UNIVERSITY of CALIFORNIA
Santa Barbara

Superconducting flux qubits for high-connectivity quantum annealing without lossy dielectrics

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Physics

by

Christopher M. Quintana

Committee in charge:

Professor John Martinis, Chair

Professor David Weld

Professor Mark Srednicki

September 2017

The dissertation of Christopher M. Quintana is approved:

Professor David Weld

Professor Mark Srednicki

Professor John Martinis, Chair

September 2017

Acknowledgements

It goes without saying that this work would not have been possible without the many contributions from everyone in the Martinis group. I sincerely thank all of you who have helped me along the way (and will hopefully continue helping me in the future!). John, I still remember how excited I was the moment you told me that I could join your group. I could tell from when I first heard you speak that you really wanted to build something, which is exactly what is happening now at Google. Jason, I wouldn't have been prepared to join this group if it weren't for the many serious learning experiences I had in your lab. Dan, thanks for your never-ending enthusiasm about all things lab-related – the first time I witnessed this was meeting you at APS in 2013 and was one of the factors that led to me coming here. Both you and John have exemplified how to filter through jargon and get to the core concepts in order to understand things better. You also encouraged me to never be afraid to ask questions when something isn't clear, and demonstrated the importance of documentation, which is largely the purpose of this thesis.

Tony and Yu, you have both been great mentors and friends over the past years and for that I am very grateful. You are both always looking out for the well-being of all the grad students. Tony, your ability to solve problems and come up with new and great ideas never ceases to amaze me. The silicon "timeline of sadness" has drastically turned itself around thanks to your perseverance and leadership. And even though you are so serious about your work, you are still always able to joke around and it's always a pleasure having a conversation with you. Yu, I don't think anyone else would have been better suited to take the lead on the flux qubit project. I admire your dedication and your ability to 'do it all', from design to fab to coordinating with theorists to thinking about the overall vision for the project, even when you had "jet leg." Most of all, thanks for introducing me to great Chinese food. I hope that I've taught you some things as well – I can think of one example when I taught you what "lol" means, to which you replied "laughing out loud? dude, go back to work, no lol". You have made me lol quite a bit.

Dvir, your incredible analytical skills have brought and continue to bring a lot to the fluxmon project. We are extremely lucky to have you as a core member of the fluxmon team. Mostafa, it was great working with you as well. Your attention to detail is truly impressive. Hartmut, I am glad that you brought our group under Google's wing with John. And thank you for being an optimist about quantum annealing – this is part of what motivates us to come into work every day. Andre, it's always great discussing the possible microscopics of flux noise with you and I always learn a lot from our meetings. I'm glad you'll be continuing to work with us this year. Thanks also to Sergio, Vadim, and others on the theory team who have taught us about quantum annealing.

Andrew and Brooks, you are probably the two nicest guys I have ever met. Andrew, you paved the way for doing "simulations in real life" (or as the rest of us call them, "experiments") to make silicon qubits not suck. We also really benefited from you and Tony turning Yu's original SiO_2 crossovers into robust dielectric-less crossovers. Thank you also for periodically quoting SpongeBob without shame. Brooks, I'm still not convinced you're not a wizard after seeing you re-thread the wirebonder in less than one second.

And thanks for sharing your espresso machine with the group before we got one of our own. Brooks and Josh, thanks for paving the way to superconducting bump bonds!

Jimmy and Amit, thank you for all the times you guys helped me and Yu with code, electronics, and various other things. Your continued patience and willingness to share your valuable skills with others is widely appreciated. Jimmy, just for you I will not write "here, we show" anywhere in this thesis. Jimmy and Matthew, I apologize in advance for any "wonton destruction" when we start to move the flux qubit code into master. Evan and Amit, it is sometimes scary how much knowledge you guys have about just about anything, and the group has on many occasions benefited from your ability to figure out why things are broken. Ted, you are also a pretty good guy, even though some of your puns are unforgivable. We can't wait to incorporate your latest work onto our carrier wafers. Also, if you're reading this, please bring in leftover smoked meats into work more often! It's been great having you around from tuning paramps to smoking cigars in San Antonio. Jiiiiiiiiri, it's always a blast hanging out with you at March Meeting too.

I have had the privilege of sharing an office with Jim, and have since developed a more (or less?) refined taste for punnery. Jim, you also clearly set the bar for being a great lab member and for scientific prolificness. Julian, the efficiency and speed with which you continue to progress with Xmons blows my mind. Keep doing what you're doing. Ben, your work and Peter's with the jump table and ADC for flux noise measurements was crucial for our noise studies. Between that and all the strange hours I've seen you in lab, your simultaneous dedication to your work and your family is impressive. Charles, I could tell from the first time I met you that you were a 'badass' both inside and outside the lab. You never fail to impress with gmon qubits and dancing skills. I hope that you and the rest of the graduate students decide to stay with us at Google. Rami, the real reason I came here was your depiction of a nice walk in Santa Barbara in your 2013 APS talk. Your flavor of humor always makes conversations more interesting. Sorry the flux noise reduction fab didn't work out, but there will be more experiments to come! Austin, the entire Xmon project rests on having a clear quantitative goal for error correction from you. I really admire your no-nonsense attitude. Thanks for thinking about connectivity graphs for the fluxmon. Pedram, whenever I needed something for the DRs or wanted a whiff of cologne I knew all I had to do was come to your office. Thanks to everyone else at Google SBA who also contributed to the experimental infrastructure.

Ethan and Austin W, sorry I haven't kept in touch as much as I would have liked to. I'm really glad we were able to go on two memorable trips together in the last few years (the sun truly is a deadly lazer), and there will be more to come. You guys are genuinely amazing people and have influenced me greatly. Udit, it still means a lot to me that you've kept in touch ever since high school. It is always a memorable time when I visit you in New York. Keep being awesome and making a difference in the world. Last but not least, thank you to my family for always believing in me. I couldn't have asked for more supportive parents. I know that you gave a lot so that I could get to this point, and I literally owe you my life. Alyssa, I am so proud of you as well. Almost certainly, the next time you visit Santa Barbara it won't be 90 degrees outside.

Curriculum Vitæ

Christopher M. Quintana

Education

(2017) Ph.D., Physics, University of California, Santa Barbara

(2013) A.B., Physics, *summa cum laude*, Princeton University
Senior thesis: "Josephson Parametric Amplification for Circuit QED"

(2009) Nyack High School, Nyack, NY, *valedictorian*

Honors and Awards

NSF Graduate Research Fellowship, NSF GRFP (2014 - 2017)

Yzurdiaga Fellowship, UCSB Physics Department (2013-2014)

Elected to Membership in the Phi Beta Kappa Society (2013)

Best Engineering Physics Independent Work Award, Princeton University (2013)

Kusaka Memorial Prize in Physics, Princeton Physics Department (2012 and 2013)

Shapiro Prize for Academic Excellence, Princeton University (2011)

American Physical Society Minority Scholar, APS (2009-2011)

Publications

"Cavity-mediated entanglement generation via Landau-Zener interferometry", **C. M. Quintana**, K. D. Petersson, L. W. McFaul, S. J. Srinivasan, A. A. Houck, J. R. Petta *Phys. Rev. Lett.* **110**, 173603 (2013)

"Characterization and reduction of microfabrication-induced decoherence in superconducting quantum circuits", **C. M. Quintana**, et al. *Appl. Phys. Lett.* **105**, 062601 (2014)

"Observation of topological transitions in interacting quantum circuits", P. Roushan, C. Neill, Y. Chen, M. Kolodrubetz, **C. Quintana**, N. Leung, M. Fang, et al. *Nature* **515**, 241-244 (2014)

"Fast charge sensing of a cavity-coupled double quantum dot using a Josephson parametric amplifier", J. Stehlik, Y.-Y. Liu, **C. M. Quintana**, C. Eichler, T. R. Hartke, J. R. Petta *Phys. Rev. Appl.* **4**, 014018 (2015)

"Measuring and suppressing quantum state leakage in a superconducting qubit", Zijun Chen, Julian Kelly, **Chris Quintana**, et al. *Phys. Rev. Lett.* **116**, 020501 (2015)

"Observation of Classical-Quantum Crossover of $1/f$ Flux Noise and Its Paramagnetic Temperature Dependence", **C. M. Quintana**, Yu Chen, D. Sank, A. Petukhov et al. *Phys. Rev. Lett.* **118**, 057702 (2017)

"Tunable inductive coupling of superconducting qubits in the strongly nonlinear regime", Dvir Kafri, **Chris Quintana**, Yu Chen, Alireza Shabani, John M. Martinis, Hartmut Neven *Phys. Rev. A* **95**, 052333 (2017)

"Characterization and reduction of capacitive loss induced by sub-micron Josephson junction fabrication in superconducting qubits", A. Dunsworth, A. Megrant, **C. Quintana**, Z. Chen, et al. *Appl. Phys. Lett.* **111**, 022601 (2017)

"Fabrication and characterization of silicon oxide scaffolded aluminum airbridges for superconducting microwave devices", A. Dunsworth, **C. Quintana**, Z. Chen, et al. *in preparation* (2017)

"Qubit compatible superconducting interconnects", B. Foxen, J. Y. Mutus, E. Lucero, R. Graff, A. Megrant, Yu Chen, **C. Quintana**, B. Burkett, J. Kelly, E. Jeffrey, Y. Yang, A. Yu, et al. *arXiv:1708.04270* (2017)

Abstract

Superconducting flux qubits for high-connectivity quantum annealing without lossy dielectrics

by

Christopher M. Quintana

Quantum annealing can potentially be used to find better solutions to hard optimization problems faster than purely classical hardware. To take full advantage of quantum effects such as tunneling, a physical annealer should be comprised of qubits with a sufficient degree of quantum coherence. In addition, to encode useful problems, an annealer should provide a dense physical connectivity graph between qubits. Towards these goals, we develop superconducting "fluxmon" flux qubits suitable for high-connectivity quantum annealing without the use of performance-degrading lossy dielectrics. We carry out in-depth studies of noise and dissipation, and of qubit-qubit coupling in the strongly nonlinear regime. We perform the first frequency-resolved measurements extracting both the quantum and classical parts of the $1/f$ flux noise intrinsic to superconducting devices, and observe the classical-to-quantum crossover of the noise. We also identify atomic hydrogen as a magnetic dissipation source. We then implement tunable inter-qubit coupling compatible with high connectivity, and provide direct spectroscopic measurements of ultra-strong coupling between qubits. Finally, we use our system to explore quantum annealing faster than the system thermalization time, a previously unaccessed regime.

Contents

Chapter 1

Introduction

In this chapter, we give context to our work by describing quantum annealing and how it might be useful for computational optimization. We provide a basic overview of the current state of the complex, continually evolving (and sometimes controversial) field of quantum annealing with superconducting quantum bits (qubits). We also consider how one might improve the performance of existing quantum annealers.

1.1 Quantum-enhanced optimization

Many important computational tasks in the modern world boil down to optimization problems. These problems include resource allocation, vehicle routing, job scheduling, engineering design, and machine learning, among many other economically and scientifically significant tasks. A large class of these problems takes the form of discrete optimization, for which one needs to examine a finite but combinatorially large set of

potential solutions. One famous and easy to understand example is the traveling salesman problem: given a collection of N interconnected cities, what is the shortest possible tour that visits each city exactly once? For this and many other industrially relevant problems, the set of possible solutions is so enormous that exhaustive search is practically impossible,[1] and furthermore no efficient solution algorithm to find the true global minimum is known. In computer science, 'efficient' usually means requiring resources that are polynomial in the problem size, and in practice means that it should be a very low-power polynomial.

Fortunately, a solution to an optimization problem doesn't have to be the absolute best possible solution in order to be practically useful. Instead, approximate algorithms and heuristic techniques are often used to find a 'good enough' solution with a reasonable amount of computational time and resources. In certain cases, one can take advantage of structure in the problem to make it easier to find solutions. Sometimes, it is possible to come up with heuristic algorithms specific to the problem at hand to obtain approximate solutions in an acceptable amount of time. In such cases, the quality of solution can be quantified by an approximation ratio, which is the ratio of the actual cost function of the found state to the lowest possible cost function. For example, there exists an $O(N^3)$ algorithm to obtain an approximate answer to the travelling salesman problem within a factor of 1.5 of the optimal solution length [1], and there are other heuristics that can typically reach solutions that are to within $\sim 1\%$ of optimal for a million cities.

[1] Amazingly, a quick factorial calculation shows that, assuming it takes a computer one nanosecond to evaluate and compare the distance of a given tour, and assuming no parallelization, for just 27 cities it would take as long as the apparent age of the universe to try all possible tours.

When little is known about the structure of the problem at hand, or it is not possible to take advantage of any known structure, generic metaheuristic algorithms are often used. A universal issue that such techniques need to address is the propensity to get stuck in local optima that are far from the global optimum, as is common in algorithms using gradient descent or analogous techniques. A commonly used metaheuristic that addresses this issue and enables good sampling of a large search space is known as simulated annealing (SA), first proposed by Kirkpatrick *et al.* [2]. This probabilistic technique is inspired by the physical process of annealing in metallurgy and glassblowing, whereby a system is heated up and then allowed to slowly cool down. This process helps the system settle into an overall lower energy state, reducing the number of defects and improving material properties. SA mimics this process by introducing an artificial temperature that induces transitions analogous to thermal transitions, allowing the system to escape from local minima and converge to a more global optimum. More concretely, SA works as follows. An arbitrary initial state is chosen and a temperature is assigned to the system. Then, a new state of the system is proposed (usually, this means flipping one or more of the bits representing the possible solutions), and the energy difference ΔE (i.e., difference in cost function) between the proposed and current state is computed. If $\Delta E < 0$, the new state is accepted. If $\Delta E > 0$, the state is accepted with finite probability according to the Boltzmann factor $\exp[-\Delta E/kT]$. This update rule is an example of what is known as a Metropolis algorithm [3], a general Monte Carlo method for sampling probability distributions that are hard to sample directly. As this stochastic cycle is

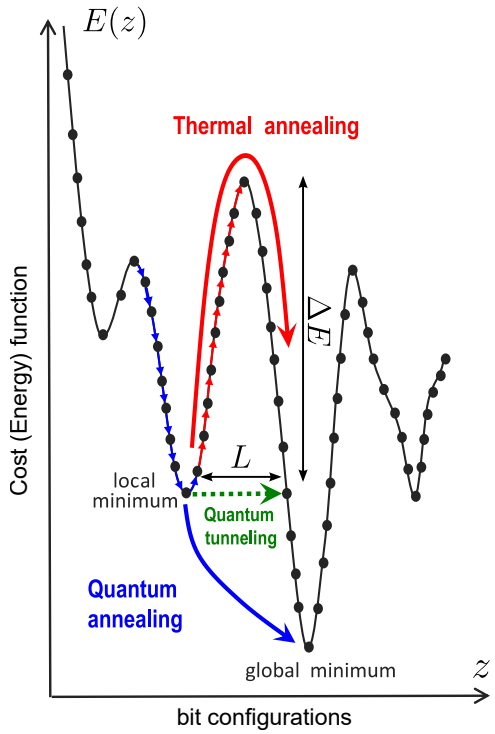


Figure 1.1: Illustration of escape from a local minimum via thermal vs. quantum annealing. Figure adapted from Ref. [4].

repeated, the temperature is gradually lowered to zero, and the whole process is repeated if needed, until one converges on a low energy solution. The lowering of the temperature should be done slowly enough to avoid "quenching" the system into a high energy local minimum, analogous to how quenching a real glass can lead to metastable configurations and internal stresses that degrade its reliability. Simulated annealing naturally provides a useful tradeoff between exploration of the entire space and concentration on potentially promising regions.

Quantum annealing (QA) as a metaheuristic was inspired by SA, and is in some sense an extension of SA. In QA, one replaces (or supplements) thermal fluctuations with "quantum fluctuations" and attempts to use quantum tunneling as an additional pathway

to find good solutions [5, 6, 7]. In a quantum system, moving across an energy landscape is not restricted to hopping over energy barriers, but classically forbidden pathways that "tunnel through" barriers are possible as well. If one were to encode the cost function in the Hamiltonian of a quantum system, one might be able to more efficiently find solutions by introducing a transverse field (i.e., a term that does not commute with the classical potential energy part of the problem Hamiltonian) that drives quantum tunneling transitions. Mathematically, the rate of quantum tunneling through an energy barrier can be compared to the rate of thermal hopping at temperature T as illustrated in Fig. 1.1. In general, in contrast to the thermal hopping rate

$$\Gamma_T \propto \exp\left[-\Delta E/kT\right], \tag{1.1}$$

which only depends on the height ΔE of the barrier (at a given temperature), the quantum tunneling rate through an energy barrier also depends on the width L of the barrier,

$$\Gamma_Q \propto \exp[-L\sqrt{\Delta E}/\Gamma], \tag{1.2}$$

where Γ is the strength of the tunneling field.[2] Therefore, one expects quantum tunneling to be more effective in cases where there are tall, narrow energy barriers. On the other hand, quantum tunneling will be ineffective compared to thermal hopping when there are very broad barriers.

Quantum annealing is usually implemented with tunable two-level quantum systems (qubits) with an Ising Hamiltonian having local fields for each 'spin' (qubit) and pairwise

[2]To understand this formula, think of the action under the barrier in the WKB approximation, with Γ acting as an inverse measure of the mass in the kinetic energy term.

two-body interactions between the spins. It turns out that such Hamiltonians naturally represent quadratic unconstrained binary optimization problems (QUBOs), a very powerful class of problems that can encode a large, generic class of discrete optimization problems.[3] QUBOs are problems with a cost function that is quadratic in the bits $x_i \in \{0, 1\}$,

$$E(x_1, ..., x_n) = \sum_{i<j=1}^{n} J_{ij} x_i x_j + \sum_{i=1}^{n} h_i x_i, \tag{1.3}$$

where the goal is to find the optimal configuration of bits,

$$\vec{x}^* = \text{argmin}\{E(x_1, ..., x_n)\} \tag{1.4}$$

for the problem defined by a chosen set of J_{ij}'s and h_i's. We can see the direct correspondence with an Ising Hamiltonian for quantum spins,

$$H_P = H_{\text{Ising}} = \sum_{i,j} J_{ij} \sigma_i^z \sigma_j^z + \sum_i h_i \sigma_i^z, \tag{1.5}$$

where σ^z is the standard 2×2 Pauli Z matrix. Namely, finding the ground state of the Ising Hamiltonian (1.5), which is diagonal in the z basis, would solve the QUBO problem (1.3). Similarly, finding low-lying states of the Ising Hamiltonian will yield approximate solutions to the corresponding QUBO problem. It is known that solving the Ising spin glass problem H_{Ising} on arbitrary connectivity graphs is in general NP-hard, meaning it is at least as hard as exponential problems like the travelling salesman problem. However, the hope is not that quantum annealing will necessarily turn an exponential scaling into a polynomial scaling in N, but more realistically that it might change the coefficients

[3] General discrete variables can be encoded with integers and therefore bits, and higher-order polynomials can be encoded with quadratic terms plus ancilla bits with constraints, leading to an extra polynomial overhead.

within the exponent, leading to polynomial but still practically significant speedups.

The "annealing" part of QA is typically implemented by initializing the system of spins in the ground state of a "driver" Hamiltonian, H_D, which is easy to prepare. Usually one chooses

$$H_D = \sum_{i=1}^{N} \sigma_i^x, \tag{1.6}$$

for which the ground state is a uniform superposition over all the possible Ising states (σ_z eigenstates) and is easy to prepare in practice. This simply corresponds to each spin aligning with its own transverse field, i.e., a field in the X direction. The driver Hamiltonian is then dialed down while the problem Hamiltonian is dialed up, i.e., the time-dependent Hamiltonian over the course of the anneal is chosen according to the annealing schedules $A(t)$ and $B(t)$

$$H(t) = A(t)H_D + B(t)H_P, \quad t \in [0, t_f] \tag{1.7}$$

where $A(t)$ decreases monotonically and $B(t)$ increases monotonically, such that $A(0)/B(0) \gg 1$ and $A(t_f)/B(t_f) \ll 1$. Decreasing the quantum fluctuations $A(t)$ over time is analogous to how thermal fluctuations are decreased over time in classical SA by lowering the temperature over time. This procedure is similar to the well-known adiabatic quantum computing algorithm introduced by Farhi *et al.* [8]. Adiabatic quantum computing is a simpler, noiseless closed-system version of quantum annealing that can be understood through the adiabatic theorem: if we prepare the system in the known ground state of an "easy" Hamiltonian and slowly change the Hamiltonian to take the form of the problem Hamiltonian H_P, the adiabatic theorem tells us that the system should remain in its

ground state throughout the computation as long as the adiabaticity condition [9, 4]

$$|\langle\psi_0(t)|\dot{H}(t)|\psi_1(t)\rangle| \ll \frac{|E_1(t) - E_0(t)|^2}{\hbar} \tag{1.8}$$

is satisfied throughout the process, where $|\psi_0(t)\rangle$ and $|\psi_1(t)\rangle$ are the ground and first excited states of the quantum system with eigenenergies $E_0(t)$ and $E_1(t)$. Essentially, for a uniform sweep rate (which is the best one can do a priori without further knowledge of the structure of the Hamiltonian) this means that the total evolution time must be longer than the inverse square of the minimum energy gap in order to avoid Landau-Zener transitions out of the ground state,

$$T_{\text{adiabatic}} \propto \frac{1}{(\text{minimum gap})^2}. \tag{1.9}$$

Unfortunately, for generic problems (especially for "hard" problems of interest), the minimum gap will become exponentially small as the number of bits grows [10].[4] Therefore, to use quantum annealing to find approximate solutions in polynomial time, the system will in general need to leave its ground state at some point. Furthermore, as we will see, a real quantum annealer will operate at a finite temperature with a substantial amount of environmental noise, leading to further transitions out of the ground state and requiring an open systems approach to describe it. Therefore, the adiabatic picture of quantum annealing is not particularly useful for real-world quantum annealers.

In general, large optimization problems will have both tall, narrow energy barriers and wide energy barriers. Statistically, if one starts in a random state, it will have an energy

[4]Intuitively, one can convince oneself of this fact by noting that the maximum possible energy above the ground state energy grows linearly or quadratically with the number of bits (depending on the range of spin-spin couplings), whereas the total number of possible energy levels in between them increases exponentially.

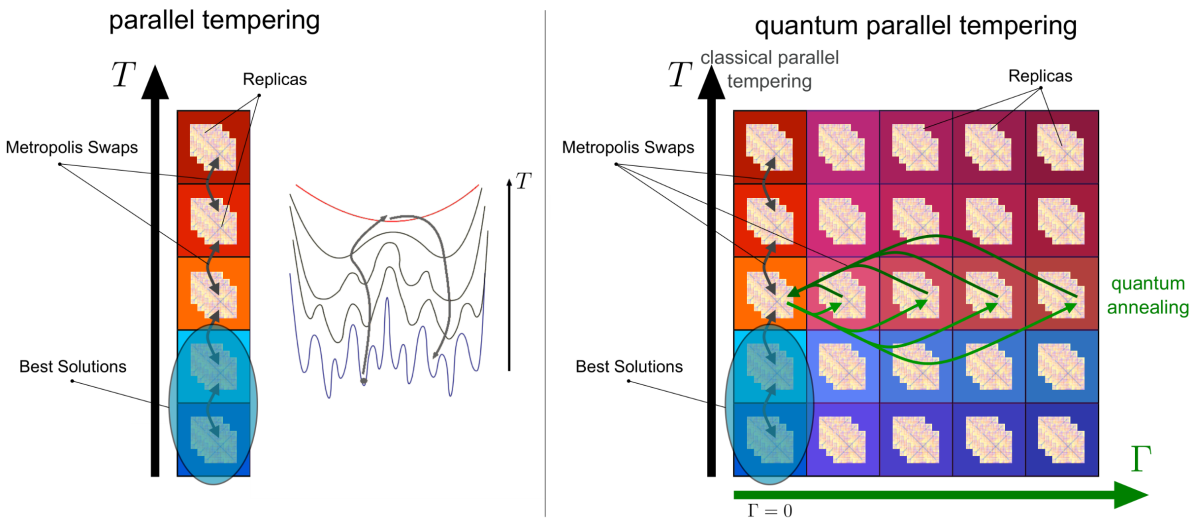


Figure 1.2: Schematic illustration of quantum parallel tempering, a potential "hybrid" way to take advantage of both thermal and quantum transitions. Figure adapted from Ref. [11].

significantly above the ground state energy, and finding a lower energy state would be easy for quantum tunneling because such a state would on average be separated by only a few spin flips, corresponding to a narrow barrier. However, once the energy of a state is low enough, the next lowest state will in general be many spin flips away, corresponding to a very wide barrier for which thermal annealing would be more effective. In other words, one might expect quantum tunneling to be useful for quickly finding low-lying states, but not necessarily the true ground state. For these reasons, it would be desirable to have a metaheuristic that combines he 'best of both worlds' of thermal hopping transitions (1.1) and quantum tunneling transitions (1.2). A particularly promising proposal of such a 'hybrid' method is what has recently been dubbed "quantum parallel tempering" by Neven *et al.* [11], with related ideas explored in the context of data clustering in Ref. [12]. This proposal is motivated by the technique of parallel tempering, which is an improved,

more state-of-the-art version of SA.[5] In classical parallel tempering, one has many copies of the problem at hand, each undergoing its own Metropolis update dynamics at its own fixed temperature after being randomly initialized. Periodically, configurations are exchanged between neighboring replicas according to a similar Metropolis rule. This way, if a cold replica is stuck in a bad state, it can be heated up while another replica is cooled, and conversely good states will diffuse towards the colder replicas. The idea of quantum parallel tempering is to supplement this approach with a second replica dimension besides temperature, namely quantum tunneling strength. Now, configurations from cold replicas can be fed into the quantum device, which is initialized into the state output from the classical algorithm. Then, the transverse field is dialed up from zero and then back down after a variable amount of time, allowing groups of spins to tunnel through tall, narrow barriers behind which they may have previously been stuck. This process, illustrated in Fig. 1.2, may be a very efficient way to reach new basins of attraction that would not have otherwise been explored by the classical part of the algorithm. If quantum-enhanced optimization heuristics such as this one are able to find improved solutions to hard problems by even just a few percent, from an economic perspective the consequences could be very significant [14], especially when the effects are compounded over time.

[5] A variant of parallel tempering [13] was the winning algorithm of the unweighted random MAX-SAT portion of last year's international SAT competition. SAT (Boolean satisfiability problem) is a (if not *the*) canonical NP-hard problem in computer science. In particular, all problems in NP can be mapped to it in polynomial time, a result known as the Cook-Levin theorem.

1.2 Quantum annealing with superconducting qubits

Superconducting qubits [15] are suitable candidates for implementing quantum annealing. These solid-state electrical circuits, based on superconductors and Josephson tunnel junctions, are particularly promising because they can be fabricated using the scalable techniques of integrated circuits, have parameters that can be engineered, and can be readily tuned and controlled *in situ* during experiments. D-Wave Systems Inc., a Canadian private company, has in fact been building and selling programmable quantum annealers over the past six years. The D-Wave processors are comprised of niobium-based superconducting flux qubits operating at millikelvin temperatures [16, 17, 18], with the current-generation D-Wave 2000Q$^{\text{TM}}$ processor[6] having 2000 qubits with a limited "Chimera" connectivity graph, illustrated in Fig. 1.3. These large-scale analog quantum machines are very impressive engineering feats that have required sophisticated systems engineering to construct.

As will be described in more detail in Chapter 2, a flux qubit is a type of superconducting qubit based on loops of superconducting current interrupted by Josephson tunnel junctions, and (in certain limits of operation) the two states $|\uparrow\rangle$ and $|\downarrow\rangle$ of the artificial Ising spin correspond to current circulating either clockwise or counterclockwise through these loops.[7] In a flux qubit, the superconducting phase difference across the qubit's Josephson junction(s) (which is related to the flux threading the qubit loop) acts

[6] https://www.dwavesys.com/d-wave-two-system

[7] In the next chapter we will describe in detail exactly how one can implement qubits suitable for quantum annealing using the building blocks of capacitors, inductors, and Josephson junctions (nonlinear kinetic inductors), by using the techniques of circuit quantization and the quantum nature of the macroscopic superconducting phase.

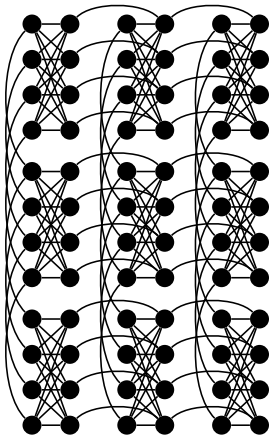


Figure 1.3: Graphical representation of a 3×3 patch of unit cells within the D-Wave Chimera unit tile topology. Dots represent qubits and lines represent inter-qubit couplings. Each unit cell contains eight qubits forming the complete bipartite graph $K_{4,4}$, with sparse connections to neighboring unit cells such that each qubit is connected to at most 6 other qubits.

as a quantum mechanical degree of freedom residing in a double-well potential energy landscape, with $|\uparrow\rangle$ and $|\downarrow\rangle$ corresponding to the phase being localized in the left or right wells.[8] In this two-level flux basis, the flux qubit Hamiltonian takes the form

$$H = -\frac{1}{2}(\Delta\sigma^x + \varepsilon\sigma^z). \tag{1.10}$$

Qualitatively, the height of the barrier in the double-well potential tunes the annealing driver Hamiltonian H_D through the tunneling term Δ, and is controlled by the flux through a secondary DC SQUID loop of the qubit. As we will see later, this energy barrier physically arises from the Josephson tunneling energy. The local field coefficients h_i of σ_i^z in the problem Hamiltonian (1.5) are controlled by flux threading the main qubit loop, inducing an energy offset ε between the left and right wells of the double-well potential, arising from the magnetic field energy stored in the qubit's inductive

[8]One typically assumes that the intra-well energy level spacing is large so that levels beyond the lowest two are negligible.

loop. Sign- and magnitude-tunable $\sigma^z\sigma^z$ couplings J_{ij} between these artificial spins are provided by flux-tunable superconducting rf-SQUID transformers which act as effective mutual inductances, coupling flux from the loop of one flux qubit to the loop of another.

It is known that superconducting qubits, and in particular D-Wave's qubits, undergo substantial unwanted coupling with their environments, with single-qubit relaxation and dephasing times orders of magnitude shorter than the annealing times in D-Wave processors. Environmental coupling makes real quantum annealers challenging to model and has in the past led to some debate over their degree of "quantumness". Coupling to the environment has at least two significant effects on quantum tunneling within an otherwise closed quantum system. Dephasing between energy levels due to low-frequency noise causes tunneling to no longer be described by coherent swapping between the initial and final states, but rather by an incoherent rate equation with a quadratically reduced tunneling rate. Namely, the tunneling occurs on a timescale of order Δ^2/W rather than Δ, where Δ is the associated tunnel splitting and W is a measure of the rate of dephasing between energy levels. Dissipation also suppresses quantum tunneling rates, as it tends to localize the system's wavefunction in local minima. Furthermore, if the dissipation is strong enough, the environment induces an effective energy barrier for tunneling, exponentially suppressing the tunneling rate. The severity of these effects becomes greater for multiqubit tunneling transitions, growing as the number of qubits that need to simultaneously tunnel [19]. D-Wave's qubits have strong low-frequency flux noise and high-frequency dielectric dissipation, but despite this, there is clear evidence

that the device does harness quantum-mechanical effects, particularly incoherent quantum tunneling, which occurs in both single- and multi-qubit devices [20, 17, 19]. For example, reasonable phenomenological models of low-frequency environmental noise can be used to model the measured tunneling rates, which for quantum tunneling is generally observed to become independent of temperature below a certain temperature [21, 22].

Several empirical results have recently been uncovered concerning how quantum tunneling might be relevant as a computational resource in a real quantum annealer with noise. Using the D-Wave machine, Boixo *et al.* [19] clearly show that multiqubit quantum tunneling can play a computational role in a crafted, proof-of-concept class of "weak-strong cluster" problems with tall, narrow energy barriers through which multiple qubits must simultaneously tunnel. This multiqubit tunneling dominates over thermal hopping, leading to better solutions at lower physical device temperatures. Denchev *et al.* [23] extended the weak-strong cluster study, showing that finite-range quantum tunneling can be used to obtain a computational speedup over simulated annealing in larger sets of problems that contain networks of weak-strong clusters coupled to each other in a glassy (random) manner. In particular, for a fixed annealing time of 20 μs (the fastest time that was available on the D-Wave device), a scaling speedup[9] relative to SA was observed, with the D-Wave device finding the ground state with 99 % probability $\sim 10^8$

[9]It could be argued that this is not a genuine limited quantum speedup, defined as "a speedup obtained when compared specifically with classical algorithms that correspond to the quantum algorithm in the sense that they implement the same algorithmic approach, but on classical hardware" [24]. This is because the annealing time was not optimized for each problem size, since the D-Wave hardware had a minimum allowed annealing time. Very recently, a class of problems was found for which the D-Wave annealing time could be optimized for each problem size, and a genuine limited quantum speedup was observed (although only a very modest one) [25]. However, there were still other classical algorithms that do much better than QA for this class of problems.

times faster for problems with ~ 1000 qubits compared to when SA is run on a single core. However, path-integral quantum Monte Carlo (QMC) techniques [26] were able to solve the problem with the same scaling in problem size as D-Wave, yet D-Wave was still able to gain a prefactor of $\sim 10^8$ in speed. Even though this is not a scaling speedup, it is still a very substantial effect – it took a significant fraction of Google's server fleet to run the classical algorithms in this study, costing around $1M to generate Fig. 4 in Ref. [23] [27]. However, it must be stressed that this result by no means demonstrates an unqualified quantum speedup – there exist specialized classical cluster-finding algorithms, such as Selby's algorithm [28], which are able to solve most problems on D-Wave's Chimera graph faster than any of the above methods, including D-Wave, by taking advantage of tree-structured subgraphs. The hope however is that such cluster-finding algorithms will become ineffective in future quantum annealers with a denser, more highly connected graph, making techniques like QMC the real long-term classical competitor. There is also hope that a quantum annealer with less unwanted environmental coupling may take advantage of not just incoherent but also coherent tunneling, and may be able to yield a scaling speedup relative to QMC. Furthermore, there exist topological obstructions to QMC such that the scaling is no longer the same as incoherent tunneling when there are multiple topologically inequivalent paths for tunneling [29] (although, this is only true for closed boundary condition QMC, not open-boundary condition QMC [26]. Also, it remains to be seen how generic such obstructions are).

A crucial open question is how one might build improved quantum annealers that

have a chance of providing a genuine computational speedup. Given the above, there are several ways one might be able to do so. To improve device performance and take full advantage of quantum effects like coherent tunneling, superposition, and entanglement, a key task is to reduce the amount of noise and dissipation in flux qubits. D-Wave's qubits rely on conventional multilayer fabrication with amorphous dielectrics, which along with their limited control scheme enabled them to quickly scale up to $\sim$ 1000 qubits; however, it is well-known from studies with superconducting resonators and gate-based superconducting qubits that these amorphous dielectrics add a significant amount of background dissipation to quantum circuits, in addition to introducing individual strongly coupled two-level charge defects, especially in the GHz frequency regime. It therefore quite possible that building a quantum annealer without lossy dielectrics could lead to significantly different qualitative behavior. In fact, the D-Wave device is typically modeled by assuming that it is fully thermalized with its environment up to some 'freeze-out' point in the middle of the anneal where the coupling matrix elements become small [30, 19], but this thermalization stage would become irrelevant with longer thermalization times and/or faster annealing. This brings us to a second aspect in which existing annealers could be improved. It is often observed that the optimal annealing time on the D-Wave device is the fastest annealing time available (around $10\,\mu$s due to engineering constraints). The ability to anneal significantly faster would therefore allow improved performance, especially if one were able to anneal faster than the system thermalization time.

Besides improving noise and dissipation, it is necessary to increase the density of the connectivity graph beyond that of existing quantum annealers, so that classical algorithms like Selby's can no longer make "cheap cuts" and will therefore become ineffective. Of course, a fully connected graph would be theoretically ideal, but given the limited finite size of flux qubits and inter-qubit couplers, one must make compromises and come up with a good connectivity graph given a limited number of qubits that can connect to any given qubit. It is not known what the optimal graph is given the physical constraints of the qubits and couplers, but some figures of merit are likely small diameter (i.e., it takes as few hops as possible to get from one side of the graph to the other), large treewidth (meaning it is hard to transform the graph into a tree structure that classical algorithms like Selby's could take advantage of), and large conductance (meaning the graph is 'well-knit' so it is not easy to cut the graph without incurring penalties, analogous to there being many bridges between two cities. High conductance graphs also give more freedom in how to embed problems that don't natively fit on the graph, and are more robust to faulty device elements such as broken couplers). Finally, it may also be desirable to have sparse long-range connections in the graph, to mimic the lateral connectivity structure of the neocortex, which may be the only quasi-2D arrangement of elements that has a non-zero spin-glass transition, meaning that hard problems can natively live in it [27]. However, at present it is unclear how to make such long-range connections between flux qubits while maintaining appreciable coupling strength. Finally, improving control errors in specifying h_i and J_{ij} over the course of the anneal is also required to

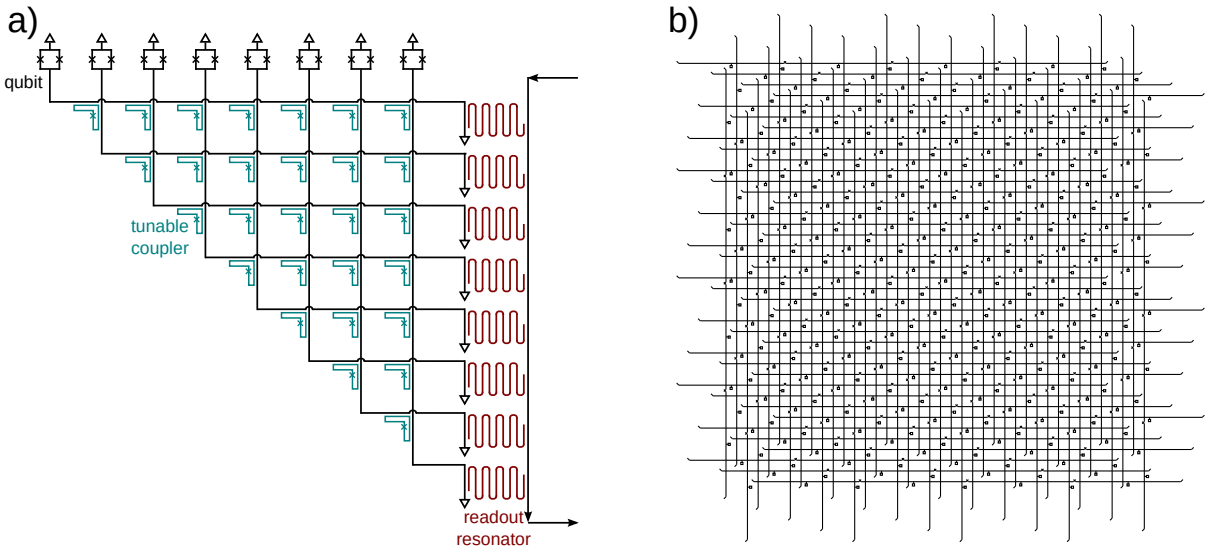


Figure 1.4: Possible small-scale and scalable architectures for the fluxmon qubit. Here, the schematic indicates the physical layout (unlike in Fig. 1.3), so the lines represent qubits and their intersections represent couplings. (a) Fully connected architecture for small-scale demonstrations. (b) Proposed tangled fabric of fluxmons by for medium-scale applications (couplers not shown) [31].

improve the performance of existing annealers. A large part of this must be achieved via control hardware improvements, but improvement can also be gained through better device modeling.

In this work, we develop a variant of the flux qubit, called the "fluxmon", that is suitable for high-connectivity quantum annealing without lossy dielectrics. In particular, we show that a scalable multi-layer architecture without lossy dielectrics is possible by building flux qubits with superconducting air-gap crossovers within one chip and superconducting bump bonds between chips. We accordingly observe substantially reduced microwave dissipation relative to D-Wave's qubits. We also demonstrate strong tunable coupling between qubits, and introduce more sophisticated yet scalable modeling for inter-qubit coupling that goes beyond linear theory. Furthermore, the fluxmon design is compatible with high connectivity graphs, and we expect that one fluxmon can in

principle couple to up to 10 to 20 others simultaneously while maintaining large coupling strengths. With the fluxmon, we can hope to demonstrate small-scale fully connected graphs such as the one shown in Fig. 1.4(a), and larger scale weaving networks shown in Fig. 1.4(b) that will likely be improvements over the existing Chimera architecture.

An outline of this thesis is as follows. In Chapter 2, we briefly review the basics of superconducting quantum circuits and Josephson junctions, and describe the theory behind the fluxmon circuit and how it can be used as a flux qubit. We provide analyses of circuit models with varying degrees of sophistication that can be used to model real devices to high accuracy. In Chapter 3, we show how the fluxmon is physically implemented, and describe in detail how it can be measured and manipulated. In particular, we will show that we can implement fast initialization of the $|L\rangle$ and $|R\rangle$ states, meaning our hardware is capable of implementing 'annealing in reverse' as required for quantum parallel tempering. We also show how our high-bandwidth experimental setup along with lower-dissipation qubits allow the demonstration of single-qubit annealing faster than the system thermalization time, a previously unaccessed regime. In Chapter 4, we perform an in-depth study of noise and dissipation in the fluxmon, and in particular implement the first measurement that extracts both the "classical" and "quantum" parts of the $1/f$ flux noise intrinsic to superconducting qubits. In particular, we are able to observe for the first time the classical-quantum crossover of flux noise, restricting the possible microscopic physical mechanisms that could be responsible for it. Remarkably, we also observe atomic hydrogen as a source of magnetic dissipation at 1.42 GHz. We show that indepen-

dent measurements of noise and dissipation can roughly predicted incoherent quantum tunneling rates between flux qubit wells. In Chapter 5, we describe the theory behind tunable couplers for fluxmon qubits, and in particular provide a scalable, nonlinear Born-Oppenheimer analysis of the circuit necessary when the coupling is very strong, allowing us to model our qubits to a higher degree of accuracy than the linear models typically used to model the D-Wave device. We also discuss the important tradeoff between the ability to couple strongly to many qubits and the ability to have low sensitivity to noise. Finally, in Chapter 6, we implement ultrastrong tunable coupling between two fluxmon qubits in a way that in principle extends to coupling to many other qubits at once. We perform the first direct spectroscopic measurements of tunable ultrastrong coupling between superconducting qubits, and demonstrate two-qubit annealing faster than the system thermalization time.

Chapter 2

The fluxmon qubit: design, theory and modeling

In this chapter, we review the essential physics of superconducting circuits and introduce a qubit circuit called the "fluxmon," a tunable gap flux qubit designed for high-connectivity quantum annealing without lossy dielectrics. We will show how this qubit circuit can in principle be used for quantum annealing. We will see that there are various levels at which one can model the fluxmon qubit, and the complexity of the model will depend on how accurately one wishes to model the experimental device at hand. For example, given that the fluxmon is implemented with a long segment of superconducting transmission line, we quantify the accuracy of a continuous transmission line model versus a more efficiently computable and scalable lumped-element model.

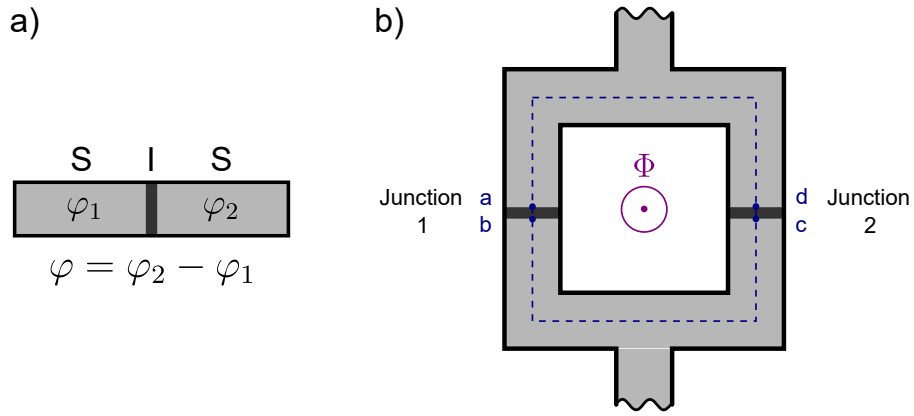


Figure 2.1: a) Josephson junction. Gray represents superconducting metal, and the black strip represents a tunnel junction. (b) DC SQUID comprised of two Josephson junctions in parallel.

2.1 Superconducting phase and the Josephson effect

Before we delve into the details of the fluxmon qubit and how it can be used for quantum annealing, we need to understand the basics of superconducting circuits, and in particular how Josephson junctions and DC SQUIDs work. We will see how a Josephson junction acts as a nonlinear inductor, giving rise to the nonlinearity necessary for an otherwise harmonic superconducting circuit system to form a viable qubit. We will then see how a DC SQUID can effectively act as a single Josephson junction whose critical current can be changed via an applied magnetic flux, giving the tunability needed for quantum annealing with flux qubits.

2.1.1 Josephson junctions

The Josephson junction is arguably the most important building block of superconducting qubits. As illustrated in Fig. 2.1, Josephson junctions are formed when a superconductor is interrupted by a layer of insulator (in our case, $\sim 1\,\text{nm}$ of AlO_x) that is thin enough to

allow appreciable tunneling of superconducting electrons. The microscopics of Josephson junctions have been studied in many other works, so we take as our starting point the well-known BCS theory, put forward by Bardeen, Cooper and Schrieffer in 1957 [32], and the Josephson tunneling effects, predicted by Brian Josephson in 1962 [33, 34]. In the standard BCS theory of superconductivity, within the superconductor, Cooper pairs of electrons (with mass $2m_e$, charge $2e$, and net spin zero) condense into an overlapping common ground state[1] that can be described by a macroscopic wavefunction

$$\Psi(\boldsymbol{r}) = \sqrt{n_{\mathrm{cp}}(\boldsymbol{r})}e^{i\varphi(\boldsymbol{r})}, \tag{2.1}$$

where $n_{\mathrm{cp}} = |\Psi(\boldsymbol{r})|^2$ is the density of Cooper pairs and is usually taken to be constant within a uniform superconductor. This fact is important because it means we can ignore all the internal degrees of freedom of the superconductor and just consider one degree of freedom: the macroscopic phase φ. The current and voltage across a Josephson junction are then related to the phase difference across it via the famous Josephson relations [33, 34],

$$I = I_c \sin\varphi, \tag{2.2}$$

$$V = \frac{\Phi_0}{2\pi}\frac{d\varphi}{dt}, \tag{2.3}$$

[1]This ground state is separated from electron-like and hole-like quasiparticle excitations by an energy gap Δ (corresponding to $T_c \sim 1.1\,\mathrm{K}$ in our thin film Al), meaning it will take an energy of 2Δ to break a Cooper pair. At millikelvin temperatures and low applied voltages we can assume that all tunneling across the junction is superdconducting, described by the Josephson relations given here. Later on we will consider the dissipative effects of a small nonequilibrium occupation of quasiparticles that may be present in real superconducting qubit systems.

where I_c is the critical current of the junction (the maximum non-dissipative current that can flow through the junction, determined by the geometry and materials of the junction and inversely proportional to the normal-state tunnel resistance) and $\varphi = \varphi_2 - \varphi_1$ is the phase difference across the junction.[2] An important consequence of the Josephson effects can be seen by differentiating the first Josephson equation with respect to time and then substituting into the second one, giving

$$V = \frac{\Phi_0}{2\pi} \frac{1}{I_c \cos\varphi} \frac{dI}{dt} \equiv L_J \frac{dI}{dt}, \tag{2.4}$$

where L_J is called the Josehpson inductance.[3] Note that the Josephson inductance is nonlinear and can even be negative, a crucial fact relevant to the operation of flux qubits and inter-qubit couplers. In the small-phase limit the junction will behave like a linear inductor with $L = L_{J0} = \frac{\Phi_0}{2\pi I_c}$.

To obtain a quantum model for superconducting qubits, we would like to have a Hamiltonian describing the energy associated with a Josephson junction, and therefore the time evolution of systems involving Josephson junctions. Ignoring the capacitance across the Josephson junction for now, we can find the energy stored in the junction by integrating the usual expression,

$$\begin{aligned} U &= \int_0^t I(t')V(t')\,dt' = \int_0^t I_c \sin\varphi \frac{\Phi_0}{2\pi}\frac{\partial\varphi}{\partial t'}\,dt' = \frac{\Phi_0}{2\pi} I_c \int_{\varphi(0)}^{\varphi(t)} \sin\varphi\, d\varphi \\ &\rightarrow -E_J \cos\varphi, \end{aligned} \tag{2.5}$$

[2] We have assumed the junction/weak link has zero spatial thickness. If not, then we must technically use a gauge-invariant phase difference instead, $\varphi = \varphi_2 - \varphi_1 - \frac{2\pi}{\Phi_0}\int_1^2 \boldsymbol{A}(\boldsymbol{r}) \cdot d\mathbf{l}$.

[3] Also note that L_J is the differential inductance $d\Phi/dI$, not the "secant" inductance Φ/I.

where we ignored an inconsequential offset term in the last step and we have defined the characteristic Josephson energy

$$E_J = \frac{\Phi_0}{2\pi} I_c. \tag{2.6}$$

So, viewing φ as a position-like coordinate, we see that a Josephson junction adds a nonlinear cosine potential energy term to the system Hamiltonian.[4] This derivation has been classical, but we can promote the phase variable to an operator to obtain a quantum Hamiltonian for the Josephson junction,

$$\hat{H}_{JJ} = -E_J \cos\hat{\varphi}. \tag{2.7}$$

Note that the cosine term can also be written as $\cos\hat{\varphi} = [\exp(+i\hat{\varphi}) + \exp(-i\hat{\varphi})]/2$ where the operators $\exp(\pm i\hat{\varphi})$ can be interpreted as charge number displacement operators which couple states differing by one Cooper pair across the junction. This can be seen explicitly by writing the Josephson term in the charge basis,[5]

$$E_J \cos\hat{\varphi} = \sum_{n\in\mathbb{Z}} \frac{E_J}{2} \left(|n+1\rangle \langle n| + |n\rangle \langle n+1| \right), \tag{2.8}$$

where $|n\rangle$ are the discrete eigenstates of the charge operator with n Cooper pairs on an electrode. This gives a more intuitive interpretation of the tunneling of Cooper pairs across the insulating barrier. However, we will almost always find it more convenient to work in the continuous flux basis.

The justification for treating φ as a quantized variable lies in the BCS theory. Exper-

[4] We have ignored the intrinsic capacitance across the junction, which adds a kinetic term to the Hamiltonian. For our purposes this is usually negligible, but we will see how to deal with such terms in section 2.2.3.

[5] We haven't defined the charge basis yet, but we will derive it from canonical commutation relations in section 2.2.1. For now, believe that the charge operator corresponds to the canonical momentum to the flux/phase, $\hat{q} = -i\frac{d}{d\varphi}$.

imental confirmation of the quantum nature of the macroscopic phase difference was first demonstrated in the 1980s through quantum tunneling experiments [21, 35, 22], which observed the tunneling rate out of a potential well to be independent of temperature below a certain temperature, in accordance with quantum theory, and more directly in 1985 by Martinis *et al.*, where quantized energy levels were observed in agreement with theory that quantized $\varphi \rightarrow \hat{\varphi}$ [36]. Since then, there has been remarkable success of the quantized superconducting phase in predicting a wide variety of superconducting quantum experiments.

2.1.2 The DC SQUID as a tunable Josephson Junction

The fluxmon utilizes two Josephson junctions in a parallel arrangement known as a DC SQUID. The name "DC SQUID" comes from the fact that, historically, this circuit operates based on the DC Josephson effect, taking advantage of the resulting interference effects that occur through two parallel paths each containing a Josephson junction. For reasons that will soon be apparent, DC SQUIDs are often used as very sensitive magnetometers. A particularly interesting application of DC SQUIDs is magnetoencephalography, where arrays of SQUIDs inside a helmet-shaped vacuum flask are used to infer neural activity via currents in the brain, making use of both the high sensitivity and temporal resolution of the SQUID [37]. In this section we show how a DC SQUID behaves as a single Josephson junction with a critical current that is tunable by an applied magnetic flux.

The DC SQUID operates based on a phenomenon in superconductors known as fluxoid quantization, which is usually just called "flux quantization" (the technical distinction between these two names will be explained shortly). Flux quantization arises from the requirement that the macroscopic superconducting phase φ must be single-valued (modulo 2π) around any closed path within a superconductor. This imposes a constraint between the phases across any junctions within a superconducting loop and the magnetic flux threading the area defined by that loop. This constraint is enforced by the supercon-ducting current within the loop. By the DC Josephson equation and Kirchoff's rule, the current flowing across the DC SQUID is given by $I_{\mathrm{SQ}} = I_{j1} + I_{j2} = I_{c1}\sin\varphi_1 + I_{c2}\sin\varphi_2$. We will now show how flux quantization imposes a constraint between φ_1, φ_2, and applied magnetic flux, reducing the DC SQUID to a circuit element with a single effective phase.

More concretely, consider the DC SQUID circuit in Fig. 2.1(b) with two Josephson junctions in parallel. Using the macrosopic wavefunction (2.1), the supercurrent density at any point within the superconductor is given by

$$\mathbf{J}_s = \mathrm{Re}[q\langle\Psi|\mathbf{v}|\Psi\rangle] = \frac{n_{\mathrm{cp}}q}{m}(\hbar\nabla\varphi(\boldsymbol{r}) - q\mathbf{A}), \tag{2.9}$$

where we have used the usual quantum mechanical velocity operator

$$\mathbf{v} = \frac{1}{m}(\mathbf{p} - q\mathbf{A}), \tag{2.10}$$

with $m = 2m_e$ and $q = -2e$ the mass and charge of a Cooper pair, and $\mathbf{A}$ is the magnetic vector potential. Rearranging terms we have that

$$\nabla\varphi(\boldsymbol{r}) = \frac{m}{n_{\mathrm{cp}}q\hbar}\mathbf{J}_s + \frac{q}{\hbar}\mathbf{A}. \tag{2.11}$$

Next, consider integrating $\nabla\varphi(\boldsymbol{r})$ around the loop of the DC SQUID, taking the integration path denoted by the dashed blue line in Fig. 2.1(b). Taking the integral one labeled segment at a time, we have

$$\varphi_b - \varphi_a = -\varphi_1 + \frac{q}{\hbar}\int_a^b \mathbf{A}\cdot d\boldsymbol{l} \tag{2.12}$$

$$\varphi_d - \varphi_c = \varphi_2 + \frac{q}{\hbar}\int_c^d \mathbf{A}\cdot d\boldsymbol{l} \tag{2.13}$$

$$\varphi_c - \varphi_b = \frac{m}{n_{\mathrm{cp}}q\hbar}\int_b^c \mathbf{J}_s\cdot d\boldsymbol{l} + \frac{q}{\hbar}\int_b^c \mathbf{A}\cdot d\boldsymbol{l} \tag{2.14}$$

$$\varphi_a - \varphi_d = \frac{m}{n_{\mathrm{cp}}q\hbar}\int_d^a \mathbf{J}_s\cdot d\boldsymbol{l} + \frac{q}{\hbar}\int_d^a \mathbf{A}\cdot d\boldsymbol{l}, \tag{2.15}$$

where φ_1 and φ_2 are the gauge-invariant phase differences across junctions 1 and 2, both defined with an orientation looking towards the top of the circuit. The change in macroscopic phase around a closed loop must be zero modulo 2π, so we must therefore have

$$2\pi n = \oint \nabla\varphi(\boldsymbol{r})\cdot d\boldsymbol{l} = \varphi_2 - \varphi_1 + \frac{m}{n_{\mathrm{cp}}q\hbar}\oint \mathbf{J}_s\cdot d\boldsymbol{l} + \frac{q}{\hbar}\oint \mathbf{A}\cdot d\boldsymbol{l}, \tag{2.16}$$

with n an integer. We recognize the line integral of $\mathbf{A}$ around the loop as the total magnetic flux threading the DC SQUID loop (Stokes' theorem), and also recognize $|q|/\hbar = \Phi_0/(2\pi)$, where

$$\Phi_0 = \frac{h}{2e} \approx 2.068\times 10^{-15}\,\mathrm{Wb} \tag{2.17}$$

is the magnetic flux quantum. Rearranging (2.16) then yields

$$\varphi_1 - \varphi_2 = 2\pi n + \frac{2\pi}{\Phi_0}\Phi + \frac{2\pi}{\Phi_0}\frac{m}{n_{\mathrm{cp}}q^2}\oint \mathbf{J}_s\cdot d\boldsymbol{l}. \tag{2.18}$$

Here, Φ is the sum of any external magnetic flux applied through the loop and the magnetic flux generated by any geometric magnetic inductance within the SQUID loop.

Let us consider a few special cases of Eq. (2.18). We first consider a loop without Josephson junctions (i.e., $\varphi_1 = \varphi_2 = 0$). Defining the *fluxoid* as originally introduced by London as [34]

$$\Phi' = \Phi + \frac{m}{n_{\mathrm{cp}} q^2} \oint \mathbf{J}_s \cdot d\boldsymbol{l}, \tag{2.19}$$

we see that Eq. (2.18) reduces to what is known as *fluxoid quantization* around a superconducting loop: the fluxoid associated with any superconducting loop is quantized in units of Φ_0. Next, further assume that the thickness of the loop metal is sufficiently larger than the penetration depth so that the current density vanishes inside of the superconductor along the path of integration. In this case, the fluxoid is simply equal to the magnetic flux threading the loop, $\Phi' = \Phi$. In such a loop, we have the phenomenon of *flux quantization* proper: the magnetic flux through the loop is quantized in units of Φ_0. Note that physically, Φ_0 is comparable to the amount of flux threaded by the Earth's magnetic field through a $\sim 50\,\mu\mathrm{m}^2$ loop, one of many reasons that magnetic shielding is crucial in SQUID devices, whose loop sizes are typically of this order.

Usually, in superconducting circuits the thickness of the metal is comparable to or larger than the London penetration depth λ, so that the term $\frac{m}{n_{\mathrm{cp}} q^2} \oint \mathbf{J}_s \cdot d\boldsymbol{l}$ is nonzero but small. In this case the magnetic flux is actually quantized in units of roughly $\Phi_0(1 + 2\lambda^2/(rd))^{-1}$ (for an idealized cylindrical geometry [38]), where d is the width of the loop's metal and r is the radius of the loop. For typical qubit SQUID dimensions we expect

this correction to be $\lesssim 1\%$, but as we will see shortly, we can incorporate this correction into an effective inline kinetic inductance within the DC SQUID (which is in fact the physical origin of this term). Kinetic inductance is associated with the inertia (i.e., kinetic energy) of the Cooper pairs that carry the superconducting current. In normal (ohmic) metals, it is usually negligible at low frequencies because the kinetic energy of charge carriers is dissipated on the timescale of scattering (think imaginary part of the Drude model conductivity). In superconductors however, the kinetic inductance can be quite important. Classically, the concept of kinetic inductance can be understood by comparing the expression for the energy stored in an inductor by a current ($\frac{1}{2}LI^2$) with the kinetic energy associated with a current ($\sum \frac{1}{2}m_{q_i}v_{q_i}^2$) – both are proportional to the current squared. So, even though the kinetic energy is not stored in the magnetic field of any geometric inductance, it can still be treated like an inductance. The kinetic energy associated with a superconducting Cooper pair current is given by the volume integral

$$E_{\mathrm{k}} = \int_V \frac{1}{2} n_{\mathrm{cp}} m v^2 \, dV \approx \frac{1}{2} \frac{l}{A} \frac{m}{n_{\mathrm{cp}} q^2} I_s^2, \tag{2.20}$$

where l and A are the length and cross-sectional area of the superconducting wire and I_s is the superconducting current through the wire. From Eq. (2.20) we can identify the kinetic inductance[6]

$$L_k = \frac{l}{A} \frac{m}{n_{\mathrm{cp}} q^2}. \tag{2.21}$$

Looking back at the definition of the fluxoid (2.19), we then see that the extra term added to the total magnetic flux is simply the flux associated with the kinetic inductance

[6]In our simple derivation we have ignored variation of the cross-sectional area over the wire and also the exact current distribution of the cross-sectional area, but this does not matter for the result.

of (2.21), $\Phi_k = L_k I$. This makes the fluxoid easier to interpret, as later on we will see how to deal with inline inductances within the DC SQUID loop. An important fact to keep in mind is that although the flux Φ itself might not always be quantized in units of exactly one Φ_0, the response of the DC SQUID is still exactly Φ_0-periodic in any applied external magnetic flux. This is a crucial property that will be used later on for calibrating mutual inductances between various qubit loops and their bias lines.

Now let us add back the two Josephson junctions to the DC SQUID loop, as was drawn in Fig. 2.1, but for simplicity ignore any inline geometric or kinetic inductances due to the superconducting wires themselves. For simplicity fix the quantized flux at $n = 0$ (this assumption is not necessary though). In this case the flux is simply equal to the applied magnetic flux, $\Phi = \Phi_{\text{ext}}$. The flux quantization condition (2.18) then reduces to

$$\varphi_1 - \varphi_2 = \frac{2\pi}{\Phi_0}\Phi_{\text{ext}}. \tag{2.22}$$

Defining the new variables

$$\begin{aligned} \varphi &\equiv \frac{\varphi_1 + \varphi_2}{2}, \\ \delta &\equiv \varphi_1 - \varphi_2, \end{aligned} \tag{2.23}$$

flux quantization dictates that $\delta = (2\pi/\Phi_0)\Phi_{\text{ext}}$. Using this constraint, we can rewrite

the DC SQUID Hamiltonian in a more useful form using trigonometric identities:

$$\begin{aligned} H_{\mathrm{SQUID}} &= -E_{J1}\cos\varphi_1 - E_{J2}\cos\varphi_2 \\ &= -E_{J1}\cos\left(\varphi + \frac{\delta}{2}\right) - E_{J2}\cos\left(\varphi - \frac{\delta}{2}\right) \\ &= -E_{J1}\left(\cos\varphi\cos\frac{\delta}{2} - \sin\varphi\sin\frac{\delta}{2}\right) - E_{J2}\left(\cos\varphi\cos\frac{\delta}{2} + \sin\varphi\sin\frac{\delta}{2}\right) \\ &= -E_J^{\mathrm{sum}}\left(\cos\left(\pi\frac{\Phi_{\mathrm{ext}}}{\Phi_0}\right)\cos\varphi - \chi\sin\left(\pi\frac{\Phi_{\mathrm{ext}}}{\Phi_0}\right)\sin\varphi\right), \end{aligned} \tag{2.24}$$

where

$$E_J^{\mathrm{sum}} \equiv E_{J1} + E_{J2} \tag{2.25}$$

and

$$\chi \equiv \frac{E_{J1} - E_{J2}}{E_{J1} + E_{J2}} \tag{2.26}$$

is a measure of junction asymmetry, which is ideally small (in the submicron AlO_x junctions used in this work, a few percent or less with standard fabrication techniques). In the SQUID Hamiltonian (2.24), we interpret φ as the operator for the SQUID degree of freedom, while δ is treated as a classical number. This is because by flux quantization, δ is constrained to equal the external bias flux, which behaves as a classical variable. Later when we consider inline inductance and capacitance within the DC SQUID loop we will have to treat δ as an operator as well to have an accurate description of the system, although if the inline inductances are small enough this can be ignored because the extra degrees of freedom are very fast and can be thought of as always remaining in

their classical ground states.[7]

If we ignore junction asymmetry, we thus see that an ideal DC SQUID will behave as a single effective Josephson junction (with φ acting as the phase across the junction) with a flux-tunable Josephson energy

$$E_J^{\text{eff}} = E_J^{\text{sum}} \cos\left(\pi \frac{\Phi_{\text{ext}}}{\Phi_0}\right). \tag{2.27}$$

Often in the literature this is written as $E_J^{\text{sum}} \left|\cos\left(\pi \frac{\Phi_{\text{ext}}}{\Phi_0}\right)\right|$, but as we will see in the context of the fluxmon qubit, there is a crucial observable difference between the positive and negative signs for this term (this is in contrast to other DC SQUID-based qubits like the split transmon, where the DC SQUID is itself not part of a DC loop).

2.2 Fluxmon circuit model and circuit quantization

The fluxmon circuit is schematically illustrated in Fig. 2.2. It consists of a segment of superconducting transmission line (typically CPW or close to CPW) that is shorted on one end ($x = 0$) and terminated with a DC SQUID (acting as a tunable Josephson junction) on the other ($x = \ell$). The aim of this section is to show how such a circuit can serve as a tunable gap flux qubit that is able to encode the Hamiltonian necessary for quantum annealing. There are various levels of approximations at which one can describe the fluxmon circuit. For now, we will ignore any geometric inductances within the loop forming the DC SQUID. This is justified if they are small enough such that the

[7] We'll discuss this idea further in the context of the Born-Oppenheimer approximation in a later section.

resulting modes within the DC SQUID have a characteristic frequency much higher than the operating frequencies of the qubit. Later on we will add finite DC SQUID inline inductances in order to quantify their effects. For now let us also assume that the two junctions of the DC SQUID are identical. We will later add in the important effect of undesired junction asymmetry. As shown in 2.1.1, we can then treat the DC SQUID simply as a single Josephson junction whose Josephson energy is tunable by the external flux bias Φ_{SQ} threaded through it, according to $E_J(\Phi_{\mathrm{SQ}}) = E_J^{\max} \cos(\pi\Phi_{\mathrm{SQ}}/\Phi_0)$, where $\Phi_0 = h/2e$ is the magnetic flux quantum.

The inductance of the main body of the fluxmon is distributed throughout a rather long length of a few millimeters in order to accommodate simultaneous coupling to many other qubits at once, a desirable feature for quantum annealing that will be discussed in chapters 5 and 6. This idea has also been used by the company D-Wave with microstrip-based flux qubits [16]. We chose a coplanar waveguide geometry for the fluxmon body largely due to its compatibility with low dielectric dissipation, as demonstrated in CPW resonators [39] and in the Xmon transmon qubit, whose capacitor geometry is quite close to that of a CPW [40]. One can model a CPW line (or any transmission line) as having a uniform distributed series inductance per unit length $\mathcal{L}$ and a uniform distributed parallel capacitance per unit length $\mathcal{C}$, both of which are independent of frequency.[8] If the CPW

[8]The physical justification for such a model can be found by looking at the electric and magnetic field distributions in an arbitrary uniform TEM transmission line geometry and equating the time-averaged electric and magnetic energy per unit length with the corresponding circuit quantities [41], meaning the two models are dynamically equivalent. Note however that a real CPW supports "quasi-TEM" waves, and the assumption of a constant wave velocity ignores any frequency dispersion that can occur with a CPW geometry on a finite dielectric substrate. At the operating frequencies we will consider though ($f \ll 1\,\mathrm{THz}$), such effects are negligible [42].

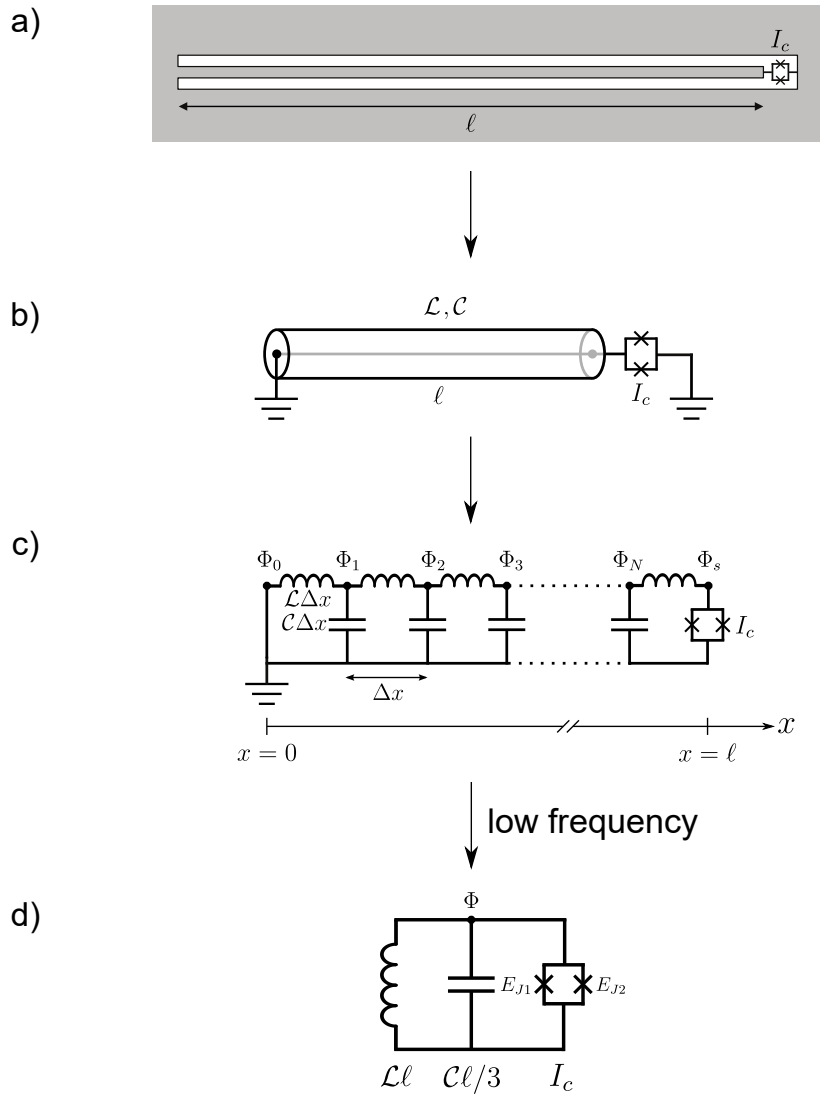


Figure 2.2: Fluxmon circuit. As is standard, Josephson junction circuit elements are drawn with an "×" symbol.

segment of the fluxmon is short enough (i.e., if the relevant operating frequencies are low enough), one would expect that it behaves as a lumped element inductance and capacitance. This intuition turns out to be correct, as can be shown using standard transmission line theory.[9] The input admittance looking into the shorted transmission line through a port located at the position of the DC SQUID is given by [41]

$$Y_{\text{in}}(\omega) = \frac{1}{iZ_0 \tan(\omega\ell/v)} \tag{2.28}$$

$$\approx -i\sqrt{\frac{\mathcal{C}}{\mathcal{L}}}\left(\frac{1}{\omega\ell\sqrt{\mathcal{L}\mathcal{C}}} - \frac{\omega\ell\sqrt{\mathcal{L}\mathcal{C}}}{3}\right) \tag{2.29}$$

$$= \frac{1}{i\omega\mathcal{L}\ell} + i\omega\frac{\mathcal{C}\ell}{3}, \tag{2.30}$$

where $Z_0 = \sqrt{\mathcal{L}/\mathcal{C}}$ is the characteristic impedance of the line and $v = 1/\sqrt{\mathcal{L}\mathcal{C}}$ is the frequency-independent speed of wave propagation, and where in the second line we have taken the low-frequency limit $f \ll f_{\lambda/4} = v/(4\ell)$, which is $\approx 20\,\text{GHz}$ in typical designs. We can immediately recognize Eq. (2.30) as the admittance of a lumped element parallel LC circuit, with effective inductance $\mathcal{L}\ell$ and capacitance $\mathcal{C}\ell/3$. Physically, the reduction in C from $\mathcal{C}\ell$ to $\mathcal{C}\ell/3$ comes from the fact that one end of the CPW is shorted, meaning that there must be a voltage profile that shrinks to zero at $x = 0$, so the full capacitance per unit length does not contribute equally over the entire length of the CPW.

The accuracy of the low-frequency expression (2.30) relative to the full transmission

[9] From the telegraph equations, which describe forward and backwards propagating waves along a transmission line, one can derive that the input impedance (defined as net voltage over current) seen looking into a finite transmission line of characteristic impedance $Z_0 = \sqrt{\frac{\mathcal{L}}{\mathcal{C}}}$ and terminated with a load impedance Z_L at $x = \ell$ from the input is $Z_0 \frac{Z_L + iZ_0 \tan k\ell}{Z_0 + iZ_L \tan k\ell}$, with $k = 2\pi/\lambda$ for a lossless line [41].

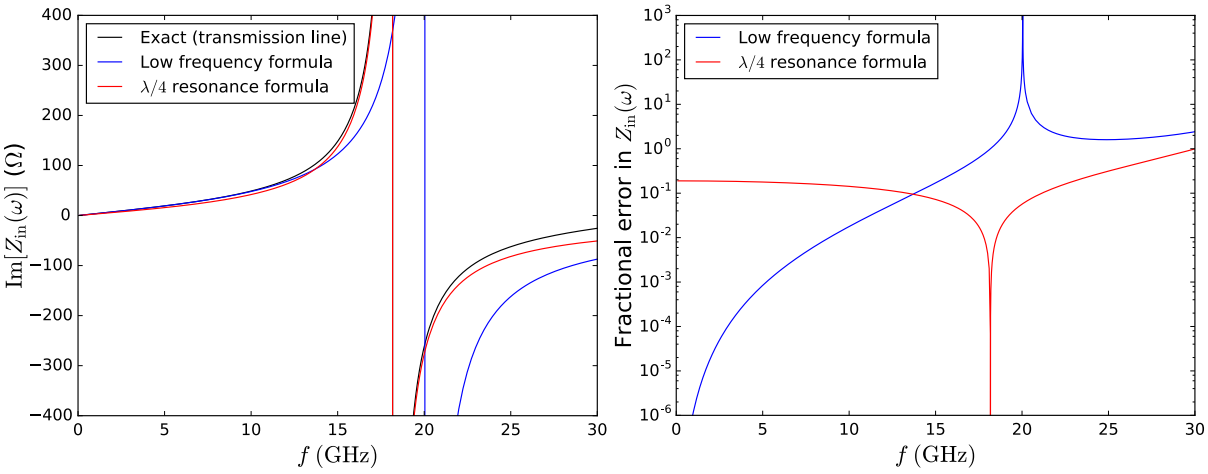


Figure 2.3: Exact vs. approximate (parallel LC) expressions for transmission line segment impedance vs. frequency, for $\ell = 1750\,\mu$m. Near the $\lambda/4$ resonance, the $\lambda/4$ resonance formula from Pozar [41] works very well. Below $\frac{1}{2}f_{\lambda/4} \approx 10\,$GHz, the fractional error in the low frequency formula (2.30) is less than one part in one hundred, and below $\frac{1}{4}f_{\lambda/4} \approx 5\,$GHz it is less than one part in one thousand.

line admittance (2.28) is shown in Fig. 2.3. Below $\frac{1}{2}f_{\lambda/4} \approx 10\,$GHz, the fractional error in the low frequency formula (2.30) is less than one part in one hundred, and below $\frac{1}{4}f_{\lambda/4} \approx 5\,$GHz it is less than one part in one thousand. In section 2.2.2 we discuss the consequences of this deviation, as well as the consequences of using a 1D model that ignores higher modes of the transmission line segment. The reader may skip the analysis in section 2.2.2 unless they are interested in predicting the fluxmon spectrum to high accuracy, but note that such an analysis is crucial to understanding the upper limit on how long we can make the fluxmon's CPW while also having a valid lumped LC model at the frequencies we care about. As a notational note, we will denote the true harmonic quarter-wave resonance frequency by $f_{\lambda/4}$, and the harmonic resonance according to the low-frequency model by $f_{LC} \equiv \frac{1}{2\pi}\frac{1}{\sqrt{(\mathcal{L}\ell)(\mathcal{C}\ell/3)}}$. These two values differ by only a factor of $\frac{2\sqrt{3}}{\pi} \approx 1.10$. Although f_{LC} doesn't correspond to any physically real resonance of the

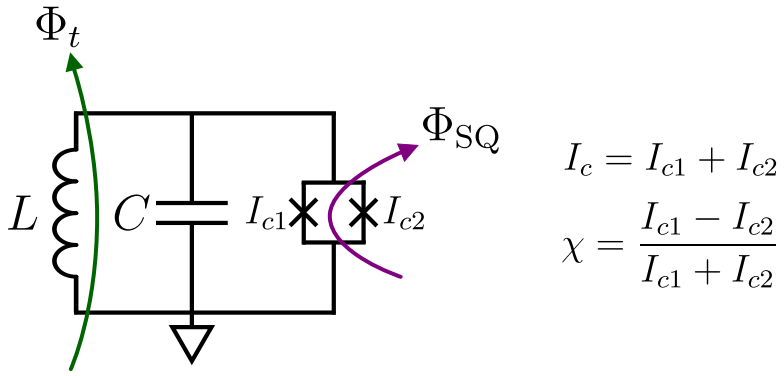


Figure 2.4: 1D fluxmon circuit model consisting of a lumped element inductor, capacitor, and DC SQUID acting as a tunable Josephson junction. The flux node for the one-dimensional Hamiltonian is at the top of circuit, while the bottom node of the circuit is taken to be a reference ground.

system, it is still a useful quantity for parameterizing the overall energy scale within the low-frequency 1D fluxmon model, the primary circuit model we will use for the fluxmon.

2.2.1 1D circuit model

With the CPW approximated as a lumped parallel LC circuit as in Fig. 2.2d and ignoring any inline geometric inductances within the DC SQUID, the fluxmon has one degree of freedom (taken to be the node at the top of the circuit) and the Hamiltonian takes a one-dimensional form. Such a 1D model is particularly easy to work with and desirable to use whenever it is applicable. We will now study this model in depth and show how such a circuit can be used as a tunable gap flux qubit.

2.2.1.1 1D Hamiltonian and control knobs

To control the qubit formed by this circuit, as illustrated in Fig. 2.4 we add an external flux bias Φ_t to the main loop to effectively bias the phase across the DC SQUID. We

also add an external flux bias $\Phi_{\rm SQ}$ to the DC SQUID loop to control the effective critical current of the SQUID. The subscript of Φ_t stands for "tilt", the reason for which will be explained shortly in the context of the conventional flux qubit double-well potential.[10]

Let us first consider the ideal case of symmetric junctions with identical critical currents $I_{c1} = I_{c2}$ (i.e., $\chi = 0$). The 1D fluxmon Hamiltonian is then

$$\hat{H}_{\rm 1D} = \frac{(\hat{\Phi} - \Phi_t)^2}{2L} + \frac{\hat{Q}^2}{2C} - E_J^{\rm max} \cos\left(\pi \frac{\Phi_{\rm SQ}}{\Phi_0}\right) \cos\left(2\pi \frac{\hat{\Phi}}{\Phi_0}\right) \tag{2.31}$$

where $E_J^{\rm max} = E_{J1} + E_{J2} = \frac{\Phi_0}{2\pi} I_{c1} + \frac{\Phi_0}{2\pi} I_{c2}$ and $\hat{\Phi}$ and $\hat{Q}$ are the canonically conjugate[11] flux and charge operators across the compound junction satisfying $[\hat{\Phi}, \hat{Q}] = i\hbar$. This commutation relation can be deduced using canonical quantization of LC circuits, with flux and charge being directly analogous to position and momentum. The first term in (2.31) represents the potential flux energy stored in the inductor. The second term is the kinetic charge energy within the capacitor. The third term is the Josephson potential energy arising from the flux across the effective Josephson junction formed by the DC SQUID. It is convenient to re-write the Hamiltonian (2.31) in terms of the dimensionless phase and charge operators $\hat{\varphi} = \frac{2\pi}{\Phi_0}\hat{\Phi}$ (understood to have units of radians) and $\hat{q} = \frac{1}{2e}\hat{Q}$, with $[\hat{\varphi}, \hat{q}] = i$,

$$\hat{H}_{\rm 1D} = E_L \frac{(\hat{\varphi} - \varphi_t)^2}{2} + 8E_C \frac{\hat{q}^2}{2} + \beta E_L \cos(\hat{\varphi}), \tag{2.32}$$

[10]Note that in the actual physical implementation of the circuit, the flux bias Φ_t through the main CPW inductor is a gradiometric flux bias, with the flux having opposite sign on either side of the CPW center trace when looking into the plane of the CPW. This is because the corresponding current flow needs to be the CPW mode of the transmission line (current flowing down the center trace in one direction and flowing back symmetrically on either side of the ground plane, as illustrated in Fig. 3.1).

[11]The fact that charge and flux are canonical coordinates in electrical oscillators follows through a direct analogy of the equations of motion for flux and charge to position and momentum in a harmonic oscillator formed by a mass on a spring.

where we have also defined the useful parameters

$$E_L = \frac{1}{L}\left(\frac{\Phi_0}{2\pi}\right)^2, \tag{2.33}$$

$$E_C = \frac{e^2}{2C}, \tag{2.34}$$

$$\beta = -\frac{2\pi}{\Phi_0} I_c L = -L/L_J = -E_J/E_L. \tag{2.35}$$

From now on we will usually drop the hat symbols from the flux and charge operators, but it will be clear from context that they are operators.

To gain a bit more intuition, we can rewrite the 1D Hamiltonian yet again, this time in a dimensionless form,

$$\frac{H_{1\mathrm{D}}}{E_L} = \frac{(\varphi - \varphi_t)^2}{2} + \beta\cos\varphi + \left(\frac{Z_{LC}}{R_K/8\pi}\right)^2 \frac{q^2}{2}, \tag{2.36}$$

where

$$Z_{LC} = \sqrt{\frac{L}{C}} \tag{2.37}$$

is the fluxmon's characteristic impedance and

$$R_K = \frac{h}{e^2} \approx 25.8\,\mathrm{k\Omega} \tag{2.38}$$

is the resistance quantum. The dimensionless form (2.36) provides an intuitive interpretation of the characteristic impedance Z_{LC}: since it sets the scale of the charge term in the Hamiltonian, and the charge acts as the non-commuting momentum to the flux (phase), Z_{LC} must set the quantum uncertainty in the phase wavefunction; i.e., it sets the width of the wavefunction in the flux basis. In fact, as will be shown shortly, in the harmonic

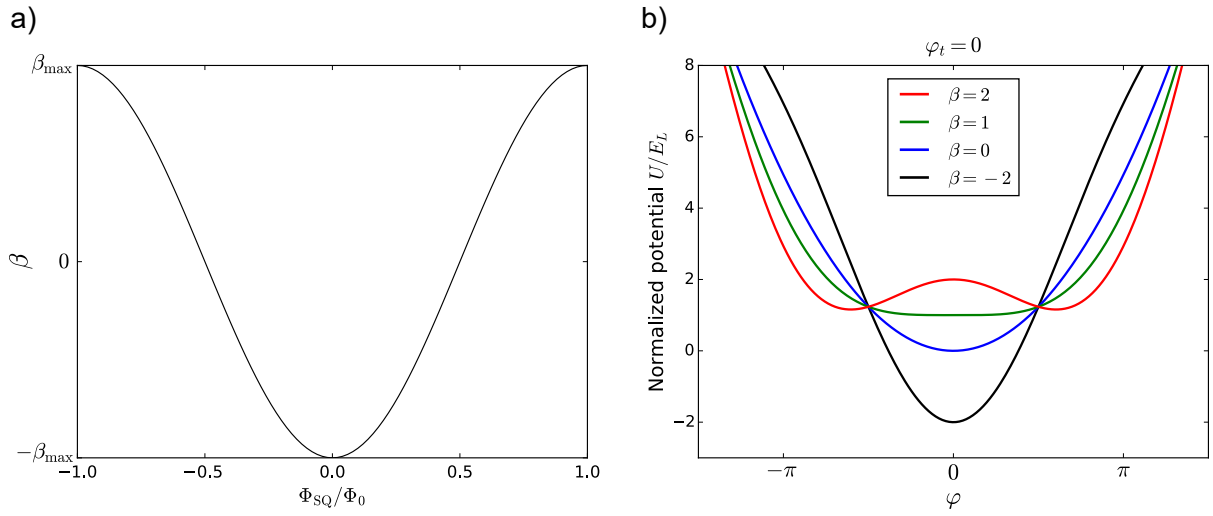


Figure 2.5: (a) Plot of $\beta(\Phi_{\text{SQ}}) = \beta_{\max}\cos(\pi\Phi_{\text{SQ}}/\Phi_0)$ vs. Φ_{SQ}, ranging from $-\beta_{\max}$ to $\beta_{\max}$. At zero physical applied flux biases in the system, the potential "barrier" is negative, leading to the convention of negative β at $\Phi_{\text{SQ}} = 0$. (b) Plot of the dimensionless potential energy $U(\varphi)$ vs. φ for different values of β. $\beta = 0$ corresponds to the harmonic limit, and $\beta > 1$ corresponds to a double-well potential.

limit there is a direct proportionality between the width of the Gaussian ground state wavefunction and Z_{LC}/R_K. Besides leading to generically quantum behavior, as we will see later on this finite width has important implications for coupling strength between qubits, and when considered in the context of the Born-Oppenheimer approximation for inter-qubit couplers gives quantum corrections to the otherwise classical ground state energy of the couplers.

The meaning of the important parameter $\beta = \beta_{\max}\cos(\pi\Phi_{\text{SQ}}/\Phi_0)$ becomes clear[12] when we plot the dimensionless potential energy part of the Hamiltonian, $U(\varphi)/E_L = \frac{1}{2}(\varphi - \varphi_t)^2 - \beta\cos\varphi$. As shown in Fig. 2.5(b), β is a measure of the barrier height in

[12]Note that in this work, the sign convention for β is such that β is negative when there is zero physical applied flux bias in the system, $(\Phi_{\text{SQ}}, \Phi_t) = (0, 0)$, which is reflected in Fig. 2.5(a). The reason for this convention is so that when $\Phi_t = 0$, positive β means "positive barrier" in the potential energy landscape, as can be intuitively understood in Fig. 2.5(b). In the literature there is often an alternative definition of tilt flux $\Phi_t \to \Phi_t - \Phi_0/2$ as in a conventional flux qubit, but we will not use it here.

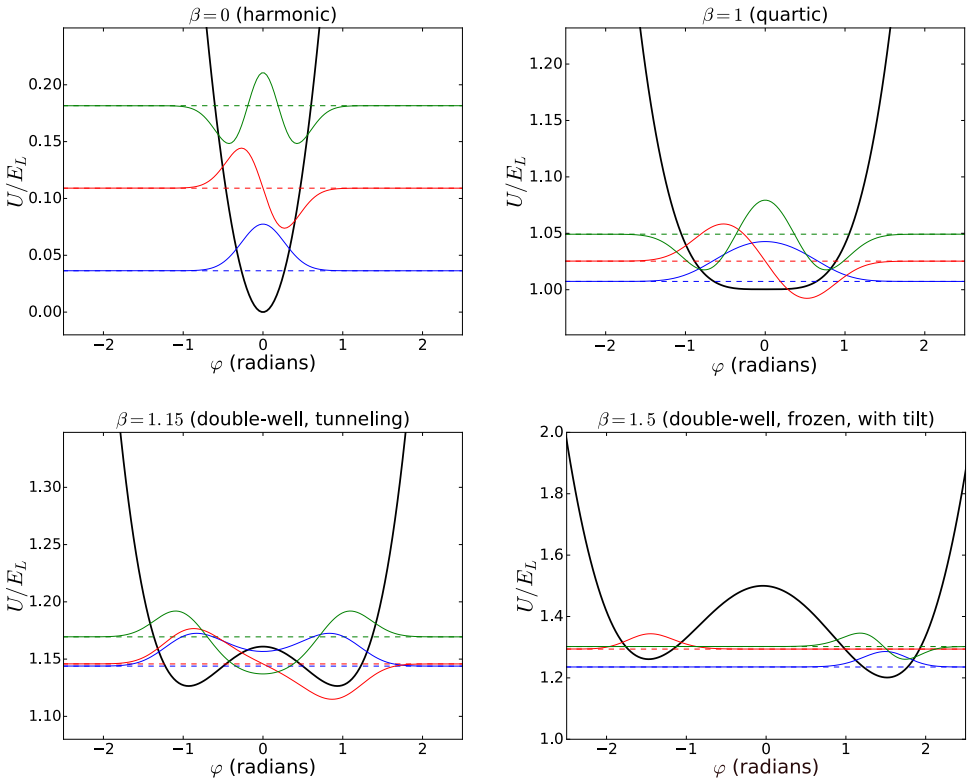


Figure 2.6: Plot of numerically computed fluxmon wavefunctions in the 1D model for various values of β, illustrating four qualitative regimes (harmonic, pure quartic, double-well with substantial tunneling, and double-well with suppressed tunneling). The blue, red, and green curves correspond to the wavefunctions of the ground, first excited, and second excited energy eigenstates. The dashed lines are the corresponding eigenenergies, and the black curve is the rescaled potential.

the double-well potential. When $\beta > 1$, we have a double-well potential. When $\beta < 1$, we have a single-well anharmonic potential. When $\beta = 0$ ($\Phi_{\text{SQ}} = \Phi_0/2$), the Josephson term vanishes and the 1D fluxmon becomes a harmonic oscillator. Note that β can also be negative, yielding an anharmonic potential with an anharmonicity that has opposite sign from that when $0 < \beta < 1$. The eigenenergies and wavefunctions for the first three energy eigenstates are illustrated for different values of β in Fig. 2.6.

The 1D Hamiltonian can be easily diagonalized to high accuracy by discretizing it over the interval $[\varphi_t - 2\pi, \varphi_t + 2\pi]$ and numerically diagonalizing the resulting matrix. Explic-

itly, this means considering discrete flux points φ_i with $i = 1, 2, ..., N$. The corresponding dimensionless potential energy matrix is then the diagonal matrix whose entries are given by $U_{nm}/E_L = \delta_{n,m}\left[\frac{(\varphi_i-\varphi_t)^2}{2} + \beta\cos\varphi_i\right]$. The matrix for the kinetic energy is found by using the relation $\hat{q} = -i\frac{d}{d\varphi}$ (this follows from the commutation relation $[\hat{\varphi}, \hat{q}] = i$). Since the difference quotient for the second derivative of a function f with respect to φ is $[f(\varphi+\delta\varphi)-2f(\varphi)+f(\varphi-\delta\varphi)]/(\delta\varphi)^2$, the discrete matrix for the kinetic energy becomes $T_{nm}/E_L = \frac{1}{2}\frac{8E_C}{E_L}\frac{1}{(\delta\varphi)^2}(2\delta_{n,m} - \delta_{n,n\pm1})$. Putting these together, the discrete matrix for the 1D fluxmon Hamiltonian over the discretized flux φ_i is

$$\frac{H_{nm}^{\text{1D,disc.}}}{E_L} = \delta_{n,m}\left[\frac{(\varphi_i-\varphi_t)^2}{2} + \beta\cos\varphi_i\right] + \frac{1}{2}\frac{8E_C}{E_L}\frac{1}{(\delta\varphi)^2}(2\delta_{n,m} - \delta_{n,n\pm1}). \tag{2.39}$$

To obtain an accuracy of 2 MHz (significantly less than the observed qubit linewidth) in $f_{10} \equiv (E_1 - E_0)/h$ and a fractional accuracy of $< 5 \times 10^{-5}$ for the off-diagonal matrix elements $\varphi_{10} \equiv \langle 0|\hat{\varphi}|1\rangle$ within the regime of interest, $N = 400$ points suffices and is still sufficiently fast. Using $N = 200$ is appreciably faster but will lead to errors up to 10 MHz in f_{10} in the regime of interest. We note that the discretization technique works efficiently when the Hamiltonian in one-dimensional but, as we will see later on, this technique becomes quickly intractable for circuits with more degrees of freedom, and so other techniques will be used to diagonalize more complex circuits.

Fig. 2.7(a) shows the numerically obtained f_{10} vs. the flux biases Φ_{SQ} and Φ_t within the 1D model for typical qubit parameters of $f_{LC} = 20\,\text{GHz}$ and $Z_{LC} = 75\,\Omega$. Note that the spectrum is Φ_0-periodic in tilt bias Φ_t and $2\Phi_0$-periodic in the DC SQUID flux

bias $\Phi_{\rm SQ}$ (this is because the DC SQUID term is set by the interference between two junctions that respond to the SQUID flux with opposite signs). To help understand the different regions of the spectrum, in Fig. 2.7(b) we have indicated the regions of the 2D bias space where the potential in monostable vs. bistable, along with vertical lines indicating where $|\beta| = 1$. The red/blue regions are where the potential is classically bistable/monostable respectively. More precisely, bistable points are the biases for which the equation $U'(\varphi) = \varphi - \varphi_t - \beta \sin\varphi = 0$ has three solutions [two local minima and one local maximum for $U(\varphi)$)]. We note that the bistable regime can be reached from the origin either by applying $\sim \Phi_0/2$ to the tilt bias, or instead by applying $\sim \Phi_0$ to the SQUID bias. The former is how a conventional flux qubit obtains a double-well potential, and can be thought of as moving the parabolic part of the potential relative to the cosine part. The latter instead corresponds to moving the cosine part relative to the parabolic part to achieve the same result. This is often useful in practice, allowing us to restrict the available flux range of the tilt bias line to protect the qubit from noise and dissipation.

In Fig. 2.8(a) we plot the numerically computed spectra f_{10} and f_{21} vs. β for a symmetric potential (i.e., zero tilt, $\Phi_t = 0$). Note that at $\beta = 0$, the anharmonicity $\eta \equiv f_{21} - f_{10}$ vanishes since the fluxmon potential becomes perfectly harmonic there, and as claimed earlier the anharmonicity at zero tilt therefore changes sign when β crosses through zero. When β is sufficiently larger than 1, f_{10} at zero tilt quickly approaches zero, in fact exponentially so. This can be understood through the WKB approximation for the inter-well tunneling in the double-well regime as discussed later on. We note

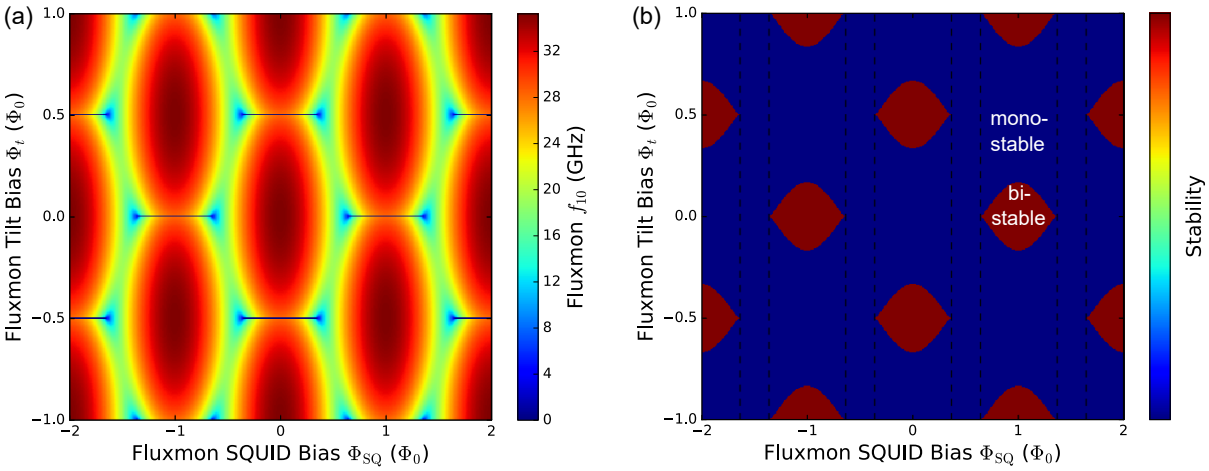


Figure 2.7: (a) Fluxmon $f_{10} = (E_1 - E_0)/h$ vs. SQUID and tilt flux biases according to the 1D Hamiltonian, for a typical fluxmon with $f_{LC} = 20\,\text{GHz}$, $Z_{LC} = 75\,\Omega$, and $\beta_{\max} = 2.4$. (b) Regions of classical monostability (blue) and bistability (red). Dashed vertical lines indicate where $|\beta(\Phi_{\text{SQ}})| = 1$.

that the value of β at which the gap starts to decrease exponentially depends on the impedance. Increasing the impedance increases the value of β at which the gap starts to shrink exponentially. This is shown in the logarithmic plot of $\Delta = hf_{10}(\varphi_t = 0)$ vs. β in Fig. 2.8(b). This makes intuitive sense given that this parameter determines the width of the wavefunction. A higher impedance also gives a broader crossover region between the harmonic (single-well) and tunneling (double-well) regimes.

In Fig. 2.8(c) we plot the numerically computed dimensionless flux and charge matrix elements $\varphi_{10} \equiv |\langle 0|\hat{\varphi}|1\rangle|$ and $q_{10} \equiv |\langle 0|\hat{q}|1\rangle|$ at zero tilt. As will be discussed in subsequent chapters, these quantities determine how strongly the fluxmon couples to its environment through the flux or charge degree of freedom. By the same physics, these quantities also determine how strongly the fluxmon will couple to other quantum circuits through a mutual inductance or capacitance.

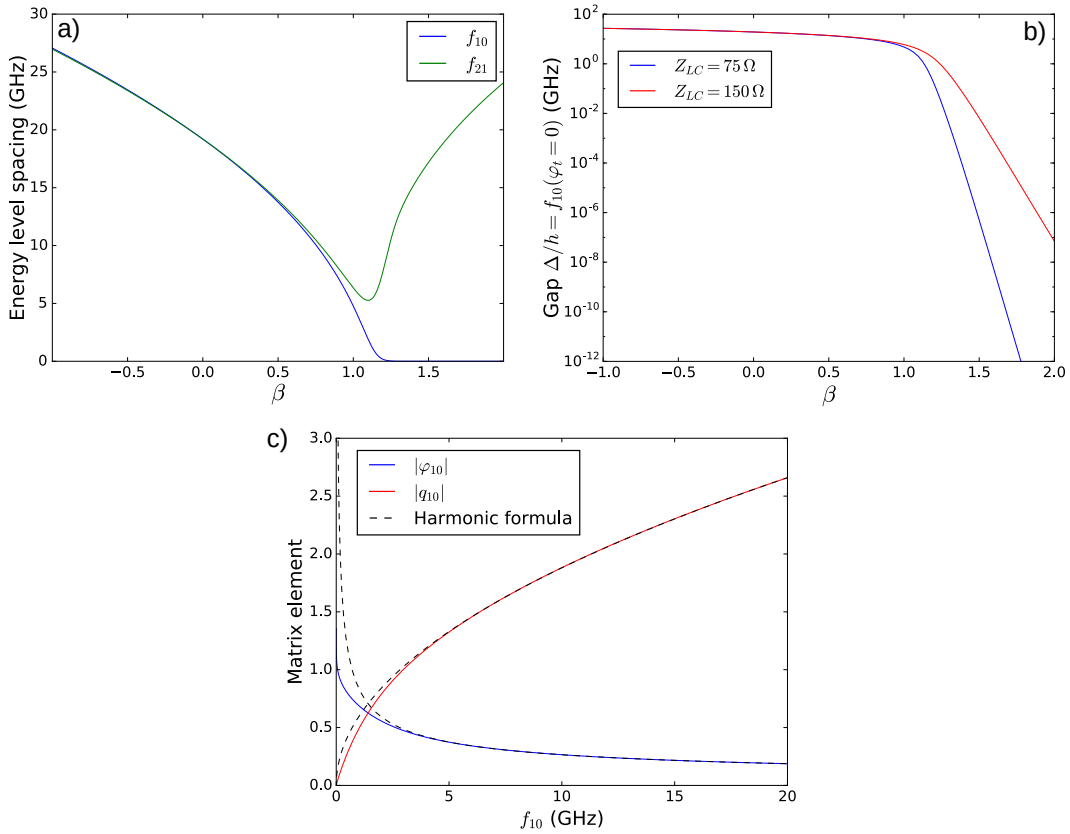


Figure 2.8: (a) Plot of fluxmon spectral frequencies f_{10} and f_{21} vs. β at zero tilt for typical design parameters $f_{LC} \approx 20\,\mathrm{GHz}$ and $Z_{LC} \approx 100\,\Omega$. (b) Logarithmic plot of the flux qubit gap Δ/h (i.e., f_{10} at zero tilt) vs. beta for two different values of impedance Z_{LC}. (c) Plot of the dimensionless flux and charge matrix elements at zero tilt obtained via numerical diagonalization of the 1D circuit model, along with the prediction using the analytical harmonic limit predictions (2.43) and (2.44), which break down for large qubit nonlinearity β (small Δ).

2.2.1.2 Harmonic regime $\beta \ll 1$

Some intuition can be gained about the fluxmon spectrum and matrix elements in the harmonic regime $|\beta| \ll 1$. Consider a symmetric potential with $\varphi_t = 0$. In this regime we can ignore all quartic and higher order terms in the potential. In this case the potential takes the quadratic form $U(\varphi) \approx E_L(1-\beta)\frac{\varphi^2}{2}$, yielding a β-tunable curvature with characteristic frequency of oscillation $\omega = \omega_{LC}\sqrt{1-\beta}$, where $\omega_{LC} = 2\pi f_{LC}$ is the LC oscillation frequency in the harmonic limit $\beta = 0$.

We can compare the numerically computed matrix elements from the previous section to the result obtained by treating the fluxmon as a harmonic oscillator with the same natural frequency $\omega = 2\pi f_{10}$ and mass (i.e., capacitance C). The harmonic formula for the flux and charge matrix elements can be found by identifying[13]

$$\begin{aligned} \hat{\Phi} &= \Phi_{\text{zpf}}(\hat{a}^\dagger + \hat{a}), \\ \hat{Q} &= iQ_{\text{zpf}}(\hat{a}^\dagger - \hat{a}), \end{aligned} \tag{2.40}$$

where

$$\Phi_{\text{zpf}} = \sqrt{\frac{\hbar}{2\omega C}} = \sqrt{\frac{\hbar Z}{2}}, \tag{2.41}$$

$$Q_{\text{zpf}} = \sqrt{\frac{\hbar\omega C}{2}} = \sqrt{\frac{\hbar}{2Z}}. \tag{2.42}$$

[13]This correspondence follows from direct comparison of the harmonic part of the Hamiltonian (2.31) to the well-studied Hamiltonian for a harmonic oscillator $H_{\text{H.O.}} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2$, where as is standard the ladder operators are $\hat{a}_\pm = \frac{1}{\sqrt{2\hbar m\omega}}(\mp i\hat{p} + m\omega\hat{x})$ [43].

Using the standard algebra of creation and annihilation operators we have the following useful matrix elements,

$$\left|\varphi_{10}^{\text{harmonic}}\right| = \frac{2\pi}{\Phi_0}\sqrt{\frac{\hbar}{2\omega C}} = \frac{2\pi}{\Phi_0}\Phi_{\text{zpf}}, \tag{2.43}$$

$$\left|q_{10}^{\text{harmonic}}\right| = \frac{1}{2e}\sqrt{\frac{\hbar\omega C}{2}} = \frac{1}{2e}Q_{\text{zpf}}. \tag{2.44}$$

These harmonic formulas are (perhaps surprisingly) quite accurate down to $f_{10} \sim 1\,\text{GHz}$ ($\beta \sim 1.15$) for a typical impedance of $\sim 100\,\Omega$. For typical fluxmon parameters there is only a 6 percent error at $f_{10} = 2\,\text{GHz}$, where the potential is very anharmonic [$\beta = 1.1$, $(f_{21} - f_{10})/f_{10} = 1.6$]. We will find this approximate harmonic behavior useful in Chapter 4 where we can use it to help us interpret dissipation from environmental flux noise above $\sim 1\,\text{GHz}$ in terms of a phenomenological power law model.

We can also compute the width of the ground state wavefunction in the harmonic limit $\beta = 0$, a useful quantity to know. Since the wavefunction is a Gaussian, the width is simple to compute: for a harmonic LC oscillator with frequency ω_{10} and capacitance C, we have

$$\langle 0|\varphi^2|0\rangle = |\langle 0|\varphi|1\rangle|^2 = 4\pi\frac{Z}{R_K}, \tag{2.45}$$

which when written in this dimensionless form is very easy to remember. Note the equality of the width of the wavefunction and the flux matrix element φ_{10} in the harmonic limit. For a typical fluxmon impedance Z_{LC} of $100\,\Omega$, we see that the width of the wavefunction is much less than 2π. This is consistent with the general property in

superconducting qubits that when $E_C \ll E_J, E_L$ the wavefunction is better described in terms of their flux because there is a low quantum uncertainty in the phase whereas there is a large uncertainty in the charge [44].

2.2.1.3 Two-level flux qubit approximation

It is often convenient to express the fluxmon Hamiltonian in the language of a conventional flux qubit Hamiltonian [45, 46]. Within the two-level approximation, a conventional flux qubit is usually described by the tunnel coupling Δ between wells and the energy detuning ε between the minima of the two wells. The energy detuning is usually taken to be linear in applied tilt flux, with the constant of proportionality being the expected persistent current I_p in each of the wells. To arrive at such a Hamiltonian from the 1D fluxmon Hamiltonian, we first make a two-level approximation by projecting the Hamiltonian $H(\beta, \Phi_t)$ (where $\beta(\Phi_{\mathrm{SQ}})$ and Φ_t are parameters) into the two lowest energy levels $\{|g(\beta,0)\rangle, |e(\beta,0)\rangle\}$ of $H(\beta, 0)$, the full Hamiltonian at zero tilt, and treat the applied tilt flux as a perturbation to first order,

$$H(\beta,\Phi_t)\Big|_{\{|g(\beta)\rangle,|e(\beta)\rangle\}} \approx H(\beta,0) + \Phi_t \frac{\partial H(\beta,0)}{\partial \Phi_t}\Big|_{\{|g(\beta)\rangle,|e(\beta)\rangle\}} \tag{2.46}$$

$$= H(\beta,0) + \Phi_t \frac{\Phi}{L}. \tag{2.47}$$

We can interpret the term $\frac{\Phi}{L}$ as the persistent current operator, which is multiplied by Φ_t to give the correct units of energy. The projection onto this two-level subspace should be a good approximation as long as the linear term in Φ_t is small compared to the transition

to the second excited state of the full Hamiltonian. The zero-tilt eigenstates $|g(\beta,0)\rangle$ and $|e(\beta,0)\rangle$ are states of even and odd parity. They are also the even and odd superpositions of the left and right well states, which have their fluxes localized in the left and right well of the double-well potential,

$$\begin{aligned} |g(\beta,0)\rangle &= \frac{1}{\sqrt{2}}(|L\rangle + |R\rangle), \\ |e(\beta,0)\rangle &= \frac{1}{\sqrt{2}}(|L\rangle - |R\rangle), \end{aligned} \tag{2.48}$$

where $|L\rangle$ and $|R\rangle$ are illustrated in Fig. 2.9(a). In this work we will refer to the basis $\{|g(\beta)\rangle, |e(\beta)\rangle\}$ (also written as $\{|0(\beta)\rangle, |1(\beta)\rangle\}$) as the parity basis, and the basis $\{|L\rangle, |R\rangle\}$ as the flux basis. Denoting the energy gap between $|g\rangle$ and $|e\rangle$ at zero tilt by Δ, then in terms of Pauli matrices the two-level flux qubit Hamiltonian (2.47) becomes

$$H \approx -\frac{1}{2}(\Delta\sigma_x + \varepsilon\sigma_z), \tag{2.49}$$

where

$$\varepsilon = 2I_p\Phi_t \tag{2.50}$$

is linearly proportional to the applied tilt flux with constant of proportionality equal to the persistent current

$$\begin{aligned} I_p &= \frac{1}{L}|\langle R|\Phi|R\rangle| \\ &= \frac{1}{L}|\langle g(\beta,0)|\Phi|e(\beta,0)\rangle|. \end{aligned} \tag{2.51}$$

Physically, Δ corresponds to the tunnel coupling between left and right flux states that lifts the ground state degeneracy between the otherwise degenerate states associated with the left and right wells. But the Hamiltonian (2.49) still can still be (approximately) used even when the potential in monostable. Mathematically we have

$$\Delta/h = \min_{\varphi_t} f_{10}, \tag{2.52}$$

or colloquially just "the gap" or "f_{10} at zero tilt." In the previous section we saw an approximate expression for Δ in the harmonic limit (i.e., for a slightly anharmonic oscillator), and in the next section we will see how to obtain an expression for it in the double-well limit of small Δ using the WKB approximation. The name "persistent current" for I_p comes from the fact that, as shown explicitly in the next section, it is equal to the expected current flowing through the qubit when the qubit is localized in one of the wells of a double-well potential. Even though this interpretation does not hold in the single-well regime, the quantity $I_p = \frac{1}{L}|\langle g(\beta,0)|\Phi|e(\beta,0)\rangle|$ is still often called the persistent current, and in particular does not vanish in the limit $\beta \to 0$.

An illustration of the meaning of ε, Δ and the states $|L\rangle$ and $|R\rangle$ is shown in Fig. 2.9(a), and the corresponding energy level diagram in 2.9(b). Under the two-level approximation (2.49), we have

$$f_{10} = \frac{1}{h}\sqrt{\Delta^2 + \varepsilon^2}, \tag{2.53}$$

meaning a hyperbolic relation between the system's energy gap and applied tilt flux.

