UNIVERSITY of CALIFORNIA
Santa Barbara

Operating Flux-Tunable Superconducting Qubits with High Fidelity

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

Doctor of Philosophy

in

Physics

by

Brooks Riley Foxen

Committee in charge:

Professor John Martinis, Chair

Professor Mark Sherwin

Professor Matthew Fisher

December 2019

The dissertation of Brooks Riley Foxen is approved:

Professor Mark Sherwin

Professor Matthew Fisher

Professor John Martinis, Chair

December 2019

To my parents, Michele and Aaron Campbell,
to my partner, Anya Foxen,
and to my pets, Abby and Baggins

Acknowledgements

This work would not have been possible without John Martinis—thank you for building such an incredible lab as well as the teaching and resources necessary for me to complete this work! Similarly, I am indebted to all the members of the UCSB Martinis group and many of the more recent Google additions. Thank you all for your kindness, helpfulness, and for building and sharing the infrastructure this work is build upon.

Pedram, your brought me into this lab on the majorana project and you saw me out with the fSim project. Your love of physics is truly inspiring. It is a joy to see you thrilled by data, day after day, and an even greater joy to (now) often see you promptly pack up at the end of the day to go for a jog, leaving the office at a reasonable time (even if you still return for the night shift). Your dedication and love of physics reminds me how lucky I am to be a part of this group.

Josh, it was a pleasure working with you on the bump bond project, and I would not have made it through the tunable resonator experiment without you. Your friendship and all the runs/bike rides helped keep me going in those middle years, and I am truly thankful for your constant kindness and prospective. Thank you for coming to visit me over in Broida!

Charles, thank you for taking the time to bring me up to speed and mentoring me through out the fSim gate project. It was great learning from you and made an awesome final project.

Jimmy and Andrew, class of 2012 forever! I have asked you both for help more times than I can count over the past seven years. Thank you both for always being generous with your time and for your willingness to teach!

Mom and dad, thank you for the many years of love, support, and encouragement it took to get me here, I could not have done it without you.

Anya, I am so thankful that we had each other through these past several years and I look forward to many more to come!

Abby, thank you for being a good dog and loving me even when graduate school yielded fewer walks than you expected.

Baggins than you for channeling my anxieties... maybe that helped.

Curriculum Vitæ

Brooks Riley Foxen

Education

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

2007 B.S., Physics, University of Texas, Austin

2004, Grapevine High School, Grapevine

Professional Experience

2008 Applications Engineer, National Instruments

2010 Product Support Engineer, National Instruments

First author publications

"High speed flux sampling for tunable superconducting qubits with an embedded cryogenic transducer", B. Foxen, J.Y. Mutus, E. Lucero, E. Jeffrey, D. Sank, R. Barends, K. Arya, B. Burkett, Yu Chen, Zijun Chen, B. Chiaro, A. Dunsworth, A. Fowler, C. Gidney, M. Giustina, R. Graff, T. Huang, J. Kelly, P. Klimov, A. Megrant, O. Naaman, M. Neeley, C. Neill, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, T.C. White, and John M. Martinis *submitted August 2018*

"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, 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, John M. Martinis *Quantum Science and Technology* vol 3, 1 (2017)

Abstract

Operating Flux-Tunable Superconducting Qubits with High Fidelity

by

Brooks Riley Foxen

The experimental challenge of today's quantum computing engineers is to choose a physical qubit system and whittle away at the multifaceted orders of magnitude improvement needed to build a practical quantum computer. In the first part of this thesis I will provide a general introduction to superconducting qubits and the isolated cryogenic environment in which they operate with an eye towards both the particular design requirements and the challenges we face in accommodating many more qubits in the future. Next, I will present a series of three experiments specifically oriented towards system-level improvements for flux-tunable superconducting qubits. In the first experiment, we develop a new metrology tool to characterize the on-chip settling of magnetic flux waveforms. We then used this technique to develop a new PCB-based packaging solution thereby increasing our package-to-chip wiring limit by at least a *factor of ten* enabling control of 10x more qubits on a single chip. In the second experiment, we provide a fabrication process for superconducting interconnects which allow for the three dimensional integration of our qubits and a direct *factor of 2* improvement in qubit connectivity moving from linear chains to two dimensional grids of qubits enabling n^2 more complex circuits. Finally we

implement a hardware-efficient 2-qubit fermionic simulation gateset, proposed to study quantum chemistry, using DC flux control on an adjustable coupling gmon transmon device. This first realization of the complete fSim gateset, of which CZ is a member, yielded nearly a *2x improvement* over the best reported Pauli error for a CZ for a solid state system to date, 0.41%, and implements an arbitrary photon conserving and low leakage two qubit unitary operation with *a factor of 4* times higher fidelity than a minimally universal gateset using single qubit rotations with only a CZ enabling 4-8x more coherent fSim operations. In total, these improvements have increased the computational complexity of our quantum processor by a *factor of 400-800* on the road towards building a practical quantum computer.

Contents

Chapter 1

Superconducting qubits

1.1 Introduction

Quantum computers present a potentially revolutionary computational technology. We expect that, once developed, quantum computers will be able to perform some types of computations that are intractable on both existing and even any future classical supercomputers. The precise nature of these applications is somewhat unknown due to the difficultly of proving that solutions are optimal for a given computational problem. Never the less, there are limited examples of a provable speedup for quantum algorithms [1, 2, 3, 4], and many more grey areas where promising quantum algorithms exceed the best known classical algorithm for suspected hard classical problems. Feynman is often credited with formalizing the idea of building a quantum computer in 1982 due to the difficulty of simulating quantum mechanics on classical computers [5]. And today, still, some of the most exciting potential near-term applications involve just that—simulating

quantum dynamics to advance research into materials, energy, and beyond.

Over the past five years, interest in quantum computing has surged. In 2014, superconducting qubits demonstrated performance near the threshold necessary for error correction which indicated that some level of brute-force scaling might be productive [6]. A November 2018 industry report from the Boston Consulting Group identified more than $700 million in private funding for quantum computing startups and nearly $6 billion in funding from government led initiatives across the globe since 2012 [7]. Expanding on this report in 2019, they explored the potential commercial value of quantum computing [8]. They broadly consider three types of quantum computers with increasing increasing complexity and computational ability that roughly align with the expected scientific development path. In the near term, they suppose that Noisy Intermediate Scale Quantum (NISQ, [9]) computers may begin to be able to solve materials simulations using error mitigation techniques—they estimate problems of this sort may be worth $2 to $5 Billion USD annually. With the addition of some level of error correction, when quantum computers have begun to realize some large scale quantum advantage, BCG estimates the addressable computational value at $25 to $50 Billion USD annually. Finally, they assess the value of full-scale fault tolerant quantum computation to be in the range of $450 to $850 Billion USD annually! In October 2019 quantum supremacy was achieve with a 53 qubit device performing a well defined computation in 200 seconds that would take 10,000 years on the largest supercomputer in the world, an achievement likely to spark further interest an investment into quantum computing [10].

However, many technical challenges remain before the commercial value of quantum computation may be realized. Current estimates indicate that 10^6 to 10^8 qubits will be necessary to build a moderately sized error corrected quantum computer. Even if we were to ignore the many unsolved challenges in scaling up, the cost of the existing qubit infrastructure, likely $10,000 to $25,000 per qubit, is prohibitive. For these reasons, quantum computing researchers must continue to pursue improvements in the quality of the physical qubits and gates to reduce the overall number necessary to achieve error correction and reductions in the cost of the corresponding infrastructure.

1.2 What is a quantum computer?

The goal of quantum computing is to harness quantum mechanics to perform computational tasks. The details of how and why that is desirable are subtle, but we can start by considering the instructive, self-referential use case. Quantum mechanics is widely accepted as the most successful quantitative theory ever produced. As such, being able to simulate quantum mechanics in a computer is quite desirable when it comes to understanding chemistry and materials problems. Arbitrarily sized simulations of quantum mechanics might allow for the design of materials like high temperature superconductors, chemical reaction catalysts, more efficient batteries and solar cells, or even help identify candidate drugs [11, 12, 13]. The trouble is that simulating quantum particles requires an exponential overhead of classical resources. That is, simulating n quantum particles requires, in general, storing and computing the evolution of 2^n amplitudes on a classical

computer (that is assuming no *a priori* knowledge of the system that allows some of these amplitudes to be neglected). While this is easy enough for small n, by the time you reach $n = 57$, it requires roughly an exabyte (10^{18} bytes) of storage—more storage than any single supercomputer today. Regardless of the available classical computational resources, with this exponential scaling, there surely exists some such n where one may start to wonder if it might be easier to just control a relatively small number of quantum particles directly. And in this way, one may stumble in to a desire for a quantum computer...

The most basic element of a quantum computer is a quantum bit, or qubit, for those in the know. The designation "qubit" refers to the information content of a quantum system with two states, and not any particular physical implementation of a qubit. While a classical system with two states exists in either one or the other (0 or 1), a qubit may exist in a superposition of both states simultaneously $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers. The vector $|\Psi\rangle$ is the quantum mechanical wavefunction, the square of which indicates the probability of finding the qubit in either the $|0\rangle$ state (probability $|\alpha|^2$) or the $|1\rangle$ state (probability $|\beta|^2$) if the state of the qubit is measured. Normalization, or the requirement that we find the qubit in either state during measurement, requires that $|\alpha|^2 + |\beta|^2 = 1$. After normalization, we may consolidate the remaining three degrees of freedom into one amplitude and two phases, $|\Psi\rangle = e^{i\phi_g}(\cos\theta/2|0\rangle + e^{i\dot{\phi}}\sin\theta/2|0\rangle)$. The global phase, ϕ_g, is inconsequential for a single qubit, further simplifying the state to,

$$|\Psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \tag{1.1}$$

This representation of an arbitrary qubit superposition may be connected to the Bloch Sphere, where the state $|\Psi\rangle$ is represented as a point on a sphere where θ and ϕ denoted by the azimuthal and equatorial angles [14].

To expand the complexity of this system, a second qubit may be added. An arbitrary wavefunction for 2 qubits will involve four basis states,

$$|\Psi\rangle = \alpha|00\rangle + \beta|01\rangle + \delta|10\rangle + \gamma|11\rangle \tag{1.2}$$

After constraints, it takes 2^n complex amplitudes to describe the wavefunction of an n-qubit system. This is still a bit abstract as most physical quantum systems, like atoms, have many more than just two possible states. Fortunately, with a general enough set of gate operations, a system of qubits may be assembled into a universal quantum computer [15, 16, 17].

So, the task of building a computer consists of choosing a physical system that may be restricted to, and controlled within, a finite number of (at least two) states. Most often this means just two states since controlling the evolution of a three or more level system is likely more difficult than just two levels. Some systems are inherently two-level, like the spin of an electron, or the polarization of a photon. Other systems, like the orbital structure of atoms or the resonant circuits used in superconducting qubits have more than two levels and care must be taken to evolve the state within only the desired subspace of states. After identifying a physical qubit implementation, the task

of building a quantum computer, as outlined by DiVincenzo, is to perform sufficient single and two qubit operations to implement a universal gate set, state preparation, and measurement all in a scalable architecture [18]. The devil, of course, lies in the details, a few of which are considered in this thesis; primarily, the quality and choice of two qubit gates (Chapter 6) and general system scalability concerns for flux tunable superconducting qubits (Chapters 3, 4, and 5).

1.3 Superconducting qubits

Superconducting qubits are non-linear electrical resonators—we fabricate ours using aluminum on sapphire, or more recently silicon. When a qubit is excited by placing a photon of energy into the resonator, the excitation produces currents which oscillate back and forth in the circuit. Superconductivity reduces the resistive losses of the currents flowing in the metal, but it is impossible to completely isolate the resonator circuit. The oscillating currents may have some weak interaction with lossy residues from fabrication, lattice imperfections in the substrate, or even a direct coupling to other qubits or resonators intentionally built into the system. Due to these interactions, eventually the energy will leak out of the resonator. This limits the qubit coherence, but also provides a straight forward method of initializing the state of the qubit—just wait long enough and photons will leak out of the qubit (assuming that the qubit's temperature is cold enough that it is not excited thermally). We excite the qubits with an electronic microwave frequency pulse, and, for frequency tunable qubits, we can control the frequency by applying a

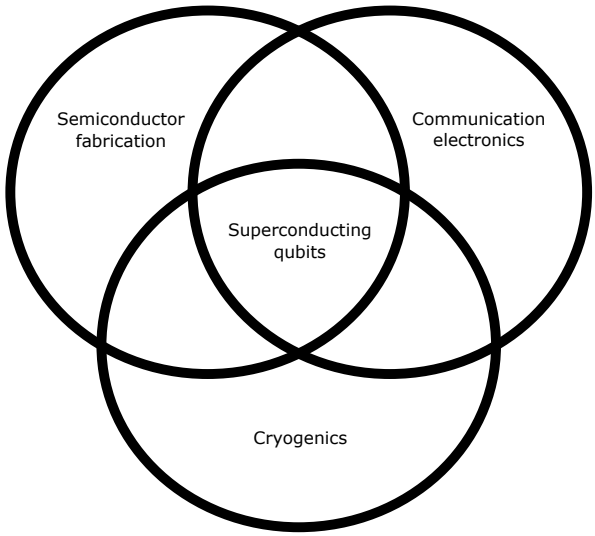


Figure 1.1: Superconducting qubits are an attractive candidate for building a quantum computer because the core technologies they rely on have already been scaled enormously on an individual basis.

magnetic field to the qubit, which we modulate with a control current. To measure the state of the qubit, we couple a linear resonator to each qubit such that the frequency of the linear resonator is dependent on the presence or absence of a photon in the qubit. We then perform a reflective microwave measurement, driving the linear resonator at an appropriate frequency, to distinguish the 0 and 1 photon states.

Superconducting qubits are an attractive candidate system for building a quantum computer because they have been developed on the periphery of advanced technologies that have been scaled enormously on an individual basis (Figure 1.1). Their aluminum-on-silicon fabrication is largely based on semiconductor fabrication techniques developed for CMOS aluminum on silicon fabrication that has received trillions of dollars in funding

over many decades and now routinely produces processors with several billion transistors and memory storage devices with tens of trillions of transistors each [19, 20]. The control and readout of superconducting qubits require high frequency electronics, but fortunately, qubits may be designed to operate in standard telecommunication bandwidths thus benefiting from the corresponding development of high quality and low(er) cost microwave components and design practices. Millikelvin temperature are perhaps the most exotic requirement for operating superconducting qubits, but these low temperatures are readily attainable. We are able to purchase off-the-shelf systems capable of handling the cooling needs of a couple hundred cables and a dozen qubit readout amplifiers thus far supporting as many as 72 qubits or 54 qubits and 88 couplers [10, 21]. While many design challenges remain, it is worth noting that particle physicists have already commissioned and built very large cryogenic cooling systems. The largest such cooling system, at the Large Hadron Collider, has a cooling capacity of 144 kW at 4.5 K and 20 kW of cooling power at 1.8 K, nearly 300,000x more cooling power than our existing cryostat[22, 23, 24].

1.4 Thesis overview

As I mentioned earlier, most of this thesis focuses on concerns related to the scalability of frequency tunable superconducting qubits. In Chapter 2 we will first review the superconducting circuits used in this thesis. In Chapter 3, we will explore some of the general experimental infrastructure requirements for superconducting qubit processors ranging from cryogenics, to the control electronics, and packaging. While fixed

frequency qubits require only low-current microwave control signals on-chip, frequency tunable qubits which offer more control, require several hundred MHz base band control waveforms with 1-2 mA of current. In Chapter 4 we develop a metrology tool for sampling on-chip flux waveforms using a frequency tunable resonator that allowed us to develop a near-term-scalable PCB based chip mount and directly model the complex DC to MHz flux bias transfer function on a superconducting chip. In Chapter 5 we develop and characterize superconducting interconnects used to assemble flip chip hybrid qubit devices, compatible with the mA flux bias currents. Finally, in Chapter 6 we use flux bias control of both gmon qubits and their couplers to implement a continuous set of two qubit gates with an average error for an arbitrary excitation-preserving 2-qubit unitary less than the previously reported best CZ_ϕ.

Chapter 2

Superconducting Circuits

In this chapter we will briefly review the theory of some relevant superconducting circuits. Typical quantum processor consists of a relatively small number of components since we deal with only reactive circuit elements due to the extremely small cooling power available at mK temperatures. This means circuit design is limited to distributed inductors or capacitors formed by various transmission lines of varied geometries, lumped element parallel plate capacitors, the lumped non-linear inductance offered by Josephson junctions, or the flux-controllable non-linear inductance of a SQUID. Without Josephson junctions to provide a non-linear circuit element, superconducting quantum computing would not be possible as all circuits would produce linearly spaced energy levels that cannot be addressed independently.

With any introduction of superconducting qubits, it is difficult to avoid describing the quantum harmonic oscillator both because superconducting qubits are electrical an-

harmonic oscillators and because the harmonic oscillator is one of the few analytically solvable problems in quantum mechanics. The values of the inductance and capacitance of the oscillator circuit may be adjusted to trade one noise sensitivity for another. Transmon qubits have gained favor because they are insensitive to charge noise that seems to be unavoidable with existing fabrication techniques. Here we will introduce frequency tunable transmon qubits in four steps by first describing a superconducting harmonic oscillator circuit, then a Josephson junction followed by a SQUID which allows for a flux tunable circuit element, and finally frequency tunable transmons. Each of these circuits plays an important role in the design of superconducting quantum processors, and with the exception of a single Josephson junction, are later used in the experiments described in this thesis.

2.1 LC oscillator

Here we follow the Lagrangian formalism to derive the energy level structure of the quantum harmonic oscillator. In Figure 2.1 we have a parallel LC circuit, and we begin by writing down the kinetic and potential energies of the inductor and capacitor using the charge, Q, as our position coordinate,

$$U = \frac{Q^2}{2C} \tag{2.1}$$

$$T = \frac{L\dot{Q}^2}{2}. \tag{2.2}$$

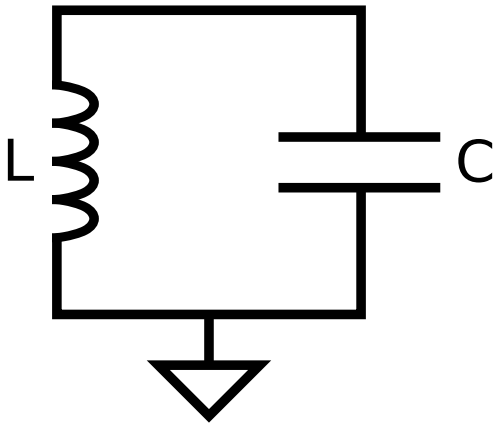


Figure 2.1: Simple LC oscillator composed of an inductor in parallel with a capacitor.

Combining these, we write down the Lagrangian of the circuit,

$$\mathcal{L} = T - U = \frac{L\dot{Q}^2}{2} - \frac{Q^2}{2C}. \tag{2.3}$$

We then apply the Euler-Lagrange formula,

$$\frac{\partial \mathcal{L}}{\partial q_j} - \frac{d}{dt}\Big(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\Big) = 0. \tag{2.4}$$

Using our charge coordinate $Q = q_j$, we obtain,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial Q} - \frac{d}{dt}\Big(\frac{\partial \mathcal{L}}{\partial \dot{Q}}\Big) &= 0 \\
\frac{Q}{C} - \frac{d}{dt}(L\dot{Q}) &= 0 \\
\frac{Q}{C} - L\ddot{Q} &= 0 \\
\frac{Q}{C} - L\ddot{Q} &= 0 \\
Q\omega^2 - \ddot{Q} &= 0
\end{aligned} \tag{2.5}$$

where ω is the resonance frequency, $1/\sqrt{\mathrm{LC}}$. We can then produce the conjugate coordinate to Q by computing $\mathrm{p_x} = \partial\mathcal{L}/\partial\dot{\mathrm{q}}_\mathrm{j}$, or, in our case:

$$p_x = \frac{\partial \mathcal{L}}{\partial \dot{Q}} = L\dot{Q} = \Phi. \tag{2.6}$$

where Φ is the flux through the inductor. Since Q and Φ are canonically conjugate coordinates, they obey the canonical commutation relation $[\Phi, \mathrm{Q}] = i\hbar$. We then perform a Legendre transformation to convert the Lagrangian into a Hamiltonian:

$$\begin{aligned}
H = &\quad p_j \cdot \dot{q} - \mathcal{L} \\
H = &\quad \dot{Q}\Phi - \tfrac{L\dot{Q}^2}{2} + \tfrac{Q^2}{2C} \\
H = &\quad \dot{Q}\Phi - \tfrac{\dot{Q}\Phi}{2} + \tfrac{Q^2}{2C} \\
H = &\quad \tfrac{\Phi^2}{2L} + \tfrac{Q^2}{2C}
\end{aligned} \tag{2.7}$$

Now we have the familiar harmonic oscillator Hamiltonian which can be rewritten in terms of raising and lowering operators, a and $a^\dagger$, as

$$H = \hbar\omega\Big(a^\dagger a + \frac{1}{2}\Big) \tag{2.8}$$
$$\tag{2.9}$$

where,

$$a^\dagger = \sqrt{\frac{1}{2\hbar Z}}(\Phi - iZQ) \tag{2.10}$$
$$a = \sqrt{\frac{1}{2\hbar Z}}(\Phi + iZQ) \tag{2.11}$$

and $Z = \sqrt{L/C}$ is the circuit impedance. Similarly, Q and Φ may be rewritten in terms of a and $a^\dagger$,

$$Q = i\sqrt{\frac{\hbar}{2Z}}(a^\dagger - a) = iQ_{zpf}(a - a^\dagger) \tag{2.12}$$
$$\Phi = \sqrt{\frac{\hbar Z}{2}}(a^\dagger + a) = \Phi_{zpf}(a + a^\dagger) \tag{2.13}$$

Where $Q_{zpf} = \sqrt{\langle 0|Q^2|0\rangle} = \sqrt{\hbar/2Z}$ is the zero point fluctuation of the charge operator, and $\Phi_{zpf} = \sqrt{\hbar Z/2}$ is the zero point fluctuation of the flux operator.

The harmonic oscillator energy levels are all evenly spaced by an energy, $\hbar\omega$. As such, the harmonic oscillator is not a suitable qubit circuit—it is impossible to confine the system to any two levels. Any attempt to excite a single energy level will immediately populate a coherent state involving many energy levels [25]. Alas, this effort is not in vain. Linear harmonic oscillators may be fabricated with discrete inductors and capacitors or as

distributed transmission line stubs using either shorted or open termination to form $\lambda/4$ or $\lambda/2$ resonators. Such transmission line resonators form useful ancillary qubit readout circuitry [26, 27, 28] for dispersive measurement. These resonators also require only a single layer of lithography to fabricate and they may be characterized with a standard Vector Network Analyzer. Distributed transmission line resonators are sensitive to many of the same loss mechanisms as transmon qubits, so they may often be used as proxies for qubit performance subject to new fabrication techniques, packaging, etc, as we do in Chapter 3. Additionally, since actual superconducting qubit circuits are just non-linear resonators, we will pick this derivation back up in Section 2.4.

2.2 Josephson junctions

While the LC oscillator discussed above is useful for building ancillary qubit circuitry and for general materials characterization, the equal spacing of energy make it impossible to operate a linear LC oscillator as a qubit. In order to make a quantum circuit with addressable energy levels tat can be used as a qubit, we need a non-linear superconducting circuit element. Fortunately, such an element exists: the Josephson junction (JJ). Figure 2.2 depicts a JJ constructed by placing a weak link between two superconducting islands. In our case, this weak link is a thin, insulating, oxide barrier, but JJs may be formed by sandwiching a normal metal, insulator, or even just a superconductor with a lower critical current between two robust superconductors.

In this section, we follow the the classic Feynman lectures [29] to derive the current-

phase and voltage-phase relations for a JJ [30]. We can begin by defining Ψ_L and Ψ_R as the wave functions of the superconducting state in the left and right superconductors. The dynamics of these wavefunctions are determined by coupled Schrödinger equations

$$\begin{aligned} i\hbar\frac{d\Psi_L}{dt} &= U_L\Psi_L + K\Psi_R \\ i\hbar\frac{d\Psi_R}{dt} &= U_R\Psi_R + K\Psi_L \end{aligned} \tag{2.14}$$

where K is a constant representing the coupling across the JJ barrier, and U_L and U_R are the lowest energy states on either side. In the absence of an applied voltage, U_L and U_R would be equal, but here we consider the case of an applied voltage $V = U_L - U_R$, defined such that the voltage on either size is $\pm V/2$.

$$\begin{aligned} i\hbar\frac{d\Psi_L}{dt} &= \frac{qV}{2}\Psi_L + K\Psi_R \\ i\hbar\frac{d\Psi_R}{dt} &= -\frac{qV}{2}\Psi_R + K\Psi_L \end{aligned} \tag{2.15}$$

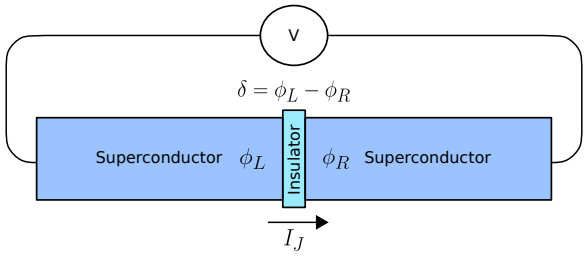


Figure 2.2: A Josephson junction consisting of two superconducting islands separated by a thin insulating barrier.

To solve these equations we make the ansatz,

$$\begin{aligned} \Psi_L &= \sqrt{\rho_L} e^{\imath\phi_L} \\ \Psi_R &= \sqrt{\rho_R} e^{\imath\phi_R} \end{aligned} \tag{2.16}$$

where Ψ_L and Ψ_R are the wavefunctions of the electrons on the left and right superconducting islands, ρ_L and ρ_R are the electron densities, and ϕ_L and ϕ_R are the superconducting phases. We then substitute equations 2.16 into equations 2.14 and equate the real and imaginary parts of each obtaining the system of four equations:

$$\begin{aligned} \dot{\rho_L} &= +\frac{2}{\hbar} K \sqrt{\rho_L \rho_R} \sin(\phi_R - \phi_L) \\ \dot{\rho_R} &= -\frac{2}{\hbar} K \sqrt{\rho_L \rho_R} \sin(\phi_R - \phi_L) \end{aligned} \tag{2.17}$$

$$\begin{aligned} \dot{\phi_L} &= +\frac{K}{\hbar} \sqrt{\frac{\rho_R}{\rho_L}} \cos(\phi_R - \phi_L) - \frac{qV}{2\hbar} \\ \dot{\phi_R} &= +\frac{K}{\hbar} \sqrt{\frac{\rho_L}{\rho_R}} \cos(\phi_R - \phi_L) + \frac{qV}{2\hbar} \end{aligned} \tag{2.18}$$

The first two equations (2.17) are just a statement of charge conservation indicating that the change in electron density on the left and right islands are equal and opposite ($\dot{\rho}_L = -\dot{\rho}_R$). This makes sense for a tunneling current, I_J, moving electrons from one island to the other. The magnitude of I_J, $|\dot{\rho_L}|$, may be simplified to the standard current-phase relation,

$$\begin{aligned} I_J &= \frac{2K}{\hbar}\sqrt{\rho_L \rho_R}\sin\delta \\ I_J &= I_0 \sin\delta \end{aligned} \tag{2.19}$$

where $\delta = \phi_R - \phi_L$ is the difference in the superconducting phase on either side of the barrier and I_0 is the maximum tunneling current, characteristic of the junction. Electrostatic forces keep the electron density fairly uniform, so ρ_L and ρ_R are both approximately equal to the normal electron density, ρ_0, which means $I_0 \approx 2K\rho_0/\hbar$. We can then take the difference of the second two equations (2.18) to obtain the voltage-phase relationship,

$$\begin{aligned} \dot{\phi_R} - \dot{\phi_L} &= \frac{qV}{\hbar} \\ \dot{\delta} &= \frac{qV}{\hbar} \\ \implies V &= \frac{\hbar}{2e}\frac{\partial\delta}{\partial t} \\ V &= \frac{\Phi_0}{2\pi}\frac{\partial\delta}{\partial t} \end{aligned} \tag{2.20}$$

where $q = 2e$ is the total charge of a cooper pair, and $\Phi_0 = h/2e$ is the magnetic flux quantum. Equations 2.19 and 2.20 are the standard Josephson Junction current-phase and voltage-phase relations. If we take the time derivative of equation 2.19 and solve for $\partial\delta/\partial t$, we obtain:

$$\begin{aligned} \frac{\partial I_J}{\partial t} &= I_0\cos(\delta)\frac{\partial\delta}{\partial t} \\ \frac{\partial\delta}{\partial t} &= \frac{\partial I_J}{\partial t}\frac{1}{I_0\cos(\delta)} \end{aligned} \tag{2.21}$$

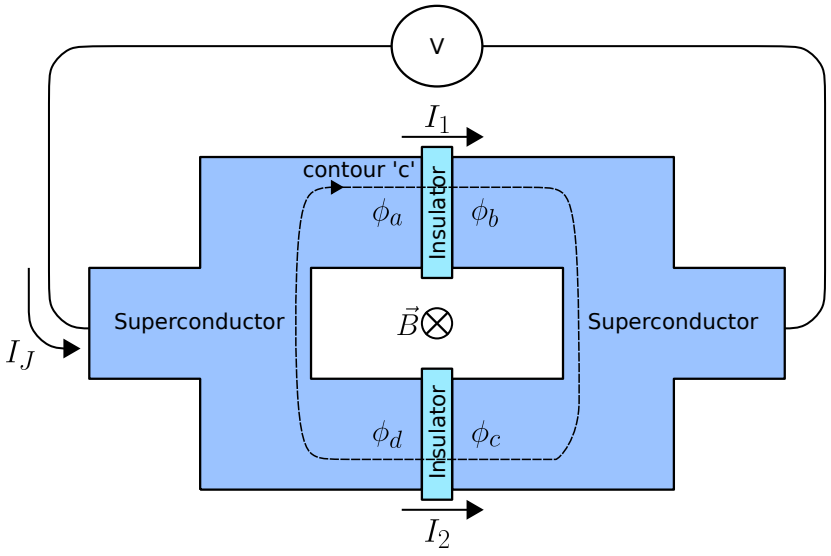


Figure 2.3: A superconducting quantum interference device (SQUID) composed of two Josephson junctions.

which we can then substitute into 2.20 to obtain the expression,

$$V = \underbrace{\frac{\Phi_0}{2\pi}\frac{1}{I_0\cos(\delta)}}_{\equiv L_J}\frac{\partial I_J}{\partial t} \tag{2.22}$$

The voltage in equation 2.22 takes the form of an inductance where the Voltage is proportional to the change in current giving us an expression for the so-called Josephson inductance $L_j = \Phi_0/2\pi I_0\cos(\delta)$.

2.3 SQUID

Before we explore how a single Josephson junctions may be used to create a qubit, we will first consider a circuit with two JJs which is the core of the magnetic-flux-tunable

circuits used in this thesis. Figure 2.3 provides a schematic drawing of a Superconducting QUantum Interference Device, or SQUID. A SQUID is composed of two Josephson junctions joined into a loop by two superconducting electrodes. If we assume that both junctions have the same critical current[1], I_c, then,

$$\begin{aligned} I_1 &= I_c \sin(\delta_1) \\ I_2 &= I_c \sin(\delta_2) \end{aligned} \tag{2.23}$$

where δ_1 and δ_2 are the phase differences across the first and second junctions. Summing these currents, we obtain an expression for the total current through the SQUID,

$$\begin{aligned} I_J &= I_1 + I_2 \\ I_J &= I_c(\sin(\delta_1) + \sin(\delta_2)) \\ I_J &= 2I_c \cos\left(\frac{\delta_1 - \delta_2}{2}\right) \sin\left(\frac{\delta_1 + \delta_2}{2}\right) \end{aligned} \tag{2.24}$$

We can then consider the difference in the superconducting phase along the closed contour c going around the loop clockwise starting at the 11 o'clock position.

$$\begin{aligned} \oint_c \bigtriangledown\phi \cdot d\vec{\ell} &= 2\pi n \\ &= (\phi_b - \phi_a) + (\phi_c - \phi_b) + (\phi_d - \phi_c) + (\phi_a - \phi_d) \end{aligned} \tag{2.25}$$

Where we know the integral of the phase for a closed loop must be 2π-periodic (e.g. $= 2\pi n$ for integer values of n, since the wavefuction $\Phi = |\rho|e^{i\phi}$ must be single valued).

[1]While all the qubits used in this thesis are nominally symmetric, it is worth mentioning that asymmetric SQUIDs have been suggested as a way to trade some frequency tunability for a reduced flux sensitivity and consequently improved dephasing [31].

We can then write the phase differences across the junction as gauge invariant phases,

$$\begin{aligned}(\phi_b - \phi_a) &= -\delta_1 - \frac{2\pi}{\Phi_0}\int_a^b \vec{A}\cdot d\vec{\ell} \\ (\phi_d - \phi_c) &= \delta_2 - \frac{2\pi}{\Phi_0}\int_c^d \vec{A}\cdot d\vec{\ell}\end{aligned} \tag{2.26}$$

and for the phase differences within the superconductor we can use the supercurrent equations,

$$\begin{aligned}(\phi_c - \phi_b) &= \int_b^c \bigtriangledown\phi\cdot d\vec{\ell} = -\Lambda\int_b^c \vec{J}\cdot d\vec{\ell} - \frac{2\pi}{\Phi_0}\int_b^c \vec{A}\cdot d\vec{\ell} \\ (\phi_a - \phi_d) &= \int_d^a \bigtriangledown\phi\cdot d\vec{\ell} = -\Lambda\int_d^a \vec{J}\cdot d\vec{\ell} - \frac{2\pi}{\Phi_0}\int_d^a \vec{A}\cdot d\vec{\ell}\end{aligned} \tag{2.27}$$

where $\Lambda = \frac{m}{nq^2}$. Then, combining equations 2.26 and 2.27 into equation 2.25, we have

$$\begin{aligned}2\pi n = -\delta_1 - \frac{2\pi}{\Phi_0}\int_a^b \vec{A}\cdot d\vec{\ell} + \delta_2 - \frac{2\pi}{\Phi_0}\int_c^d \vec{A}\cdot d\vec{\ell} \\ - \Lambda\int_b^c \vec{J}\cdot d\vec{\ell} - \frac{2\pi}{\Phi_0}\int_b^c \vec{A}\cdot d\vec{\ell} - \Lambda\int_d^a \vec{J}\cdot d\vec{\ell} - \frac{2\pi}{\Phi_0}\int_d^a \vec{A}\cdot d\vec{\ell},\end{aligned} \tag{2.28}$$

and collecting the $\int \vec{A}\cdot d\vec{\ell}$ terms, this simplifies to

$$\delta_2 - \delta_1 = 2\pi n + \underbrace{\frac{2\pi}{\Phi_0}\oint_c \vec{A}\cdot d\vec{\ell}}_{\equiv \Phi\, through\, SQUID} + \underbrace{\Lambda\int_b^c \vec{J}\cdot d\vec{\ell} + \Lambda\int_d^a \vec{J}\cdot d\vec{\ell}}_{\vec{J}=0\, inside\, S.C.}, \tag{2.29}$$

by assuming we can take the contour, c, to be deep inside the superconductor where $\vec{J} = 0$ this reduces to,

$$(\delta_2 - \delta_1) = 2\pi n + \frac{2\pi}{\Phi_0}\Phi \tag{2.30}$$

where Φ is the magnetic flux through the SQUID loop. This flux quantization means we can control the effective critical current of the SQUID with an applied flux. Substituting equation 2.30 into equation 2.24,

$$\begin{aligned} I_J &= 2I_c \cos\left(\frac{2\pi n}{2} + \frac{\pi\Phi}{\Phi_0}\right) \sin\left(\frac{2\pi n}{2} + \frac{2\pi n}{2\Phi_0}\Phi + \frac{2\delta_1}{2}\right) \\ I_J &= \underbrace{2I_c \cos\left(\frac{\pi\Phi}{\Phi_0}\right)}_{effective\, I_c} \sin\left(\delta_1 + \frac{\pi\Phi}{\Phi_0}\right) \end{aligned} \tag{2.31}$$

Comparing the SQUID current phase relation (2.31) with that of the single junction (2.19) we see that the behavior of a SQUID is just like that of a single junction with an effective critical current that may be adjusted from some maximum value for integer values of trapped flux down to zero at half-integer flux quanta with an externally applied magnetic flux[2]. In the next section we will see how to use such an element to form a qubit.

[2]Note, if the two junctions have different critical currents, the tunable range is only over the range $|I_{j1} - I_{j2}|$ to $|I_{j1} + I_{j2}|$

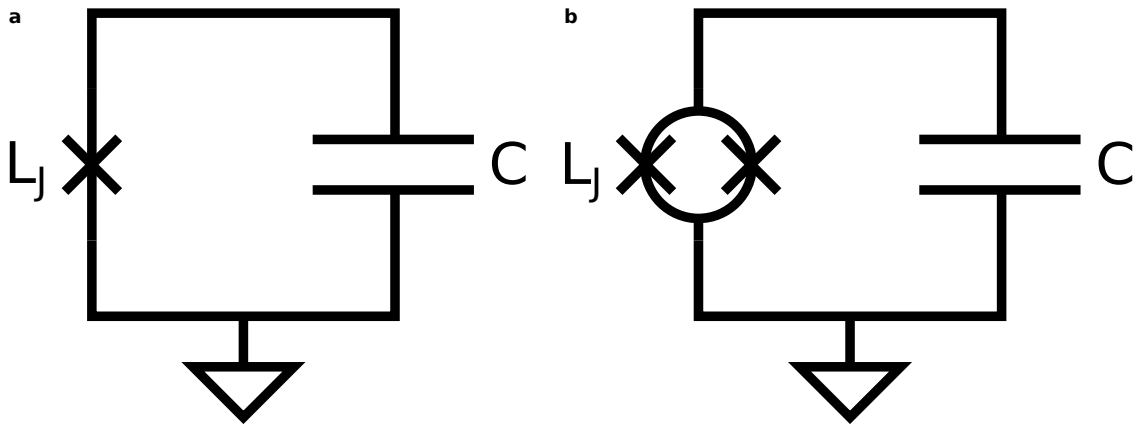


Figure 2.4: Transmon circuits. **a**, Fixed frequency transmon qubit and **b**, flux tunable transmon qubit.

2.4 Transmon qubits

In order to produce a quantum circuit with addressable energy levels, we can replace the inductance in the oscillator discussed in Section 2.1, with one Josephson junction (Section 2.2), or pair of Josephson junctions forming a SQUID (Section 2.3), as shown in Figure 2.4. As we saw in Section 2.3, a SQUID behaves like a single junction, so the following derivation holds for either circuit if you take I_c to be the single junction critical current, or the effective, flux-tuned, critical current of the SQUID (equation 2.31). Here, we begin by restating the JJ voltage and current relations.

$$\begin{aligned} V(t) &= \frac{\hbar}{2e}\frac{d\delta}{dt} \\ I(t) &= I_c \sin(\delta) \end{aligned} \tag{2.32}$$

where δ is the difference in the superconducting phase across the junction and I_c is the

junction critical current, above which the junction exhibits an Ohmic resistance (e.g. a linear resistance where $V = IR$). We can then integrate the IV product to compute the potential energy of the JJ circuit element

$$\begin{aligned} U =& \int_0^t I(t)V(t)dt \\ =& \frac{\hbar I_c}{2e} \int \sin(\delta) \frac{\partial \delta}{\partial t} dt \\ =& \frac{\hbar I_c}{2e} \int \sin(\delta)\, d\delta \\ =& -\frac{\hbar I_c}{2e} \cos(\delta) \\ =& -\frac{\Phi_0 I_c}{2\pi} \cos(\delta) \\ =& -E_j \cos(\delta) \end{aligned} \tag{2.33}$$

where $\Phi_0 = h/2e$ is the magnetic flux quantum, and $E_j = \Phi_0 I_c / 2\pi$ is the Josephson energy. Combining this with the capacitive energy we have the circuit Hamiltonian

$$H = \frac{Q^2}{2C} - E_J \cos(\delta) \tag{2.34}$$

We can rewrite the first term using the number operator, $n = Q/2e$, which counts the number of Cooper pairs on the capacitor, and the charging energy, $E_c = e^2/2C$.

$$H = 4E_c n^2 - E_J \cos(\delta) \tag{2.35}$$

where we may expand $\cos(\delta) = 1 - \delta^2 + \delta^4/24 + O(\delta^6)$, yielding an approximate harmonic oscillator Hamiltonian (ignoring the offset energy from the constant term),

$$H = 4E_c n^2 + E_J \delta^2 + O(\delta^4) \tag{2.36}$$

with energy levels given by $E_n = \sqrt{8E_cE_J}(n + 1/2)$, for integer values of n. In order to determine the qubit non-linearity, we can consider the next term in the expansion, $E_J\delta^4/24$ as a perturbation. We begin by rewriting δ in terms of the ladder operators a and $a^\dagger$[32],

$$\delta = \Big(\frac{2E_C}{E_J}\Big)^{\frac{1}{4}}(a^\dagger + a) \tag{2.37}$$

Then we can compute the effect of the δ^4 term using perturbation theory,

$$\begin{aligned}
\Delta E_j &= \langle j|\frac{-E_J\delta^4}{24}|j\rangle \\
&= -\frac{E_J}{24}\langle j|\frac{2E_C}{E_J}(a^\dagger + a)^4|j\rangle \\
&= -\frac{E_C}{12}\langle j|(a^\dagger + a)^4|j\rangle \\
&= -\frac{E_C}{12}\langle j|(a^\dagger a^\dagger + a^\dagger a + aa^\dagger + aa)^2|j\rangle \\
&= -\frac{E_C}{12}\langle j|(a^\dagger a^\dagger a^\dagger a^\dagger + a^\dagger a^\dagger a^\dagger a + a^\dagger a^\dagger aa^\dagger + a^\dagger a^\dagger aa \\
&\qquad + a^\dagger aa^\dagger a^\dagger + a^\dagger aa^\dagger a + a^\dagger aaa^\dagger + a^\dagger aaa \\
&\qquad + aa^\dagger a^\dagger a^\dagger + aa^\dagger a^\dagger a + aa^\dagger aa^\dagger + aa^\dagger aa \\
&\qquad + aaa^\dagger a^\dagger + aaa^\dagger a + aaaa^\dagger + aaaa|j\rangle \\
&= -\frac{E_C}{12}\langle j|(a^\dagger a^\dagger aa + a^\dagger aa^\dagger a + a^\dagger aaa^\dagger + aa^\dagger a^\dagger a + aa^\dagger aa^\dagger + aaa^\dagger a^\dagger|j\rangle
\end{aligned} \tag{2.38}$$

between the last two lines, we have ignored all terms with an unequal number of raising and lowering operators, as they would not map $\langle j|$ to $|j\rangle$. We then use the fact that $a^\dagger|j\rangle = \sqrt{j+1}|j+1\rangle$ and $a|j\rangle = \sqrt{j}|j-1\rangle$ to simplify further,

$$
\begin{aligned}
\Delta E_j = -\frac{E_C}{12}(&\langle j|a^\dagger a^\dagger aa|j\rangle \\
&+\langle j|a^\dagger aa^\dagger a|j\rangle \\
&+\langle j|a^\dagger aaa^\dagger|j\rangle \\
&+\langle j|aa^\dagger a^\dagger a|j\rangle \\
&+\langle j|aa^\dagger aa^\dagger|j\rangle \\
&+\langle j|aaa^\dagger a^\dagger|j\rangle) \\
\Delta E_j = -\frac{E_C}{12}(&\sqrt{j}\langle j|a^\dagger a^\dagger a|j-1\rangle \\
&+\sqrt{j}\langle j|a^\dagger aa^\dagger|j-1\rangle \\
&+\sqrt{j+1}\langle j|a^\dagger aa|j+1\rangle \\
&+\sqrt{j}\langle j|aa^\dagger a^\dagger|j-1\rangle \\
&+\sqrt{j+1}\langle j|aa^\dagger a|j+1\rangle \\
&+\sqrt{j+1}\langle j|aaa^\dagger|j+1\rangle) \\
\Delta E_j = -\frac{E_C}{12}(&\sqrt{j(j-1)}\langle j|a^\dagger a^\dagger|j-2\rangle \\
&+\sqrt{j^2)}\langle j|a^\dagger a|j\rangle \\
&+\sqrt{(j+1)^2}\langle j|a^\dagger a|j\rangle \\
&+\sqrt{j^2}\langle j|aa^\dagger|j\rangle \\
&+\sqrt{(j+1)^2}\langle j|aa^\dagger|j\rangle \\
&+\sqrt{(j+1)(j+2)}\langle j|aa|j+2\rangle) \\
\Delta E_j = -\frac{E_C}{12}(&\sqrt{j(j-1)^2)}\langle j|a^\dagger|j-1\rangle \\
&+\sqrt{j^3)}\langle j|a^\dagger|j-1\rangle \\
&+\sqrt{(j+1)^2 j}\langle j|a^\dagger|j-1\rangle \\
&+\sqrt{j^2(j+1)}\langle j|a|j+1\rangle \\
&+\sqrt{(j+1)^3}\langle j|a|j+1\rangle \\
&+\sqrt{(j+1)(j+2)^2}\langle j|a|j+1\rangle) \\
\Delta E_j = -\frac{E_C}{12}(&\sqrt{j^2(j-1)^2)}\langle j||j\rangle \\
&+\sqrt{j^4)}\langle j||j\rangle \\
&+\sqrt{(j+1)^2 j^2}\langle j||j\rangle \\
&+\sqrt{j^2(j+1)^2}\langle j||j\rangle \\
&+\sqrt{(j+1)^4}\langle j||j\rangle \\
&+\sqrt{(j+1)^2(j+2)^2}\langle j||j\rangle)
\end{aligned}
\tag{2.39}
$$

and finally we simplify all the eigenvalues,

$$\begin{aligned}\Delta E_j &= -\frac{E_C}{12}(j(j-1)+j^2+(j+1)j+j(j+1)+(j+1)^2+(j+1)(j+2)) \\ &= -\frac{E_C}{12}(j^2-j+j^2+j^2+j+j^2+j+j^2+2j+1+j^2+3j+2) \\ &= -\frac{E_C}{12}(6j^2+6j+3) \\ &= -\frac{E_C}{4}(2j^2+2j+1)\end{aligned} \tag{2.40}$$

So, the j^{th} energy level is given by,

$$E_J = j\sqrt{8E_JE_C} - \frac{E_C}{4}(2j^2+2j+1) \tag{2.41}$$

We can take the difference of the lowest two energy levels,

$$\begin{aligned}E_1 - E_0 &= \left(\sqrt{8E_JE_C} - \frac{E_C}{4}(2+2+1)\right) - \frac{E_C}{4} \\ &= \sqrt{8E_JE_C} - E_C\end{aligned} \tag{2.42}$$

and similarly for the difference between the j^{th} and $(j-1)^{th}$ states,

$$\begin{aligned}E_j - E_{j-1} =& \left(j\sqrt{8E_JE_C} - \frac{E_C}{4}(2j^2+2j+1)\right) \\ &- \left((j-1)\sqrt{8E_JE_C} - \frac{E_C}{4}(2(j-1)^2+2(j-1)+1)\right) \\ =& \sqrt{8E_JE_C} - \frac{E_C}{4}((2j^2+2j+1)-(2j^2-4j+2+2j-2+1)) \\ =& \sqrt{8E_JE_C} - \frac{E_C}{4}((2j^2+2j+1)-(2j^2-2j+1)) \\ =& \sqrt{8E_JE_C} - \frac{E_C}{4}(4j) \\ =& \sqrt{8E_JE_C} - jE_C\end{aligned} \tag{2.43}$$

Thus the energy levels of the transmon are addressable with an anharmonicity equal to E_C. Practically speaking, the anharmonicity is an extremely important parameter in our circuit design as sets a "speed limit" on how fast we can perform single qubit microwave gates equal to $\tau = h/E_c$. Beyond this limit, faster gates would require spectral content overlapping with the next energy level resulting in leakage from the qubit subspace to a higher energy level or levels. For this reason, a large E_C desirable. However, when E_C is large, the shunt capacitance is small making us sensitive to charge noise—e.g. noise due to charge carriers entering or leaving the capacitor. Charge noise has proven unavoidable with existing fabrication techniques ultimately limiting the coherence of charge qubits. This led to the development of the more successful Transmons qubit which uses a fairly sizable shunt capacitance ($\approx 80\, fF$) yielding a non-linearity of just 240 MHz. In Chapter 4 we use a very large capacitance to create an approximately harmonic tunable resonator.

