Multipartite Entanglement in Rabi Driven Superconducting Qubits

by

Marie Lu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Irfan Siddiqi, Chair
Professor Hartmut Haeffner
Professor Birgitta Whaley

Fall 2022

Multipartite Entanglement in Rabi Driven Superconducting Qubits

Abstract

Multipartite Entanglement in Rabi Driven Superconducting Qubits

by

Marie Lu

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Irfan Siddiqi, Chair

In harnessing quantum advantages for computation, there is a need for developing high fidelity operations on qubits. An algorithm can be broken down into single qubit operations and multi-qubit entangling gates. However, as the leading quantum processors today are limited to 50-100 qubits and each qubit is sensitive to decoherence noise (often referred to as NISQ era devices), running algorithms with long gate depth is difficult. Understanding the errors that plague existing gates and also expanding the dictionary of available gates is an important part of building a quantum computer. In this thesis we demonstrate two multiqubit gate experiments.

In the first experiment we demonstrate a multiqubit entangling gate for superconducting qubits on an all-to-all connected processor that draws upon the advantages of Rabi driven qubits. We also take inspiration from the ion qubit community by using a Mølmer-Sørensen-like interaction through the use of a shared coplanar waveguide (CPW) resonator driven superconducting qubits. We perform sensitivity analysis to understand the parameters that limit our gate fidelities.

In the second experiment we introduce and demonstrate a technique for scalable RB of many universal and continuously parameterized gate sets, using a class of circuits called randomized mirror circuits. The technique can be applied to a gate set containing an entangling Clifford gate and the set of arbitrary single-qubit gates, as well as gate sets containing controlled rotations about the Pauli axes. We use our technique to benchmark universal gate sets on four qubits, including a gate set containing a controlled-S gate and its inverse, and we investigate how the observed error rate is impacted by the inclusion of non-Clifford gates. We also show that our technique scales to many qubits with experiments on a 27-qubit IBM Q processor. We use our technique to quantify the impact of crosstalk on this 27-qubit device, and we find that it contributes approximately 2/3 of the total error per gate in random many-qubit circuit layers.

To all the people who fill my heart. Peace, love, stay hydrated.

Contents

List of Figures

List of Tables

Acknowledgments

This thesis was brought into the world through the kindness of many supportive hands and strokes of good luck. First, I would like to thank my advisor, Irfan, for the opportunity to be in this lab and his expert guidance. I also thank my thesis committee for their time reviewing this manuscript and providing feedback.

Thank you to my parents who are, in my eyes, heroes of the immigrant experience. They were the first people to encourage in me excitement and curiosity about the world because that was how they felt coming to the US. Most importantly, through their own actions, they taught me that no matter how hard a problem is, if I put my mind to it, I can figure it out.

I'm thankful for the mentors I've had along the way. I'm grateful for Mr. Doyle, my high school English teacher for helping me fall in love with reading and with people. Because at the end of the day no matter what I'm doing, it all comes back to the "human experience". Thank you to Rozina Tenenbaum for opening the world of music to me and showing me the value of patience. Thank you to all of my research advisors from high school through undergrad, who opened the doors of their labs to me: Bill Flounders, Jimmy Chang, Anna Szucs, Alex Rider, Professor Giorgio Gratta, and Professor Rana Adhikari. Thanks in particular to David Moore, my postdoc mentor during all four years of undergrad, who laid the groundwork of the way I approach experiments today. David, your unconditional patience and soft-spoken wisdom I sometimes did not deserve. I'm still sorry for breaking that expensive optical fiber.

This brings me to QNL and its incredible graduate students and postdocs. Thank you Sydney Schreppler for being the best mentor I could have asked for in my early years of grad school and continuing to show up for me even after your time at Berkeley. My lab motto is "What would Sydney do?", and to be honest, it has yet to fail me. Thank you Jean-Loup for bringing your openness of thought, insightfulness, and enthusiasm as a collaborator. I'd like to thank William Livingston, John Mark Kreikebaum, and Andrew Eddins for sharing their knowledge on all aspects of lab from cryogenics to measurement. They were always the first people I'd turn to for outside input and generously always put down whatever they were working on to help me. Also thank you to JM for managing the cleanroom in all its glory and all the folks who helped make the final samples for my experiment: Larry Chen and Christian Jünger. I'd like to thank MIT Lincoln Labs for providing the JTWPAs used in almost all of the measurements here. Thank you Brad Mitchell, Trevor Chistollini, Noah Goss for bring such joy and kindness to my lab experience. My initial years in QNL overlapped with Vinay Ramasesh, Shay Hacohen-Gourgy, Emmanuel Flurin, James Colless, Machiel Blok, and Kevin O'Brien laid the foundations for the kinds of 2D architectures used in this thesis. Without them I'd be trying to do my experiments in a 3D cavity and who knows how the crosstalk might have damaged my graduation date.

Thank you to my incredible theory collaborators at Sherbrooke University: Joachim Cohen, Alexandru Petrescu, and Professor Alexander Blais. And also to Chiara Pelletti and Alexei Marchenkov at Bleximo for designing the shielding in the fridges.

Thank you to Naomi for your sisterhood and always being a source of love and support and general shenanigans. To the SF Crown Cleaner Trio and its members Evan and Dylan

for encouraging me to rediscover playing music again after an 8 year hiatus and being a second fam. To Luna, Vikram, Alex for being my OG friends in the Bay Area. Thank you board game nights for the new friendships. I love you all so much.

Thank you to (#NotSponsored) City Dance and Mission City Swing for enabling my dance addiction, to Blick's for supplying me with discount art materials, to Berkeley Bowl for their produce aisles and, in particular the mushroom section. To Asian skin care products, kpop, soft boiled eggs, Megan the Stallion, therapy, and cats.

Chapter 1

Introduction

Computers are an invaluable part of our lives, integrated into everything from our handheld phones to supercomputers that help run some of the most powerful computing centers. However there are still some problems that even most power computers today cannot solve. Theory developments have shown that exchanging the transistors in classical computers for new materials with quantum properties could help make previous intractable problems possible.

However this change in the fundamental building blocks of the computer necessitate a complete rework of all the higher layers of the computing stack, from gates all the way to the language that interfaces with users. There are many models of quantum computers. Among the most popular of models is the gate based quantum computer. Must like a classical computer, algorithms are mapped into operations on a register of N qubits. Another popular model is to use quantum bits to model more complex materials and simulate the detailed dynamics of systems that typically can only be measured for bulk behaviour.

Since the proposal of these new methods of computing, huge resources have been invested into developing all aspects of the quantum computer, from the base level bits to the algorithms and languages that users might interface with. The end goal is general fault tolerant quantum computations. The devices that are being made now are currently in the NISQ era, which stands for Noisy Intermediate Scale Quantum devices. The types of challenges to be faced depend on the qubit platform. Currently, various types of materials being considered for the hardware, each with its own advantages challenges. One of the most important yet conflicting properties of a qubit is its relationship to its environment. On one hand, a qubit cannot be to strongly coupled to its surroundings. Stray electromagnetic fields, cosmic rays, and material defects can all be sources that destroy the information stored in the bit. However, a completely isolated qubit sitting in vacuum is also useless in terms of computation. The qubit must have some interaction with the environment to allow its state to be manipulated and measured and to interact with other qubits in order to perform operations. These conflicting needs is one of the hardest parts of constructing a qubit. Among the contenders are superconducting qubits, atoms, ions, photons, quantum dots, nitrogen vacancy centers, and majoranas. Atoms, for instance could be great for storing information because they do

not interact with the environment very much, minimizing sources of decay and decoherence. But at the same time this same property makes it hard to control and interact with for faster gates. Superconducting qubits are also a top contender with its fast gate times, but it has lower connectivity between qubits and requires the processor to be at mK temperatures, which can be a challenge when scaling up to millions of qubits. Photons are great a traveling long distances in fiber optic cables and do not require superconducting temperatures but one of the challenges is the amount of loss they experience going from the fibers to chips. One of the visions for quantum computers is that there might not be one single type of bit to do everything, but rather we might have to employ a hybrid approach where some types of bits are used for storing information and others are used for computation or building far reaching networks.

The idea of observing quantum behavior in a macroscopic entity was first in the 1980s and carried out in experiment in 1986 when a layer of insulating material was sandwiched between two superconductors. Unlike other qubit proposals, superconducting qubits are macroscopic. This device, known as a Josephson junction, exhibited tunnel behavior of its electrons when cooled to superconducting temperatures. This component later became the fundamental building block of one of the most popular superconducting qubits used today, the transmon. Through focused efforts various versions of the qubit similar to the transmon that all utilize the Josephson Junction have been realized, each with increasing lifetimes up to hundreds of microseconds. These growing lifetimes along with the flexibility of having individual qubit control and readout and qubit structure design are all part of what make superconducting qubits so attractive.

At the same time, there is work to be done on superconducting qubits before achieving fault tolerance. For instance, despite the recent improvements, we must continue to grow qubit lifetimes, reduce qubit crosstalk, understand material defects that cause qubits to couple to unwanted two-level-systems, and many other technical engineering challenges that come with scaling up. Some of these include figuring out how make a dilution refrigerator that is large enough and powerful enough to cool down 1 million superconducting qubits. In this thesis, we will focus on the area of how to improve the native gates. This encompasses understanding the sources of error plaguing existing gates, which will touch on in the last chapter of this thesis. The majority of this thesis, though, will be dedicated to expanding the dictionary of gates, in an attempt to provide more options when breaking down an algorithm into a series of gates. Superconducting qubits lack the higher connectivity between qubits as experienced by other platforms such as ion qubits. Superconducting qubits are printed on the 2-D surface of a silicon wafer which means that generating couplings between neighboring qubits is easy but connecting a qubit on one side of the chip to one physically farther away is difficult. In addition, high connectivity typically exacerbates crosstalk challenges. For this reason, the vast majority of super conducting gates are two qubit gates. Entanglement between qubit numbers greater than two is achieved often through cascaded pairwise interactions.

In this thesis look towards other platforms for advantages and inspiration. Specifically, we work with superconducting qubits but draw inspiration from the ion qubit and NMR

communities. Ion qubits have natural all-to-all connectivity due the laser that hold them all in place. This allows them efficiently to do multi-qubit entangling operations with high fidelity. Furthermore, the NMR community has developed dynamical decoupling techniques that help isolate qubits from noise. We combine these advantages with the superconducting qubit system to expand the toolbox of available gates to prepare a maximally entangled state.

1.1 Structure of Thesis

The rest of this thesis is structured as follows:

In chapter 2, we discuss some of the fundamental building blocks of superconducting qubit systems. This includes the various components that make up a qubit and also commonly used state control techniques. We also cover some concepts that are revisited later in the thesis. In particular this includes the Rabi dressed frame and dynamical decoupling, as the final version of the gate is operated in the dressed frame.

In chapter 3, we introduce the ion Mølmer-Sørensen gate. The chapter begins with key components of the ion gate and motivates why such a gate is advantageous. We then discuss the differences in the ion and superconducting platforms that make a direct translation of the gate infeasible. Finally, we bring back some concepts from chapter 2, such as the Rabi drive, that allow us to adapt the ion gate to superconducting qubit systems. We show that the Rabi drive is not only a tool to help us adapt the gate but also a helpful feature that provides the gate numerous advantages.

Chapter 4 introduces the hardware used to implement the gate. This includes processor design, dilution fridge wiring details, and amplifiers used. Hopefully this section will be useful reference to anyone who wishes to would like to design microwave components or run a dilution refrigerator.

Chapter 5 delves into the details of characterizing the chip and calibrating for the gate. This section is where we call upon the spinlocking technique from the NMR community previously introduced in Chapter 2. We pair the spinlocking measurement the with resonator drives used for the gate to obtain single qubit resonances, which we use to extract qubit-resonator couplings and photon numbers–important parameters for the gate fidelity. Additionally, we detail calibration procedures for the Rabi drive powers and resonator drive phases.

In chapter 6, we show implementations of the gate for 2, 3, and 4 qubits and describe how we characterize the gate. This section could be useful for anyone hoping to learn how to implement basic tomographic methods. We show results from state and process tomography. We detail how we analyze the effect of the Rabi drive on each qubit appears as a global phase on the gate. In addition, we explore the various factors that could be limiting the gate and propose an alternative version of this gate that could solve several of its challenges.

Finally in chapter 7, we switch gears from focusing on the intricacies of errors on specific gate to looking at methods of characterizing errors on a broader scale. We discuss the

limitations in the current status quo of validation techniques, such as Randomized Benchmarking. We motivate a new kind of circuit construction as a way of making randomized benchmarking more scalable while ensuring its ability to capture noise. We demonstrate that randomized mirror circuits perform just as well as standard Randomized benchmarking by using both methods to characterize four qubits on our eight-qubit processor. Furthermore, we use the benchmarking data to quantify the amount of crosstalk error by fitting to error models. Finally, we demonstrate that Mirror Randomized Benchmarking is scalable by demonstrating it on IBM's 27 qubit processor.

Chapter 2

Superconducting Circuits and Techniques

A quantum computer is composed of two-level systems that are used to store information and serve as a quantum memory. Unlike a classical bit, the quantum bit has a phase component and is expressed as

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2.1}$$

where $\alpha = \cos\frac{\theta}{2}$ and $\beta = e^{i\phi}\sin\frac{\theta}{2}$ are allowed to be complex values and $|\alpha|^2$ and $|\beta|^2$ represent the probability of being in the $|0\rangle$ and $|1\rangle$ states, respectively. These states are often visualized on the Bloch sphere. The angles ϕ and θ determine the azimuthal and polar angles on the Bloch sphere. The north and south pole represent the $|0\rangle$ and $|1\rangle$ states with $\alpha = 1$ or $\beta = 1$, respectively. The coordinates that represent the state on the Bloch sphere are

$$\vec{a} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \tag{2.2}$$

States with coordinates that land on the surface of the Bloch sphere are called pure states. States that are inside the sphere are called mixed states. Imperfect qubits in the NISQ era are not fixed in position on the Bloch sphere. There are mechanisms due to a qubit's interactions with its environment that can cause the Bloch vector to change unexpectedly.

The purpose of the rest of this section is to introduce the basic building blocks that going into superconducting circuits. There are a few key properties that a qubit must have in order to function for computation. These include

1. two discrete states that serve as the computational basis
2. a method to manipulate the state of the qubit
3. a method to readout the state of the qubit
4. a coupling mechanism to connect qubits

The sections below should touch on each of these aspects.

2.1 Quantum LC Circuit

The most important feature of a qubit is distinctly addressable quantized energy levels. Photon qubits use the choice of photon pathway, as measured by a detector, as the qubit states. Many qubit platforms use energy levels of electron orbitals or nuclear spins found in naturally occurring atoms and ions. However, naturally occurring energy levels in atoms are often chosen among of forest of energy levels. The complex level structure provides opportunities for unwanted transitions. There have been efforts to develop artificial atoms where the level structure is simplified.

One of the simplest models of quantized energy levels is the quantum harmonic oscillator. We show that this model is equivalent to the quantum LC circuit, which is a lumped element representation of artificially made components commonly found in superconducting circuit systems. However, we also show it is not quite sufficient to be a qubit, although it will reappear later in this chapter as a useful element for readout and qubit-qubit coupling.

The LC circuit can be first classically described the the sum of it's energies

$$H = \frac{1}{2}CV^2 + \frac{1}{2}LI^2 \tag{2.3}$$

as we sum the energies from both the capacitor and the inductor. We can rewrite this in terms of the charge and magnetic flux by substituting in $V = Q/C = \dot{\Phi}$ and $I = \Phi/L$:

$$H = \frac{Q^2}{2C} + \frac{\Phi^2}{2L} \tag{2.4}$$

The first step of quantizing this circuit is to place hats on H, Q, and Φ to turn them into operators that have the commutation relation $[\Phi, Q] = i\hbar$. From here, it is noticeable that the equation has a very similar form to the quantum harmonic oscillator ($\hat{H} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2$, allowing us to follow the second quantization procedure of substituting in analogous ladder operators

$$a = \frac{1}{\sqrt{\hbar\omega}}\left(\frac{\hat{\Phi}}{2L} - i\frac{\hat{Q}}{2C}\right) \tag{2.5}$$

$$a^\dagger = \frac{1}{\sqrt{\hbar\omega}}\left(\frac{\hat{\Phi}}{2L} + i\frac{\hat{Q}}{2C}\right) \tag{2.6}$$

where $\omega = 1/\sqrt{LC}$ and have the commutation relation $[a, a^\dagger] = 1$. Making these substitutions, one arrives at precisely the form form of the Hamiltonian for the quantum harmonic oscillator, as expected

$$\hat{H} = \hbar\omega(a^\dagger a + \frac{1}{2}) \tag{2.7}$$

While this does conveniently have a clear level structure, with equal energy spacing, it is inconvenient to use as qubit. The equal energy level spacing makes it impossible to uniquely

Figure 2.1: Cross-section visualization of josephson junction

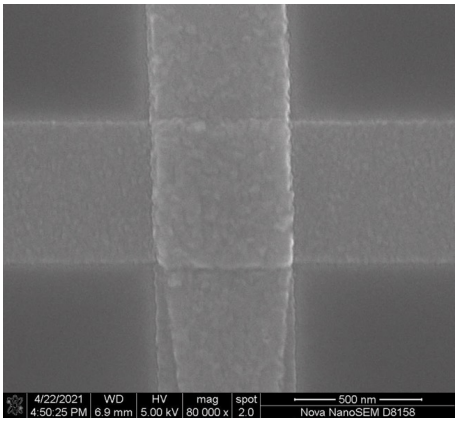

Figure 2.2: Image of actual Josephson junction after fabrication. Two strips of aluminum are evaporated on with an oxidation step in between. Top view.

isolate to energy levels to address as a qubit. Any photons that excites the $0 \longrightarrow 1$ transition will also be able to excite up the ladder. A coherent tone from a generator will drives all of the transitions simultaneously, resulting in a probability of occupying each state in a Poissonian distribution. In order to select out just a pair of energy levels, we need energy levels with unique energy separations. The LC circuit must have some additional nonlinearity to generate anharmonicity.

2.2 Josephson Relations

We add a nonlinear element to the circuit by replacing the linear inductor with a new element called the Josephson junction. The Josephson junction is composed of two superconducting layers (typically aluminum) that sandwich an insulating layer, often an oxide. When aluminum is cooled down below the T_C, 1.2K, the electrons experience an attractive potential

to the lattice phonons that overcomes the repulsive Coulomb potential that normally repels electrons from each other. Thus, the electrons with opposite momenentum at the Fermi surface pair up into Cooper pairs, forming a composite boson. These pairs condense into a Bose-Einstein superfluid condensate and carry charge through the metal without dissipation. The cooper pairs are also able to tunnel from one superconducting layer to the other layer, through the insulating oxide. Here we derive the Josephson relations that describe the dynamics of cooper pairs tunneling in Josephson junctions. Starting from Schrödingers equation, we obtain a set of coupled equations for the bulk wave functions of the Cooper pairs on either side of the insulating barrier:

$$\begin{aligned} i\hbar\frac{\partial\Psi_1}{\partial t} &= U_1\Psi_1 + K\Psi_2 \\ i\hbar\frac{\partial\Psi_2}{\partial t} &= U_2\Psi_2 + K\Psi_1 \end{aligned} \tag{2.8}$$

where $\Psi_i = \sqrt{n_i}e^{i\phi_i}$. Here, U_i are the energies of the cooper pairs, also known as the chemical potential. Since each cooper pair contains two electrons, The energy is $2eV$, where V is the potential across the junction $V_2 - V_1$. For convenience, we choose 0V as the middle between V_1 and V_2. Thus the above equations become

$$\begin{aligned} i\hbar\frac{\partial\Psi_1}{\partial t} &= eV\Psi_1 + K\Psi_2 \\ i\hbar\frac{\partial\Psi_2}{\partial t} &= -eV\Psi_2 + K\Psi_1 \end{aligned} \tag{2.9}$$

We substitute our expression of Ψ_i into Eq. (2.9) to obtain

$$\begin{aligned} i\hbar\left[\frac{1}{2\sqrt{n_1}}\dot{n_1}e^{i\phi_1} + \sqrt{n_1}(i\dot{\phi_1})e^{i\phi_1}\right] &= eV\sqrt{n_1}e^{i\phi_1} + K\sqrt{n_2}e^{i\phi_2} \\ i\hbar\left[\frac{1}{2\sqrt{n_2}}\dot{n_2}e^{i\phi_2} + \sqrt{n_2}(i\dot{\phi_2})e^{i\phi_2}\right] &= -eV\sqrt{n_2}e^{i\phi_2} + K\sqrt{n_1}e^{i\phi_1} \end{aligned} \tag{2.10}$$

which simplifies to

$$\begin{aligned} i\hbar\left[\frac{1}{2\sqrt{n_1}}\dot{n_1} + \sqrt{n_1}(i\dot{\phi_1})\right] &= eV\sqrt{n_1} + K\sqrt{n_2}e^{i(\phi_2-\phi_1)} \\ i\hbar\left[\frac{1}{2\sqrt{n_2}}\dot{n_2} + \sqrt{n_2}(i\dot{\phi_2})\right] &= -eV\sqrt{n_2} + K\sqrt{n_1}e^{-i(\phi_2-\phi_1)} \end{aligned} \tag{2.11}$$

Expressing the exponential quantities in terms of sines and cosines

$$\begin{aligned} \frac{1}{2n_1}\dot{n_1} + (i\dot{\phi_1}) &= \frac{eV}{i\hbar} + \frac{K}{i\hbar}\sqrt{\frac{n_2}{n_1}}[\cos(\phi_2-\phi_1) + i\sin(\phi_2-\phi_1)] \\ \frac{1}{2n_2}\dot{n_2} + (i\dot{\phi_2}) &= -\frac{eV}{i\hbar} + \frac{K}{i\hbar}\sqrt{\frac{n_1}{n_2}}[\cos(\phi_2-\phi_1) - i\sin(\phi_2-\phi_1)] \end{aligned} \tag{2.12}$$

We equate the real and imaginary parts on either side of the equations to obtain

$$\begin{aligned}
\dot{n_1} &= \frac{2K}{\hbar}\sqrt{n_1 n_2}\sin(\phi_2 - \phi_1) \\
\dot{n_2} &= -\frac{2K}{\hbar}\sqrt{n_1 n_2}\sin(\phi_2 - \phi_1) \\
\dot{\phi_1} &= -\frac{eV}{\hbar} - \frac{K}{\hbar}\sqrt{\frac{n_2}{n_1}}\cos(\phi_2 - \phi_1) \\
\dot{\phi_2} &= \frac{eV}{\hbar} - \frac{K}{\hbar}\sqrt{\frac{n_2}{n_1}}\cos(\phi_2 - \phi_1)
\end{aligned} \tag{2.13}$$

We subtract the pairs of equations for $\dot{\phi_i}$ and substitute in $\phi = \phi_2 - \phi_1$

$$I = e\dot{n_1} = -e\dot{n_2} = \frac{2eK}{\hbar}\sqrt{n_1 n_2}\sin\phi = I_c \sin\phi \tag{2.14}$$

$$\dot{\phi} = \frac{2eV}{\hbar} \tag{2.15}$$

which are the Josephson relations. J_C is the maximum current that can pass through the junction without dissipation. Together these equations describe the behavior of the electrons going through the thin insulator. The first Josephson relation informs us that a current can flow through the barrier even with no applied voltage, just from having a phase difference between the two superconducting islands. This is called the DC Josephson effect. The second relation tells us that with a voltage applied, the phase difference will change as a function of time. This will cause the current through the barrier to oscillate, hence it's name, the AC Josephson effect[1].

We will derive the expression for the induction and show that, unlike that of a regular inductor, it is nonlinear. We start by taking the time derivative of the first Josephson relation:

$$\frac{\partial I}{\partial t} = I_C \cos\phi \frac{\partial \phi}{\partial t} \tag{2.16}$$

Recall that $\frac{\partial \phi}{\partial t}$ is known from the second Josephson relation. Making the substitution gives

$$\frac{\partial I}{\partial t} = \frac{2\pi I_C V}{\Phi_0}\cos\phi \tag{2.17}$$

where we have used the definition of the flux quantum $\Phi_0 = \frac{h}{2e}$. Recall that $V = L\frac{\partial I}{\partial t}$ so rearranging the previous expression gives

$$V = \frac{\Phi_0}{2\pi I_C \cos\phi}\frac{\partial I}{\partial t} \tag{2.18}$$

[1]Dearest reader, thank you for spending your precious time on this mediocre text. This footnote has no purpose other than to provide useless trivia about my third great love: kpop. If you see this footnote, feel free to let me know at marielu94@gmail.com. It will tickle me greatly. Kpop fun fact 1/7: The first known usage of the term kpop appeared in on *Billboard* in 1999.

and thus

$$L = \frac{\Phi_0}{2\pi I_C \cos\phi} \tag{2.19}$$

2.3 Quantization of the Qubit Hamiltonian

Now that we have a nonlinear element, the Josephson junction, we return to the task of generating a level structure with unique energy level spacing. We can arrange the JJ and capacitor together in a variety of combinations, each producing a unique type of qubit. As mentioned before, designing a qubit is often a balancing acting between allowing for control and protecting from noise. Different structures of Josephson junctions and capacitor combinations are protected from different kinds of noise. In Fig. 2.3 we highlight some of the common types of superconducting qubits. At the time of writing this thesis, we mainly use the fixed transmon, tunable transmon, flux qubit, and fluxonium in the lab. The transmon is an modification of the original cooper pair box with a large shunting capacitor to reduce the qubit's sensitivity to charge noise [38]. There are both tunable and fixed frequency transmons. There is trade-off for this capacitor, however. The increased capacitance also slightly reduces the anharmonicity, which exacerbates frequency crowding issues. Tunable transmons have the advantage of being able to better avoid frequency crowding since their frequencies are adjustable. However, the flux coupling that allows for tuning also introduces a new pathway for flux noise, typically causing shorter lifetimes. The Xmon is similar to transmons but has different capacitor geometries. Alternatively, fluxonium is an exciting candidate for future work. Instead of shunting with a large capacitor, it has a shunting inductor composed of a chain of small Josephson junctions. This protects the qubit from charge noise, just like the transmon, while maintaining a large anharmonicity. Furthermore the inductive loop allows the frequency to be tuned. This typically makes the qubit sensitive to flux noise as well, but the qubit has a flux noise insensitive sweet spot when half of a magnetic flux quantum is threaded through the loop. Here it exhibits long coherence times. In the following derivation and the rest of this thesis, we will be using the fixed frequency transmon. To better understand the level structure, we start with the familiar Hamiltonian for a harmonic oscillator

$$H = \frac{Q^2}{2C} + \frac{\Phi^2}{2L} \tag{2.20}$$

but for our circuit we replace a linear inductor with a nonlinear inductor, the Josephson junction.