In Fig. 2.10(a) we plot f_{10} vs. tilt bias for different values of Δ, comparing a full numerical diagonalization of (2.32) to the two-level flux qubit approximation (2.49), with

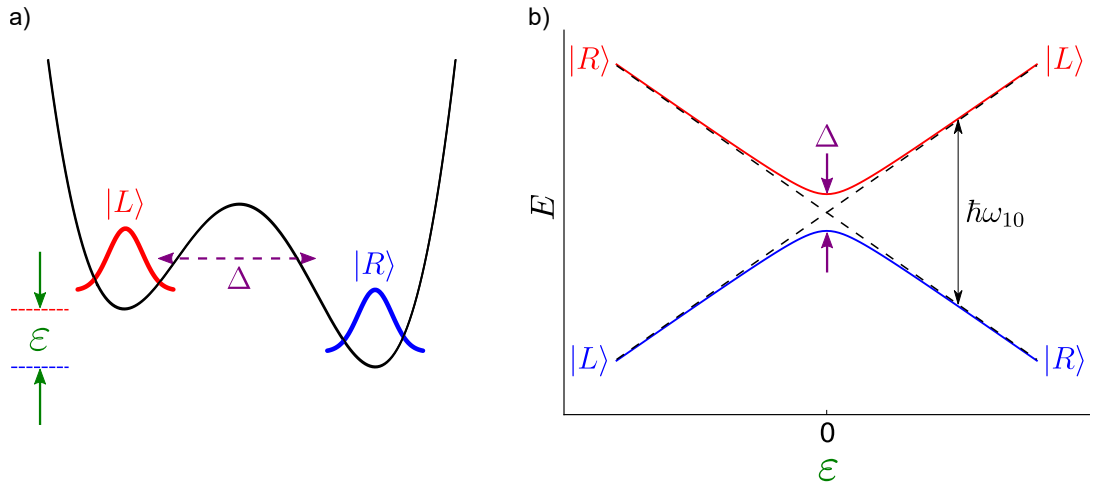


Figure 2.9: (a) Illustration of the meaning of ε, Δ, and $|L/R\rangle$ in the double-well regime of a flux qubit. Δ is the tunnel splitting that lifts the degeneracy at zero tilt, and ε is the tilt bias giving an energy detuning between the well minima. (b) Energy levels in two-level flux qubit approximation according to (2.49). Blue and red are the energies of the $|0\rangle$ and $|1\rangle$ energy eigenstates.

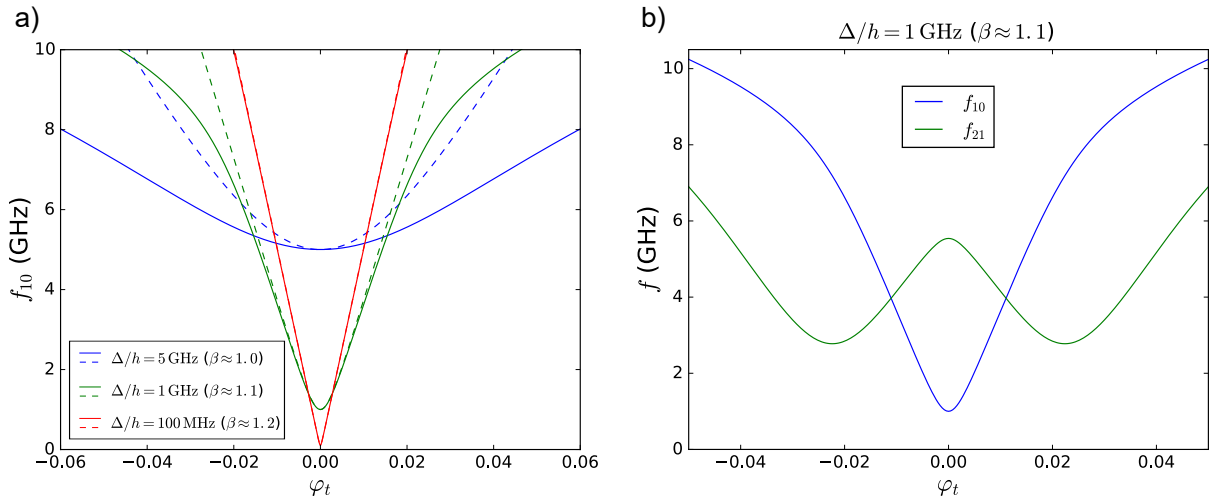


Figure 2.10: (a) Comparison of f_{10} vs. tilt bias between using numerical diagonalization of the 1D Hamiltonian vs. using the two-level flux qubit approximation of (2.49), using the same standard circuit parameters as in Fig. 2.7. Solid lines correspond to numerical diagonalization and dashed lines are equation (2.53). (b) Plot of f_{10} and f_{21} vs. tilt bias for the same value of $\Delta/h = 1$ GHz.

I_p computed numerically as $\frac{1}{L}|\langle 0|\Phi|1\rangle|$ at zero tilt, and Δ/h computed numerically as f_{10} at zero tilt. Note that the discrepancy between these curves doesn't necessarily imply that a two-level approximation itself is bad, but rather that using the zero-tilt basis with a tilted potential instead of the lowest two energy eigenstates of the tilted potential is bad. In fact, as shown in Fig. 2.10(b), even though the two-level flux qubit approximation predicts f_{10} accurately, there are certain tilt biases where $f_{21} = f_{10}$, which would be important if one were trying to isolate the $|0\rangle \rightarrow |1\rangle$ transition. We see that the approximation is better for larger beta (equivalently, smaller $\Delta/h = \min_{\varphi_t} f_{10}$) because it makes more sense to define two wells with a well-defined energy detuning between them.

2.2.1.4 Double-well limit $\beta > 1$

The conventional flux qubit picture is a qubit with a double-well potential, where the two qubit states are defined by the supercurrent flowing clockwise or counter-clockwise through a superconducting loop (or equivalently, the macrosopic flux pointing up or down) [45, 46]. This led to the original name "persistent current qubit" [47, 48].

An approximate expression for the persistent current can be found by noting that in the double-well limit, the wavefunction should be localized at the potential minima of the left and right wells. If we ignore the width of the $|L\rangle$ and $|R\rangle$ wavefunctions and treat them as delta functions, we would have $I_p = \frac{1}{L}\frac{\Phi_0}{2\pi}\varphi_m$, where φ_m is the value of φ that minimizes the potential, i.e., where $\partial V/\partial\varphi = 0$, leading to the transcendental equation $\varphi_m = \beta \sin\varphi_m$ (here we have assumed zero tilt bias consistent). One can obtain an expression for φ_m^2 to second order in $(\beta - 1)/\beta$ by expanding the sine to fifth order,

obtaining $\varphi_m^2 \approx 6[(\beta-1)/\beta] + (9/5)[(\beta-1)/\beta]^2$. Multiplying φ_m by $\frac{1}{L}\frac{\Phi_0}{2\pi}$ then gives an accurate expression for I_p in the double-well regime. This expression is accurate to $\sim 1\%$ at $\beta = 2$, but in general for better accuracy it should be computed numerically.

In the symmetric double-well limit ($\beta > 1$, $\varphi_t = 0$), the tunnel coupling Δ (f_{10} at zero tilt) can be approximately computed using the WKB approximation [43] by calculating the action S under the barrier. In this semi-classical approximation,[14] valid when the potential barrier is sufficiently wide and high, the wavefunction penetrating under the barrier decays as $e^{-S(\varphi)/\hbar}$, and the associated tunneling rate should roughly equal the "attempt frequency" to escape one well times the probability to be found "under the barrier". More explicitly, as derived in various textbooks on quantum mechanics [49], we expect to have a tunnel splitting of

$$\Delta \approx \frac{\hbar\omega_m}{\pi} \exp\left[-\frac{1}{\hbar}\int_{x_a}^{x_b} |p(x)|\, dx\right], \tag{2.54}$$

where ω_m is the frequency of oscillation at the bottom of one of the wells, $p(x) = \sqrt{2m[E - V(x)]}$ is the momentum (which is in this case imaginary), and the limits of integration are the classical turning points on either side of the potential barrier. In our system, we must replace x by Φ and m by C, so that $p(x)$ will then have units of charge and the integral units of action. This integral is not analytically solvable and doesn't necessarily provide much insight when numerically integrated. However,

[14] "Semi-classical" refers to de Broglie wavelength much shorter than the characteristic dimensions of the system, and formally means the logarithm of the wavefunction can be expanded to a low power in $\hbar$. This terminology is also applied to penetration under a barrier even though this is a purely quantum effect. Here the de Broglie "wavelength" is the length scale of exponential decay under the barrier. Conversely, it applies to bound states with high quantum number whose energies are much higher than the minimum of a potential well, for which the de Broglie wavelength is interpreted as usual.

we can gain an accurate and somewhat insightful WKB formula for Δ as follows: If β is not much greater than 1, we will have $\omega_m \approx \sqrt{2}\omega_{LC}\sqrt{\beta-1}$. To lowest order, the potential barrier height from the potential minimum to the potential maximum is $E_m = \frac{3}{2}E_L(\beta-1)^2/\beta$, and the characteristic width of the barrier between the minima is roughly $\Phi_m = \frac{\Phi_0}{2\pi}\varphi_m \approx \frac{\Phi_0}{2\pi}\sqrt{6(\beta-1)/\beta}$. The classical turning points are higher in energy than the potential minima by an amount equal to the zero point energy of the oscillator associated with the wells, $\frac{1}{2}\hbar\omega_m$. If we approximate the integrated action under the barrier as $\sqrt{2C\left(E_m - \frac{1}{2}\hbar\omega_m\right)}\Phi_m$, then plugging these approximations into (2.54) yields

$$\Delta(\beta) \approx a\frac{\sqrt{2}}{\pi}\hbar\omega_{LC}\sqrt{\beta-1}\exp\left[-b\sqrt{\frac{9}{4}\frac{E_L}{E_C}\frac{(\beta-1)^3}{\beta^2} - 3\sqrt{\frac{E_L}{E_C}\frac{(\beta-1)^3}{\beta^2}}}\right] \tag{2.55}$$

where a and b are dimensionless fit parameters to compensate for our approximate math, with best-fit $a = 4.6$ and $b = 1.4$. A comparison of the WKB result (2.55) with numerical diagonalization is shown in Fig. 2.11.

Some intuition can be gained from (2.55) by noting that the dominant term in the exponential is roughly equal to $E_m/\hbar\omega_t \approx \frac{3}{4\sqrt{2}}\sqrt{\frac{E_L}{E_C}}(\beta-1)^{3/2}/\beta$ (to lowest order), where $\omega_t \approx \omega_m/\sqrt{2}$ is the oscillation frequency of the inverted potential barrier. We can therefore interpret the argument of the exponential as the number of bound states that would be contained in the inverted potential well, n_{well}, giving the rule of thumb that the tunneling rate is equal to the well's oscillation frequency times $\exp[-n_{\text{well}}]$. This is no accident, as the WKB approximation is intimately tied together with the semi-classical Bohr-Sommerfeld quantization rule, which states that for eigenstates with high enough

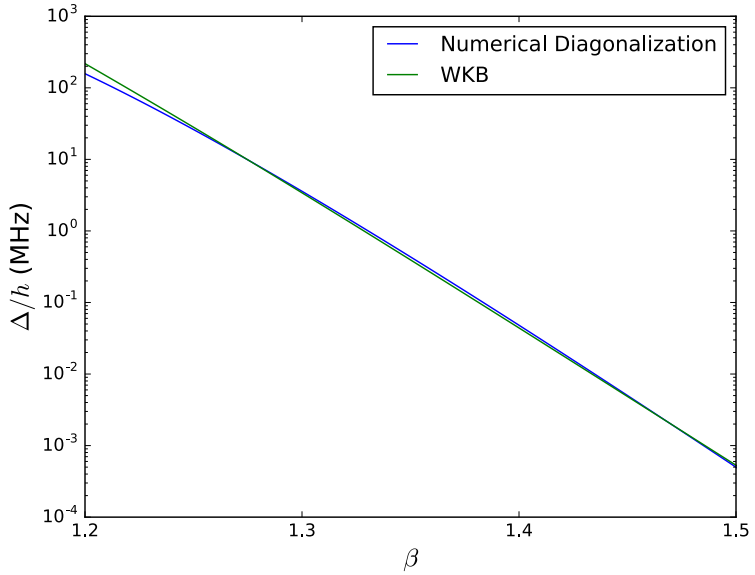


Figure 2.11: Comparison of WKB result (2.55) with numerical diagonalization.

quantum number in a potential well, we should have $\oint p(x)\,dx = 2\int_{x_a}^{x_b} p(x)\,dx = 2\pi n\hbar$, where the integral is taken over a full period of classical motion [49]. Intuitively, this states that the particle's wavefunction vanishes at the classical turning points and must have an integer number of nodes so that its phase constructively interferes over one full period of oscillation. For such wavefunctions, the quantum number n is the number of nodes in the wavefunction, so that the de Broglie wavelength is small for large n. Considering the maximum possible energy of a bound state within the inverted potential barrier, this condition implies that $n_{\text{well}} \approx \frac{2}{2\pi\hbar}\int_{x_a}^{x_b} \sqrt{2m[U_{\text{inverted}}(x) - E_{\min}]}\,dx$. Since this Bohr-Sommerfeld integral is almost equal to the integral in the WKB result (2.54) for an inverted potential barrier, the aforementioned rule of thumb naturally follows.

2.2.1.5 Incorporating junction asymmetry

So far we have assumed perfect symmetry between the two Josephson junctions within the DC SQUID, i.e., identical critical currents $I_{c1} = I_{c2}$. In reality, with available fabrication techniques, for the small (sub-micron) junctions desirable for qubits, there will be of order $1\% - 5\%$ variation in critical current between nominally identical junctions (but work is in progress to improve this number). With junction asymmetry, we must account for the energy in each Josephson junction separately and the 1D Hamiltonian (2.32) generalizes to

$$H_{\rm 1D}^{\rm asym.} = 8E_C\frac{q^2}{2} + E_L\frac{(\varphi-\varphi_t)^2}{2} - E_{J_1}\cos(\varphi+\varphi_{\rm SQ}/2) - E_{J_2}\cos(\varphi-\varphi_{\rm SQ}/2). \quad (2.56)$$

With some algebra, it can be shown that the contribution of the DC SQUID to the Hamiltonian [i.e., Eq. (2.24)], can be rewritten as

$$H_{\rm SQ} = -E_{J\Sigma}\cos\left(\pi\frac{\Phi_{\rm SQ}}{\Phi_0}\right)\sqrt{1+\chi^2\tan^2\left(\pi\frac{\Phi_{\rm SQ}}{\Phi_0}\right)}\cos(\varphi-\varphi_0), \quad (2.57)$$

where $\varphi_0 = \arctan(\chi\tan[\pi\Phi_{\rm SQ}/\Phi_0])$. We see that there are two effects of junction asymmetry. The first is to renormalize $E_J^{\rm max} = E_{J\Sigma}$ and also give it a slightly different functional form vs. $\Phi_{\rm SQ}$, but this is a rather small effect for the fluxmon qubit. In particular, Eq. (2.57) implies that the effective E_J cannot be tuned exactly to zero (this is known from flux-tunable transmon qubits, since it prevents the ability to tune the transmon f_{10} to arbitrary low values [44, 50]). For the fluxmon, this means that the qubit can never exactly reach the harmonic limit $\beta = 0$, although it can get quite close. In practice this is not an issue since annealing applications don't require the ability to

obtain this limit. The second effect is of much greater consequence in the fluxmon, which is that flux applied to the DC SQUID loop gets translated into an effective shift in tilt flux,

$$\Delta\Phi_t = \frac{\Phi_0}{2\pi}\tan^{-1}(\chi\tan[\pi\Phi_{\mathrm{SQ}}/\Phi_0]). \tag{2.58}$$

This effect can be quite large even for small junction asymmetries, and is important for the calibration of the 'zero tilt' bias vs. Δ in real devices, as will be discussed in the next chapter. Note that the potential energy is conveniently still one-dimensional even with asymmetric junctions, due to flux quantization.

2.2.2 Modeling the fluxmon at high frequencies: full transmission line analysis

As mentioned earlier in this chapter in the context of Eq. (2.28), at frequencies that are not sufficiently low, the impedance of the CPW deviates from that given by the low frequency lumped parallel LC model. A long fluxmon is desirable from the standpoint of qubit connectivity for quantum annealing, but this is at odds with the desired simple lumped element model (among other design considerations) since f_{LC} decreases with CPW length. This could cause the 1D model to break down for two reasons. Firstly, as already mentioned, the expression for impedance vs. frequency may deviate from that of the low frequency LC model (2.30). Secondly, even if the operating frequencies are low enough so that (2.30) is an accurate expression for $Y(\omega)$ at the qubit frequency, the 1D LC model still ignores the possibility of SQUID-mediated coupling to higher modes of

the transmission line segment. Because the ability to predict and model qubit properties over a wide operating regime is crucial for annealing, it is important to quantify these results in order to determine the acceptable range of fluxmon design parameters that supports a scalable circuit analysis. In particular, we must be able to bound the errors in the qubit spectrum and matrix elements incurred by using a lumped element as opposed to full transmission line model. We will also see whether or not it is possible to use a lumped element model but with a slightly modified L and C to reduce these errors to acceptable levels.

To solve the quantum transmission line model, we will use a strategy similar to that used to model a common implementation of Josephson parametric amplifiers [51, 52, 53], which is an identical circuit apart from the CPW being open instead of shorted at $x = 0$. We need to go beyond these previous treatments by computing the system's spectrum via a full quantum Hamiltonian and also introduce a DC flux bias for the CPW. The idea will be to break down the transmission line into a ladder of infinitesimally small inductors and capacitors, as shown in Fig. 2.2c, and then later take the continuum limit. Afterwards we will break the Hamiltonian up into a linear plus a nonlinear part, allowing us to define a mode decomposition for the system with which to express the Hamiltonian as a sum of infinitely many nonlinearly coupled harmonic oscillators, a form which can be numerically diagonalized.

We first need to obtain the classical Lagrangian of our circuit. We begin by breaking the CPW segment of length ℓ into N LC circuits of length $\Delta x = \ell/N$, as illustrated

in Fig. 2.2(c). As is standard, at each node n at position $x_n = n\Delta x$ (i.e., between the n^{th} inductor and capacitor), we define the flux variable $\Phi_n(t) = \int_{-\infty}^{t} V(x_n, t')dt'$. By definition, the voltage at the n^{th} capacitor node will then be $V(x_n, t) = \dot{\Phi}_n(t)$. By Faraday's law the current through the n^{th} inductor satisfies $\frac{dI}{dt} = -\frac{1}{\mathcal{L}\Delta x}(\dot{\Phi}_n(t) - \dot{\Phi}_{n-1}(t))$, which when integrated with respect to time yields the current across the n^{th} inductor as $\frac{1}{\mathcal{L}\Delta x}(\Phi_{n-1}(t) - \Phi_n(t))$. Now using the electric and magnetic energy stored in the capacitors and inductors we can write down the discretized classical Lagrangian for the bulk of the CPW segment,

$$L_{\text{CPW}}^{\text{disc.}} = \left(\frac{\Phi_0}{2\pi}\right)^2 \sum_{n=1}^{N} \left[\frac{1}{2}\mathcal{C}\Delta x\dot{\varphi}_n^2(t) - \frac{1}{2\mathcal{L}\Delta x}\left(\varphi_n(t) - \varphi_{n-1}(t)\right)^2\right], \tag{2.59}$$

where we have defined the dimensionless phase variables φ_n related to flux through the relation $\varphi_n = \frac{2\pi}{\Phi_0}\Phi_n$ as usual and Φ_0 is the flux quantum.

Next we must consider the end of the CPW containing the SQUID at $x = \ell$, which we treat as a lumped element (this reasonably assumes that the SQUID's physical extent is much smaller than the wavelength associated with CPW wave propagation at the frequencies under consideration). The contribution from the SQUID contains the Josephson potential energy $-E_J(\Phi_{\text{SQ}})\cos\varphi_s$, where $E_J = 2E_{J_{1,2}}\cos(\pi\Phi_{\text{SQ}}/\Phi_0)$ is the effective Josephson energy of the DC SQUID (for now the junctions are still assumed identical) and $\varphi_s = (\varphi_{s,1} + \varphi_{s,2})/2$ is the effective Josephson phase across the DC SQUID. If we account for the capacitance of the junctions (but still ignore the DC SQUID inline inductances), there is also a kinetic term of the form $\left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2}C_J\dot{\varphi}_s^2$, where $C_J = 2C_{j_{1,2}}$ is the total

capacitance within the DC SQUID.[15] Finally, there is also one remaining CPW inductor between node N and the SQUID, contributing the term $-\left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2\mathcal{L}\Delta x}\left(\varphi_s(t)-\varphi_N(t)\right)^2$. Finally, since the CPW plus SQUID forms a single closed inductive loop, we can allow for an external flux bias through this loop, such that by flux quantization we have for the flux across the effective junction

$$\varphi_s \to \varphi_J = \varphi_s + \varphi_x. \tag{2.60}$$

Putting all the terms together, we obtain the full discretized Lagrangian,

$$\begin{aligned} L_{\text{full}}^{\text{disc.}} = &\left(\frac{\Phi_0}{2\pi}\right)^2 \sum_{n=1}^{N}\left[\frac{1}{2}\mathcal{C}\Delta x\dot{\varphi}_n^2(t) - \frac{1}{2\mathcal{L}\Delta x}\left(\varphi_n(t)-\varphi_{n-1}(t)\right)^2\right] \\ &- \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2\mathcal{L}\Delta x}\left(\varphi_s(t)-\varphi_N(t)\right)^2 + \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2}C_J\dot{\varphi}_s^2 + E_J(\Phi_{\text{SQ}})\cos(\varphi_s+\varphi_x), \end{aligned} \tag{2.61}$$

Taking the continuum limit, the full system Lagrangian is then

$$L_{\text{full}} = \left(\frac{\Phi_0}{2\pi}\right)^2 \int_0^\ell dx\ \left[\frac{1}{2}\mathcal{C}\dot{\varphi}^2 - \frac{1}{2\mathcal{L}}\varphi'^2\right] + \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2}C_J\dot{\varphi}_s^2 + E_J(\Phi_{\text{SQ}})\cos(\varphi_s+\varphi_x). \tag{2.62}$$

In the continuum limit, the phase field in the bulk CPW segment will obey the wave equation, as is usual for lossless transmission lines [41]. To see this formally from the

[15] More precisely, we actually have the sum $\left(\frac{\Phi_0}{2\pi}\right)^2 \sum_{i=1,2} \frac{1}{2}\left(\frac{C_J}{2}\right)^2 \dot{\varphi}_{s,i}^2$. To equate this with $\left(\frac{\Phi_0}{2\pi}\right)^2 \frac{C_J}{2}\dot{\varphi}_s^2 = \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{C_J}{2}(\dot{\varphi}_{s,1}+\dot{\varphi}_{s,2})^2/4$, we use the fact that under a static DC SQUID flux bias Φ_{SQ}, through flux quantization we have the constraint $\dot{\varphi}_{s,1}-\dot{\varphi}_{s,2}=0$, allowing us to make the substitution $\dot{\varphi}_s^2 = (\dot{\varphi}_{s,1}^2+\dot{\varphi}_{s,2}^2)/2$.

Lagrangian, we can write down the classical bulk Hamiltonian

$$\begin{aligned}\mathcal{H}_{\mathrm{CPW}}^{\mathrm{disc}} &= \sum_n \dot{\varphi}_n \frac{\partial L_{\mathrm{CPW}}}{\partial \dot{\varphi}_n} - L_{\mathrm{CPW}} \\ &= \left(\frac{\Phi_0}{2\pi}\right)^2 \sum_n \left[\frac{1}{2}\mathcal{C}\Delta x \dot{\varphi}_n^2(t) + \frac{1}{2\mathcal{L}\Delta x}\left(\varphi_n(t) + \varphi_{n-1}(t)\right)^2\right]. \end{aligned} \tag{2.63}$$

Introducing the dimensionless charge on the n^{th} capacitor, $q_n = \frac{\partial L}{\partial \dot{\varphi}_n} = \mathcal{C}\Delta x \dot{\varphi}_n(t)$, as the conjugate momentum of the phase, Hamilton's equations of motion $\dot{q}_n = -\frac{d\mathcal{H}}{d\varphi_n}$ then give $\mathcal{C}\Delta x \ddot{\varphi}_n = \frac{\varphi_{n+1} - 2\varphi_n + \varphi_{n+1}}{\mathcal{L}\Delta x}$. Dividing both sides by $\mathcal{C}\Delta x$ and recognizing the right hand side as a difference quotient for the second spatial derivative, we obtain in the continuum limit of large N (i.e., $\Delta x \to dx$)

$$\frac{\partial^2 \varphi(x,t)}{\partial t^2} = v^2 \frac{\partial^2 \varphi(x,t)}{\partial x^2}, \tag{2.64}$$

where $v = 1/\sqrt{\mathcal{L}\mathcal{C}}$ is the speed of wave propagation as we stated earlier without proof in Eq. (2.28). For later reference we also write down the classical Hamiltonian in the continuum limit from (2.62),

$$\mathcal{H}_{\mathrm{full}} = \left(\frac{\Phi_0}{2\pi}\right)^2 \int_0^\ell dx \, \left[\frac{1}{2}\mathcal{C}\dot{\varphi}^2 + \frac{1}{2\mathcal{L}}\varphi'^2\right] + \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2} C_J \dot{\varphi}_s^2 - E_J(\Phi_{\mathrm{SQ}}) \cos(\varphi_s + \varphi_x). \tag{2.65}$$

Before we proceed to quantizing the system, we need to perform a mode decomposition of the possible solutions to this wave equation that obey the boundary conditions at $x = 0$ and $x = \ell$. By definition, these modes must be harmonic, and so will not contain any of the nonlinearities of the DC SQUID. We therefore include the *linear* part of the DC

SQUID's response in the boundary condition at $x = \ell$. Unfortunately, as we'll see, this approach only works for $\beta < 1$, because the mode frequencies become imaginary when the net inductance of the system becomes negative. Presumably there is a more clever approach that works for $\beta > 1$.

2.2.2.1 Diagonalization of transmission line model using mode decomposition (for $\beta < 1$)

Here we will attempt to diagonalize the full transmission line model Hamiltonian (2.65) using a mode decomposition. There will be considerable algebra involved, but it will be worth it at the end because we will arrive at a semi-intuitive picture of a frequency-dependent inductance and capacitance for the CPW that wouldn't be obvious from standard transmission line theory. The boundary condition at $x = 0$ dictates that the voltage ($\propto \dot{\varphi}$) must vanish there. This means that the eigenfunctions of our harmonic basis must take the form (in phasor notation) $\varphi(x,t) \propto e^{i\omega_n t} \sin(k_n x)$, with $\omega_n = v k_n$ as dictated by the wave equation. Keeping the quadratic part of the cosine term in the discrete Lagrangian (2.61), the Euler-Lagrange equation for the DC SQUID phase, $\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{\varphi}_s}\right] - \frac{\partial L}{\partial \varphi_s} = 0$, becomes $\left(\frac{\Phi_0}{2\pi}\right)^2 C_s \ddot{\varphi}_s + E_J(\Phi_{\mathrm{SQ}})\varphi_s + \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{\varphi_s - \varphi_N}{\mathcal{L}\Delta x} = 0$. Recognizing the last term as the difference quotient for the spatial derivative of $\varphi(x,t)$ at $x = \ell$ and identifying

$$\varphi(\ell, t) = \varphi_s(t) \tag{2.66}$$

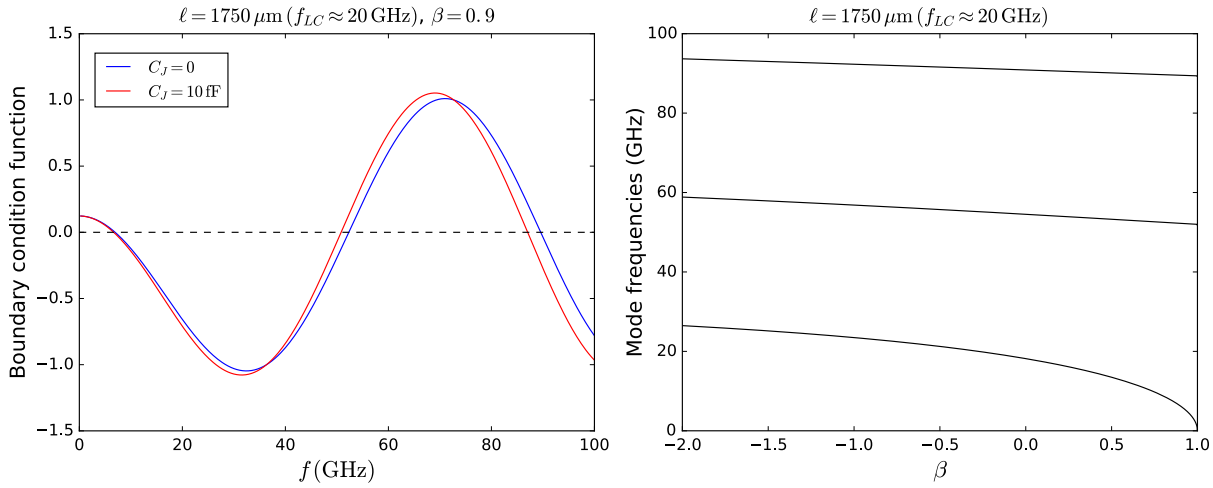


Figure 2.12: (a) Visual solution of boundary condition function (2.68) at $\beta = 0.9$ for two different values of C_J for a $f_{LC} = 20$ GHz fluxmon. (b) Mode frequencies vs. β for a $f_{LC} = 20$ GHz fluxmon. Note that the first mode frequency vanishes as $\beta \to 1$ from below.

in the continuum limit, we arrive at the classical boundary condition

$$\left(\frac{\Phi_0}{2\pi}\right)^2 C_J \ddot{\varphi}(\ell, t) + E_J(\Phi_{\text{SQ}})\varphi(\ell, t) + \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{\mathcal{L}} \varphi'(\ell, t) = 0. \tag{2.67}$$

Inserting the ansatz eigenfunction then yields a transcendental equation for $\omega_n = v k_n$, which we divide through by $\left(\frac{\Phi_0}{2\pi}\right)^2 \frac{k_n}{\mathcal{L}}$ to make it dimensionless:

$$-\mathcal{L} C_J v \omega_n \sin\left(\omega_n \ell / v\right) + \frac{v}{\omega_n} \left(\frac{2\pi}{\Phi_0}\right)^2 \mathcal{L} E_J(\Phi_{\text{SQ}}) \sin\left(\omega_n \ell / v\right) + \cos\left(\omega_n \ell / v\right) = 0. \tag{2.68}$$

We can check the limiting case of infinite Josephson inductance ($E_J \to 0$, corresponding to an open boundary) and zero junction capacitance, which yields $\cos(k_n \ell) = 0$, defining the harmonics of a quarter-wave resonator, as would be expected. The presence of the linear part of the Josephson inductance however causes the spacing between mode frequencies to become nonuniform. Fig. 2.12(a) visually shows the solution to this equation for a $f_{LC} = 20$ GHz qubit for $C_J = 0$ and for $C_J = 10\,\text{fF}$.

Now that the modes and their frequencies are defined, we decompose the phase field in the basis comprised of them,

$$\varphi(x,t) = \frac{2\pi}{\Phi_0}\sqrt{\frac{2}{\mathcal{C}\ell}}\sum_n q_n(t)\sin k_n x, \tag{2.69}$$

where $q_n(t)$ are the time-dependent mode amplitudes that will later be quantized. Because the spatial eigenfunctions $\sin(k_n x)$ are not necessarily orthogonal, what follows involves a bit of algebra, some steps of which we omit for brevity. First, we substitute the decomposition (2.69) into the full system Lagrangian (2.62) and use the boundary condition (2.68) to write the following overlap integrals as

$$\int_0^\ell dx\ \sin k_n x \sin k_m x = \frac{\ell}{2}M_n\delta_{nm} - \mathcal{L}C_J v^2 \sin k_n\ell \sin k_m\ell, \tag{2.70a}$$

$$\int_0^\ell dx\ k_n k_m \cos k_n x \cos k_m x = \frac{\ell}{2}k_n^2 M_n\delta_{nm} - \mathcal{L}\left(\frac{2\pi}{\Phi_0}\right)^2 E_J(\Phi_{\rm SQ})\sin k_n\ell \sin k_m\ell, \tag{2.70b}$$

where

$$M_n \equiv 1 + \frac{1}{\ell}\left[\frac{1}{k_n^2}\left(\frac{2\pi}{\Phi_0}\right)^2 \mathcal{L}E_J(\Phi_{\rm SQ}) + \mathcal{L}C_J v^2\right]\sin^2(k_n\ell). \tag{2.71}$$

The reason for introducing the variables M_n will become clear soon. Next, we divide the boundary contribution of the DC SQUID Lagrangian into a linear plus a nonlinear part $L_s = L_s^{\rm lin.} + L_s^{\rm nonlin.}$,

$$L_s^{\rm lin.} = \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{1}{2}C_J\dot{\varphi}_s^2 - E_J(\Phi_{\rm SQ})\frac{\varphi_s^2}{2}, \tag{2.72a}$$

$$L_s^{\rm nonlin.} = E_J(\Phi_{\rm SQ})\left[\cos\varphi_s + \frac{\varphi_s^2}{2}\right], \tag{2.72b}$$

so that the system Lagrangian becomes

$$
\begin{aligned}
L_{\text{full}} &= L_{\text{CPW}} + L_s^{\text{lin.}} + L_s^{\text{nonlin.}} \\
&= \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{\mathcal{C}}{2} \int_0^\ell dx \, \left(\frac{2\pi}{\Phi_0}\right)^2 \frac{2}{\mathcal{C}\ell} \left[\sum_{n,m} \dot{q}_n \dot{q}_m \sin k_n x \sin k_m x\right] \\
&\quad - \left(\frac{\Phi_0}{2\pi}\right)^2 \frac{\mathcal{C}}{2} \int_0^\ell dx \, \left(\frac{2\pi}{\Phi_0}\right)^2 \frac{2}{\mathcal{C}\ell} v^2 \left[\sum_{n,m} q_n q_m k_n k_m \cos k_n x \cos k_m x\right] \\
&\quad + \frac{\mathcal{L}}{\ell} \sum_{n,m} \left[-C_J v^2 \sin k_n \ell \sin k_m \ell \, \dot{q}_n \dot{q}_m + v^2 \left(\frac{2\pi}{\Phi_0}\right)^2 E_J(\Phi_{\text{SQ}}) \sin k_n \ell \sin k_m \ell \, q_n q_m\right] \\
&\quad + \frac{C_J}{\mathcal{C}\ell} \sum_{n,m} \sin k_n \ell \sin k_m \ell \, \dot{q}_n \dot{q}_m - E_J(\Phi_{\text{SQ}}) \left(\frac{2\pi}{\Phi_0}\right)^2 \frac{1}{\mathcal{C}\ell} \sum_{n,m} \sin k_n \ell \sin k_m \ell \, q_n q_m + L_s^{\text{nonlin.}} \\
&= \frac{1}{2} \sum_n \left[M_n \dot{q}_n^2 - \omega_n^2 M_n q_n^2\right] + E_J(\Phi_{\text{SQ}}) \left[\cos \varphi_s + \frac{\varphi_s^2}{2}\right]. \qquad (2.73)
\end{aligned}
$$

The cancellation of off-diagonal terms between the bulk and SQUID contributions was no accident – it is a consequence of including the linear part of the SQUID response in the boundary condition defining our harmonic mode basis.

We now have a discrete mode decomposition of the otherwise continuous classical Hamiltonian,

$$
\mathcal{H} = \frac{1}{2} \sum_n \left[M_n \dot{q}_n^2 + \omega_n^2 M_n q_n^2\right] - E_J(\Phi_{\text{SQ}}) \left[\cos \varphi_s + \frac{\varphi_s^2}{2}\right], \qquad (2.74)
$$

with φ_s given by Eq. (2.69) with $x = \ell$. The system can therefore be viewed as a collection of harmonic oscillators that are coupled by the nonlinear part of the SQUID's potential. We can interpret M_n as the dimensionless masses of these oscillators. Noting that q_n and $M_n \dot{q}_n \equiv p_n$ are canonically conjugate coordinates, and defining the creation

and annihilation operators through the standard relations

$$\hat{q}_n = \sqrt{\frac{\hbar}{2M_n\omega_n}}(\hat{a}_n^\dagger + \hat{a}_n), \tag{2.75a}$$

$$\hat{p}_n = i\sqrt{\frac{\hbar M_n\omega_n}{2}}(\hat{a}_n^\dagger - \hat{a}_n), \tag{2.75b}$$

we can canonically quantize the system into a set of nonlinearly coupled quantum harmonic oscillators:

$$\hat{H} = \sum_n \hbar\omega_n \left(\hat{a}_n^\dagger \hat{a}_n + \frac{1}{2}\right) + V\left(\frac{2\pi}{\Phi_0}\sqrt{\frac{2}{\mathcal{C}\ell}}\sum_n \sqrt{\frac{\hbar}{2M_n\omega_n}}\left(\hat{a}_n^\dagger + \hat{a}_n\right)\sin k_n\ell + \varphi_x\right), \tag{2.76}$$

with

$$V(\varphi) \equiv E_J(\Phi_{\mathrm{SQ}})\left[\cos\varphi + \frac{1}{2}\varphi^2\right]. \tag{2.77}$$

This is the final result for the fluxmon Hamiltonian within the full transmission line model.

Before diagonalizing the Hamiltonian (2.76), we can try and gain some intuition behind these modes. Firstly, we can define an effective inductance and capacitance for each mode by equating the harmonic contribution to the 1D lumped element Hamiltonian $\frac{1}{2L_n}\Phi_s^2 + \frac{1}{2C_n}Q_s^2$, leading to

$$L_n = \frac{2}{\mathcal{C}\ell}\frac{\sin^2(k_n\ell)}{\omega_n^2 M_n}, \qquad C_n = \frac{\mathcal{C}\ell}{2}\frac{M_n}{\sin^2(k_n\ell)}. \tag{2.78}$$

It will be illustrative to isolate the CPW's contribution to L_n and C_n, so we subtract out the known linear inductance and capacitance of the SQUID to define

$$L_n^{\mathrm{CPW}} = \left(1/L_n - 1/L_J\right)^{-1}, \qquad C_n^{\mathrm{CPW}} = C_n - C_J. \tag{2.79}$$

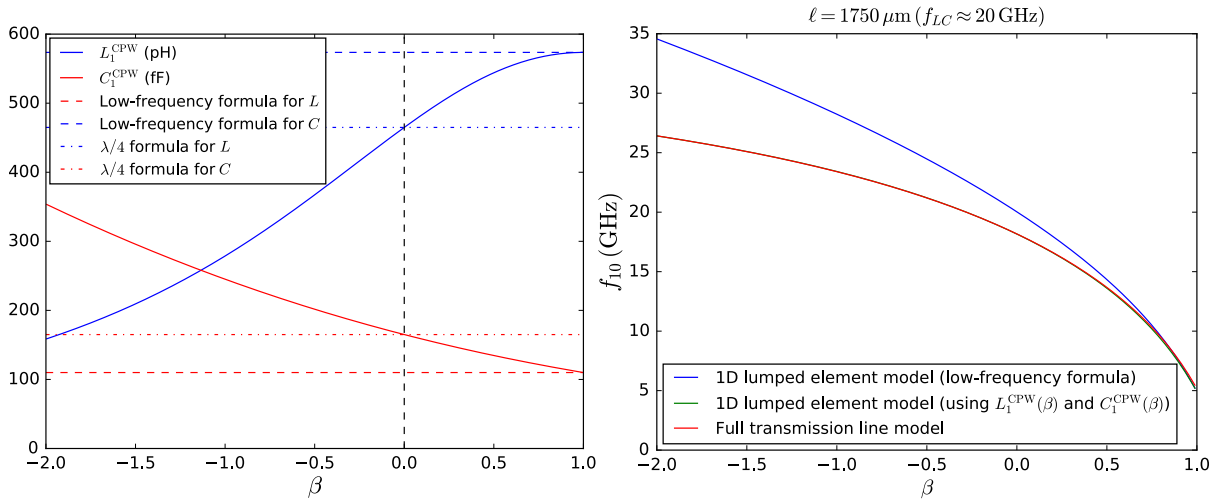


Figure 2.13: a) Effective CPW inductance and capacitance L_1^{CPW} and C_1^{CPW} for the first mode, vs. β. In the low-frequency limit these tend to the low-frequency lumped values in Eq. (2.30). At the harmonic limit $\beta = 0$, they agree with the standard $\lambda/4$ resonator values from [41]. b) f_{10} at zero tilt versus $\beta \equiv$ computed for three different models: 1D lumped element model (low-frequency formula) (blue), a hybrid 1D lumped element model (green) where L and C are given by $L_1^{\mathrm{CPW}(\beta)}$ and $C_1^{\mathrm{CPW}(\beta)}$ instead of the low-frequency formulas $\mathcal{L}\ell$ and $\mathcal{C}\ell/3$, and the full transmission line model (red).

The parameters L_n^{CPW} and C_n^{CPW} in a sense constitute the closest approximation of the transmission line to a lumped parallel LC circuit as a function of frequency. In Fig. 2.13a, we plot the effective CPW inductance and capacitance of the first mode vs. $\beta(\Phi_{\mathrm{SQ}})$, for a standard fluxmon with $f_{LC} = 20\,\mathrm{GHz}$ and $\beta_{\max} \approx 2$. In the limit of small frequencies $\beta \to 1$ (remember, we can't go above $\beta = 1$ with this method), the effective CPW inductance and capacitance approach the low-frequency "fluxmon" limit given in Eq. (2.30), $L_1^{\mathrm{CPW}} \to \mathcal{L}\ell$, $C_1^{\mathrm{CPW}} \to \mathcal{C}\ell/3$, as shown in the figure. In the harmonic limit $\beta = 0$ ($\Phi_{\mathrm{SQ}} = 0.5$) we should expect the system to behave as a $\lambda/4$ resonator, which according to standard transmission line theory [41] will be a lumped LC oscillator with $L = (8/\pi^2)\mathcal{L}\ell$ and $C = \mathcal{C}\ell/2$. This agrees with the expression for L_1^{CPW} and C_1^{CPW} at $\beta = 0$, as can be seen in the figure.

Fig. 2.13b shows f_{10} vs. $\beta(\Phi_{\mathrm{SQ}})$ at zero tilt from numerical diagonalization of the full transmission line model Hamiltonian (2.76). The numerical diagonalization is performed by writing the Hamiltonian in the basis of harmonic oscillator states for each mode, keeping 50 or so levels for the lowest mode and a few levels for higher modes, which do not contribute as much (if β is not too close to 1, one can of course afford to use substantially less than 50 levels for the first mode). For negative β, the low-frequency fluxmon model is significantly off. At $\beta = 0$ the transmission line model correctly predicts $f_{10} = f_{\lambda/4}$. As f_{10} gets lower, the low-frequency 1D fluxmon model becomes better and better. It is not clear what the exact discrepancy is in the region we care about since we can't go above $\beta = 1$ with the treatment described above.

We also plot the result of diagonalizing a modified 1D fluxmon using $L = L_1^{\mathrm{CPW}}(\beta)$ and $C = C_1^{\mathrm{CPW}}(\beta)$, the result of which is quite close to the full transmission line model for all values of β in the plot. This simply means that it is a very good approximation to only consider the first mode of the coupled mode system, and that for a given β we can treat the CPW as a lumped parallel LC circuit with an L and C given by $L = L_1^{\mathrm{CPW}}(\beta)$ and $C = C_1^{\mathrm{CPW}}(\beta)$.

A better transmission line model that works for $\beta > 1$ is work under progress. However, at low frequencies where we operate the fluxmon, the lumped element model for the CPW is a good enough approximation, although it tells us that we need to be careful about reducing f_{LC} too much below 20 GHz. From now on we assume the lumped-element model for the CPW, and consider other important modifications to the 1D

fluxmon model.

2.2.3 Incorporating inductance and capacitance within the DC SQUID

A more refined model for the fluxmon is one that takes into account the inline inductance and capacitance present in any physical implementation of a DC SQUID. For example, for typical designs the skinny wires comprising the DC SQUID are of length $\sim 10\,\mu$m and will contribute of order 10 pH of inductance, and the Josephson junctions will have a capacitance across them of order 1 - 10 fF. We need to quantify how the modified circuit changes the fluxmon spectrum and matrix elements. We outline two methods to do so, based on work done jointly with Dvir Kafri. One takes into account the inline inductances and capacitances exactly, making the Hamiltonian three-dimensional. The other, more practical approach will be to transform the 3D Hamitonian into an approximate 1D Hamiltonian by absorbing the junction capacitance into the main capacitance and treating the inline inductances as a modification of the $\cos\varphi$ term. This is also a more practical approach because it is much more efficient to compute, and furthermore adds only half as many extra fit parameters to the model when it is used to fit experimental data compared to the 3D model. Either of these two approaches allows us to model the fluxmon spectrum to within 10 MHz of accuracy throughout the regime of operation, which happens to roughly be the experimental uncertainty in the spectrum. Thus, assuming these models can also precisely predict other relevant quantities (such as the

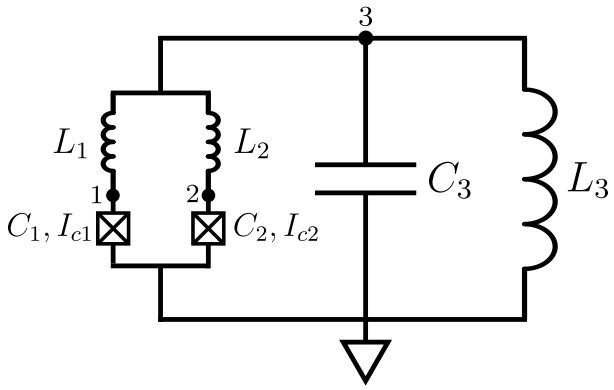


Figure 2.14: 3D fluxmon circuit model including inline inductances and capacitances within the DC SQUID. It is customary to draw a box around a Josephson junction circuit element (i.e., ⊠ instead of just ×) to indicate the inclusion of parallel junction capacitance. Three nodes are labeled 1, 2, and 3, with the bottom of the circuit being ground.

persistent current matrix element, discussed below), they should be sufficient for use in quantum annealing.

2.2.3.1 Full 3D Model

We account for the DC SQUID inline inductances and capacitances using the circuit model shown in Fig. 2.14. There are three degrees of freedom in this circuit. Note that we are free to include all inline linear inductances within the DC SQUID as drawn above the junctions, even if there is inductance on either side of the junctions. Because the current conservation equations are symmetric under interchange of the inductors and the junctions, it doesn't matter if the inductance is drawn above or below the junction. Furthermore, if there is inductance on both sides, we can choose to lump them together on one side of the junction. This is because the current flowing through the branch above the junction must be equal to the current flowing below the junction, and so feels a net inline lumped inductance equal to the sum of the inductances above and below the

junction.

Before writing down the 3D circuit Hamiltonian, we must define our node fluxes and flux biases. Let the three degrees of freedom be the fluxes Φ_1, Φ_2, Φ_3 at the nodes labeled 1, 2, 3 respectively. Our convention for currents in all branches is that all currents flow towards ground. We then define two external flux biases for the two loops going through L_3 and either of the two branches of the DC SQUID as Φ_{1x} and Φ_{2x}. Explicitly, we have

$$\Phi_{1x} = \text{Flux through loop } [3 \xrightarrow{L_3} \text{ ground } \xrightarrow{J_1} 1 \xrightarrow{L_1} 3], \tag{2.80}$$

$$\Phi_{2x} = \text{Flux through loop } [3 \xrightarrow{L_3} \text{ ground } \xrightarrow{J_2} 2 \xrightarrow{L_2} 3] \tag{2.81}$$

These external loop fluxes are related to what we usually think of as the SQUID and tilt flux biases according to

$$\Phi_t = \frac{\Phi_{1x} + \Phi_{2x}}{2}, \quad \Phi_{\text{SQ}} = \Phi_{1x} - \Phi_{2x}. \tag{2.82}$$

Flux quantization requires that $\Phi_{1x} = \Phi_{L_1} + \Phi_1 - \Phi_3$ and $\Phi_{2x} = \Phi_{L_2} + \Phi_2 - \Phi_3$, from which we obtain the fluxes through each of the linear inductors,

$$\Phi_{L_1} = \Phi_3 - \Phi_1 + \Phi_t + \Phi_{\text{SQ}}/2, \tag{2.83}$$

$$\Phi_{L_2} = \Phi_3 - \Phi_2 + \Phi_t - \Phi_{\text{SQ}}/2, \tag{2.84}$$

$$\Phi_{L_3} = \Phi_3. \tag{2.85}$$

One can easily check that these flux quantization conditions reduce to the usual condition

without inline capacitors by setting $\Phi_{L_1} = \Phi_{L_2} = 0$ and subtracting the second equation from the first above.

By inspection we can immediately write down the system Lagrangian as

$$\begin{aligned} L_{3\mathrm{D}} =& \frac{1}{2}C_1\dot{\Phi}_1^2 + \frac{1}{2}C_2\dot{\Phi}_2^2 + \frac{1}{2}C_3\dot{\Phi}_3^2 \\ & - \frac{1}{2L_1}(\Phi_1 - \Phi_3 - \Phi_t - \Phi_{\mathrm{SQ}}/2)^2 - \frac{1}{2L_2}(\Phi_2 - \Phi_3 - \Phi_t + \Phi_{\mathrm{SQ}}/2)^2 - \frac{\Phi_3^2}{2L_3} \\ & + \frac{\Phi_0}{2\pi}\left(I_{c1}\cos(2\pi\Phi_1/\Phi_0) + I_{c2}\cos(2\pi\Phi_2/\Phi_0)\right). \end{aligned} \tag{2.86}$$

It can be easily checked that applying the Euler-Lagrange equations $\frac{d}{dt}\frac{\partial L}{\partial \dot{\Phi}_i} - \frac{\partial L}{\partial \Phi_i} = 0$ to this Lagrangian yields the correct circuit equations of motion (i.e., Kirchoff's current conservation equations). From (2.86) we can obtain the canonical momenta $Q_i = \frac{\partial L}{\partial \dot{\Phi}_i}$ and write down the system Hamiltonian $H = \sum_i \dot{\Phi}_i Q_i - L$,

$$\begin{aligned} H_{3\mathrm{D}} =& \sum_{i=1}^{3} 8E_{C_i}\frac{q_i^2}{2} + E_{L_1}\frac{(\varphi_1 - \varphi_3 - \Phi_t - \Phi_{\mathrm{SQ}}/2)^2}{2} + E_{L_2}\frac{(\varphi_2 - \varphi_3 - \Phi_t + \Phi_{\mathrm{SQ}}/2)^2}{2} + E_{L_3}\frac{\varphi_3^2}{2} \\ & - E_{J_1}\cos(\varphi_1) - E_{J_2}\cos(\varphi_2), \end{aligned} \tag{2.87}$$

where we have defined the dimensionless flux variables as usual (e.g., $\varphi_i = \frac{2\pi}{\Phi_0}\Phi_i$), along with the capacitive, inductive, and Josephson energies as usual. We can write this in an equivalent, more familiar form by shifting φ_3 by the constant φ_t and φ_1/φ_2 by $\pm\varphi_{\mathrm{SQ}}/2$

to obtain

$$H_{\mathrm{3D}} = \sum_{i=1}^{3} 8E_{C_i}\frac{q_i^2}{2} + E_{L_1}\frac{(\varphi_1-\varphi_3)^2}{2} + E_{L_2}\frac{(\varphi_2-\varphi_3)^2}{2} + E_{L_3}\frac{(\varphi_3-\varphi_t)^2}{2}$$
$$- E_{J_1}\cos(\varphi_1+\varphi_{\mathrm{SQ}}/2) - E_{J_2}\cos(\varphi_2-\varphi_{\mathrm{SQ}}/2). \tag{2.88}$$

Now that we have a Hamiltonian for the 3D fluxmon circuit, we can compare its spectrum and matrix elements to those of the 1D model. Unlike the case with the 1D model, it is not practical to diagonalize (2.87) by discretizing the wavefunction in three-dimensional flux space, because the resulting matrices would be much too large. Instead, we can choose a basis in which to expand the wavefunction, and then keep the first N basis levels, with N a small enough number to allow for relatively fast numerical diagonalization while maintaining high precision and sufficient accuracy. A very common technique is to expand in the basis of tensor products of the harmonic oscillator states for each degree of freedom, as we did with the modes of the transmission line fluxmon model. An efficient way to do this is to perform a change of basis into the normal modes of the harmonic part of the Hamiltonian, and then diagonalize using the harmonic oscillator states in this basis. Explicitly, we can write (2.87) as the sum of a linear plus nonlinear

part,

$$H_{\rm 3D} = \frac{1}{2}\begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ q_1\\ q_2\\ q_3\end{pmatrix}^T \begin{pmatrix} E_{L_1} & 0 & -E_{L_1} & 0 & 0 & 0\\ 0 & E_{L_2} & -E_{L_2} & 0 & 0 & 0\\ -E_{L_1} & -E_{L_2} & E_{L_1}+E_{L_2}+E_{L_3} & 0 & 0 & 0\\ 0 & 0 & 0 & 8E_{C_1} & 0 & 0\\ 0 & 0 & 0 & 0 & 8E_{C_2} & 0\\ 0 & 0 & 0 & 0 & 0 & 8E_{C_3}\end{pmatrix}\begin{pmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ q_1\\ q_2\\ q_3\end{pmatrix} + H_{\rm 3D}^{\rm nonlin.}, \tag{2.89}$$

and use the techniques of classical mechanics to transform to a normal mode basis in which the quadratic form matrix in (2.89) is diagonal [54]. We then truncate this basis to diagonalize the full Hamiltonian (2.87). The matrix elements needed to construct the matrix for $H_{\rm 3D}^{\rm nonlin.}$ can be found analytically using the identities

$$\cos\left(\sum_k c_k\varphi_k + \varphi_t\right) = \mathrm{Re}(e^{i\varphi_t}e^{ic_1\varphi_1}\otimes e^{ic_2\varphi_2}\otimes e^{ic_3\varphi_3}) \tag{2.90}$$

(valid because operators in different tensor product spaces commute) and the following identity [55, 56] for harmonic oscillator basis states

$$\langle j|e^{ir(\hat{a}+\hat{a}^\dagger)}|k\rangle = (i)^{3j+k}\sqrt{\frac{j!}{k!}}e^{-\frac{r^2}{2}}r^{k-j}L_j^{(k-j)}(r^2), \tag{2.91}$$

where $L_n^\alpha(z)$ are the generalized Laguerre polynomials.

In Fig. 2.15 we plot f_{10} vs. DC SQUID bias at zero tilt for the 3D model as well as for a naive "1D simple" model that ignores the presence of the inline inductors and capacitors. As can be seen, there is a significant difference between the models (of order 1 GHz in f_{10} and 0.1 in $\varphi_{10} = \langle 0|\hat{\varphi}|1\rangle$) throughout much of the operating regime. This simply means that the introduction of the inline inductances and capacitances has a non-negligible effect for the physically realistic values of $L_1 = L_2 = 20\,\mathrm{pH}$ and $C_1 = C_2 = 5\,\mathrm{fF}$.

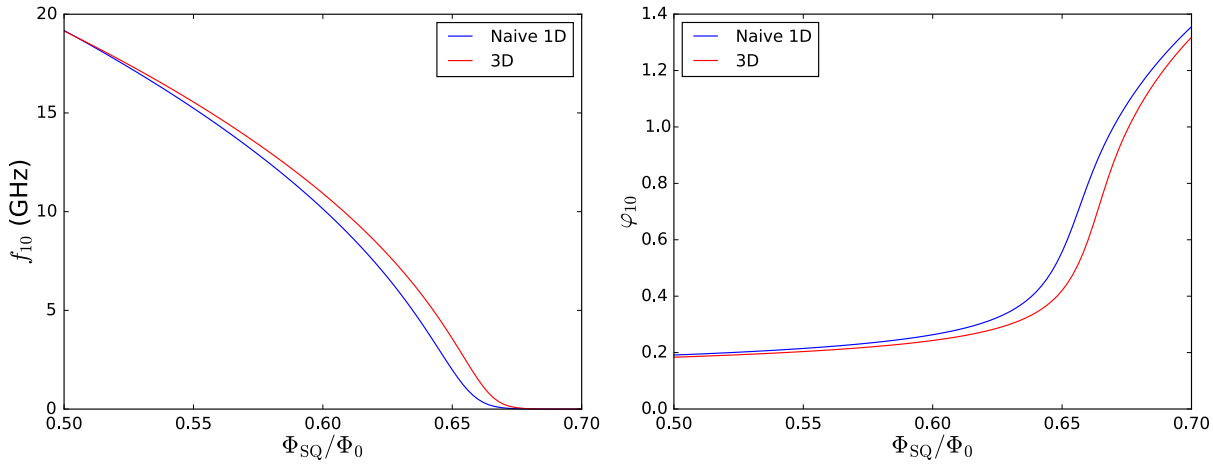
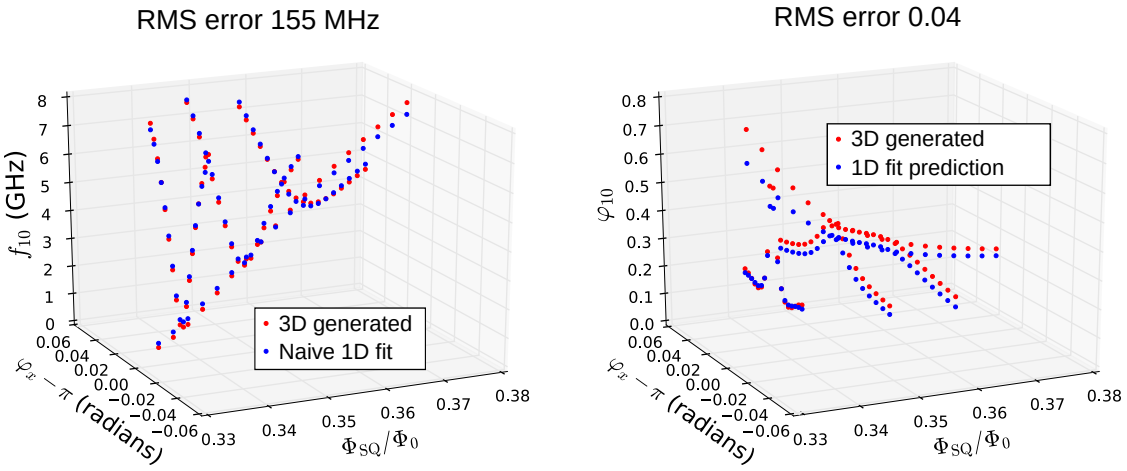


Figure 2.15: 3D vs. 1D simple model for $L = L_3 = 624\,\text{pH}$, $C = C_3 = 110\,\text{fF}$, $L_1 = L_2 = 20\,\text{pH}$, $C_1 = C_2 = 5\,\text{fF}$ and $I_{c1} = I_{c2} = 0.64\,\mu\text{A}$ ($\beta_{\max} = 2.4$). The numerical diagonalization used the first 50 harmonic oscillator states for the lowest normal mode, and 5 states for each of the other two normal modes.

One may ask whether or not one can model the 3D circuit by an effective 1D model with modified L, C and $\beta_{\max}$ ($\propto I_c^{\max}$), in which case a full 3D model would not be necessary for the fitting of experimental data. In Fig. 2.16a, we plot the result of finding the values of L, C, and $\beta_{\max}$ that minimize the RMS error in f_{10} over a set of flux biases (Φ_{SQ}, Φ_t) covering a good portion of the fluxmon operating range (for simplicity we have assumed zero junction asymmetry χ). As can be seen, the best possible 1D model still has an RMS error of 155 MHz in f_{10}, and an RMS error of 0.04 in the flux matrix element φ_{10}. The fitted values of L and C also do not have any apparent physical meaning. We can go a step further by allowing for a flux offset in Φ_{SQ} to be an extra fitting parameter. Although there is no obvious physical intuition for including such a shift, this is not an unreasonable thing to do because when we fit experimental spectrum data, we do not know the exact applied flux and must often include an offset as a fit parameter anyway. The result of this fit is shown in Fig. 2.16b. Here, the RMS error in f_{10} is much lower

a) 1D simple: fitting $L, C, I_c^{\max}$



b) 1D simple: fitting $L, C, I_c^{\max}, \delta\Phi_{\rm SQ}$

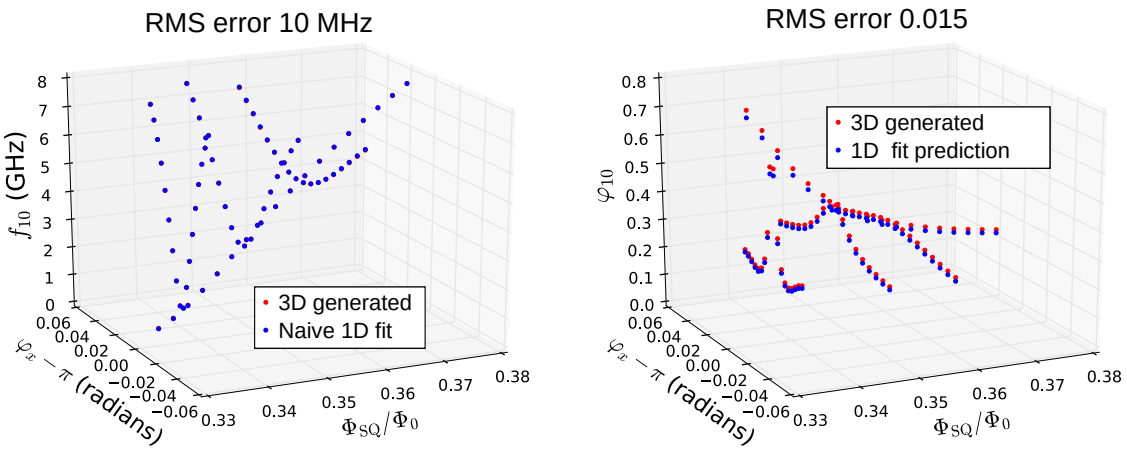


Figure 2.16: a) Result of fitting numerically generated 3D fluxmon spectrum to a 1D simple model, with fitting parameters L, C, and $I_c^{\max}$. The 3D model parameters used to generate the 3D spectrum are the same as in Fig. 2.15: $L_3 = 624\,\text{pH}$, $C_3 = 110\,\text{fF}$, $L_1 = L_2 = 20\,\text{pH}$, $C_1 = C_2 = 5\,\text{fF}$ and $I_{c1} = I_{c2} = 0.64\,\mu\text{A}$ ($\beta_{\max} = 2.4$) The fitted 1D simple values are L =509 pH, C =152 fF, and $\beta_{\max} = 2.24$. b) The result of fitting the same 3D data to a 1D simple model that includes a constant offset $\delta\Phi_{\rm SQ}$ in SQUID flux bias as an additional fit parameter. Fit values were L =602 pH, C =131 fF, $\beta_{\max} = 2.57$, and $\delta\Phi_{\rm SQ} = .019\,\Phi_0$.

(10 MHz), but there is still a few percent RMS error in the predicted φ_{10}, a parameter crucial for predicting coupling strengths between qubits. The fitted value for L is also closer to its original value than it was in the previous fit. If one is happy with a percent error in these values, then such a fitting might be sufficient. However, if one wishes to fit the fluxmon spectrum to better than 10 MHz without using a 3D model (which may be too slow to use for fitting), or, importantly, to predict φ_{10} to within a percent, a more refined 1D model is necessary.

2.2.3.2 Efficient 1D "no caps" model: Born-Oppenheimer approximation

From the above section, it is clear that in order to be able to predict the matrix element φ_{10} to better than a few percent and f_{10} to better than 10 MHz, fitting the fluxmon's spectrum to the 1D simple model with a flux offset is not sufficient. However, using the full 3D model to fit the data instead would be computationally inefficient. Fortunately, there is a middle ground. There is a more physically realistic and more accurate 1D model than the one obtained from fitting to the 1D simple model. The idea is to realize that at low frequencies (much below the intra-SQUID resonances associated with the inline capacitances), the inline capacitors C_1 and C_2 don't play much of a dynamical role. Because of this, one might consider lumping the inline capacitors into the CPW capacitance, $C_3 \rightarrow C_3 + C_1 + C_2$, but keep the inline inductances where they are, as illustrated in Fig. 2.17. This would make the potential one-dimensional again. It might not be immediately obvious that removing the inline capacitors but leaving the inductors makes the potential one-dimensional. But this can easily be seen by the fact that each

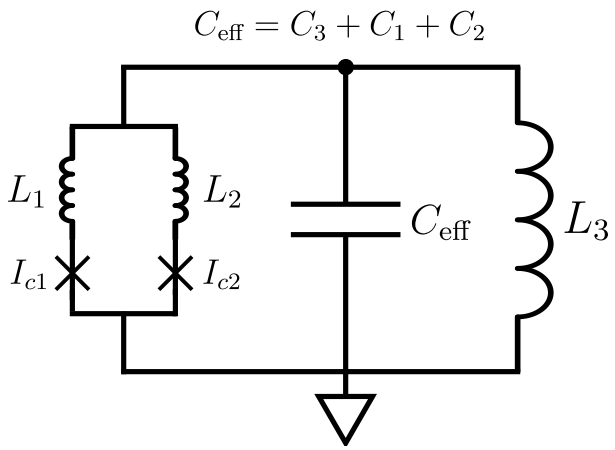


Figure 2.17: 1D "no caps" model absorbing the DC SQUID capacitances into the main capacitance and keeping the inline DC SQUID inductances.

node between the inline inductance and its junction contributes one current conservation equation, while flux quantization within the DC SQUID loop adds a third constraint equation. Therefore there are three linear constraint equations in four unknowns (or four constraint equations in five unknowns if you include the CPW flux as well), leading to one degree of freedom. This model is one-dimensional and therefore more efficient to use, and furthermore has fewer fit parameters than the full 3D model. We will see below that this allows us to model the underlying 3D circuit to very high accuracy.

The rigorous justification behind the reduction to the 1D no caps model with $C_{\text{eff}} = C_3+C_1+C_2$ is known as the Born-Oppenheimer approximation [57]. The Born-Oppenheimer approximation recognizes that one or more degrees of freedom are "fast" compared to a "slow" degree of freedom, in the sense that their characteristic resonance frequencies are much higher. Because of this, the fast degrees of freedom should be able to respond almost instantaneously to changes in the slow degree of freedom, meaning that as a function of the slow degree of freedom, the fast degrees of freedom should remain close

to their equilibrium values, i.e., the values that minimize the system's energy given the value of the slow degree of freedom. If this is the case, then the energy of the system is only a function of the slow degree of freedom, i.e., the presence of the fast degrees of freedom simply adds an effective potential to the Hamiltonian of the slow degree of freedom, with that effective scalar potential simply being the ground state energy of the fast degree(s) of freedom. This approximation also implies that the wavefunction of the slow degree of freedom is separable from that of the fast degrees of freedom. The Born-Oppenheimer approximation is commonly utilized in molecular physics and quantum chemistry, where it is used to more efficiently calculate the low energy spectrum of systems where slow nuclei evolve adiabatically with respect to the fast electrons. The Born-Oppenheimer approximation will also be utilized in Chapter 5 in the context of two slow qubits inductively coupled through a fast coupler [55].

We choose the fast degree of freedom in the 3D fluxmon circuit to be the fluxes across each of the inline inductances, φ_{L_1} and φ_{L_2}. We choose to eliminate φ_{L_1} and φ_{L_2} rather than φ_1 and φ_2 because there is a more substantial overlap of φ_1 and φ_2 with the slowest normal mode of the linear part of the system compared to φ_{L_1} and φ_{L_2} (by linear part, we mean when $I_{c1} = I_{c2} = 0$). We also expect the flux across these inductors to always be small. We therefore start by re-writing the Hamiltonian (2.87) using a new set of coordinates including φ_{L_1} and φ_{L_2}. The change of variables needs to be canonical in order to have the usual commutation relations apply, so we identify a suitable canonical

transformation $(\varphi_1, \varphi_2, \varphi_3, q_1, q_2, q_3) \to (\overline{\varphi}_1, \overline{\varphi}_2, \overline{\varphi}_3, \overline{q}_1, \overline{q}_2, \overline{q}_3)$ with

$$\begin{aligned}
\overline{\varphi}_1 &= \varphi_1 - \varphi_3 - \varphi_{1x} = -\varphi_{L_1} \\
\overline{\varphi}_2 &= \varphi_2 - \varphi_3 - \varphi_{2x} = -\varphi_{L_2} \\
\overline{\varphi}_3 &= \varphi_3 \\
\overline{q}_1 &= q_1 \\
\overline{q}_2 &= q_2 \\
\overline{q}_3 &= q_3 + q_1 + q_2,
\end{aligned} \tag{2.92}$$

where $\varphi_{jx} = \frac{2\pi}{\Phi_0}\Phi_{jx} = \varphi_t - (-1)^j \varphi_{\mathrm{SQ}}/2$. The change in q_3 was necessary to ensure that all of the canonical commutation relations $[\overline{\varphi}_i, \overline{q}_j] = i\delta_{ij}$ hold.

In terms of these new canonical variables, the Hamiltonian takes the form

$$\begin{aligned}
H_{3\mathrm{D}}(\overline{\varphi}_1, \overline{\varphi}_2, \overline{\varphi}_3, \overline{q}_1, \overline{q}_2, \overline{q}_3) =& 8E_{C_1}\frac{\overline{q}_1^2}{2} + 8E_{C_2}\frac{\overline{q}_2^2}{2} + 8E_{C_3}\frac{(\overline{q}_3 - \overline{q}_1 - \overline{q}_2)^2}{2} \\
&+ E_{L_1}\frac{\overline{\varphi}_1^2}{2} + E_{L_2}\frac{\overline{\varphi}_2^2}{2} + E_{L_3}\frac{\overline{\varphi}_3^2}{2} \\
&- E_{J_1}\cos(\overline{\varphi}_1 + \overline{\varphi}_3 + \varphi_{1x}) - E_{J_2}\cos(\overline{\varphi}_2 + \overline{\varphi}_3 + \varphi_{2x}).
\end{aligned} \tag{2.93}$$

We now perform the Born-Oppenheimer approximation by eliminating $\overline{\varphi}_1$ and $\overline{\varphi}_2$; i.e., we must minimize $H_{3\mathrm{D}}(\overline{\varphi}_1, \overline{\varphi}_2, \overline{\varphi}_3, \overline{q}_1, \overline{q}_2, \overline{q}_3)$ with respect to $\overline{\varphi}_1, \overline{\varphi}_2, \overline{q}_1, \overline{q}_2$. Minimizing over the charge terms $\overline{q}_1$ and $\overline{q}_2$ as a function of $\overline{q}_3$ means solving $\frac{\partial H_{3\mathrm{D}}}{\partial \overline{q}_1} = \frac{\partial H_{3\mathrm{D}}}{\partial \overline{q}_2} = 0$,

yielding the solutions

$$\overline{q}_1^{\text{equilib.}}(\overline{q}_3) = \frac{E_{C_2}E_{C_3}}{E_{C_1}E_{C_2} + E_{C_1}E_{C_3} + E_{C_2}E_{C_3}}\overline{q}_3,$$
$$\overline{q}_2^{\text{equilib.}}(\overline{q}_3) = \frac{E_{C_1}E_{C_3}}{E_{C_1}E_{C_2} + E_{C_1}E_{C_3} + E_{C_2}E_{C_3}}.\overline{q}_3 \tag{2.94}$$

Plugging in these solutions yields $8E_{C_{\text{eff}}}\overline{q}_3^2$ for the charging part of the Hamiltonian, where $E_{C_{\text{eff}}} = \frac{e^2}{2C_{\text{eff}}}$ is obtained from the effective capacitance

$$C_{\text{eff}} = C_3 + C_1 + C_2. \tag{2.95}$$

This constitutes the justification for lumping the inline capacitances together with the CPW capacitance as claimed by the diagram for the 1D no caps model in Fig. 2.17.

Next, to minimize over the flux terms $\overline{\varphi}_1$ and $\overline{\varphi}_2$, we must solve $\frac{\partial H_{3\text{D}}}{\partial \overline{\varphi}_1} = \frac{\partial H_{3\text{D}}}{\partial \overline{\varphi}_2} = 0$, leading to the transcendental equations

$$\overline{\varphi}_i + \beta_i \sin(\overline{\varphi}_i + \overline{\varphi}_3 + \varphi_{ix}) = 0, \tag{2.96}$$

where

$$\beta_i \equiv E_{J_i}/E_{L_i}. \tag{2.97}$$

For now we simply state that the solution to this transcendental equation is

$$\min_{\varphi}\left(\frac{\varphi^2}{2} + \beta\cos(\varphi + x)\right) \equiv \beta\cos_\beta(x), \tag{2.98}$$

where as shown in Ref. [55]

$$\cos_\beta(\varphi) = 1 + \sum_{\nu\geq 1}\frac{2J_\nu(\nu\beta)}{\beta\nu^2}(\cos(\nu\varphi) - 1). \tag{2.99}$$

Alternatively, one can easily solve the transcendental equation numerically. The meaning

of the $\cos_\beta$ function and its related derivative $\sin_\beta(x)$ will be discussed in more intuitive detail in Chapter 5 in the context of coupled qubits. For now we note that since the inline inductances are small, β_1 and β_2 are much less than 1, meaning that the $\cos_\beta(x)$ function will be fairly close to the function $\cos(x)$ that would occur in the complete absence of the inline inductors.

Now that $\overline{\varphi}_3$ and $\overline{q}_3$ form the only degree of freedom after the minimization is performed, we can refer to them simply as φ and q. Putting the charge and flux terms together we then obtain the one-dimensional Born-Oppenheimer Hamiltonian,

$$H_{\text{B.O.}} = 8E_{C_{\text{eff}}}\frac{q^2}{2} + E_{L_3}\frac{(\varphi-\varphi_t)^2}{2} - E_{J_1}\cos_{\beta_1}(\varphi+\varphi_{\text{SQ}}/2) - E_{J_2}\cos_{\beta_2}(\varphi-\varphi_{\text{SQ}}/2), \quad (2.100)$$

where we have shifted the flux by φ_t to put it in a more familiar form to compare to H_{1D} of Eq. (2.32).

We compare the result of diagonalizing $H_{\text{B.O.}}$ vs. diagonalizing the full 3D model in Fig. 2.18(a) after direct substitution of the circuit parameters (i.e., no fitting of any parameters). As can be seen, the 1D no caps model does much better than the 1D simple model without fitting, meaning that the Born-Oppenheimer approximation yields a much more physically accurate 1D model. In Fig. 2.18(b), we plot the result of the best-fit 1D no caps model. That is, the parameters L_3, C_{eff}, $I_c^{\max}$ and $\beta_1 = \beta_2$ were allowed to vary to minimize the RMS error in f_{10}, and notably there is no ad hoc $\delta\Phi_{\text{SQ}}$ flux offset. The best-fit RMS error is extremely small, only 0.5 MHz. Notably, this fitting is also able to predict the matrix element φ_{10} very accurately, to one part in 1000. In addition we

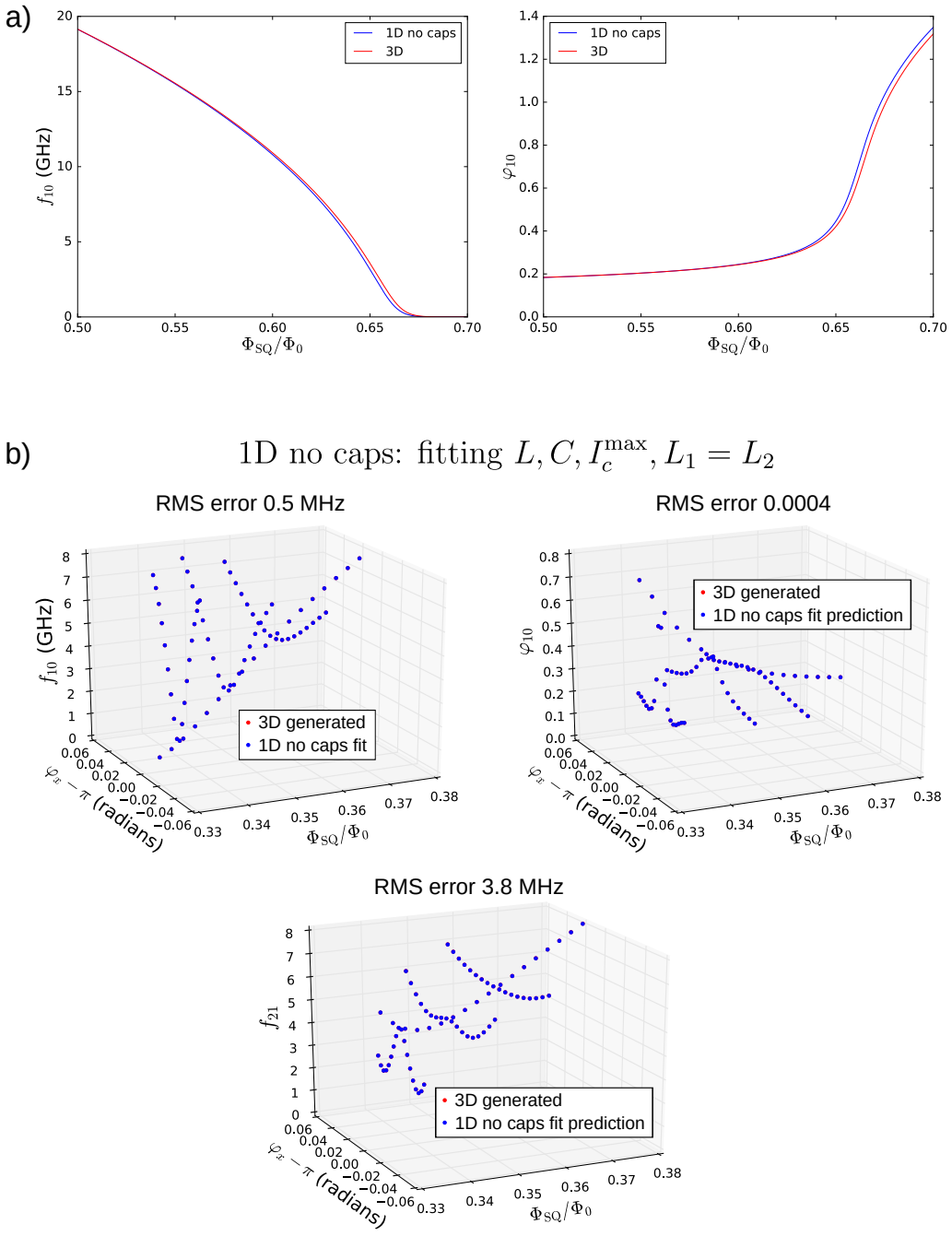


Figure 2.18: a) Comparison of full 3D and 1D no caps model upon direct substitution of 3D circuit parameters (plot is at zero tilt bias). We see that even before fitting, the effective 1D no caps model does much better than the 1D simple model. (b) Result of best-fit 1D no caps parameters using the same generated 3D data as in Fig. 2.16. Best-fit parameters are $L_3 = 625.05\,\text{pH}$, $C_{\text{eff}} = 119.9\,\text{fF}$, $\beta_{\text{max}} = 2.38$, $L_1 = L_2 = 20.0\,\text{pH}$.

have plotted the prediction for f_{21}, which is predicted to within 3.8 MHz of RMS error. This is much improved over the 1D simple fit, where even when allowing for a fitted flux offset $\delta\Phi_{\text{SQ}}$ the RMS error in predicted f_{21} was $\sim 30\,\text{MHz}$ (not shown in the figure). Furthermore, the fitted values for L_3, L_1, L_2, and C_{eff} are very close to what would be expected based on the parameters of the full 3D circuit used to generate the 3D model data, again showing that this is a very physically realistic model.

We conclude this section by noting that the Born-Oppenheimer approximation we used was a classical Born-Oppenheimer approximation, i.e., one that minimized the *classical* ground state energy of the fast degrees of freedom. A more accurate match in Fig. 2.18(a) could be achieved by accounting for the zero-point energy of the fast degrees of freedom in addition to the classical minimum. The techniques for accounting for the zero-point energy in the Born-Oppenheimer approximation are discussed in Chapter 5 in the context of coupled qubits. However there is no need to account for these effects in the simple 1D no caps model since with the classical Born-Oppenheimer approximation we have more that enough accuracy to fit experimental data and predict matrix elements to within desired accuracy.

Chapter 3

The fluxmon qubit: implementation, measurement, and operation

In this chapter, we describe the physical design and implementation of the fluxmon qubit, and describe in detail how it can be measured and operated both theoretically and experimentally. In particular, we will outline some of the basic procedures for calibrating single-fluxmon devices, and explore the ability of our circuit models to effectively model the observed fluxmon energy spectrum and extract physical device parameters. We will also look at the physics of single-qubit annealing as a case study.

3.1 Qubit design and fabrication

The fluxmon qubit is physically implemented by a segment of coplanar waveguide (CPW)[1] shunted with a DC SQUID comprised of two AlO_x Josephson junctions. All active metal on the device is superconducting aluminum,[2] which has for example the advantage of low kinetic inductance and insensitivity to normal state resistivity, but disadvantage of lower superconducting gap and higher sensitivity to vortex-inducing stray magnetic fields (see Chapter 4). The metal layers are deposited on a low-loss crystalline dielectric substrate, either sapphire (early devices) or silicon (current and future devices[3]). A 100 nm thick base layer of Al defines the CPW, ground planes, and control lines. This base layer is deposited on a solvent-cleaned substrate in a high-vacuum electron beam evaporation system, and the CPW and control lines are dry etched after photolithography via an inductively coupled plasma with a BCl_2/Cl_3 chemistry. Later on the Josephson junctions are connected to the CPW via lift-off after electron beam lithography. All junctions are made with a Dolan bridge double-angle evaporation process, similar to the method described in Ref. [58]. This step involves an Ar ion milling step to remove the native aluminum oxide of the base layer in order to make galvanic superconducting contact

[1] Although, as we will see in chapter 6, newer fluxmon devices based on a more scalable flip-chip architecture can have most of their return current flow in a different plane above the qubit on a different chip, quasi-similar to a microstrip geometry, and so cannot be considered a pure coplanar waveguide.

[2] Actually, there is also a Ti/Au layer that is deposited using an optical lithography step for the sole purpose of later use as an alignment layer for e-beam lithography. None of the gold features are in galvanic contact with the aluminum layers. Also, in the flip-chip architecture devices later on, superconducting TiN pads and inter-chip superconducting In bumps are used as well.

[3] All else equal, silicon is preferable for several reasons. For one, there is already an enormous amount of knowledge and experience with silicon in the integrated circuit industry over the past decades. Silicon more readily allows for some potentially useful substrate-modifying techniques, such as substrate trenching, wafer-thinning, and through-substrate vias.

between the layers. However, this milling process turns out to induce dielectric loss via substrate surface amorphization [59, 60] (see section 4.2.2.4). This mill-induced damage is much more severe on silicon than on sapphire, so for all fluxmon samples on silicon the galvanic contact between CPW and junction metal is made in a separate, subsequent "bandage" step that protects the substrate from the ion mill, as detailed in Ref. [60]. The fluxmon junctions themselves are similar in size to Xmon junctions (slightly larger in area, but substantially smaller than $1\,\mu\mathrm{m}^2$ in order to minimize two-level state defects within the junction [61]), but with a critical current density that is $10-20$ times larger. The reason for this larger critical current density is that in order to obtain the double-well potential energy characteristic of a flux qubit, the Josephson inductance needs to be on the order of the geometric inductance, which is set by the CPW geometry. Because of this, the DC SQUID normal state resistance is of order a couple hundred Ω compared to $\sim 5\,\mathrm{k}\Omega$ typical of Xmon transmon qubits.

Besides its low dielectric loss, the CPW-based design was chosen to allow a long loop of distributed inductance and appreciable persistent current, so that strong coupling to many qubits at once (as in Fig. 1.4) is possible for quantum annealing applications. This is in contrast to a conventional flux qubit [47] or a high-coherence capacitively-shunted flux qubit [62], where the inductance-like term is provided by two or more large lumped-element Josephson junctions in addition to a smaller junction, and the double-well potential is formed within a two-dimensional phase space [47, 63]. In this conventional design, it is possible to achieve longer qubit coherence times, but it is quite

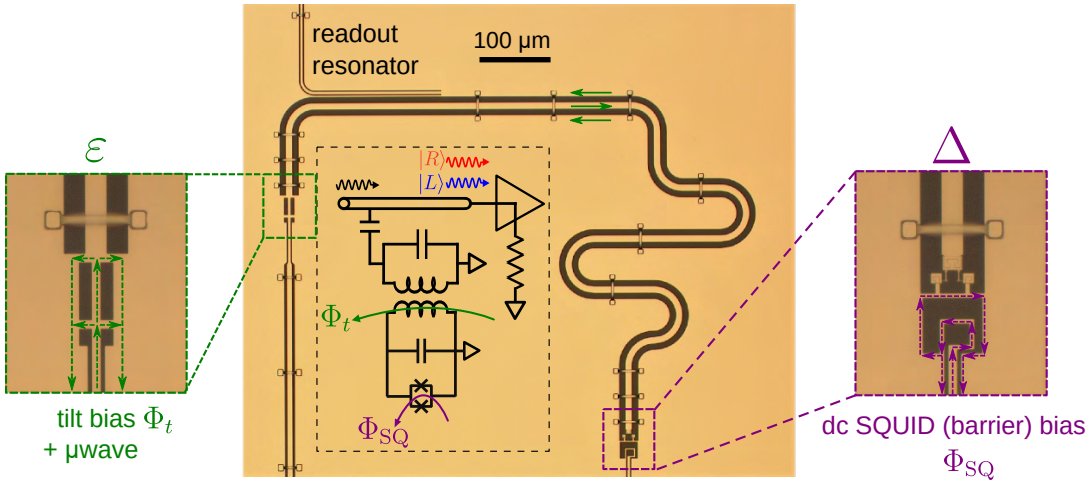


Figure 3.1: Example micrographs of single-fluxmon device on sapphire substrate with flux bias lines. Beige color is aluminum metal, dark regions are sapphire substrate where the Al is etched away. Current flow for tilt and SQUID flux bias lines illustrated with green and purple respectively. The main qubit loop and bias utilize a gradiometric CPW current. Qubit is inductively coupled to a CPW readout resonator with $f_r \sim 7 - 8$ GHz.

difficult to couple multiple flux qubits together without galvanically connecting them, let alone controllably couple to multiple qubits at once.[4] In contrast, for the fluxmon, the potential is inherently one-dimensional (within the lumped-element approximation for the CPW), and more importantly the inductance is spread over a macroscopically long physical distance to accommodate strong coupling to many other qubits at once. This is one feature in common with D-Wave's qubits, where the inductance is similarly distributed over a transmission line, but with a microstrip geometry that necessitates the use of performance-degrading lossy dielectrics within the qubit.

As shown in Fig. 3.1, the fluxmon's CPW is shorted on one end and shunted with a DC SQUID on the other. The CPW has inductance and capacitance per unit length

[4]The capacitively shunted flux qubit has just begun to achieve appreciable coupling strengths between two qubits [64] (but with degraded coherence relative to a single qubit), although it is unclear how to extend this to multiple simultaneous qubits except possibly through a proposed complicated network of nested coupler circuits.

$\mathcal{L}$ and $\mathcal{C}$ (accurately derivable from standard CPW formulae based on conformal mapping [65]), with length ℓ such that at frequencies below its $\lambda/4$ resonance, it acts as a lumped-element parallel $L = \mathcal{L}\ell \approx 600\,\text{pH}$ and $C = \mathcal{C}\ell/3 \approx 100\,\text{fF}$, as shown in Chapter 2. The high-quality CPW capacitance dominates over the junction capacitance ($\sim 5\,\text{fF}$), minimizing dielectric loss, a shared feature with the C-shunted flux qubit [62]. The linear CPW inductance replaces the need for additional large junctions, and makes the potential intrinsically one-dimensional. Shunting the CPW with a DC SQUID adds a tunable nonlinear term, allowing the potential to be varied from a harmonic single well to a conventional flux qubit double-well [47]. The potential's "tilt" is tuned by an external flux bias Φ_t^x threading the CPW mode of current flow (color-coded green in the figure), while the barrier height is tuned via the DC SQUID external flux Φ_{SQ}^x (color-coded purple in the figure), yielding a tunable Josephson term parameterized by $\beta = (2\pi/\Phi_0)^2 E_J L = \beta_{\max}\cos(\pi\Phi_{\text{SQ}}^x/\Phi_0)$, where $\beta_{\max}$ is typically designed to be ~ 2.5. Grounds are connected throughout the circuit with Al airbridges[5] to control linear inductive crosstalk and coupling to spurious modes. All qubits measured in this thesis have $f_{LC} \approx 20\,\text{GHz}$, unless otherwise specified.

The precise geometry and placement of the flux bias lines and airbridges are crucial for obtaining the desired mutual inductances to the qubit loops and low geometric crosstalk (we will see this is even more true for coupled fluxmon devices where there are many bias

[5]On all devices with sapphire substrates in this thesis, the airbridges are fabricated via a resist reflow technique as detailed in Ref. [66], while on all silicon samples, the airbridges are formed by a more robust SiO_2 scaffold and HF vapor release technique as developed in Ref. [67] (however, the geometry of the latter style of airbridge requires that the release is done as the very last step of the fab). In later, coupled-qubit devices, these airbridges play a crucial role by forming part of the qubit itself as galvanic hopping crossovers (see Chapter 6).

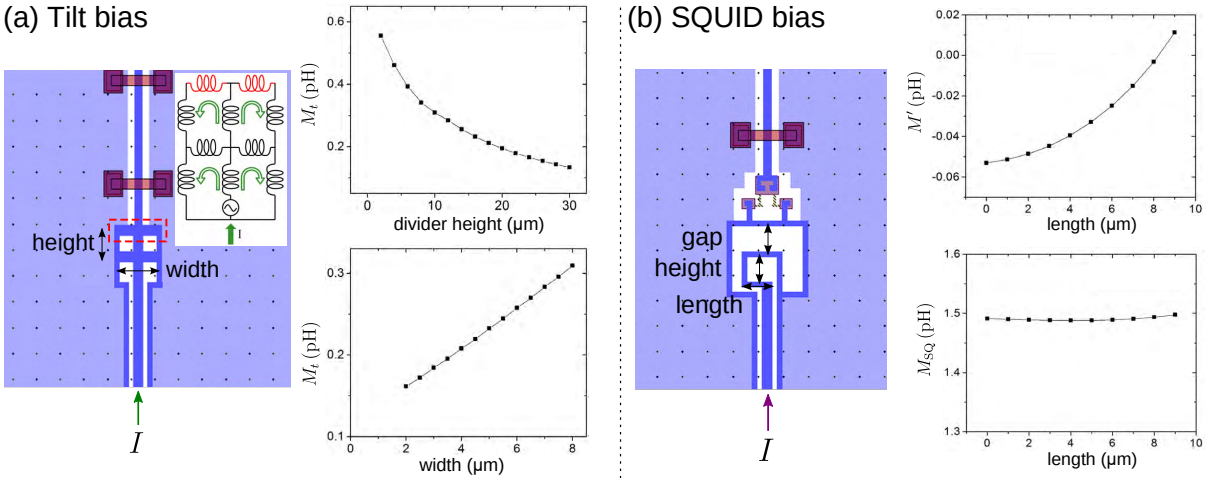


Figure 3.2: Design and Sonnet simulation results for typical single-fluxmon bias line mutual inductances. (a) Tilt bias line, showing current divider scheme in inset. (b) SQUID bias line. The gap controls the primary coupling to the SQUID while the length is used to null the net coupling to the CPW mode. The height allows for tolerance in the dimensions of the gap and length. The plotted simulation results are for a height equals gap equals 10 μm.

lines and loops that come into close proximity with each other). There are various ways one might go about designing these bias lines. For the tilt bias line, we chose a symmetric current divider design, schematically shown in Fig. 3.2(a). This was done to enable very low (a fraction of a pH) mutual inductances to the qubit tilt loop in order to preserve coherence (see Chapter 4 for quantitative descriptions of dissipation and dephasing in the qubit induced by the presence of the flux bias lines). Furthermore, as can be seen from the simulation results, the resulting M_t is not overly sensitive to the dimensions of the current divider. Typically, we must have $M_t \lesssim 0.4\,\text{pH}$ to avoid any effects on T_1 or steady-state stray population. This forms a trade-off with control, as it is in general easier to calibrate the qubit and easier to reduce the effects of crosstalk if one were to more strongly couple each flux bias line to its intended loop. We can be more lenient in the SQUID bias and allow for larger M_{SQ}, as the fluxmon's flux sensitivity to SQUID bias is

smaller (except at the special point of zero tilt) and, unlike the tilt bias line, the SQUID line is filtered at microwave frequencies (details in section 3.3.1). Accordingly we will typically design the SQUID mutual inductances to be $M_{\mathrm{SQ}} \approx 1\,\mathrm{pH}$. However, we must be careful not to couple the SQUID bias line to the tilt loop (i.e., the CPW mode of current flow), as this can introduce significant crosstalk (i.e., non-orthogonality between controls) and in addition couple excess noise into the tilt loop. To control this crosstalk mutual inductance $M_{\mathrm{SQ}\to\mathrm{t}}$ (often referred to simply as M'), we need to eliminate any asymmetry in the bias line return current geometry so as not to induce CPW-mode flux. One way to do so is as shown in the right inset of Fig. 3.1, where a more weakly coupled secondary loop can adjust its dimensions so as to cancel out any CPW-mode flux induced by the primary SQUID bias current above, using Sonnet simulation software to guide the design. In practice, this strategy has allowed us to consistently obtain $M' \lesssim 0.1\,\mathrm{pH}$, consistent with the Sonnet predictions. As $\mu_0 \sim 1\,\mathrm{pH}/\mu\mathrm{m}$, this means that careful layout and simulation on the $1\,\mu\mathrm{m}$ length scale were needed to engineer these coupling structures.

3.2 Qubit measurement: theory

At the end of quantum annealing, all qubit potentials will be in their double-well limit, and the result of the annealing is determined by whether the qubit ended up occupying either its left or right well. It is therefore natural to design a readout scheme that distinguishes the left and right well ground states $|L\rangle$ and $|R\rangle$. However, it is also desirable to be able to faithfully read out the $|0\rangle$ and $|1\rangle$ states of the qubit when the

potential is not in the double-well regime. In this section, we describe a way to do this that involves faithfully transferring the ground and first excited states of the system to the left and right well ground states of a double-well potential and performing the same type of readout after this projection process is complete.

Since the states $|L\rangle$ and $|R\rangle$ are clearly distinguishable by their expected fluxes, one might consider using a SQUID to read out the qubit by detecting the magnetic flux induced in the SQUID by the qubit. The conventional switching current SQUID method is indeed a very common readout method, where a nearby SQUID is biased to detect the current at which it switches to its voltage state. A potential issue with this readout is that when a readout SQUID switches to its normal voltage state, it will generate quasiparticles on the chip which can detrimentally affect the qubit during subsequent operation unless one waits a very substantial amount of time (up to of order 1 ms) for the quasiparticles to diffuse and decay via recombination [68]. In addition, using additional SQUIDs makes the fabrication process more complex.

To address all of the above, we implement a variation of microwave-based dispersive readout inspired by phase qubits [69, 70, 71]. In dispersive readout, a harmonic oscillator (here, a CPW electromagnetic resonator with typical $\omega_r/(2\pi) \sim 8\,\mathrm{GHz}$, referred to as a readout resonator) is weakly coupled to the qubit, and the frequency of that oscillator depends on the state of the qubit [72, 73]. The readout resonator protects the qubit from the dissipative readout line entering the chip from a $50\,\Omega$ environment, and furthermore multiple readout resonators can be coupled to a single transmission line, allowing for

frequency-multiplexed readout of multiple qubits using a single input/output line. It also reduces on-chip dissipation because the resonator energy dissipates off-chip in the $50\,\Omega$ environment.[6]

It might seem at first that microwave-based readout couldn't possibly distinguish the left from the right well of a double-well potential, because there is complete symmetry between left and right. This would indeed be true for a symmetric potential (i.e., zero tilt). In order to distinguish the two wells we must add a tilt bias to the potential so that the two minima have different curvature. In that case the oscillations associated with the left and right wells will have different frequencies, shifting the readout resonator differently. In the following two sections, we will first describe the projection process that transfers the qubit energy states $|0\rangle$ and $|1\rangle$ into the left and right wells, and then describe the dispersive readout used to distinguish the two wells $|L\rangle$ and $|R\rangle$.

3.2.1 Double-well projection process

As mentioned above, for device characterization we require a method to transfer the $|0\rangle$ and $|1\rangle$ energy eigenstates to the left and right wells of a double-well potential, so that the states can be read out using the same method that will be used to read out $|L\rangle$ and $|R\rangle$ at the end of quantum annealing. A graphical depiction of a scheme to do this is shown in Fig. 3.3(a)/(b). After performing any qubit state manipulations with microwave pulses to prepare a state within the $\{|0\rangle, |1\rangle\}$ qubit subspace, we first apply a pre-determined amount of tilt to the potential. The amount of pre-tilt is determined

[6]This isn't always strictly true, since the resonator will also have intrinsic microwave dissipation.

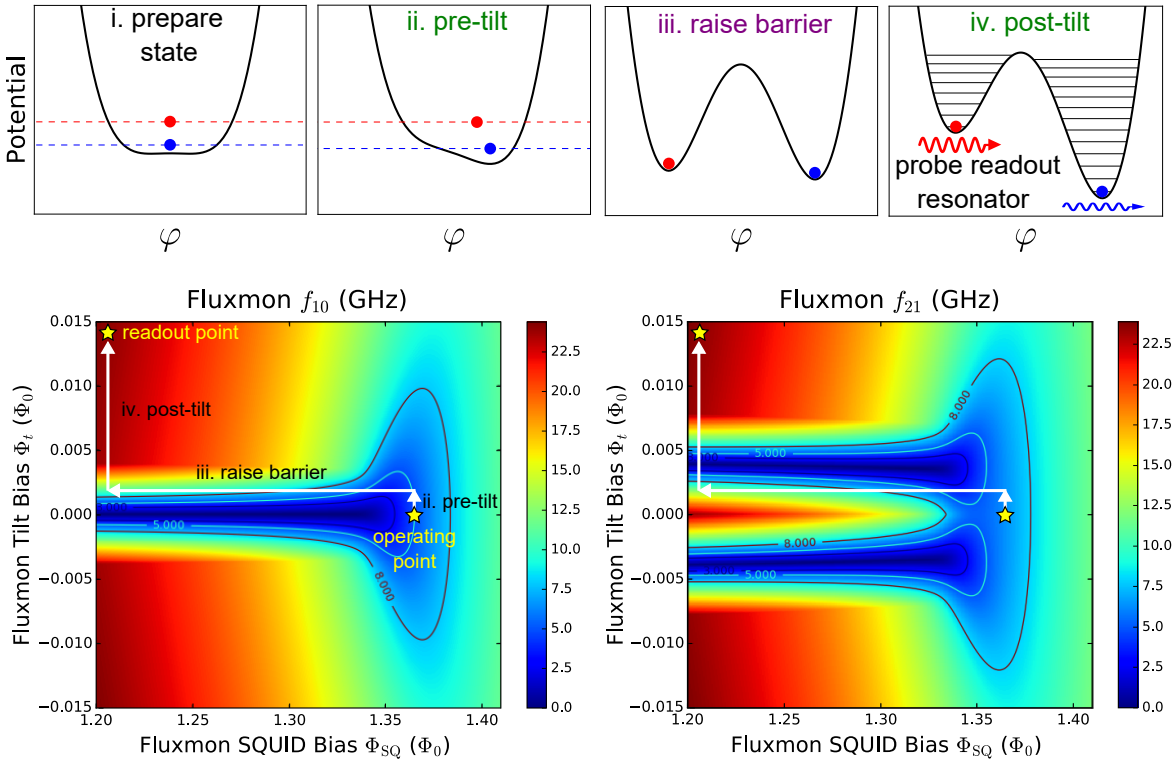


Figure 3.3: Projection process: state preparation, measurement, and dispersive readout.

by the requirement that upon raising the barrier, the $|1\rangle$ and $|0\rangle$ energy eigenstates are adiabatically transferred to the localized left and right well ground states with high fidelity according to the adiabatic theorem. If the pre-tilt is too large, the well with lower energy will be so deep that both the ground and first excited states of the system reside in the same well. For typical fluxmon parameters this would happen at a critical pre-tilt of $\Phi_t^{\text{crit}}/\Phi_0 \approx .004$ (this pre-tilt depends on Z_{LC}). On the other hand if the pre-tilt is too small, the barrier must be raised very slowly in order to remain adiabatic, putting a constraint on the speed and increasing errors due to energy relaxation and thermalization with the environment during the raising of the barrier. In practice, using a pre-tilt of $\sim \Phi_t^{\text{crit}}/2$ works well, giving a minimum gaps (both f_{10} and f_{21}) of a few GHz for the

whole projection process as can be seen in Fig. 3.3. This allows for fast, nanosecond-scale flux pulses to still be adiabatic all the way through the raising of the barrier, for both the ground and first excited states. One can think of this process as a special case of quantum annealing with a single qubit, but for both the ground and first excited states.

Once the barrier is raised and the tunneling between wells is essentially turned off, we need to apply a larger post-tilt to the potential so that the two wells have distinguishable curvatures for dispersive microwave readout (described in detail below). The post-tilt pulse is nonadiabatic (since the ground state in one well will cross the energy levels of the weakly tunnel-coupled states within the other well), which is what's desired to keep the left/right well ground states in the ground state of the left/right well, and so can also be performed fast (nanosecond timescales). Since the ground states of both wells are quasi-stable (the tunneling times can be made much longer than the time needed for the readout resonator probe pulse, as detailed in the next section), the readout pulse can be made very long (easily a microsecond or more), enabling high fidelity readout without requiring the use of quantum-limited paramps that are currently necessary for the readout of transmon qubits for surface code quantum computing [74, 75].

The observant reader might notice that when this projection process is used to read out the $|1\rangle$ state, the qubit's frequency f_{10} will cross the frequency $f_r \sim 8\,\mathrm{GHz}$ of the readout resonator, which could lead to loss of population from the qubit to the resonator and therefore an error in the readout scheme. However, the design parameters of the system (described in a later section) are such that the system will undergo a full

nonadiabatic transition through the qubit-resonator avoided crossing. The probability of making a Landau-Zener transition [76, 77] through the avoided level crossing is given by [78]

$$P_{\mathrm{LZ}} = e^{-[2\pi(\Delta E_{\mathrm{min}})^2/4\hbar\nu]}, \tag{3.1}$$

where ΔE_{min} is the minimum gap between the qubit and resonator energy levels and $\nu = |d(E_1 - E_2)/dt|$ is the "level velocity" of the uncoupled energies. For fast enough pulses then, the qubit should not lose population to the resonator. For a realistically fast 8 ns barrier raising pulse at a pre-tilt bias of $\Phi_t^{\mathrm{crit}}/2$, the minimum gap is $\sim 1\,\mathrm{MHz}$, so that Eq. (3.1) gives a prediction of a loss of less than 1×10^{-5} of the qubit population.[7]

3.2.2 Dispersive Readout

There are two complementary ways to understand dispersive readout. One is a full quantum description that shows how the two different well states will entangle with different coherent pointer states of the resonator that can be distinguished by a classical apparatus. The other simply views the two wells as classical (approximately) harmonic oscillators which each shift the readout resonator's resonance frequency by different amounts due to the different effective inductances they present to the resonator. These two approaches agree in practically relevant limits, as we will discuss. In the following, we will consider both descriptions in order to arrive at a more holistic understanding.

[7] Note that this minimum gap is much less than the $g_{qr} = 100\,\mathrm{MHz}$ discussed in the next section, which would have been the minimum gap if the qubit crossed the resonator at zero tilt, in which case the qubit would lose about 20% of its excited state population. Instead, the qubit energy crosses the resonator away from zero tilt, where in the fluxmon's energy basis the interaction with the resonator is no longer transverse but mostly longitudinal.

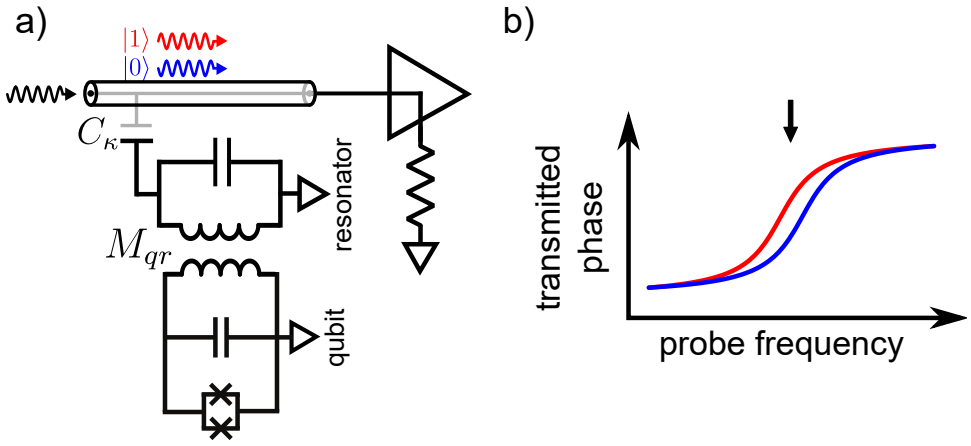


Figure 3.4: (a) Circuit diagram of dispersive readout using a readout resonator inductively coupled to the qubit and capacitively coupled to a transmission line. (b) Illustration of two different curves in the transmitted phase due to scattering from a resonator with two distinct qubit-induced frequencies.

3.2.2.1 Quantum description of dispersive readout

It is natural to inductively couple the readout resonator to the fluxmon qubit, as represented by the mutual inductance M_{qr} in the lumped-element circuit diagram of Fig. 3.4(a). The interaction Hamiltonian is, for small coupling strengths,[8]

$$\hat{H}_{\text{int}} = M_{qr}\hat{I}_q\hat{I}_r \approx M_{qr}\frac{\hat{\Phi}_q}{L_q}\frac{\hat{\Phi}_r}{L_r}, \tag{3.2}$$

where $\hat{\Phi} = \Phi_{zp}(\hat{a}^\dagger + \hat{a})$ for each of qubit and resonator as in (2.40). We can then expand the system Hamiltonian $\hat{H} = \hat{H}_q + \hat{H}_r + \hat{H}_{\text{int}}$ using the basis of uncoupled qubit eigenstates,

$$\hat{H} = \hbar\sum_i \omega_i |i\rangle\langle i| + \hbar\omega_r\hat{a}^\dagger\hat{a} + \hbar\sum_{i,j} g_{i,j}|i\rangle\langle j|(\hat{a} + \hat{a}^\dagger), \tag{3.3}$$

[8]In chapter 5 we will discuss the exact form of the Hamiltonian for arbitrary coupling strengths. In our system, $M_{qr}/L_r \approx 0.01$, and we can use the simple form written here.

where $|i\rangle$ denotes the i^{th} qubit energy eigenstate of energy $\hbar\omega_i$ and $\hat{a}/\hat{a}^\dagger$ act on the resonator, and we have defined the coupling matrix elements

$$\hbar g_{i,j} = \frac{M_{qr}}{L_q L_r} \Phi_{zp,r} \langle i|\hat{\Phi}_q|j\rangle. \tag{3.4}$$

The Hamiltonian (3.3) can be greatly simplified by examining the nature of the coupling matrix elements. Firstly, since during readout, the states in the left and right wells have exponentially small overlap (except for states above the barrier, which have a negligible effect on the low-lying states anyway), we can consider the eigenstates in each of the two wells separately; i.e., we ignore all inter-well matrix elements and consider separately the intra-well matrix elements for the left and right wells. Secondly, we note that since the eigenstates within one well are nearly those of a harmonic oscillator when the wells are deep (as they are during readout), the matrix elements coupling non-adjacent states within a well should vanish. We can therefore reduce (3.3) to the following generalized Jaynes-Cummings Hamiltonian [44],

$$\begin{aligned}\hat{H} \approx{}& \hbar \sum_i \omega_i |i\rangle \langle i| + \hbar \sum_j \omega_j |j\rangle \langle j| + \hbar\omega_r \hat{a}^\dagger \hat{a} \\ &+ \left(\hbar \sum_i g_{i,i+1} |i\rangle \langle i+1| \hat{a} + \text{H.c.} \right) + \left(\hbar \sum_j g_{j,j+1} |j\rangle \langle j+1| \hat{a} + \text{H.c.} \right),\end{aligned} \tag{3.5}$$

where we have explicitly broken up the sum into a sum over the index variable i to denote states within the left well and the index variable j for states of the right well.