2.5 Used in this thesis

In this chapter we have theoretically explored a linear resonator, a non-linear resonator, and a SQUID, which acts as a flux-tunable non-linear inductance. In the remaining chapters of this thesis we will use a combination of all three of these circuits. Figure 2.5 shows physical realizations a linear harmonic oscillator (**b,c**), a frequency tunable transmon qubit (**e, f**), and a nearly-linear frequency tunable resonator (**h,i**). In Chapter 3 we us linear harmonic oscillator circuits formed by $\lambda/4$ coplanar waveguide resonators as shown in Figure 2.5a and 2.5b. These resonators require just one layer of lithography

to fabricate, they are sensitive to many of the same loss mechanisms as transmon qubits, and may be characterized with a standard Vector Network Analyzer. In Chapter 4 we use a circuit consisting of a capacitively shunted SQUID that behaves as a frequency tunable resonator (Figures 2.5g, h, and i) to characterize the performance of our flux bias control electronics and wiring. While the layout of the SQUID and flux bias line geometry of the tunable resonator is similar to the transmon, the capacitance, formed by parallel plates, and SQUID critical current are much larger. The large capacitance reduces the non-linearity of the circuit and results in a nearly harmonic energy level spacing while the increased critical current reduces the minimum SQUID inductance resulting in a resonant frequency similar to the qubit. This tunable resonator circuit allows us to characterize the response of our SQUID without having to calibrate or readout qubits offering a more direct measurement of the transfer function of our control electronics as seen by qubits. In Chapters 3 and 6 we use xmon transmon and gmon transmon qubits to check the coherence of a near-term scalable PCB based package and to demonstrate a continuous set of two qubit gates (xmon circuit shown in A.3d and qubit device shown in Figures 2.5e and 2.5f).

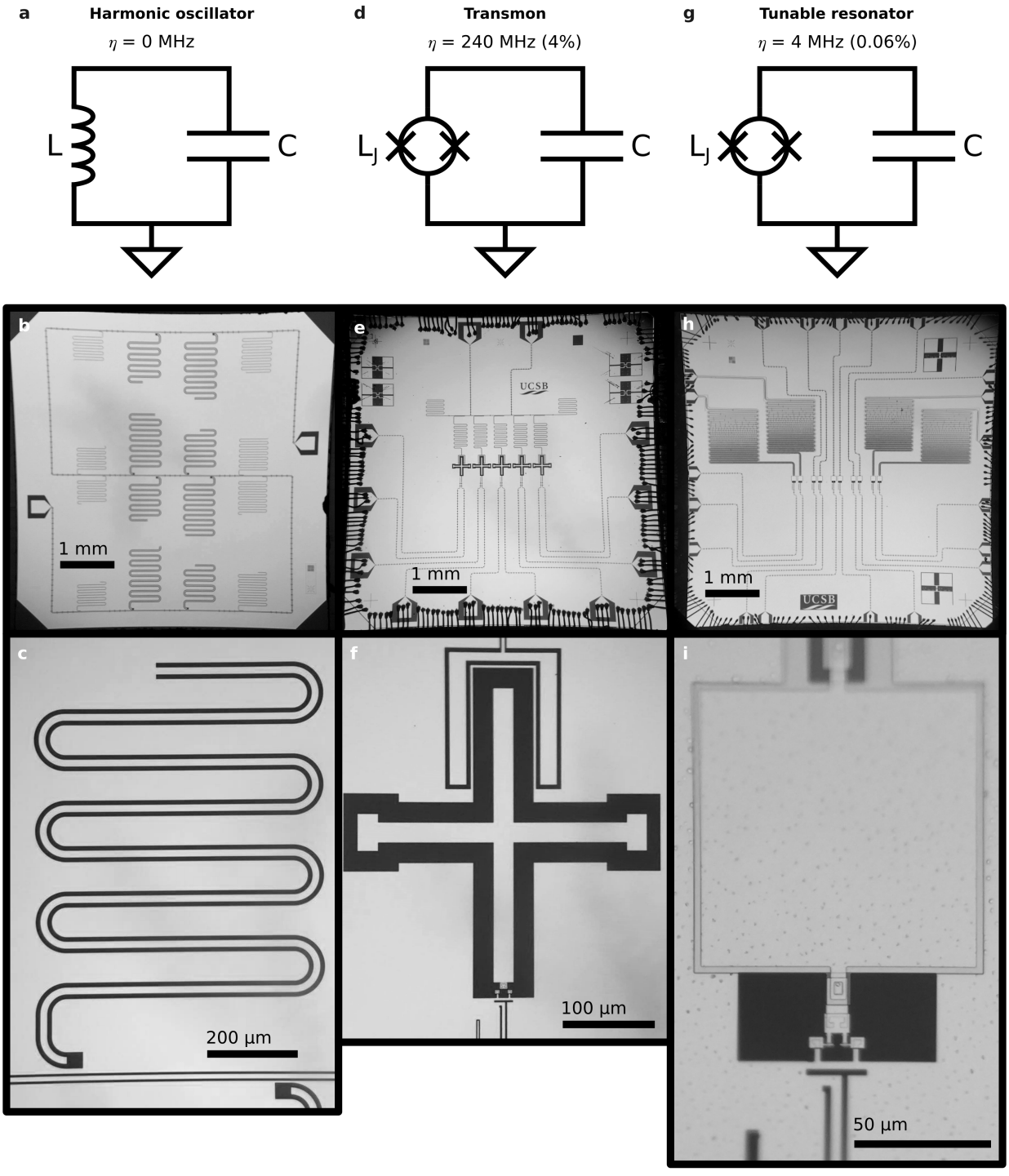


Figure 2.5: (Caption next page.)

Figure 2.5: (Previous page.) Comparison of three devices used in this thesis. **a**, Linear harmonic oscillator circuit consisting of an inductance and capacitance. **b**, Aluminum on sapphire device with sixteen coplanar waveguide resonators on one "S" shaped feed line. **c**, Zoom in of one coplanar wave guide resonator, where the L and C are both distributed along the length of the transmission line. **d**, Transmon qubit circuit where the linear inductance has been replaced by two Josephson junctions forming a SQUID which behaved as a non-linear inductance that may be tuned with an external magnet field. **e**, A five transmon device ("+" signs in the middle), coplanar waveguide resonators for dispersive readout (top center), and control wiring (left, right, and bottom). **f**, Zoom in of one transmon qubit showing the capacitance formed by the coplanar plus-shaped capacitor and the inductance formed by the SQUID loop at the bottom of the plus. **g**, Circuit diagram of a tunable nearly linear resonator. This circuit reduces the non-linearity of the qubit so that it may be driven to higher energy levels with a nearly linear resonant behavior. **h, i**, Tunable resonator device made to mimic the qubit layout, while the capacitance has been increased with a parallel plate geometry, the SQUID loop and associated control wiring are the same as the qubit design. Rather than dispersive readout, we can directly probe the impedance of this circuit with a reflective microwave phase measurement allow us to measure the time-domain flux response of our flux bias control signals.

Chapter 3

Qubit system design

Building a quantum computer is effectively the task of asserting arbitrary control over a system composed of individual quantum mechanical states. This requires two things:

1. Isolating these quantum states to avoid unintentionally exchanging energy with the surrounding environment, and

2. Coupling to these quantum states to both control their evolution and to measure their state.

In this chapter, we will consider the measures that must be taken to accomplish these goals experimentally. Where possible, we will both explain the existing infrastructure as well as give some attention to the prospect of scaling an existing systems consisting of approximately 10^1 qubits to some future error corrected quantum processor that we expect will require 10^6 to 10^8 qubits [33]. As we have seen in the previous chapter, for a superconducting transmon qubit, the two states that make up the qubit are zero or

one photons of energy oscillating in an LC resonance circuit. The necessity of superconductivity to produce low loss electrical circuits requires low temperatures, but, even more so, since these qubits operate at microwave frequencies, we must operate qubits in a thermal environment of, at most, a few tens of millikelvin to avoid thermally exciting the qubits (Section 3.1). The oscillating energy is associated with electrical currents which may generate loss if there are quasi-particles (unpaired electrons) in the superconductor (Section 3.3.3). Magnetic fields present during the superconducting transition may also induce loss by trapping flux in the superconducting film producing current vortices with normal metal cores (Section 3.3.4). If these vortices are not pinned to a defect or hole in the superconductor then these currents may generate loss as interactions move these vortices around. Additionally, if lossy dielectrics are present in the volume surrounding the qubits, then the associated electric fields may induce loss (Section 3.3.2). Furthermore, these fields may interact with other two level quantum systems (TLS) in the microscopic environment around the qubit—these TLS may couple to the qubit and cause qubit excitations to leak out.

Control of superconducting qubits is accomplished with a combination of microwave frequency electronics to excite the qubit (Section 3.2.1) and base band flux waveforms to control the qubit frequency (Section 3.2.2). We measure the state of these qubits by probing a coplanar waveguide resonator that is coupled to each qubit . When each of these control and measurement signals are coupled to the qubits, care must be taken not to limit the coherence of the qubits either actively, by injecting electronics noise, or

passively, by allowing energy to leak out of the qubit.

As one may infer from this discussion, the goal of a superconducting qubit experiment is to perform perfectly clean (residue and material defect free) fabrication of capacitively shunted SQUIDS, cool these devices to a few tens of millikelvin, and couple base band and microwave signals from room temperature electronics to these qubits at a ratio of about 2.2 electronics channels:1 qubit. Figure 3.1 provides a birds eye view of a system designed to do just that. While many divisions are possible, here the system is divided in to three parts for further discussion in this chapter. Vacuum and cryogenics are necessary to cool the quantum processor to a few tens of millikelvin. Electronics and wiring are necessary to control and measure the state of each qubit. Fabrication and packaging determine the microscopic and microscopic environment in the vicinity of our delicate qubit states and the coupling of qubits to the control and measurement electronics.

3.1 Cryogenics

Incredibly, the necessity of millikelvin temperatures is low on the list of challenges presented by expanding today's quantum processors. Call it luck or otherwise—low temperature physicists are incredibly fortunate to have one coolant, helium, that remains liquid all the way to absolute zero. At atmospheric pressure, helium is the only element that remains liquid below 14.4 K (the freezing point of hydrogen). As such, helium is the name of the game as far as continuous cooling processes go at these low temperatures. It is possible to achieve temperatures of about 1.3 K through evaporative cooling

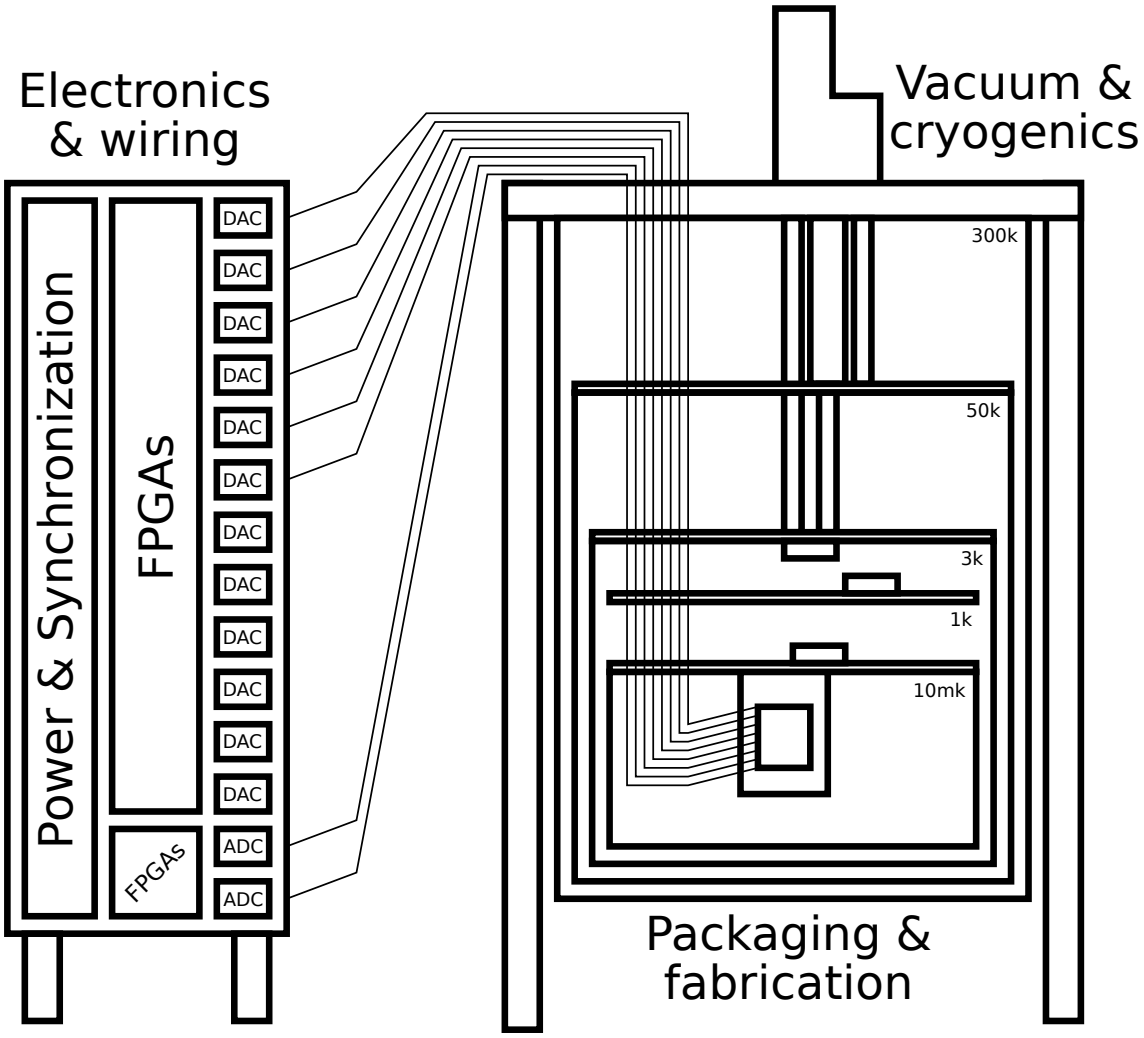


Figure 3.1: Experimental infrastructure for superconducting qubits. Here we divide the system into three parts. Vacuum and cryogenics are necessary to cool the quantum processor to a few tens of millikelvin. Electronics and wiring are necessary to control and measure the state of each qubit. Fabrication and packaging determine the microscopic and microscopic environment in the vicinity of our delicate qubit states and the coupling of qubits to the control and measurement electronics.

of He4, or 300 mK with the much more rare He3 isotope [34]. However, the cooling power of evaporative processes goes to zero as temperature goes to zero, so this is not a viable cooling process to reach 10 mK temperatures. Fortunately, another cooling process exists—He3-He4 dilution refrigeration—that functions all the way to to zero temperature and is regularly used to achieve temperatures of a few millikelvin in the lab.

Well integrated, commercially available, helium dilution cryostats, like the Janis JDry-500-QPro, provide a base temperature of 7.5 mK with a guaranteed cooling power of 14 μW at 20 mK and 500 μW at 100 mK [35]. This has proven sufficient for some of the largest superconducting qubit processors today [10]. With superconducting niobium titanium (NbTi) coaxial wiring between the 3 K and base temperature stages of the cryostat, these cryostats can support more than 200 cables with only a modest increase in the base temperature from 7.5 mK when completely empty to just 10-15 mK with 200 cables installed. Since NbTi superconducts below 10 K and the dielectric is PTFE (Teflon), these cables have an extremely low thermal conductivity when used at temperatures below 3 K [34, 36]. Due to the favorable properties of these cables, the cooling bottleneck in our existing systems is actually due to the cryogenic amplification necessary for high fidelity qubit measurement.

As of today, qubit measurement require the use of one cryogenic HEMT (high-electron-mobility transistor) amplifier for each readout transmission line which is used to measure half a dozen qubits. Each of these semiconductor amplifiers dissipate 4 to 8 mW of power on the 3 K stage of the cryostat. As the number of qubits increases, this

very quickly eats in to the cooling budget this stage which is cooled by a pulse tube cryocooler with a rated cooling capacity of 0.5 Watts at 4.2 K. If the temperature of this stage exceeds 3.5 K then the helium dilution unit will not function because it will not be able to liquefy the incoming mixture. With the existing pulse tube, we have successfully operated the cryostat with 12 HEMTs [21]. Cryomech, the pulse tube manufacturer, sells pulse tubes with 4x the cooling capacity, but eventually we will probably need a larger cooler [37].

Beyond just moving to larger existing pulse tube coolers there are a number of other possible solutions. Here we list a few in order of increasing difficulty: relocating HEMTs to a higher temperature stage (as allowed by the performance of first stage amplification), building a cryostat with more cooling power at 4 K, and designing a superconducting amplifier capable of providing enough gain so that a low-temperature HEMT is no longer necessary to preserve our measurement signal to noise ratio.

3.1.1 How cold is cold enough?

One nice feature of superconducting qubits is that their transition frequency may be engineered. As we have seen in Section 2.4 frequency tunable transmon qubits consist of a capacitively shunted SQUID. The resonant frequency of the circuit, $1/\sqrt{L_J C}$, sets the transition frequency, which may be adjusted by modifying L_J and/or C accordingly. With this in mind, the first question to address is what frequency we should choose to operate, and there is a fairly large amount of flexibility in this choice. The lowest frequency we can

operate qubits at is going to be determined by the ambient temperature of the qubits. If the temperature is too high relative to the qubit frequency then thermal energy may excite an appreciable population in our qubit. The probability of thermally exciting a qubit with a transition frequency, f_{10}, is given by the usual Boltzmann distribution,

$$P_{thermal} = \frac{e^{-hf_{10}/k_bT}}{1 + e^{-hf_{10}/k_bT}} \tag{3.1}$$

When $hf_{10} \ll k_bT$, which is where we want to operate qubits, $P_{thermal} \approx e^{-hf_{10}/k_bT}$, which is to say that thermal excitations are exponentially suppressed as the temperature is reduced below the qubit transition energy. Commercially available cryostats readily achieve base temperatures in the 5-15 mK range, but, due to the poor scaling of the electron-phonon interactions (which decreases with T^5 at mK temperatures [38, 39]), we find that our qubits typically thermalize to about 20 mK to 30 mK. Assuming an electron temperature of 30 mK, to keep the thermal population of the qubit below 0.1%, the qubit frequency should be at least 4 GHz.

As an interesting aside, it is worth noting that the observation of poorly thermalized Josephson Junctions is not unique to qubits. In fact, a series chain of Josephson Junctions forms a primary thermometer. These devices, known as a Coulomb Blockade Thermometers (CBTs), have a temperature-dependent differential conductance (dI/dV) that is dependent on only the temperature and the number of JJs in the chain. As CBTs were first developed [40, 41, 42, 43], it was observed that the differential conductance ap-

peared to saturate at temperatures below about 20 mK indicative of poor thermalization. This is notable because CBTs offer a direct measurement of the electron temperature with a qubit-compatible fabrication process whereas most thermometers would have to be bonded to the chip or package and make only phononic contact to our qubit metal layer. In particular, since qubit fabrication already requires Josephson junctions, a CBT could be added on-chip, using the same lithography and metal deposition steps as the qubit SQUID fabrication, with fairly little effort. CBT researchers have since been able to thermalize devices to temperatures as low as a few mK by employing gold "heat sinks" that mediate an electron-phonon interaction with a coupling as much as two orders of magnitude larger than in superconducting aluminum [44, 45, 46]. If this technique were deployed in superconducting qubit systems and enabled us to thermalize our qubits to 10 mK, the minimum qubit frequency could be reduced to 1.4 GHz with a 0.1% thermal population, or we could further suppress the thermal occupation of our 4-7 GHz qubits.

The upper frequency limit for operating superconducting qubits is determined by the gap energy of the superconducting material. As the qubit energy approaches the gap energy, the qubit drive may break cooper pairs and excite quasi particles. For aluminum the gap is $2\Delta/2\pi = 88$ GHz. However, for now at least, it is not advantageous to operate at higher frequencies. Doing so incurs additional cost without any particular added benefit. Higher frequency microwave electronic components (above 10 GHz) are much more expensive and difficult to design due to the smaller wavelengths involved, and the lack of commercial demand driving low cost development. As such, most superconducting

qubits operate in the in the 4-8 GHz frequency range which is basically a choice to operate with commercially available cryostats and off-the-shelf microwave electronics originally developed by and for the communications industry. However, as quantum computing grows and possibly becomes a technology driver, it may make sense to re-examine this choice if opportunities present themselves to improve either the cost and/or performance of qubit control systems through the development of some novel control scheme.

3.2 Electronics

As was shown in Figure 3.1, the control and measurement electronics may be thought of in three parts:

1. The power and synchronization backbone that is shared between all components in the system.

2. The digital back-end on each board to handle component power requirements and digital communication to a field programmable gate array (FPGA) which controls,

3. The low noise high performance analog front end.

This design, used in many large channel count electronic test, control, and measurement systems, provides a scalable data and synchronization architecture. Sharing a clock and trigger between different groups of control electronics, allows this architecture to scale synchronization to a very large number of channels. Sharing a clock and trigger can allow for the synchronization of board groups controlled by different computers which offers a

straightforward parallelization of the control computing. The digital back end of both the digital to analog converter (DAC) boards and the analog to digital (ADC) boards may share a similar power and synchronization architecture that essentially serves to synchronize an FPGA on each board. Then, these FPGAs allow for extremely flexible control of the analog front end in addition to fast and configurable custom analysis of the signals that are being generated or measured. Finally the analog front end provides the appropriate signal conditioning for a given control or measurement line.

Microwave control of superconducting qubits is well suited a frequency up-converting architecture with an IQ mixer. Qubits are sensitive to resonant frequency drives, and the main signal bandwidth requirement is that the spectral width of microwave envelop pulses must be smaller than the nonlinearity (250 MHz). For this reason, we use a 1 GS/sec DAC which minimizes the cost and data transfer rates while giving us to ability shape the spectral content of our pulses in the desired bandwidth to avoid leakage. For pulse shaping, we need a signal bandwidth on the order of the qubit non-linearity (200 MHz or so), plus, for the foreseeable future, extra bandwidth to spread out qubits and avoid TLSs so that many qubits may share the same local oscillator (LO). Because the ultimate goal is to have many qubits, it makes sense to avoid larger than necessary bandwidths to reduce the required data transfer rates. Another benefit of this architecture is that the same 1 GS/sec DACs used for microwave up conversion are also well suited to flux control when used at base band frequencies. From here we will now dive in to the critical analog specifications of these electronics.

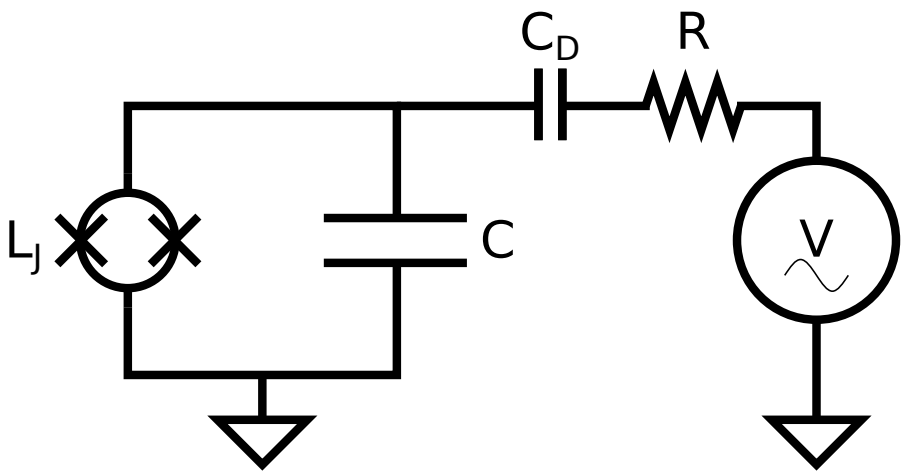


Figure 3.2: Capacitive coupling of a microwave drive line to a tunable transmon.

3.2.1 Microwave control

Proper microwave control requires electronics with a large dynamic range; low noise while the control line is idle and the ability to strongly drive the qubit to perform fast gates. Here we consider these requirements in the context of the capacitively coupled voltage source in Figure 3.2. If the drive voltage $\mathrm{V} = \mathrm{V}_0 \cos \omega \mathrm{t}$, where ω is the qubit frequency then the qubit state will oscillate between the $|0\rangle$ and $|1\rangle$ states at the Rabi frequency [47],

$$\Omega = \frac{C_d V_0 Q_{zpf}}{\hbar(C + C_d)} \tag{3.2}$$

where $Q_{zpf} = \sqrt{\hbar/2Z}$ is the zero point fluctuation of the charge operator and $Z = \sqrt{L_J/C}$ is the qubit impedance[1]. This drive voltage may be the result of either intentionally applied microwave pulses or the result of always-on electronics noise. As such,

[1]Higher impedance qubits have a larger non-linearity but also a higher sensitivity to charge noise [48].

the dynamic range, or, the ratio of the maximum drive amplitude of our electronics to the spectral noise density of our electronics will determine the ratio of the fastest and slowest achievable Rabi frequencies. Fast Rabi oscillations are used to perform qubit gates, and the slow oscillations should be appropriately "slow" compared to our qubit coherence.

We can first consider performing gates. We want to perform gates as quickly as possible, but we must keep the spectral width of our control pulses narrow so that it does not significantly overlap with the $|1\rangle \leftrightarrows |2\rangle$ transition frequency. In order to control the spectral bandwidth of our microwave drive, we convolve (multiply) the constant wave (CW) microwave drive with a smooth cosine envelope[2]. Cosine envelopes also have the nice feature that the first and second derivatives are zero at the start and end of the pulse—a favorable property when it comes to generating these wave functions with real-world electronics [49]. The spectral width of the resulting cosine envelope microwave pulse is $\omega_{\text{drive}} \pm \omega_{\text{env}}$. This is the origin of the oft quoted single qubit gate "speed limit". If cosine enveloped microwave pulses shorter than $1/\eta$ are used (where η is the qubit nonlinearity) then these pulses will have spectral power at the $|1\rangle \leftrightarrows |2\rangle$ transition frequency and leakage will occur. Shorter gates require larger drive amplitudes, and, with typical transmon non-linearities ranging up to about 250 MHz, we will consider possible gate lengths of 5 ns, 10 ns, and 20 ns corresponding to Rabi frequencies of 200 MHz, 100 MHz, and 50 MHz (multiply by 2x for cosine pulses).

[2]When using a cosine pulse an additional factor of 2 is required in the maximum drive voltage to achieve the same Rabi frequency since the integral of a cosine is 1/2 relative to a rectangular envelope.

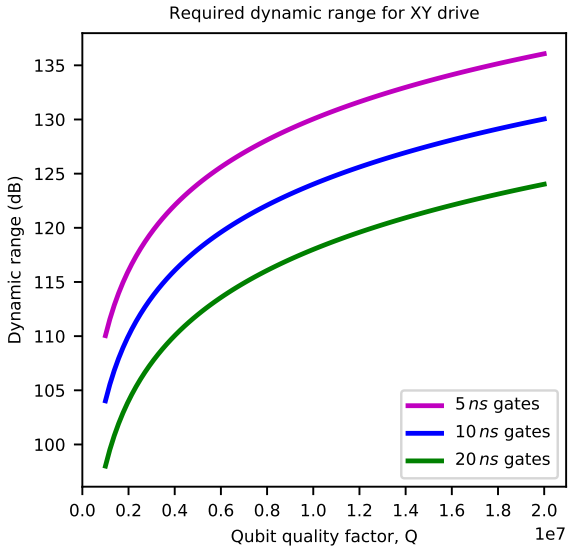


Figure 3.3: Dynamic range requirement for the qubit microwave drive electronics.

On the other side of this problem, the higher the qubit coherence, the more stringent the low-noise requirement. As qubit coherence improves, we will want to decrease electronics noise accordingly. Here, we will consider a minimum Rabi frequency that scales with the quality of our qubits. While today's qubits have quality factors in the range of 1-2 M[50, 51, 52], one may consider the requirement for controlling improved qubits in the future. In Figure 3.3 we plot the necessary dynamic range for three gate lengths as a function of the qubit quality factor, assuming a 6 GHz qubit, by assuming we want the minimum Rabi frequency to be at least 10x slower than the qubit coherence time. In reality, you must also design in some margin above these requirements to allow for a buffer in the attenuation budget due to the variation and discrete available values of attenuators.

To this point we have considered the active effects of connecting a voltage source

to our qubit, but that is not the only concern. Care must also be taken to make sure that the coupling capacitance, which also couples the qubit to a lossy 50 Ω environment, does not passively limit the qubit coherence by allowing too much energy to leak out. If the coupling is too strong, it may become a dominant source of qubit decoherence. The loaded quality factor for a qubit with a capacitively coupled microwave drive line as in Figure 3.2 is given by [53],

$$Q_L = \frac{C}{RC_d^2\omega} \tag{3.3}$$

where C in the qubit shunt capacitance, R in the impedance of transmission line and voltage source, $\mathrm{C_d}$ is the coupling capacitance to the drive line, and $\omega = 1/\sqrt{\mathrm{L_J C}}$ is the qubit frequency. Q_L sets a limit on the coherence of our qubits. The qubit capacitance, frequency, and the drive line and source impedances are relatively fixed by other constraints, but we can freely vary the coupling capacitance, C_d. As fabrication improvements increase the qubit's internal quality factor, we will need to reduce the drive line capacitance $\mathrm{C_d}$ to avoid limiting the coherence of the qubits. To avoid damping the qubit we can make sure the loaded quality factor is considerably larger than the internal quality factor of the qubit by keeping C_d sufficiently small. Taking $Q_L > 10Q_i$, we can find value of C_d such that the control coupling will not passively decohere our qubits.

$$10Q_i < \frac{C}{RC_d^2\omega} \implies \tag{3.4}$$

$$C_d < \sqrt{\frac{C}{10RQ_i\omega}} \tag{3.5}$$

Now that we have electronics with an acceptable dynamic range and we know the maximum allowable coupling capacitance, the final task is to determine the correct amount and placement of attenuation—one wonderful feature about working at cryogenic temperatures is that Johnson noise is hardly a problem! While thermal electronics noise often presents a floor for room temperature electronics, that is not so much the case for cryogenic systems operating at microwave frequencies. The usual Johnson noise is linearly dependent on temperature dependent, given by [54],

$$S_V = 4RK_bT \tag{3.6}$$

where R is the resistance, K_b is the Boltzman constant, and T is the temperature. In the same year, Nyquist derived the more general quantum result, where the spectral noise density is given by [55],

$$S_V = \frac{2Rhf}{e^{hf/K_bT} - 1} \tag{3.7}$$

When $K_bT \gg hf$, Nyquist's version collapses to Johnson's rule. At room temperature this condition fails in the Terahertz regime. In our case, 6 GHz corresponds roughly to a temperature of 300 mK. Below this temperature, the noise drops off exponentially. So, by placing additional attenuation on the base temperature (10 mK) stage of our cryostat, we can overcome a very large amount of noise ultimately achieving whatever absolute

noise level is required by our qubits. Thus, the key requirement of our electronics is their over all dynamic range!

3.2.2 Frequency control

As we saw in Section 2.3, the frequency of flux-tunable qubits may be varied by changing the magnetic flux through a SQUID. This allows for the freedom to choose the idle frequency of the qubit which is particularly useful for avoiding TLS defects in the qubit spectrum. Likewise, as in Chapter 6, dynamic flux control may be used to perform either single or two qubit gates by temporarily changing the qubit frequency and/or relative detuning of a pair of qubits. These benefits come at the expense of exposing the qubits to an additional noise source (flux noise), dealing with DC coupled pulse distortion, and accommodating the necessary milliamps DC control currents at millikelvin temperatures in the cryostat. Figure 3.4 shows a typical qubit frequency versus applied flux response for a transmon qubit. Following from 2.22, the qubit frequency is $\propto \sqrt{\cos \Phi/\Phi_0}$.

Typically this flux control is achieved with an inductively coupled flux bias line terminated near the SQUID as shown in Figure 3.5. The shorted transmission line shares a mutual inductance, M, with the SQUID loop, and a mutual inductance, M', with the larger qubit loop such that currents applied to the bias line result in a magnetic fields through the SQUID and qubit circuit. Generally speaking, M' is an undesirable parasitic mutual inductance as it couples to the qubit mode and will decohere the qubit. The loaded quality factor of a qubit circuit inductively coupled to a drive line as in Figure

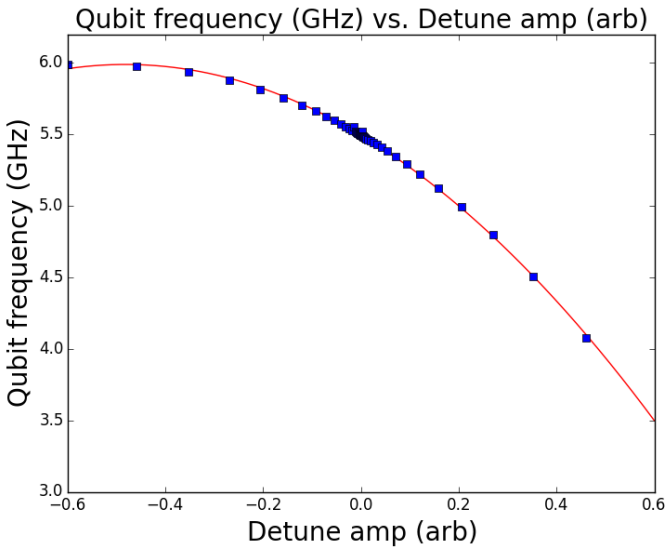


Figure 3.4: Qubit frequency vs flux response for a tunable transmon qubit.

3.5 is given by[53],

$$Q_L = \frac{R'}{Z_q} \tag{3.8}$$

$$\approx \frac{R}{M'^2\omega_q^3 C}. \tag{3.9}$$

For a 50 Ω line inductively coupled to a 6 GHz transmon qubit with a capacitance of 80 fF, a loaded quality factor of 1 M to 10 M corresponds roughly to a parasitic mutual inductance, M', of 1 pH to 3 pH. Fortunately, this parasitic mutual inductance my be designed to be very small, much smaller than 1 pH, by taking advantage of symmetry in the qubit layout[3].

The desired mutual inductance, M, is between the bias line and the SQUID loop

[3]In the case of flip chip hybrid devices, as in Chapter 5, if the bias line and qubit are on different chips, one must be careful to not be overly dependent on alignment for minimizing M'.

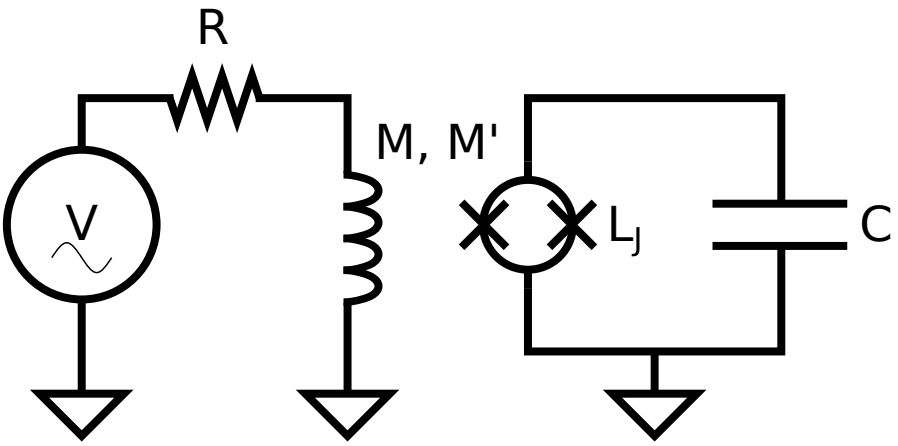


Figure 3.5: Inductive coupling of a flux bias line for a flux-tunable transmon qubit. The bias line shares a mutual inductance, M, with the SQUID loop, and, if improperly designed, may share a mutual inductance with the larger qubit loop, M'.

and allows us to change the Josephson inductance and consequently the qubit frequency. While this coupling does not introduce a decoherence channel for the qubit, it does expose the SQUID to broadband electrical noise which translates to noise in the qubit frequency. This frequency jitter results in dephasing of the qubit. The magnitude of this frequency jitter depends on both the flux bias of the qubit and the characteristic noise spectrum. The flux sensitivity is the derivative of Figure 3.4, $d\omega_{10}/d\Phi$, and ranges from near zero 0 at zero bias to roughly $10\,\mathrm{GHz}/\Phi_0$ over a typical range of operating biases. In the worst case, where the qubit evolves freely in the presence of this noise for the duration of an algorithm, the qubit is exposed to the full noise spectrum. However, in typical algorithms, microwave gates effectively flip the sign of the electrical noise fairly rapidly. The result is that the noise spectrum is effectively windowed by these operations and the qubit is effectively exposed to a fairly narrow frequency band of noise centered at a frequency determined by the time between microwave rotations[50]. With typical

10-25 ns gates, the center frequency is above 10 MHz where the 1/f electronics noise has fallen below the white noise floor. Assuming a spectral noise density, S_V, with a white noise spectrum, the magnitude of this dephasing, $1/T_2^*$, is given by[53],

$$\frac{1}{T_2^*} = \frac{1}{4}\Big(\frac{d\omega_{10}}{d\Phi}\frac{M}{R}\Big)^2 S_V \tag{3.10}$$

We now have two competing interests. On one hand, we want to use a large mutual inductance to reduce the necessary control currents—larger currents mean more heat dissipation in the cryostat—with enough qubits this may eventually be an issue. On the other hand, the qubit dephasing increases with the square of the mutual inductance. Here, we fall back on the typical cryostat design for guidance. Since considerably more cooling power is available at 3 K than at lower temperatures, we can consider the effect of 3 K noise on dephasing. With 3 K noise ($\approx 0.1\text{nV}/\sqrt{\text{Hz}}$), and a mutual inductance of 2 pH, T_2^* would be limited to than 300 μs. Since we need to be able to apply a substantial fraction of about 1 Φ_0 of magnetic flux to the SQUID for calibration and control, with a 2 pH mutual, this requires a fairly substantial $\approx$ mA of current. While there may be some room to increase M to lower the current requirement, it seems infeasible to go below 0.1 mA, and even that would likely require attenuation at temperatures lower than 3 K. As such, the development of superconducting interconnects in Chapter 5 allows for maximum flexibility in designing quantum processors since they can handle the necessary flux control currents without introducing heat dissipation on the processor.

In addition to the challenges presented by these relatively large control currents in a cryogenic environment, precise shaping of flux control pulses is very challenging. Whereas microwave control requires careful pulse shaping over a fractional bandwidth ($\delta f/f$) of about 0.05, qubit frequency control requires careful pulse shaping over many orders of magnitude in frequency from DC to several hundred MHz. It is difficult to design a system that functions optimally over such a wide range and the result is long flux settling tails on nominally short control pulses. Measuring these flux settling tails is the objective of Chapter 4.

3.3 Packaging

Among the less glamorous challenges in quantum computing is the practical consideration of how to package superconducting qubits to both shield them from the environment and facilitate bringing in control wiring for each qubit. The quantum states of superconducting qubits occupy a non-negligible volume that must be considered when engineering the environment in which qubits are housed. Shielding, packaging, and fabrication must all be optimized to preserve qubit coherence while enabling the necessary control coupling to and from qubits. While I did not spend much time optimizing the fabrication used to improve qubit coherence, I did spend considerable time testing our qubit packaging.