We make a change of variables here using $\Phi = \Phi_0\phi$. Now to find the qubit Hamiltonian,

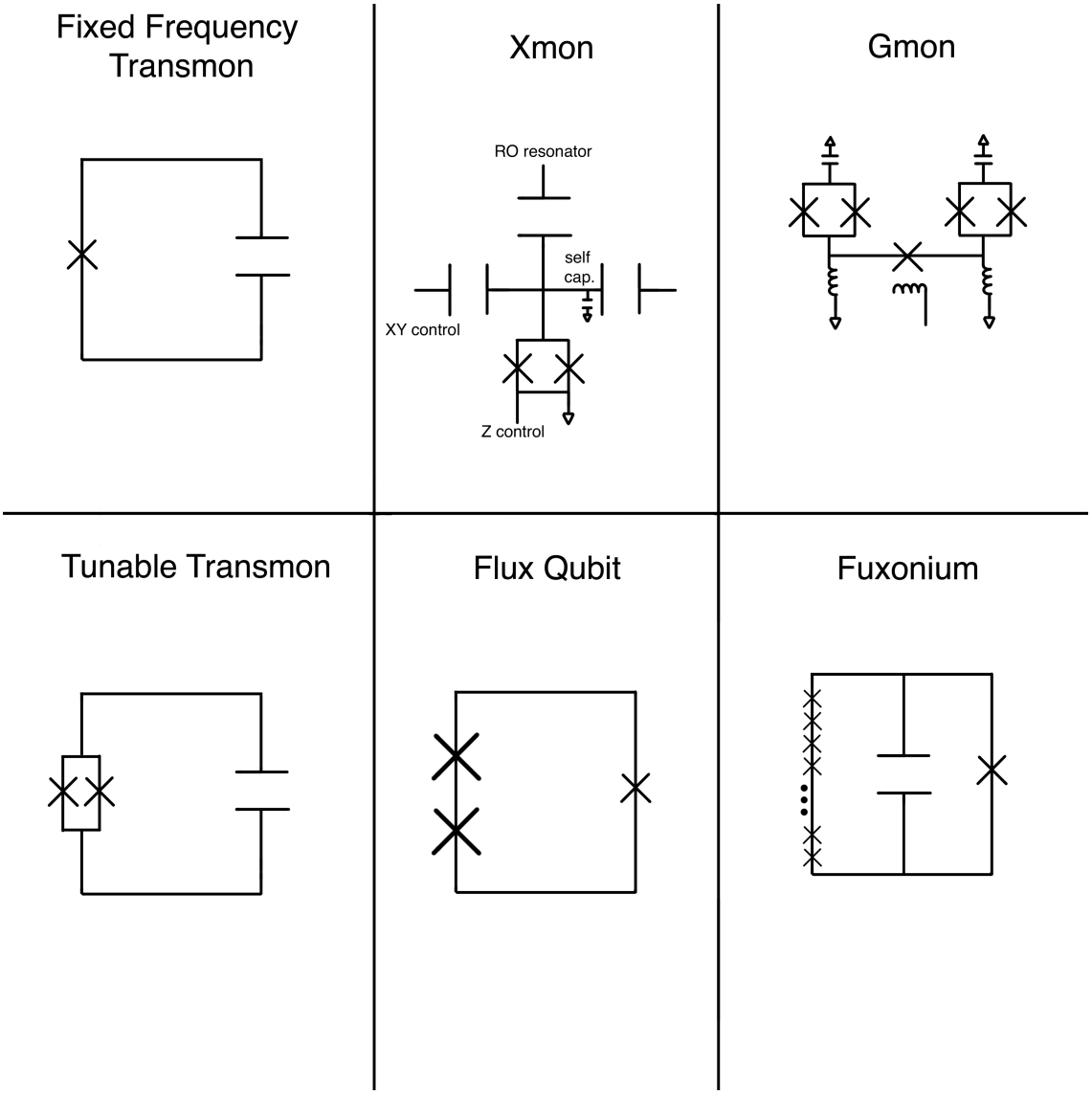

Figure 2.3: A table of some of the common types of superconducting qubits.

recall that the power $P = IV$ so to find the energy we integrate [6]

$$\begin{aligned}\hat{H} = \int IV dt &= \int I_c \sin\left(\frac{\Phi}{\Phi_0}\right)\frac{\hbar\dot{\Phi}}{2e}dt \\ &= \frac{I_c\hbar}{2e}\int \sin\left(\frac{\Phi}{\Phi_0}\right)d\Phi \\ &= -\frac{\hbar I_c\Phi_0}{2e}\cos\left(\frac{\Phi}{\Phi_0}\right) \\ &= -E_J\cos\left(\frac{\Phi}{\Phi_0}\right)\end{aligned}$$

Now the total qubit Hamiltonian is

$$\hat{H} = \frac{Q^2}{2C} - E_J\cos\left(\frac{\Phi}{\Phi_0}\right) \tag{2.21}$$

with $E_J = \frac{\hbar I_c\Phi_0}{2e}$. In the first quantization step, we promote the conjugate variables Q and Φ to operators

$$\hat{H} = \frac{\hat{Q}^2}{2C} - E_J\cos\left(\frac{\hat{\Phi}}{\Phi_0}\right) \tag{2.22}$$

and make the following substitutions:

$$\begin{aligned}\hat{n} &= \frac{\hat{Q}}{2e} \\ \hat{\phi} &= \frac{2\pi\Phi}{\Phi_0}\end{aligned}$$

such that the Hamiltonian becomes

$$\hat{H} = \frac{4e^2}{2C}\hat{n}^2 - E_J\cos\hat{\phi} \tag{2.23}$$

and use the substitution $E_C = \frac{e^2}{2C}$ such that

$$\hat{H} = 4E_C\hat{n}^2 - E_J\cos\hat{\phi} \tag{2.24}$$

Similar to the quantum harmonic oscillator derivation, we perform the second quantization by defining $\hat{n}$ and $\hat{\phi}$ in term of ladder operators:

$$\begin{aligned}\hat{n} &= in(\hat{c}^\dagger - \hat{c}) \\ \hat{\phi} &= \varphi(\hat{c}^\dagger + \hat{c}) \\ n &= \left(\frac{E_J}{32E_C}\right)^{\frac{1}{4}} \\ \varphi &= \left(\frac{2E_C}{E_J}\right)^{\frac{1}{4}}\end{aligned}$$

With this substitution, the Hamiltonian becomes

$$\hat{H} = -4E_C n^2(\hat{c}^\dagger - \hat{c})^2 - E_J \cos\big(\varphi(\hat{c}^\dagger + \hat{c})\big) \tag{2.25}$$

This is the general Hamiltonian for a Josephson junction and a capacitor, commonly known as the Cooper Pair Box. The values of the E_J and E_C can be adjusted to protect from charge noise. In this lab, we use the transmon regime [38], where $E_J >> E_C$ so $\varphi << 1$ and φ is a small parameter. This allows us to Taylor expand the cos term:

$$\hat{H} = -4E_C n^2(\hat{c}^\dagger - \hat{c})^2 - E_J\left[1 - \frac{\varphi^2(\hat{c}^\dagger + \hat{c})^2}{2!} + \frac{\varphi^4(\hat{c}^\dagger + \hat{c})^4}{4!} + ...\right] \tag{2.26}$$

$$\approx -4E_C n^2((\hat{c}^\dagger)^2 + \hat{c}^2 - \hat{c}^\dagger\hat{c} - \hat{c}\hat{c}^\dagger) - E_J\left[1 - \frac{1}{2}\varphi^2((\hat{c}^\dagger)^2 + \hat{c}^2 + \hat{c}^\dagger\hat{c} + \hat{c}\hat{c}^\dagger) + \frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4\right] \tag{2.27}$$

Note that $4E_C n^2 = 4E_C\sqrt{\frac{E_J}{32E_C}} = \sqrt{\frac{16E_C E_J}{32}} = \sqrt{\frac{E_C E_J}{2}}$ and $\frac{1}{2}E_J\varphi^2 = \frac{1}{2}E_J\sqrt{\frac{2E_C}{E_J}} = \sqrt{\frac{2E_C E_J}{4}} = \sqrt{\frac{E_J E_C}{2}} = 4E_C n^2$. Using these identities, we simplify the Hamiltonian

$$\begin{aligned}
\hat{H} - E_J + &\approx 2\sqrt{\frac{E_C E_J}{2}}(\hat{c}^\dagger\hat{c} + \hat{c}\hat{c}^\dagger) - E_J\frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4 \\
= -E_J + &\approx 2\sqrt{\frac{E_C E_J}{2}}(\hat{c}^\dagger\hat{c} + 1 + \hat{c}^\dagger\hat{c}) - E_J\frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4 \\
= -E_J + &\approx 2\sqrt{\frac{E_C E_J}{2}}(2\hat{c}^\dagger\hat{c} + 1) - E_J\frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4
\end{aligned}$$

We remove the constant offsets and simplify

$$\begin{aligned}
\hat{H} = -E_J + &\approx 4\sqrt{\frac{E_C E_J}{2}}\hat{c}^\dagger\hat{c} - E_J\frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4 \\
= -E_J + &\approx \sqrt{8E_C E_J}\hat{c}^\dagger\hat{c} - E_J\frac{\varphi^4}{24}(\hat{c}^\dagger + \hat{c})^4
\end{aligned}$$

Next, using the rotating wave approximation, we drop any terms that have unequal numbers of $\hat{c}^\dagger$ and $\hat{c}$:

$$\hat{H} = -E_J + \approx \sqrt{8E_C E_J}\hat{c}^\dagger\hat{c} - E_J\frac{\varphi^4}{24}((\hat{c}^\dagger)^2\hat{c}^2 + \hat{c}^2(\hat{c}^\dagger)^2 + \hat{c}^\dagger\hat{c}^2\hat{c}^\dagger + \hat{c}(\hat{c}^\dagger)^2\hat{c})$$

Using commutation relation $[\hat{c}, \hat{c}^\dagger] = 1$, we combine the various terms and simplify. We arrive at

$$\hat{H} = \left(\omega_0 + \frac{1}{2}\right)\hat{c}^\dagger\hat{c} + \frac{\delta}{2}\left(\hat{c}^\dagger\hat{c}\right)^2 \tag{2.28}$$

Applying the definition $\hat{c} = \sum_j \sqrt{j+1}\,|j+1\rangle\langle j|$, we rewrite the Hamiltonian in terms of $|j\rangle$, it's various energy levels:

$$\hat{H} = \sum_j \left[\left(\omega_0 + \frac{\delta}{2}\right) j + \frac{\delta}{2} j^2\right] |j\rangle\langle j| \tag{2.29}$$

We now see that the energy levels are anharmonically spaced, allowing us to select the bottom two energy levels to be the computational basis of the qubit.

2.4 Controlling Qubit Population

Now that we have a qubit, the next step is to show that we can manipulate the state of the qubit. One drive that appears repeatedly in the rest of this manuscript is the Rabi drive. This is a semi-classical process. When an electric field, a classically generated pulse, is applied to the qubit closely resonant with the qubit's transition frequency, ω_q, the qubit will experience populations swaps between the ground and excited states. To see this we start with the Hamiltonian of the qubit and a drive

$$\hat{H}_T = \hat{H}_0 + \hat{H}_d \tag{2.30}$$
$$\hat{H}_0 = \omega_q \sigma_z \tag{2.31}$$
$$\hat{H}_d = \vec{d} \cdot \vec{E} \tag{2.32}$$

where $\vec{d} = e\vec{r}$ is the dipole of the qubit and $\vec{E} = (\mathcal{E}e^{-i\omega_d t} + \mathcal{E}^* e^{i\omega_d t})\hat{\epsilon}$. Recall that the qubit does not have an intrinsic dipole moment[2] so the only dipole matrix elements that appear are the ones that involve different energy levels: $\mu_{ij} = \langle j|\,\vec{r}\cdot\vec{\epsilon}\,|i\rangle$ for $i \neq j$. Expanding $\hat{H}_d$ we have

$$\begin{aligned}
\hat{H}_d &= \left(\mu_{ge}\,|e\rangle\langle g| + \mu_{eg}\,|g\rangle\langle e|\right)\left(\mathcal{E}e^{-i\omega_d t} + \mathcal{E}^* e^{i\omega_d t}\right) \\
&= \left(\mu_{ge}\sigma_+ + \mu_{eg}\sigma_-\right)\left(\mathcal{E}e^{-i\omega_d t} + \mathcal{E}^* e^{i\omega_d t}\right) \\
&= \mu_{ge}\sigma_+\mathcal{E}e^{-i\omega_d t} + \mu_{eg}\sigma_-\mathcal{E}e^{-i\omega_d t} + \mu_{ge}\sigma_+\mathcal{E}^* e^{i\omega_d t} + \mu_{eg}\sigma_-\mathcal{E}^* e^{i\omega_d t}
\end{aligned}$$

If we take H_T to the interaction picture by using the transformation $U = e^{i\omega_q\sigma_z t}$ we obtain

$$\tilde{H}_T = \mu_{ge}\sigma_+\mathcal{E} + \mu_{eg}\sigma_-\mathcal{E} + \mu_{ge}\sigma_+\mathcal{E}^* e^{2i\omega_d t} + \mu_{eg}\sigma_-\mathcal{E}^* e^{2i\omega_d t}$$

We apply the rotating wave approximation (RWA) and drop the terms oscillating at $2\omega_d$ and return back to the original frame:

$$\hat{H}_T = \omega_q\sigma_z + \Omega\sigma_+ e^{-i\omega_d t} + \Omega^*\sigma_- e^{i\omega_d t} \tag{2.33}$$

[2]Kpop fun fact 2/7: Psy's video for Gangnam Style was the first Youtube video to reach 1 billion views.

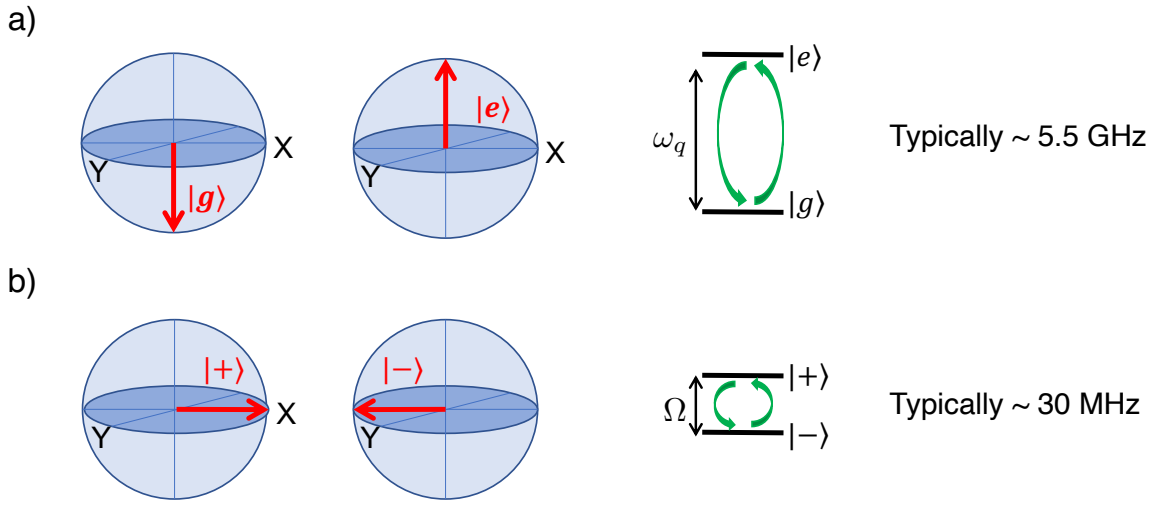

Figure 2.4: a) Original qubit basis on the Bloch sphere b) Dressed states as new basis states that result from a resonant qubit drive.

where we have made the replacement $\Omega = \mu\mathcal{E}$. If we now go to the frame of the drive, then the Hamiltonian becomes

$$\hat{H}_T = \delta\sigma_z + \Omega\sigma_+ + \Omega^*\sigma_- \tag{2.34}$$

where $\delta = \omega_q = \omega_d$. We can put this into the Schrödinger equation and use a general form of the wave function $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$. This gives us a set of coupled equations for c_1 and c_2

$$i\dot{c_0} = \Omega c_1 \tag{2.35}$$
$$i\dot{c_1} = c_0\Omega^* - \delta c_1 \tag{2.36}$$

We will consider these coupled equations in two regimes.

2.4.1 Near detuned regime

We set $\delta = 0$ and take second derivative of the coupled equations above to obtain

$$i\ddot{c_0} = \Omega\dot{c_1} \tag{2.37}$$
$$i\ddot{c_1} = \dot{c_0}\Omega^* \tag{2.38}$$

and substituting in our coupled equations from above gives

$$\ddot{c_0} = -|\Omega|^2 c_0 \tag{2.39}$$
$$\ddot{c_1} = -|\Omega|^2 c_1 \tag{2.40}$$

This tells us that the amplitudes evolve as sines and cosines as follows:

$$|\Psi\rangle = \cos(|\Omega|t)\,|1\rangle + i\frac{\Omega}{|\Omega|}\sin(|\Omega|t)\,|0\rangle \tag{2.41}$$

The frequency of the oscillations depends on Ω which is proportional to the strength of the electric field, an experimentally controllable parameter. Thus by choosing the length and amplitude of the applied microwave field, we can select the final state of the qubit.

In addition, another useful picture to interpret the Hamiltonian in is the frame of the drive. We apply $U = e^{i\omega_d\sigma_z t}$ to obtain

$$\hat{H} = (\omega_q - \omega_d)\sigma_z + \Omega\sigma_x \tag{2.42}$$

We re-diagonalize the Hamiltonian and solve for new basis states which are superpositions of the original $|g/e\rangle$ states:

$$|\pm\rangle = \frac{|g\rangle \pm |e\rangle}{\sqrt{2}} \tag{2.43}$$

Most importantly, in the limit of $\delta << \Omega$, the energy separation between these states is given by

$$\epsilon_\pm = -\frac{\delta}{2} \mp \sqrt{\left(\frac{\delta}{2}\right)^2 + \Omega^2} \tag{2.44}$$

with $\delta = \omega_q - \omega_d$. This basis is typically called the Rabi Dressed basis and can form a new effective qubit with energy levels that are experimentally tunable, as they largely depend on Ω. In experiments, we aim for $\delta = 0$.

2.4.2 The far detuned regime $\omega_d >> \delta >> \Omega$

We note here that above, we have dealt with the near detuned regime when the drive frequency is close to the natural qubit transition. We also touch upon the far detuned regime when the drive is very far from the qubit frequency. This will be technically relevant in later parts of this thesis. When doing gates on qubits, there are often many tones present, not only driving the qubit directly, that could be at a wide range of frequencies. It is important to understand how these tones affect the qubit.

When the drive frequency is far, we use a method called adiabatic elimination. Intuition tells us that the populations will rarely transfer to the excited state, meaning $|c_0(t)| \approx 1$ and $|c_1(t)| \approx << 1$. Given this insight, we set $c_0(t) = 1$ in Eq. (2.36) such that $\dot{c_1}(t) = i\delta c_1(t) + i\Omega$. We an integrate this expression to get the population in the excited state as a function of time

$$c_1(t) = -\frac{\Omega}{\delta}(1 - e^{i\delta t}). \tag{2.45}$$

Now we do something similar for the excited state population. Since our detuning is big, $i\delta c_1$ is large, meaning $c_1(t)$ will be oscillating very fast, overall averaging out to zero. Hence we

set $\dot{c_1}(t) = 0$. This gives the relation $c_1 = \frac{\Omega^*}{\delta} c_0$. We can substitute this back into Eq. (2.36) for $i\dot{c_0} = |\Omega|^2/\delta c_0$ – an easily integrable expression:

$$c_0(t) = c_0(t=0)e^{-i\frac{|\Omega|^2}{\delta}t} \tag{2.46}$$

The lower state energy appears as if it has undergone an energy shift by $\frac{|\Omega|^2}{\delta}$, an effect known as the AC Stark shift. Intuitively speaking, applying a light tone to the qubit that is far from the qubit's transition frequency, shifts qubit frequency by $\frac{|\Omega|^2}{\delta}$. The stronger this of resonant drive, the more the qubit frequency moves. The direct of this shift is set by whether the drive tone is above or below the qubit frequency. Later in this thesis when we discuss how to operate the gate, we will have two sideband tones that are a few GHz higher than the qubit frequency. These will shift the qubit frequency down when turned on, an effect we must calibrate for to have accurate qubit pulses.

2.5 Coupling Qubits to Resonators

2.5.1 cQED

As we've seen earlier, control of the qubit is through classical electronics, and similarly, measurement of the qubit state must use classical electronics too. In order to readout the qubit, we must map the qubit state onto something that we can probe with classically. To understand this process, we look to the (cQED) cavity quantum electrodynamics–and its extension circuit quantum electrodynamics–a field pioneered by Serge Haroche in the 1970s. Here we bring back the harmonic oscillator and LC circuit equivalent from earlier in this chapter. We will show that it is useful for readout.

cQED describes the interaction between a two-level atom and electromagnetic field in a cavity. The cavity is composed of two mirrors facing one another in a Fabry-Perot geometry. The low loss reflective walls allow electromagnetic waves to bounce between them resonantly and form standing waves. If one of the mirrors is slightly transparent, photons in the cavity will remain in the cavity for a long time, but a small percentage will escape to allow for detection of processes in the cavity at a rate κ.

We can make an analogous setup described the the same physics by using an artificial atom, such as a superconducting qubit, and a coplanar waveguide resonator. These 1D transmission lines printed onto dielectric substrates, typically made from superconducting material like niobium or aluminum. The conducting track is printed with return conductors on either side, hence the word 'coplanar' in the name.

To show how we can map the qubit state to this printed resonator, we begin with the Hamiltonian of the combined system:

$$\hat{H} = \frac{1}{2}\omega_q + \omega_r a^\dagger a + g\sigma_x(a^\dagger + a) \tag{2.47}$$

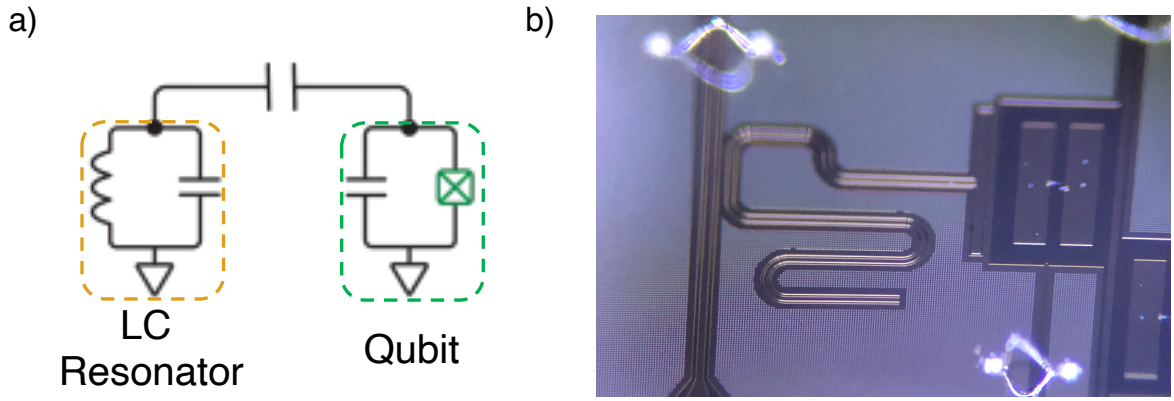

Figure 2.5: a) Qubit and resonator as lumped elements. b) photo of fabricated qubit and resonator on a processor.

e expand the σ_x term into $\sigma_+ + \sigma_-$ and then go into the interaction picture defined by $U = e^{-i(\omega_q \sigma_z + \omega_r a^\dagger a)t}$. In the interaction picture the Hamiltonian is

$$\begin{aligned} H &= g\left[e^{i(\omega_q+\omega_r)t}\sigma_+ a^\dagger + e^{i(\omega_q-\omega_r)t}\sigma_+ a + e^{-i(\omega_q-\omega_r)t}\sigma_- a^\dagger + e^{-i(\omega_q+\omega_r)t}\sigma_- a\right] \\ &= g\left[e^{i(\omega_q-\omega_r)t}\sigma_+ a + e^{-i(\omega_q-\omega_r)t}\sigma_- a^\dagger\right] \end{aligned}$$

. Now going back to the original frame, we have the Jaynes-Cummings Hamiltonian:

$$H = \frac{1}{2}\omega_q + \omega_r a^\dagger a + g(\sigma_+ a + \sigma_- a^\dagger) \tag{2.48}$$

Note that the generalized form including all the transmon levels is

$$H = \sum_j \frac{\omega_j}{2} |j\rangle\langle j| + \omega_r a^\dagger a + g_{j,j+1}\left(|j+1\rangle\langle j| a + |j\rangle\langle j+1| a^\dagger\right) \tag{2.49}$$

For superconducting platforms, and in particular for readout, we typically operate the system in the dispersive regime, $g_{j,j+1} << \omega_q - \omega_r$. We transform the generalized Hamiltonian using $U = e^{\sum_j \lambda_i\left(|j+1\rangle\langle j|a + |j\rangle\langle j+1|a^\dagger\right)}$ where $\lambda_i = \frac{g_{i,i+1}}{\omega_q - \omega_r}$ are small parameters. We make use of the identity

$$e^{\lambda T} H e^{-\lambda T} = H - \lambda[T, H] + \frac{\lambda^2}{2}[T,[T,H]] + ... \tag{2.50}$$

and expand the Hamiltonian in terms of small parameter λ_i

$$H = \sum_j \frac{\omega_j}{2} |j\rangle\langle j| + \omega_r a^\dagger a + \chi_{j,j+1} |j+1\rangle\langle j| + \chi_{01} a^\dagger a |0\rangle\langle 0| + \sum_{j=1}(\chi_{j-1,j} - \chi_{j,j+1}) a^\dagger a |j\rangle\langle j|$$

We define $\chi_{ij} = \frac{g_{ij}^2}{\omega_q - \omega_r}$. Truncating to the two level system gives

$$H = (\omega_q - \chi_{01})\sigma_z + (\omega_r - \chi_{12}/2 + \chi\sigma_z)a^\dagger a \tag{2.51}$$

and $\chi = \chi_{01} - \frac{\chi_{12}}{2}$. Note that the resonator frequency has gained a qubit state dependent term $\chi\sigma_z$, implying that the resonator frequency will shift depending on whether the qubit is in the ground or excited state. This is particularly useful because by monitoring the frequency of the resonator coupled to the qubit, we can indirectly infer the state of the qubit.

2.6 Introduction to Noise and Decoherence

All quantum objects are subject to decoherence. While we could put our efforts into completely isolating a qubit in vacuum to have longer coherence times, that would also be a rather useless qubit. Access the qubit through drive lines, coupling or readout resonators means the qubit cannot be isolated in a vacuum. As such, the qubit and the surrounding quantum objects like coplanar waveguide resonators are subject to decay through a variety of uncontrolled degrees of freedom in its environment and within itself. These can be expressed as noise fluctuations in the qubit Hamiltonian itself. Each noise source is described by its quantum spectral density

$$S_\lambda(\omega) = \frac{1}{2\pi}\int \langle\lambda(0)\lambda(t)\rangle e^{i\omega t}dt \tag{2.52}$$

While classical noise spectrums describe the magnitude of the noise at a frequency ω, quantum noise spectrums describes a system's ability to absorb and emit energy at $\hbar\omega$ according to the decay rates given by Fermi's Golden Rule. Positive (negative) values of ω are associated with absorption (emission).

Given we typically operate our processors in the $k_bT << \hbar\omega_q$ limit, there are two main types of decoherence we will be focusing on: relaxation processes that are mainly dependent on $S_\lambda(\omega = \omega_q)$ and dephasing processes that depend on $S_\lambda(|\omega| << \omega_q)$. Relaxation is described by the decay time T_1 which is the timescale for a qubit to decay from the excited state to the ground state, described by the diagonal components of the qubit density matrix. Deocoherence (T_2), on the other hand, is obtained from the free induction decay of a Ramsey measurement. This is associated with the off-diagonal elements of the density matrix. For Gaussian noise sources that are weakly coupled and have short correlation times, a qubit's decay is described by the Bloch-Redfield equations:

$$\begin{pmatrix} 1 + (|\alpha|^2 - 1)e^{-\Gamma_1 t} & \alpha\beta^* e^{i\Delta\omega t}e^{-\Gamma_2 t} \\ \alpha^*\beta e^{-i\Delta\omega t}e^{-\Gamma_2 t} & |\beta|^2 e^{-\Gamma_1 t} \end{pmatrix} \tag{2.53}$$

where Γ_1 is the longitudinal relaxation rate and describes the depolarization along the qubit quantization axis:

$$\Gamma_1 = \frac{1}{T_1} = \Gamma_{1\downarrow} + \Gamma_{1\uparrow} \tag{2.54}$$

Longitudinal noise fluctuates the qubit vector field and changes the Bloch vector precession rate on the equator. This contributes to Γ_ϕ that feeds into

$$\Gamma_2 = \frac{1}{T_2} = \frac{\Gamma}{2} + \Gamma_\phi \tag{2.55}$$

While this decay is exponential in the presence of white noise, it can be non-exponential when there is other noise like 1/f noise or any other noise that is singular at $\omega = 0$. In this situation, instead of describing the decay with $e^{-\Gamma_2 t}$, the decay function becomes

$$f(t)e^{\Gamma_1 t} \tag{2.56}$$

where $f(t)$ is some other decay function.

Some common sources of noise include two level fluctuators (TLF) in the qubit junction tunnel barriers that cause changes in E_J and also relaxation events.Another potential source of noise is fluctuation in the superconducting phase variable φ. While this is more of an issue for tunable or flux qubits rather than fixed frequency qubits (as in this thesis), we make not of it here because it is commonly seen on other processors. This fluctuation can be caused by deviations in the magnetic field surrounding the qubit or from magnetic vortices in the superconducting electrodes on the device.