Dispersive readout is performed when the resonator's frequency ω_j is far detuned from any of the qubit frequencies $\omega_{i,i+1} = \omega_{i+1} - \omega_i$ relative to the coupling matrix

elements. This is the so-called dispersive regime. With the generalized Jaynes-Cummings Hamiltonian (3.5) in hand, it is also typical to make a rotating wave approximation (RWA) to make the calculations easier to perform and interpret. The calculation goes as follows. First, we make it a proper Jaynes-Cummings Hamiltonian by making the rotating wave approximation, which ignores terms such as $\hat{\sigma}_{-}\hat{a}$ and $\hat{\sigma}_{+}\hat{a}^{\dagger}$. This approximation is based on the fact that in a rotating frame these terms have a rapid oscillatory time dependence that should average out to zero compared to the dynamics of the excitation-preserving terms. Such an approximation is safe when the couplings g are not too large, when the detunings $\Delta_i = \omega_{i,i+1} - \omega_r$ are not larger than the $\omega_{i,i+1}$ themselves, and for a small mean photon occupation of the resonator, and no accidental resonances between higher non-RWA-coupled states (we will have more to say about this later) [72, 79]. Next, we assume that $\Delta_i \gg g_{i,i+1}$, where $\Delta_i = \omega_{i,i+1} - \omega_r$ is the detuning between the i^{th} sequential qubit frequency and the resonator's frequency, but not so large that the rotating wave approximation breaks down. One can then perform a change of basis that eliminates the interaction term to second order in $g_{i,i+1}/\Delta_i$. This can be done using the canonical transformation $\hat{H}' = \hat{D}\hat{H}\hat{D}^{\dagger}$ with $\hat{D} = \exp[\hat{S} - \hat{S}^{\dagger}]$ and generating operator $\hat{S} = \sum_i \frac{g_{i,i+1}}{\Delta_i}\hat{a}\,|i+1\rangle\langle i|$ [44]. After applying the Baker-Campbell-Hausdorff relation one

obtains, up to order $g_{i,i+1}^2/\Delta_i^2$, the "dispersive" Hamiltonian [44]

$$\begin{aligned}\hat{H}_{\text{disp.}} \approx & \sum_i \hbar\omega_i \left|i\right\rangle \left\langle i\right| + \sum_j \hbar\omega_j \left|j\right\rangle \left\langle j\right| + \hbar\omega_r \hat{a}^\dagger \hat{a} \\ & + \sum_i \hbar\chi_{i,i+1} \left|i+1\right\rangle \left\langle i+1\right| + \sum_j \hbar\chi_{j,j+1} \left|j+1\right\rangle \left\langle j+1\right| \\ & - \hbar\chi_{01,L} \hat{a}^\dagger \hat{a} \left|0_L\right\rangle \left\langle 0_L\right| + \sum_{i=1}^{\infty} \hbar(\chi_{i-1,i} - \chi_{i,i+1}) \hat{a}^\dagger \hat{a} \left|i\right\rangle \left\langle i\right| \\ & - \hbar\chi_{01,R} \hat{a}^\dagger \hat{a} \left|0_R\right\rangle \left\langle 0_R\right| + \sum_{j=1}^{\infty} \hbar(\chi_{j-1,j} - \chi_{j,j+1}) \hat{a}^\dagger \hat{a} \left|j\right\rangle \left\langle j\right|, \end{aligned} \tag{3.6}$$

with

$$\chi_{i,i+1} = \frac{g_{i,i+1}^2}{\omega_{i,i+1} - \omega_r} \tag{3.7}$$

[we have dropped negligible two-photon transition terms from (3.6)]. Although the dispersive Hamiltonian (3.6) might look complicated at first, there is a very simple way to interpret it for the purposes of dispersive readout. The first line is simply the bare (uncoupled) Hamiltonians of the qubit and resonator, where again we use the indices i/j to denote states in the left/right well. The second line describes an effective "dispersive shift" in the qubit energy levels by the amount $\hbar\chi_{i,i+1}$, which can be interpreted as a vacuum fluctuation-induced Lamb shift of the qubit spectrum. This term does not have any important effects for the purposes of readout. The last two lines contain terms proportional to $\hat{a}^\dagger\hat{a}$ times a qubit state projection operator. Since the bare resonator Hamiltonian is proportional to $\hat{a}^\dagger\hat{a}$, these terms can be interpreted as a qubit state-dependent dispersive shift in the resonator's resonance frequency ω_r. In particular, at

the end of the double-well projection process (or at the end of annealing), the qubit will be occupying either the left or right well ground states $|0_L\rangle$ or $|0_R\rangle$, meaning we only need to consider the first term on each of the two last lines. Therefore, depending on which well the qubit ended up populating, the readout resonator's frequency will be dispersively shifted by either the amount $-\chi_{01,L}$ or $-\chi_{01,R}$. For short, we will refer to the magnitudes of the left and right dispersive shifts simply as $\chi_L = g_L^2/\Delta_L$ and $\chi_L = g_L^2/\Delta_L$, where g_L/g_R are the $0,1$ coupling matrix elements for the left and right wells respectively and Δ_L/Δ_R are the detunings between the resonator's frequency and the left/right well resonance frequencies (technically their $0 \to 1$ transition frequencies) respectively. It follows that the difference in the readout resonator's frequency between the qubit being the left and right wells is[9]

$$\begin{aligned}\chi &\equiv \omega_{r,R} - \omega_{r,L} \\ &= \frac{g_L^2}{\Delta_L} - \frac{g_R^2}{\Delta_R} \\ &\approx \left(\frac{g}{\Delta}\right)^2 \delta\omega_w,\end{aligned} \tag{3.8}$$

where in the second line we have approximated $g_L \approx g_R \equiv g$ and $\Delta_L \approx \Delta_R \equiv \Delta$ and defined

$$\delta\omega_w \equiv \omega_R - \omega_L \tag{3.9}$$

as the difference between right and left well frequencies. Later in this section we will

[9]Note that this definition of χ differs by a factor of 2 from the usual definition in dispersive readout literature [44], where in the case of a transmon the qubit ground and excited state shift the natural frequency of the resonator equally up and down by $\pm\chi$, in which case $\omega_{r,|0\rangle} - \omega_{r,|1\rangle} = 2\chi$.

discuss the requirements of how large g and $\delta\omega_w$ need to be to obtain a large enough dispersive shift χ in the readout resonator. Later we will find it useful to have a formula for g between two oscillators in the harmonic limit,

$$\begin{aligned} g_{\text{harm.}} &= \frac{1}{\hbar}\frac{1}{2}\left(\frac{\Phi_0}{2\pi}\right)^2 \frac{M}{L_1 L_2}\frac{\sqrt{Z_1 Z_2}}{R_K/8\pi} \\ &= \frac{1}{2}\sqrt{\left(\frac{M}{L_1}\right)\left(\frac{M}{L_2}\right)\omega_1\omega_2}, \end{aligned} \tag{3.10}$$

obtained by combining (3.4) and (2.43).

To complete our quantum description of the dispersive measurement process, we consider what happens when the system is probed with a coherent state, which is what naturally builds up in the resonator when it is probed with a CW microwave tone.[10] To compute the dynamics, we first consider the time-evolution of uncoupled basis states $|0_L\rangle|n\rangle$ and $|0_R\rangle|n\rangle$ under the dispersive interaction Hamiltonian, where $|n\rangle$ is the n^{th} Fock state of the resonator:

$$\begin{aligned} |0_L\rangle|n\rangle &\to e^{-iH_{\text{int.}}^{\text{disp.}}t/\hbar}|0_L\rangle|n\rangle = e^{-i\chi_L nt}|0_L\rangle|n\rangle, \\ |0_R\rangle|n\rangle &\to e^{-iH_{\text{int.}}^{\text{disp.}}t/\hbar}|0_R\rangle|n\rangle = e^{-i\chi_R nt}|0_R\rangle|n\rangle, \end{aligned} \tag{3.11}$$

meaning each basis state gets its own phase shift proportional to the photon number with a sign dependent on the qubit state. Using this, we next consider what happens for

[10] It can be shown that a classical sinusoidal drive "implements" a displacement operator of the vacuum state, yielding a coherent state to high accuracy. Technically the readout resonator inherits a small amount of nonlinearity from the qubit so this reasoning is approximate, but holds very well for small enough mean photon number in the resonator.

a coherent state $|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$ in the resonator instead of a single Fock state:

$$|0_L\rangle |\alpha\rangle \to e^{-iH_{\text{int.}}^{\text{disp.}} t/\hbar} |0_L\rangle |\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} e^{-i\chi_L n t} \frac{\alpha^n}{\sqrt{n!}} |0_L\rangle |n\rangle = |0_L\rangle \left|e^{-i\chi_L t}\alpha\right\rangle,$$
$$|0_R\rangle |\alpha\rangle \to e^{-iH_{\text{int.}}^{\text{disp.}} t/\hbar} |0_R\rangle |\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} e^{-i\chi_R n t} \frac{\alpha^n}{\sqrt{n!}} |0_R\rangle |n\rangle = |0_R\rangle \left|e^{-i\chi_R t}\alpha\right\rangle. \quad (3.12)$$

We can then see that the two different well ground states will become entangled with two coherent states with distinct phases. As is well-known, coherent states are eigenstates of dissipation (i.e., of $\hat{a}$) and therefore act as dynamically stable pointer states even when inevitably coupled to an environmental bath. When the entanglement "collapses" to one classically observable coherent state or the other due to decoherence, the macroscopically distinguishable measurement result will therefore correspond to the qubit being in either the left or right well ground state as desired. However, in practice, the fluxmon itself fully decoheres (due to dephasing between wells on the order of nanoseconds) before the measurement is complete. Similarly, dissipation may also act on a timescale faster than the time taken by a readout probe tone, so that dissipation will also tend to localize the qubit in one well or the other. Therefore, when we probe the readout resonator we actually have a mixed state of qubit-resonator $\left|0_{L/R}\right\rangle \left|\alpha_{L/R}\right\rangle$ product states instead of a coherent entangled state, so in a sense the "measurement" actually occurs while the barrier is raised, before the readout probe pulse. However, the fact that this occurs does not affect the fidelity of the measurement if we only need to measure occupation probabilities and not phases, since these are in the above sense classical states.

We note that the above analysis relied on making the rotating wave approximation in addition to the dispersive limit approximation, i.e., it required that

$$g \ll \Delta \ll \omega_{L/R} + \omega_r. \tag{3.13}$$

The second inequality is there so that the effect of the counter-rotating terms in the original Rabi Hamiltonian have a negligible effect compared to that of the RWA terms. However, for reasons discussed below, for typical fluxmon device biased at its readout point, we have $g/(2\pi) \approx 100\,\mathrm{MHz}$, $\omega_L/(2\pi) \approx \omega_R/(2\pi) \approx 20\,\mathrm{GHz}$ and $\omega_r \approx 8\,\mathrm{GHz}$. While the first inequality of (3.13) clearly holds, the second one does not hold very well. Fortunately, it is still possible to perform the same dispersive readout to a good approximation, but with a modified magnitude for the dispersive shift χ. As detailed in Ref. [80], we consider adding back in the two counter-rotating coupling terms to the Hamiltonian for a two-level system coupled to a resonator,

$$H_{\text{non-RWA}} = H_0 + \hbar g(\sigma_- a^\dagger + \sigma_+ a + \sigma_+ a^\dagger + \sigma_- a). \tag{3.14}$$

Using a modified unitary transformation $D = e^{\lambda(\sigma_- a^\dagger - \sigma_+ a) + \overline{\lambda}(\sigma_+ a^\dagger - \sigma_- a)}$, where $\lambda = g/(\omega_q - \omega_r)$ and $\overline{\lambda} = g/(\omega_q + \omega_r)$, we obtain a modified dispersive Hamiltonian [80]

$$H_{\text{disp.}}^{\text{non-RWA}} = -\frac{1}{2}\hbar\omega_q\sigma_z + \hbar\omega_r a^\dagger a + \frac{1}{2}\hbar g^2\left(\frac{1}{\Delta} + \frac{1}{\omega_q + \omega_r}\right)\sigma_z(a^\dagger + a)^2, \tag{3.15}$$

valid to second order in λ and $\overline{\lambda}$. In contrast to the RWA case, the coupling is now proportional to $(a^\dagger + a)^2$ as opposed to simply $a^\dagger a$, and the coupling coefficient is modified by an amount proportional to $\overline{\lambda}$. Nonetheless, we can still interpret (3.15) as a qubit state-dependent frequency shift in the classical limit: since the term $(a^\dagger + a) \propto x^2$ for

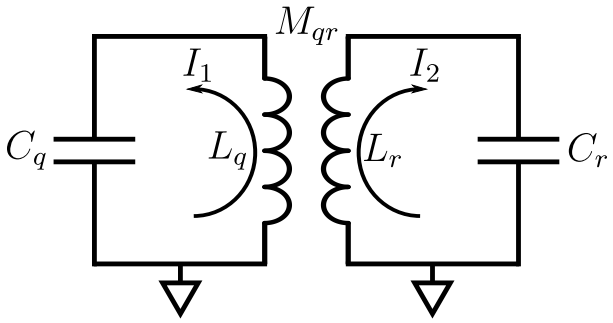


Figure 3.5: Circuit for two inductively coupled classical LC oscillators.

the harmonic oscillator with position coordinate x and mass m, when the qubit is the ground state this amounts to a well-defined shift in the "spring constant" k and therefore frequency of the harmonic oscillator according to

$$\begin{aligned}\omega_r \to \sqrt{\frac{k}{m}} &= \omega_r \sqrt{1 - \frac{2g^2}{\omega_r}\left(\frac{1}{\Delta} + \frac{1}{\omega_q + \omega_r}\right)} \\ &\approx \omega_r - g^2\left(\frac{1}{\Delta} + \frac{1}{\omega_q + \omega_r}\right). \end{aligned} \tag{3.16}$$

Including the counter-rotating terms modifies the fluxmon-induced dispersive shift by up to $\sim 50\%$.

3.2.2.2 Classical oscillator description of the dispersive shift

There are two ways to classically understand the qubit-induced dispersive shift. The first can be seen by considering the classical normal modes of two coupled LC harmonic oscillators corresponding to "qubit" and resonator, as shown in Fig. 3.5. The dispersive shift then corresponds to a slight hybridization of the otherwise uncoupled qubit and resonator modes. This assumes the qubit wells are harmonic, which is a reasonably good

assumption for the deep wells used for readout. Kirchoff's rules determine the coupled equations of motion of the system,

$$\begin{aligned}\frac{Q_1}{C_q} + L_q\frac{dI_1}{dt} + M_{qr}\frac{dI_2}{dt} = 0,\\ \frac{Q_2}{C_r} + L_r\frac{dI_2}{dt} + M_{qr}\frac{dI_1}{dt} = 0,\end{aligned} \tag{3.17}$$

which upon the substitutions $I_1 = \frac{dQ_1}{dt}$, $I_2 = \frac{dQ_2}{dt}$ yields

$$\begin{aligned}\ddot{Q}_1 + \alpha_1\ddot{Q}_2 + \omega_q^2 Q_1 = 0,\\ \ddot{Q}_2 + \alpha_2\ddot{Q}_1 + \omega_r^2 Q_2 = 0,\end{aligned} \tag{3.18}$$

where

$$\alpha_1 \equiv \frac{M_{qr}}{L_q}, \qquad \alpha_2 \equiv \frac{M_{qr}}{L_r} \tag{3.19}$$

and $\omega_q \equiv 1/\sqrt{L_q C_q}$ and $\omega_r \equiv 1/\sqrt{L_r C_r}$ as usual. Seeking eigenmode solutions of the form $Q_1 = Ae^{i\omega t}$, $Q_2 = Be^{i\omega t}$ yields the determinant equation for the eigenfrequencies ω,

$$\begin{vmatrix}\omega_q^2 - \omega^2 & -\alpha_1\omega^2\\ -\alpha_2\omega^2 & \omega_r^2 - \omega^2\end{vmatrix} = 0. \tag{3.20}$$

We can simplify the solution if we assume that ω_q is not too far detuned from ω_r, similar to the second RWA assumption in the previous section. This allows us to make the linearized substitution $\omega_{q/r} - \omega^2 = (\omega_{q/r} + \omega)(\omega_{q/r} - \omega) \approx 2\omega_{q/r}(\omega_{q/r} - \omega)$. Applying this substitution and dividing through by $\omega^2 \approx \omega_{q/r}^2$ yields the simplified determinant

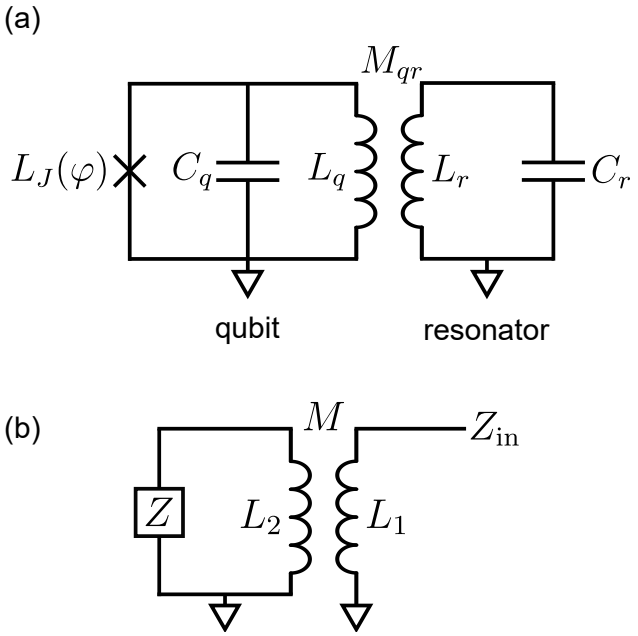


Figure 3.6: (a) Circuit for calculating the classical inductance shift induced in the readout resonator by the qubit's changing Josephson inductance. (b) Generally useful auxiliary circuit for calculating the impedance seen through a mutual inductance.

equation

$$\begin{vmatrix} \frac{2(\omega_q-\omega)}{\omega_q} & -\alpha_1 \\ -\alpha_2 & \frac{2(\omega_r-\omega)}{\omega_r} \end{vmatrix} = 0, \tag{3.21}$$

which has solutions

$$\omega_\pm = \frac{\omega_q+\omega_r}{2} \pm \frac{1}{2}\sqrt{\omega_q^2 - (2-\alpha_1\alpha_2)\omega_q\omega_r + \omega_r^2}. \tag{3.22}$$

In the dispersive limit $\Delta \gg g$, where $g = \frac{1}{2}(1-\alpha_1\alpha_2)^{-1}\sqrt{\alpha_1\alpha_2\omega_q\omega_r} \approx \frac{1}{2}\sqrt{\alpha_1\alpha_2\omega_q\omega_r}$ in the harmonic limit, we can expand (3.22) in the limit of large qubit frequency to obtain

$$\omega_- \approx \omega_r - \frac{1}{4}\frac{\alpha_1\alpha_2\omega_q\omega_r}{\omega_q-\omega_r} = \omega_r - \frac{g^2}{\Delta}. \tag{3.23}$$

Under these approximations then, the classical treatment reproduces the quantum result for the ground state-induced dispersive shift in the rotating-wave approximation.

A second way to approximately understand dispersive readout classically is by treating

the dispersive shift as the result of an effective shift in the inductance presented to the readout resonator, in which case the resonator essentially acts as a classical AC probe of the curvature of the qubit's potential [70]. To do this, for simplicity we treat the qubit's Josephson junction as a classical, linear inductor whose inductance L_J is a function of equilibrium phase bias, $L_J(\varphi)$ [given by Eq. (2.4)], as illustrated in Fig. 3.6(a). If the qubit's plasma frequency is sufficiently larger than the resonator probe frequency, the effect of the qubit is to present a φ-dependent inductance to the readout resonator. To see this, consider the input impedance Z_{in} seen by a circuit coupled through a mutual inductance to an arbitrary impedance as depicted in Fig. 3.6(b). Basic circuit analysis shows that

$$Z_{\text{in}}(\omega) = i\omega L_1 + \frac{\omega^2 M^2}{i\omega L_2 + Z}. \tag{3.24}$$

We can apply this rule to the coupled qubit-resonator circuit of Fig. 3.6(a) by asking what the input impedance is looking into the qubit from the resonator's inductor. The answer is that the qubit effectively adds an impedance of

$$\Delta Z = \frac{\omega^2 M_{qr}^2}{i\omega L_q + \frac{1}{\frac{1}{i\omega L_J(\varphi)} + i\omega C_q}} \tag{3.25}$$

to the impedance of the resonator's inductance. For large qubit junction plasma frequencies relative to the resonator probe frequency, we can ignore the effect of qubit capacitance, so that ΔZ reduces to a shift in the resonator's inductance given by

$$\Delta L_r = -\frac{M_{qr}^2}{L_q + L_J(\varphi)}. \tag{3.26}$$

To lowest order then, the resonator's resonance frequency will shift as

$$\omega_r \to \omega_r' = \frac{1}{\sqrt{(L_r + \Delta L_r)C_r}} \approx \omega_r \left(1 - \frac{\Delta L}{2L_r}\right) \approx \omega_r - \frac{2g^2}{\omega_q} \approx \omega_r - \frac{2g^2}{\Delta} \tag{3.27}$$

in the limit of large qubit frequency. This result is asymptotically the same as the quantum dispersive shift (3.16) accounting for the counter-rotating terms, as required in the limit we have assumed. Depending on whether the qubit is classically localized in its left or right well, the Josephson inductance and hence ΔL will depend on the qubit state, leading to a qubit-state-dependent frequency shift in the resonator. Of course, this analysis ignored details such as the nonlinearity of the Josephson inductance and the modified expressions for the well frequencies as opposed to the bare harmonic qubit frequency (although they are fairly close for the fluxmon at the readout point), but it illustrates the essential points.

3.2.2.3 Measuring the dispersive shift: microwave scattering

Now that we've established that the readout resonator undergoes a shift in frequency χ that serves as a pointer to the qubit state, we must consider how to actually measure this frequency shift in the lab and how to optimize the measurement signal given certain practical design constraints. At this point, we can treat the readout resonator classically, and look at how to probe frequency shifts in a classical LC resonator. The technique we will use is to weakly capacitively couple[11] the readout resonator to a $50\,\Omega$ transmission

[11] One reason we choose capacitive as opposed to inductive coupling between resonator and transmission line is so that the resonator can be made a $\lambda/4$ CPW resonator as opposed to a $\lambda/2$ resonator, which saves space on the chip for a given resonator frequency (capacitive coupling to a transmission line resonator is strongest at an open end, while inductive coupling is strongest at a shorted end).

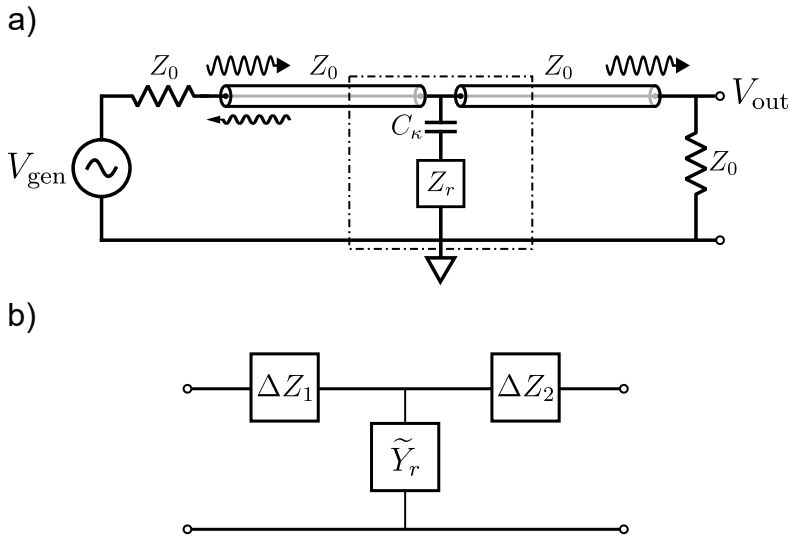


Figure 3.7: (a) Microwave scattering circuit for detection of the dispersive shift. The dashed box in the middle represents where the off-chip transmission line meets the on-chip transmission line, with the possibility of a small impedance defect where they meet as described in the text. (b) Two-port network representation of the circuit, with possible impedance mismatches ΔZ_1 and ΔZ_2 on either side of the resonator added. Source and load impedances are implicitly assumed to be Z_0.

line, and measure the scattering of a probe tone through the transmission line. This scheme has the advantage of allowing multiple qubits to be read out using a single input-output line on the chip [70, 75], and can in fact be multiplexed in frequency space (we will see this feature in action when we discuss two-qubit annealing experiments in Chapter 6).

Accordingly, we consider the microwave scattering problem drawn in Fig. 3.7(a), where we have lumped the readout resonator into the lumped element impedance Z_r of an LC resonator. Experimentally, it is convenient to measure transmission, so we want to compute the expected forward scattering coefficient S_{21} as a function of frequency. For a two-port microwave circuit, given a matched source and load impedance $Z_S = Z_L =$

$Z_0 = 50\,\Omega$, the forward scattering coefficient S_{21} is given by[12]

$$S_{21} = 2\,\frac{V_{\text{out}}}{V_{\text{gen}}}\bigg|_{Z_S=Z_L=Z_0} \tag{3.28}$$

To compute S_{21} in a convenient way, we can use some standard machinery from microwave circuit analysis called the $ABCD$ matrix.[13] We first express the circuit as in Fig. 3.7(b), where $\widetilde{Y}_r = 1/\widetilde{Z}_r = 1/(Z_{C_\kappa} + Z_r)$ is the admittance of the coupling capacitance in series with the resonator (sometimes called the "loaded" resonator admittance), and we have added ΔZ_1 and ΔZ_2 to allow for small inline impedance mismatches on either side of the resonator that can occur in a real system [39]. For example, bad microwave design of the sample mount, sloppy wirebonding, or a photolithographic defect can lead to an impedance mismatch at the interface between the chip and the coax cable going to the sample box. We compute the $ABCD$ transmission matrix parameters[14] since they have the nice property that they multiply in a cascade of two-port networks [41]. From Pozar [41], for the circuit in Fig. 3.7 we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 + Z_1\widetilde{Y}_r & Z_1 + Z_2 + Z_1 Z_2 \widetilde{Y}_r \\ \widetilde{Y}_r & 1 + Z_2\widetilde{Y}_r \end{pmatrix}, \tag{3.29}$$

[12]This is a standard definition, with the factor of 2 intuitively there so that $S_{21} = 1$ (equal voltage drop over source and load) when transferring directly from the source to a matched load. In this case, transmission is unity in the sense that there are no microwave reflections and maximum available power is transferred).

[13]The alternative is to directly compute $2V_{\text{out}}/V_{\text{gen}}$ using Kirchoff's rules, but we will find the $ABCD$ matrix method easier for predicting the effect of standing waves due to possible impedance mismatches between the on-chip and off-chip transmission lines.

[14]For a two-port network, the $ABCD$ transmission matrix is defined according to $V_1 = AV_2 + BI_2$, $I_1 = CV_2 + DI_2$ [41], and therefore multiplies in series when cascading two two-port networks together.

from which we can compute S_{21} according to [41]

$$\begin{aligned} S_{21} &= \frac{2}{A + B/Z_0 + CZ_0 + D} \\ &\approx \frac{2}{2 + |Z|e^{i\phi}\widetilde{Y}_r}, \end{aligned} \tag{3.30}$$

where

$$Z \equiv (1/(2Z_1) + 1/(2Z_2))^{-1} = |Z|e^{i\phi} \tag{3.31}$$

with $Z_1 = Z_0 + \Delta Z_1$ and $Z_2 = Z_0 + \Delta Z_2$, and for simplicity we have normalized S_{21} to be unity off resonance.

Following Ref. [81], we consider $\widetilde{Z}_r$ near resonance. The resonance frequency ω_0 of the combined series capacitor and resonator differs from the bare $\lambda/4$ resonance frequency by $\omega_0 - \omega_{\lambda/4} \approx -\omega_{\lambda/4}\sqrt{\frac{2}{\pi Q_c}}$. Here,

$$Q_c = \pi/(2Z_0^2 C_\kappa^2 \omega_{\lambda/4}^2) \tag{3.32}$$

is the coupling quality factor characterizing the rate of energy loss through the coupling capacitor assuming an ideal $Z_0 = 50\,\Omega$ environment. Allowing the readout resonator to have an internal quality factor Q_i (classically this means allowing for a large parallel resistance R_p within the LC resonator to make it an RLC circuit, with $Q_i = \omega_0 R_p C$), we have that the series impedance of the coupling capacitor and readout resonator is $\widetilde{Z}_r \approx \frac{Q_c}{2Q_i} Z_0(1+2iQ_i\delta x)$, where $\delta x = (\omega - \omega_0)/\omega_0$ is a fractional detuning. With $\widetilde{Y}_r = 1/\widetilde{Z}_r$ in hand, using (3.30) we obtain

$$S_{21} = \left(1 + \frac{Q_i}{Q_c^*} e^{i\phi} \frac{1}{1 + i2Q_i\delta x}\right)^{-1}, \tag{3.33}$$

where $Q_c^* = \frac{Z_0}{|Z|}Q_c$ is a re-scaled coupling quality factor[15] due to the impedance mismatches, agreeing with Ref. [39]. In the IQ plane, the resonance traces out a circle starting at 1 below resonance and returning to 1 above resonance, with a possible phase rotation about 1 induced by any impedance mismatches. The full expression 3.33 is useful for fitting experimental resonator data, which often contains a small phase mismatch angle ϕ. For example, Fig. 3.8 shows the measured S_{21} of a readout resonator using a vector network analyzer (VNA), along with a fit to Eq. (3.33), with extracted $f_0 = 7.87578\,\text{GHz}$, $Q_i = 270,000$, $Q_c^* = 55,000$, and $\phi = -11°$. The designed coupling capacitance C_κ to the feedline was 2.05 fF (according to finite element simulation of the chip) for a designed κ of $\sim 1/(1\,\mu\text{s})$ (the reason for this design is discussed below), which when plugged into (3.32) is consistent to within 5 percent, meaning we can predict coupling capacitance during the design stage very well.

For choosing design parameters it will be useful to gain some intuition behind Eq. (3.33). For simplicity, we ignore the impedance mismatches so that the phase offset ϕ is zero, in which case we can rewrite (3.33) in the form [81]

$$\begin{aligned} S_{21} &= \frac{S_{21}^{\min} + 2iQ_l\delta x}{1 + 2iQ_l\delta x}, \\ &= \frac{S_{21}^{\min} + 4Q_l^2(\delta x)^2}{1 + 4Q_l^2(\delta x)^2} + 2i(1 - S_{21}^{\min})\frac{Q_l\delta x}{1 + 4Q_l^2(\delta x)^2} \end{aligned} \tag{3.34}$$

where $1/Q_l = 1/Q_i + 1/Q_c$ is the total loaded quality factor and $S_{21}^{\min} = Q_c/(Q_i + Q_c)$.

[15]This interpretation assumes that ΔZ_1 and ΔZ_2 do not vary over the bandwidth of the resonator, which is a reasonable assumption unless there is another resonator of the same frequency coupled to the same transmission line, which would of course be a bad design choice.

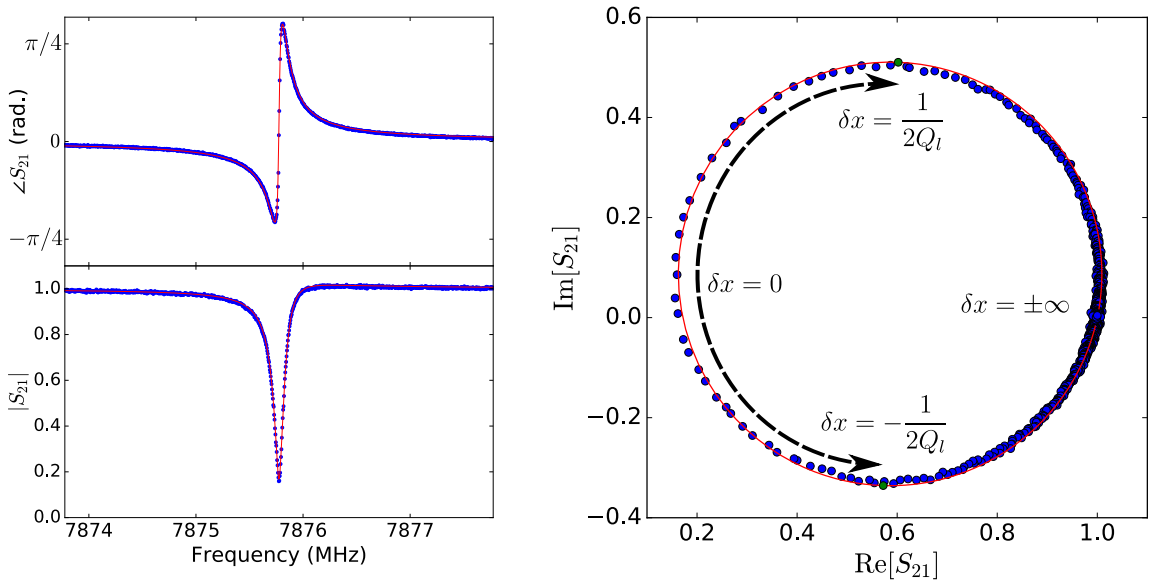


Figure 3.8: VNA measurement of S_{21} for a readout resonator at 7.876 GHz on a typical fluxmon device. The qubit is far detuned in its deep single-well regime ($\beta < -1$) during this measurement. Probe power corresponds to a mean photon number of $n_{\text{photon}} \sim 100$ in the resonator.

Therefore, the points $\delta x = \pm 1/2Q_l$ are on diametrically opposite sides of the S_{21} circle, where the imaginary part has maximum magnitude. This means that if we have the equality $\chi = \kappa$ (dispersive shift equals the resonator decay rate), where

$$\kappa = \frac{\omega_r}{Q_l}, \tag{3.35}$$

then the readout signal will be maximal for such a frequency shift. This means that for a given resonator linewidth κ, it is not necessary to have $\chi > \kappa$ since that will not increase the optimal visibility anymore. In the next section we will use these properties to come up with a set of design parameters for qubit readout.

One subtlety to be aware of is that as the on-chip transmission line gets longer (necessary for larger devices that multiplex many readout resonators at once), sloppy wirebonding with too few or too long wirebonds to the chip becomes more of an issue by causing

Q_c to vary strongly with position along the feedline due to the generation of standing waves, as detailed below. This might explain some previous anomalous Q_c dependence on position not due to stray coupling capacitance seen at one time in our group. Luckily the solution is to use low-impedance wirebonding, which requires minimal extra effort with modern automatic wirebonders, assuming that the chip mount is designed reasonably.

To see the effect of feedline position dependence, we generalize the above treatment so that in addition to the lumped-element wirebond impedances ΔZ_1 and ΔZ_2, we allow for a length of on-chip transmission line segment between the wirebonds and the resonator, with length ℓ_1 to the left and ℓ_2 to the right. Using the $ABCD$ matrix for a transmission line segment [41]

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \cos\beta\ell & iZ_0 \sin\beta\ell \\ i\frac{1}{Z_0}\sin\beta\ell & \cos\beta\ell \end{pmatrix}, \tag{3.36}$$

where $\beta = 2\pi/\lambda$ at the frequency in question, performing a similar (but messier) analysis yields S_{21} of the same form as (3.33) but with

$$Q_c^* = \left| \frac{Z_1 Z_2}{(Z_1+Z_2)Z_0} \frac{(Z_1+Z_2)Z_0\cos[\beta(\ell_1+\ell_2)] + i[Z_1Z_2+Z_0^2]\sin[\beta(\ell_1+\ell_2)]}{(Z_1\cos\beta\ell_1 + iZ_0\sin\beta\ell_1)(Z_2\cos\beta\ell_2 + iZ_0\sin\beta\ell_2)} \right| Q_c \tag{3.37}$$

As a concrete example, consider that a typical wirebond has a self-inductance of roughly 1 nH/mm and a length of $\sim$ 1 mm for an inductance of $\sim$ 1 nH, which could have a significant effect at 8 GHz (these numbers are based on a wire in free space and may be different in the proximity of a ground plane). A realistic length for a meandering on-chip feedline in our devices is $\ell = \ell_1 + \ell_2 \approx 18$ mm, which is roughly equal to 1.1λ at 8 GHz.

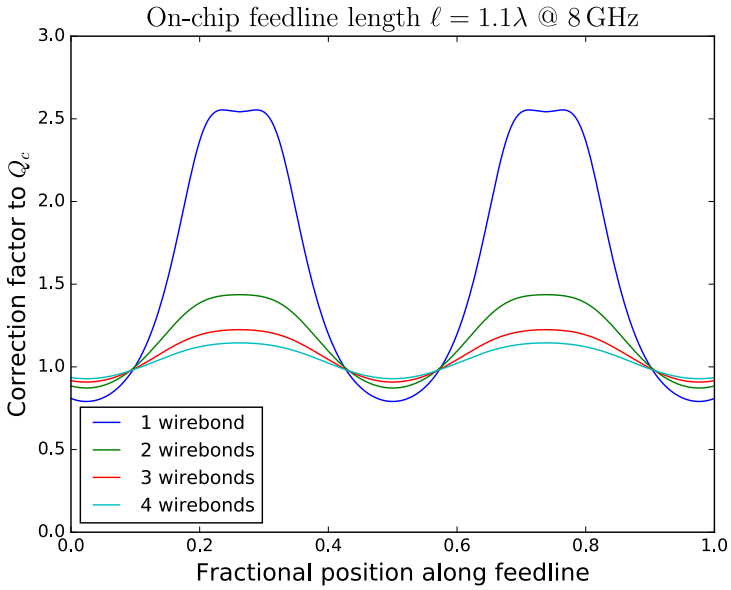


Figure 3.9: Plot showing how, for a long on-chip feedline length, the resonator's coupling quality factor Q_c can be strongly modified by standing waves due to a high-impedance wirebond connections at the chip edges, leading to variation in effective Q_c along the feedline. This plot assumes an inductance of 1 nH per wirebond. Our devices typically use 4 or 5 wirebonds in parallel for the connections to the chip pads.

In Fig. 3.9, we plot the correction factor to Q_c for such a system as a function of the position of the readout resonator along the feedline. We show predictions for 1, 2, 3, and 4 parallel wirebonds at the chip connections. As can be seen, using multiple wirebonds for low impedance chip connections is (among other microwave engineering considerations) crucial to have a controllable Q_c in a large system. As we will see in the next section, Q_c determines our ability to measure the qubit and affects how the resonator acts as a filter between the qubit and the 50 Ω environment, so this is an important consideration. We used 4 or 5 wirebonds for all devices measured in this thesis work, for example the bonds shown in Fig. 3.13.

3.2.2.4 Design parameters for dispersive readout

There are various practical constraints on the design parameters g, Δ, κ, $\delta\omega_w$. Here we find a suitable set of design parameters that obey these constraints while yielding a good readout signal that distinguishes the two qubit well states. Of course, for a good signal to noise ratio we need the qubit-induced dispersive shift to be comparable to or larger than the linewidth of the resonator itself, meaning $\chi \geq \kappa$. According to our discussion in the previous section, if we probe the resonator at $\omega = (\omega_{r,L} + \omega_{r,R})/2$, for a given resonator κ it is sufficient to achieve the equality $\chi = \kappa$ to optimally distinguish the two resonances. κ needs to be large enough so that the measurement takes a reasonable amount of time: no matter how strong the dispersive shift is, κ determines the amount of time necessary to ring up the resonator (i.e., to build up a coherent state in the resonator). In other words, we need to measure for longer than the inverse linewidth of the resonator in order to resolve the frequency shift. A measurement time of order 1 μs is convenient for the electronics in our lab and won't appreciably limit the repetition rate of experiments, so we will assume $\kappa = 1/(1\,\mu\text{s})$ and see if this allows for an acceptable readout signal within the rest of our constraints. We will assume $f_{LC} = 20\,\text{GHz}$ and $Z_{LC} \sim 100\,\Omega$ for the fluxmon. We will also assume the highest readout resonator frequency that is easily measurable with our electronics, which is $\sim 8\,\text{GHz}$. These two numbers determine the detuning $\Delta \approx 15\,\text{GHz}$ in the calculation of the dispersive shift.

Next we need to consider how to achieve $\chi = \left(\frac{g}{\Delta}\right)^2 \delta\omega_w = 1/(1\,\mu\text{s})$. There are two practical constraints on how big we allow g to be. One is that g determines the physical

length of the mutual inductive coupling segment between the qubit and resonator. We want to minimize this footprint so that as much of the fluxmon's CPW can be used for coupling to other fluxmons as possible. Let's say that the maximum allowed inductive coupling length is 5% of the fluxmon's CPW length, meaning roughly $100\,\mu$m. This corresponds to a mutual inductance of $M_{qr} \approx 10\,\text{pH}$. We must also be aware of another possible limit on g, which is the induced dissipation and dephasing in the qubit due to its coupling to the resonator. Specifically, we consider induced dissipation when the qubit is biased in the operating regime $\Delta/h \lesssim 6\,\text{GHz}$. This dissipation through the resonator is usually called the "Purcell effect" [82, 44, 83], though it can also been understood classically as a filter of the $50\,\Omega$ environment [83, 84, 85]. In the dispersive limit, the Purcell damping is given by[16]

$$\frac{1}{T_1^{\text{Purcell}}} \approx \left(\frac{g}{\Delta}\right)^2 \kappa. \tag{3.38}$$

We would like to ensure that $T_1^{\text{Purcell}} > 100\,\mu\text{s}$ at $f_{10} = 6\,\text{GHz}$. Given the κ and Δ above, this means $g/2\pi \lesssim 100\,\text{MHz}$ at $6\,\text{GHz}$, which translates to $M_{qr} \lesssim 10\,\text{pH}$. Conveniently, this happens to be the same constraint on M_{qr} as our physical length constraint above.

We also in principle need to consider dephasing between qubit energy eigenstates induced by coupling to the resonator. In particular, fluctuation in photon number ('shot noise') in the resonator could dephase the qubit due to a fluctuating ac Stark shift in the qubit's energy level spacing [72]. As studied in Refs. [86, 87], for a thermal state in the resonator, photon shot noise will dephase the qubit with roughly a white[17] noise spectrum,

[16] Intuitively, the Purcell decay rate induced in the qubit is the resonator's decay rate times the fraction of the first excited eigenstate of the joint system that "lives" in the qubit.

[17] Really, it is a Lorentzian noise spectrum with cutoff frequency determined by the resonator's decay

leading to an roughly exponential decay of the off-diagonal terms in the density matrix over time. For an effective temperature $T_{\rm eff}$ in the resonator, the induced dephasing rate between qubit eigenstates is predicted to be [87]

$$\frac{1}{T_{\varphi 1}^{\rm th}} \approx \frac{\kappa}{2} \mathrm{Re}\left[\sqrt{\left(1+\frac{2i\chi}{\kappa}\right)^2+\left(\frac{8i\chi n_{\rm th}^{\rm eff}}{\kappa}\right)}-1\right], \tag{3.39}$$

where $n_{\rm th} = 1/(\exp[\hbar\omega_r/k_B T_{\rm eff}]-1)$ is the stray thermal photon number in the resonator. Effective resonator temperatures of up to 60 mK are often observed,[18] which at zero tilt would lead to a dephasing time of $\sim 200\,\mu$s in the qubit, vastly longer than the limit of ~ 100 ns due to intrinsic flux noise in the qubit.[19]

Assuming the acceptable $g/2\pi \approx 100$ MHz, we now need to see how large β and $\delta\omega_w$ (i.e., the post-tilt in the readout sequence of Fig. 3.3) need to be to obtain a large enough dispersive shift χ. There is a limit on how large the post-tilt can be (and thus how large $\delta\omega_w$ can be) because the shallow well will eventually become so shallow that its metastable ground state will tunnel out of it and into the deeper well. As a rule of thumb, we require that this tunneling time should be no faster than ~ 1 ms ($\sim$ 1000x

rate [88], which looks white at low frequencies.

[18]This number comes from indirect measurements of the steady state resonator population using similarly fabricated Xmon qubit devices that use dispersive readout. Similarly "hot" resonators are also seen in state of the art devices from other groups [62]. The reasons for these temperatures, which are substantially higher than the dilution refrigerator mix plate temperatures, are not completely understood, and may involve bad thermalization of dissipative components and/or radiation from the output lines of the chip.

[19]It is not completely fair to compare these two dephasing times, since one is a true dephasing rate whereas the other is a Gaussian instead of exponential decay of the coherence with time. In reality, we should compare the noise spectral density of the two noise sources over all frequencies. Even when we do this, photon shot noise is negligible compared to the $1/f$ flux noise at all frequencies (including all the way up to the resonator decay bandwidth of ~ 1 MHz). In other types of qubits, such as the C-shunted flux qubit with a non-tunable gap [62], photon shot noise can actually become a limit on dephasing at zero tilt at the level of $\sim 100\,\mu$s dephasing times, though this is unlikely to be of practical importance for quantum annealing.

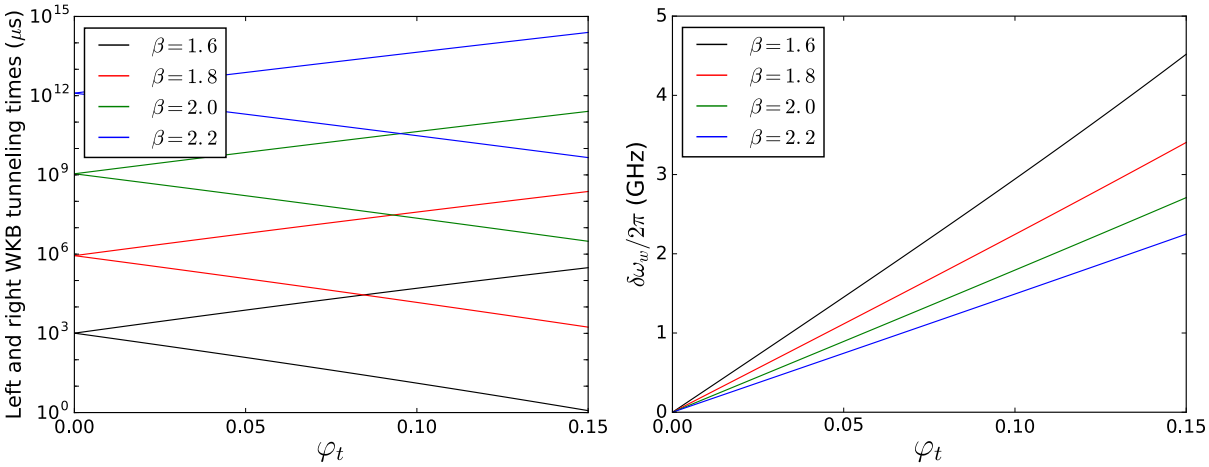


Figure 3.10: (a) Tunneling times out of the shallow and deep wells according to a numerical generalization of the WKB result (2.55), as a function of β and tilt flux bias, for a fluxmon with $f_{LC} = 20\,\text{GHz}$ and $Z_{LC} = 100\,\Omega$. (b) Difference in left and right well frequencies, $\delta\omega_w$ as a function of β and tilt flux bias.

longer than the measurement time), so that a negligible amount of population is lost from the shallow to the deep well during the readout pulse. In Fig. 3.10(a) we plot the predicted WKB tunneling times[20] out of the left and right wells as a function of φ_x, for different values of β. In order to have $\chi = \kappa$ with the given values of g and Δ, we would like $\delta\omega_w/2\pi \approx 3\,\text{GHz}$. Looking at Fig. 3.10(b), we see that this is possible while also having the tunneling time out of the shallow well be longer than 1 ms if we have $|\varphi_x| \approx 0.15$ with $\beta \gtrsim 1.8$. Note that the tunneling times will depend exponentially on Z_{LC} in addition to β, so that the actual optimal readout parameters will highly depend on the actual parameters of the system.

[20]The actual tunneling times are likely to be much longer since a real system has dissipation, which can strongly suppress the tunneling rate by keeping the wavefunction (more precisely, the density matrix) localized in one well [89, 90, 91]. In fact, even if the ground state of the upper well is resonant with an excited state within the lower well, the resonant tunneling rate will be largely suppressed because the tunneling will be incoherent [92] (we will study this effect in chapter 4). Nevertheless, to have a conservative design we compute and plot predicting tunneling rate out of left and right well based on a naive WKB calculation.

Note that in our discussion, we did not consider the possibility of transitions out of the shallow well induced by the off-resonant photons of the resonator drive itself. Such effects are difficult to predict and model, but in the next section we will see empirically that while these effects are certainly present at large enough tilts, we can find a suitable range of resonator drive powers and post-tilt biases that allow faithful readout without these effects.

3.3 Qubit measurement: experiment

In this section we summarize some of the basic procedures involved in bringing up single-qubit measurement and calibration. In particular we will see data illustrating the double-well projection measurement described in section 3.2.1 and the qubit state-dependent dispersive shfit described in section 3.2.2, as well as the calibration of qubit frequency vs. flux biases and calibration of nonlinear crosstalk due to junction asymmetry. We will also look at single-qubit annealing, (which we use as a calibration tool for multi-qubit devices in Chapter 6) in detail as a case study.

3.3.1 Experimental setup

A schematic diagram of the wiring used for qubit measurements is summarized in Fig. 3.11, and physical photographs of the setup inside the fridge (a Janis model JDry-500-QPro ^{3}He - ^{4}He cryogen-free dilution refrigerator) are shown in Fig. 3.12. Each qubit has two coaxial bias lines entering the aluminum sample box, each of which is wirebonded to

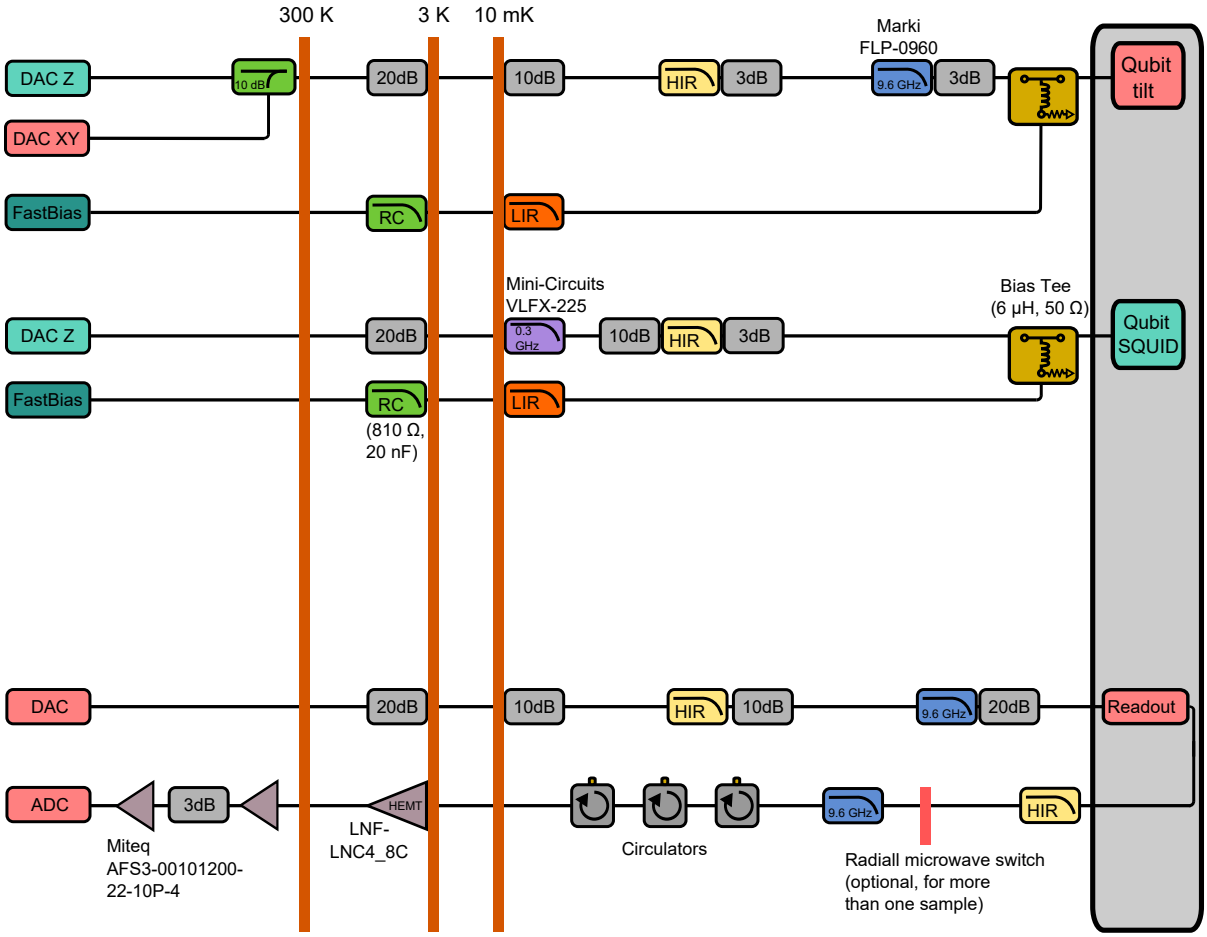


Figure 3.11: Wiring diagram for single-qubit measurement, illustrating the number of coaxial lines per qubit in the fridge and attenuation/filtering components in each of the lines as described in the text. A single readout line is used to multiplex measurement of many qubits in one microwave line. LIR and HIR stand for high- and low-frequency infrared filters [93]. The tilt bias line has microwave signals coupled into it at room temperature.

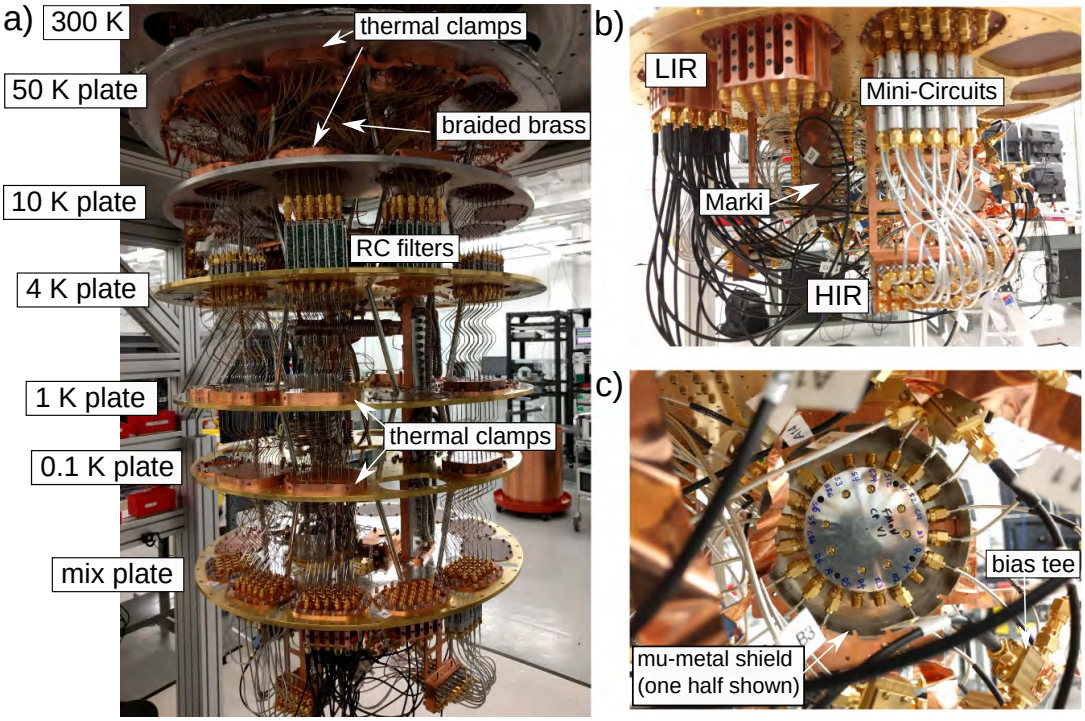


Figure 3.12: (a) All cables between room temperature and sample are coaxially shielded. On the microwave and fast flux bias lines, we use SC-086/50-CN-CN from Coax Co. between 300 K and 4 K, and SC-086/50-NbTi-CN (niobium-titanium inner conductor, copper-nickel outer) between 4 K and the mix plate. On the DC lines, we use braided brass coaxial cables between 300 K and 4 K. (b) Picture showing various components mounted on the mix plate. (c) Aluminum sample box mounted on hanging plate at 10 mK visible in center. Two-piece Amumetal 4K "mu-metal" shield encloses qubit box (only half of shield is shown in image). Nonmagnetic SMA connectors are used inside this shield.

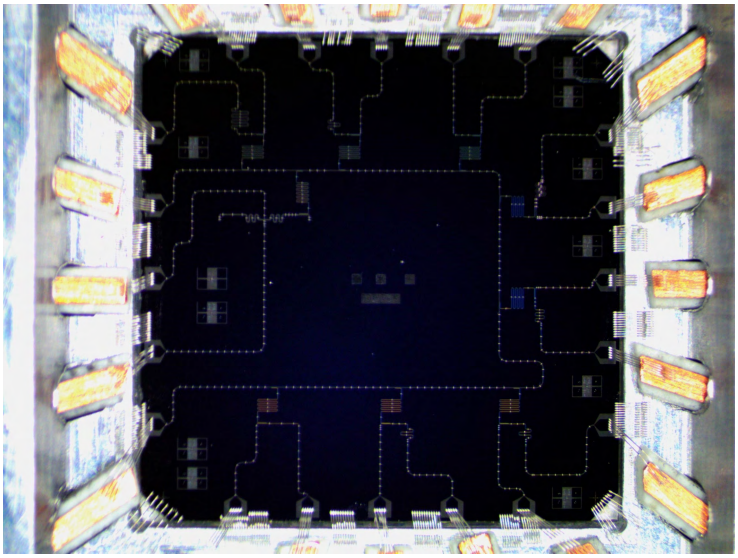


Figure 3.13: $1 \times 1\,\mathrm{cm}^2$ die wirebonded into an aluminum box with recessed cavity above and below (lid not shown). Copper traces on a PCB strip enter the box through tunnels, and are wirebonded to the on-chip bond pads using several wirebonds in parallel (here, four or five). Ground planes of the chip are wirebonded to the box ground around the edge of the chip in between the bond pads as uniformly as possible to provide a good microwave path to ground, and to help reduce crosstalk and spurious chip modes [94].

a CPW line on the chip for applying SQUID and tilt flux bias currents through the on-chip mutual inductances depicted in Fig. 3.1. Each of these lines receives the combined signal of a DC bias current and a higher bandwidth bias current through a bias tee at the mix plate of the fridge. The signal for the DC bias lines is provided by a low noise DC voltage bias source, which is converted to a current source by a circuit board RC filter at 4 K. For these DC lines, at the mix plate there is a microwave- and infrared-absorbing ECCOSORB® filter (with a cutoff of a few hundred MHz) [93] to protect from inline IR radiation coming from 4K and above [95]. The DC bias lines then go through the inductive port of a bias tee where they combine with their corresponding higher bandwidth line before entering the qubit sample box. The higher bandwidth line

on the SQUID bias has a bandwidth of $\sim 100\,\mathrm{MHz}$, and is controlled by a GHz DAC waveform generator [96] that is low-pass filtered to $\sim 100\,\mathrm{MHz}$. For the tilt bias line, the high bandwidth line is even wider ($\sim 10\,\mathrm{GHz}$ bandwidth) so that microwave pulses can be applied to drive qubit transitions. The envelopes for these microwave pulses are also controlled by a GHz DAC waveform generator, via IQ mixing [96, 97]. The microwave drive is coupled into this line at room temperature via a directional coupler, where is is combined with the signal from another GHz DAC that is low-pass filtered to $\sim 100\,\mathrm{MHz}$. On both SQUID and tilt lines, the higher bandwidth lines also have an IR filter, but with a much higher cutoff frequency ($\gtrsim 20\,\mathrm{GHz}$, to avoid loss at the signal frequencies). The amount of attenuation and filtering on each of the coaxial lines within the fridge is chosen according to various tradeoffs between having enough range in the electronics to flux bias and drive the qubits, and having low enough noise entering the qubit chip (we will go through calculations of this noise in Chapter 4).

3.3.2 Single-qubit bringup

When a sample is cooled down, we do not know a priori the dc transfer function between applied bias currents and received fluxes in the qubit's SQUID and tilt loops. In order to calibrate the on-chip mutual inductances and any flux offsets due to ambient magnetic fields during the cooldown, we can probe the response of the readout resonator vs. the applied qubit bias currents. An example of this data is shown in Fig. 3.14(b) (which we colloquially call a "fireball scan"), where we probe the readout resonator at 1.5 MHz below

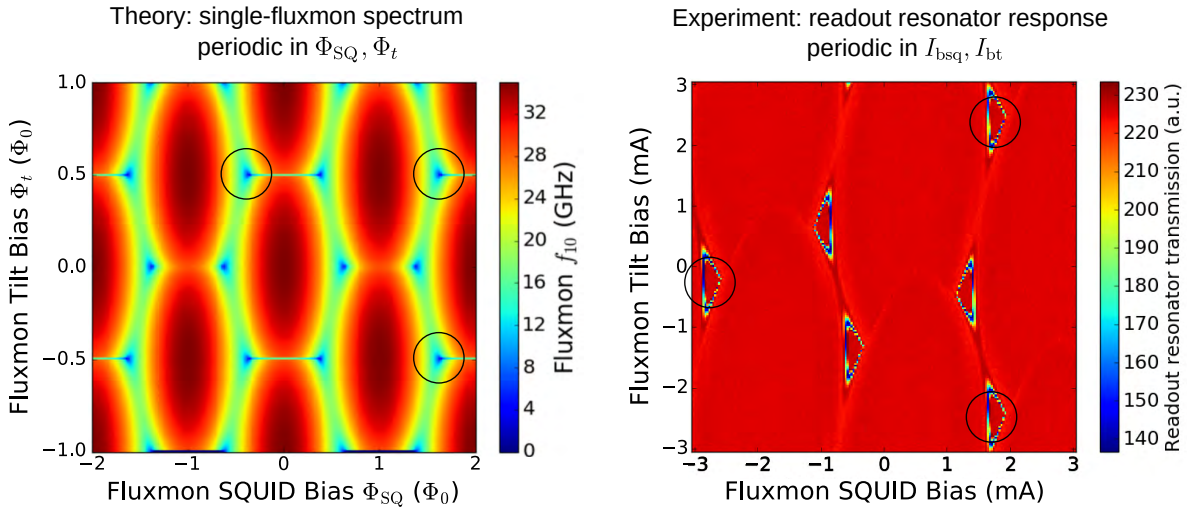


Figure 3.14: (a) Theoretical plot showing periodicity of fluxmon spectrum with respect to its flux biases, which is to be mapped to experimental data. (b) Experimental data from a "fireball scan" used to obtain the intra-qubit mutual inductance matrix and flux offsets. The qubit's readout resonator is probed at a fixed frequency (corresponding to roughly 1 MHz below f_r at zero qubit flux bias, where the qubit's f_{10} is largest). In this device there was substantial geometric crosstalk from squid bias to qubit tilt loop. Extracted parameters were $M_t = 0.42\,\text{pH}$, $M_{\rm SQ} = 0.92\,\text{pH}$, $M' = -0.24\,\text{pH}$, $M'' = 0.0005\,\text{pH}$.

its frequency at nominally zero flux bias. The shift in resonance frequency corresponds to the "Lamb shift" part of the dispersive shift with the qubit in its ground state. As we saw in Fig. 2.7, the qubit spectrum is periodic in its flux biases, meaning that so will be the response of the readout resonator.[21] By mapping the periodicity and offsets of these features in the data to theory using an affine linear transformation, we can extract the qubit bias mutual inductances $M_{\rm SQ}$ and M_t, as well as any flux offsets $\Delta\Phi_{\rm SQ}$ and $\Delta\Phi_t$. Explicitly, we have

$$\begin{pmatrix}\Phi_{\rm qsq}\\ \Phi_{\rm qt}\end{pmatrix} = \begin{pmatrix}M_{\rm bqsq\to qsq} & M_{\rm bqt\to qsq}\\ M_{\rm bqsq\to qt} & M_{\rm bqt\to qt}\end{pmatrix}\begin{pmatrix}I_{\rm bqsq}\\ I_{\rm bqt}\end{pmatrix} + \begin{pmatrix}\Phi^{\rm offset}_{\rm qsq}\\ \Phi^{\rm offset}_{\rm qt}\end{pmatrix} \tag{3.40}$$

[21] The faint, upward-pointing "McDonald's arch"-like features are a hysteretic effect that occurs in the bistable regime when the qubit's potential is tilted so far that it dumps the qubit population from one metastable well to the other. The point in the opposite direction when the scan vs. tilt bias is taken in the opposite direction.

where when talking about a single qubit we will often use the shorthand terminology $M_t \equiv M_{\mathrm{bqt}\to\mathrm{qt}}$, $M_{\mathrm{SQ}} \equiv M_{\mathrm{bqsq}\to\mathrm{qsq}}$, $M' \equiv M_{\mathrm{bqsq}\to\mathrm{qt}}$ and $M'' \equiv M_{\mathrm{bqt}\to\mathrm{qsq}}$. The particular data shown here is for a qubit within our first iteration of a coupled qubit device, and so has a relatively high M' of $-0.24\,\mathrm{pH}$. Ideally we would like to have $M' \lesssim 0.1\,\mathrm{pH}$, which is about the best we can predict using finite element simulations of the chip design. This number has since been improved to $M' \lesssim 0.1\,\mathrm{pH}$ (data not shown) within a two-qubit device.

We would like to dc bias the qubit at zero tilt (for optimal dephasing) somewhere in the range of $\Delta = 4-6\,\mathrm{GHz}$, as this is an ideal frequency range for applying microwave pulses for qubit characterization. This range of biases will be inside one of the "fireballs," where f_{10} is below the readout resonator frequency $f_r \approx 8\,\mathrm{GHz}$. Once a dc bias point is chosen, it will be fixed for the duration of all experiments, and the faster, 100 MHz bandwidth "z" lines are used to change the qubit bias from its dc point during experiments. In order to find a proper DC bias point, we must first bring up the qubit readout. To do this, we need to calibrate the dispersive shift of the resonator when the qubit is in the left and right wells in the double-well regime. One way to do this is as follows. First, we estimate where $\beta \approx 2$ at zero tilt (which occurs outside the fireball, in the direction in which the fireball "points"[22]). We then prepare the qubit in either the left or right well and probe the readout resonator transmission vs. frequency, obtaining the two resonance curves shown in Fig 3.15. To prepare one well or the other, we start by initializing in the

[22]The reason it points towards the bistable regime is because the qubit frequency becomes a narrower and narrower hyperbola vs. tilt bias as β increases.

single-well regime, followed by a pulse on the qubit's tilt bias line to tilt in one direction or the other, after which we raise the barrier and allow for dissipation to localize the qubit at the ground state of the corresponding well. For this particular data we used a readout probe power of $\sim -120\,\mathrm{dBm}$ at the chip feedline, corresponding to an average number of $\langle n_{\mathrm{photon}} \rangle \sim 100$ photons in the resonator.[23] This data is for the same resonator that was measured by VNA in the previous section (Fig. 3.8).

The particular measurement in this figure was performed with $\beta = 1.9 \pm 0.1$ at a post-tilt bias of $\Phi_t \approx 0.3\,\Phi_0$. With these parameters, numerics predict that $g/(2\pi) = 62\,\mathrm{MHz}$ here, with $\delta\omega_w/(2\pi) \approx 4.1\,\mathrm{GHz}$ and $\Delta/(2\pi) \approx 14\,\mathrm{GHz}$, so according to eq. (3.8) we should expect $\chi/(2\pi) \approx 0.08\,\mathrm{MHz}$. Experimentally, we observe $\chi/(2\pi) = 0.12\,\mathrm{MHz}$, within a factor of two of the prediction. This number depends strongly on β, so this is within the range of experimental uncertainty. We note that the observed $\chi = 2\pi \times 0.12\,\mathrm{MHz}$ is also fairly close to the desired design value of $\chi = \kappa \approx 1/(1\,\mu\mathrm{s})$ as per the discussion in the previous section, meaning that a near-optimal readout signal should be possible.

Next, we need to fix the resonator probe frequency and vary the post-tilt bias in order to maximize the difference in signals between left and right wells, ultimately allowing for single-shot readout of the qubit state. An example of such data, which we refer to as "readout branch" data, is shown in Fig. 3.16. In this experiment, we again prepare the

[23] This number is an estimate determined as follows, using the estimated attenuation of the coaxial lines in the fridge. If the $Z_0 = 50\,\Omega$ feedline is probed with microwave power P_{probe}, then it can be shown that the steady state power of the traveling wave within the $\lambda/4$ resonator is given by $P_{\mathrm{internal}}/P_{\mathrm{probe}} \approx \frac{2}{\pi}\frac{Q_l^2}{Q_c}$ [98], meaning the rms standing wave voltage at the resonator's voltage antinode is $V_r = 2\sqrt{Z_0 P_{\mathrm{internal}}}$ so that the total energy in the resonator is $\frac{1}{2}\mathcal{C}_r V_r^2 \ell_r$. Dividing this energy by $\hbar\omega_r$ defines the mean photon number $\langle n \rangle$ in the resonator. This same estimate is used for the resonator measurements in Refs. [39, 99, 59] by our group. Note that quantum-mechanically, for a coherent state $|\alpha\rangle$ in the resonator, we will have $\langle n \rangle = |\alpha|^2$ with a quantum uncertainty of $\Delta n = \sqrt{\langle n \rangle}$ consistent with Poisson statistics.

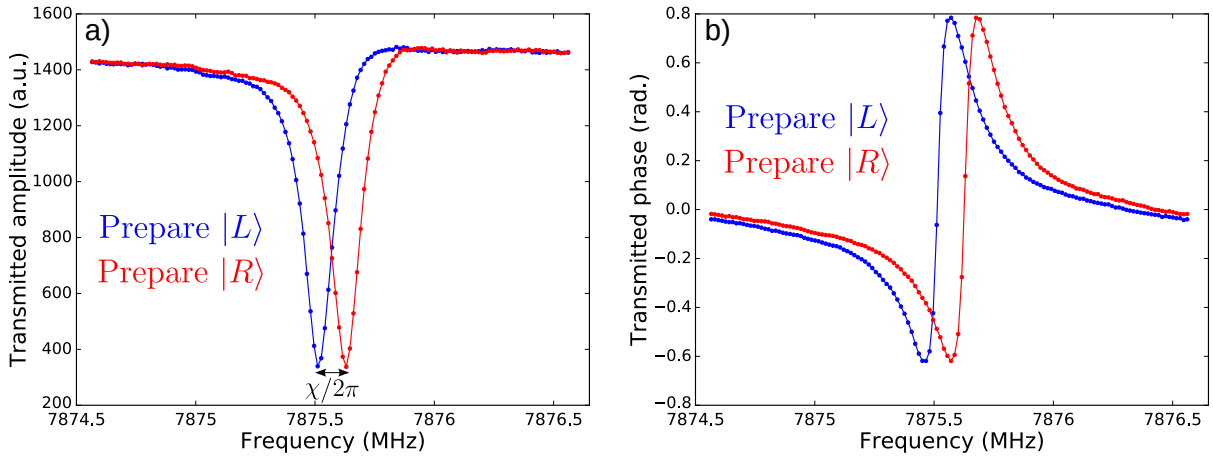


Figure 3.15: Readout resonator resonance curves with qubit prepared in left and right wells ($\beta \approx 1.9$, $\Phi_t \approx 0.3\,\Phi_0$). Readout pulse length was $5\,\mu$s with an on-chip power of $-120\,$dBm (~ 100 steady-state intra-resonator photons).

qubit in either its left or right well in the double-well regime, but now probe the readout resonator at a fixed frequency. The figures show all the single-shot datapoints, without any averaging over experimental runs. We can see that there is a range of post-tilt biases for which there is perfect separation between the distributions of measured amplitude or phase corresponding to the two qubit states. This indicates both very high fidelity in preparing the left/right well states and very high fidelity in distinguishing these two states. This can be seen explicitly in Fig. 3.18(a), where we plot the raw statistics of the readout in the IQ plane for preparing the left vs. right well. To calibrate single-shot measurement, we draw a line dividing the plane midway between the midpoints of these two clouds, calling the result either $|L\rangle$ or $|R\rangle$. Both the $|L\rangle/|R\rangle$ state preparation fidelities and separation fidelity (the probability of correctly identifying a given prepared state) of the Gaussian clouds in the IQ plane are very high (can be made 99.9 or even higher with modest effort). Note in particular that this shows we are able to faithfully

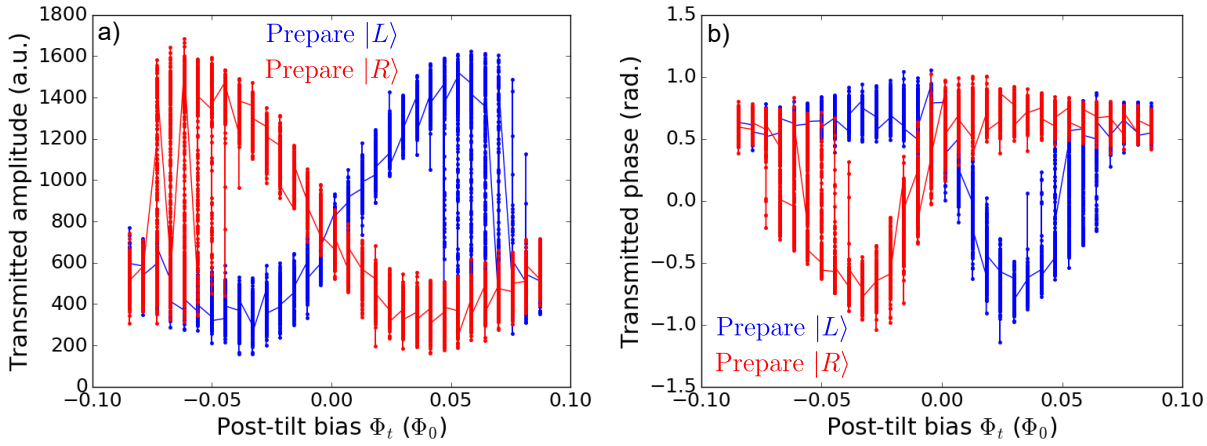


Figure 3.16: Experimental data for the readout branch experiment of resonator response vs. prepared well, showing individual stats without averaging over experiments (300 stats for each of $|L\rangle$ and $|R\rangle$). $\beta \approx 1.9, \Phi_t \approx 0.3\,\Phi_0$. At large post-tilts, errors due to tunneling from the higher, metastable well to the lower well are increasingly present. At even larger post-tilts, both states end up in the same well. (a) Looking at transmitted amplitude while probing readout resonator midway between its two resonances, i.e. at $\left(f_{r,|L\rangle} + f_{r,|R\rangle}\right)/2$. (b) Looking at transmitted phase while probing readout resonator at $f_{r,|R\rangle}$.

prepare, in of order one microsecond or less, the classical Ising states that usually occur at the end of standard quantum annealing. This means that our system should be capable of running annealing in reverse for quantum parallel tempering, as was discussed in Chapter 1.

Now that we know what barrier and post-tilt bias to use for distinguishing the left and right well states, we can try and implement the double-well projection scheme for reading out the $|0\rangle$ and $|1\rangle$ eigenstates at the dc bias point (which is in the single-well regime). There are a few more prerequisite steps before we can do measurements like Rabi oscillations between these two states. The first is to perform what we call an "S curve" measurement with the qubit nominally prepared in its ground state. Namely, we perform the double-well projection sequence of Fig. 3.3 starting in the $|0\rangle$ state, as

a function of pre-tilt bias, measuring the probability of ending up in the left or right well. This yields a typical dataset shown by the blue S-shaped curve in Fig. 3.17. At zero tilt, the probability of ending up in the left or right well is $\frac{1}{2}$. Far from zero tilt, there is a very high probability of ending up in that respective well. The width of the S-curve is set by a combination of ramp rate (dictating both nonadiabatic effects and thermalization along the trajectory) and flux noise (broadening), as discussed later in section 3.4 in the context of single-qubit annealing. Note that in the data one can see a region within $\sim 0.003\,\Phi_0$ of zero tilt in which the probability of ending up the desired well differs from unity by a percent or more, leading to a 'step' in the S-curve.[24] We attribute this step to imperfect ground state preparation due to a finite steady-state stray qubit population; i.e., we actually started out with some population of $|1\rangle$. We will study the steady-state qubit population and its relation to thermal occupation and nonequilibrium noise in Chapter 4.

Within the stray population step, the $|0\rangle$ and $|1\rangle$ states end up in different wells. But when the tilt is large enough, both states end up in the same well. This can be seen more explicitly by looking at what happens when we prepare the $|1\rangle$ state with a pi pulse before performing the readout sequence. A pi pulse is simply a microwave

[24] Accounting for the precise position of this step is subtle. Using the design parameters of the system (the fluxmon's f_{LC} and Z_{LC}, which are verified to reasonable accuracy through the measured qubit spectrum below 10 GHz), using the 1D model (2.32) one would predict that the edge of this step should occur at $\Phi_t = 0.0038\,\Phi_0$, rather than the (consistently) observed $0.0030\,\Phi_0$. This discrepancy is not accounted for by including the inline SQUID inductances in the 1D model (in fact, this shifts things in the wrong direction). Interestingly, using the fluxmon's quarterwave CPW L and C instead of the "fluxmon" L and C (closer to the effective L and C one would obtain from the full transmission line fluxmon model at high frequencies [Fig. 2.13]), one would predict the boundary at $0.0029\,\Phi_0$, much closer to the observed value. Unfortunately, we don't yet have a way to deal with the full transmission line model in the double-well regime, so this cannot be theoretically verified as of yet. We also note that the location of this tipping point is fairly independent of f_{LC} and mostly depends on Z_{LC}.

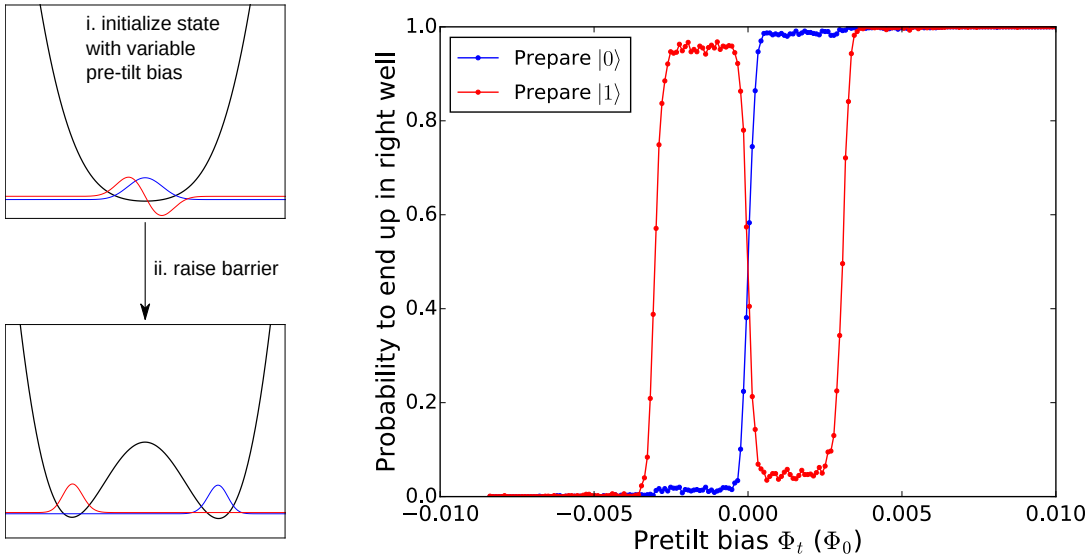


Figure 3.17: S curve data without and with pi pulse (i.e., preparing $|0\rangle$ and $|1\rangle$) to calibrate the pre-tilt part of the double-well projection process. Ramp time for barrier raising pulse was 10 ns, and pi pulse length was 10 ns. f_{10} at the dc bias point used for state preparation was 6 GHz. Steady-state stray population is $\sim 1.3\,\%$.

pulse applied to the qubit, resonant with the qubit's $|0\rangle \rightarrow |1\rangle$ transition frequency f_{10}, with an amplitude and length such that the qubit ends up in the excited state at the end of the pulse.[25] The result of performing the S-curve experiment immediately after a pi pulse is the red curve of Fig. 3.17. By choosing a pre-tilt in the middle of this step, we can faithfully distinguish between the $|0\rangle$ and $|1\rangle$ states using the double-well projection sequence.[26] With the optimal pre-tilt chosen, we can look at the fidelity of the readout process for both states. Fig. 3.18(b) shows the raw IQ statistics for preparing and measuring both states. We can see that there are some red dots in the blue cloud and

[25]This is called a "pi pulse" because it rotates the qubit's zenith angle θ by π in the Bloch sphere representation of the qubit state, $|\psi\rangle = \cos\left(\frac{\theta}{2}\right) + e^{i\phi}\sin\left(\frac{\theta}{2}\right)$. It can be calibrated by looking at Rabi oscillations of the qubit population vs. pulse frequency and amplitude [100, 97].

[26]Unfortunately, simultaneously distinguishing the $|2\rangle$ state as well (which would be useful for some experiments) is not feasible with this method, since the $|2\rangle$ state will end up in the same well as the $|0\rangle$ state when the barrier is raised and will be indistinguishable from the ground state (the intra-well dispersive shift is not large enough to distinguish the intra-well states, and even if it was, dissipation would very quickly collapse all intra-well states into its respective well ground state).

vice versa. We attribute this to state preparation error. There is a $\sim 1-2\,\%$ steady state stray population error leading to error in preparing $|0\rangle$. There is an even larger infidelity in preparing the $|1\rangle$ state, which we attribute to two sources: dephasing on the timescale of the pi pulse length,[27] and energy relaxation during the first stages of the double well projection trajectory, each of which are estimated to contribute of order a few percent of error. As mentioned earlier, we do not expect any appreciable nonadiabatic error during the projection sequence once a near-optimal pre-tilt is chosen.

After calibrating the readout process, we can begin to look at the functional form of the qubit's spectrum vs. flux bias. Such spectroscopic measurements provide information about the qubit's excited state(s) and are crucial for device calibration in the region where f_{10} and f_{21} are in the range of $\sim 0.5 - 10\,\mathrm{GHz}$ at degeneracy (we will sometimes refer to this region as the "microwave regime"). An example dataset showing spectroscopy vs. tilt flux bias is shown in Fig. 3.19, for two different values of SQUID flux bias corresponding to $\Delta/h = 5\,\mathrm{GHz}$ and $\Delta/h = 1\,\mathrm{GHz}$. For spectroscopy experiments, a microwave tone is applied to the qubit as a function of frequency and flux bias, followed

[27]The ensemble-averaged pi-pulse infidelity arising from gaussian frequency noise is calculated as follows. For a driving Hamiltonian of the form $H/\hbar = \frac{\omega_q}{2}Z + g\cos(\omega_d t)X$ in the lab frame (which is appropriate for an inductive drive), by looking at the rotating frame of the qubit $\left|\tilde{\psi}(t)\right\rangle = e^{i\omega t Z/2}\left|\psi(t)\right\rangle$ one can show that the time evolution operator for $\left|\tilde{\psi}(t)\right\rangle$ in this frame is $U(t) = e^{i\left[\frac{\omega_q-\omega_d}{2}Z+2\pi f_R X\right]t}$ [100], where f_R is the on-resonance Rabi frequency. In particular, for a perfectly resonant drive, a pulse length of $\tau = 1/4f_r$ would form a perfect pi pulse, meaning if we started in $|0\rangle$ we would end up in $|1\rangle$ with probability 1. If we calibrate our pi pulse to have the correct length for a nominally perfect on-resonance pi pulse, and the qubit then drifts in frequency by the amount $\delta\omega$, then the probability of ending up in $|1\rangle$ when applying the nominal pi pulse after the drift will be $P_{|1\rangle}(\delta\omega) = \frac{16\pi^2 f_r^2 \sin^2\left[\sqrt{16\pi^2 f_r^2+\delta\omega^2}/8 f_r\right]}{16\pi^2 f_r^2+\delta\omega^2}$. If the qubit frequency drift follows a Gaussian distribution with standard deviation σ, then the ensemble-averaged probability of the qubit ending up in the $|1\rangle$ state is $\int d(\delta\omega)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\delta\omega)^2}{2\sigma^2}}P_{|1\rangle}(\delta\omega)$. For a pi pulse length of 8 ns and $\sigma/2\pi \approx 10\,\mathrm{MHz}$ (obtained from CW spectroscopy, or equivalently from $T_{\varphi 2} \approx 30\,\mathrm{ns}$, at the zero tilt dc bias point) we would expect an ensemble-averaged pi pulse infidelity of 2 - 3 percent.

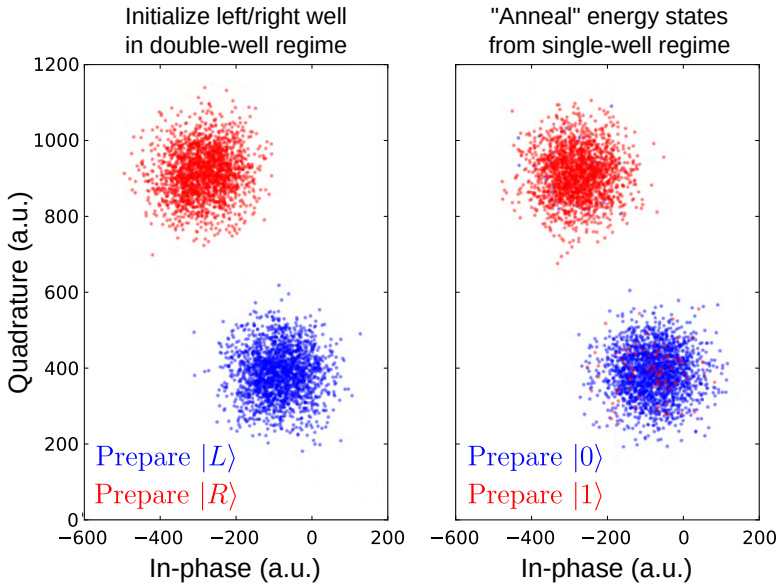


Figure 3.18: Raw single-shot IQ demodulations ("readout cloud" data) showing 2000 stats each of preparing each state. (a) Preparing the left and right well states $|L\rangle$ and $|R\rangle$ directly in the double-well regime using flux pulses. (b) Preparing the $|0\rangle/|1\rangle$ states at the single-well dc bias point without/with a pi pulse and performing the projection sequence. Measurement infidelity is due to state initialization errors. For this dataset, the total $|0\rangle$ state fidelity was 98.5 %, while the total $|1\rangle$ state fidelity was 95.5 %, both limited by state preparation.

by the readout sequence. The result is that there will be peaks in qubit population where the system is able to absorb energy from the spectroscopy tone, corresponding (usually) to the qubit's f_{10} transition frequency. In the data of Fig. 3.19, we see the characteristic hyperbola shape, corresponding to the theory curves plotted earlier in 2.10.

During device operation, we often need to be able to remain at zero tilt bias (or constant tilt flux bias) as one changes the DC SQUID bias. This is equivalent to the calibration of the total intra-qubit crosstalk from SQUID to tilt (a combination of linear and nonlinear crosstalk). For example, as described in the next chapter, it is useful to measure T_1 (the qubit's $|1\rangle \rightarrow |0\rangle$ energy relaxation time) at zero tilt as a function of $\Delta(\Phi_{\mathrm{SQ}})$. In addition and more generally, such a calibration needs to be done in order

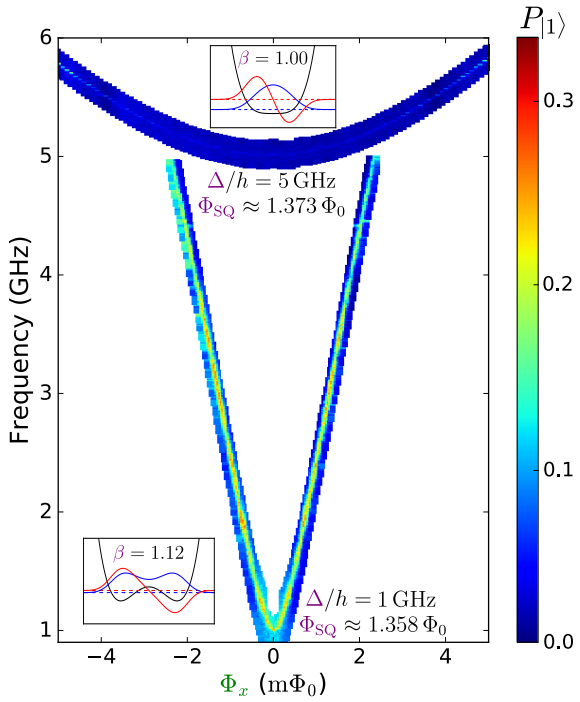


Figure 3.19: Qubit spectroscopy vs. tilt bias for two different values of $\Delta(\Phi_{\mathrm{SQ}})$. A microwave probe of varying frequency is applied to the qubit's driveline and the qubit's tilt bias is stepped.

to be able to perform quantum annealing with quasi-orthogonal controls. The most immediately obvious way to calibrate zero tilt is to measure spectroscopy vs. tilt bias just like we did in Fig. 3.19. The applied tilt flux that gives us the minimum f_{10} can be then identified with zero tilt, and we can repeat this measurement for many different values of Φ_{SQ}. The problem with doing the zero-tilt calibration this way is that it is very costly – it is a three-dimensional scan since we need to scan versus frequency versus tilt bias versus SQUID bias, and we would therefore like a more scalable measurement. A second way to do this calibration reduces the dimensionality of the scan by one dimension: instead of scanning the spectroscopy tone's frequency, we can keep a fixed probe frequency as we step the tilt bias. The result should be that we obtain a signal when the qubit's frequency

is resonant with the probe frequency. As a function of tilt bias, this means we should expect two spectroscopy peaks on either side of zero tilt. Assuming that the crosstalk from the qubit's tilt bias to the SQUID is negligible (as we always observe to be the case, thanks to the high symmetry of the bias line and smallness/remoteness of the SQUID loop), the location of zero tilt is then simply the midpoint of these two peaks. Performing this fixed-frequency probe scan vs. SQUID and tilt biases should then ideally give us zero tilt vs. SQUID bias. This method works fairly well, but only for $\Delta/h \gtrsim 500\,\mathrm{MHz}$. For smaller gaps, it becomes quite hard to drive the qubit when it is away from zero tilt due to the rotation of the energy basis relative to the axis of the inductive drive. One can extend the measurement down to $\Delta/h \sim 100\,\mathrm{MHz}$ (the lowest frequency one can go without incurring nonadiabatic error as one crosses zero tilt) using a different piece of data: the steady-state population of the qubit in the absence of a spectroscopic drive will be significantly above zero when $f_{10} \lesssim k_B T/h$. This stray population peak at small Δ lets us infer where the lowest frequency is from where the highest stray population is, and therefore where zero tilt is. To obtain zero tilt vs. SQUID bias at even lower Δ, one has to use other calibration techniques. For example, once β is large enough to be in the bistable double-well regime, one can perform the readout branch experiment (Fig. 3.16) vs. SQUID bias. The point where the readout resonator gives an equal response for both $|L\rangle$ and $|R\rangle$ will be at zero tilt (for such a calibration we would take the average response rather than look at all the individual stats like in Fig. 3.16).

An example of zero-tilt calibration data in the microwave regime ($100\,\mathrm{MHz} < \Delta/\mathrm{h} <$

8 GHz) is shown in Fig. 3.20. In this scan, the SQUID flux bias is stepped on the horizontal axis, and the tilt flux bias is scanned along the vertical axis. For a given value of SQUID bias, a microwave tone of fixed frequency is applied to the qubit's drive line. Ideally, this probe frequency is fixed for all SQUID biases, leading to a spectroscopic signal corresponding to a contour of constant f_{10} that should be symmetric about zero tilt. A slight complication arises in practice where the probe frequency usually needs to change for different SQUID biases. There are two considerations in choosing the probe frequency. For a given Δ, the probe frequency needs to be larger than Δ/h, because the minimum qubit f_{10} vs. tilt bias is equal to Δ. On the other hand, when Δ is small and the probe frequency is significantly larger than Δ/h, it becomes hard to drive the qubit. Furthermore, for large Δ, the range over which the tilt must be scanned to reach a spectroscopy peak grows to be quite large, since the characteristic hyperbola of f_{10} vs. tilt widens with larger Δ. To deal with all of these factors, we usually choose the probe frequency via the following heuristic. First, we do a spectroscopy vs. SQUID bias scan, with the tilt bias chosen to be roughly at zero tilt. For each SQUID bias, we then obtain a qubit frequency which we know must be greater than or equal to Δ for that SQUID bias. We then choose the probe frequency to be on the order of 100 MHz higher than this, so that the tilt scan range doesn't need to be too large. This works well for $\Delta \gtrsim 1$ GHz. However, for small Δ, the spectroscopy peaks begin to disappear, as seen on the left hand side of the data, but as mentioned above, a stray population peak starts to appear at zero tilt, which we can use to find zero tilt for the smallest Δ's.

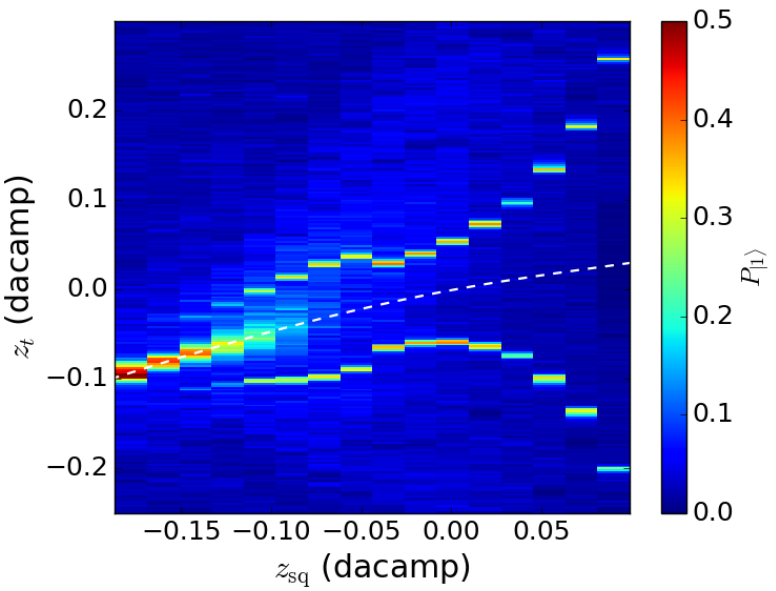


Figure 3.20: Microwave regime zero-tilt calibration scans taken at base temperature. Horizontal axis is SQUID bias in raw experimental DAC units (called dacamps) relative to the DC operating point (plotted range corresponds to $\sim 0.04\,\Phi_0$ in SQUID flux [0.3 in β]). Vertical axis is raw tilt bias in dacamps (plotted range corresponds to $\sim 0.015\,\Phi_0$ in tilt flux). Left side of plot corresponds to $\Delta/h \approx 100\,\mathrm{MHz}$, right hand to $\Delta/h \approx 7\,\mathrm{GHz}$. In this scan the spectroscopic probe frequency was chosen to vary between 3 and 7 GHz from left to right. The dashed white line is the result of a fit of zero tilt vs. SQUID bias [Eq. (2.58)]. Δ is increasing from left to right for this dataset. On the right hand side of the plot, the midpoint of the two spectroscopic peaks is used to infer zero tilt. On the very left side of the plot ($\Delta/h \lesssim 1\,\mathrm{GHz}$), the two spectroscopic peaks lose signal strength, and the stray population peak is used instead to infer zero tilt. To calibrate zero tilt at even lower $\Delta/h \lesssim 100\,\mathrm{MHz}$, another method such as readout branch must be used.

With zero tilt calibrated in the microwave regime, we can then do scans such as spectroscopy vs. SQUID bias at zero tilt, adding an appropriate compensation bias on the tilt as the SQUID bias is varied to keep the qubit at zero tilt. Usually we check that the zero-tilt calibration actually worked properly by doing variable-frequency spectroscopy vs. tilt (like in Fig. 3.19) at one or two values of Δ around the nominal zero tilt point. These short check scans do not take much extra time.

Given the various complications of this scan, one might imagine that an easier way

to calibrate zero tilt in the microwave regime is to simply look at the readout resonator response vs. qubit flux bias. This would indeed be much simpler and take less time, but unfortunately the response of the resonator to qubit bias is quite broad in the microwave regime such that there is not enough sensitivity to find zero tilt to within sufficient accuracy (typically this method would give a tilt error such that there is a $\sim 100\,\mathrm{MHz}$ deviation in f_{10} from its true zero-tilt value at $\Delta/h \sim 1\,\mathrm{GHz}$). It may be possible to change the design parameters of the readout to make such a scan possible with sufficient accuracy, but as of yet this has not been investigated in detail. However, in Chapter 6 we will find that looking at the readout resonator response vs. qubit biases is in fact extremely useful (and accurate enough) to distinguish linear geometric crosstalk in the system from intrinsic nonlinear "crosstalk" (nonlinear qubit-coupler interactions).

If the junction asymmetry is small (~ 1 percent), this nonlinear crosstalk can often be approximated as linear within a finite region of SQUID bias space, and then can be lumped in with the geometric crosstalk correction. Unfortunately, this is not always possible, even with a small range of bias. D-Wave circumvented the problem of junction asymmetry by making what they call a "compound-compound Josephson junction" (CCJJ) [16], which they claim was necessary to be able to operate devices with more than a few qubits. This structure replaces each junction of the DC SQUID with a DC SQUID, so that the two effective junctions of the main loop can be flux-tuned to be equal, eliminating the junction asymmetry. This of course comes at the expense of extra control knobs. The other advantage of such a structure is that not just β but β_{max} can

be made uniform between qubits, which aids D-Wave in the operation of multiqubit devices without individual time-dependent control lines for all qubit knobs. It is not clear yet whether or not the fluxmon architecture will require a similar type of compound-compound structure, since we plan to keep individual qubit control, but this needs to be considered as part of an overall systems optimization at the level of many qubits.

3.3.3 Modeling the measured fluxmon spectrum

An important part of device modeling and calibration is the ability to predict the qubit's spectrum as a function of its flux biases. To test our ability to do this, we can take spectroscopy data versus SQUID and tilt biases, and try to fit the resulting spectrum to a model. We require that the model and resulting fit parameters are physical, so that the model not only predicts the measured data but also can predict the spectrum reasonably well outside of the range of measurement, in addition to predicting with reasonable accuracy the flux and charge matrix elements that come into play later on for qubit-qubit coupling.

In Fig. 3.21, we show data where a few "cuts" through the two-dimensional flux bias space are taken. These include both horizontal and vertical cuts (constant tilt and constant SQUID biases), along with the special curvilinear cut at zero tilt (this requires that the zero-tilt calibration of Fig. 3.20 be performed beforehand). For this data, the intra-qubit geometric mutual inductance matrix was already precisely calibrated and the geometric crosstalk was already compensated, so we fix these numbers and do not include

them as free fit parameters. In particular this means that all crosstalk from SQUID to tilt remaining in the data is due purely to junction asymmetry. However, the remaining parameters (qubit inductances, capacitance, junction critical currents, and absolute flux offsets) are not measurable and are treated as fit parameters. We perform fits to two one-dimensional models (a fit to a three-dimensional model is not practical or necessary): the "1D simple" model described by eq. (2.56), and the Born-Oppenheimer "1D no caps" model described by eq. (2.100) which takes into account finite geometric inductance in the DC SQUID. Note that the 1D simple Hamiltonian is a special case of the 1D no caps Hamiltonian but with the inline betas β_1 and β_2 set to zero, and with a renormalized capacitance modulo the junction capacitances, as discussed in section 2.2.3.2. For the 1D no caps model for simplicity we assumed equal inline SQUID inductances $L_1 = L_2$ ($\beta_1 = \beta_2$).

The quality of the fits and the fitted parameters for both models are summarized in the Fig. 3.21 and the accompanying table 3.1. With the 1D simple model, we are able to fit the spectrum (which encompasses a wide range of frequencies and flux biases) to within $\sim 30\,\mathrm{MHz}$ RMS error, while with the 1D no caps model we can fit the data to within $\sim 20\,\mathrm{MHz}$ RMS error, which is actually roughly the average experimental uncertainty in the data.[28] More importantly, we find that the extracted qubit parameters for the 1D no caps model are more physically realistic, in that they much more closely match

[28]The linewidth of the qubit itself is $\sim 100\,\mathrm{MHz}$ away from zero tilt due to flux-noise limited dephasing (see next chapter for explicit extraction of linewidth), so this adds significant experimental uncertainty itself. The quality of the data is also not as good away from zero tilt due to it being harder to get a signal, so the resulting RMS error of our fit is actually quite good all things considered.

	Expected	1D simple fit	1D no caps (BO) fit
L (L_3)	685 pH	626 pH	680 pH
C (C_{eff})	108 fF	141 fF	118 fF
$I_c^{\text{sum}} = I_{c1} + I_{c2}$	1.3 μA	1.63 μA	1.32 μA
$\chi = \frac{I_{c1}-I_{c2}}{I_{c1}+I_{c2}}$	± 0.05	-0.02	-0.03
L_1, L_2	~ 20 pH	N/A	27 pH

Table 3.1: Expected (based on Sonnet simulations and room-temperature junction resistance measurements) and fitted parameters for the to two models to the spectroscopy data. The 1D no caps model yields more physically realistic parameters. The physically expected capacitance in this table does not include junction capacitance, which we estimate should contribute 5 - 10 fF, making the match to the 1D no caps fit even more remarkably close.

the expected physical inductance, capacitance, and critical currents of the device based on Sonnet simulations[29] and room temperature junction resistance measurements (which give the junction critical current through the Ambegaokar-Baratoff relation [34, 101]). The resulting best-fit parameters for both models are displayed in table 3.1. One can see that the fitted parameters for the 1D no caps model are remarkably close to the physically expected parameters.

As a further check, in Fig. 3.22 we plot the subset of data from Fig. 3.21(b) at zero tilt as a function of z_{sq}, along with an extra piece of experimental data, f_{21}. f_{21} is measured by probing the system with a higher power spectroscopy tone at the second-order two-photon transition $|0\rangle \rightarrow |2\rangle$, i.e. at $f = f_{20}/2$, and then inferring f_{21} from the $f_{20}/2$ and f_{10} data.[30] We can then plot the prediction (not a fit) for f_{21} vs. z_{sq}

[29]The small inline SQUID inductances are difficult to simulate accurately, but we expect very roughly around 10 - 20 pH for each branch (a reasonable rule of thumb for inductive wires is 1 pH per micron of wire, more or less depending on proximity of ground planes and weakly dependent on wire thickness).

[30]Earlier, we mentioned it is not possible to distinguish the $|2\rangle$ and $|0\rangle$ states when using the double-well projection readout scheme. This is true, but we can rely on decay of the $|2\rangle$ state to the $|1\rangle$ state before performing the readout, which provides a good enough signal to perform spectroscopy of the $|2\rangle$ state.

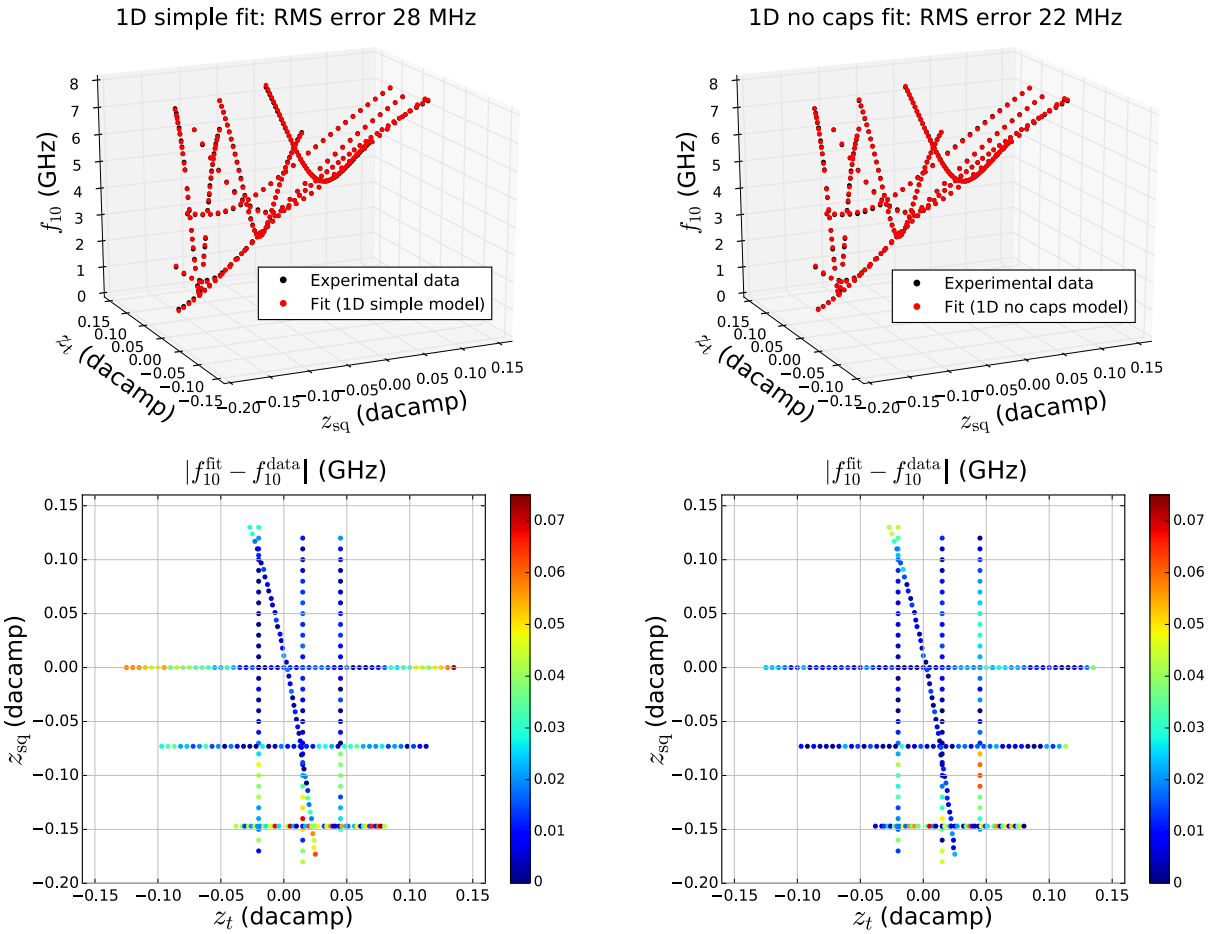


Figure 3.21: Fitting of experimental single-qubit spectroscopy data to two different 1D models: the "1D simple" model described by eq. (2.56), and the Born-Oppenheimer "1D no caps" model described by eq. (2.100) which takes into account finite geometric inductance in the DC SQUID.

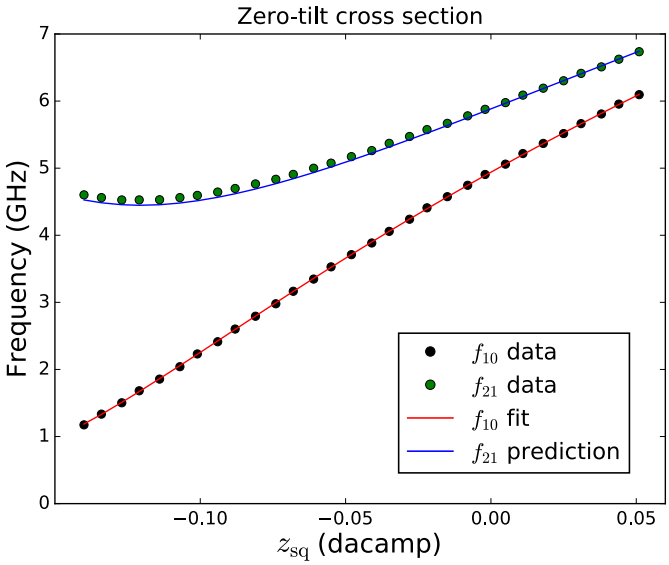


Figure 3.22: Spectroscopy of f_{10} and f_{21} at zero tilt (black and green), with 1D no caps fit for f_{10} from Fig. 3.21(b) (red) and resulting prediction (not a fit) for f_{21} (blue). The f_{10} data is a subset of the data from Fig. 3.21.

based on the fitting of the f_{10} data. The result is shown in Fig. 3.22. Our ability to fairly closely predict f_{21} after fitting only for f_{10} is a nice sanity check and more evidence that our model and fitting are physically realistic. More sophisticated fitting that also incorporates f_{21} may lead to better model, but this needs to be studied some more. Also, we note that the qubit measured in this study was not an isolated qubit, but a qubit within a coupled-qubit device with the coupler and other qubit biased away such that they shouldn't influence the qubit in question.

3.4 Understanding the S-curve: Single-qubit annealing

The S curve experiment of Fig. 3.17 (without pi pulse) is interesting because it is the simplest possible case of quantum annealing, using a single qubit with a local field but no inter-qubit couplings. Furthermore, we will find that the S curve experiment is an extremely useful tool for calibrating flux crosstalk in a multiqubit system (Chapter 6), and also for measuring low-frequency flux noise intrinsic to the qubit (Chapter 4). Therefore, it would be worthwhile to try and understand what determines the width and shape of the S curve.

For an $\sim 8\,\text{ns}$ linear ramp speed of the SQUID bias (i.e., the speed at which we raise the barrier in the schematic pulse sequence in Fig. 3.17), we typically observe a full width (defined as the interval between 5 and 95 percent probability to end up in the right well) of $\sim 600 - 700\,\mu\Phi_0$, as can be seen in Fig. 3.23. In the next chapter, we will see that low-frequency flux noise contributes a Gaussian broadening over experimental runs with $\sigma \approx 40 - 50\,\mu\Phi_0$, meaning that the noise would smear out[31] a perfectly sharp step function to a full width of $\sim 3\sigma \approx 120 - 150\,\mu\Phi_0$. Therefore, broadening from low-frequency noise is not enough to explain the majority of the S curve width. Furthermore, when we increase the ramp time to 40 ns, the width of the S curve actually decreases, as can be seen in Fig. 3.23. To understand this, we need to look at the dynamics of the S curve trajectory, which we will find involves nonadiabatic physics. What's even more

[31] The convolution of a Gaussian with a step function is $\text{erf}\left(\frac{x}{\sqrt{2}\sigma}\right)$, which has a width of $\sim 3\sigma$.

