro Grajales Dau, Michael C. Hamilton, Sean D. Harrington, Jeremy Hilton, Markus Hoffmann, Sabrina Hong, Trent Huang, Ashley Huff, Justin Iveland, Evan Jeffrey, Mária Kieferová, Seon Kim, Paul V. Klimov, Fedor Kostritsa, John Mark Kreikebaum, David Landhuis, Pavel Laptev, Lily Laws, Kenny Lee, Brian J. Lester, Alexander Lill, Wayne Liu, Aditya Locharla, Erik Lucero, Trevor McCourt, Matt McEwen, Xiao Mi, Kevin C. Miao, Shirin Montazeri, Alexis Morvan, Matthew Neeley, Charles Neill, Ani Nersisyan, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Rebecca Potter, Chris Quintana, Pedram Roushan, Kannan Sankaragomathi, Kevin J. Satzinger, Christopher Schuster, Michael J. Shearn, Aaron Shorter, Vladimir Shvarts, Jindra Skruzny, W. Clarke Smith, Marco Szalay, Alfredo Torres, Bryan W.K. Woo, Z. Jamie Yao, Ping Yeh, Juhwan Yoo, Grayson Young, Ningfeng Zhu, Nicholas Zobrist, Yu Chen, Anthony Megrant, Julian Kelly, and Ofer Naaman. Readout of a quantum processor with high dynamic range Josephson parametric amplifiers. *Applied Physics Letters*, 122:14001, 1 2023.

[70] Mingkang Xia, Chao Zhou, Chenxu Liu, Param Patel, Xi Cao, Pinlei Lu, Boris Mesits, Maria Mucci, David Gorski, David Pekker, and Michael Hatridge. Fast superconducting qubit control with sub-harmonic drives, June 2023. arXiv:2306.10162 [quant-ph].

[71] T. Yamamoto, K. Inomata, M. Watanabe, K. Matsuba, T. Miyazaki, W. D. Oliver, Y. Nakamura, and J. S. Tsai. Flux-driven Josephson parametric amplifier. *Applied Physics Letters*, 93:42510, 7 2008.

[72] Chao Zhou. *Superconducting Quantum Routers, Modules, Gates, and Measurements Based on Charge-pumped Parametric Interactions*. University of Pittsburgh ETD,

University of Pittsburgh, January 2024. Num Pages: 210 Publisher: University of Pittsburgh.

[73] X. Zhou, V. Schmitt, P. Bertet, D. Vion, W. Wustmann, V. Shumeiko, and D. Esteve. High-gain weakly nonlinear flux-modulated Josephson parametric amplifier using a squid array. *Physical Review B - Condensed Matter and Materials Physics*, 89:214517, 6 2014.