Before my time in this group began, a two piece machined aluminum package shaped as a 12 sided prism as in Figure 3.6b had become the *de facto* standard. Signal lines are brought inside the package to where the chip is mounted by drilling a hole from the

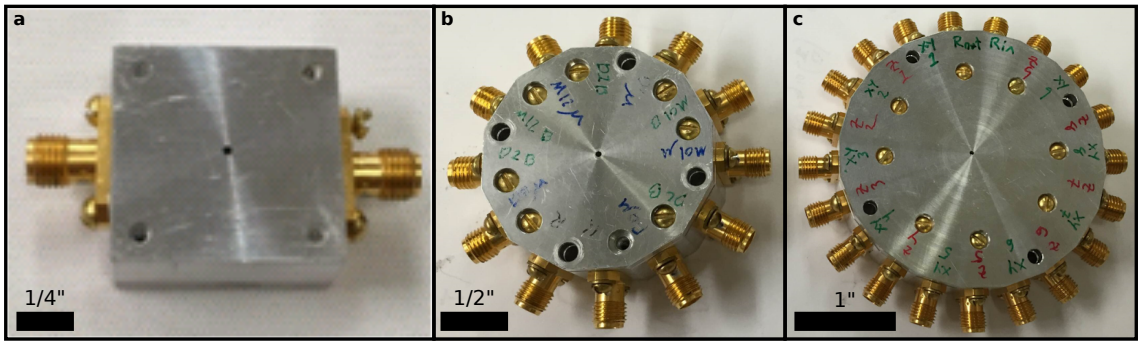


Figure 3.6: Classic UCSB Martinis group machined aluminum packaging for superconducting circuits. **a** 2, **b** 12, and **c** 28 port variants. Each package is made from two pieces of 6061 Aluminum, the base and a lid. The base houses a cavity for the superconducting chip and signal lines are routed to the interior of the package via holes drilled in the base of the package. Narrow PCB inserts, soldered to SMA panel mount connectors, are attached to the exterior of the package. Ultimately the high aspect ratio of the necessary through holes for the signal lines reaches the machining limit somewhere in the low 30's of holes.

perimeter of the box and inserting a PCB trace with dimensions appropriate to maintain a 50 Ohm transmission line. Over the years, variants of this package accommodating 2 (Figure 3.6a) to 20 (Figure 3.6c) signal lines were machined to accommodate amplifiers, resonators, and larger multi-qubit processors. Packages of these type housed basically every superconducting resonator, amplifier, or qubit device performed in our lab up until about 2016. This package was designed to shield superconducting devices from stray photons, to screen out magnetic fields, and to allow $50\,\Omega$ connections for control and readout of qubits—and it was quite a successful design! We have measured resonators and qubits with quality factors exceeding 1M [56, 51], achieved readout fidelity of 99.8% [57], single qubit gate fidelities consistently exceeding 99.9% [58], and two qubit CZ gates at 99.5% [47]. What this box was *not* designed to do to was facilitate the hundreds of signal lines necessary to interface control and measurement electronics with a large

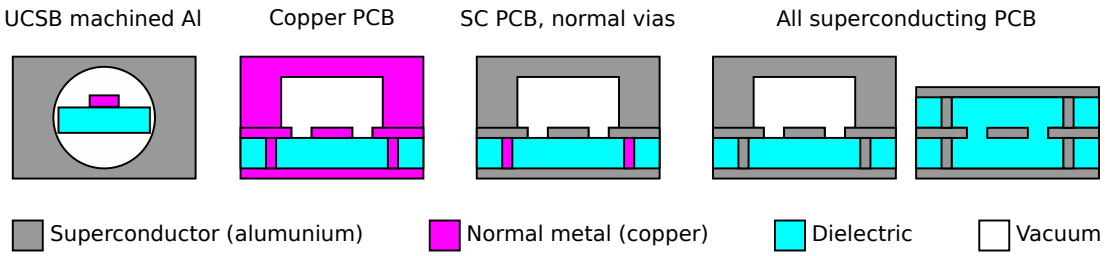


Figure 3.7: Comparison of transmission line cross sections for the machined aluminum package and a similar printed circuit board (PCB) variant. The machined aluminum package has a superconducting aluminum ground plane and a normal metal (copper) signal line. The left three packages show three PCB-based variants we tested before ultimately settling on the three layer design on the far right. We found that introducing normal metals in the ground plane, even just the vias, was enough to introduce settling time constants of $\approx 100\,\mu$s (see Chapter 4

quantum processor. In this section I will describe what effectively became our acceptance testing for a near-term-scalable printed circuit board based package similar to the ones now used at Google. In Figure 3.7 shows cross sections of the package signal lines in the machined aluminum package and several PCB-based alternatives. While we would have ideally liked to use a standard copper PCB, we found that introducing normal metals to the ground plane added long settling tails to our flux control pulses and we ultimately settling on an aluminum plated copper design as presented on the far right of Figure 3.7.

In the following sections we will first review coplanar waveguide resonators and why they are used as a proxy for qubit performance in many scenarios. We will then use these resonators to check the infrared photon and magnetic field shielding offered by the new printed circuit board chip mount and confirm that the coherence of resonators and qubits is not degraded. We will then look at the non-ideal flux bias settling in these packages—this measurement instigated the project described in Chapter 4. Finally, we

will take a fun look at one of my favorite tools I have worked with as a graduate student: the fully automatic wire bonder!

3.3.1 Coplanar waveguide resonators as qubit proxies

Successfully running an experiment in a cryogenic environment requires a high degree of care and preparation. It takes roughly a day to cool the cryostat down from room temperature to its base temperature and it takes roughly a day to warm back up. Furthermore, while cold, the experiment is fairly opaque. The superconducting device, cryogenic amplifiers, and all of the associated wiring become a black box with relatively few inputs in output and many potential points of failure—a challenging environment to troubleshoot. This inaccessibility instills a certain aversion to change in the heart of many low temperature experimental physicists. As such, the ultimate acceptance test for any qubit-adjacent hardware modifications (or really, anything inside the cryostat) is to perform an A/B test confirming the performance of a known qubit device before and after any changes. Making as few simultaneous changes as possible greatly simplifies the debugging process. Unfortunately, changing a particular component sometimes necessitates additional changes. In the case of testing different packages, while it is possible to re-package a single die multiple times, this process is invasive enough that, if a subsequent result was in the negative, repackaging could not be ruled out as a culprit. In fact, in this series of experiments, we chose to accept chip to chip variation using dies that were fabricated at the same time on the same 3" wafer rather than repackaging the same die

multiple times[4]. While characterization times have improved substantially[5], fabricating and preparing qubits remains a somewhat complicated, multi-layer process. Because of these challenges, coplanar waveguide (CPW) resonators have been successfully developed as a proxy for qubit performance when it comes to characterizing materials and fabrication processes [51, 56].

Figure 3.8a and c show a side by side comparison of sixteen CPW resonators and five transmon qubits. Both are resonant circuits operating at a frequency in the 4-7 GHz frequency range. While the qubits are frequency tunable across this range, the CPW resonators operate as a single frequency. The qubit capacitor and CPW resonator are fabricated using the same substrate preparation, metal deposition, and lithography procedures allowing for the preparation of similar microscopic "environments.". As such, both devices will experience similar materials-related loss due to surface contaminants, substrate defects, etc. Similarly, since both circuits use superconducting aluminum, both circuits are sensitive to the presence of quasi-particles (unpaired electrons that produce resistive losses), and loss due to the normal core of vortices caused by ambient electric fields. While resonator performance is not completely indicative of qubit performance, they are a good initial test since fabricating resonators may be done with a single layer of lithography (instead of 5+ for qubits), and measuring the quality factor may be done with simple S-parameter measurements using a standard Vector Network Analyzer. Qubits,

[4]The concern with repackaging the same die is that removing wire bonds from the die will, if the wire bond has made good contact, tear off the thin aluminum film leaving a marred metal layer that cannot be re-bonded to in the same places.

[5]Five years ago, a fairly basic characterization of a five qubit device could take 2 weeks. Now, a more thorough characterization can be completed in about 30 minutes.

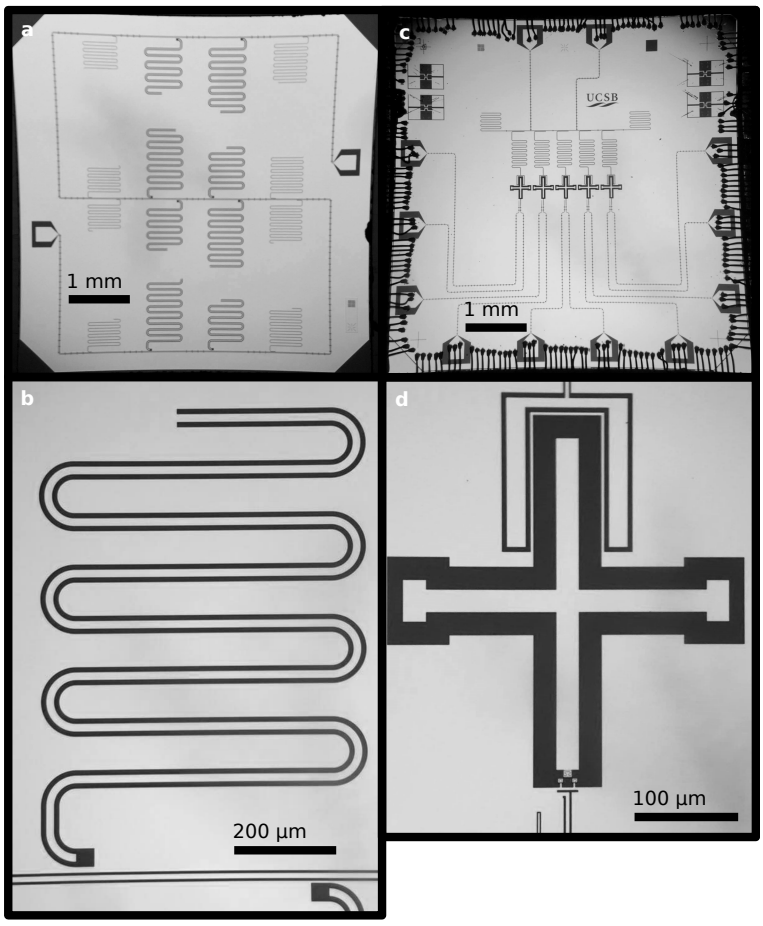


Figure 3.8: Side by side comparison of coplanar waveguide (CWP) resonators and qubits. **a** Sixteen CPW resonators. **c** Five transmon qubit chip. **b** Zoom in of one CPW resonator, and, **d** zoom in of one transmon qubit. Due to their similar geometry both of these resonant circuits are sensitive to many of the same loss mechanism.

on the other hand, require careful calibration of both control and readout before coherent can be measured.

3.3.2 Coherence check

A critical check experiment for any potential qubit packaging solution is to measure coherence times. Superconducting qubits are quantum resonators; if we excite one of these qubits by placing one photon of energy into the qubit it will oscillate back and forth at the qubit frequency. Each time the energy oscillates, there is some probability that the photon will leak out of the qubit due to interactions with the environment around the physical circuit elements. The equal probability to leak out of the qubit with each cycle results in an exponential decay in the probability of the photon remaining in the qubit. One may consider the quality factor of the qubit which is the number of oscillations required before the probability that the photon is still in the qubit has been reduced by a factor of $1/e \approx 0.368$. To convert this to a coherence time, one may multiply the quality factor by the period of oscillation (the qubit frequency), which is designed to be 4-7 GHz. Typically, interactions with the microscopic environment around the qubit limit coherence. Considerable effort has gone in to developing fabrication processes that produce resonant circuits with quality factors of 1 M and beyond [56, 51]. Much of this research is focused on the fabrication techniques used to prepare and fabricate these aluminum on sapphire (or silicon) circuits, but the package we place these circuits in also plays a non-negligible part in forming the relevant "environment" encompassing these

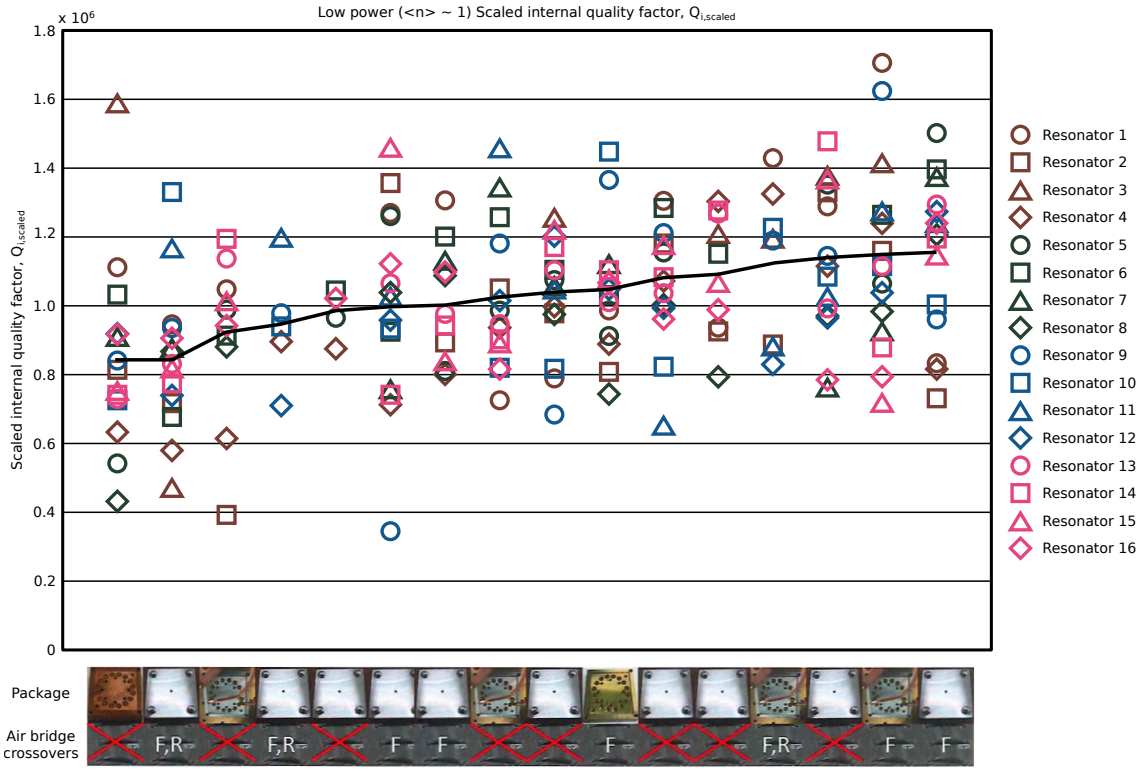


Figure 3.9: Scaled low-power internal quality factors of resonators in a variety of packages with and without air bridges. There was no discernible trend and all of the results were pretty close to the expected fabrication limits. In this instance, we scaled the quality factors to account for resonator geometry to allow for a comparison between resonators in each package, we also subtracted off the impact of the high power quality factor (which varied significantly, see Figure 3.10 for each of these devices.

circuits.

To test our packaging, we began by measuring low-power ($n \approx 1$) internal quality factors of superconducting quarter wave coplanar waveguide resonators, with and without feedline crossovers as shown in Figure 3.9. Here we plot a scaled low power internal quality factor of up to 16 resonators per die, testing dies with and without feedline and/or resonator air bridge crossovers. The internal quality factor was adjusted for resonator geometry and for the variation in high power quality factor, which varied considerably

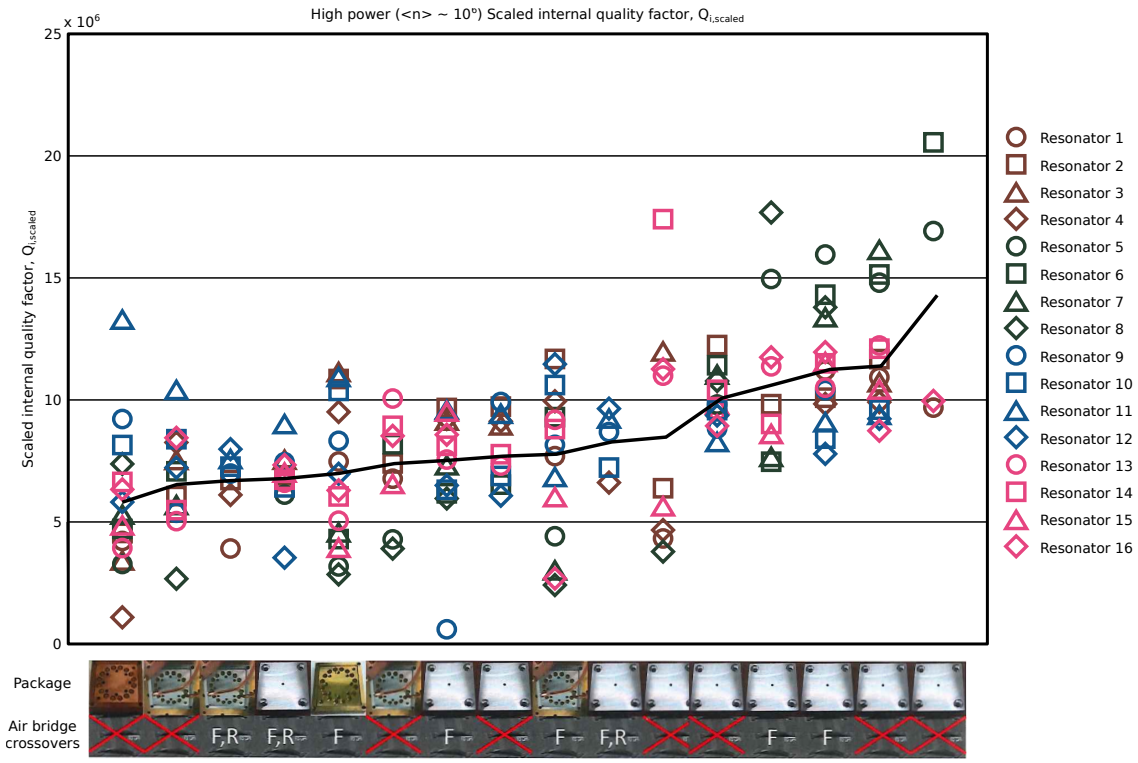


Figure 3.10: Scaled high-power internal quality factors of resonators in a variety of packages with and without air bridges. In this instance, we scaled the quality factors to account for resonator geometry to allow for a comparison between resonators in each package.

for these samples (see Figure 3.10. While there is considerable variation in the low power internal quality factor, the performance was generally good (at the limit of our fabrication at the time), and it did not seem dependent on the particular package used,

After confirming a reasonable CPW resonator quality factors for each of these packages, we moved on to the real acceptance test—to confirm that the package did not degrade the performance of our qubits[6]. For this experiment, we fabricated 5-qubit test devices as shown in Figure 3.8b that were the same as devices previously measured in the machined aluminum package, but with the signal bond pads routed to accommodate the four-per-side layout of the PCB based package rather than our traditional three-per-side of the machined aluminum box. Due to limited wiring in our shared dilution refrigerator, I was only able to wire up three qubits in each of two packages on two separate cooldowns in our wet dilution cryostat. We used swap spectroscopy to characterize the qubit coherence over a frequency range of roughly 3 GHz on the low end up to the maximum of each qubit, ≈ 6 GHz. In Figure 3.11 and we report the average and standard deviation of all measured qubits and frequencies. In all cases the coherence times were acceptably comparable to coherence times in the machined aluminum box.

[6]At the time, using electron beam evaporated aluminum on sapphire fabrication, our expected coherence times were roughly 20 μs[56].

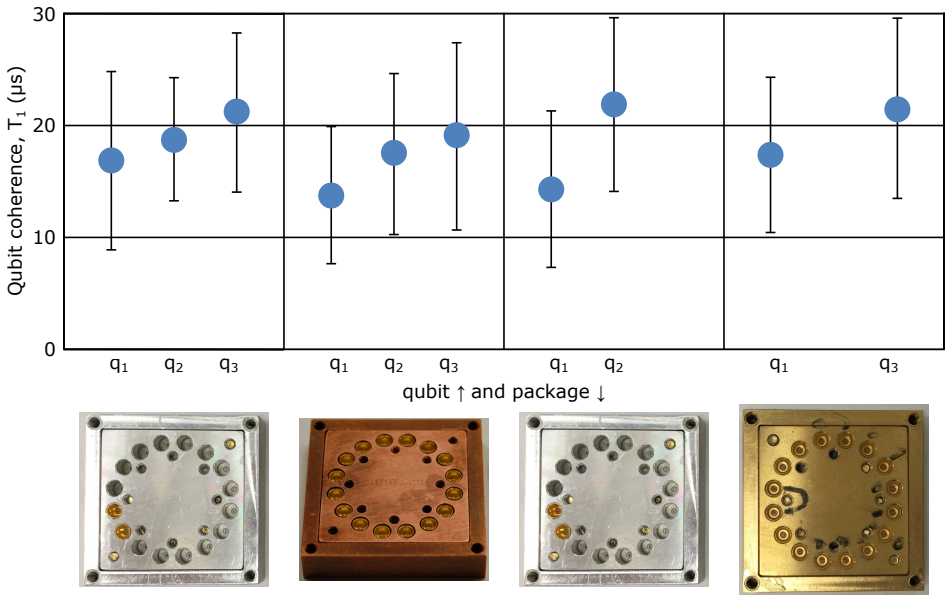


Figure 3.11: We tested the coherence of 2 to 3 qubits in 4 different packages (2 in each cooldown). We found that the average values were consistent with the performance of qubits in our standard machined aluminum package.

3.3.3 Blackbody radiation

This section closely follows the work in reference [59] which was performed using the machined aluminum chip mount. In that work, multistage infrared shielding for our Adiabatic Demagnitization Cryostat (ADR) was developed to reduce the quasiparticle density in our superconducting aluminum circuits. This shielding was designed to block or filter black body radiation from higher temperature stages in the cryostat from reaching inner cavity of the lowest temperature stage where our chips are wirebonded within the machined aluminum package. Here, we reproduce that experiment using the new PCB-based package to confirm that our cryostat shielding is sufficient. We use similar superconducting coplanar waveguide resonators (with higher internal quality factors

than in the original work) factors allowing us to place a smaller, lower-bound on the quasiparticle density.

Quasiparticles are single electrons within a superconductor that have not joined the Cooper pair condensate. As such, these unpaired electrons induce loss in our resonant circuits as they move through the material. The amount of loss is proportional to the quasiparticle density, n_{qp}, [60, 61],

$$\frac{1}{Q} = \frac{\alpha}{\pi}\sqrt{\frac{2\Delta}{hf_r}}\frac{n_{qp}}{D(E_F)\Delta} \tag{3.11}$$

where Δ is the superconducting energy gap, $D(E_F)$ is the two-spin density of states, and α is the kinetic inductance fraction [62]. The rate equation for the total number of quasiparticles, N_{qp} is given by,

$$\frac{\partial N_{qp}}{\partial t} = \frac{P}{\Delta} + G - RN_{qp}^2 \tag{3.12}$$

where P is the total power absorbed from photons with an energy large enough to break at least one Cooper pair (e.g. $hf > 2\Delta$), G is the thermal generation term due to pair breaking by phonons [63] and R is a material-dependent recombination constant. When the absorbed power, P, exceeds the thermal background, the first term dominates and the quasiparticle density become proportional to $\sqrt{P/\Delta}$. For aluminum, $2\Delta = 88\,\mathrm{GHz}$, which means that it can absorb 96% of the radiation emitted by a black body at a temperature of 4.2 K [59]. As we mentioned in Section 3.1, it is typical for cryostats using a pulse tube cryocooler to have a 3-4 K cooling stage, which means that the sample

must be placed inside of a light tight shield that is held at an even lower temperature to avoid black body radiation from the higher temperature stages of the cryostat.

In Figure 3.12 we test the effectiveness of our infrared shielding using superconducting resonators housed in the new, PCB-based package. Outside the sample enclosure, we use the same filters, light-tight shield, and absorptive black coating on the inside of the shield all of which is inside an Adiabatic Demagnitization Refrigerator (ADR) of the same model as in reference [59]. After a magnet cycle of the ADR, with the cryostat at its base temperature, the sample is at 50 mK and the next cryostat stage is at about 3 K. We begin measuring the high-power ($< n_{\text{photons}} > \approx 10^6$) internal quality factor of eight coplanar waveguide resonators. Four of these resonators are 3 μm center trace-2 μm gap resonators with frequencies near 4 GHz and the other four 15μm-10μm resonators with frequencies near 6 GHz. We then turn off the helium compressor used to drive the pulse tube cryocooler, removing the cooling power from the 3 K stage of the cryostat so that it begins warming up due to absorbed blackbody radiation and heat conducted through the physical connections to higher temperature states. We continue repetitively measuring the high-power internal quality factor of all eight resonators while the outer stage warms up to 20 K, the highest temperature we could reach before the sample stage temperature exceeded 150 mK where we would begin to expect additional loss in the resonators due to thermal quasiparticles. As shown in Figure 3.12 we observe resonator loss, independent of the cryostat stage temperature, as low as $1.8 \cdot 10^{-7}$—about a factor of two better than the lower bound in the original experiment. In this case we attribute the improved

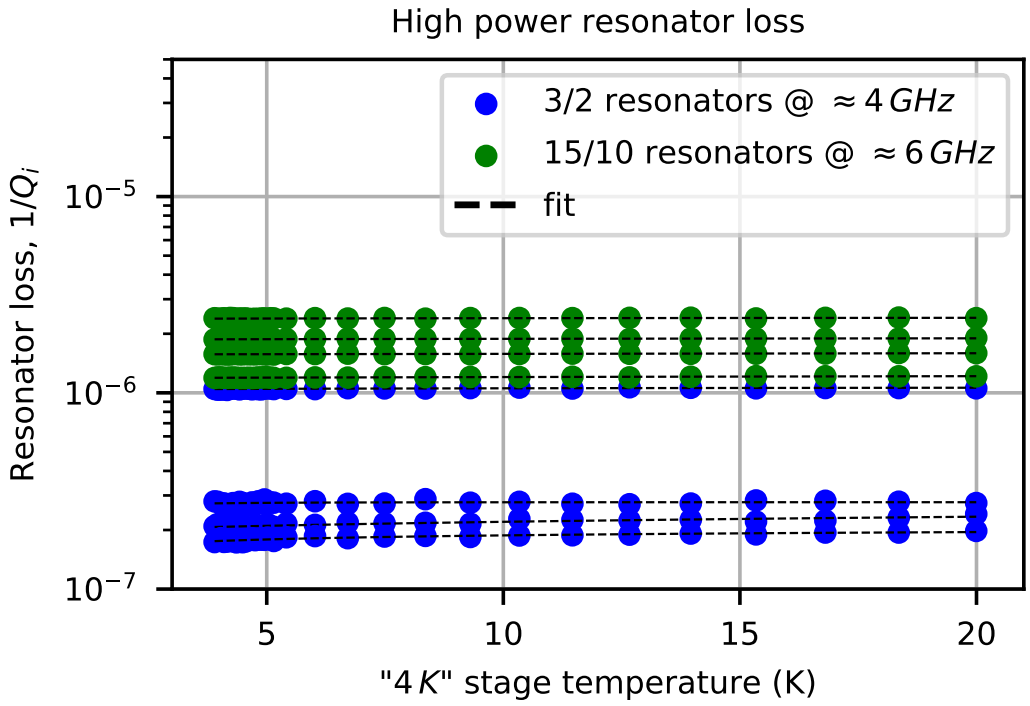


Figure 3.12: Testing the infrared shielding provided by our cryostat and the new PCB based packaging. We measure the high power ($< n > \approx 10^6$) internal quality factor, Q_i of eight superconducting coplanar waveguide resonators (four 3 μm center trace-2 μm gap resonators with frequencies near 4 GHz and four 15μm-10μm resonators with frequencies near 6 GHz) as a function of the temperature of the next-coolest cryostat stage. We find that the resonator loss, $1/Q_i$, is independent of the cryostat temperautre indicating that the infrared shielding properties of the cryostat infrastructure and package are sufficient to support loss as small as $1.8 \cdot 10^{-7}$.

bound to improved resonator fabrication techniques rather than any improvement in the IR-shielding due to the change in packaging. From this result we conclude that the light tight shield and IR signal-line filters we have built into our cryostat, in addition to either the machined aluminum or PCB package, sufficiently shields our superconducting devices from stray high-temperature photons.

3.3.4 Magnetic field shielding

This section closely follows reference [64] where magnetic field cools of coplanar waveguide resonators were used to characterize vortex-related loss.

High ambient magnetic fields may induce loss in superconducting qubits. If a thin-film superconductor is cooled through its critical temperature in an ambient field $\vec{B}_{cool}$, perpendicular to the film's surface, it is energetically favorable for the superconductor to trap flux in vortices if $\vec{B}_{cool} \geq \Phi_0/W^2$ where W is length of the side of the largest square that may be drawn within an area of the thin film superconductor [65]. Since the superconducting order parameter must vanish at the vortex core [66], the normal core produces dissipation in response to currents flowing past the core [67].

Here, we measure the loss ($1/Q_i$) of coplanar waveguide resonators to detect the presence of these normal-core vortices. By measuring this loss as a function of an applied field, we can determine the ambient magnetic field strength perpendicular to the chip. In Figure 3.13 we measure the loss of CPW resonators as a function of $\vec{B}_{cool}$. In Figure 3.13a we show a typical measurement for large magnetic fields. At low fields, the additional loss

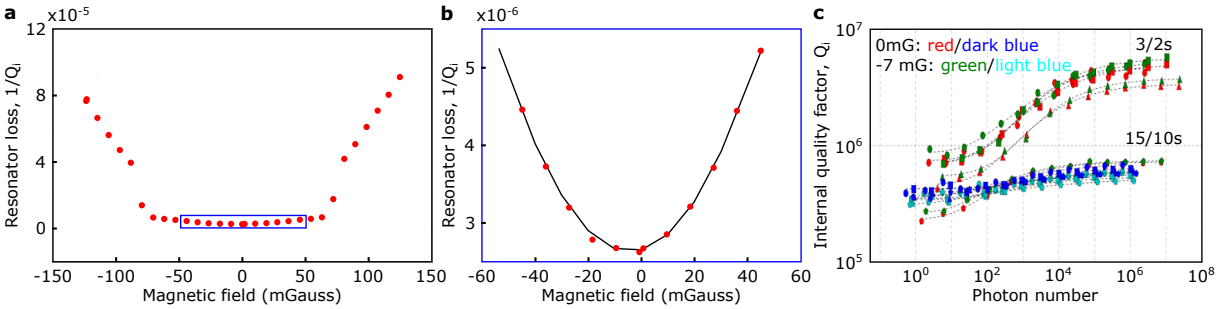


Figure 3.13: Field cooling of resonators to confirm that the new packages are non-magnetic.

is fairly small as vortices form only in the ground plane. Then, when the characteristic vortex spacing reaches the width of the resonator center trace (3 or 15 μm), the loss increases much more rapidly—applying these larger fields is a good check experiment. In Figure 3.13b, we zoom in on the loss at small fileds where we are able to fit a parabola to the loss to identify the ambient field as the x-offset of the parabola. We repeated this measurement for eight resonators in the machined aluminum box, and the PCB box with a brass, copper, or aluminum shell. The results are summarized in Table 3.1. In all cases, with either a normal (brass or copper) shield or a superconducting aluminum shield, the ambient field was less than 10 mGauss which; small enough to have a negligible effect on the quality factor of our resonators. To confirm that this effect is negligible, we perform a power-sweep measurement of the internal quality factor of all eight resonators with and without applying a compensating field of 7 mGauss to zero $\vec{B}_{cool}$ in Figure 3.13c. In each case the change in the internal quality factor with and without the corrective field is much less than the initial spread in internal quality factors between each type (geometry) or resonator.

Table 3.1: $\vec{B}_{cool}$ measured by minimizing the loss of coplanar waveguide resonators as a function of the applied field. Each chip has 8 resonators, four 3 μm center trace-3 μm gap resonators near 3.9 GHz, and four 15 μm-10 μm resonators near 6.1 GHz. Here we report the applied field offset that minimized the loss for each bank of resonators.

Resonator geometry (μm)	Avg freq (GHz)	Machined Al (mGauss)	Al shell (mGauss)	Cu shell (mGauss)	brass shell (mGauss)
3-2	3.9	-1.5 ± 0.1	-5.2 ± 0.1	-7.9 ± 0.2	-6.7 ± 0.3
15-10	6.1	-1.4 ± 0.4	-4.5 ± 0.3	-7.2 ± 0.3	-7.2 ± 0.2

3.3.5 Flux bias settling

The final check experiment for the PCB-based package was to measure the flux bias settling time used for our qubit and tunable coupler control. Using the same five qubit test devices as in the previous section, we performed qubit detuning spectroscopy to characterize the settling response. In this experiment, we apply a flux bias pulse to detune the qubit, wait a variable amount of time, drive a variable frequency microwave spectroscopy pulse to excite the qubit, and then detune back to the qubit idle frequency to read out the state of the qubit. This allows us to extract the qubit frequency as a function of time following the start of the detuning pulse. A typical response using a gold plated copper PCB in an aluminum shell is plotted in Figure 3.14 where the fit line corresponds to an exponential decay settling with a time constant of over 40 μs! Given that typical gate lengths are on the order of 10-40 ns, this is an unacceptably long settling time. We repeated this test for gold plated copper printed circuit boards in copper, brass, and aluminum shells, but the results, which are summarized in Table 3.2, were similar. Ultimately this result led to the development of the tunable resonator metrology tool

Table 3.2: Here we summarize the settling response of gold plated copper printed circuit boards with either a brass, copper, or aluminum shell enclosing the PCB and chip. Missing quibits had wiring failures and were not measurable.

Shell	qubit 1	qubit 2	qubit 3
Brass	86 μs	—	112 μs
Copper	87 μs	118 μs	115 μs
Aluminum	43 μs	128 μs	—

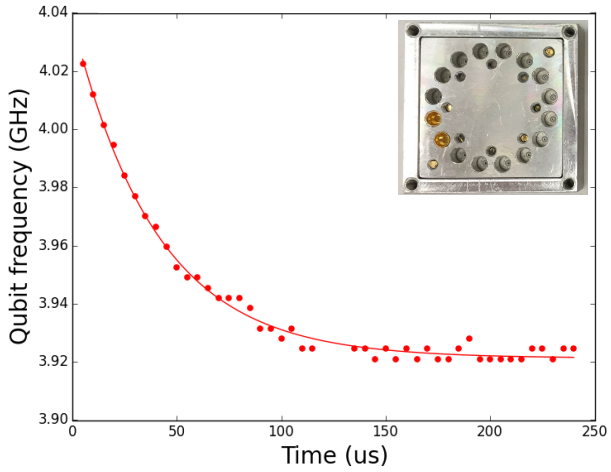


Figure 3.14: Qubit detuning spectroscopy using the gold plated copper PCB in an aluminum shell.

that is the subject of Chapter 4 to better sample our flux control waveforms on-chip.

A comparison of the responsive machined aluminum package and the printed circuit board packages seemed to indicate that the long settling times arose as a consequence of having return currents in normal metal versus a superconducting metal. We later confirmed that aluminum plated copper circuit boards recover the response of our machined aluminum package offering a near-term scalable packaging solution. These tests are described in more detail in Chapter 4.

3.3.6 Wire bonding

Once we have packaged our qubits, we must make many electrical connections between our control electronics, which have coaxial transmission line output, and our qubit chip, which has coplanar waveguide transmission lines leading from the edge of the chip to the qubits. A tried and true method for making these connections is to use edge mount PCB connectors that covert coaxial lines to planar geometries and then to use wire bonds to connect PCB traces to coplanar transmission lines on a crystaline substrate.

Wedge wire bonding may be performed with fine wire ranging 12.5 μm to 75 μm in diameter or heavy wire as thick as 600 μm. The wire is threaded through through a wire bonding tip so that wire may be routed under the wedge tip. To place a bond, the wedge presses the wire onto the surface to be bonded and uses a pizeoelectric actuator to vibrate the wedge at a frequency in the 50-200 kHz frequency range. The frequency, amplitude, and duration of the vibration may be adjusted in addition to the touch-down force the wedge applies to the wire to ensure the formation of a strong weld between the wire and the surface. The tool then moves the bond head to the second bond location and repeats the bonding process before tearing off the wire to complete the connection electrical connection.

Typically wire bonding uses gold, silver, copper, or aluminum wire. For the purposes of most of the integrated cricuit community those metals are listed in order of decreasing desirability and cost. Gold and silver have higher conductivity and do not corrode making them more desirable for typical electronic applications. Somewhat amusingly, aluminum

wire bonding was developed as a low-cost alternative to sliver of gold wire bonding. The surface oxide and higher resistivity of aluminum make it less desireable for non-superconducting circuits, but in our case, it is advantageous. The sonication used to form the weld is sufficient to break the surface aluminum oxide and make a galvanic connection without forming a tunnel barrier.

Industrial level wire bonding tools have been designed to operate 24 hours a day 7 days a week can place as many as six bonds per second, requiring only the replacement of consumables, like relatively inexpensive wedge tips, at intervals exceeding kilometers of wire. There are a variety of wedge tip options that vary the angle as which the wire approaches the tip depending on the necessary clearances around the wire bonds. You can trade the speed of a 45° wedge for the utility of a 90° deep-access wedge tip which, for some tools, may require a different bond head as well. When wire bonds are used in many integrated circuit contexts, the wires are embedded in epoxy for reliability. In the context of superconducting qubits, we do not want to embed the bonds in epoxy because it may be lossy, or actually shear off wire bonds due to the relative thermal expansion coefficients. Fortunately, with proper welding parameters, we have found Al-Al wire bonding to be quite robust to thermal cycles down to 10 mk (no problems after 10 or more thermal cycles of the same device), though care must be taken to avoid mechanical deformation of the wire bonding package during installation to the cryostat.

The bond footprint is determined by both the shape of the wire bonding wedge and the diameter of the wire. Typically the wedge foot deforms a length of wire about 75 μm

long by about 50%. For the smallest commercially used aluminum wire, (17.78μm or $7/10^{ths}\,mil$ diameter), the bond footprint is roughly rectangular and about 50μm by 75 μm. These dimensions fall somewhere between the minimum line widths of typical PCB processes (2 mil or 50 μm) and standard thin film optical lithography which are about 1 μm, but this is a reasonable level of fan-out.

Aluminum wire bonding is a reliable and established industrial packaging solution for integrated circuits, and, at least mechanically speaking, one of the less-scary aspects of scaling existing superconducting qubit systems from tens of qubits to thousands of qubits. Having pursued a variety of *ad hoc* wire bonding solutions throughout my time as a graduate student, I have confidence that with proper tooling, it would not be unreasonable to place 10,000-100,000 wire bonds on a single die. With existing tools able to place up to 6 bonds per second, even 100,000 bonds would take just a few hours which is not unreasonable. Standard wire spools are 1000 ft (300 m) in length, which is probably enough for about 600k wire bonds. My guess is that, if necessary, wire bonding is probably a viable solution all the way to somewhere between 100,000 and 1,000,000 wire bonds per device. Beyond this limit, we should probably look towards more scalable lithographically patterned ball grid arrays or bump bonds (as in Chapter 5. At some point the serial nature of wire bonding will limit the production of quantum processors.

On the several pages that follow, I have included images captured with my phone camera through inspection microscopes offering a sampling of various wire bonding jobs I have completed. The precision and reliability of wire bonds is truly impressive, and,

should it come down to it, it is not unreasonable to consider multi-row wire bonding solutions. It is also worth noting that the basic wire bonding foot print is roughly 10-20x smaller than that of a standard coplanar transmon capacitor—and 10-20 lines may be enough to connect the necessary signal lines to each qubit with a nice number of caging ground wires. Combining this with superconducting through silicon vias a few microns in diameter [68], wire bonding could be a viable packing solution for superconducting qubits well into the future.

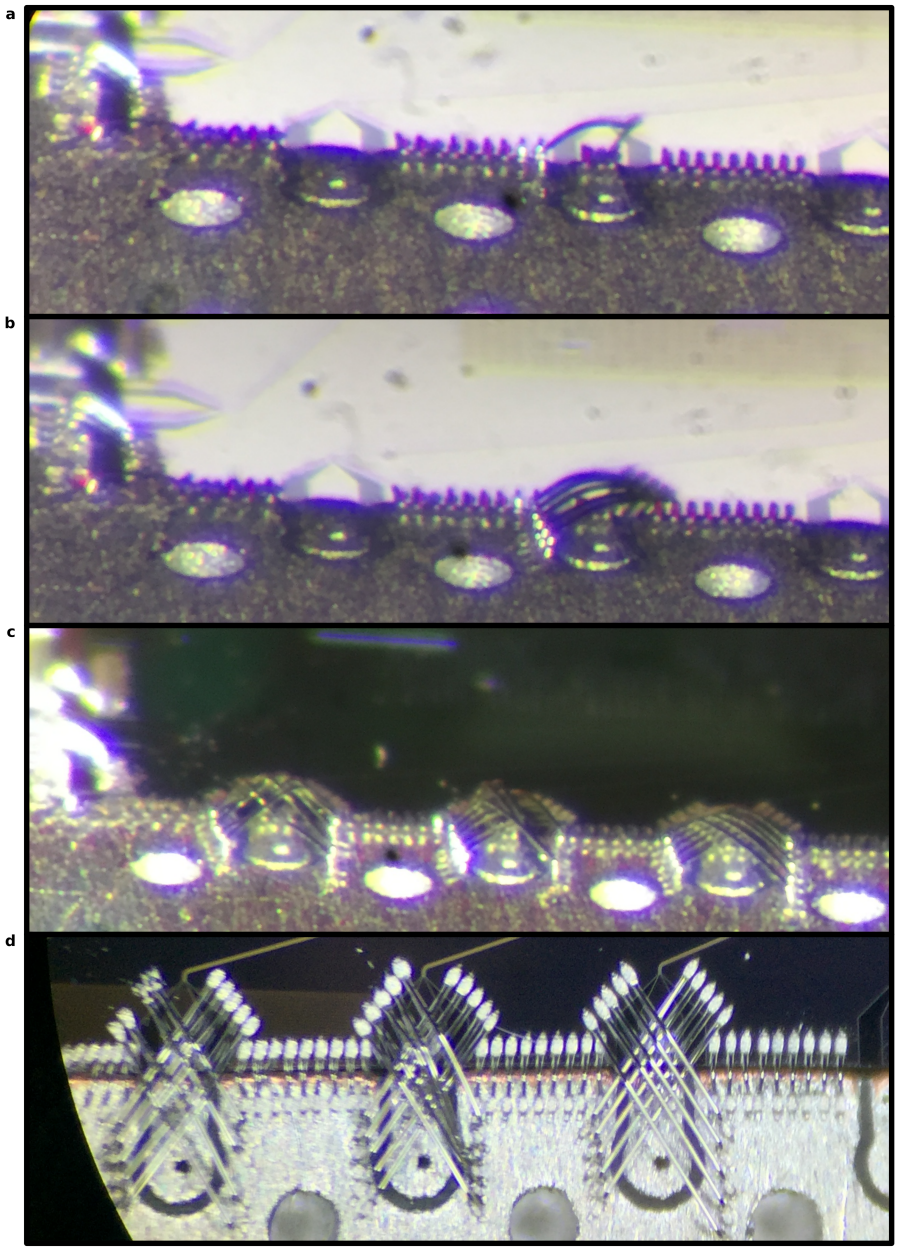


Figure 3.15: Demonstration using wire bonds to 'cage' signal lines in ground plane bonds. Should it become necessary to approximate a shielded wire bonding geometry, wires can be placed with positioning and shaping that is reproducible down the the level of a few microns.

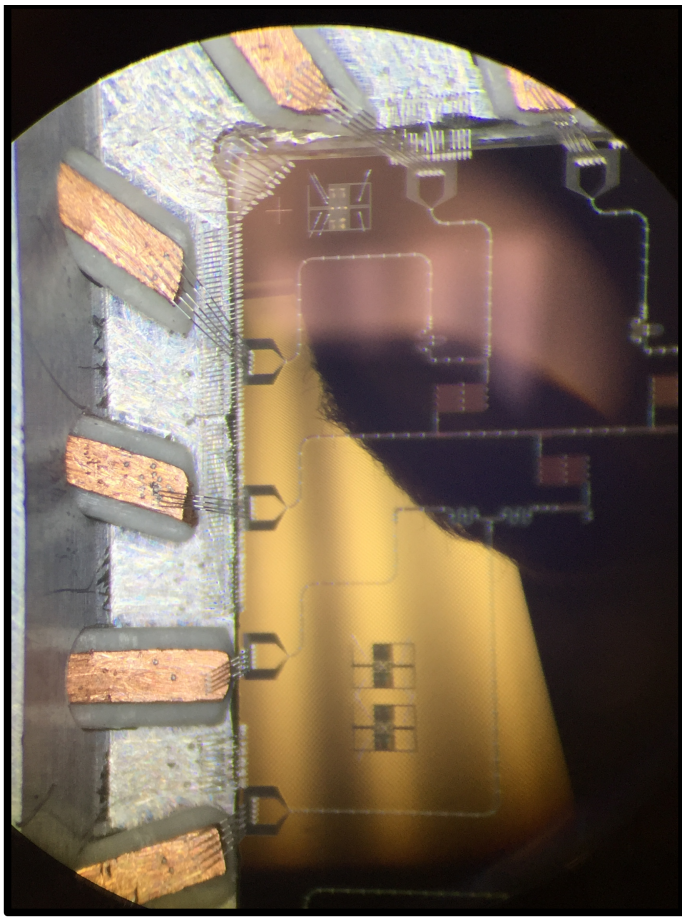


Figure 3.16: Demonstration using nice new wire bonder in an old-style package. With a nice wire bonder we were able to place grounding wires continuously in front of the copper PCB insert that had been cut too short.