All of the sources listed above contribute low frequency noise.

2.6.1 T_1

Using Fermi's Golden Rule, T_1 has contributions from both relaxation events and excitation events:

$$\begin{aligned}\Gamma_1 &= \Gamma_{rel} + \Gamma_{ex} \\ &= |\langle e|H_{int}|g\rangle|^2\rho(\hbar\omega_q) + |\langle g|H_{int}|e\rangle|^2\rho(-\hbar\omega_q)\end{aligned}$$

This can then be rearranged and written in terms of the noise spectrum

$$\Gamma_1 = \left|\left\langle e\left|\frac{\partial H_{int}}{\partial\lambda}\right|g\right\rangle\right|^2 \rho(\hbar\omega_q) + \left|\left\langle g\left|\frac{\partial H_{int}}{\partial\lambda}\right|e\right\rangle\right|^2 \rho(-\hbar\omega_q)$$

In the $k_bT << \hbar\omega_q$ limit, the excitation process is exponentially suppressed so $\Gamma_1 = \Gamma_{rel}$

2.6.2 T_2 and Dynamical Decoupling Methods

For noise that deviates from Bloch-Redfield theory we provide an example for Gaussian noise. The accumulated noise in the qubit state relative between the ground and excited states is described by

$$\varphi = \int_0^t \omega_q dt' = \langle\omega_q\rangle t + \delta\varphi(t) \tag{2.57}$$

This fluctuates according to changes in the qubit frequency due to noise sources

$$d\varphi = \frac{\partial \omega_q}{\partial \lambda} \int_0^t dt' \delta\lambda(t') \tag{2.58}$$

For noise generated by a large number of fluctuators, that are weakly coupled to the qubit, the overall statistics are Gaussian. When we average over all realization of the Gaussian distributed noise, we get the decay function

$$e^{-\chi_N} \tag{2.59}$$

$$\chi_N(t) = \frac{t^2}{2} \sum_\lambda (\frac{\partial \omega_q}{\partial \lambda})^2 \int_{-\infty}^{\infty} d\omega S(\omega) g_N(\omega, t) \tag{2.60}$$

where $g_N(t)$ is a filter function. The simplest measurement to capture the decay function is a Ramsey measurement. This is where a $\pi/2$ pulse prepares the qubit in a superposition state of $|g\rangle$ and $|e\rangle$ and then it is allowed to evolve for time t before a second $\pi/2$ pulse is applied to map the remaining bloch vector into the z basis for measurement. An alternative version of this pulse sequence that is also an example of dynamical decoupling is called the CPMG sequence which also inserts a series of N π pulses in between the $\pi/2$ pulses. When N=0, $g_{N=0} = \text{sinc}^2(\omega t/2)$, but in general,

$$g_N(t) = \left(\frac{1}{\omega t}\right)^2 \left| 1 + (-1)^{N+1} e^{i\omega t} + 2 \sum_n^N (-1)^n e^{i\omega \delta_n t} \cos(\omega \tau_\pi / 2) \right|^2 \tag{2.61}$$

We use $\delta_n \in [0, 1]$ to be the normalized center of the jth π pulse relative to position of the $\pi/2$ pulses and τ_π is the length of each π pulse. As you increase the number of π pulses, the filter function peak moves to higher frequencies, which acts to reduce the contribution from low frequency, thus increasing coherence times. In the limit of continuous π pulses, we have a Rabi drive, which is the basis of the spinlocking pulse sequence discussed in the next section. However, this is only beneficial if the noise spectrum is dominated by low frequency terms. The filter function always integrates to have the same area, regardless of N, so if the noise spectrum is white or if it is dominated by higher frequency terms, then performing CPMG will provide no change or make things worse, respectively.

2.7 Spinlocking and Dressed Frame Lifetimes

When a drive resonant with the qubit transition is applied to the qubit, the Hamiltonian can be re-diagonalized such that the basis states $|\pm\rangle$ are superpositions of the original $|g/e\rangle$ basis states, as described in the section earlier on driving the qubit. In this Rabi dressed frame, we can define new lifetimes $T_{1\rho}$ and $T_{2\rho}$ that are analogous to T_1 and T_2 of the bare qubit. These describe the longitudinal and transverse decay of the dressed qubit [78].

The $T_{1\rho}$ lifetime can be measured using the spinlocking sequence, derived from traditional NMR measurements and is an extension of the CPMG pulse sequence referenced earlier. The pulse sequence consists of a $\frac{\pi}{2}_x$ pulse followed by a Rabi drive around the y-axis followed by a final $\frac{\pi}{2}_x$ pulse. This is the limit of using an infinite number of π pulses in the CPMG scheme. The $T_{2\rho}$ lifetime can be measured using by observing the decay of the envelope of Rabi drive response. Both measurements should produce envelopes that follow exponential decay function if the noise is purely Gaussian. However, qubits are often subject to 1/f noise due to TLS's coupled to the qubit in at surfaces and interfaces. These arise from defects in the material and result in bistable fluctuators that tunnel between interfaces and cause telegraphing noise. The key difference between these two schemes is the type of noise they are senstive to. $T_{2\rho}$ is associated with transverse decay in the dressed frame and is sensitive to low frequency fluctuations in the qubit drive which may cause non-exponential decay envelopes that are more complicated to fit. It is also difficult to sample enough points to fit the sinusoid when looking at the qubit response over long periods of time. In contrast, $T_{1\rho}$ dynamically decouples from the 1/f noise as it has the drive axis collinear with the Bloch vector sitting in either the $|+\rangle$ or $|-\rangle$ states such that the qubit decays longitudinally from the spinlocked state such that the decay function is purely exponential and easy to fit. One may add an artificial detuning–similar to a Ramsey experiment–to assist with fitting the data. Furthermore, the decay is dominated by noise at the Rabi drive frequency, instead of being sensitive to broadband noise. For this reason, the $T_{1\rho}$ time is typically longer than T_1 and T_2 times.

We will revisit these measurements as the STAR gate we introduce is done in the Rabi Dressed frame. The implication is that the main limiting qubit lifetimes are the dressed frame lifetimes $T_{1/2\rho}$ which are longer due to dynamical decoupling. This will be an advantage of the gate. Furthermore, these spinlocking measurements will also be developed into calibration procedures for parameters important to the gate.

Chapter 3

Multiqubit STAR Gate

In this chapter we motivate the advantages of a more highly connected qubit system, particularly in the era of NISQ devices, and describe a multiqubit entangling gate for transmons on an all-to-all connected platform that derives inspiration from the ion Mølmer-Sørensen gate. We will begin by detailing the Mølmer-Sørensen gate. We then discuss a direct translation of this gate to the superconducing platform. We discuss why the direct translation does not produce a high fidelity gate due to transmon crosstalk, frequency crowding issues, and fabrication limitations. Finally, we introduce a technique that allows us to emulate many advantages of ion qubits and successful adapt the Mølmer-Sørensen gate.

Most of the gate development on superconducting qubits has been in achieving high fidelity two-qubit gates [51, 35, 74, 57]. This is because superconducting qubit platforms often lack the higher connectivity between qubits present on other platforms, as it is hard to place $O(n^2)$ qubits in an area that fits $O(n)$ qubits [47]. Thus many multi-qubit entanglement schemes have nearest neighbor couplings and use cascaded pairwise interactions such as the demonstration of 20 qubit entanglement on the IBM systems [54]. However, in the NISQ era where qubit lifetimes are shorter and gates are not at the fault-tolerant threshold, it is advantageous to shorten the gate depth. There are some works that show multi-qubit entanglement with three and more qubits on more highly connected systems. For instance, the Pan group demonstrated entanglement between 10 tunable qubits coupled to a shared resonator [76]. The qubits are tuned to be close to resonance with each other and the shared resonator frequency, allowing the qubits to do a SWAP interaction through the shared resonator. However, this demonstration uses tunable qubits which require a tuning line in addition to a drive line, increasing the footprint per qubit. This makes scaling up even more challenging as the community is already dealing with crosstalk issues between neighboring lines. Furthermore, tunable qubits expose the qubit to $1/f$ flux noise and typically have shorter lifetimes than fixed frequency qubits. In general, higher connectivity processors do invite in its own challenges, such as higher crosstalk, correlated errors, and increased complexity in spectator qubit dynamics on superconducting qubit platforms. However, the benefits still make it worth exploring. We attempt to address some of the challenges exposed by previous demonstrations in our implementation of the STAR gate.

3.0.1 The Mølmer-Sørensen Gate

We now present the ion gate that inspires the multiqubit STAR gate. Single-operation multiqubit entanglement is most commonly found in ion qubit experiments where, using the Mølmer-Sørensen gate and similar interactions, they have demonstrated GHZ state preparation of up to 14 ions [53]. Ion qubit are suspended in free space using electromagnetic fields and the traps provide natural all-to-all connectivity due to the shared phonon mode. They are able to selectively entangle any subset of qubits, providing greater flexibility when decomposing an algorithms into a sequence of gates [46]. Their most popular multiqubit entangling gate is the Mølmer-Sørensen gate [73, 72, 52].

There are a few advantages of the gate.

1. the gate was designed to be insensitive to the vibrational mode the ions are coupled to. The key to this is as long as the starting and ending vibrational mode are the same, the gate can be operated from any phonon mode, eliminating the need to always cool ions to the lowest temperature possible.

2. The gate is composed of multiple transition paths that interfere to cancel any dependence of the gate's rates and frequencies on vibrational quantum numbers.

3. The gate is that is scalable, in theory, to entangle as many qubits as can fit in the ion trap.

4. There are two operating modes. In the near detuned regime, the vibrational modes are significantly populated. This produces a faster gate. Conversely, in the far-detuned regime, the gate is slower and two photon transitions are virtual. This does make the gate less sensitive to any losses in the coupling mode, however.

Demonstrations of the gate have shown over 99.9% fidelity on 2 qubit gates and at maximum have entangled 14 qubits in one operation. Scaling of the ion gate is currently limited by correlated errors among the ions such as fluctuations in the homogeneous magnetic field. Entangled states of N qubits under the presence of such noise decays N^2 time faster than for a single qubit. Furthermore it becomes harder and harder to hold more ions in a single trap as adding more qubits makes the spectrum of the phonon modes more complex and inter-ion spacing decreases. Finally the coupling strength of the ions to the phonon mode, described by the Lambe-Dicke parameters scales as $\sqrt{N}$ which means that as more ions are added, the coupling decreases relatively, slowling down all qubit gates, making the qubits more prone to decoherence, B-field noise, and laser frequency fluctuations. This provides a limit on how many qubits can be entangled.

The gate is driven by red and blue sidebands driven at frequencies $\nu_r = \omega_q - \omega_{ph}$ and $\nu_b = \omega_q + \omega_{ph}$ where ω_q is the frequency of the ion qubit and ν_{ph} is the frequency of the shared phonon mode, as shown in Fig. 3.1. These sidebands drive 2 photon transitions that carry population from the $|gg\rangle$ state to the $|ee\rangle$ state. If one applies the sidebands for half the time, an entangled state $\Psi = \frac{|gg\rangle + |ee\rangle}{\sqrt{2}}$ is formed. Key to the entanglement is to

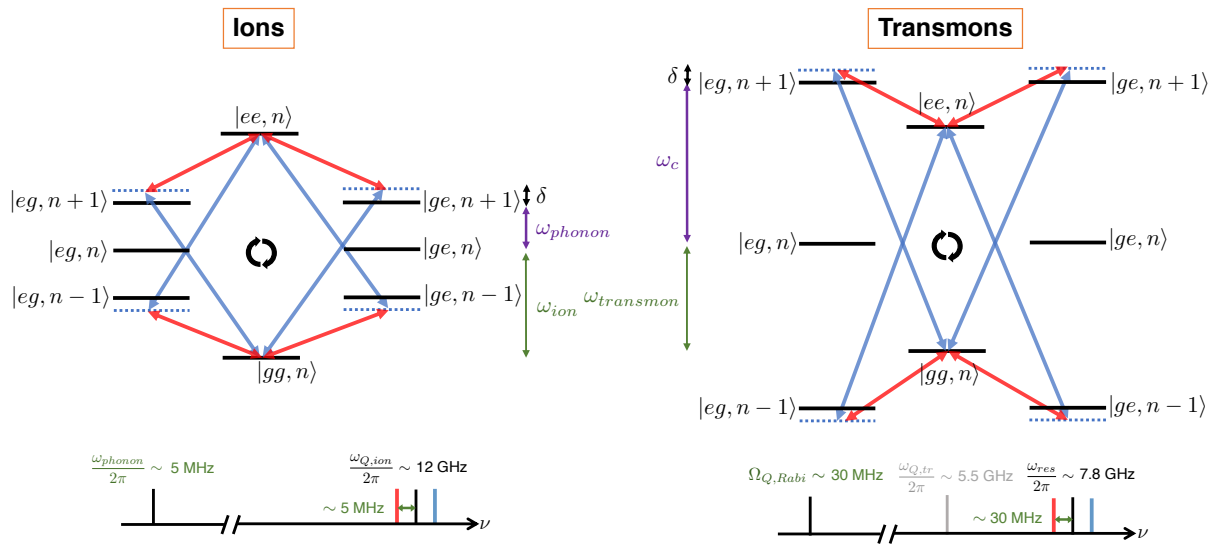

Figure 3.1: The original ion Mølmer-Sørensen gate on the left versus the direct transmon adaptation. Energy scaling differences are shown on the bottom for each platform. In both systems, g/e label the qubit states. For the ion platform, n labels the vibrational mode, whereas for the transmon system, n is the number of photons in the shared resonator. The gate is driven by red and blue sidebands that are at the qubit plus and minus the coupling mode energy.

select a proper detuning δ of the sidebands from the intermediate states of the two photon transitions. These are the energy levels that are $n \pm 1$ in phonon mode. This ensures that the intermediate states are not populated at the time of the gate.

Looking at the phonon mode in the phase space picture provides additional insight the importance of δ. At the beginning of the gate the the phonon mode starts at the origin of the phase space diagram. During the course of the gate, the phonon makes a circle in phase space as it entangles with the ions and, at the gate time, returns back to the origin, having completely detangled from the ions, leaving all the ions entangled with each other.

3.1 A Direct Transmon Adaptation

While it is possible to draw an analogous level diagram with identical sidebands for superconducting qubits, there are several difference between ions and superconducting qubits that make this direction adaptation infeasible. First, while ions couple using the vibrational modes, superconducting qubits typical couple using the electromagnetic mode of a

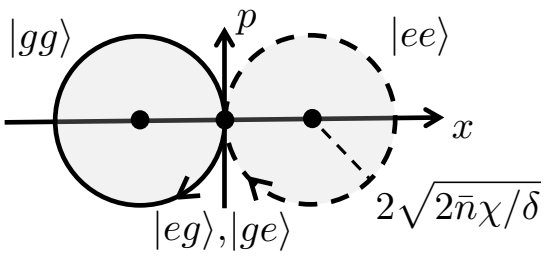

Figure 3.2: The phonon/resonator mode makes a circle in phase space.

coplanar waveguide (CPW) resonator, causing an energy scaling difference between the two platforms. Ions have qubits in the GHz regime whereas the coupling mode frequency is in the MHz range. In superconducting qubits, the coupling does not come from shared phonon modes. Rather it is generated by the electromagnetic mode of resonators, which have frequencies in the GHz regime, just like the qubits themselves. Mapping the ion gate directly to the transmon platform would result in sidebands that were over 10 GHz apart. Given the most resonators on superconducting platforms have bandwidths that are at most MHz wide, the sidebands would be largely be filtered out by the linewidth of whatever resonator they are being sent into. Second, ion qubits are largely identical in qubit frequency, up to slight deviations due to the trap environment. In contrast transmon frequencies are at the mercy of fabrication inaccuracies. In addition, due to crosstalk in qubit control lines, qubits on the same chip cannot have transition frequencies in case driving one qubit might excite the $g \longrightarrow e$ or $e \longrightarrow f$ transitions in another qubit. One possible method of dealing with this is to use tunable qubits, but tunable qubits require an additional tuning line beyond the usual drive line and thus have a larger per qubit footprint. Tunability also makes the qubit vulnerable to flux noise and as a result, tunable qubits have shorter lifetimes than fixed frequency qubits. For these reasons superconducting qubits are intentionally printed at different frequencies.

This spread in qubit frequencies shifts the energy levels in the diagram shown in Fig. 3.1 such that the detunings of the red and blue sidebands from the intermediate states are not all equal. As mentioned in the previous section, the choice of δ is important in ensuring that the side states are not populated and that qubits are properly disentangled from the shared mode at the time of the gate. On the superconducting qubit platform, for a fixed set of frequencies for the red and blue sidebands, δ to the side states wold all be unequal, causing misalignment in the rates of all the two photon transition paths. This would be damaging to the entanglement fidelity.

3.2 Multiqubit STAR Gate: A Rabi Driven Transmon Adaptation

Here we present the multiqubit STAR (Sideband Tone Assisted Rabi driven) gate. The concept of this work was developed by Sydney Schreppler, a postdoc at QNL. The derivations were fleshed out by our collaborators Joachim Cohen and Alexandru Petrescu from Professor Alexandre Blais's lab at Université de Sherbrooke.

To solve the technical challenges of unequal qubit frequencies Mølmer-Sørensen gate to a superconducting platform presented above, we use the dressed states of Rabi driven qubits as a new effective qubit. We apply a drive resonant with the ground to excited state transition to each of the participating qubits. As shown in previous chapters, re-diagonalizing the combined Hamiltonian of the qubit and drive produces new basis state $|\pm\rangle$. The energy level separation between these states is set by the power of the drive, an experimentally tunable parameters. This solves both challenges of the direct transmon translation of the MS gate. First, we create a new effective qubit out of each native qubit with energy levels that we can experimentally tune to be identical. Second, typically Rabi drives are kept within the MHz regime so we have also successfully lowered the qubit energy scale down to be much smaller than the energy scale of the coupling mode (GHz). Now, like the ion situation, we can have $\omega_b - \omega_r << (\omega_b + \omega_r)/2$, making them much easier to send into a shared resonator.

The red and blue sidebands are similar to that of the traditional Mølmer-Sørensen gate. The sidebands are at $\omega_r = \omega_c - \Omega + \delta$ and $\omega_b = \omega_c + \Omega + \delta$ where ω_c is the frequency of the coupling element, the cavity mode, and Ω is the strength of the Rabi drive, an experimentally tunable parameter. We point out that just like the ion gate case, the sidebands have a slight detuning δ from being resonant with the intermediate states of the two photon transitions such that these states are not fully populated. By carefully selecting the correct value of δ which we will calibrate for in Chapter 4, one should ideally have no population in the intermediate states at the gate time. As indicated by the thick black arrows in Fig. 4.6a, these sidebands drive two-photon transitions that result in population swaps between $|{+}{+}, n\rangle$ and $|{-}{-}, n\rangle$ if the qubits are prepared in either $|{+}{+}, n\rangle$ or $|{-}{-}, n\rangle$. We note that one may also initialize the qubits in $|{-}{+}, n\rangle$ and generate population swaps with $|{+}{-}, n\rangle$.

In the frame rotating at the qubit frequencies and at $\omega_c + \delta$ for an n-qubit gate, the dispersive Hamiltonian describing the system is given by [5]

$$\mathbf{H}_I = -\delta \mathbf{a}^\dagger \mathbf{a} + \sum_{k=1}^{n} (\Omega_R/2)\boldsymbol{\sigma}_{x_k} - \chi_k \boldsymbol{\sigma}_{z_k} \mathbf{a}^\dagger \mathbf{a} + \mathbf{H}_{\text{sb}}(t). \tag{3.1}$$

Here, $\mathbf{a}$ is the annihilation operator of the resonator, and $\boldsymbol{\sigma}_{l_k}$, $l = x, y, z$, are the Pauli matrices of qubit k. The effect of the sideband term $\mathbf{H}_{\text{sb}}(t)$ acting on the resonator mode is to displace the field of the resonator such that $\mathbf{a} = \mathbf{d} + \alpha(t)$, with $\alpha(t) = \sqrt{2\bar{n}}\cos(\Omega_R t + \varphi_\Delta)$, where $\bar{n}$ is the mean resonator photon number due to the two sideband tones, and φ_Δ is the phase difference of the two tones. After performing the displacement transformation and going into the qubit frame at the Rabi frequency [please see Supplement for details], the

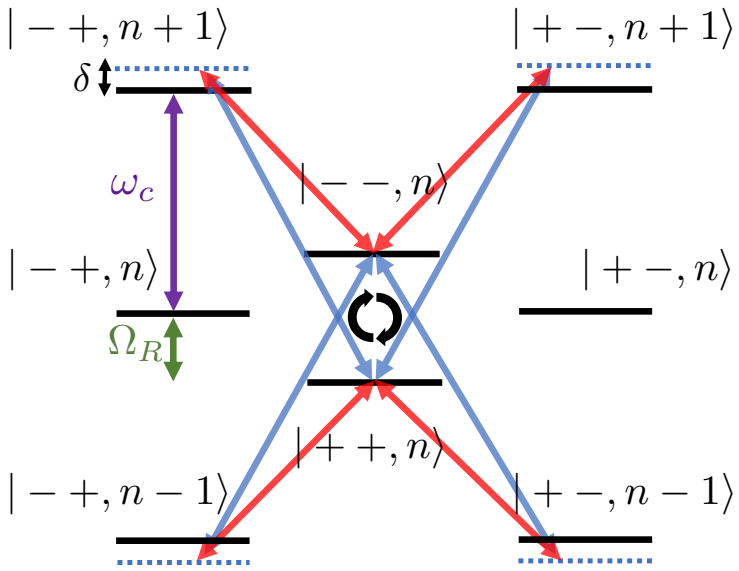

Figure 3.3: Driven transmon adaptation with Rabi dressed qubit levels. Instead of labeling the levels with g/e, they are now labeled with $\pm$, representing superpositions of the original basis states. The new effective qubit energy is Ω_R, which is qubit drive strenght dependent. The energy hierarchy has been flipped compared to the ion gate. Here the coupling energy is higher than the qubit energy. Thus the blue (red) sidebands are at the resonator frequency plus (minus) Ω_R.

Hamiltonian reduces to

$$\mathbf{H}_R = -\delta \mathbf{d}^\dagger \mathbf{d} - 2\sqrt{2\bar{n}}\chi \mathbf{J}_{zy}^{\varphi_\Delta}(\mathbf{d}+\mathbf{d}^\dagger) + \mathbf{H}_{\text{err}}, \tag{3.2}$$

where $\mathbf{J}_{\varphi_\Delta} = \cos(\varphi_\Delta)\mathbf{J}_z - \sin(\varphi_\Delta)\mathbf{J}_y$, $\mathbf{J}_l = \sum_k \boldsymbol{\sigma}_{l_k}/2$, $l = x, y, z$ are the generalized spin operators, and $\mathbf{H}_{\text{err}} = \mathbf{A}_1 e^{i\Omega_R t} + \mathbf{A}_1 e^{2i\Omega_R t} + \mathbf{h.c}$ are spurious oscillating terms [17]. The effect of these terms can be neglected in the limit of large Rabi frequency, $\delta \sim \chi\sqrt{\bar{n}} < \chi\bar{n} \ll \Omega_R$. In this parameter regime, the Hamiltonian of Eq. (3.2) can be mapped to the Hamiltonian originally proposed by Mølmer and Sørensen in the context of trapped ions.

Here, we give a brief review of the working principle of the STAR gate[3]. During the gate, the qubits entangle with the resonator field, resulting in a non-trivial operation on the qubits $U = e^{i\frac{\pi}{2}\mathbf{J}^2_{\varphi_\Delta}}$. The origin of this non-trivial phase comes from the qubit-state-dependent paths that the resonator describes in phase-space. To see this, one can cast the Hamiltonian of Eq. (3.2) in the form $\mathbf{H}_R = -\delta(\mathbf{d}^\dagger - \boldsymbol{\alpha}^*)(\mathbf{d}-\boldsymbol{\alpha}) - \boldsymbol{\alpha}^2$, with $\boldsymbol{\alpha} = 2\chi\sqrt{2\bar{n}}\mathbf{J}_{\varphi_\Delta}/\delta$. The first term describes a harmonic oscillator of frequency δ centered on the qubit-state dependent $\boldsymbol{\alpha}$, while the second term $\boldsymbol{\alpha}^2 = 8\bar{n}\chi^2\mathbf{J}^2_{\varphi_\Delta}/\delta^2$ describes a qubit-qubit interaction

[3]Kpop fun fact 3/7: The dance for 'On' by BTS was choreographed by a 19 year old choreographer Sienna LaLau.

term. Note that the two terms commute. Initializing the resonator field $\mathbf{d}$ in the vacuum, the field state remains in a coherent state and revolves around the qubits-dependent vacuum positions $\boldsymbol{\alpha}$, as depicted in Fig. 4.6b. After one period of evolution $T = 2\pi/\delta$, the field state comes back to its initial position (the vacuum), and the qubits and the resonator disentangle. The qubits undergo a non-trivial evolution U generated by the last term $-\boldsymbol{\alpha}^2$, with $U = \exp\left[i\pi(8\bar{n}\chi^2/\delta^2)\mathbf{J}^2_{\varphi_\Delta}\right]$. When $\delta = 2\sqrt{\bar{n}}\chi$, this implements the entangling gate $U = e^{i\frac{\pi}{2}\mathbf{J}^2_{\varphi_\Delta}}$. With two qubits, U takes the simple form $U = e^{i\frac{\pi}{2}\boldsymbol{\sigma}^{\varphi_\Delta}_{zy,1}\boldsymbol{\sigma}^{\varphi_\Delta}_{zy,2}}$ up to a global phase factor, where $\boldsymbol{\sigma}^{\varphi_\Delta}_{zy,k} = \cos(\varphi_\Delta)\boldsymbol{\sigma}_{z_k} - \sin(\varphi_\Delta)\boldsymbol{\sigma}_{y_k}$. During the gate, the entanglement of the qubits with the resonator makes the gate fidelity sensitive to the resonator photon loss channel. In addition to $\chi\sqrt{\bar{n}} \ll \Omega_R$, we therefore require that the gate rate is much larger than κ, i.e $\kappa \ll \chi\sqrt{\bar{n}}$.

3.3 A more detailed derivation of gate Hamiltonian

We now derive the version of the Hamiltonian we use in our simulations. We include resonator losses and qubit decay times, but we assume that the transmon is a two-level system, neglecting higher states. This makes the simulation less accurate for Rabi drives that approach the anharmonicity, where f states are very likely to be populated. Furthermore, we assume that the resonator only has one mode that the qubits are coupled to, whereas in reality, the resonator has many modes. These modes could provide spurious interactions that affect qubit couplings. However, for our experiment we assume that these other modes are several GHz away and will not, at least, be populated by the sidebands. Let us consider two transmon qubits coupled to a resonator through a Jaynes-Cumming Hamiltonian, two resonant Rabi drives applied to the transmons, and two sideband tones applied to the resonator. The Hamiltonian of the driven system can be written in the following form [60],

$$\begin{aligned}\mathbf{H}(t)/\hbar = \tilde{\omega}_c\mathbf{a}^\dagger\mathbf{a} + \sum_{k=1}^{2}\tilde{\omega}_k\mathbf{b}_k^\dagger\mathbf{b}_k \\ - \sum_{k=1}^{2}E_J^k\left(\cos\left(\frac{\boldsymbol{\Phi}_k}{\phi_0}\right) + \frac{1}{2}\frac{\boldsymbol{\Phi}_k^2}{\phi_0^2}\right) \\ + \sum_{k=1}^{2}\Omega_{R_k}(t)i(\mathbf{b}_k^\dagger - \mathbf{b}_k) \\ + i(\epsilon_r(t) + \epsilon_b(t))(\mathbf{a}^\dagger - \mathbf{a}),\end{aligned} \tag{3.3}$$

where

$$\boldsymbol{\Phi}_k = \phi_k^a(\mathbf{a} + \mathbf{a}^\dagger) + \sum_{k'=1}^{2}\phi_{k,k'}^b(\mathbf{b}_{k'} + \mathbf{b}_{k'}^\dagger).$$

Here we note $\mathbf{a}$ (resp. $\mathbf{a}^\dagger$) and $\mathbf{b}_k$ (resp. $\mathbf{b}_k^\dagger$) the annihilation (resp. creation) operator of the resonator and qubit k, $\tilde{\omega}_c$ and $\tilde{\omega}_k$ the dressed frequencies of the resonator and qubit

k respectively, E_J^k the Josephson energie of qubit k, $\phi_0 = \hbar/2e$ the superconducting flux quantum. The drives applied to the resonator (resp. the transmons) also weakly drive the transmons (resp. the resonators) through the hybridization. However, if the frequencies of the system are well separated, these terms can be neglected.