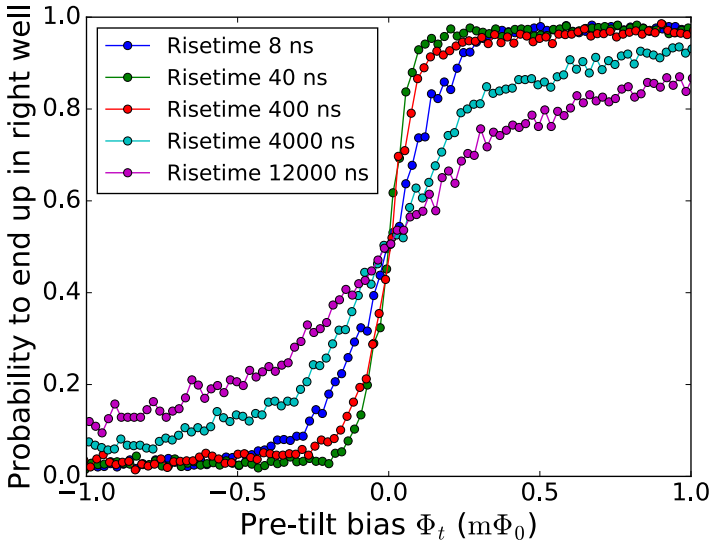


Figure 3.23: S curve measured for several different SQUID (barrier) bias ramp times: 8 ns, 40 ns, 400 ns, 4000 ns, and 12000 ns. We observe a non-monotonic dependence of the width of the transition region on ramp rate. The 40 ns ramp leads to a narrower curve than the 8 ns ramp due to a reduction in nonadiabatic error, but the 4000 and 12000 ns ramps lead to a much broader curve due to thermalization errors.

puzzling (at first) is that there is a non-monotonic dependence of the width on ramp time, as can be seen in Fig. 3.23: for very long ramp times, the width of the S curve actually starts to increase drastically.

Fig. 3.24 illustrates how the S curve experiment can be viewed as a single-qubit annealing experiment. The system is prepared in its ground state in the single-well regime (left hand side of plot) with a variable amount of pre-tilt bias Φ_t. Then, the barrier is raised by ramping the DC SQUID flux bias (towards the right hand side of the plot), with the qubit ending up in one of the two classical states of a double-well potential. We plot the predicted energy gap of the system along this annealing trajectory, for several different values of tilt bias Φ_t. For small tilt biases, the minimum energy gap

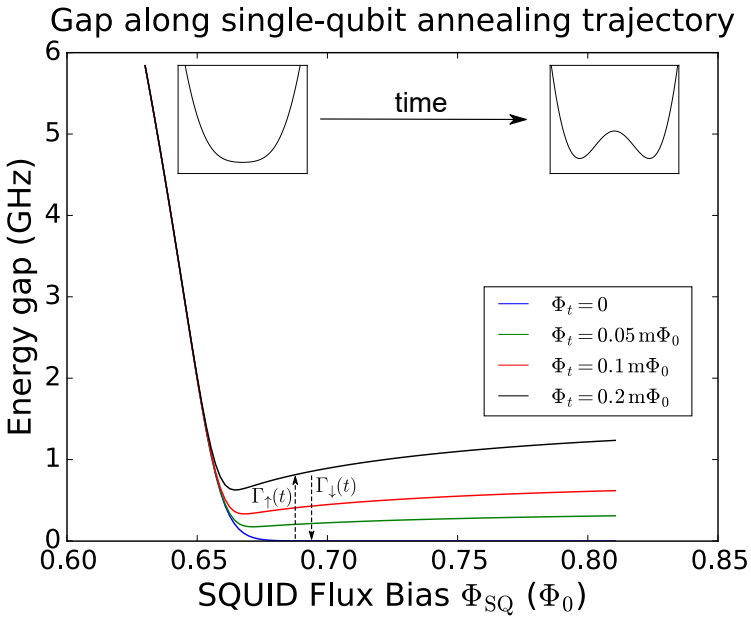


Figure 3.24: Numerically predicted energy gap of the system along the single-qubit annealing trajectory, for several different values of tilt bias Φ_t. Time goes from left to right. The presence of noise or dissipation at high frequencies can drive transitions $\Gamma_{\uparrow/\downarrow}$ out of and back into the ground state throughout the annealing.

along the annealing trajectory becomes arbitrarily small, and for increasing tilt biases monotonically increases. In the absence of noise then, we can guess the following: for very fast ramps, the system will undergo a Landau-Zener transition at larger gaps, leading to broadening of the S curve. Conversely, for very slow ramps, the system should always be adiabatic, leading to an arbitrarily narrow S curve in the absence of an environment. However, thermalization to the environment may explain the broadening at very long ramp times.

To test these hypotheses quantitatively, we first simulated the time-dependent Schrödinger dynamics of the system in the absence of an environment according to the two-level

Hamiltonian

$$H(t) = -\frac{1}{2}\Delta(t)\sigma_x - \frac{1}{2}\varepsilon(t)\sigma_z, \tag{3.41}$$

with $\varepsilon(t)$ computed as $2I_p(t)\Phi_t$, with the system prepared in the ground state of $H(0)$ at $t = 0$. We used a linear ramp for the SQUID flux bias. The results are shown by the dashed lines in Fig. 3.25 for various ramp rates. We note that the results of these two-level simulations were within a few parts in 1000 of the result of the much more expensive simulation of the full discretized flux wave function dynamics under the continuous lab-frame Hamiltonian

$$H(t) = E_L\frac{(\hat{\varphi} - \varphi_t(t))^2}{2} - 4E_C\frac{\partial^2}{\partial\hat{\varphi}^2} + \beta(\Phi_{\mathrm{SQ}}(t))E_L\cos(\hat{\varphi})\,. \tag{3.42}$$

We see that the simulation predicts the experimental widths of the 8 ns ramp and 40 ns ramps fairly well, meaning that this picture is likely correct. However, note that the simulation drastically fails to predict the experimental data for the longest ramp times, as the simulation predicts an almost perfectly sharp transition as expected from adiabaticity.

A possible reason for broadening at very long ramp times is thermalization to the environment, which can induce upwards and downwards transition rates $\Gamma_\uparrow(t)$ and $\Gamma_\downarrow(t)$ between energy eigenstates, as illustrated in Fig. 3.24. These transition rates to not necessarily have to obey detailed balance if the environment is not in thermal equilibrium (see next chapter), and can be computed according to Fermi's golden rule [equation (4.80)] if the environmental noise spectral density at the transition frequency is known (this will be explained in more detail in the next chapter). It turns out that we are able

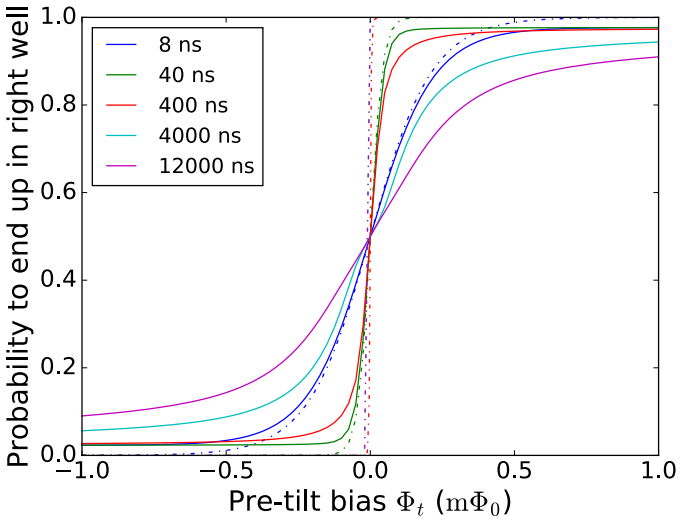


Figure 3.25: Numerical simulation of single-qubit annealing with (solid lines) and without (dashed lines) thermalization to the environment. Annealing fast avoids thermalization. When the environment was included, the up/down transition rates $\Gamma_{\uparrow/\downarrow}(t)$ were calculated along the annealing trajectory according to Eq. (3.43) based on a phenomenological fit to the measured environmental noise spectrum [Fig. 4.23(c)]. The simulations consider only the lowest two levels of the qubit. Note that the simulations come reasonably close to predicting the actual widths and shapes of the measured S curves. The 40 ns ramp yields a narrower S curve than the 10 ns ramp, but the 4000 ns ramp with environment yields a much wider transition than the extremely sharp one that would occur in the absence of thermalization.

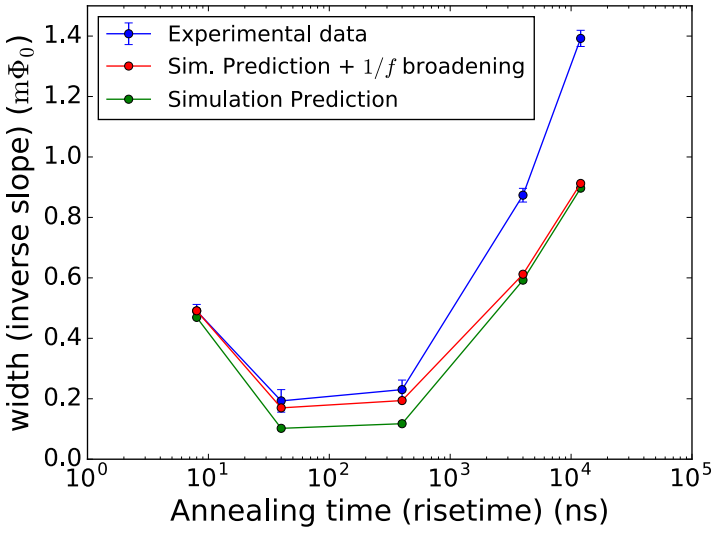


Figure 3.26: Comparison of extracted S curve widths (defined as the inverse of the fitted slope near zero tilt) between the experimental data versus the simulations of Fig. 3.25. For the simulations, we also show the effect of adding $1/f$ broadening (convolution with a Gaussian with $\sigma = 0.05\,\mathrm{m}\Phi_0$.)

to measure the two-sided environmental noise spectrum $S_\Phi(f)$ between $\sim 500\,\mathrm{MHz}$ and $\sim 6\,\mathrm{GHz}$ (see Chapter 4), allowing us to come up with a phenomenological model for the noise versus frequency as in Fig. 4.23(c). This noise model has an effective temperature of $\sim 30\,\mathrm{mK}$ for $f \lesssim 3\,\mathrm{GHz}$. We can then plug this noise model into our single-qubit annealing trajectory to compute the up and down transition rates as a function of time according to Fermi's golden rule,

$$\Gamma_{\downarrow/\uparrow}(t) = \frac{1}{\hbar^2}\frac{1}{L^2}|\langle 0|\hat{\Phi}|1\rangle(t)|^2 S_\Phi(\pm f_{10}(t)). \tag{3.43}$$

We incorporate the environmentally induced transitions into our simulation by generalizing it to a Lindblad master equation [102, 103] for the qubit's density matrix,

$$\dot{\rho} = \frac{1}{i\hbar}[H,\rho] + \sum_k \left(L_k \rho L_k^\dagger - \frac{1}{2}\{\rho, L_k^\dagger L_k\} \right). \tag{3.44}$$

This is a standard approach for weakly coupled Markovian environments, where the Lindblad superoperators L_k are introduced to represent stochastic transitions between energy levels.[32] It is valid to use this approach under the same conditions we will discuss in the context of Fermi's golden rule in the next chapter. This will be the case everywhere except possibly right near zero tilt, where the minimum gap becomes smaller than the linewidth of the qubit. We note that this same master equation model was also successfully used to quantitatively predict Landau-Zener transition probabilities in a superconducting qubit system with dissipation [104]. We use two collapse operators

[32] Physically, this type of equation is derived from the unitary evolution of the joint system + environment and subsequently tracing over the environment's degrees of freedom.

corresponding to qubit excitation and decay,

$$L_{\uparrow}(t) = \sqrt{\Gamma_{\uparrow}(t)}\sigma_{+}, \qquad L_{\downarrow}(t) = \sqrt{\Gamma_{\downarrow}(t)}\sigma_{-}. \tag{3.45}$$

In the beginning of the actual experiment, the system is not initialized entirely into its ground state, but rather has a small percentage occupation of the first excited state given by a stray population p_{stray} typically around 1 %. Accordingly, we initialize our master equation simulation with the qubit in the mixed state $\rho_i = (1-p_{\text{stray}})\left|0\right\rangle\left\langle 0\right| + p_{\text{stray}}\left|1\right\rangle\left\langle 1\right|$ to mimic the actual experiment. This is the reason why the S curve does not go all the way to zero/one for large negative/positive pretilts. We also note that the full machinery of the Lindblad master equation is not actually necessary when the dynamics are adiabatic (which is almost everywhere for long ramp times), because then the final populations at the end of the anneal are simply determined by integrating the transition rates $\Gamma_{\uparrow/\downarrow}(t)$ over time (in other words there is a rate equation for two probabilities without any off-diagonal terms in the density matrix).

As a technical note, to actually numerically solve the master equation (3.44), we need to work in the instantaneous energy basis, because this is the basis in which the transitions happen for a weakly coupled environment. However, the resulting Hamiltonian then becomes $H(t) = -\frac{1}{2}hf_{10}(t)\sigma_z$ (where f_{10} is the instantaneous gap), which when integrated gives bogus results. The reason is that when we use the instantaneous energy basis, we have taken a frame of reference "rotating" with the system. In reality, the actual equation of motion obeyed by the qubit in this time-dependent basis is given by an effective

Hamiltonian that picks up an extra nonadiabatic geometric correction term [105],

$$H^{\text{eff}}(t) = -\frac{1}{2}hf_{10}(t)\sigma_z - \hbar g(t)\sigma_y, \tag{3.46}$$

where

$$g(t) \equiv \langle 1(t)|\partial_t|0(t)\rangle, \tag{3.47}$$

with the derivative taken in the lab frame. We use this effective Hamiltonian for our master equation simulation (we verified that this effective Hamiltonian gives the same results as (3.41) in the absence of the environment).

The results of the simulations with environment are shown by the solid lines in Fig. 3.25. For the 8 and 40 ns ramps, including the environment barely makes a visible difference to the width. However, for the 4000 ns and longer ramps, there is an enormous difference due to thermalization to the environment, causing the S curve to be greatly widened.

In summary, we have seen that we can view the single-qubit annealing process as a tradeoff between nonadiabaticity and thermalization, and have used this insight to understand the experimental S curve data for different ramp rates. Going much faster than the thermalization timescale (in particular, faster than the fastest anneal time of $5\,\mu$s available on current D-Wave hardware) can substantially change the success rate of the single-qubit annealing.

Chapter 4

Noise, dissipation, and observing the classical-quantum crossover

Superconducting flux qubits in a quantum annealer cannot be treated as a quantum system isolated from its environment. For the same reasons that superconducting qubits can be easily coupled together (as we'll study in Chapters 5 and 6), they can also unintentionally couple to baths of microscopic defects, or to a poorly engineered microwave electromagnetic background. In particular, the fluxmon is unfortunately very sensitive to flux noise, even more so than other types of superconducting qubits (or in a more positive light, it enables the use of the fluxmon as a sensitive tool to study flux noise and perhaps figure out how to reduce it). For readers familiar with flux-tunable transmon qubits, we note that even at zero tilt the fluxmon's flux sensitivity is 100× larger than typical transmon qubits, and away from degeneracy it can be another factor of 10 larger!

In quantum annealing, fundamentally different behavior can occur in the presence of environmental noise and dissipation; this is the reason that existing real-world devices are called "quantum annealers" as opposed to "adiabatic quantum computers." For example, dissipation can suppress quantum tunneling rates, which is one motivation for the construction of a quantum annealing device without the use of dissipative dielectric materials. Low frequency flux noise, whose precise microscopic origin is still largely unknown, leads to dephasing, which, for small enough tunneling matrix elements, reduces otherwise coherent quantum tunneling into incoherent quantum tunneling with a quadratically suppressed tunneling rate, potentially undoing an otherwise present quantum speedup. In addition, low-frequency flux noise can lead to parameter drift translating to programming errors in any quantum annealer if the drifts cannot be continually re-calibrated on the timescale of the noise.

In this chapter, we will quantify how the fluxmon qubit is coupled to its environment. In particular, we look at noise and dissipation induced by flux noise and dielectric loss, and their effects on quantum tunneling between flux qubit wells. In most cases, the effects of noise and dissipation can be described by a properly defined spectral density that captures the statistics of the environmental degrees of freedom. This description is valid when the noise source is Gaussian, which in most (but not all) cases we expect to be true as long as the environment consists of a large collection of individual fluctuators (think central limit theorem). We will study noise, dissipation, and spectral densities in depth both theoretically and experimentally. Using various experimental techniques we will extract

the fluxmon's environmental flux noise spectrum over a wide range of frequencies. In particular, by extracting the flux noise spectrum over a range of frequencies near the energy scale set by the device temperature ($\sim 1\,\mathrm{GHz}$), we are able to observe for the first time the crossover between the classical and quantum regimes of flux noise. The fact that the noise follows a $1/f^{\alpha}$ spectrum all the way to such a high frequency restricts the possible mechanisms responsible for the noise. Finally, we will see how these independent noise measurements can be used to approximately predict the performance of quantum tunneling between flux qubit wells.

4.1 Dephasing and low-frequency $1/f$ noise

The fluxmon qubit forms a continuous dc loop from one side of its dc SQUID to the other through the ground plane, making it insensitive to charge noise at low frequencies. On the other hand, it will still be sensitive to any low-frequency noise in magnetic flux. For several decades, low-frequency magnetic flux noise has been observed in superconducting circuits, including SQUIDs and superconducting qubits [106, 107, 108, 109, 110, 111, 112, 113]. This flux noise is observed to have an approximately $1/f$ spectrum at low frequencies, meaning a frequency-dependent power spectral density of the form $S_{\Phi}(f) \propto 1/f^{\alpha(f)}$ with α close to 1 at low temperatures [110, 114, 115, 116, 117, 118, 119, 62, 71]. Given the ubiquity and importance of this noise, it is perhaps surprising that its microscopic origin has yet to be determined.

A commonly accepted view is that the $1/f$ noise probably arises from a bath of

telegraph-like magnetic fluctuators, with a wide and properly distributed range of relaxation times (we will explain what this means momentarily), which couple random fluxes to the qubit loops. These fluctuators might be spatially localized, weakly interacting single spins or spin clusters in the vicinity of the qubit, or collective, spatially distributed dissipative modes associated with the diffusion of magnetic moments around the qubit. One of the experimental results of this chapter is that the latter phenomenon, known as "spin diffusion," is unlikely to be the source of the $1/f$ flux noise, at least not over the entire frequency range over which we observe it.

We can briefly give a generic argument for what it means for $1/f$ noise to originate from an ensemble of fluctuators, and why this type of origin may in fact be expected. A telegraph fluctuator is any physical quantity that fluctuates between two states with a constant probability per unit time for transitioning from one state to the other (i.e., a Poissonian process). Physically, this fluctuator may be a subsystem with two stable configurations separated by a potential energy barrier, where transitions between the two configurations are possible when enough energy is supplied. As shown in Ref. [120], by analyzing the resulting differential equation dictating transitions between the two states, the autocorrelation function is exponential in time with time constant τ, leading to a Lorentzian spectral density[1] of noise for a single fluctuator,

$$S(\omega) \propto \frac{\tau}{\omega^2\tau^2/4 + 1}. \tag{4.1}$$

[1]Right now we use 'proportional to' symbols for spectral density, but in a later section we will carefully define the single- and double-sided noise power spectral densities of a random process. In fact, to be sure that all factors of 2 and 2π are correct, we will actually derive the Wiener-Khinchin theorem later in this chapter.

Following Dutta and Horn [121], if we have a collection of independent fluctuators with a distribution of characteristic times $D(\tau)$, the sum of the noise from all fluctuators will then take the form

$$S(\omega) \propto \int d\tau D(\tau) \frac{\tau}{\omega^2 \tau^2 / 4 + 1}. \tag{4.2}$$

If the two stable states of each fluctuator are separated by a potential energy barrier of height ΔE, then if the fluctuation is thermally activated we would expect the transition rate to roughly be

$$\frac{1}{\tau} \propto \frac{1}{\tau_0} \exp(-\Delta E / k_B T), \tag{4.3}$$

where T is the temperature. If the fluctuations are instead driven by quantum tunneling between configurations ("quantum fluctuations"), there would (nominally) no longer be a temperature dependence, but we would still obtain an exponential dependence of τ on ΔE consistent with the WKB approximation (2.54). It is physically reasonable to expect a more or less uniform distribution of activation energies ΔE, at least over a limited range, which translates to a distribution $D(\tau) \propto 1/\tau$ over a very wide range of τ. Suppose this range is defined by $\tau_1 \ll \tau \ll \tau_2$. Then, for $\frac{1}{\tau_2} \ll \omega \ll \frac{1}{\tau_1}$, we can take the limits of integration for τ in (4.2) to be 0 and ∞, yielding

$$S(\omega) \propto \frac{1}{\omega}. \tag{4.4}$$

Using this argument, one may convince oneself that $1/f$ noise should actually be expected to be fairly ubiquitous. Note that we have not specified the expected temperature dependence of the noise, as this depends on other assumptions, a point to which we will

come back later. In this section, we describe simple experiments to measure the total noise power of the low frequency noise, followed by more involved experiments to extract frequency-resolved information about the noise spectrum.

4.1.1 Integrated effect of noise: T_2, broadening, and programming errors

The effect of low-frequency flux noise on the fluxmon depends very strongly on where the qubit is being biased during an experiment. For example, at the special bias of zero tilt, which gives a symmetric potential, the first-order sensitivity of f_{10} to tilt flux bias vanishes (intuitively, at this special point the states are insensitive to dephasing from flux noise because the energy eigenstates all have the same expected value of flux), and the dephasing due to noise in the SQUID flux bias will dominate. However, away from zero tilt the sensitivity to tilt flux bias very quickly becomes dominant. This effect is most drastic in the regime of large persistent current, say $\Delta/h \lesssim 1\,\mathrm{GHz}$, where the sensitivity can be of order $1000\,\mathrm{GHz}/\Phi_0$, around 1000 more sensitive than tunable transmon qubits used for gate-based quantum computing [40]. These points are illustrated quantitatively in Fig. 4.1, where we plot the sensitivities to each flux bias as a function of flux biases. During quantum annealing, however, the flux bias of the qubit will not be fixed, and will in general deviate from zero tilt, making sensitivity to tilt flux noise a primary concern.

An easily observable effect of $1/f$ noise is the loss of phase coherence between energy eigenstates, which is equivalent to a decay in time of the off-diagonal phase elements

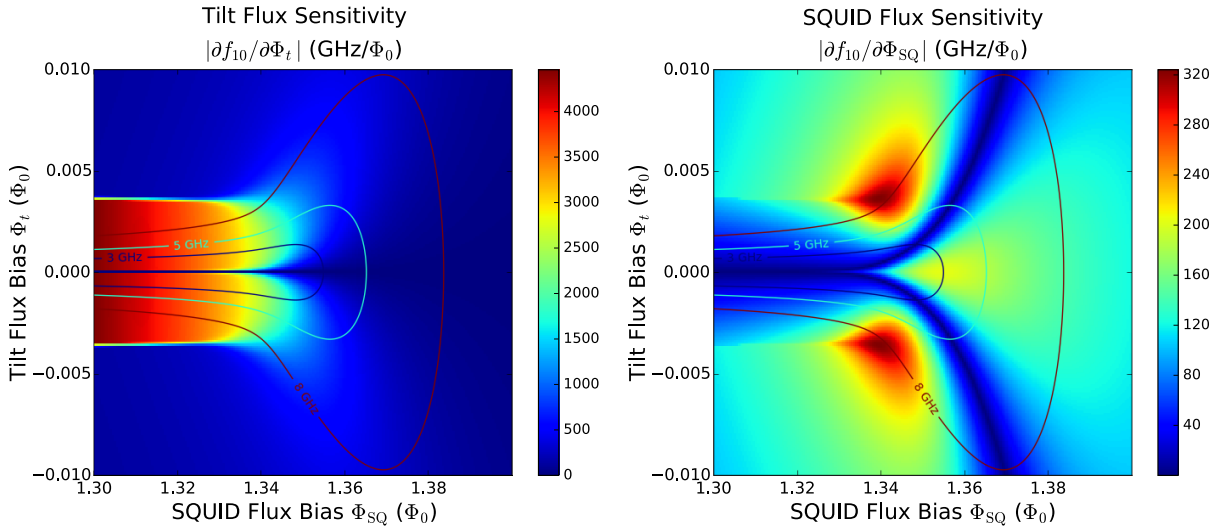


Figure 4.1: Sensitivity of qubit transition frequency f_{10} to SQUID and tilt flux biases, in units of GHz/Φ_0, with contours of constant f_{10} drawn in at $f_{10} = 3$, 5, and 8 GHz for reference. Note that away from degeneracy (zero tilt), the sensitivity to tilt flux can be much higher than the sensitivity to SQUID flux. These numbers can be compared with a typical Xmon transmon flux sensitivity of $\sim 1\,\mathrm{GHz}/\Phi_0$). The sudden jump in tilt bias sensitivity at $\Phi_t \approx .004\,\Phi_0$ corresponds to when the first two eigenstates stop being in different wells and occupy the same well.

in the ensemble-averaged qubit density matrix. The characteristic time it takes for the phase to decay gives information about the integrated noise power (modulo some filter function unique to the pulse sequence), and the precise functional form of the decay of the phase coherence vs. time gives some information about the shape of the noise spectral density vs. frequency. Similarly, in frequency space, the noise will lead to a broadening of the qubit's energy transition in spectroscopy, and the precise shape of the spectroscopic peak will give some information about the functional form of the noise power vs. frequency.

The simplest time-resolved experiment to measure phase decay is the Ramsey pulse sequence [122], illustrated in Fig. 4.2(a). A simplified explanation of this experiment is as

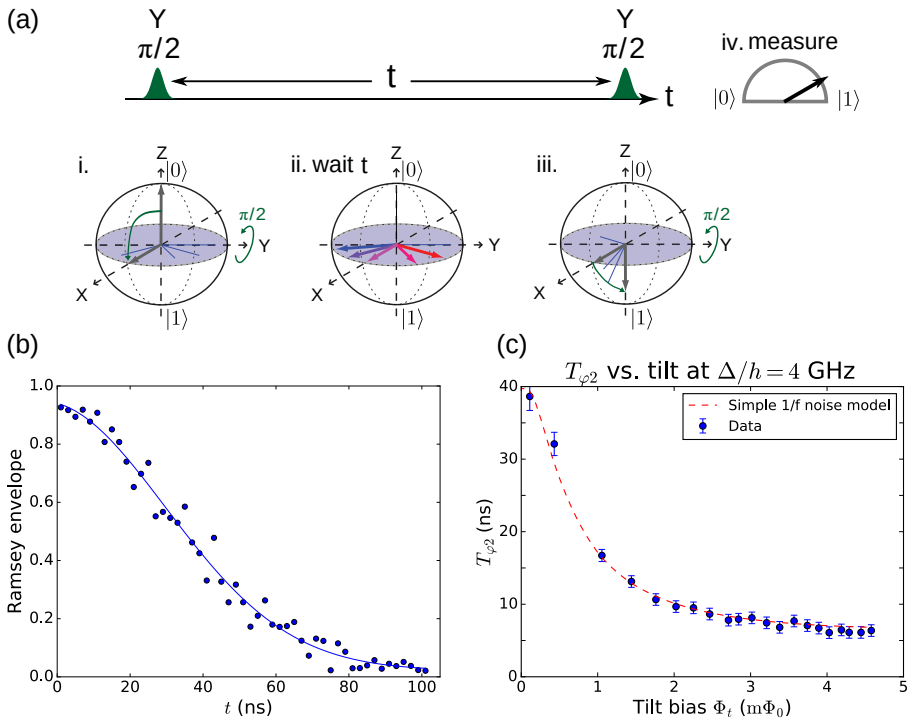


Figure 4.2: (a) Illustration of pulse sequence and Bloch sphere trajectory of a simple time-resolved Ramsey phase coherence experiment. (b) Typical Ramsey decay envelope measured for a fluxmon qubit at zero tilt and $f_{10} \approx 5\,\text{GHz}$. For this dataset, we extract $T_{\varphi 2} \approx 40\,\text{ns}$ and $T_{\varphi 1} \approx 300\,\text{ns}$. (c) Extracted $T_{\varphi 2}$ vs. tilt bias at a gap of $\Delta/h = 1\,\text{GHz}$. The functional form of the degradation away from zero tilt is consistent with two relatively uncorrelated $1/f$ flux noise sources in the fluxmon's SQUID and tilt loops.

follows. After initializing the qubit in its ground state $|0\rangle$, a $\pi/2$ pulse[2] is applied to bring the qubit to the superposition state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ on the equator of the Bloch sphere, in a frame rotating with the resonant drive. We then wait for a variable delay time t. In the rotating frame of the qubit, the relative phase of the $|0\rangle$ and $|1\rangle$ state components should not change, so that when a second $\pi/2$ pulse is applied, the qubit should end up at the north (or south, depending on the sign of the pulse) pole of the Bloch sphere. If there is any drift in qubit frequency during the delay time, this will lead to the accumulation of an azimuthal phase $\delta\varphi$ on the equator, meaning the probability of ending the correct state at the end is reduced. For an ensemble average of random phase drifts, this will act to shrink the magnitude of the resulting Bloch vector, corresponding to dephasing. A fully dephased qubit will end up at the center of the Bloch sphere (although in the very long time limit [compared to T_1 energy decay] it will also move vertically on the Bloch sphere end up in the ground state).

In practice, the drive frequency is chosen to be slightly detuned from the true qubit frequency, so that oscillatory "fringes" at the detuning frequency can be observed, making it easier to distinguish true phase decay from an unintentional small fringe frequency. In addition, we typically measure the decay of the envelope function deduced from two-axis tomography with both $X_{\pm\pi/2}$ and $Y_{\pm\pi/2}$ pulses (which rotate the basis of measurement) in order to make the measurement more robust, defining the Ramsey envelope visibility

[2] A $\pi/2$ pulse is similar to the π pulse we used in the previous chapter, but with nominally half the duration (or amplitude).

as

$$V = \frac{1}{2}\sqrt{\left(\langle\sigma_x\rangle - \langle -\sigma_x\rangle\right)^2 + \left(\langle\sigma_y\rangle - \langle -\sigma_y\rangle\right)^2}, \tag{4.5}$$

where the expectation values are taken when the Bloch vector is on the equatorial plane, before the second $\pi/2$ pulse.

It can be shown that for phase fluctuations that are Gaussian distributed, the ensemble-averaged Ramsey envelope visibility is only a function of $\langle\delta\varphi^2(t)\rangle$ given by $V(t) = \exp\left[-\frac{1}{2}\langle\delta\varphi^2(t)\rangle\right]$ [123]. In a real experiment, we have

$$V(t) = A\exp\left[-\frac{1}{2}\langle\delta\varphi^2(t)\rangle - \frac{t}{2T_1}\right] + B, \tag{4.6}$$

where $A \approx 1$ accounts for readout and state preparation error, $B \approx 0$ accounts for nonzero steady state stray population, and T_1 accounts for exponential energy decay from $|1\rangle$ to $|0\rangle$. For the fluxmon though, T_1 is much longer than the dephasing time $T_{\varphi 2}$ at the operating points at which we perform Ramsey, so we can ignore this contribution for most purposes. Note that the Ramsey decay time (without spin echo) is often called T_2^* in the literature.

Fig. 4.2(b) shows a measured Ramsey envelope vs. delay time for a typical fluxmon qubit on a sapphire substrate, biased at zero tilt and $f_{10} = \Delta/h \approx 5\,\mathrm{GHz}$. Fitting the accumulated phase error in (4.6) to the form

$$\frac{1}{2}\langle\delta\varphi^2(t)\rangle = \frac{t}{T_{\varphi 1}} + \left(\frac{t}{T_{\varphi 2}}\right)^2, \tag{4.7}$$

we typically extract $T_{\varphi 1} \gg T_{\varphi 2}$, meaning that the Ramsey envelope is nearly Gaussian. As shown in Ref. [123], this is consistent with predominantly $1/f$ noise in the frequency

range relevant to the Ramsey experiment. Explicitly, assuming a $1/f$ flux noise spectrum $S_\Phi(f) = \frac{S_\Phi^*}{f}$, one can compute a resulting $T_{\varphi 2}$ of [115]

$$\frac{1}{T_{\varphi 2}} \approx \frac{2\pi}{\sqrt{2}} \frac{df_{10}}{d\Phi} \sqrt{S_\Phi^*} \sqrt{\ln\left(\frac{0.4}{f_l t}\right)} \approx 3.5 \cdot 2\pi \frac{df_{10}}{d\Phi} \sqrt{S_\Phi^*}, \tag{4.8}$$

where f_l is a low-frequency cutoff is given by the total experimental data acquisition time (i.e., including all repetitions of an ensemble average), typically of order 1 minute, and t is of order $T_{\varphi 2}$ itself. This means that given a $T_{\varphi 2}$ and flux sensitivity, we can extract an equivalent $1/f$ flux noise amplitude, specified by its value at 1 Hz. For example, for the zero-tilt data shown in Fig. 4.2, the qubit had a $T_{\varphi 2}$ of 40 ns and a sensitivity to SQUID flux of $\sim 200\,\mathrm{GHz}/\Phi_0$, implying an equivalent flux noise amplitude of $6\,\mu\Phi_0/\sqrt{\mathrm{Hz}}$ in the DC SQUID loop. The quantitative utility of this metric however is unclear, since the slope α in the $1/f^\alpha$ noise power law could easily deviate from 1, and the actual noise power in the DC SQUID loop at 1 Hz is actually measured to be less, more like $2\,\mu\Phi_0/\sqrt{\mathrm{Hz}}$.

From the flux sensitivity landscape of Fig. 4.1 (or more intuitively from the hyperbolic form of f_{10} vs. tilt seen in Fig. 2.10(a)), we should expect that the dephasing time should get worse as we move away from zero tilt bias, assuming that the flux noise in the tilt loop is at least as bad as the noise in the DC SQUID loop. In Fig. 4.2(c), we plot the extracted $T_{\varphi 2}$ vs. tilt bias at $\Delta/h = 4\,\mathrm{GHz}$, which shows that this is indeed the case. Away from zero tilt, the $T_{\varphi 2}$ roughly saturates to a constant value determined by the slope of the f_{10} hyperbola away from zero tilt. For $\Delta/h = 4\,\mathrm{GHz}$, this worst-case $T_{\varphi 2}$ is typically around ~ 9 ns, with some samples a bit lower down to ~ 6 ns, and occurs at a

flux sensitivity of $\sim 450\,\mathrm{GHz}/\Phi_0$. Using the known sensitivies to SQUID and tilt flux, we can fit the curve to a model with uncorrelated $1/f$ noise in the SQUID and tilt loops, which fits the data fairly well. The fitted flux noise power in the tilt loop is typically 3 - 5 times larger than that in the SQUID loop, which makes rough qualitative sense given the geometries of the loops (specifically, the effective length over wire thickness, which for independent spin fluctuators should scale linearly with the noise power [111]).

As can be seen in Fig. 4.1(a), the tilt flux sensitivity in the small-gap, double-well regime can be substantially higher, up to a few thousand GHz/Φ_0. Unfortunately, this is such a large flux sensitivity that $T_{\varphi 2}$ is expected to be only $\sim 1-3\,\mathrm{ns}$, too small to reliably measure (primarily because we can't perform the $\pi/2$ pulses that fast). However, it is still possible to do a measurement of dephasing here, but in the frequency domain instead of the time domain, by looking at the spectroscopic linewidth of the qubit. An example dataset of linewidth extraction vs. tilt bias for $\Delta/h = 1\,\mathrm{GHz}$ is shown in Fig. 4.3(a) (raw data) and (b) (extracted linewidth). One can see that the linewidth (half width at half maximum) starts fairly narrow, around 10 MHz, and then very quickly saturates to $\sim 100\,\mathrm{MHz}$ away from zero tilt, consistent with a $T_{\varphi 2}$ of 1 - 2 ns as expected. We note that D-Wave infers this noise through the linewidth of macrosopic resonant quantum tunneling between flux qubit wells (an experiment which we will discuss later in section 4.3). We note that the method used here is more direct than the MRT linewidth method, as it directly gives the broadness of the qubit transition in units of energy, rather than having to convert from a flux linewidth to an equivalent energy linewidth.

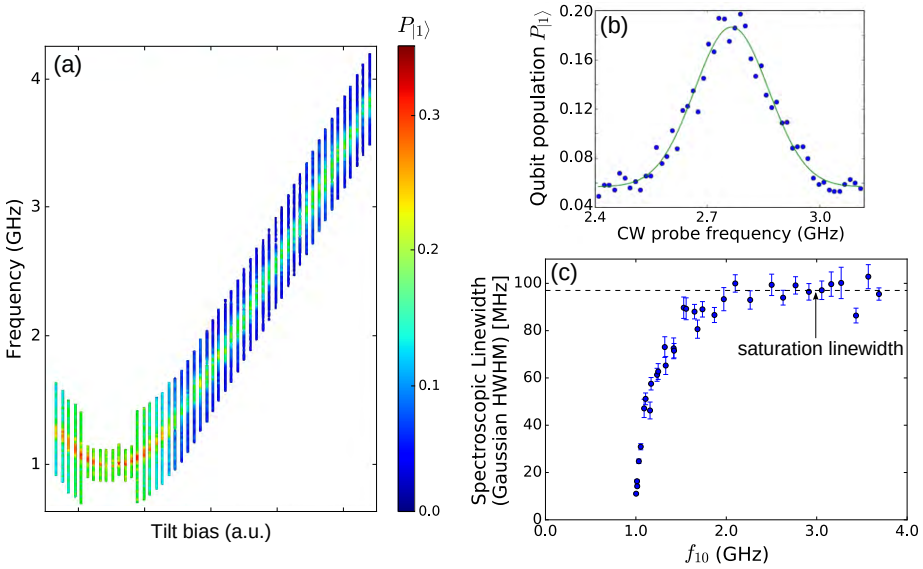


Figure 4.3: (a) Raw spectroscopy vs. tilt data. Note that the linewidth at zero tilt (reddish color on green background) is much narrower than the linewidth away from zero tilt (greenish color on blue background). The colors differ because the background thermal population is a function of frequency, whereas the linewidths differ because of the variation in flux sensitivity vs. frequency. (b) Example of Gaussian fit to lineshape used to extract qubit linewidth at $f_{10} = 2.75\,\text{GHz}$. (c) Extracted linewidth versus frequency (tilt bias).

The broadening of the linewidth due to tilt flux noise can be related to programming errors in an implementation of quantum annealing. Namely, each time an annealing experiment is run, one has to specify the values of h and J in the notation of the Ising Hamiltonian Eq. 1.5 (remember, in the double-well limit, from Eq. (2.50) we have $h = \varepsilon/2 = I_p\Phi_t$). Flux noise that is slower than the repetition rate of the experiment can be treated as a DC offset to the value of h and J that is different for each experimental run, potentially changing the answer that the annealer gives at the end. The exact lower cutoff for the programming noise integral is expected to roughly be the frequency at which one can re-calibrate the device parameters. A linewidth of 100 MHz will then correspond to roughly of order 100 MHz uncertainty in programming h, which on a full

scale of 3 GHz would correspond to a few percent. When considering quantum annealing applications, one should keep in mind that even a small change in h and J can completely change the ground state when the system is large, since the energy levels will come closer and closer together for large systems.

A common family of techniques, primarily developed in the field of nuclear magnetic resonance (NMR) [124, 125, 126] and more recently applied to superconducting qubits, are "spin echo" techniques in which a sequence of pulses (usually mostly pi pulses) are inserted into the algorithm in order to "cancel out" low frequency noise, by periodically reversing the sign of unwanted phase accumulation. These techniques work well for gate-based implementations where all qubit operations are performed relative to a fixed flux bias point, but it is not clear how one would extend these techniques to flux qubits in quantum annealing where the qubit bias constantly changes over an enormously complicated energy landscape with couplings to many other qubits. Furthermore, in the later stages of annealing, tunneling is very slow making it unclear how to even apply something analogous to echoing. However, this and other potential forms of error correction would be an area worth further study.

4.1.1.1 Geometry dependence of qubit linewidth

It is reasonable to expect that the physical geometry of the fluxmon may have an effect on the level of flux noise. For example, a simple geometric argument can roughly explain the often (but not universally) observed scaling of flux noise in SQUIDs with SQUID geometry [110, 113]. Imagine that we want to calculate the flux noise through a circular

loop of skinny metallic wire, and that the source of the noise is a collection of microscopic spin-$\frac{1}{2}$ magnetic dipole fluctuators uniformly distributed on the wire's surface. We will refer to this loop of wire as the qubit loop. Let us further assume that the fluctuators are independent from one another. Call the thickness (cross-sectional radius) of the qubit wire r, and the radius of the qubit loop R. The flux threaded through the qubit loop by one of the spins at first seems like it would be a complicated calculation, but it can be easily solved using reciprocity. To see what this means intuitively, we can pretend as if the microscopic spin is a tiny current loop with an area A_s and current I_s, yielding its magnetic moment $\vec{\mu}_s = I\vec{A}_s$. If a current I_q flows in the qubit loop, the flux it threads through the imaginary spin loop is $\Phi_s = M_{q\to s} I_q$. Conversely, if a current I_s flows in the tiny spin loop, it induces a flux in the qubit loop $\Phi_q = M_{s\to q} I_s$. Reciprocity is the statement that $M_{s\to q} = M_{q\to s} \equiv M$. Identifying Φ_s with $A_s B(\vec{r})$, where $B(\vec{r})$ is the magnetic field generated by a test current I_q in the big loop, minimal algebra then shows that the flux induced in the big metallic loop is

$$\Phi_q = \mu_s \frac{B(\vec{r})}{I_q}. \tag{4.9}$$

For the ideal toroidal wire geometry, the magnetic field at the surface of the wire generated by a test current I is $\propto \frac{I}{r}$ (via Ampere's law and symmetry), meaning the flux threaded from spin to qubit is $\propto \frac{1}{r}$. If there is a uniform density of independently fluctuating spins, when integrated over the surface area of the wire their flux will then add incoherently to yield a flux noise power proportional to $\langle \Phi^2 \rangle \propto \frac{R}{r}$. Using instead the thin-film geometry more appropriate for a qubit wire fabricated with thin-film deposition would

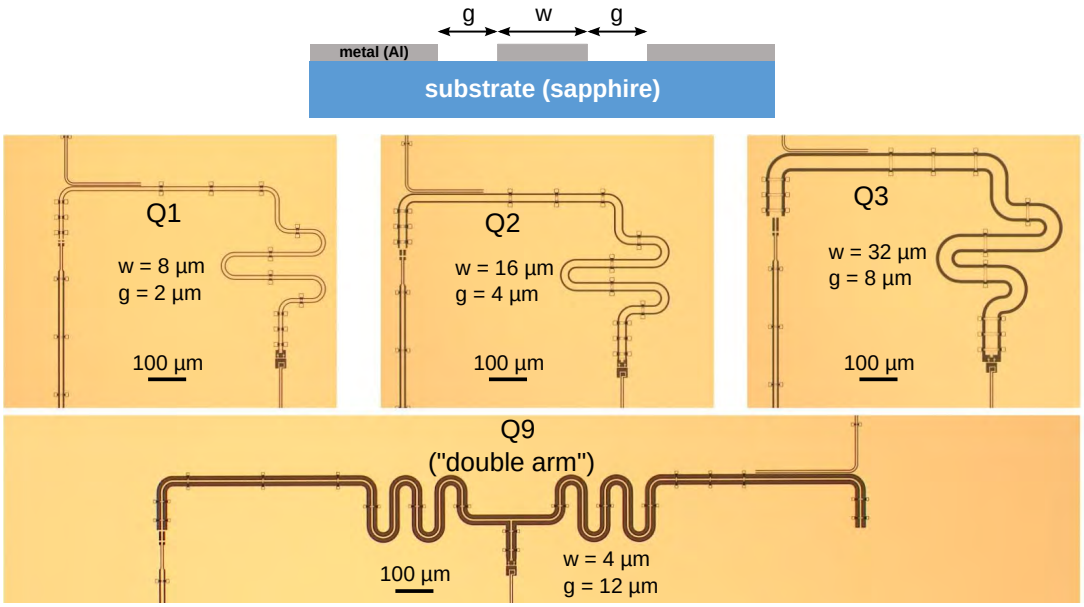


Figure 4.4: Width variation sample. Subset of four fluxmon qubits on the same chip with differing CPW geometries. The geometries were chosen so that all qubits all have $f_{LC} \approx 20\,\mathrm{GHz}$ and $Z_{LC} \approx 75\,\Omega$. Note that the CPW characteristic impedance for the "double arm" geometry is different, but the lumped element impedance for both CPW arms in parallel ends up being the same as the single-arm qubits.

yield a slightly modified result by a logarithmic factor, $\langle \Phi^2 \rangle \propto \frac{R}{W}\left[\frac{\ln(2bW/\lambda^2)}{2\pi} + 0.27\right]$, where here W is the width of the thin-film wire [110]. The CPW geometry of the fluxmon is more complicated than this, but numerical COMSOL simulations of the CPW current flow show that for a uniform independent spin model we should still expect a noise power proportional to ℓ/W, where here ℓ is the length of the CPW, and this fact does not materially depend on the superconductor penetration depth, at least for the dimensions considered.

As an initial test of geometry dependence of the fluxmon's coherence, we fabricated a sample with several qubits of different CPW widths on the same chip and measured these qubits during the same cooldown. Four of these qubits are shown in Fig. 4.4.

Qubit	Z_{LC} (Ω)	w (μm)	g (μm)	Saturation linewidth HWHM (MHz)	Equivalent tilt flux noise ($\mu\Phi_0$)
1	74	8	2	110 ± 8	54
2	74	16	4	100 ± 6	50
3	74	32	8	80 ± 8	40
5	108	20	20	80 ± 7	46
6	108	10	10	80 ± 6	46
7	150	4	12	90 ± 7	63
9	74	4	12	90 ± 8	43

Table 4.1: Table of measured saturation linewidths vs. qubit geometry. There is at most a very weak trend, if any, with overall qubit CPW width.

Qubits 1 through 3 are standard fluxmon qubits with equal f_{LC} and equal Z_{LC}, but with different overall scales for the CPW width and gap (keeping their ratios constant). The other qubit shown is a "double-arm" variant that has the same lumped-element f_{LC} and Z_{LC} as the others but a different CPW gap to width ratio (the math works out because the two arms of the qubit are in parallel). There were also three more single-arm qubits not shown in the picture.

In table 4.1, we display the measured saturation linewidths [as defined in Fig. 4.3(c)]. We observe only a very weak improvement, if any, in the linewidth as the overall CPW width increases. Note that, assuming all the noise comes from independent fluctuators on the CPW wiring, we would have expected a 50 % reduction in the linewidth between Q_1 and Q_3, but instead observe a $\sim 25\,\%$ reduction. We would also expect a $\sqrt{2}$ difference between Q_5 and Q_6 but we observe no measurable difference. We note that various sanity checks were performed to make sure the broadening is intrinsic to the qubit and not from bias line noise (an explicit frequency-resolved plot of bias line noise vs. measured flux

noise will be shown in a later section), so this does not explain the lack of scaling. Another possible explanation is that much of the low-frequency flux noise in the qubit's tilt loop actually comes from spin fluctuators located on the skinny DC SQUID wiring, since this wiring shares a boundary with both the CPW (tilt) and SQUID loops (see inset of of device picture in Fig. 3.1). However, later data shows devices in which the SQUID flux noise was improved while the tilt flux noise was not, making this explanation unlikely.

A second possible explanation of the weak scaling with size is that the fluxmon is in a regime where independent spin fluctuators are less dominant than interacting fluctuators, which would invalidate the simple scaling argument we gave earlier. For example, according to a recent theoretical study, if there is an inhomogeneous density of spins on the qubit surface undergoing spin diffusion [127], this inhomogeneous distribution of spins can yield an exponent of $\alpha < 1$ in the frequency dependence and a corresponding weaker than linear dependence of of the noise on overall width of the CPW geometry. However, as we will see in the next section, the slope α of the tilt flux noise is observed to be fairly close to 1, so this description would not completely capture the observed noise.

It is also possible that the level of flux noise in fluxmon devices is low enough that it is no longer dominated by surface spin defects, but perhaps spins in the bulk of the substrate. This is also consistent with the fact that when we try various materials surface treatments, the noise has only gotten worse, not better. However, it is still possible (and very much hoped) that more careful surface treatments and an improved vaccuum environment can reduce the noise further if it indeed originates from chemisorbed or

physisorbed surface spins [128]. In the following section we will see more conclusive results when looking at the frequency-resolved flux noise on a different sample varying the CPW length instead of the width, where we will see a more clear-cut linear scaling (although those results do not distinguish between surface and bulk fluctuators).

4.1.2 Method for frequency-resolved measurement of quasistatic noise

A single dephasing number is not extremely useful for modeling the performance of qubits. This point is understood quite well for gate-based quantum computation, where noise on the timescale of gates has a qualitatively different effect than lower-frequency noise which can be "echoed" away. Furthermore, noise at the frequency of the qubit has yet another qualitatively different effect which is related to dissipation. In this section, we describe how to measure the frequency resolved noise power spectral density in the low frequency range $f \lesssim 10\,\mathrm{KHz}$. In a later section, we will study very high frequency ($f \sim 1\,\mathrm{GHz}$) noise and its effects on qubit coherence, and at the end of this chapter we will look at the effect of both these types of noise on quantum tunneling.

Frequency-resolved measurements of flux noise in SQUIDs have been around for a while [106, 107, 111]. Frequency-resolved measurements of flux noise in SQUID-based superconducting qubits have seen several developments in the past decade. Bialczak *et al.* measured flux noise at very low frequencies by directly measuring the fluctuating resonant response of a phase qubit in the frequency domain. The MIT/Lincoln Labs group has

pioneered various measurements to probe flux noise at higher frequencies [114, 116, 118]. For the fluxmon, since the noise in the main qubit loop in general dominates the qubit decoherence, we will focus specifically on the measurement of tilt flux noise. We refer to noise slower than the experimental repetition rate as quasistatic. To measure the quasistatic flux noise in Φ_t^x (i.e., noise in ε), we use a pulse sequence similar to that used in Ref. [112] combined with the signal processing techniques used in Refs. [115, 116]. This allows us to obtain data both well below and well above 1 Hz, the latter being achieved by directly processing a binary sequence of single-shot measurements.

The measurement works as follows. We treat each experimental repetition as if it had a static flux offset, and repeatedly measure a function $f(\Phi_t^x(t))$ that is sensitive to fluctuations in Φ_t but not to fluctuations in Φ_{SQ}. We do this by performing the pulse sequence illustrated in Fig. 4.5(a). We initialize the qubit in its ground state (by energy relaxation) in the single-well regime at zero tilt. We then symmetrically raise the barrier (this is equivalent to single-qubit annealing) so that in the absence of noise in Φ_t^x there would be probability 0.5 to end up in the left or right well, completely uncorrelated with any previous or future measurement. Deviations from $P = 0.5$ correspond to deviations away from zero tilt in Φ_t^x. We calibrate this experiment by measuring $P_{|R\rangle}$ (the probability of ending up in the right well) as a function of applied tilt bias Φ_t^x, as shown in Fig. 4.5(b). This is precisely the S-curve experiment (single-qubit annealing) described in section 3.4. Since we accurately know the applied flux, this curve gives a direct calibration $P_{|R\rangle}(\Phi_t^x)$ between physical flux and probability, which (as detailed below) can be used

to convert between probability noise and flux noise as long as the flux excursions are small enough that they remain on the linear part of the curve (close to $P_{|R\rangle} = 0.5$). The exact functional form of this curve depends on the ramp rate due to factors including non-adiabatic transitions, the integrated low-frequency flux noise, and thermalization, but it is possible to choose a ramp rate such that the flux fluctuations remain within the linear part of the curve. We do this by choosing a ramp rate that is fast enough to give enough broadening from nonadiabatic transition such that flux noise contributes a small but still measurable contribution to the width of the curve. We can verify this by looking at the raw measurement statistics of the noise measurement when nominally parked at $P_{|R\rangle} = 0.5$. For example, Fig. 4.5(c) gives a histogram of probabilities obtained by averaging every 100 consecutive single-shot samples, along with a Gaussian fit. We also checked experimentally that the quasistatic sensitivity to the barrier bias, $dP_{|R\rangle}/d\Phi^x_{\mathrm{SQ}}$, was less than $\frac{1}{100} \cdot dP_{|R\rangle}/d\Phi^x_t$, which is certainly negligible upon taking the square when comparing the relative contributions of incoherent flux noise.

For flux noise data below $\sim 0.1\,\mathrm{Hz}$, we measure a time series of $P_{|R\rangle}$ (each value being the average of a few hundred consecutive stats) over a total period of $\sim$ 24 hours and use the same signal processing techniques as in Ref. [115] to obtain $S^+_{\Phi^x_t}(f)$. For the data above $\sim 0.1\,\mathrm{Hz}$, we instead obtain $S^+_{\Phi^x_t}(f)$ by processing time series of $N \approx 1$ million single-shot measurements taken with a regularly spaced sampling interval δt (ranging from $10 - 100\,\mu\mathrm{s}$) and use a variant of the signal processing techniques used in Ref. [116]. We refer to this method as a "1-bit detector measurement" because it involves keeping

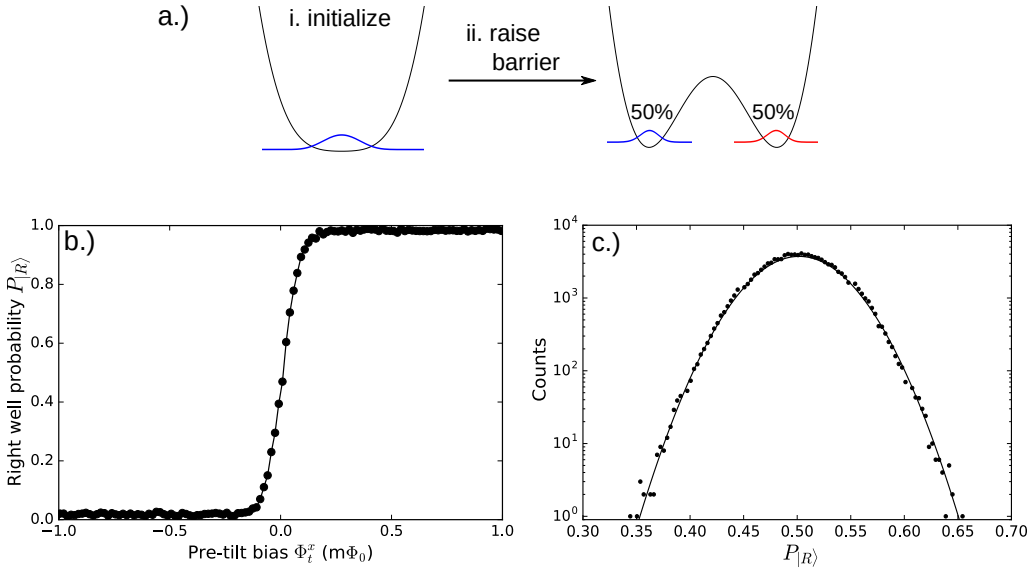


Figure 4.5: (a) Illustration of pulse sequence for the quasistatic flux noise measurement. (b) Calibration curve giving probability to end up in the right well as a function of external tilt flux bias. Note that this is simply the S-curve described in Fig. 3.17 without a pi-pulse. (c) Semi-log histogram of probabilities (obtained by averaging every 100 consecutive single-shot measurements) showing a Gaussian distribution with a standard deviation small enough to be within the linear part of the flux-probability curve.

all single-shot measurement results without explicitly computing probabilities. Given an underlying sequence of flux fluctuations $\{\delta\Phi_n\}$ in the qubit, we obtain a finite probabilistic binary sequence $\{x_n\}$ of length N, with $x_n \in \{-1, +1\}$ corresponding to each single-shot measurement, where $-1/+1$ are assigned to the outcomes $|L\rangle/|R\rangle$ respectively. The probability P_{x_n} of obtaining -1 or $+1$ is related to the underlying sequence $\{\delta\Phi_n\}$ according to

$$\begin{aligned} P_{x_n}(x) = &\delta(x-1)\left[\frac{1}{2} + \frac{dP}{d\Phi_t^x}\delta\Phi_n\right] \\ &+ \delta(x+1)\left[\frac{1}{2} - \frac{dP}{d\Phi_t^x}\delta\Phi_n\right], \end{aligned} \tag{4.10}$$

where for short we now use P to denote $P_{|R\rangle}$ and we have assumed a linear probability-flux curve. Defining the DFT coefficients as

$$\tilde{x}_k \equiv \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/N}, \tag{4.11}$$

then the expected value for the periodogram (PSD estimate) of $\{x_n\}$ can be related to the underlying power spectral density of $\{\delta\Phi_n\}$ according to

$$\begin{aligned}
\langle|\tilde{x}_k|^2\rangle &= \sum_{n,m=0}^{N-1} \langle x_n x_m\rangle e^{-i2\pi(n-m)k/N} \\
&= \sum_{n,m=0}^{N-1} \left[4\left(\frac{dP}{d\Phi_t^x}\right)^2 \delta\Phi_n\delta\Phi_m + \delta_{n,m}\right] e^{-i2\pi(n-m)\frac{k}{N}} \\
&= 4\left(\frac{dP}{d\Phi_t^x}\right)^2 \left|\sum_{n=0}^{N-1} \delta\Phi_n e^{-i2\pi nk/N}\right|^2 + N \\
&= 4\left(\frac{dP}{d\Phi_t^x}\right)^2 \left|\widetilde{\delta\Phi}_k\right|^2 + N,
\end{aligned} \tag{4.12}$$

where in the second line we have used the relation for the correlation

$$\langle x_n x_m\rangle = \begin{cases} 4\left(\frac{dP}{d\Phi_t^x}\right)^2 \delta\Phi_n\delta\Phi_m & n \neq m \\ 1 & n = m \end{cases}$$

computed from (4.10). We can then convert to a physical single-sided flux noise PSD

after assigning a sampling interval time δt to the sequence according to

$$\begin{aligned} S_\Phi^+(f) &= \frac{2T}{N^2}\left|\widetilde{\delta\Phi}_{k=fT}\right|^2 \\ &= \frac{2\delta t}{N}\frac{1}{4\left(\frac{dP}{d\Phi_t^x}\right)^2}\langle|\tilde{x}_{k=fT}|^2\rangle - 2\frac{1}{4\left(\frac{dP}{d\Phi_t^x}\right)^2}\delta t\,. \end{aligned} \tag{4.13}$$

where $T = N\delta t$ is the total acquisition time for the dataset of N samples. The normalization convention for S_Φ^+ is chosen so that the total power in the signal is obtained by integrating over positive frequencies only. It is normalized such that the deterministic signal $\Phi(t) = A\sin(2\pi ft)$, with f in the baseband of the sampled DFT, has total power $A^2/2$ as physically expected. The ability to correctly extract the underlying $S_\Phi^+(f)$ with these formulas was verified by numerically feeding in an artificially generated noise source into a numerical simulation of our measurement, which was in turn fed into the data analysis software.

Equation (4.13) tells us that the PSD of our measurement sequence $S_x^+(f)$ will be a combination of the underlying $S_\Phi^+(f)$ plus a white noise floor. This noise floor makes sense because if there were no flux noise at all, we would expect shot noise from a perfectly uncorrelated probability of 0.5 for each measurement result. This white noise floor can be substantial: for typical parameters of $\delta t = 10 - 100\,\mu$s and $dP/d\Phi_t^x \approx 2000/\Phi_0$, it has an equivalent flux noise amplitude of order $1\,\mu\Phi_0/\sqrt{\text{Hz}}$. Since this is the typical strength expected for the intrinsic $1/f$ flux noise of the device, this means the white noise floor will dominate the signal above $\sim 1\,$Hz.

Fortunately, there is a way to process the data that allows one to drastically reduce

the shot noise floor and infer $S_\Phi^+(f)$ from the measurement of $S_x^+(f)$ without noticeably distorting the underlying flux noise signal [116]. The idea is inspired by the technique of using two separate detectors to sample a signal and computing the cross spectrum to throw away the detectors' contribution to the noise. By breaking $\{x_n\}$ into two interleaved series and taking the cross-spectral density (CSD) of the two sub-series, in the limit of infinite statistics this would completely eliminate the white noise floor, as shown below. The intuition behind this is that the interleaving removes the zero-delay autocorrelation term because the point $\delta\Phi_n$ never "sees itself" in the sum. However, the actual datasets are finite, so as discussed below this cancellation will not be perfect. Regarding the underlying $1/f$ signal itself, at least at low enough frequencies, the two interleaved flux noise signals should be approximately equal because the noise is highly correlated, and so the spectrum should not become distorted. This will be shown numerically below.

Mathematically, we define the two subsequences

$$x'_n = x_{2n}, \;\; x''_n = x_{2n+1} \tag{4.14}$$

$n = 0, 1, 2, ..., M-1$, where $M = N/2$. We define a CSD for these interleaved sequences multiplied by a particular frequency-dependent phase factor:

$$\left\langle \tilde{x}'_k \left(\tilde{x}''_k\right)^* \right\rangle e^{i2\pi k/N} \tag{4.15}$$

$$= \sum_{n,m=0}^{M-1} \langle x'_n x''_m \rangle e^{i2\pi(n-m)k/M} e^{i2\pi k/2M}$$

$$= 4\left(\frac{dP}{d\Phi_t^x}\right)^2 \sum_{n,m=0}^{M-1} \delta\Phi_{2n}\delta\Phi_{2m+1} e^{-i2\pi(n-m-\frac{1}{2})k/M}, \tag{4.16}$$

where the extra phase factor in line (4.15) corrects for the phase shift arising from the time domain offset between the two interleaved sequences. This phase offset ensures that the interleaved CSD of a deterministic sinusoidal signal with integer wave index q is itself real.

Equation (4.16) says that if we perform the interleave and take the CSD in the absence of underlying flux noise, we would observe absolutely no noise (since $\delta\Phi_n$ is identically zero). Of course, the only reason the noise cancellation is exact is due to the use of the expectation value operator, $\langle\cdot\rangle$, which means an average over an ensemble of infinitely many realizations of $\{x_n\}$ generated by a given $\{\delta\Phi_n\}$. In reality, we deal with finite sequences ($N \approx 10^6$), and instead of (4.15), we can only compute the CSD coefficients $\tilde{x}'_k \left(\tilde{x}''_k\right)^* e^{i2\pi k/N}$ for a finite number of finite sequences. The simplest way to understand the effect of interleaving on the white noise floor in the actual experiment, then, is to think about a random walk of phasors in the complex plane. First, let us consider how we can reduce the noise floor within a single dataset, and then we will consider averaging over multiple datasets. When $\{\delta\Phi_n\}$ is identically zero (no flux noise), each CSD coefficient of the 1-bit detector measurement will itself a Gaussian distributed complex random variable, with uniformly distributed phase and some distribution of magnitude that is peaked close to $\delta t/(dP/d\Phi_t^x)^2$. Thus, the white noise floor will not actually be reduced unless we perform some sort of averaging, either across frequency bins or across datasets. Since it is informative to plot flux noise on a log-log scale, it is natural to use a logarithmic averaging scheme where the number of neighboring frequency bins whose CSD coefficients

are averaged together is proportional to frequency. In other words, the number of bins per decade of frequency used in the averaging is constant, as we will do for the experimental quasistatic flux noise data in the next section. Since the number of points N averaged in a frequency bin is chosen to scale as f, and sum of N random phasors scales in magnitude as $\sqrt{N}$, then upon taking the average over each frequency bin we would expect the white noise floor power to decrease as $1/\sqrt{N}$, meaning as $1/\sqrt{f}$. Thus, the act of interleaving and taking a coherent logarithmic frequency average within a given dataset *amounts to a* $1/\sqrt{f}$ *filter for the white noise floor*. This can be seen from the slope of the observed filtered white noise floor in the numerical simulation of Fig. 4.6.

However, we find that the filtered suppression of the white noise floor from averaging over frequency bins within a single dataset alone is not enough to bring the white noise floor below the qubit flux noise signal at high frequencies. Because of this, we further average the CSD coherently over K datasets, with K a few hundred, before taking the real part. This gives a factor of $\sqrt{K} \approx 10$ further reduction in the filtered white noise floor power without distorting the underlying (presumably) correlated flux noise signal, according to the numerical simulation of the measurement shown in Fig. 4.6. This simulation uses artificially generated $1/f^{\alpha}$ noise signals with $\alpha = 1$ and $\alpha = 0.9$ and magnitude at 1 Hz equal to $5\,\mu\Phi_0/\sqrt{\text{Hz}}$, the value extracted from experiment. We note that the discretely generated artificial $1/f$ noise extends an order of magnitude higher in frequency than the sampling frequency of the simulated 1-bit detector measurement. This was done purposefully to make sure there is no influence from aliasing. We find that

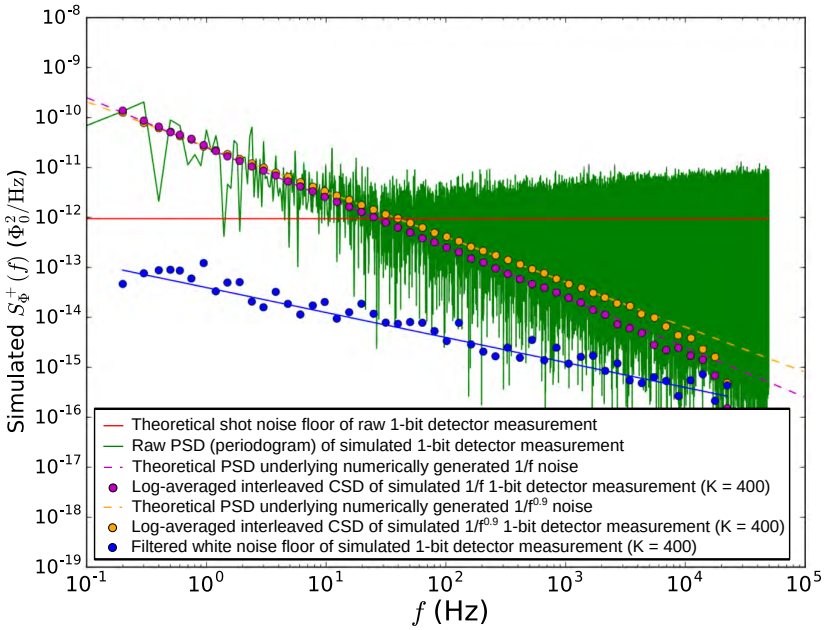


Figure 4.6: Numerical simulation of the 1-bit detector flux noise measurement for artificially generated $1/f$ and $1/f^{0.9}$ noise. The red line shows the expected white noise level for the raw 1-bit detector measurement without interleaving. Green shows the raw PSD of the simulated 1-bit detector measurement for a single realization of the numerically generated $1/f$ noise. The dashed magenta line shows the theoretical PSD used to generate the artificial $1/f$ noise. Magenta circles show the result of the simulated 1-bit detector measurement using the same processing that was used on the experimental data (interleaved + log-averaged over frequency, averaged over $K = 400$ realizations of the numerically generated $1/f$ noise). Orange: same but for numerically generated $1/f^{0.9}$ noise. Blue circles show the white noise floor of the simulated measurement after interleaving and averaging, with the blue line its theoretically expected level.

the interleaving technique greatly reduces the effects of aliasing that would otherwise be present from the substantial amount of $1/f$ noise that is likely present above the Nyquist frequency of the measurement. From the simulation, we see that the 1-bit detector measurement after interleaving and coherently averaging over datasets would faithfully reproduce the $1/f$ power spectrum without any significant distortions except for the highest half-decade of frequency. The filtered white noise is well below the simulated $1/f$ noise (below the highest half-decade of frequency), which means it is even further below the experimentally measured flux noise (because of the significant "bump" observed starting at $10 - 100\,\mathrm{Hz}$ in the experimental data).[3] As an extra consistency check, we note that the extracted flux noise data in the experiment was not materially changed (apart from the highest factor of 2 in frequency) whether we coherently averaged over 250 or 500 datasets.

4.1.3 Experimental flux noise spectrum: geometry dependence

In Fig. 4.7, we show the experimental results of measuring the quasistatic flux noise on two samples: the width variation sample described in an earlier section (Fig. 4.4), and also a length variation sample, shown in Fig. 4.8. Consistent with the linewidths in table 4.1, in Fig. 4.7(a) we are unable to extract a trend of flux noise power vs. qubit CPW width. However, in Fig. 4.7(b), we see a clear trend with CPW length, roughly (but not exactly) a factor of two increase in flux noise power for each factor of two increase

[3]We note that the "bump" in noise around 100 Hz in the experimental data is unchanged when the pulse tube compressor of the dry dilution refrigerator is turned off.

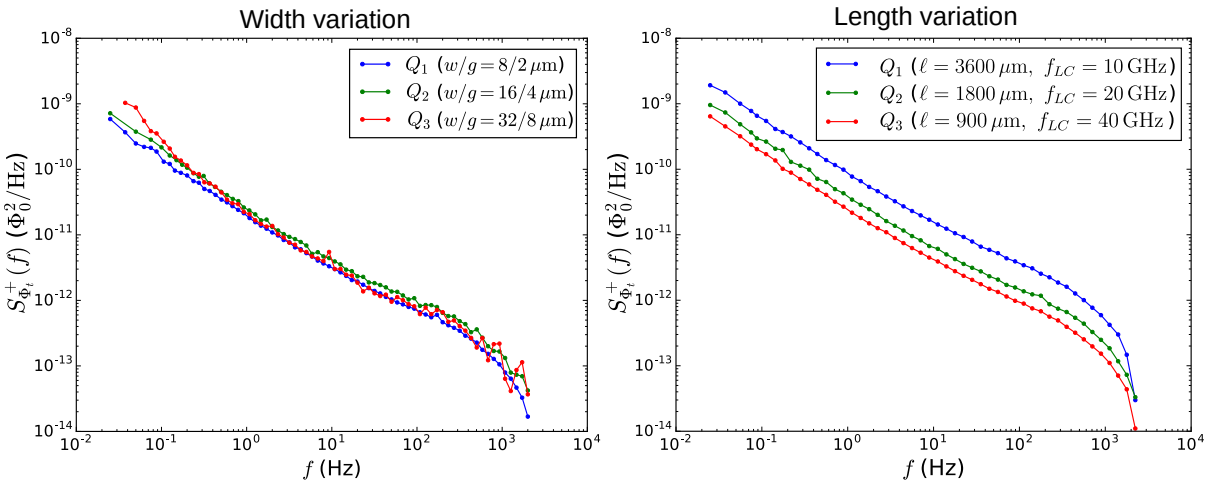


Figure 4.7: Experimentally measured frequency-resolved flux noise for two sets of fluxmon qubits. This particular data was taken using 800,000 points at a sampling interval of 100 μs and coherently averaging over $K \approx 250$ datasets as per the analysis procedure described in the previous section. (a) Width variation sample (constant length of 1750 μm, see Fig. 4.4). (b) Length variation sample (dimensions shown in Fig. 4.8).

in CPW length. This rough scaling with CPW length is consistent with most of the tilt flux noise originating in the CPW (this scaling doesn't depend on whether or not the noise comes from the surface or bulk, since the CPW is uniform in the dimension of its length). However, there is caveat in making this conclusion, because the DC SQUID design is also different between the qubits; namely, in order to ensure good readout for each qubit, $\beta_{\max}$ was designed to be constant, meaning that an increase in qubit length implies a proportional decrease in qubit junction width, and hence a skinnier vertical SQUID wire (this segment is just 5 μm long). If the noise was dominated by the skinny vertical SQUID wire, this could also explain the data. However, as mentioned before, we think that this explanation is unlikely (see the discussion of silicon fluxmons later on). It might be possible to directly test this by designing a length variation sample where the DC SQUID resistance (and therefore the DC SQUID geometry) remains constant. This

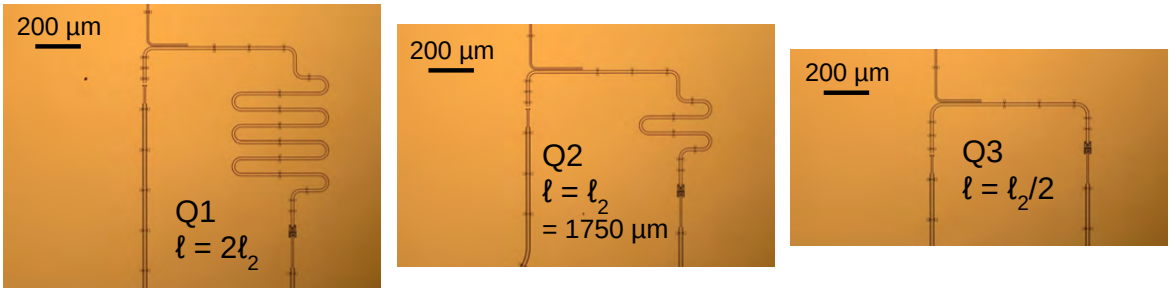


Figure 4.8: Length variation sample. Optical micrographs of the three qubits of different lengths measured on the same chip during the same cooldown. Q_1, Q_2 and Q_3 have $f_{LC} \approx 10, 20,$ and 40 GHz respectively. All qubit CPWs had cross-sectional dimensions $w, g = 8, 2\,\mu$m, and $Z_{LC} = 75\,\Omega$.

would necessitate a significant change in the readout parameters if it is possible at all.

For the width variation sample, the noise power at 1 Hz is roughly $2.1 \pm 0.3 \times 10^{-11}\,\Phi_0^2/\text{Hz}$, or equivalently $4.6 \pm 0.2\,\mu\Phi_0/\sqrt{\text{Hz}}$, for all qubits. For the length variation sample, the noise powers for qubits 1, 2, and 3 were $9.0 \times 10^{-11}\,\Phi_0^2/\text{Hz}$ ($9.5\,\mu\Phi_0/\sqrt{\text{Hz}}$), $3.9 \times 10^{-11}\,\Phi_0^2/\text{Hz}$ ($6.3\,\mu\Phi_0/\sqrt{\text{Hz}}$), and $2.5 \times 10^{-11}\,\Phi_0^2/\text{Hz}$ ($5.0\,\mu\Phi_0/\sqrt{\text{Hz}}$) respectively. We see that the difference in noise power between the shortest qubit (Q_3) and middle qubit (Q_2) is not quite a factor of two, which is either the result of experimental variation or a reflection of the fact that as the noise originating on the CPW metal decreases, the noise contributions from elsewhere (such as the DC SQUID wire) become more noticeable. However, quoting a number at 1 Hz is not very useful, as the effect of the noise depends on its power integrated over many orders of magnitude in frequency, and is therefore also strongly dependent on the slope α in the $1/f^\alpha$ power law. We typically extract an α between 0.9 and 1.0 below ~ 10 Hz and $\alpha \approx 0.65 - 0.7$ above 10 Hz (up to the artificial data processing roll-off at ~ 1 KHz).

We emphasize that the width dependence data was taken with all qubits a single chip

during the same cooldown, as was the length dependence data. This is important unless we are able to have statistics over many fab runs/cooldowns, because the noise power can vary as much as a factor of ~ 2 between fab runs for nominally the same qubit design, and can vary on order of 10 - 30% on a given qubit between cooldowns.

4.2 Dissipation and high-frequency noise

Low frequency noise is by no means the only way the environment can couple to the fluxmon qubit. Noise at a frequency that matches a transition frequency between two energy levels of the system can drive unwanted incoherent transitions between them. A well-known consequence of this in superconducting qubits is T_1 decay and leakage into non-computational states, two of the primary limits on performance in transmon qubits for gate-based quantum computing [129, 130]. Furthermore, dissipation not only at GHz frequencies but also at much lower frequencies can be strongly relevant to slower processes like quantum tunneling, as first studied in the 1980's in Josephson junction and SQUID devices [21, 131, 22], which is relevant for quantum annealing applications. Intuitively, dissipation keeps quantum systems trapped at the bottom of potential energy wells by damping the oscillation associated with that well, or more rigorously by reducing the zero-point variance of the flux and localizing the density matrix [91]. One might at first think that this is what one actually wants for quantum annealing, since the ground state of the system is the goal anyway. However, this way of thinking is misleading, as this will trap the system as a whole in a local, not global ground state. Technically if one waits

long enough, the energy relaxation could in fact pull the system to its global ground state, but this can take an exponentially long time and therefore cannot be used for solving hard computational problems in practice. High frequency noise can also excite the system out of the ground state, although in quantum annealing this could be a good or bad thing depending on the potential landscape and when the transition occurs during the anneal. Dissipation can also help bring the system back to its ground state if it is excited by a Landau-Zener transition during the annealing process, or conversely if the system was excited before the Landau-Zener transition, it will end up back in the ground state afterwards. In other words, excitation could perhaps sometimes help in quantum annealing. This is obvious in the case of thermally assisted tunneling, especially for wide barriers where quantum tunneling is bad and thermal activation will actually be faster [132, 23]. However, if low-frequency dissipation is too substantial, it imposes extra barriers that suppress quantum tunneling, a serious concern for building quantum annealers.

Physically, high-frequency noise is intimately related to dissipation, which can sometimes lead to potentially confusing language. There is however a rigorous definition of both classical and quantum noise and their relation to dissipation, a concept on which we will elaborate in this section. For a system in thermal equilibrium, there is a simple relationship between noise and dissipation, namely the fluctuation-dissipation theorem which covers both classical and quantum noise. We will discuss what this theorem means as well in this section. However, as we will see later on when we measure dissipation in

the fluxmon, things can become less clear when the environment is not in equilibrium (or when its consists of several environments with their own quasi-equilibria).

4.2.1 Dissipation in quantum circuits

Accounting for dissipation in linear, classical circuits is relatively straightforward. There, dissipation can be represented by a set of (possibly frequency-dependent) resistors. Any classical linear circuit element is characterized by an impedance, which can always be written as the series combination of a lossless reactance and one lumped element resistor, i.e. $Z = R + iX$. It is also possible to instead express dissipation as one lumped parallel resistor, a duality that is analogous to the Thévenin-Norton equivalence for voltage and current sources [133], as illustrated in Fig. 4.9(a). In the small-dissipation limit, it is easily derived that the equivalent R_p is

$$R_p = |Z|^2/R_s \tag{4.17}$$

by computing the impedance of the circuit on the left and taking the small-dissipation limit $R_s \ll |Z|$. A handy way to remember this is that both representations should give the same loss tangent [97],

$$\frac{R_s}{|Z|} = \frac{|Z|}{R_p}, \tag{4.18}$$

where low dissipation corresponds to either a low series or high parallel resistance. In the following we will choose to represent dissipation in parallel, because admittances add in parallel and are therefore easier to incorporate into our circuit diagrams. This representation also often arises naturally. We can then treat the damping of a classical

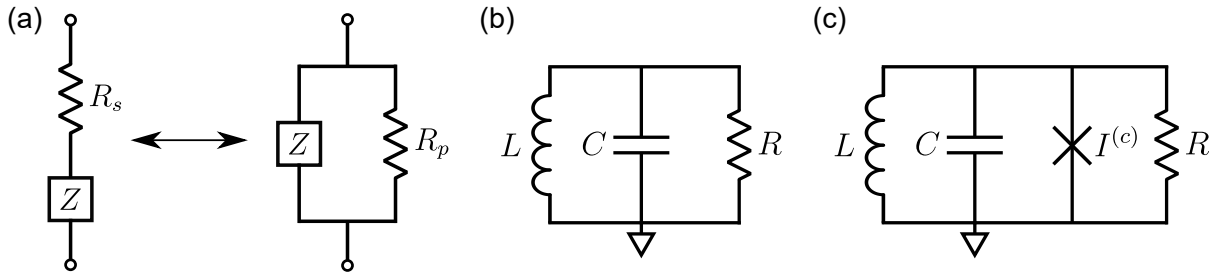


Figure 4.9: (a) "Thévenin-Norton" equivalent representations of dissipation in classical linear circuits as series or parallel resistance. (b) Classical parallel RLC circuit. A linear LC oscillator (once rung up) will have its energy exponentially damped vs. time according to $E \propto \exp(-t/RC)$, where R is assumed to be in parallel. (c) Nonlinear qubit circuit with lumped parallel dissipation.

linear oscillator, such as the LC circuit in Fig. 4.9(b), by adding a parallel admittance. As commonly studied in classical circuits, if such an oscillator is rung up and oscillating at its resonance frequency, its energy will decay over time with a time constant $\tau = RC$.

It is not immediately obvious how to extend this "resistor" representation of dissipation to quantum circuits, for two reasons. One is that circuits with Josephson junctions are nonlinear, unlike the linear LC circuit considered in Fig. 4.9(b). Second (and perhaps more fundamental), the Hamiltonian formalism that we use for quantum circuits leads to inherently reversible dynamics, making dissipation at first seem unnatural. It will turn out that the first concern is not difficult to deal with once we figure out how to deal with the second concern, because even if the qubit itself is not linear, the environment that represents dissipation usually is. To address the second concern, the resolution is to add more degrees of freedom the system corresponding to the "environment," such that when looking at the qubit, these extra degrees of freedom are traced over. The dynamics of the joint system can still be unitary, but the effective dynamics of the qubit density matrix become irreversible.

In the following, we will explain the Caldeira-Leggett model for quantum dissipation, where a dense bath of lossless elements can effectively lead to irreversibility. This model replaces an arbitrary dissipative environment with a dense collection of bosonic modes (in the context of circuits, LC oscillators). It turns out that using bosonic modes is very generic, even if the actual microscopic nature of the environment does not consist of bosons, under the assumption that the bath is in the regime of linear response (it is only ever slightly perturbed so as not to undergo any type of saturation) and that its fluctuations are Gaussian [134]. For example, even if the qubit is coupled to a fermionic bath (or a bath of two-level systems within its dielectric materials, which we will discuss in section 4.2.2), in the unsaturated regime we can still apply a bosonic model like Caldeira-Leggett to describe the resulting dissipation. If we further assume the environment is in thermal equilibrium, we will see how to use the Caldeira-Leggett model to derive an important result called quantum fluctuation-dissipation theorem. Finally, we will use Fermi's golden rule to relate noise and dissipation, which will be important in understanding the experimental measurements of dissipation and noise in the fluxmon qubit, and to predict the dissipation induced by various known dissipation sources such as bias lines.

4.2.1.1 A classical intuition: infinite transmission line as a resistor

We have in fact already dealt with a well-known example where a real admittance can arise from an infinite collection of infinitesimal lossless elements, namely an infinitely long transmission line with uniform characteristic impedance $\sqrt{\mathcal{L}/\mathcal{C}}$. Mathematically,

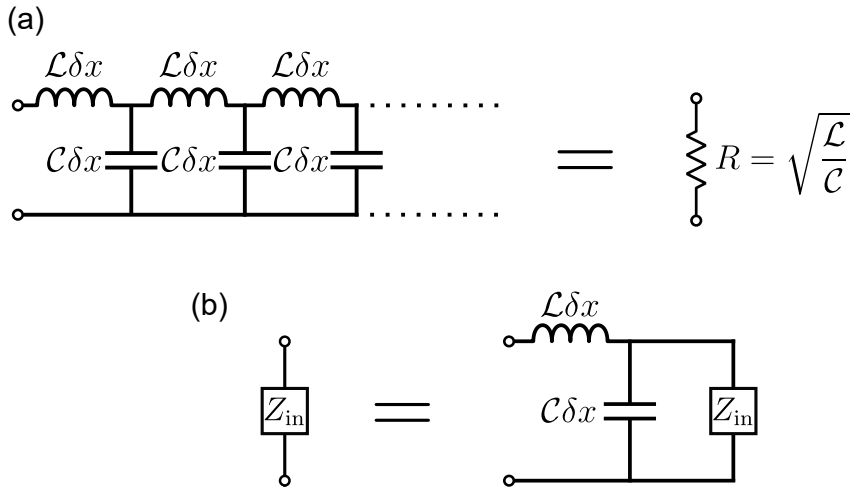


Figure 4.10: (a) Ladder model for a transmission line, equivalent to a resistor in the continuum and infinite length limits. (b) Recursive trick to obtain the input impedance of the ladder network in the limit of infinite rungs.

one can arrive at the input impedance of the network in a non-rigorous way if one notes that adding one more rung of $L = \mathcal{L}\delta x$ and $C = \mathcal{C}\delta x$ should not change the result [see Fig. 4.10(b)], leading to the recursive relation $Z_{\text{in}} = Z_L + (Z_C || Z_{\text{in}})$ which is easily solved. Physically, the "continuum limit" occurs when the wavelength of wave propagation is much longer than the spacing δx between ladder elements. In this limit, we obtain the simple result $Z_{\text{in}} = \sqrt{\mathcal{L}/\mathcal{C}}$.

However, mathematically it is puzzling how the result could be a real number, if for any finite number of rungs on the ladder the impedance is purely imaginary. Physically, this result appears to make sense since any energy carried by an incident wave at any frequency on such an ideal transmission line will suffer no reflections and continue to propagate to infinity, meaning its energy is irreversibly lost from the point of view of the source, which is indistinguishable from a load resistor matched to the source impedance. Mathematically however, there is still something to be desired in this explanation. As

we will see shortly, the resolution is that the concept of impedance assumes an infinite time limit, and this needs to be properly incorporated into the infinite transmission line, a process which involves a second limit, namely the limit of infinite modes. As we will see, this is essentially achieved by switching the order of the two limits, which can be visualized by adding finite damping to each rung before taking the limit of a lossless transmission line.

In the following, we will extend this idea to model dissipation in quantum circuits. The idea will be that irreversibility comes in the limit of infinitely many modes in the system. All lossless Hamiltonian systems will have a Poincaré recurrence time associated with them [54]. What we are doing here is taking a limit when this recurrence time is infinity, or at least larger than any physically relevant timescale. In the language of quantum theory, this corresponds to having energy levels become arbitrarily close together, meaning any periodicities associated with these transitions tends to infinity. This is not a new idea, and in fact analogous models have been around for a very long time to explain the exponential decay of a two-level atom into the continuum of electromagnetic vacuum modes. In quantum optics, the Weisskopf-Wigner theory of spontaneous emission assumes that the atom interacts via a Jaynes-Cummings Hamiltonian with each mode of the electromagnetic field [135, 136]. In the vacuum limit, these modes form a continuum, and so the coherent "swapping" of photons between the atom and each mode interfere in such a way as to induce exponential leakage of the photon to the continuum (within a Markov approximation). Another example using a similar concept was

introduced by Feynman and Vernon [137]. In a similar spirit to these earlier models, the Caldeira-Leggett model discussed below extends this idea to linear dissipative systems in a way that is easily applicable to quantum circuits, by mapping the electromagnetic environment to a dense bath of weakly coupled LC oscillators.

4.2.1.2 Resistors in quantum circuits: Caldeira-Leggett model

The problem at hand is the following: how can we write down a Hamiltonian for an arbitrary environmental admittance $Y(\omega)$ (with an ideal resistor $Y = 1/R$ being a special case)? This is the essence of the Caldeira-Leggett sum-of-oscillators model, which originates in Ref. [89] (which studied the effect of dissipation on quantum tunneling) and was applied to the circuit model in Ref. [138]. This model is illustrated by the indistinguishability of the two circuits of Fig. 4.11. Once we have a Hamiltonian for the "resistor" part of the circuit in terms of harmonic oscillators, we will be able to straightforwardly compute useful quantities such as the quantum thermal distribution of noise generated by such an environment and how a qubit coupled to that environment will be damped or excited. We will outline a derivation that was first introduced by Devoret *et al.* [138], but will try to introduce some of the relevant mathematical tools (such as the concept of impedance at complex frequencies) in a more physically intuitive way as well as correct a few minor errors along the way.

Suppose we have an environmental admittance $Y(\omega)$ as illustrated on the left hand side of Fig. 4.11, which we want to map to the collection of oscillators depicted in the middle circuit of Fig. 4.11 (it is convenient to assume that the oscillators are parallel

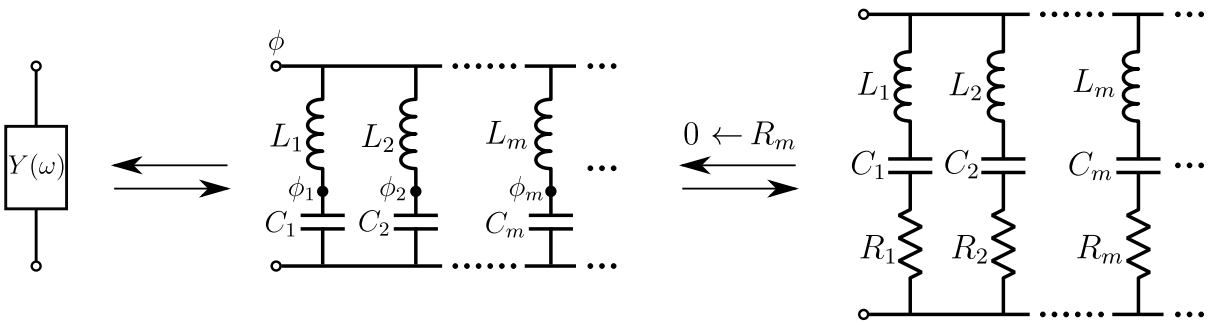


Figure 4.11: Illustration of the Caldeira-Leggett model for an arbitrary admittance $Y(\omega)$ as an infinite collection of lossless oscillators. The node at the bottom of the circuit is ground.

combinations of series LC oscillators). The task is the find the distribution of L_m, C_m that will yield an input admittance $Y(\omega)$ across the terminals in the continuum limit. However, each oscillator contributes a purely imaginary admittance to the system,

$$Y_m(\omega) = \left[i\omega L_m + \frac{1}{i\omega C_m} \right]^{-1}, \tag{4.19}$$

the same problem we had in our earlier discussion of the transmission line. The resolution is to realize that something degenerate happens exactly on resonance, where the admittance diverges, i.e. at $\omega = \omega_m = 1/\sqrt{L_m C_m}$. Properly accounting for this pole will allow us to proceed. Devoret *et al.* addresses this problem by generalizing the definition of impedance to complex frequencies and taking the limit as the imaginary part of the frequency goes to zero, i.e., $Y[\omega] = \lim_{\eta \to 0^+} \int_{-\infty}^{+\infty} dt\, \widetilde{Y}(t) \exp[i(\omega + i\eta)t]$, where the time-domain linear response function $\widetilde{Y}(t)$ is defined by the response of the current to an applied voltage, $I(t) = \int_{-\infty}^{\infty} dt'\, \widetilde{Y}(t') V(t - t')$ (this is what we mean when we say impedance is defined in the infinite time limit). A more physical way to arrive at the same result is to add a finite amount of damping to each of the modes, represented by the resistors R_m in the circuit on the right hand side of Fig. 4.11, and then later taking

the limit of zero mode damping. One way to understand this is that R_m gives a finite response time to each oscillator, allowing us to first take the limit of infinite oscillators before taking the limit of infinite time at the end of the calculation. Within the finite mode damping model, the expression (4.19) is modified to

$$Y_m(\omega) = \left[i\omega L_m + \frac{1}{i\omega C_m} + R_m \right]^{-1}. \tag{4.20}$$

Defining the oscillator admittance $y_m = \sqrt{C_m/L_m}$ and quality factor $Q_m = 1/(y_m R_m)$, then to leading order in $\omega - \omega_m$ this becomes

$$Y_m(\omega) \approx -i\frac{y_m}{2} \frac{1}{\frac{\omega-\omega_m}{\omega_m} - \frac{i}{2Q_m}}. \tag{4.21}$$

Next, we can take the limit of zero mode damping by taking $Q_m \to \infty$. When we take this limit, Y_m becomes proportional to $1/(\omega - \omega_m)$, but since there is an i in front of the expression, it still looks like the impedance will be imaginary in this limit as well; however, this is not the case. When taking the $Q_m \to \infty$ limit, we must apply the following well-known formula from the theory of distributions [139],

$$\lim_{\varepsilon \to 0} \frac{1}{(x - x_0) \pm i\varepsilon} = \text{p.v.} \left(\frac{1}{x - x_0} \right) \mp i\pi\delta(x - x_0). \tag{4.22}$$

where p.v is the Cauchy principal value.[4] Note that this picks up a factor of i in the second term. Applying this with $\varepsilon \propto 1/Q_m$ to one of our environmental oscillators near resonance yields

$$Y_m(\omega) = -i\frac{y_m}{2} \left[\text{p.v.} \left(\frac{\omega_m}{\omega - \omega_m} \right) + \text{p.v.} \left(\frac{\omega_m}{\omega + \omega_m} \right) \right] + \frac{\pi}{2} y_m \omega_m \left[\delta(\omega - \omega_m) + \delta(\omega + \omega_m) \right], \tag{4.23}$$

[4]The Cauchy principal value distribution is defined on a smooth test function $f(x)$ according to $\left[\text{p.v.}\left(\frac{1}{x}\right)\right](f) \equiv \lim_{\varepsilon \to 0^+} \int_{|x| \geq \varepsilon} \frac{f(x)}{x} dx$.

and in particular we now have the real contribution

$$\mathrm{Re}[Y_m(\omega)] = \frac{\pi}{2} y_m \omega_m \left[\delta(\omega - \omega_m) + \delta(\omega + \omega_m)\right]. \tag{4.24}$$

Note that we added a delta function not only at ω_m but also at $-\omega_m$, which is required for $\widetilde{Y}(t)$ to be a causal function in the time domain (this arises because mathemetically there are really two resonance frequencies, leading to another expansion of the form (4.21) about $-\omega_m$). This will turn out to be important when we do a quantum calculation of the noise and dissipation induced by $Y(\omega)$.

We can now use this result to create any impedance $Y(\omega)$ via a dense comb of delta functions according to the construction

$$\omega_{m\neq 0} = m\Delta\omega \tag{4.25}$$

$$y_{m\neq 0} = \frac{2}{\pi m}\mathrm{Re}[Y(m\Delta\omega)] \tag{4.26}$$

(from which L_m and C_m can be deduced), with

$$Y(\omega) \approx \sum_m \mathrm{Re}[Y(\omega_m)]. \tag{4.27}$$

When the delta function comb is dense enough, the real part of $Y_m(\omega)$ will dominate the imaginary contributions from the other oscillators at $\omega = \omega_m$ (since it is a delta function), meaning the admittance will be purely real, as was sought from the start.

The Caldeira-Leggett sum-of-oscillators model does not necessarily capture the microscopic internal workings of the dissipative environment $Y(\omega)$, but since its linear response is indistinguishable, for the sake of calculation we can assume that it indeed represents

the environment. In particular, we can now write down the Hamiltonian for the $Y(\omega)$ part of the circuit, by reading it off the Caldeira-Leggett circuit diagram,

$$\hat{H}_{\text{env}} = \sum_m \left[\frac{\hat{Q}_m^2}{2C_m} + \frac{(\hat{\Phi}_m - \Phi)^2}{2L_m} \right], \tag{4.28}$$

where the node fluxes $\phi_m = (2\pi/\Phi_0)\Phi_m$ are defined in Fig. 4.11 and $q_m = 1/(2e)Q_m$ is the corresponding conjugate charge on C_m.

In an upcoming section, we will apply this Hamiltonian model to calculate the decay of the fluxmon to an arbitrary dissipative environment via a Fermi's golden rule calculation. To do this within the Hamiltonian formalism, ultimately we need to write down an interaction Hamiltonian between the qubit and the reservoir of oscillators, assume some initial state for the qubit, and then compute resulting expected dynamics under some assumption about the distribution of occupied oscillator states in the bath (usually but not always this will be the assumption of thermal equilibrium). Below, we compute the expected double-sided "quantum" spectral density of current and voltage fluctuations across the leads of $Y(\omega)$ in thermal equilibrium, and will then show how this can be used to derive the resulting dissipation (both relaxation and excitation) of the qubit via Fermi's golden rule. In the next section we assume thermal equilibrium, but we will see in the following section that the Fermi's golden rule part of the calculation does not rely on any such assumption.

4.2.1.3 Fluctuation-dissipation theorem

The aim of this section is the compute the spectral density of current or charge fluctuations induced across the leads of the dissipative environment $Y(\omega)$ in thermal equilibrium, which will turn out to be intimately related to the dissipation induced by the environment on whatever it is coupled to. We do this by pretending the environment is described by an equivalent collection of Caldeira-Leggett oscillators with Hamiltonian (4.28) as derived above. We first consider current fluctuations induced across the leads when they are short-circuited. In this case, ϕ is fixed at a constant value of zero (the phase at the ground node), and the parallel oscillators are then all independent, so we can consider one oscillator at a time,

$$\hat{H}_m = \frac{\hat{Q}_m^2}{2C_m} + \frac{\hat{\Phi}_m^2}{2L_m} \tag{4.29}$$

$$= \hbar\omega_m \left(\hat{a}_m^\dagger \hat{a}_m + \frac{1}{2}\right) = \frac{1}{2}\hbar\omega_m \left(\hat{a}_m^\dagger \hat{a}_m + \hat{a}_m \hat{a}_m^\dagger\right), \tag{4.30}$$

which has the usual creation and annihilation operators defined by $\hat{\Phi}_m = \Phi_{\mathrm{m,zp}}(\hat{a}_m^\dagger + \hat{a}_m)$ and $\hat{Q}_m = iQ_{\mathrm{m,zp}}(\hat{a}_m^\dagger - \hat{a}_m)$. We will compute the thermal spectrum by first calculating the time correlation function of the charge (since charges add in parallel) and then invoking the Wiener-Khinchin theorem, which states that the spectral density is just the Fourier transform of the time correlation function (stay tuned for a derivation of this theorem in a later section). Because there is a fundamental and important difference between taking the spectral density of an operator as opposed to a scalar function of time, we will

explicitly keep writing hats on all operators in this section.

To compute the time correlation function of charge fluctuations, we use the fact that the time dependence of any observable operator in the Heisenberg picture is given by

$$\hat{A}(t) = e^{i\frac{\hat{H}}{\hbar}t}\hat{A}(0)e^{-i\frac{\hat{H}}{\hbar}t}. \tag{4.31}$$

Next, we need to compute the expected time-correlation function over a thermal ensemble for the LC oscillators. For any operator, the thermal expectation value is given by

$$\langle\hat{A}\rangle = \frac{1}{Z}\mathrm{tr}\left[\hat{A}e^{-\beta\hat{H}}\right], \tag{4.32}$$

where $\beta = 1/(k_BT)$ and $Z = \mathrm{tr}\left[e^{-\beta\hat{H}}\right]$ is the partition function. Using the identity (4.31), we can simply use linearity to compute

$$\langle\hat{Q}_m(t)\hat{Q}_m(0)\rangle = Q_{m,zp}^2\left(\langle\hat{a}_m^\dagger\hat{a}_m\rangle e^{+i\omega_m t} + \langle\hat{a}_m\hat{a}_m^\dagger\rangle e^{-i\omega_m t}\right). \tag{4.33}$$

Next, taking the thermal average (4.32) of the operators

$$\langle\hat{a}_m^\dagger\hat{a}_m\rangle = \bar{n}(\omega_m) = \frac{1}{e^{\beta\hbar\omega_m}-1} = \frac{1}{2}\left[\coth\left(\frac{\hbar\omega_m}{2k_BT}\right) - 1\right] \tag{4.34}$$

$$\langle\hat{a}_m\hat{a}_m^\dagger\rangle = \bar{n}(\omega_m) + 1 = \frac{1}{2}\left[\coth\left(\frac{\hbar\omega_m}{2k_BT}\right) + 1\right] \tag{4.35}$$

(simply the Planck distribution for the mode in question), we compute the net charge

correlation function as a sum over all the independent oscillators,

$$\begin{aligned}\langle\hat{Q}(t)\hat{Q}(0)\rangle &= \sum_m \langle\hat{Q}_m(t)\hat{Q}_m(0)\rangle \\ &= \sum_m \frac{\hbar}{2z_m}\int_{-\infty}^{\infty} d\omega\, \left[\bar{n}(\omega_m)\delta(\omega+\omega_m) + (\bar{n}(\omega_m)+1)\delta(\omega-\omega_m)\right] e^{-i\omega t} \\ &= \sum_m \int_{-\infty}^{\infty} d\omega\, \frac{\hbar}{\pi\omega}(\bar{n}(\omega)+1)\mathrm{Re}[Y_m(\omega)]e^{-i\omega t} \\ &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\hbar\frac{\mathrm{Re}[Y(\omega)]}{\omega}\left[\coth\left(\frac{\hbar\omega}{2k_BT}\right)+1\right]e^{-i\omega t}, \end{aligned} \tag{4.36}$$

where we used $q_{m,zp} = \sqrt{\hbar/2z_m}$ with $z_m = y_m^{-1}$, and in the second to last line we used the identity $-\bar{n}(-\omega) = \bar{n}(\omega)+1$. Since the double-sided spectral density is the Fourier transform of the time correlation function,[5] from this integral we can immediately read off the spectral density of charge fluctuations,

$$S_{QQ}(\omega) = \hbar\frac{\mathrm{Re}[Y(\omega)]}{\omega}\left[\coth\left(\frac{\hbar\omega}{2k_BT}\right)+1\right]. \tag{4.37}$$

Since charge and current are related by a time derivative, we can also read off the equivalent spectral density of current fluctuations,

$$S_{II} = \hbar\omega\mathrm{Re}[Y(\omega)]\left[\coth\left(\frac{\hbar\omega}{2k_BT}\right)+1\right]. \tag{4.38}$$

This result is an example of the quantum fluctuation-dissipation theorem. It is a circuit version of the more general theorem for any system with a linear response, made famous by Kubo *et al.* [140]. It is called this because it is relating the noise $S(\omega)$ to

[5]Explicitly, $S_{xx}(\omega) = \int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\langle x(\tau)x(0)\rangle$, known as the Wiener-Khinchin theorem, a fact which we will derive later in section 4.4. For now we simply accept this statement to be true.

the dissipative part of the system's response, $Y(\omega)$. More generally, the fluctuation-dissipation theorem states that the equilibrium fluctuations of a system are related to the dissipative force driving it back to equilibrium in response to a non-equilibrium perturbation.

We derived (4.37) and (4.38) carefully to be valid for both positive and negative ω. We observe that $S(\omega)$ does not need to be the same at positive and negative frequencies, but at this point it is not clear what this means exactly. We can see that the origin of this asymmetry comes from the fact that we are taking the correlation of operators, which could fail to commute unlike regular numbers. In the next section, we will see using Fermi's golden rule that the noise at positive and negative frequencies correspond to the environment absorbing and emitting photons respectively. But before we delve into that calculation, we can get an idea of why this might be by looking at the various limits of (4.38),

$$S_{II}(\omega) = \begin{cases} 0, & \hbar\omega \ll -k_B T \\ 2k_B T \mathrm{Re}[Y(\omega)], & |\hbar\omega| \ll k_B T \\ 2\hbar\omega \mathrm{Re}[Y(\omega)], & \hbar\omega \gg k_B T \end{cases} \tag{4.39}$$

In the high temperature limit, we obtain the usual Johnson noise spectral density for the current noise of a resistor [there is a factor of 2 difference from the usual expression $4k_B T/R$ because we have defined a double-sided spectral density, whereas with classical circuits one usually just consders a single-sided spectrum, equivalent to $S^+(\omega) \equiv S(\omega) + S(-\omega)$]. In this classical regime, S_{II} is the noise that would be felt by a perfect ammeter hooked up across the resistor. Conversely, in the low-temperature limit, there is noise at positive frequencies but not negative frequencies. The noise at positive frequencies

is proportional to the zero-point fluctuation energy $\frac{1}{2}\hbar\omega$. We can then guess that the $\omega < 0$ part has to do with photon emission by the admittance while $\omega > 0$ corresponds to absorption of energy by the admittance. We will see shortly that this intuition is correct even when the environment is not in thermal equilibrium.

4.2.1.4 Dissipation as quantum noise: Fermi's golden rule

In the previous section we derived the spectral density of current or charge fluctuations induced by $Y(\omega)$. In particular, we kept the current and voltage as operators, leading to a spectral density that can be different at positive and negative frequencies. In this section, we will show why we performed such a calculation and assign physical meanings to the noise spectrum at positive and negative frequencies, namely the generation of relaxation and excitation of the qubit. We will tie together the two languages by interpreting the fluctuation-dissipation theorem simply as a statement about the system obeying detailed balance.

To do this, we will perform an analysis using perturbation theory similar to Refs. [141, 142] to derive Fermi's golden rule in a form that is useful for us. We start with an uncoupled qubit Hamiltonian $H_Q = -\frac{1}{2}\hbar\omega_{10}\sigma_z$ and environment Hamiltonian H_E, which could have a very large number of degrees of freedom, such as a Caldeira-Leggett environment. We suppose quite generally that the environment is coupled to the qubit via a σ_x term (where σ_z is the energy eigenbasis) according to the interaction Hamiltonian

$$H_{\text{int}} = \hbar g \hat{F}(t) \otimes \sigma_x. \tag{4.40}$$

This linear form for the coupling is justified in the weak-coupling limit. In particular the system should only weakly disturb the environment and vice-versa. Choosing σ_x as opposed to σ_y makes no real difference, and usually we will have one or the other and we can perform a change of basis to make it equivalent to σ_x. We do not consider coupling σ_z since it commutes with H_Q and does not drive transitions.

For a classical noise source, F will be just a scalar number, but in general for a quantum environment it will be an operator acting on the environment E. By moving to a rotating frame defined by the time-evolution operator of the uncoupled system,

$$U(t) = U_E(t) \otimes U_Q(t) = \exp\left[-\frac{i}{\hbar} H_E t\right] \otimes \exp\left[-\frac{i}{\hbar} H_Q t\right], \tag{4.41}$$

H_Q and H_E vanish and the interaction term becomes

$$\widetilde{H}_{\text{int}} = U^\dagger(t) H_{\text{int}} U(t) = \hbar g \hat{F}(t) \otimes \left(e^{i\omega_{10} t}\sigma_+ + e^{-i\omega_{10} t}\sigma_-\right). \tag{4.42}$$

First we consider what happens if the system starts in a pure product state, with the environment in some specific state $|m_E\rangle$ and the qubit in the ground state,

$$|\Psi(0)\rangle = |m_E\rangle \otimes |g\rangle. \tag{4.43}$$

We treat the interaction Hamiltonian as a perturbation, and write down the Schrödinger evolution for the whole system to first order in this perturbation,

$$|\Psi(t)\rangle = |\Psi(0)\rangle - \frac{i}{\hbar}\int_0^t dt'\, \widetilde{H}_{\text{int}}(t')\,|\Psi(0)\rangle. \tag{4.44}$$

If the environment starts in a specific state, the probability of populating the qubit excited state $|e\rangle$ as a function of time will then be

$$\sum_n |\langle n_E|\langle e|\Psi(t)\rangle|^2 = \sum_n g^2 \left|\int_0^t dt'\, \langle n_E|\hat{F}(t')\,|m_E\rangle\, e^{i\omega_{10} t'}\right|^2. \tag{4.45}$$

In reality, the initial state of the environment will be described by a density matrix over some ensemble (which could be thermal but doesn't have to be),

$$\rho_E = \sum_m \rho_{mm} \left|m\right\rangle \left\langle m\right|, \tag{4.46}$$

so the probability for the qubit to become excited over the ensemble is given by summing the above expression over all possible initial states of the environment,

$$\begin{aligned}
p_{|1\rangle}(t) &= g^2 \sum_m \rho_{mm} \sum_n \left| \int_0^t dt' \left\langle n_E|\hat{F}(t')\left|m_E\right\rangle e^{i\omega_{10}t'}\right.\right|^2 \\
&= g^2 \sum_m \sum_n \rho_{mm} \int_0^t \int_0^t dt'dt'' \left\langle n_E|\hat{F}(t')\left|m_E\right\rangle \left\langle m_E|\hat{F}(t'')\left|n_E\right\rangle e^{i\omega_{10}(t'-t'')}\right. \\
&= g^2 \sum_m \rho_{mm} \int_0^t \int_0^t dt'dt'' \left\langle m_E|\hat{F}(t'')\hat{F}(t')\left|m_E\right\rangle e^{i\omega_{10}(t'-t'')}\right. \\
&= g^2 \int_0^t \int_0^t dt'dt'' \left\langle \hat{F}(t'')\hat{F}(t')\right\rangle e^{i\omega_{10}(t'-t'')} \\
&= g^2 \int_0^t \int_{t'}^{t+t'} dt'd\tau \left\langle \hat{F}(t'+\tau)\hat{F}(t')\right\rangle e^{-i\omega_{10}\tau}
\end{aligned} \tag{4.47}$$

where in the last line we changed variables to $\tau = t'' - t'$. Assuming the autocorrelation time is much shorter than the timescale of the experiment (i.e., a "memoryless" environment, also known as a Markovian approximation, in which the spectral density doesn't vary strongly with frequency), we can extend the limits of integration for τ to positive and negative infinity without materially changing the answer, leading to

$$p_{|1\rangle}(t) = \left(\frac{g}{\hbar}\right)^2 \int_0^t S_{FF}(-\omega_{10}), \tag{4.48}$$

where

$$S_{FF}(\omega) \equiv \int_{-\infty}^{\infty} dt\, \langle \hat{F}(t)\hat{F}(0)\rangle e^{i\omega t} \tag{4.49}$$

is the exact same double-sided spectral density we considered when we analyzed the current and charge fluctuations Caldeira-Leggett model. Therefore, in the short-time limit, we arrive at the following transition rate out of the qubit ground state,

$$\Gamma_{\uparrow} = g^2 S_{FF}(-\omega_{10}). \tag{4.50}$$

Performing the same analysis but with the qubit starting in the excited state similarly yields a downwards transition rate,

$$\Gamma_{\downarrow} = g^2 S_{FF}(+\omega_{10}). \tag{4.51}$$

This shows the physical meaning behind the spectral density at positive and negative frequencies: at positive/negative frequencies it corresponds to the environment absorbing/emitting energy from/to the qubit. In the case of a classical noise source, where $S(\omega) = S(-\omega)$, one can think of the noise simply as an incoherent drive of the qubit, i.e., an incoherent sum of Rabi driving terms, leading over an ensemble average to a diffusive walk away from the initial state towards the center of the Bloch sphere [123].

These formulas are simply a statement of Fermi's golden rule (FGR). In fact, another way to write them is

$$\Gamma_{\downarrow,\uparrow} = \frac{2\pi}{\hbar} \sum_{nm} \rho_{nn}^{\mathrm{env}} |\langle n, 1|\hat{H}_{\mathrm{int}}|m, 0\rangle|^2 \delta(\mathcal{E}_n - \mathcal{E}_m \pm E_{10}), \tag{4.52}$$

which looks the same as how Fermi's golden rule is usually presented in terms of an effective density of states. In this interpretation, $\mathrm{Re}[Y(\omega)]$ determines the effective density

of Caldeira-Leggett oscillator modes.