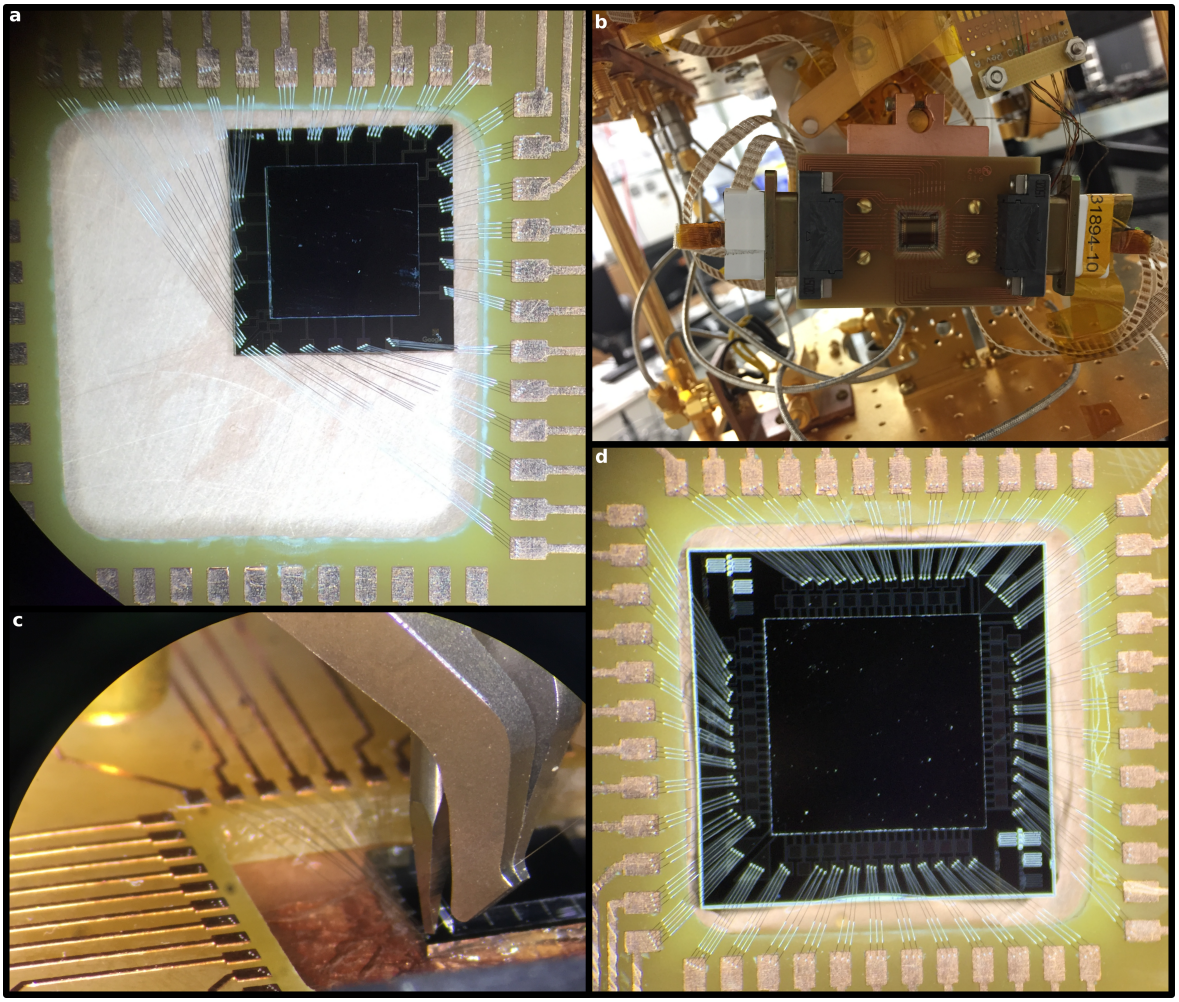


Figure 3.17: Wire bonding for DC testing. Here we wire bonded chip that was too small for out sample mount which necessitated very long wire bonds.

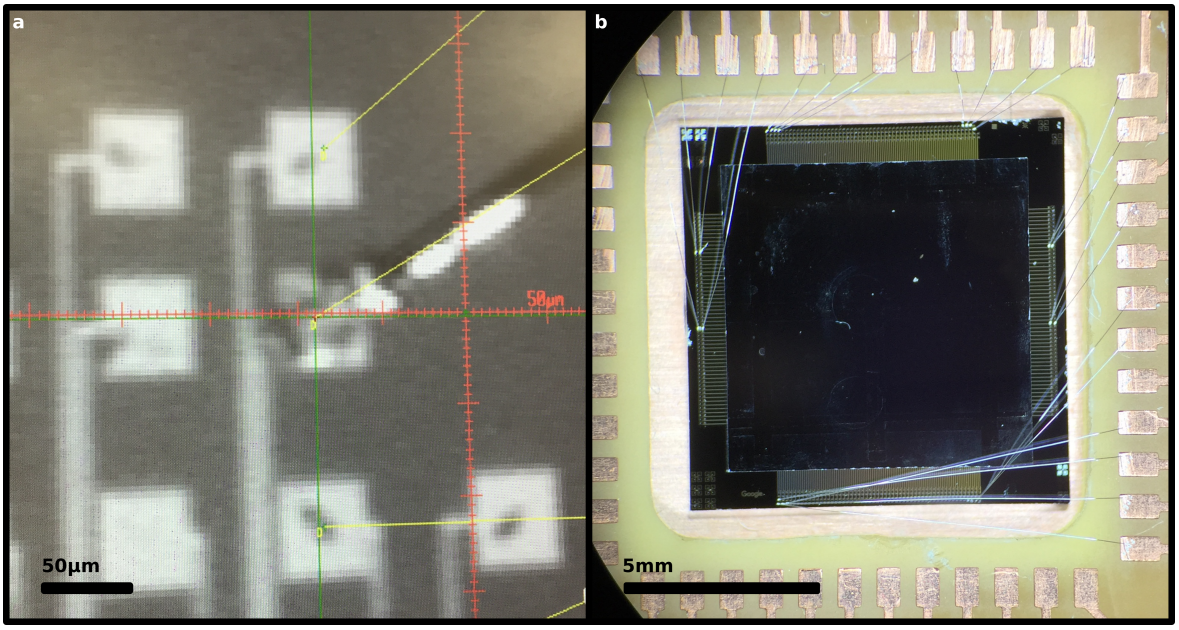


Figure 3.18: Wire bonding with 50μm pads. This demonstrates that it is possible to place wires on relatively small pads. The wire bonder also has a feature to allow the bonds to be rotated so it would be possible to place a bond on a smaller rectangular bad regardless of the orientation of the wire bonding direction.

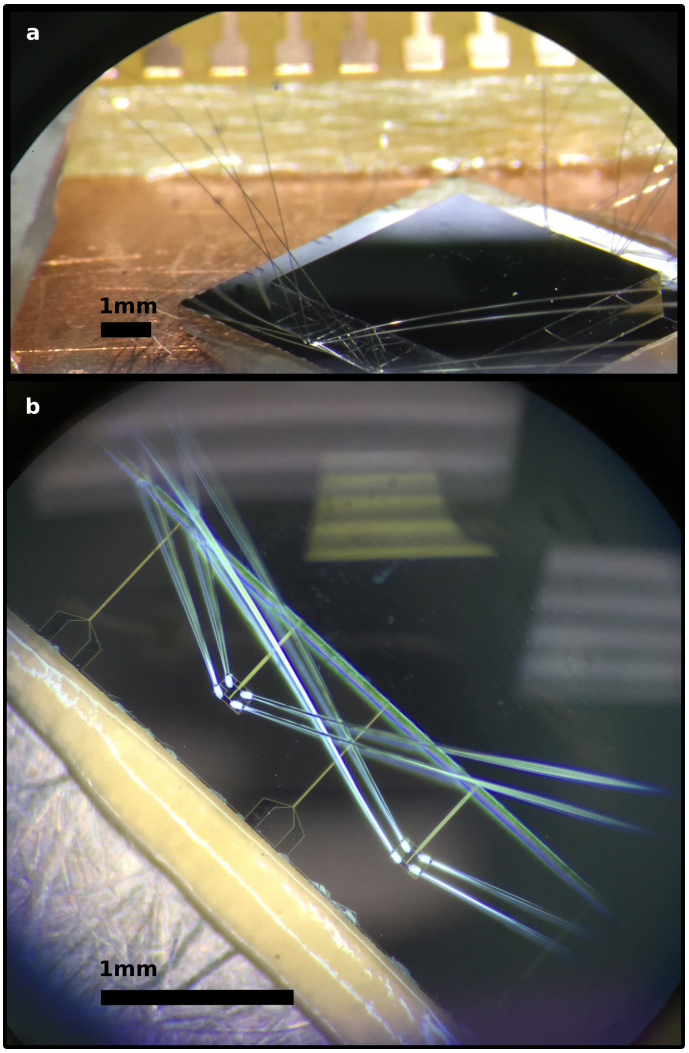


Figure 3.19: Wire bonding with 50μm pads with a 6 mm chip for reference.

Chapter 4

High speed on-chip flux measurement

We develop a high speed on-chip flux measurement using a capacitively shunted SQUID as an embedded cryogenic transducer and apply this technique to the qualification of a near-term scalable printed circuit board (PCB) package for frequency tunable superconducting qubits. The transducer is a flux tunable LC resonator where applied flux changes the resonant frequency. We apply a microwave tone to probe this frequency and use a time-domain homodyne measurement to extract the reflected phase as a function of flux applied to the SQUID. The transducer response bandwidth is 2.6 GHz with a maximum gain of $1200^\circ/\Phi_0$ allowing us to study the settling amplitude to better than 0.1%. We use this technique to characterize on-chip bias line routing and a variety of PCB based packages and demonstrate that step response settling can vary by orders of magnitude in both

settling time and amplitude depending on if normal or superconducting materials are used. By plating copper PCBs in aluminum we measure a step response consistent with the packaging used for existing high-fidelity qubits.

4.1 Introduction

Superconducting qubits have risen to a position of prominence in the field of quantum computation as they have demonstrated a system-level performance near the fault-tolerant threshold [**?**]. The demonstration in ref. [**?**] was an important milestone, and further performance gains above and beyond the error correction threshold will exponentially lower the required resources [69]. While qubit decoherence remains the largest source of system infidelity, control error in two-qubit gates is a close second [47]. Frequency tunable superconducting qubits, which use a superconducting quantum interference device (SQUID) loop as a magnetic-flux-tunable non-linear inductance, perform entangling operations by bringing qubit frequencies near resonance with shaped flux pulses. The highest fidelity two-qubit entangling gates have used a fast-adiabatic approach which is theoretically capable of realizing a CZ gate with error less than 10^{-4} [49]. Experimental implementations of this gate have thus far achieved error rates of $5.4 \cdot 10^{-3}$ per gate, where two qubit randomized benchmarking and purity measurements in that system indicate qubit decoherence accounts for about 60% of the gate error while control error is estimated to account for most of the remaining 40% [47]. Characterizing the control error is therefore a priority for quantum computation with superconducting

qubits.

Flux control pulses for qubits are often calibrated *in situ* to account for line-specific impedance varaitions due to connectors, wire bonds, and other imperfections that alter these flux pulses before they reach the qubits [70, 71]. However, the time resolution of these measurements is limited by the width of microwave spectroscopy pulses used to probe the qubit frequency, which are typically 5-20 ns. In this paper we demonstrate a time-domain, qubit-free, GHz bandwidth measurement for flux pulses using a capacitively shunted SQUID as a flux-tunable resonator to realize a cryogenic magnetic flux to microwave phase transducer. In this transducer, flux pulses shift the resonator frequency which in turn modulates the phase of a microwave tone reflecting off the resonator. Similar circuits have been operated as quantum limited amplifiers optimized for flux sensitivity at the cost of bandwidth [72]; we have instead optimized for a large bandwidth at the expense of flux sensitivity by operating without amplification and using an impedance transformer to reduce the resonator's quality factor. In contrast to experiments that measure flux settling with qubits, this measurement may be performed with just a single microwave source, a two channel oscilloscope, and an arbitrary waveform generator (AWG). We achieve a DC-2.6 GHz flux waveform measurement bandwidth which may be used to measure pulse settling to better than 0.1% of a step with an absolute flux sensitivity of $0.2\,\mathrm{m}\Phi_0$.

In section 2 of this paper, we provide a simple analytic model of the tunable resonator device. In section 3, we describe the measurement setup and DC calibration of the

tunable resonator as a flux-to-phase transducer. In section 4, we verify the performance of the tunable resonator by measuring the settling response in a package we can compare with previous qubit results. Finally, in section 5, we apply this measurement technique to characterize a variety of device packages yielding a near-term scalable PCB-based package and gaining some insight into how to further improve the performance of our qubit flux bias lines. Notably, in A.1 we bound the error introduced by microwave reflections and offer an alternate measurement technique that uses the tunable resonator to characterize microwave reflections in a cryogenic environment with a sensitivity better than -30 dB.

4.2 Theory of operation

The circuit used in this experiment consists of an on-chip impedance transformer terminated by a lumped element capacitor in parallel with a two-junction SQUID, as shown in figure 4.1a. This circuit and its fabrication are the same as the Impedance Matched Parametric Amplifier (IMPA) in reference [73]. We choose a junction critical current of $2\,\mu$A ($\mathrm{I_c} = 2 \cdot 2\,\mu\mathrm{A}$) with a 4 pF parallel plate capacitor giving a maximum resonance frequency of 8.7 GHz and making the device compatible with the same microwave components used for typical 4-8 GHz qubit circuits. The impedance transformer maximizes the flux measurement bandwidth by reducing the resonator's quality factor. Since $\mathrm{Q} = \omega_\mathrm{r}/\Delta\omega_\mathrm{r} = \mathrm{Z_0}\omega_\mathrm{r}\mathrm{C}$, the resonator bandwidth $\Delta\omega_r = 1/(Z_0 C)$ is independent of the applied flux. With $\mathrm{Z_0} = 15\,\Omega$ the tunable resonator is able to operate over a 2.6 GHz bandwidth which is more than sufficient to characterize flux control pulses that typically

have a bandwidth of several hundred MHz.

The flux bias line, highlighted in red in figure 4.1a, has the same geometry as used in qubit circuits and shares a mutual inductance of 1-2 pH with the SQUID loop, highlighted blue, which converts current in the bias line to a magnetic flux through the SQUID loop. This flux Φ changes the inductance of the tunable resonator as given by the Josephson inductance,

$$\mathrm{L_j}(\Phi) = \frac{\Phi_0}{2\pi \mathrm{I_c} |\cos(\frac{\pi\Phi}{\Phi_0})|} = \frac{\Phi_0}{2\pi \mathrm{I_c^*}(\Phi)} \tag{4.1}$$

where $\mathrm{I_c}$ is twice the critical current of each of the two, nominally identical, Josephson junctions that compose the SQUID, $\Phi_0 = h/2e$ is the magnetic flux quantum, and $\mathrm{I_c^*}(\Phi)$ is the SQUID's flux-dependent effective critical current.

The SQUID's effective critical current goes to zero as the applied flux approaches $\frac{1}{2}\,\Phi_0$. As such, there will always be a flux where the current from the microwave tone will break the small amplitude limit invalidating equation 4.1—this sets the upper limit for the tunable resonator's flux operating range. Simple analysis of the circuit in figure 4.1b can be used to compute the frequency dependent resonator impedance, as a function of the flux through the SQUID loop.

$$\mathrm{Z_r}(\omega, \Phi) = (\mathrm{L_j}(\Phi) || \mathrm{C_s}) = \frac{i\omega \mathrm{L_j}(\Phi)}{1 - \omega^2 \mathrm{L_j}(\Phi)\mathrm{C}} = \frac{i\omega \mathrm{L_j}(\Phi)}{1 - \omega^2/\omega_\mathrm{r}(\Phi)^2} \tag{4.2}$$

From here it is straightforward to compute the complex coefficient of reflection for the tunable resonator, $\Gamma_\mathrm{r} = (\mathrm{Z_r} - \mathrm{Z_0})/(\mathrm{Z_r} + \mathrm{Z_0})$. Since $\mathrm{Z_r}$ is purely imaginary the magnitude of Γ_r is 1 and all of the flux dependence is contained in the reflection angle $\angle\Gamma_\mathrm{r}$ given by

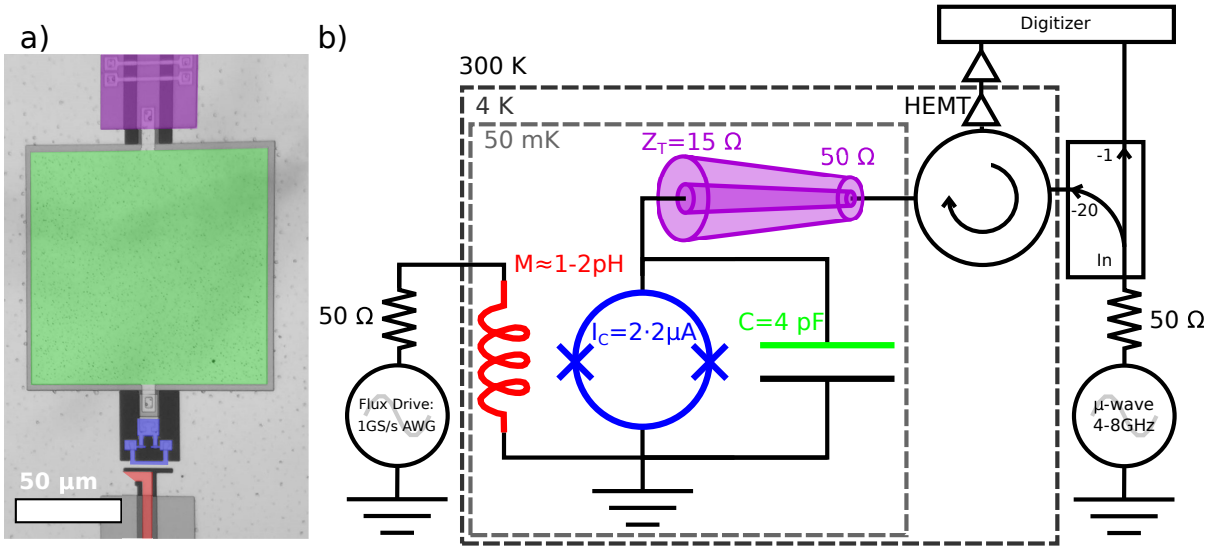


Figure 4.1: Device and circuit diagram. a) Micrograph image of the tunable resonator which consists of a 2-junction SQUID (highlighted blue) shunted by a parallel plate capacitor (green). Microwaves are reflected off of the tunable resonator through a galvanicly coupled 50 Ω to 15 Ω impedance transformer (magenta). Magnetic flux is applied to the SQUID through a mutual inductance with the adjacent flux bias line (red). b) Schematic diagram of the turnable resonator and measurement system. Flux pulses are generated by a FPGA-based arbitrary waveform generator on the left. These flux pulses change the frequency of the tunable resonator which modulates the phase of the reflected probe tone. The probe tone is supplied by a variable frequency microwave source; it reflects off of the tunable resonator is directed via a cryogenic microwave circulator through amplifier stages and towards a digitizer.

$$\begin{aligned} \angle\Gamma_r &= \arctan\left(\frac{\mathrm{Im}(\Gamma_r)}{\mathrm{Re}(\Gamma_r)}\right) = \arctan\left(\frac{-2\mathrm{Im}(Z_r)Z_0}{Z_r^2+Z_0^2}\right) \\ &= \arctan\left(\frac{2Z_0}{\omega L_j(\Phi,I_c)}\frac{((1-\omega^2 L_j(\Phi,I_c)C)^2)^2-Z_0^2)}{(1-\omega^2 L_j(\Phi,I_c)C)}\right). \end{aligned} \tag{4.3}$$

Where ω is the frequency of the incoming probe tone, and Z_0, I_c, and C are our circuit design parameters. Equation 4.3 summarizes the operation of the tunable resonator. If a single frequency probe tone ω is reflected off of the tunable resonator, then it functions

as a flux-to-phase transducer, translating the flux waveform applied to the SQUID to the microwave frequency domain as a phase modulation of the reflected probe tone.

The transducer's inherent frequency up-conversion enhances its functionality as an on-chip flux transducer. For the purpose of calibrating a qubit flux bias line, we would like to measure the transfer function of the cabling, connectors, wirebonds, *etc.* between the generator and a cryogenic SQUID. This is complicated by the fact that both our generation and digitization instruments are at room temperature; we can only directly measure the round-trip transfer function. However, since the SQUID upconverts the signal bandwidth, this means we capture the baseband transfer function from room temperature to the SQUID and the microwave frequency transfer function on the return path. Since dispersive effects are larger for our incoming baseband flux waveform than the outgoing phase modulated microwave signal, we are more sensitive to these effects in the desired portion of the signal path.

Finally, since the transducer's flux-to-phase response is nonlinear, we are able to operate at different flux biases to take advantage of the transducer's gain and characterize our system. By operating at a flux bias where the transducer is is maximally sensitive to flux (e.g. when the resonator and probe frequencies are equal), we are able to characterize irregularities in the flux drive line with a high precision. Alternatively, if we operate where the transducer is minimally sensitive to flux (e.g. with the resonator detuned from the probe frequency), then the transducer may be used to characterize systematic phase errors due to reflections in the microwave cabling. For a discussion of these errors, see

A.1.

4.3 Measurement setup and transducer calibration

A simplified measurement schematic is shown in figure 4.1b. The tunable resonator is cooled to 50 mK in an adiabatic demagnetization refrigerator (ADR). Flux waveforms are produced at room temperature by a FPGA controlled 1 GS/s DAC that together form an arbitrary waveform generator (AWG). We place a 220 MHz Gaussian low pass filter on the output of the AWG to prevent ringing, and stagger attenuators from room temperature down to the 4 K stage of the ADR to reduce noise and reflections while bringing the full scale AWG output level down to 1.5-2.0 Φ_0 on-chip, thereby using the AWG's full dynamic range.

A variable frequency microwave source (Hittite HMC-T2220) provides the probe tone, which we split at room temperature with a directional coupler; the transmitted port is used as the local oscillator reference for our phase measurement and the coupled (-20 dB) port is sent into the cryostat to reflect off of the tunable resonator. After the probe tone reflects off of the tunable resonator, a cryogenic microwave circulator directs the reflection through a HEMT amplifier at 4 K and towards the room temperature amplification and digitizer. The probe frequency may be chosen arbitrarily, but practically, it should be in the pass band of the HEMT amplifier (4-8 GHz). It is best if the tunable resonator frequency can sweep through the probe frequency to maximize the phase shift and result-

ing flux sensitivity[1]. With a maximum resonator frequency of 8.7 GHz, we found using a probe frequency of 6.4 GHz to be an optimal operating point.

We measure the microwave signal with an effective homodyne receiver that is phase-locked to the AWG and microwave source through a shared 10 MHz clock. To reconstruct the baseband phase waveform, we digitize both the local oscillator reference and the reflection with a 40 GS/s oscilloscope (Keysight DSA90804A) and digitally demodulate the signal. Alternatively, in section 5 we use an IQ mixer to demodulate the reflection so that we may digitize at baseband frequencies and accommodate the oscilloscope's single-trace memory limit while sampling at longer times. In A.2 we compare digital and hardware demodulation of this tunable resonator measurement to qubit-based settling measurements. Hardware demodulation is quite precise (just a little less accurate), allows for sparse sampling at long times, and requires a significantly less expensive oscilloscope since only several hundred MHz (rather than 4-8 GHz) of bandwidth are needed. Digital demodulation offer a superior performance with respect to flatness and settling due to the oscilloscope's superior AC *vs.* DC response.

As shown in figure 4.2a, we calibrate the tunable resonator as a flux-to-phase transducer by applying DC fluxes to the SQUID and measuring the reflected phase $\angle\Gamma$. For a probe frequency of 6.4 GHz, we use equation 4.3 to fit $\angle\Gamma$ as a function of the applied flux and find fit parameters within 10% of the designed parameters for junction critical

[1]Since neither our AWG or microwave source accept external triggers, we are also required to chose a flux waveform repetition period that is a multiple of the probe tone. This way the AWG may be used to trigger the oscilloscope and the phase-locked probe tone will remain synchronized throughout the experiment.

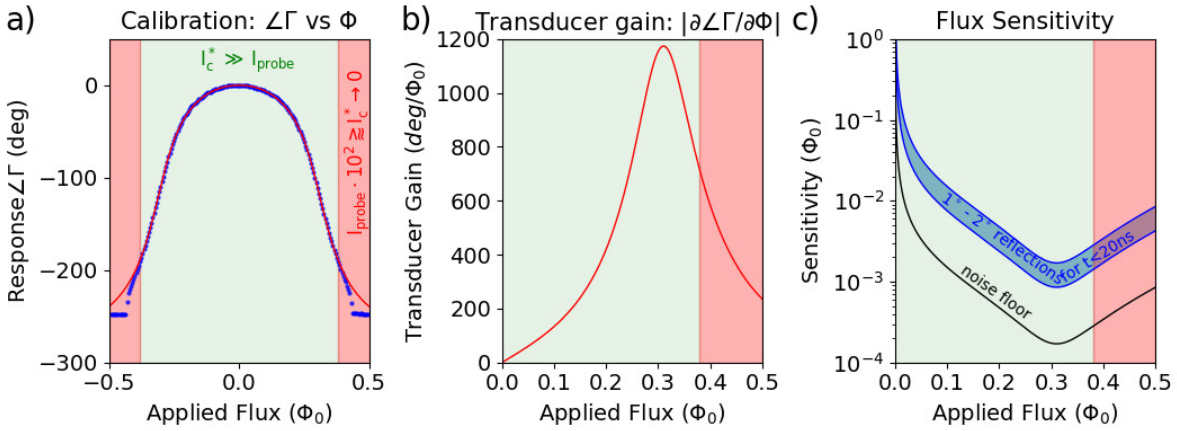


Figure 4.2: Transducer calibration, gain, and flux sensitivity. a) Reflection angle ($\angle\Gamma$) vs. applied DC flux; this curve serves as the transducer's flux-to-phase calibration curve. For a probe frequency of 6.4 GHz, we fit this data using equation 4.3, and find good agreement with the designed parameters for SQUID critical current $I_c = 1.8\,\mu$A (2 μA designed), transformer impedance, $Z_{0,fit} = 14.8\,\Omega$ (15 Ω) and shunt capacitance, $C_{fit} = 3.8$ pF (4 pF). As the flux approaches $\frac{1}{2}\Phi_0$ the effective critical current of the SQUID is reduced to zero; the fit diverges from the data at $|0.38\,\Phi_0|$ because the probe tone drive power is no longer in the linear limit relative to the SQUID critical current. b) Transducer gain as a function of applied flux showing a maximum at $0.31\,\Phi_0$; this plot is the magnitude of the derivative of the fit in a). c) Flux sensitivity vs. applied flux computed by dividing our phase measurement noise by the transducer gain in b). Our measurement sensitivity in terms of flux is limited by our phase measurement noise of about 0.25°. For short times (less than twice the propagation time from the resonator to the digitizer, approximately 20 ns in our case), the blue region represents an additional systematic error of 1°-2° due to transient reflection of the probe tone discussed in A.1.

current, $I_c = 1.8\,\mu$A (designed = 2 μA), transformer impedance, $Z_0 = 14.8\,\Omega$ (15Ω), and capacitance, C = 3.8 pF (4 pF). While a phase shift of 360° is typically expected when sweeping through a resonance, the phase shift we measure is reduced due to the finite (and large) resonator bandwidth relative to its flux-tunable frequency range. In order to keep the transducer's flux-to-phase relationship single-valued, we limit the operating range to a single $\frac{1}{2}\Phi_0$ bias range (e.g. $0 - \frac{1}{2}\Phi_0$), with the actual upper operating limit set by the point where the probe current begins driving the SQUID non-linearly. In figure

4.2a, this non-linear behavior may be observed as the point where the fit deviates from the measured phase at a flux of $|0.38\,\Phi_0|$. In this case, the probe power was chosen to be as large as possible while still allowing for operation at $0.31\,\Phi_0$ where the transducer gain is at a maximum as shown in figure 4.2b.

Since the transducer converts flux-to-phase, our phase measurement uncertainty may be converted to a flux uncertainty by dividing by the transducer gain in 4.2b. Immediately after a large change in flux, reflections in the microwave side of the measurement dominate our phase measurement error as described in A.1. We observe -30 dB to -36 dB microwave reflections, typical of SMA connections and other microwave components, that result in time dependent phase errors of 1° to 2°. Since second-reflections from the same components generate -60 dB reflections (0.05° errors), their effect dies off in twice the signal propagation time from the tunable resonator to the digitizer (about 20 ns). Since the transducer converts flux-to-phase, the noise in our phase measurement may be scaled to a flux sensitivity by dividing our phase noise by the transducer gain in 4.2b. When digitally demodulating the settling response, the phase measurement noise is limited by the $20\,\mathrm{ps}_{\mathrm{pk-pk}}$ synchronization jitter between our microwave source and AWG. By averaging 50k oscilloscope traces we are able to reduce this jitter to 0.25° for a 6.4 GHz probe frequency. Further averaging would reduce this noise, but time-averaging samples often becomes more economical when looking at long times. In figure 4.2c, we plot the flux sensitivity due to a 0.25° phase measurement error which is better than $0.2\,\mathrm{m}\Phi_0$ for a small range of applied fluxes around $0.31\,\Phi_0$. We also plot the systematic error in flux

due to 1° to 2° microwave reflections at short times.

4.4 Step response

We measure the settling response of the flux bias line, by using the AWG to generate a rising edge that has been digitally low pass filtered with a 220 MHz Gaussian profile to prevent ringing. We choose to end the step at a flux of $0.31\,\Phi_0$, where the transducer is most sensitive to flux, and start the step at $0.08\,\Phi_0$ to maximize the amplitude of the step and increase the fractional sensitivity of the measurement. In figure 4.3 we plot the measured flux waveform settling, A(t) (normalized to a step amplitude of 1), and fit it heuristically with three exponential time constants:

$$\mathrm{A(t)} = 1 - (\alpha_0 \mathrm{e}^{-\mathrm{t}/\tau_0} + \alpha_1 \mathrm{e}^{-\mathrm{t}/\tau_1} + \alpha_2 \mathrm{e}^{-\mathrm{t}/\tau_2}) \tag{4.4}$$

We find that the settling response is fit well for settling amplitudes of 0.48, 0.04, and 0.01 for time constants of 0.73 ns, 7.9 ns, and 53.5 ns respectively. Previous qubit based measurements with the same model AWG found the settling to be well fit by two time constants of 5 ns and 100 ns attributed to reflections and an L/R time constant associated with a bias tee [**?**]. In the present measurement, we have removed the bias tee, so the longest time constant in our system has been reduced as expected. Additionally, since we are also able to measure at much shorter times than qubit experiments allow, we are able to fit the rising edge with a time constant of 0.73 ns. Since the settling is well-fit by exponential decays, it is likely that L/R time constants, rather than skin effect losses that would result in a $1/\sqrt{t}$ dependence, are the dominant source of pulse distortion

in our system. This step response measurement's agreement with previous qubit-based results confirms that the tunable resonator technique is sufficient for characterizing and evaluating the transfer function of high performance qubit flux bias lines.

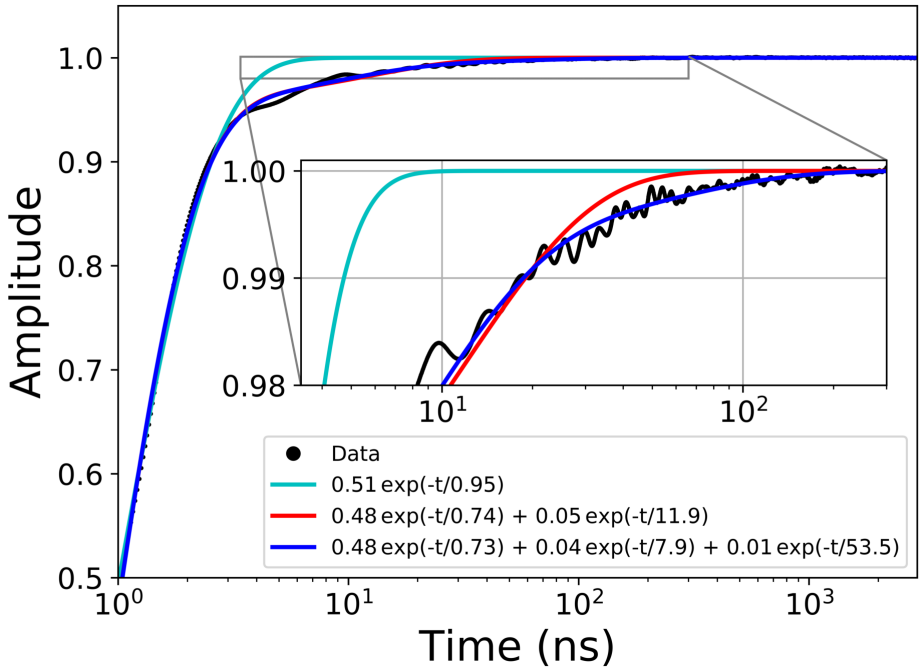


Figure 4.3: Fitting the measured step response using equation 4.4 with 1, 2, or 3 exponential decay rates. Previous measurements with the same AWG and similar cabling, packaging, and on-chip layouts found time-constants of 5 ns and 100 ns attributed to reflections and a L/R time constant from a bias tee. In this measurement, we find the response to be well fit by three time constants of 0.73 ns, 7.9 ns and 53.5 ns.

While this settling performance is imperfect, the settling amplitudes and times are sufficiently good that we may pre-distort control waveforms with relatively small correction amplitudes. It is theoretically possible to compensate for any transfer function, but larger settling amplitudes require larger compensation amplitudes which effectively reduce the signal to noise ratio of the AWG. Additionally, if the settling times (τ's) require

compensation at times longer than our desired gate times (20-40 ns) then it is difficult to design discretized, settling-compensated flux pulses that may be loaded into FPGA memory and sequenced arbitrarily to perform gate operations. For both of these reasons it is important to understand the source of these settling times.

4.5 Package design and device layout

Most qubit and resonator measurements from our group have been performed using a machined aluminum package. This package consists of two pieces of 6061 aluminum (a base and a lid) and was designed to provide a continuous ground plane and shielding from external magnetic fields. Electrical signal connections are made by drilling holes from the perimeter of the base towards an interior chip cavity and inserting narrow copper printed circuit board (PCB) strips with dimensions appropriate to maintain a 50 Ω impedance. On the exterior of the package, these PCB traces are soldered to the center pin of normal-metal panel-mount SMA connectors attached directly to the aluminum package. On the interior of the package, these PCB traces are wire bonded to the device under test (DUT) and ground plane wirebonds connect from the DUT to the bulk of the aluminum package. The key point for later is that in this package the signal tranmission lines have a normal metal (copper) center trace and a superconducting aluminum ground plane.

Despite this package's solid performance, its major disadvantage is that the physical machining necessary to incorporate additional signal lines is not scalable beyond a couple dozen qubits. In fact, package design and scaling up qubit systems is an active area of

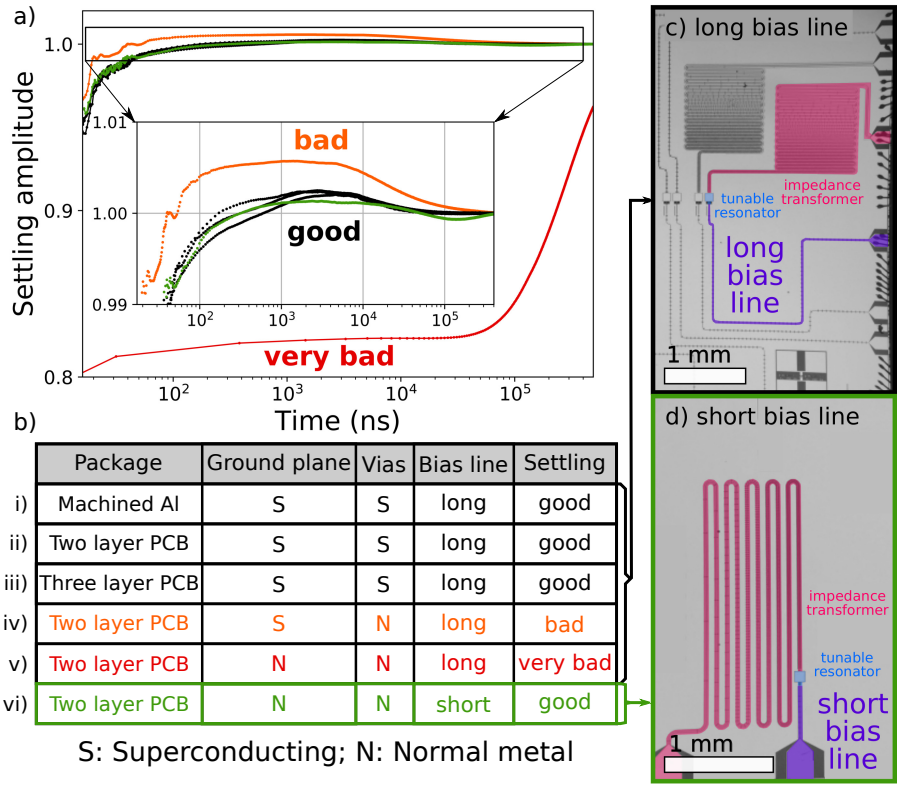
research for many superconducting qubit groups [68, 74, 75]. PCB-based packages, able to implement additional signal lines in lithography rather than with physical machining, offer a path forward, but we have found that traditional copper PCBs perform poorly with considerable settling of 10-20% occurring for hundreds of microseconds after a stepped input signal. When appreciable settling is still occurring after hundreds of microseconds, directly digitizing the microwave reflections at GHz frequencies becomes impractical due to oscilloscope memory limitations. As discussed in section 3 and A.2, we chose to use hardware demodulation to compare the performance of various packages. As a result, we are able to digitize several hundred microseconds of settling data in a single trace at the cost of some amplitude accuracy—fortunately the precision of both techniques is similar as we find relative measurements reproducible within about $2 \cdot 10^{-3}$.

In figure 4.4a we plot the measured step response of a variety of packages which can be summarized as having one grouping of good settling responses (plotted together in black/green), one response that is a little bad (orange), and one that is very bad (red). The small differences between the traces in the 'good' group are artifacts of hardware demodulation (see A.2)—the actual response of all these these packages is similar to the response in figure 4.3. This grouping of four equivalently good traces includes:

1. Our standard machined aluminum package with normal metal signal lines

2. A two layer aluminum-plated copper PCB

3. A three layer aluminum-plated copper PCB, and finally



b)

	Package	Ground plane	Vias	Bias line	Settling
i)	Machined Al	S	S	long	good
ii)	Two layer PCB	S	S	long	good
iii)	Three layer PCB	S	S	long	good
iv)	Two layer PCB	S	N	long	bad
v)	Two layer PCB	N	N	long	very bad
vi)	Two layer PCB	N	N	short	good

S: Superconducting; N: Normal metal

Figure 4.4: Measured step response of various chip packages and on-chip flux bias line routings. a) Hardware demodulated step response of several different packages with two different on-chip bias line designs—the table in b) serves as a legend. b) Summary of the package and chip configurations for the traces in a). c) Micrograph image of a standard 'long' on-chip flux bias line routing. d) Micrograph image of a 'short,' symmetric on-chip flux bias line.

6. A two layer normal metal PCB with a 'short' on-chip bias line routing.

The bad (orange) trace, which has an additional settling of $5 \cdot 10^{-3}$ over several hundred microseconds, is a two-layer aluminum-plated copper PCB with copper vias, and the very bad (red) trace, which has a settling of about 20% over a ms or more is a two layer gold-plated copper PCB.

There are three conclusions to draw from these datasets. First of all, since the response of (i), (ii), and (iii) are the same, we see that the aluminum-plated copper PCBs match

the fast settling response of our standard machined aluminum package, so we have a near-term scalable PCB-based packaging solution for our frequency tunable qubits. Secondly, we can compare the packaging differences between the good, bad, and very bad traces to gain insight as to the cause of the long settling time in the copper PCB. In these tests, each package is connected to normal-metal coaxial cabling on its exterior, and aluminum wire bonds connect the package to a tunable resonator chip on its interior. If we use a standard two-layer copper PCB with copper vias, we measure the very bad (red) trace. If we then plate this PCB with aluminum, but leave the vias copper, we measure the bad (orange) trace, where the settling amplitude has been reduced from 0.2 to $5 \cdot 10^{-3}$. Then, if the vias are also plated in aluminum, we measure the good response (black). Again noting that the center trace of the good machined aluminum box is copper, we can conclude that these settling times are likely caused by redistribution of ground plane currents in normal metals. This current redistribution occurs as the transmission line geometry changes from coplanar waveguides on-chip, to a microstrip geometry in the package, and then to normal metal coaxial connectors as the transmission line leaves the packages. It seems plausible that some of the remaining settling in these good packages is due to the transition to the normal metal cables outside the superconducting package.

The third and perhaps the most surprising result is that changing the on-chip bias line routing can have a significant impact on the bias line settling. The very bad (red) trace is a two-layer gold-plated copper PCB package and a typical 'long' flux bias line layout, shown highlighted purple in figure 4.4c. If we use the same normal metal package

with the 'short' bias line layout in figure 4.4d, we measure the good (green) trace in figure 4.4a—indistinguishable from the long bias line layout in the good, superconducting, packages. Both the 'long' and 'short' bias lines have a $50\,\Omega$ coplanar waveguide geometry with superconducting aluminum crossovers every $100\,\mu$m to connect the ground planes and short out slot-line modes. Since both bias lines have superconducting crossovers connecting the ground planes, we do not fully understand why this layout change seems to resolve the settling issue in a normal metal package. However, this indicates that on-chip layout may be able to correct for or prevent non-idealities in the package or other cabling by modifying the return current distribution.

4.6 Conclusions

We have presented a time-domain measurement technique capable of measuring flux waveforms on a superconducting chip from DC to hundreds of MHz with a theoretical bandwidth limit of 2.6 GHz. This transducer has a maximum flux-to-phase gain approaching $1200^\circ/\Phi_0$ that, given our phase measurement noise, translates to a flux sensitivity of $0.2\,\mathrm{m}\Phi_0$. We have benchmarked this technique using our standard cryogenic packaging and then applied this technique to develop a suitable PCB-based package for frequency tunable superconducting qubits. We find that for an arbitrary chip layout, packages with normal metal in the return path may introduce large settling times. This settling may be mitigated by either using superconducting aluminum-plated copper PCBs or by modifying the on-chip layout of these bias lines to avoid return current asymmetry.

Contributions

Tunable resonator devices were made at the UC Santa Barbara Nanofabrication Facility, a part of the NSF funded National Nanotechnology Infrastructure Network. The PCB-based chip mount was a reproduction of a design provided by Northrop Grumman Corporation.

Chapter 5

Superconducting interconnects

We present a fabrication process for fully superconducting interconnects compatible with superconducting qubit technology. These interconnects allow for the three dimensional integration of quantum circuits without introducing lossy amorphous dielectrics. They are composed of indium bumps several microns tall separated from an aluminum base layer by titanium nitride which serves as a diffusion barrier. We measure the whole structure to be superconducting (transition temperature of 1.1 K), limited by the aluminum. These interconnects have an average critical current of 26.8 mA, and mechanical shear and thermal cycle testing indicate that these devices are mechanically robust. Our process provides a method that reliably yields superconducting interconnects suitable for use with superconducting qubits.

5.1 Introduction

As superconducting qubit technology grows beyond one dimensional chains of nearest neighbor coupled qubits [**?**], arbitrarily sized two dimensional arrays are the logical next step towards both surface code error correction and more complex high fidelity quantum circuits [69]. Following on the demonstration of prototypical two dimensional arrays [76, 77, 78], these superconducting interconnects were critical to the implementation of the Sycamore quantum processor used to demonstrate quantum supremacy [10]. For example, frequency tunable Xmon transmon qubits on the interior of a two dimensional array require coupling to four nearest neighbor qubits and a readout resonator as well as individually addressable XY drive lines and an inductively coupled flux line [79]. Routing these control wires with a single layer of base wiring and crossovers is not scalable beyond a few-deep array of qubits. Multilayer fabrication with embedded routing layers is a natural solution [80], but integrated dielectric layers on a qubit wafer introduce additional decoherence to the qubits [81]. This individual addressability problem can be solved by separating the device into two chips, a dense wiring chip that allows for lossy dielectrics and a pristine qubit chip with only high coherence materials. Combining these two chips to form a hybrid device provides the advantages of both technologies.