Noting that $\phi_k^a \ll \phi_{j,j}^b$, the dressed mode $\mathbf{a}$ shares a much smaller part of the non-linearity than the dressed modes $\mathbf{b}_k$. This is why we refer to the b modes as the qubit modes and the a modes as the cavity modes. In the transmon regime, zero-point phase fluctuations of the modes are small, and we can limit the expansion of the cosine to fourth order. Assuming that the frequencies of the system are well separated, and $|\tilde{\omega}_k - \tilde{\omega}_l| \gg \Omega$, where Ω is the common Rabi frequency, the Hamiltonian simplifies to

$$\begin{aligned}\mathbf{H}(t)/\hbar = &\omega_c \mathbf{a}^\dagger \mathbf{a} + \sum_{k=1}^{2} \omega_k \mathbf{b}_k^\dagger \mathbf{b}_k - K_k {\mathbf{b}_k^\dagger}^2 {\mathbf{b}_k}^2 \\ &- \mathbf{a}^\dagger \mathbf{a} \Big(\sum_{k=1}^{2} \chi_k \mathbf{b}_k^\dagger \mathbf{b}_k \Big) - \sum_{j \neq k}^{2} \chi_{j,k} \mathbf{b}_j^\dagger \mathbf{b}_j \mathbf{b}_k^\dagger \mathbf{b}_k \\ &+ \sum_{k=1}^{2} \Omega_{R_k}(t) i (\mathbf{b}_k^\dagger - \mathbf{b}_k) \\ &+ i(\epsilon_r(t) + \epsilon_b(t))(\mathbf{a}^\dagger - \mathbf{a}), \end{aligned} \tag{3.4}$$

where ω_a and ω_k are the renormalized frequencies, χ_k, $\chi_{j,k}$ and K_k are respectively the cross Kerr coefficient between the resonator and qubit k, the cross Kerr coefficient between qubit j and qubit k, the self Kerr coefficient of qubit k. The Rabi drives and the sideband tones take the forms $\Omega_{R_k}(t) = \Omega_k \cos(\omega_k t)$ and $\epsilon_{r/b}(t) = \epsilon_r \cos\big((\omega_c + \Omega_{r/b})t + \varphi_{r/b}\big)$, where $\Omega_{r/b}$ is the detuning of the tone from the resonator frequency.

One can write $\Omega_r = -\Omega_{SB} + \delta$, and $\Omega_b = \Omega_{SB} + \delta$, where δ is a small detuning compared to Ω_{SB}. We eliminate the drive term in the Hamiltonian by applying the displacement on the cavity

$$\begin{aligned}\mathbf{a} =& \mathbf{d} + \alpha(t), \\ \alpha(t) =& \frac{\epsilon_r e^{-i\varphi_r}}{2i} \left(\frac{e^{-i(\omega_a - \Omega_{SB} + \delta)t}}{-\Omega_{SB} + \delta} + \frac{e^{i(\omega_a - \Omega_{SB} + \delta)t}}{2\omega_a - \Omega_{SB} + \delta} \right) \\ &+ \frac{\epsilon_b e^{-i\varphi_b}}{2i} \left(\frac{e^{-i(\omega_a + \Omega_{SB} + \delta)t}}{\Omega_{SB} + \delta} + \frac{e^{i(\omega_a + \Omega_{SB} + \delta)t}}{2\omega_a + \Omega_{SB} + \delta} \right) \end{aligned}$$

We choose ϵ_b and ϵ_r such that $|\epsilon_b|/(\Omega_{SB} + \delta) = |\epsilon_r|/(-\Omega_{SB} + \delta) = \sqrt{\bar{n}}$. As $|\epsilon_{r/b}| \ll 2\omega_a \pm \Omega_{SB} + \delta$, one can discard the terms having $2\omega_a \pm \Omega_{SB} + \delta$ in the denominator. The amplitude of the classical field becomes simply $\alpha(t) = \sqrt{2\bar{n}} \cos(\Omega_{SB} t + \phi_\Delta) e^{-i((\omega_c + \delta)t + \phi_\Sigma)}$, where $\phi_\Sigma = \frac{\phi_r + \phi_b + \pi}{2}$ and $\phi_\Delta = \frac{\phi_b - \phi_r}{2}$. We drop the phase factor $e^{-i\phi_\Sigma}$, as it can be eliminated by the transformation $\mathbf{a} \to \mathbf{a} e^{i\phi_\Sigma}$.

Moving to the frame rotating at the frequencies $\omega_a+\delta$ and ω_k, and discarding the counter rotating terms, the Hamiltonian becomes

$$\begin{aligned}\mathbf{H}(t)/\hbar = &-\delta\mathbf{d}^\dagger\mathbf{d} - \sum_{k=1}^{2} K_k {\mathbf{b}_k^\dagger}^2 {\mathbf{b}_k}^2 \\ &- (\mathbf{d}^\dagger\mathbf{d} + \sqrt{2\bar{n}}\cos(\Omega_{SB}t+\phi_\Delta)(\mathbf{d}^\dagger+\mathbf{d}) \\ &+ 2\bar{n}\cos^2(\Omega_{SB}t+\phi_\Delta))\left(\sum_{k=1}^{2}\chi_k\mathbf{b}_k^\dagger\mathbf{b}_k\right) \\ &- \sum_{j\neq k}^{2}\chi_{j,k}\mathbf{b}_j^\dagger\mathbf{b}_j\mathbf{b}_k^\dagger\mathbf{b}_k \\ &+ \sum_{k=1}^{2}\Omega_k\cos(\omega_k t)i(\mathbf{b}_k^\dagger e^{i\omega_k t} - \mathbf{b}_k e^{-i\omega_k t}). \end{aligned} \tag{3.5}$$

From this more complete form of the Hamiltonian, we note two possible limitations to the performance of the gate. First, the cross-Kerr or ZZ interaction, which becomes a resonant XX interaction when the Rabi drives are on. Secondly, another limitation is due to the finite ratio of the Rabi frequency over the anharmonicity of the transmons, Ω_k/K_k.

Assuming that the anharmonicity of the transmons remains larger than the Rabi frequency, i.e $K_k/\Omega_k \gg 1$, we can make a two-level approximations and project the above Hamiltonian on the ground and excited states of the transmons. In addition, we neglect the cross-Kerr between the qubits mediated by the resonator, and perform a rotating-wave approximation on the Rabi drive terms, leading to

$$\begin{aligned}\mathbf{H}(t)/\hbar = &-\delta\mathbf{d}^\dagger\mathbf{d} + \sum_{k=1}^{2}\frac{\Omega_k}{2}\boldsymbol{\sigma}_{x_k} \\ &- (\mathbf{d}^\dagger\mathbf{d} + \sqrt{2\bar{n}}\cos(\Omega_{SB}t+\phi_\Delta)(\mathbf{d}^\dagger+\mathbf{d}) \\ &+ 2\bar{n}\cos^2(\Omega_{SB}t+\phi_\Delta))\left(\sum_{k=1}^{2}\tilde{\chi}_k\boldsymbol{\sigma}_{z_k}\right), \end{aligned} \tag{3.6}$$

where $\tilde{\chi}_k = \chi_k/2$. In the following and in the main text (under Circuit QED Implementation), we take the definition $\chi_k := \tilde{\chi}_k$. Setting $\Omega_k = \Omega_{SB}$ and $\chi_k = \chi$ for all k, and going into the frame rotating at the Rabi drive frequency, the Hamiltonian becomes that of eq.(3.2),

$$\mathbf{H}_R = -\delta\mathbf{d}^\dagger\mathbf{d} - 2\sqrt{2\bar{n}}\chi\mathbf{J}_{zy}^{\varphi_\Delta}(\mathbf{d}+\mathbf{d}^\dagger) + \mathbf{H}_{\text{err}}(t), \tag{3.7}$$

where $\mathbf{J}_{\varphi_\Delta} = \cos(\varphi_\Delta)\mathbf{J}_z - \sin(\varphi_\Delta)\mathbf{J}_y$, $\mathbf{J}_l = \sum_k \boldsymbol{\sigma}_{l_k}/2$, $l = x, y, z$ are the generalized spin operators, and $\mathbf{H}_{\text{err}} = \mathbf{A}_1 e^{i\Omega_{SB}t} + \mathbf{A}_1 e^{2i\Omega_{SB}t} + \mathbf{h.c}$ represents spurious oscillating terms [17].

$\mathbf{H}_{\rm err}(t)$ can be neglected as long as one satisfies $\|\mathbf{B}_{1,2}\| \ll \Omega_R$. The dominant term in $\mathbf{H}_{\rm err}(t)$ comes from the term in third line of eq. (3.6), and scales as $\chi\bar{n}e^{i\Omega t}$. This leads to a renormalization of the Rabi frequency that is taken into account in the simulations. This term is responsible for the fidelity saturation in Fig.6.6b. When the Rabi frequency is set to $\Omega_R = 2\Omega_{SB}$, this term becomes resonant and leads to large oscillation of $\langle\sigma_X\rangle$ seen at $\Omega_R = 60$ MHz in Fig. 5.3.

3.3.1 Master Equation

All simulations are obtained using the following master equation:

$$\begin{aligned}\frac{d\boldsymbol{\rho}}{dt} =& -\frac{i}{\hbar}[\mathbf{H}(t),\boldsymbol{\rho}] \\ &+\sum_k \mathcal{D}\big[\sqrt{1/2T_{1,\rho,k}}\boldsymbol{\sigma}_{z_k}\big](\boldsymbol{\rho}) \\ &+\sum_k \mathcal{D}\big[\sqrt{1/2T_{2,\rho,k}}\boldsymbol{\sigma}_{z_x}\big](\boldsymbol{\rho}) \\ &+\mathcal{D}\big[\sqrt{\kappa}\mathbf{a}\big](\boldsymbol{\rho}),\end{aligned} \tag{3.8}$$

where $\mathcal{D}\big[\mathbf{M}\big](\boldsymbol{\rho}) = \mathbf{M}\boldsymbol{\rho}\mathbf{M}^\dagger - (\mathbf{M}^\dagger\mathbf{M}\boldsymbol{\rho} + \boldsymbol{\rho}\mathbf{M}^\dagger\mathbf{M})/2$. Here, $T_{1,\rho,k}$ and $T_{2,\rho,k}$ are the spinlocking times of qubit k, and we use the Hamiltonian of Eq. (3.6).

Chapter 4

Experimental Setup

In this section we describe all the techniques used to build a system that allows for maximal accuracy in controlling the qubits and detecting signals that come back from the chip for readout. This includes several low temperature methods such as using dilution refrigerators to keep the processor far under the T_C, amplification choices, plus wiring and shield methods. These help to keep stray thermal photons from disturbing the qubits and also help boost the quality of the signal that is coming back form the chip. In addition, we touch upon design choices for the processor itself such that we are able to increase control of the qubit and also maximize the sensitivity of the signal to the qubit state for readout while minimizing sources of unwanted external coupling.

4.1 Low temperature methods

The superconducting processors must be operated at temperatures lower than the T_C of the materials used. For our processors, we use niobium on the ground plane and aluminum for the junctions. The T_C of niobium is 9.7 K and 1.2 K for aluminum. However, it is not enough just to operate slightly lower than these temperatures. Ideally if we do not send control pulses to a qubit, we would want the qubit to be idling in the ground state. But if the surrounding environment temperature is too warm, there could be stray photons that excite the qubit [33]. Typically we design the qubit to have a ground to excited state transition energy that is much larger than the energy of the thermal photons in its surrounding environment.

$$E_{ge} = hf_{ge} >> k_B T \tag{4.1}$$

where typically $f_{ge} \sim$ 5GHz and $T \sim$ 15mK. Ideally, the qubit is perfectly in equilibrium with its fridge base plate environment, and thus the probability of the qubit being in the excited state for a given temperature is described by the Bose-Einstein distribution

$$P_e = \frac{1}{e^{\hbar\omega_q/k_b T} - 1} \tag{4.2}$$

With the values given above, this corresponds to an idle excited state population of $10^{-5}\%$. In reality, we usually measure excited state populations that are orders of magnitude above this value–anywhere from a few percent to 20% [33]. There are many channels that thermal photons may reach the qubit and we will cover the methods [13, 39] used to mitigate the leakage in this section:

1. Using a dilution fridge
2. Design choices of elements on the quantum processor
3. Wiring the fridge and filter choices
4. Heralding

However, even assuming all of these techniques have been perfectly implemented, many research groups report that there is still thermal population that could be due to hot quasiparticles [33]. In addition we include a few sections on general hardware choices such as

1. processor fabrication
2. amplification techniques
3. Rabi drive stabilization

4.1.1 Basic principles of a ^{3}He/^{4}He refrigerator

Dilution refrigerators are composed of multiple stages as shown in Fig. 4.1, with each descending stage colder than the previous one. The coldest one is typically at 10-20 mK. This is where the sample is attached and thermalized. Coaxial cables that bring signals from room temperature down to the sample on the base stage are carefully designed with specific materials at each stage to help thermalize them properly at every stage. We discuss this in more detail in the wiring section.

The highest stage is the 70K stage. Historically in wet fridges, the helium mixture was pre-cooled to 77K by a bath of liquid nitrogen, giving this stage its name. However, with advances in fridge technology, most fridges now are closed circuit dry fridges (wet fridges are only used for vibration sensitive experiments). Two-stage pulse tube coolers first bring the top plate down to 50K and the next stage down to 4K. We now explain the basic principles of evaporative cooling using just ^{4}He, a key concept for the rest of this section. First we have a small pot of ^{4}He that is constantly fed from the main bath of helium. A pump a used to reduce the pressure inside the pump to approximately 0.1 mbar. The reduced pressure accelerates the evaporation of He particles from the surface of the liquid, causing cooling. The limit of lowest achievable temperature is set by the exponentially decreasing vapor pressure. At low enough pressure there will not be enough helium to evaporate to cool the setup further. This method can cool the system to 1.2K. However, modern fridges don't

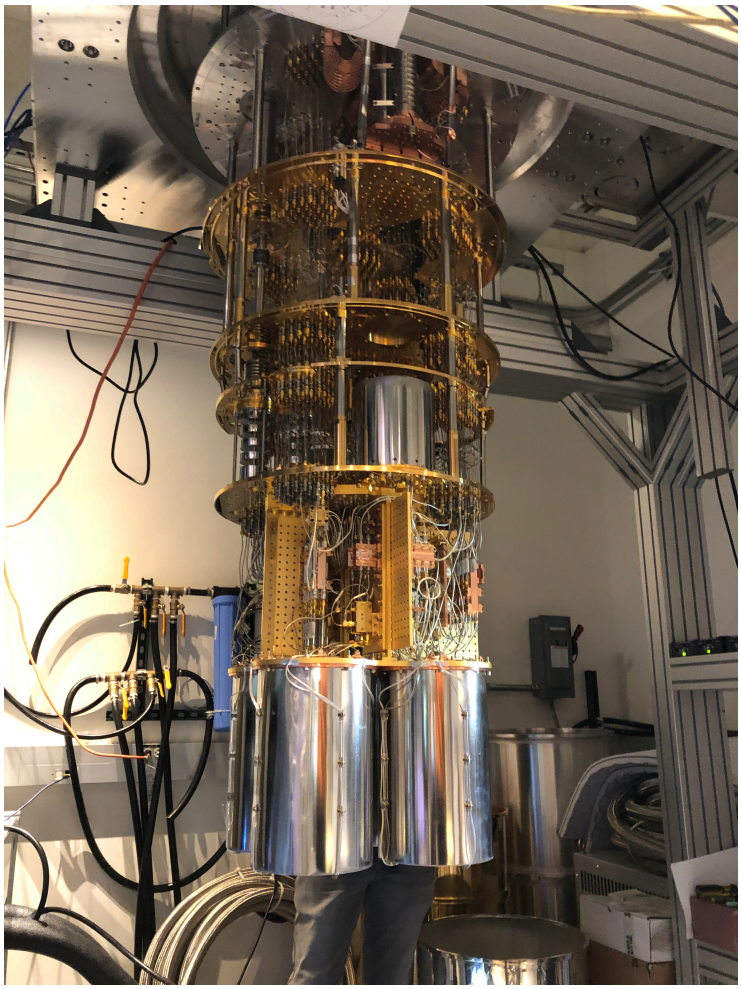

Figure 4.1: Example of the inside of a dilution refrigerator. This is the Bluefors fridge in Campbell 109.

just use one type of helium. It is the mixture of ^{3}He and ^{4}He. Here are the key differences between ^{3}He ^{4}He:

1. ^{3}He is lighter, which means that the binding energy between molecules is smaller. This means it has higher vapor pressure than its heavier counterpart. We can repeat the evaporative cooling process above with ^{3}He, allowing us to cool to 0.25 K.

2. Pure ^{4}He has a nuclear spin of $l = 0$, meaning that it obeys Boson statistics. It undergoes superfluid transition at 2.7 K.

3. In contrast, ^{3}He has $l = 1/2$, obeys Fermi statistics. The Pauli Exclusion principle suppresses the superfluid transition until much lower temperatures. Then the ^{3}He pair up into bosons and obey Boson statistics, forming a superfluid.

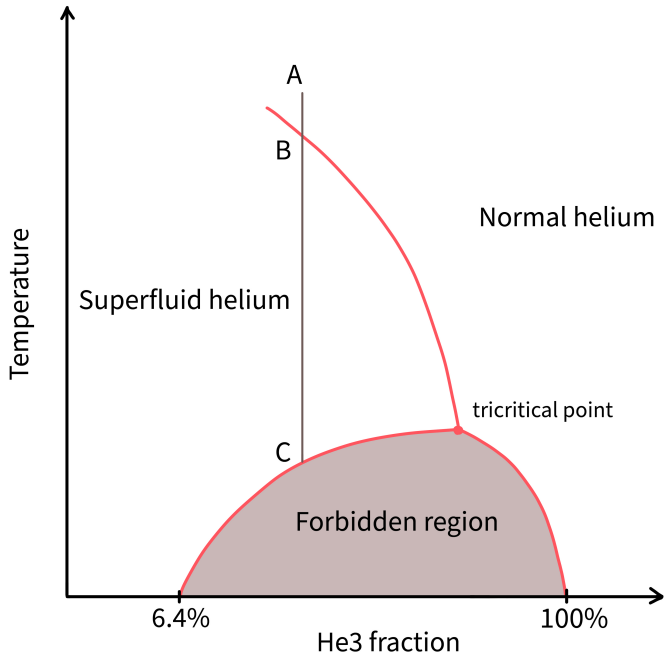

Figure 4.2: Phase diagram of helium

When ^{3}He and ^{4}He are mixed together, the superfluid transition temperature depends on the ratio of the concentrations, according to the phase diagram Fig. 4.2. If we follow the path drawn, we see that if we start at point A, we are in the normal phase. Cooling the temperature down takes us to point B, where we enter the superfluid phase. Cooling it further still takes the mix to a state where it is split between a ^{3}He rich phase floating on top of a ^{3}He poor phase (the dilute phase that contains mostly ^{4}He) as shown in Fig. 4.3. The key part is that the dilute phase has a finite solubility of ^{3}He at 6.4%.

If we remove ^{3}He from the dilute phase, then molecules of ^{3}He from the top layer will go into the bottom layer to fill the vacant spots to keep the concentration at 6.4%. This is similar to the evaporative cooling process we described earlier. This is all happening in the mixing chamber of the fridge that is located on the base plate. A line that is connected to the mixing chamber from the Still pumps on the bottom of the mixing chamber and is what pulls these ^{3}He molecules from the dilute phase on the bottom. A return line takes the ^{3}He and feeds it back to the ^{3}He rich layer in the mixing chamber. The still is typically heated such that it is not cooled too much by the ^{3}He that is pulled up from the mixing chamber. Otherwise the vapour pressure will get too low and the cycling of ^{3}He will stop. We usually toggle the heater to a level that keeps the still at 700-800 mK.

The reason that this dilution process is more effective than evaporative cooling is that the

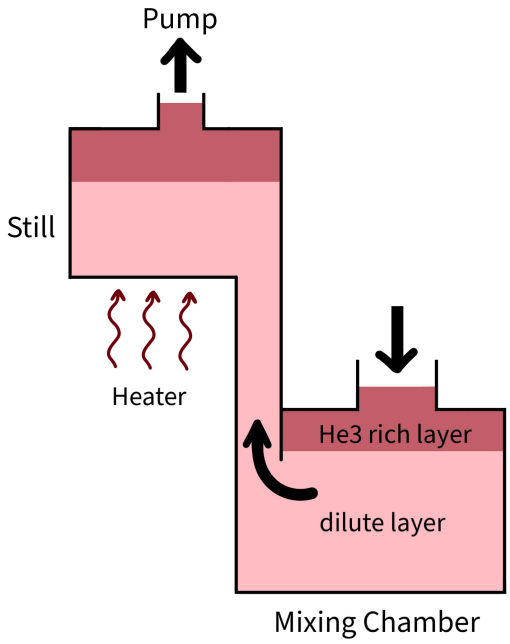

Figure 4.3: Illustration of how cooling works from the mixing of the isotopes of helium

solubility of ^{3}He in ^{4}He is constant. This is analogous to the vapour pressure in evaporative cooling except the vapour pressure decreases over time. This technique allows us to reach base temperatures of 10 mK.

4.2 Wiring

The dilution refrigerator has lines that pass through each stage down, connecting from the room temperature environment down to the mixing chamber at 15 mK. These lines include coaxial drive lines that carry qubit drive signals, readout pulses, DC lines for flux tuning, pumps for amplifiers, and also readout lines that carry return signals from the chip that carry information about the qubit state. As mentioned previously, each of these channels are critical to running the processor but are also sources of thermal noise that can lead to qubit dephasing, quasiparticles, and reduced lifetimes. We use thermalization of the lines of cables, attenuators, and other microwave components at each stage of the dilution refrigerator to mitigate the loss[4]. We also add filters with stop-bands outside of the qubit frequency ranges to further suppress thermal radiation or 1/f noise from electronics. Because

[4]Kpop fun fact 4/7: The singer Rain auditioned and was rejected by the company JYP 18 times before he was accepted on the 19th try. He went on to make 7 albums and win dozens of awards. Never give up folx.

this experiment used only fixed frequency qubits, we will neglect details about DC line wiring and focus on coax lines for qubit control and readout. We refer the reader to more details at [40].

There are two types of heat loads on the lines: passive and active. Passive loads refers to heat energy which is gained by the stages of the fridge due to conduction through the microwave lines. Between stages, the heat flow to stage i with temperature T_i from stage $i-1$ for a single coax cable is

$$P_i = \int_{T_{i-1}}^{T_i} dT \frac{\rho_O(T)A_O + \rho_d(T)A_d + \rho_c(T)A_c}{L_i} \tag{4.3}$$

where ρ_O, ρ_d, ρ_c and A_O, A_d, and A_c are thermal conductivities and cross-sections of the outer conductor, dielectic, and center conductor respectively. The length of the cable between the stages is denoted by L_i. To minimize this, we use stainless steel cables, a material with low thermal conductivity to connect between the stages of the fridge up until the mixing chamber base. Because stainless steel (SS-SS) wires are very touch to bend and shape, an alternative that is sometimes used is Cupronickle (CuNi), which has a 20-30% higher passive load than SS-SS cables. For the sample at base, we use superconducting wire (typically copper) since everything should be at the same temperature.

In addition to passive loads, the attenuation on the lines and the qubit control and readout signals add an active load. Attenuation is needed on the RF lines to reduce room temperature blackbody radiation, also known as Johnson noise. The thermal photons come from room temperature and propagate down the lines through the lower temperature stages. The photon occupation number at the base temperature is approximately 10^{-3} and we typically need at least 60 dB of attenuation is spread along the lines to lower the thermal radiation to match the mixing chamber environment level. This forces us to use much higher power from the room temperature generators to have enough power at the qubit after it passes through the attenuation. To calculate how we spread this attenuation among the stages, we use the following formula, which describes the photon occupation n_i at each stage i in terms of the attenuation A_i, the temperature T_i, at the frequency of interest:

$$n_i(\nu) = \frac{n_{i-1}(\nu)}{A_i} + \frac{A_i - 1}{A_i} \frac{1}{e^{h\nu/k_B T} - 1} \tag{4.4}$$

Knowing that we would like to have $n_{\text{mixing chamber}} = 10^{-3}$, we can use the equation above to calculate solve for n_{still} with a guess value for the attenuation A_i and chain together the calculation stage by stage up. It turns out the putting 20 dB on each stage works pretty well. Each 20 dB attenuator is designed to transmit 1% of the signal and 99% of of the radiation is dissipated in the attenuator. We carefully thermalize the attenuators at each stage to keep. The cooling power of each stage is large enough such that the stage just acts as an infinite cold bath for the attenuation. For a Bluefors XLD400, the cooling power at each stage is 30 W, 1.5 W, 40×10^{-3} W, 200×10^{-6} W, and 19×10^{-6} W at the 50K, 4K, Still, CP, and MXC stages, respectively.

A diagram of the wiring used for our final experiment setup in the Campbell Bluefors fridge (Blizzard) is shown in Fig. 4.4. The qubit drive lines are labeled by Q_i. These drive lines have 20 dB attenuators that are anchored and thermalized to each stage. Once we reach the base stage, we use additional filters as an added layer of protection from noise. The high pass filter has a cutoff around 4 GHz, just below the qubit frequencies at around 5 GHz. The low pass filter has a cutoff around 10 GHz. These filters allow qubit and readout pulses through but limit noise at other frequencies. We omit the lowpass filter on the shared resonator in line because the we require high power for the gate sidebands and there would be a decent amount of attenuation at around 8 GHz.

4.3 Amplifiers

In the wiring diagram in Fig. 4.4, there are two amplifiers, both located on the readout out line, that are key to allowing us to detect the qubit state. They amplify the return signal that has been reflected off of the readout resonators on the chip. There is a fundamental limit to how well we can detect the qubit state due to the uncertainty principle and the zero-point energy fluctuations. Furthermore, for superconducting qubit measurements, we are dealing with measurements close to this quantum limit at the mixing chamber, but we must also pass this signal back to room temperature for acquisition while maintaining enough signal to over noise floor of room temperature electronics. For these reasons, the signal carrying the qubit state is not just a point in phase space. It has an area of uncertainty around it that follows a Gaussian distribution, one for each of ground and excited states. The goal of the amplifiers is to maximize the separation between the two Gaussians while also minimizing contributions the size of the Gaussians at each amplification stage.