If we apply these FGR equations to the special case of an environment in thermal equilibrium through Eq. (4.38), we can gain some intuition behind the fluctuation-dissipation theorem. Intuitively, at low temperatures the environmental oscillators on resonance with the qubit are in their ground state, and so can only absorb energy, whereas at high temperatures they will both emit and absorb energy. The up and down transition rates are related through

$$\frac{\Gamma_{\uparrow}^{\text{thermal}}}{\Gamma_{\downarrow}^{\text{thermal}}} = \exp(-\hbar\omega_{10}/k_B T), \tag{4.53}$$

which is precisely the condition of detailed balance required for thermal equilibrium to hold.

4.2.1.5 Qubit decay in the presence of a resistor

We now have the tools to compute the effect of a resistor on a qubit. First, we can look at the effect of a resistor in the quantum version of an LCR resonator (Fig. 4.9(b)) as a consistency check, by applying the Caldeira-Leggett model. Here, the resistor is not shorted, but has its leads placed across the oscillator. Here, we treat the LC part of the oscillator is the "qubit' and R is the environment. The interaction part of the system Hamiltonian (4.28) with $\hat{\Phi}$ being the qubit degree of freedom is $\left(\sum_m \frac{\hat{\Phi}_m}{L_m}\right)\hat{\Phi} = \left(\sum_m \hat{I}_m\right)\hat{\Phi}$ and so when tracing over the environmental oscillators' thermal state we simply get an effective current noise in the qubit, with the same value as in Eq. (4.38) (assuming the qubit doesn't perturb the environment too much). The current noise from the resistor couples

to the qubit's σ_x degree of freedom with a coupling coefficient $\langle 0|(d\hat{H}/dI)|1\rangle = \langle 0|\hat{\Phi}|1\rangle$, which for a harmonic LC oscillator is given by (2.43), $\sqrt{\hbar\sqrt{L/C}/2}$. Using Eqs. (4.51) and (4.38) in the low-temperature (quantum) limit, the decay rate of a linear quantum LC oscillator[6] is then $\Gamma_\downarrow = 1/RC$, which matches the expected decay time constant from classical circuits.

However, in general, the "qubit" part of the circuit will of course also have a Josephson junction in parallel, leading to the circuit illustrated in Fig. 4.9(c). This won't change the noise source or the form of the interaction Hamiltonian in the previous paragraph, but will change how the interaction Hamiltonian is able to drive transitions by changing the matrix element $g = |\langle 1|\hat{\Phi}|0\rangle|$ in Eq. 4.51. Namely, the flux matrix element will not be the harmonic value and the decay rate of the fluxmon (or any other type of superconducting qubit) will be given by

$$\Gamma_\downarrow = \frac{1}{\hbar^2}|\langle 0|\hat{\Phi}|1\rangle|^2 S_{II}(\omega_{10}) \tag{4.54}$$

$$= \frac{2\omega_{10}}{\hbar}|\langle 0|\hat{\Phi}|1\rangle|^2 \mathrm{Re}[Y(\omega_{10})] \tag{4.55}$$

$$\approx \frac{\mathrm{Re}[Y(\omega_{10})]}{C}, \tag{4.56}$$

where the last line holds only when the matrix element is close to its harmonic limit value. We note that this result, in which the quantum transition rate is proportional to $\mathrm{Re}[Y(\omega)]$, was derived rigorously by Esteve *et al.* in Ref. [90] through a related method.

[6] Technically, we are computing the $|1\rangle \to |0\rangle$ Fock state decay rate, but this turns out to be the same as the decay rate of the average photon number of a coherent state under reasonable assumptions [143].

In upcoming sections we will use this equation to predict the contribution from known sources of dissipation, such as the flux bias lines, to the fluxmon's T_1.

4.2.1.6 Flux vs. charge noise: a quantum Thévenin-Norton equivalence

The observant reader may have made the following observation: if dissipation can be equivalently viewed as a series (Thévenin) or parallel (Norton) resistance [at least for the lumped-element picture of Fig. 4.9(a)], then since the qubit capacitor is linear we could equally view dissipation as a "Thévenin" series resistor across the qubit capacitor with the small series resistance $R_s = (1/\omega C)^2/R_p$, and the environmental oscillators will couple to the qubit's charge degree of fredom, producing an effective voltage noise

$$\begin{aligned}\Gamma_{\downarrow} &= \frac{1}{\hbar^2}|\langle 0|\hat{Q}|1\rangle|^2 S_{VV}(\omega_{10}) \\ &= \frac{2\omega_{10}}{\hbar}|\langle 1|\hat{Q}|0\rangle|^2 \mathrm{Re}[Z(\omega_{10})] \end{aligned} \tag{4.57}$$

with $R = R_s$. Since the two representations of dissipation are the same, this must be equal to the decay rate (4.55) with $R = R_p$. This is quite interesting, because when we equate these two expressions it seems to imply there must be a simple relationship between the flux and charge matrix elements of the fluxmon, namely that $|\langle 0|\hat{\Phi}|1\rangle|^2 = 1/(\omega_{10}C)^2|\langle 0|Q|1\rangle|^2$. We can in fact derive this independently from the qubit Hamiltonian with a quantum-mechanical calculation. The fluxmon qubit, or more generally any type of superconducting qubit, has a qubit Hamiltonian

$$H = \frac{Q^2}{2C} + V(\Phi), \tag{4.58}$$

where V is a nonlinear potential formed by some combination of biased inductors and Josephson junctions. Using the identity

$$[H, \Phi] = \frac{1}{2C}[Q^2, \Phi] = -i\hbar\frac{Q}{C}, \tag{4.59}$$

we have the relation

$$\begin{aligned}\langle 0|[H, \Phi]|1\rangle &= \langle 0|(H\Phi - \Phi H)|1\rangle \\ &= -\hbar\omega_{10}\langle 0|\Phi|1\rangle \\ &= i\hbar\langle 0|Q|1\rangle/C,\end{aligned} \tag{4.60}$$

from which we deduce

$$|\langle 0|\hat{\Phi}|1\rangle|^2 = 1/(\omega_{10}C)^2|\langle 0|Q|1\rangle|^2 \tag{4.61}$$

as sought. Similarly to how the Thévenin-Norton equivalence relies on the linearity of the capacitor in the classical circuit argument, we needed to have a quadratic contribution of the capacitance to the Hamiltonian ($\propto Q^2$) in order for the commutation relations to work out. The result is that the effect of flux (or current) noise is indistinguishable from the effect of charge (or voltage) noise that has a power spectrum differing by two powers in ω. We should keep this in mind when modeling dissipation vs. frequency for a given qubit. If one does not have any other physical information about the source of the dissipation, it is equivalent to model the dissipation as either flux noise or charge noise with an appropriate noise spectrum.

4.2.2 Known sources of dissipation and high-frequency noise

In this section we discuss known sources of dissipation and high-frequency noise in the fluxmon qubit. These sources include ohmic dissipation from the bias line, Purcell decay through the readout resonator, and dielectric loss. We also discuss stray transitions driven by classical noise from room temperature. Afterwards, we will show measurements of dissipation in the fluxmon qubit, which we find to be dominated by a previously unmeasured dissipation source that we attribute to flux noise.

4.2.2.1 Damping from flux bias lines

The tilt flux bias line is a high-bandwidth line that presents an impedance of $50\,\Omega$ behind the mutual inductance M_t of the bias line for all qubit frequencies up to $\sim 10\,\mathrm{GHz}$ (depending on the exact cutoff frequency of the filtering in the line). This is illustrated by the circuit in Fig. 4.12(a). Using the usual expression for the impedance seen across the qubit's inductor leads in the presence of mutual inductance [equation (3.24) for the general circuit of Fig. 3.6(b)] in the limit of small bias line inductance L_b yields an effective resistance $R_s = (\omega M_t)^2/R$ in series with the qubit inductor. Using Eq. (4.18) with $Z = i\omega L_q$, we can convert this to an equivalent parallel resistance as in Fig. 4.12 with

$$R_p = \left(\frac{L_q}{M_t}\right)^2 R. \tag{4.62}$$

Therefore, the effect of the bias line is to present a transformed resistance R_p much larger than $50\,\Omega$ in parallel with the qubit. We can then apply Eq. (4.55) with $\mathrm{Re}[Y(\omega)] = 1/R_p$

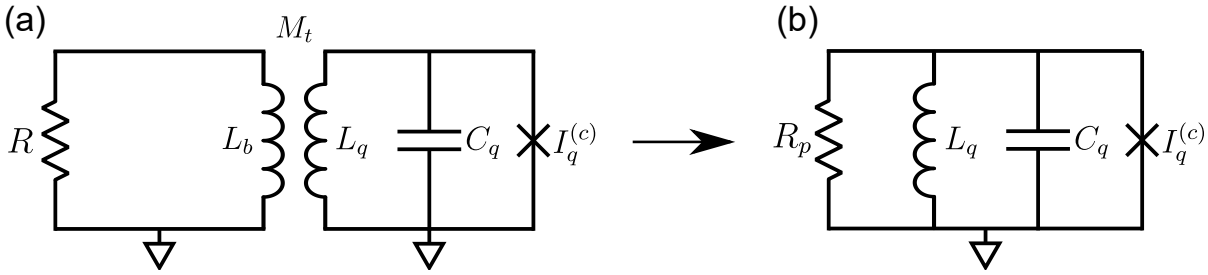


Figure 4.12: Illustration of qubit damping through a mutual inductance to a resistive environment. (a) Actual dissipative bias line circuit. (b) Equivalent parallel resistance to the qubit.

to obtain the T_1 limit imposed by the bias line.

In Fig. 4.13, we plot the resulting T_1 limit vs. qubit frequency at zero tilt for two different values of M_t. To get a sense of scale, we note that the transformed parallel resistance R_p is typically a couple hundred MΩ. If the fluxmon were harmonic at every frequency, we would expect the T_1 limit to be independent of frequency. However, at low frequencies, the flux matrix element deviates from the harmonic value [as was shown earlier in Fig. 2.8(c)], leading at low frequencies to a larger T_1 limit than would be predicted in the harmonic limit. For typical fluxmon design parameters this deviation starts to become significant below $\sim 2\,\mathrm{GHz}$. Also note that away from zero tilt (not shown), the flux matrix element strictly decreases, so the T_1 limit will strictly improve away from zero tilt. In order to avoid excess dissipation beyond the intrinsic fluxmon dissipation (which is measured in the next section), we need to restrict $M_t \lesssim 0.5\,\mathrm{pH}$, assuming a $50\,\Omega$ environment.

We tested the prediction of eqs. (4.62) and (4.55) by designing one fluxmon qubit that was very strongly coupled to its flux bias line. For a qubit with a measured $M_t = 2.6\,\mathrm{pH}$, we observed a T_1 of 210 ns at zero tilt at $f_{10} = 5.9\,\mathrm{GHz}$. Theoretically, we would expect

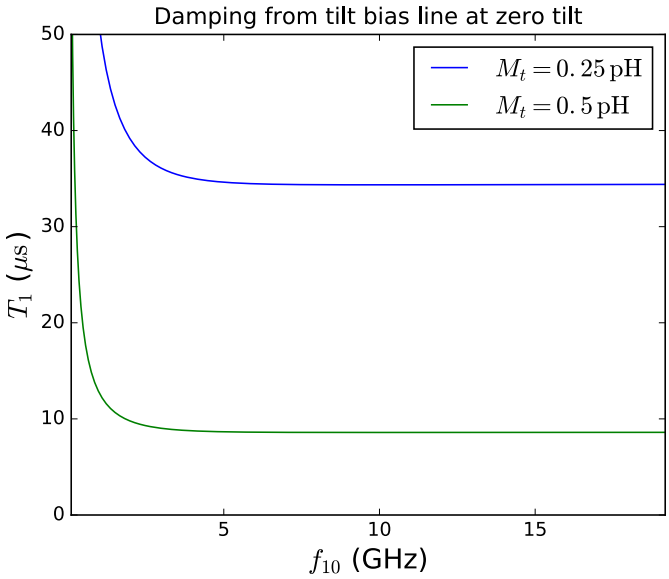


Figure 4.13: Predicted fluxmon T_1 limit from tilt bias line damping, for two different values of M_t. Note that the damping rate depends quadratically on M_t. The right hand side of the plot corresponds to the harmonic limit $\beta = 0$, where the T_1 limit is independent of f_{10}.

a T_1 limit of 275 ns from the flux bias line with the measured M_t and a $50\,\Omega$ microwave environment. However the expected intrinsic T_1 (from other sources of dissipation) is around $1.5\,\mu$s as measured on other devices with very weak coupling to the bias line, and when this expected contribution is subtracted out we are left with a contribution of 230 ns from the bias line, within $\sim 10\%$ of what was actually measured.

We can also consider the SQUID bias line with mutual inductance M_{SQ} as a source of dissipation. Typically we have attenuators between the qubit SQUID and the 100 MHz low-pass filter on the SQUID line, meaning the qubit will see an approximately $50\,\Omega$ environment behind M_{SQ} at microwave frequencies. However, at zero tilt, the SQUID bias tunes σ_z in the energy basis (which is the parity basis), so it will not drive transitions, similar to the case for a transmon qubit with a SQUID bias line.[7] But away from zero

[7]This is not strictly true if there is junction asymmetry, where noise/dissipation in the SQUID bias

tilt, the energy basis starts to rotate towards the flux basis, and the SQUID line can start to drive transitions. We can compute the expected decay rate through the SQUID line as a function of flux bias using Fermi's golden rule,

$$\begin{aligned}\Gamma_{\downarrow}^{\mathrm{SQ}} &= \frac{1}{\hbar^2}\left|\left\langle 0\left|\frac{\partial\hat{H}}{\partial\Phi_{\mathrm{SQ}}}\right|1\right\rangle\right|^2 S_{\Phi_{\mathrm{SQ}}}(\omega_{10}) \\ &= \frac{1}{\hbar^2}\left(\frac{\partial\beta(\Phi_{\mathrm{SQ}})}{\partial\Phi_{\mathrm{SQ}}}\right)^2 E_L^2|\langle 0|\cos\hat{\varphi}|1\rangle|^2 M_{\mathrm{SQ}}^2\frac{2\hbar\omega_{10}}{R},\end{aligned} \tag{4.63}$$

where $R \approx 50\,\Omega$. Note that for the SQUID bias line, the relevant matrix element is not φ_{10} but $(\cos\varphi)_{10}$, which by symmetry vanishes at zero tilt, consistent with our qualitative discussion above. In Fig. 4.14 we plot the T_1 limit from the SQUID bias line vs. both SQUID and tilt flux biases. Since to first order the T_1 limit diverges to infinity at zero tilt, we have artificially truncated the color scale at $T_1 = 100\,\mu$s.

4.2.2.2 Damping from readout circuit (Purcell effect)

As mentioned in section 3.2.2, the qubit can be damped through its coupling to the readout resonator, which mediates a second-order coupling between the qubit and a $50\,\Omega$ transmission line environment (in addition, the resonator could have an additional intrinsic dissipation source as well). This decay mechanism has been well-studied in superconducting qubits coupled to resonators that are in turn coupled to transmission lines [72, 144, 83] (the name historically comes from a study by Purcell of how the

can partially transfer to effective noise/dissipation in the tilt bias via the differential transfer function 4.109

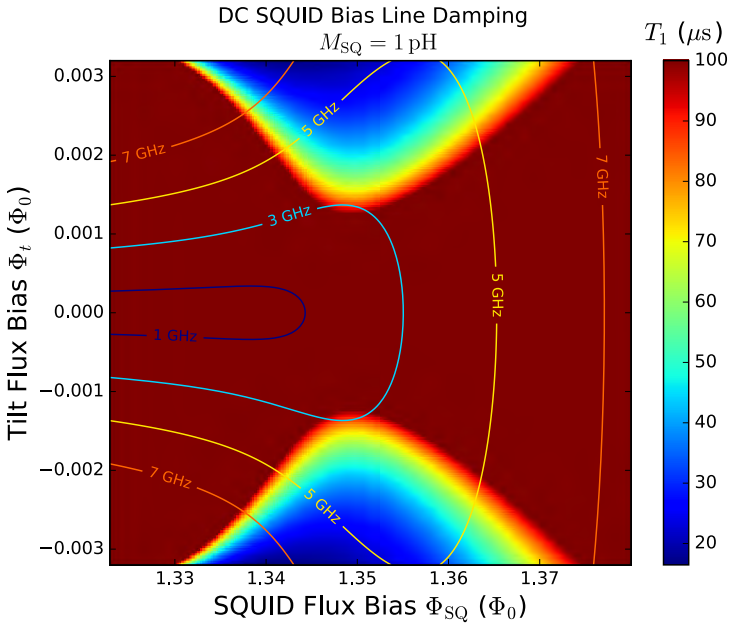


Figure 4.14: T_1 limit induced by DC SQUID bias line for $M_{\mathrm{SQ}} = 1\,\mathrm{pH}$. Contours of constant frequency $f_{10} = 1, 3, 5$, and $7\,\mathrm{GHz}$ are drawn for reference. Because the T_1 limit diverges at zero tilt, we truncate the color scale at $T_1 = 100\,\mu\mathrm{s}$ for the sake of illustration.

spontaneous decay rate of an atom is altered in a cavity vs. vacuum [82]). For transverse coupling to a single-mode resonator with coupling strength

$$g_{qr} = \frac{M_{qr}}{L_q L_r} \Phi_{zp,r} \langle 0 | \hat{\Phi}_q | 1 \rangle, \tag{4.64}$$

in the dispersive regime $|\Delta| \equiv |\omega_r - \omega_q| \gg g$, the Purcell decay rate is given by Eq. (3.38), where $\kappa = \omega_r / Q_r$ is the total decay rate of the resonator. However, this expression will break down when the qubit is very close to the resonator $\Delta \sim g$. We can slightly generalize it to allow a weak violation of the dispersive assumption as follows. By looking at the exact Jaynes-Cummings hybridized eigenstates between the qubit and resonator mode [72], one can compute what fraction of the qubit-like excited state population "lives

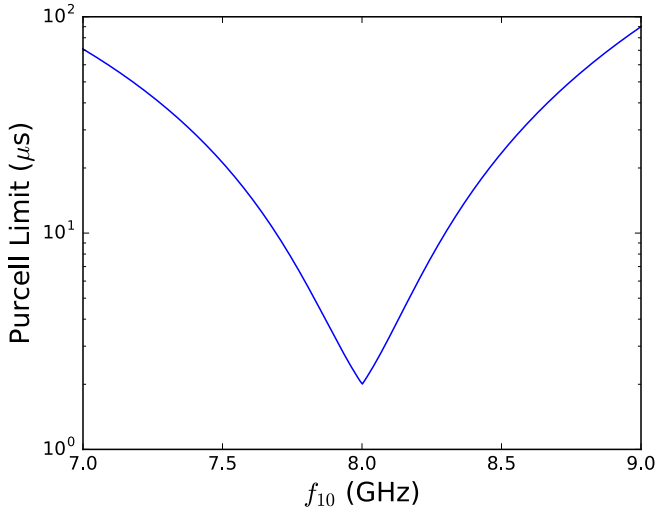


Figure 4.15: Predicted Purcell decay rate (4.65) vs. qubit frequency at zero tilt for standard readout design parameters ($\omega_r/(2\pi) = 8\,\text{GHz}$, $g_{qr}/(2\pi) \approx 100\,\text{MHz}$ (this is a function of frequency), $1/\kappa_r = 1\,\mu\text{s}$). Exactly on resonance, half of the qubit population "lives in" the resonator, leading to an effective decay rate of $\kappa_r/2$, but this is not a very useful quantity since very close to resonance, the qubit population will undergo semi-coherent swapping with the resonator rather than a pure exponential decay. In other words a photon that is emitted from the qubit can be re-absorbed, so we would not be in the regime where Fermi's golden rule applies.

in" the resonator, leading to the more accurate expression

$$\frac{1}{T_1^{\text{Purcell}}} = \sin^2\left[\frac{1}{2}\arctan\left(\frac{2g}{\Delta}\right)\right]\kappa_r, \tag{4.65}$$

which reduces to (3.38) in the dispersive limit. Alternatively, since the resonator and transmission line environment are linear, we can just use linear circuit theory on their combined admittance as seen from the qubit in order to compute T_1 through Eq. (4.55)/(4.56). This approach has the advantage that it can easily capture the higher modes of the CPW readout resonator [83], and is also more useful for understanding and designing more complicated readout circuit designs relevant for transmon qubits, such as "Purcell filters" [84, 85].

In Fig. 4.15 we plot the predicted Purcell decay rate vs. qubit frequency at zero tilt for standard readout design parameters ($\omega_r/(2\pi) = 8\,\mathrm{GHz}$, $g_{qr}/(2\pi) \approx 100\,\mathrm{MHz}$ [this is a function of frequency], $1/\kappa_r = 1\,\mu\mathrm{s}$). We conclude that Purcell decay is negligible unless the qubit is biased very close (a few hundred MHz) to the resonator. In practice, this will not happen, except for when the qubit bias is swept through the resonator frequency, in which case the dominant source of population loss will usually be Landau-Zener swapping [Eq. (3.1)] rather than Purcell decay.

4.2.2.3 Distributed CPW loss in the fluxmon qubit

It is reasonable to expect that the CPW portion of the fluxmon qubit could contribute intrinsic dissipation that is distributed evenly throughout its length. For example, it is well-known that superconducting CPW resonators suffer from dielectric loss due to charged tunneling defects distributed within thin lossy interfacial dielectric layers [61, 145, 59]. In the case of dielectric loss, the typical distribution of tunneling defects over frequency are such that one observes an approximately constant quality factor $Q = \omega T_1$. This constant-Q behavior is also roughly observed in smaller lumped element qubits like the phase qubit and transmon [61, 40]. Typically, we look at the performance of distributed CPW resonators to predict loss in qubits fabricated with similar materials and geometries. For example, in the Xmon transmon qubit [40], whose capacitor is essentially a very short (lumped-element) segment of CPW transmission line (and whose matrix elements are very nearly those of a harmonic oscillator), we expect that the T_1 of the qubit should be limited to at most the same Q as CPW resonators fabricated with

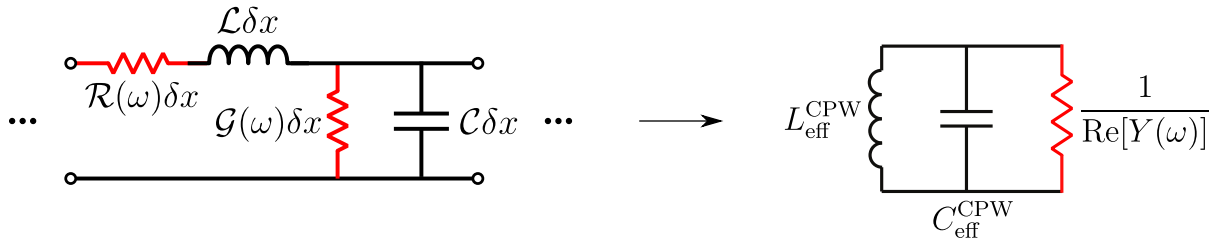


Figure 4.16: Infinitesimal ladder model for transmission line with two types of distributed dissipation, and lumped element approximation for a finite segment of transmission line.

the same geometry. However, in the fluxmon there is also the possibility of inductive in addition to capacitive losses in the CPW.

To understand the effect of CPW dissipation in general on the fluxmon beyond just dielectric loss, we need to find the effective $\text{Re}[Y(\omega_{10})]$ presented by the CPW segment. To do this, we need to include distributed dissipation in our circuit analysis. We will compare this to the effect of CPW dissipation in $\lambda/4$ resonators and Xmon qubits. One of the conclusions will turn out to be that at a given frequency the fluxmon has the same sensitivity to dielectric loss as its $\lambda/4$ resonator counterpart, but a much higher sensitivity to dissipation from a series resistance such as magnetic loss, due to an effective impedance transform of the distributed dissipation.

To model distributed linear dissipation, we must revisit our ladder model for the fluxmon's transmission line segment in Fig. 2.2(c). Linear dissipation in a uniform transmission line can be modeled by adding a series resistance per unit length $\mathcal{R}$ or parallel admittance $\mathcal{G}$ to each ladder rung,[8] as shown in Fig. 4.16. $\mathcal{R}$ is a resistance per unit length that could represent conductor loss from quasiparticles or a lossy permeability

[8]This circuit is not to be confused with our earlier model of a resistor as a transmission line in Fig. 4.10. There, we had a physical resistor and an abstract transmission line. Here, we now have a physical transmission line with abstract resistors.

(magnetic dissipation) associated with a.c. current flowing in the inductors. $\mathcal{G}$ is a conductance per unit length that could represent stray electrical conduction between center trace and ground plane, or a lossy a.c. permittivity (dielectric loss) from dipole relaxation associated with a.c. voltage across the capacitor dielectric. In general, both $\mathcal{R}$ and $\mathcal{G}$ can be frequency-dependent. For example, dielectric loss within the standard TLS model corresponds to $\mathcal{G}(\omega) = \omega\mathcal{C}\tan\delta \propto \omega$ (constant Q). Also note that the distributed dissipation will also in general create a small amount of dispersion in the transmission line, which we can ignore given that the fluxmon is so short.

We need to find the input impedance of the CPW as seen from the fluxmon's Josephson junction. We can generalize the input impedance formula (2.28) by using the generalized telegraph equations with dissipation [41], leading to a complex propagation constant

$$\gamma = \sqrt{(\mathcal{R}(\omega) + i\omega\mathcal{L})(\mathcal{G}(\omega) + i\omega\mathcal{C})} \tag{4.66}$$

and a generalized complex characteristic impedance

$$\widetilde{Z}_0 = \sqrt{\frac{\mathcal{R}(\omega) + i\omega\mathcal{L}}{\mathcal{G}(\omega) + i\omega\mathcal{C}}}. \tag{4.67}$$

Using the fact that the fluxmon CPW is terminated with a short at $x = \ell$, the input admittance is then

$$\begin{aligned} Y_{\text{in}}^{\text{CPW}}(\omega) &= \frac{1}{\widetilde{Z}_0}\frac{e^{\gamma\ell} + e^{-\gamma\ell}}{e^{\gamma\ell} - e^{-\gamma\ell}} \\ &\approx i\omega\frac{\mathcal{C}\ell}{3} + \frac{1}{i\omega\mathcal{L}\ell} + \frac{1}{3}\mathcal{G}(\omega)\ell + \frac{\mathcal{R}(\omega)\ell}{\ell^2\mathcal{L}^2\omega^2}, \end{aligned} \tag{4.68}$$

where in the second line we have taken the limits of a short transmission line and small

dissipation,

$$\frac{\mathcal{G}}{\mathcal{C}}, \frac{\mathcal{R}}{\mathcal{L}} \ll \omega \ll \frac{1}{\ell\sqrt{\mathcal{L}\mathcal{C}}}. \tag{4.69}$$

The first two terms of (4.68) are imaginary and correspond precisely to the effective lumped-element fluxmon capacitance and inductance that we found in Eq. (2.30), while the second two terms are real and will cause dissipation.

At this point we make several observations. Because the CPW capacitance $\mathcal{C}\ell$ and $\mathcal{G}\ell$ are both rescaled by a factor of $\frac{1}{3}$, we expect dissipation from dielectric loss to be the same in the fluxmon as in a $\lambda/4$ resonator. Intuitively, this is because the voltage profile along the CPW affects the electric participation of the lossless and lossy parts of the capacitance in the same way. More explicitly, we can compute the resulting fluxmon T_1 using Eq. (4.56) with $C = \frac{1}{3}\mathcal{C}\ell$, and then do a similar analysis to compute the T_1 that would occur in a $\lambda/4$ CPW resonator of the same frequency, and in an Xmon qubit of the same frequency (the Xmon has an open boundary condition at $x = \ell$. The resonator has only one possible frequency $\omega = \omega_{\lambda/4}$ at a given CPW length, whereas the Xmon and fluxmon can have varying ω_{10} with the same CPW by varying its Josephson inductance). The results are summarized in table 4.2. We can see that the contribution to T_1 from $\mathcal{G}$ (such as dielectric loss) is the same across all types of circuits at a given frequency. However, the contribution to T_1 from $\mathcal{R}$ differs drastically between the three circuits. For instance, the Xmon qubit has no first-order contribution from $\mathcal{R}$, intuitively because the CPW acts as a purely lumped-element capacitor. For the fluxmon qubit, the contribution from $\mathcal{R}$ is drastically re-scaled at low frequencies from the T_1 that a $\lambda/4$ resonator of

	$C_{\text{eff}}^{\text{CPW}}$	$\text{Re}\left[Y^{\text{CPW}}(\omega_{10})\right]$	$\frac{1}{T_1}$	Assumptions
$\lambda/4$ Resonator	$\frac{1}{2}\mathcal{C}\ell$	$\frac{1}{2}\mathcal{G}\ell + \frac{1}{2}\frac{\mathcal{R}\ell}{Z_0^2}$	$\frac{\mathcal{G}}{\mathcal{C}} + \frac{\mathcal{R}}{\mathcal{L}}$	$\omega_{10} = \omega_{\lambda/4} = \frac{\pi}{2}\frac{1}{\ell\sqrt{\mathcal{LC}}}$
Xmon (single-arm)	$\mathcal{C}\ell$	$\mathcal{G}\ell$	$\frac{\mathcal{G}}{\mathcal{C}}$	$\frac{\mathcal{G}}{\mathcal{C}}, \frac{\mathcal{R}}{\mathcal{L}} \ll \omega_{10} \ll \frac{1}{\ell\sqrt{\mathcal{LC}}}$
Fluxmon	$\frac{1}{3}\mathcal{C}\ell$	$\frac{1}{3}\mathcal{G}\ell + \frac{1}{3}\frac{\mathcal{R}\ell}{Z_0^2}\left(\frac{\omega_{LC}}{\omega_{10}}\right)^2$	$\frac{\mathcal{G}}{\mathcal{C}} + \frac{\mathcal{R}}{\mathcal{L}}\left(\frac{\omega_{LC}}{\omega_{10}}\right)^2$	$\frac{\mathcal{G}}{\mathcal{C}}, \frac{\mathcal{R}}{\mathcal{L}} \ll \omega_{10} \ll \frac{1}{\ell\sqrt{\mathcal{LC}}}$

Table 4.2: Summary of the effect of distributed transmission line dissipation on three types of CPW-based quantum circuits: a $\lambda/4$ resonator, an Xmon qubit, and the fluxmon qubit. The dissipation $\mathcal{R}(\omega_{10})$ and $\mathcal{G}(\omega_{10})$ are in general functions of frequency. Note the drastic frequency dependence of the decay rate from $\mathcal{R}$ unique to the fluxmon qubit.

the same frequency would have.[9] One way to understand this frequency dependence is to view it as a series-parallel impedance transformation analogous to a generalized Thévenin-Norton representation of dissipation from Eq. (4.17), where the impedance of the inductors (which dominates over the series resistor impedance) $|Z_L|^2 \propto \omega^2$ determines the effective real part of the parallel circuit admittance via a factor of $1/\omega^2$.

If there is series dissipation $\mathcal{R}$ physically present in the system, the $\left(\frac{\omega_{LC}}{\omega_{10}}\right)^2$ frequency-dependence would explain why the fluxmon's T_1 is observed to be so much lower than the T_1 of $\lambda/4$ resonators and Xmon qubits in the frequency range of $3 - 6\,\text{GHz}$, despite the fact that the CPWs of all these circuits are made with the same materials and cross-sectional geometries (and hence the same $\mathcal{R}$ and $\mathcal{G}$). Since we operate the fluxmon at $5 < \omega_{LC}/\omega_{10} < 20$, the fluxmon's T_1 could be 10 to 100 times shorter than an equivalent $\lambda/4$ resonator. One possible source of distributed series resistance is a non-zero quasiparticle

[9]Note that the $1/\omega_{10}^2$ dependence of $1/T_1$ won't actually hold all the way down to $\omega_{10} = 0$, since this would violate the first inequality in (4.69) that we assumed to get the second line of (4.68). Rather, in the true zero-frequency limit, the contribution from $\mathcal{R}$ will saturate to the intuitive result $1/(\mathcal{R}\ell)$. However, for physically realistic levels of dissipation, this limit is irrelevant.

contribution to the current, which produce a surface impedance $Z_s = R_s + iX_s$ with a dissipative part R_s that limits resonator Q to $\frac{1}{Q} = \frac{\alpha}{\pi}\sqrt{\frac{2\Delta}{\hbar\omega}}x_{\mathrm{qp}}$ [146, 147, 95], where $\alpha \approx 0.05$ is the kinetic inductance fraction, Δ is the superconducting gap and x_{qp} is the fractional density of (presumably non-equilibrium) quasiparticles, or in the language of Fig. 4.16,

$$\mathcal{R}_{\mathrm{qp}}(\omega) \approx \frac{\alpha\mathcal{L}}{\pi}\omega\sqrt{\frac{2\Delta}{\hbar\omega}}x_{\mathrm{qp}}. \tag{4.70}$$

As argued later on in section 4.4, we expect $x_{\mathrm{qp}} \lesssim 1 \times 10^{-7}$, and even for a conservative estimate of $x_{\mathrm{qp}} = 1 \times 10^{-6}$, using table 4.2 this implies a fluxmon T_1 of over $20\,\mu$s at $f_{10} = 5\,$GHz, making quasiparticle current in the CPW an unlikely explanation for the low T_1 (in section 4.4 we also perform an experimental check by introducing magnetic vortices to reduce quasiparticle occupation). Furthermore, we would expect to see a linear scaling with kinetic inductance fraction, which we do not see when varying the overall CPW cross-sectional width [see Fig. 4.22(a)]. We conclude that there must be some other source of $\mathcal{R}$ that we hypothesize to be intrinsic to the geometric inductance of the transmission line, to which the fluxmon at low frequencies will be much more sensitive than resonators.

4.2.2.4 Dielectric loss

Dielectric loss is a source of dissipation in superconducting circuits that has been well studied over the past decade. This type of dissipation is believed to originate from charged dipole moments within any amorphous dielectrics present in the system, which

then couple to the qubit's electric field and absorb its energy. These dipoles likely arise from microscopic configurational defects, which can be modeled as two-level systems (TLSs) with two quasi-stable states separated by a barrier in a double-well potential [148]. In the limit of many defects weakly coupled to the electric field, this leads to a finite loss tangent for the dielectric constant, $\tan\delta = \frac{\mathrm{Im}[\epsilon]}{\mathrm{Re}[\epsilon]}$, that is roughly constant over frequency. This will in turn damp resonators and qubits to a roughly frequency-independent Q [61]. This TLS-induced loss tangent is generally dependent on power and temperature, increasing to a saturating value at low power and temperature (the regime relevant for qubits) as the TLS's become unsaturated,[10] leading to much more loss than would otherwise be expected [61].

When we design and fabricate superconducting circuits, we must keep dielectric loss in mind. The bulk crystalline substrate (usually sapphire or silicon) generally has a negligible loss tangent, and because we know about dielectric loss in amorphous dielectrics, we do not purposefully grow or deposit any dielectric films during the qubit fabrication process (besides the Josephson junctions themselves). We note that this constraint makes the fabrication of superconducting quantum integrated circuits even more challenging, in terms of scaling up to complicated large-scale systems. In fact, D-Wave systems chose to keep standard lossy dielectrics in their fabrication process in order to facilitate quick scalability [16], at the expense of substantial dielectric dissipation. According to a great many experiments over the past decade, these standard amorphous dielectrics have loss

[10] Another way to think of this is that at high powers, the linear response of the TLS becomes power-broadened and therefore has a weaker effect [149].

tangents on the order of 1×10^{-3} or more at qubit operating conditions [150], meaning that D-Wave's qubits likely have a $T_1 \sim 20\,\text{ns}$ at 5 GHz.[11] This is in addition to the numerous strongly coupled defects that are likely to be present due to the strong electric fields permeating any dielectrics near the Josephson junction. It will be interesting to see if and how a quantum annealer without lossy dielectrics, and hence greatly reduced background high-frequency dissipation and substantially fewer strongly coupled defects, will perform differently than the existing D-Wave devices. In particular, we expect that dielectric loss will strongly affect the "thermalization stage" of quantum annealing [19] (see also the section on fast two-qubit annealing in Chapter 6).

It turns out that just because we don't purposefully use amorphous dielectrics, it doesn't mean we are completely insensitive to dielectric loss. Once other sources of dissipation are mitigated (such as magnetic vortices and quasiparticles), superconducting CPW resonators are typically limited to a quality factor of 100,000 - 1,000,000 and show a power dependence in their quality factors consistent with dielectric loss [151, 39]. They also show a temperature-dependent frequency shift in accordance with TLS theory [152, 153]. Based on the scaling of the electric participation ratio with resonator cross-sectional geometry, there is very strong evidence that there is unwanted residual lossy dielectric located at the thin ($\sim 1\,\text{nm}$ thick) interfaces between the device substrate, metal, and vacuum [145]. Furthermore, it has been shown that careful in-situ surface cleaning before metal deposition can improve the quality factors of the resulting

[11] This number is an educated guess, as D-Wave does not report measurements of T_1, since their standard experimental setup does not allow microwave control.

resonators [39]. Assuming a reasonable dielectric permittivity $\epsilon_r \sim 10$ for the interfacial layers, the substrate-vacuum and substrate-metal interfaces participate roughly equally and much more than the metal-vacuum interfaces [145, 59], and a loss tangent of $\sim 1 \times 10^{-3}$ is the right order of magnitude for a $\sim 1\,\mathrm{nm}$ thick interface to explain the observed loss. For any physical capacitance, the effect of dielectric loss is determined by the electric participation of its lossy interfaces,

$$\tan\delta = \sum_i p_i \tan\delta_i, \tag{4.71}$$

where the energetic participation ratio (sometimes called filling factor) of each interfacial layer is [153, 145]

$$p_i = \frac{\int_{V_i} \epsilon_i |\vec{E}(\vec{r})|^2 d\vec{r}}{\int_V \epsilon(\vec{r}) |\vec{E}(\vec{r})|^2 d\vec{r}}. \tag{4.72}$$

The participation of the different interfaces of various capacitor geometries for $\epsilon_i = 10$ is well understood through the COMSOL simulations of Wenner *et al.* [145]. In Fig. 4.17 we plot the result of similar simulations but as a function of ϵ_i. We find that the relative participation of the three types of CPW interfaces depends strongly on the interfacial ϵ_r, consistent with the theory in [145]. This means that resist residue, which can have ϵ significantly smaller than 10,[12] is particularly dangerous at the substrate-metal interface. This could explain why conventional transmon qubits, whose capacitor electrodes were fabricated with electron-beam lithography (EBL) and lift-off at the same time as their

[12]For example, in Ref. [59], we measure the dielectric constant of PMMA to be ~ 2.5 at $\sim 5\,\mathrm{GHz}$ at millikelvin temperatures, by looking at the frequency shifts of CPW resonators spin-coated in a known thickness of resist.

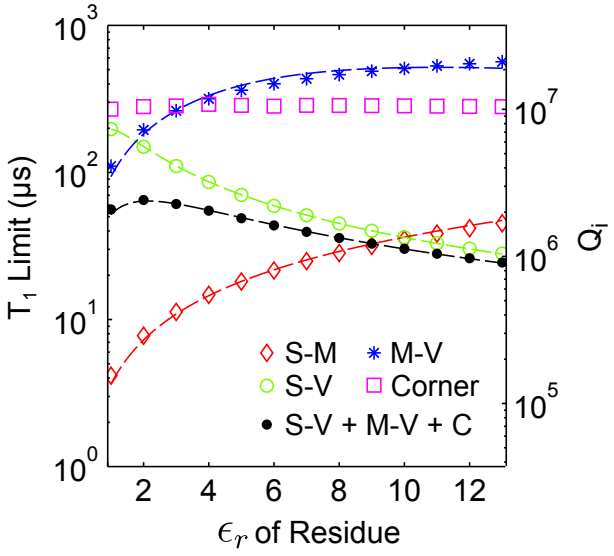


Figure 4.17: Markers: COMSOL simulations of $T_1 = Q/\omega$ limit at 6 GHz due to loss at the different types of CPW interfaces (considering separately "corners" where the interface types intersect [145]) as a function of ϵ_r of the interface, assuming a $w, g = 15, 10\,\mu$m untrenched CPW resonator with 3 nm thick interfaces of $\tan\delta = 2 \times 10^{-3}$, with $\epsilon_r^{\text{substrate}} = 10$. Lines are fits to a simple model [145]. Figure adapted from Ref. [59].

Josephson junctions, had significantly lower T_1 than Xmon qubits, whose capacitors are formed with a subtractive etch process following a clean metal deposition.

To test this hypothesis and look for physical signatures of these interfacial layers, we performed a controlled study comparing the internal quality factors Q_i of CPW resonators fabricated with conventional-transmon-style lift-off versus a pure etch, both on the same chip, as follows (full details of the fabrication process can be found in the supplement of Ref. [59]). Photolithography and a dry etch are used to define $\lambda/2$ CPW resonators coupled to a feedline [Fig. 4.18(a)] in an Al film that was e-beam-deposited on a solvent-cleaned sapphire substrate. During the etch, the entire CPW structure is defined for purely etched control resonators, whereas the full ground plane slot of width $w + 2g$ is etched away for "lift-off resonators"; the center traces of these resonators are defined

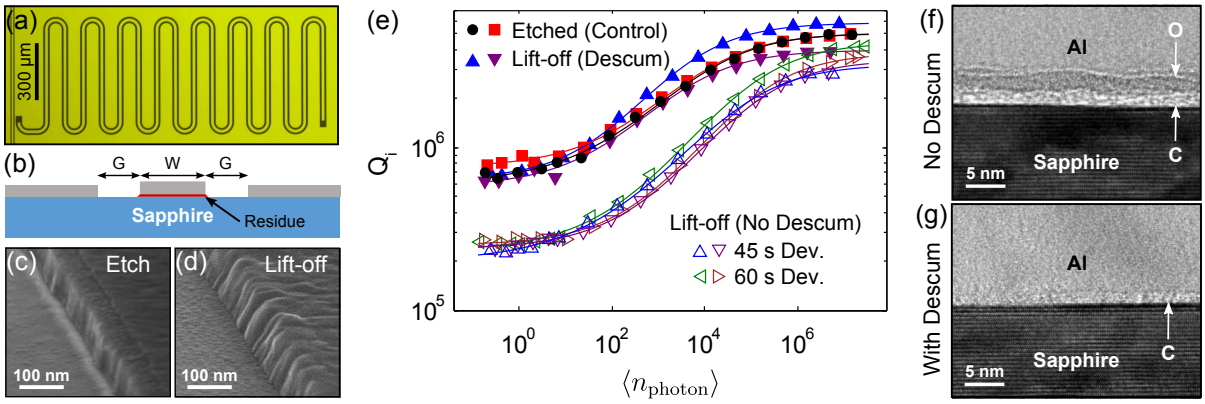


Figure 4.18: Etch versus lift-off, and descum versus no descum: low-power Q_i is degraded using lift-off without descum. (a) Optical micrograph of a "hanging" $\lambda/2$ CPW resonator capacitively coupled to a feedline (left). (b) Schematic of lift-off resonator cross section. (c/d) SEM image of center trace edge of etched/lift-off resonator. (e) Plot of Q_i versus mean photon population for resonators made with lift-off with and without a pre-deposition descum, as well as for etched control resonators. Different marker types are distinct resonators. Solid lines are guides to the eye. (f/g) Edge-on cross-sectional HRTEM image of S-M interface that saw processing similar to that of the lift-off resonators without/with descum. Elemental peaks in C and O from qualitative EELS scans across the interfaces are indicated. Figure adapted from Ref. [59].

later using PMMA-based EBL and lift-off [Fig. 4.18(b)], mimicking traditional transmon fabrication. After e-beam resist development, the wafer is optionally treated with a downstream oxygen ash descum before center trace deposition, during which the heated substrate sees purely chemical cleaning, but not ions or plasma. The wafer is then transferred to the same high-vacuum evaporator that provided the initial ground plane, and the lift-off resonator center traces deposited and excess metal lifted off in N-Methyl-2-pyrrolidone (NMP) at 80 °C (3 hrs.) and cleaned in IPA.

The resonator chip is wirebonded into an aluminum sample box, which is mounted on the 50 mK stage of an adiabatic demagnetization refrigerator equipped with sufficient filtering and shielding so that radiation and magnetic vortex losses are negligible [95, 39]. All resonators had $w, g = 15, 10\,\mu$m with frequencies near 6 GHz. Using a feedline [39]

allows us to reproducibly extract Q_i (as described in section 3.2.2) for multiple lift-off and etched resonators on the same chip. The resulting internal resonator quality factors are shown in Fig. 4.18(e). The decrease and saturation of Q_i at low powers for all resonators is the signature of TLS-dominated loss for both the control (fully etched) and degraded (lift-off) resonators. A clear difference (factor of 3) is observed in low-power Q_i between the etched resonators and the lift-off resonators without descum. As seen in Fig. 1(e), the descum increases the low-power Q_i back to or slightly below that of the control resonators. These measurements suggest that the edge profile of the resonators [Fig. 1(c/d)] had a negligible effect on loss at this level of dissipation. It is also apparent that roughness of the S-M interface had a minimal effect on loss: the substrate under the center trace of the lift-off resonators was previously etched,[13] and is three times rougher than that under the center trace of the control resonators [0.3 versus 0.1 nm RMS roughness as measured by atomic force microscopy (AFM)].

To help understand the increased loss in the lift-off resonators, which we attribute to a contaminated S-M interface, we use cross-sectional high-resolution transmission electron microscopy (HRTEM) to examine the S-M interfaces of samples that saw similar[14] processing to the center traces of the lift-off resonators without/with the descum [Fig. 4.18(f/g)]. With no descum, we observe two sublayers at the S-M interface. Directly

[13]The BCl_3/Cl_2 ICP dry aluminum etch used to define the resonators etches ~ 4 nm into the sapphire substrate.

[14]Minor differences arise due to fabricating multiple interfaces on a single TEM sample. The non-ashed S-M interface in the TEM sample saw processing temperatures up to 160 °C, whereas the ashed S-M interface in the TEM sample and in the resonators only saw processing temperatures up to 115 °C. As such, it is not certain if the upper AlO_x sublayer was present in the non-ashed lift-off resonators. In addition, the ashed S-M interface in the TEM sample saw an initial coating of e-beam resist and subsequent strip before a second coating for e-beam lithography.

above the substrate is a film of average thickness 1.6 nm, presumably residual resist polymer, which shows a peak in carbon content when probed with electron energy loss spectroscopy (EELS). Above this, a $\sim 2\,\mathrm{nm}$ layer with similar phase contrast to aluminum oxide is observed, accompanied by a peak in oxygen content when probed with EELS. This layer is likely formed by a reaction of the unpassivated Al with resist and/or solvent residue either as the metal is evaporated onto the substrate, or during a later processing step when the wafer is heated. As such, it may contain contaminants such as hydrogen that may increase dielectric loss through the introduction of GHz-frequency TLS defects [61, 154, 155, 156]. Oxide contamination from residue may be relevant to previous experiments finding that thermally oxidized submicron Josephson junctions are made significantly more stable by cleaning the substrate with oxygen plasma before metal deposition [157, 158].

The S-M interface of the descummed substrate on the other hand shows a decreased average thickness[15] of carbon-containing residue and no observed peak in oxygen content. We note that our data is not sufficient to determine whether or not the decrease in carbon residue is in direct proportion to the decrease in resonator loss.

In situ descum techniques such as ion beam cleaning may perform similarly to the downstream ash explored here. However, as this involves physical bombardment, a cleaning which if too aggressive might damage the substrate quality at the interface. To test

[15] AFM scans of the substrate after e-beam resist exposure and development also reveal residual resist granules with widths of $\sim 10 - 100\,\mathrm{nm}$ and heights $\sim 2 - 20\,\mathrm{nm}$, even significantly above the e-beam clearing dose exposure, consistent with literature on PMMA [159]. However, we expect these granules to be negligible sources of dielectric loss compared with residual films, as the space they fill is negligible (areal fraction $\sim 0.5\,\%$ and equivalent uniform thickness $\lesssim 0.1\,\mathrm{nm}$). Downstream ashing or UV-Ozone cleaning mostly removes these granules.

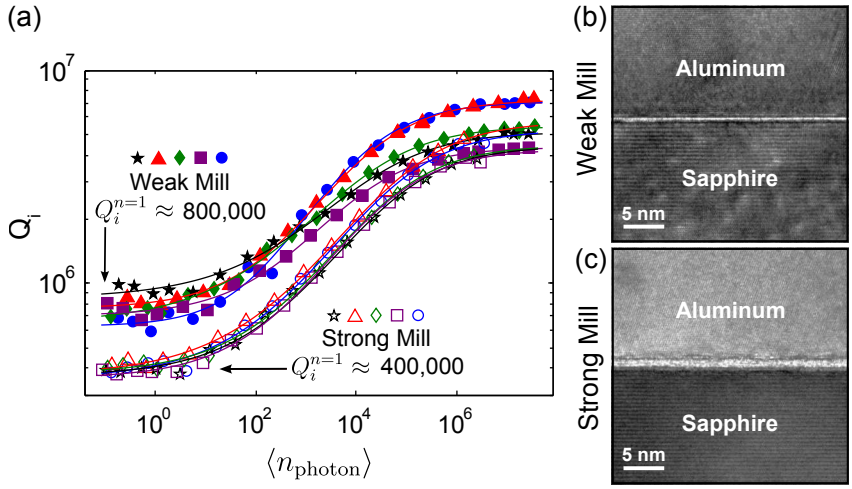


Figure 4.19: Comparison of weak and strong ion beam treatment, the latter inducing resonator degradation. (a) Q_i of etched CPW resonators whose bare substrates saw weak or strong *in situ* ion milling before base metal deposition (five resonators of each). (b/c) Cross-sectional HRTEM of S-M interface for weak/strong ion mill, showing thicker disordered interfacial layer for strong mill. Figure adapted from Ref. [59].

this hypothesis independently from questions of resist residue contamination, we fabricate etched superconducting $\lambda/4$ resonators whose substrates saw different strengths of *in situ* argon ion beam cleaning prior to the base aluminum deposition: a weak clean (beam energy 200 eV, dose $\sim 5 \times 10^{15}\,\mathrm{cm}^{-2}$) and a stronger mill (beam energy 400 eV and dose $\sim 5 \times 10^{17}\,\mathrm{cm}^{-2}$). The stronger parameters are identical to those used to etch away native AlO_x in the fabrication of Xmon and fluxmon qubits [40, 71] and similar to those used for substrate preparation in previous planar superconducting resonator experiments [160]. The resulting resonator quality factors are shown in Fig. 4.19(a), and display a power dependence consistent with TLS-dominated loss at low powers. We observe a clear difference (factor of 2) between the low-power internal quality factors, with the *stronger* ion beam yielding a lower Q_i.

Fig. 4.19(b/c) shows cross-sectional HRTEM images of the S-M interface for the

weak/strong ion beam treatments. The strong mill creates a $\sim 1.2\,\mathrm{nm}$ interfacial layer, significantly thicker than the weakly-treated interface of unresolvable thickness. EELS reveals no measurable elemental peaks at either interface, including Ar, C, and O. We do not believe the uniform interface is an artifact of surface roughness, as AFM scans reveal no change in roughness between a bare and a strongly milled wafer,[16] consistent with literature on sapphire [161]. We therefore attribute the excess loss to TLS induced by sapphire amorphization. Using finite-element COMSOL simulations [145], assuming a relative permittivity $\epsilon_r = 10$ for this layer we extract[17] an effective TLS loss tangent $\delta^0_{\mathrm{TLS}} \sim 1 \times 10^{-2}$. We note that recent work by Dunsworth *et al.* shows that this same ion mill on a silicon substrate can induce a significantly thicker damaged layer of similar or worse loss tangent [60], one of numerous reasons that fabrication of high quality superconducting qubits on silicon had been challenging for quite some time.

To our knowledge, our study of interfacial dielectric loss is the first systematic experiment that was able to both "see" with microscopy the cross-sectional interface getting worse and correlate that with degrading Q, showing that these thin lossy layers really do exist (although in the case of Fig. 4.18 we weren't able to tell whether the residue itself or the resulting under-oxide layer dominated the loss). In any case, a take-away message from our study of dielectric loss in resonators is that choice of fabrication methods can affect qubit coherence. We hope that this is also true for flux noise if the dominant

[16] In addition, the electron-transparent TEM sample was $50 - 100\,\mathrm{nm}$ thick, and no height variation of 1 nm was observed over this length scale with AFM. The interfacial layer was also uniform over a micron of TEM imaging.

[17] We only attribute participation to this layer at the S-M interface and not the partially etched S-V interface.

fluctuators reside on the qubit's surfaces – there is some evidence that surface treatment and/or vacuum conditions may be able to reduce or at least influence the functional form of intrinsic low-frequency flux noise [128]. However, to our knowledge no surface treatment has yet made the lowest noise samples have lower noise, only samples that had relatively high noise to begin with. This is on contrast with the case of dielectric loss, where it was shown that you can improve the Q of the best resonators by carefully chosen *in situ* surface treatments and growth conditions [39]. However, even for the best devices, the power-dependence of Q_i indicates that dielectric loss is still a (if not the) dominant source of loss in resonators.

We conclude this section by noting that our study of dielectric loss in CPW resonators allows us to isolate the contribution of dielectric loss (associated with $\mathcal{G}$ in table 4.2) to the fluxmon's T_1, since the fluxmon's capacitance is dominated by its CPW segment. For a $w, g = 15, 10\,\mu$m CPW, we therefore expect a fluxmon Q of $\sim$ 800k, while for a smaller dimension $w, g = 10, 2$ that we have been using for more recent fluxmon devices (favored for allowing denser circuits and also stronger inductive couplings given other device optimizations), we expect a fluxmon Q of $\sim$ 300k. Note that this is for the average "background" dielectric dissipation in the loss tangent limit of many defects. In a later section we will re-examine this assumption to see whether or not we also expect any individually resolvable strongly-coupled TLS defects to be present in the specific circuit of the fluxmon qubit.

4.2.2.5 Induced transitions from room temperature noise

There is always unwanted noise generated by our control electronics. Even if we heavily attenuate the output of the electronics at room temperature, there will still be room temperature Johnson noise entering the fridge. Typically we will attenuate all lines by 20 dB between room temperature and 3K (via lumped resistive attenuators at the 3K stage) since this temperature drop corresponds to a factor of 100 in classical Johnson noise power, and then below that stage we attenuate the remaining noise as much as practically possible (determined by factors such as resistor heating and range of the flux bias electronics), typically an extra 16 dB at the mix plate. For our flux bias lines, we need a large signal amplitude, and this noise cannot necessarily be ignored.

Noise entering the fridge from room temperature will act as an incoherent classical noise source on the qubit at the qubit's frequency [remember this means $S(\omega_{10}) = S(-\omega_{10})$], leading to equal up and down transition rates $\Gamma_{\uparrow}^{\text{noise}} = \Gamma_{\downarrow}^{\text{noise}} \equiv \Gamma_{\uparrow\downarrow}^{\text{noise}}$. If the qubit is initialized in the ground state, it will increase its population initially at a rate of $2\Gamma_{\uparrow\downarrow}^{\text{noise}}$ until it reaches a steady-state stray population of $\frac{1}{2}$. In, reality, the qubit will have some intrinsic source of dissipation, so the actual steady-state stray population is determined by a combination of the noise and the intrinsic dissipation,

$$p_{\text{stray}} = \frac{\Gamma_{\uparrow\downarrow}^{\text{noise}} + \Gamma_{\uparrow}^{\text{intrinsic}}}{2\Gamma_{\uparrow\downarrow}^{\text{noise}} + \Gamma_{\uparrow}^{\text{intrinsic}} + \Gamma_{\downarrow}^{\text{intrinsic}}}. \tag{4.73}$$

To compute $\Gamma_{\uparrow\downarrow}$, suppose we have a noise source at room temperature entering the fridge. Because we attenuate the microwave line heavily before it enters the fridge, we can

assume that the noise is simply Johnson noise from a room temperature $50\,\Omega$ resistor,[18] which has an open-circuit single-sided white spectral density of

$$S_{VV}^{+} = 4k_BTR \approx 1\,\frac{(\mathrm{nV})^2}{\mathrm{Hz}}. \tag{4.74}$$

This noise is then attenuated by matched attenuators in the fridge, and converts into a current noise at the qubit flux bias port (because the qubit bias line has a small [few pH] inductance and negligible dissipation, we can treat the bias line termination as a short). Because the attenuators are matched to $50\,\Omega$, the Johnson noise from the room temperature resistor will be attenuated by the same amount and so the single-sided current noise flowing across the bias line's short will be

$$S_{II}^{+} = \frac{S_{VV}^{+}}{(50\,\Omega)^2}A, \tag{4.75}$$

where the power attenuation factor is

$$A = 10^{-\frac{\text{attenuation in dB}}{10}}. \tag{4.76}$$

Note that this is different by a factor of 4 from the current noise that would occur across the load if the load were a matched $50\,\Omega$ instead of a short.

To compute $\Gamma_{\uparrow\downarrow}$ of the qubit, we simply apply the Fermi golden rule result (4.50)/(4.51) with $F = I$. It remains to find the form of the driving Hamiltonian associated with the flux bias current in order to extract g of the interaction Hamiltonian $\hbar g F(t)\sigma_{x/y}$ as in (4.40). This is simple – classically, the energy stored in the mutual inductance between

[18] It is empirically justified to ignore any GHz frequency room temperature electronics noise because we tested that changing the attenuation between the electronics and the fridge does not change the observed stray population in the qubit, at least for a sample with $M_t = 0.2\,\mathrm{pH}$. Typically this attenuation is large, around $\sim 20\,\mathrm{dB}$ between the electronics and fridge.

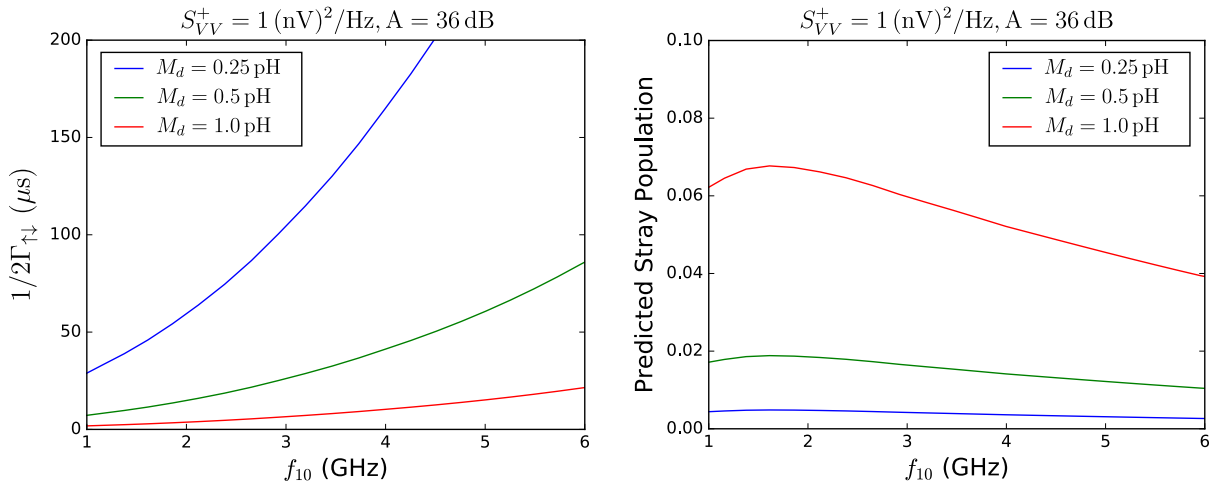


Figure 4.20: (a) Predicted bias line transition rate $1/(2\Gamma_{\uparrow\downarrow})$ vs. qubit frequency at zero tilt, for room temperature Johnson noise and 36 dB of in-fridge attenuators, for three different values of tilt bias drive mutual inductance M_d. (b) Predicted contribution of room temperature noise to stray population [Eq. (4.73)] based on the measured intrinsic T_1 of the qubit (which mostly decreases with decreasing frequency).

the drive line and qubit is $-M_d I_q I_d$, where I_d is the drive line current, which upon quantizing the qubit leads to

$$H_d = -M_d \frac{\hat{\Phi}}{L} I_d(t) \tag{4.77}$$

$$= -\left(\frac{M_d}{L}\right) \langle 0|\hat{\Phi}|1\rangle I_d(t)\sigma_x. \tag{4.78}$$

Reading off g then allows us to compute

$$2\Gamma_{\uparrow\downarrow} = \left[\frac{1}{\hbar}\left(\frac{M_d}{L}\right)|\langle 0|\varphi|1\rangle|\frac{\Phi_0}{2\pi}\right]^2 \frac{S^+_{VV}}{(50\,\Omega)^2} A, \tag{4.79}$$

where the factor of two is due to the fact that we have specified the Johnson noise with a single-sided spectrum.

In Fig. 4.20, we plot the expected stray population (which includes an extra frequency-dependent cable attenuation [around ~ 5 dB] in addition to the lumped element attenu-

ators shown in the wiring diagram of Fig. 3.11). We conclude that the predicted stray population is almost negligible, but will quickly become significant if the tilt bias line mutual inductance M_d is too large. To compromise between microwave noise and tilt flux bias range, we typically design devices to have $M_d \approx 0.3\,\mathrm{pH}$, which should give us at most $\sim 1\%$ stray population. An interesting observation is that for the previously mentioned sample that was purposely over-coupled at $M_d = 2.6\,\mathrm{pH}$, the observed stray population at $f_{10} = 5.9\,\mathrm{GHz}$ was 0.24, corresponding to an effective temperature of $\sim 200\,\mathrm{mK}$. Based on this number and the measured T_1 of 210 ns, we predict that room temperature noise should only have contributed $\lesssim 0.05$ to this number, meaning that the majority of the stray population must have originated from somewhere inside the fridge outside of the qubit box, perhaps due to hot attenuators. Note that we don't need to consider stray population from noise in the SQUID line, because this line is low-pass filtered to $\sim 100\,\mathrm{MHz}$ at the lowest stage of the fridge and so is only relevant for lower frequency noise.[19]

4.2.3 Measurement of dissipation in the fluxmon qubit

To experimentally measure T_1 and p_{stray} in the fluxmon qubit, we use the method of "swap spectroscopy" [162], whose pulse sequence is shown in Fig. 4.21(a). First, a π-pulse is calibrated at the qubit's d.c. operating point (typically $f_{10} \approx 5\,\mathrm{GHz}$ [$\beta \approx 1$], at zero tilt). To perform a swap spectroscopy experiment, the qubit is first excited using a π-pulse,

[19] Although, it is possible that this filter does not work above the spec of 20 GHz, in which case some high frequency noise might be entering the chip and exciting high frequency modes. At frequencies well above 20 GHz though the HIR filters should be dissipating any such noise.

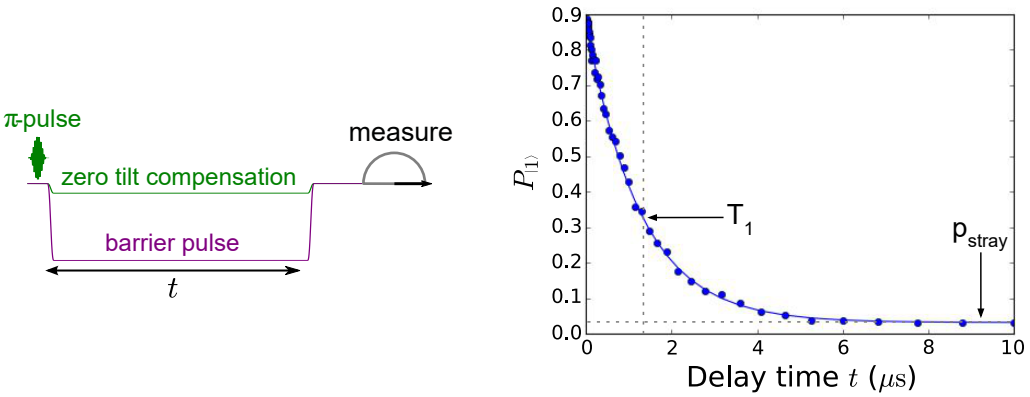


Figure 4.21: (a) Pulse sequence for measuring T_1 at a flux bias point different from the DC operating point via detuning pulses (this type of sequence is often called "swap spectroscopy" [162]). Typically, the tilt bias pulse is chosen to put the qubit at zero tilt bias, where flux noise is purely transverse to the energy eigenbasis. (b) Example dataset showing the exponential decay of $P_{|1\rangle}$ vs. time to a steady state stray population. T_1 and p_{stray} are extracted through a fit to a single exponential plus a constant offset.

and then a SQUID (barrier) pulse immediately detunes f_{10} to a different frequency. This detuning pulse is adiabatic yet much shorter than T_1. During the barrier pulse, in the presence of crosstalk we can optionally add a compensating tilt bias pulse to keep the qubit at zero tilt, where flux and charge noise are both purely transverse to the qubit's energy basis. We wait for a variable amount of time t at the detuned bias point, before tuning back to the original bias and then performing readout. After measuring $p_{|1\rangle}(t)$, we extract T_1 and p_{stray} by fitting to $p_{|1\rangle}(t) = p_0 \exp(-t/T_1) + p_{\text{stray}}$. An example time trace and fit is shown in Fig. 4.21(b).

4.2.3.1 Geometry dependence of T_1

In Fig. 4.22, we show experimental data of T_1 measured for fluxmon qubits of varying geometries. In Fig. 4.22(a), we show the results for three fluxmons on the same chip[20]

[20]It is important to compare qubits on the same chip because if the dissipation is due to material defects, it can and does vary somewhat from fabrication run to fabrication run.

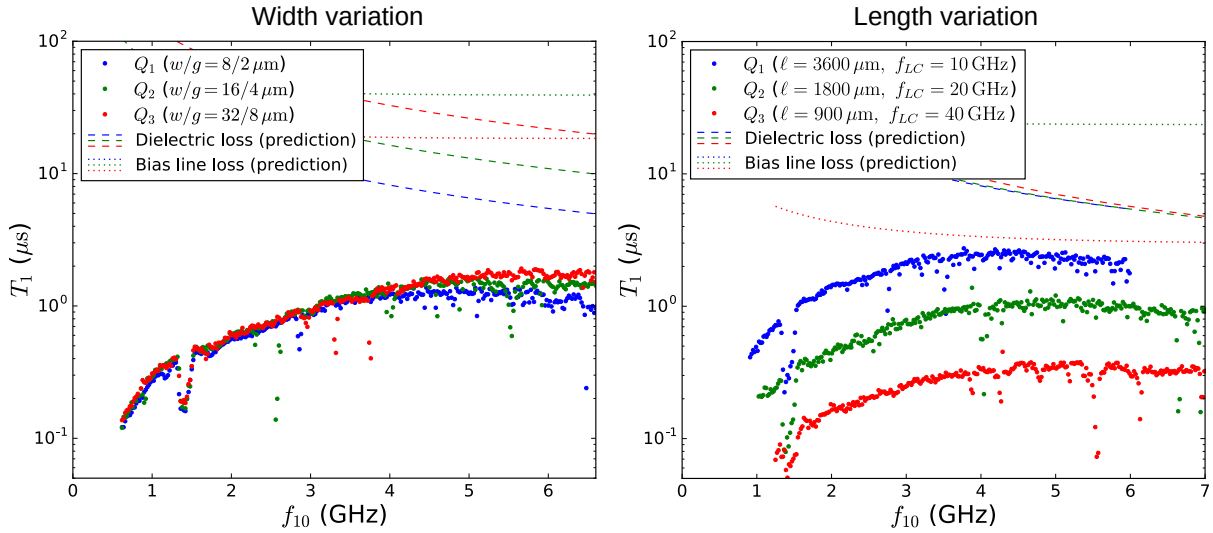


Figure 4.22: Measured T_1 vs. f_{10} at zero tilt via swap spectroscopy for three different fluxmon qubits on the same chip, for two different sapphire chips. (a) Width variation sample: fluxmons have different overall CPW cross-sectional scales (see device images in Fig. 4.4), but all qubits still had the same net L and C. (b) Length variation sample (Fig. 4.8): The length of the fluxmon CPW was varied while keeping the cross-sectional width constant at $w, g = 8, 2\,\mu$m. Therefore, the different qubits have different f_{LC} but the same Z_{LC}. All datasets were taken with a uniform frequency step size of 20 MHz.

with varying CPW widths, with a factor of two from one qubit width to the next (device image in Fig. 4.4). All qubits had the same length of CPW and the same L and C, and in addition the same Josephson junction critical currents, so from a lumped circuit perspective they are all equivalent except for their overall cross-sectional CPW dimension. We observe that over most of the frequency range $0.5 - 7$ GHz, there is no measurable trend in T_1 vs. CPW width. At high frequencies we see a slight width dependence, with narrow qubits having somewhat lower T_1. If the fluxmon were entirely limited by dielectric loss, we would expect a factor of 4 difference between Q_1 and Q_3 according to the interfacial energy participation ratio argument discussed in section 4.2.2, which is not observed. The prediction for dielectric loss (based on CPW resonator measurements) is shown with

color-coded dashed lines, and as can be seen the expected T_1 limit for the widest qubit is more than ten times higher than the observed T_1. The slight suppression in T_1 for the narrower qubits could be explained by dielectric loss, but this by no means explains the bulk of the dissipation at high frequencies, and certainly not at low frequencies. We also plot the expected bias line damping in color-coded dotted lines. Similarly, no other known source of loss explains the data, especially at low frequencies (see supplemental section 4.4 for an experimental check of quasiparticle dissipation and loss from magnetic vortices).

In Fig. 4.22(b), we show T_1 data for a sample containing three fluxmons of different length, with a factor of two difference from one qubit length to the next (device image in Fig. 4.8). In particular this means that f_{LC} varies by a factor of two from one qubit to the next, with the longer qubit having $f_{LC} = 10\,\mathrm{GHz}$, the middle qubit having $f_{LC} \approx 20\,\mathrm{GHz}$, and the shorter qubit having $f_{LC} \approx 40\,\mathrm{GHz}$. According to table 4.2, since the per-unit-length nature of the CPW is the same for all qubits, if the loss were dominated by capacitive dissipation represented by $\mathcal{G}(\omega)$ (such as dielectric loss), at a given frequency we would expect no difference in T_1 between these three qubits. On the other hand, if the dissipation originated in inductive loss represented by $\mathcal{R}(\omega)$, at a given frequency we should expect a factor of 4 difference in T_1 from one qubit to the next[21] (note again that this is in contrast with the case of $\lambda/4$ resonators, where we would not expect a change in T_1 with varying length, because f_{10} is always set to equal $f_{\lambda/4}$).

[21] Really it is more like a factor of 3.5, because we derived those formulas in the harmonic limit, which doesn't precisely hold at low frequencies, where "low" is relative to f_{LC}.

Experimentally, we observe a factor of ~ 3, which suggests we are dominated by some sort of inductive loss (there is a small contribution expected from dielectric loss and bias line damping). In a different language of dissipation as noise, one can arrive at the same conclusion by noting that the longer qubits should have twice as much flux noise power contributed by the CPW, which upon using Eq. (4.80) leads to the same result as table 4.2. This is a hint that we might learn something by interpreting the loss as due to high frequency flux noise, which is precisely what we'll do in the next section.

However, as already mentioned when we discussed the length dependence of the low frequency flux noise, there is a caveat in this interpretation, because in the length variation sample the Josephson junction critical currents were also varied by a factor of 2 from qubit to qubit to keep the same $\beta_{\max}$ for all qubits for readout purposes. The longer qubits had narrower junctions, and so it is possible that the increased dissipation came from the narrowing of the very short electrodes within the DC SQUID (think in terms of the R/W expression for flux noise power discussed in section 4.1.1). However, we do not think this explains it because we have measured qubits with different junction sizes and the same f_{LC} (albeit on different fab runs) and have not seen a systematic effect. A length variation test where the junction electrode dimensions are constant would help definitively answer this question. In either case, flux noise is still a consistent explanation of the observed dissipation. However, if the noise is really coming from the CPW, a mystery that still remains is why there is no width dependence to the T_1 as would be expected due decreasing magnetic participation of interfaces, analogous to how we

see a width dependence with dielectric loss due to decreasing electric participation of interfaces. As in our earlier discussion of low frequency noise, it is possible that we are in a regime where surface magnetic defects are not the dominant source of flux noise, in which case it might be hard to reduce the level of noise. This is an important question that needs to be studied further.

In both samples, there are various "defects" in the spectrum where T_1 drops significantly below the background over a narrow bandwidth. Many of these defects are narrow ($\sim 10\,\mathrm{MHz}$ is the narrowest we can observe due to the dephasing-limited qubit linewidth) and occur at random frequencies. One such feature is significantly wider ($\sim 100\,\mathrm{MHz}$) and always occurs at 1.4 GHz for all qubits. The likely source of these features are dielectric TLS defects in the qubit's relatively large junction area and weakly bound hydrogen atoms on the bulk of the qubit surface, respectively. Both of these will be discussed in detail soon.

4.2.3.2 Interpretation of dissipation as high-frequency flux noise: classical-quantum crossover

In the past decade, superconducting qubits have extended the measurement of flux noise to increasingly wider frequency ranges, showing a $1/f^{\alpha(f)}$ power spectrum from $f \approx 10^{-5}\,\mathrm{Hz}$ to $f \approx 1\,\mathrm{GHz}$ with α close to 1 at low temperatures [110, 114, 115, 116, 117, 118, 119, 62], and a big question that remains is, what is the high frequency cut-off of this noise? Although previous frequency-resolved measurements of the $1/f$ noise have used a variety of experimental methods, they have extracted only a single quantity,

the symmetrized spectrum $S_\Phi^+(f)$, to describe the noise. However, as is well known, a quantum environment can generate different noise spectra at positive and negative frequencies, $S_\Phi(-f) \neq S_\Phi(f)$, with $S_\Phi^+(f) \equiv S_\Phi(f) + S_\Phi(-f)$. The asymmetric quantum part $S_\Phi^-(f) \equiv S_\Phi(f) - S_\Phi(-f)$ is a measure of the environmental density of states. In thermal equilibrium, the two spectra are related through the fluctuation-dissipation theorem, $S_\Phi^+(f) = \coth(hf/2k_BT)S_\Phi^-(f)$ [138]. In the classical regime $f \ll 2k_BT/h$, S_Φ^+ dominates and S_Φ^- is negligible unless one is sensitive to weak dissipation. Conversely, in the quantum limit of high frequencies one has $S_\Phi^+ \approx S_\Phi^-$, meaning a single spectrum suffices unless one is sensitive to small levels of thermal noise. But when $f \sim k_BT/h$ or if there is non-equilibrium noise, a single spectrum no longer characterizes the environment. Solid-state qubits are uniquely well-suited to measure environmental quantum noise due to their strong environmental coupling and faithful individual qubit readout [163, 164]. There is evidence that $1/f$ noise in superconducting qubits contributes to relaxation in the quantum regime [114, 62], but there has been no frequency-resolved experimental distinction between S_Φ^+ and S_Φ^- or between equilibrium and non-equilibrium noise. Furthermore, the transition from classical to quantum flux noise has yet to be observed or understood.

Experimentally, obtaining $S_\Phi^-(f)$ in the classical or crossover regime is challenging, requiring high fidelity readout of the qubit population over a range of qubit frequency $f_{10} \lesssim 2k_BT/h$, outside typical qubit operating conditions. However, as we've seen, the fluxmon qubit's measurement scheme allows us to measure dissipation in this regime.

In this section, we use T_1 and stray population data to extract the full two-sided noise spectrum over a range of f containing $2k_BT/h \sim 1\,\mathrm{GHz}$. We observe a classical-quantum crossover, which we find to coincide with a transition from $1/f$ to quasi-ohmic dissipation. Remarkably, we find that $S_\Phi^+(f) \propto 1/f$ at 1 GHz, with a magnitude close to that extrapolated from the $1/f$ noise below 1 Hz. The level of $1/f$ noise at high and low frequencies changes similarly between samples (see section 4.4), providing evidence they may originate from the same physical source. Below the crossover we find that the environment is close to thermal equilibrium. We measure the T-dependence of $S_\Phi^\pm(f)$, and discover a paramagnetic $1/T$ scaling in S_Φ^-. Finally, we show that the small S_Φ^- in the classical regime has an important effect by predicting the reorganization energy during incoherent tunneling between flux qubit wells [20], a crucial quantity for modeling quantum annealers.

To better quantify dissipation in the fluxmon qubit and interpret it as equilibrium and/or non-equilibrium flux noise, we need to measure both the $|1\rangle \to |0\rangle$ and $|0\rangle \to |1\rangle$ transition rates as a function of f_{10}. It will be illustrative to convert these transition rates into an effective double-sided flux noise spectrum as follows. At zero tilt bias, the potential energy function is symmetric, meaning the energy eigenstates $|0\rangle$ and $|1\rangle$ have even and odd parity. Using this parity basis in the two-level approximation, the qubit Hamiltonian can be written as $\hat{H} = -\frac{1}{2}[\Delta(\Phi_{\mathrm{SQ}}^x)\sigma_z + \varepsilon(\Phi_t^x)\sigma_x]$, where the gap Δ/h is f_{10} at zero tilt, and $\varepsilon = 2I_p\Phi_t^x$. Here, $I_p \equiv \frac{1}{L}|\langle 0|\hat{\Phi}|1\rangle|(\sim 0.1 - 0.5\,\mu\mathrm{A}$ for Δ/h between 1 and 7 GHz) is the persistent current. At zero tilt, flux noise in the main qubit loop (i.e., in ε)

at f_{10} induces incoherent transitions between energy eigenstates according to the Fermi golden rule relations (4.50)/(4.51),

$$\Gamma_{\downarrow/\uparrow} = \frac{1}{\hbar^2}\frac{1}{L^2}|\langle 0|\hat{\Phi}|1\rangle|^2 S_\Phi(\pm f_{10}). \tag{4.80}$$

This implies a two-rate equation

$$\begin{pmatrix}\dot{p}_{|0\rangle}\\ \dot{p}_{|1\rangle}\end{pmatrix} = \begin{pmatrix}-\Gamma_\uparrow & \Gamma_\downarrow\\ \Gamma_\uparrow & -\Gamma_\downarrow\end{pmatrix}\begin{pmatrix}p_{|0\rangle}\\ p_{|1\rangle}\end{pmatrix}. \tag{4.81}$$

As can be seen from the eigenvalues of this linear system of differential equations, if prepared in either of its energy states the qubit will relax to a steady state population $p_{\text{stray}} = \Gamma_\uparrow/(\Gamma_\downarrow + \Gamma_\uparrow)$ at a rate $1/T_1 = \Gamma_\downarrow + \Gamma_\uparrow$. From T_1 and p_{stray} we can then extract both $S_\Phi^+(f)$ and S_Φ^- at the qubit frequency,

$$S_\Phi^+(f_{10}) = \frac{\hbar^2 L^2}{T_1|\langle 0|\hat{\Phi}|1\rangle|^2}, \tag{4.82}$$

$$S_\Phi^-(f_{10}) = [1 - 2p_{\text{stray}}(f_{10})]S_\Phi^+(f_{10}). \tag{4.83}$$

We measure T_1 and p_{stray} vs. f_{10} at zero tilt bias by varying the DC SQUID (barrier) bias, and use the data to extract $S_\Phi^\pm(f)$ after numerically computing $\langle 0|\hat{\Phi}|1\rangle$. This ability to tune Δ is crucial, as it allows us to vary f_{10} while remaining at zero tilt so as to be sensitive only to transverse noise in ε. It also allows measurement over a wide frequency range within a valid two-level approximation. A typical dataset for T_1 and p_{stray} vs. Δ/h is shown in Fig. 4.23 (a) and (b).

A typical dataset converted to spectral densities [Eqs. (4.82) and (4.83)] is shown in Fig. 4.23(c). In addition, the extracted effective temperature T_{eff} is plotted in Fig.

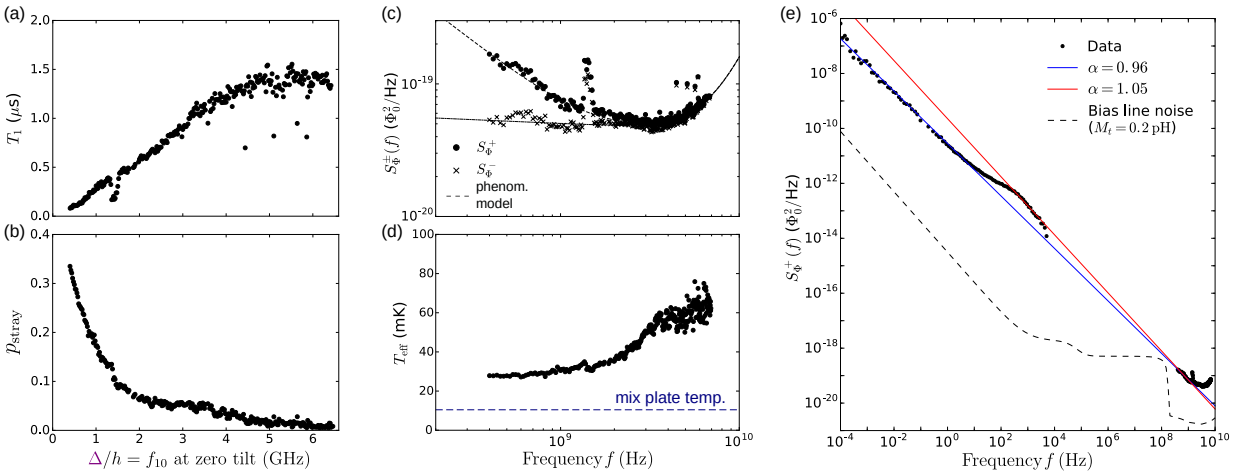


Figure 4.23: (a) T_1 versus f_{10} at zero tilt measured with swap spectroscopy, on a linear scale. (b) p_{stray} versus f_{10} at zero tilt on a linear scale. (c) $S_\Phi^{\pm}(f)$ extracted from (a) and (b) and numerical evaluation of $\langle 0|\hat{\Phi}|1\rangle$, on a log-log scale. (d) Effective temperature T_{eff} of the data in (b). (e) Low-frequency quasistatic flux noise together with $S_\Phi^+(f)$ from (c) on the same axes, on a log-log scale. Also shown is predicted noise entering through the qubit's tilt bias line (from known sources of thermal noise and room temperature electronics noise). Figure adapted from Ref. [71].

4.23(d), where $\exp(-hf_{10}/k_BT_{\mathrm{eff}}) \equiv \Gamma_\uparrow/\Gamma_\downarrow$. We observe several interesting features. First, below $\sim$ 1 GHz, $S_\Phi^+(f)$ follows a $1/f^\alpha$ law, with $\alpha \approx 1$. Remarkably, in Fig. 4.23(e) we find that extrapolating this power law to frequencies below 1 Hz predicts the magnitude of the quasistatic $1/f$ noise surprisingly closely. In thermal equilibrium, $S_\Phi^-(f)$ should scale as $f^{1-\alpha}$ at low frequencies, meaning a constant $S_\Phi^-(f)$ for $1/f$ noise, which is roughly what we observe. This is strong experimental evidence that $1/f$ noise extends all the way up to $\sim$ 1 GHz. This suggests, but does not show, that a single physical source is generating the noise over 10 orders of magnitude. At frequencies below the classical-quantum crossover, the noise appears to be described by a single $T_{\mathrm{eff}} \approx 30$ mK, suggesting thermal equilibrium of the low-frequency environment, but with $T > T_{\mathrm{fridge}}$ [Fig. 4.23(d)]. We believe the peak in dissipation at 1.4 GHz may be due to coupling to

the hyperfine transition of weakly bound hydrogen atoms on the qubit surface [165], as discussed in the next section.

Secondly, we find that the classical-quantum crossover is accompanied by a transition from $1/f^{\alpha}$ to super-ohmic dissipation, meaning noise for which $S_{\Phi}^{+}(f) \approx S_{\Phi}^{-}(f) \propto f^{\gamma}$ with $\gamma \geq 1$. If we fit to the phenomenological thermodynamic model $S_{\Phi}(\omega) = A\omega/|\omega|^{\alpha}[1 + \coth(\hbar\omega/[2k_B T_A])] + B\omega|\omega|^{\gamma-1}[1 + \coth(\hbar\omega/[2k_B T_B])]$, then $\alpha = 1.05$ and γ between 2.5 and 3 fits our data best. Purcell decay is negligible over the measured frequency range. We note that $\gamma = 3$ gives a frequency dependence for T_1 that is indistinguishable from ohmic charge noise ($S_Q^{-} \propto \omega$), the high-frequency model used in Ref. [62]. Dielectric loss [61] would give $\gamma = 2$, but based on similarly fabricated Xmon qubits and airbridges we estimate a limit of $T_1 \sim 20\,\mu$s at 5 GHz. Allowing for a finite high-frequency cutoff for the $1/f^{\alpha}$ noise could yield a different best-fit $\gamma \approx 1.5$ (see section 4.4). If we instead simply ignore the $1/f$ part and only fit the data above ~ 3.5 GHz, an ohmic flux noise model with $\gamma \approx 1$ describes the data reasonably well, with the net dissipation represented by a frequency-independent parallel resistance $R \approx 20\,\mathrm{M}\Omega$. Dissipation from the tilt flux bias line would similarly have $\gamma = 1$, but with $T_1 \approx 40\,\mu$s. Since the observed level of dissipation is not seen in the Xmon transmon qubit [40], which is comprised of a similar capacitance and Josephson inductance (albeit with a critical current density 10 times smaller) but negligible geometric inductance, we hypothesize that the quasi-ohmic noise is magnetic in nature. The presence of unexplained ohmic dissipation ($\gamma = 1$) was seen in microstrip-based flux qubits [166], but with a much stronger magnitude

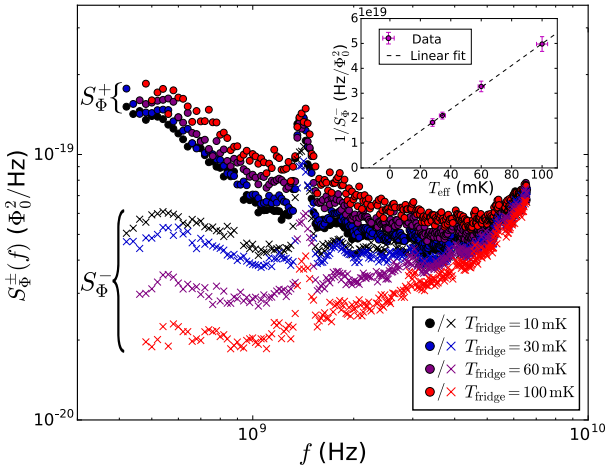


Figure 4.24: Temperature dependence of $S_\Phi^{\pm}$. $S_\Phi^-(f = 500\,\text{MHz})$ shows a $1/(T_{\rm eff} + T_0)$ dependence, as explicitly plotted in the inset. Figure adapted from Ref. [71].

($R \approx 20\,\text{k}\Omega$). Given that the extracted γ depends on whether the $1/f$ noise is included in the model, it could be that this earlier result was the combined effect of $1/f$ and super-ohmic dissipation. We note that at high frequencies, $T_{\rm eff}$ ranges from $50 - 80\,\text{mK}$, meaning a higher T for the quasi-ohmic bath and/or the presence of non-equilibrium noise, making it difficult to model.

4.2.3.3 Temperature dependence of the $1/f$ noise: paramagnetism

We investigate the nature of the $1/f$ noise further by looking at its temperature dependence. The temperature-independence of the classical low-frequency noise at millikelvin temperatures and the $1/T$ dependence of the static susceptibility in SQUIDs are evidence of a paramagnetic origin [111] for the noise. Here, we uncover dynamical evidence for this conclusion. As shown in Fig. 4.24, we find that S_Φ^- displays an approximately $1/T$ dependence below the classical-quantum crossover point, which through the fluctuation-dissipation theorem implies that $\chi''(\omega)$, the imaginary (absorptive) part of the environ-

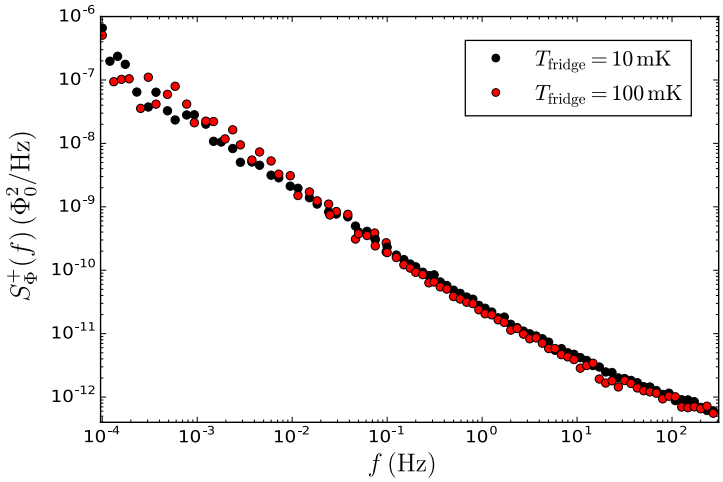


Figure 4.25: Quasistatic flux noise measurement is not noticeably changed between fridge temperatures 10 mK and 100 mK.

ment's dynamic susceptibility $\chi(\omega) = \chi'(\omega) + i\chi''(\omega)$, has a paramagnetic $1/T$ scaling, as discussed below. In comparison, S_Φ^+ displays only a very slight temperature dependence, consistent with the fact that we see no measurable temperature dependence in the quasistatic flux noise ($f < 10\,{\rm kHz}$), as shown explicitly in Fig. 4.25. In the inset to Fig. 4.24 we explicitly plot $1/S^-(500\,{\rm MHz})$ versus $T_{\rm eff}$, implying that $\chi''(\omega, T) \propto 1/(T + T_0)$, with $T_0 \approx 10\,{\rm mK}$. This functional form might be taken as evidence for paramagnetic spins that would behave antiferromagnetically at lower temperatures. We also note that a model with a temperature-dependent high-frequency cutoff of a few $k_B T_{\rm eff}/h$ (consistent with spin-phonon or spin-spin interactions) fits the crossover region vs. f and T somewhat better than one with a fixed or infinite cutoff (see 4.4).

As shown in Fig. 4.25, we do not see any systematic temperature dependence of the quasistatic noise at low frequencies or of the Ramsey decay time $T_{\varphi 2}$, consistent with a picture of a paramagnetic environment with temperature-independent relaxation times.

In what follows, we give a simple, generic argument for how the two-sided spectrum $S_\Phi^\pm(\omega)$ is related to the dynamic susceptibility $\chi(\omega) = \chi'(\omega) + i\chi''(\omega)$ of the magnetic environment, in order to clarify the discussion above. In particular, we outline how $S_\Phi^-(\omega)$ may be related to the absorptive part of the environment's linear response, and that its $1/T$ dependence is consistent with a paramagnetic environmental susceptibility.

The most common assumption in models for $1/f$ noise is that the noise comes from a collection of dynamical fluctuators, each characterized by an exponential decay with a single relaxation time. The $1/f$ scaling then arises from an exponentially broad distribution of these relaxation times. The fluctuators could be represented by weakly interacting spin clusters of various sizes on the qubit surface, where the relaxation time is exponentially dependent on cluster size. An alternative model invokes a spin diffusion mechanism where collective diffusion modes play the role of individual fluctuators, and the broad distribution of relaxation times arises from the highly non-uniform distribution of the magnetic field around the superconducting metal of the flux qubit loop.

Whatever the microscopic source of these fluctuators may be, each fluctuator will have an effective magnetic moment operator M_n that will couple flux into the qubit according

to $\Delta\Phi = g_n M_n$. The two-sided flux noise spectrum in the qubit can then be written as

$$\begin{aligned} S_\Phi(\omega) &= \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \langle \hat{\Phi}(\tau)\hat{\Phi}(0)\rangle \\ &= \int_{-\infty}^{\infty} d\tau e^{i\omega\tau} \left\langle \sum_{n,m} g_n M_n(\tau) g_m M_m(0) \right\rangle \\ &\approx \sum_n |g_n|^2 S_{M_n}(\omega), \end{aligned} \tag{4.84}$$

where S_{M_n} is the full two-sided spectral density of magnetization noise, and in the last line we have assumed negligible correlations between fluctuators (spin clusters/diffusion modes). We now use the fluctuation-dissipation theorem [140], which states that the equilibrium fluctuations of each moment are related to the dissipative part of its response to a non-equilibrium perturbation according to

$$S_{M_n}(\omega, T) = \hbar(1 + \coth[\hbar\omega/2k_B T])\chi_n''(\omega, T), \tag{4.85}$$

where $\chi(\omega, T) = \chi'(\omega, T) + i\chi''(\omega, T)$ is the frequency-domain linear response function (dynamical susceptibility) to a magnetic field. Inserting this relation above yields

$$S_\Phi(\omega, T) = \sum_n \hbar|g_n|^2 \left[1 + \coth\left(\frac{\hbar\omega}{2k_B T}\right)\right] \chi_n''(\omega, T). \tag{4.86}$$

Since $\chi''(\omega, T)$ is an odd function of ω, looking at the antisymmetric part shows that the experimental $1/T$ dependence of $S_\Phi^-(f)$ below the classical-quantum crossover is consistent with a dynamical susceptibility χ_n'' that scales as $1/T$ for all fluctuators. This is consistent with previous measurements of the $1/T$ dependence of the static susceptibility in SQUIDs [111], assuming a temperature-independent distribution of relaxation times.

4.2.3.4 Defects in the spectrum

We saw in the swap spectroscopy data (Fig. 4.22) that there were several "defects" in the qubit's T_1 spectrum that are random from qubit to qubit. Here, we argue that the majority of these defects are likely from TLS defects on the surface of the thin electrodes leading to the Josephson junctions. These are the same type of TLSs we discussed earlier in the context of dielectric loss, but we are now considering the effect of individual defects rather than just the average background loss tangent that arises in the continuum limit. It is known that TLS defects can reside in the thin oxide of the qubit's Josephson junction; for example, in phase qubits, individual junction defects could be observed and quantified [61]. These defects have been well-studied and the standard model is that in the AlO_x, TLS's have a per-volume distribution over dipole moment and frequency of [61]

$$\frac{d^2n}{d\omega dp} = \frac{\rho_0}{2\pi}\frac{\sqrt{1-p^2/p_{\max}^2}}{p}, \tag{4.87}$$

with $p_{\max} \approx 6\,\text{debye}$ and $\rho_0 \times (2\,\text{nm}) \approx 0.5/\mu\text{m}^2/\text{GHz}$, where we have plugged in 2 nm as a typical AlO_x thickness. Each TLS couples to the qubit through the qubit's electric field through a dipole interaction Hamiltonian $H_{\text{int}} = \vec{p}\cdot\vec{E} \approx \hbar g_{qd}(a^\dagger b + ab^\dagger)$. Assuming Markovian decoherence for both qubit and defect, in the incoherent defect limit $\Gamma_q \ll \Gamma_d$, $\Gamma_d > 4g$, where Γ_q and Γ_d are the intrinsic qubit and defect decoherence rates, the defect will induce exponential relaxation of the qubit according to [40]

$$\Gamma_\downarrow(\omega_{10}) = \frac{2g^2\Gamma_d}{\Gamma_d^2 + (\omega_{10}-\omega_d)^2}, \tag{4.88}$$

where ω_{10} is the qubit frequency.

We note that we can obtain the continuum limit by integrating this equation over all defects with the distribution (4.87), which yields a T_1 that is equivalent to having a frequency-independent loss tangent within the TLS-hosting volume,[22]

$$\tan\delta = \frac{\rho_0}{h}\frac{p_{\max}^2}{6\epsilon}. \tag{4.89}$$

We will shortly use this result later as a sanity check when we perform a more realistic Monte Carlo simulations with discrete collections of defects.

However, for small sub-micron junctions, like those of the Xmon and fluxmon qubits, it is believed that junction defects are for the most part statistically avoided [150], and defects on the electrodes near the junctions are a more likely culprit. Results from our group over the past few years with sub-micron junctions in Xmon, gmon, and fluxmon qubits suggest that junction areas greater than $\sim 0.2\,\mu\text{m}^2$ have significantly more strong defects per bandwidth, but it is not known for sure whether these defects reside on the thin electrodes leading up to the junction (which generally have the same width as the junctions) or within the junctions themselves. However, the former may be more likely, since if a defect were located in the junction itself, it would feel electric fields on the order of a few thousand V/m, which for a dipole moment of 6 debye means a coupling of $\sim 100\,\text{MHz}$, meaning they would almost all strongly violate the incoherent defect limit assumption and exhibit coherent swapping with the qubit, unless there was some preference for junction defects to be aligned parallel with the plane of the junction.

[22]Note interestingly that in this continuum limit, the resulting loss tangent is actually independent of the distribution of Γ_d, thanks to the Lorentzian integrand.

Of course it is also possible that the defects in these junctions simply have smaller dipole moments than expected. We observe that most defects do not show any coherent swapping, although once in a while (roughly once per two samples) we do see a coherent defect (usually on samples with relatively large junctions) that could indeed be coming from inside the junctions. The fluxmon qubits measured in this work have relatively large junctions compared to standard Xmon junction designs, around $0.15\,\mu\text{m}^2$ or more in measured area,[23] leading to more defects with T_1 below $\sim 1\,\mu\text{s}$ than in the best Xmon qubits.[24]

Of course, the other possibility is that the defects reside in the CPW interfaces that we considered in the resonator experiments of section 4.2.2. However, because there is roughly an equal number of defects with roughly equal strength for both narrow and wide qubits, we believe that this is unlikely to be the case: an on-resonance defect will damp the qubit beyond the background damping at an excess rate proportional to the electric field squared [see Eq. (4.88)], which between qubits 1 and 3 of the length variation sample should be over a factor of 10 change, but the defects are observed to be similar between them.

As an extra sanity check to further confirm this conclusion, we perform a Monte Carlo simulation of defects in the CPW interfaces, making some reasonable assumptions about the TLS distribution. If we assume the same TLS distribution as AlO_x and "calibrate"

[23] We can reduce this in the future. This was a consequence of having different qubit designs on the same chip for testing. When the qubits are uniform, we can choose a single small junction size and adjust the oxidation accordingly. Note however that there is a conflict between having small junction size and small junction asymmetry, a tradeoff that we will need to work on.

[24] Xmon qubits have a similar lumped element capacitance $\sim 100\,\text{fF}$, so this a fair comparison to make.

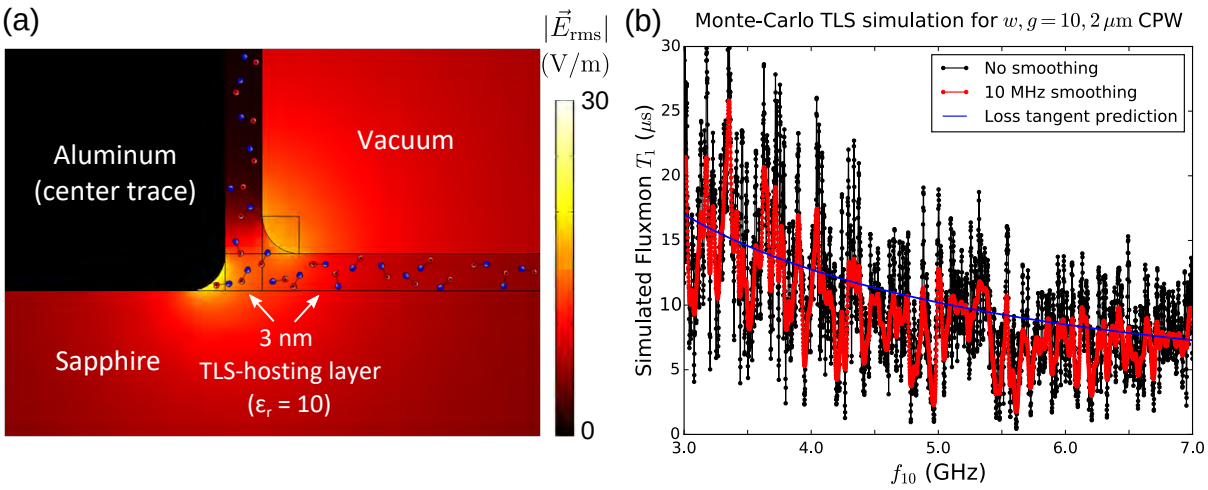


Figure 4.26: Simulation of vacuum-interface TLS defects in the fluxmon qubit, assuming a 3 nm thick interfaces with the same dielectric constant and TLS distribution as AlO_x. (a) COMSOL simulation of quasi-static electric field strength along as cross-section of the fluxmon's CPW near the SQUID end. Over 100,000 defects in the simulation. Dipole defect cartoons are just for show and not to scale. Sub-angstrom meshing was used near the metal corners, although we found that putting a 3 nm rounding on the corners (more realistic) didn't make any observable difference in the simulation results compared to a perfectly sharp corner. (b) Result of Monte Carlo simulation of defects using experimentally motivated parameters for the distribution of TLS dipole moments and linewidths as discussed in the main text. The black data is the expected T_1 with a 1 MHz spacing in frequency. The red data is the same but with a 10 MHz Gaussian smoothing to mimic the effect of the qubit's dephasing-limited linewidth. The blue curve is the expected T_1 based on the many-defect loss tangent formula (4.89) and the simulated participation ratio (4.72), showing good agreement.

the thickness of the substrate-vacuum interface to $\sim 3\,\mathrm{nm}$, we find that the simulation reproduces both the experimentally measured CPW resonator Q and the typical variation in Q across different $\lambda/4$ CPW resonators.[25] With this justification, we show the result of the analogous simulation for the fluxmon qubit CPW, summarized in Fig. 4.26. The simulation uses the electric field at the interface as simulated by COMSOL (the same technique used in Ref. [145]), and places defects throughout the interfacial volume according to the distribution (4.87). Since we are not in the continuum limit, the intrinsic linewidths of the defects matter for the simulation. We use the experimentally measured range of 30 - 150 ns for Γ_d [167, 40]. The simulation results are shown for a $w, g = 10, 2\,\mu\mathrm{m}$ CPW fluxmon. The result in red shows that we would not expect any defects below 1 μs, but of course this depends on several assumptions such as maximum TLS dipole moment and TLS coherence time. Nevertheless, for larger CPW dimensions, the excess loss from worst-case defects should be much better, which it isn't. We also note that the simulation shows that for a relatively narrow $w, g = 10, 2\,\mu\mathrm{m}$ fluxmon we may expect roughly one to two defects over a 4 GHz bandwidth to be slightly coherent $4g \gtrsim \Gamma_d$, which could possibly explain the occasionally observed coherent defect.

A third possibility is that the defects are on the underside of the airbridge crossovers of the qubit CPW; however, a simple parallel-plate capacitor calculation shows that the electric field at these defects would be only $\sim 0.1\,\mathrm{V/m}$, not large enough to have a noticeable effect. We therefore conclude that the TLSs likely reside in the junction

[25] The resonator simulations were for a $w, g = 24, 24\,\mu\mathrm{m}$ CPW and had over 1 million defects, yielding a mean Q of 1.5 million and ~ 25 percent expected variation of Q over frequency.

area. Unfortunately it is very hard to simulate the electric field at these junction-area electrodes, as it would require a very fine mesh on a large 3D simulation, but qualitatively we can see why this might be, as these electrodes will have strong electric fields over a larger area, meaning a given defect is more likely to find itself in a large field of order 10 V/m.

So far we have only considered dielectric defects. But there is in principle the possibility of magnetic defects as well. In the following section, we amazingly find one such defect coming from a known physical source, the ground state hyperfine transition of an ensemble of physisorbed atomic hydrogen.

4.2.4 Adsorbed hydrogen as a dissipation source: silicon vs. sapphire

The mysterious feature at 1.4 GHz in the swap spectroscopy data, where there is a peak in dissipation, is present at the same frequency and similar strength in all sapphire-substrate qubits of all geometries. There is reason to believe that this feature is something physically intrinsic to the materials of the qubit, because it persists even when significantly changing the filtering and attenuation on all the coaxial lines going to the qubit box (including the addition of attenuation on the output of the readout line), and is furthermore independent of chip size and box size, and so is unlikely to be due to coupling to an environmental electromagnetic mode.

Recently, using on-chip ESR techniques, de Graaf *et al.* [165] uncovered compelling

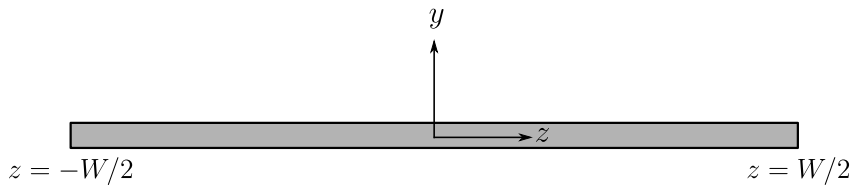


Figure 4.27: Cross section of center trace of fluxmon CPW (treated as a thin film) and definition of coordinates. Fluxmon current flows into the page (x direction), generating a magnetic field in the z direction. $W \approx 10\,\mu$m is the width of the CPW center trace.

evidence for coupling of superconducting resonators to the ground state hyperfine transition of surface hydrogen, whose presence is claimed to be due to physical adsorption as a by-product of water dissociation [165, 168, 169]. It so happens that the ground state hyperfine transition frequency of hydrogen is 1.42 GHz, the frequency of the feature observed in our fluxmon qubits. Below, we provide a rough calculation showing what surface density of weakly bound hydrogen atoms would be needed to explain the peak in dissipation observed in the fluxmon.

We suppose that there is a distribution of hydrogen atoms weakly bound to the surface of the fluxmon's CPW segment, so that the ground state hyperfine levels are roughly those of an isolated H atom. We will later consider instead what happens if there defects are located on the substrate-air interface rather than any of the metal interfaces. First, we write down the interaction Hamiltonian between the fluxmon and a single such atom (electron plus proton) at a lateral position z ($-W/2 \leq z \leq W/2$) on the surface of the

CPW center trace (Fig. 4.27),

$$\begin{aligned}\hat{H}_{\text{int}} &= -\boldsymbol{\mu}_e \cdot \mathbf{B}(z) - \boldsymbol{\mu}_p \cdot \mathbf{B}(z) \\ &= -\mu_B \hat{B}_z(z)\sigma_e^z - \frac{g_p e}{2m_p}\frac{\hbar}{2}\hat{B}_z(z)\sigma_p^z,\end{aligned} \tag{4.90}$$

where we have assumed the magnetic field generated by the qubit at the atom is in the z direction (see Fig. 4.27). The magnetic field generated by the qubit on the surface of the metal at position z is related to the qubit current operator according to

$$\hat{B}_z(z) = \frac{\mu_0 \hat{I}}{\pi W}\frac{1}{\sqrt{(1-\frac{2z}{W})(1+\frac{2z}{W})}} \tag{4.91}$$

This equation is valid as long as z is further than $\sim \lambda^2/2t \approx 50\,\text{nm}$ from the edge of the CPW strip [170]. Note that it is normalized so that effectively half the current flows on the bottom and half flows on the top side of the CPW. Since the CPW center trace dominates the surface magnetic participation, we have ignored the CPW ground planes.

We claim that the qubit's magnetic field (which is uniform over the hydrogen atom) will couple to the transition between the singlet hyperfine ground state $|S\rangle$ and the $m = 0$ triplet state $|T_0\rangle$ 1.4 GHz above. We will want to do a Fermi's golden rule calculation to calculate the induced $|e\rangle \to |g\rangle$ transition rate in the qubit. This requires computing the matrix element between initial and final states, $|i\rangle = |e\rangle\,|S\rangle$ and $|f\rangle = |g\rangle\,|T_0\rangle$, where $|S\rangle$ and $|T_0\rangle$ are the singlet and $m = 0$ triplet states

$$|S\rangle = \frac{1}{\sqrt{2}}(|\uparrow_e\rangle|\downarrow_p\rangle - |\downarrow_e\rangle|\uparrow_p\rangle), \tag{4.92}$$
$$|T_0\rangle = \frac{1}{\sqrt{2}}(|\uparrow_e\rangle|\downarrow_p\rangle + |\downarrow_e\rangle|\uparrow_p\rangle),$$

where e and p stand for electron and proton. We compute for a H atom at position z

$$\langle f|\hat{H}_{\rm int}|i\rangle(z) = \langle g|\hat{I}|e\rangle\mu_B\frac{\mu_0}{\pi W}\frac{1}{\sqrt{(1-\frac{2z}{W})(1+\frac{2z}{W})}}, \tag{4.93}$$

where we have dropped a second term due to the proton because it is much smaller (by a factor of $\sim \frac{m_e}{m_p}$). We recognize $\langle g|\hat{I}|e\rangle = \frac{1}{L}\langle g|\hat{\Phi}|e\rangle$ as the qubit persistent current I_p, which at 1.4 GHz is approximately $0.3\,\mu$A.