A hybrid device is composed of a "base substrate" bonded to a "top chip." Hybridization allows for improved impedance matching between chips as compared to wirebonds and the close integration of incompatible fabrication processes. A qubit hybrid would also benefit from the availability of straightforward capacitive, inductive, or galvanic

coupling of electrical signals between the base substrate and top chip through the use of parallel plate capacitors and coupled inductors. Hybrid devices have become ubiquitous in the semiconductor industry, finding applications in everything from cell phones to the Large Hadron Collider [82]. Cryogenic applications are fewer; bolometer arrays for sub-millimeter astronomy [83, 84] and single flux quantum devices [85, 86] have utilized this technology. Low resistance cryogenic bump bonds [87, 88] and superconducting bump bonds that proximitize normal metals have also been fabricated [89]. Here we present a novel bump bond metal stack up consisting of all superconducting materials with the intent of achieving maximal flexibility in designing flux tunable qubit circuits where mA control currents are necessary.

In order to maintain compatibility with our existing qubit architecture, bump bond interconnects for a superconducting qubit hybrid must meet these requirements:

1. Bumps must be compatible with qubit fabrication (e.g., aluminum on silicon).

2. If interconnects will be used in routing control signals (rather than just as ground plane connections and chip spacers), fabrication yield must be high. e.g., With a 99.9% yield, a device with 700 interconnects on control lines would yield all lines ($0.999^{700} =$) 50% of the time.

3. Interconnects must continue to perform electrically and mechanically after cooling from 300 K to 10 mK.

4. Bonding must be accomplished at atmospheric pressure without elevated process

temperatures to avoid altering Josephson junction critical currents through annealing [90].

5. Interconnects must superconduct to provide a lossless connection between chips and avoid local heating.

6. The critical current of the interconnects must exceed 5 mA to enable applications in current-biased flux lines.

To satisfy condition (1) above and to extend our wire-routing capabilities through known multi-layer techniques, bumps must provide a connection between aluminum wiring on both the base substrate and top chip. This design consideration will allow us to connect our qubit fabrication to a dense, multi-layer, wire routing device based on standardized complementary metal-oxide-semiconductor (CMOS) fabrication techniques. Known bump bonding materials that also superconduct include indium and various soldering alloys. Indium is a natural choice because high purity sources are readily available, it can be deposited in many μm thick layers by thermal evaporation, it has a relatively high critical temperature of 3.4 K, and room temperature indium bump bonding is an industrially proven technology [91]. However, since aluminum and indium form an intermetallic [92], under bump metalization (UBM) it is necessary to act as a diffusion barrier. Fortunately, titanium nitride, fulfills our UBM requirements as it is a well known diffusion barrier (used in CMOS fabrication) with a T_c as high as 5.64 K and has also been shown to be a viable high-coherence qubit material [93, 94].

5.2 Device fabrication and layout

Figure 5.1 shows a minimal, qubit compatible, asymmetric bump bond process used here for DC characterization. The base substrate has a full aluminum/titanium nitride/indium metal stack and, for simplicity, the top chip has just a single layer of indium wiring (which allowed us to avoid the complication of processing with two die sizes in every fabrication run while still testing all the necessary metal interfaces). In this case, as current flows between the base substrate and top chip, it passes through one aluminum/titanium nitride interface, one titanium nitride/indium interface, and one indium/indium interface. Actual qubit hybrids would be symmetric, with aluminum wiring and titanium nitride UBM on both chips, which adds one aluminum/titanium nitride interface and one titanium nitride/indium interface to the metal stack for each interconnect.

For the base substrate, we first blanket deposit 100 nm of aluminum through e-beam evaporation–the same base wiring material used in qubit fabrication [95]. The base wiring, shown in figure (5.1a), is defined with optical lithography and a $BCl_3 + Cl_2$ plasma dry etch (although lift-off defined aluminum base wiring has been used with similar results). Then, (5.1b) titanium nitride pads are defined in lift-off resist and the device is placed into a sputter chamber where an *in situ* ion mill (see B.3 for ion milling parameters) removes the native oxide from the aluminum (5.1b) before titanium nitride is reactively sputtered in argon and nitrogen partial pressures (5.1c). After titanium nitride lift-off, the indium pillars are defined in lift-off resist and, then (5.1d), in a third vacuum chamber, another in situ ion mill (B.3) is used to remove oxide and contaminants from the titanium

nitride surface, before depositing indium in a thermal evaporator with the substrate cooled to 0 °C (5.1e). The right side of Figure 5.1e shows the single layer of indium lift off used to define indium wiring on the top chip. For the devices we characterized here, we deposited 5 μm of indium on the substrate and separately deposited 2 μm of indium on the top chip.

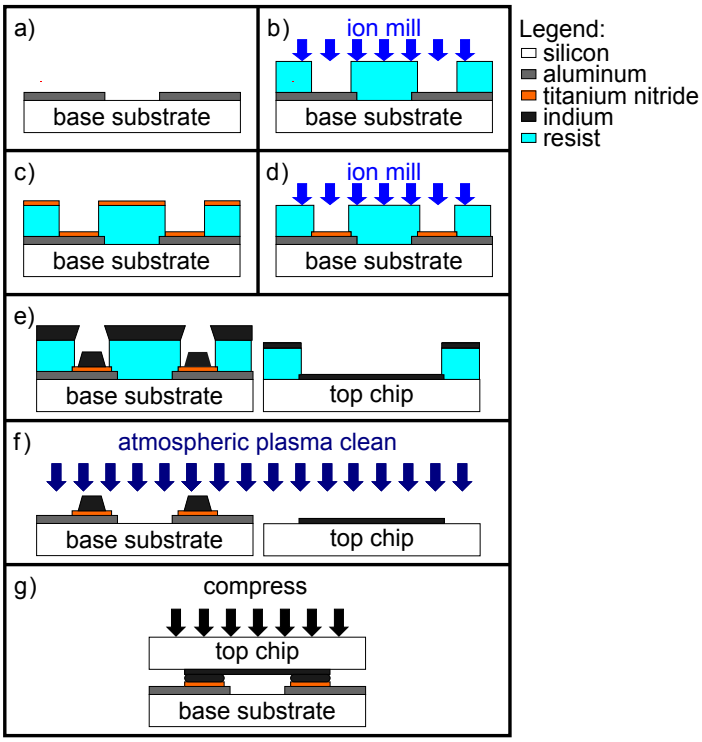


Figure 5.1: Hybrid fabrication process; (a-d) describe steps specific to the base substrate and (e-g) are common to both the base substrate and top chip. a) On a silicon substrate, a base electrode is defined in 100 nm of e-beam evaporated aluminum by a $BCl_3 + Cl_2$ plasma dry etch. b) The native aluminum oxide is removed by an ion mill at locations defined by lift-off resist. c) In the same vacuum chamber as b), 50-80 nm of titanium nitride is sputter deposited from a pure titanium source in argon and nitrogen partial pressures. d) After lift-off of the titanium nitride and patterning new resist, oxide and contaminants are removed from the titanium nitride by an ion mill at locations defined by lift-off resist. e) In the same vacuum chamber as d), 2-10 μm of indium is deposited by thermal evaporation on both the base substrate and top chip. f) After lift-off of the indium, an atmospheric plasma is used to clean and passivate the surface of both devices a few minutes before bonding. g) The base substrate and top chip are aligned and compressed together at room temperature to complete the hybrid.

After both the base substrate and top chip have been fabricated, an atmospheric plasma surface treatment (with a mix of of hydrogen, helium and nitrogen) is used to remove surface oxide and passivate the surface of the indium a few minutes before the two chips are bonded together (5.1f). This surface treatment is critical to making good indium-to-indium contact during bonding without reflowing the indium [96]. We then flip over the top chip, align the two devices, and compress the dies together using a SET FC-150 flip-chip bonder (5.1g). Bonding is performed at room temperature with a typical bonding force of 10-20 N per mm^2 of bump area for 15 μm diameter bumps (2-5 grams/bump), which results in a compression of roughly 40-60% the total height of the two indium depositions. Inspection with an edge gap tool indicates that typically the tilt between the base substrate and top chip is parallel within ± 0.5 mRad, and inspection with an infrared microscope indicates that the xy alignemnt is typically within $\pm 2\,\mu$m.

Choosing an appropriate bump geometry is subject to several constraints. First, it is desirable to have a chip-to-chip separation of at least several microns so that the impedance of a 2 μm wide, 50 Ω coplanar waveguide transmission line is not dramatically changed by the presence of an overhead ground plane. Providing sufficient separation allows designs to be insensitive to the final chip-to-chip separation and for a smooth impedance transition as transmission lines travel under the edge of the top chip. In order to achieve a desired separation of 2-10 μm post-compression, 2-10 μm of indium must be deposited on both the base substrate and top chip. When depositing such thick layers of material, especially a high mobility material like indium, sidewall deposition can result

in a considerable constriction of the bump feature size. 15 μm diameter bumps were chosen as they have a width to height aspect ratio of 3:2 at the thickest intended bump height; for more information on thick indium deposition see B.2. Secondly, the titanium nitride UBM footprint must be large enough so that, after compression, indium does not contact aluminum directly. Given the post-compression alignment accuracy of our flip chip bonder ($\pm$ 2 μm) and an expected 50% compression, we find that 30 μm square titanium nitride pads are sufficient for 15 μm diameter indium pillars.

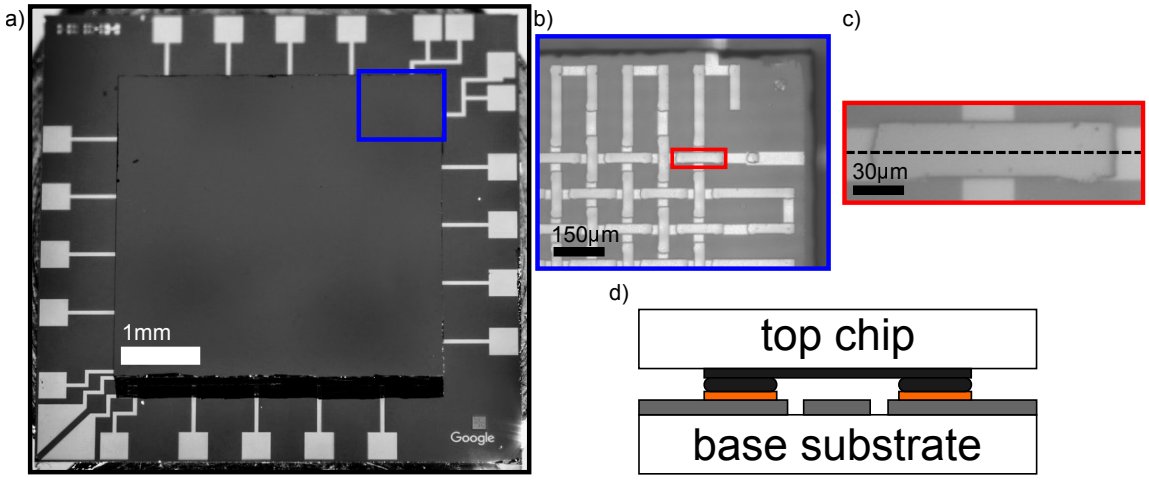


Figure 5.2: Design of the bump bond DC characterization hybrid. a) Photograph of a hybrid device with a 6 mm x 6 mm base substrate and a 4 mm x 4 mm top chip. b) Infrared micrograph looking through the top chip of the hybrid device. The woven pattern of test circuit can be seen, and bumps are located on either side of the crossings to connect the base wire from the base substrate to the top chip and back. c) Zoomed in infrared micrograph of a single indium bar on the top chip with interconnects at either end. d) Cross-sectional diagram of the device along the dotted line in c).

The devices characterized here consist of a 6 mm x 6 mm base substrate and a 4 mm x 4 mm top chip shown in Figure 5.2. In order to electrically characterize a large number of interconnects, we place 1620, 15 μm diameter, circular indium bumps on the base substrate

and $30\,\mu$m x $150\,\mu$m indium bars on the top chip to connect pairs of bumps into a series chain of 1620 chip-to-chip interconnects. At each end of the chain, and every 90 interconnects along the chain, we wire bond to pads on the perimeter of the chip. This wiring configuration allows us to make four-wire resistance measurements by applying an excitation current to any 90 interconnect subsection (or number of subsections of the device) while measuring the voltage across that subsection/s with other leads. Each section of 90 interconnects consists of three rows or columns that extend across the entire top chip, spread over an area of roughly $2\,\text{mm}^2$. By weaving these rows and columns of together, as shown in figure 5.2b, we are able to ascertain whether or not electrical failures are spatially correlated. For instance, if one subsection arranged in the rows fails to superconduct or has a suppressed critical current, but none of the columns show the same behavior, it is likely that there are no spatially correlated failures. However, if one section of rows and one section of columns fails, then the intersection indicates a region of interest for failure analysis such as electron energy loss spectroscopy (EELS), focused ion beam (FIB) cross sections, post-shear inspection, or inspection with an optical or infrared microscope.

5.3 Electrical characterization

We perform low temperature four-wire electrical measurements in an adiabatic demagnetization refrigerator (ADR) down to 50 mK using a lock-in amplifier, ammeter, source measure unit (SMU) and a matrix switch to rapidly characterize a large number of de-

vices. Twisted pair wiring and shielding is used to reduce parasitic coupling between the current excitation leads and voltage sense leads. Common mode voltage correction is implemented with the matrix switch which also allows us to quickly switch between measurements. For a detailed look at the measurement system as well as the the resistance and critical current measurements discussed below, see B.1.

This setup allows us to make a resistance measurement of the device in its superconducting state. Using common mode compensation and the lock-in amplifier with a several mA sinusoidal test current, we are typically able to bound the resistance of a series chain of 1620 interconnects to be less than $5\,\mu\Omega$ below 1.1 K, which is an average resistance of $3\,\mathrm{n}\Omega$ per interconnect. Figure 5.3a shows a typical resistance versus temperature curve for a full 1620 interconnect chain and a 2 interconnect test structure on the same device. At 1.1 K we observe a clear transition to a superconducting state when the resistance of 1620 interconnects in series falls more than 7 orders of magnitude to a few $\mu\Omega$. The resistance measured below 1.1 K is roughly the same for both 1620 interconnects and the 2 interconnect test structure which indicates that this measurement is likely limited by system parasitics or measurement electronics rather than by an actual resistance or the inductance of the device. In figure 5.3b we use a SMU to assess the critical current of each of the eighteen 90 interconnect subsections on three hybrid devices. The average critical current for each subsection is 26.8 mA, with a number of subsections above 30 mA and a single subsection with a suppressed critical current of 10.3 mA. This data represents 4860 interconnects, 100% of which superconduct with a critical current above 10 mA.

Furthermore, at least 98% of the interconnects have a critical current above 24.5 mA. Since there was only one section of rows (and no columns) with a suppressed critical current, it is likely that a single interconnect could be responsible for the lower critical current. The high yield of this process and lack of spatially correlated failures indicate that parallel interconnects can be used to further increase the critical current and/or to serve as precautionary redundant connections (though we yielded 100% on these 3 test devices and have had similar yields across several generations of test devices). The average room temperature resistance of these 90 interconnect subsections is 47.7 Ω with a standard deviation of 2 Ω indicating reasonable bump uniformity. Typically we find that a room temperature resistance <1 Ω/interconnect (including the aluminum and indium base wiring used to chain them together) indicates that the flip chip bonding was successful. We find that insufficient compression or a bad material interface results in a resistance higher than 1 Ω per interconnect. [Do we have better data now?]

5.4 Mechanical characterization

Several mechanical tests were performed on a different generation of hybrids consisting of a 10 mm x 10 mm substrate and a 6 mm x 6 mm square chip. These devices had about four thousand 20 μm diameter circular bump bonds spread fairly evenly over the 36 mm^2 area of the top chip. In order to characterize the mechanical strength of these interconnects, we performed destructive die shear strength tests (in accordance with MIL-STD-883) in which a force is applied to the edge of the top chip, parallel to the face of the chip (e.g.,

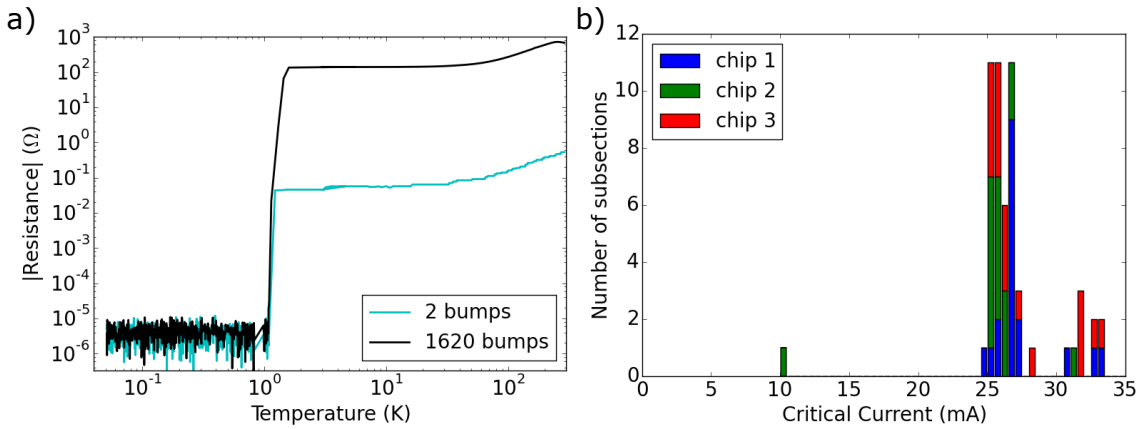


Figure 5.3: Electrical device characterization. a) Typical four-wire resistance measurement versus temperature for a chain of 1620 interconnects and a 2 interconnect test structure on the same device from room temperature to 50 mK. A superconducting transition can be seen at 1.1 K where the resistance of both the 1620 and 2 interconnect structures fall to a few $\mu\Omega$. For the 1620 long chain, this measurement demonstrates a superconducting resistance more than 7 orders of magnitude lower than its normal state resistance at 3 K. b) Histogram of critical currents for each of the eighteen 90-interconnect subsections on three different chips. The average critical current is 26.8 mA with >98% of the subsections above 24.5 mA

force was applied in the plane of the page as the chip is shown in figure 2a), until the top chip separates from the substrate. Four devices were tested; three separated at 35 N and one exceeded the limits of the tool at 49.9 N, all of which are more than sufficient to ensure that devices are robust enough for handling. Finally, thermal cycling was performed on a device that had been previously confirmed to be fully superconducting below 1.1 K. One hundred thermal cycles from -80 °C to 45 °C were performed with a 23 minute dwell at both -80 °C and 45 °C and a 20 °C/min ramp rate for transitions. After 100 thermal cycles (and unknown conditions during round-trip ground shipping to our off-site lab) the sample was cooled back down to 50 mK. All interconnects on the device still remained superconducting, although the critical current was reduced to 1-5 mA in

most subsections down from 20-25 mA in the initial characterization of this device. The reason for the reduced critical current is not known, but it is worth noting that, in a more typical use case, the devices measured in figure 5.3 were cycled from room temperature to 50 mK and back as many as three times in our ADR (approximately 0.2°C/min average warming/cooling rate) with no measurable impact on the critical current.

5.5 Conclusions

The flip chip hybrid devices we have developed offer a viable solution to control signal routing in two-dimensional high-coherence circuits. These interconnects, consisting of a titanium nitride diffusion barrier and indium bumps, serve as electrical interconnects between two planar devices with aluminum wiring. This fabrication process opens the door to the possibility of the close integration of two superconducting circuits with each other or, as would be desirable in the case of superconducting qubits, the close integration of one high-coherence qubit device with a dense, multi-layer, signal-routing device. Furthermore, these interconnects have a typical critical current above 25 mA which is an order of magnitude larger than the largest typical DC control currents used to flux-tune superconducting qubits. Limited by the aluminum, these bumps are fully superconducting below 1.1 K, and below this critical temperature, we are able to estimate the resistance of each bump to $< 3\,\mathrm{n}\Omega$. These high yield, mechanically robust, and high critical current electrical interconnects are ready to be implemented into more complex circuits including two dimensional arrays of nearest neighbor coupled flux-tunable superconducting qubits.

Chapter 6

Continuous Set of Two-qubit Gates

Quantum algorithms offer a dramatic speedup for computational problems in machine learning, material science, and chemistry. However, any near-term realizations of these algorithms will need to be heavily optimized to fit within the finite resources offered by existing noisy quantum hardware. Here, taking advantage of the strong adjustable coupling of gmon qubits, we demonstrate a continuous two-qubit gate set that can provide a 3x reduction in circuit depth as compared to a standard decomposition. We implement two gate families: an iSWAP-like gate to attain an arbitrary swap angle, θ, and a CPHASE gate that generates an arbitrary conditional phase, ϕ. Using one of each of these gates, we can perform an arbitrary two-qubit gate within the excitation-preserving subspace allowing for a complete implementation of the so-called Fermionic Simulation, or fSim, gate set. We benchmark the fidelity of the iSWAP-like and CPHASE gate families as well as 525 other fSim gates spread evenly across the entire fSim(θ, ϕ) parameter

space achieving purity-limited average two-qubit Pauli error of 3.8×10^{-3} per fSim gate.

6.1 Introduction

Quantum computing is a potentially transformative technology, but challenges remain in identifying a path towards solving practical problems with a quantum advantage [5]. Continued progress towards this goal may be made on many fronts including qubit coherence or scalability [32, 68], measurement or gate fidelities [26, 57], and algorithmic improvements that reduce the required circuit depth through either compilation or innovation [97]. In superconducting qubit systems, single-qubit gates are usually a factor of two or more lower error than two-qubit gates. Consequently, a typical strategy has been to demonstrate a minimally universal gate set consisting of arbitrary single-qubit rotations and a single two-qubit gate [98]. This is an efficient approach for some algorithms, e.g. surface code error correction, which compiles optimally with such a gate set [33]. However, many noisy intermediate-scale quantum (NISQ, [9]) algorithms require a more diverse set of two-qubit gates. An implementation of these gates could take the place of six to eight single-qubit gates and three CZ_ϕ gates per arbitrary two-qubit gate required with an optimal decomposition into a standard minimally-universal gate set [99].

To maximize our compilation efficiency, we need the largest two-qubit gate set that may be implemented with high fidelity using our tunable coupling gmon qubit architecture [100]. A general two-qubit unitary gate allows independent control over the strength of $\sigma_X\sigma_X$, $\sigma_Y\sigma_Y$, and $\sigma_Z\sigma_Z$, coupling between qubits requiring both DC and microwave

control of gmon qubits. However, models of interacting particles typically conserve the number of excitations corresponding to a simpler model where the $\sigma_X\sigma_X$ and $\sigma_Y\sigma_Y$ couplings have equal coefficients. This reduces the number of control parameters from three to two and eliminates the need for microwave control during an algorithm. This set of excitation-conserving gates has been appropriately termed the Fermionic Simulation, or fSim, gate set since it maps electron conservation in a chemistry problem to photon conservation in the qubits [13]. An fSim gate can be defined with two control angles, θ, the $|01\rangle \leftrightarrow |10\rangle$ swap angle, and, ϕ, the phase of the $|11\rangle$ state with a matrix representation in the $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ basis given by:

$$\mathrm{fSim}(\theta, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\,\theta & -i\sin\,\theta & 0 \\ 0 & -i\sin\,\theta & \cos\,\theta & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix} \tag{6.1}$$

We use this as both a convenient definition and a useful model for describing general two-qubit gates resulting from arbitrary flux control of gmon qubit frequencies and their coupling. Notably, promising low-depth algorithms using this gate set have been proposed including the quantum approximate optimization algorithm [101] and an algorithm for linear-depth circuits simulating the electronic structure of molecules [13]. Additionally, algorithms performed with just Z rotations and fSim gates enable both post selection and zero noise extrapolation [102] error mitigation techniques, further improving this gate sets prospects on NISQ processors.

In this letter we first demonstrate the strong flux tunable coupling between gmon

qubits which we use to perform fast two-qubit gates. Then, to describe our calibration and control strategy, we use shallow circuits to illuminate the natural correspondence of the coupled transmon Hamiltonian and the fSim gate set. We use cross-entropy benchmarking (XEB, [103]) to characterize two linearly independent and continuous families of entangling gates: the iSWAP-like family corresponding to $\mathrm{fSim}(\theta, \phi \propto \theta^2)$, and the CPHASE family corresponding to $\mathrm{fSim}(\theta \approx 0°, \phi)$. We then combine these two continuous gate sets to calibrate and benchmark 525 fSim gates spread evenly across the entire (θ, ϕ) parameter space.

6.2 Strong coupling with gmon qubits

The quantum processors used in this work each consist of four gmon transmon qubits in a chain, sandwiched together with three couplers. Both the qubit frequencies and the coupling between qubits can be independently controlled, providing multiple advantages over fixed coupling designs [100, 48, 104]. Firstly, since we can turn off the coupling with any detuning, both qubits may idle and perform single-qubit gates while operating closer to their flux insensitive point. This improves dephasing and decreases our sensitivity to flux settling tails. Secondly, since entangling gates are performed by bringing the qubit states near resonance, idling the qubits closer together means that gates require much smaller dynamic detunings, reducing the effect of flux settling tails [105, 106]. Thirdly, since the on/off coupling ratio is not dependent on the maximum qubit-qubit detuning,

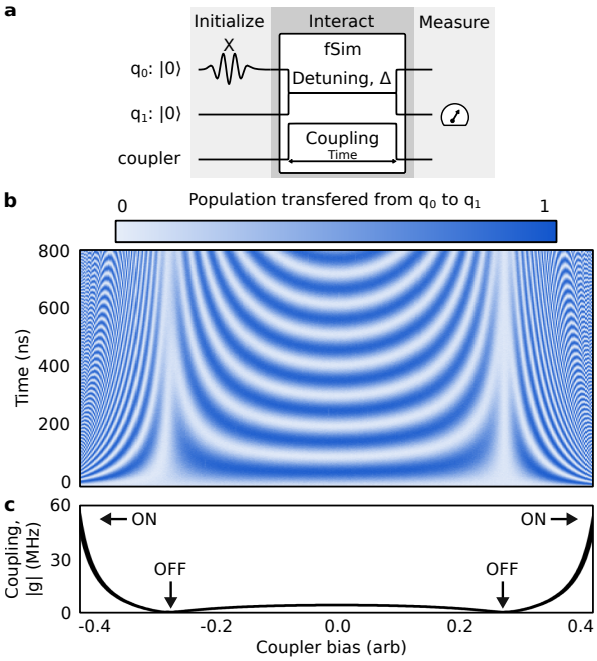


Figure 6.1: Demonstration of the tunable coupling between gmon qubits. **a**, Pulse sequence used to measure swapping as a function of coupler bias. We initialize one qubit, perform an fSim gate, defined by a set of three flux pulses that control the qubit frequencies and the coupling between the qubits, and measure the population of the other qubit. **b**, We vary the fSim gate as a function of the length and amplitude of the coupler pulse to measure the swap rate as a function of the bias amplitude. **c**, By taking the Fourier transform of the oscillations in **b**, we extract the coupling strength, $|g|$, as a function of coupler bias.

we are able to increase the overall coupling strength allowing for faster gates and reduced decoherence error.

In Figure 6.1 we characterize the qubit-qubit coupling strength as a function of the coupler flux bias by measuring the Rabi swap rate. We use the pulse sequence in Figure 6.1a where we initialize one qubit, apply an fSim gate, and measure the population transferred to the other qubit. Each fSim gate is defined by the amplitude and length of three nominally rectangular flux bias pulses, two of which control the qubit frequencies and set their relative frequency detuning, Δ, while the third pulse controls the relative

coupling strength between the qubits, g. In Figure 6.1b we repeat this pulse sequence using the qubit flux biases to place them on resonance ($\Delta = 0\,\text{MHz}$) while varying the coupler bias amplitude and the shared duration of all three pulses. By taking the Fourier transform of the oscillating population data in 6.1b we extract the Rabi swap rate as a function of coupler bias which is equivalent to twice the qubit-qubit coupling, $|g|$, which is plotted in Figure 6.1c. We measure $|g| = 6\,\text{MHz}$ when the coupler is biased to zero Φ_0, and a coupling exceeding 50 MHz as the coupler bias approaches $\Phi_0/2$. The coupling changes sign between these two regions ensuring we can turn the coupling off. During general operation, we idle and perform single-qubit gates with the coupler at the "OFF" bias and make excursions to stronger couplings ("ON" region) to perform fSim gates. In this work we use $g_{\max} \approx 45\,\text{MHz}$ which is three times stronger than is typical for fixed coupling devices allowing for three times faster gates.

6.3 Coupled Transmon physics and the fSim gate set

In the absence of a resonant microwave drive, coupled transmon qubits naturally evolve within the excitation-preserving subspace. The specific time evolution is determined by three parameters: the qubit nonlinearity, η, the qubit-qubit frequency detuning and the coupling between qubits, g. While η is fixed at 240 MHz by our qubit capacitance, the gmon architecture allows for time-dependent control of both Δ and $|g|$ using DC to $\approx 200\,\text{MHz}$ bandwidth flux waveforms. The qubit center frequency, $(f_{q1} + f_{q2})/2$, is a free parameter that may be used to avoid coupled two level system (TLS) defects

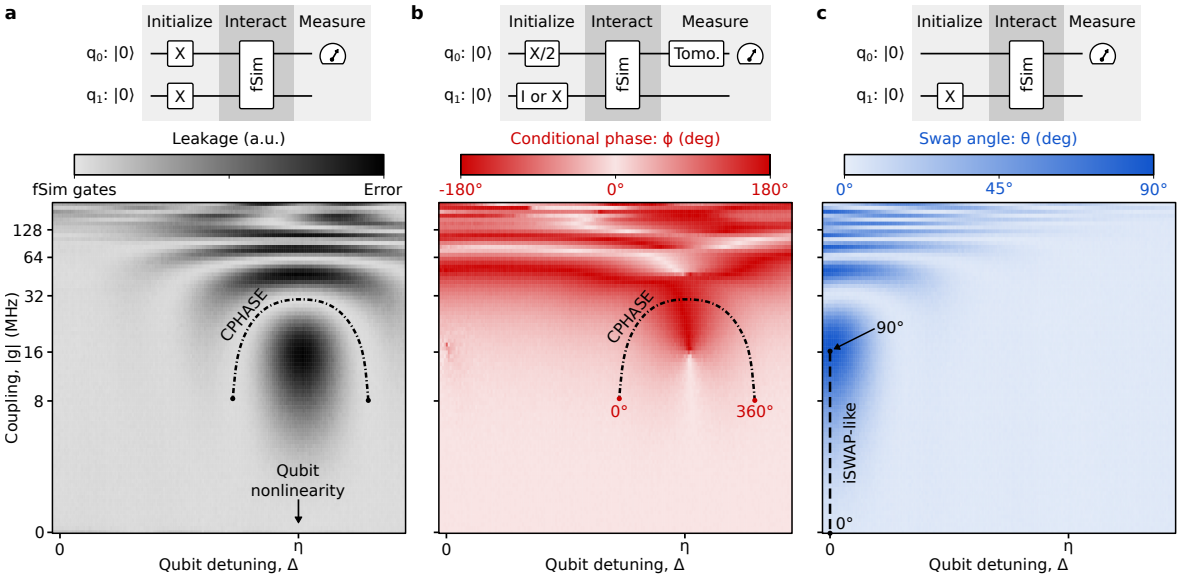


Figure 6.2: Exploring the parameter space of two-qubit gates. Each pixel represents one experiment. We use a set of 15 ns rectangular current bias waveforms to perform some fSim unitary by setting the qubit-qubit detuning, Δ, and the coupling strength, g. **a**, To identify the low-leakage gates described by the fSim model, we measure leakage by initializing the $|11\rangle$ state and measuring the $|02\rangle$ state. When the detuning is near the qubit nonlinearity, we observe the expected Rabi oscillations. **b**, We measure the conditional phase, ϕ, by performing a Ramsey experiment where we initialize one qubit with an X/2 gate and perform tomography to measure the difference in accumulated phase (ϕ) with and without initializing the other qubit to the $|1\rangle$ state. By choosing combinations of Δ and g as indicated by the CPHASE dash-dotted line (chosen as the low-leakage coupling strength from **a**), we are able to achieve any $\phi : [-180°, 180°]$. **c**, We measure the swap angle, θ, by initializing the $|01\rangle$ state and measuring the $|10\rangle$ state. By placing the qubits on resonance and varying the coupling strength along the iSWAP-like dashed line, we are able to achieve any $\theta : [0°, 90°]$.

present in the frequency spectrum of either qubit [107, 108, 109]. For simplicity, we limit our fSim control pulses to synchronous, nominally rectangular waveforms defined by four parameters: a shared length, typically 13 ns to 15 ns and three control amplitudes that set the differential qubit detuning and the coupling strength. While further pulse shaping may improve gate performance in the future, these basic waveforms were sufficient to approach the decoherence limit of our qubits which have a T_1 of $25.3 \pm 7.3\,\mu$s (T_1 versus frequency data in Appendix C.3.2).

The full fSim control model describes any low-leakage two-qubit unitary evolution with five parameters, θ and ϕ, which we have already discussed, in addition to three parameters describing single-qubit phases as detailed in Appendix C.1. Here, we focus on the parameters that describe the two-qubit interaction and use the three experiments described in Figure 6.2 to measure leakage to the $|02\rangle$ state and map out the ϕ and θ control landscape (A complete unitary tomography procedure is outlined in Appendix C.3.3). Each experiment follows the same pattern: initialize a relevant state, apply fSim control pulses, and then perform either population or tomographic measurements to extract the desired qubit's population or phase. Within the fSim model, leakage is the dominant error. In Figure 6.2a, we map out leakage by initializing $|11\rangle$ and measuring the $|0\rangle$ population of the lower frequency qubit as a proxy for leakage in the higher frequency qubit. In Figure 6.2b we explore the ϕ parameter space by performing a Ramsey experiment where we take the difference in the accumulated phase with and without the second qubit initialized to the $|1\rangle$ state. The conditional phase is accumulated

more quickly when the $|11\rangle$ and $|02\rangle$ states are near resonance. Finally, in Figure 6.2c we explore the θ parameter space by initializing one qubit, and measuring the $|1\rangle$ population of the other qubit after the fSim gate. The Rabi oscillation physics explored with these measurements is reproduced with fairly rudimentary numerics in appendix C.2, but these experiments serve to demonstrate our fSim control strategy.

6.4 Benchmarking iSWAP-like and CPHASE gates

The data presented in Figure 6.2 provides a map for implementing an arbitrary fSim—each pixel defines a set of three control amplitudes, and any control amplitudes yielding low-leakage should result in a high fidelity gate described by the fSim control model (Eq. 6.1). While it may be possible to perform an arbitrary fSim gate with a single set of flux pulses using either very strong coupling or more complex control waveforms, we have chosen to implement an arbitrary fSim gate as a composition of two continuous gate families using simple rectangular control pulses to minimize the gate length. The first gate family completes a diabatic $|11\rangle \rightleftarrows |02\rangle$ swap to perform a gate with an arbitrary conditional phase, ϕ, using control amplitudes denoted by the dot-dashed line labeled 'CPHASE' in Figures 6.2a and 6.2b. The dominant control angle in the CPHASE gate family is the conditional phase, but, we do accumulate a small swap angle θ due to the strong coupling necessary to perform a fast CPHASE gate ($\theta \leq 5^\circ$ for a 13 ns CPHASE gate—this may be reduced considerably by increasing the gate duration by ≈ 5 ns). The second gate family places the qubits on resonance ($\Delta = 0$ MHz) and uses a variable coupling strength to reach

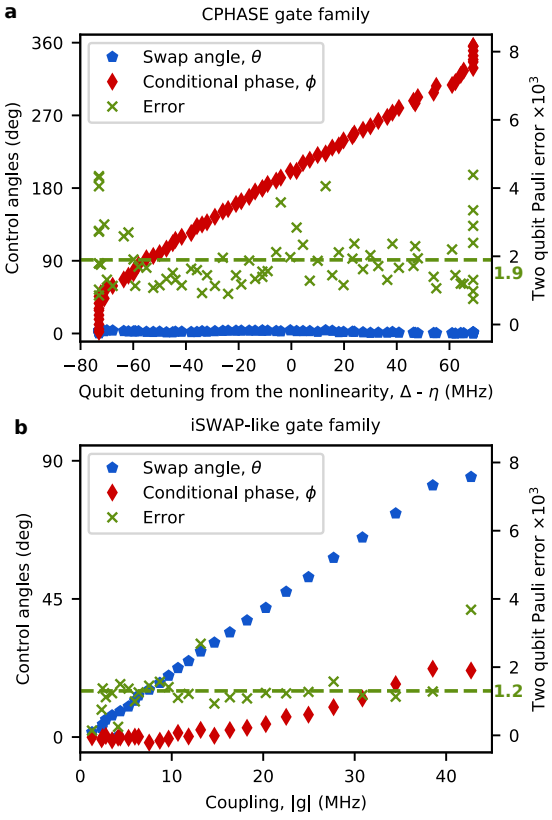


Figure 6.3: Characterizing the iSWAP-like and CPHASE gate families with cross-entropy benchmarking. We plot the optimized fSim control angles, θ and ϕ, on the left y-axes and the Pauli gate error per two-qubit gate on the right y-axes, conservatively assuming 7.5×10^{-4} single-qubit Pauli gate errors. **a**, Characterization of the CPHASE gate family corresponding to fSim($\theta \approx 0°, \phi$). Each gate is 15 ns long, consisting of control pulses that vary the qubit detuning, Δ, around the qubit nonlinearity, η, with a coupler bias amplitude chosen to complete one full swap: $|11\rangle \rightarrow |02\rangle \rightarrow |11\rangle$. We measure an average two-qubit Pauli error of 1.9×10^{-3} for the CPHASE family. **b**, Characterization of the iSWAP-like gate family corresponding to fSim($\theta, \phi \propto \theta^2$). Each gate is 13 ns long, consisting of control pulses that place the qubits on resonance and vary the coupling strength, $|g|$, to achieve an arbitrary swap angle θ between the $|01\rangle$ and $|10\rangle$ states. We measure an average two-qubit Pauli error of 1.2×10^{-3} for the iSWAP-like family.

any swap angle, θ, using control amplitudes along the dashed line labeled "iSWAP-like" in Figure 6.2c. We have deemed this gate family "iSWAP-like" since the swap angle varies from $\theta : [0°, 90°]$ and because this gate accumulates a conditional phase $\phi \propto \theta^2$ due to the dispersive interaction with the $|02\rangle$ and $|20\rangle$ states. Both of these gates are a subset of the fSim group individually, and, compiled together, they can reach the full fSim parameter space.

In Figure 6.3 we characterize both the iSWAP-like and CPHASE gate families using cross-entropy benchmarking (XEB) [103]. On the left axes in Figure 6.3a and b, we plot the optimized values of θ and ϕ for a range of CPHASE and iSWAP-like gates, and on the right y-axes we plot the Pauli error per two-qubit gate (see Appendix C.3.2).

6.5 Benchmarking fSim gates

In Figure 6.4a we present the Pauli error of 525 distinct fSim(θ, ϕ) gates where the values of θ and ϕ have been constrained to be exactly the value indicated by the xy-coordinates at the center of each pixel (where *ex situ* optimization has been used only to optimize the single-qubit phases). Each 28 ns long fSim gate in Figure 6.4 is a composition of a 15 ns CPHASE gate followed by a 13 ns iSWAP-like gate. While the fSim fidelity is largely independent of the values of θ and ϕ there are a few features of note. As discussed in Appendix C.4.3, we most-directly calibrated line cuts at $\theta = 0°, 90°$ and $\phi = 180°$. The regions of higher error where ϕ is near $0°$ $(360°)$ involve the most extrapolation from the directly calibrated gates. Secondly, there is a faintly-visible indication of a band of

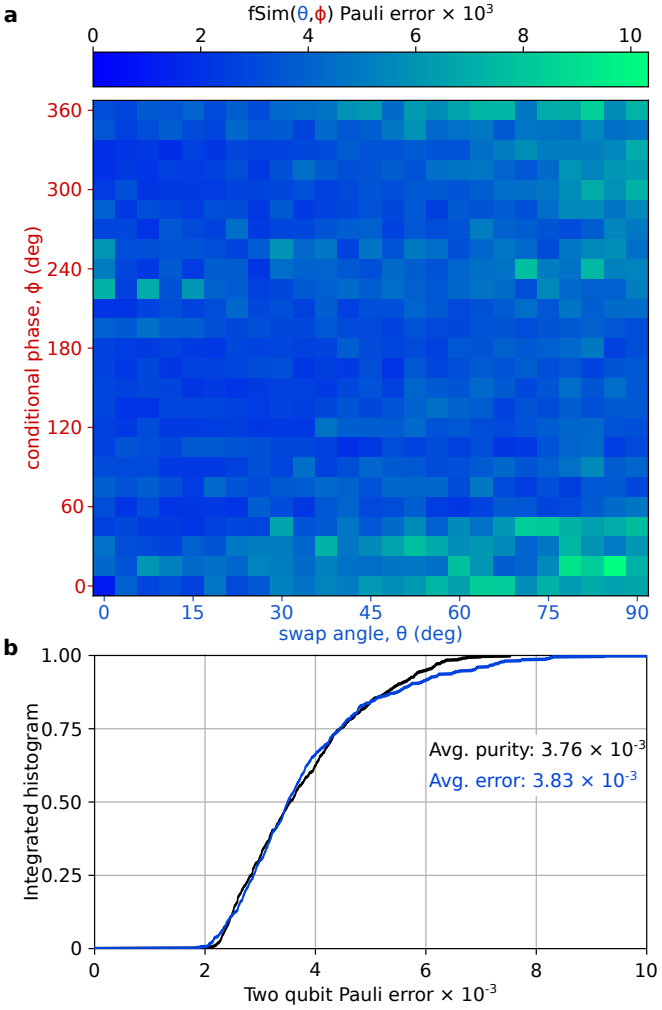


Figure 6.4: Benchmarking the fSim gate set. Using XEB, we measure the Pauli error per cycle and subtract off a conservative estimate for the single-qubit Pauli gate errors of 7.5×10^{-4}. **a**, We plot the two-qubit Pauli error for 525 fSim gates where θ and ϕ have been constrained to a grid. **b**, Histogram of both the error and purity for the gates presented in **a**. Here we confirm a purity (coherence) limited average error for our fSim gates of 3.83×10^{-3}.

higher error near $\phi \approx 240^\circ$ which we believe is due to a weakly interacting TLS defect in the spectrum of one of the qubits—in the future we hope to be able to avoid such defects by using a symmetric detuning. In Figure 6.4b we histogram these results in addition to the purity [110] per fSim and confirm a purity-limited average Pauli error of 3.83×10^{-3} per fSim gate.

Table 6.1: Summary of named fSim gate errors from cross-entropy benchmarking.

Gate	fSim	Pauli error
CZ_ϕ	fSim$(0^\circ, 180^\circ)$	4.5×10^{-3}
SWAP	fSim$(90^\circ, 180^\circ)$	4.8×10^{-3}
$\sqrt{SWAP}$	fSim$(45^\circ, 180^\circ)$	6.1×10^{-3}
iSWAP	fSim$(90^\circ, 0^\circ)$	8.6×10^{-3}
$\sqrt{iSWAP}$	fSim$(45^\circ, 0^\circ)$	8.3×10^{-3}

6.6 Conclusions

We have implemented continuous iSWAP-like and CPHASE gate families with average Pauli error rates of 1.2×10^{-3} and 1.9×10^{-3} respectively. These fast (13-15 ns) gates take advantage of the strong, tunable, qubit-qubit coupling offered by our gmon transmon qubit architecture achieving error rates more than a factor of two lower than the best previously reported two-qubit gates for superconducting qubits [111]. Additionally, we have combined these two gate sets to demonstrate a complete implementation of the two-qubit fSim gate set with an average Pauli error of 3.83×10^{-3} per gate. This direct implementation of the fSim gate offers roughly an additional factor of four in compilation efficiency for promising NISQ algorithms over a traditional minimally-universal gate set.