To better understand amplifiers, we introduce a couple of parameters that amplifiers are often evaluated by. The first is the noise temperature, and related, the noise power. A resistor at temperature T has Johnson noise, which comes from the vibration of the electrons due to their kinetic energy. The electrons jostling causes the voltage to fluctuate as well with a root mean square value given by

$$V_{rms} = \sqrt{\frac{4hf\Delta fR}{e^{hf/k_bT} - 1}} \tag{4.5}$$

which reduces to

$$V_{rms} = \sqrt{hfk_bT\Delta fR} \tag{4.6}$$

in the limit of high temperature. This is also known as the Rayleigh-Jeans limit. The Thevenin equivalent of a noisy resistor is a noiseless resistor plus a generator with a given voltage. With this picture, we can convert the voltage above to an rms power

$$N_{rms} = \left(\frac{V_{rms}}{2R}\right)^2 R = k_BT\Delta f. \tag{4.7}$$

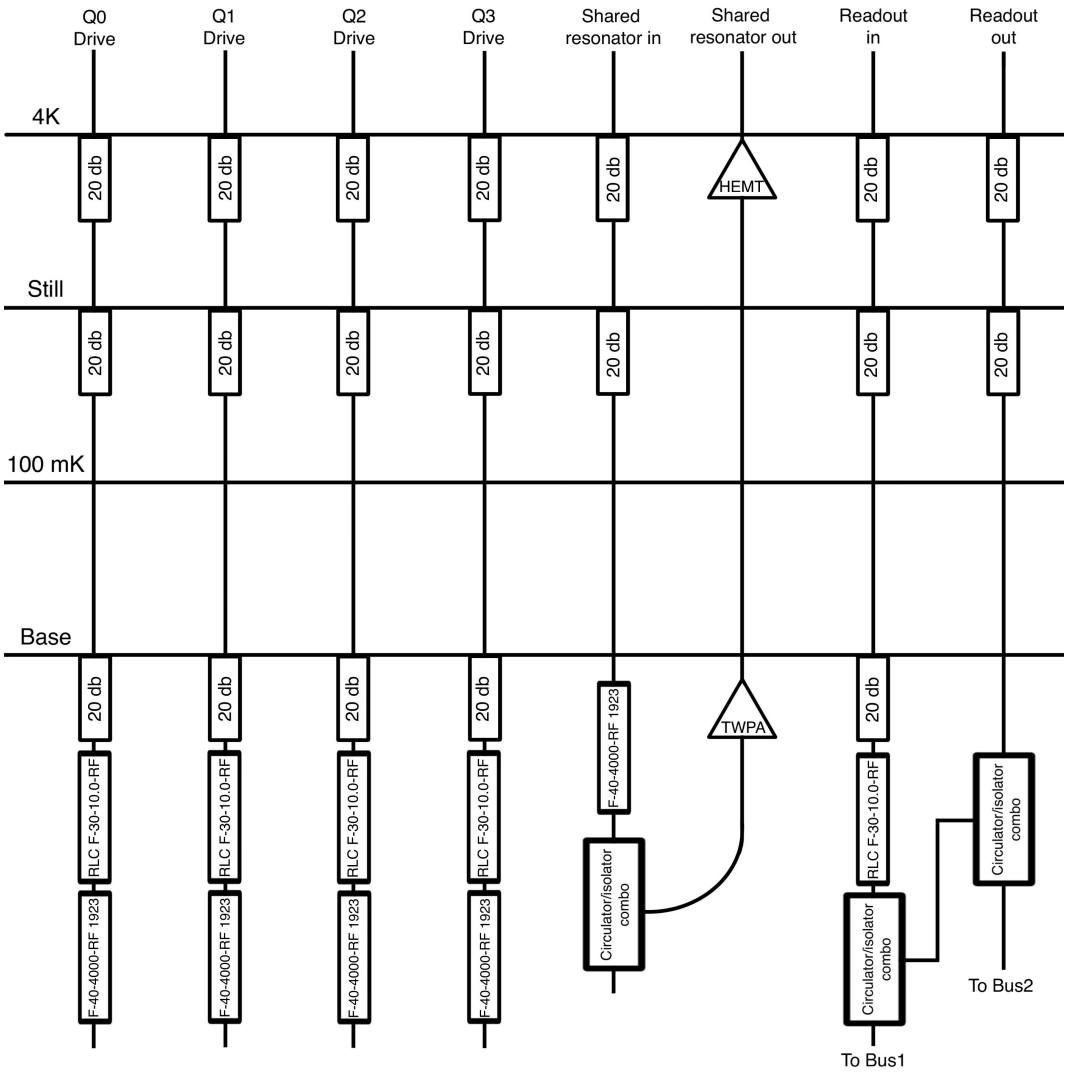

Figure 4.4: Wiring diagram for all filters and attenuators in the fridge. We note that there is no attenuation on the readout line and we have two amplifiers: one at 4K and one at base.

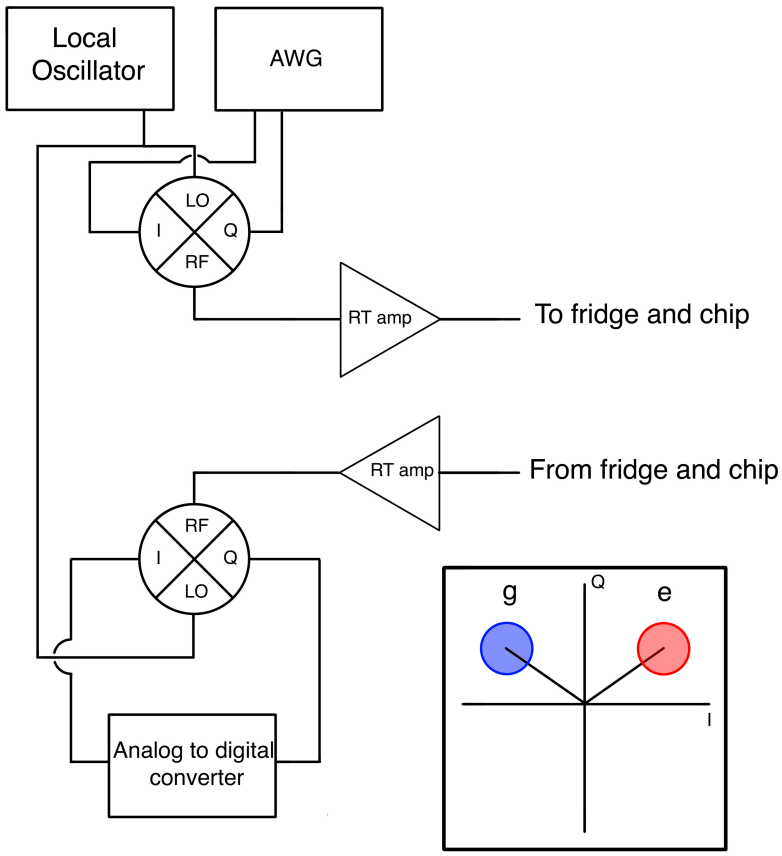

Figure 4.5: Room temperature setup to generate readout signals and demodulate signal returning from the chip. The local oscillator (LO, typically around 6 GHz) is mixed with signals from an Arbitrary Waveform Generator (AWG) in the hundreds of MHz regime. This generates sidebands from the mixer at the resonator frequencies. These signals are sent to the chip and reflected off of the readout resonators for multiplexed readout. The signals at come back with a qubit state dependent frequency shift. Returning from the fridge, the signal is demodulated again by using the mixer in reverse and remixing with the LO. Finally it is captured by an analog to digital converter (ADC) to be processed by the computer for another demodulation step. We perform an FFT of the signal captured by the ADC and record the power of at the frequencies corresponding to the readout resonators as a function of time to obtain a time trace. We integrate this time trace to obtain a single point in the IQ plane, as shown in the inset in the bottom left. Each blob in the inset is the result of thousands of measurements.

When the noise power white as it is here, meaning is not dependent on the frequency, it can be characterized with an equivalent noise temperature obtained from rearranging the above equation

$$T_e = \frac{N}{Gk_B\Delta f} \tag{4.8}$$

where G is the gain of an amplifier.

Another useful descriptor is the noise figure which is a measure of how much the signal to noise ratio decreases between the input and output of the amplifier:

$$F = \frac{S_i/N_i}{S_o/N_o} \geq 1. \tag{4.9}$$

Given that $S_o = GS_i$, $N_i = k_BT_0\Delta f$, and $N_o = k_B\Delta f(T_o + T_e)$ where T_o and T_e are the equivalent noise temperature before the amplifier and of the amplifier, respectively, we can rewrite F

$$F = 1 + \frac{T_e}{T_0} \geq 1. \tag{4.10}$$

.

We typically use multiple amplifiers in a chain, and in this situation, the first amplifier is the most important. Each amplifier increases the amount of signal. However, there will be some degradation of signal to noise ratio through each segment of the chain. We can calculate the overall dynamics using the concepts we built up above. Consider two amplifiers with gain G_i, noise figures F_i, and equivalent noise temperatures T_{ei}. The noise power after the first amplifier is

$$N_1 = G_1k_BT_0\Delta f + G_1k_BT_{e1}\Delta f. \tag{4.11}$$

The noise power after the second stage is

$$N_o = G_2N_1 + G_2k_BT_{e2}\Delta f = G_1G_2k_B\Delta f(T_{cas} + T_0) \tag{4.12}$$

$$T_{cas} = T_{e1} + \frac{1}{G_1}T_{e2}. \tag{4.13}$$

We can convert the noise power to a noise figure

$$F_{cas} = F_1 + \frac{1}{G_1}(F_2 - 1). \tag{4.14}$$

Extending this method for more than two amplifiers we have:

$$T_{cas} = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1G_2} + ... \tag{4.15}$$

$$F_{cas} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2} + ... \tag{4.16}$$

These equations show that the performance of the system is dominated by the first amplifier. The first amplifier should have the lowest noise figure and as much gain as possible.

The later stages have much smaller impact. For this reason, much research has gone into developing amplifiers for the base stage, closest to the processor.

For a description of the room temperature setup, see Fig. 4.5. The base stage amplifier is called a Josephson traveling wave parametric amplifier (JTWPA) on the base stage that is on the readout out line. This is a wide bandwidth cryogenic amplifier developed at QNL in 2015 [42]. The TWPA is made from a chain of a couple thousand Josephson junctions. In contrast to other amplifiers, such as the JPA that amplify over a few tens of MHz, the TWPA amplifies with 20 dB of gain over a wide bandwidth–over 4 GHz. The JTWPA is operated using a pump tone that travels down the line of Josephson junctions along with the signal, facilitating four-wave mixing [61]. The pump is coupled into the signal line using a directional coupler right before entering the TWPA. Due to its sensitive to changes in the magnetic field, the JTWPA is always encased in a MuMetal shield and thermalized to the mixing chamber base.

The second amplifier that the return signal passes through is the Low Noise Factory LNC4 8A High Electron Mobility Transistor (HEMT) LNF-LNC4_8C s/n 1666H. This amplifier is attached to the 4K plate of the fridge. The HEMT provides 40 dB of gain by generally achieves noise temperatures at least 10 times the quantum limit.

4.3.1 Heralding

Despite best efforts with a dilution fridge in conjunction with best wiring and shielding practices, typically we find some excited state population. To compensate for this, we implement heralding to reduce the effect of spurious excited state population due to thermal photons. Heralding has been developed on many qubit platforms including trapped ion, photonic, and quantum dot systems. This work was demonstrated on the superconducting qubit platform by Johnson et al. in 2012 [34]. Heralding is implemented by inserting an additional readout pulse before the start of the intended qubit manipulation and final measurement. This allows us to check what the state of the qubit is before each measurement. For instance, a T_1 measurement is done by preparing a series of pulse sequences where you apply a π pulse to the qubit followed by a wait time $n\Delta T$ and end with a pulse sent to the readout resonator to probe the qubit state. Here, n indexes into the series. To add heralding, we would insert an additional readout pulse before each π pulse. The signals from both readout pulses would be captured by the analog to digital converter the collects the data. After all measurements have been collected, we check each result from the inserted readout measurements. Then in post processing, we throw out any results for which this initial readout measured that the qubit was in the excited state. Depending on the temperature at base and the quality of the processor shielding, the percentage thrown out can be anywhere between 1% and 10%. This technique reduces the effect of an incorrect initial state.

4.4 Device Design

We designed a 4-qubit fixed frequency transmon device ($E_J/E_C = 70$), with all-to-all connectivity through a shared coplanar waveguide resonator. Each qubit has its own qubit control line and readout resonator ($\lambda/4$ resonator). The readout resonators on the left and right are coupled to a readout bus ($\lambda/2$) for multiplexed readout. There were a few parameters on the chip that we paid special attention to.

1. We prioritized simulating the couplings of the qubits and the shared resonator (χ), making them as close to equal as possible. While it would have been easier to have equal coupling if we placed the qubits on the left and right directly opposite of each other in a symmetric fashion, we also wanted to avoid direct qubit-qubit coupling. Placing qubits directly across from one another would have maximized the overlap of their dipoles. This is why you see the qubits slightly offset from each other to minimize this.

2. In practice, we found that χ varied quite a bit from chip to chip. On some chips, the simulations matched the couplings very well. However, on other chips, we had some qubits that matched the simulations and other chips that were up to 2x larger in χ, despite all of the chips coming from the same design. One possible explanation is that the wirebonding and packaging environment could be a huge variable. To help prevent bulk chip modes, we always bond over the shared resonator down the middle. Since this was always done by hand (by me), there could be a lot of difference per chip.

3. We also tried to minimize the value of κ_{ext} of the shared resonator. We cannot make $\kappa_{\text{ext}} = 0$ because we need to be able to send the red and blue sidebands in. However, we also need to minimize the chance of photons leaking out of the resonator during the gate. Unfortunately, κ_{int}, the internal loss of the resonator, is dependent on the fabrication process, material quality, and packaging environment. These are harder to control for. There have been demonstrations of niobium resonators with $> 10^6$ in quality factor, but these chips have not undergone the qubit evaporation process, which is believed to reduce resonator quality factors.

4. We made several iterations of the chip testing different magnitudes of χ. A larger χ contributes to a faster gate, ideally helping fight limiting coherence times. However, when we made chips with higher χ, we also found that κ increased as well, that there was a trade off between the two parameters. This was unexpected. One possible explanation is that the qubits coupled to the shared resonator also serve as a source of absorption and loss for any photons in the shared resonator. However, we only tested a couple of chips with higher coupling and a more detailed, systematic study should be made.

5. We optimized the qubit control line coupling to have both strong enough coupling to have Rabi drives of at least 40 MHz, but also avoid strong T_1 decays.

Parameter	Ideal Value	Experimental Value
ω_{ge}^0	5.2 GHz	5.24 GHz
ω_{ge}^1	5.4 GHz	5.37 GHz
ω_{ge}^2	5.8 GHz	5.69 GHz
ω_{ge}^3	5.6 GHz	5.48 GHz
Shared resonator χ_0	500 KHz	380 KHz
Shared resonator χ_1	500 KHz	410 KHz
Shared resonator χ_2	500 KHz	718 KHz
Shared resonator χ_3	500 KHz	815 KHz
ω_c	7.84 GHz	7.82 GHz
Shared resonator κ_{ext}	30 KHz	20 KHz
Shared resonator κ_{int}	0 KHz	180 KHz
RO resonator 0	6.2 GHz	6.4 GHz
RO resonator 1	6.4 GHz	6.5 GHz
RO resonator 2	6.8 GHz	6.8 GHz
RO resonator 3	6.6 GHz	6.6 GHz

Table 4.1: Summary of commonly referenced chip parameters

All simulations were done in ANSYS HFSS, a 3D electromagnetic (EM) simulation software for designing and simulating high-frequency electronic products.

In Table 4.1, we list some relevant frequencies. We point out that the two qubit gate was done on Q0 and Q1, while the 3-qubit gate used Q0, Q1, and Q2. Q2 has much stronger coupling relative to Q0 and Q1, which we believe to be one of the limitations on the gate. Our target values are in the middle column and the actual experimental values in the last column.

4.5 Device Fabrication

Here we broadly summarize the steps for fabricating the sample.

1. Clean silicon wafer substrate. Some other labs use sapphire.

2. Deposit a layer of niobium. This becomes the ground plane of the processor. Technically it is possible to use aluminum as well and one might expect to use aluminum especially given demonstrations of million Q 3D cavities that use aluminum. However, somehow people have gotten better results in our lab with niobium. Also niobium has a larger bandgap and a higher T_C which is beneficial when it comes to avoiding quasiparticles.

3. The niobium is then coated with two layers of resist and patterned using a Raith e-beam writer. The Raith exposes the resist to electrons. The resist under any exposed

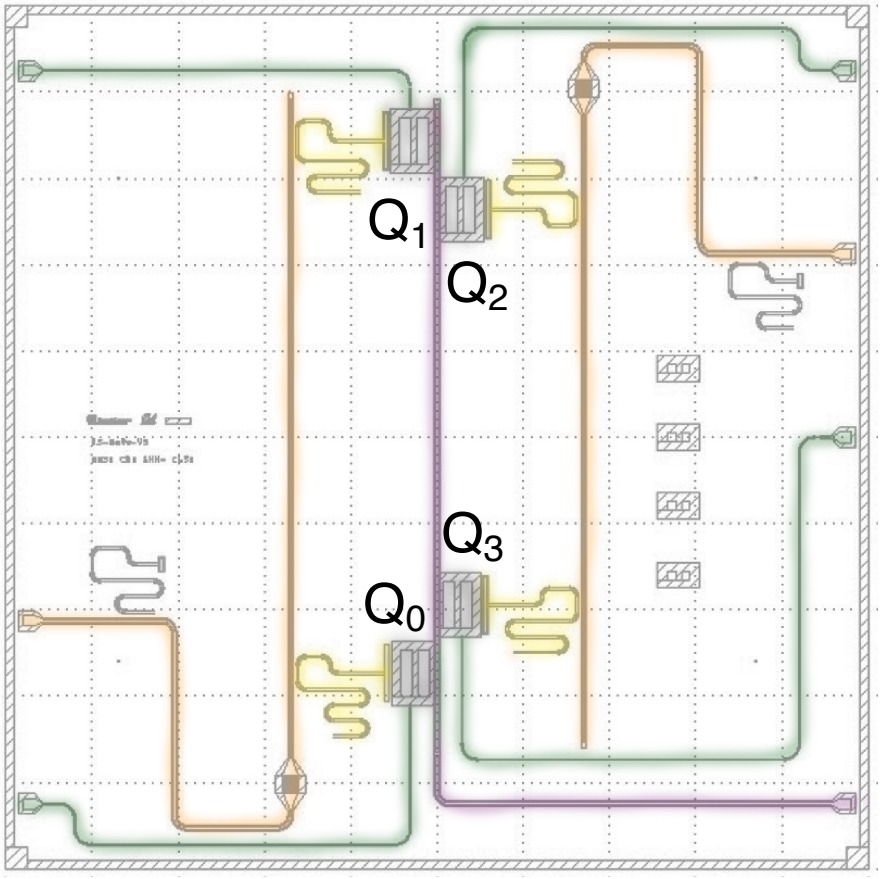

Figure 4.6: Chip with 4 fixed frequency qubits (grey boxes) coupled to a shared resonator in the middle (purple). We send the sidebands into this line. Each qubit has its own qubit control line (green). This is where we send any state preparation or tomography pulses and the Rabi drive. Each qubit also has its individual readout resonator (yellow) coupled to two readout buses (orange) on the left and right for multiplexed readout. Each readout bus also has an interdigitated capacitor that helps form a Purcell filter to limit qubit decays through the readout lines.

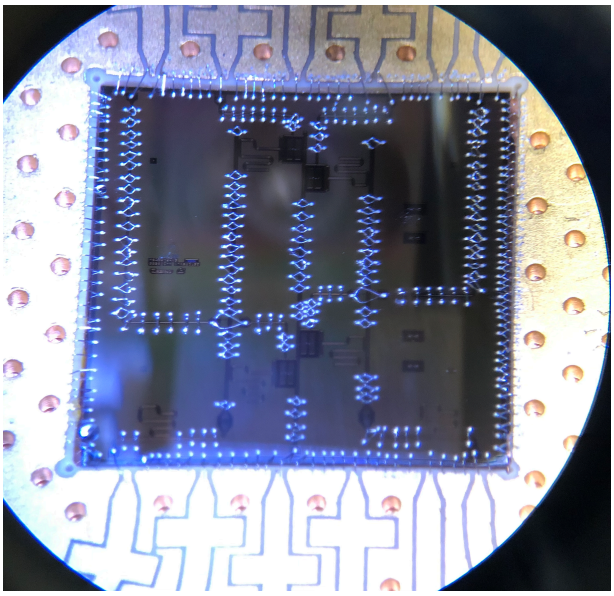

Figure 4.7: Image of chip after wirebonding. We wirebond all edges of the chip to the ground plane of the PCB. Whenever possible, we use two wirebonds to connect each signal trace to the chip bond pads. In addition, we bond over any long control lines or resonators such as the Purcell filter readout bus and the shared resonator in the middle.

areas will dissolve away in the development process leaving exposed niobium. This allows us to then etch away these parts of the niobium.

4. We then repeat the resist and patterning process for the Josephson junctions.

5. The wafer, now with resist on top and the junction pattern exposed is placed into the Plassys. Then aluminum is evaporated onto the entire sample covering everything. A liftoff process is performed after where all the areas of aluminum with resist underneath are shed from the wafer, leaving only the areas where the aluminum was laid directly on top of the sample, forming the junctions.

6. We note that the junction evaporation process is actually a three step process where a layer of aluminum is evaporated, followed by an oxidation step where the chamber is filled with oxygen and an oxide is allowed to form on the surface, followed by a final aluminum evaporation step to finish the Josephson junction.

We note here that since niobium is a superconducting material too, one could make junctions out of niobium. However, the oxide layer for niobium is notoriously difficult to work with. There is a lot of dielectric loss, resulting in much shorter lifetimes for qubits. Experiments that use a niobium junction have to use a tri-layer process (Nb-Al-AlOx).

After fabrication, the wafer of 64 chips is probed to measure their room temperature resistance, which allows us to predict their frequency once cooled down to superconducting

temperatures. The room temperature resistance can be related to the critical current of the qubit through the Ambegaokar-Baratoff formula:

$$I_C = \frac{\pi\Delta(T)}{2eR_n}\tanh\frac{\Delta(T)}{2k_bT} \tag{4.17}$$

which becomes

$$I_C = \frac{\pi\Delta(T=0)}{2eR_n} \tag{4.18}$$

in the limit of $T << T_C$. From here, one can use the following equations that we derived in Chapter 2 to calculate the qubit frequency from I_C:

$$\omega_q = \sqrt{8E_JE_C} - E_C \tag{4.19}$$

$$E_C = \frac{e^2}{2C_\Sigma} \tag{4.20}$$

$$E_J = \frac{\Phi_0 I_C}{2\pi} \tag{4.21}$$

The first few years, we probed the wafers by hand, but for the last fabrication round, we used the Micromanipulator P200L semi-automatic probe station to obtain the room temperature resistance of Josephson junctions on a standard 8 x 8 array of samples. The tool uses pattern recognition to track its location on the wafer and then shifted the stage for the probe to access each qubit site. Two probe tips made from tungston carbide and are used to punch through the hard native Nb oxide of each qubit capacitor paddle simultaneously and a voltage is applied to measure the resistance going through the junction.

After probing, the wafer is diced and a selected set of chip is cleaned. We package each chip in a boxes made from oxygen-free copper, which has the best thermal conductivity at low temperatures. The box has a PCB inside with traces that connect SMA ports on the outside of the box to the chip in the middle of the PCB. The middle of the PCB has a cutout that is designed to fit our chips exactly. We GE varnish to glue the chip to two ledges below the cutout that support the chip. The ledges also create an 0.5 mm airgap between the chip and the bottom of the box to reduce the effective dielectric constant of the chip mode. This increases the chip mode frequency to be higher than the frequency range of the circuit elements. We wirebond the ground plane of the chip to the ground plane of the box and also use wirebonds to connect all of the drive lines on the chip to lines on the PCB. Typically we use two wirebonds per bondpad on the chip. Finally, we also wirebond across any major control lines and resonators on the chip itself to prevent slotline modes.

4.6 Rabi Drive Stabilization

The key feature of the gate in this thesis is the Rabi drive and the tunability of the dressed basis states. However, this flexibility also requires precise control of this parameter from

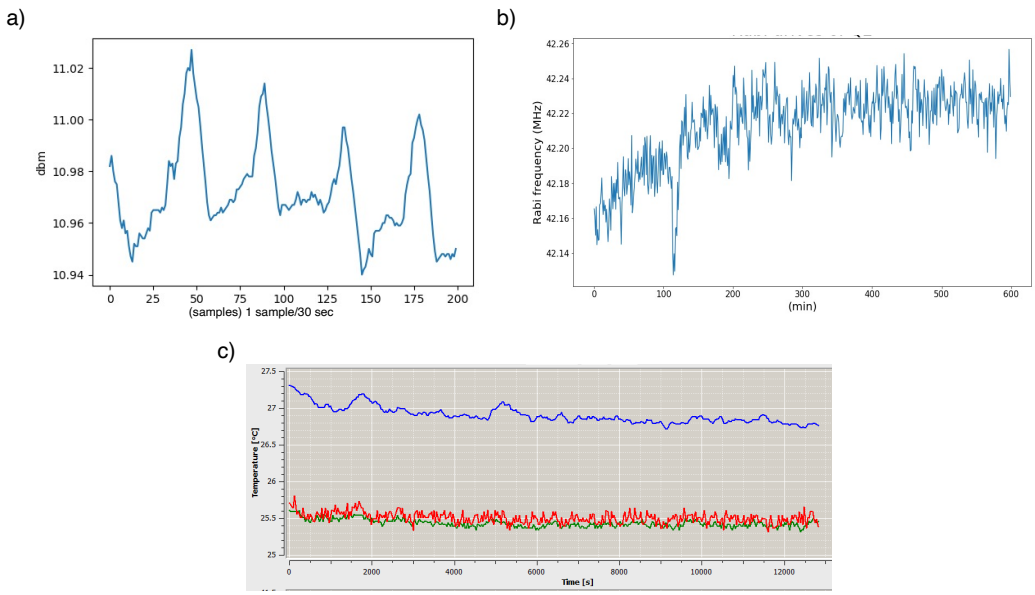

Figure 4.8: a) Fluctuations in power of an Agilent generator in the laboratory. Samples of the power output for constant generator settings are taken once every 30 seconds. b) Rabi response of a single qubit with the water cooled racks installed. Rabi drive is able to be stabilized to within 60 KHz of target. We see an exponential rise upwards at the beginning. That is when we first close the enclosure doors and let the temperature inside the racks equilibrate. We see a sharp dip at a little just past 100 min. We quickly open the doors and see a sharp drop in the Rabi drive. c) temperature inside racks over several hours.

the experiment. One challenge that we experienced early on was the fluctuation in qubit drive power due to the room temperature electronics. We found that the generators and amplifiers all fluctuated in the output based on the temperature of the room itself. With an air conditioning unit that turned on and off every 25-30 minutes, we saw the the Rabi drive fluctuated on a period of 30 minutes with maximum difference up to 1 MHz, making any Rabi calibration completely obsolete by the time we were ready to measure the gate. This is especially detrimental when our sideband detunings are small–on the order of a few MHz, such that any 1 MHz shift in the Rabi drives is significant compared to the sideband detunings. In addition, broader average room temperature would cycle with the day and night and the amount of sunlight in the room. It was very hard to take data overnight. To solve the temperature drift issue, we collaborated with Martin Int'l Enclosures to design water cooled racks that kept the generator environment stable and helped us stabilize the Rabi frequency to within 60 KHz.

Chapter 5

Implementation

In this section we describe the experimental details of tuning up and implementing the gate.

5.1 Pulse Sequence

The pulse sequence for this gate is shown in Fig. 5.1 and is implemented as follows:

1. Turn on sidebands and send them into the shared resonator. We start the sidebands 10μs before the qubit pulses because the shared resonator has a narrow width and we need to fill the shared resonator to a steady state before beginning the qubit pulses. If the shared resonator is not in a steady state, the Stark shift on the qubits will be changing during the course of the gate and that will affect the quality of all the pulses.

2. The qubits have been sitting in the ground state while the resonator is being filled. Once the resonator is in steady state, we do a state preparation pulse. Typically if we are just trying to make a GHZ state with the gate, we will prepare in either the $|+\rangle$ or $|-\rangle$ state. We do this by sending a $\frac{\pi}{2}_x$ pulse to the qubit. However, during process tomography, we have other state preparation pulses as we are preparing a variety of initial states that span the two qubit subspace. More about this in the gate characterization sections.

3. Then to do the gate, we must bring the qubits close to resonance with the sidebands and that requires us to turn on the Rabi drive. Note that the axis of the Rabi drive is always perpendicular to the axis of the $\frac{\pi}{2}_x$ pulses for state preparation. This means that for the states $|\pm\rangle$, the Bloch vector does not continue to spin around the YZ plane. Instead, it may slowly evolve from one basis state to another basis state.

4. After the duration of the gate (which is just the length of the Rabi drive) we finish with a final qubit pulse that maps any population we are trying to measure back to the Z-axis for readout.