Now we are ready to use Fermi's golden rule to compute the decay rate induced in the qubit by a large collection of hydrogen atoms. We need to know the density of states of hyperfine splittings. Since the peak in qubit dissipation at 1.4 GHz has a linewidth of 150 MHz, significantly wider than the dephasing-limited qubit linewidth of ~ 20 MHz, we assume the hyperfine splittings are spread out over ~ 100 MHz, presumably due to an inhomogeneous weak interaction with the AlO_x surface.[26] This gives us the density

[26] A naive interpretation Ref. [171] suggests that this spread might be possible if the hydrogen is weakly bound to O atoms with a mean distance of ~ 4.5 Bohr radii with an inhomogeneous spread of ~ 0.2 Bohr radii. Further theoretical study of the surface physics is needed to determined whether or not this spread is physically realistic.

of states $\rho(E_{10})$ to use in Fermi's golden rule,

$$\Gamma_{|e\rangle\to|g\rangle} = \frac{2\pi}{\hbar}\int_{-W/2+50\,\mathrm{nm}}^{W/2-50\,\mathrm{nm}} dz\,|\langle f|\hat{H}_{\mathrm{int}}|i\rangle(z)|^2\rho(E_{10}), \tag{4.94}$$

where for simplicity we use a length cutoff of 50 nm $\approx \lambda$ from the CPW edge in order to have a convergent integral. Using the true current distribution that correctly treats the edge is not expected to significantly change the answer. Using the fact that the fluxmon has a length of 2 mm (i.e., the dimension going into the page in Fig. 4.27), we back out that to get the experimentally observed excess dissipation $\Gamma_{|e\rangle\to|g\rangle} \approx 1/(200\,\mathrm{ns})$, we must have an areal density of $\sim 1\times10^{16}/\mathrm{m}^2$, or about 1 H atom per 10 nm, on the AlO_x surface. We note that this number is reasonably consistent with the spin defect densities of $\sim 1-5\times10^{17}/\mathrm{m}^2$ typically inferred from low-frequency flux noise measurements in SQUIDs [172, 111]. This is remarkable and perhaps a coincidence, given that these two numbers are based on measurements at frequencies separated by about 10 orders of magnitude.

Remarkably, in experiment, we only see a dip in T_1 at 1.4 GHz on sapphire samples, not on silicon samples. For example, Fig. 4.28 shows a typical swap spectroscopy dataset on a silicon fluxmon. This clear difference tells us that the substrate plays an important role and that the hydrogen is likely located there. This means one of two things – the hydrogen only forms on the bottom side of the metal (i.e., at the substrate-metal interface), or in the CPW gaps on the substrate-vacuum interface. If the former is true, this simply increases the necessary spin density by a factor of 2, and if the latter is true it would increase by roughly a factor of 5, so is still physically feasible. However, we

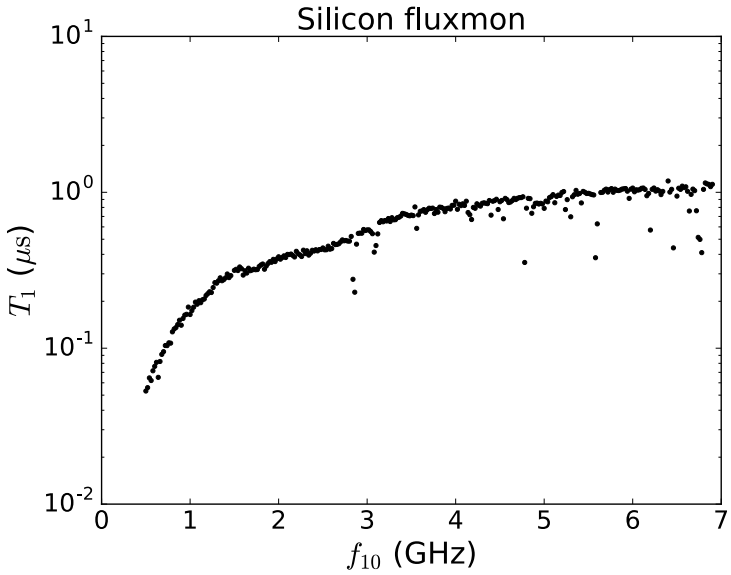


Figure 4.28: Typical swap spectroscopy T_1 data on silicon, showing an absence of any hydrogen feature at 1.4 GHz.

believe that the latter is more likely to be true given he theoretical studies in Ref. [169] and the recent experimental data from de Graaf *et al.* [165]. We note that it's possible that the surface hydrogen is still there on silicon, but is interacting more strongly with the silicon than it did sapphire, which could smear out the dissipation over a wider range of frequencies, possibly explaining the slightly suppressed T_1 between 1.5 and 3 GHz in Fig. 4.28. We also note that the background level of dissipation and the low-frequency quasistatic tilt flux noise did not change systematically between sapphire and silicon. In fact, within the margin of error and sample-to-sample variation of the measurement, the low frequency tilt flux noise had the same magnitude (with silicon being perhaps slightly better) and exponent α between sapphire and silicon, as can be seen in the data shown in Fig. 4.29, although we need a larger sample size to make a firmer conclusion.

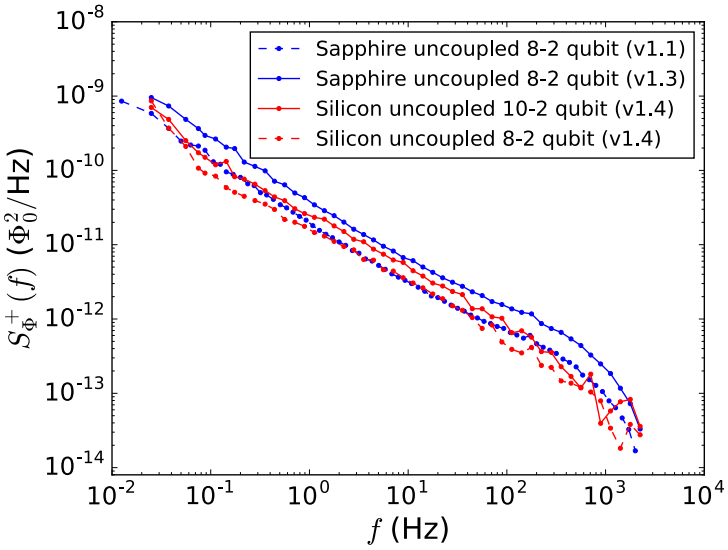


Figure 4.29: Comparison of low-frequency flux noise for fluxmon qubits of similar geometries on different samples, with two qubits on sapphire and two qubits on silicon. Given the sample-to-sample variation and small sample size, we cannot definitively conclude whether the noise is any better on silicon. All of these qubits were isolated qubits meaning they are not coupled to any other qubits through a coupler circuit.

4.3 Macroscopic resonant quantum tunneling

As described in Chapter 1, quantum tunneling is an important computational resource for a quantum annealer. Here, we study incoherent quantum tunneling between the two wells of the flux qubit potential to simulate what will happen in a quantum annealer solving a problem with a small energy gap, and in particular study how the tunneling rates depend on noise and dissipation. Using the independent T_1 and flux noise measurements from the previous sections, we will show that we can roughly predict the performance of quantum tunneling, in particular the linewidth of resonant quantum tunneling and the associated environmental "reorganization energy" penalty. We will find that the measured tunneling rates are consistent with the measured S_Φ^- of Fig. 4.23 extending deep into the classical regime, further justifying the interpretation of dissipation as flux noise.

We look at the effect of dissipation on incoherent macroscopic resonant tunneling (MRT) between the lowest states of the left and right flux qubit wells, which is illustrated in the inset of Fig. 4.30(b). In the regime of large β, the tunnel coupling Δ/h is much smaller than the linewidth of the energy levels, meaning that quantum tunneling will be incoherent, that is, described by a rate rather than by coherent swapping of population between the wells. For the level of damping and dephasing in our system, at temperatures below $T_{\rm cr} \approx 200\,{\rm mK}$ the escape rate from one well to the other near resonance should be dominated by quantum tunneling [173, 22, 91, 92]. Rewriting the result of [92], the tunneling rate as a function of tilt bias energy ε is predicted to be[27]

$$\Gamma_{L\to R}(\varepsilon) = \frac{\Delta^2}{4\hbar^2}\int_0^\infty dt\, e^{I_A(t)}\cos[\varepsilon t/\hbar - I_B(t)], \tag{4.95}$$

where

$$\begin{aligned} I_A(t) &= \int_{f_l}^{f_h} df\, \frac{(2I_p)^2}{(hf)^2} S_\Phi^+(f)\cos(2\pi f t), \\ I_B(t) &= \int_{f_l}^{f_h} df\, \frac{(2I_p)^2}{(hf)^2} S_\Phi^-(f)\sin(2\pi f t), \end{aligned} \tag{4.96}$$

and f_l and f_h are appropriate low and high frequency cutoffs. Assuming the integrated noise is dominated by frequencies smaller than the resonant tunneling linewidth W/h, then near its peak Eq. (4.95) can be approximated as a Gaussian,

$$\Gamma(\varepsilon) = \sqrt{\frac{\pi}{8}}\frac{\Delta^2}{\hbar W}\exp\left[-\frac{(\varepsilon-\varepsilon_p)^2}{2W^2}\right], \tag{4.97}$$

[27] See the supplemental section 4.4 for a derivation of the equivalence of these equations and those in Ref. [92].

where [92]

$$W^2 = 4I_p^2 \int_{f_l}^{f_h} df\, S_\Phi^+(f),$$
$$\varepsilon_p = 4I_p^2 \int_{f_l}^{f_h} df\, S_\Phi^-(f)/(hf) \tag{4.98}$$

are the linewidth and reorganization energy. Intuitively, the linewidth W is simple to understand, as it is essentially just integrating the drift in energy level between some frequency cutoffs determined by the details of the experiment. In addition to this noise broadening, we see that we expect the two Gaussian peaks to offset from zero tilt in opposite directions depending on whether the tunneling is from right to left or left to right [20]. Physically, this offset energy ε_p for the maximum tunneling rate physically corresponds to the reorganization energy of the environment that must be absorbed upon tunneling. More intuitively, if a very low frequency environment dominates, this energy bias offset simply corresponds to the polarization of environmental spins in response to the qubit's magnetic field. More generally, we can understand ε_p as the classical energy absorbed by the environment during a tunneling event as follows. Using (4.38) to write $S_\Phi^-(\omega) = L^2 S_I^-(\omega) = L^2 2\hbar\omega \mathrm{Re}[Y(\omega)]$ and recognizing $2I_p \approx \frac{1}{L}\Delta\Phi$, where $\Delta\Phi$ is the change in qubit flux during a tunneling event, we can rewrite (4.98) as

$$\varepsilon_p = \int_0^\infty \frac{d\omega}{2\pi}\, 2(\Delta\Phi)^2 Y(\omega), \tag{4.99}$$

where for simplicity we are for now ignoring the exact low and high frequency cutoffs. Intuitively, if we model the tunneling as an instantaneous change in flux $\Delta\Phi$ and therefore

a voltage spike $V(t) = \Delta\Phi\delta(t)$, the single-sided power spectrum of this voltage spike is constant over frequency with a magnitude of $2(\Delta\Phi)^2$. Therefore, the integration in (4.99) corresponds to integrating in the frequency domain the power dissipation due to $Y(\omega)$ in response to the voltage spike. Note that Eq. (4.99) does not contain $\hbar$ and therefore represents the classical energy dissipated during a tunneling event. As a concrete example, consider the special case of an environment consisting of a single low frequency resonant mode represented by a series inductance L_r and capacitance C_r, with oscillation frequency $\omega_r = 1/\sqrt{L_r C_r}$. Since the resonance period $1/\omega_r$ is much longer than the tunneling voltage pulse $V(t)$ (i.e., the characteristic tunneling time), then the current through the environmental inductor due to the voltage pulse adds coherently across the pulse,

$$I_{L_r} = \frac{1}{L_r}\int dt\, V(t) = \frac{\Delta\Phi}{L_r}, \tag{4.100}$$

corresponding to an energy

$$E_r = \frac{1}{2} I_{L_r}^2 L_r = \frac{(\Delta\Phi)^2}{2L_r} \tag{4.101}$$

transferred to the environmental mode. This agrees with the expression (4.99) for ε_p when we substitute the admittance of a single mode from Eq. (4.24), $\mathrm{Re}[Y_r(\omega)] = \frac{\pi}{2L_r}\delta(\omega-\omega_r)$. For an arbitrary collection of low-frequency modes, through ε_p it is possible to observe the integrated effect of dissipation and therefore the integrated effect of S_Φ^- in the deep classical regime.

We measure the small-time[28] incoherent tunneling rate $\Gamma(\varepsilon)$ using the pulse sequence

[28]By small-time, we mean we only consider tunneling times for which the probability of having tunneled is less than ~ 0.25, which is where the probability vs. time curve is linear.

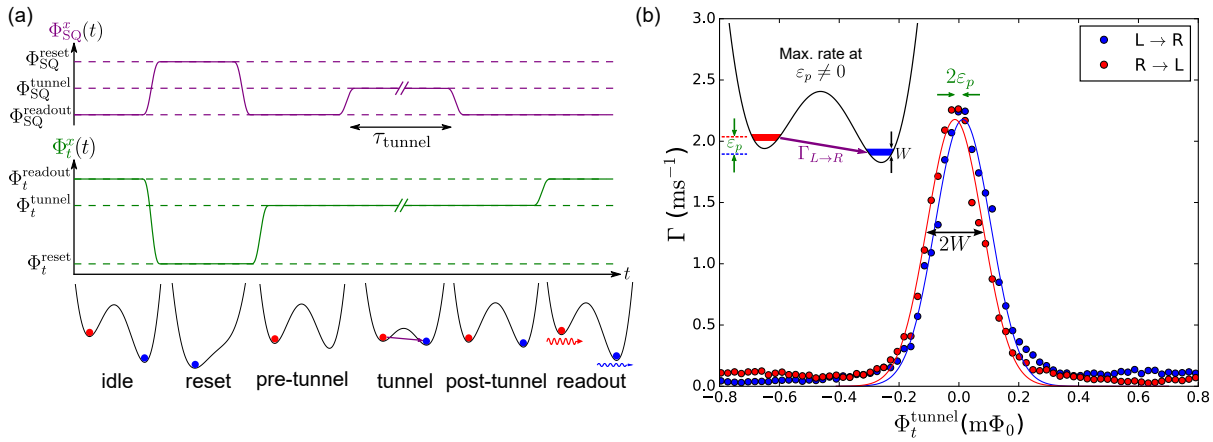


Figure 4.30: (a) Pulse sequence for the MRT experiment. (b) MRT data with $\Delta/h \approx 0.5\,\mathrm{MHz}$, showing a small offset in the tilt bias that maximizes tunneling, consistent with the measured S^-_Φ. Figure adapted from Ref. [71].

in Fig. 4.30(a). We prepare either the left or right well ground state with a high barrier, and as a function of tilt bias lower the barrier to $\beta \approx 1.5$ ($\Delta/h \approx 1\,\mathrm{MHz}$) and measure the incoherent tunneling rate to the other well. A typical dataset is shown in Fig. 4.30(b). Fitting the tunneling peaks to Gaussians, over multiple datasets we extract $\varepsilon_p/(2I_p) = 7 \pm 3\,\mu\Phi_0$ and $W/(2I_p) = 80 \pm 20\,\mu\Phi_0$. Above base temperature W is not changed within the margin of error, but ε_p becomes too small to reliably measure.

We can compare W and ε_p to that expected from directly integrating $S^\pm_\Phi(f)$ according to (4.98), interpolating a $1/f$ power law between the noise measured at low and high frequencies [Fig. 4.23(e)]. Including the ohmic noise leaves the tunneling rate virtually unaffected near the peak even when integrating (4.95) up to $f_h = 10\,\mathrm{GHz}$, the oscillation frequency of the inverted potential barrier, the natural high frequency cutoff [174, 175]. The natural low-frequency cutoff for W and ε_p is the peak tunneling rate itself, $\sim 10^3\,\mathrm{Hz}$ (see section 4.4 for a discussion of both the low and high frequency cutoffs). However,

there is additional broadening of W due to quasistatic noise averaged over experimental repetitions, which amounts to extending the low frequency cutoff for W down to the inverse total data acquisition time (see supplemental section 4.4). Using these cutoffs we predict $W/(2I_p) \approx 50\,\mu\Phi_0$ and $\varepsilon_p/(2I_p) \approx 4\,\mu\Phi_0$, within a factor of two of the measured values. For low-frequency noise in thermal equilibrium, one would expect [20]

$$T = W^2/(2k_B\varepsilon_p). \tag{4.102}$$

Plugging in our measured W and ε_p yields $T_{\text{eff}} \approx 60\,\text{mK}$, higher than the 30 mK deduced in Fig. 4.23(d). However, this may be explained by the extra broadening of W from quasistatic noise. Subtracting out our estimation of this contribution yields instead 20 mK. An alternative explanation for the inferred effective temperature being high is that the environment did not have enough time to thermalize and fully polarize itself in response to the qubit state being initialized in one well or the other (we used an initialization time of a few microseconds, whereas D-Wave, who often reports T_{eff} matching the fridge temperature, uses several hundred microseconds). A more quantitative study of the role of the environmental response time in this experiment is needed. We note that the effect of ε_p on tunneling becomes stronger for for multiqubit tunneling [19], since the net reorganization energy will grow and suppress tunneling more easily. This is again a primary reason why we want to build an annealer with low-dissipation qubits.

To conclude this chapter, we have used the fluxmon to measure flux noise over a range of frequencies about $2k_BT/h$, separately extracting the symmetric and antisymmetric components $S_\Phi^{\pm}(f)$ and observing the classical-quantum crossover. We find that

S_Φ^- displays a paramagnetic temperature dependence below the crossover, and that S_Φ^+ follows a $1/f$ power law whose magnitude is consistent with that of the $1/f$ flux noise near 1 Hz. The fact that the noise spectrum has a $1/f$ shape near the crossover indicates that the underlying magnetic fluctuators have a distribution of relaxation times that extends to at least 1 GHz, possibly hinting towards spin clustering as opposed to spin diffusion [176, 177] (see supplemental section 4.4), which would also be consistent with the correlated low-frequency inductance fluctuations observed in SQUIDs that were postulated to arise from fluctuations in spin cluster relaxation times [178]. Recent evidence [128] that adsorbed molecular O_2 (spin-1) may play a dominant role in flux noise could also support this conclusion, as spin-orbit induced magnetic anisotropy could break conservation of total spin and allow clusters to locally transfer energy and angular momentum to the lattice. Finally, we showed that the measured noise and dissipation can approximately predict incoherent quantum tunneling rates between flux qubit wells, which has direct implications for quantum annealing applications.

4.4 Further details on coherence and experimental checks

4.4.0.1 Consistent definition of $S_\Phi^+(f)$ at low and high frequencies

Because the low and high frequency flux noise are measured by very different methods, we must be careful to have consistent definitions of $S_\Phi^+(f)$ at low and high frequencies.

In other words, in Fig. 2(e) of the main text, we must be sure we are plotting the same physical quantity at low and high frequencies, without any discrepant factors of 2 or 2π. Such a discrepancy could for example affect the best-fit value of α in an interpolating power law between the two frequency ranges.

To infer $S_\Phi^+(f)$ at low frequencies, we measure the discrete time sequence Φ_n, where Φ_n is a classical real number, over N discrete time steps indexed by $n \in \{0, 1, ..., N-1\}$ and separated by the physical sampling interval δt (meaning total data acquisition time $T = (N-1)\delta t \approx N\delta t$). We then estimate the single-sided PSD by computing

$$S_\Phi^+(f) = \frac{2T}{N^2}\langle|\tilde{\Phi}_{k=fT}|^2\rangle, \tag{4.103}$$

where the DFT coefficients Φ_k are defined by

$$\tilde{\Phi}_k \equiv \sum_{n=0}^{N-1} \Phi_n e^{-i2\pi nk/N}. \tag{4.104}$$

and $\langle\cdot\rangle$ denotes an ensemble average since $\tilde{\Phi}_k$ is itself a random variable.

At high frequencies, we instead infer the flux noise spectral density through Fermi's golden rule [141] using the fact that $1/T_1 = \Gamma_\uparrow + \Gamma_\downarrow$, which implies the relation

$$\begin{aligned} S_\Phi^+(f) &\equiv S_\Phi(f) + S_\Phi(-f) \\ &= \frac{\hbar^2}{T_1}\frac{1}{|\langle 0|\frac{d\hat{H}}{d\Phi_t^x}|1\rangle|^2} = \frac{\hbar^2 L^2}{T_1}\frac{1}{|\langle 0|\hat{\Phi}|1\rangle|^2}, \end{aligned} \tag{4.105}$$

where [141]

$$S_\Phi(\omega) \equiv \int_{-\infty}^{\infty} d\tau e^{i\omega\tau}\langle\Phi(\tau)\Phi(0)\rangle. \tag{4.106}$$

Here, Φ could be an operator, but for simplicity we can assume it's a real number, since it is sufficient to check whether or not the two definitions (4.103) and (4.105) for S_Φ^+ coincide for a classical incoherent flux noise source acting on the qubit. Showing this will turn out to be equivalent to deriving the Wiener-Khinchin theorem for a stationary stochastic process.

First, we can write (4.104) in the continuum limit $N \to \infty, \delta t \to 0$ with T held constant, so that $\Phi(t = n\delta t) = \Phi_n$ and $\tilde{\Phi}_{k=ft} \to \frac{1}{\delta t}\int_0^T \Phi(t)e^{-i2\pi ft}dt$. Since $\Phi(t)$ is a random process, so is $\tilde{\Phi}(f)$, so we keep the $\langle\cdot\rangle$ before taking the limit $T \to \infty$. We can obtain its expectation value by taking the limit $T \to \infty$ after plugging the continuum limit expression into (4.103),

$$\begin{aligned} S_\Phi^+(\omega) &= \lim_{T\to\infty} \frac{2}{T}\left\langle \left| \int_0^T \Phi(t)e^{-i\omega t}dt \right|^2 \right\rangle \\ &= \lim_{T\to\infty} \frac{2}{T}\int_0^T \int_0^T dt\, dt'\, e^{i\omega t}e^{-i\omega t'}\langle\Phi(t)\Phi(t')\rangle \\ &= \lim_{T\to\infty} \frac{2}{T}\int_0^T \int_0^T dt\, dt'\, e^{-i\omega(t-t')}\langle\Phi(t-t')\Phi(0)\rangle, \end{aligned} \tag{4.107}$$

where in the last step we have assumed that $\Phi(t)$ is a stationary process. To continue, we note that the integrand [call it $f(t, t')$] inside the double integral is a function of $\tau \equiv t' - t$ alone, meaning that $f(\tau) = f(t - t')$ is constant along lines of constant τ defined by the equation $t' = t + \tau$ within the t-t' plane. We can therefore convert the double integral into a one-dimensional integral in $f(\tau)$, by integrating the diagonal "slices" formed by such lines across the two-dimensional domain of integration. The

domain of integration is the square $[0,T]\times[0,T]$ in the t-t' plane, which is covered by diagonal strips parameterized by τ ranging from $-T$ to T. The area of each diagonal strip corresponding to τ with infinitesimal width $d\tau$ is $\sqrt{2}(T-|\tau|)\frac{d\tau}{\sqrt{2}} = (T-|\tau|)d\tau$, so we can convert $\int_0^T\int_0^T dt\,dt'\,f(t-t')$ to $\int_{-T}^T d\tau\,f(\tau)$, meaning that (4.107) becomes

$$\begin{aligned} S_\Phi^+(\omega) &= \lim_{T\to\infty}\frac{2}{T}\int_{-T}^{T} d\tau\, e^{i\omega\tau}\langle\Phi(\tau)\Phi(0)\rangle(T-|\tau|) \\ &= \lim_{T\to\infty} 2\int_{-T}^{T} d\tau\, e^{i\omega\tau}\langle\Phi(-\tau)\Phi(0)\rangle\left(1-\frac{|\tau|}{T}\right) \\ &= 2\int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\langle\Phi(\tau)\Phi(0)\rangle, \end{aligned} \tag{4.108}$$

where in the last line we have used the property that the autocorrelation function $\langle x(t+\tau)x(t)\rangle$ is an even function of τ. Comparing this to (4.106) and (4.105) shows that the two definitions of S_Φ^+ are indeed equivalent.

4.4.0.2 Flux noise at high and low frequencies changes similarly between samples

Fig. 4.31 shows low and high frequency flux noise data for nominally the same qubit on two different chips, with the second chip seeing extra post-processing in the form of a downstream oxygen ash clean and the application of a (nominal) monolayer of perfluorodecyltrichlorosilane (FDTS) via molecular vapor deposition. The second sample also sat covered in photoresist for 6 months longer than the first sample. The flux noise power just below 1 GHz changes by a factor of 1.6, while the flux noise power around 1

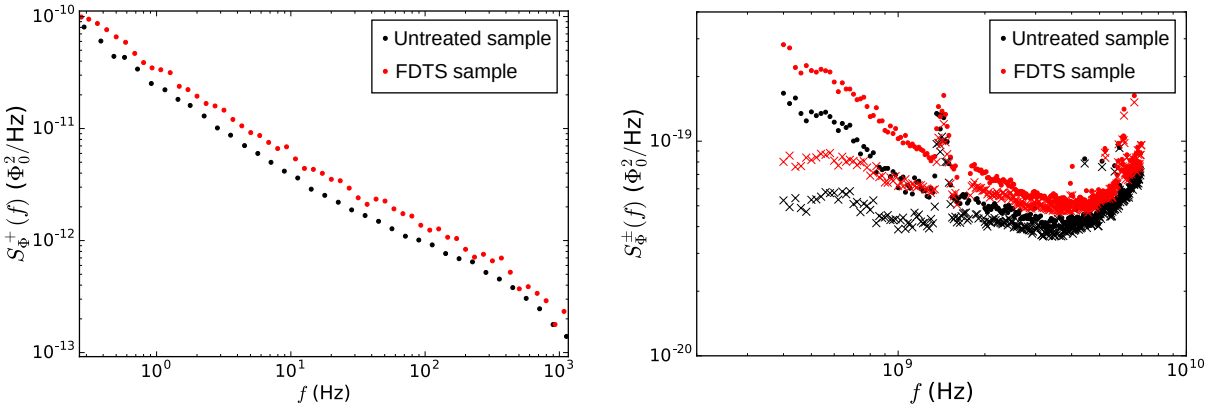


Figure 4.31: Comparison of flux noise at low and high frequencies between two different samples at base temperature. The samples were nominally identical apart from differing fabrication post-treatments.

Hz changes by a factor of approximately 1.5. Consistent with this we also observe that the Ramsey decay time $T_{\phi 2}$ away from zero tilt was $\sim 30\%$ lower on the FDTS sample when measured at a point with the same sensitivity of f_{10} to tilt flux (this degradation was reproducible at several bias points).

4.4.0.3 Checking for distortion of extracted $S^{\pm}_{\Phi_t^x}(f)$ from nonlinear crosstalk

In the main text, to deduce the high frequency flux noise $S^{\pm}_{\Phi_t^x}(f)$ in the main qubit loop we measure T_1 at zero tilt, because at degeneracy only flux noise in $\varepsilon(\Phi_t^x)$ and not noise in $\Delta(\Phi_{\rm SQ}^x)$ would induce transitions between qubit energy eigenstates. However, this is no longer strictly true if the two junctions in the DC SQUID are not perfectly symmetric, which is to be expected due to fabrication imperfections. It can be shown [44] that if the junction asymmetry is $d \equiv \frac{E_{J1}-E_{J2}}{E_{J1}+E_{J2}}$, any flux threading the DC SQUID loop will lead to an offset in tilt flux according to the nonlinear relation of Eq. (2.58), implying that noise

in $\Phi^x_{\rm SQ}$ leads to noise in Φ^x_t according to the differential transfer function

$$\frac{d\Phi^x_t}{d\Phi^x_{\rm SQ}} = \frac{d}{2}\frac{1}{d^2\sin^2(\pi\Phi^x_{\rm SQ}/\Phi_0)+\cos^2(\pi\Phi^x_{\rm SQ}/\Phi_0)}. \tag{4.109}$$

By measuring $\Delta\Phi^x_t$ as a function of $\Phi^x_{\rm SQ}$ and subtracting out any contribution from linear geometric crosstalk between barrier and tilt bias lines, we can extract this intrinsic nonlinear transfer function experimentally. For the data in the main text, we obtain $(\frac{d\Phi_t}{d\Phi_{\rm SQ}})^2 < .005$ over the range of $\Delta(\Phi^x_{\rm SQ})$ measured, corresponding to a junction asymmetry of $\sim 1\%$. This suggests that dissipation from incoherent flux noise is likely dominated by noise in Φ^x_t and not $\Phi^x_{\rm SQ}$, since the $1/f$ noise in $\Phi^x_{\rm SQ}$ should be less than or comparable to the noise in Φ^x_t (using conventional Ramsey experiments [115] at zero tilt we obtain the noise in the DC SQUID flux $\Phi^x_{\rm SQ}$ has magnitude $1-2\,\mu\Phi_0/\sqrt{\rm Hz}$ at 1 Hz, compared to $\sim 5\ \mu\Phi_0/\sqrt{\rm Hz}$ for Φ^x_t). However, this does not exclude the possibility of the relative strength of noise in Φ^x_t and $\Phi^x_{\rm SQ}$ changing greatly between 1 Hz and GHz, but given that we see similar T_1's over several samples with different junction asymmetries d, this seems not to be the case. Note that this analysis does not rule out noise from surface spins fluctuating on the wiring of the DC SQUID, because this wire is shared by both barrier and tilt loops; rather, it only implies that noise from such fluctuators would affect the T_1 data via induced noise in Φ^x_t and not due to induced noise in $\Phi^x_{\rm SQ}$.

4.4.0.4 Checking for dissipation from non-equilibrium quasiparticles

Quasiparticle dissipation has been observed in other superconducting qubit systems [68, 179]. The frequency dependence of the fluxmon T_1 at zero tilt below $\sim 3\,\mathrm{GHz}$ is also consistent with quasiparticle dissipation if one were to assume a large enough population of non-equilibrium quasiparticles in the system. We give theoretical and experimental arguments as to why this is unlikely to be a dominant effect in our system, including a test of the effects of magnetic vortices.

Quasiparticles with energy near the superconducting gap Δ can absorb energy from the qubit when they tunnel across one of the Josephson junctions, while "hot" quasiparticles with energies more than $\hbar\omega_{10}$ from the gap can excite the qubit. If the energies of all quasiparticles influencing the qubit are sufficiently less than 2Δ, then for an arbitrary quasiparticle occupation distribution $f(E)$, the decay and excitation rates induced on the qubit are given by [180]

$$\Gamma_{i\to f} = \sum_{j=1,2} \left| \langle f | \sin \frac{\hat{\varphi}_j}{2} | i \rangle \right|^2 S^j_{\mathrm{qp}}(\omega_{if}), \tag{4.110}$$

where j indexes the two Josephson junctions of the fluxmon and

$$S^j_{\mathrm{qp}}(\omega) = \frac{32 E_{Jj}}{\pi\hbar} \int_0^\infty dx\, \rho((1+x)\Delta)\rho((1+x)\Delta + \hbar\omega) \times f[(1+x)\Delta](1 - f[(1+x)\Delta + \hbar\omega]) \tag{4.111}$$

is the double-sided quasiparticle current spectral density, with $\rho(E) = E/\sqrt{E^2 - \Delta^2} \approx 1/\sqrt{2(E-\Delta)/\Delta}$ the normalized quasiparticle density of states. The formula (4.111)

works for $\omega < 0$ by simply swapping the arguments of f and replacing ω with $-\omega$. In a non-tunable gap flux qubit (i.e., a single-junction fluxmon), at zero tilt the junction would be biased at π, meaning that quasiparticle dissipation ($\propto |\langle 0| \sin \frac{\hat{\varphi}}{2} |1\rangle|^2$) would vanish (physically, this is due to destructive interference between electron-like and hole-like tunneling [181]). However, for the gap-tunable fluxmon, even though the effective dynamical phase $\hat{\varphi} = (\hat{\varphi}_1 + \hat{\varphi}_2)/2$ can be biased at π at zero tilt, the phase of the individual junctions are not. Instead, flux quantization dictates that $\hat{\varphi}_1 = \hat{\varphi} - \pi\Phi_{\mathrm{SQ}}/\Phi_0, \hat{\varphi}_2 = \hat{\varphi} + \pi\Phi_{\mathrm{SQ}}/\Phi_0$, meaning that the matrix element can be non-zero and quasiparticle dissipation can occur even at zero tilt.

While a thermal distribution of quasiparticles would be much too small to explain the observed dissipation, this does not rule out the possibility of non-equilibrium quasiparticles. A reasonable model for computing the distribution of non-equilibrium quasiparticles is outlined in Refs. [147, 68, 182]. Here, quasiparticles are assumed to be injected at some energy or range of energies well above the gap and, via electron-phonon scattering and recombination, relax to some steady-state distribution that is essentially independent of the injection energy as long as the injection energy is high enough. Using the steady state equations outlined in Ref. [68], we can calculate the expected distribution of nonequilibrium quasiparticles as a function of injection rate and phonon temperature T_{ph}, and then using (4.110) and (4.111) numerically calculate the resulting up and down transition rates to obtain the quasiparticle-induced T_1 and stray population. We choose an injection rate that leads to a quasiparticle density that best matches the measured T_1

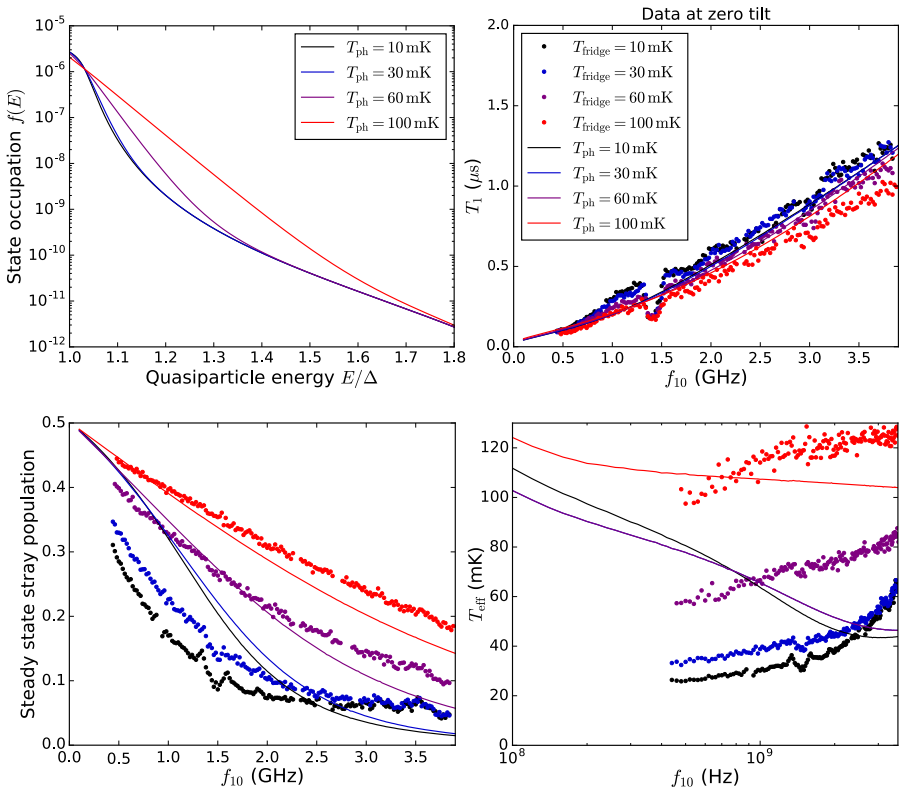


Figure 4.32: Data vs. fit to nonequilibrium quasiparticle theory. Although a quasiparticle density can be chosen large enough to roughly match T_1 vs. frequency, a simultaneous fit to stray population is not possible within this model.

data while choosing the phonon temperature to be equal to the fridge temperature. We assume the same injection rate for all four fridge temperatures used in the main paper. The injection energies were between 2.1Δ and 2.2Δ, though changing this energy range does not materially affect the result. The results are summarized in Fig. 4.32, where we plot the computed $f(E)$, T_1 and stray population/effective temperature induced on the qubit on top of the T_1 and stray population data used in Fig. 3 of the main paper. The best fit yields a quasiparticle density $n_{\rm qp} \equiv 2D(E_F)\int_\Delta^\infty \rho(E)f(E)\,dE$ of $3.5/\mu{\rm m}^3$, or a fractional quasiparticle density $x_{\rm qp} = n_{\rm qp}/n_{cp} = n_{\rm qp}/(D(E_F)\Delta) = 1.3\times10^{-6}$, where $D(E_F)$ is the density of electron states at the Fermi energy.

We make several observations about this calculation. While the fits do match the T_1 data reasonably well below $\sim 4\,\mathrm{GHz}$, the stray population is quite off and does not have nearly a large enough dependence on temperature at low frequencies. We also see that the quasiparticle-induced qubit effective temperature does not approach a constant value at low frequencies like the data appears to do, and like the stray populations, the predicted effective temperatures do not match the data well at all, especially for the lowest three temperatures. Furthermore, there are several experimentally-based estimates substantially below $x_{\mathrm{qp}} = 1.3 \times 10^{-6}$ in the superconducting qubit literature, for example 1×10^{-8}, 4×10^{-8}, and 3×10^{-7} in Refs. [183, 184, 185], and 2.2×10^{-7} in Ref. [186] (where the nonequilibrium quasiparticle density explained both T_1 and stray population), so it is not unreasonable that x_{qp} is similarly low in our system, especially given the stringent level of light-tight filtering used in our setup [95]. In addition, we note that if the non-equilibrium quasiparticle density was indeed $x_{\mathrm{qp}} = 1.3 \times 10^{-6}$, the T_1 limit imposed on our standard Xmon qubits would be $7\,\mu\mathrm{s}$ at $3\,\mathrm{GHz}$, but Xmons consistently achieve T_1 a factor of 10 higher than this at $3\,\mathrm{GHz}$, and furthermore show a defect structure and opposite overall scaling in frequency from that predicted from quasiparticles. We have no reason to believe the quasiparticle density would be any higher in the fluxmon device given that the device is fabricated with identical materials and chip mount, and the fridge wiring is essentially identical to that used for Xmon experiments. If the quasiparticles were being introduced by the long (few μs) readout pulses used for the fluxmon, we would expect T_1 to depend on the repetition time between experiments. We do not

see any change in T_1 as we vary the repetition time from $50\,\mu$s to $1000\,\mu$s.

One more experimental check we can perform is to see if T_1 improves after inducing magnetic vortices into the Al film of the sample. Abrikosov vortices, which have quasi-normal cores, form when a thin-film superconductor is cooled through T_c in the presence of a magnetic field. It is well-known that such vortices trap quasiparticles, and have even been shown to significantly decrease quasiparticle-induced dissipation in superconducting qubits [185]. In particular, for a thin-film Al transmon qubit vortices were shown to significantly decrease quasiparticle-induced dissipation with a modest applied field of $\sim 10\,$mG [185]. To check if there is any improvement in T_1 to be gained from the presence of vortices, we added a magnetic coil to the setup directly outside the qubit box, which we used to apply a several different magnetic fields between -30 and $30\,$mG to the qubit chip as it cooled through its superconducting transition. The result of these magnetic field cools are shown in Fig. 4.33. We see only degradation of qubit T_1 with applied magnetic field. This data suggests that quasiparticles are not playing a dominant role in qubit dissipation in the fluxmon device.

4.4.0.5 Low and high frequency noise cutoffs for Macroscopic Resonant Tunneling Rates

In the main text, it was argued that the low-frequency cutoff f_l in the noise integral for the MRT tunneling linewidth W should be the inverse of the total experimental data acquisition time, whereas the f_l for the integral for the reorganization energy ε_p should be the tunneling rate near maximum tunneling. The latter physically makes sense as the

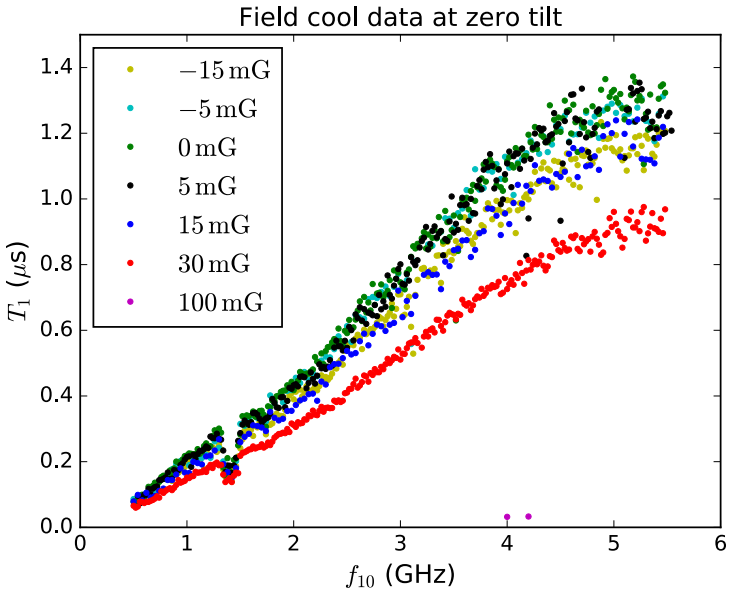


Figure 4.33: Field cool data. Inducing vortices only degrades T_1.

reorganization energy should not depend on the time at which the experiment was performed. In other words, only dissipation at frequencies higher than the slowest timescale of the tunneling can affect this energy. However, different instances of the tunneling experiment may have different quasistatic flux offsets $\delta\varepsilon = 2I_p\delta\Phi_t^x$ as discussed earlier in the context of the low frequency flux noise measurement, which can give additional broadening of the MRT tunneling peak. For simplicity, let us take the Gaussian approximation to the lineshape described in the main text, $\Gamma(\varepsilon) = \sqrt{\frac{\pi}{8}}\frac{\Delta^2}{\hbar W}\exp\left[-\frac{(\varepsilon-\varepsilon_p)^2}{2W^2}\right]$, so that what is actually measured after averaging is the quantity

$$\overline{\Gamma}(\varepsilon) = \int_{-\infty}^{\infty} d(\delta\varepsilon)\Gamma(\varepsilon+\delta\varepsilon)p(\delta\varepsilon), \tag{4.112}$$

where $p(\delta\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\delta\varepsilon)^2}{2\sigma^2}}$ describes the Gaussian distribution of quasistatic flux fluctuations. Performing the integration yields

$$\overline{\Gamma}(\varepsilon) = \frac{\pi}{8} \frac{\Delta^2}{\sqrt{W_0^2 + \sigma^2}} e^{-\frac{(\varepsilon - \varepsilon_p)^2}{2\left(W_0^2 + \sigma^2\right)}}, \tag{4.113}$$

which shows that the original W_0 (obtained from integrating down to f_l equal to the maximum tunneling rate) is broadened by the r.m.s. quasistatic flux fluctuations via addition in quadrature. Since $W^2 = 4I_p^2 \int_{f_l}^{f_h} df\, S_\Phi^+(f)$, this amounts to extending f_l down to the inverse total experimental averaging time.

For the high frequency cutoff to the tunneling rate integral, we use the oscillation frequency of the inverted potential barrier. Previous theoretical and experimental work in macroscopic quantum tunneling has shown this to be the physical high-frequency cutoff [174, 131, 187, 175]. Nevertheless, we find that the tunneling rate near resonance according to full integral in Eq. (3) of the main text is not materially affected by this high frequency cutoff.

4.4.0.6 $1/f^\alpha$ scaling near the crossover and high-frequency cutoff

Here we discuss the $1/f^\alpha$ form of the flux noise and its modifications due to a possible high frequency cutoff for the relaxation times of the magnetic fluctuators. We argue that while our data clearly shows that such a cutoff must be of order $k_B T/h$ or higher, a model where the cutoff is a few times $k_B T$ fits the temperature dependence data slightly better than one with a much higher cutoff.

To obtain a $1/f^\alpha$ scaling, we assume a single relaxation time τ_n for each fluctuator.

The dynamic susceptibility of a single such fluctuator is given by a standard Drude formula $\chi_n(\omega,T) = \frac{\chi_n(0,T)}{1+i\omega\tau_n}$, meaning $\chi_n''(\omega,T) = \chi_n(0,T)\frac{\omega\tau_n}{1+\omega^2\tau_n^2}$. For paramagnetic spins, $\chi_n(0,T) \propto 1/T$. In the limit of many fluctuators with different relaxation times, we can convert (4.86) into an integral over τ with an effective weight for each τ:

$$S_\Phi(\omega) = \hbar\left[1+\coth\left(\frac{\hbar\omega}{2k_BT}\right)\right]\int_{\tau_{\min}}^{\tau_{\max}}\rho(\tau)\frac{1}{T}\frac{\omega\tau}{1+\omega^2\tau^2}, \tag{4.114}$$

where $\tau_{\min}$ and $\tau_{\max}$ are lower and upper cutoffs for the relaxation times, and we have included a uniform $1/T$ factor in the integrand under the assumption that all the fluctuators are paramagnetic.[29] In the classical limit,[30]

$$S_\Phi^+(\omega \ll k_BT/\hbar) \propto k_B\int_{\tau_{\min}}^{\tau_{\max}} d\tau\rho(\tau)\frac{\tau}{1+\omega^2\tau^2}, \tag{4.115}$$

$$S_\Phi^-(\omega \ll k_BT/\hbar) \propto \frac{\hbar}{T}\int_{\tau_{\min}}^{\tau_{\max}} d\tau\rho(\tau)\frac{\omega\tau}{1+\omega^2\tau^2}. \tag{4.116}$$

As before, from this we can see that our experimental data below the classical-quantum crossover is consistent with an environment of magnetic fluctuators with a paramagnetic static susceptibility $\chi(0)$, under the assumption that $\rho(\tau)$ is independent of temperature.

If we postulate $\rho(\tau) \propto \frac{1}{\tau}$ between τ_1 and τ_2, then performing the integration leads to $S_\Phi^+(\omega) \propto 1/\omega$ assuming $\frac{1}{\tau_2} \ll \omega \ll \frac{1}{\tau_1}$ ($\omega_{\min} \ll \omega \ll \omega_{\max}$), which is the usual picture of $1/f$ noise in the classical limit. However, more precisely performing the full integration

[29] We have neglected a possible dependence of the static susceptibility $\chi(0,\tau,T)$ on τ because it should be very weak. For instance, if we consider superparamagnetic clusters with different values for the total spin S that fluctuate by tunneling through anisotropy barriers $U \propto S^2$, then for a given S, $\chi(0) \propto \frac{1}{3}S(S+1)$ while $\log\tau \propto U \propto S^2$. Thus $\chi(0,\tau,T)$ depends on τ only logarithmically and can be safely replaced by its average value.

[30] We note that in the fully classical limit $\hbar \to 0$, S_Φ^- will vanish while $\chi''(\omega)$ (the dissipation) does not need to vanish.

(4.114) without assuming anything about ω relative to T or $\tau_{\min/\max}$ yields for the full spectrum

$$S_\Phi(\omega) \propto \frac{\hbar}{T}\left[1+\coth\left(\frac{\hbar\omega}{2k_BT}\right)\right]\tan^{-1}(\omega\tau)\Big|_{\tau_{\min}}^{\tau_{\max}} \tag{4.117}$$

As long as $\tau_{\min}$ and $\tau_{\max}$ don't depend exponentially on T, the temperature dependence of S_Φ^- will again be given by that of the static susceptibility. Let us assume that $\omega_{\min} \ll k_BT/\hbar$ (justified by the presence of $1/f$ flux noise well below 1 GHz) and look at the shape of the classical-quantum crossover for different $\omega_{\max}$. Fig. 4.34 shows that there are three qualitative types of scaling behavior of $S_\Phi^\pm(f)$ around the crossover. If $\omega_{\max} \ll k_BT/\hbar$, then $S_\Phi^+(f)$ would approach the crossover from below as $1/f^2$, inconsistent with the data. On the other hand, if $\omega_{\max} \gg k_BT/\hbar$, then $S_\Phi^+(f)$ will turn into white noise just above the crossover. There is also an intermediate regime $\omega_{\max} \approx 3k_BT/\hbar$ where $S_\Phi^+(f)$ is very close to $1/f$ for all frequencies except for a slight deviation at the crossover point.

In Fig. 4.35, we compare the phenomenological thermodynamic power law model used in the main text,

$$S_\Phi^{\text{phen.}}(\omega) = A\frac{\omega}{|\omega|^\alpha}[1+\coth(\hbar\omega/[2k_BT_A])] \tag{4.118}$$

$$+ B\omega|\omega|^{\gamma-1}[1+\coth(\hbar\omega/[2k_BT_B])], \tag{4.119}$$

to the finite high-frequency cutoff model (4.117). The former implicitly assumes that $\omega_{\max} \gg k_BT/\hbar$, so that any deviation of S_Φ^+ from a perfect $1/f$ scaling near the classical-quantum crossover would be due to temperature alone. Although this seems to fit our

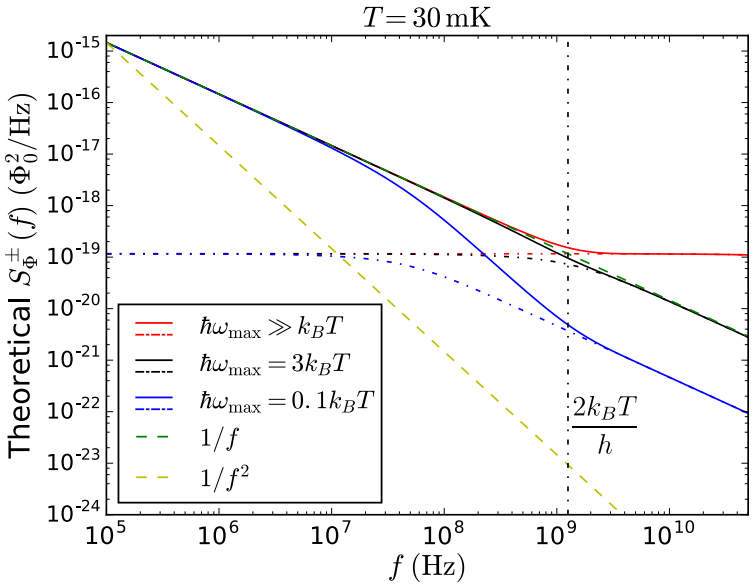


Figure 4.34: $1/f$ noise according to the finite high frequency cutoff model (4.117) for different values of $\omega_{\max}$, showing the three qualitative 'flavors' the noise scaling near the classical-quantum crossover. Dash-dotted lines are S_Φ^- while solid lines are S_Φ^+. The 'smoothest' transition through the cutoff where $S_\Phi^+(f)$ remains close to $1/f$ for all frequencies is achieved for $\omega_{\max} \approx 3k_B T/\hbar$.

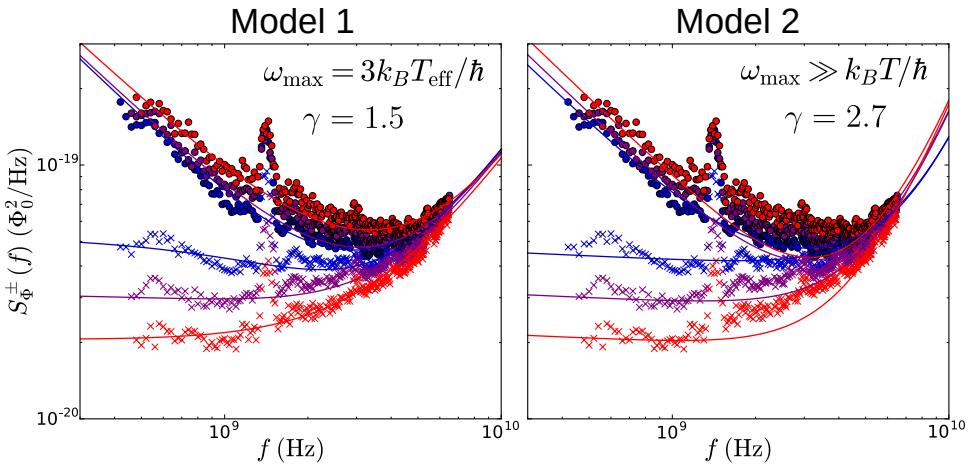


Figure 4.35: Comparison of two models with the constraint that the ohmic exponent γ be the same at all temperatures. The model with a finite, temperature-dependent $\omega_c \approx 3k_B T_{\mathrm{eff}}$ fits the temperature dependence data somewhat better than the phenomenological model with $\alpha = 1.05$ used in the main text at base temperature. Here, T_{eff} is the temperature deduced from the data below the crossover. However, there is no significant difference between the fits just looking at 30 mK alone.

data at base temperature, we find that incorporating a finite $\omega_{\max}$ fits the higher temperature data slightly better under the conditions of the fit (constant α and γ for all temperatures). Namely, we use the "intermediate" type of crossover, $\omega_{\max} = 3k_B T_{\text{eff}}$, where T_{eff} is the limiting effective temperature deduced from the stray population data below the crossover. A high frequency cutoff that scales linearly with temperature might have a natural physical meaning. For instance, if the inverse relaxation times τ_n^{-1} are determined by spin-phonon interactions, both τ_n^{-1} and $\omega_{\max}$ will be proportional to the number of phonons with energy close to the typical Zeeman splittings of the fluctuators, which scales linearly with T (since the Zeeman splittings should be $\ll k_B T$ even for clusters).

4.4.0.7 Implications of high frequency cutoff for spin diffusion

We mention one more mechanism that would give a frequency-dependent α due to a finite $\omega_{\max}$, but with a different functional form. Namely, we consider spin diffusion, which was proposed by Faoro and Ioffe [176] to explain $1/f$ noise in SQUIDs and further explored in the context of D-Wave flux qubits by Lanting *et al.* [188]. We conclude that given the $1/f$ scaling of $S_\Phi^+(f)$ near 1 GHz, spin diffusion is unlikely to be the source of the $1/f$ noise near the classical-quantum crossover unless i.) the spin diffusion constant is several orders of magnitude higher than estimates in the literature [176, 188] or ii.) the spin density is substantially inhomogeneous, leading to a shallower power law at high frequencies, in which case a separate physical mechanism needs to be invoked for the $1/f$ power law at lower frequencies.

Within the spin diffusion model, the total spin is conserved and spin excitations will diffuse around the surface of the superconducting qubit metal, generating flux noise by coupling to a non-uniform distribution of the magnetic field. In Ref. [176], the mechanism of diffusion was proposed to be a Ruderman-Kittel-Kasuya-Yosida(RKKY) interaction [189] mediated through the superconductor. The diffusion equation for the coarse-grained magnetization is

$$\partial M_\alpha(\boldsymbol{r},t)/\partial t = \mathcal{D}\nabla^2 M_\alpha(\boldsymbol{r},t), \tag{4.120}$$

where $\mathcal{D}$ is the diffusion coefficient and $\alpha = x, y, z$. Eq. (4.120) can be solved using the Laplace transform $M_\alpha(\boldsymbol{r}) \propto e^{-\Gamma_n t}\varphi_n(\boldsymbol{r})$, which leads to the eigenvalue problem

$$\nabla^2\varphi_n(\boldsymbol{r}) = -\Gamma_n\varphi_n(\boldsymbol{r}), \tag{4.121}$$

with periodic boundary conditions on the surface of the qubit metal. Solving for the eigenmodes $\varphi_n(\boldsymbol{r})$ allows one to express the dynamic magnetic susceptibility from the Green's function for Eq. (4.120):

$$\chi''_{\alpha\beta}(\boldsymbol{r},\boldsymbol{r}',\omega) = \delta_{\alpha\beta}\chi(0,T)\sum_n \varphi_n(\boldsymbol{r})\frac{\omega\Gamma_n}{\omega^2+\Gamma_n^2}\varphi_n(\boldsymbol{r}'). \tag{4.122}$$

The flux noise then becomes a sum of Lorentzians corresponding to each diffusion mode

$$S_\Phi(\omega) = \hbar\omega\chi(0,T)\left[1+\coth\left(\frac{\hbar\omega}{2k_BT}\right)\right]\sum_n b_n^2\frac{\Gamma_n}{\omega^2+\Gamma_n^2}, \tag{4.123}$$

where $b_n^2 = \sum_\alpha b_{\alpha n}^2$, $b_{\alpha n} = \int d\boldsymbol{r} b_\alpha(\boldsymbol{r})\varphi_n(\boldsymbol{r})$, are coupling factors describing how the local moment couples to the qubit's magnetic field and by reciprocity the qubit loop itself. Therefore the spin diffusion model naturally leads to a broad set of relaxation times $\tau_n = 1/\Gamma_n$ with a temperature-independent distribution function $\rho(\tau_n) = b_n^2$ given by the

form factors. The temperature dependence of S_Φ^- will then arise solely from that of the static magnetic susceptibility $\chi(0,T)$.

In the limit of a flat wire (thin film), one would expect a $1/f^{\alpha=1}$ power law for $\omega_{\min} \ll \omega \ll \omega_{\max}, k_B T/\hbar$. But allowing for a finite aspect ratio or for inhomogeneity of the spin density on the surface, one can have an exponent $\alpha \neq 1$. In particular, if the spins are concentrated near edges we can have $\alpha < 1$ [127]. An analytic approximation for the noise summation is [127]

$$S_\Phi(\omega) = A\left[1 + \coth\left(\frac{\hbar\omega}{2k_BT}\right)\right]\frac{\hbar\omega}{k_BT}\int_0^\infty dx\,\frac{x^{3-2\alpha}e^{-\frac{x}{\sqrt{\omega_{\max}}}}}{\omega^2+x^4}. \tag{4.124}$$

One would expect the high frequency cutoff $\omega_{\max}$ for Γ_n to be given by $\omega_{\max} \approx \mathcal{D}/\ell^2$, where ℓ is the smallest dimension associated with the qubit geometry, which for our device should be the metal thickness of $\sim 100\,\text{nm}$. The highest estimates in the literature for $\mathcal{D}$ are $10^8 - 10^9\,\text{nm}^2/\text{s}$ [176], which would imply a physically expected cutoff of $\omega_{\max}/(2\pi) \sim 100\,\text{kHz}$.[31] Therefore, for spin diffusion to be relevant to the classical-quantum crossover, we would need either a much larger $\mathcal{D}$ and/or $\alpha < 1$. If $\alpha < 1$, this would mean the quasistatic flux noise (which shows $\alpha = 1$ below 10 Hz) is not spin diffusion noise, but it may be possible to have different physical mechanisms in different frequency ranges with the noise power still scaling similarly between samples at high and low frequencies.

[31] Conversely, a low-frequency cutoff of $\sim 10\,\text{Hz}$ would be expected, which would mean the noise below 1 Hz would need to have a different physical mechanism, unless the diffusion constant is much lower than expected. The "bump" in flux noise observed at intermediate frequencies in the main paper could potentially be due to spin diffusion, assuming it is not suppressed by spin relaxation.

Chapter 5

Tunable coupling for quantum annealing: theory and design

In this chapter we study the theory and design of tunable inductive coupling between fluxmon qubits. We start with a simple classical linear circuits understanding of coupling, where a "coupler" circuit containing a Josephson junction is treated as a flux transformer loop that provides an effective tunable mutual inductance between the qubits. This provides a simple, intuitive picture of coupling and other local field effects that is often but not always sufficient to describe a system of inductively coupled qubits. We will then refine our analysis to account for both nonlinearity and capacitance in the coupler circuit using the Born-Oppenheimer approximation following the analysis of Ref. [55]. We will see that there are nonlinear and quantum effects arising from the coupler circuit that can become non-negligible for large coupling strengths and for non-zero characteristic

impedance of the coupler. We will furthermore provide a semi-intuitive understanding of when such nonlinear effects are important using an effective finite difference picture for qubit-qubit coupling. The nonlinear treatment will allow us to model ultra-strong coupling between qubits (a regime where the coupling strength is an appreciable fraction of the qubit frequency itself), which could be useful for quantum annealing, and which we will experimentally measure in the following chapter.

Tunable inductive coupling between superconducting qubits is not a new idea.[1] Tunable inductive coupling circuits have been implemented before between flux qubits [191, 192, 193], phase qubits [194], between qubits and resonators [195], and recently between gmon transmon qubits [196]. The essence of our approach to implementing tunable coupling is quite similar to the approach taken by D-Wave [197], using an rf-SQUID coupler as a tunable mutual inductance, but with some technical differences in physical implementation, and also differences in control. For example, we will implement arbitrary time-dependent control of individual couplers, a feature that is not currently available on D-Wave devices. Furthermore, we will go beyond the analysis previously used for these couplers, ultimately using a more accurate nonlinear Born-Oppenheimer analysis as opposed to a linear Born-Oppenheimer analysis [198]. With this we hope to be able to operate multiple qubits in the nonlinear regime at the same time, enabling stronger coupling and more accurate modelling and control of the annealing process without having to perform an unscalable full diagonalization of the system.

[1] Proposals for tunable capacitive coupling also exist [190], but tunable inductive coupling is much simpler to implement in practice, and we don't want to deal with charge noise.

5.1 Inductive coupling from a linear circuits perspective

In this section we will approach tunable inductive coupling from a linear circuits perspective, which provides the simplest way to understand and calibrate coupled qubit devices. This means that we consider the coupling circuit as a tunable linear mutual inductance, even though in reality it contains nonlinearity due to its Josephson junctions and also has finite capacitance. We will still allow for nonlinearity within the qubits, however. This kind of linear treatment is sufficiently accurate when the couplings are not too large and when the nonlinearity of the coupler circuit is not too high. For now we also ignore any capacitance within the coupler's Josephson junction or main body of the coupler. Later we will allow for finite capacitance and study the resulting quantum and dispersive effects.

5.1.1 Direct coupling through a mutual inductance

Before delving into the Josephson coupler circuit, it is instructive to first consider a system of two qubits coupled through a direct geometric mutual inductance, as illustrated in Fig. 5.1. Intuitively, the interaction energy of this circuit is simply $MI_1I_2 \approx M\frac{\Phi_1}{L_1}\frac{\Phi_2}{L_2}$, giving rise to the simple interaction Hamiltonian $M\hat{I}_1\hat{I}_2 \approx M\frac{\hat{\Phi}_1}{L_1}\frac{\hat{\Phi}_2}{L_2}$. We will see that this intuition is correct in the limit of weak coupling. We'll now perform a more rigorous analysis to obtain the Hamiltonian in terms of canonical flux and charge variables, and see

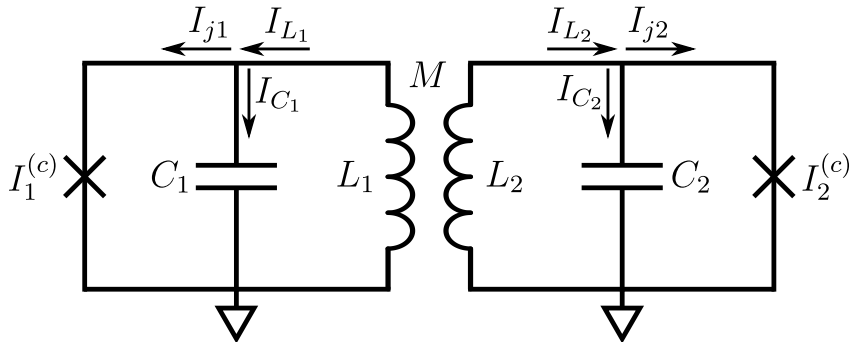


Figure 5.1: Circuit diagram and current conventions for two flux qubits coupled through a direct geometric mutual inductance.

that the above Hamiltonian is only approximate up to some renormalization terms that are negligible when $M/L \ll 1$. There will also turn out to be a similar renormalization in the Hamiltonians of each of the individual qubits as well.

We begin by writing down the classical current and flux equations,

$$\text{current conservation} \begin{cases} I_{L_1} - I_{C_1} - I_{j1} = 0 \\ I_{L_2} - I_{C_2} - I_{j2} = 0 \end{cases}, \tag{5.1}$$

$$\text{flux relations} \begin{cases} \Phi_{L_1} = L_1 I_{L_1} + M I_{L_2} \\ \Phi_{L_2} = L_2 I_{L_2} + M I_{L_1} \end{cases}, \tag{5.2}$$

$$\text{flux relations (inverted)} \begin{cases} I_{L_1} = \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left(\frac{\Phi_{L_1}}{L_1} - \frac{M}{L_1 L_2}\Phi_{L_2}\right) \\ I_{L_2} = \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left(\frac{\Phi_{L_2}}{L_2} - \frac{M}{L_1 L_2}\Phi_{L_1}\right) \end{cases}, \tag{5.3}$$

$$\text{flux quantization} \begin{cases} \Phi_{L_1} + \Phi_{j1} = \Phi_{t1} \\ \Phi_{L_2} + \Phi_{j2} = \Phi_{t2} \end{cases}, \tag{5.4}$$

where the last set of equations distinguishes the fluxes across the inductor and Josephson junction of each qubit according to that qubit's external tilt flux bias. Using the relations

$I = \dot{Q}$ and $\Phi = \int V' dt'$ along with the DC Josephson relation, we can rewrite the current conservation equations (5.1) as

$$\begin{aligned} I_{L_1} - C_1 \ddot{\Phi}_{j1} - I_1^{(c)} \sin(2\pi \Phi_{j1}/\Phi_0) = 0, \\ I_{L_2} - C_2 \ddot{\Phi}_{j2} - I_2^{(c)} \sin(2\pi \Phi_{j2}/\Phi_0) = 0. \end{aligned} \tag{5.5}$$

Substituting the inverted flux relations (5.3) and the flux quantization conditions (5.4) lets us write everything in terms of the fluxes Φ_{j1} and Φ_{j2} across the Josephson junctions (which from now on we denote simply as Φ_1 and Φ_2),

$$\begin{aligned} \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left(\frac{\Phi_{t1} - \Phi_1}{L_1} - \frac{M}{L_1 L_2}(\Phi_{t2} - \Phi_2)\right) - C_1 \ddot{\Phi}_1 - I_1^{(c)} \sin(2\pi \Phi_1/\Phi_0) = 0, \\ \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left(\frac{\Phi_{t2} - \Phi_2}{L_2} - \frac{M}{L_1 L_2}(\Phi_{t1} - \Phi_1)\right) - C_2 \ddot{\Phi}_2 - I_2^{(c)} \sin(2\pi \Phi_2/\Phi_0) = 0. \end{aligned} \tag{5.6}$$

Upon inspection we see that these are the Euler-Lagrange equations corresponding to the Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{coupled}} =& \frac{1}{2} C_1 \dot{\Phi}_1^2 + \frac{1}{2} C_2 \dot{\Phi}_2^2 - \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left[\frac{(\Phi_1 - \Phi_{t1})^2}{2L_1} + \frac{(\Phi_2 - \Phi_{t2})^2}{2L_2}\right] \\ &+ \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \frac{M}{L_1 L_2} (\Phi_1 - \Phi_{t1})(\Phi_2 - \Phi_{t2}) \\ &+ \frac{\Phi_0}{2\pi} I_1^{(c)} \cos(2\pi \Phi_1/\Phi_0) + \frac{\Phi_0}{2\pi} I_2^{(c)} \cos(2\pi \Phi_2/\Phi_0). \end{aligned} \tag{5.7}$$

Using the junction fluxes Φ_i as our generalized coordinates, we see that the usual flux and charge across the junction are the canonical variables; i.e., the canonical momenta

are still the charges

$$p_i \equiv \frac{\partial \mathcal{L}}{\partial \dot{\Phi}_i} = C_i \dot{\Phi}_i = Q_i. \tag{5.8}$$

The Hamiltonian $H = \sum_i \dot{\Phi}_i Q_i - \mathcal{L}$ of two qubits coupled through a mutual inductance is then

$$\begin{aligned} H_{\text{coupled}} =& \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} \\ &+ \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} \left[\frac{(\Phi_1 - \Phi_{t1})^2}{2L_1} + \frac{(\Phi_2 - \Phi_{t2})^2}{2L_2}\right] \\ &- \left(1 - \frac{M^2}{L_1 L_2}\right)^{-1} M \frac{(\Phi_1 - \Phi_{t1})}{L_1} \frac{(\Phi_2 - \Phi_{t2})}{L_2} \\ &- E_{J1} \cos(2\pi \Phi_1/\Phi_0) - E_{J2} \cos(2\pi \Phi_2/\Phi_0). \end{aligned} \tag{5.9}$$

We note that the second two lines of the Hamiltonian could have been derived less rigorously by substitution of the inverted flux relation into the classical formula for the energy stored in two mutually coupled inductances, $U = \frac{1}{2} L_1 I_1^2 + \frac{1}{2} L_2 I_2^2 - M I_1 I_2$, although the derivation of this formula along with the change to flux variables would have required about the same amount of work.

We can immediately read off several features of the coupled Hamiltonian (5.9). The first and most obvious is the introduction of an interaction term between qubits, which at zero tilt is proportional to $\Phi_1 \Phi_2$ as expected. The coefficient in front of the interaction term is negative, which is just an artifact of how we chose our current and flux conventions. Other equally valid definitions with reversed currents through one of the inductor will yield a term of the same magnitude but opposite sign (also keep in mind

that M can be negative, depending on the physical implementation of the coupling segment). However, the magnitude of the interaction term differs from $M\frac{\Phi_1}{L_1}\frac{\Phi_2}{L_2}$ by a factor of $\left(1-\frac{M^2}{L_1L_2}\right)^{-1}$, which is a result of the fact that $\Phi_i \neq \frac{I_i}{L_i}$ in the coupled system. In addition to the interaction term at zero tilt, there are two "local field" effects. The first is that the effective inductance of each qubit has been rescaled by a factor of $\left(1-\frac{M^2}{L_1L_2}\right)$, though this effect is second order in M. In the next section where we consider each qubit coupled to a coupler circuit, we will see that there can actually be large positive induced inductance shifts due to the possibility of the coupler having a negative net inductance. Note that we can equally well write the Hamiltonian in terms of the fluxes across the qubit inductors rather than the junctions via the canonical transformation $\Phi_i \rightarrow \Phi_i - \Phi_{ti}$, which gets rid of the "cross terms,"

$$\begin{aligned} H_{\text{coupled}} =& \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} \\ &+ \left(1-\frac{M^2}{L_1L_2}\right)^{-1}\left[\frac{\Phi_1^2}{2L_1}+\frac{\Phi_2^2}{2L_2}\right] \\ &- \left(1-\frac{M^2}{L_1L_2}\right)^{-1} M\frac{\Phi_1}{L_1}\frac{\Phi_2}{L_2} \\ &- E_{J1}\cos(2\pi[\Phi_1+\Phi_{t1}]/\Phi_0) - E_{J2}\cos(2\pi[\Phi_2+\Phi_{t2}]/\Phi_0). \end{aligned} \tag{5.10}$$

When $M^2/(L_1L_2) \ll 1$ (the largest coupling we will need in experiment corresponds to $M_{\text{eff}} \approx 30\,\text{pH}$ between qubits, corresponding to $M_{\text{eff}}^2/(L_1L_2) \approx 0.002$), we can safely say that within the linear approximation, the effect of coupling through a mutual inductance

M is to introduce the interaction Hamiltonian

$$H_{\text{int}} \approx M \frac{\hat{\Phi}_1}{L_1} \frac{\hat{\Phi}_2}{L_2}, \tag{5.11}$$

or in the language of quantum annealing means a ZZ coupling (in the flux basis) strength of

$$\begin{aligned} J &\approx \frac{M}{L_1 L_2} |\langle 0|\hat{\Phi}_1|1\rangle| |\langle 0|\hat{\Phi}_2|1\rangle| \\ &= M I_{p1} I_{p2}. \end{aligned} \tag{5.12}$$

5.1.2 Josephson coupler circuit as a tunable mutual inductance

We now consider how to make a tunable mutual inductance by using an intermediate Josephson coupler loop. Physically, the coupler is actually another flux qubit, but with smaller inductance and capacitance, and operated at a much higher characteristic frequency so that it always remains in its ground state. This is known as the Born-Oppenheimer (BO) approximation. For now, we will treat the coupler's Josephson junction as a tunable linear inductance, valid in the linear limit, and ignore its capacitance so that the coupler plays no dynamical role other than being a linear flux transformer between qubits. Later on in the next section we will revisit the validity of these assumptions, and include modifications both due to coupler nonlinearity and coupler capacitance (including the fact that the coupler might not remain in its ground state).

We consider the circuit in Fig. 5.2, showing two qubits coupled through a coupler

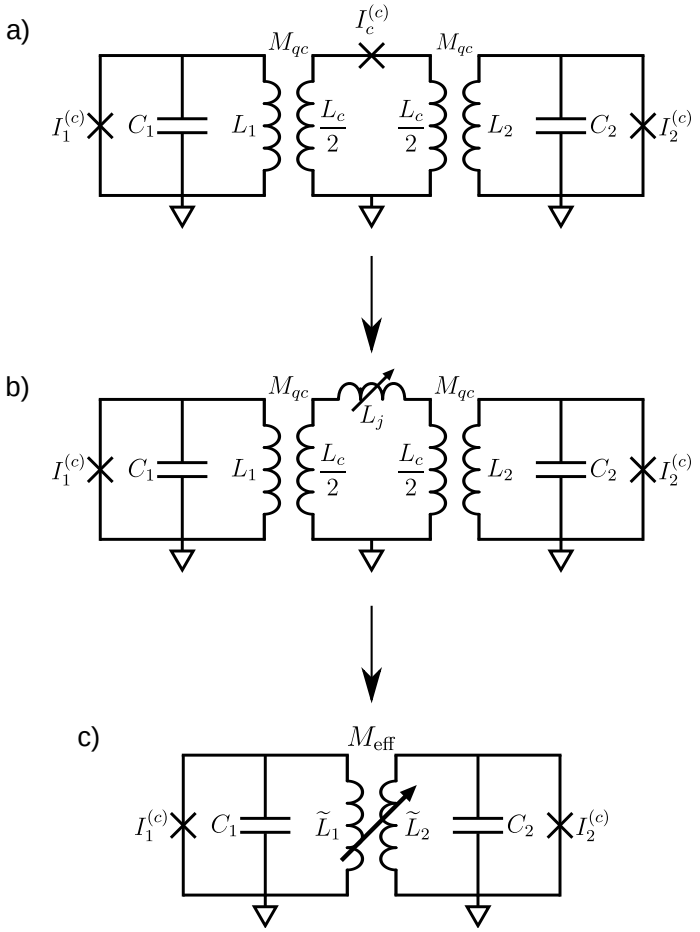


Figure 5.2: Linear circuit representation of two flux qubits coupled via a coupler, which is a flux transformer loop containing a Josephson junction (or DC SQUID). In the linear approximation the coupler's junction acts as a tunable linear inductance, leading to a tunable effective mutual inductance between qubits mediated by the coupler loop.

circuit. We treat the coupler's junction as a tunable linear inductor [Eq. (2.4)]

$$L_j(\varphi_{cj}) = \frac{\Phi_0}{2\pi} \frac{1}{I_c^{(c)} \cos(\varphi_{cj})}, \tag{5.13}$$

where φ_{cj} is the flux across the coupler's junction, so that the coupler acts as a flux transformer loop with tunable inductance. To find L_j as a function of the applied coupler flux bias φ_{cx}, we need to solve for φ_{cj} as a function of φ_{cx}. To do this, we first write down the current-flux relationships

$$\Phi_{L,c} = L_c I_c + \sum_k M_k I_k \tag{5.14}$$

and

$$\Phi_{L,k} = L_k I_k + M_k I_c, \tag{5.15}$$

where I_c is the current flowing through the coupler loop and I_k are the currents flowing through the qubit inductors. The third relation we need is flux quantization over the coupler's loop, which dictates that

$$\Phi_{L,c} + \Phi_{cj} = \Phi_{cx}. \tag{5.16}$$

Solving for I_k in (5.15) yields

$$I_k = \frac{\Phi_{L,k}}{L_k} - \alpha_k I_c, \tag{5.17}$$

where

$$\alpha_k = \frac{M_k}{L_k} \tag{5.18}$$

are the inductive coupling efficiencies of the qubits. Substituting this expression for the

qubit currents back into (5.14) yields the flux through the coupler's inductor,

$$\Phi_{L,c} = (L_c - \alpha_k M_k) I_c + \sum_k \alpha_k \Phi_{L,k}, \tag{5.19}$$

which can be rewritten as

$$\varphi_{L,c} = -\beta_c \frac{I_c}{I_c^{(c)}} + \sum_k \alpha_k \varphi_k, \tag{5.20}$$

where

$$\begin{aligned} \beta_c &\equiv -\frac{2\pi}{\Phi_0} I_c^{(c)} \left(L_c - \sum_k \alpha_k M_k \right) \\ &\equiv -\frac{2\pi}{\Phi_0} I_c^{(c)} \tilde{L}_c \end{aligned} \tag{5.21}$$

is the effective nonlinearity of the coupler and $\tilde{L}_c$ is a renormalized coupler inductance (note that in our notation, β_c is positive when the Josephson inductance [and hence the critical current] is negative). Next, the classical DC Josephson relation says that we must simultaneously have $I_c = I_c^{(c)} \sin \varphi_{cj} = I_c^{(c)} \sin(\varphi_{cx} - \varphi_{L,c})$, where we have used the coupler's flux quantization condition (5.16). Substituting (5.20) and dividing through by $I_c^{(c)}$ yields a transcendental equation for the coupler current as a function of φ_{cx},

$$\frac{I_c}{I_c^{(c)}} = \sin \left(\varphi_{cx} - \sum_k \alpha_k \varphi_k + \beta_c \frac{I_c}{I_c^{(c)}} \right). \tag{5.22}$$

Equating the argument of the sine on the right hand side with φ_{cj} yields a transcendental equation relating φ_{cj} and φ_{cx},

$$\varphi_{cj} - \beta_c \sin \varphi_{cj} = \varphi_x, \tag{5.23}$$

where

$$\varphi_x \equiv \varphi_{cx} - \alpha_1\varphi_1 - \alpha_2\varphi_2 \tag{5.24}$$

is the total flux bias felt by the coupler including the influence of the qubits. We formally solve this transcendental equation[2] by introducing the function $\sin_\beta(x)$ such that

$$\sin_{\beta_c}(\varphi_x) \equiv \sin(\varphi_{cj}), \tag{5.25}$$

which along with (5.23) implies the relations

$$\sin_{\beta_c}(\varphi_x) = \sin(\varphi_x + \beta_c \sin_{\beta_c}(\varphi_x)) \tag{5.26}$$

and

$$\varphi_{cj} = \varphi_x + \beta_c \sin_{\beta_c}(\varphi_x). \tag{5.27}$$

Physically, the motivation for this notation is that the current through the Josephson junction is then given by a function of the applied bias that looks almost like the Josephson relation,

$$I_c = I_c^{(c)} \sin_{\beta_c}(\varphi_x) = I_c^{(c)} \sin(\varphi_{cj}), \tag{5.28}$$

and in fact reduces to the simple Josephson relation with $\varphi_{cx} = \varphi_{cj}$ in the $\beta_c \to 0$ limit of a loop with just a Josephson junction. An analytical expression for $\sin_\beta(\varphi)$ is derived in Ref. [55],

$$\sin_\beta(\varphi) = 2\sum_{k\geq 1} \frac{J_k(k\beta)}{k\beta} \sin(k\varphi), \tag{5.29}$$

[2]Note that in Ref. [55] (the work on which we base our nonlinear analysis in the following sections), some sign conventions are different, with $\varphi_{cj} \to \varphi_{cj} + \pi$, $\varphi_{cx} \to \varphi_{cx} + \pi$, $\beta_c \to -\beta_c$. Our convention is "physical" in that it doesn't include a π-shift, so that at zero physical applied flux in the coupler we have negative β_c, which matches our convention used for qubits. However, note that equation (5.23) as well as equations (5.25) and (5.27) are invariant under this change of convention. A useful identity for convincing oneself of this is $\sin_\beta(\varphi) = -\sin_{-\beta}(\varphi + \pi)$.

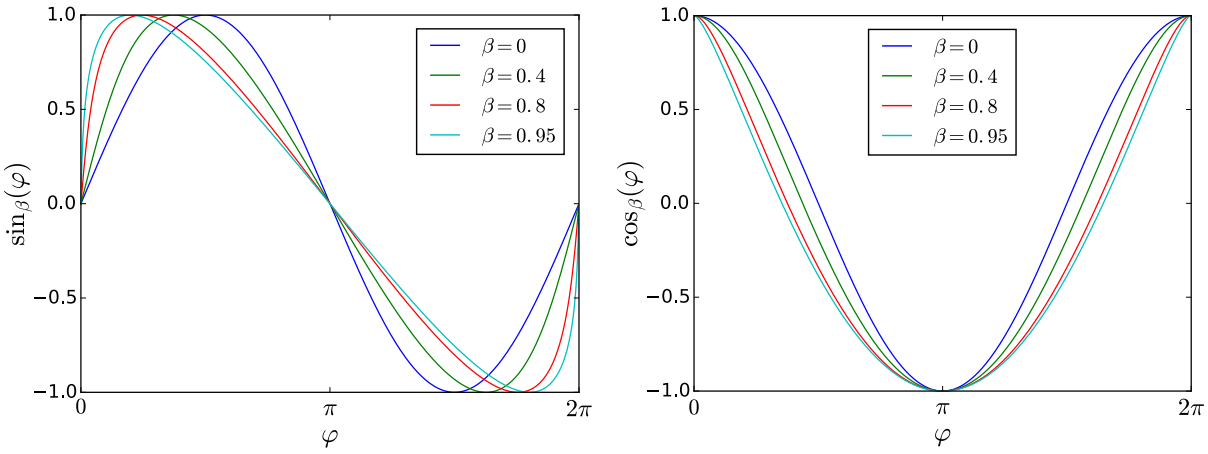


Figure 5.3: Plots of the $\sin_\beta(\varphi)$ and $\cos_\beta(\varphi)$ functions for different values of β.

where $J_k(x)$ is the Bessel function of the first kind, but usually a numerical solution is more practical, especially in the highly nonlinear limit $\beta \approx 1$. The $\sin_\beta$ function is illustrated in Fig. 5.3(a), and its integral $\cos_\beta$ in Fig. 5.3(b). We note that $\sin'_\beta(0) = \frac{1}{1-\beta}$, so the coupler will become more and more sensitive as $\beta_c \to 1$. However, although this first derivative and therefore the first-order coupler susceptibility increases without bound in this limit, as shown in the next section, going beyond first order will keep all quantities finite, including M_{eff}.

Next, consider what happens when a current I_{q1} flows through the inductor of one of the qubits, say qubit 1. This will induce a flux $\Delta\Phi_{cx} = -M_{qc}I_{q1}$ in the coupler loop. According to Eq. (5.28), this will cause the current I_c flowing in the coupler to change to first order by the amount

$$\Delta I_c = \frac{\partial I_c}{\partial \Phi_{cx}}\Delta\Phi_{cx} = -\frac{2\pi}{\Phi_0}I_c^{(c)}\frac{\cos(\varphi_{cx} + \beta_c \sin_{\beta_c}(\varphi_{cx}))}{1 - \beta_c \cos(\varphi_{cx} + \beta_c \sin_{\beta_c}(\varphi_{cx}))}M_{qc}I_{q1}, \tag{5.30}$$

where we have used the identity

$$\partial_{\varphi_{cx}} \sin_{\beta_c}(\varphi_{cx}) = \frac{\cos(\varphi_{cx} + \beta_c \sin_{\beta_c}(\varphi_{cx}))}{1 - \beta_c \cos(\varphi_{cx} + \beta_c \sin_{\beta_c}(\varphi_{cx}))}, \tag{5.31}$$

which can be obtained by differentiating Eq. (5.26). This change in current will then be felt by qubit 2 through its mutual inductance to the coupler according to $\Delta\Phi_{q2} = M_{qc}\Delta I_c$. It follows that we have an effective mutual inductance between qubits 1 and 2 of

$$M_{\text{eff}}(\varphi_{cx}) = \frac{\Delta\Phi_{q2}}{I_{q1}} = \frac{M_{qc}^2}{\tilde{L}_c}\frac{1}{1-\frac{1}{\beta_c\cos(\varphi_{cx}+\beta_c\sin_{\beta_c}(\varphi_{cx}))}}, \tag{5.32}$$

where we have rearranged some terms and used the definition of β_c. In the case where the mutual inductances between qubit and coupler are not equal for both qubits, the same analysis leads to

$$\begin{aligned} M_{\text{eff}}(\varphi_{cx}) &= \frac{M_{ca}M_{cb}}{\tilde{L}_c}\frac{1}{1-\frac{1}{\beta_c\cos(\varphi_{cx}+\beta_c\sin_{\beta_c}(\varphi_{cx}))}} \\ &\equiv M_{ca}M_{cb}\chi_c, \end{aligned} \tag{5.33}$$

where

$$\chi_c = 1/(\tilde{L}_c + L_j), \tag{5.34}$$

is the inverse effective coupler inductance, often referred to as the coupler's susceptibility [197, 199]. In the language of linear circuit theory, this is equivalent to treating the junction as a linear inductance with $L = L_j(\varphi_{cx})$ as illustrated in Fig. 5.2. In fact, this is a second, equivalent way to obtain the result (5.32) without explicitly taking the derivative $\frac{\partial I_c}{\partial\Phi_{cx}}$ [one can simply plug in the expression (5.27) for $\varphi_{cj}(\varphi_{cx})$ into the formula (5.13) for the differential Josephson inductance $L_j(\varphi_{cj})$].

We now consider two sources of "nonlinear crosstalk" (perhaps a better name is "nonlinear biasing") due to the influence of the coupler circuit on each of the qubits.

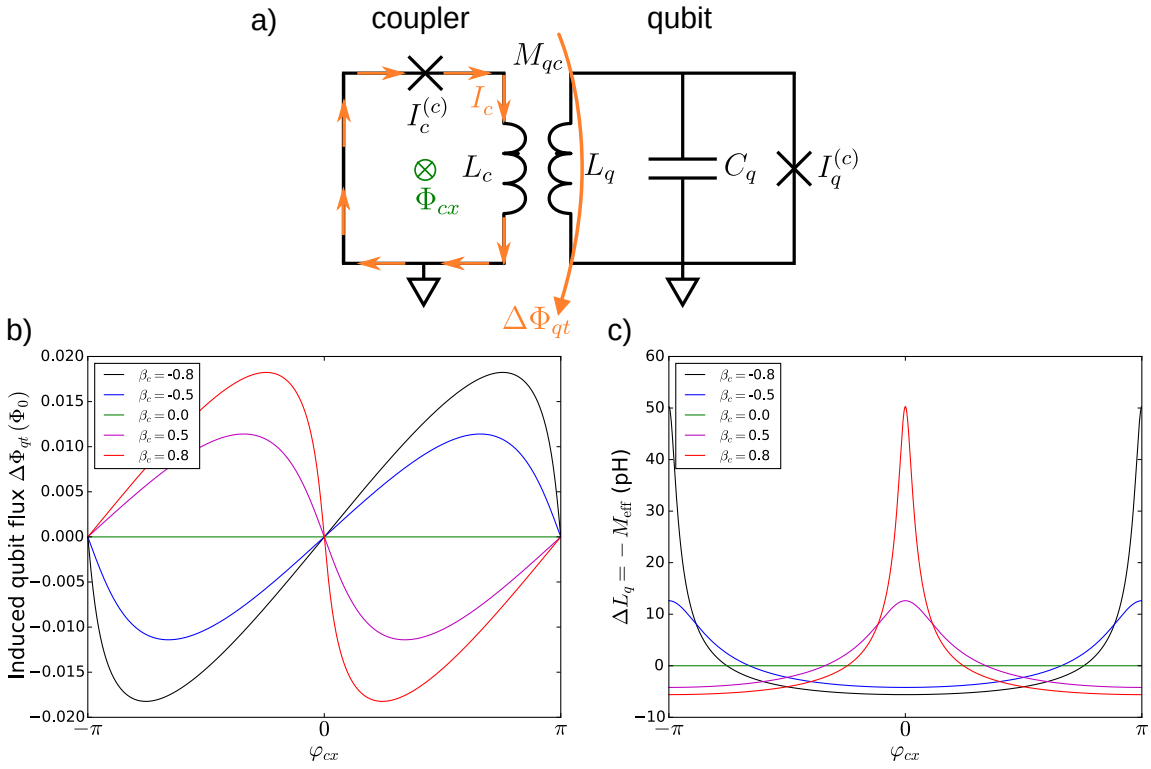


Figure 5.4: (a) Illustration of flux bias induced on qubit by the coupler's current. (b) Plot of coupler-induced nonlinear flux crosstalk versus coupler tilt bias, for typical design parameters and different values of β_c. (c) Plot of coupler-induced qubit inductance shift versus coupler bias, for typical design parameters and different values of β_c.

Namely, we consider two "induced local field" effects: in section 5.1.2.1 we look at the effective flux bias induced on the qubit by the coupler, and then in section 5.1.2.2 we look at the effective inductance shift of the qubit due to the coupler. Both of these effects are functions of the coupler's flux bias and must be carefully calibrated to operate a real device, as we'll see in the measurements of Chapter 6.

5.1.2.1 Coupler-induced nonlinear flux bias on qubit

One consequence of having a steady state current I_c flowing in the coupler loop is that it will impart a tilt flux bias on each of the qubits given by

$$\Delta\Phi_{qt} = M_{qc}I_c = M_{qc}I_c^{(c)}\sin_{\beta_c}(\varphi_{cx}). \tag{5.35}$$

This flux crosstalk, which is a nonlinear function of the coupler bias, is illustrated in Fig. 5.4(a). A plot of this induced flux vs. coupler bias for typical design parameters is shown in Fig. 5.4(b). This flux crosstalk is very substantial – for example, $20\,\mathrm{m}\Phi_0$ of tilt bias corresponds to an enormous $\varepsilon/h \approx 50\,\mathrm{GHz}$ for a persistent current of $0.5\,\mu\mathrm{A}$! The need to calibrate for a nonlinear function with such a large magnitude is one reason it maybe desirable to use a dc SQUID instead of a single junction in the coupler. With a dc SQUID, β_c can be tuned to both positive and negative values via the coupler SQUID flux bias Φ_{csq}, while keeping the coupler tilt bias at zero, where the induced flux on the qubit vanishes. Such a scheme is illustrated in Fig. 5.5(a), and the associated nonlinear flux crosstalk vs. both coupler SQUID and tilt biases are shown in Fig. 5.5(b) [the curves in Fig. 5.4(b) correspond to the vertical cuts of matching color in the 2D plot of Fig. 5.5(b)]. By operating along the horizontal cut $\Phi_{cx} = 0$, we can still obtain sign- and magnitude-tunable coupling with zero induced qubit flux bias. Physically, this is because the control current for the coupler can be localized within the coupler's DC SQUID [purple arrows in Fig. 5.5(a)], so unless there is any direct stray mutual inductance from the coupler's small dc SQUID loop to the qubits (more on this point in a real device next chapter), there will be no coupler-induced nonlinear flux crosstalk. The other advantage of using

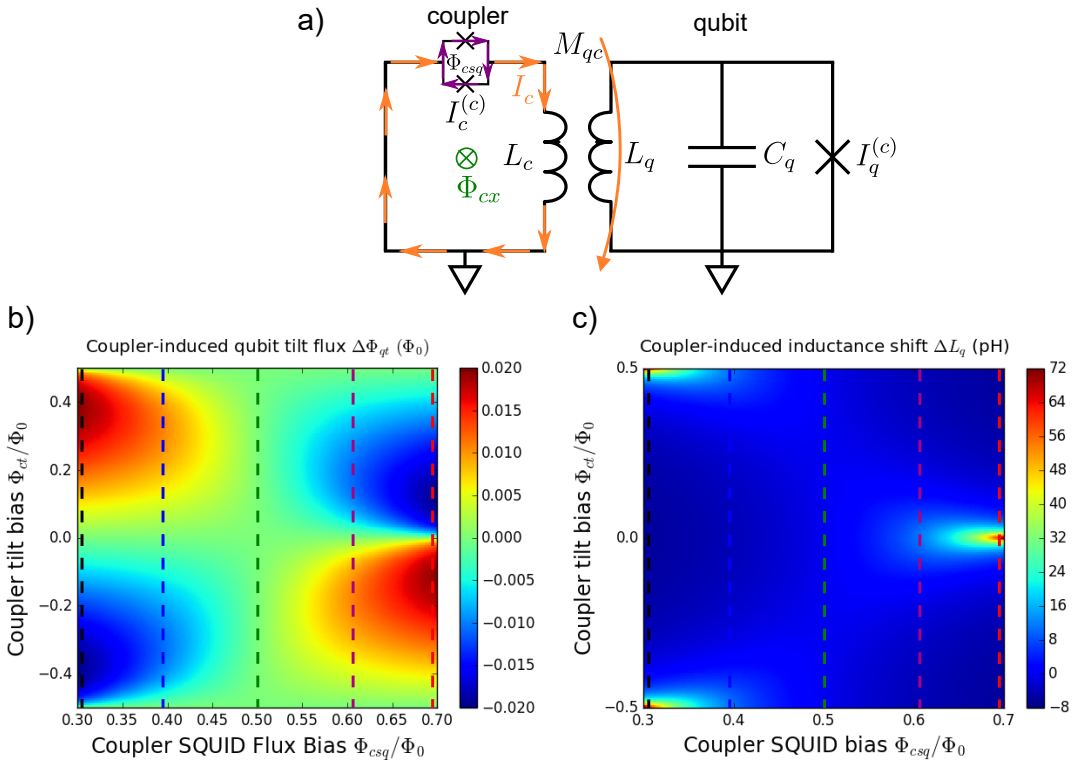


Figure 5.5: (a) Illustration of coupler-qubit nonlinear crosstalk with a DC SQUID rather than a single junction in the coupler. (b) Coupler-induced qubit tilt bias shift $\Delta\Phi_{qt}$ vs. coupler SQUID and tilt biases. (c) Coupler-induced inductance shift ΔL_q vs. coupler SQUID and tilt biases.

a dc SQUID for the coupler is that β_c can be made arbitrarily close to 1 *in situ*, as long as $\beta_c^{\max}$ is larger than 1. Compare this to a single-junction coupler, where in order to obtain strong coupling one would have to fabricate a junction with critical current corresponding to β_c close to but not exceeding 1, which is difficult in practice.

The fact that flux in the coupler loop influences the flux in the qubit can also have consequences if there is noise in the coupler flux. For example, $1/f$ noise originating in the coupler can translate to $1/f$ noise in the qubit if the sensitivity of $\Delta\Phi_{qt}$ to φ_{cx} is high enough. To quantify this effect, in Fig. 5.6 we plot the differential flux transfer function $d\Phi_{qt}/d\Phi_{cx}$ versus β_c at zero coupler tilt ($\varphi_{cx} = 0$), where the flux sensitivity of

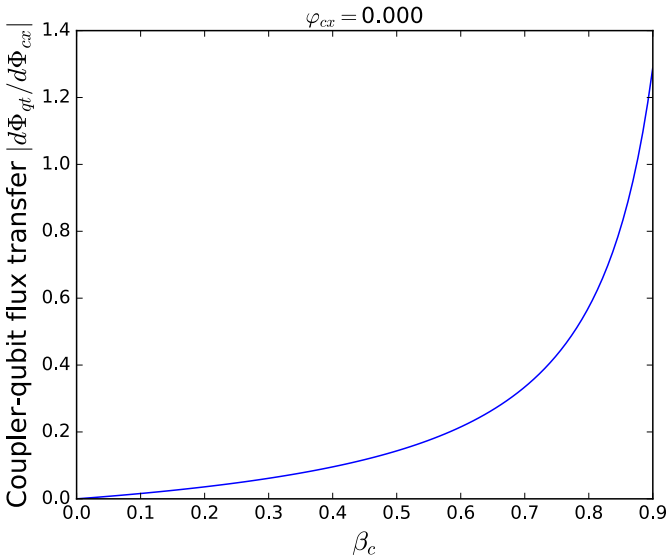


Figure 5.6: Differential flux transfer function from coupler to qubit at zero coupler tilt (maximal transfer), vs. coupler β_c.

the qubit to the coupler is largest [i.e., each point on this plot is the slope of one of the curves in Fig. 5.4(b) at $\varphi_{cx} = 0$]. As can be seen, for large β_c (where strongest coupling occurs), this transfer function can be around unity, meaning that the qubit will fully feel any flux noise in the coupler. If one qubit is to be strongly coupled to many others at the same time, this poses some concern for coherence in a quantum annealer, especially if the coupler's flux noise is comparable to or greater than the qubit's. We will have to learn how to trade off coupling strength and robustness to noise, and how to optimize the geometry of the couplers to reduce the flux noise intrinsic to them and therefore excess flux noise induced in the qubits (see measurements in next chapter). For example, we might not need to operate all of the couplers simultaneously at high β_c, or maybe we will never even need to exceed $\beta_c \sim 0.5$, since this number still allows for 2 GHz of coupling at $\Delta_q/h = 1\,\text{GHz}$, exceeding our "annealing region" rule of thumb.

Note that this differential flux transfer function also applies at GHz frequencies, so that any dissipative flux noise in the coupler will cause similar dissipative flux noise in the qubit. Since at strongest coupling each qubit might feel the flux noise of up to 20 couplers (in an optimistically high-connectivity architecture), both the low frequency and dissipative flux noise in a coupler should be roughly $\sqrt{20} \approx 5$ lower than that of a qubit. This poses an engineering challenge both in terms of flux noise intrinsic to the coupler and noise coming down from the coupler bias lines. Dissipation at GHz frequencies from the coupler can be solved by adding filtering to the bias line, but this does not fix low frequency noise.

5.1.2.2 Coupler-induced inductance shift in qubit

The coupler will also induce an effective shift in each qubit's self-inductance, as illustrated in Fig. 5.7. This added inductance can be calculated by considering the input impedance as seen from qubit's inductor [i.e., the same problem considered in Fig. 3.6(b)], leading to

$$\begin{aligned} \Delta L_q &= -\frac{M_{qc}^2}{L_c + L_j(\beta_c, \varphi_{cx})} \\ &= -\frac{M_{qc}^2}{L_c} \frac{1}{1 - \frac{1}{\beta_c \cos(\varphi_{cx} + \beta_c \sin_{\beta_c}(\varphi_{cx}))}} \end{aligned} \tag{5.36}$$

Note that within our linear analysis, ΔL_q has the same magnitude but opposite sign as M_{eff} of Eq. (5.33) (in the symmetric $M_{ca} = M_{cb} = M_{qc}$). Fig. 5.4(c) shows a plot of $\Delta L_q = -M_{\text{eff}}$ as a function of φ_{cx} for different values of β_c, and Fig. 5.5(c) shows a 2D

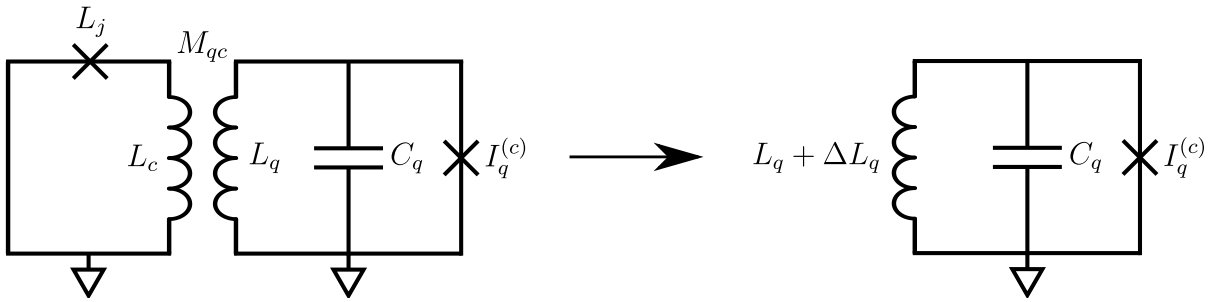


Figure 5.7: Linear circuit description of the effective inductance shift felt by a qubit due to the presence of a coupler.

plot of ΔL_q vs. coupler SQUID and tilt biases. Note that ΔL_q doesn't vanish at any particular tilt bias, since this depends on β. It does however vanish for all values of φ_{cx} when $\beta_c = 0$.

The coupler-induced inductance shift leads to a shift in qubit frequency that depends on coupler bias. An example of this is shown in Fig. 5.7 as a function of β_c, with the qubit and coupler are both at zero tilt. We see that for large coupler nonlinearities, this freuqency shift can be substantial, a few GHz at $\beta_c = 0.9$ for typical design parameters. It is possible to compensate for this frequency shift by adjusting the DC SQUID bias of the qubit to keep Δ constant versus coupler bias. However, while this can exactly compensate the frequency shift, after the compensation there will be a slight dependence of the resulting qubit I_p on coupler bias. This is because we are attempting to compensate a shift in geometric inductance with a shift in Josephson inductance, and these are not equivalent in the Hamiltonian. Fig. 5.7(b) shows the resulting qubit I_p vs. β_c after such a compensation for a qubit frequency of 4 GHz. This shift is fairly small, even for large coupler nonlinearities, and so might even be able to be ignored. We note that to address this subtlety, D-Wave actually designed "L-tuners" in their qubits, where a large

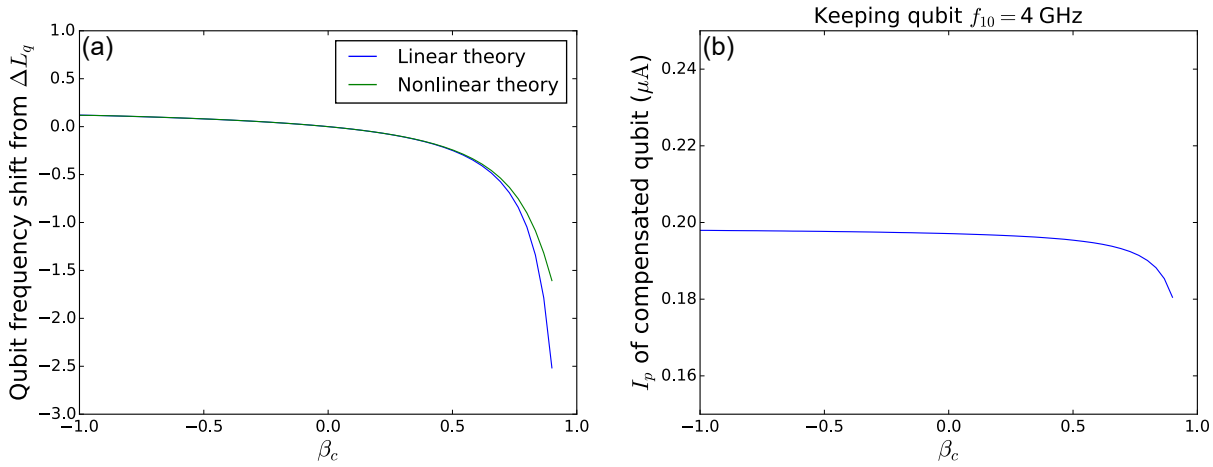


Figure 5.8: (a) Shift in qubit frequency due to coupler-induced inductance shift $\Delta L_q(\beta_c)$, at zero qubit and coupler tilt. (b) Resulting qubit persistent current after compensating the qubit's DC SQUID bias as a function of β_c to negate the effect of ΔL_q and keep a constant $f_{10} = 4\,\text{GHz}$ at zero tilt.

Josephson junction adds an extra tuning knob for the main linear qubit inductance. We do not a present plan to implement this extra knob, but we will need to at least consider it as an option when we scale up and start to apply our qubits to quantum annealing in a large system.

Furthermore, since ΔL_q is a function of coupler flux, flux noise in the coupler can translate to noise in the qubit inductance. Since the qubit inductance effects both ε and Δ, this translates to noise in both ε and Δ, although away from zero tilt the flux noise from direct flux crosstalk (i.e., Fig. 5.6) will dominate. In particular, a fluctuating inductance shift can dephase the qubit even when the qubit is at degeneracy, although qubit dephasing at zero tilt is not likely a dominant effect for annealing applications, for which the qubit will not remain at zero tilt. Nevertheless, to be thorough, in Fig. 5.9 we plot the resulting qubit sensitivity to the coupler flux biases solely due to the coupler-induced inductance shift.

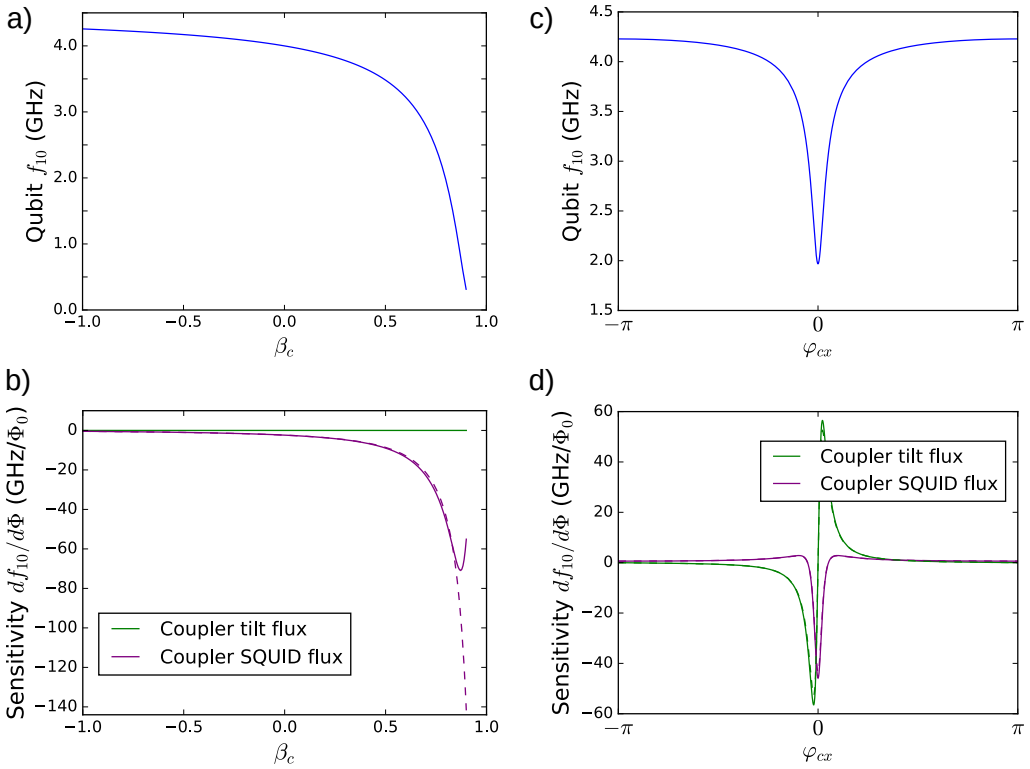


Figure 5.9: Qubit sensitivity to coupler fluxes at zero qubit tilt due to coupler-dependent inductance shift, with uncoupled qubit $\Delta/h = 4\,\mathrm{GHz}$. (a) Qubit frequency f_{10} vs. coupler β_c with the qubit kept at zero tilt. (b) Corresponding sensitivities of the qubit frequency to coupler flux biases, for both coupler tilt (green) and SQUID (purple). $L_c = 600,\mathrm{pH}$, $M_{qc} = 80\,\mathrm{pH}$ ($\alpha_i = 0.1$), and $\beta_c^{\max} = 1.5$ is assumed. Dashed lines are with "inductance compensation" in the qubit, meaning that for each coupler bias the qubit SQUID bias is compensated to keep the qubit frequency at $f_{10} = 4\,\mathrm{GHz}$. Solid lines are without such compensation. (c) Same as (a) but versus coupler tilt bias. In this case, the qubit is kept at zero tilt by compensating for the nonlinear coupler-induced flux bias. (d) Same as (b) but versus coupler tilt bias.

In summary, within a fully linear treatment of the coupler circuit without coupler capacitance, we can model the system as two qubits coupled via an effective mutual inductance, with coupler-induced inductance and flux shifts $\Delta L_{qi}(\varphi_{cx})$ and $\Delta\Phi_{ti}(\varphi_{cx})$ in each qubit according to the Hamiltonian

$$\begin{aligned} H^{\text{lin.}}_{\text{tunable}} =& \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} \\ &+ \left(\frac{1}{1+\frac{\Delta L_1(\varphi_{cx})}{L_1}}\right)\frac{\Phi_1^2}{2L_1} + \left(\frac{1}{1+\frac{\Delta L_2(\varphi_{cx})}{L_2}}\right)\frac{\Phi_2^2}{2L_2} \\ &- M_{\text{eff}}(\varphi_{cx})\frac{\Phi_1}{L_1}\frac{\Phi_2}{L_2} \\ &- E_{J1}\cos(2\pi[\Phi_1 + \Phi_{t1} + \Delta\Phi_{t1}(\varphi_{cx})]/\Phi_0) \\ &- E_{J2}\cos(2\pi[\Phi_2 + \Phi_{t2} + \Delta\Phi_{t2}(\varphi_{cx})]/\Phi_0). \end{aligned} \tag{5.37}$$

Before concluding this section, we note that there is an equivalent method of tunable coupling that uses a galvanic connection between the qubits and coupler. Namely, it can easily be shown that the two circuits in Fig. 5.10(a) are equivalent, and so then the two circuits in Fig. 5.10(b) are also equivalent. The galvanic coupling therefore essentially acts to boost α_i to unity. However, this extra coupling does not come for free, since the linear model will break down more easily and there will also be correspondingly more crosstalk in the system. Some qubit architectures may need to use a galvanic architecture to obtain strong enough coupling for quantum annealing applications [64], but the fluxmon architecture has more than enough coupling without needing any galvanic connections.