Appendix A

Fast Flux Measurement Appendices

A.1 Measuring microwave reflections with the tunable resonator

At the core of the experiment described in the main text, we use the tunable resonator as a flux to microwave phase transducer. DC (or time-dependent) flux applied to the SQUID translates to a corresponding phase (or phases) of the reflected microwave tone. Ideally this allows us to sample the on-chip flux waveform and deduce the transfer function of the flux wiring. In reality, microwave reflections that occur after the probe tone has reflected off the tunable resonator will induce time-dependent phase changes which are a source of systematic error in our flux measurement. In this appendix we discuss an alternative experiment with the tunable resonator in which we use it to characterize the location and amplitude of the dominant reflections in our microwave measurement chain. This

technique allows us to bound the systematic error in our flux measurement and may also be used to characterize the impedance matching of microwave components in a cryogenic environment at individual frequencies in future experiments.

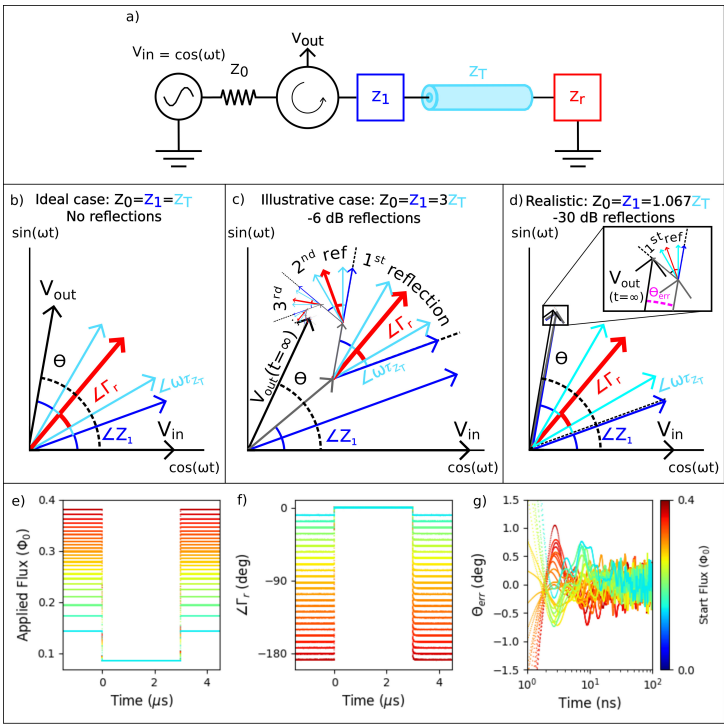


Figure A.1: Effect of microwave reflections on our phase measurement. a) Schematic of our phase measurement considering just two non-ideal components, Z_1 and Z_T, in addition to our tunable resonator Z_r, and an ideal microwave source and circulator each with impedance Z_0. b) Phasor diagram of the ideal case where $Z_0 = Z_1 = Z_T$ and there are no unintended reflections (e.g. the only reflection is off of Z_r). $\mathbf{V_{out}}$ is just a rotation of $\mathbf{V_{in}}$. c) Illustrative case considering very large, -6 dB reflections. It takes $\mathbf{V_{out}}$ many reflections to settle to a final value. d) Realistic case considering -30 dB reflections. In this case the phasor diagram nearly collapses to the ideal case in b), but the first reflection still contributes some error Θ_{err}. e) Informed by the model discussed in a)-d), we use the transducer's DC flux-to-phase calibration curve to choose an array of flux waveforms that correspond to transducer reflection angles that span at least 180 degrees as shown in f). By measuring the phase response of these steps, we vary the relative angle of the reflection and sample over at least one maxima. Finally, in g) we plot Θ_{err} for 50 ns after changing the flux. We estimate Θ_{err} as the difference between each phase waveform and the average of all the phase waveforms. At 3 ns we observe the dominant reflections where the reflection causes a spread in our phase measurement of +1.1 to -0.8 = 1.9 degrees which corresponds to a reflection amplitude of -30 to -36 dB.

The schematic in figure A.1a provides a model for understanding the effect of these microwave reflections. In this model we consider two non-ideal components, Z_1 and Z_T, our tunable resonator Z_r, and a microwave source and circulator each with impedance Z_0. Z_1 may represent any component in the actual system that causes reflections such as SMA connectors, the circulator, HEMT, *etc.* and Z_T is a section of cabling between these components. The actual cryostat wiring consists of more than a dozen components, connections, and cables which we do not model explicitly. We instead look at the aggregate effect of all these reflections and bound their impact on our phase measurement.

We first consider the ideal case where $Z_0 = Z_1 = Z_T$ and there are no unintended microwave reflections—only the reflection off of the tunable resonator Z_r at $\angle\Gamma_r$. In figure A.1b we use a phasor diagram to plot the progression of the probe tone $\mathbf{V_{in}}$ as it first leaves the microwave source, reflects off our transducer and finally propagates to our digitizer as $\mathbf{V_{out}}$. In this picture, our homodyne detector measures the magnitude and phase of $\mathbf{V_{out}}$. Since there are no reflections and all of these components are assumed to be lossless, $|\mathbf{V_{out}}| = |\mathbf{V_{in}}|$ and $\hat{\mathbf{V}}_{\mathbf{out}} \cdot \hat{\mathbf{V}}_{\mathbf{in}} = \cos(\Theta)$ where Θ is the sum of the accumulated phases. In our system, all the component impedances are constant except for when we change Z_r by applying flux. This means that for a fixed probe tone frequency, $\Theta = \angle\Gamma_r + const.$, and the constant term gets rolled into the global phase offset of the homodyne measurement leaving only the phase we are trying to measure.

If there are any mismatched impedances, microwave reflections complicate the transducer's transient response. In A.1c we consider the illustrative case of very large, -6 dB,

reflections with $Z_0 = Z_1 = 3\,Z_T$. As $\mathbf{V_{in}}$ leaves the generator and passes through Z_1, it reflects off of Z_T. This reflection effectively shifts the tunable resonator's response from the origin since half of $\mathbf{V_{in}}$ reflects back to the homodyne detector without ever interacting with Z_r. After this initial reflection, multiple reflections occur between Z_1 and Z_r until the steady-state standing wave ratio is reached and $\mathbf{V_{out}}$ stabalizes to $\mathbf{V_{out}}(t = \infty)$. If we consider changing $\angle\Gamma_r$ from one angle to another by applying a flux pulse to the tunable resonator, these reflections propagate through the system nudging the phase of $\mathbf{V_{out}}$ a little bit in time every 2τ where τ is the propagation time through the transmission line Z_T. With -6 dB reflections, it is clear from our diagram that $\mathbf{V_{out}}$ is still changing appreciably after 5 or more reflections. In figure A.1d we plot the more realistic case of -30 dB reflections with $Z_0 = Z_1 = 1.067\,Z_T$. In this case if the first reflection adds perpendicular to $\mathbf{V_{out}}$, the transient phase error Θ_{err} is given by $\arctan(10^{-30\,dB/20\,dB}) = 1.8$ degrees. Second reflections are negligible because a second perpendicular reflection from the same component would induce an error of $\arctan(10^{-60\,dB/20\,dB}) = 0.06$ degrees, well below the 0.25 degree noise floor of our homodyne detector.

We can now apply this reflection model and use the tunable resonator transducer to estimate the amplitude and position of the dominant reflection. In the actual microwave measurement chain there are likely a number of small reflections with unknown propagation times and reflection angles. However, we have control of $\angle\Gamma_r$ which is always present in the sum of angles that determine the relative angle of the reflection. If we rapidly change from some initial angle $\angle\Gamma_{r,i}$ to some final angle $\angle\Gamma_{r,f}$ then there will be

a transient error as the first reflection propagates through the system. If we vary the starting angle $\angle\Gamma_{r,i}$ over 180 degrees while holding $\angle\Gamma_{r,f}$ fixed, then regardless of the propagation time and reflection angles in the system, we will sample at least one, and maybe both of the reflection angles that maximize Θ_{err}. Ideally we would sample over 360 degrees, but we can only vary the transducer's reflection angle over about 200 degrees. By sampling over 180 degrees we can bound Θ_{err} within a factor of two.

In figure A.1e we plot an array of flux waveforms that we apply to the tunable resonator to achieve the reflection angle waveforms in figure A.1f based on our transducer's DC calibration (figure 4.2 in the main text). The color scale in e,f and g is the same and corresponds to the initial flux value for each waveform. We choose to end all of the flux waveforms near $0.08\,\Phi_0$ where the transducer is relatively insensitive to flux—this means that the small amplitude settling from our imperfect flux transfer function will produce a negligible change to the reflection angle and allow us to distinguish reflections from flux settling. In figure A.1g we plot Θ_{err} (estimated as the deviation of each $\angle\Gamma_r$ waveform from the average of all the $\angle\Gamma_r$ waveforms) for 50 ns after changing the reflection angle. At t = 3 ns the spread in the waveforms is largest and ranges from +1.1 to -0.8 degrees and we continue to see features above the noise until about 20 ns. This indicates that the dominate reflection in our system occurs approximately 1.5 ns from the tunable resonator and that this reflection has an amplitude of -30 to -36 dB corresponding to Θ_{err} =0.85 to 1.9 deg, depending on whether we assume Θ_{err} is measured over 1 or two maximums.

In summary, reflection in the microwave side of this experiment produce time-dependent

phase changes that are a source of systematic error in our transducer-based flux measurement. We use the tunable resonator to estimate the dominant reflection to be 1.5 ns from the tunable resonator with a magnitude of -30 to -36 dB, resulting in up to 1.8 degrees of phase error. These reflections settle out in about 20 ns due to the signal propagation time between the tunable resonator and the oscilloscope. This technique could also be used to characterize microwave components placed in the homodyne measurement chain anywhere from room temperature down to the cryogenic tunable resonator.

A.2 Qubit detuning spectroscopy

One way to characterize on-chip flux settling is to perform frequency detuning spectroscopy with qubits. A typical pulse sequence is shown in figure A.2a. The qubit is first prepared in the ground state. At time t = 0 a frequency detuning flux pulse is applied to change the qubit frequency. After a variable amount of time we drive the qubit with a microwave pulse to try and excite the qubit before detuning back to the qubit readout frequency for measurement. By varying the frequency of the microwave pulse and the delay time between the start of the flux pulse and the microwave pulse and frequency of the microwave pulse, we are able to measure the qubit frequency in time and map that back to an applied flux.

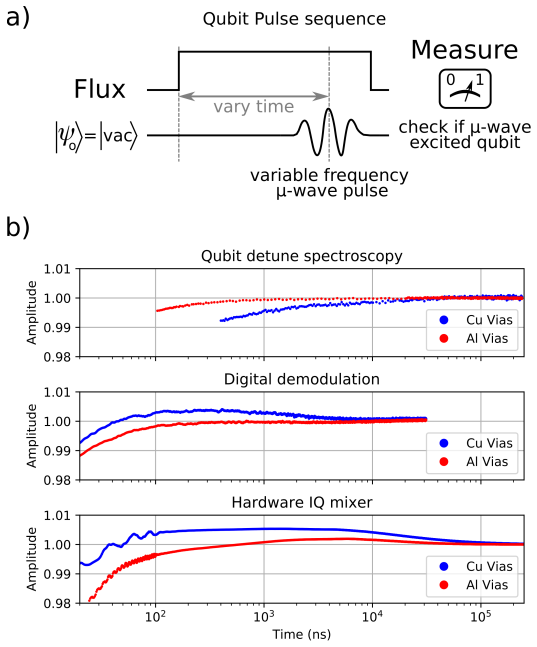


Figure A.2: Measurement of flux bias line settling comparing qubit detuning spectroscopy to hardware and digitally demodulated data from the tunable resonator. a) Typical control pulse sequence for a qubit detuning spectroscopy measurement. The qubit is initially prepared in the ground state. At t = 0 a frequency detuning flux pulse is applied to the SQUID to change the qubit frequency. After a variable time delay, we drive a microwave pulse to try and excite qubit, and then to measure the state of the qubit. By repeating this process for different delays and microwave frequencies, we are able to identify the qubit frequency as a function of time and scale this back to an applied flux using calibration data linking bias to qubit frequency. b) We plot the measured response of both the three layer aluminum PCB with aluminum vias (red) and the two layer PCB with copper vias (blue). In each case, we are able to reproduce very small additional settling amplitude of about 0.005 over 10's of μs. While the use of hardware demodulation allows us to measure at longer times, sampling a DC step response subjects us to the DC settling response of our oscilloscope which imparts a settling of 0.002 over 10's of μs even with the good aluminum via package. The fact that the aluminum via package has a very flat response in both the qubit data and digitally demodulated tunable resonator data indicates that our AWG has a very flat output. Finally, one may note that the copper via device undershoots in the qubit data but overshoots with the tunable resonator data (middle and bottom). This is likely due to the fact that the on-chip layout of these lines was not identical between the qubit and tunable resonator devices.

In figure A.2b we plot the measured flux step response of the nominally good three-layer aluminum PCB with aluminum vias and the two-layer aluminum PCB with copper vias that exhibits a settling amplitude of about $5 \cdot 10^{-3}$ over $10 - 100\,\mu$s. The key result is that regardless of the measurement technique used, we are able to reproduce a similar difference in settling behavior between these two PCBs despite it being less than 1%. The next observation is that the good aluminum via package (red) has a very flat settling in both the qubit and digital demodulation datasets where-as the hardware demodulated data shows a settling behavior of $2 \cdot 10^{-3}$ over 10's of μs even on the aluminum via package. While using hardware demodulation allows us to sample at a lower rate and measure at longer times (accommodating the limited oscilloscope memory), we are subjected to the oscilloscope's impulse response, which is not the case for both the qubit and digitally demodulated datasets. The fact that both the qubit and digitally demodulated data have a very flat response indicates that our AWG performs well with respect to settling. Finally, notice that the copper via data under-shoots in the qubit data and overshoots in the tunable resonator data for both digital and hardware demodulation. This is likely because the exact on-chip bias line routing differences between the qubit data and the tunable resonator data; while we have not studied this carefully, we have seen evidence across multiple devices in different packages that certain on-chip routings overshoot while others undershoot. This is not necessarily surprising given the stark difference in settling between the long and short flux bias lines discussed in figure 4.4 in the main text.

A.2.1 Device comparison

Here we provide a micrograph image comparing the long bias line tunable resonator device to the qubit device used in this chapter.

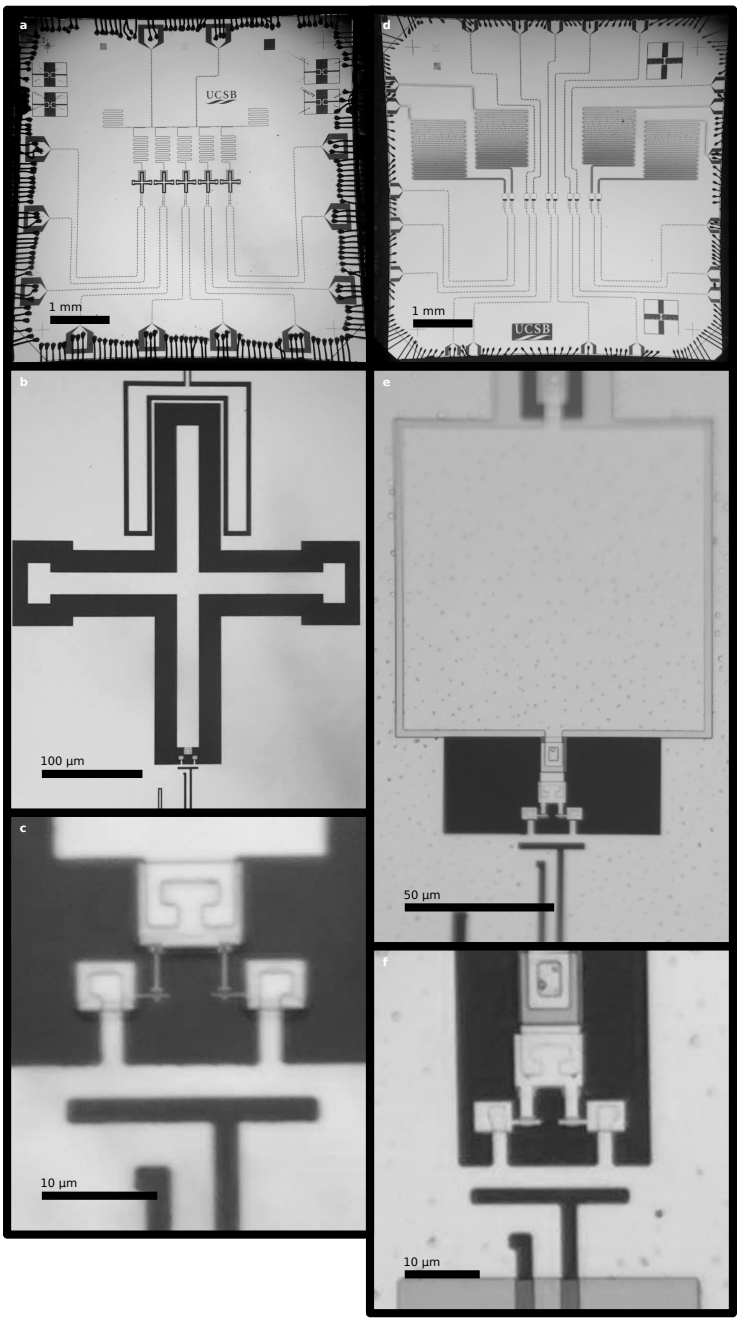


Figure A.3: Micrograph images of the qubit chip (left) and tunable resonator chip (right) used in this chapter. The SQUID layouts are nominally identical. The tunable resonator has a larger Josephson junction to allow a larger critical current. To keep the resonant frequency of the tunable resonator near the qubit frequency, we need a much larger capacitance (qubit: 80 fF, tunable resonator: 4 pF) so we use a parallel plate capacitor.

Appendix B

Superconducting Interconnect Appendices

B.1 Electrical characterization measurement setup

Bounding the resistance of a suspected superconducting device requires the ability to very accurately measure the excitation current and the resulting voltage drop across the device. Even when care is taken to use appropriate signal wiring and grounding, as shown and described in Figure B.1, DC based measurements are subject to thermoelectric voltages, broadband noise, and measurement ranges optimized for non-zero resistance materials where finite voltages are expected. An AC excitation and a digital lock-in amplifier can be used to mitigate these effects, but the resulting measurement is not without difficulties. A lock-in amplifier implements mixing and filtering to extract the

signal amplitudes both in phase and 90-degrees out of phase with a reference tone at the specific reference frequency. Both the in phase voltage (V_x) and quadrature voltage (V_y) across a device may be extracted if the sinusoidal excitation is used as the lock-in amplifier reference.

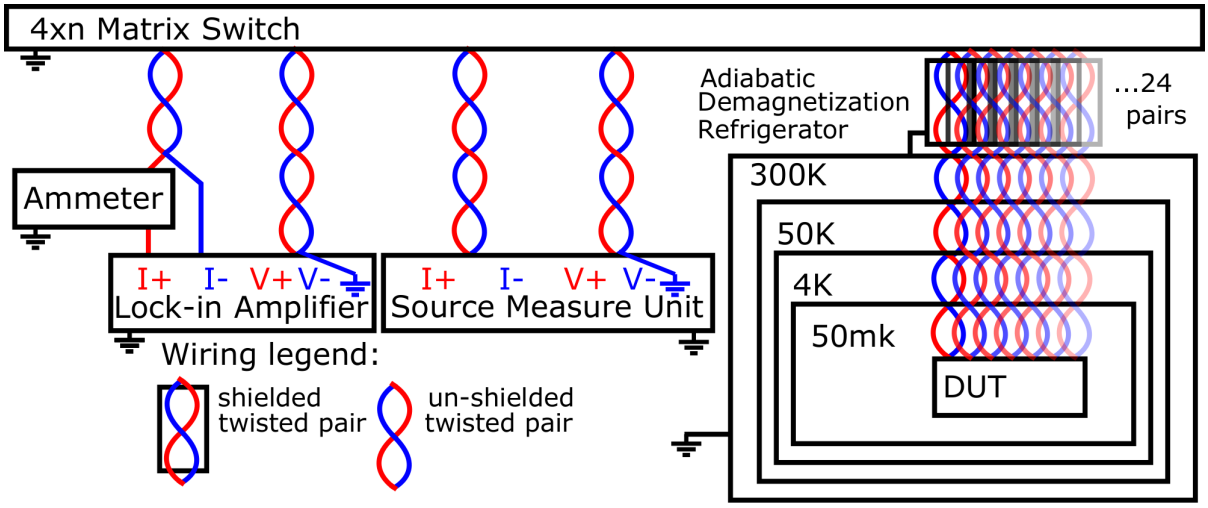


Figure B.1: Schematic of the measurement setup. A 4 by n matrix switch is used to route two sense and two excitation lines from various measurement equipment, including a lock-in amplifier and a source measure unit, to the bump bond devices. Both the measurement equipment and DUT are connected to columns of the matrix switch and the rows are used to connect any column to any other column. The measurement equipment chassis are all grounded to a common surge protector. The twisted pair shielding is grounded at the ADR, and floating at the matrix switch. The four-wire measurement ground is provided by the negative excitation terminal of either lock-in amplifier or source measure unit.

Figure B.2 shows a model of the 4-wire measurement circuit used to perform a bounding resistance measurement. The voltage excitation signal is an adjustable frequency sine-wave generator with a $50\,\Omega$ output impedance provided by the lock-in amplifier, a Stanford Research Systems model SR830. An ACrms ammeter, a Keysight Technologies model 34461A, is placed in-line with the positive voltage lead of the sine-wave generator

to measure the excitation current, which is set by the amplitude of voltage waveform and the approximately 50 Ω lead resistance in the I_+ and I_- leads. This lead resistance is dominated by the niobium titanium wiring used in our cryostat all the way from 300 K to the 50 mK stage and varies by 10-20% channel-to-channel. This lead resistance variation is why the excitation current is measured directly with the ammeter rather than inferring it from the excitation voltage.

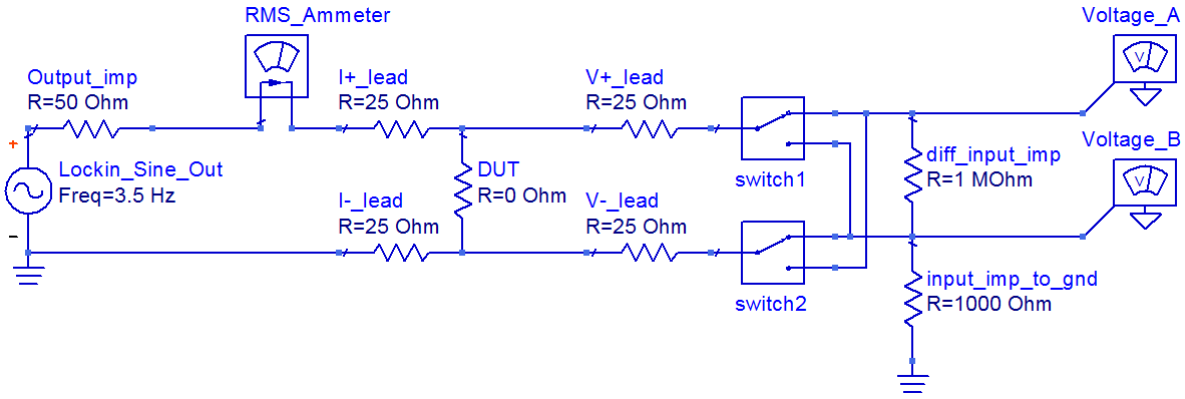


Figure B.2: A model of the 4-wire measurement circuit used to perform the bounding resistance measurement.

It is important to note that this wiring configuration results in a common-mode voltage at the sample approximately equal to half of the excitation voltage due to the voltage divider created by the excitation leads. The lock-in amplifier used here has a common mode rejection ratio (CMRR) of 100 dB meaning that common mode (CM) voltages may leak into the differential voltage measurement attenuated by 10^5. Without further common mode compensation (and independent of the excitation voltage) the 50 Ω lead resistance and CMRR specification would limit the measurement accuracy as follows. The excitation current is approximately equal to the excitation voltage divided

by the total of the voltage source output impedance and the sum of the two excitation leads:

$$I_{ex} = V_{ex}/R_{lead+source} \tag{B.1}$$

Since the excitation leads are approximately equal, the common mode voltage on both sense leads will be approximately $V_{ex}/2$ and the lock-in amplifiers CMRR specifies how much of this voltage may leak into its differential voltage measurement:

$$V_{cmleakage} = V_{cm} * CMRR = \frac{V_{ex}}{2} * CMRR = V_{ex} * 5 * 10^{-4} \tag{B.2}$$

Combining B.1 and B.2 we find that the common mode leakage and lead resistance would limit our measurement to a minimum of $500\,\mu\Omega$.

$$R_{min} = \frac{V_{cmleakage}}{I_{ex}} = \frac{V_{ex}}{2} * CMRR * (\frac{V_{ex}}{R_{leads+source}})^{-1} = 500\,\mu\Omega \tag{B.3}$$

Compensation for this common mode voltage leakage is accomplished by taking two voltages measurement using the switches shown in figure B.2. While holding the excitation signal constant, the switches are used to reverse the polarity of the voltage sense leads and two measurements are recorded:

1. Differential Voltage_A-Voltage_B + common mode leakage, and

2. -(Differential Voltage_A-Voltage_B) + common mode leakage

The sum of these two measurements is two times the common mode leakage and the dif-

ference is two times the differential voltage of interest. Figure B.3 shows these two voltage measurements as well as the computed common mode (CM) and differential voltages for the V_x and V_y signals measured across a 1620 bump structure. This data confirms that the lock-in amplifier is meeting both its common mode rejection specification of >100 dB, as well as its input noise specification of $6\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$–since a 0.3 s time-constant was used, the input noise should be $< 11\,\mathrm{nV}_{rms}$.

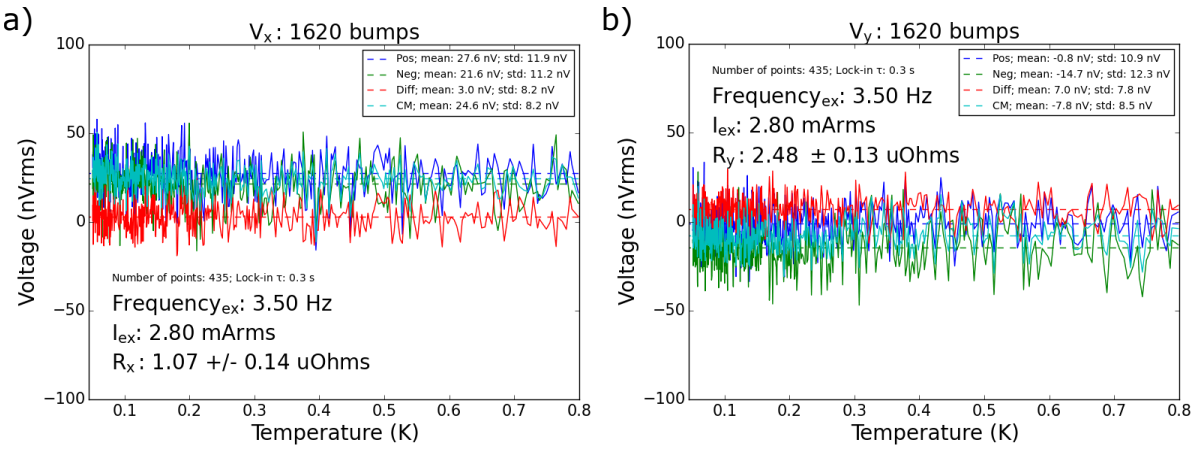


Figure B.3: Typical Vx and Vy voltages traces with positive and negative lead polarity. The differential voltage computed from the difference of the positive and negative polarity measurements is near zero and the common mode voltage computed from the sum is consistent with the lock-in amplifiers CMRR specification of 100dB (we usually see 110-140 dB).

Since the lock-in amplifier is making an AC measurement, care must be taken to make sure that parasitic inductances and capacitances do not affect the measurement. Firstly, it is very important to use twisted pair wiring for at least one pair of, and preferably both, the sense +/- leads and the excitation +/- leads (grounded shielding should also be used where possible to further reduce mutual inductances of the sense

and excitation leads and to further reduce electromagnetic noise pick up). By utilizing twisted pairs in this configuration, the mutual inductance between these leads is reduced considerably–without twisted pairs it is easy to end up with a several μH or more mutual inductance between the sense and excitation leads which will end up looking like an in-phase voltage (or real resistance). Secondly, the voltage signal from the inductance of the sample as well as other parasitic inductors and capacitors should be proportional to the frequency of the excitation voltage. To reduce the impact of such signals and parasitics, measurements were made using a low frequency excitation signal, typically $<$ 10 Hz. Finally, we typically find that the resulting differential voltage is proportional to both the excitation voltage as well as the excitation frequency indicating that the signal we are measuring is due to a system parasitic and is not just electronics noise. The frequency dependence in particular hints that this load is primarily not resistive, but even in taking the conservative approach of assuming it is all resistive, this measurement system is able to limit the resistance of a series chain of 1620 bumps to be several $\mu\Omega$. The fact that the differential voltage measured across 1620 bumps is the same as for 2 bumps is further evidence that the measurement is dominated by a cabling parasitic and not an actual resistance in the sample, or even the samples inductance.

Measuring the critical current of these interconnects in an Adiabatic Demagnetization Refrigerator (ADR) required some optimization to run efficiently. An ADR uses a helium compressor to cool a superconducting magnet and sample stage down to 3-4 K. Then, to cool the sample down to 50 mK, the current in the superconducting magnet is ramped up

over 10-15 minutes to align magnetic dipoles in a salt crystal. After a 30-45 minute soak time in this magnetic field, the salt crystal and sample stage are thermally disconnected from the rest of the system and the magnetic field from the superconducting magnet is ramped down over 10-15 minutes. When the magnetic field approaches zero, the dipoles in the salt crystal begin to mis-align, pulling heat out of the system and cooling the sample stage down to about 50 mK. If too much heat is added to the system then the 60-75 minutes magnet cycle must be repeated to cool back down. The superconducting devices tested here have a critical current $> 25\,\mathrm{mA}$, and once a subsection is driven normal by exceeding the critical current, the sample stage of the cryostat heats up from 50 mK to 3 K in a about a second if the current is not reduced. In order to efficiently characterize many devices, care was taken to avoid unnecessarily heating the cryostat.

In order to limit the heat dissipation of the sample in the cryostat, a Keysight Technologies model B2901A source measure unit (SMU) was used. A SMU is a combination source (with a configurable current or voltage set point) integrated with a meter (configurable for current and/or voltage). In this case, a current set point is used and the voltage across the sample is measured–as the current is increased, if the voltage across the sample jumps above the noise level, then the sample has transitioned to a normal state. SMUs are fast, accurate, and offer a number of features that enabled us to make hundreds of critical current measurements in a single ADR magnet cycle. Firstly, this SMU offers pulsed operation where the source provides a timed current pulse and the measurement aperture is synchronized to occur just after the excitation has settled. We

found that we were able to achieve good results using just a 2 ms long current pulse with a 0.4 ms measurement aperture window. Furthermore, this SMU offers a voltage protection feature where the source terminals are physically disconnected inside the unit with a relay if the source compliance condition (compliance voltage in this case) is reached. Since a superconducting material is being measured, the compliance voltage at the device should be 0 V. A compliance limit of 5 mV was set, and we found that these results were in good agreement with measurement where we did not pulse the excitation.

B.2 Material and interface characterization

Aluminum is deposited using e-beam evaporation in a vacuum chamber with a base pressure of 1e-7 mBar. 100 nm of aluminum is deposited at a rate of 1 nm/s. Structures were patterned and etched using standard lithographic techniques and $BCl_3 + Cl_2$ chemistry in an inductively coupled plasma etcher. (Other samples have yielded using both wet etches and lift-off defined structures.)

The titanium nitride under bump metalization (UBM) is used as a diffusion barrier between indium and aluminum as both are known to be very reactive metals [92]. To achieve a dense film with low oxidation and Tc above 3 K, we employ a substrate bias during deposition [112]. A 50-80 nm titanium nitride film is grown using a reactive sputter (150 W power) from a pure titanium target in 3 mTorr of argon and nitrogen (48 sccm and 1.75 sccm flows, respectively).

The resulting films are found to be nearly stoichiometric, but slightly nitrogen rich

using X-ray photoelectron spectroscopy (XPS) (Figure B.4) and Rutherford backscattering spectroscopy (RBS) (Table B.1). Moreover interdiffusion of aluminum into the titanium nitride is absent.

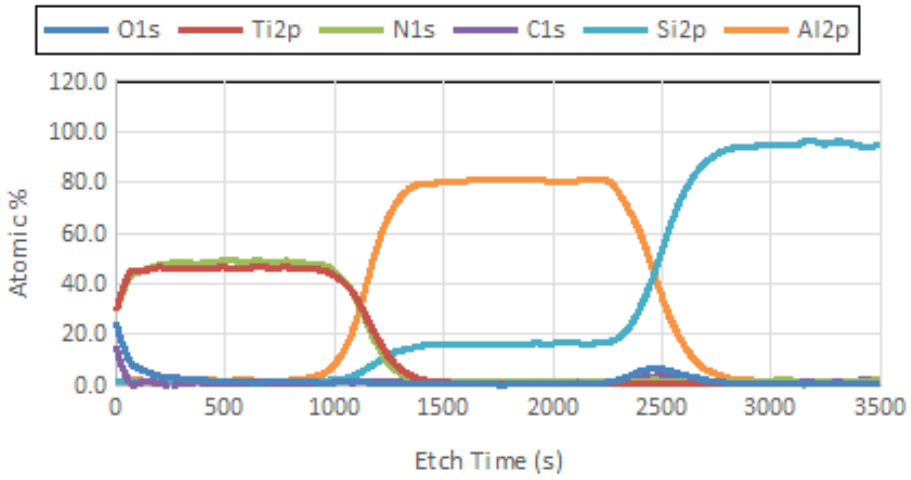


Figure B.4: XPS data for a layer of titanium nitride on aluminum on a silicon substrate.

Table B.1: RBS data for a 320 Å titanium nitride layer on 1000 Å of aluminum on a silicon substrate.

	"RBS" Thickness [Å]	Atomic Concentration [at%]						Assumed Density [at/cc]
		N	**Si**	**Ti**	**Al**	**W**	**Ar**	
Layer 1	320	53.5	-	46.5	-	-	-	1.07E23
Layer 2	10	-	-	-	29.7	3.7	66.6	3.79E22
Layer 3	1000	-	-	-	100	-	-	6.02E22
Bulk	-	-	100	-	-	-	-	5.00E22

Since titanium nitride is employed as a diffusion barrier, it is deposited as square pads beneath the indium bumps. These pads are defined in lift-off, using a single layer of positive photoresist and a MIB (AZ Developer) developer to prevent etching and roughening of the underlying aluminum during developer rinse. To make good electrical contact between the titanium nitride and the underlying aluminum, the patterned aluminum wafer is ion milled in-situ before sputter deposition. Mill parameters are shown in B.3 (120 s

mill time).

After the titanium nitride is lifted off, the wafer is patterned again using a lift-off polarity and a thick positive resist. Circular apertures are opened using a MIF developer (since the aluminum is encapsulated by corrosion resistant titanium nitride). The wafer is loaded into a thermal evaporator with a base pressure below 1E-7 Torr. To remove any contaminants and insulating oxides, the wafer is ion milled *in situ*, then allowed to cool on the water cooled chuck (held at 0 C). Indium is deposited at rates exceeding 2 nm/s to prevent a constriction of the lithographically defined apertures by crystallite growth. Figure B.5 is a SEM of typical crystallite growth that occurs when slow deposition rates are used. Indium lift-off is performed in a heated NMP bath.

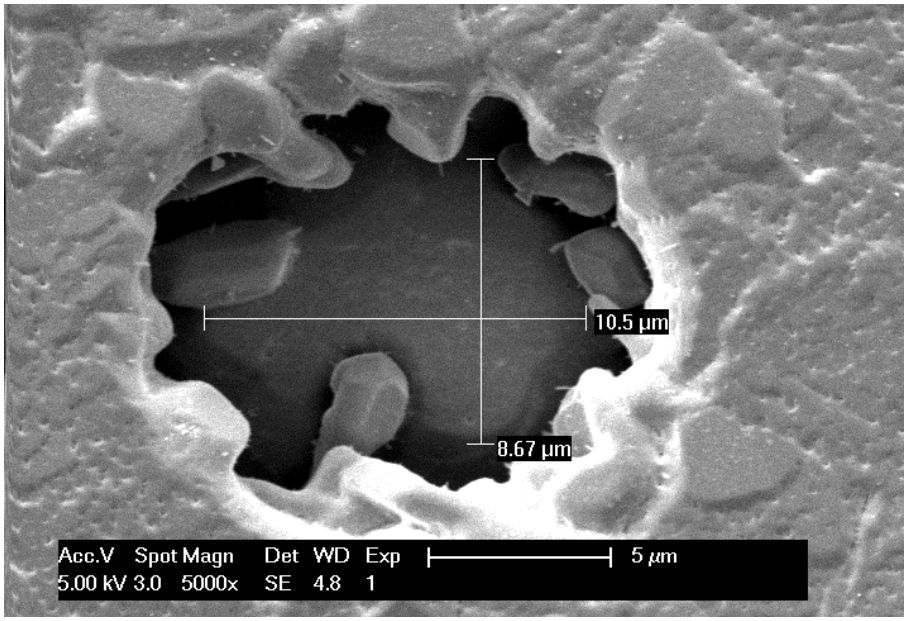


Figure B.5: SEM image of indium crystallite growth over a 15 μm diameter hole during a slow (<1 nm/sec) indium deposition.

The entire material stack up has been characterized using focused ion beam (FIB)

cross sections and electron energy loss spectroscopy (EELS), as shows in figure B.6, to determine the composition of the layers and, most importantly, their interfaces. Crucially, no indium-aluminum interdiffusion exists across the titanium nitride barrier. However, intermittent oxide contamination (up to 15% by atomic percent) at the titanium nitride/indium interface and titanium nitride/aluminum interface has been measured on various samples, although this seems to have little affect on yield or critical current.

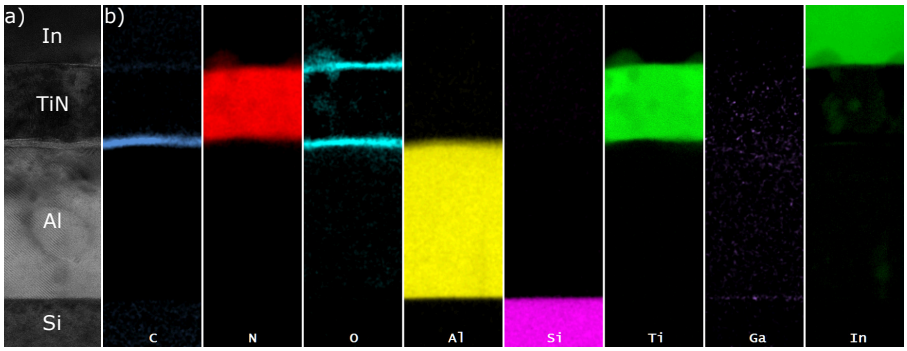


Figure B.6: Electron energy loss spectroscopy of a focused ion beam cross-section of one interconnect. a) SEM image of the focused ion beam cross section of an indium bump on a titanium nitride diffusion barrier with aluminum base wiring on a silicon substrate. b) Electron energy loss spectroscopy of the sample shows in a). This confirms the titanium nitride to be a sufficient diffusion barrier as there is no indium to aluminum contamination. Oxide contamination can be seen at both titanium nitride interfaces, but this does not seem to affect the critical current of these interconnects. The carbon present at the aluminum/titanium nitride interface is due to redeposition of lift off photoresist during the ion mill of the aluminum a and the gallium present is from the focused ion beam used to cut the cross section.

B.3 Ion mill parameters

Table B.2: *In situ* ion mill parameters used to clean aluminum surface before depositing titanium nitride. Ion mill time is 120 s.

	Cathode	Discharge	Beam	Accelerator	Neutralizer	Emission
Voltage (V)	7.3	40.0	399	79	18.9	n/a
Current (A)	10.8	0.48	0.055	0.0031	17.2	0.118

Table B.3: *In situ* ion mill parameters used to clean titanium nitride surface before depositing indium. Ion mill time is 90 s.

	Cathode	Discharge	Beam	Accelerator	Neutralizer	Emission
Voltage (V)	9.3	40.0	600	120	10.4	n/a
Current (A)	14.9	1.52	0.12	0.004	0.0122	0.116

Appendix C

fSim Gate Appendices

C.1 fSim control model

A generic representation of a Fermionic Simulation (fSim) gate corresponding to a two-qubit photon conserving unitary requires five parameters. We may separate out the single and two-qubit parameters as follows: a $|01\rangle \leftrightarrow |10\rangle$ swap angle, θ, a $|11\rangle$ state conditional phase, ϕ, and three single qubit phases, Δ_+, Δ_-, and $\Delta_{-,off}$ yielding a generic fSim parameterization,

$$\mathrm{fSim}(\theta, \phi, \Delta_+, \Delta_-, \Delta_{-,off}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i(\Delta_+ + \Delta_-)} \cos\theta & -ie^{i(\Delta_+ - \Delta_{-,off})} \sin\theta & 0 \\ 0 & -ie^{i(\Delta_+ + \Delta_{-,off})} \sin\theta & e^{i(\Delta_+ - \Delta_-)} \cos\theta & 0 \\ 0 & 0 & 0 & e^{i(2\Delta_+ + \phi)} \end{pmatrix} \tag{C.1}$$

We are interested in performing a two-qubit gate, which is independent of the single-qubit rotations. Therefore, we can focus on the matrix where Δ_+, Δ_-, and $\Delta_{-,off}$are all zero, leading to the notation,

$$\mathrm{fSim}(\theta,\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -i\sin\theta & 0 \\ 0 & -i\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix} \tag{C.2}$$

used to designate an arbitrary gate within the excitation preserving subspace.

C.2 fSim gate numerics

The qubit dynamics presented in the main paper (Figure 6.2) are well described by numerics simulating two interacting qutrits (e.g. a pair of coupled three-level anharmonic oscillators) evolving with a time dependent detuning, $\Delta(t)$, and coupling, $g(t)$. We truncate the full two-qutrit Hamiltonian limiting our simulation to states with 1 or 2 excitations. Operating with the basis $|01\rangle, |02\rangle, |10\rangle, |20\rangle, |11\rangle$, the Hamiltonian describing the system is given by:

$$\mathrm{H(g},\Delta,\eta) = \begin{pmatrix} 0 & g & 0 & 0 & 0 \\ g & \Delta & 0 & 0 & 0 \\ 0 & 0 & \Delta & \sqrt{2}g & \sqrt{2}g \\ 0 & 0 & \sqrt{2}g & 2\Delta+\eta & 0 \\ 0 & 0 & \sqrt{2}g & 0 & \eta \end{pmatrix} \tag{C.3}$$

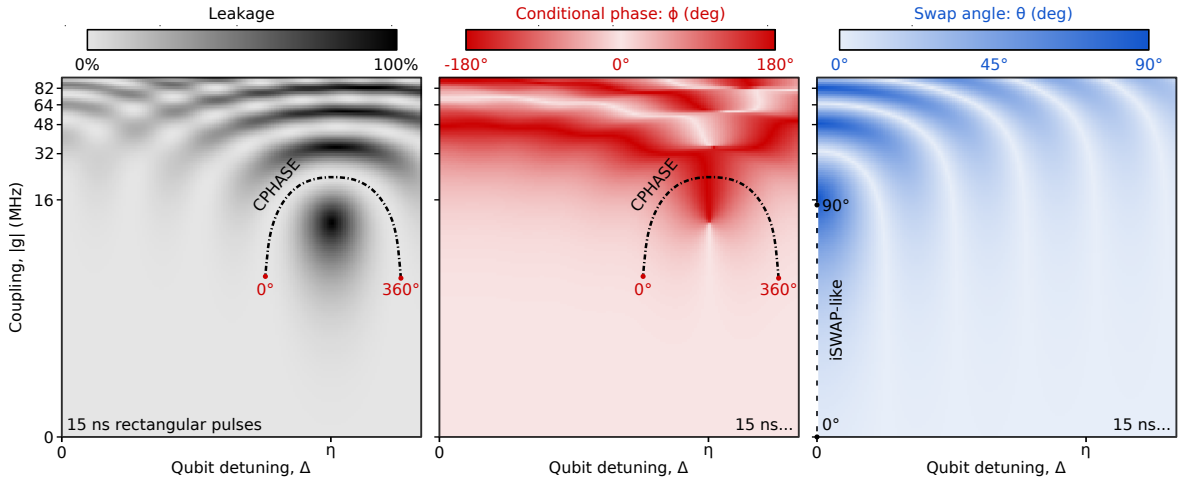


Figure C.1: Numeric simulation of two interacting qutrits reproducing the data from our experiments in Figure 6.2 of the main text. We simulate qubits with a fixed nonlinearity (240 MHz) with 15 ns long rectangular control pulses defining the qubit detuning, Δ, and their coupling, g.

where η is the nonlinearity of each qubit, which we assume is the same for both qubits (240 MHz). Using this model, we may estimate the unitary operation enacted by arbitrary time-domain control of the coupling strength and the qubit detuning by discretizing these time domain control waveforms and performing a time ordered integral of H(t).