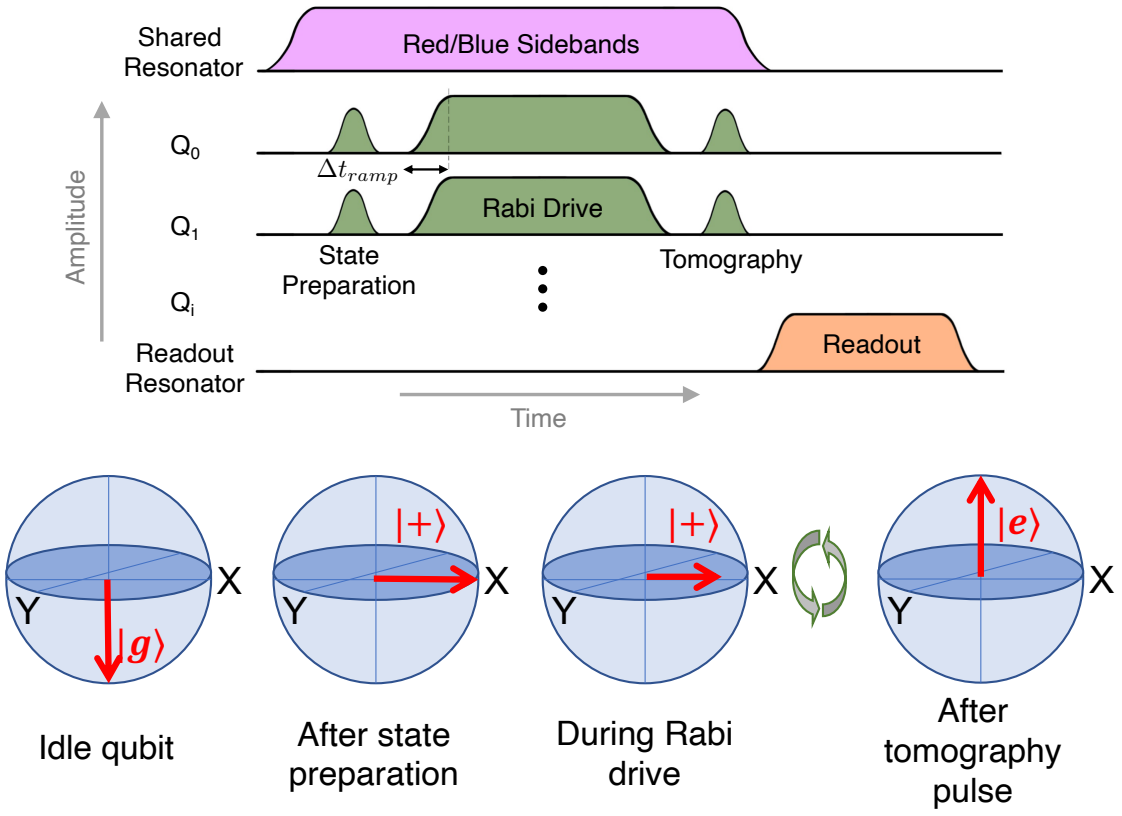

Figure 5.1: Pulse sequence. The qubit begins idle, typically sitting in the ground state. With the sidebands on, a state preparation pulse is applied to the qubit to prepare it in an eigenstate of the dressed frame. This is usually a $\pm\pi/2$ pulse around the x-axis. However, without a Rabi drive, the $|+\rangle$ and $|-\rangle$ states are degenrate. Thus we turn on a Rabi drive about the y-axis to split this degeneracy. Since the drive axis is parallel with the Bloch vector, the state does not precess. After the appropriate interaction time, a final qubit pulse is applied to map the chose component of the Bloch vector to the z-axis for measurement.

The gate is scaled by adding a Rabi drive to any qubit coupled to the shared resonator that you wish to participate in the gate. Qubits without a Rabi drive are completely out of resonance with the sidebands by at least 2 GHz, and therefore just experience a Stark shift as the sit idly.

We point out here that the pulse sequence is reminiscent of the spinlocking pulse sequence that we discussed in the introduction when we set the state preparation and tomography pulses to both be $\frac{\pi}{2}_x$. This gives us some intuition about system.

1. The gate is inherently built to dynamically decouple from low frequency noise. The lifetimes most relevant to the gate will be the dressed frame lifetimes $T_{1\rho}$ and $T_{2\rho}$, which are obtained from measurements like the spinlocking measurement. The spinlocking pulse sequence is a dynamical decoupling method. Especially in the presence of non-Gaussian noise, like 1/f noise, these lifetimes are typically longer than the bare frame T_1 and T_2 values.

2. William Oliver's group at MIT has used spinlocking measurements as a form of noise detection. Sweeping the Rabi drive frequency allows them to adjust the qubit's sensitivity to the noise spectrum in frequency space. Changes in the decay function give insight into the type of noise the qubit experiences.

3. The first and last $\pi/2$ pulses are just pulses that map between the bare to dressed qubit frames. The Rabi drive in the middle splits the degenerate energy levels of the $|+\rangle$ and $|-\rangle$ states qubit, turning the qubit into a dressed qubit.

5.2 Calibration

To tune up a new gate, we had to develop calibration protocols for the most important parameters of the gate: δ, Ω, and ϕ_{sb}. For inspiration, we look to the ion community to see how they calibrate. When the ion experiments tune up the Mølmer-Sørensen gate, they measure the individual qubit responses to the sidebands and tune the sidebands until each qubit has an equal oscillation rate between $|g\rangle$ and $|e\rangle$ due to individual sideband drives. Given that we are in the driven frame, we would like to generate similar single qubit population flops using sidebands between the $|+\rangle$ and $|-\rangle$.

We point out in the section above that the pulse sequence for the gate resembles the classic NMR spinlocking pulse sequence implemented on multiple qubits simultaneously, under the presence of sidebands. We take inspiration from past experiments from our lab. In 2012, Kater Murch used a Rabi driven qubit coupled to a dissipative cavity and additional sidebands to show that one could cool the qubit to either the $|+\rangle$ or $|-\rangle$ state. The experiment used a spinlocking-like pulse sequence with sidebands, similar to our experiment. The key difference between their system and ours is that the decay of the cavity is engineered to be much stronger in their system compared to the qubit-cavity coupling. Conversely, in our system, our qubit-resonator couplings are higher, giving the qubit an opportunity to

exchange with the cavity multiple times before the resonator decays. We also expect a cooling dynamic to be present in our system set by κ, but because we are in the high χ/κ regime, we also expect that the sidebands will drive population swaps between the $|+\rangle$ and $|-\rangle$ states before the decay pins you to a final state. These population swaps are exactly the single qubit drives that are analogously done in the ion systems for calibration. We will show how we harness these for Hamiltonian spectroscopy and gate calibration.

5.2.1 Calibration of χ and $\bar{n}$.

In this section we describe the single dressed qubit state population swaps we can engineer using individual sidebands. We then show that this type of dressed qubit control allows us to calibrate the parameters most important to our system: χ, $\bar{n}$, and κ. The combination of χ and $\bar{n}$ sets the gate time and is used to calculate the sideband detuning δ necessary for the gate [56]. Whereas κ, the decay rate of photons from the shared resonator, sets a limit on the fidelity of the gate. We discuss this below when we analyse the sources of errors.

We provide an overview of our characterization procedure in Fig. 5.2. The level diagram for the single qubit interaction with sidebands is shown in Fig. 5.2. We apply a state-preparation pulse to initialize the qubit(s) of interest, followed by a Rabi drive around the x axis to bring the qubits close to resonance with the sidebands. The strength of the Rabi drive sets the energy level spacing in Fig. 5.2 and also how close to resonance the dressed qubits are with the sidebands. For instance in Fig. 5.3, we sweep Ω_R with sideband frequencies fixed at $\nu_{res} + \Omega_R - \delta$ and $\nu_{res} - \Omega_R - \delta$ for $\Omega_R = 30$MHz and detuning $\delta = 2.5$ MHz. The qubit is in resonance with the red (blue) sideband when $\Omega_R = 32.5$ (27.5) MHz. When Ω_R is far from resonance, the dynamics is set by a $T_{1,\rho}$ limited decay of the driven qubit [78], where $T_{1,\rho}$ is the dressed frame analog of T_1 qubit relaxation. In contrast, when Ω_R is close to resonance, excitation swaps will occur in addition to the $T_{1,\rho}$ limited decay. We use excitation swaps between the Rabi driven qubit and resonator combined with a Stark shift measurement to calibrate χ and $\bar{n}$. Where $\chi\bar{n} \gg \kappa$, these exchanges occur at a faster time scale than the loss out of the shared resonator. The red (blue) sideband drives population swaps between $|+,0\rangle$ ($|-,0\rangle$) and $|-,1\rangle$ ($|+,1\rangle$), shown in Fig. 5.3, which is a cutaway of the full chevron in Fig. 5.3 for Q_0. The rate of these oscillations is set by $\chi\sqrt{\bar{n}}$, and the observed exponential decay is set by κ, cooling the driven qubit to either $|+\rangle$ or $|-\rangle$ [56]. This contrasts with previous works studying the resonance features in the low χ/κ that show similar cooling effects without population swaps [56, 3, 36]. The frequency of the oscillations in Fig. 5.4 combined with the Stark shift of the qubit - which has a $\chi\bar{n}$ dependence - for different sidebands power allow us to calibrate χ and $\bar{n}$. We explain this in detail in the Supplement, along with other calibration measurements for the Rabi drive power Ω_i and the sideband phase. We note that there is an additional resonance feature at 60 MHz, marked by the black dotted line in Fig. 5.3. These are generated by higher order terms in the Hamiltonian due to the presence of the Rabi drive. In general, as discussed later, $\omega_b - \omega_r$ should be maximized as to push this higher frequency resonance away from the Rabi drive frequencies for the gate.

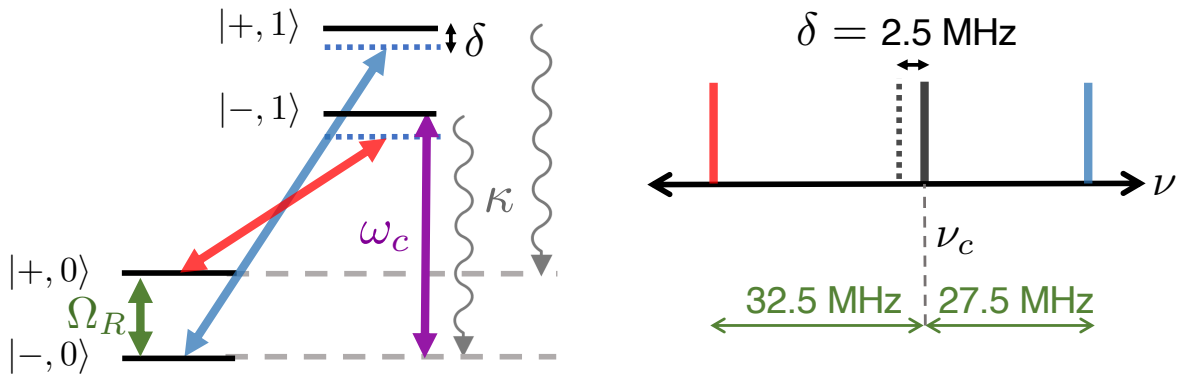

Figure 5.2: Single Rabi driven qubit level diagram with example of sideband positions for calibrations. For a fixed set of sideband frequencies, there will be two resonances depending, one for each sideband at specific Rabi drive frequencies and qubit initializations. The red (blue) sideband will drive transitions between the $|+,0\rangle$ ($|-,0\rangle$) and $|-,1\rangle$ ($|+,1\rangle$) states when the qubit is initialized in the $|+\rangle$ state and the Rabi drive applied to the qubit induces a Rabi frequency that is equal to the frequency separation of the red (blue) sideband from the shared resonator frequency.

We use the combination of two measurements to calibrate the coupling strength of the qubit to sidebands: the spinlocking measurements with sidebands and qubit Stark Shift measurements as a function of sideband power. Both measurements produce oscillations that we fit to extract the frequency. The resulting frequencies for the spinlocking and starkshift measurements exhibit different scalings with respect sidebands power[5]. In the spinlocking sequence, we repeat the measurement for several different red sideband powers, recording the resonant population swaps between $|+0\rangle$ and $|-1\rangle$. We fit the population evolution and extract the frequency of the oscillations, which has a $\chi\sqrt{\bar{n}}$ dependency, and sets the strength of the couplings between the dressed qubit and sidebands, as shown in Fig. 5.4. This $\chi\sqrt{\bar{n}}$ value sets the gate detuning and also gate time. We also measure the Stark shift of the bare qubit frequency as a function of the sideband strength using a Ramsey measurement, for which we expect a $\chi\bar{n}$ dependence. Combining these two relations, we obtain the χ's.

5.2.2 Rabi Drive

We repeat Rabi drive measurements and sweep the amplitude of the qubit pulse. The population swaps between $|g\rangle$ and $|e\rangle$ for each Rabi drive measurement are fit with a sin function and the frequency is extracted. We see an increase in the frequency of qubit population

[5]Kpop fun fact 5/7: Tablo, singer and rapper in Epik High, is a Stanford University graduate. He acquired a bachelors and a masters in English in just 3.5 years. He only started to learn English after moving to Canada from Korea in 2nd grade with his family.

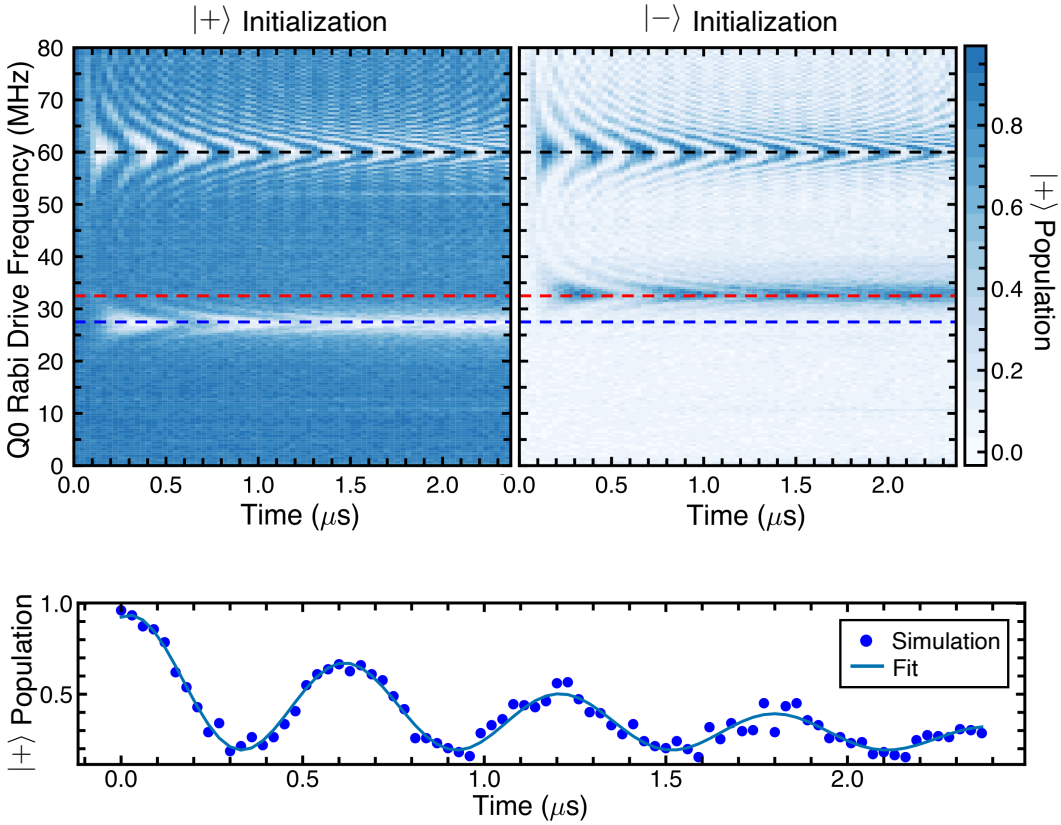

Figure 5.3: Chevron obtained from repeating the spinlocking sequence with sidebands on at various Rabi drive powers. Each horizontal cut is obtained from sweeping the Rabi drive duration in the spinlocking sequence. The bottom is an example of a linecut at the blue dashed line where the qubit is resonant with the blue sideband.

swaps with the power of the drive power. We then plot the frequencies as a function of the amplitude of the pulse, producing a linear dependence (Fig. 5.5a). The linear fit is used to interpolate between the data points to select the amplitudes required for specific Rabi drive frequencies.

5.2.3 Sidebands Phase calibration

The phase difference of the sidebands and their relative phase compared to the qubit drive determines whether the Hamiltonian is an XX or combination of XX and YY with respect to the $|\pm\rangle$ basis. Due to the fact that the Rabi pulse we use for the gate has a cosine edge, there an extra accumulation of phase difference between the sidebands and the qubit drive depending on the length of the cosine edge. We calibrate the sideband phase such that we implement the XX interaction. One can tune the gate angle

$$\varphi_{\Delta}^{\text{eff}} = \varphi_{\Delta}(t_r) - \Omega_R t_r = \varphi_{\Delta}(0) + \Omega_R t_r \tag{5.1}$$

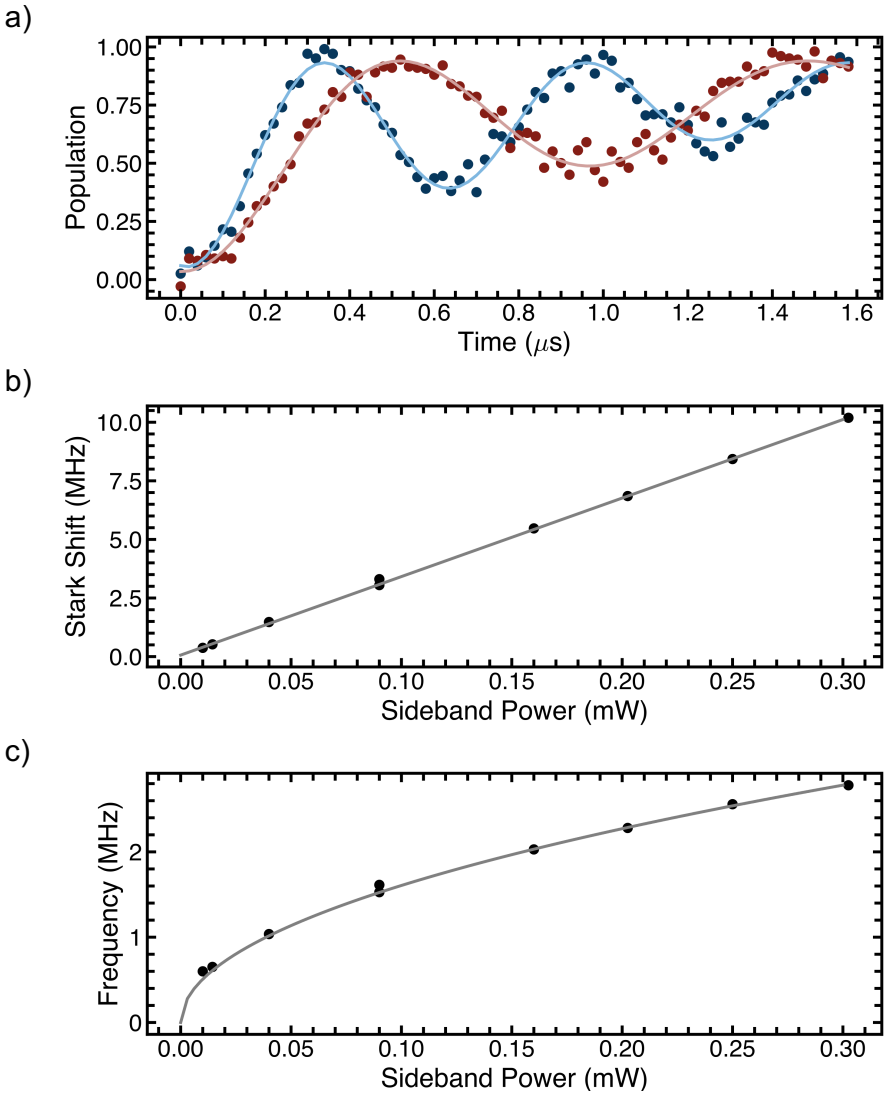

Figure 5.4: Example of χ calibration procedures. a) Examples of dressed qubit resonant population swaps at different sideband powers. The fitted frequency is plotted in c), following a $\sqrt{\bar{n}}$ relation. For the same sideband powers used in a) and c), we measure the Stark shift on qubit, which is plotted in b) and is linearly related to $\bar{n}$.

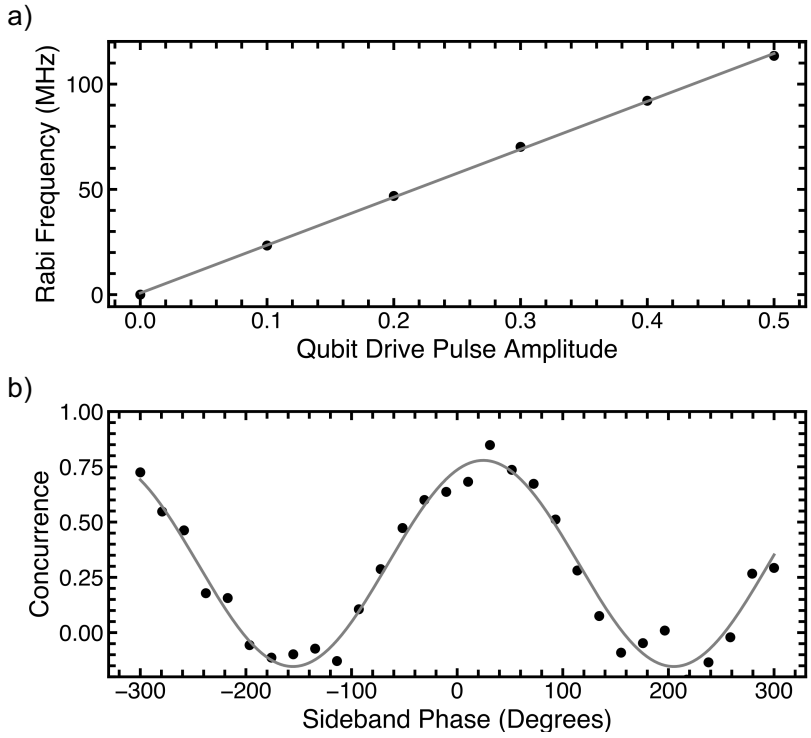

Figure 5.5: a) Fitted Rabi frequencies as a function of the amplitude of the qubit pulse used to do the Rabi drive. b) Sideband phase calibration. We plot the concurrence at the gate time as a function of the sideband phase.

by adjusting the sideband phase difference $\varphi_\Delta(0)$. The angle is calibrated by initializing the qubits in $|00\rangle$, running the gate and looking at the concurrence at the gate time as a function of the sideband phase. Without a cosine edge in the pulse, the minimum of the concurrence should occur in the plot at $\varphi = 0$. However, as in Fig. 5.5b), the minimum is offset. We set the experimental effective angle $\varphi_\Delta^{\text{eff}}$ to be 217 degrees, which is when the concurrence is minimum.

5.3 Conclusion

While these calibration procedures have been adequate to tune up our gate for a first demonstration, we highly recommend refining these techniques for future experiments aiming for higher fidelity. One recommendation is to develop a way to better measure how well the qubit is detangled from the resonator at the time of the gate. Ideally, the calibrations of χ and $\bar{n}$ that feed into the chosen value of δ should account for this, but we don't currently have any method of quantifying this. One possible method is to look at the single and two

qubit purity of the state

$$\gamma = \mathrm{Tr}\{\rho^2\} \tag{5.2}$$

at the time of the gate. For a fully entangled two-qubit state, the individual qubit states should be mixed (purity is low). However, the purity of a properly entangled state should be high. If the two qubit purity is lower than expected, one possible reason is that the qubits are still somewhat entangled with the resonator. In ion experiments, they look for a minimum of the odd qubit states (when starting in an even state) while fine tuning parameters. This allows them to optimize the gate time and detuning to ensure the resonator returns to its original state. Perhaps this is a good diagnosic to go on in the future.

Chapter 6

Gate Characterization

6.1 Introduction to Tomographic Methods

The development of gates and novel state generation methods requires us to have accurate ways of characterizing these methods so that we can both diagnose the types of errors that systems could be experiencing and also as a measure of performance. Performance must be evaluated to determine whether we are satisfying the requirements for quantum error correction (QEC). Several methods of characterization are available:

1. Quantum state tomography
2. Quantum process tomography
3. Randomized benchmarking
4. Gate set tomography

In this thesis, we will make use of state and process tomography in this chapter and randomized benchmarking in the next chapter. We devote the rest of this introductory section to providing some background each of these methods.

Quantum state tomography reconstructs the unknown state ρ by measuring its individual components $\langle\langle k|\rho\rangle\rangle$, where k are often the Pauli basis vectors. However, one does not have to use the Pauli basis vectors and can use some other set of measurement operators E_j as long as they span the Hilbert-Schmidt space with $j = 1, ..., d^2$. In general

$$p_j = \langle\langle E_j|\rho\rangle\rangle = \sum_k \langle\langle E_j|k\rangle\rangle\langle\langle k|\rho\rangle\rangle \tag{6.1}$$

Experimentally, we obtain p_j by measuring E_j a large number of times. Typically for our experiments that's on the order of $N = 1000$. For each measurement, the state is collapsed onto an eigenstate and we obtain a measurement outcome value $m_{ij} = 0$ or 1. We then solve for ρ using the equation above by applying matrix inversion:

$$|\hat{\rho}\rangle\rangle = \langle\langle E_j|k\rangle\rangle^{-1}|m\rangle\rangle \tag{6.2}$$

where $|m\rangle\rangle$ is the vector of measurement averages and $|\hat{\rho}\rangle\rangle$ is the estimated version of the true density matrix.

Quantum process tomography is an extended version of state tomography. We now wish to characterize a gate, G, not just a single state. This is done by measuring the Pauli Transfer Matrix

$$\langle\langle k|G|l\rangle\rangle \tag{6.3}$$

. We now need d^4 measurements of the probabilities

$$p_{ij} = \langle\langle E_j|G|\rho_i\rangle\rangle \tag{6.4}$$
$$= \langle\langle E_j|k\rangle\rangle\langle\langle k|G|l\rangle\rangle\langle\langle l|\rho_i\rangle\rangle \tag{6.5}$$

As before, both E_j and ρ_i are set by the experimenter and so $\langle\langle E_j|k\rangle\rangle$ and $\langle\langle l|\rho_i\rangle\rangle$ are known. We can put these two known quantities together as S in the equation above:

$$\vec{p} = S\vec{r} \tag{6.6}$$

Once again, we invert this equation to solve for $\vec{r} = S^{-1}\vec{p}$ where $\vec{p}$ is a vector of measurements.

In practice, process tomography amounts to performing state tomography on the results of the gate for a whole set of initial states that span the N-qubit subspace.

6.1.1 Interpreting the PTM

There are many ways to express the composition of a gate. Of the many expressions there is the process matrix, the pauli transfer matrix, the Choi-Jamiolkowski, etc, etc. Each form gives a slightly different insight into the same process. For our experiment we chose to use the PTM.

The Pauli Transfer Matrix for a gate is a d^2 by d^2 matrix defined as a gate's action on the Pauli matrices. The values of its entries are restricted to be from -1 to 1. In this representation, density matrices are written as vectors–denoted by $|\rho\rangle\rangle$ and each entry is defined as

$$\rho_i = \langle\langle i|\rho\rangle\rangle = Tr\{P_i\rho\} \tag{6.7}$$

and P_i are the Pauli matrices.

6.2 Characterizing the STAR gate

We use a combination of the tomographic techniques described above to characterize the gate. For the two qubit gate, we use both state and process tomography. For the three and four-qubit gates, we use state tomography.