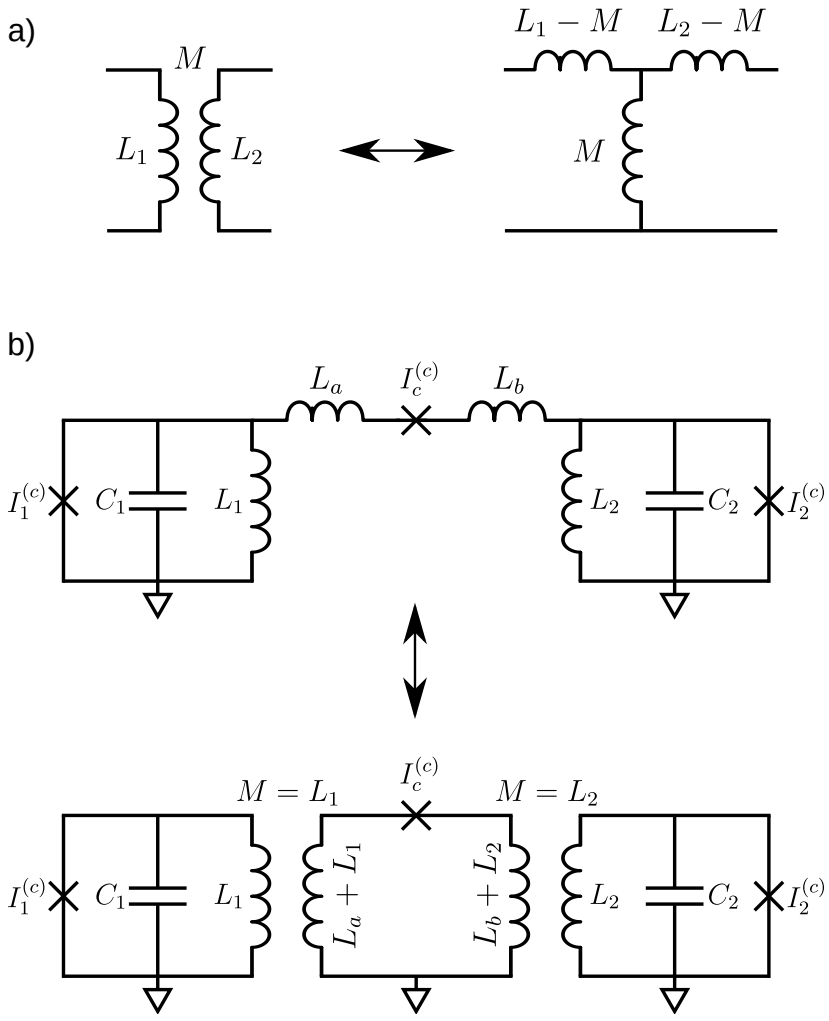


Figure 5.10: (a) Equivalent galvanic circuit replacement for a mutual inductance. (b) Equivalent galvanic and non-galvanic coupler circuits.

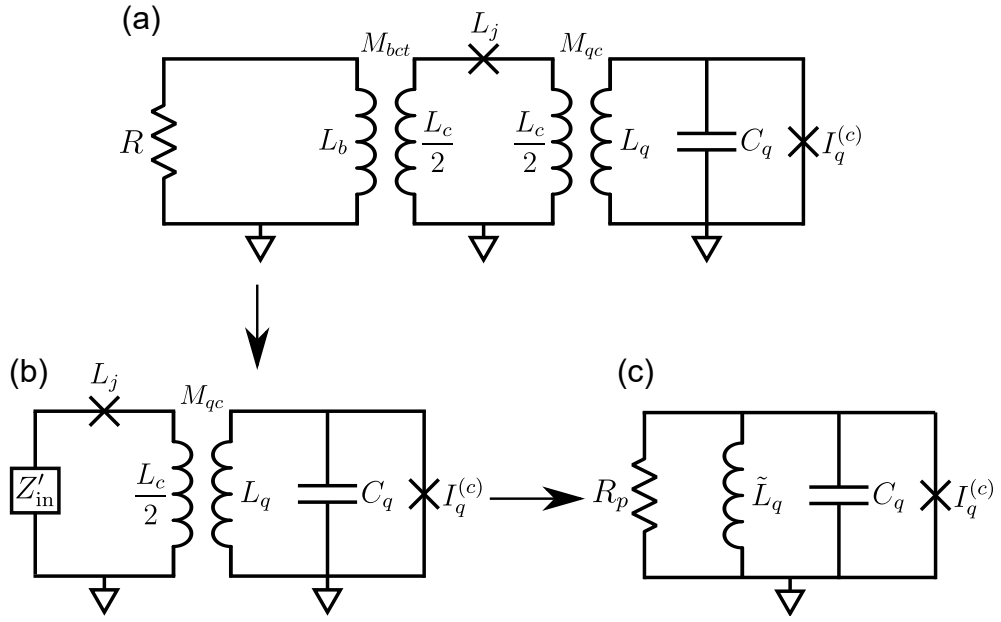


Figure 5.11: Circuit model for calculating qubit damping through the coupler's bias line into the $R = 50\,\Omega$ environment.

5.1.2.3 Qubit damping from coupler flux bias line

When designing a device, we must be aware of any excess qubit dissipation to the $50\,\Omega$ environment through the coupler's bias lines. This is especially important if one qubit is to be coupled to many couplers within a high connectivity quantum annealer. To compute the qubit damping through the coupler's main bias line, we consider the circuit[3] in Fig. 5.11. The strategy is to find an equivalent parallel resistance as seen from across the qubit's terminal due to the presence of the coupler and its bias line. We will use the identity (3.24) for the transformed impedance seen through a mutual inductance twice, once through the coupler's bias line mutual inductance M_{bct} and once through M_{qc}.

First, we compute the impedance $Z'_{\rm in}$ seen by the coupler looking into the bias line,

[3]The fact that we have depicted the coupler's inductance as divided into two equal segments is not of any fundamental significance. We would obtain the same answers dividing it up any other way.

$Z'_{\rm in} = i\omega L_c/2 + \frac{M_{bct}^2\omega^2}{i\omega L_b + R} \approx i\omega L_c/2 + M_{bct}^2\omega^2/R$. Next, we compute the impedance seen across the qubit's inductor looking into M_{qc}. Applying the same rule again yields for this impedance $Z''_{\rm in} = i\omega L_q + \frac{M_{qc}^2\omega^2}{i\omega L_c/2 + L_j + Z'_{\rm in}} = i\omega L_q + M_{qc}^2 \frac{M_{bct}^2\omega^2/R - i\omega L_c - i\omega L_j}{(L_c+L_j)^2 + M_{bct}^4\omega^2/R^2}$. In the regime we are concerned with, we will always have $L_c + L_j \gg M_{bct}^2\omega/R$ (the latter term will typically be around $\sim 1\times10^{-5}$ pH), so we can rewrite this as $Z''_{\rm in} \approx i\omega L_q - i\omega\frac{M_{qc}^2}{L_c+L_j} + \frac{M_{qc}^2 M_{bct}^2\omega^2}{R(L_c+L_j)^2}$, where we have separated out the imaginary and real parts representing a series inductance ΔL_q and resistance R_s. It will be convenient to convert the series resistance to an equivalent parallel resistance,

$$R_p \approx \frac{(\omega L_q)^2}{R_s} = R\left(\frac{L_q}{M_{qc}}\right)^2\left(\frac{L_c + L_j(\beta_c, \varphi_{cx})}{M_{bct}}\right)^2. \tag{5.38}$$

Note that this equation is intuitive, with the parallel resistance felt by the qubit being the original resistance times two voltage division factors through each of the two mutual inductances. We can then use Eq. (4.55) to compute the qubit T_1 imposed by this parallel admittance. In Fig. 5.12 we plot the induced qubit T_1 vs. β_c at zero coupler tilt (maximum coupling strength), for 1, 5, and 10 couplers, with $M_{bct} = 1\,$pH and $M_{qc} = 80\,$pH.[4] If this dissipation becomes an issue, in principle we can modify the coupler bias line to be higher impedance at GHz frequencies via a filter, which would still allow for desired coupler flux pulses with a bandwidth of $\sim 100\,$MHz.

[4]It is useful but not necessary to have the coupler's bias mutual inductance a bit larger than the qubit's, so we can achieve a larger range of fluxes in order to see the nonlinear flux crosstalk, which as described in the next chapter is very useful for calibration of a coupled qubit device.

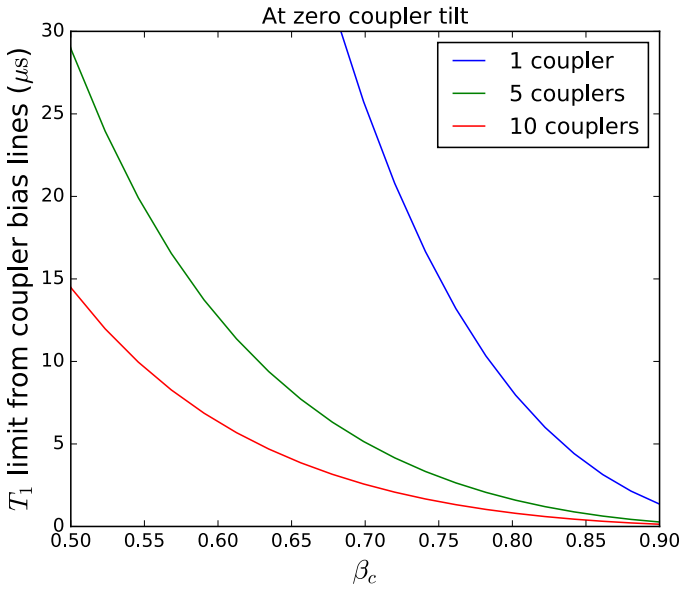


Figure 5.12: Plot of single-qubit T_1 limit imposed by $50\,\Omega$ coupler bias lines vs. β_c at zero coupler tilt for different number of couplers coupled to a given qubit. T_1 evaluated in qubit's harmonic limit, at zero tilt. Calculation assumes $M_{bct} = 1\,\text{pH}$ and $M_{qc} = 80\,\text{pH}$ ($\alpha_i = 0.1$).

5.2 Coupling in the nonlinear regime

So far, we have been using linear analysis for the coupler, which will eventually break down due to nonlinearity in the coupler's junction. Here, we derive the exact inter-qubit interaction Hamiltonian that properly takes into account the full behavior of the coupler's junction. This full nonlinear analysis will lead to significant departures from the linear results of the previous section for $\beta_c \gtrsim 0.75$ (the actual value at which linear treatment breaks down depends on the system parameters, especially α_i which are around 0.1 in our initial designs). In particular, this non-perturbative treatment will allow one to write down the inter-qubit coupling strength all the way up to $\beta_c = 1$ without any divergences. However, in a real system, even this analysis will deviate from reality at large β_c, due to neglected capacitance in the coupler present in any physical implementation,

a complication that we defer until later.

5.2.1 Non-perturbative nonlinear analysis of coupler circuit

The exact Hamiltonian of the coupler circuit of 5.2(a) can be found by writing down the true current equations of motion for each of the qubits and the coupler. For the qubits, we have the usual current conservation equations

$$\frac{\Phi_0}{2\pi} C_k \ddot{\varphi}_k + I_k + I_k^{(c)} \sin(\varphi_k + \varphi_{kx}) = 0. \tag{5.39}$$

Inserting equation (5.17) for I_k with Eq. (5.28) substituted for I_c yields

$$\frac{\Phi_0}{2\pi} \left(C_k \ddot{\varphi}_k + \frac{\varphi_k}{L_k} \right) + I_k^{(c)} \sin(\varphi_k + \varphi_{kx}) - \alpha_k I_c^{(c)} \sin_{\beta_c}(\varphi_{cx} - \alpha_1 \varphi_1 - \alpha_2 \varphi_2) = 0, \tag{5.40}$$

which is in terms of the qubit flux variables alone. Upon inspection we see that these current equations are the classical Euler-Lagrange equations of motion corresponding to the Hamiltonian

$$\hat{H}^{\text{nonlin.}}_{\text{tunable}} = \hat{H}_1 + \hat{H}_2 + E_{J_c} \cos_{\beta_c}(\varphi_{cx} - \alpha_1 \hat{\varphi}_1 - \alpha_2 \hat{\varphi}_2) \tag{5.41}$$

$$\equiv \hat{H}_1 + \hat{H}_2 + E_c(\hat{\varphi}_x), \tag{5.42}$$

where

$$\cos_\beta(\varphi_x) \equiv 1 - \int_0^{\varphi_x} \sin_\beta(\varphi') d\varphi' \tag{5.43}$$

and $\hat{H}_k$ are the uncoupled qubit Hamiltonians. For reference, an illustration of the $\cos_\beta(\varphi)$ function is plotted in Fig. 5.3(b). Since its first derivative is the $\sin_\beta$ function,

its second derivative at $\varphi = 0$ will diverge in the limit $\beta \to 1$. A somewhat more direct but less rigorous way to arrive at the result for the interaction term in (5.41) is to simply identify $H_{\rm int}$ with the extra energy stored in the coupler circuit, which is the sum of its inductive and junction energies [with the inductance replaced by the "renormalized" inductance $\tilde{L}_c$ defined in (5.21)],

$$\begin{aligned} E_{\rm c}(\varphi_x) &= \frac{1}{2}\tilde{L}_c I_c^2 + E_{J_c}\cos(\varphi_{cj}) \\ &= E_{\tilde{L}_c}\left[\frac{1}{2}\left(\beta_c \sin_{\beta_c}(\varphi_x)\right)^2 + \beta_c \cos(\varphi_x + \beta_c \sin_{\beta_c}(\varphi_x))\right]. \end{aligned} \tag{5.44}$$

Differentiating the right hand side with respect to φ_x and comparing the result to the identity (5.31) for $\partial_{\varphi_x}\sin_{\beta_c}(\varphi_x)$ lets us recognize the right hand side as $E_{\tilde{L}_c}\int -\beta_c \sin_{\beta_c}(\varphi_x) = E_{\tilde{L}_c}\beta_c \cos_{\beta_c}(\varphi_x)$, leading to

$$E_c(\varphi_x) = E_{J_c}\cos_{\beta_c}(\varphi_{cx} - \alpha_1\varphi_1 - \alpha_2\varphi_2). \tag{5.45}$$

The interaction Hamiltonian contains both an interaction between qubits and local field effects similar to those described in the previous section. We can compute any matrix element of the interaction Hamiltonian within the subspace of the two lowest levels of each uncoupled qubit according to[5]

$$\hbar g_{\bar{\eta}} = \frac{1}{4}\mathrm{tr}\left[\sigma_{\bar{\eta}}\hat{\mathrm{H}}_{\rm int}\right], \tag{5.46}$$

[5]This formula for the matrix elements comes from the standard Hilbert-Schmidt inner product on operators [100], properly normalized for a two-qubit space, since $\mathrm{tr}[\sigma_\alpha\sigma_\beta]/2 = \delta_{\alpha\beta}$ for the single-qubit Pauli matrices.

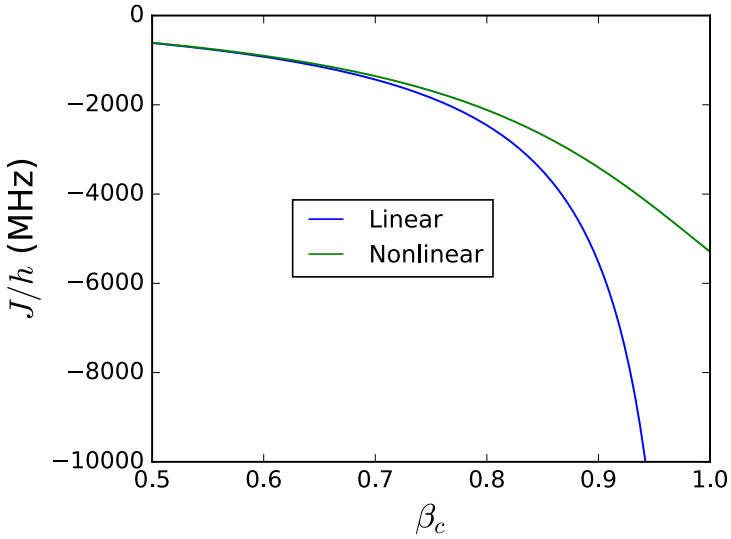


Figure 5.13: Plot of qubit-qubit transverse coupling strength $J = g_{xx}$ for the linear and full nonlinear theories at zero qubit and coupler tilts ($\varphi_{cx} = 0$, i.e., maximum negative coupling). Both qubit uncoupled gaps are $\Delta_i/h = 4\,\mathrm{GHz}$.

where[6]

$$\hat{H}_{\mathrm{int}} = \hbar \sum_{\bar{\eta}} g_{\bar{\eta}} \sigma_{\bar{\eta}}, \tag{5.47}$$

and

$$\sigma_{\bar{\eta}} = \sigma^{(1)}_{\eta_1} \otimes \sigma^{(2)}_{\eta_2}. \tag{5.48}$$

For example, the coupling strength between qubits (which is transverse in the qubits' parity basis) is g_{xx}, the coefficient of the $\sigma^{(1)}_x \sigma^{(2)}_x$ term, usually called J in quantum annealing literature.

We can compare the various matrix elements predicted by the full nonlinear theory to those predicted by the linear theory, by using the linear approximation for H_{int} in (5.46),

[6]Technically the left hand side is $\hat{P}_q \hat{H}_{\mathrm{int}} \hat{P}_q$, where $\hat{P}_q = \hat{P}_1 \otimes \hat{P}_2$ is the projection operator onto the qubit subspace. Explicitly, $\hat{P}_j = |0\rangle\langle 0|_j + |1\rangle\langle 1|_j$. Then, as shown in Ref. [55], an explicit Fourier series for any coefficient is $\hbar g_{\bar{\eta}} = E_{\tilde{L}_c} \sum_{\nu \neq 0} B_\nu e^{i\nu\varphi_{cx}} \prod_{j=1}^{k} c^{(j)}_{\eta_j}(\nu\alpha_j)$, where $B_\nu = \frac{J_\nu(\beta_c \nu)}{\nu^2}$ and we have defined the coefficients $c_I(s) = \frac{\langle 0|e^{-is\hat{\varphi}}|0\rangle + \langle 1|e^{-is\hat{\varphi}}|1\rangle}{2}$, $c_x(s) = \frac{\langle 0|e^{-is\hat{\varphi}}|1\rangle + \langle 1|e^{-is\hat{\varphi}}|0\rangle}{2}$, $c_y(s) = i\frac{\langle 0|e^{-is\hat{\varphi}}|1\rangle - \langle 1|e^{-is\hat{\varphi}}|0\rangle}{2}$, $c_z(s) = \frac{\langle 0|e^{-is\hat{\varphi}}|0\rangle - \langle 1|e^{-is\hat{\varphi}}|1\rangle}{2}$. and

namely expanding to second order in $\alpha_k\varphi_k$,

$$\hat{H}_{\rm int}^{\rm lin.} = E_c'(\varphi_{cx})(\alpha_1\hat{\varphi}_1 + \alpha_2\hat{\varphi}_2) + \frac{1}{2}E_c''(\varphi_{cx})(\alpha_1\hat{\varphi}_1 + \alpha_2\hat{\varphi}_2)^2, \tag{5.49}$$

which is (to leading order) equivalent to our previous linear treatment. For example, the coupling within this approximation is the cross-term $M_1M_2\chi_c\hat{I}_1\hat{I}_2$, where

$$\chi_c = \frac{1}{\tilde{L}_c}\frac{E_c''(\varphi_{cx})}{E_{\tilde{L}_c}}, \tag{5.50}$$

which is equivalent to the linear result 5.33. In Fig. 5.13 we plot $J = g_{xx}$ vs. β_c at zero coupler tilt bias for the nonlinear and linear theories. Both theories agree in the regime of weak coupling. The most drastic change from using the full Hamiltonian (5.41) is that the coupling strength no longer diverges as $\beta_c \to 1$, unlike in the linear theory, which breaks down drastically above $\beta_c \sim 0.8$. This agrees with physical intuition. Even though the second derivative of the interaction term diverges as $\beta_c \to 1$, taking into account the higher derivatives of the $\cos_{\beta_c}$ smooths out the "effective" second derivative felt by the qubits, leaving the interaction strength finite (a more rigorous explanation of this point will be provided in the next section where we will discuss one interpretation of coupling as a finite difference approximation to the second derivative). In Fig. 5.14 we plot the various Pauli coefficients (5.46), including g_{xx}. Of particular interest is the appearence of a non-stoquastic[7] interaction term of the form $\sigma_z\sigma_z$ (in the parity basis) that is absent in any linear theory. This non-stoquastic term becomes non-negligible (but is still fairly small) above $\beta_c \sim 0.8$, where the linear theory breaks down. Non-stoquastic terms

[7] A non-stoquastic Hamiltonian is one that has both postive and negative off-diagonal matrix elements, leading to the 'sign problem' that prevents efficient simulation of the system using stochastic integration techniques like quantum Monte Carlo.

are of somehwat controversial theoretical interest for adiabatic quantum computing and quantum annealing since they may be necessary to ever observe an exponential speedup over classical methods. For example, for certain classes of problems introducing non-stoquastic terms can lead to exponential speedup by changing a first-order quantum phase transition into a second-order one [200], although this example was for a very complicated Hamiltonian with higher order qubit interactions than 2-body. It has also been shown that in a two-local spin architecture adding such interactions is necessary to make adiabatic quantum computing universal and QMA-complete [201, 202]. However, physically implementing these terms in a tunable way with large magnitude in a practical device seems difficult if at all possible, at least with the fluxmon architecture, and the actual benefit of non-stoquastic terms in an actual quantum annealer with noise, or when one is not expected to remain near the ground state at all times, is unclear.

5.2.2 Simple finite difference picture of nonlinearity

In the linear theory, the response of the coupler to a change in qubit flux is described with a first order susceptibility, which is proportional to the second derivative of the energy stored in the coupler. It would be nice to have a similarly intuitive description of the coupling in the nonlinear regime. One way to reproduce the nonlinear theory to reasonable accuracy is by replacing this second derivative with a finite difference of the coupler-induced flux, which in particular will not diverge as $\beta_c \to 1$. In this way, we can understand the effect of nonlinearity as an effective average of the otherwise diverging

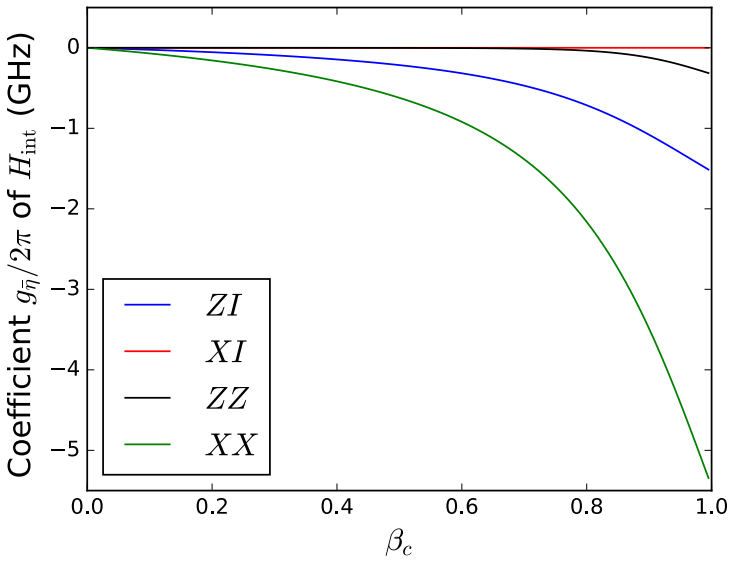


Figure 5.14: Two-qubit Pauli coefficients of the interaction Hamiltonian with uncoupled $\Delta_1/h = \Delta_2/h = 4\,\text{GHz}$. Note in particular the lack of divergence in g_{xx}, and the presence of a small non-stoquastic coupling term g_{zz} above $\beta_c \sim 0.8$ that is absent in a linear treatment of the coupler.

second derivative over a finite interval of flux. This also gives us an intuitive criterion for predicting when the linear theory breaks down and nonlinear theory becomes necessary, namely whether or not the curvature of the coupler energy appreciably changes when feeling a finite qubit-induced shift in flux bias.

To see what we mean by this, let us suppose that the qubits are in the double-well limit and have a low impedance so that their wavefunctions in the flux basis are approximately delta functions centered at the left and right well minima. This means that a change in the flux state of one qubit will induce a well-defined change in coupler flux, which in turn influences the second qubit. Explicitly, we assume the qubit energy wavefunctions

at zero tilt are

$$\langle\varphi|0\rangle = \sqrt{\frac{\delta(\varphi+\varphi_m)+\delta(\varphi-\varphi_m)}{2}},$$
$$\langle\varphi|1\rangle = \sqrt{\frac{\delta(\varphi+\varphi_m)-\delta(\varphi-\varphi_m)}{2}}. \tag{5.51}$$

We can then compute the coupler-mediated coupling strength,

$$\begin{aligned}\hbar g_{xx} &= \langle 00|\hat{H}_{\rm int}|11\rangle \\ &= \iint d\varphi_1 d\varphi_2 \langle 0|\varphi_1\rangle\langle\varphi_1|1\rangle\langle 0|\varphi_2\rangle\langle\varphi_2|1\rangle E_c\left(\varphi_{cx}-\alpha(\varphi_1+\varphi_2)\right) \\ &= \frac{E_c(\varphi_{cx}+2\alpha\varphi_m)-2E_c(\varphi_{cx})+E_c(\varphi_{cx}-2\alpha\varphi_m)}{4}\end{aligned} \tag{5.52}$$

where for simplicity we have assumed symmetry bewteen the qubits. From basic calculus, one can immediately recognize (5.52) as the expression of the finite difference approximation to the second derivative [multiplied by $(2\alpha\varphi_m)^2$], which is the same result as one would obtain by replacing the coupler susceptibility with the average of the second derivative of the coupler energy over the interval $[\varphi_{cx}-2\alpha\varphi_m, \varphi_{cx}+2\alpha\varphi_m]$. This finite difference deviates from the second derivative method precisely when the second derivative changes appreciably over the interval $[\varphi_{cx}-2\alpha\varphi_m, \varphi_{cx}+2\alpha\varphi_m]$, requiring the use of an average as opposed to a derivative at one single point. Replacing the delta functions with localized Gaussian wavefunctions will change the answer slightly, but this simple treatment gives a somewhat intuitive picture of what happens in the nonlinear regime under reasonable assumptions.

5.3 Accounting for coupler capacitance degree of freedom

In our analysis of the coupler circuit so far, we have treated the coupler's current/flux as a well-defined classical variable. In reality, the coupler's flux must also be quantized just like the qubits', and in particular if there is finite coupler capacitance (as is physically inevitable) the coupler will have an associated zero-point motion. In an ideal scenario, we wouldn't need to worry about the dynamics of the coupler itself, and would always be able to factor out the wavefunction of the coupler from the wavefunction of the two qubits. That way, the coupler just provides an effective interaction between the qubits and we wouldn't neeed to explicitly consider the dynamical role of the coupler itself. A sufficient condition for this to occur is that the coupler always remains in its ground state. This is the principle behind the Born-Oppenheimer approximation, which was previously discussed in the context of section 2.2.3.2. In this section, we will show how the Born-Oppenheimer approximation can be used to derive an effective interaction Hamiltonian between qubits in the presence of coupler capacitance, in the limit where the coupler's characteristic frequencies are much higher than that of the qubits'. We will find two effects of coupler capacitance, a contribution from the coupler's zero-point energy from a finite coupler impedance, as well as a "dispersive shift" from a finite coupler frequency. We will also explore the regime of applicability of the Born-Oppenheimer approximation.

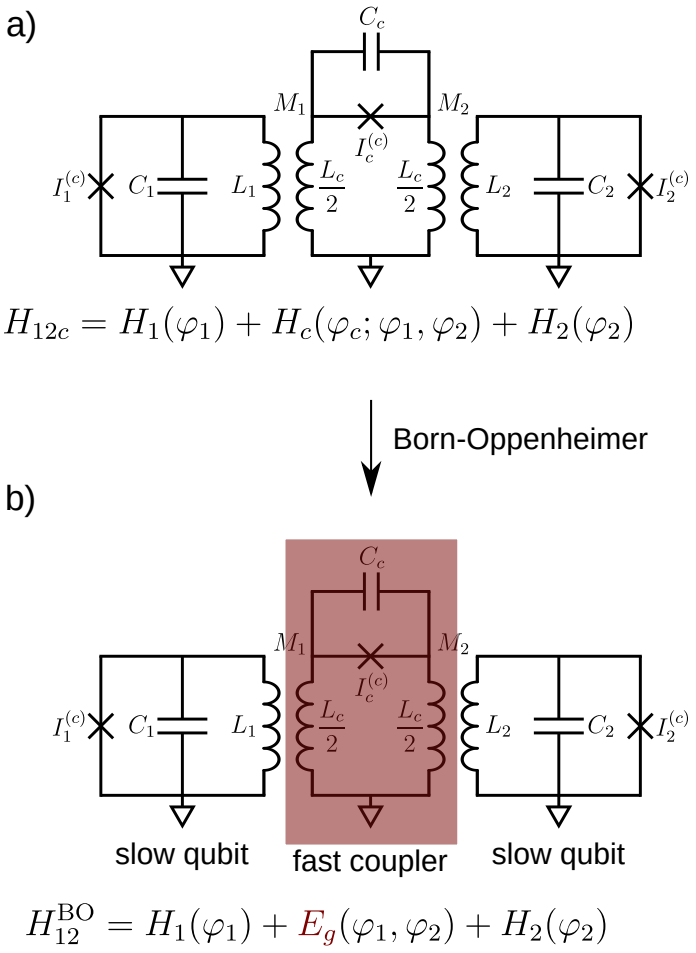


Figure 5.15: Illustration of the Born-Oppenheimer approximation as applied to two qubits coupled through a coupler with finite capacitance. The coupler is taken to be the "fast" degree of freedom and is assumed to always be in its ground state no matter what state the "slow" qubits are in. Therefore, the coupler's Hamiltonian is replaced by its ground state energy, reducing the system to from three-dimensional to an effectively two-dimensional.

5.3.1 Quantum correction to the Born-Oppenheimer approximation

We consider the circuit in Fig. 5.15(a), which is the same as that in Fig. 5.2(a) but with capacitance in the coupler. A similar analysis of the current equations of motion as before (this time taking into account coupler capacitance, details found in Ref. [55])

yields full three-dimensional system Hamiltonian

$$\begin{aligned}\hat{H}_{12c} &= \frac{\hat{Q}_c^2}{2C_c} - E_{J_c}\cos(2\pi\hat{\Phi}_c/\Phi_0) + \frac{\left(\hat{\Phi}_c - \Phi_{cx} + \alpha_1\hat{\Phi}_1 + \alpha_2\hat{\Phi}_2\right)^2}{2\tilde{L}_c} + \hat{H}_1 + \hat{H}_2 \\ &\equiv \hat{H}_c + \hat{H}_1 + \hat{H}_2, \end{aligned} \tag{5.53}$$

where $\Phi_c = \Phi_{cj}$ is the flux across the coupler's junction and

$$\tilde{L}_c = L_c - \alpha_1 M_1 - \alpha_2 M_2 \tag{5.54}$$

is the renormalized coupler inductance just like in Eq. (5.21), and $\hat{H}_1$ and $\hat{H}_2$ are the Hamiltonians of the uncoupled qubits in the absence of the mutual inductances M_1 and M_2. This Hamiltonian for the full system and has three degrees of freedom, one for the coupler and each of the two qubits. The Born-Oppenheimer approximation adiabatically eliminates[8] the coupler's degree of freedom, meaning we assume the coupler is "fast" in that it responds instantaneously to the "slow" qubit fluxes and always remains in its instantaneous ground state, analogous to the slow nuclei evolving adiabatically relative to the fast electrons allowing for the nuclear degree of freedom to be treated as a parameter within the electronic Hamiltonian. What is left will then be an effective interaction Hamiltonian between qubits taking the form of a potential energy precisely given by the instantaneous ground state energy of the coupler. For this approximation to be valid, the characteristic frequency $\omega_{10,c} \sim \frac{1}{\sqrt{\tilde{L}_c C_c}}\sqrt{1-\beta_c}$ of the coupler must be much bigger than the qubits' characteristic frequencies. For typical coupler designs (see next chapter) we

[8]Technically, "adiabatic elimination" is not exactly the same thing since we are not accounting for virtual transitions into excited states.

will have $L_c \approx 500 - 600\,\text{pH}$ and $C_c \sim 20 - 30\,\text{fF}$, meaning $f_{LC,c} \approx 40 - 50\,\text{GHz}$, meaning qualitatively the Born-Oppennheimer approximation should hold even for fairly large β_c (we will revisit this claim quantiatively later).

To make the Born-Oppenheimer approximation, the coupler-dependent part of the Hamiltonian, $H_c(\varphi_c; \varphi_1, \varphi_2)$ in (5.53), must be replaced by its ground state energy $E_g(\varphi_1, \varphi_2)$ with the qubit fluxes treated as parameters, transforming 5.53 into an effective two-qubit Hamiltonian

$$\hat{H}_{12}^{\text{BO}} = \hat{H}_1(\hat{\varphi}_1) + \hat{H}_2(\hat{\varphi}_2) + E_g(\hat{\varphi}_1, \hat{\varphi}_2). \tag{5.55}$$

This idea is illustrated in Fig. 5.15(b). Note that in our nonlinear analysis in section 5.2, where we referred to E_g simply as "the energy stored in the coupler" E_c in (5.41), we were performing a version of the Born-Oppenheimer approximation in disguise, where the coupler's ground state energy was simply equal to the well-defined classical potential energy stored in the coupler. This classical energy is one of essentially two contributions to the coupler's ground state energy, the other being its zero point energy about its classical minimum. We must therefore generalize this treatment by including a kinetic energy term for the coupler.

As outlined in Ref. [55], we can write the coupler-dependent part of the full Hamiltonian as

$$\begin{aligned} \hat{H}_c &\equiv \hat{H}_{12c} - \hat{H}_1 - \hat{H}_2 \\ &= E_{\tilde{L}_c} \left(4\zeta_c^2 \frac{\hat{q}_c^2}{2} + U(\hat{\varphi}_c; \varphi_x) \right), \end{aligned} \tag{5.56}$$

where

$$\zeta_c \equiv \frac{2\pi e}{\Phi_0}\sqrt{\frac{\tilde{L}_c}{C}} = 4\pi\frac{\tilde{Z}_c}{R_K} \tag{5.57}$$

is a measure of the coupler's effective impedance, which determines the width of the coupler ground state wavefunction and therefore the relevance of quantum corrections to the ground state energy, and

$$U(\varphi_c;\varphi_x) = \frac{(\varphi_c-\varphi_x)^2}{2} + \beta_c\cos(\varphi_c) \tag{5.58}$$

is a dimensionless coupler potential normalized to $E_{\tilde{L}_c}$. Note that this potential is a function of the qubit flux biases through the parameter

$$\varphi_x \equiv \varphi_{cx} - \alpha_1\varphi_1 - \alpha_2\varphi_2. \tag{5.59}$$

The classical contribution to the coupler's ground state energy is simply the minimum value of the coupler's potential,

$$U_{\min}(\varphi_x) = \min_{\varphi_c} U(\varphi_c;\varphi_x) \equiv \frac{(\varphi_c^{(*)}-\varphi_x)^2}{2} + \beta_c\cos(\varphi_c^{(*)}), \tag{5.60}$$

where $\varphi_c^{(*)}$ is the value of the coupler flux that minimizes the potential, where we have

$$\partial_{\varphi_c}U(\varphi_c;\varphi_x)\big|_{\varphi_c=\varphi_c^{(*)}} = \varphi_c^{(*)} - \varphi_x - \beta_c\sin(\varphi_c^{(*)}) = 0. \tag{5.61}$$

We note that performing this minimization leads to the classical current equation of motion of the system in the limit of large coupler plasma frequency $\tilde{L}_cC \to 0$ [55]. Also note that $\varphi_c^{(c)}$ is the same as φ_{cj} used in the classical nonlinear analysis in the previous section 5.2.1, where we have already solved this same transcendental equation, allowing

us to write

$$\begin{aligned} U_{\min}(\varphi_x) &= \frac{(\beta_c \sin_{\beta_c}(\varphi_x))^2}{2} + \beta_c \cos(\varphi_x + \beta_c \sin_{\beta_c}(\varphi_x)) \\ &= \beta_c \cos_{\beta_c}(\varphi_x), \end{aligned} \tag{5.62}$$

which is the exact same energy as $E_c/E_{\tilde{L}_c}$ as we derived from (5.44) in section 5.2.1, where the second line follows from the same argument that led to Eq. (5.45). We also note that at this point we have derived the identity (2.98) used in section 2.2.3.2, where we originally introduced the $\cos_\beta$ function for the Born-Oppenheimer approximation in the context of the 1D no caps fluxmon model.

Next, we need to consider the remaining quantum contribution to the ground state energy (i.e., the "zero-point energy" contribution), formally defined as

$$\begin{aligned} U_{\text{ZPE}}(\varphi_x) &= E_g/E_{\tilde{L}_c} - U_{\min}(\varphi_x) \\ &= \langle \psi_{g,c} | \left(4\zeta_c^2 \frac{\hat{q}_c^2}{2} + U(\hat{\varphi}_c; \varphi_x) - U_{\min}(\varphi_x) \right) | \psi_{g,c} \rangle, \end{aligned} \tag{5.63}$$

where $|\psi_{g,c}\rangle$ is the coupler ground state corresponding to the qubit fluxes φ_1 and φ_2. In general, this zero-point energy must be computed through numerical diagonalization. However, if ζ_c is small, as is the case experimentally (our first coupled fluxmon device is estimated to have $\zeta_c \approx 0.05$ [see next chapter]), we can obtain an approximate analytical expression for it by expanding about the potential minimum to second-order and using the expression for the zero-point energy of the resulting harmonic oscillator. Explicitly,

this approximation leads to

$$U_{\mathrm{ZPE}} \approx \frac{1}{2}\sqrt{4\zeta_c^2 U''(\varphi_c^{(c)};\varphi_x)}, \tag{5.64}$$

where

$$U''(\varphi_c;\varphi_x) \equiv \partial_{\varphi_c}^2 U(\varphi_c;\varphi_x). \tag{5.65}$$

This constitutes the harmonic approximation for the coupler's zero-point energy. As shown in Ref. [55], an analytic Fourier series can be derived for U_{ZPE} in this approximation. The Born-Oppenheimer approximation is now complete, with the effective interaction Hamiltonian given by

$$H_{\mathrm{int}}^{\mathrm{BO}}(\varphi_1,\varphi_2)/E_{\tilde{L}_c} = U_{\mathrm{min}} + U_{ZPE}. \tag{5.66}$$

We will plot the effect of the quantum correction along with the effect of a second type of correction that we will now discuss.

5.3.2 Qubit-coupler hybridization: "dispersive shift" correction

The Born-Oppenheimer approximation by construction neglects any hybridization of the qubit and coupler, which can become significant when the detuning between the coupler frequency and qubit frequencies becomes comparable to or small relative to the coupling g_{qc} between qubit and coupler. This is usually understood as a "dispersive shift" in the language of circuit QED, wherein the qubit-like first excited state of the system has a small component that "lives in" the resonator (the coupler) it is coupled to [72]. In the language of harmonic oscillators ($\beta = 0$ limit for qubits and coupler), the normal modes of the system become slightly rotated, with some of the qubit current flowing through

the coupler and vice versa.

Usually, the effect of the dispersive shift is summarized in a statement such as the qubit frequency undergoes a shift by an amount g^2/Δ, where Δ is the detuning between the qubit and coupler (or resonator). However, the physics of the dispersive shift can be incorporated into our Born-Oppenheimer model in a more generically useful and accurate way. The key is to realize that the coupler flux φ_c was not the correct variable to eliminate, because it is not truly the fastest degree of freedom in the system. In the harmonic limit, the fastest degree of freedom would not be the coupler flux itself, but rather the normal mode coordinate associated with the highest frequency mode of oscillation, which is mostly the coupler flux but with a slight hybridization with the qubit flux. In general, $\beta \neq 0$ and we will not be in the limit of harmonic normal modes, but we can still find the "fastest" degree of freedom to eliminate as follows. Instead of the coupler flux φ_c, we can eliminate the coupler flux relative to its classical minimum point $\varphi_c - \varphi_c^*$ (which is a function of the qubit fluxes). It turns out that to first order this is equivalent to the qubit-like degree of freedom inheriting a fraction of the coupler capacitance according to [203]

$$C_q \to C_q + \frac{\alpha^2}{(1-\beta_c)^2} C_c. \tag{5.67}$$

We can compare the full system spectrum predicted by our BO analysis with and without the quantum and dispersive shift corrections to a full numerical diagonalization of the system. In Fig. 5.16, we plot the low-lying spectrum of the system for both

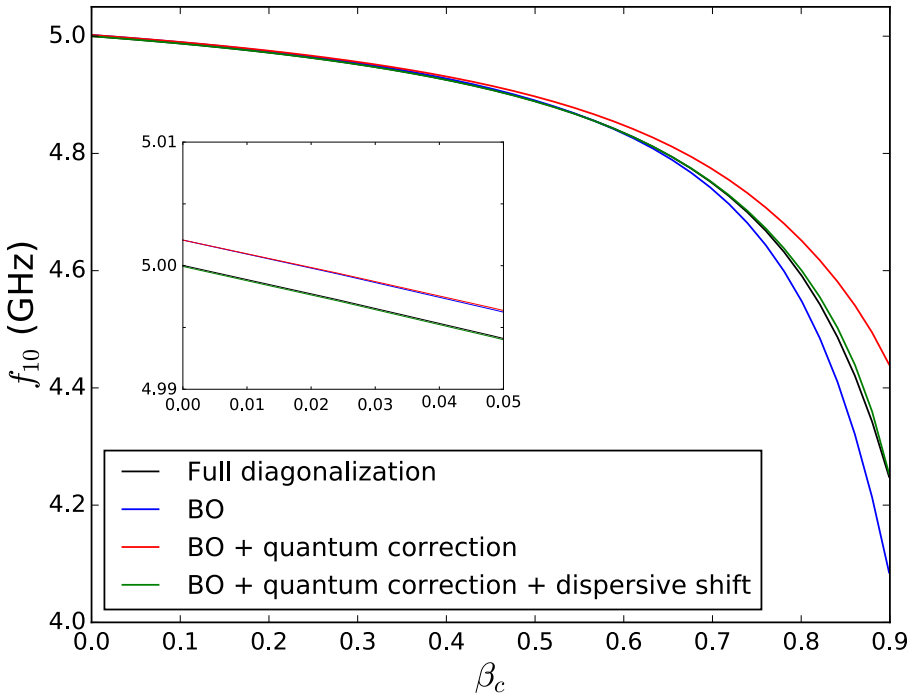


Figure 5.16: Comparison of predicted qubit(-like) f_{10} for four different models (other qubit biased at $\beta = 0$): full 3D diagonalization, Born-Oppenheimer, Born-Oppenheimer with quantum correction, and finally Born-Oppernheimer with quantum correction and dispersive shift correction as per the description in this section. Including the dispersive shift correction gives a very close match to full diagonalization all the way to very high coupler nonlinearity, corresponding to an uncoupled coupler $f_{10} \approx 15\,\mathrm{GHz}$. Standard fluxmon parameters were used, with coupling parameters $\alpha_i = 30\,\mathrm{pH}$ ($\alpha_i \approx 0.04$), coupler impedance $\zeta_c \approx 0.04$ and coupler harmonic $f_{10}(\beta_c = 0) = 40\,\mathrm{GHz}$. Adding the dispersive shift correctly predicts a shift in f_{10} when $\beta_c = 0$, which was not previously captured (see inset, where the black and green curves are so close together that it is hard to see).

theories at zero coupler tilt and qubit at degeneracy, with qubit nonlinearity chosen to be equal to $\beta_q = 1.0$ for one qubit and $\beta_q = 0$ for the other. We can see that the BO treatment breaks down at high nonlinearities, and the quantum correction plus dispersive shift is necessary to accurately model the system. Note in particular that the dispersive shift correction correctly predicts a shift in qubit freuqency even when the coupler is the harmonic limit $\beta_c = 0$, which would not have been predicted by the simple inductance shift model discussed in section 5.1.2.2.

Chapter 6

Interacting fluxmons: achieving ultra-strong tunable coupling

In this chapter, we implement a properly designed coupler circuit to realize tunable ultra-strong coupling between two fluxmon qubits without reducing qubit coherence. The circuit described here can in principle be used to couple one fluxmon to up to ten or more others simultaneously, as desired for high-connectivity quantum annealing applications. We will first describe the physical design of the coupler device, and then walk through some of the basic aspects of device calibration, including calibration of crosstalk (both geometric and nonlinear) and extraction of circuit model parameters from experimental data. We then demonstrate (to our knowledge) the first spectroscopic measurement of ultra-strong tunable coupling between two superconducting qubits. This regime of qubit-qubit interactions is not only useful for future quantum annealing applications, but also

enables us to study some of the interesting physics that arise in the regime of non-perturbative coupling. Finally, we use a two-fluxmon device as a "mini-annealer," and demonstrate that annealing fast relative to thermalization timescales (faster than can be done on commercial quantum annealing devices) can significantly change the nature of the two-qubit annealing process.

6.1 Physical design and implementation of coupled qubits

The first intuition one might have for coupling two fluxmon qubits is to achieve a mutual inductance by routing the qubits' CPW segments parallel and close to one another for a certain length (with a coupler circuit in between to make the coupling tunable). However, such a scheme is limited to a single plane and, given the constraints of the fluxmon geometry, would likely limit circuit complexity. We would like a coupling geometry that is compatible with high connectivity graphs such as those in Fig. 1.4; i.e., small couplers at the "intersections" of long qubits. One possibility, given that we have airbridge crossovers at our disposal, is to use a scheme in which qubits hop over each other perpendicularly, enabling these types of high-connectivity graphs. If the qubit intersections are designed with sufficient symmetry, the currents flowing down one qubit's CPW will not couple flux into the CPW mode of the other qubit, and there would be no direct coupling between qubits. Coupling can then be induced via a tunable coupler circuit physically located

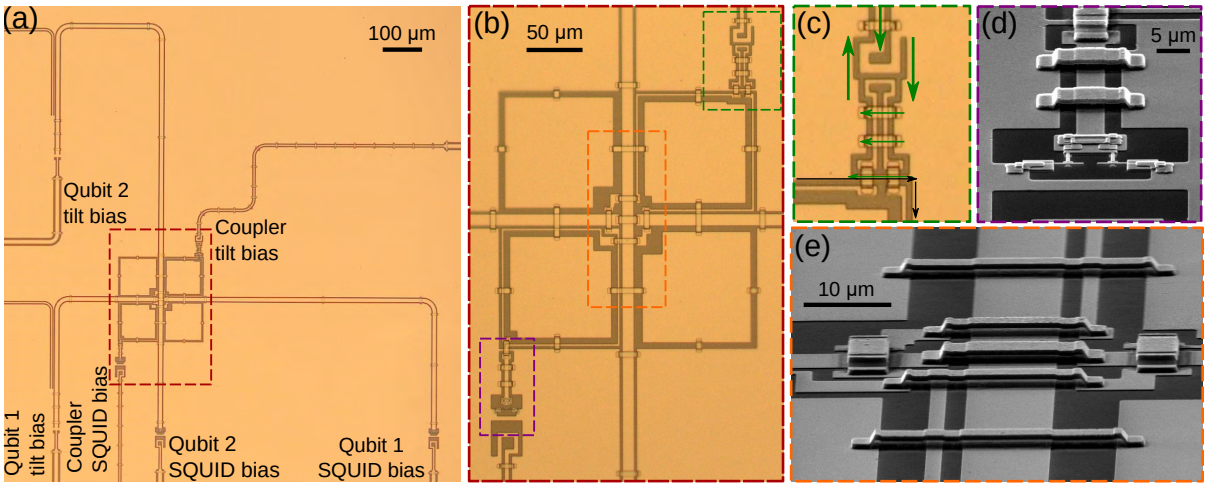


Figure 6.1: (a) Optical micrograph of two-qubit coupled fluxmon device on a silicon substrate. Qubit 1, qubit 2, and the coupler each have their own main (tilt) and DC SQUID biases. (b) Zoom-in of the coupler at the intersection of the two qubits. The horizontal qubit jumps over the vertical qubit via an airbridge crossover [67]. The coupler loop is a gradiometric figure-eight loop, jumping over each qubit twice to cover two diagonally opposite quadrants of their intersection. Ground plane pads in the middle of the coupler loop are connected to ground via airbridges that hop over the coupler loop. (c) Illustration of current flow in the coupler's tilt bias line. Airbridges help direct the flow of the bias return current (green) away from the small fraction of the qubit's ground plane current (black) to reduce crosstalk. (d) Angled SEM image of coupler's DC SQUID, through which copuler connects to ground. (e) Angled SEM image of the intersection point of the qubits and coupler, giving a better view of the airbridge hopovers.

near the intersection point of the two qubits, in one or more of the four quadrants defined by the intersection, as was schematically indicated by the blue couplers in Fig. 1.4(a).[1] We will use this scheme, as shown in the device photo in Fig. 6.1. The question that remains is how to actually implement the coupler.

The first way one might think of implementing the coupler is as a loop in one of the

[1]In reality, the perfect cancellation of stray direct inductive coupling between the qubits does not happen, but that is o.k. as long as this stray coupling is small enough, because then it can be cancelled by the tunable coupling mediated by the coupler. On the other hand, there will always be a stray amount of capacitive coupling between qubits, which for typical devices we simulate to be $1-2\,\mathrm{fF}$ (this capacitance is dominated by the proximity of the CPWs in the same plane, not by the capacitance of the vacuum-gap hop-over airbridge to the hopped-over qubit). This capacitive coupling will yield a stray YY coupling in the lab frame of $\sim 50\,\mathrm{MHz}$ that cannot be exactly cancelled by the tunable XX coupling, although it can be approximately cancelled in the rotating wave approximation (which holds much of the time), under which both types of couplings look like $XX+YY$.

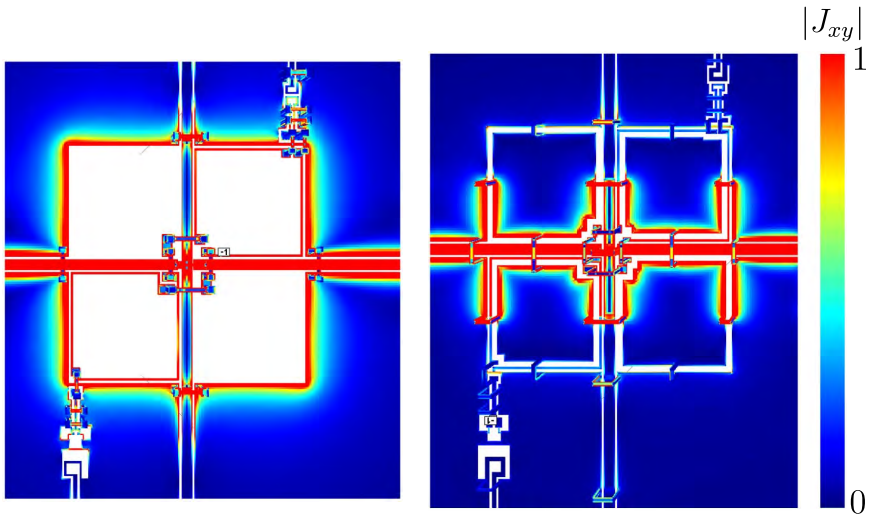


Figure 6.2: Sonnet simulation comparing the horizontal qubit's return current without and with ground plane pads plus crossovers. The latter allows a more localized return current.

quadrants, as was drawn schematically in Fig. 1.4(a). However, this design has several disadvantages, one of which is that it is difficult to bring in bias lines for both the coupler SQUID and tilt without there being substantial crosstalk. Instead, we have chosen to have the loop extend over two quadrants, which need to be diagonally opposite – otherwise, the induced qubit-qubit coupling would have opposite signs for the two quadrants, leading to zero coupling. In order to cover two quadrants, we utilize a handful of series airbridge crossovers, four for the coupler to jump over the qubits and itself, another two at the tilt bias to make the biasing easier, and another one near the coupler's SQUID, where the coupler connects to ground. We also fill in ground plane pads within the coupler loop, which are connected to ground through airbridges that hop over the coupler. This reduces the available coupling strength (essentially due to reduced effective flux pickup area), but due to alternate current paths greatly reduces control crosstalk and stray qubit-qubit coupling. This point is illustrated in Fig. 6.2, showing how with the ground

plane pads plus crossovers, the qubit return current can be more closely localized to the qubit since it does not need to flow around the coupler. Note that we also chose to cut out slots in the unused two quadrants as well. This symmetry helps to minimize stray qubit-qubit coupling.

The equivalent circuit diagram for our coupler system is exactly what we showed in Fig. 5.15 (or, ignoring coupler capacitance, Fig. 5.2), but with all junctions replaced by tunable junctions in the form of DC SQUIDs.

6.2 Calibration of coupled fluxmon device

Before we can explore qubit-qubit coupling, it is necessary to calibrate the device. An annoying but very important aspect of this is to calibrate bias line crosstalk. To compare the importance of crosstalk calibration to other superconducting qubit platforms, we note that the fluxmon has an almost 1000× larger loop size than the Xmon's to pick up stray flux, and in addition can have 1000× higher flux sensitivity $df_{10}/d\Phi$. Careful design can reduce bias line crosstalk substantially, but at least with the current architecture there will still be a moderate amount of residual crosstalk present. After calibrating the full crosstalk matrix, we go on to characterize qubit-qubit coupling and two-qubit annealing.

The crosstalk matrix at first looks daunting – even though we only have two qubits and one coupler, there are a lot of potential crosstalk terms. Fortunately, as we'll see, many of these elements are so small that they can be ignored (usually these are the ones with a DC SQUID as the pickup loop), but some (usually the ones with qubit tilt

as the pickup loop) must be precisely measured and calibrated. The table in Fig. 6.3 shows the full 12×6 inductance matrix containing both self-inductances and crosstalk parameters, with the row representing the source and column representing the pickup loop. The reason the matrix is not 12×12 is because we only consider the bias lines to be sources, not pickups. Conceptually, this matrix is easier to understand if one breaks it up into two 6×6 matrices, one with bias lines as sources and qubit loops as pickups (in this matrix, all entries are independent), and another with qubit loops as sources and qubit loops as pickups (this is a symmetric matrix).

6.2.1 Measurement of coupler $\rightarrow$ qubit flux crosstalk

We already discussed how to measure intra-qubit crosstalk in section 3.3.2 for single-qubit devices, so here we will talk about something new that arises due to the coupler, namely crosstalk from the coupler bias lines to the qubit. As we saw in the previous chapter, in addition to any linear bias line crosstalk there are also two intrinsic nonlinear but periodic sources of crosstalk from coupler to qubit: the coupler-induced flux and coupler-induced inductance shift. This slightly complicates the extraction of the bias line crosstalk, but we can use the fact that the nonlinear crosstalk is a periodic function of the coupler biases, allowing us to subtract out the linear geometric crosstalk background. This is simple to do assuming the coupler tilt bias has a full range larger than one period (one flux quantum).

The first way one might think of measuring the crosstalk is to measure the shift in

pickup / source	Q1 SQUID	Q1 Tilt	Coupler SQUID	Coupler	Q2 SQUID	Q2 Tilt
Q1 SQUID	L_{q1s}	$M_{\mathrm{q1s}\to\mathrm{q1t}}$	$M_{\mathrm{q1s}\to\mathrm{csq}}$	$M_{\mathrm{q1s}\to\mathrm{ct}}$	$M_{\mathrm{q1s}\to\mathrm{q2s}}$	$M_{\mathrm{q1s}\to\mathrm{q2t}}$
Q1 SQUID Bias	M_{q1s}	$M_{\mathrm{b1s}\to\mathrm{q1t}}$	$M_{\mathrm{b1s}\to\mathrm{csq}}$	$M_{\mathrm{b1s}\to\mathrm{ct}}$	$M_{\mathrm{b1s}\to\mathrm{q2s}}$	$M_{\mathrm{b1s}\to\mathrm{q2t}}$
Q1 Tilt	$M_{\mathrm{q1t}\to\mathrm{q1s}}$	L_{q1}	$M_{\mathrm{q1t}\to\mathrm{csq}}$	$M_{\mathrm{q1t}\to\mathrm{ct}}$	$M_{\mathrm{q1t}\to\mathrm{q2s}}$	$M_{\mathrm{q1t}\to\mathrm{q2t}}$
Q1 Tilt Bias	$M_{\mathrm{b1t}\to\mathrm{q1s}}$	M_{q1t}	$M_{\mathrm{b1t}\to\mathrm{csq}}$	$M_{\mathrm{b1t}\to\mathrm{ct}}$	$M_{\mathrm{b1t}\to\mathrm{q2s}}$	$M_{\mathrm{b1t}\to\mathrm{q2t}}$
Coupler SQUID	$M_{\mathrm{csq}\to\mathrm{q1s}}$	$M_{\mathrm{csq}\to\mathrm{q1t}}$	L_{csq}	$M_{\mathrm{csq}\to\mathrm{ct}}$	$M_{\mathrm{csq}\to\mathrm{q2s}}$	$M_{\mathrm{csq}\to\mathrm{q2t}}$
Coupler SQUID Bias	$M_{\mathrm{bcsq}\to\mathrm{q1s}}$	$M_{\mathrm{bcsq}\to\mathrm{q1t}}$	M_{csq}	$M_{\mathrm{bcsq}\to\mathrm{ct}}$	$M_{\mathrm{bcsq}\to\mathrm{q2s}}$	$M_{\mathrm{bcsq}\to\mathrm{q2t}}$
Coupler	$M_{\mathrm{ct}\to\mathrm{q1s}}$	$M_{\mathrm{ct}\to\mathrm{q1t}}$	$M_{\mathrm{ct}\to\mathrm{csq}}$	L_{c}	$M_{\mathrm{ct}\to\mathrm{q2s}}$	$M_{\mathrm{ct}\to\mathrm{q2t}}$
Coupler Bias	$M_{\mathrm{bct}\to\mathrm{q1s}}$	$M_{\mathrm{bct}\to\mathrm{q1t}}$	$M_{\mathrm{bct}\to\mathrm{csq}}$	M_{ct}	$M_{\mathrm{bct}\to\mathrm{q2s}}$	$M_{\mathrm{bct}\to\mathrm{q2t}}$
Q2 SQUID	$M_{\mathrm{q2s}\to\mathrm{q1s}}$	$M_{\mathrm{q2s}\to\mathrm{q1t}}$	$M_{\mathrm{q2s}\to\mathrm{csq}}$	$M_{\mathrm{q2s}\to\mathrm{ct}}$	L_{q2s}	$M_{\mathrm{q2s}\to\mathrm{q2t}}$
Q2 SQUID Bias	$M_{\mathrm{b2s}\to\mathrm{q1s}}$	$M_{\mathrm{b2s}\to\mathrm{q1t}}$	$M_{\mathrm{b2s}\to\mathrm{csq}}$	$M_{\mathrm{b2s}\to\mathrm{ct}}$	M_{q2s}	$M_{\mathrm{b2s}\to\mathrm{q2t}}$
Q2 Tilt	$M_{\mathrm{q2t}\to\mathrm{q1s}}$	$M_{\mathrm{q2t}\to\mathrm{q1t}}$	$M_{\mathrm{q2t}\to\mathrm{csq}}$	$M_{\mathrm{q2t}\to\mathrm{ct}}$	$M_{\mathrm{q2t}\to\mathrm{q2s}}$	L_{q2}
Q2 Tilt Bias	$M_{\mathrm{b2t}\to\mathrm{q1s}}$	$M_{\mathrm{b2t}\to\mathrm{q1t}}$	$M_{\mathrm{b2t}\to\mathrm{csq}}$	$M_{\mathrm{b2t}\to\mathrm{ct}}$	$M_{\mathrm{b2t}\to\mathrm{q2s}}$	M_{q2t}

Figure 6.3: Full inductance matrix between qubit loops and bias lines. There are 72 total elements, 57 of which are independent. 15 are desired design parameters (green), 12 of which are independent. Then there are three unwanted self-inductances for the DC SQUIDs. The rest (black) are undesired crosstalk parameters, 30 of which represent bias line crosstalk, which should be calibrated before the rest of the circuit parameters can be precisely extracted. Three of the bias line crosstalk parameters (red) also have a nonlinear contribution arising from intra-qubit/intra-coupler junction asymmetry.

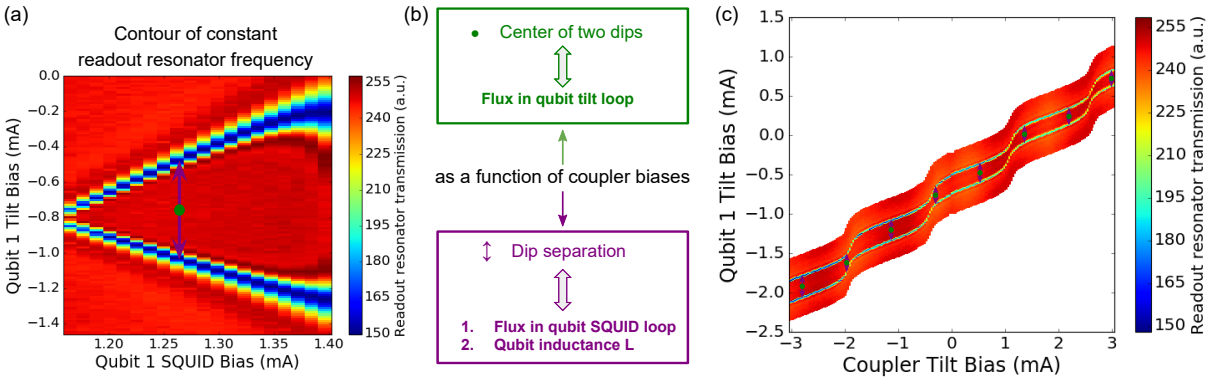


Figure 6.4: Illustration of the "double dip" experiment, which involves tracking two dips in readout resonator transmission at a fixed probe frequency versus qubit tilt. (a) Calibration dataset vs. intra-qubit biases, yielding contour of constant resonator frequency. (b) Summary of the information learned by measuring a vertical cut of the plot in (a) versus coupler biases: the center of the two dips gives the tilt flux induced in the qubit, and the separation of the dips is a function of both the flux induced in the qubit SQUID from linear crosstalk and the coupler-induced inductance shift.

qubit f_{10} as a function of coupler biases. However, this is quite complicated because when we wiggle one of the coupler's biases, the qubit frequency will change for four different reasons: geometric crosstalk from coupler bias to qubit tilt loop, geometric crosstalk from coupler bias to qubit SQUID loop, a nonlinear flux bias from the coupler's persistent current to the qubit tilt loop, and a coupler-induced shift in qubit inductance. Furthermore, if there are defects in the spectrum, or when the qubit frequency goes out of the easily measurable range of $2 - 7\,\mathrm{GHz}$, the automated tracking in the data scan may lose track of the qubit.

Here we instead use a measurement that is able to robustly handle the above concerns and gives us the crosstalk from both of the coupler bias lines to both the tilt and SQUID loops of one of the qubits in a single type of dataset (i.e., the four elements $M_{\mathrm{bct}\to\mathrm{q1t}}$, $M_{\mathrm{bct}\to\mathrm{q1s}}$, $M_{\mathrm{bcsq}\to\mathrm{q1t}}$, and $M_{\mathrm{bcsq}\to\mathrm{q1s}}$). The way it works is as follows. First, we assume that

the intra-qubit crosstalk for the qubit in question is small or has already been calibrated away. We also tune the other qubit away to $\beta \lesssim 0$ ($f_{10} \gtrsim 20\,\mathrm{GHz}$). Then, for the qubit in question, we perform the same readout resonator response scan vs. flux biases as we did in Fig. 3.14(b), and zoom in on one of the "hot spots" where we can see a detailed contour of constant readout resonator frequency. An example of this data is shown in Fig. 6.4(a), which serves as a calibration dataset for the crosstalk measurement. For each qubit SQUID bias, taking a vertical slice of this dataset reveals two dips in resonator transmission. These dips tell us two pieces of information: the midpoint of the two dips tells us where zero qubit tilt is, and the separation between the dips is a function of both the flux in the qubit's DC SQUID and the coupler-induced inductance shift ΔL_q of the qubit. If we then pick a fixed qubit SQUID bias and take a vertical cut as a function of the coupler tilt bias, we will obtain the data shown in Fig. 6.4(c). Here, as a function of coupler tilt bias, both the midpoint of the two dips and the separation between the two dips changes, each according to a periodic function superposed on top of a linear drift. By fitting to the periodic function plus linear drift, we can then isolate the geometric crosstalk from coupler tilt bias to both qubit loops.

To make this clear, we first look at the extracted midpoint of the two dips as a function of coupler tilt bias, plotted in Fig. 6.5 for different values of coupler SQUID bias. By fitting these curves to the functional form

$$y(x) = ax + b + c\sin_{\beta_c}(2\pi[x - x_0]/d), \tag{6.1}$$

we can isolate the linear drift contribution from the slope a, giving us $M_{\mathrm{bct}\to\mathrm{qt}}$. In Fig.

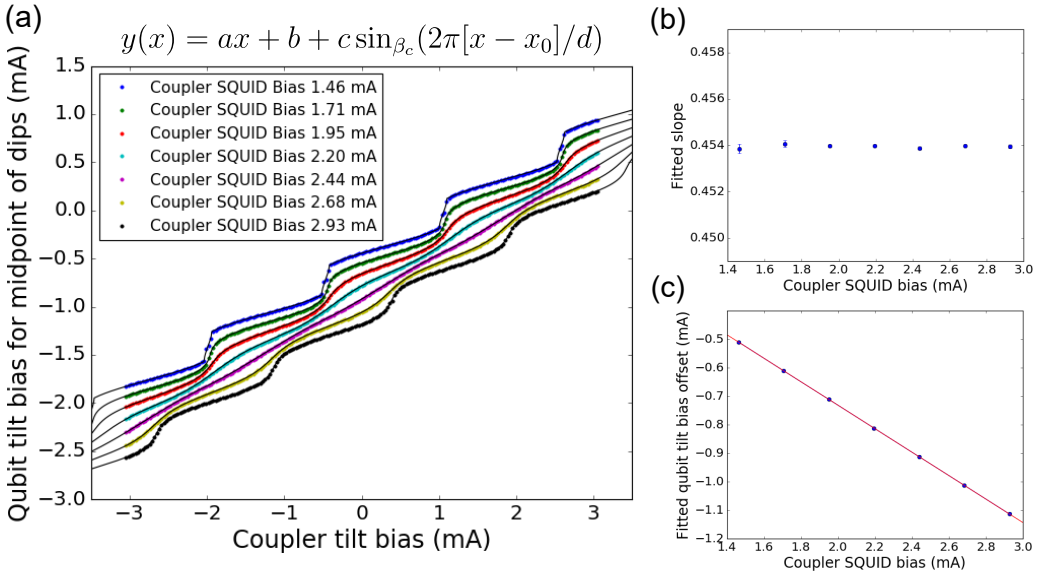


Figure 6.5: (a) Extracted midpoint of resonator transmission dips versus coupler tilt bias, for different values of coupler SQUID bias. Each curve is fit to the functional form of equation (6.1), allowing us to isolate the linear crosstalk from coupler biases to the qubit tilt loop. (b) Extracted slope a for each curve, allowing the extraction of $M_{\mathrm{bct}\to\mathrm{qt}}$. (c) Linear fit to the extracted vertical offset b as a function of coupler SQUID bias, giving $M_{\mathrm{bcsq}\to\mathrm{qt}}$.

6.5(b), we show the extracted slope versus coupler SQUID bias, each point corresponding to one of the curves in Fig. 6.5(a). All curves give the same slope to significantly less than one part in 1000, showing that this measurement is robust. In addition, by looking at the extracted vertical offsets b as a function of coupler SQUID bias, we can fit a slope to b versus coupler SQUID bias to give us $M_{\mathrm{bcsq}\to\mathrm{qt}}$, as shown in Fig. 6.5(c). We note that the data shown in Fig. 6.5 was from our first coupler sample, which had significant crosstalk due to a missing airbridge, for the sake of illustration ($M_{\mathrm{bct}\to\mathrm{qt}} = -0.198\,\mathrm{pH}$, $M_{\mathrm{bcsq}\to\mathrm{qt}} = 0.175\,\mathrm{pH}$). Subsequent designs with proper airbridge symmetry significantly improved these numbers (see Fig. 6.9).

Next, from the same datasets [i.e., Fig. 6.4(c)], we instead look at the separation of the two dips. This is plotted in Fig. 6.6 versus the same coupler tilt biases, for the

$$y(x) = ax + b + \frac{c}{1 - \frac{1}{\beta_c \cos\left(2\pi[x-x_0]/d + \beta_c \sin_{\beta_c}(2\pi[x-x_0]/d)\right)}}$$

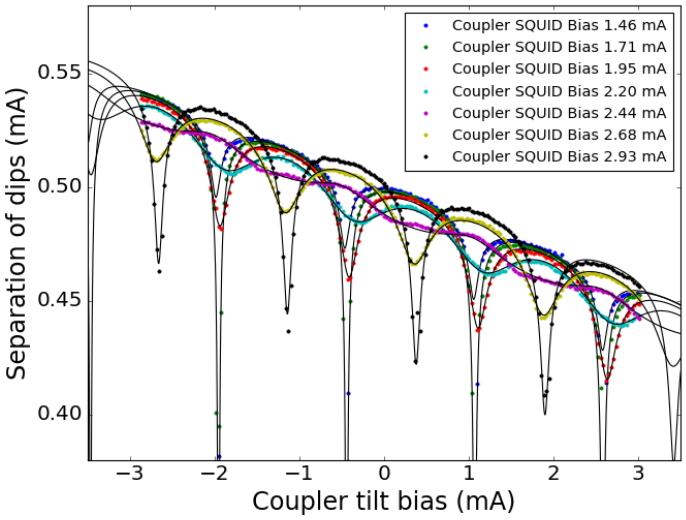


Figure 6.6: Extracted separation of resonator transmission dips versus coupler tilt bias, for different values of coupler SQUID bias. Each curve is fit to the functional form of equation (6.2), allowing us to isolate the linear crosstalk from coupler biases to the qubit SQUID loop.

same various coupler SQUID biases. Since the dip separation is a linear function of qubit SQUID flux and also of qubit inductance (for small inductance shifts), we can fit these curves to the functional form

$$y(x) = ax + b + \frac{c}{1 - \frac{1}{\beta_c \cos\left(2\pi[x-x_0]/d + \beta_c \sin_{\beta_c}(2\pi[x-x_0]/d)\right)}}, \tag{6.2}$$

representing both linear crosstalk from coupler bias to qubit SQUID loop as well as the nonlinear periodic effect of the coupler-induced inductance shift. Note that the fits are not perfect when the coupler induces a very large inductance shift on the qubit, because the linear theory breaks down somewhat, but this does not matter for the purposes of this measurement, as we are only trying to extract the linear background from the nonlinear periodic contribution. From the linear slope a, we can extract $M_{\mathrm{bct}\to\mathrm{qs}}$, and from the vertical offset b versus coupler SQUID bias, we can extract $M_{\mathrm{bcsq}\to\mathrm{qs}}$.

6.2.2 Measurement of intra-coupler geometric crosstalk

Similarly to how each of the qubits has a 2×2 intra-qubit crosstalk matrix (3.40), the coupler also has an intra-coupler crosstalk matrix

$$\begin{pmatrix} \Phi_{\text{csq}} \\ \Phi_{\text{ct}} \end{pmatrix} = \begin{pmatrix} M_{\text{bcsq}\to\text{csq}} & M_{\text{bct}\to\text{csq}} \\ M_{\text{bcsq}\to\text{ct}} & M_{\text{bct}\to\text{ct}} \end{pmatrix} \begin{pmatrix} I_{\text{bcsq}} \\ I_{\text{bct}} \end{pmatrix} + \begin{pmatrix} \Phi_{\text{csq}}^{\text{offset}} \\ \Phi_{\text{ct}}^{\text{offset}} \end{pmatrix}. \tag{6.3}$$

Ideally, we could calibrate this matrix using the same "fireball" scan we used for the intra-qubit crosstalk (Fig. 3.14). However, the coupler does not have its own dedicated readout resonator. Nevertheless, it turns out it is still possible to do a fireball scan using one of the qubits' readout resonators. The reason for this is that there will be a second-order coupling between the coupler and the readout resonator mediated by the qubit, as illustrated in Fig. 6.7(a). This second-order interaction is beyond just a direct stray geometric mutual inductance between the coupler and readout resonator, which we can bound to be less than 0.1 pH (which as shown in Fig. 6.7(c) would not be enough to explain the observed dispersive shift of the resonator due to the coupler). Explicitly, if g_{qc} is the transverse coupling strength between the qubit and coupler, and g_{qr} is the transverse coupling strength between the qubit and its readout resonator, then the coupler and resonator will be coupled via the effective transverse coupling strength [204]

$$g_{cr}^{\text{eff}} \approx \frac{1}{2} g_{qc} g_{qr} \left(\frac{1}{\omega_c - \omega_q} + \frac{1}{\omega_r - \omega_q} \right). \tag{6.4}$$

This formula can be obtained by adiabatically eliminating the two first-order interaction terms in the system Hamiltonian, and can be interpreted as virtual photon exchange.

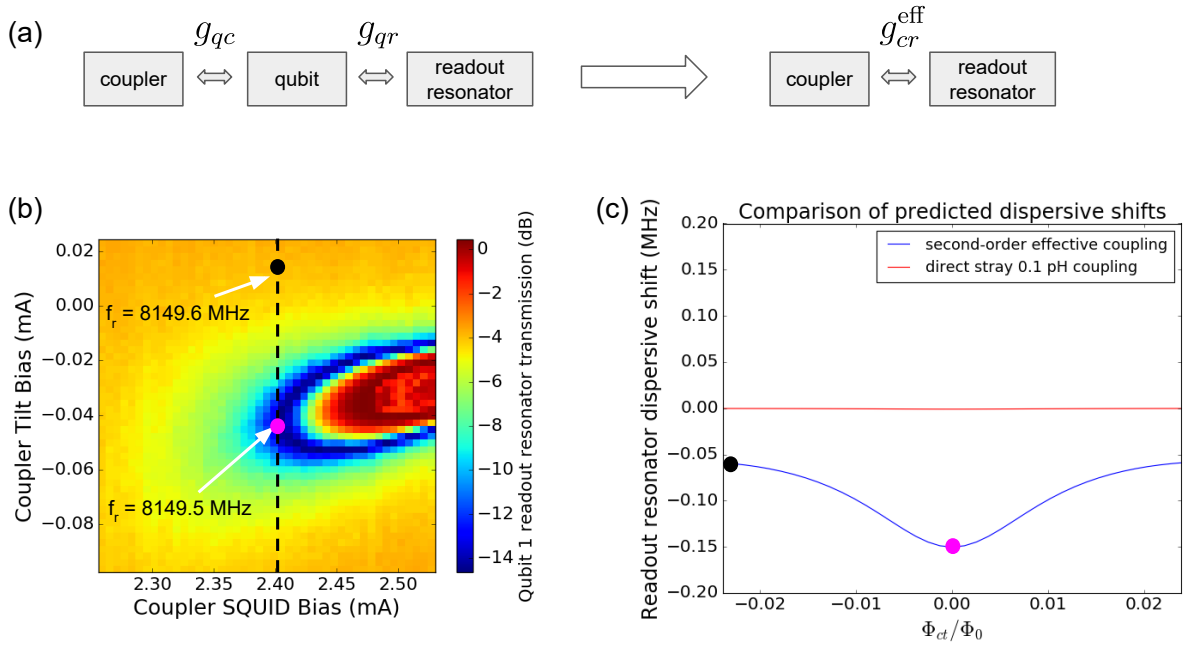


Figure 6.7: (a) Illustration of "virtual" second-order interaction between coupler and readout resonator mediated by a qubit. (b) Zoom-in on measured readout resonator response vs. coupler biases, with dispersively shifted resonator frequency at two points indicated. (c) Theoretical prediction for resonator dispersive shift due to the coupler as a function of coupler tilt bias, for a coupler gap of 10 GHz. The predicted dispersive shift matches the data.

As shown in Fig. 6.7(b) and (c), this formula fairly accurately predicts the dispersive shift of the resonator in response to the coupler. We can then use this second-order dispersive shift to do a "fireball" scan for the coupler, as shown in Fig. 6.8. By mapping the theoretical periodicity of the resonator response to theory, we can extract the intra-coupler bias line crosstalk matrix (6.3). Note that during this scan, the qubit whose readout resonator is being used should be biased very far from the resonator, usually at $\beta = -\beta_{\max}$, so that the coupler dominates the response.

That concludes the discussion of bias line crosstalk calibration. There are other types of bias line crosstalk whose calibration data we do not bother to show. One is qubit to qubit crosstalk, which can be measured in the same way as coupler to qubit crosstalk

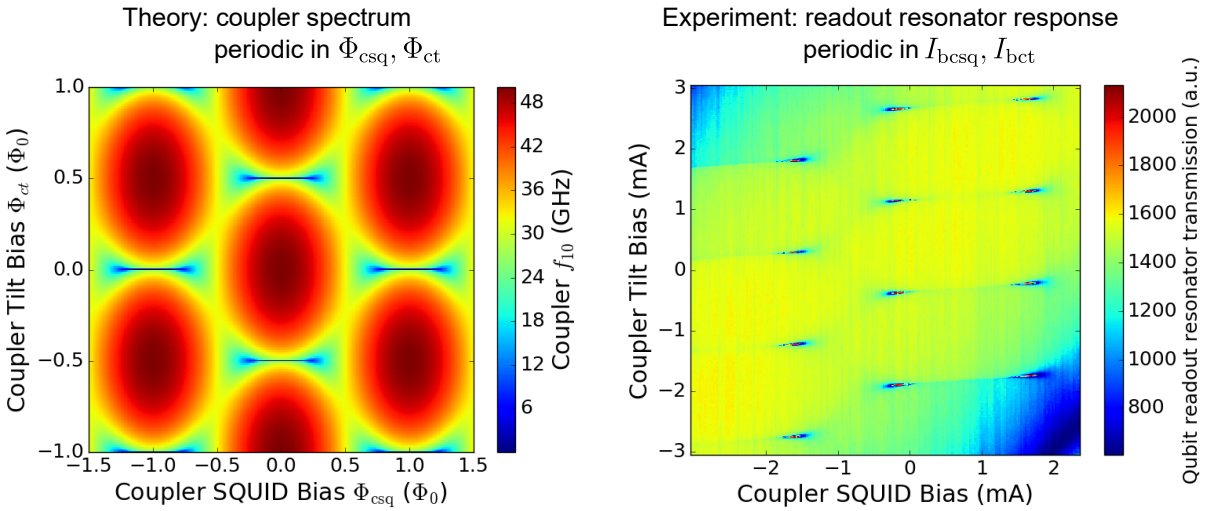
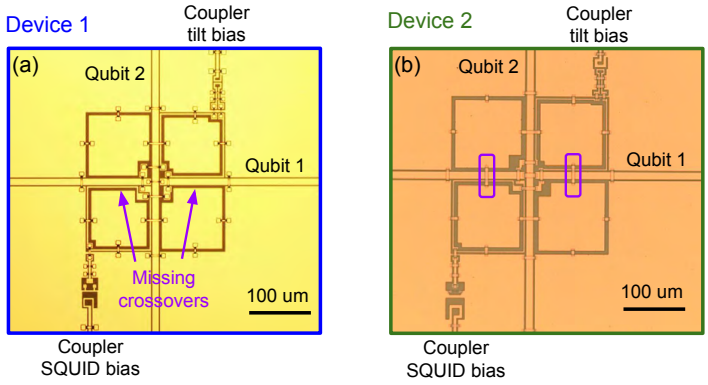


Figure 6.8: (a) Theoretical plot showing periodicity of coupler spectrum with respect to its flux biases, which is to be mapped to experimental data. Note the similarity to Fig. 3.14(a), but with generally higher frequencies for the coupler. (b) Experimental data from the coupler fireball scan. The qubit's readout resonator is probed at a fixed frequency (corresponding to roughly 0.1 MHz below f_r at zero qubit flux bias). Extracted parameters are $M_{ct} = 1.36\,\text{pH}$, $M_{\rm csq} = 0.635\,\text{pH}$, $M_c' = -0.103\,\text{pH}$, $M_c'' = 0.0\,\text{pH}$.



(c)

pickup / source	Q1 Tilt	Q2 Tilt	Coupler Tilt
Coupler SQUID Bias	0.175 pH 0.061 pH	0.036 pH 0.019 pH	-0.103 pH 0.011 pH
Coupler Tilt Bias	-0.198 pH -0.060 pH	0.107 pH -0.052 pH	

Figure 6.9: Comparison of an important subset of the crosstalk matrix with/without missing bridges. Crosstalk from coupler biases is drastically improved for the device with proper crossover symmetry. In addition, the direct stray coupling $M_{\rm q1t\to q2t}$ between qubits went down from $\sim$3 pH to $< 0.5\,\text{pH}$.

	Q1 SQUID	Q1 Tilt	Coupler SQUID	Coupler	Q2 SQUID	Q2 Tilt
Q1 SQUID Bias	.917 pH (design .85)	-.241 pH (design .05)	-.002 pH	.125 pH	-.003 pH	-.133 pH (design .04)
Q1 Tilt Bias	.0003 pH	.425 pH (design .42)	.001 pH	-.040 pH	-.003 pH	.001 pH
Coupler SQUID Bias	.001 pH	.175 pH (design .18)	.635 pH (design .66)	-.103 pH (design -.17)	.004 pH	.036 pH
Coupler Bias	-.003 pH	-.198 pH (design .04)	0.0000 pH	1.360 pH (design 1.35)	-.018 pH	.107 pH (design .02)
Q2 SQUID Bias	.002 pH	-.029 pH	-.0006 pH	.024 pH	1.762 pH (design 1.9)	-.136 pH (design .05)
Q2 Tilt Bias	-.004 pH	-.057 pH	.0004 pH	-.005 pH	-.009 pH	.446 pH (design .38)

Figure 6.10: Measured bias line crosstalk matrix for the first coupled fluxmon sample. Diagonal (desired) elements come within 0.1 pH of the Sonnet simulations during the design stage, but the off-diagonal (unwanted) elements can sometimes be as much as 0.2 pH different from prediction.

using the double dip scan, by putting the coupler at $\beta_c \lesssim 0$ and varying the flux biases of the other qubit. The other is qubit to coupler crosstalk. For this type of crosstalk, we can't use an analogous double dip measurement as a function of qubit bias to obtain the crosstalk to the coupler, because the effect of the qubit on the readout resonator is much greater than the small effect of the coupler on on the resonator. Instead, we can do something a little less accurate but still good enough, namely, we use the known qubit periodicity and jump by this known periodicity (to null out any effects from nonlinear flux crosstalk from qubit's persistent current to the coupler), and then measure the discrete movement of the coupler feature [such as the feature shown in Fig. 6.7(b)]. Finally, there is also the issue of intra-coupler nonlinear crosstalk due to coupler junction asymmetry. A way to measure this will be discussed shortly.

In Fig. 6.10 we plot the full measured bias line crosstalk matrix. For the rest of this

chapter, we will assume that this matrix has been corrected for, i.e., from now on we will use new coordinates such that there is no bias line crosstalk (to within the accuracy of our calibrations). We note that the crosstalk involving the coupler bias lines was greatly improved upon adding two airbridge crossovers that were accidentally omitted in our first device, as illustrated in Fig. 6.9.

6.2.3 Extraction of coupler parameters

A lot can be deduced about the coupler simply by precisely measuring the nonlinear flux bias it induces on the qubit(s). A more sensitive way to measure the coupler-induced flux bias on a qubit is to use single-qubit annealing as a flux detection scheme. This way, we are limited by the linewidth of the S curve rather than the relatively broad readout resonator response. By performing the S curve experiment as a function of coupler bias, we can detect how the degenerate point of the S curve moves. In practice, we don't actually need to measure over a whole range of qubit pre-tilt fluxes to trace out the whole S curve, but rather we can perform a binary search over pre-tilt biases until we converge at a probability of 0.5 to end up in each well. In practice, we can converge to zero tilt to within $\sim 10\,\mu\Phi_0$ with at most 10 queries.

An example dataset of induced qubit flux vs. coupler biases is shown in Fig. 6.11(a). For this data, the other qubit was detuned away to $\beta \lesssim 0$. Note the striking similarity to the theoretical prediction of Fig. 5.5(b). Taking a vertical cut corresponds to the $\sin_{\beta_c}$ function describing the nonlinear flux crosstalk for a given β_c. Three such cuts are

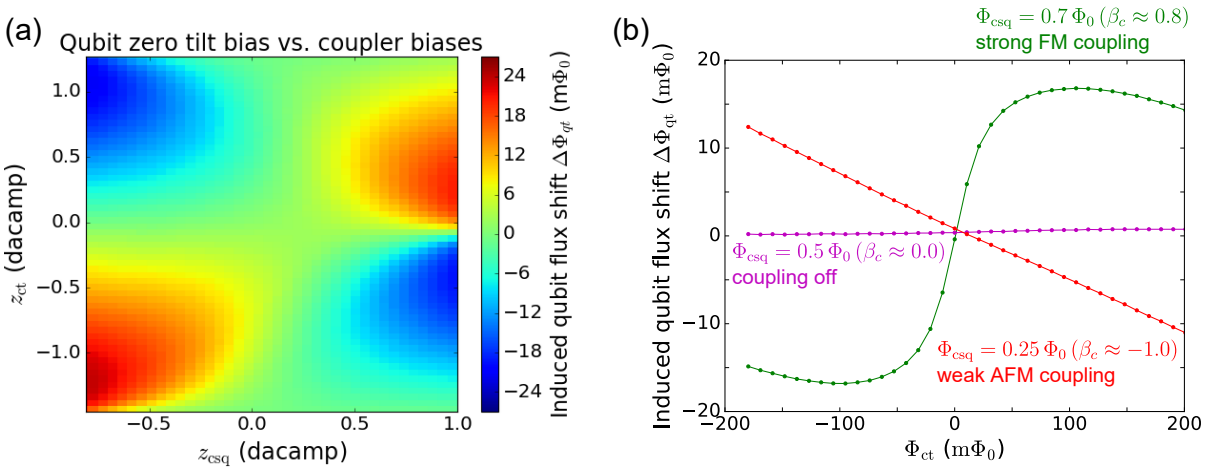


Figure 6.11: (a) Experimental dataset of qubit flux shift vs. coupler biases. Each datapoint is the result of a binary search of the qubit's S-curve for zero qubit tilt. (b) Three vertical cuts illustrating three types of coupler susceptibilities, namely those corresponding to strong ferromagnetic, off, and weak antiferromagnetic coupling.

shown in Fig. 6.11(b). The slope of these curves is a measure of coupler susceptibility, since it is a measure of the current flowing in the coupler in response to the coupler's flux bias. At zero coupler tilt, the green curve should correspond to strong ferromagnetic (FM) coupling between qubits. The flat magenta curve indicates that the coupling is nominally off. Finally, the red curve should correspond to weak antiferromagnetic (AFM) coupling. Note that this slope doesn't actually tell us the absolute coupling between qubits, because there may be stray direct coupling between the qubits not mediated by the coupler. We will look at measuring the absolute qubit-qubit coupling in the next section via two methods, spectroscopy and "sequential annealing" [197].

Upon close inspection, we observe that the three curves of Fig. 6.11(b) [and the rest of the vertical cuts of the data in Fig. 6.11(a)] do not all intersect at exactly the same point, which would nominally be the point (0, 0). It appears as though when we change the coupler SQUID bias, it induces a shift in the coupler tilt bias (this is not from bias

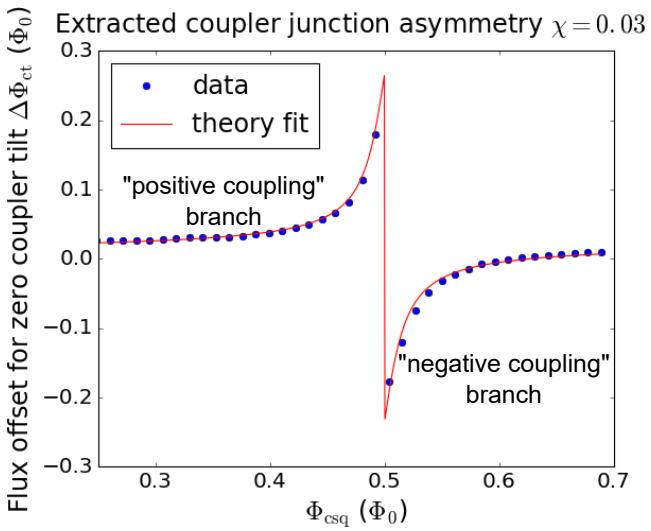


Figure 6.12: Measured shift in coupler tilt as a function of coupler SQUID bias, and a fit to theory for nonlinear crosstalk due to coupler junction asymmetry.

line crosstalk, which we have already calibrated). To understand what is going on more quantitatively, we can look at how much we must shift the coupler tilt bias for all of the curves such that they all end up intersecting at zero tilt. The result is shown in Fig. 6.12, including a fit to equation (2.58) for junction asymmetry in the coupler, indicating a modest junction asymmetry of $\chi \approx 0.03$ explains the data.

We can also use these flux shift curves to extract other coupler parameters. In Fig. 6.13 we show the result of measuring the coupler-induced flux shifts for both qubits and fitting all the datasets at the same time with only four free parameters: M_{ca}, M_{cb}, L_c, and

	Extracted	Designed
M_{ca}	92 pH	88 pH
M_{cb}	74 pH	71 pH
L_c	577 pH	614 pH
$I_c^{(c),\max}$	0.85 μA	0.79 μA

Table 6.1: Expected (designed and simulated) vs. extracted coupler parameters.

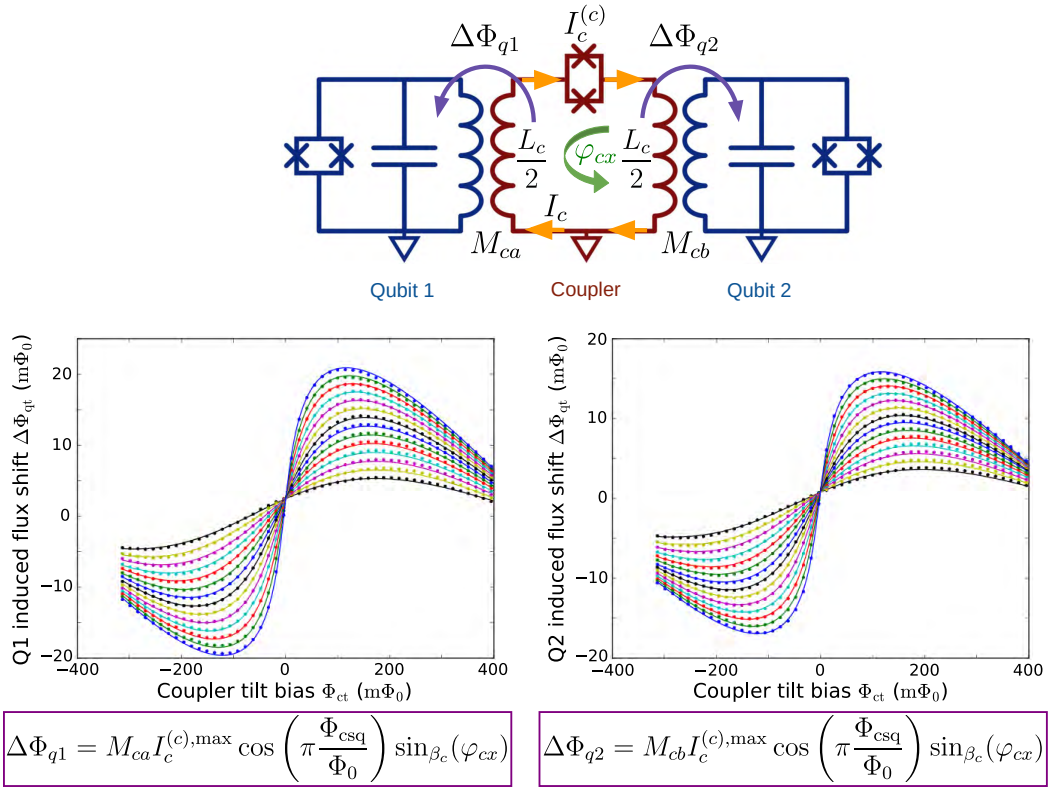


Figure 6.13: Measured coupler-induced flux shifts in both qubits along with simultaneous fit to all the data for the coupler parameters M_{ca}, M_{cb}, L_c, and $I_c^{(c),\max}$.

$I_c^{(c),\max}$. In this figure, the lines are not connecting the dots, but are the results of this fitting. The fitted coupler parameters are displayed in table 6.1, along with the design parameters from Sonnet simulations and a room temperature resistance measurement. Remarkably, the fitted parameters are quite close to the physically expected values. We note that getting such a good match required us to have already precisely calibrated the bias line mutual inductances and crosstalk elements. We also note that if we only fit the data for one of the qubits instead of two, the fitting is much more likely to get stuck in an unphysical local minimum for the coupler parameters.

6.3 Measurement of ultra-strong tunable coupling

We are almost ready to measure qubit-qubit coupling mediated by the coupler. In the following, we will see a few different ways of measure this coupling all the way into the ultra-strong regime. The regime of ultra-strong coupling can be roughly defined as when the qubit-qubit coupling strength J is a substantial fraction of the qubit frequencies themselves, such that certain typically valid approximations, such as the rotating wave approximation of the Jaynes-Cummings Hamiltonian, are no longer valid. As already mentioned in the previous chapter, coupling between flux qubits is not new. There are even some results that can claim to be ultrastrong coupling, although these results are either not tunable or not between two qubits, or did not allow a spectroscopic measurement. For example, Ref. [205] demonstrates moderate (12% of qubit frequency) ultra-strong coupling between a qubit and transmission line resonator. Ref. [206] achieved ultra-strong coupling between a resonator and qubit using a galvanically shared segment of kinetic inductance. Very recently, semi-tunable ultrastrong coupling was achieved between a qubit and oscillator [207], again using a galvanic connection. D-Wave's qubits should have similar tunable coupling strengths to ours, but the coupling strength has not been directly measured via spectroscopy. Here, we will directly observe tunable ultrastrong coupling between qubits, without using a galvanic connection. Furthermore, our implementation will allow simultaneous coupling between one qubit and up to 10 or more other qubits at once.

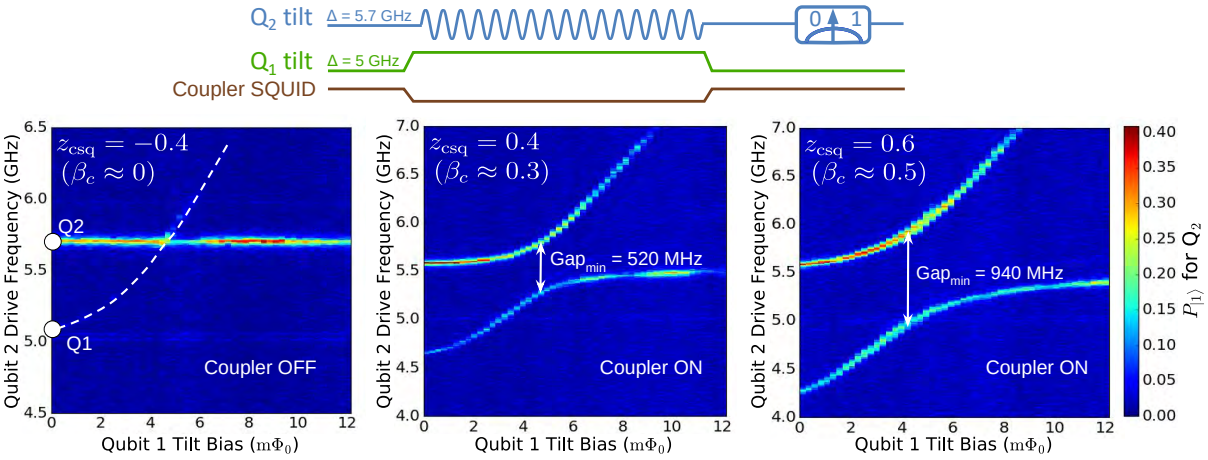


Figure 6.14: Simple spectroscopy versus tilt bias for various coupler biases. Here, the magnitude of the avoided level crossing away from zero tilt does not directly give the qubit-qubit coupling strength J. (a) Coupling off. (b) Small coupling. (c) Moderately large coupling.

6.3.1 Spectroscopy versus tilt bias

A simple way one might think of spectroscopically measuring coupling between two flux qubits is as follows. Assuming Δ is large enough to measure via spectroscopy, park one qubit (let's say Q_2) at zero tilt. Next, put the other qubit (Q_1) at zero tilt with a somewhat smaller Δ. Then, step Q_1's tilt bias while sweeping a spectroscopy tone over frequency, until an anticrossing bewteen the two energy transitions corresponding to the $|01\rangle$ and $|10\rangle$ states becomes visible. This experiment is implemented in Fig. 6.14.

A limitation of this method is that because the avoided level crossing in general occurs away from zero tilt, the coupling is no longer purely transverse to the qubits' energy bases, and so the coupling strength J is not simply equal to half the avoided level crossing as usual. Instead, the Hamiltonian is really (in the parity basis)

$$H = \frac{1}{2}\left(\Delta_1 ZI + \Delta_2 IZ + \varepsilon_1 XI + \varepsilon_2 IX\right) + JXX, \tag{6.5}$$

from which it can be shown, with a little algebra, that

$$J = \frac{\Delta E_{\min}}{2}\sqrt{\frac{\Delta_2^2}{\Delta_1^2 - \left(\frac{\Delta E_{\min}}{2}\right)^2}} \geq \frac{\Delta E_{\min}}{2}, \tag{6.6}$$

where $\Delta E_{\min}$ is the minimum energy gap of the avoided level crossing. This means that unless other parameters, such as the coupler-induced inductance shift, are calibrated, $\Delta E_{\min}/2$ will underestimate the actual coupling strength J. For an idea of how important this rotation of basis effect is, we note that for the third dataset of Fig. 6.14 the actual coupling strength is roughly forty percent larger than 470 MHz (half the minimum gap). Furthermore, this technique won't extend well to the double-well regime, since there even a little bit away from zero tilt the energy basis will be completely rotated into alignment with the flux basis, meaning that the coupling is longitudinal in this basis and there will be no avoided crossing (although there will instead by a frequency shift in one qubit dependent on the state of the other qubit, but as we know in the double-well regime it would be very difficult to prepare the two different states of one qubit away from zero tilt). We would like a measurement that enables a more direct and confident extraction of J. In the next two sections, we will consider what measurements will allow a more clear cut extraction of J when the qubits are in the single-well or double-well regime.

6.3.2 Spectroscopy at zero tilt: J from level splitting

The solution to indirectness of the previous method is to arrange for the avoided crossing to happen without the $\varepsilon_1 XI$ and $\varepsilon_2 IX$ terms, i.e., with both qubits at zero tilt. This is directly analogous to how the coupling between two transmons can be measured via an avoided crossing, since transmons have a symmetric potential; i.e., they have no local X field (in the energy basis) in the lab frame. In this case, the Hamiltonian is simply

$$H = \frac{1}{2}\Delta_1 ZI + \frac{1}{2}\Delta_2 IZ + JXX \tag{6.7}$$

$$= \begin{pmatrix} \frac{1}{2}(\Delta_1+\Delta_2) & 0 & 0 & J \\ 0 & \frac{1}{2}(\Delta_1-\Delta_2) & J & 0 \\ 0 & J & \frac{1}{2}(-\Delta_1+\Delta_2) & 0 \\ J & 0 & 0 & -\frac{1}{2}(\Delta_1+\Delta_2) \end{pmatrix}. \tag{6.8}$$

With this Hamiltonian, if we sweep over the qubit frequencies (which now can only be changed via Δ_1/Δ_2), the minimum gap of the avoided level crossing will then occur when $\Delta_1 = \Delta_2$, and will be equal to $2J$. Explicitly,

$$J = \frac{\Delta E_{\min}}{2}. \tag{6.9}$$

This is true even if we keep the counter-rotating terms of the matrix (6.8) that are typically discarded in the rotating wave approximation.[2] This approximation amounts to removing the two entries that connect the $|00\rangle$ and $|11\rangle$ states, as they don't conserve

[2] A straightforward diagonalization shows that $E_{10} = -J + \sqrt{\Delta^2 + J^2}$, $E_{20} = J + \sqrt{\Delta^2 + J^2}$ (the counter-rotating terms shift the ground state energy level relative to the two higher levels).

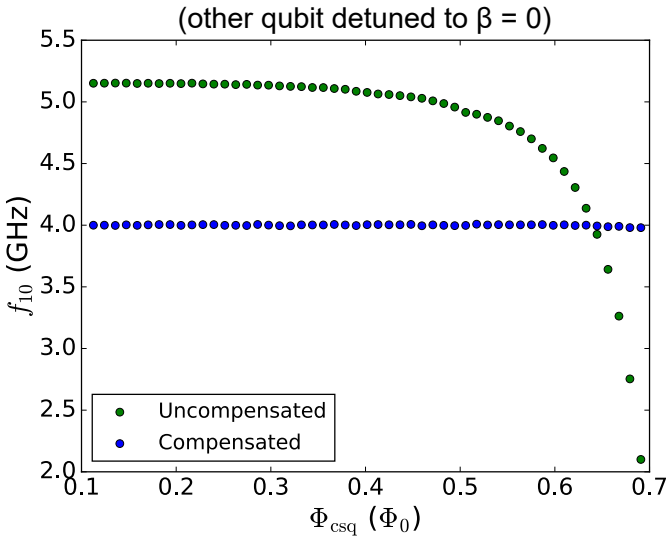


Figure 6.15: Measurement and compensation of qubit frequency shift due to coupler-induced inductance shift.

excitation number. However, as we'll see when the coupling is very strong, in order to predict the center of the avoided level crossing and not just the difference between the two levels, we will need to take these counter-rotating terms into account.

Experimentally, the question becomes how to arrange for $\Delta_1 = \Delta_2$ at zero tilt with the coupling turned on. If we operate the coupler at zero tilt, we can avoid any coupler-induced flux bias in the qubits. But we can't avoid the coupler-induced inductance shift in the qubits. This means that every time we change the coupler bias, we must compensate the qubits' SQUID biases to keep $\Delta_1 = \Delta_2$ constant. We therefore require a calibration dataset before we can measure the coupling, for example the dataset shown in Fig. 6.15. Here, the green points show the measured qubit frequency as a function of coupler SQUID bias, in accordance with the expected coupler-induced inductance shift. During this scan, the other qubit is held far detuned at $\beta = 0$ ($f_{10} \approx 20\,\mathrm{GHz}$) so it does not influence the measurement of the qubit in question. Note that, as expected, the

coupler induces a positive frequency shift for $\Phi_{\rm csq} < 0.5$ ($\beta_c < 0$), and a larger negative frequency shift for $\Phi_{\rm csq} > 0.5$ ($\beta_c > 0$). We can then achieve the desired constant qubit Δ of 4 GHz by compensating for the frequency shift as a function of coupler bias. This is done by adjusting the qubit's SQUID bias until Δ reaches its desired value, while also keeping the qubit at zero tilt (which requires compensating for intra-qubit nonlinear crosstalk). We then repeat this same experiment for the other qubit, with the original qubit detuned to $\beta = 0$.

Since each qubit is now at $\Delta = 4\,\text{GHz}$ when the other qubit is detuned to $\beta = 0$, we can then bring the two qubits together and observe the avoided level crossing with $\Delta_1 = \Delta_2 = 4\,\text{GHz}$ and $\varepsilon_1 = \varepsilon_2 = 0$. In Fig. 6.16 we plot the measured spectroscopic signal at this avoided level crossing, as a function of coupler SQUID bias. In other words, for each coupler bias, a vertical slice through the data gives two peaks whose separation is equal to $2J$. On the right-hand plot of the figure, we plot the measured position of the two peaks (blue and green points), corresponding to f_{10} and f_{20} of the coupled system. The horizontal axis of this plot actually extends further than in the raw data plot on the left, going all the way to $\beta_c \approx 0.8$.

We make several observations about this avoided level crossing data. The first is that the coupling can be tuned from positive through zero to negative. Here, 'zero' means that the avoided level crossing vanishes to within the linewidth of the qubit, which sets a limit on what we can measure. We also note that the location of zero coupling does not coincide with $\Phi_{\rm csq} = 0.5$ as one would expect (the location of $\Phi_{\rm csq} = 0.5$ was determined by the bias

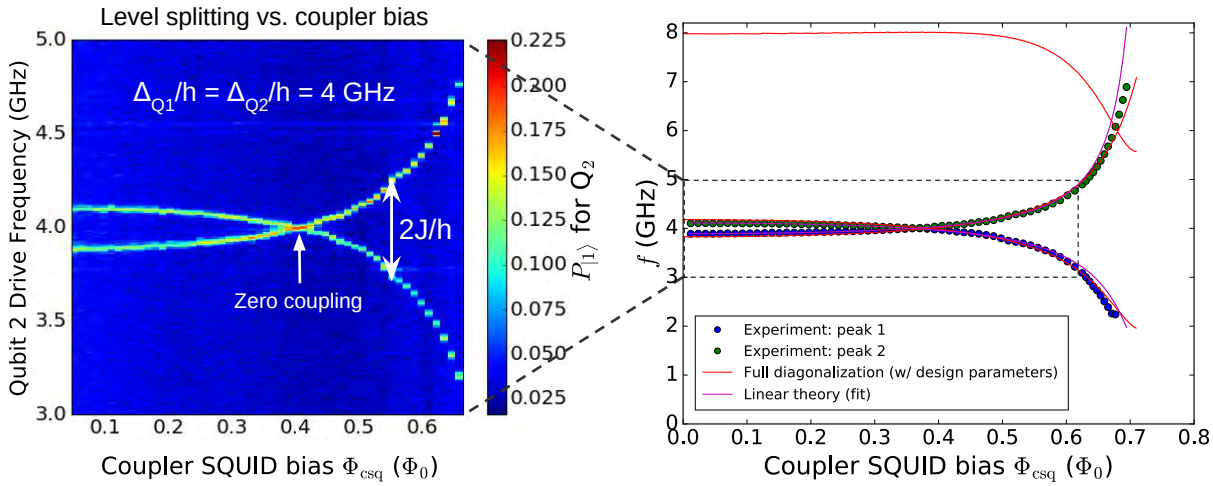


Figure 6.16: Measurement of avoided level crossing between qubits at zero tilt, as a function of coupler SQUID bias. At each SQUID bias, the qubit-qubit coupling strength is equal to half the energy gap. In the plot on the right, we compare the extracted energy levels to theory.

at which the coupler-induced flux shift vanishes in Fig. 6.11). The explanation for this is that there is an appreciable direct stray qubit-qubit coupling via the mutual inductance $M_{\mathrm{q1t}\to\mathrm{q2t}}$. Ideally, this stray coupling wouldn't be so big, but we can trace most of it back to the missing airbridge in the device as illustrated in Fig. 6.9(a). Sonnet simulation of the design predicted a stray mutual inductance of $M_{\mathrm{q1t}\to\mathrm{q2t}} = 3.4\,\mathrm{pH}$, which would imply a stray coupling of $J/2\pi = 205\,\mathrm{MHz}$ when $\Delta_1 = \Delta_2 = 4\,\mathrm{GHz}$, remarkably close to the observed 215 MHz, meaning that Sonnet is also useful for predicting stray coupling. New devices with proper airbridge symmetry are predicted to have $M_{\mathrm{q1t}\to\mathrm{q2t}} < 0.5\,\mathrm{pH}$.

We can compare the measured spectrum to theory. The red line is the prediction from full diagonalization of the qubit + coupler + qubit system [i.e., the 3D Hamiltonian 5.53)] using only predicted parameters from Sonnet and room temperature resistance measurements. The only experimental data used for this prediction is the calibrated horizontal axis. Despite the fact that this is not a fit to the data, it matches remarkably

well. We can see if we can do better by fitting to the data. Since the full Hamiltonian would take too long to fit to (remember, each point is a binary search through bias space for the qubit SQUID biases which give $\Delta_i = 4\,\mathrm{GHz}$), we can simply fit to linear theory, resulting in the magenta curve. This fits the data well for small coupling strengths, but noticeably deviates at the highest coupling strengths. This isn't surprising since the linear theory starts to break down at $\beta_c \approx 0.8$. Still, this gives a decent fit over much of the coupler's operating range.

As mentioned earlier, besides the linear approximation breaking down at large coupling strengths, other approximations start to break down as well. The rotating wave approximation (RWA) starts to break down, one symptom of which is that for large couplings, the midpoint of the two spectroscopy peaks deviates noticeably from 4 GHz, as expected from keeping the counter-rotating terms of (6.8). For the largest coupling shown, the midpoint is actually 4.5 GHz. Another assumption that breaks down is that the coupling is purely XX in (6.7). A small non-stoquastic ZZ coupling appears, as predicted in Fig. 5.14, although we have not implemented any measurements to try and detect this small non-stoquastic component. More significant is the breakdown of the two-level approximation for the fluxmon qubits. In the full diagonalization theory plot (red curves), we show not just the first two but also the third transition frequency corresponding to $|00\rangle \rightarrow |11\rangle$. We can see that this level actually gets pushed down from $\Delta_1/h + \Delta_2/h = 8\,\mathrm{GHz}$, and even crosses the upper single-excitation curve.[3] The

[3]Since direct excitation of this transition is a "forbidden" transition, we do not expect to see this intersection in spectroscopy, and expect this to be an actual level crossing rather than an avoided level crossing. We do however see a slight broadening of the upper single-excitation spectrum close to where

explanation is that there is an avoided level crossing between $|02\rangle$ and $|11\rangle$, where $|2\rangle$ denotes the second excited state of one of the fluxmons. We need to be careful about accounting for these higher levels when modeling devices with large couplings.

6.3.3 Sequential annealing: coupling in the double-well regime

The measurement of Fig. 6.16 is fairly challenging to calibrate, and as discussed earlier will not work when the qubits are in the double-well regime, so it would be nice if we had another method to characterize the coupling strength. Here, we will instead measure coupling by inferring the effective mutual inductance M_{eff} mediated by the coupler. The idea is to use a "source" qubit as a source of persistent current, and then detect the flux induced in a "detector" qubit. This idea and the pulse sequence behind this measurement, called "sequential annealing" [197], are shown in Fig. 6.17(a) and (b). For a given coupler bias, we measure the difference in measured detector qubit flux between when the source qubit is prepared in either its left or right persistent current state. Essentially, the only piece of information we need to know to extract M_{eff} is the source qubit's persistent current. This is called sequential annealing because first the source qubit is annealed into a known persistent current state, and then the detector qubit is annealed with variable tilt bias to zero out its received flux. In particular, no microwave pulses are needed for this pulse sequence (except for probing the readout resonators at the end).