In Figure C.1 we qualitatively reproduce the experimental results in Figure 6.2 by simulating 15 ns rectangular control pulses defining both g and Δ. In Figure C.2 we illustrate the broadening effect that using shorter pulse lengths has on the Rabi interactions of both the $|01\rangle \leftrightarrow |10\rangle$ and $|11\rangle \leftrightarrow |02\rangle$ interactions by simulating rectangular pulses that are 10 ns, 15 ns, and 20 ns long. In Figures C.2 and C.3, we have omitted points where the leakage exceeds a 1% threshold which identifies the parameter space where we can perform fSim gates with low error. Experimentally we have chosen to implement our CPHASE gates with 13 ns long rectangular pulses with a 1 ns pad on either

side—when we made the gate length shorter, leakage increased (data not shown). Here, in Figure C.2a, we qualitatively see that the width of the 1% leakage band where we perform the CPHASE gate begins to pinch off and the $|2\rangle$ state Rabi interaction reaches all the way to the on-resonance iSWAP-like parameter space (dotted white line) when the gate length is 10 ns. Both these results qualitatively reproduce what we observed experimentally when attempting iSWAP-like gates shorter than 11 ns or the CPHASE gate shorter than 13 ns. Finally, in Figure C.3 we simulate the effect of smoothing the control pulses by simulating 20 ns long coupler pulses that are rectangular, rectangular with 3 ns Gaussian smoothing, and cosine shaped (all detuning pulses are rectangular and have the same length). Here we see that smoothing reduces the extent of leakage from the second and third $|11\rangle \leftrightarrow |02\rangle$ swap lobes expanding the available low-error fSim control space. This indicates that pulse smoothing may be an important consideration of any future fSim implementation that aims to perform an arbitrary fSim using a single coupler pulse instead of the two discrete rectangular pulses we have used in this work.

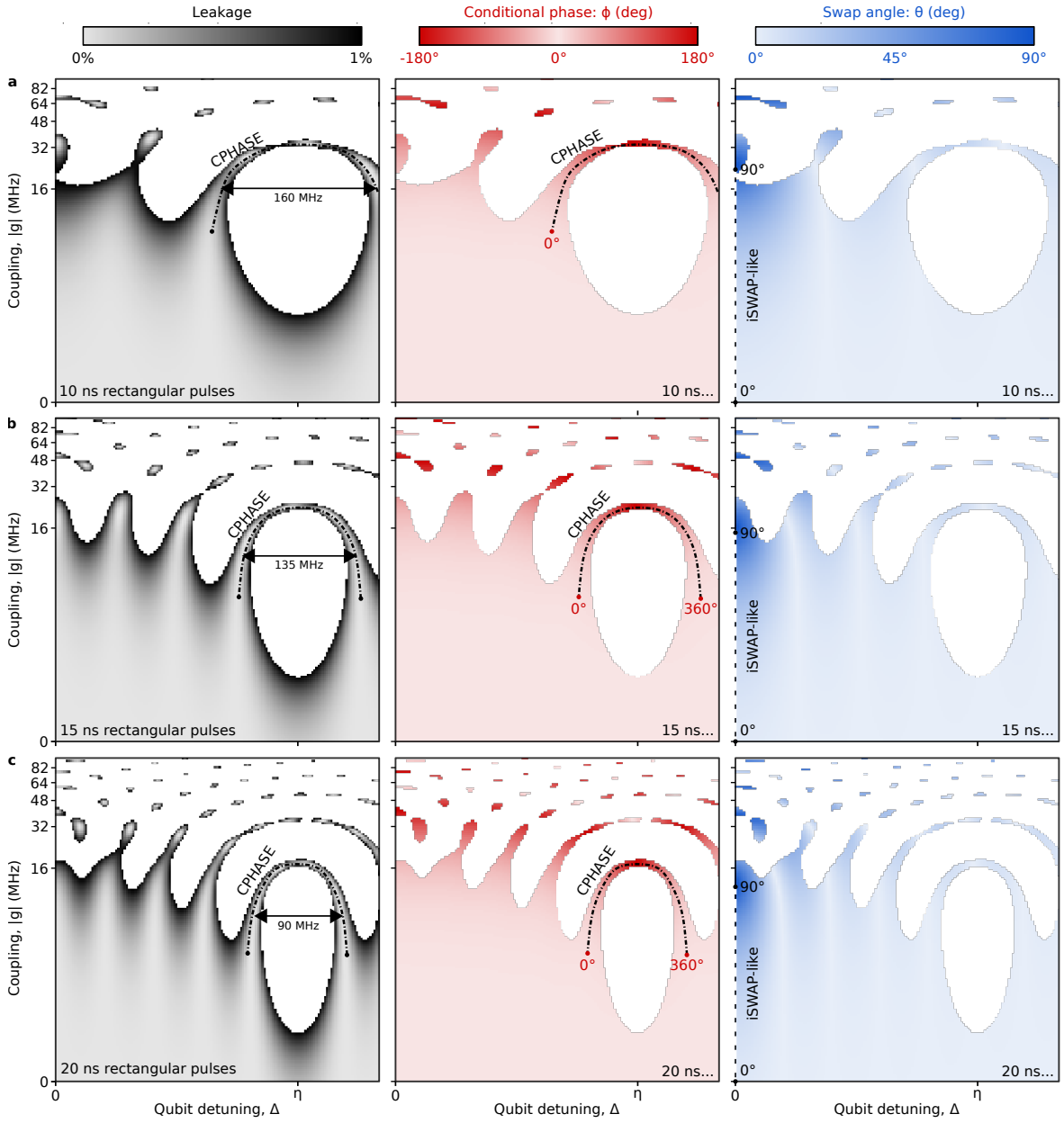


Figure C.2: Numeric simulation of **a**, 10 ns, **b**, 15 ns, and **c**, 20 ns rectangular control pulses showing the fSim parameter space where leakage is less than 1% (white regions are where leakage exceeded this threshold). Experimentally we chose to perform our CPHASE gate with 13 ns long pulses and the iSWAP-like gate with 11 ns control pulses (both of which had 1 ns pads on either side)—as we found that shorter implementations of either gate increased leakage and the overall gate error. Here, these numerics demonstrate that for 10 ns long gates, the low-leakage lobe where we perform the CPHASE gate narrows considerably and the $|2\rangle$ state Rabi interaction reaches the on-resonance iSWAP-like line cut near $\theta = 90°$, both of which agree with our experimental results.

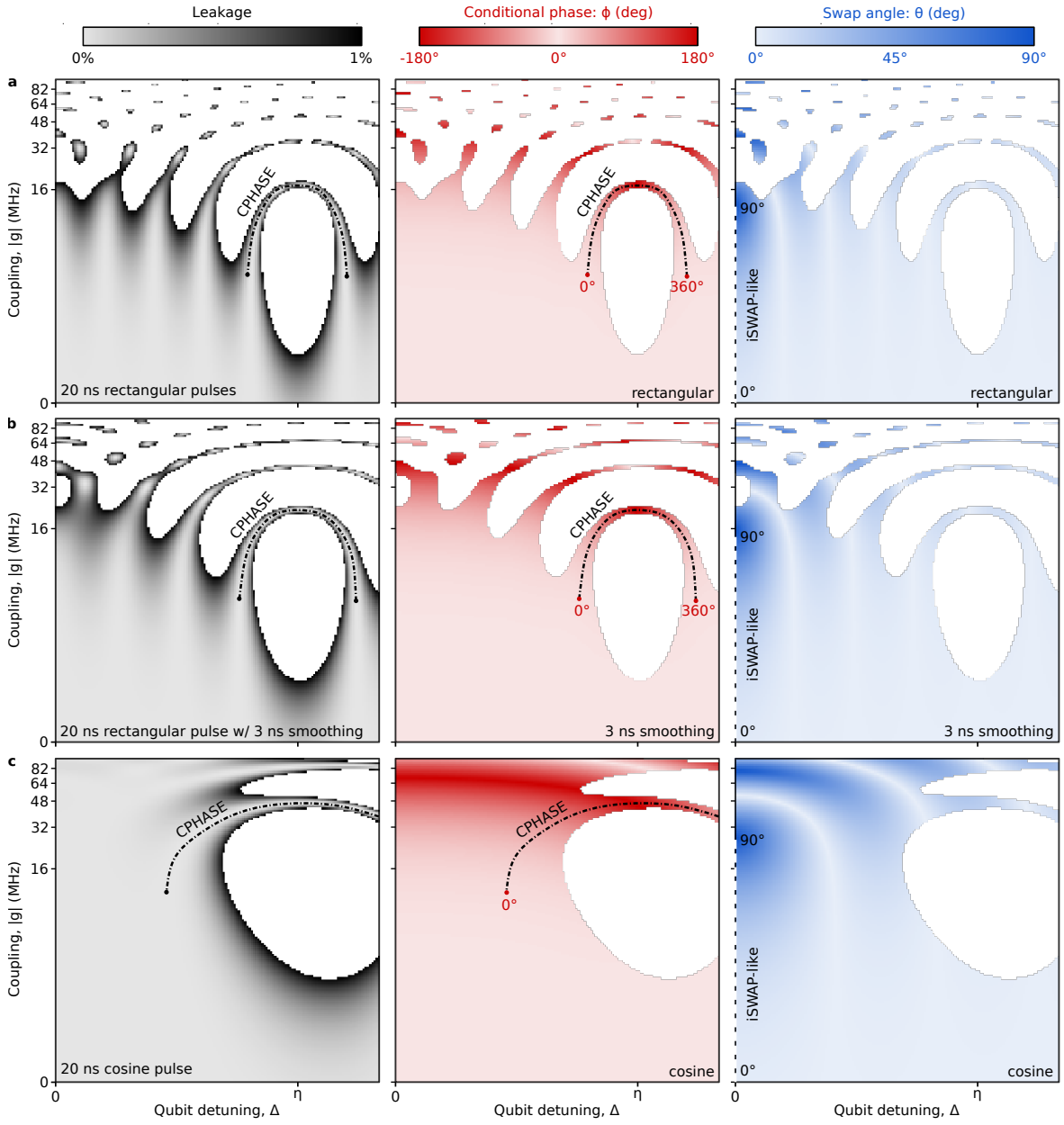


Figure C.3: Numeric simulation of **a**, a 20 ns rectangular coupler pulse, **b**, a 3 ns rise time rectangular pulse, and **c**, cosine coupler pulse showing the fSim parameter space where leakage is less than 1%. We observe that as the coupler pulses become more smooth, the fSim parameters space where leakage is less than 1% expands considerably. This indicated that pulse shaping and or smoothing may play an important role in any future implementation of the fSim gate set that aims to implement the gate set with a single pulse.

C.3 Gate characterization

We use a variety of techniques to characterize the performance of our single and two-qubit gates which we detail in this section. In lieu of full process tomography, we use depth one population based measurements to perform unitary tomography to quickly assess the unitary operation performed by a given set of control pulses. We then turn to benchmarking techniques that amplify gate errors and allow for the characterization of small error rates. We use Clifford based benchmarking to characterize our single-qubit microwave gates and cross-entropy benchmarking (XEB) to characterize our two-qubit entangling gates.

C.3.1 Computing and reporting Pauli error rates

Before jumping in to gate characterization, a quick aside on Pauli error rates. We report Pauli error rates which are independent of the Hilbert space dimension and thus add linearly as the circuit's Hilbert space grows. In the past, many have reported average single and two-qubit error, e_r, as exponential decay constants of a sequence fidelity, $F = Ae^{me_r} + B$ where A and B are fit parameters to compensate for state preparation and measurement (SPAM) errors, m is the number of gate repetitions in the sequence, and e_r is the error per cycle. The Pauli error, e_p, is related to e_r by the dimension of the Hilbert space:

$$e_p = e_r \times \left(1 + \frac{1}{D}\right) \tag{C.4}$$

where $D = 2^n$ is the dimension of the Hilbert space for an n-qubit gate. We note that this results in an increase in the reported error by a factor of 1.5 for single-qubit gates ($n = 1$) and a by a factor of 1.25 for two-qubit gate errors ($n = 2$).

When performing two-qubit XEB, we measure the exponential decay constant per cycle, $e_{r,\text{cycle}}$ where each cycle consists of the application of one single-qubit gate per qubit and one fSim entangling gate involving both qubits. In order to extract the error per fSim gate, we can convert this to a Pauli error per cycle, $e_{p,\text{cycle}}$, and subtract off the two single-qubit Pauli gate errors, e_{p,q_1} and e_{p,q_2}, which we estimate using single-qubit Clifford based randomized benchmarking.

$$e_{p,\text{2q}} = e_{p,\text{cycle}} - (e_{p,\text{q}_1} + e_{p,\text{q}_2}) \tag{C.5}$$

For simplicity, all two-qubit Pauli errors have been computed assuming single-qubit Pauli errors of 7.5×10^{-4} per gate per qubit consistent with our typical single-qubit error rates immediately following a successful run of our standard single-qubit gate calibration procedure (see Appendix C.3.2, below).

$$e_{p,\text{two\_qubit}} = e_{p,\text{cycle}} - (2 \times 7.5 \times 10^{-4}) = e_{p,\text{cycle}} - 1.5 \times 10^{-3} \tag{C.6}$$

C.3.2 Single-qubit coherence and gates

Qubit coherence, in conjunction with gate duration, places a lower bound on both our single and two-qubit gate error rates. In Figure C.4 we characterize T_1 for four qubits

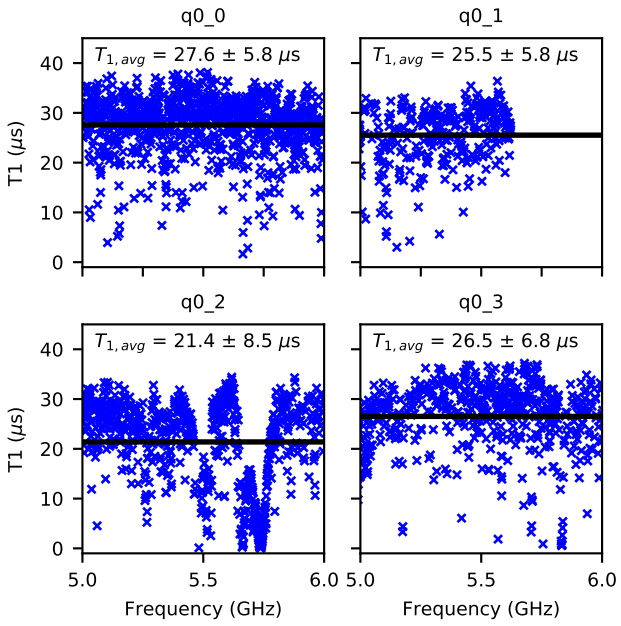


Figure C.4: Swap spectroscopy for four qubits from 5 to 6 GHz characterizing the qubit T_1 as a function of qubit frequency. For all four qubits on this chip over the available frequencies in the range of 5-6 GHz we find an average $T_1 = 25.3 \pm 7.3\mu$s.

over a frequency range of 5 to 6 GHz. To perform this measurement we calibrate single-qubit gates, readout, and flux bias frequency control for a given qubit idle frequency. We then excite the qubit to the $|1\rangle$ state and detune the qubit to another frequency for a variable amount of time before detuning back to the idle frequency for readout. For each detuned frequency, T_1 is extracted as an exponential decay of the population over time, $P|1\rangle \propto Ae^{-t/T_1} + B$, where A and B are fit parameters to compensate for state preparation and measurement errors. We find $T_1 = 25.3 \pm 7.3\,\mu$s averaging data from all four qubits over a frequency range of $5 - 6$ GHz. Since $f_{\max}$ for the second qubit was anomalously low, we averaged data for this qubit from $5 - 5.61$ GHz.

We use single-qubit purity [110] and Clifford-based randomized benchmarking [113,

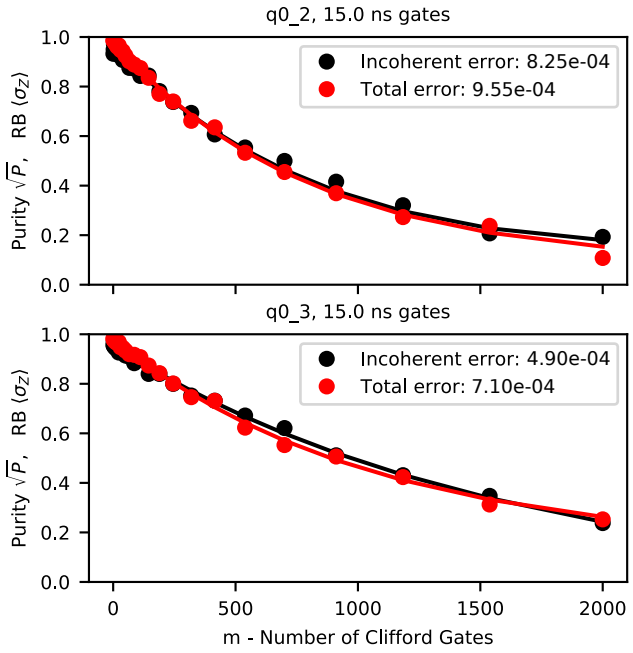


Figure C.5: Representative single-qubit Clifford-based randomized benchmarking results used to characterize the average error of our single-qubit gates. With a typical calibration, the single-qubit Pauli errors for both qubits are usually in the range of $5 - 10 \times 10^{-4}$. When computing the two-qubit gate error from the XEB per cycle error throughout this paper, we assume a moderately conservative error of 7.5×10^{-4} per single-qubit gate.

114] to characterize the average error of our single-qubit gates. In Figure C.5 we present representative results for a pair of qubits demonstrating purity-limited (incoherent error-limited) performance. These gate errors drift over time, but immediately following a successful run of our standard calibration procedure we typically observe single-qubit error rates at or slightly higher than the 7.5×10^{-4} level [58]. As such, we use this estimate in computing two-qubit error rates throughout this paper. These error rates are consistent with the coherence limit, for $\mathrm{T_{gate}} = 15\,\mathrm{ns}$ and $\mathrm{T_1} = 30\,\mu\mathrm{s}$, giving $e_{p,\mathrm{inc}} \approx 1.5 \times \mathrm{T_{gate}}/3\mathrm{T_1}) = 2.5 \times 10^{-4}$, with the remainder of the error coming from leakage and $\mathrm{T_2}$ [115].

C.3.3 Unitary tomography

Section II of the main text describes shallow circuits used to characterize leakage and the two-qubit control parameters, θ and ϕ. Here, we detail the procedure used to directly measure all the non-zero matrix elements composing an arbitrary photon conserving unitary operation and the algebra used to convert these matrix elements into the five fSim control parameters (in Eq. C.1). We use the resulting fSim model to compute the XEB sequence fidelity which we may then use as a cost function to optimize some, or all, of the fSim model parameters.

Table C.1: Summary of the two-qubit unitary tomography measurement sequences. Here, $\{u_{11}, u_{12}, u_{22}, u_{21}\}$ are the complex matrix elements of the two-qubit unitary in $[|01\rangle, |10\rangle]$ subspace. The two additional measurements ($u_{12,\text{excited}}$ and $u_{22,\text{excited}}$) are repeated measurements of u_{12} and u_{22} but with the other qubit placed into the excited state. This additional information is used to construct the conditional phase, ϕ.

Matrix element	Initial state	Measure qubit
u_{11}	(x, 0)	0
u_{12}	(0, x)	0
u_{22}	(0, x)	1
u_{21}	(x, 0)	1
$u_{21,\text{excited}}$	(1, x)	0
$u_{22,\text{excited}}$	(1, x)	1

In order to efficiently characterize the unitary operation performed by a given set of control pulses, we initialize and measure a set of circuits as summarized in Table C.1. If we consider a general photon conserving unitary the non-zero matrix elements will take the form:

$$\mathrm{U} = \begin{array}{c} \begin{array}{cccc} |00\rangle & |01\rangle & |10\rangle & |11\rangle \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & u_{11} & u_{12} & 0 \\ 0 & u_{21} & u_{22} & 0 \\ 0 & 0 & 0 & u_{33} \end{pmatrix} \end{array} \begin{array}{c} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{array} \tag{C.7}$$

Where u_{nm} denotes a non-zero element. We measured u_{nm} by initializing excited qubit in the basis ket of column m with an $X/2$ gate, and measuring the expectation value of $\sigma_x + i\sigma_y$ of the excited qubit in the basis ket denoted by row n. e.g. for u_{21} we initialize the left qubit, apply the fSim gate, and then measure $\sigma_x + i\sigma_y$ of the right qubit—this is the complex value of u_{21}. This procedure works for the single excitation subspace (e.g. n, m in $[1, 2]$), but u_{33} is computed from repeated measurements of $u_{12,\mathrm{excited}}$ and $u_{22,\mathrm{excited}}$ where the previously uninitialized qubit is instead placed into the $|1\rangle$ state as summarized in Table C.1. This procedure is similar to process tomography, but requires considerably fewer measurements to characterize the fSim matrix. We note that an optimal measurement sequence would require only $2n$-1 circuits (for a $n \times n$ matrix) [116]. Even with several thousand repetitions of each circuit, characterizing the matrix with this method takes only a few seconds. Our series of six circuits is intentionally over-complete to avoid singular behavior when some matrix elements are small. In table C.2 we list the conversion matrix elements to the five parameters of our fSim control model. These are useful measurements for building an fSim model, but we cannot characterize small gate errors ($\approx 10^{-3}$) using this method due to the limitations of state preparation and measurement (SPAM) errors which are a few percent.

Table C.2: Computing fSim model parameters from the results of our unitary tomography protocol. The "condition" column is present because we compute $u_{33} = u_{22,\text{excited}}/u_{11}^*$ or $u_{33} = u_{12,\text{excited}}/u_{21}^*$ depending on if u_{11} or u_{21} is larger to ensure the result is non-singular. ψ_{10} is the phase difference accumulated between the two qubits over the gate duration.

fSim parameter	Value	condition
θ	$\arctan(\lvert u_{12}\rvert/\lvert u_{11}\rvert$	none
ϕ	$\Delta_+ - \angle(u_{12,\text{excited}} \times u_{21})$	$\lvert u_{21}\rvert > \lvert u_{11}\rvert$
ϕ	$\angle(u_{22}) - \angle(u_{22,\text{excited}})$	$\lvert u_{21}\rvert < \lvert u_{11}\rvert$
Δ_+	$\angle(-u_{11} \times u_{21})$	$\lvert u_{21}\rvert > \lvert u_{11}\rvert$
Δ_+	$\angle(u_{11} \times u_{22})$	$\lvert u_{21}\rvert < \lvert u_{11}\rvert$
Δ_-	$2 \times \angle(u_{11}) - \Delta_+$	none
$\Delta_{-,\text{off}}$	$-2\left(\angle(-u_{12}/i) + \psi_{10}\right) + \Delta_+$	$\psi_{10} = (\omega_{q_1} - \omega_{q_0}) * t_{gate}$

C.3.4 Cross-entropy error benchmarking

Cross-entropy benchmarking (XEB) is a powerful technique for characterizing the error of an arbitrary gate [103]. It is particularly useful when implementing non-Clifford gates like the continuous fSim gate set we use here. XEB uses a repetitive gate sequence to amplify small errors where each cycle consists of a random single-qubit gate from the set {X/2, Y/2, ±X/2±Y/2} applied to each qubit followed by the fSim gate we are benchmarking. We extract the error per cycle as an exponential decay in the XEB sequence fidelity, $\mathcal{F}_{\text{XEB}}$. The sequence fidelity is computed using the cross-entropy between two probability distributions P and Q, $S(P,Q) = -\sum_i p_i ln(q_i)$, by comparing the expected, measured, and incoherent probability distributions for a given gate sequence,

$$\mathcal{F}_{\text{XEB}} = \frac{S(P_{\text{incoherent}}, P_{\text{expected}}) - S(P_{\text{measured}}, P_{\text{expected}})}{S(P_{\text{incoherent}}, P_{\text{expected}}) - S(P_{\text{expected}})} \tag{C.8}$$

The numerator is the difference between the measured and expected cross-entropy and the denominator serves as a normalization so that $\mathcal{F}_{\text{XEB}}$ takes a value from [0, 1]. We then use $1 - \mathcal{F}_{\text{XEB}}$ as a cost function to optimize the five parameters of our fSim control

model. For a given random sequence, we compute the expected probability distribution using perfect single-qubit gate models and the fSim model obtained from our unitary tomography experiment (see Appendix C.3.3). Since, the sequence fidelity is dependent on the single and two-qubit gate models used in the cross-entropy calculation, we can use $1 - \mathcal{F}_{\text{XEB}}$ as a cost function to optimize some or all of our fSim gate model parameters, a process termed *ex situ* optimization.

C.3.5 RB vs XEB

As a sanity check, one may ask that we compare the result of Clifford based randomized benchmarking (RB) and cross-entropy benchmarking (XEB). Clifford based RB requires an inversion gate, inverting a random gate sequence to map the total ideal gate sequence starting in the $|0\rangle$ state back to $|0\rangle$. For most of the fSim gates, the inversion gate is non-trivial, but, for the special case of a $\text{CZ}_\phi = \text{fSim}(0°, 180°)$, which is part of the Clifford gate set, this comparison is possible.

In Figure C.6a we perform single-qubit Clifford based randomized benchmarking (gate sequence inset), extracting average single-qubit Pauli errors $e_{p,q1} = 0.7 \times 10^{-3}$ and $e_{p,q2} = 0.9 \times 10^{-3}$. In Figure C.6b we perform two-qubit Clifford based randomized benchmarking with and without an interleaved CZ_ϕ gate (sequences inset), extracting a Pauli error per CZ_ϕ of 4.1×10^{-3}. Then, in Figure C.6c we use XEB to measure the per cycle error of the CZ_ϕ + two single-qubit gates obtaining $e_{p,\text{cycle}} = 5.7 \times 10^{-3}$. If we then sum the Clifford based errors for each SQ gate and the CZ_ϕ $(0.7 + 0.9 + 4.1) \times 10^{-3} = 5.7 \times 10^{-3}$

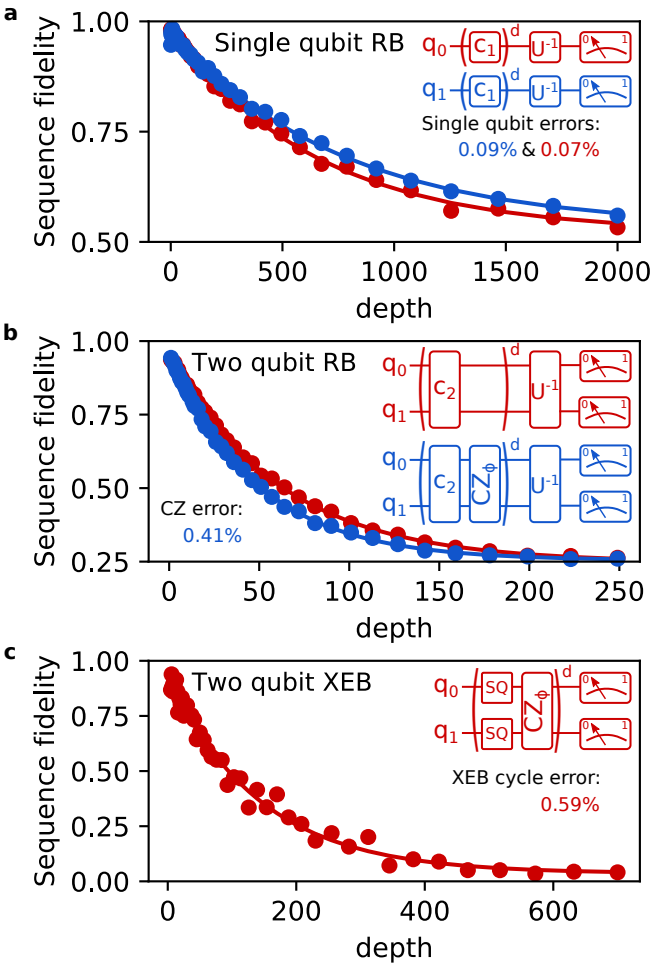


Figure C.6: Comparison of Clifford-based randomized benchmarking (RB) and cross-entropy benchmarking (XEB). **a**, Single-qubit Clifford based randomized benchmarking measuring average Pauli errors of 0.09% and 0.07% for each qubit. **b**, Two-qubit Clifford based randomized benchmarking with (blue) and without (red) an interleaved CZ_ϕ gate, allowing us to extract the Pauli error per $CZ + \phi$ of 0.41%. **c**, Two-qubit cross-entropy benchmarking where each cycle includes two single-qubit gates and a CZ_ϕ gate yielding a Pauli error per cycle of 0.59%. Here we find that the sum of the single and two-qubit errors measured with Clifford based RB (0.09% + 0.07% + 0.41% = 0.57%) corresponds well to the XEB error per cycle (0.59%).

we find good agreement with the XEB error per cycle $e_{p,\text{cycle}} = 5.9 \times 10^{-3}$.

C.3.6 Error budgeting

In this section, we use various techniques to provide a more thorough budget of our XEB per cycle errors. As we have discussed, XEB measures the total error per cycle, $e_{p,\text{cycle}}$. This includes coherent and incoherent errors for one single-qubit gate per qubit and one fSim gate. We use single-qubit Clifford-based randomized benchmarking to characterize the average total error for single-qubit gates, we use purity benchmarking to characterize incoherent error of both the single-qubit and fSim gates, and we use $|2\rangle$ state readout in conjunction with XEB to characterize per cycle leakage (which is included in the incoherent error). Here we focus on the two-qubit gate errors by assuming purity-limited single-qubit Pauli gate errors of 7.5×10^{-4} as described in Appendix C.3.2—this effectively means we subtract 1.5×10^{-3} from $e_{p,\text{cycle}}$ to obtain $e_{p,2\text{q}}$ for both error and purity measurements..

In Figure C.7a we perform Purity benchmarking for each XEB gate sequence and obtain an average Purity of 3.76×10^{-3} per fSim gate. In Figure C.7b we plot $e_{p,2\text{q,unitary\_tomography}}$, the Pauli error per fSim gate using the fSim gate model obtained from unitary tomography. The average $e_{p,2\text{q,unitary\_tomography}}$ is 5.07×10^{-3} indicating a coherent error of 1.31×10^{-3} per fSim. In Figure C.7c we perform *ex situ* optimization of our fSim gate model to reduce the coherent error by changing the three single-qubit detuning model parameters. We hold the values of θ and ϕ fixed to the sampling grid, but allow the

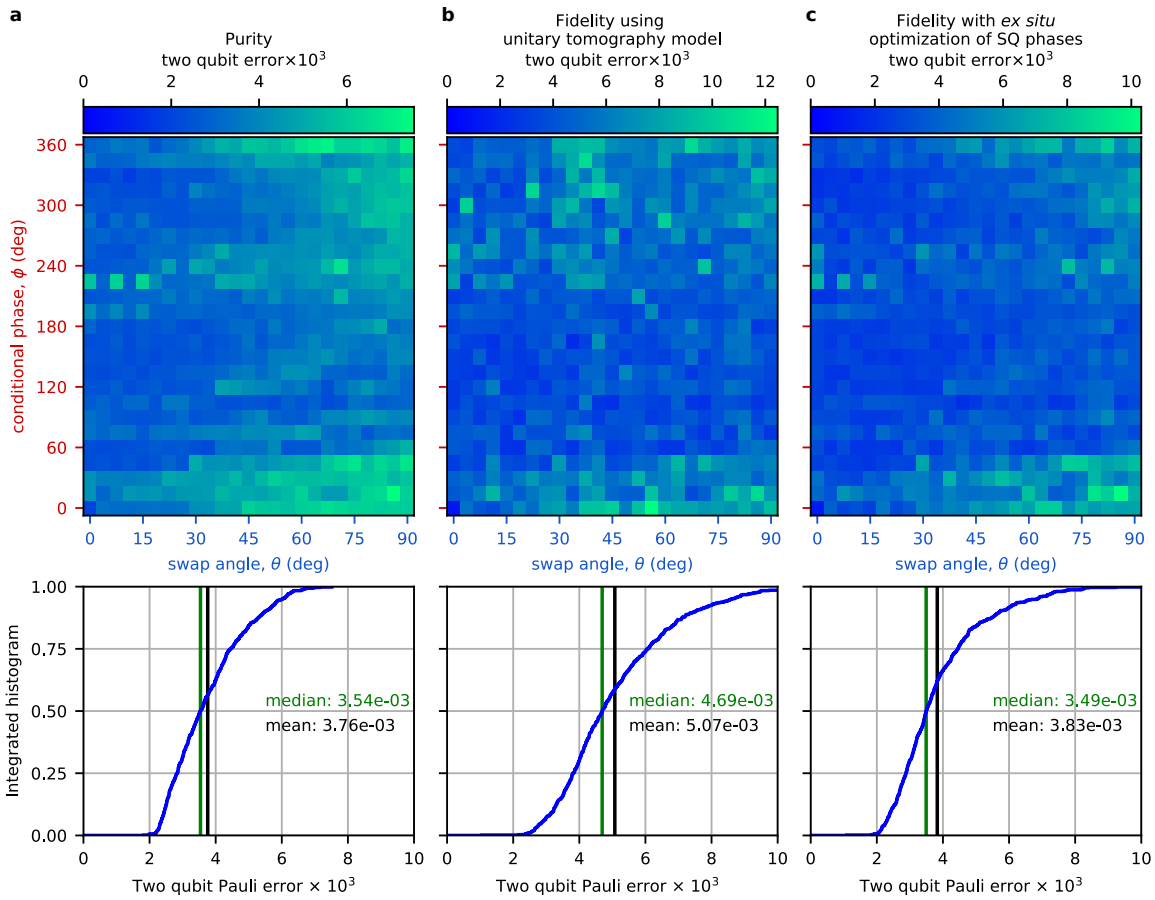


Figure C.7: Comparison of purity benchmarking and cross-entropy benchmarking with and without a constrained *ex situ* optimization of the fSim control angles. **a**, Purity benchmarking. **b**, XEB error per gate using the fSim gate model obtained from unitary tomography (Appendix C.3.3). **c**. XEB error after a constrained *ex situ* optimization of the fSim gate parameters where θ and ϕ were held fixed to the grid and the single-qubit phases were optimized.

single-qubit phases in the fSim model to be optimized. With this improved gate model coherent error is nearly eliminated. The average error $e_{p,\mathrm{2q,ex\,situ}}$ is 3.83×10^{-3} reducing the average coherent error to 7×10^{-5} per gate.

We characterize leakage by directly measuring the $|2\rangle$ state population as a function of the XEB sequence depth. In Figure C.8 we perform this measurement for a line cut of fSim control pulses that sweep the coupler bias on either side of the low-leakage bias used to perform a CPHASE gate. We find leakage to be minimized to a value of $5-6\times10^{-4}$ for a range of coupler biases spanning nearly 10 "clicks" of our 13-bit bipolar DAC ($2/2^{13} \approx 0.0002$).

In total, these metrics indicate that we have achieved incoherent-error-limited gates with fairly low leakage (if necessary, leakage may be reduced further by optimizing the gate length). Additionally, we find that we are able to perform the desired fSim(θ, ϕ) gate we want without incurring additional coherent error. A critical component in achieving these results was eliminating the non-gate-like behaviors induced by long settling tails on our flux bias pulses. As such, we will now detail the procedure used to calibrate our flux control pulses.

C.3.7 Unitary overlap

The unitary overlap of two unitary matrices, e.g. some target fSim, U_{target}, and the actual fSim, U_{actual}, is defined as $Tr(U_{\mathrm{target}} \cdot U_{\mathrm{actual}})/D$, where D is the dimension of the Hilbert space. The unitary overlap is related to the Pauli error, $e_p = 1-(Tr(U_{\mathrm{target}} \cdot U_{\mathrm{actual}})/D)^2$.

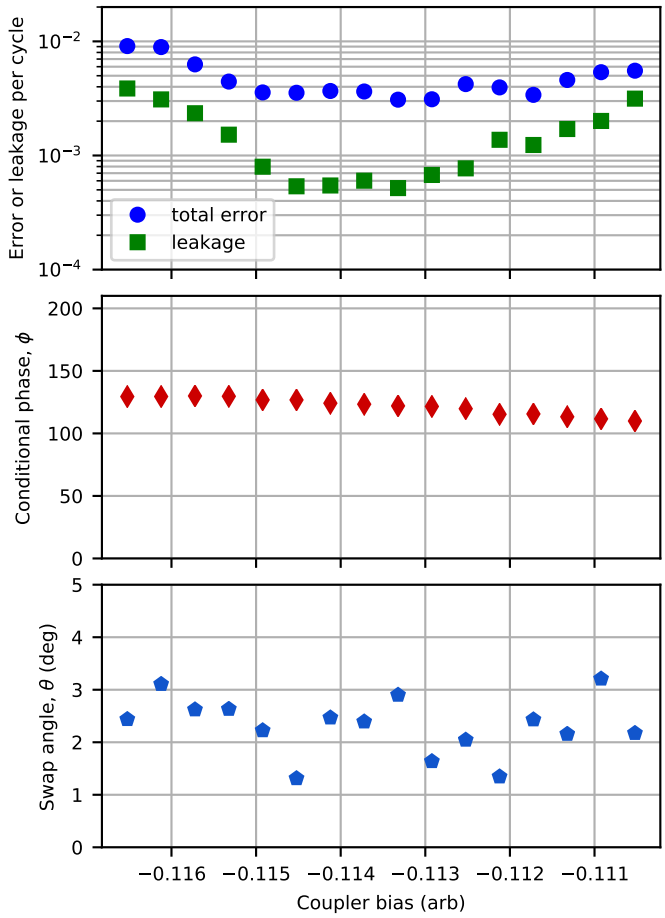


Figure C.8: Plot of leakage and total XEB error per cycle for a 13 ns CPHASE gate as a function of the coupler bias. The increment on the x-axis is twice the minimum increment of our DAC ($2/2^{14}$, e.g. a 14-bit bipolar DAC ≈ 0.0002). In this case we find that leakage reaches a minimum of $5 - 6 \times 10^{-4}$ for a range of coupler amplitudes approximately $10x$ our minimum DAC adjustment.

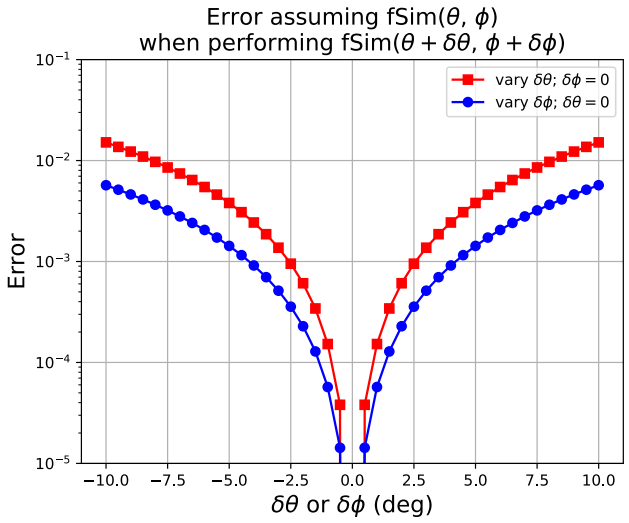


Figure C.9: We may choose to interpret some $\text{fSim}_{\text{actual}} = \text{fSim}(\theta + \delta\theta, \phi + \delta\phi)$ as some $\text{fSim}_{\text{target}} = \text{fSim}(\theta, \phi)$, by accepting additional coherent error. For small deviations in either θ or ϕ the error is proportional to the square of the deviation.

The Pauli error in an fSim gate for small deviations in either θ or ϕ is proportional to the square of the deviation angle. In Figure C.9 we plot the additional coherent error incurred if you assume some actual $\text{fSim}_{\text{actual}} = \text{fSim}(\theta+\delta\theta, \phi+\delta\phi)$ is instead some target $\text{fSim}_{\text{target}} = \text{fSim}(\theta, \phi)$. This plot indicated that a deviation of either 2.5° in θ or 4° in ϕ with result in an additional coherent error of 1×10^{-3}. In our case (Figure C.7), after a constrained optimization where θ and ϕ were fixed to a grid, our average error was approximately 1×10^{-4} higher than the purity limit which corresponds to a deviation of about 1° in either θ or ϕ.

C.4 Control pulse calibration

In a world without flux settling tails, we would be able to implement an arbitrary fSim gate with a fidelity that is the sum of the requisite CPHASE and iSWAP-like gates by just merging the control pulses into a composite fSim gate. Unfortunately, due to flux settling tails, further calibration, described in C.4.3, was required. The keystones of this calibration were two-fold: 1) When performing two flux control based gates back to back (e.g. 2 ns separation), adjust the amplitude of the second pulse based on the first. 2)When implementing a composite gate, perform a CPHASE gate followed by the iSWAP-like gate so that bleed through is well behaved; in the reverse order, bleed through of the iSWAP-like coupler pulse into the CPHASE gate pulses will result in leakage to the $|2\rangle$ state which is an error in the fSim model. Using these two principles, we were able to implement a robust calibration of the complete fSim gate set.

As we have demonstrated numerically in Appendix C.2, our desired implementation of the fSim gate set is possible with less than 1% error using simple rectangular control pulses. Unfortunately, the system transfer function (electronics and wiring) is imperfect and cannot produce these ideal waveforms exactly. Fortunately, as explored numerically in Figure C.3, our fSim implementation is mostly sensitive to the integral of our control pulses rather than the shape. This likely remains true unless the spectral content of our flux control pulses approaches the qubit frequency. However, we must be very careful to ensure our control pulses do not bleed into each other which requires careful calibration of our flux bias settling tails.

We can consider settling non-idealities at two time scales: 1) pulse distortion during the duration of a gate (roughly 15 ns), and 2) pulse settling that occurs after the intended gate duration. Distortion at short times may, for instance, make it difficult to place the qubits exactly on resonance during a gate—this may make it difficult to achieve a swap angle, θ, of 90° swap amplitude (Rabi oscillation amplitude = $g^2/(g^2 + \pi\Delta^2/2) = 1$ if and only if the qubits are on resonance), but fortunately these distortions do not have a huge impact on the rest of the fSim parameter space. Due to the periodic nature of Rabi oscillations the resulting fSim is mostly dependent on the integral of the control pulses. Pulse settling that occurs outside the intended gate interval means that adjacent gates will bleed in to each other. If the tails are relatively short (a few ns), it is possible to mitigate this error just by placing a short idle time between gates. Pulse settling at longer times is particularly nefarious because it becomes no longer feasible to pad gates with idle times and setting times of 5-1000+ ns have been observed in superconducting qubit systems. If left uncompensated, the performance of the m^{th} 15 ns long gate would be dependent on the preceding 1-60+ gates. This runs contrary to the entire notion of gate-based local operations and certainly would not fit within our static fSim control model used with XEB. As such, it is this long-time settling in particular that requires a careful calibration to enable the sensible control strategy employed throughout this letter.

The full fSim gate calibration happens in three stages. In the first stage, we calibrate the electronics to eliminate the long-time settling flux settling. In the second stage, we

describe the calibration procedure for the CPHASE and iSWAP-like gate sets. Then, for the fSim gate family, we perform further calibrations of the composite fSim gates to achieve the best possible gate performance by adjusting the control amplitude of the second pulse dependent on the first rather than adding longer buffer times between flux pulses.

C.4.1 Electronics calibration

On this device there are a total of seven flux bias lines, four for the qubits and three for the couplers. Each channel is driven by a dedicated 1 GS/s, 14-bit DAC controlled by an FPGA to form an arbitrary waveform generator. Each line uses nominally identical cabling, attenuation, and filtering from room temperature down to the sample's chip mount. To compensate for non-idealities in each line, we first measure the qubit's response to a flux pulse, fit the response using three exponential decay time constants, and then use this model to pre-distort our control pulses as in previous work [**?** 106]. This allows us to directly measure and compensate for the transfer function of each qubit's flux bias wiring. Implementing a similar *in situ* calibration of the coupler bias lines is the subject of on-going work. For now, we have found it sufficient to simply apply the average of the two adjacent qubit settling models to the coupler. The pulse calibration parameters for the pair of qubits and the coupler used to benchmark the fSim gateset are summarized in Table C.3.