First, we study the two-qubit population evolution over time to extract the gate time. To perform the gate, we implement the full pulse sequence in Fig. 5.1 on any desired subset of our qubits. We use Q_0 and Q_1, two qubits with the most similar shared resonator dispersive

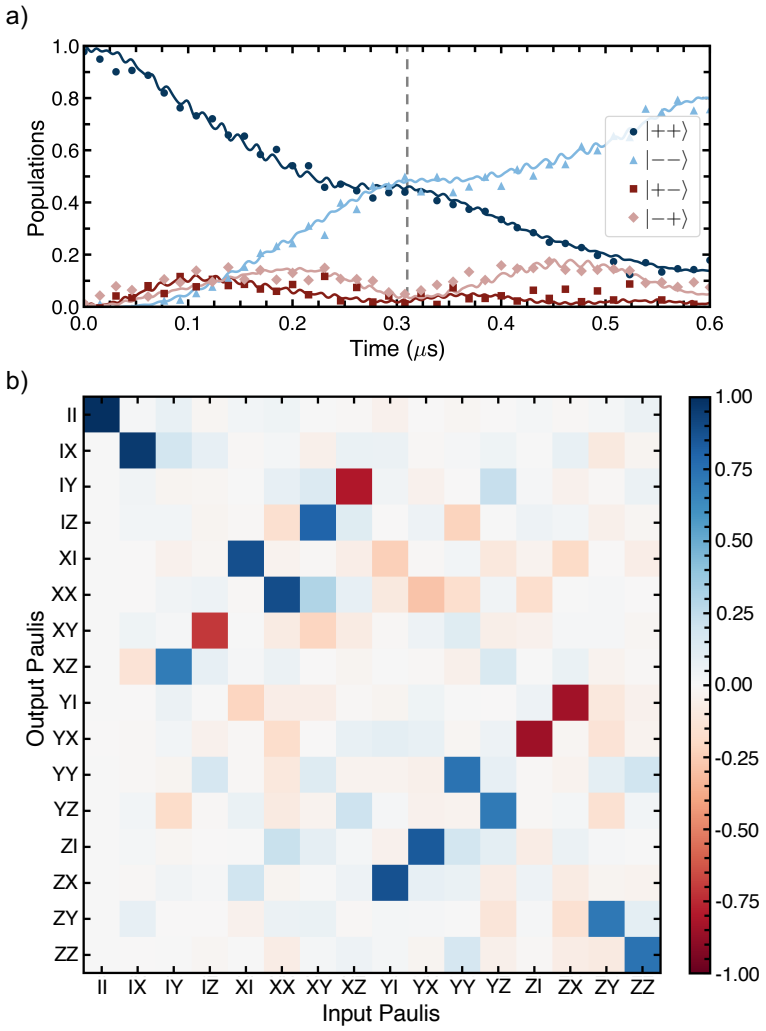

Figure 6.1: **Characterizing two-qubit gate.** a) Gate evolution for two qubits. Experimental data (points) together with simulations of Eq. (3.8) (solid lines) with $\Omega_0/2\pi = 30.55$ MHz, $\Omega_1/2\pi = 29.92$ MHz, $\chi_0/2\pi = 380KHz$, $\chi_1/2\pi = 410$ KHz, $\kappa = 180$ KHz. b) PTM of the experimental data in the $|g/e\rangle$ basis, at $T_{gate} = 310$ ns, obtained measuring 16 different initial states. The process fidelity is 81.6%. The average fidelity to the target bell states 91.8%. The solid lines that fit each population are from simulations.

couplings χ. The results for our choice of $|{+}{+}\rangle$ initial state are summarized in Fig. 6.1. Maximum entanglement occurs when the $|{+}{-}\rangle$ and $|{-}{+}\rangle$ populations reach a minimum and the $|{+}{+}\rangle$ and $|{-}{-}\rangle$ populations cross, as indicated by the vertical dashed line in Fig. 6.1a. The gate time at that point is 310 ns, which refers to the length of the Rabi drive applied to each qubit. We prepare each of the 4 Bell states and perform state tomography at that

gate time. The state fidelity [59], defined as

$$F = \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}, \tag{6.8}$$

ranges from 87% to 95% with an average of 91%. We do not fully understand the spread in the fidelities. While we suspected unequal couplings or inaccurate Rabi drives to be potential reasons, both of these failed to cause large spreads in the fidelity in simulation.

We further characterize our gate using quantum process tomography (QPT) [11, 63], which is achieved by preparing 16 different input states and performing state tomography on each output state. A convenient set of input states that spans the two-qubit subspace is $\{|+\rangle, |-\rangle, |g\rangle, |i-\rangle\} \otimes \{|+\rangle, |-\rangle, |g\rangle, |i-\rangle\}$, where $|i-\rangle = (|+\rangle - i\,|-\rangle)/\sqrt{2}$. The measurement basis is chosen as $\sigma_m \otimes \sigma_n$ where $m, n = 0, x, y, z$, and $\sigma_0 = I$. Since we measure the qubits in the σ_z basis, we apply single qubit tomographic pulses to measure the expectation values of $\sigma_i \otimes \sigma_j$. We present our results as a Pauli Transfer Matrix in Fig. 6.1b, which maps input Pauli state vectors to output Pauli state vectors. We then perform QPT on the two-qubit entangling interaction with process fidelity of 81.6% between qubits 0 and 1. The average state fidelity of all 16 initial states is 91.8%. We note that the qubit Rabi drive adds a global phase in the Rabi driven qubit frame that we unwind using techniques described in the Supplement, resulting in the PTM shown. We also perform QPT on the identity operation with process fidelity 93.5% as a baseline check of state preparation and measurement (SPAM) errors (see Fig. 6.4 in Supplements).

The most important feature of the gate is its scalability, which we demonstrate by generating a three-qubit GHZ state on Q_0, Q_1, Q_2:

$$\Psi_{\mathrm{GHZ}} = \frac{|ggg\rangle + |eee\rangle}{\sqrt{2}}. \tag{6.9}$$

We note here that for odd numbers of qubits, the Hamiltonian allows us to generate entanglement in either the $|\pm\rangle$ basis or directly in the $|g/e\rangle$ basis, depending on the initial state and sideband phase. Given that the two bases have a direct mapping between them, we choose an initial state and sideband phase combination that directly entangles in the $|g/e\rangle$ basis for simplicity. We use a $Y_{\frac{\pi}{2}}$ pulse to prepare each qubit in the $|+\rangle$ state and then apply a σ_x Rabi drive to each qubit for 217 ns. Using individual qubit pulses to do state tomography, we find state fidelity 90.5%, similar to our two qubit fidelities. The density matrix is shown in Fig. 6.2. We note that the three-qubit gate time is faster than the two-qubit gate time because the speed is proportional to the average χ_i and $\sqrt{\bar{n}}$. We used the same sideband power for both gates, but the third added qubit had a much higher coupling than the first two qubits, thus raising the gate speed.

Finally, we attempted 4-qubit GHZ state preparation and achieved state fidelity 66% in 200 ns. The density matrix is shown in the Fig. 6.3. Similar to before, the gate time is even shorter because the final added qubit has the highest χ of all qubits on our processor. The 4-qubit fidelity is mainly limited by the large spread in χ and crosstalk between Q_2 and Q_3. The Q_3 qubit frequency is very close to the $e \longrightarrow f$ transition frequency for Q_2. When both

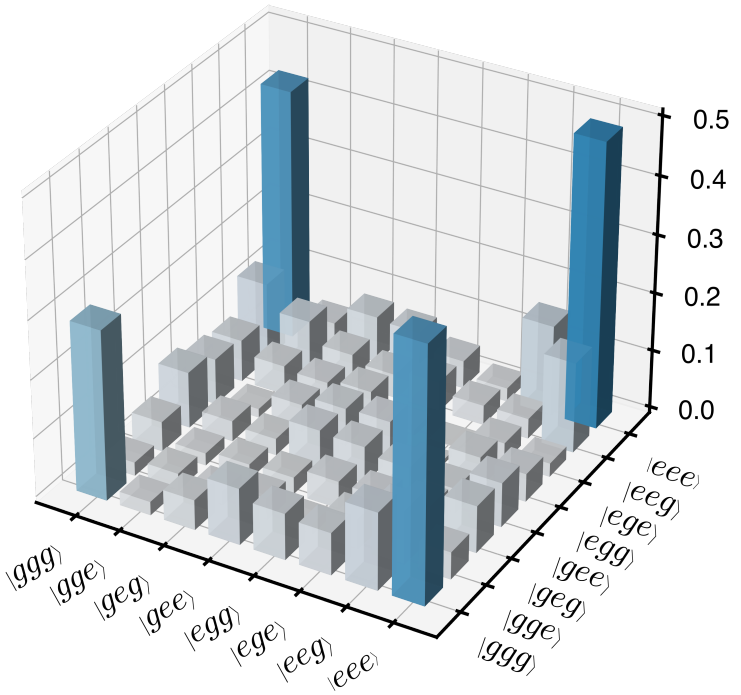

Figure 6.2: **Three-qubit entanglement.** Density matrix for three qubits measured at gate time of 217 ns. The bars represent the amplitude of the density matrix elements. We prepare the qubits in $|+++\rangle$ state and use sideband phase calibrated to be 180 degrees offset from the value used for the two-qubit interactions such that we entangle directly in the bare qubit basis. Note here that the qubit states are labeled with $|g/e\rangle$ rather than $|\pm\rangle$. Entanglement in the $\pm$ basis can be done by preparing in the $|i-\rangle$ state and using the same sideband phase as in the 2 qubit measurements. We chose this for convenience.

qubits are Rabi driven, as during the gate, we see significant f state population for Q_2. To mitigate this, we had to lower Ω_R for each qubit to 20 MHz, which further limits fidelity. We discuss this in our error analysis.

6.3 Unwinding Global Phase from Rabi Drive

The gates applied on undriven transmon qubits commonly take place in the frame rotating at the mode frequencies. In the case of Rabi driven qubits, the system undergoes the MS gate in the frame rotating at the Rabi frequencies. As the latter is not kept constant throughout the entire pulse sequence and as the tomography is done in the original undriven frame, one needs to track and unwind the phase accumulated to characterize the gate. Without

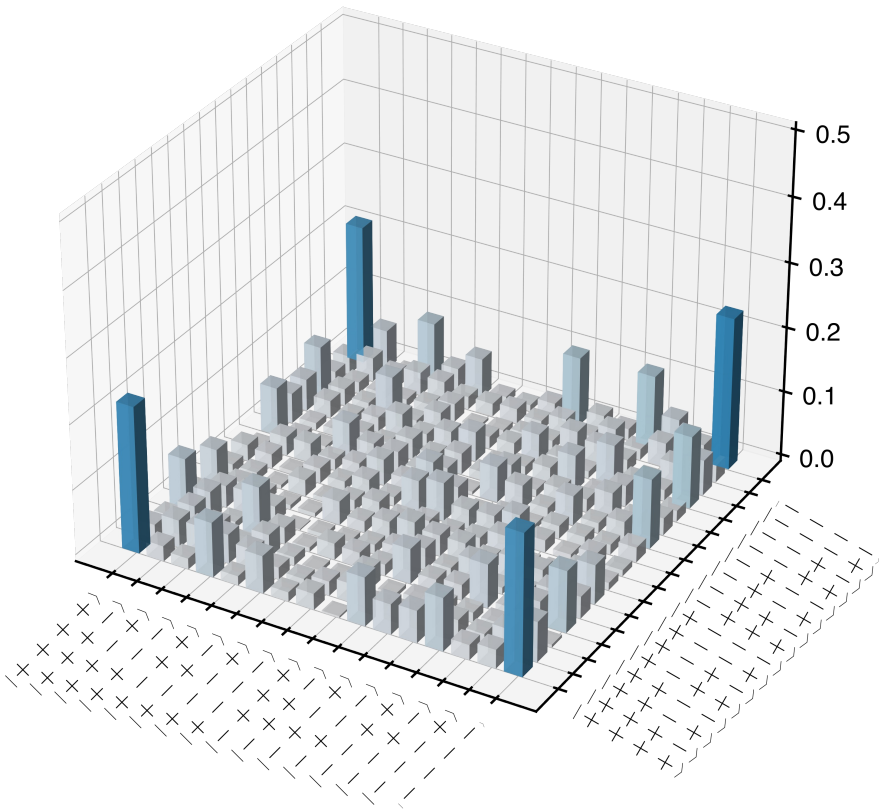

Figure 6.3: **Four-qubit entanglement.** Density matrix for four qubits measured at gate time of 200 ns. The bars represent the amplitude of the density matrix elements. We prepare the qubits in $|++++\rangle$ state and prepare a GHZ state

unwinding the phase, one will notice XI and IX single qubit rotation terms in the PTM that arise from the Rabi drives.

The Rabi drive pulse, of duration t_p, consists a square amplitude pulse[6], of duration t_{sq} during which the entangling operation is happening, sandwiched between two cos edge ramps, of duration t_r. The gate occurs during the square pulse, and one has $t_p = t_{sq} + 2t_r$.

Neglecting decoherence, the propagator corresponding to the full Rabi drive pulse can be expanded as $U_{t_p,0} = U_{\text{down}} U_{\text{gate}} U_{\text{up}}$, with $U_{\text{down}} = U_{t_{sq}+2t_r,t_{sq}+t_r}$, $U_{\text{gate}}(t_{sq}) = U_{t_{sq}+t_r,t_r}$, and $U_{\text{up}} = U_{t_r,0}$. Note that the populations in Fig. 6.1 are plotted as a function of t_{sq}.

As the ramp time t_r is much shorter than the inverse gate rate, the qubit-resonator coupling is (almost) resonant only during the square pulse. Consequently, U_{down} and U_{up}

[6]Kpop fun fact 6/7: Min Yoongi, a member of one of the biggest Kpop groups in history, was as a part-time delivery worker before he debuted. We love humble beginnings.

consists of local operations on each qubit around the X axis (accumulated phase), given by $U_{\text{up/down}} = \exp(-i\Omega_R t_r \mathbf{J}_x/2)$, where $\mathbf{J}_x = (\boldsymbol{\sigma}_{x_1} + \boldsymbol{\sigma}_{x_2})/2$. The factor $1/2$ in $U_{\text{up/down}}$ comes from the integral of the cos edge pulse. Here, we have assumed that the Rabi frequencies of the qubit are the same.

The propagator $U_{\text{gate}}(t_{sq})$ expands as $U_{\text{gate}}(t_{sq}) = U_R(t_{sq})U^R_{\text{gate}}(t_{sq})$, where $U_R(t_{sq}) = \exp(-i\Omega_R t_{sq}\mathbf{J}_x)$ represents the single qubit phase accumulated during the gate, and

$$U_{\text{R,gate}}(t_{sq}) = U^R_{t_{sq}+t_r,t_r} \tag{6.10}$$

is the time evolution operator generated by the Hamiltonian in Eq. (3.2). Since the two sideband tones are detuned from one another by $2\Omega_R$, the effective sideband phase difference φ_Δ of the gate depends on the time t_r at which the gate starts. More precisely, one can write $\varphi_\Delta(t_r) = \varphi_\Delta(0) + 2\Omega t_r$.

The global evolution operator can be cast in the convenient form

$$U_{t_p,0} = U_{\text{down}}U_R(t_{sq})U_{\text{up}}(U^\dagger_{\text{up}}U^{\varphi_\Delta(t_r)}_{\text{R,gate}}(t_{sq})U_{\text{up}}) \tag{6.11}$$

. From the expression of $U_{\text{up/down}}$, we note that this operator applies a rotation on the qubits around the $\mathbf{J}_x$ axis by an angle $\Omega_R t_r$, and its effect on $U^{\varphi_\Delta(t_r)}_{\text{R,gate}}(t_{sq})$ is merely to change the angle of the gate from $\varphi_\Delta(t_r)$ to $\varphi^{\text{eff}}_\Delta = \varphi_\Delta(t_r) - \Omega_R t_r$. Furthermore, the effective winding operator takes the form $U_{\text{R,eff}}(t_{sq}) = U_{\text{down}}U_R(t_{sq})U_{\text{up}} = \exp(-i\Omega_R(t_{sq}+t_r)\mathbf{J}_x)$, leading to $U_{t_p,0} = U_{\text{R,eff}}(t_{sq})U^{\varphi^{\text{eff}}_\Delta}_{\text{R,gate}}(t_{sq})$. By applying the unwinding operator $U^\dagger_{\text{R,eff}}(t_{sq})$ to the the measured density matrix, one recovers the density matrix resulting from the MS gate evolution $U^{\varphi^{\text{eff}}_\Delta}_{\text{R,gate}}(t_{sq})$.

6.4 Discussion and Errors

We now turn to a study of the following sources of infidelity: state preparation and measurement (SPAM) errors, shared resonator decay (κ), spread in qubit-shared resonator couplings ($\Delta\chi$), δ calibration errors, pulse shaping, and the effects of a lower Rabi drive rate (Ω_R). Each error listed here is included in our simulations aside from SPAM errors.

The first major source of infidelity is due to state preparation and measurement (SPAM) error. Process tomography of the identity process resulted in process fidelity of 93.5%. Due to the nonzero κ of the shared resonator, we start the sideband pulse before the state preparation pulse and perform all qubit operations, including the tomography pulses, under the presence of sidebands. However, the presence of sidebands decreases our qubit lifetimes and pulse quality (see Supplement).

Assuming no SPAM errors, simulations based on Eq. (3.8) (details found in the Supplementary Materials) suggest κ as one of the leading sources of error (a complete summary of simulated accumulated error can be found in Table 6.2). The effect of κ is shown in the decay of oscillations in Fig. 5.3d. During the course of the gate, the resonator makes a circle

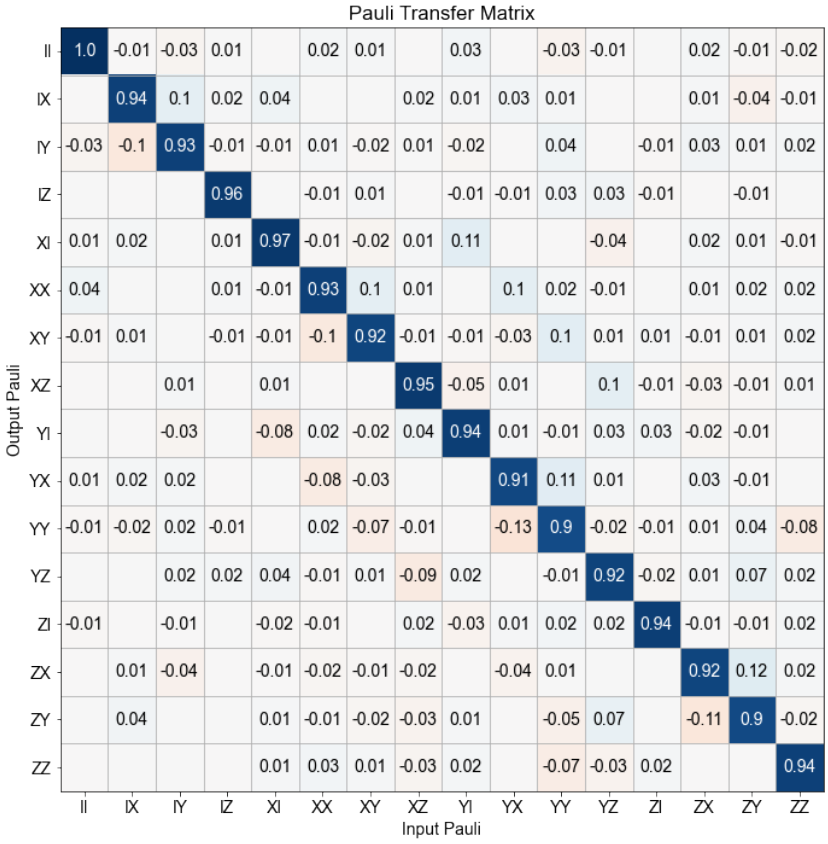

Figure 6.4: Process tomography of the identity operation

in phase space (Fig. 3.2) and changes its photon population by $\Delta n = 0.5$ as the qubits swaps excitations with the resonator. This allows the qubits to each simultaneously entangle and then disentangle with the resonator, leaving the qubits entangled with one other. However, while the resonator is populated, photons may decay from the resonator. We have designed the external κ of the resonators to minimize the loss of photons during the gate. As such, the resonator loss is dominated by internal loss. We highlight this source of error first because we believe it is the most technically challenging to solve. While there have been demonstrations of high quality factor CPW resonators on chips without qubits [1], this has yet to be recreated on chips with qubits present and the additional fabrication steps may still lower the quality factor.

A second source of error that becomes more important with increasing qubit number is the spread of cross-Kerr couplings, $\Delta\chi$, between the qubits. We define $\Delta\chi = \frac{\max(\chi)-\min(\chi)}{\text{avg}(\chi)}$. Our simulations show that for a two qubit gate, the contribution of $\Delta\chi$ to the infidelity is overshadowed by κ. The coupling strengths χ_i sets the rate of the each segment of the two photon transition paths. If different qubits have different coupling strengths, then the various

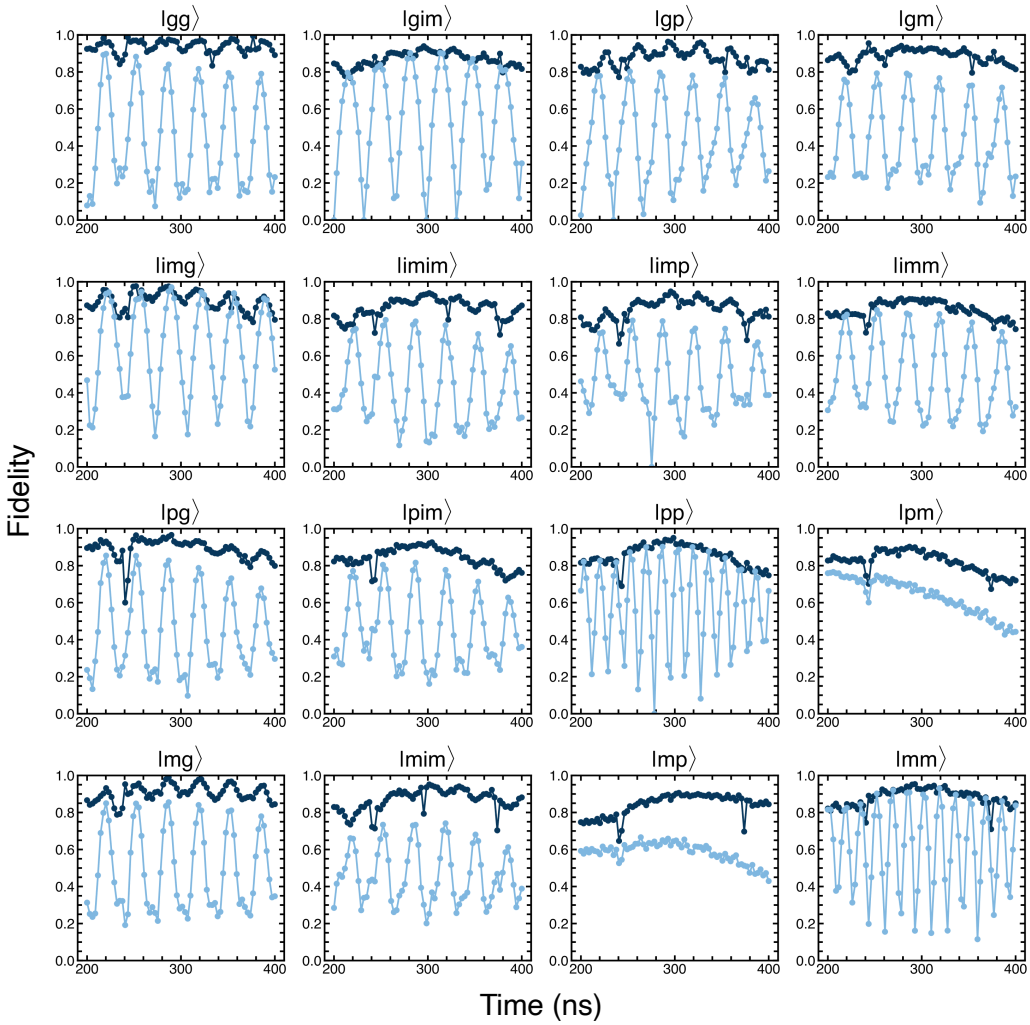

Figure 6.5: Fidelities to the target states for 16 different initial states labelled in the title of each subplot, used to construct the PTM of figure 3. The light blue points are the fidelities without removing the Rabi drive, exhibiting oscillations at 30 MHz for most initial states, at 60 MHz for $|+,+\rangle, |-,-\rangle$ and no oscillations for $|+,-\rangle, |-,+\rangle$. The dark blue points are the processed data which have been unwound. The points at t=300 ns give an average fidelity of 92%.

paths will have different rates of population transfer.Since the gate is the interference of all the various paths, the rates need to line up for to minimize the populations in the side states and achieve the highest fidelity possible. To verify this experimentally, we measure the two-qubit state fidelity between two qubits (Q_0, Q_2) that have the largest $\Delta\chi$ among the three qubits used for our multi-qubit gate. We obtain a state fidelity of 93% for preparing $\frac{|{+}{+}\rangle+|{-}{-}\rangle}{\sqrt{2}}$

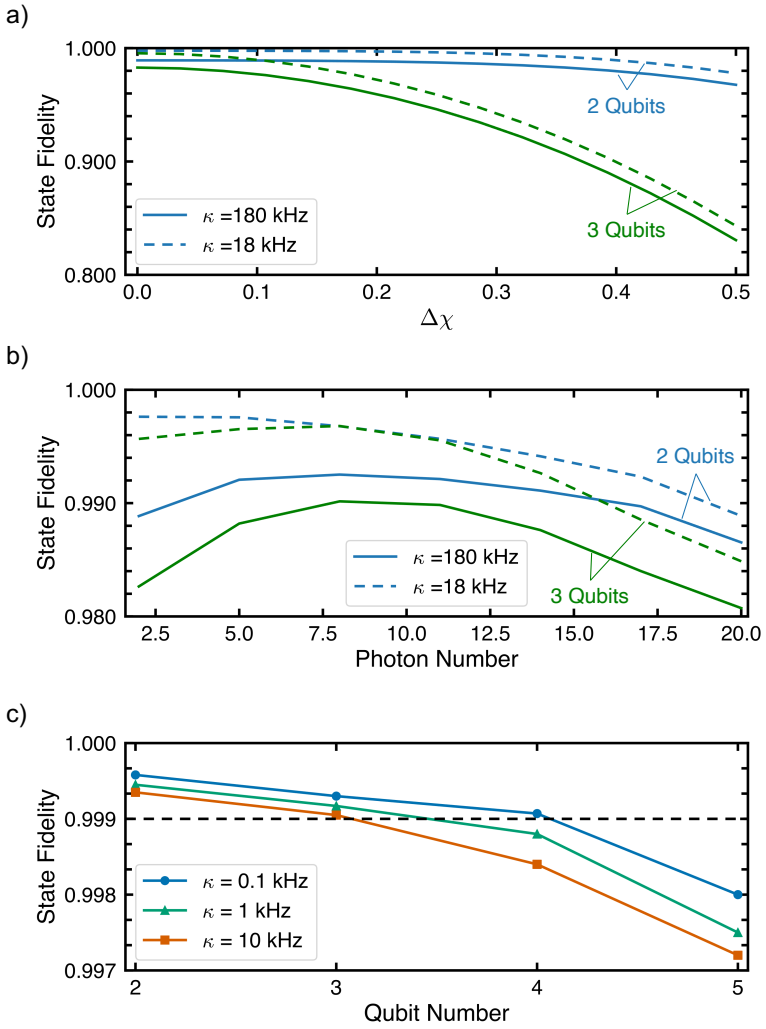

Figure 6.6: **Error analysis using simulations of Eq. (3.8)** a) For a system composed of two (blue) or three (green) qubits, the fidelity of the gate is shown as a function of $\Delta\chi$. The two different styles of lines represent different κ values. We exclude qubit decays in order to isolate the contribution of error due to $\Delta\chi$ and use the average of the $\chi_{\text{av}}/2\pi = 500$ KHz. b) State fidelity as a function of the mean photon number $\bar{n}$. A higher photon number raises the gate speed which helps mitigate the effect of lifetimes but also raises the contribution of counter rotating terms that cause the fidelity to saturate at approximately 10 photons. c) Scaling of entanglement fidelity as function of N, the number of qubits, for the best possible achievable chip. The black dashed line marks the 0.999 fidelity threshold. We set Ω_R =150 MHz and $\chi_k = 1$MHz while varying κ. All three figures take qubit lifetimes to the industry best values.

Error Term	Infidelity (2-qubits)	Infidelity (3-qubits)	Infidelity (4-qubits)
$\Omega = 30$ MHz	0.14%	0.27%	3.7% ($\Omega_R = 20$ MHz)
$\kappa = 180$ KHz	1.7%	1.9%	4.34%
$\Delta\chi$	2.14%	6.8%	26.8%
Qubit lifetimes	2.7%	7.5%	27.2%
δ_{sb} miscalibration by 10%	4%	11%	30.5%

Table 6.1: **Error budget.** Accumulated infidelity with each added term, descending, using parameter values that are currently found on our chip. The last item the table includes an error in δ_{sb} is the only speculative error term, where we quote a 10% miscalibration to give a intuitive sense of scaling, but we are not currently measuring this value.