The resulting experimental data for the detector flux shift $\Delta\Phi_t^d$ is plotted in Fig. 6.17(c). Note that each point in the 2D plot is the result of a binary search through the

we predict this intersection to happen, which could very well be due to crossing with this third transition.

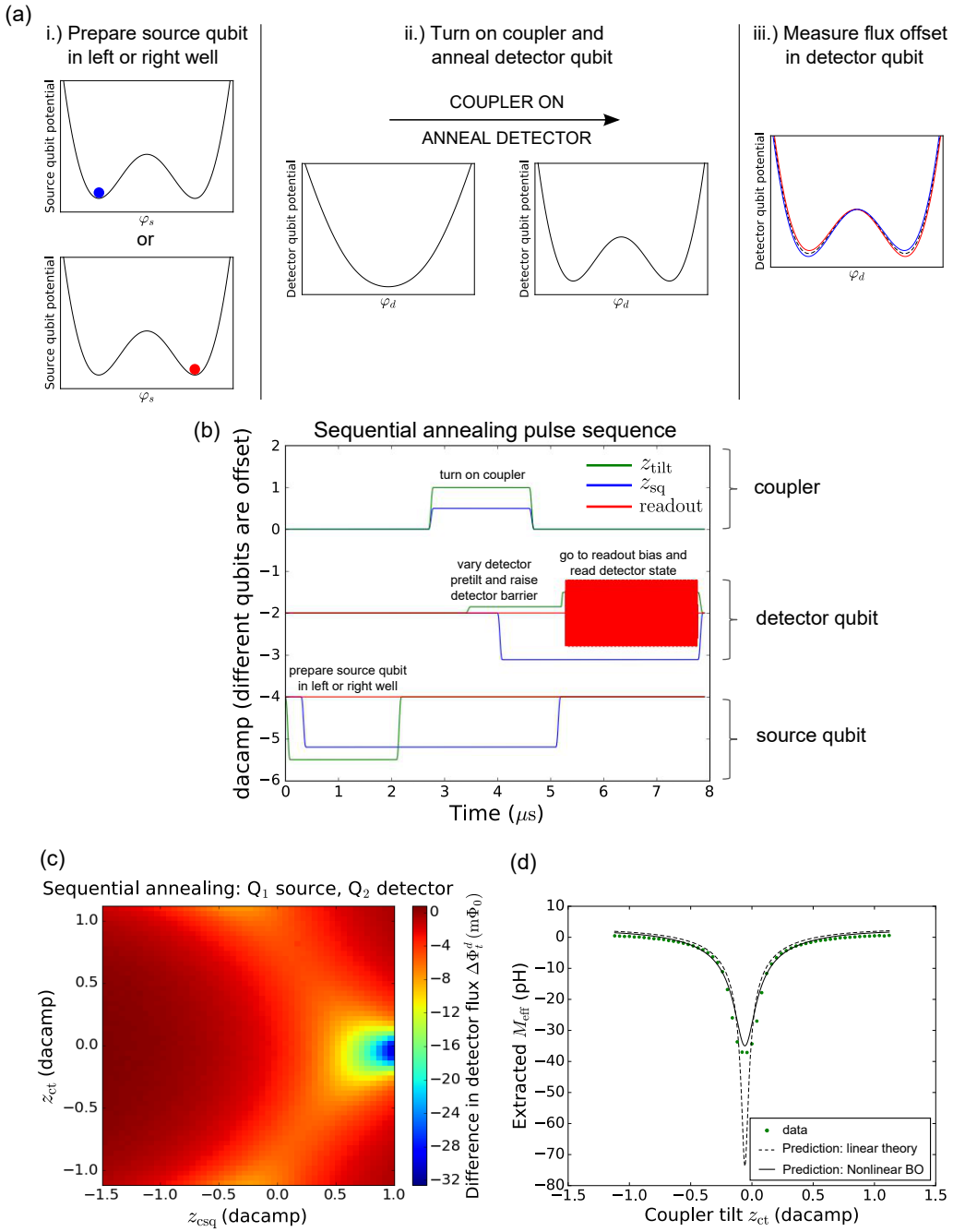


Figure 6.17: Sequential annealing experiment for measuring the inter-qubit M_{eff}. (a) Illustration of sequential annealing. The source qubit is prepared in either of its persistent current states. The coupler is turned on and the detector qubit is annealed from single- to double-well potential, allowing the detection of the flux induced by the source qubit's persistent current mediated by the inter-qubit coupling. (b) Pulse sequence of the sequential annealing experiment. (c) Experimentally measured $\Delta\Phi_t^d$, the difference in detector qubit flux between the two prepared states of the source qubit, versus coupler biases. (d) Extracted M_{eff} vs. coupler tilt along a vertical cut of the data from (c) at the largest $\beta_c \approx 0.85$.

detector qubit's S curve for probably 0.5 of ending up in each well, similar to the data in Fig. 6.11. If we plug in the source qubit's persistent current of $I_p^s \approx 0.87\,\mu$A, we can extract $M_{\rm eff}$ from the data according to the relation

$$\Delta\Phi_t^d = 2M_{\rm eff} I_p^s. \tag{6.10}$$

We can compare the extracted $M_{\rm eff}$ to theory, as shown in Fig. 6.17(d) vs. coupler tilt for the largest coupler nonlinearity measured, $\beta_c \approx 0.85$., using the designed coupler and qubit parameters. We plot two theory predictions, one using linear theory (section 5.1.2) for the coupler and the other using the nonlinear theory (section 5.2.1). We see that for this coupler nonlinearity, the nonlinear theory is much more successful at predicting the effective mutual inductance mediated by the coupler. We emphasize that $M_{\rm eff} = 40\,$pH corresponds to an extremely large coupling, and with the fluxmon design this very large coupling can occur between one qubit and each of many other qubits simultaneously. For example, when the qubit Δ's are 1 GHz, this corresponds to a coupling strength of almost 7 GHz, more coupling than we need!

6.4 Qubit coherence vs. coupler bias

In order for the fluxmon architecture to be scalable for quantum annealing, we need to ensure that qubit coherence doesn't degrade between an isolated qubit and a qubit connected to a coupler. We first look at dissipation. In Fig. 6.18, we plot the measured T_1 of an uncoupled (isolated) qubit and of a coupled qubit with the coupler off, both on the same chip. We see no significant degradation in T_1 over the range of 500 MHz

to 7 GHz with the coupler off, meaning that the geometry change of the qubit due to the presence of the coupler has not affected the qubit's dissipation. We can then look at what happens to one of the qubits when we turn the coupler on, with the other qubit detuned away to $\beta \approx 0$. In Fig. 6.18(b), we show the resulting T_1 vs. coupler β_c with the qubit at $f_{10} = 4\,\text{GHz}$. Note that because the coupler induces a β_c-dependent frequency shift in the qubit, for each β_c we have compensated the qubit's SQUID bias to maintain a constant f_{10}. We see that over much of the coupler's operating range, the qubit's T_1 is not affected, until $\beta_c \gtrsim 0.6$. This means we can achieve fairly large coupling strengths before seeing excess dissipation in the qubit. We can fit the degradation in T_1 to a model in which the coupler has an intrinsic T_1 that damps the qubit, analogous to the Purcell effect for resonators. We find that assuming an intrinsic coupler T_1 of 300 ns explains the data fairly well, and roughly makes sense given that the coupler has very skinny wires and is probably limited by both dielectric loss and flux noise to a lower T_1 than the qubit. Although the degradation in T_1 is small for a single coupler, we do need to consider what will happen in the future when we turn on the coupling of multiple couplers connected to a single qubit, in which case the excess loss will be amplified in proportion to the number of couplers. Improving the coupler's intrinsic T_1 in the future should help alleviate this effect.

Next, we can look at low-frequency noise and dephasing versus coupler bias. The primary effect of the coupler on qubit dephasing will be through noise coupled to the qubit's tilt loop. In Fig. 6.19, we plot the measured low frequency flux noise in the qubit

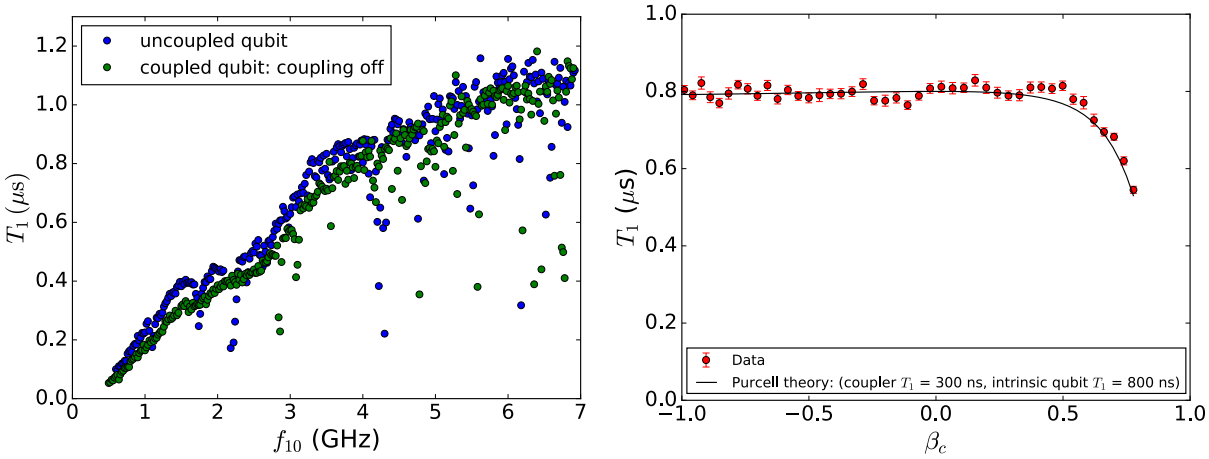


Figure 6.18: (a) T_1 at zero tilt versus frequency for an uncoupled (isolated) and coupled (with coupling off) qubit on the same chip. (b) Measured qubit T_1 versus coupler β_c with qubit f_{10} fixed at 4 GHz.

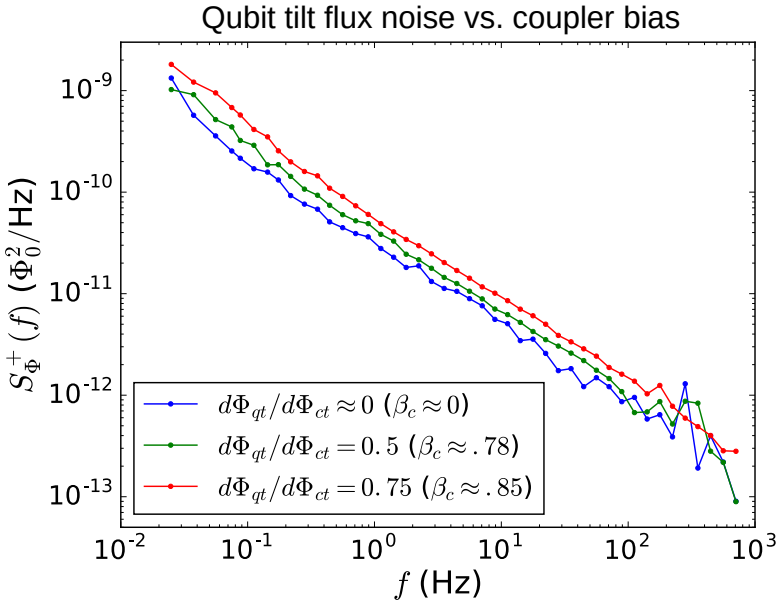


Figure 6.19: Low frequency flux noise for different coupler susceptibilities. Flux noise is degraded for very strong large coupler susceptibilities.

loop for three different coupler biases, with the other qubit detuned away to $\beta \lesssim 0$. With the coupler turned off, the flux noise is within the typical range we see for uncoupled qubits. We also see no change in Ramsey dephasing times at or away from zero tilt with the coupler off. We can start to push up the coupler nonlinearity, which enhances the susceptibility and therefore the differential flux transfer function $d\Phi_{qt}/d\Phi_{ct}$ from coupler to qubit (the quantity we plotted theoretically in Fig. 5.6), and see when the qubit starts to feel flux noise from the coupler. The green curve corresponds to a measured $d\Phi_{qt}/d\Phi_{ct} = 0.5$ ($\beta_c \approx 0.78$), for which we observe roughly a 50% increase in the qubit's flux noise power. We can push the sensitivity up even further in the red curve, where $d\Phi_{qt}/d\Phi_{ct} = 0.75$ ($\beta_c \approx 0.85$). The amount by which the flux noise degraded when increasing the sensitivity grew approximately as the square of the sensitivity, as expected. From the data and known sensitivity, we can then deduce the flux noise intrinsic to the coupler. We extract that the flux noise power intrinsic to the coupler is approximately a factor of two higher than than in the qubit, which again makes sense given the very high aspect ratio of the coupler's skinny wires. We note that these coupler susceptibilities are significantly higher than we would need to use for coupling qubits, but as with dissipation it may start to have an appreciable effect on qubit coherence when many couplers coupled to a single qubit are tuned to high susceptibility at once.

6.5 Demonstration of fast two-qubit annealing

We now put our coupler to use and perform the simplest case of quantum annealing with quadratic terms: quantum annealing with two coupled qubits. Before we can attempt this experiment, we need to make sure that individual qubit readout still works when the other qubit is read out simultaneously. To test simultaneous qubit readout, we prepare one qubit in one of its persistent current states, and the other qubit in one of its persistent current states. We then dispersively measure what state qubit 1 is in, with or without a simultaneous dispersive measurement of what state qubit 2 is in. As shown in Fig. 6.20, the readout fidelity is unchanged whether or not we probe just one or both qubits. During this experiment, the coupler is off, which is generic enough for our purposes because at the end of the annealing all qubits will be in stable states and we can turn off the coupling before sending the readout probes. It will be interesting to see how many qubits can be simultaneously read out in this fashion. Since we don't rely on a near-quantum-limited parametric amplifier, it may be possible to read out a great many qubits this way before we start running into problems such as induced transitions or amplifier saturation.

In Fig. 6.21, we illustrate the pulse sequence for a two-qubit annealing experiment. Here, both qubits are initialized in (or at least nearly in) their ground states within their monostable potentials ($\beta \approx 0$), with the coupler turned on, and some variable amount of tilt bias (local field) applied to each qubit. After the system settles in its ground state, the qubit barriers are both uniformly raised to the double-well regime, after which the coupling is turned off and the qubits are brought to their respective readout post-tilts

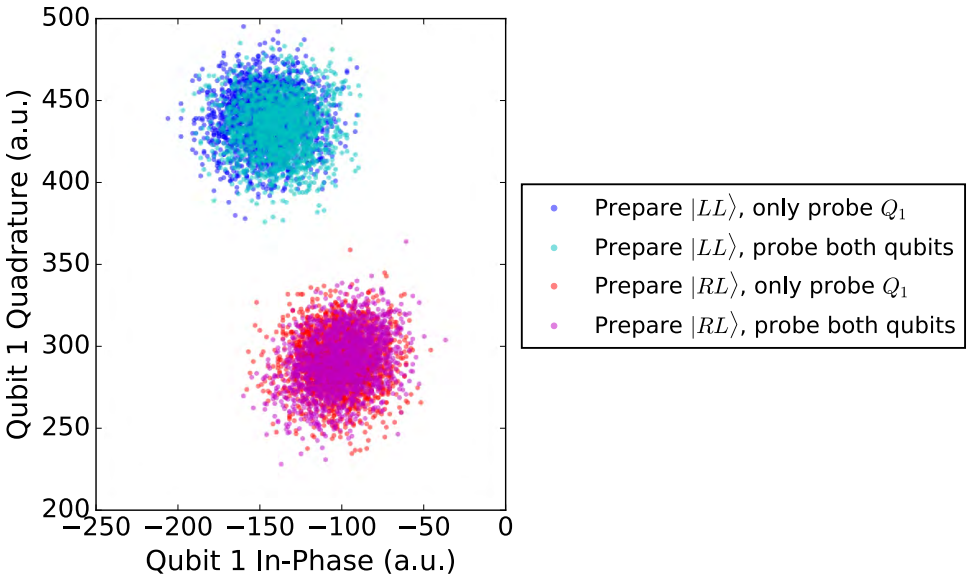


Figure 6.20: Experimental check of the robustness of two-qubit readout. For the chosen readout parameters, there is no effect on readout fidelity due to interference from simultaneous readout of another qubit.

for simultaneous readout. This experiment can be thought of as a generalization of the single-qubit S curve experiment, where there are now two qubits and a coupling term between them. We note that this is not an implementation of "controlled" annealing as defined in Ref. [208], since we are fixing the flux biases rather than the ratio of h and J as would be desired in a computational annealer (doing "controlled" annealing would require a calibration of I_p along the trajectory, which we have not yet implemented). Explicitly, the time-dependent Hamiltonian over the course of the anneal is, in the two-level flux basis,

$$H(t) = \frac{1}{2}\Delta_1(t)XI + \frac{1}{2}\Delta_2(t)IX + h_1(t)ZI + h_2(t)IZ + J(t)ZZ, \tag{6.11}$$

where $\Delta_1(t)$ and $\Delta_2(t)$ monotonically decrease from $\sim f_{LC} \approx 20\,\text{GHz}$ to ~ 0 and $h_1(t) = \frac{1}{2}\varepsilon_1(t) = I_{p1}(t)\Phi_{t1}$, $h_2(t) = \frac{1}{2}\varepsilon_2(t) = I_{p2}(t)\Phi_{t2}$, and $J(t) = M_{\text{eff}}I_{p1}(t)I_{p2}(t)$.

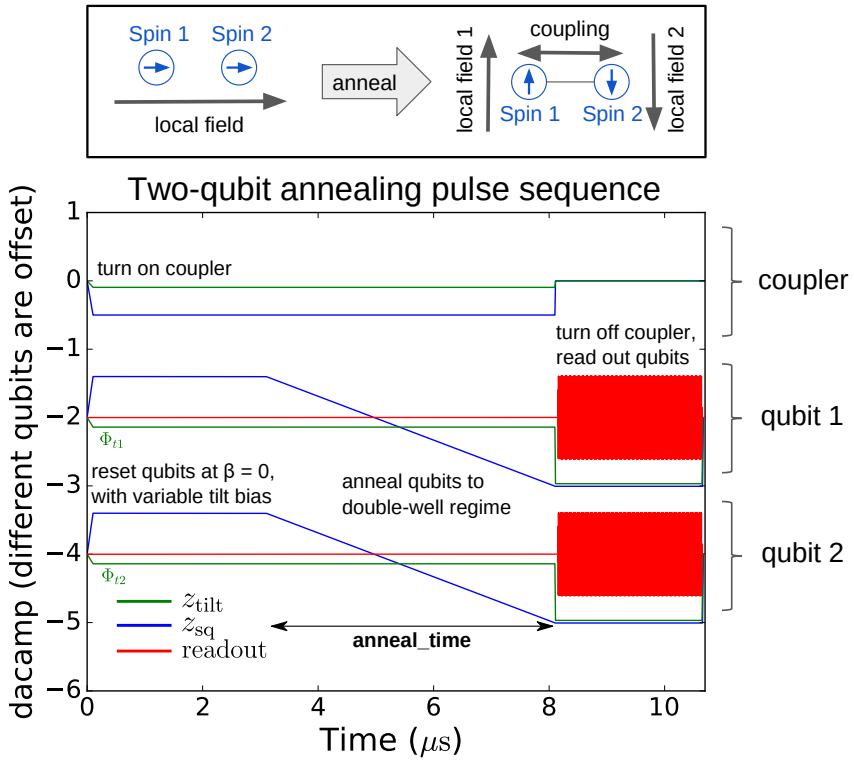


Figure 6.21: Cartoon illustration and actual pulse sequence for the two-qubit annealing experiment.

This "fixed" tilt bias annealing is sufficient to demonstrate the essential physics of the annealing process we wish to study.

We first consider what happens when we perform two-qubit annealing in the trivial case without any coupling, i.e., with the coupler biased such that $M_{\text{eff}} = 0$ between qubits. The resulting experimental data is shown in Fig. 6.22, for two different annealing ramp times, 5000 ns in (a) and 100 ns in (b). In the plots on the left, we show the measured probability for observing each of the four possible states $\{|LL\rangle, |LR\rangle, |RL\rangle, |RR\rangle\}$ at the end of the anneal, as a function of the two qubit tilt biases (i.e., local Z fields) during the anneal. We have also drawn in black boundaries where the state of maximum probability switches from one state to another, corresponding to ground state degeneracies. These

boundaries can be thought of as defining a ground state stability diagram. In the plots on the right, we show the probability of the most likely state as a function of the qubit tilt biases. Here, the areas of minimum probability coincide with the phase diagram boundaries drawn in the plots on the left. The locations of the phase boundaries are simple to understand when there is no coupling. Assuming adiabaticity, each qubit simply follows its local field, meaning that the four quadrants of the (Φ_{t1}, Φ_{t2}) plane correspond to the four different ground states. Because one qubit does not feel the flux of the other, taking a horizontal/vertical cut of the data simply corresponds to an S-curve experiment (single-qubit annealing) for Q_1/Q_2 (c.f. section 3.4). We can then understand the difference between slow and fast annealing according to the same physics we used to describe single-qubit annealing. In particular, when annealing slow relative to thermalization times, thermalization errors are significant. Annealing faster than the thermalization times but still slow enough not to incur much non-adiabatic transition makes the errors significantly less for the majority of annealing paths.

Next, we consider two-qubit annealing with nonzero qubit-qubit coupling J. In Fig. 6.23, we plot the resulting experimental data for a moderately strong amount of FM coupling, $M_{\text{eff}} \approx -5\,\text{pH}$. There are several interesting aspects to this data. First, we can try and understand the location of the phase boundaries, which is now a bit more complicated due to an extra feature added by the FM coupling. Qualitatively, since the coupling is ferromagnetic, it makes sense that the quadrants corresponding to the ferromagnetic states $|LL\rangle$ and $|RR\rangle$ have grown in area, while the quadrants for the

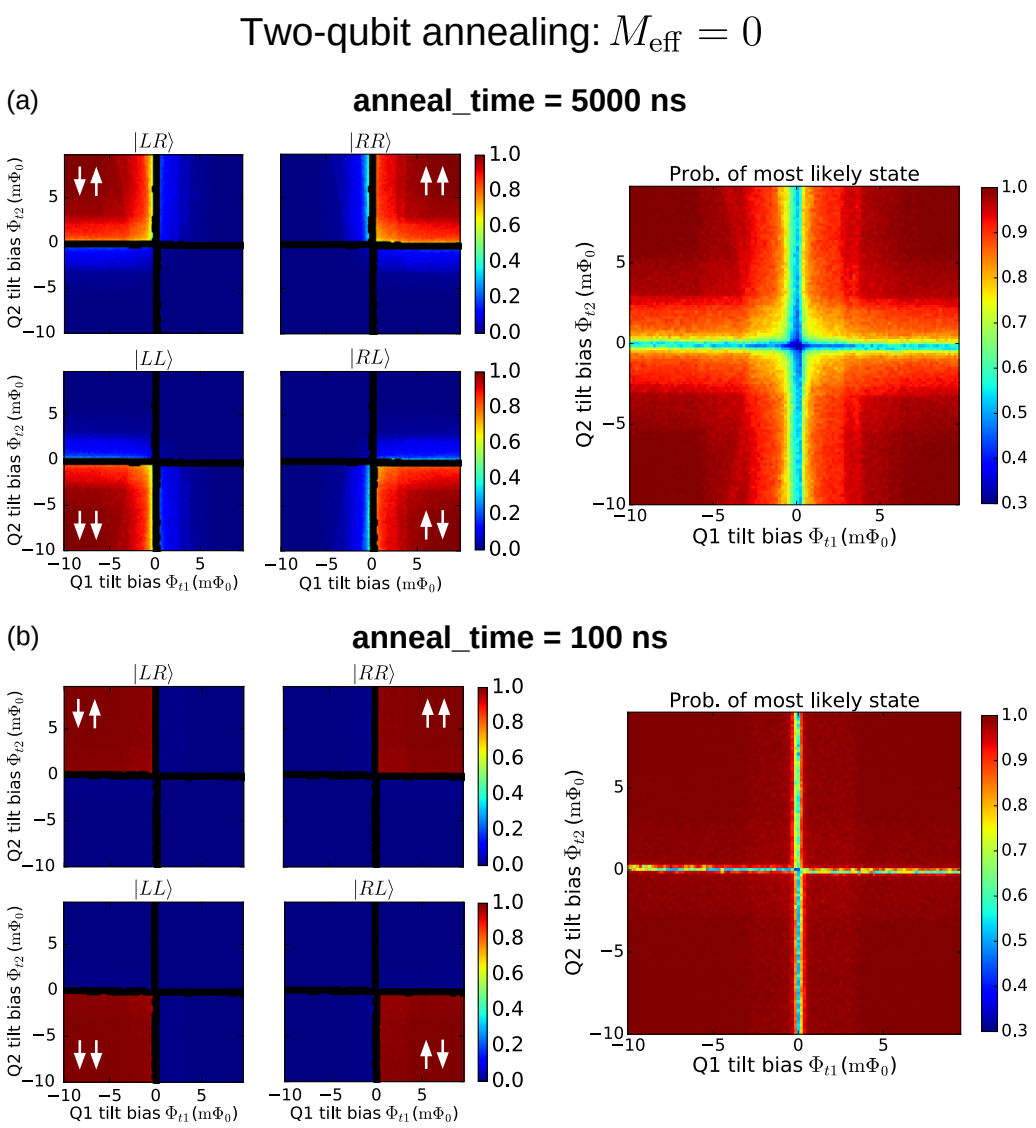


Figure 6.22: Experimental two-qubit annealing results for coupling turned off ($M_{\text{eff}} \approx 0\,\text{pH}$) for two different annealing rates: (a) 5000 ns and (b) 100 ns. In each of (a) and (b), the left collection of plots shows the measured probabilities for the four classical states at the end of the annealing, with boundaries drawn in, and the right plot shows the probability of the most likely of the four possible states.

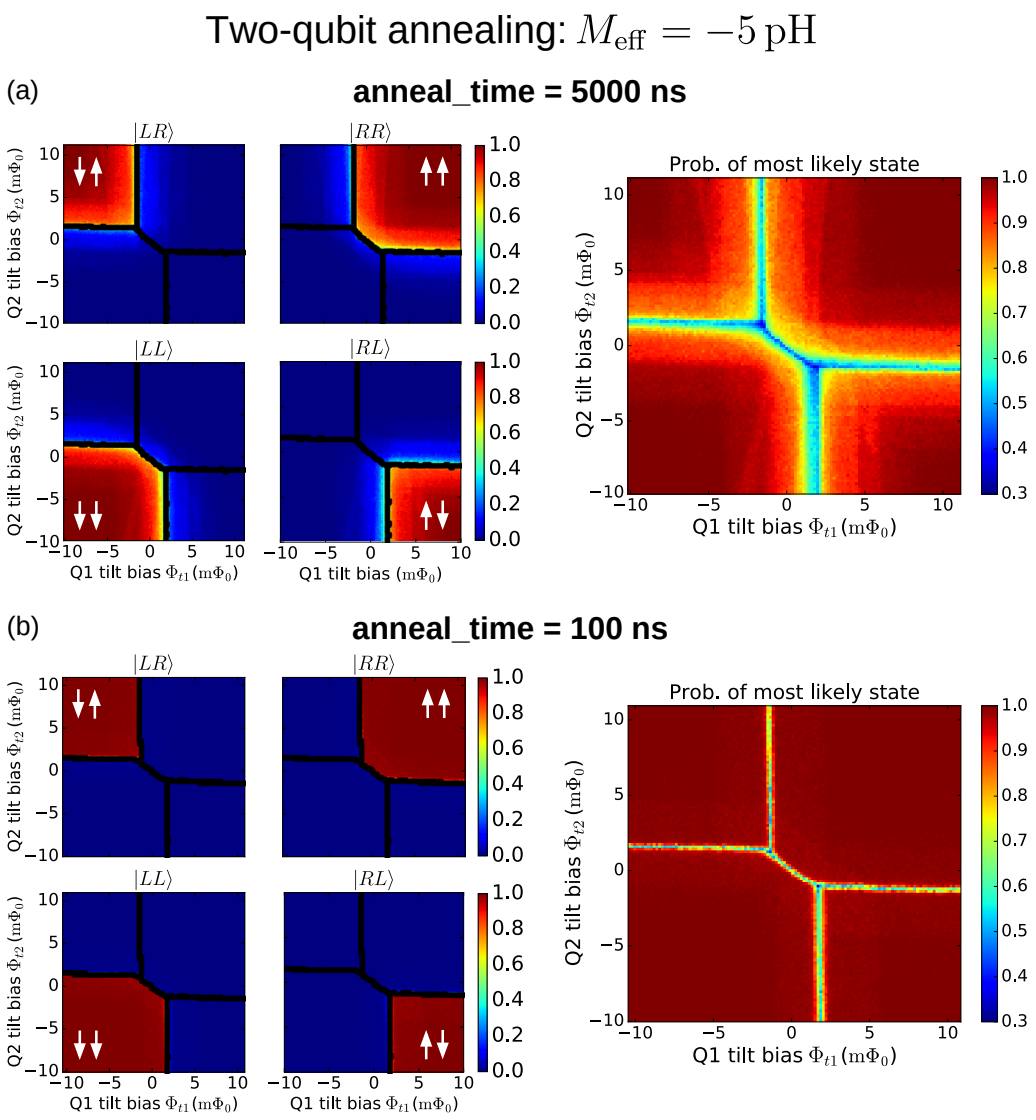


Figure 6.23: Experimental two-qubit annealing results for $M_{\text{eff}} = -5\,\text{pH}$ for two different annealing rates: (a) 5000 ns and (b) 100 ns. In each of (a) and (b), the left collection of plots shows the measured probabilities for the four classical states at the end of the annealing, with boundaries drawn in, and the right plot shows the probability of the most likely of the four possible states.

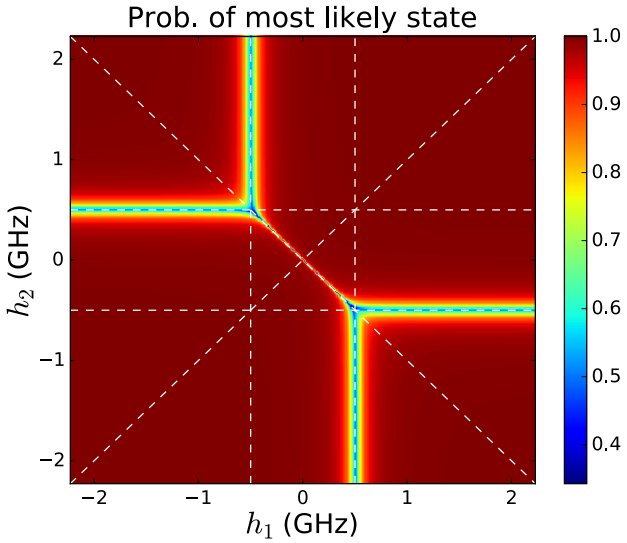


Figure 6.24: Numerically computed ground state probability of most likely two-qubit flux state $\{|LL\rangle, |LR\rangle, |RL\rangle, |RR\rangle\}$ at a fixed moment in time. This probability is plotted versus h_1 and h_2 with $\Delta/h = 100\,\text{MHz}$ and $J/h = -0.5\,\text{GHz}$, according to numerical diagonalization of (6.11). The widths of the horizontal and vertical boundaries correspond single-qubit tunneling elements and are of order Δ, while the width of the narrower diagonal boundary corresponds to a two-qubit tunneling element, and is of order Δ^2.

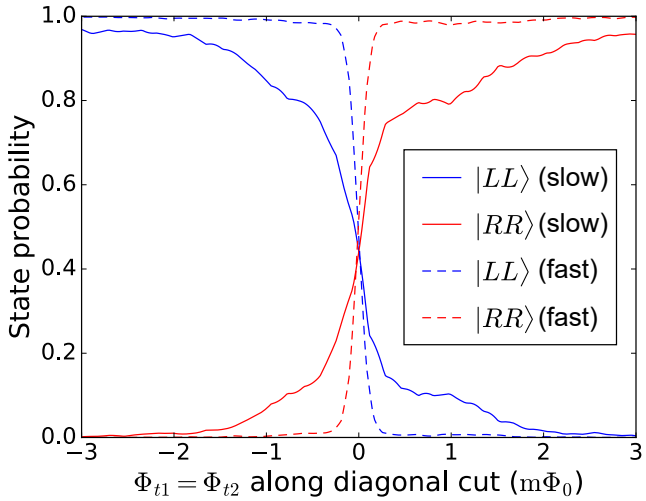


Figure 6.25: Diagonal cuts through the diagonal phase boundary in the experimental data of Fig. 6.23, showing the qualitatively different crossover from $|LL\rangle$ to $|RR\rangle$ across the boundary.

antiferromagnetic states $|LR\rangle$ and $|RL\rangle$ have shrunk in area. Since the Hamiltonian is very nearly diagonal in the flux basis when the dynamics freezes near the end of the annealing (i.e., an Ising Hamiltonian with vanishingly small transverse field), the location of the phase boundaries is determined by the classical energies of the states. For each set of flux biases, there are four such energies determined by whether each "spin" is up or down. These energies are summarized in table 6.2. A phase boundary occurs whenever two of these energies become degenerate and also happen to be the lowest energy, which defines various line segments in the (h_1, h_2) plane. In Fig. 6.24, we compute the ground state and plot the theoretical probability of the most likely classical state for a finite tunneling of $\Delta_1/h = \Delta_2/h = 100\,\text{MHz}$, which corresponds to a somewhat arbitrary point in time during the anneal. We also plot the lines of degeneracy in the (h_1, h_2) plane defined by equating every pair of the rows of table 6.2. Note that these lines define the degeneracies for all four of the energy levels, but experimentally we only detect the degeneracies of the lowest energy level, turning the lines into line segments. As can be seen, the vertical and horizontal shifts of the horizontal and vertical phase boundaries, respectively, should be equal to $\pm J$ (or $\pm J/I_p$ in units of flux) relative to the origin. However, because this represents the ground state of the system at single moment in time, this is an over-simplified picture that does not accurately describe the outcome of the experiment plotted in Fig. 6.23. Instead, we need to consider the full time-dependent trajectory of the Hamiltonian 6.11.

To better understand the data, it is convenient to think of the annealing as occurring

σ_{z1}	σ_{z2}	Energy
$+1$	$+1$	$h_1 + h_2 + J$
$+1$	-1	$h_1 - h_2 - J$
-1	$+1$	$-h_1 + h_2 - J$
-1	-1	$-h_1 - h_2 + J$

Table 6.2: Classical energies of the two-spin Ising Hamiltonian for the four possible states.

along a path within the $\left(\frac{h_1}{|J|}, \frac{h_2}{|J|}\right)$ plane. In Fig. 6.26, we plot the various inequalities defining different regions of the two-qubit ground state stability diagram within this plane. The thick solid black lines indicate the location of phase boundaries (ground state degeneracies). The reason it is better to think within these re-scaled axes is because under our "fixed" annealing schedule, the ratios $\frac{h_i}{J}$ will be time-dependent. Because $h(t) \propto I_p$ whereas $J(t) \propto I_p^2$, at the beginning of the annealing, the ratio $\left|\frac{h}{J}\right|$ will monotonically increase over the anneal, meaning the annealing trajectory will move radially inward in the $\left(\frac{h_1}{|J|}, \frac{h_2}{|J|}\right)$ plane.

As a case study, we consider a trajectory corresponding to the particular annealing schedule with $(\Phi_{t1}, \Phi_{t2}) = (2\,\mathrm{m}\Phi_0, -1\,\mathrm{m}\Phi_0)$. This trajectory is illustrated in Fig. 6.27, and corresponds to a point within the $|RR\rangle$-dominated region of Fig. 6.23 near the $|RR\rangle\,/\,|RL\rangle$ phase boundary. Note in particular that this trajectory crosses a phase boundary in the $\left(\frac{h_1}{|J|}, \frac{h_2}{|J|}\right)$ plane, as do most trajectories that start within the shaded red regions of Fig. 6.26. This is simple to understand intuitively: at the beginning of the anneal, J is relatively small, so spin 2 will want to align with its local field. When J becomes larger, the coupling energy becomes significant and the system can lower its energy by flipping spin 2 (the relative strength of h_i and J over time can be seen in

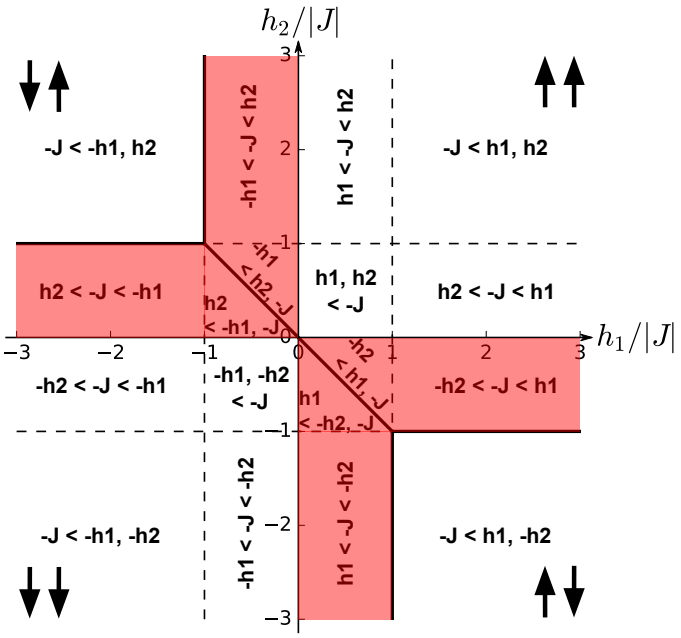


Figure 6.26: Inequalities defining various regions of the two-qubit ground state stability diagram in the $\left(\frac{h_1}{|J|}, \frac{h_2}{|J|}\right)$ plane with FM coupling ($J < 0$). Solid thick black lines indicate the location of phase boundaries (ground state degeneracies). The shaded red regions are regions whose ground state has been changed from the case with $J = 0$.

Fig. 6.27(b), along with the components of the ground state in Fig. 6.27(d), where we can see the ground state changing from $|\uparrow\downarrow\rangle$ to $|\uparrow\uparrow\rangle$ in the middle of the anneal). This corresponds to the system tunneling out of a local minimum into a global minimum in the middle of the annealing, analogous to the weak-strong cluster problem of Ref. [19].

The crossing of the phase boundary can be described in two different ways. One way is to note that the crossing of the boundary coincides with the minimum energy gap (for this problem around 200 MHz) as illustrated in Fig. 6.27(c), meaning there is an avoided level crossing between the ground state and first excited state. The success rate of reaching the correct state on the other side of the avoided level crossing is then determined by the probability of a non-adiabatic transition occurring in accordance with standard Landau-

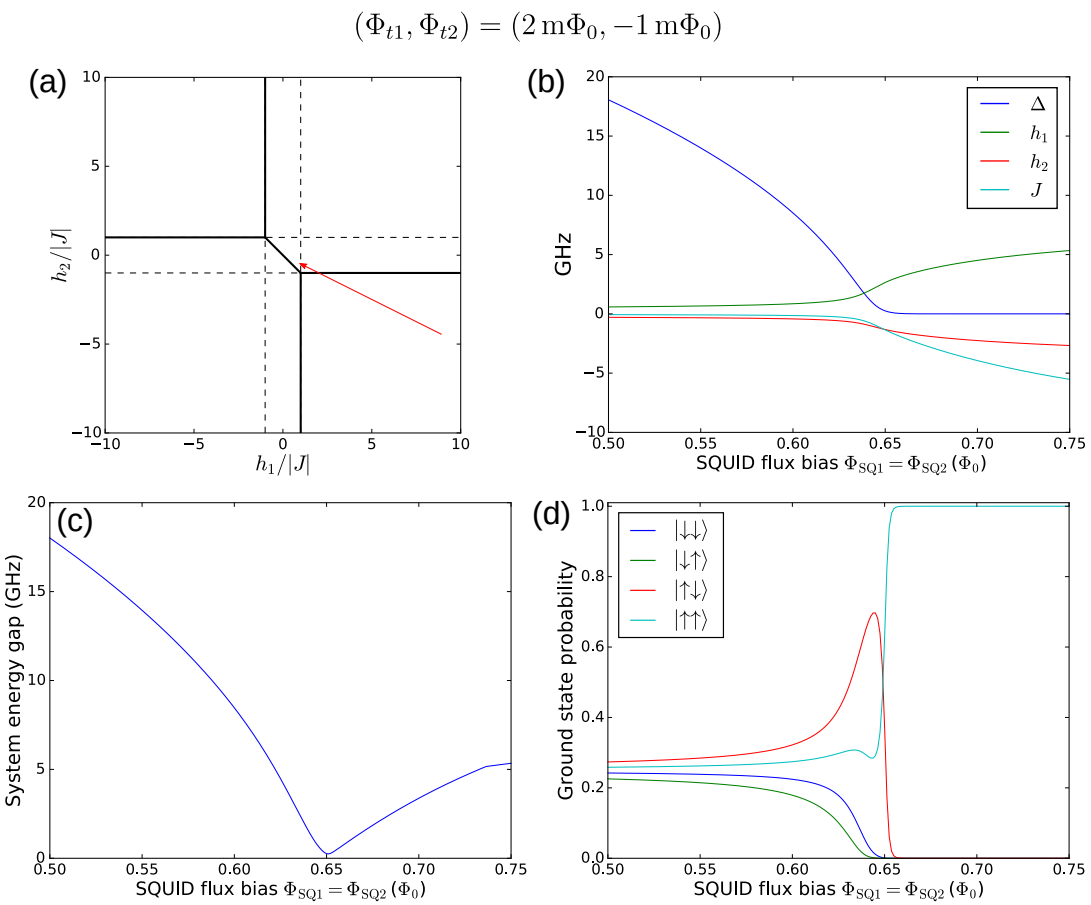


Figure 6.27: Numerical simulation of the particular annealing schedule with $(\Phi_{t1}, \Phi_{t2}) = (2\,\mathrm{m}\Phi_0, -1\,\mathrm{m}\Phi_0)$. (a) Red arrow indicates trajectory of the annealing schedule within the $\left(\frac{h_1}{|J|}, \frac{h_2}{|J|}\right)$ plane. (b) Time-dependent Hamiltonian coefficients of (6.11) throughout the fixed two-qubit annealing schedule: $\Delta(t) \equiv \Delta_1(t) = \Delta_2(t)$, $h_1(t)$, $h_2(t)$, and $J(t)$. (c) Energy gap of the two-qubit system along the annealing trajectory. (d) Probability of the four classical flux basis states within the ground state along the annealing trajectory.

Zener physics, ignoring environmental effects. A second way to understand the crossing is through quantum tunneling. In order for the second spin to flip, the system must move from a false local minimum to a the true global minimum in the middle of the anneal. This can be directly visualized by plotting the two-dimensional potential energy landscape of the two-qubit system at different points during the annealing, as done in Fig. 6.28. The two degrees of freedom in this potential are the phase variables for each qubit, φ_1 and φ_2. As illustrated in Fig. 6.28, the original global minimum continuously deforms into a local minimum, and the system must tunnel through a potential barrier by flipping spin 2 to reach the global minimum at the end.

Looking back at the experimental data in Fig. 6.23, we can see that there is a large qualitative difference in the resulting probabilities between Fig. 6.23(a) and 6.23(b). We note in particular that $5\,\mu$s is the fastest annealing rate available on D-Wave's current commercial quantum annealers. This slower annealing time incurs substantially more inter- and intra-well thermalization errors and substantially broadens the single- and two-qubit phase boundaries. In a computational annealer, these thermalization errors would lead to a substantially reduced probability of reaching the ground state of the problem Hamiltonian over much of the (h_1, h_2) space.

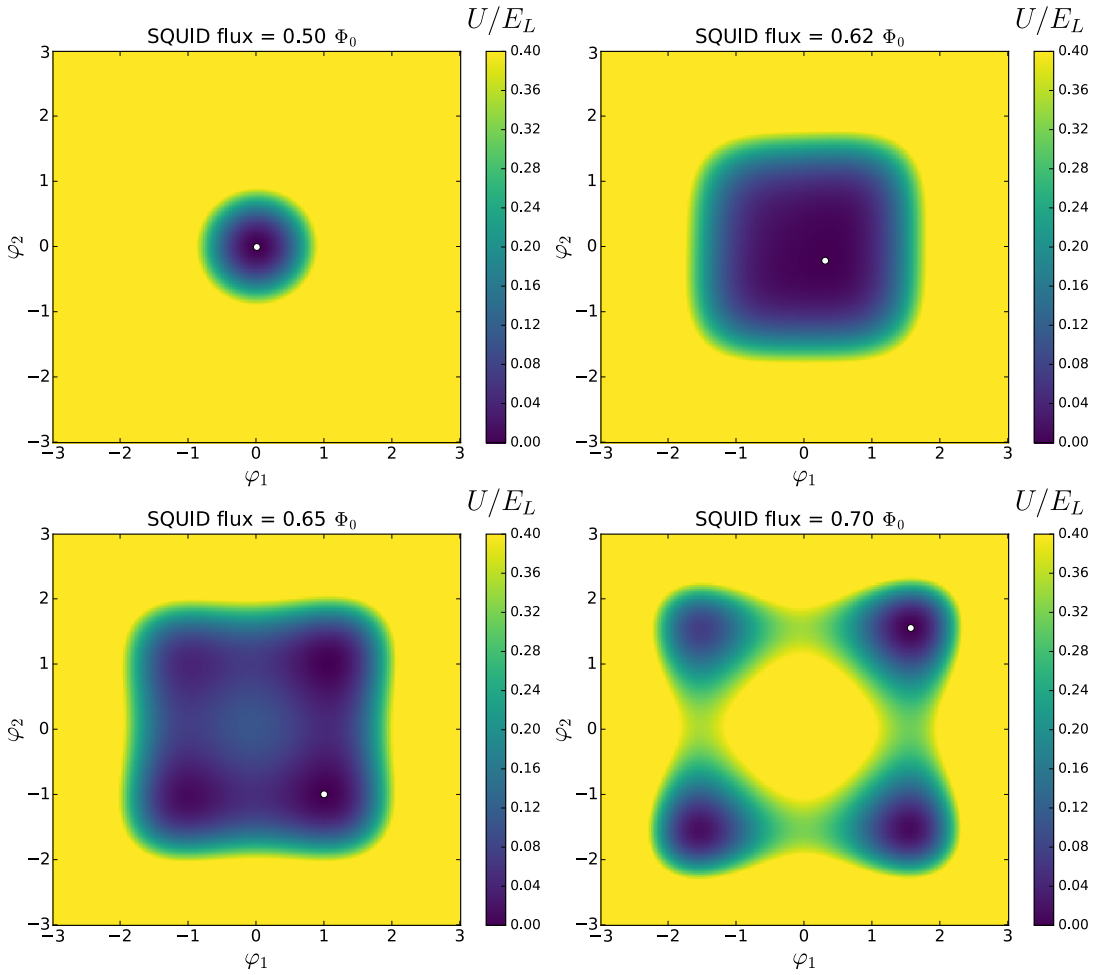


Figure 6.28: Snapshots of the two-qubit potential landscape $U(\varphi_1, \varphi_2)/E_{L1}$ throughout the particular anneal with $(\Phi_{t1}, \Phi_{t2}) = (2\,\mathrm{m}\Phi_0, -1\,\mathrm{m}\Phi_0)$. White dot indicates the location of the global minimum of the potential, which we define as zero for the sake of illustration. Color scale is truncated at an upper limit of 0.4 so that the features are more easily visible. The original global minimum continuously deforms into a local minimum, and the system must tunnel by flipping spin 2 to reach the global minimum at the end.

6.6 Scaling up: Flip-chip architecture and future directions

The planar fluxmon geometry as implemented so far is, in practice, not very scalable to a complex many-qubit device. This is especially true with a CPW geometry with a shared ground plane, through which nonlocal return currents can cause significant crosstalk. When scaling up the number of qubits, it also quickly becomes unclear how to even bring in all the bias lines for the device, let alone have low crosstalk between them. A way to potentially solve these issues is to utilize the third dimension with some sort of via-like structures. As we know, conventional fabrication with lossy dielectrics is not compatible with low dissipation, so we need to be careful about how we proceed. Although we have superconducting airbridges at our disposal, which have helped us to access the third dimension over short distances via hop-overs, this is not enough to achieve more complex fluxmon devices beyond a handful of qubits.

One promising path towards a multilayer architecture without lossy dielectrics is using flip-chip bump bonding techniques. Bump bonds are thick ($\gg 1\,\mu$m) "bumps" of soft metal that can form a connection between metallic layers on two separate chips by squishing them in between the chips. This typically leads to a chip-to-chip separation of order 1 - 10 microns. This is coarsely illustrated by the cartoon in Fig. 6.29(a). One strategy would then be to have bias circuitry on one chip and qubit circuitry on the other. The bias circuitry chip could even be a conventional multi-layer fab with

lossy dielectrics, as long as the qubit chip is shielded from the lossy elements. Bump bonding is fairly common in classical IC and MEMS devices, but having the bumps be superconducting is much less common and much more difficult. We implement a version of the superconducting indium bump bond process developed in Ref. [209], which uses thin-film TiN as a diffusion barrier between the indium bumps and aluminum circuitry.

High-angle SEM micrographs of the first flip-chip fluxmon test device are shown in Fig. 6.29(b). For this test device, the bias lines are connected via wirebonds to the carrier chip, and then go through bump bonds to reach the qubit chip (but we have also made subsequent samples for which the bias lines are completely out-of-plane of the qubit chip, on the carrier chip). For inductive couplings between elements on the qubit chip (such as between qubit and readout resonator, or bias line and qubit), a continuous carrier chip ground plane reduces the mutual inductance, so we cut slots in the flip-chip near these couplings to retain the coupling strength [an example of this can be seen for qubit B in Fig. 6.29(c)]. The chip-to-chip separation for this device was measured to be between 3 and 3.5 μm. 3 μm was the minimum acceptable gap because that is the gap at which the impedance of the fluxmon CPW with ground plane directly above it starts to significantly change (in general, the presence of a flip-chip ground plane will decrease the inductance and increase the capacitance). We measured two types of isolated fluxmons in this test device, as illustrated in Fig. 6.29(c). One qubit had no ground plane on the flip chip directly above the qubit, and also no bump bonds in close vicinity to the qubit. The second type did have ground plane on the carrier chip, leading to a large portion of

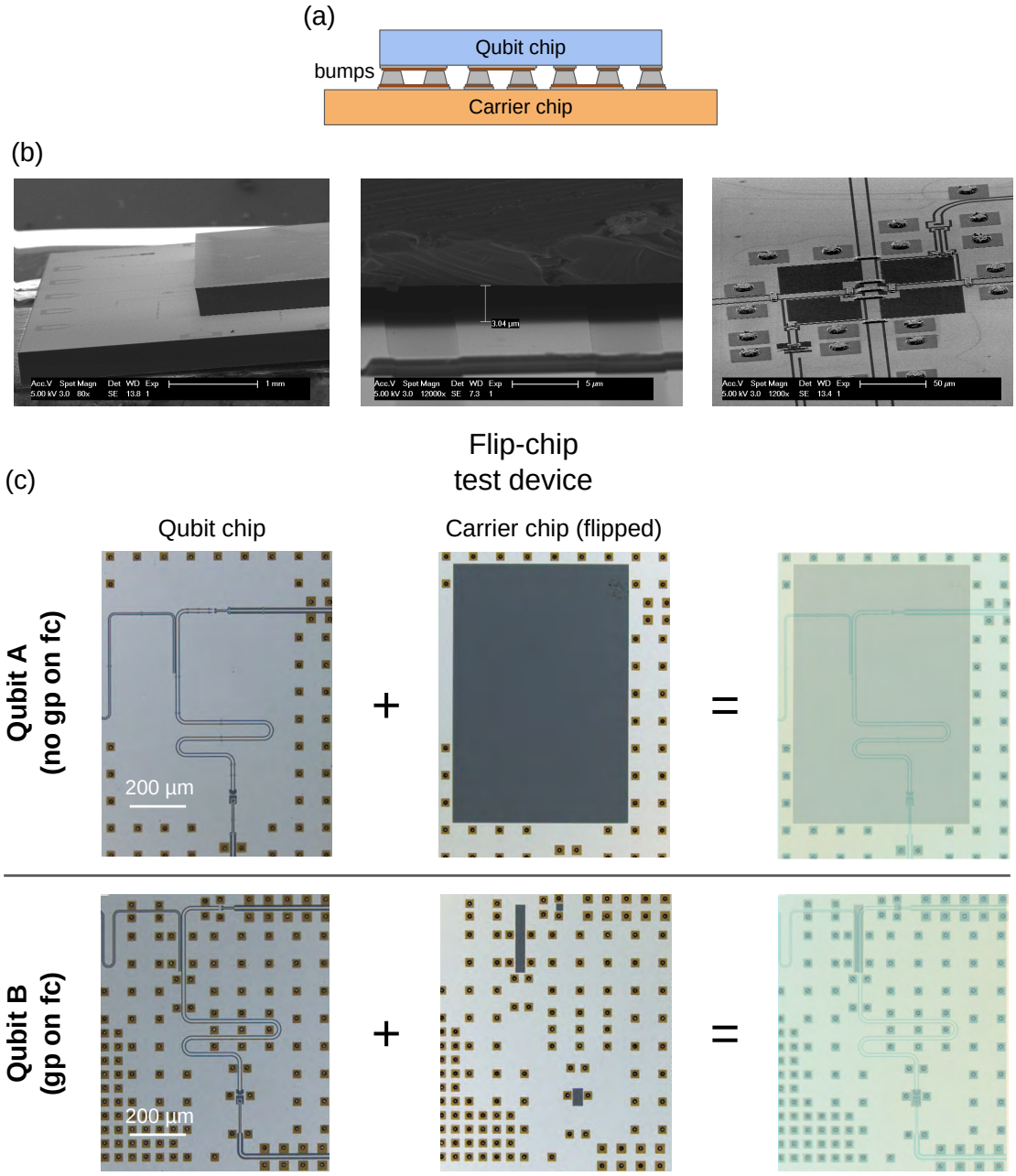


Figure 6.29: (a) Cartoon illustration of the bump bond technique for the flip-chip architecture. (b) High-angle SEM micrographs of the first flip-chip fluxmon test device. The first two pictures show the flip-chip. The picture on the right is the intersection of two coupled qubits on the qubit chip, with indium bump bonds on TiN pads visible. (c) Two types of isolated qubits tested on the flip-chip test device, one without ground plane directly above the qubit or bumps near the qubit, and the other with both. For both of these qubits, the bias lines were routed from the carrier chip to the qubit chip through bump bonds.

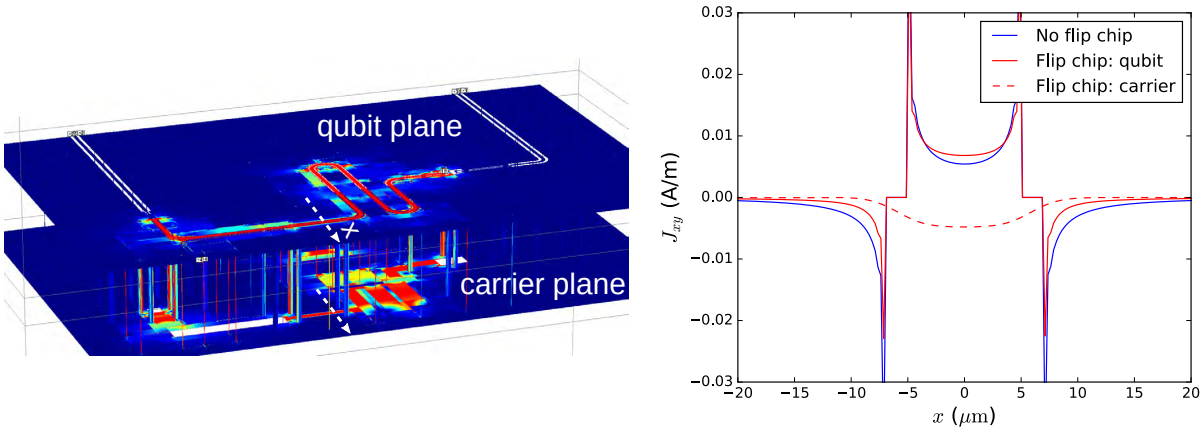


Figure 6.30: Simulation of fluxmon current distribution in the presence of a connected ground plane directly above the qubit. In contrast with the single-plane CPW fluxmon, the quasi-CPW/quasi-microstrip fluxmon has a substantial portion ($\sim$ 60 %) of its return current in another plane.

the fluxmon's return current being localized above the qubit on the carrier chip rather than on the sides of the qubit on the qubit chip as usual. The differing return current distributions simulated in Sonnet are shown in Fig. 6.30.

There are two other main requirements that the flip chip architecture must satisfy besides just allowing more complexity. One of these requirements is low crosstalk. In the right-most picture of Fig. 6.29(b), we can compare the coupler design to our previous planar design in Fig. 6.1(b). In the flip-chip device, each quadrant of the coupler structure is a 60 μm$\times$60 μm loop, compared to the of the coupler was reduced from 100 μm$\times$100 μm loops of the planar device, meaning a factor of 3 reduction in areal footprint while keeping roughly the same M_{ca} and M_{cb}. If we were to have made the same reduction in coupler size in the planar design without flip-chip, the crosstalk would have been much higher given that coupler bias lines would come even closer to the qubits. However, we observe that the crosstalk from the coupler bias lines to the qubits is the same or better than in

our best planar device. The fact that we maintained the same level of crosstalk while making the design significantly more compact is one of many benefits of utilizing a flip-chip. This was possible because of the alternative current paths for the return current on the flip chip. Note also that we were able to partially shield the coupler loop with ground plane on the flip chip to help with crosstalk, but didn't need to fill the entire coupler loop with a ground plane island as we did in Fig. 6.1(b), allowing us to maintain the same qubit-coupler coupling strength as the previous device as well.

The other requirement for the flip-chip architecture is that it retains at least the same level of coherence as the planar architecture. On the first test device we observed a roughly 30% reduction in T_1 at zero tilt relative to typical planar devices for both qubit A and qubit B, as well as for the pair of coupled qubits. Because qubit A and B have very different participations of the return current through the bump bonds and carrier chip, it is unlikely that the bumps themselves played a role in this reduction. Rather, it is possible that other systematic differences in the fabrication process led to this change. This is likely the explanation, since in a second flip-chip sample with the bias lines completely on the carrier chip, we see the usual T_1's consistent with single-chip silicon devices. In addition, we see no measurable degradation in T_2 or low-frequency flux noise in this device relative to non-flip chip devices. This makes sense since, based on surface magnetic participation ratios, we would expect a flip chip qubit with ground plane above it to have only a 20% increase in flux noise power, small enough that it wouldn't be noticed in a low-frequency noise measurement with a small sample size. Given the

maintained coupling strength and coherence, we are optimistic about moving forward with the flip-chip architecture to build more complex medium-scale quantum annealing devices.

Chapter 7

Conclusion

We have demonstrated a flux qubit suitable for high-connectivity quantum annealing without the use of lossy dielectrics. Removing amorphous dielectrics allowed for a $\sim 100\times$ improvement in high-frequency microwave dissipation compared to commercial flux qubits fabricated with lossy dielectrics. However, at lower frequencies we found that flux noise dominates the dissipation, and in particular found that $1/f$ flux noise from a paramagnetic environment can extend up to GHz frequencies. Nevertheless, compared to flux qubits in currently available commercial quantum annealers, we find a relatively small energy offset ε_p describing the environmental energy penalty for incoherent resonant quantum tunneling between flux qubit wells, indicating that multi-qubit tunneling should not be as significantly suppressed by dissipation when scaling up. However, work still needs to be done to improve the low-frequency noise in order to increase the chances for coherent tunneling to occur in an annealer. Other features of our hardware include

the ability to "anneal in reverse" for quantum parallel tempering, and microwave spectroscopy that enables precise calibration of the qubit using a physically realistic model. In addition to the fluxmon itself, we also developed tunable couplers enabling ultra-strong inter-qubit coupling that can be described by a scalable nonlinear Born-Oppenheimer theory. With the reduced coupler footprint made possible by a flip-chip architecture, the fluxmon is in principle able to strongly couple to up to 20 other qubits at once, as desired for a dense connectivity graph. Finally, our control system allowed us to perform one- and two-qubit quantum annealing experiments faster than the system thermalization time. We found that in this regime, annealing fast drastically reduces thermalization errors, leading to improved annealing success rates for almost all one- and two-qubit problem instances (whenever non-adiabatic effects were not dominant). With faster annealing and longer thermalization times in the fluxmon architecture, it may be possible to largely eliminate the "thermalization stage" observed in currently available commercial quantum annealers.

Bibliography

[1] Nicos Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Management Sciences Research Report 388, Graduate School of Industrial Aministration, CMU, 1976.

[2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

[3] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[4] Vadim N. Smelyanskiy, Eleanor G. Rieffel, Sergey I. Knysh, Colin P. Williams, Mark W. Johnson, Murray C. Thom, G. Macready, William, and Kristen L. Pudenz. A Near-Term Quantum Computing Approach for Hard Computational Problems in Space Exploration. *arXiv:1204.2821*, 2012.

[5] P. Ray, B. K. Chakrabarti, and Arunava Chakrabarti. Sherrington-Kirkpatrick model in a transverse field: Absence of replica symmetry breaking due to quantum fluctuations. *Phys. Rev. B*, 39:11828–11832, Jun 1989.

[6] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll. Quantum annealing: A new method for minimizing multidimensional functions. *Chem. Phys. Lett.*, 219(5):343 – 348, 1994.

[7] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse Ising model. *Phys. Rev. E*, 58:5355–5363, Nov 1998.

[8] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem. *Science*, 292(5516):472–475, 2001.

[9] M. Born and V. Fock. Beweis des Adiabatensatzes. *Z. Phys*, 51:165–180, 1928.

[10] Thomas Jörg, Florent Krzakala, Guilhem Semerjian, and Francesco Zamponi. First-Order Transitions and the Performance of Quantum Algorithms in Random Optimization Problems. *Phys. Rev. Lett.*, 104:207206, May 2010.

[11] Hartmut Neven. Opening Remarks: Why We Believe Quantum Annealing Will Succeed. In *Adiabatic Quantum Computing Conference 2016*, 2016.

[12] Kenichi Kurihara, Shu Tanaka, and Seiji Miyashita. Quantum annealing for clustering. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 321–328. AUAI Press, 2009.

[13] Zheng Zhu, Chao Fang, and Helmut G. Katzgraber. *borealis* - A generalized global update algorithm for Boolean optimization problems. *arXiv:physics.comp-ph/1605.09399*, 2016.

[14] Masoud Mohseni, Peter Read, Hartmut Neven, Sergio Boixo, Vasil Denchev, Ryan Babbush, Austin Fowler, Vadim Smelyanskiy, and John Martinis. Commercialize quantum technologies in five years. *Nature Comment*, 2017.

[15] Michel H. Devoret and John M. Martinis. Implementing Qubits with Superconducting Integrated Circuits. *Quantum Information Processing*, 3(1):163–203, Oct 2004.

[16] R. Harris, J. Johansson, A. J. Berkley, M. W. Johnson, T. Lanting, Siyuan Han, P. Bunyk, E. Ladizinsky, T. Oh, I. Perminov, E. Tolkacheva, S. Uchaikin, E. M. Chapple, C. Enderud, C. Rich, M. Thom, J. Wang, B. Wilson, and G. Rose. Experimental demonstration of a robust and scalable flux qubit. *Phys. Rev. B*, 81:134510, Apr 2010.

[17] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, May 2011.

[18] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, and J. Whittaker. Architectural Considerations in the Design of a Superconducting Quantum Annealing Processor. *IEEE Transactions on Applied Superconductivity*, 24(4):1–10, Aug 2014.

[19] Sergio Boixo, Vadim N Smelyanskiy, Alireza Shabani, Sergei V Isakov, Mark Dykman, Vasil S Denchev, Mohammad H Amin, Anatoly Yu Smirnov, Masoud Mohseni, and Hartmut Neven. Computational multiqubit tunnelling in programmable quantum annealers. *Nature communications*, 7, 2016.

[20] R. Harris, M. W. Johnson, S. Han, A. J. Berkley, J. Johansson, P. Bunyk, E. Ladizinsky, S. Govorkov, M. C. Thom, S. Uchaikin, B. Bumble, A. Fung, A. Kaul, A. Kleinsasser, M. H. S. Amin, and D. V. Averin. Probing Noise in Flux Qubits via Macroscopic Resonant Tunneling. *Phys. Rev. Lett.*, 101:117003, Sep 2008.

[21] Richard F. Voss and Richard A. Webb. Macroscopic Quantum Tunneling in 1-μm Nb Josephson Junctions. *Phys. Rev. Lett.*, 47:265–268, Jul 1981.

[22] Michel H. Devoret, John M. Martinis, and John Clarke. Measurements of Macroscopic Quantum Tunneling out of the Zero-Voltage State of a Current-Biased Josephson Junction. *Phys. Rev. Lett.*, 55:1908–1911, Oct 1985.

[23] Vasil S. Denchev, Sergio Boixo, Sergei V. Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven. What is the Computational Value of Finite-Range Tunneling? *Phys. Rev. X*, 6:031015, Aug 2016.

[24] Troels F. Rønnow, Zhihui Wang, Joshua Job, Sergio Boixo, Sergei V. Isakov, David Wecker, John M. Martinis, Daniel A. Lidar, and Matthias Troyer. Defining and detecting quantum speedup. *Science*, 345(6195):420–424, 2014.

[25] Tameem Albash and Daniel A Lidar. Evidence for a Limited Quantum Speedup on a Quantum Annealer. *arXiv:1705.07452*, 2017.

[26] Sergei V. Isakov, Guglielmo Mazzola, Vadim N. Smelyanskiy, Zhang Jiang, Sergio Boixo, Hartmut Neven, and Matthias Troyer. Understanding Quantum Tunneling through Quantum Monte Carlo Simulations. *Phys. Rev. Lett.*, 117:180402, Oct 2016.

[27] Hartmut Neven. How can Artificial Intelligence benefit from Quantum Resources? In *The Science of Consciousness Conference 2016*, 2016.

[28] Alex Selby. Efficient subgraph-based sampling of Ising-type models with frustration. *arXiv:1409.3934*, 2014.

[29] Evgeny Andriyash and Mohammad H. Amin. Can quantum Monte Carlo simulate quantum annealing? *arXiv:1703.09277*, 2017.

[30] Mohammad H. Amin. Searching for quantum speedup in quasistatic quantum annealers. *Phys. Rev. A*, 92:052323, Nov 2015.

[31] A. G. Fowler, C. Quintana, and Yu Chen. Adiabatic architecture framework. Internal note, 2016.

[32] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of Superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.

[33] B.D. Josephson. Possible new effects in superconductive tunnelling. *Phys. Lett.*, 1(7):251, 1962.

[34] Michael Tinkham. *Introduction to Superconductivity*. Dover, second edition, 2004.

[35] L. D. Jackel, J. P. Gordon, E. L. Hu, R. E. Howard, L. A. Fetter, D. M. Tennant, R. W. Epworth, and J. Kurkijärvi. Decay of the Zero-Voltage State in Small-Area, High-Current-Density Josephson Junctions. *Phys. Rev. Lett.*, 47:697–700, Aug 1981.

[36] John M. Martinis, Michel H. Devoret, and John Clarke. Energy-Level Quantization in the Zero-Voltage State of a Current-Biased Josephson Junction. *Phys. Rev. Lett.*, 55:1543–1546, Oct 1985.

[37] Wikipedia. Magnetoencephalography — Wikipedia, The Free Encyclopedia, 2016. [Online; accessed 28-December-2016].

[38] J. Bardeen. Quantization of Flux in a Superconducting Cylinder. *Phys. Rev. Lett.*, 7:162–163, Sep 1961.

[39] A. Megrant, C. Neill, R. Barends, B. Chiaro, Yu Chen, L. Feigl, J. Kelly, Erik Lucero, Matteo Mariantoni, P. J. J. O'Malley, D. Sank, A. Vainsencher, J. Wenner, T. C. White, Y. Yin, J. Zhao, C. J. Palmstrøm, John M. Martinis, and A. N. Cleland. Planar superconducting resonators with internal quality factors above one million. *Appl. Phys. Lett.*, 100(11):113510, 2012.

[40] R. Barends, J. Kelly, A. Megrant, D. Sank, E. Jeffrey, Y. Chen, Y. Yin, B. Chiaro, J. Mutus, C. Neill, P. O'Malley, P. Roushan, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis. Coherent Josephson Qubit Suitable for Scalable Quantum Integrated Circuits. *Phys. Rev. Lett.*, 111:080502, Aug 2013.

[41] D. M. Pozar. *Microwave Engineering*. John Wiley & Sons, third edition, 2005.

[42] G. Hasnain, A. Dienes, and J. R. Whinnery. Dispersion of Picosecond Pulses in Coplanar Transmission Lines. *IEEE Transactions on Microwave Theory and Techniques*, 34(6):738–741, Jun 1986.

[43] David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson, 2nd edition, 2005.

[44] Jens Koch, Terri M. Yu, Jay Gambetta, A. A. Houck, D. I. Schuster, J. Majer, Alexandre Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A*, 76:042319, Oct 2007.

[45] Jonathan R Friedman, Vijay Patel, Wei Chen, SK Tolpygo, and James E Lukens. Quantum superposition of distinct macroscopic states. *nature*, 406(6791):43–46, 2000.

[46] J. Clarke and F. K. Wilhelm. Superconducting quantum bits. *Nature*, 453(7198):1031–1042, 2008.

[47] J. E. Mooij, T. P. Orlando, L. Levitov, Lin Tian, Caspar H. van der Wal, and Seth Lloyd. Josephson persistent-current qubit. *Science*, 285(5430):1036–1039, 1999.

[48] T. P. Orlando, J. E. Mooij, Lin Tian, Caspar H. van der Wal, L. S. Levitov, Seth Lloyd, and J. J. Mazo. Superconducting persistent-current qubit. *Phys. Rev. B*, 60:15398–15413, Dec 1999.

[49] L. D. Landau and E. M. Lifshitz. *Course of Theoretical Physics Vol 3 Quantum Mechanics*. Pergamon Press, 1958.

[50] J. A. Schreier, A. A. Houck, Jens Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B*, 77:180502, May 2008.

[51] M. Wallquist, V. S. Shumeiko, and G. Wendin. Selective coupling of superconducting charge qubits mediated by a tunable stripline cavity. *Phys. Rev. B*, 74:224506, Dec 2006.

[52] Chris Quintana. Josephson Parametric Amplification for Circuit Quantum Electrodynamics: Theory and Implementation. Bachelor's thesis, Princeton University, 2013.

[53] W. Wustmann and V. Shumeiko. Parametric resonance in tunable superconducting cavities. *Phys. Rev. B*, 87:184501, May 2013.

[54] Louis N Hand and Janet D Finch. *Analytical mechanics*. Cambridge University Press, 1998.

[55] Dvir Kafri, Chris Quintana, Yu Chen, Alireza Shabani, John M. Martinis, and Hartmut Neven. Tunable inductive coupling of superconducting qubits in the strongly nonlinear regime. *Phys. Rev. A*, 95:052333, May 2017.

[56] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Academic press, 2014.

[57] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[58] Julian S. Kelly. *Fault-tolerant superconducting qubits*. PhD thesis, University of California, Santa Barbara, March 2015.

[59] C. M. Quintana, A. Megrant, Z. Chen, A. Dunsworth, B. Chiaro, R. Barends, B. Campbell, Yu Chen, I.-C. Hoi, E. Jeffrey, J. Kelly, J. Y. Mutus, P. J. J. O'Malley, C. Neill, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis. Characterization and reduction of microfabrication-induced decoherence in superconducting quantum circuits. *Appl. Phys. Lett.*, 105:062601, 2014.

[60] A. Dunsworth, A. Megrant, C. Quintana, Zijun Chen, R. Barends, B. Burkett, B. Foxen, Yu Chen, B. Chiaro, A. Fowler, R. Graff, E. Jeffrey, J. Kelly, E. Lucero, J. Y. Mutus, M. Neeley, C. Neill, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, and John M. Martinis. Characterization and reduction of capacitive loss induced by sub-micron Josephson junction fabrication in superconducting qubits. *Applied Physics Letters*, 111(2):022601, 2017.

[61] John M. Martinis, K. B. Cooper, R. McDermott, Matthias Steffen, Markus Ansmann, K. D. Osborn, K. Cicak, Seongshik Oh, D. P. Pappas, R. W. Simmonds, and Clare C. Yu. Decoherence in Josephson Qubits from Dielectric Loss. *Phys. Rev. Lett.*, 95:210503, Nov 2005.

[62] Fei Yan, Simon Gustavsson, Archana Kamal, Jeffrey Birenbaum, Adam P. Sears, David Hover, Ted J. Gudmundsen, Danna Rosenberg, Gabriel Samach, S. Weber, Jonilyn L. Yoder, Terry P. Orlando, John Clarke, Andrew J. Kerman, and William D. Oliver. The flux qubit revisited to enhance coherence and reproducibility. *Nature Communications*, 7:12964, November 2016.

[63] Floortje Geertruida Paauw. *Superconducting flux qubits: Quantum chains and tunable qubits*. PhD thesis, TU Delft, 2009.

[64] S. J. Weber, G. O. Samach, D. Hover, S. Gustavsson, D. K. Kim, D. Rosenberg, A. P. Sears, F. Yan, J. L. Yoder, W. D. Oliver, and A. J. Kerman. Coherent coupled qubits for quantum annealing. *arXiv:1701.06544v1*, 2017.

[65] Rainee N Simons. *Coplanar waveguide circuits, components, and systems*, volume 165. John Wiley & Sons, 2004.

[66] Zijun Chen, A. Megrant, J. Kelly, R. Barends, J. Bochmann, Yu Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Y. Mutus, P. J. J. O'Malley, C. Neill, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis. Fabrication and characterization of aluminum airbridges for superconducting microwave circuits. *Applied Physics Letters*, 104(5), 2014.

[67] A. Dunsworth, A. Megrant, C. Quintana, Zijun Chen, R. Barends, B. Campbell, Yu Chen, B. Chiaro, A. Fowler, E. Jeffrey, J. Kelly, E. Lucero, J. Y. Mutus, P. J. J.

O'Malley, C. Neill, M. Neeley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, and John M. Martinis. Fabrication and characterization of silicon oxide scaffolded aluminum airbridges for superconducting microwave devices. *(in preparation)*, 2017.

[68] M. Lenander, H. Wang, Radoslaw C. Bialczak, Erik Lucero, Matteo Mariantoni, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, T. Yamamoto, Y. Yin, J. Zhao, A. N. Cleland, and John M. Martinis. Measurement of energy decay in superconducting qubits from nonequilibrium quasiparticles. *Phys. Rev. B*, 84:024501, Jul 2011.

[69] Matthias Steffen, Shwetank Kumar, David DiVincenzo, George Keefe, Mark Ketchen, Mary Beth Rothwell, and Jim Rozen. Readout for phase qubits without Josephson junctions. *Applied Physics Letters*, 96(10), 2010.

[70] Yu Chen, D. Sank, P. O'Malley, T. White, R. Barends, B. Chiaro, J. Kelly, E. Lucero, M. Mariantoni, A. Megrant, C. Neill, A. Vainsencher, J. Wenner, Y. Yin, A. N. Cleland, and John M. Martinis. Multiplexed dispersive readout of superconducting phase qubits. *Applied Physics Letters*, 101(18), 2012.

[71] C. M. Quintana, Yu Chen, D. Sank, A. G. Petukhov, T. C. White, Dvir Kafri, B. Chiaro, A. Megrant, R. Barends, B. Campbell, Z. Chen, A. Dunsworth, A. G. Fowler, R. Graff, E. Jeffrey, J. Kelly, E. Lucero, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, A. Shabani, V. N. Smelyanskiy, A. Vainsencher, J. Wenner, H. Neven, and John M. Martinis. Observation of Classical-Quantum Crossover of $1/f$ Flux Noise and Its Paramagnetic Temperature Dependence. *Phys. Rev. Lett.*, 118:057702, Jan 2017.

[72] Alexandre Blais, Ren-Shou Huang, Andreas Wallraff, S. M. Girvin, and R. J. Schoelkopf. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A*, 69:062320, Jun 2004.

[73] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature*, 431(7005):162–167, September 2004.

[74] 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 J. M. Martinis. Strong environmental coupling in a Josephson parametric amplifier. *Applied Physics Letters*, 104(26):263513, 2014.

[75] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and John M. Martinis. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519(7541):66–69, March 2015.

[76] L.D. Landau. Zur theorie der energieübertragung bei stössen. *Phys. Z. Sowjetunion*, 2:46, 1932.

[77] C. Zener. Non-Adiabatic Crossing of Energy Levels. *Proc. R. Soc. A*, 137:696–702, September 1932.

[78] S. N. Shevchenko, S. Ashhab, and F. Nori. Landau–Zener–Stückelberg interferometry. *Phys. Rep.*, 492(1):1–30, 2010.

[79] 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. *Phys. Rev. Lett.*, 117:190503, Nov 2016.

[80] David Zueco, Georg M. Reuther, Sigmund Kohler, and Peter Hänggi. Qubit-oscillator dynamics in the dispersive regime: Analytical theory beyond the rotating-wave approximation. *Phys. Rev. A*, 80:033846, Sep 2009.

[81] B. A. Mazin. *Microwave Kinetic Inductance Detectors*. PhD thesis, California Institute of Technology, 2004.

[82] E. M. Purcell. Proceedings of the American Physical Society. *Phys. Rev.*, 69:674–674, Jun 1946.

[83] A. A. Houck, J. A. Schreier, B. R. Johnson, J. M. Chow, Jens Koch, J. M. Gambetta, D. I. Schuster, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Controlling the Spontaneous Emission of a Superconducting Transmon Qubit. *Phys. Rev. Lett.*, 101, 2008.

[84] M. D. Reed, B. R. Johnson, A. A. Houck, L. DiCarlo, J. M. Chow, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf. Fast reset and suppressing spontaneous emission of a superconducting qubit. *Applied Physics Letters*, 96(20):203110, 2010.

[85] Evan Jeffrey, Daniel Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O'Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and John M. Martinis. Fast Accurate State Measurement with Superconducting Qubits. *Phys. Rev. Lett.*, 112:190504, May 2014.

[86] A. P. Sears, A. Petrenko, G. Catelani, L. Sun, Hanhee Paik, G. Kirchmair, L. Frunzio, L. I. Glazman, S. M. Girvin, and R. J. Schoelkopf. Photon shot noise dephasing in the strong-dispersive limit of circuit QED. *Phys. Rev. B*, 86:180504, Nov 2012.

[87] 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. *Phys. Rev. B*, 86:100506, Sep 2012.

[88] Jay Gambetta, Alexandre Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting. *Phys. Rev. A*, 74:042318, Oct 2006.

[89] A.O Caldeira and A.J Leggett. Quantum tunnelling in a dissipative system. *Annals of Physics*, 149(2):374 – 456, 1983.

[90] Daniel Esteve, Michel H. Devoret, and John M. Martinis. Effect of an arbitrary dissipative circuit on the quantum energy levels and tunneling of a Josephson junction. *Phys. Rev. B*, 34:158–163, Jul 1986.

[91] John M. Martinis, Michel H. Devoret, and John Clarke. Experimental tests for the quantum behavior of a macroscopic degree of freedom: The phase difference across a Josephson junction. *Phys. Rev. B*, 35:4682–4698, Apr 1987.

[92] M. H. S. Amin and Dmitri V. Averin. Macroscopic Resonant Tunneling in the Presence of Low Frequency Noise. *Phys. Rev. Lett.*, 100:197001, May 2008.

[93] Michael Fang. Development of Hardware for Scaling Up Superconducting Qubits and Simulation of Quantum Chaos. Bachelor's thesis, University of California, Santa Barbara, 2015.

[94] J. Wenner, M. Neeley, Radoslaw C. Bialczak, M. Lenander, Erik Lucero, A. D. O'Connell, D. Sank, H. Wang, M. Weides, A. N. Cleland, and John M. Martinis. Wirebond crosstalk and cavity modes in large chip mounts for superconducting qubits. *Superconductor Science and Technology*, 24(6):065001, 2011.

[95] R. Barends, J. Wenner, M. Lenander, Y. Chen, R. C. Bialczak, J. Kelly, E. Lucero, P. O'Malley, M. Mariantoni, D. Sank, H. Wang, T. C. White, Y. Yin, J. Zhao, A. N. Cleland, John M. Martinis, and J. J. A. Baselmans. Minimizing quasiparticle generation from stray infrared light in superconducting quantum circuits. *Appl. Phys. Lett.*, 99(11):–, 2011.

[96] Martinis group - http://web.physics.ucsb.edu/~martinisgroup/electronics.shtml.

[97] D. T. Sank. *Fast, Accurate State Measurement in Superconducting Qubits*. PhD thesis, University of California, Santa Barbara, 2014.

[98] Barends. R. *Photon-detecting superconducting resonators*. PhD thesis, TU Delft, 2009.

[99] C. Neill, A. Megrant, R. Barends, Yu Chen, B. Chiaro, J. Kelly, J. Y. Mutus, P. J. J. O'Malley, D. Sank, J. Wenner, T. C. White, Yi Yin, A. N. Cleland, and John M. Martinis. Fluctuations from edge defects in superconducting resonators. *Appl. Phys. Lett.*, 103(7):–, 2013.

[100] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[101] Vinay Ambegaokar and Alexis Baratoff. Tunneling Between Superconductors. *Phys. Rev. Lett.*, 11:104–104, Jul 1963.

[102] G. Lindblad. On the generators of quantum dynamical semigroups. *Comm. Math. Phys.*, 48(2):119–130, 1976.

[103] L. S. Bishop. *Circuit Quantum Electrodynamics*. PhD thesis, Yale University, 2010.

[104] C. M. Quintana, K. D. Petersson, L. W. McFaul, S. J. Srinivasan, A. A. Houck, and J. R. Petta. Cavity-Mediated Entanglement Generation Via Landau-Zener Interferometry. *Phys. Rev. Lett.*, 110:173603, Apr 2013.

[105] Walter Vinci and Daniel A. Lidar. Non-Stoquastic Interactions in Quantum Annealing via the Aharonov-Anandan Phase. *arXiv:1701.07494*, 2017.

[106] John Clarke, Wolfgang M. Goubau, and Mark B. Ketchen. Tunnel junction dc SQUID: Fabrication, operation, and performance. *Journal of Low Temperature Physics*, 25(1):99–144, 1976.

[107] Roger H. Koch, John Clarke, W. M. Goubau, J. M. Martinis, C. M. Pegrum, and D. J. van Harlingen. Flicker (1/f) noise in tunnel junction dc SQUIDS. *Journal of Low Temperature Physics*, 51(1):207–224, 1983.

[108] Frederick C. Wellstood, Cristian Urbina, and John Clarke. Low-frequency noise in dc superconducting quantum interference devices below 1 K. *Applied Physics Letters*, 50(12):772–774, 1987.

[109] F. Yoshihara, K. Harrabi, A. O. Niskanen, Y. Nakamura, and J. S. Tsai. Decoherence of Flux Qubits due to $1/f$ Flux Noise. *Phys. Rev. Lett.*, 97:167001, Oct 2006.

[110] Radoslaw C. Bialczak, R. McDermott, M. Ansmann, M. Hofheinz, N. Katz, Erik Lucero, Matthew Neeley, A. D. O'Connell, H. Wang, A. N. Cleland, and John M. Martinis. $1/f$ Flux Noise in Josephson Phase Qubits. *Phys. Rev. Lett.*, 99:187006, Nov 2007.

[111] S. Sendelbach, D. Hover, A. Kittel, M. Mück, John M. Martinis, and R. McDermott. Magnetism in SQUIDs at Millikelvin Temperatures. *Phys. Rev. Lett.*, 100:227006, Jun 2008.

[112] T. Lanting, A. J. Berkley, B. Bumble, P. Bunyk, A. Fung, J. Johansson, A. Kaul, A. Kleinsasser, E. Ladizinsky, F. Maibaum, R. Harris, M. W. Johnson, E. Tolkacheva, and M. H. S. Amin. Geometrical dependence of the low-frequency noise in superconducting flux qubits. *Phys. Rev. B*, 79:060509, Feb 2009.

[113] S. M. Anton, J. S. Birenbaum, S. R. O'Kelley, V. Bolkhovsky, D. A. Braje, G. Fitch, M. Neeley, G. C. Hilton, H.-M. Cho, K. D. Irwin, F. C. Wellstood, W. D. Oliver, A. Shnirman, and John Clarke. Magnetic Flux Noise in dc SQUIDs: Temperature and Geometry Dependence. *Phys. Rev. Lett.*, 110:147002, Apr 2013.

[114] Jonas Bylander, Simon Gustavsson, Fei Yan, Fumiki Yoshihara, Khalil Harrabi, George Fitch, David G Cory, Yasunobu Nakamura, Jaw-Shen Tsai, and William D Oliver. Noise spectroscopy through dynamical decoupling with a superconducting flux qubit. *Nature Physics*, 7(7):565–570, 2011.

[115] Daniel Sank, R. Barends, Radoslaw C. Bialczak, Yu Chen, J. Kelly, M. Lenander, E. Lucero, Matteo Mariantoni, A. Megrant, M. Neeley, P. J. J. O'Malley, A. Vainsencher, H. Wang, J. Wenner, T. C. White, T. Yamamoto, Yi Yin, A. N. Cleland, and John M. Martinis. Flux Noise Probed with Real Time Qubit Tomography in a Josephson Phase Qubit. *Phys. Rev. Lett.*, 109:067001, Aug 2012.

[116] Fei Yan, Jonas Bylander, Simon Gustavsson, Fumiki Yoshihara, Khalil Harrabi, David G. Cory, Terry P. Orlando, Yasunobu Nakamura, Jaw-Shen Tsai, and William D. Oliver. Spectroscopy of low-frequency noise and its temperature dependence in a superconducting qubit. *Phys. Rev. B*, 85:174521, May 2012.

[117] D. H. Slichter, R. Vijay, S. J. Weber, S. Boutin, M. Boissonneault, J. M. Gambetta, A. Blais, and I. Siddiqi. Measurement-Induced Qubit State Mixing in Circuit QED from Up-Converted Dephasing Noise. *Phys. Rev. Lett.*, 109:153601, Oct 2012.

[118] Fei Yan, Simon Gustavsson, Jonas Bylander, Xiaoyue Jin, Fumiki Yoshihara, David G Cory, Yasunobu Nakamura, Terry P Orlando, and William D Oliver. Rotating-frame relaxation as a noise spectrum analyser of a superconducting qubit undergoing driven evolution. *Nature communications*, 4, 2013.

[119] Fumiki Yoshihara, Yasunobu Nakamura, Fei Yan, Simon Gustavsson, Jonas Bylander, William D. Oliver, and Jaw-Shen Tsai. Flux qubit noise spectroscopy using

Rabi oscillations under strong driving conditions. *Phys. Rev. B*, 89:020503, Jan 2014.

[120] Stefan Machlup. Noise in Semiconductors: Spectrum of a Two-Parameter Random Signal. *Journal of Applied Physics*, 25(3):341–343, 1954.

[121] P. Dutta and P. M. Horn. Low-frequency fluctuations in solids: 1/f noise. *Reviews of Modern Physics*, 53:497–516, July 1981.

[122] Norman F. Ramsey. A Molecular Beam Resonance Method with Separated Oscillating Fields. *Phys. Rev.*, 78:695–699, Jun 1950.

[123] John M. Martinis, S. Nam, J. Aumentado, K. M. Lang, and C. Urbina. Decoherence of a superconducting qubit due to bias noise. *Phys. Rev. B*, 67:094510, Mar 2003.

[124] E. L. Hahn. Spin Echoes. *Phys. Rev.*, 80:580–594, Nov 1950.

[125] H. Y. Carr and E. M. Purcell. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.*, 94:630–638, May 1954.

[126] S. Meiboom and D. Gill. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments*, 29(8):688–691, 1958.

[127] A. G. Petukhov, V. N. Smelyanskiy and J. M. Martinis (unpublished).

[128] P. Kumar, S. Sendelbach, M. A. Beck, J. W. Freeland, Zhe Wang, Hui Wang, Clare C. Yu, R. Q. Wu, D. P. Pappas, and R. McDermott. Origin and Reduction of $1/f$ Magnetic Flux Noise in Superconducting Devices. *Phys. Rev. Applied*, 6:041001, Oct 2016.

[129] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S.M. Girvin, and R. J. Schoelkopf. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature*, 460(7252):240–244, 2009.

[130] Zijun Chen, Julian Kelly, Chris Quintana, R. Barends, B. Campbell, Yu Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and John M. Martinis. Measuring and Suppressing Quantum State Leakage in a Superconducting Qubit. *Phys. Rev. Lett.*, 116:020501, Jan 2016.

[131] A. J. Leggett. Quantum tunneling in the presence of an arbitrary linear dissipation mechanism. *Phys. Rev. B*, 30:1208–1218, Aug 1984.

[132] N. G. Dickson, M. W. Johnson, M. H. Amin, R. Harris, F. Altomare, A. J. Berkley, P. Bunyk, J. Cai, E. M. Chapple, P. Chavez, F. Cioata, T. Cirip, P. deBuen, M. Drew-Brook, C. Enderud, S. Gildert, F. Hamze, J. P. Hilton, E. Hoskinson, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Lanting, T. Mahon, R. Neufeld, T. Oh, I. Perminov, C. Petroff, A. Przybysz, C. Rich, P. Spear, A. Tcaciuc, M. C. Thom, E. Tolkacheva, S. Uchaikin, J. Wang, A. B. Wilson, Z. Merali, and G. Rose. Thermally assisted quantum annealing of a 16-qubit problem. *Nature Communications*, 4:1903, May 2013.

[133] Paul Horowitz, Winfield Hill, and Virgil Elings. *The art of electronics*. AIP, 1981.

[134] F Wilhelm, M Storcz, U Hartmann, and M Geller. Superconducting qubits II: Decoherence. *Manipulating Quantum Coherence in Solid State Systems*, pages 195–232, 2007.

[135] V. Weisskopf and E. Wigner. Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie. *Zeitschrift fur Physik*, 63:54–73, January 1930.

[136] M. O. Scully and M. S. Zubairy. *Quantum Optics*. Cambridge University Press, 1997.

[137] R.P Feynman and F.L Vernon. The theory of a general quantum system interacting with a linear dissipative system. *Annals of Physics*, 24:118 – 173, 1963.

[138] Michel H Devoret. Quantum fluctuations in electrical circuits. *Les Houches, Session LXIII*, 7(8), 1997.

[139] Walter Appel and Emmanuel Kowalski. *Mathematics for physics and physicists*. Princeton University Press Princeton, NJ, USA; Oxford, UK, 2007.

[140] R Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255, 1966.

[141] Robert J. Schoelkopf, A. A. Clerk, S. M. Girvin, Konrad W. Lehnert, and M. H. Devoret. Noise and measurement backaction in superconducting circuits: qubits as spectrometers of quantum noise. *Proc. SPIE*, 5115:356–376, 2003.

[142] DanielSank (https://physics.stackexchange.com/users/31790/danielsank). What information is contained in the quantum spectral density? Physics Stack Exchange. URL:https://physics.stackexchange.com/q/162180 (version: 2015-01-29).

[143] H. Wang, M. Hofheinz, M. Ansmann, R. C. Bialczak, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, J. Wenner, A. N. Cleland, and John M. Martinis. Measurement of the Decay of Fock States in a Superconducting Quantum Circuit. *Phys. Rev. Lett.*, 101:240401, Dec 2008.

[144] David I. Schuster. *Circuit Quantum Electrodynamics*. PhD thesis, Yale University, 2007.

[145] J. Wenner, R. Barends, R. C. Bialczak, Yu Chen, J. Kelly, Erik Lucero, Matteo Mariantoni, A. Megrant, P. J. J. O'Malley, D. Sank, A. Vainsencher, H. Wang, T. C. White, Y. Yin, J. Zhao, A. N. Cleland, and John M. Martinis. Surface loss simulations of superconducting coplanar waveguide. *Appl. Phys. Lett.*, 99:113513, 2011.

[146] Jiansong Gao. *The Physics of Superconducting Microwave Resonators*. PhD thesis, California Institute of Technology, 2008.

[147] John M. Martinis, M. Ansmann, and J. Aumentado. Energy Decay in Superconducting Josephson-Junction Qubits from Nonequilibrium Quasiparticle Excitations. *Phys. Rev. Lett.*, 103:097002, Aug 2009.

[148] W. A. Phillips. Tunneling states in amorphous solids. *J. Low Temp. Phys.*, 7:351–360, May 1972.

[149] J. Gao, L. R. Vale, J. A. B. Mates, D. R. Schmidt, G. C. Hilton, K. D. Irwin, F. Mallet, M. A. Castellanos-Beltran, K. W. Lehnert, J. Zmuidzinas, and H. G. Leduc. Strongly quadrature-dependent noise in superconducting microresonators measured at the vacuum-noise limit. *Applied Physics Letters*, 98(23):232508, 2011.

[150] John M Martinis and A Megrant. UCSB final report for the CSQ program: Review of decoherence and materials physics for superconducting qubits. *arXiv preprint arXiv:1410.5793*, 2014.

[151] H. Wang, M. Hofheinz, J. Wenner, M. Ansmann, R. C. Bialczak, M. Lenander, Erik Lucero, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, A. N. Cleland, and John M. Martinis. Improving the coherence time of superconducting coplanar resonators. *Appl. Phys. Lett.*, 95(23):–, 2009.

[152] W. A. Phillips. Two-level states in glasses. *Reports on Progress in Physics*, 50(12):1657, 1987.

[153] Jiansong Gao, Miguel Daal, Anastasios Vayonakis, Shwetank Kumar, Jonas Zmuidzinas, Bernard Sadoulet, Benjamin A. Mazin, Peter K. Day, and Henry G. Leduc. Experimental evidence for a surface distribution of two-level systems in superconducting lithographed microwave resonators. *Appl. Phys. Lett.*, 92(15):–, 2008.

[154] M. S. Khalil, M. J. A. Stoutimore, S. Gladchenko, A. M. Holder, C. B. Musgrave, A. C. Kozen, G. Rubloff, Y. Q. Liu, R. G. Gordon, J. H. Yum, S. K. Banerjee, C. J. Lobb, and K. D. Osborn. Evidence for hydrogen two-level systems in atomic layer deposition oxides. *Appl. Phys. Lett.*, 103(16):–, 2013.

[155] Aaron M. Holder, Kevin D. Osborn, C. J. Lobb, and Charles B. Musgrave. Bulk and Surface Tunneling Hydrogen Defects in Alumina. *Phys. Rev. Lett.*, 111:065901, Aug 2013.

[156] J. R. Jameson, D. Ngo, C. Benko, J. P. McVittie, Y. Nishi, and B. A. Young. Dielectric relaxation study of hydrogen exposure as a source of two-level systems in Al2O3. *J. Non-Cryst. Solids*, 357:2148, 2011.

[157] P. J. Koppinen, L. M. Väistö, and I. J. Maasilta. Complete stabilization and improvement of the characteristics of tunnel junctions by thermal annealing. *Appl. Phys. Lett.*, 90(5):–, 2007.

[158] I. M. Pop, T. Fournier, T. Crozes, F. Lecocq, I. Matei, B. Pannetier, O. Buisson, and W. Guichard. Fabrication of stable and reproducible submicron tunnel junctions. *J. Vac. Sci. Technol. B*, 30(010607), 2012.

[159] D. S. Macintyre, O. Ignatova, S. Thoms, and I. G. Thayne. Resist residues and transistor gate fabrication. *J. Vac. Sci. Technol. B*, 27(6):2597–2601, 2009.

[160] K. Geerlings, S. Shankar, E. Edwards, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret. Improving the quality factor of microwave compact resonators by optimizing their geometrical parameters. *Appl. Phys. Lett.*, 100(19):–, 2012.

[161] C. M. Egert. Roughness evolution of optical materials induced by ion-beam milling. *Proc. SPIE*, 1752:63–72, 1992.

[162] Matteo Mariantoni, H Wang, Radoslaw C Bialczak, M Lenander, Erik Lucero, M Neeley, AD O'Connell, D Sank, M Weides, J Wenner, et al. Photon shell game in three-resonator circuit quantum electrodynamics. *Nature Physics*, 7(4):287–293, 2011.

[163] Richard Deblock, Eugen Onac, Leonid Gurevich, and Leo P. Kouwenhoven. Detection of Quantum Noise from an Electrically Driven Two-Level System. *Science*, 301(5630):203–206, 2003.

[164] O. Astafiev, Yu. A. Pashkin, Y. Nakamura, T. Yamamoto, and J. S. Tsai. Quantum Noise in the Josephson Charge Qubit. *Phys. Rev. Lett.*, 93:267007, Dec 2004.

[165] S. E. de Graaf, A. A. Adamyan, T. Lindström, D. Erts, S. E. Kubatkin, A. Ya. Tzalenchuk, and A. V. Danilov. Direct Identification of Dilute Surface Spins on Al_2O_3: Origin of Flux Noise in Quantum Circuits. *Phys. Rev. Lett.*, 118:057703, Jan 2017.

[166] T. Lanting, M. H. S. Amin, M. W. Johnson, F. Altomare, A. J. Berkley, S. Gildert, R. Harris, J. Johansson, P. Bunyk, E. Ladizinsky, E. Tolkacheva, and D. V. Averin. Probing high-frequency noise with macroscopic resonant tunneling. *Phys. Rev. B*, 83:180502, May 2011.

[167] Yoni Shalibo, Ya'ara Rofe, David Shwa, Felix Zeides, Matthew Neeley, John M. Martinis, and Nadav Katz. Lifetime and Coherence of Two-Level Defects in a Josephson Junction. *Phys. Rev. Lett.*, 105:177001, Oct 2010.

[168] Kenneth C. Hass, William F. Schneider, Alessandro Curioni, and Wanda Andreoni. The Chemistry of Water on Alumina Surfaces: Reaction Dynamics from First Principles. *Science*, 282(5387):265–268, 1998.

[169] Yu-Huan Lu and Hsin-Tsung Chen. Hydrogen generation by the reaction of H2O with Al2O3-based materials: a computational analysis. *Phys. Chem. Chem. Phys.*, 17:6834–6843, 2015.

[170] Theodore Van Duzer and Charles William Turner. *Principles of superconductive devices and circuits*. 1981.

[171] I Tupitsyn and S Kotochigova. Hyperfine structure constants for diatomic molecules. *J. Res. Natl. Inst. Stand. Technol.*, 103:205, 1998.

[172] Roger H. Koch, David P. DiVincenzo, and John Clarke. Model for $1/f$ Flux Noise in SQUIDs and Qubits. *Phys. Rev. Lett.*, 98:267003, Jun 2007.

[173] M. Büttiker, E. P. Harris, and R. Landauer. Thermal activation in extremely underdamped Josephson-junction circuits. *Phys. Rev. B*, 28:1268–1275, Aug 1983.

[174] M. Büttiker and R. Landauer. Traversal Time for Tunneling. *Phys. Rev. Lett.*, 49:1739–1742, Dec 1982.

[175] D Esteve, J M Martinis, C Urbina, E Turlot, M H Devoret, H Grabert, and S Linkwitz. Observation of the Temporal Decoupling Effect on the Macroscopic Quantum Tunneling of a Josephson Junction. *Physica Scripta*, 1989(T29):121, 1989.

[176] Lara Faoro and Lev B. Ioffe. Microscopic Origin of Low-Frequency Flux Noise in Josephson Circuits. *Phys. Rev. Lett.*, 100:227005, Jun 2008.

[177] Hui Wang, Chuntai Shi, Jun Hu, Sungho Han, Clare C. Yu, and R. Q. Wu. Candidate Source of Flux Noise in SQUIDs: Adsorbed Oxygen Molecules. *Phys. Rev. Lett.*, 115:077002, Aug 2015.

[178] S. Sendelbach, D. Hover, M. Mück, and R. McDermott. Complex Inductance, Excess Noise, and Surface Magnetism in dc SQUIDs. *Phys. Rev. Lett.*, 103:117001, Sep 2009.

[179] Simon Gustavsson, Fei Yan, Gianluigi Catelani, Jonas Bylander, Archana Kamal, Jeffrey Birenbaum, David Hover, Danna Rosenberg, Gabriel Samach, Adam P. Sears, Steven J. Weber, Jonilyn L. Yoder, John Clarke, Andrew J. Kerman, Fumiki Yoshihara, Yasunobu Nakamura, Terry P. Orlando, and William D. Oliver.

Suppressing relaxation in superconducting qubits by quasiparticle pumping. *Science*, 354(6319):1573–1577, 2016.

[180] G. Catelani, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman. Relaxation and frequency shifts induced by quasiparticles in superconducting qubits. *Phys. Rev. B*, 84:064517, Aug 2011.

[181] Ioan M Pop, Kurtis Geerlings, Gianluigi Catelani, Robert J Schoelkopf, Leonid I Glazman, and Michel H Devoret. Coherent suppression of electromagnetic dissipation due to superconducting quasiparticles. *Nature*, 508(7496):369–372, 2014.

[182] J. Wenner, Yi Yin, Erik Lucero, R. Barends, Yu Chen, B. Chiaro, J. Kelly, M. Lenander, Matteo Mariantoni, A. Megrant, C. Neill, P. J. J. O'Malley, D. Sank, A. Vainsencher, H. Wang, T. C. White, A. N. Cleland, and John M. Martinis. Excitation of Superconducting Qubits from Hot Nonequilibrium Quasiparticles. *Phys. Rev. Lett.*, 110:150502, Apr 2013.

[183] D Ristè, C. C. Bultink, M. J. Tiggelman, R. N. Schouten, K. W. Lehnert, and L. DiCarlo. Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit. *Nature communications*, 4:1913, 2013.

[184] U. Vool, I. M. Pop, K. Sliwa, B. Abdo, C. Wang, T. Brecht, Y. Y. Gao, S. Shankar, M. Hatridge, G. Catelani, M. Mirrahimi, L. Frunzio, R. J. Schoelkopf, L. I. Glazman, and M. H. Devoret. Non-Poissonian Quantum Jumps of a Fluxonium Qubit due to Quasiparticle Excitations. *Phys. Rev. Lett.*, 113:247001, Dec 2014.

[185] C. Wang, Y. Y. Gao, I. M. Pop, U. Vool, C. Axline, T. Brecht, R. W. Heeres, L. Frunzio, M. H. Devoret, G. Catelani, L. I. Glazman, and R. J. Schoelkopf. Measurement and control of quasiparticle dynamics in a superconducting qubit. *Nature Communications*, 5:5836, December 2014.

[186] X. Y. Jin, A. Kamal, A. P. Sears, T. Gudmundsen, D. Hover, J. Miloshi, R. Slattery, F. Yan, J. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver. Thermal and Residual Excited-State Population in a 3D Transmon Qubit. *Phys. Rev. Lett.*, 114:240501, Jun 2015.

[187] John M Martinis, Michel H Devoret, Daniel Esteve, and Cristian Urbina. Measuring the time spent traversing the barrier while tunneling. *Physica B: Condensed Matter*, 152(1):159–161, 1988.

[188] T. Lanting, M. H. Amin, A. J. Berkley, C. Rich, S.-F. Chen, S. LaForest, and Rogério de Sousa. Evidence for temperature-dependent spin diffusion as a mechanism of intrinsic flux noise in SQUIDs. *Phys. Rev. B*, 89:014503, Jan 2014.

[189] M. A. Ruderman and C. Kittel. Indirect Exchange Coupling of Nuclear Magnetic Moments by Conduction Electrons. *Phys. Rev.*, 96:99–102, Oct 1954.

[190] D. V. Averin and C. Bruder. Variable Electrostatic Transformer: Controllable Coupling of Two Charge Qubits. *Phys. Rev. Lett.*, 91:057003, Jul 2003.

[191] T. Hime, P. A. Reichardt, B. L. T. Plourde, T. L. Robertson, C.-E. Wu, A. V. Ustinov, and John Clarke. Solid-State Qubits with Current-Controlled Coupling. *Science*, 314(5804):1427–1429, 2006.

[192] S. H. W. van der Ploeg, A. Izmalkov, Alec Maassen van den Brink, U. Hübner, M. Grajcar, E. Il'ichev, H.-G. Meyer, and A. M. Zagoskin. Controllable Coupling of Superconducting Flux Qubits. *Phys. Rev. Lett.*, 98:057004, Feb 2007.

[193] R. Harris, A. J. Berkley, M. W. Johnson, P. Bunyk, S. Govorkov, M. C. Thom, S. Uchaikin, A. B. Wilson, J. Chung, E. Holtham, J. D. Biamonte, A. Yu. Smirnov, M. H. S. Amin, and Alec Maassen van den Brink. Sign- and Magnitude-Tunable Coupler for Superconducting Flux Qubits. *Phys. Rev. Lett.*, 98:177001, Apr 2007.

[194] R. C. Bialczak, M. Ansmann, M. Hofheinz, M. Lenander, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis. Fast Tunable Coupler for Superconducting Qubits. *Phys. Rev. Lett.*, 106:060501, Feb 2011.

[195] M. S. Allman, F. Altomare, J. D. Whittaker, K. Cicak, D. Li, A. Sirois, J. Strong, J. D. Teufel, and R. W. Simmonds. rf-SQUID-Mediated Coherent Tunable Coupling between a Superconducting Phase Qubit and a Lumped-Element Resonator. *Phys. Rev. Lett.*, 104:177004, Apr 2010.

[196] Yu Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, A. Megrant, J. Y. Mutus, P. J. J. O'Malley, C. M. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, Michael R. Geller, A. N. Cleland, and John M. Martinis. Qubit Architecture with High Coherence and Fast Tunable Coupling. *Phys. Rev. Lett.*, 113:220502, Nov 2014.

[197] R. Harris, T. Lanting, A. J. Berkley, J. Johansson, M. W. Johnson, P. Bunyk, E. Ladizinsky, N. Ladizinsky, T. Oh, and S. Han. Compound Josephson-junction coupler for flux qubits with minimal crosstalk. *Phys. Rev. B*, 80:052506, Aug 2009.

[198] Alec Maassen van den Brink, A J Berkley, and M Yalowsky. Mediated tunable coupling of flux qubits. *New Journal of Physics*, 7(1):230, 2005.

[199] Alec Maassen van den Brink, A J Berkley, and M Yalowsky. Mediated tunable coupling of flux qubits. *New Journal of Physics*, 7(1):230, 2005.

[200] Hidetoshi Nishimori and Kabuki Takada. Exponential Enhancement of the Efficiency of Quantum Annealing by Non-Stoquastic Hamiltonians. *Frontiers in ICT*, 4:2, 2017.

[201] Jacob D. Biamonte and Peter J. Love. Realizable Hamiltonians for universal adiabatic quantum computers. *Phys. Rev. A*, 78:012352, Jul 2008.

[202] Sergey Bravyi, David P. Divincenzo, Roberto Oliveira, and Barbara M. Terhal. The Complexity of Stoquastic Local Hamiltonian Problems. *Quantum Info. Comput.*, 8(5):361–385, May 2008.

[203] D. Kafri et al. (unpublished).

[204] S. Filipp, M. Göppl, J. M. Fink, M. Baur, R. Bianchetti, L. Steffen, and A. Wallraff. Multimode mediated qubit-qubit coupling and dark-state symmetries in circuit quantum electrodynamics. *Phys. Rev. A*, 83:063827, Jun 2011.

[205] T. Niemczyk, F. Deppe, H. Huebl, E. P. Menzel, F. Hocke, M. J. Schwarz, J. J. Garcia-Ripoll, D. Zueco, T. Hummer, E. Solano, A. Marx, and R. Gross. Circuit quantum electrodynamics in the ultrastrong-coupling regime. *Nat Phys*, 6(10):772–776, October 2010.

[206] P. Forn-Díaz, J. Lisenfeld, D. Marcos, J. J. García-Ripoll, E. Solano, C. J. P. M. Harmans, and J. E. Mooij. Observation of the Bloch-Siegert Shift in a Qubit-Oscillator System in the Ultrastrong Coupling Regime. *Phys. Rev. Lett.*, 105:237001, Nov 2010.

[207] Fumiki Yoshihara, Tomoko Fuse, Sahel Ashhab, Kosuke Kakuyanagi, Shiro Saito, and Kouichi Semba. Superconducting qubit-oscillator circuit beyond the ultrastrong-coupling regime. *Nat Phys*, 13(1):44–47, January 2017.

[208] J. Johansson M. W. Johnson T. Lanting P. Bunyk E. Tolkacheva E. Ladizinsky B. Bumble A. Fung A. Kaul A. Kleinsasser S. Han R. Harris, A. J. Berkley. Implementation of a Quantum Annealing Algorithm Using a Superconducting Circuit. *arXiv:0903.3906*, 2009.

[209] B. Foxen, J. Y. Mutus, E. Lucero, R. Graff, A. Megrant, Yu Chen, C. Quintana, B. Burkett, J. Kelly, E. Jeffrey, Yan Yang, Anthony Yu, K. Arya, R. Barends, Zijun Chen, B. Chiaro, A. Dunsworth, A. Fowler, C. Gidney, M. Giustina, T. Huang, P. Klimov, M. Neeley, C. Neill, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, and John M. Martinis. Qubit compatible superconducting interconnects. *arXiv:1708.04270*, 2017.