After performing the electronics calibration we find the unitary gate interactions of

Table C.3: Summary of the settling parameters for two qubits. The average of the settling compensation for these two qubits was applied to the coupler.

	α_1 (%)	τ_1 (ns)	α_2 (%)	τ_2 (ns)	α_3 (%)	τ_3 (ns)
q_2	-0.46	858	-1.00	104	-4.94	10
q_3	-0.61	996	-0.82	94	-5.97	9
coupler (avg q_2 & q_3)	-0.53	927	-0.91	99	-5.45	10

our fSim gates to be well characterized by either unitary tomography, performed with a depth-1 circuit, or cross-entropy bench-marking using a depth N circuit where N varies from 5 to 700. This fact is illustrated by Figure C.7 panels b and c where the difference in the average error of all 525 fSim gates differs by only 1.2×10^{-3} with and without optimizing the single-qubit unitary parameters—this provides an upper bound on the effects of pulse bleed through on gate fidelity. If we consider the gate timing of the cross-entropy benchmarking sequence in Figure C.7, which used 28 ns fSim gates interleaved with 15 ns single-qubit gates, this result indicates that our settling is well compensated for at times longer than 15 ns. This result also indicates that the qubit biases are settled enough to have a minimal impact on the single-qubit gate errors. If this were not the case then we would require a circuit-depth-dependent gate model to reach the purity limit. However, the settling of the coupler bias flux signal at times less than 15 ns becomes non-negligible and merits special consideration when calibrating fSim gates composed of a CPHASE gate in close proximity to an iSWAP-like gate. So, we will first detail the calibration procedure for each of the component gate families in the next section and finish our calibration discussion with a description of composite fSim calibration procedure.

C.4.2 CPHASE and iSWAP-like calibrations

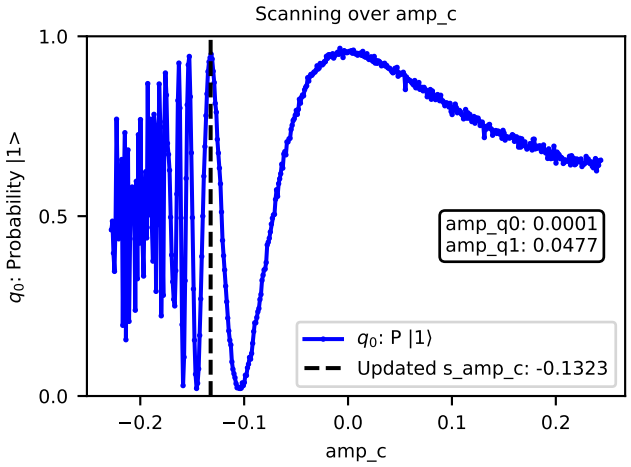


Figure C.10: CPHASE gate calibration scan. We use the pre-calibrated bias-to-qubit frequency function to choose a desired qubit-qubit detuning and then sweep the amplitude of the coupler bias, amp_c to identify the amplitude that completes a diabatic $|11\rangle \leftrightarrow |02\rangle$ swap indicated by the dotted vertical line.

We calibrate the CPHASE interaction by repeating the leakage experiment described in Figure 6.2a to fine tune the coupler bias amplitudes and to identify combinations of qubit detunings, Δ, and corresponding coupler bias amplitudes that yield low-leakage gates. We use the qubit frequency bias transfer function to choose qubit biases that set the desired qubit detuning, Δ in the vicinity of η. The frequency range around η is set by the width of the Rabi interaction which, for a fixed pulse length, is inversely proportional to gate length since shorter gates require stronger coupling, g (see Figure C.2 in Appendix C.2). We use 15 ns pulses (13 ns rectangular pulses with a 1 ns padding on either side) which makes the Rabi interaction span about 75 MHz on either side of the qubit nonlinearity, η. For each detuning in this range, we repeat the experiment in Figure 6.2a varying the coupling strength to minimize leakage. An example of the raw

data from this experiment is provided in Figure C.10 where the dotted line indicates the low-leakage coupler bias amplitude that achieves one full swap from $|11\rangle$ to $|02\rangle$ and back. We initialize the $|11\rangle$ state, apply the CPHASE control pulses, and measure the $|1\rangle$ state population of the lower frequency qubit to identify when the population has completed a full swap. Then, for each combination of Δ and the corresponding low-leakage coupler bias we repeat the experiment from Figure 6.2b to measure the conditional phase. This procedure works well for $\phi : [-130°, 130°]$ until the Rabi swap amplitude becomes small and the peak is broad along the coupling strength line-cut. At that point, we extrapolate towards the zero coupling bias while measuring the conditional phase to fill out the rest of the conditional phase control space.

In Figure C.11 we calibrate a 13 ns iSWAP-like gate (11 ns rectangular pulses, with 1 ns padding) by repeating the experiment from Figure 6.2c three times with the qubits on resonance ($\Delta = 0$ MHz) to fine-tune the pulse amplitudes needed to reach $\theta = 90°$. Then, for $0° < \theta < 90°$ we simply interpolate the coupler bias between the "OFF" bias and the $\theta = 90°$ bias. For each iSWAP-like tune up experiment we initialize one qubit to the $|1\rangle$ state, apply the iSWAP-like pulses to the qubits and coupler, and then measure the $|1\rangle$ state population of the other qubit. In Figure C.11a, we first use our pre-calibrated qubit frequency bias DC transfer functions to choose qubit bias amplitudes, amp_q0 and amp_q1, that place both qubits at the same frequency, and we sweep the coupler bias from the "OFF" bias to the maximum coupling bias to identify the amplitude that achieves exactly one a swap from the first to the second qubit corresponding to $\theta = 90°$

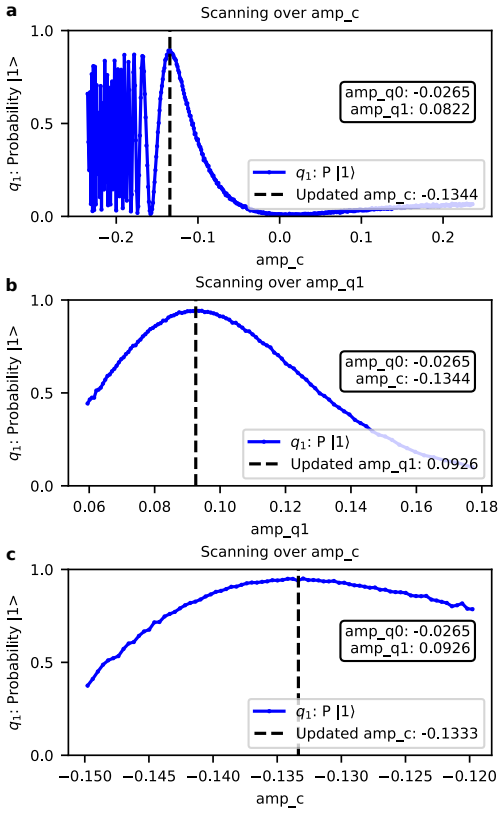


Figure C.11: iSWAP-like gate calibration for $\theta = 90°$ performed in three steps. In each experiment we initialize the $|01\rangle$ state and measure the population of the first qubit to identify the control biases to complete a full swap to $|10\rangle$ **a**, We use prior calibations to bias the qubits on-resonance and scan the coupler bias amplitude, *amp_c*, to find the bias that completes one swap. **b**, Using the new *amp_c*, we scan the bias of one qubit, *amp_q*, to place them on resonance. **c**, Using the updated qubit biases, we again scan *amp_c* to find tune the coupler bias.

(dotted line). In Figure C.11b, we repeat the experiment using the $\theta = 90°$ coupler bias from C.11a while sweeping the bias of one qubit to maximize the amplitude of the swapped population, thus placing the qubits on resonance. Finally, in C.11c, we repeat the experiment using the new qubit biases to fine-tune the coupler bias.

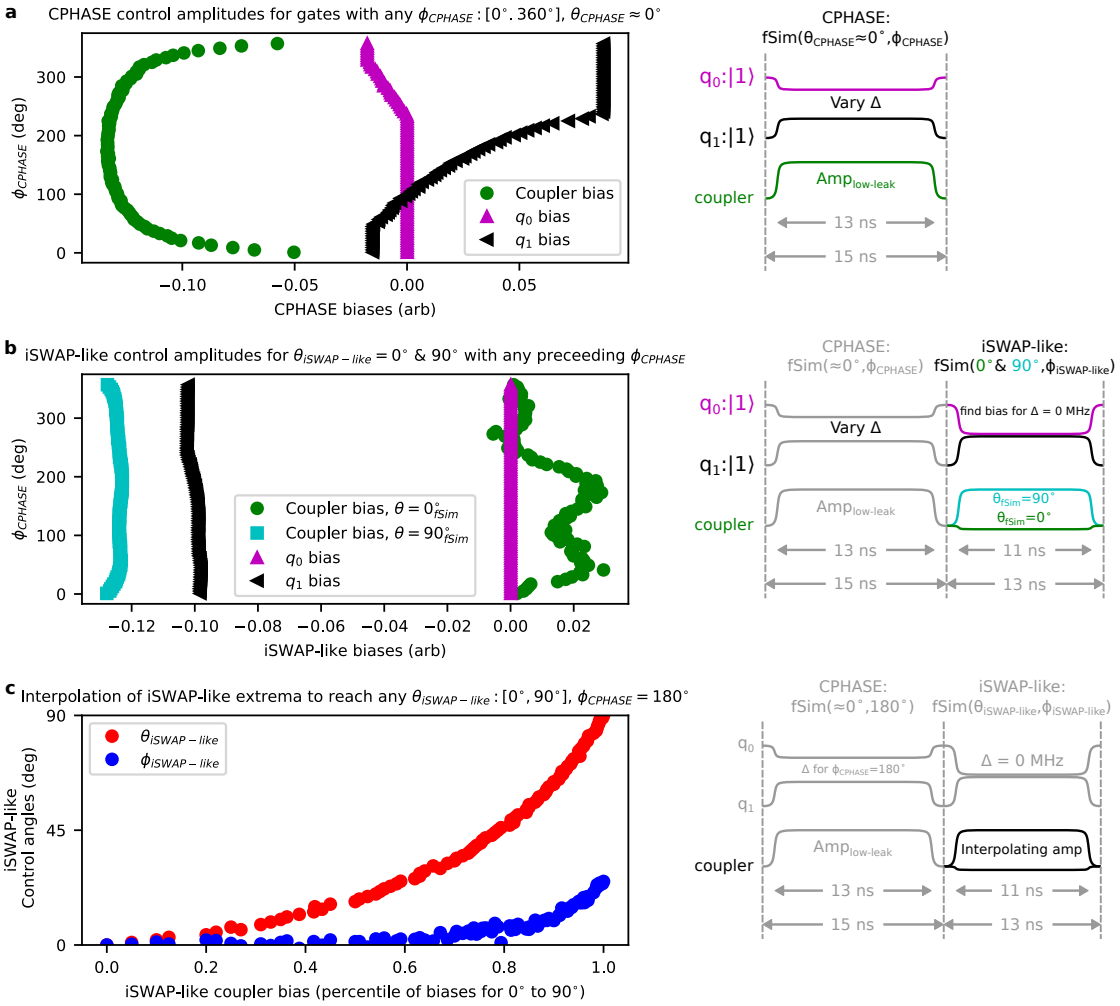


Figure C.12: Three steps to calibrating the fSim gate to account for pulse distortion of the first (CPHASE) pulses bleeding into the second (iSWAP-like) pulses. **a**, Follow the usual CPHASE calibration procedure to bring-up a full CPHASE gate family corresponding to fSim($\theta \approx 0°, \phi : [0°, 360°]$). **b**, Follow the iSWAP-like tune up procedure, but play the CPHASE control pulses before the iSWAP-like pulses. Use the sequence to identify the flux bias amplitudes that achieve fSim gates with $\theta = 0°$ and $\theta = 90°$ for each proceeding CPHASE gate. **c**, for a preceding CPHASE gate with $\phi = 180°$, bring up gates corresponding to fSim($\theta : [0°, 90°], \phi = 180°$).

C.4.3 fSim calibration

Once we have calibrated the iSWAP-like and CPHASE gates, we should nominally be able to use one of each to implement any fSim gate. Unfortunately our pulse response is imperfect at short times, as described in Appendix C.4.1. This was less of an issue for the iSWAP-like and CPHASE gates in section 6.4 because the XEB gate sequence alternated (10 ns) single and (13 ns or 15 ns) two-qubit gates; in those cases, the uncompensated flux bias settling tails resulted in a small detuning at the qubit idle frequencies. However, when we perform an fSim gate as a composition of a CPHASE followed immediately by an iSWAP-like gate, the tail of the first coupler pulse bleeds into the second coupler pulse. Even a small settling tail adding to the amplitude of the coupler pulse can drastically change the coupling during the second gate due to the large coupler flux sensitivity at strong couplings (remembering Figure 6.1c). In the future, this problem may be mitigated by identifying and removing the physical origin of these settling tails, with a more thorough *in situ* calibration procedure for the couplers, or by placing longer idle times between gates.

In this work, we deal with pulse bleed through by calibrating composite fSim gates where the amplitudes of the second set of pulses in the composite fSim sequence is dependent on the first, thus eliminating the need for excessive idle times between gates. Conveniently, the tune up procedure for each gate in the composition is the same as in the isolated iSWAP-like or CPHASE case, just with the two sets of pulses played back-to-back—this works because each experiment in our usual bring-up procedure operates

within an isolated manifold (e.g. one excitation for θ or two excitations for ϕ) when performing fSim gates. The ordering of the gates within the fSim gate is chosen to place the CPHASE gate before the iSWAP-like gate. Since both coupler pulses have the same sign, if the CPHASE coupler amplitude bleeds into the iSWAP-like coupler amplitude, this results in slightly more swapping which is easily measured and adjusted for by reducing the iSWAP-like coupler amplitude to compensate. If we ordered the gates in the reverse order, pulse bleed through would generate leakage during the CPHASE gate which is much more difficult to characterize and remove.

For the purpose of building a robust registry of gates, we erred on the side of over-calibration for this demonstration. However, we find these control parameters to be well behaved and it should be possible to sample more sparsely in the future to simplify calibration of the full fSim gate set. Figure C.12 outlines the three steps used to calibrate our composite fSim gates. In Figure C.12a, we first calibrate many CPHASE gates spaced every 1° using control pulses for just the CPHASE gate as shown on the right following the procedure outlined in Appendix C.4.2. Then, in Figure C.12b, for each preceding CPHASE gate we follow the iSWAP-like calibration procedure (also Appendix C.4.2) to identify qubit and coupler bias amplitudes to achieve both a $\theta = 0°$ and 90° gate. Finally, in Figure C.12c, for a CPHASE conditional phase, $\phi_{\text{CPHASE}} = 180°$ we tune up iSWAP-like gates for θ from 0° to 90° in 1° increments by interpolating between the min and maximum amplitudes determined in C.12b. We use this calibration to produce a spline for $\theta_{\text{iSWAP-like}} \rightarrow \%(bias_{90°} - bias_{0°})$ and another for $\theta_{\text{iSWAP-like}} \rightarrow \phi_{\text{iSWAP-like}}$.

With the fSim gate registry in hand we set out to benchmark specific fSim gates. For a given target fSim, we first look up the iSWAP-like pulse amplitudes that achieve the correct swap angle $\theta_{\text{iSWAP-like}}$, and subtract the conditional phase due to the iSWAP-like gate from the total target to choose pulse amplitudes for the desired CPHASE gate (e.g. $\phi_{\text{CPHASE}} = \phi_{\text{target}} - \phi_{\text{iSWAP-like}}$). We then performed unitary tomography (Appendix C.3.3) using the pulse amplitudes we looked up in the registry to quickly assess the resulting fSim control angles of the composite gate. If either control angle is off by more than 1°, we used the registry to adjust the corresponding iSWAP-like (θ) or CPHASE (ϕ) control amplitudes by $\pm 1°$ accordingly. This process converged to an fSim gate within 1° of both θ_{target} and ϕ_{target} for the target fSim with fewer than 9 adjustments for each of the 525 fSim gates we benchmarked. Once the unitary tomography experiment indicated the composite fSim gate produced a unitary operation near the target fSim gate, we performed purity and cross-entropy benchmarking.

C.5 System stability

As the size of quantum processors grows (number of qubits), so too does the time it takes to calibrate a device (at least until fully parallel calibrations are possible). As the system drifts from these calibrations over time, the performance of a processor will fall and calibrations must be revisited. If the required calibration time is long compared to the scale of drift, then the device becomes unusable in practice. While electronics drift with both time and temperature must be considered when designing a system, one

particularly worrisome issue is the time dependence of two level system (TLS) defects entering and leaving the qubit spectrum [109].

Here we present a promising snapshot of the stability of our system. In the process of calibrating the fSim gate set, we started by calibrating the single-qubit gates and readout. We then operated with the same single-qubit calibrations for several days while we were working on the fSim gate ultimately obtaining our primary fSim benchmarking dataset about a week after the initial single-qubit calibration. Shortly after this, a TLS showed up near one of the qubit's idle frequencies significantly limiting its coherence. Then, after about another week, we returned to the original calibration parameters to benchmark a subset of the same $\mathrm{fSim}(\theta, \phi)$ gates presented in the main text. We were pleasantly surprised to find that the original calibration was still good enough to produce high fidelity gates.

In Figure C.13 we used the two-week-old iSWAP-like and CPHASE calibrations to benchmark a less dense grid of fSim gates. While the average performance has degraded by a factor of two from the initial calibration, the average error is still less than 1%, but that is not the whole story. These fSim gates were benchmarked in a random order—if we look at a plot of the gate error as a function of time for these 91 (figure C.13b), we see a strong time dependence where the first 50 gates (benchmarked over the course of an hour) have an average error much lower than gates #50 to #80. This would seem to indicated that the two-week-old electronics calibration is stable enough to maintain high fidelity gates for weeks, and that the decreased fidelity is likely due to the residual

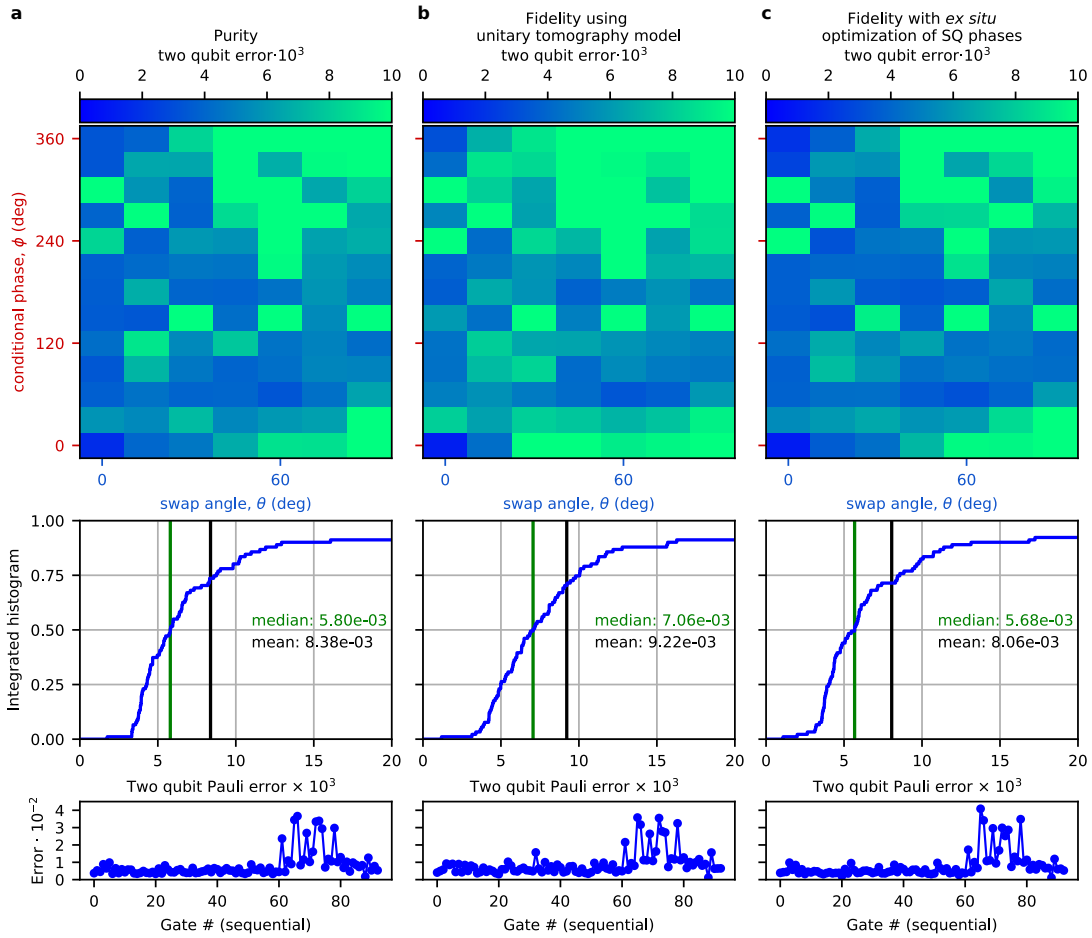


Figure C.13: Snapshot of system stability. A few days after taking the data in Figure 6.4, a TLS showed up at one of the qubit idle frequencies effectively breaking the calibration. After about another week, we returned to the original calibration and repeated the fidelity measurement on a subset of 91 fSim gates which we present here (**top** and **middle** rows). We find that the average gate fidelity had decreased somewhat, but is still above 99%. Furthermore, if we look at the gate error rates sorted in the order they were measured (**bottom** row), a strong time-dependence becomes apparent. Many of the gates presenting low errors ($\approx 5 \times 10^{-3}$) as they did after the initial calibration. It is not until gates numbered 60 to 80 or so where large errors show up. This indicates that our control electronics are stable enough to maintain a high fidelity calibration on the timescale of weeks, and that TLSs are likely the biggest threat to maintaining long term calibrations.

and/or intermittent presence of a TLS interacting with one of the qubits. In an ideal world, we would be able to prevent or remove TLS defects, but, at least presently, we do not know how to do this. Instead, relying on the stability of our electronics, an optimal strategy for maintaining up-time on a large-scale quantum processor will likely involve calibrating a number of idle frequency configurations and being able to quickly vet and switch to an old configuration if and when a TLS shows up.

Bibliography

[1] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, Nov 1994.

[2] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[3] Lov K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.*, 79:325–328, Jul 1997.

[4] Stephen Jordan. Quantum algorithm zoo, Apr 2011.

[5] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, Jun 1982.

[6] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and John M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508:500 EP –, Apr 2014.

[7] Philipp Gerbert and Frank Reuß. The next decade in quantum computing—and how to play, 2018.

[8] Philipp Gerbert and Frank Reuß. Where will quantum computers create valueand when?, 2019.

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

[10] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell,

Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[11] Fumiko Yamaguchi and Yoshihisa Yamamoto. Quantum simulation of the tj model. *Superlattices and Microstructures*, 32:343–345, 12 2002.

[12] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik, and A. G. White. Towards quantum chemistry on a quantum computer. *Nature Chemistry*, 2(2):106–111, 2010.

[13] Ian D. Kivlichan, Jarrod McClean, Nathan Wiebe, Craig Gidney, Alán Aspuru-Guzik, Garnet Kin-Lic Chan, and Ryan Babbush. Quantum simulation of electronic structure with linear depth and connectivity. *Phys. Rev. Lett.*, 120:110501, Mar 2018.

[14] F. Bloch. Nuclear induction. *Phys. Rev.*, 70:460–474, Oct 1946.

[15] David Deutsch and Roger Penrose. Quantum theory, the church–turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117, 1985.

[16] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52:3457–3467, Nov 1995.

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

[18] David P. DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik*, 48(911):771–783, 2000.

[19] Luke Durant, Olivier Giroux, Mark Harris, and Nick Stam. Inside volta: The world's most advanced data center gpu, May 2017.

[20] Samsung breaks terabyte threshold for smartphone storage with industry's first 1tb embedded universal flash storage.

[21] Julian Kelly. A preview of bristlecone, google's new quantum processor, Mar 2018.

[22] P. Lebrun. Cryogenics for the large hadron collider. *IEEE Transactions on Applied Superconductivity*, 10(1):1500–1506, March 2000.

[23] Oliver Sim Brning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[24] P Lebrun. Advances in Cryogenics at the Large Hadron Collider. Technical Report LHC-Project-Report-211. CERN-LHC-Project-Report-211, CERN, Geneva, Jul 1998.

[25] E. Schrödinger. Der stetige übergang von der mikro- zur makromechanik. *Naturwissenschaften*, 14(28):664–666, Jul 1926.

[26] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf. ac stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.*, 94:123602, Mar 2005.

[27] 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, 2004.

[28] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Approaching unit visibility for control of a superconducting qubit with dispersive readout. *Phys. Rev. Lett.*, 95:060501, Aug 2005.

[29] R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics, Vol. III: The New Millennium Edition: Quantum Mechanics*. The Feynman Lectures on Physics. Basic Books, 2011.

[30] B.D. Josephson. Possible new effects in superconductive tunnelling. *Physics Letters*, 1(7):251 – 253, 1962.

[31] M. D. Hutchings, J. B. Hertzberg, Y. Liu, N. T. Bronn, G. A. Keefe, Markus Brink, Jerry M. Chow, and B. L. T. Plourde. Tunable superconducting qubits with flux-independent coherence. *Phys. Rev. Applied*, 8:044003, Oct 2017.

[32] 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.

[33] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A*, 86:032324, Sep 2012.

[34] Frank Pobell. *Matter and Methods at Low Temperatures*. Springer Berlin Heidelberg, 2007.

[35] Applications: Systems for quantum computing.

[36] Olli V Lounasmaa. *Experimental principles and methods below 1K*. Academic Press, Orlando, FL, 1974.

[37] Pulse tube cryocoolers.

[38] M. L. Roukes, M. R. Freeman, R. S. Germain, R. C. Richardson, and M. B. Ketchen. Hot electrons and energy transport in metals at millikelvin temperatures. *Phys. Rev. Lett.*, 55:422–425, Jul 1985.

[39] F. C. Wellstood, C. Urbina, and John Clarke. Hot-electron effects in metals. *Phys. Rev. B*, 49:5942–5955, Mar 1994.

[40] J. P. Pekola, K. P. Hirvi, J. P. Kauppinen, and M. A. Paalanen. Thermometry by arrays of tunnel junctions. *Phys. Rev. Lett.*, 73:2903–2906, Nov 1994.

[41] Sh. Farhangfar, K. P. Hirvi, J. P. Kauppinen, J. P. Pekola, J. J. Toppari, D. V. Averin, and A. N. Korotkov. One dimensional arrays and solitary tunnel junctions in the weak coulomb blockade regime: Cbt thermometry. *Journal of Low Temperature Physics*, 108(1):191–215, Jul 1997.

[42] Jukka Pekola. Trends in thermometry. *Journal of Low Temperature Physics*, 135(5):723–744, Jun 2004.

[43] M. Meschke, J. P. Pekola, F. Gay, R. E. Rapp, and H. Godfrin. Electron thermalization in metallic islands probed by coulomb blockade thermometry. *Journal of Low Temperature Physics*, 134(5):1119–1143, Mar 2004.

[44] D. I. Bradley, R. E. George, D. Gunnarsson, R. P. Haley, H. Heikkinen, Yu A. Pashkin, J. Penttilä, J. R. Prance, M. Prunnila, L. Roschier, and M. Sarsby. Nanoelectronic primary thermometry below 4 mk. *Nature Communications*, 7(1):10455, 2016.

[45] M. Palma, C. P. Scheller, D. Maradan, A. V. Feshchenko, M. Meschke, and D. M. Zumbhl. On-and-off chip cooling of a coulomb blockade thermometer down to 2.8mk. *Applied Physics Letters*, 111(25):253105, 2017.

[46] L. Casparis, M. Meschke, D. Maradan, A. C. Clark, C. P. Scheller, K. K. Schwarzwlder, J. P. Pekola, and D. M. Zumbhl. Metallic coulomb blockade thermometry down to 10 mk and below. *Review of Scientific Instruments*, 83(8):083903, 2012.

[47] Zijun Chen. *Metrology of Quantum Control and Measurement in Superconducting Qubits*. PhD thesis, University of California Santa Barbara, 1 2018.

[48] Charles Neill. *A path towards quantum supremacy with superconducting qubits*. PhD thesis, University of California Santa Barbara, Dec 2017.

[49] John M. Martinis and Michael R. Geller. Fast adiabatic qubit gates using only σ_z control. *Phys. Rev. A*, 90:022307, Aug 2014.

[50] Andrew Dunsworth. *High Fidelity Entangling Gates in Superconducting Qubits*. PhD thesis, University of California Santa Barbara, Sept 2018.

[51] 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.

[52] A. Dunsworth, R. Barends, Yu Chen, Zijun Chen, B. Chiaro, A. Fowler, B. Foxen, E. Jeffrey, J. Kelly, P. V. Klimov, E. Lucero, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, H. Neven, John M. Martinis, and A. Megrant. A method for building low loss multi-layer wiring for superconducting microwave devices. *Applied Physics Letters*, 112(6):063502, 2018.

[53] Daniel Sank. *Fast, Accurate State Measurement in Superconducting Qubits*. PhD thesis, University of California Santa Barbara, Sept 2014.

[54] J. B. Johnson. Thermal agitation of electricity in conductors. *Phys. Rev.*, 32:97–109, Jul 1928.

[55] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32:110–113, Jul 1928.

[56] A. Megrant, C. Neill, R. Barends, B. Chiaro, Yu Chen, L. Feigl, J. Kelly, Erik Lucero, Matteo Mariantoni, P. J. J. OMalley, D. Sank, A. Vainsencher, J. Wenner,

T. C. White, Y. Yin, J. Zhao, C. J. Palmstrm, John M. Martinis, and A. N. Cleland. Planar superconducting resonators with internal quality factors above one million. *Applied Physics Letters*, 100(11):113510, 2012.

[57] 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.

[58] 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.

[59] R. Barends, J. Wenner, M. Lenander, Y. Chen, R. C. Bialczak, J. Kelly, E. Lucero, P. OMalley, 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. *Applied Physics Letters*, 99(11):113507, 2011.

[60] D. C. Mattis and J. Bardeen. Theory of the anomalous skin effect in normal and superconducting metals. *Phys. Rev.*, 111:412–417, Jul 1958.

[61] J. Gao, J. Zmuidzinas, A. Vayonakis, P. Day, B. Mazin, and H. Leduc. Equivalence of the effects on the complex conductivity of superconductor due to temperature change and external pair breaking. *Journal of Low Temperature Physics*, 151(1):557–563, Apr 2008.

[62] Allen Rothwarf and B. N. Taylor. Measurement of recombination lifetimes in superconductors. *Phys. Rev. Lett.*, 19:27–30, Jul 1967.

[63] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino. Quasiparticle and phonon lifetimes in superconductors. *Phys. Rev. B*, 14:4854–4873, Dec 1976.

[64] B Chiaro, A Megrant, A Dunsworth, Z Chen, R Barends, B Campbell, Y Chen, A Fowler, I C Hoi, E Jeffrey, J Kelly, J Mutus, C Neill, P J J O'Malley, C Quintana, P Roushan, D Sank, A Vainsencher, J Wenner, T C White, and John M Martinis. Dielectric surface loss in superconducting resonators with flux-trapping holes. *Superconductor Science and Technology*, 29(10):104006, aug 2016.

[65] Gheorghe Stan, Stuart B. Field, and John M. Martinis. Critical field for complete vortex expulsion from narrow superconducting strips. *Phys. Rev. Lett.*, 92:097003, Mar 2004.

[66] M. Tinkham. *Introduction to Superconductivity.* Dover Books on Physics Series. Dover Publications, 2004.

[67] C. Song, T. W. Heitmann, M. P. DeFeo, K. Yu, R. McDermott, M. Neeley, John M. Martinis, and B. L. T. Plourde. Microwave response of vortices in superconducting thin films of re and al. *Phys. Rev. B*, 79:174512, May 2009.

[68] D. Rosenberg, D. Kim, R. Das, D. Yost, S. Gustavsson, D. Hover, P. Krantz, A. Melville, L. Racz, G. O. Samach, S. J. Weber, F. Yan, J. L. Yoder, A. J. Kerman, and W. D. Oliver. 3d integrated superconducting qubits. *npj Quantum Information*, 3(1):42, 2017.

[69] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A*, 86:032324, Sep 2012.

[70] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis. Optimal quantum control using randomized benchmarking. *Phys. Rev. Lett.*, 112:240504, Jun 2014.

[71] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya, R. Barends, B. Burkett, Y. Chen, Z. Chen, A. Fowler, B. Foxen, M. Giustina, R. Graff, E. Jeffrey, T. Huang, J. Kelly, P. Klimov, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, H. Neven, and J. M. Martinis. A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science*, 360(6385):195–199, 2018.

[72] M. Hatridge, R. Vijay, D. H. Slichter, John Clarke, and I. Siddiqi. Dispersive magnetometry with a quantum limited squid parametric amplifier. *Phys. Rev. B*, 83:134501, Apr 2011.

[73] J. Y. Mutus, T. C. White, R. Barends, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, J. Kelly, A. Megrant, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, K. M. Sundqvist, A. N. Cleland, and John M. Martinis. Strong environmental coupling in a josephson parametric amplifier. *Applied Physics Letters*, 104(26):263513, 2014.

[74] J. H. Béjanin, T. G. McConkey, J. R. Rinehart, C. T. Earnest, C. R. H. McRae, D. Shiri, J. D. Bateman, Y. Rohanizadegan, B. Penava, P. Breul, S. Royak, M. Zapatka, A. G. Fowler, and M. Mariantoni. Three-dimensional wiring for extensible quantum computing: The quantum socket. *Phys. Rev. Applied*, 6:044010, Oct 2016.

[75] Nicholas T Bronn, Vivekananda P Adiga, Salvatore B Olivadese, Xian Wu, Jerry M Chow, and David P Pappas. High coherence plane breaking packaging for superconducting qubits. *Quantum Science and Technology*, 3(2):024007, 2018.

[76] Antonio D Córcoles, Easwar Magesan, Srikanth J Srinivasan, Andrew W Cross, Matthias Steffen, Jay M Gambetta, and Jerry M Chow. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nature communications*, 6, 2015.

[77] Diego Ristè, Stefano Poletto, M-Z Huang, Alessandro Bruno, Visa Vesterinen, O-P Saira, and Leonardo DiCarlo. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nature communications*, 6, 2015.

[78] Maika Takita, A. D. Córcoles, Easwar Magesan, Baleegh Abdo, Markus Brink, Andrew Cross, Jerry M. Chow, and Jay M. Gambetta. Demonstration of weight-four parity measurements in the surface code architecture. *Phys. Rev. Lett.*, 117:210505, Nov 2016.

[79] 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.

[80] 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.

[81] Aaron D. OConnell, M. Ansmann, R. C. Bialczak, M. Hofheinz, N. Katz, Erik Lucero, C. McKenney, M. Neeley, H. Wang, E. M. Weig, A. N. Cleland, and J. M. Martinis. Microwave dielectric loss at single photon energies and millikelvin temperatures. *Applied Physics Letters*, 92(11):112903, 2008.

[82] Ch. Broennimann, F. Glaus, J. Gobrecht, S. Heising, M. Horisberger, R. Horisberger, H.C. Kstli, J. Lehmann, T. Rohe, and S. Streuli. Development of an indium bump bond process for silicon pixel detectors at psi. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565(1):303 – 308, 2006. Proceedings of the International Workshop on Semiconductor Pixel Detectors for Particles and Imaging.

[83] A. J. Walton, W. Parkes, J. G. Terry, C. Dunare, J. T. M. Stevenson, A. M. Gundlach, G. C. Hilton, K. D. Irwin, J. N. Ullom, W. S. Holland, W. D. Duncan, M. D. Audley, P. A. R. Ade, R. V. Sudiwala, and E. Schulte. Design and fabrication of the detector technology for scuba-2. *IEE Proceedings - Science, Measurement and Technology*, 151(2):110–120, March 2004.

[84] G.C. Hilton, J.A. Beall, W.B. Doriese, W.D. Duncan, L.S. Ferreira, K.D. Irwin, C.D. Reintsema, J.N. Ullom, L.R. Vale, Y. Xu, B.L. Zink, W. Parkes, A.S. Bunting, C.C. Dunare, A.M. Gundlach, J.T.M. Stevenson, A.J. Walton, E. Schulte, E. Corrales, J.P. Sienicki, Dan Bintley, P.A.R. Ade, Rashmi V. Sudiwala, Adam L. Woodcraft, Mark Halpern, W. Holland, M.D. Audley, and M. MacIntosh. Fabrication of prototype imaging arrays for scuba-2. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(2):513 – 515, 2006. Proceedings of the 11th International Workshop on Low Temperature Detectors.

[85] S Narayana, V K Semenov, Y A Polyakov, V Dotsenko, and S K Tolpygo. Design and testing of high-speed interconnects for superconducting multi-chip modules. *Superconductor Science and Technology*, 25(10):105012, 2012.

[86] S. K. Tolpygo, D. Tolpygo, R. T. Hunt, S. Narayana, Y. A. Polyakov, and V. K. Semenov. Wafer bumping process and inter-chip connections for ultra-high data transfer rates in multi-chip modules with superconductor integrated circuits. *IEEE Transactions on Applied Superconductivity*, 19(3):598–602, June 2009.

[87] D. Rosenberg, D. Kim, R. Das, D. Yost, S. Gustavsson, D. Hover, P. Krantz, A. Melville, L. Racz, G. O. Samach, S. J. Weber, F. Yan, J. Yoder, A. J. Kerman, and W. D. Oliver. 3d integrated superconducting qubits. *npj Quantum Information*, 3, 2017.

[88] C. R. H. McRae, J. H. Bjanin, Z. Pagel, A. O. Abdallah, T. G. McConkey, C. T. Earnest, J. R. Rinehart, and M. Mariantoni. Thermocompression bonding technology for multilayer superconducting quantum circuits. *Applied Physics Letters*, 111(12):123501, 2017.

[89] W. O'Brien, M. Vahidpour, J. T. Whyland, J. Angeles, J. Marshall, D. Scarabelli, G. Crossman, K. Yadav, Y. Mohan, C. Bui, V. Rawat, R. Renzas, N. Vodrahalli, A. Bestwick, and C. Rigetti. Superconducting Caps for Quantum Integrated Circuits. *ArXiv e-prints*, August 2017.

[90] 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. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, 30(1):010607, 2012.

[91] M. Datta, T. Osaka, and J.W. Schultze. *Microelectronic Packaging*. New Trends in Electrochemical Technology. CRC Press, 2004.

[92] K. Wade and A.J. Banister. *The Chemistry of Aluminium, Gallium, Indium and Thallium*. Pergamon texts in Inorganic Chemistry. Pergamon Press, 1973.

[93] Josephine B. Chang, Michael R. Vissers, Antonio D. Crcoles, Martin Sandberg, Jiansong Gao, David W. Abraham, Jerry M. Chow, Jay M. Gambetta, Mary Beth Rothwell, George A. Keefe, Matthias Steffen, and David P. Pappas. Improved superconducting qubit coherence using titanium nitride. *Applied Physics Letters*, 103(1):012602, 2013.

[94] S Ohya, B Chiaro, A Megrant, C Neill, R Barends, Y Chen, J Kelly, D Low, J Mutus, P J J OMalley, P Roushan, D Sank, A Vainsencher, J Wenner, T C White, Y Yin, B D Schultz, C J Palmstrm, B A Mazin, A N Cleland, and John M Martinis. Room temperature deposition of sputtered tin films for superconducting coplanar waveguide resonators. *Superconductor Science and Technology*, 27(1):015009, 2014.

[95] 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.

[96] E. F. Schulte, K. A. Cooper, M. Phillips, and S. L. Shinde. Characterization of a novel fluxless surface preparation process for die interconnect bonding. In *2012 IEEE 62nd Electronic Components and Technology Conference*, pages 26–30, May 2012.

[97] Frederic T. Chong, Diana Franklin, and Margaret Martonosi. Programming languages and compiler design for realistic quantum hardware. *Nature*, 549:180 EP –, Sep 2017.

[98] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Phys. Rev. A*, 52:3457–3467, Nov 1995.

[99] Navin Khaneja and Steffen J. Glaser. Cartan decomposition of su(2n) and control of spin systems. *Chemical Physics*, 267(1):11 – 23, 2001.

[100] 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.

[101] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm, 2016.

[102] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 2019.

[103] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595–600, 2018.

[104] Fei Yan, Philip Krantz, Youngkyu Sung, Morten Kjaergaard, Daniel L. Campbell, Terry P. Orlando, Simon Gustavsson, and William D. Oliver. Tunable coupling scheme for implementing high-fidelity two-qubit gates. *Phys. Rev. Applied*, 10:054062, Nov 2018.

[105] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo. Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits. *Phys. Rev. Lett.*, 123:120502, Sep 2019.

[106] B Foxen, J Y Mutus, E Lucero, E Jeffrey, D Sank, R Barends, K Arya, B Burkett, Yu Chen, Zijun Chen, B Chiaro, A Dunsworth, A Fowler, C Gidney, M Giustina, R Graff, T Huang, J Kelly, P Klimov, A Megrant, O Naaman, M Neeley, C Neill, C Quintana, P Roushan, A Vainsencher, J Wenner, T C White, and John M Martinis. High speed flux sampling for tunable superconducting qubits with an embedded cryogenic transducer. *Superconductor Science and Technology*, 32(1):015012, dec 2018.

[107] P. Macha, S. H. W. van der Ploeg, G. Oelsner, E. Ilichev, H.-G. Meyer, S. Wnsch, and M. Siegel. Losses in coplanar waveguide resonators at millikelvin temperatures. *Applied Physics Letters*, 96(6):062503, 2010.

[108] 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.

[109] P. V. Klimov, J. Kelly, Z. Chen, M. Neeley, A. Megrant, B. Burkett, R. Barends, K. Arya, B. Chiaro, Yu Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, Erik Lucero, J. Y. Mutus, O. Naaman,

C. Neill, C. Quintana, P. Roushan, Daniel Sank, A. Vainsencher, J. Wenner, T. C. White, S. Boixo, R. Babbush, V. N. Smelyanskiy, H. Neven, and John M. Martinis. Fluctuations of energy-relaxation times in superconducting qubits. *Phys. Rev. Lett.*, 121:090502, Aug 2018.

[110] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia. Estimating the coherence of noise. *New Journal of Physics*, 17(11):113020, nov 2015.

[111] R. Barends, C. M. Quintana, A. G. Petukhov, Yu Chen, D. Kafri, K. Kechedzhi, R. Collins, O. Naaman, S. Boixo, F. Arute, K. Arya, D. Buell, B. Burkett, Z. Chen, B. Chiaro, A. Dunsworth, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, J. Kelly, P. V. Klimov, F. Kostritsa, D. Landhuis, E. Lucero, M. McEwen, A. Megrant, X. Mi, J. Mutus, M. Neeley, C. Neill, E. Ostby, P. Roushan, D. Sank, K. J. Satzinger, A. Vainsencher, T. White, J. Yao, P. Yeh, A. Zalcman, H. Neven, V. N. Smelyanskiy, and John M. Martinis. Diabatic gates for frequency-tunable superconducting qubits, 2019.

[112] H. M. I. Jaim, J. A. Aguilar, B. Sarabi, Y. J. Rosen, A. N. Ramanayaka, E. H. Lock, C. J. K. Richardson, and K. D. Osborn. Superconducting tin films sputtered over a large range of substrate dc bias. *IEEE Transactions on Applied Superconductivity*, 25(3):1–5, June 2015.

[113] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011.

[114] A. D. Córcoles, Jay M. Gambetta, Jerry M. Chow, John A. Smolin, Matthew Ware, Joel Strand, B. L. T. Plourde, and M. Steffen. Process verification of two-qubit quantum gates by randomized benchmarking. *Phys. Rev. A*, 87:030301, Mar 2013.

[115] P. J. J. O'Malley, J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, A. N. Cleland, and John M. Martinis. Qubit metrology of ultralow phase noise using randomized benchmarking. *Phys. Rev. Applied*, 3:044009, Apr 2015.

[116] Saleh Rahimi-Keshari, Matthew A. Broome, Robert Fickler, Alessandro Fedrizzi, Timothy C. Ralph, and Andrew G. White. Direct characterization of linear-optical networks. *Opt. Express*, 21(11):13450–13458, Jun 2013.