Qubit	T_1	$T_{2,Ramsey}$	$T_{2,echo}$	$T_{1,\rho}$	$T_{2,\rho}$
0	49.3 ± 14.1	13.2 ± 0.9	16.0 ± 0.9	50-60	28
1	57.0 ± 25.8	11.0 ± 4.7	13.9 ± 2.6	50-60	16
2	48.7 ± 5.9	15.1 ± 0.5	15.7 ± 0.3	60	45
3	23.8 ± 3.2	11.8 ± 3.4	12.9 ± 0.5	40	18

Table 6.2: Lifetimes of the 4 qubits without sidebands, all in μs. The $T_{1,\rho}$ and $T_{2,\rho}$ are shown for a Rabi frequency of 30 MHz.

, which is very similar to maximum fidelities observed between Q_0 and Q_1. However, for a three and four-qubit gate, $\Delta\chi$ carries increasing weight in the infidelity. While the three-qubit gate uses the same qubits as in our two-qubit experiments, we see a drop in state fidelity as $\Delta\chi$ as a larger effect. This takes even greater effect for the four-qubit gate, as Q_2 and Q_3 have almost twice the coupling strength as Q_0 and Q_1. We note that while the fidelity does strongly depend on $\Delta\chi$, the couplings on this chip were anomalies due to design and fabrication errors and standard fabrication techniques should allow for $\Delta\chi$ values below 15%. We emphasize that for smaller values of $\Delta\chi$, κ will become the leading source of error that stands as a technical challenge.

To further describe the effects of κ and $\Delta\chi$, we again performed simulations with Eq. (3.8) as a function of the parameters of interest. We summarize the effects of κ and $\Delta\chi$ in Fig. 6.6. In Fig. 6.6a and b, we use resonator loss values that are similar that on our sample and show fidelity dependence on key system parameters. We take qubit lifetimes to be infinite in order to isolate the effect of each parameter. Fig. 6.6a shows the fidelity as a function of $\Delta\chi$, κ, and the number of qubits. Fig. 6.6b shows expected fidelites as a function of the sideband power. While increasing $\bar{n}$ raises fidelites at first because it decreases the gate time compared to loss rates, increased $\bar{n}$ also widens the $2\Omega_R$ feature shown in Fig. 5.3, increasing the contribution of spurious counter-rotating terms in the Hamiltonian. Our simulations (Fig. 6.6c) suggest

that for leading fabrication techniques that produce resonators with quality factors over 5 million and leading qubit coherence times [1, 62], the gate can be implemented at or above the fault tolerant threshold of 0.999 fidelity for up to four qubits. Furthermore, the STAR gate can be useful for running algorithms on NISQ processors for even higher numbers of qubits.

While κ and $\Delta\chi$ are the main error sources that we are able to characterize experimentally, there are a few additional contributions that will be important to consider to optimize for the best attainable fidelities. First, we are currently not verifying how well the qubits disentangle from the shared resonator at the gate time. Ideally, a proper calibration of the δ–obtained from our measurements of χ and $\bar{n}$–should ensure this, but because we do not measure the state of the resonator, there is the possibility of an inaccurate calibration. As an example, from simulations, a miscalibration of δ by 10% on a two-qubit gate would add an additional 1.3% error. Second is the finite ramp time used in the shape of the Rabi pulse. This ramp is necessary to keep the spectrum of the pulse narrow in frequency space. However, as the drive ramps up to the required Rabi frequency, it crosses a resonance with the sidebands. These are the same resonances used in the spinlocking measurements to calibrate our chip parameters in Fig. 5.4. These spurious interactions can be mitigated using pulse optimization techniques [77, 16].

The final source of error is Rabi drive rate. As shown in the chevron plot in Fig. 5.3, there is a feature at frequency 2Ω marked by the black dotted line, at twice the Rabi drive frequency typically used for the gate at Ω_R. The frequency of the oscillations of the $2\Omega_R$ feature are at $2\times30\text{MHz} = 60$ MHz, and the width of the oscillations is set by $\chi\bar{n}$. For higher sideband powers and lower Rabi drive frequencies, the oscillations generate counter-rotating terms in the Hamiltonian that interfere with gate dynamics. In our numerical simulations, the effects on the gate can be seen in Fig. 6.7 in the Supplement (comparing 30 and 60 MHz Rabi drive gates). The fidelity of the two-qubit gate would benefit from increasing the Rabi drive to over 100 MHz, or equivalently, adding an extra cancellation tone to offset the effects of the spurious feature.

6.5 Conclusions and Outlook

In summary, we demonstrate a scalable maximally entangling gate on an all-to-all connected fixed frequency transmon processor between two, three, and four qubits. The gate is generated through bichromatic microwaves and Rabi drives applied to each participating qubit. The Rabi drive provides the advantages of reducing limitations on qubit-qubit detunings during fabrication, dynamically decoupling from noise, and allowing us to entangle any subset of qubits on the chip. For the three-qubit gate, we are able to choose whether to entangle the qubits in the Rabi dressed basis or the original qubit basis based on the sideband phase and preparation state, without the need for additional qubit pulses to map between bases. The four qubit gate is strongly limited by crosstalk, the spread in qubit-resonator couplings, and loss from the shared resonator. Looking forward, the gate is most limited by photon loss

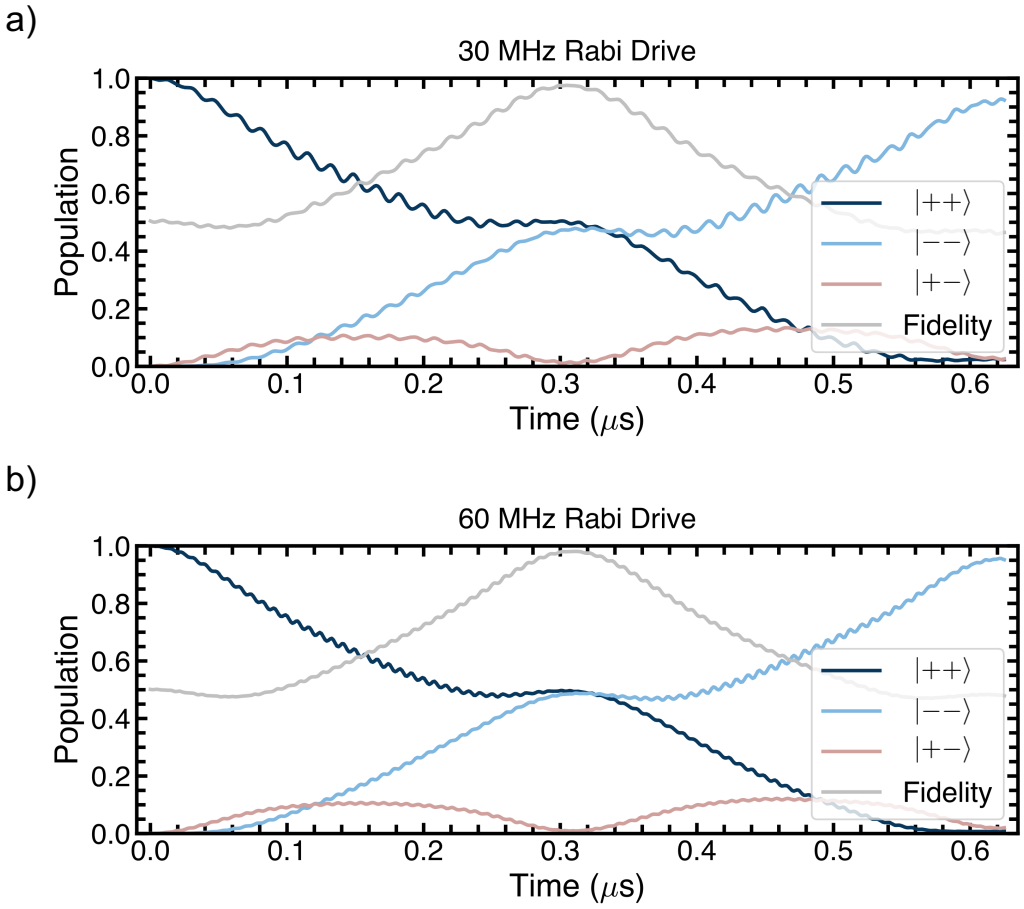

Figure 6.7: Comparing effect of counter rotating terms on the population for 30 MHz vs 60 MHz Rabi drive. We see that the small oscillation are much larger for gate 30 MHz Rabi drives. The counter-rotating terms that form the 2Ω feature in Fig. 5.3 have a larger contribution to the gate dynamics.

from the shared resonator for four and more qubits. Applying state of the art fabrication techniques will yield multi-partite gates that exceed the 0.999 fidelity threshold for up to four qubits. At the same time, for gates with higher numbers of qubits, it is also worth exploring new adaptations that couple make the gate easier to scale. Some of the major challenges of this gate are:

1. The loss out of the shared resonator is a major technical challenge.
2. The sidebands are currently turned on before the gate and the gate cannot be performed until the shared resonator is at a steady state.
3. The single qubit pulses currently always done under the presence of sidebands and are

degraded by the sidebands.

4. Because the sidebands require extra fill time, overlap with the single qubit pulses, and must be carefully phase correlated with the qubit pulses, other validation techniques like Randomized Benchmarking or Cycle Benchmarking are harder to implement. These benchmarking methods require repetitions of the gate and that is hard to do while keeping track of the phases.

5. While the physics of the gate itself lifts restrictions on the qubit frequency due to the use of the Rabi drive, the reality of crosstalk on 2D processors still limits qubit frequencies. For instance, crosstalk still prevents qubit $g \longrightarrow e$ and $e \longrightarrow f$ transitions from being too close to each other. This is not a fundamental limitation of the gate, but rather a technical challenge of the superconducting qubit platform in general.

We propose an alternative all-to-all connected scheme for fixed frequency qubits on parametrically coupled chip in a 3D flip chip architecture that has similar physics. This gate is performed in a far-detuned regime, where the shared resonator is barely populated, thus eliminating the κ loss factor. The qubits will be Rabi driven, like the current scheme, for effective qubit energies that are experimentally controllable. Instead of using sidebands to generate two photon paths that connect energy levels, we now modulate the shared tunable coupler at the difference frequency between the states of interest. Again, the Rabi drive is what allows this parametric gate to be scalable. Many other parametric schemes exist for two qubit gates. However, modulating at the frequency difference between $|ggg\rangle$ and $|eee\rangle$ in the bare frame would be several GHz of modulation, which is harder to do. Using Rabi dressed qubits reduces this frequency difference to something on the order of 60-150 MHz.

The parametric version of this gate has several advantages. Without the sidebands, tuneup will be much easier since we do not need to fill the shared resonator to steady state. Instead we will simply begin modulation of the shared resonator at the same time as the Rabi drive. This adjustment also allows the state preparation and tomography pulses to be done without an additional Stark shift on them due to sidebands. This should make for higher quality single qubit pulses, and thus, improved process tomography results.

We also suggest exploring simultaneous implementations of this gate using subsets of the qubits. We imagine that this could be achieved if the qubits are coupled to different order modes. For instance, if a shared resonator has 4 coupled qubits, one could imagine a design where two of the qubits use a lower frequency mode of the shared resonator, and two qubits couple to the next order mode. One might send two different pairs of red and blue sidebands, each addressing a different resonator mode–and thus also a different pair of qubits–potentially allowing for simultaneous STAR gate implementation.

Another future direction for this project is to implement readout from the shared resonator. This idea is derived from past work from QNL. This would involve setting up new room temperature hardware to do homodyne detection on the reflected signal that comes out of the shared resonator. This could give more insight into noises that are correlated

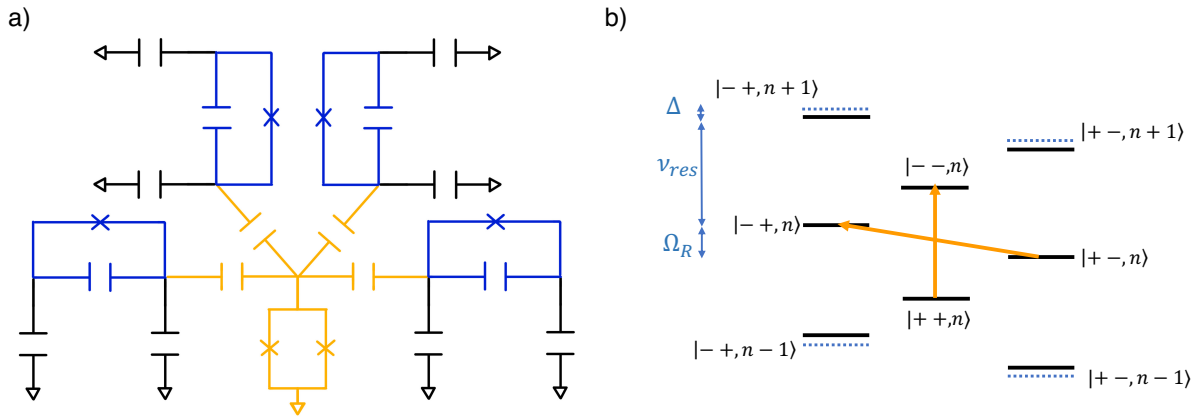

Figure 6.8: a) Example of lumped element representation of 4 fixed frequency qubits coupled to a tunable coupler. b) Level diagram for 2 qubits. The yellow arrows represent the frequency difference that one would modulate the shared coupler at to induce a transition between the indicated states.

between multiple qubits, which is especially relevant on an all-to-all connected device like this one.

A final possible direction that could be possible to explore would be to explore benchmarking and other validation techniques for multiqubit gates like the STAR gate. In Chapter 7, we explore how to better scale randomized benchmarking for two-qubit gates, but benchmarking for gates involving three or more qubits is still an unexplored territory. Typically, to make these algorithms more efficient to run, the error models assume only weight 1 and 2 errors, meaning errors are local on individual qubits or at most a neighboring qubit. In reality, there are often errors such as quasiparticles in coupling resonators that could affect multiple qubits simultaneously and these are best explored on gates that use all-to-all connectivity. Even crosstalk on a multiqubit device could have higher weights.

Chapter 7

Scalable Randomized Benchmarking

Quantum computers suffer from a diverse range of errors that must be quantified if their performance is to be understood and improved. Errors that are localized to single qubits or pairs of qubits can be studied in detail using tomographic techniques [58, 69]. However, many-qubit circuits are often subject to large additional errors, such as crosstalk [23, 70, 67, 66, 65, 49], that are not apparent in isolated one- or two-qubit experiments. There are now techniques for partial tomography on individual many-qubit circuit layers (also called "cycles"), including cycle benchmarking [20] and Pauli noise learning [26, 22, 21]. But quantum computers can typically implement exponentially many different circuit layers, and it is only feasible to characterize a small subset of them.

Randomized benchmarks [66, 65, 49, 18, 19, 44, 43, 37, 10, 14, 9, 27, 44, 43, 10, 14, 9, 27, 29, 30, 12, 28, 55, 67, 8, 4, 41, 15, 48] make it possible to quantify the rate of errors in an average n-qubit layer, by probing a quantum computer's performance on random n-qubit circuits. However, established randomized benchmarks cannot measure the performance of *universal* layer sets in the many-qubit regime, where quantum computational advantage may be possible. Those randomized benchmarks that can be applied to universal layer sets, such as standard randomized benchmarking (RB) [44, 43] and cross-entropy benchmarking (XEB) [41, 8, 4], require classical computations that scale exponentially in the number of qubits (n). XEB requires classical simulation of random circuits that are famously infeasible to simulate for more than approximately 50 qubits [4]. This is because XEB requires estimating the (linear) cross-entropy between each circuit's actual and ideal output distributions. Standard RB of a universal layer set is restricted to even smaller n, because it requires compiling and running Haar random n-qubit unitaries [44]. This compilation requires classical computations that are exponentially expensive in n, and results in circuits containing $O(2^n)$ two-qubit gates [71]. Due to the large overhead, even standard RB on Clifford gates—which has lower overheads and non-exponential scaling—has only been implemented on up to 5 qubits [66, 67, 49].

In this chapter we introduce and demonstrate a simple and scalable technique for RB of a broad class of universal gate sets. The techniques in this chapter builds up previous work showing scalable randomized benchmarking using randomized mirror circuits for Clifford

gates [66]. This work is in collaboration with the Sandia team lead by Jordan Hines and Tim Proctor, who came up with the theory and performed the data analysis. My contribution was mainly in terms of implementation of the method on our Advanced Quantum Testbed device.

The technique uses a novel kind of *randomized mirror circuits*, shown in Fig. 7.2, and advances on a recently introduced method—mirror RB (MRB)—that enables scalable RB of Clifford gates [66]. Please refer to [66] for details on demonstrating that this technique is scalable for Clifford circuits. Here we expand this technique to include universal gate sets. Our randomized mirror circuits use a layer-by-layer inversion structure that enables classically efficient circuit construction and prediction of that circuit's error-free output. Furthermore, randomized mirror circuits do not contain the exponentially large subroutines used in other RB methods, which enables MRB of even hundreds or thousands of qubits. To perform MRB on universal gate sets, we run randomized mirror circuits of varied depths (d) and compute their mean *observed polarization* [66], a quantity that is closely related to success probability. The mean observed polarization versus circuit depth is fit to an exponential decay, as shown in Fig. 7.3b. As in standard RB, the estimated decay rate is then simply rescaled to estimate the average error rate of n-qubit layers. MRB therefore preserves the core strengths and simplicity of standard RB and XEB, while avoiding the classical simulation and compilation roadblocks that have prevented scalable and efficient RB of universal layer sets.

We use MRB to study errors in two different quantum computing systems, based on superconducting qubits. Our experiments are on 4 qubits of the Advanced Quantum Testbed (AQT) [2] and on all of the qubits of a 27-qubit IBM Q quantum computer (`ibmq_montreal`) [32]. In our experiments on AQT we use MRB to quantify and compare the performance of three different layer sets on each subset of n qubits (for $n = 1, 2, 3, 4$), including a layer set containing non-Clifford two-qubit gates. In our experiments on `ibmq_montreal` we show that our method scales to many qubits by performing MRB on a universal gate set on up to 27 qubits.

Multi-qubit MRB enables probing and quantifying crosstalk, which is an important source of error in contemporary many-qubit processors [70, 67, 65, 23] that cannot be quantified by only testing one or two qubits in isolation. We quantify the contribution of crosstalk errors to the observed error rates in our experiments on AQT and further divide the error into contributions from individual layers and gates. The techniques we introduce for these analyses complement other established RB-like methods for estimating the error rates of individual gates—such as interleaved RB [45, 24, 25] and cycle benchmarking [20]. In our experiments on `ibmq_montreal`, we use MRB to study how crosstalk errors vary on this device as n increases, with n ranging from $n = 1$ up to $n = 27$. We find that crosstalk errors dominate in circuit layers on $n \gg 1$ qubits.

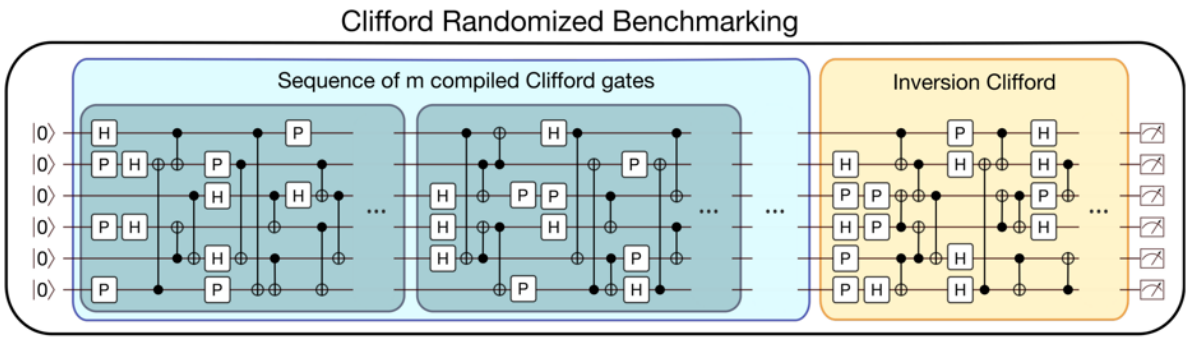

Figure 7.1: Example of a typical randomized benchmarking circuit. Randomized benchmarking is hard to scale past a few qubits because the number of two-qubit gates scales as $O(2^n)$.

7.1 Mirror Circuits

Randomized benchmarking is typically a sequence of Clifford gates applied to an initial state, such that ideally with perfect operation, it amounts to an identity operation. For instance in this image below (Fig. 7.1, we have a sequence of randomly sampled n-qubit Clifford highlighted in blue, and the inverse of everything in yellow. Unfortunately, the compilation step of turning Clifford gates into native gates makes circuits infeasible. The number of 2 qubit gates scales as 2^n.

In contrast, mirror circuits [65, 66, 48] are composed of layers of 1 gates and layers of 2 qubit gates with an inversion structure to transform an Ω-distributed random circuit into a circuit with an efficiently computable outcome, where Ω is the distribution we sample from. We now introduce a natural family of circuits—which we call Ω-distributed random circuits—that we use in our method in order to estimate ϵ_Ω. Ω-distributed random circuits are similar to the circuits used in XEB and other benchmarking routines. They are defined in terms of a customizable gate set $\mathbb{G}$ and sampling distribution Ω over that gate set. This gate set consists of one- and two-qubit gate sets $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$, and Ω is determined by two probability distributions Ω_1 and Ω_2 over n-qubit layer sets $\mathbb{L}(\mathbb{G}_1)$ and $\mathbb{L}(\mathbb{G}_2)$, respectively. An Ω-distributed random circuit with a *benchmark depth* of d is a circuit-valued random variable $\mathsf{C}_d = \mathsf{L}_{2d} \cdots \mathsf{L}_2 \mathsf{L}_1$ where the d odd-indexed layers are Ω_1-distributed and the d even-indexed layers are Ω_2-distributed. These circuits consist of interleaved layers of one and two-qubit gates, so it is useful to define a *composite layer* to be a pair of layers of the form $\mathsf{L} = \mathsf{L}_2\mathsf{L}_1$ where $\mathsf{L}_1 \in \mathbb{L}(\mathbb{G}_1)$ is a layer of one-qubit gates and $\mathsf{L}_2 \in \mathbb{L}(\mathbb{G}_2)$ a layer of one-qubit gates. We denote the set of all composite layers by $\mathbb{L}(\mathbb{G})$. An Ω-distributed random circuit of benchmark depth d then consists of d composite layers that are Ω-distributed over $\mathbb{L}(\mathbb{G})$ with $\Omega(L_2L_1) = \Omega_1(L_1)\Omega_2(L_2)$.

The detailed protocol is enumerated below. We construct a specific randomized mirror circuit on n qubits with benchmark depth d via the three-step procedure shown in Fig. 7.2.

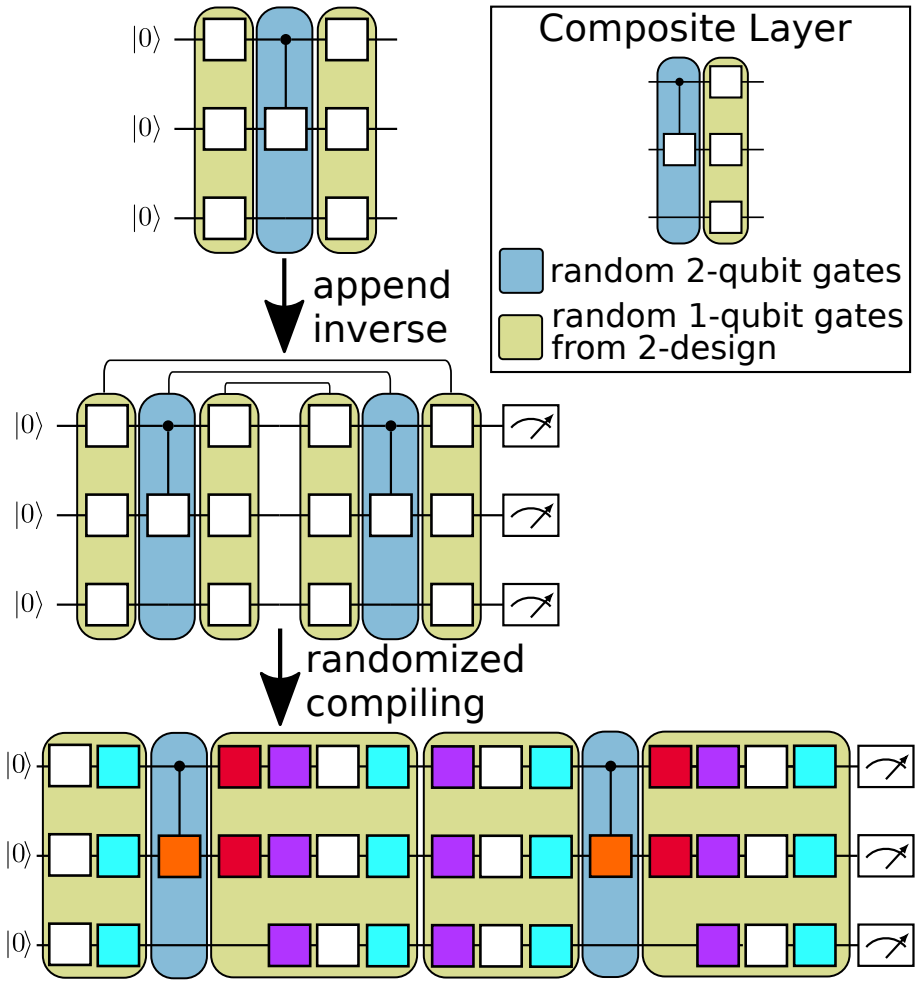

Figure 7.2: How to build up a randomized mirror circuit. We alternate layers of two qubit and single qubit gates. We then find the inverse of each layer backwards and append it to the circuit. Similar to randomized benchmarking, we want the qubit to return to the ground state if there were no errors. However, we do not want to undo all of the errors as well. We add a generalized version of an randomized compiling step since we are constructing this circuit for both Clifford and non-Clifford 2 qubit gates. In the case of Clifford gates, this step looks identical to RC

This procedure consists of first sampling a circuit C_1 consisting of an Ω-distributed random circuit preceded by an initial layer of random single-qubit gates that randomizes the state input into the circuit (enabling estimation of the circuit's fidelity using the method of Ref. [64]). We then append the inverse of C_1 to obtain C_2, a simple form of mirror (or motion-reversal) circuit whose error-free output is definite and easy to predict. Finally, C_2 is randomly compiled, to prevent systematic coherent addition or cancellation of errors between the Ω-distributed random circuit and its inverse—which is essential for reliable estimation of ϵ_Ω. The exact procedure is as follows:

1. (Sample a random circuit) Construct a circuit $C_1 = L_{d/2} L_{\theta_{d/2}} \cdots L_1 L_{\theta_1} L_0$ consisting of:

   (a) A layer L_0 sampled from Ω_1, which consists of a single-qubit gate on each qubit.

   (b) $d/2$ composite layers $L_i L_{\theta_i}$, where L_i is sampled from Ω_1, and L_{θ_i} is sampled from Ω_2.

2. (Construct simple mirror circuit) Add to the circuit C_1 the layers in step 1 in reverse order, with each layer replaced with its inverse. The result is a circuit

$$C_2 = L_0^{-1} L_{\theta_1}^{-1} L_1^{-1} \cdots L_{\theta_{d/2}}^{-1} L_{d/2}^{-1} L_{d/2} L_{\theta_{d/2}} \cdots L_1 L_{\theta_1} L_0, \tag{7.1}$$

   such that $U(C_2) = \mathbb{I}$.

3. (Randomized compiling) Construct a new circuit M by starting with C_2 and replacing layers using the following randomized compilation procedure, which reduces to standard Pauli frame randomization [75] when the two-qubit gates are all Clifford gates. The reason we do this step is that, like in RB, we want the qubits to return back to their initial state if there were no errors. At the same time, simply implementing the inverse in the second half because that could echo out coherent errors, causing us not to detect them. Thus we 'dress' each layer with additional correction gates that overall still allow us to return the qubits back to the initial state given NO errors. But if there are errors, these additional gates prevent the errors from being echoed out. For instance following our single qubit gates, we add single qubit Pauli's in bright blue (Fig. 7.2), we adjust the rotation angles of 2 the two qubit gates (orange), add single qubit Paulis on the qubits used for the 2 qubit gates after the 2 qubit gate (red) and also single qubit Paulis before 1 qubit layers. Those are in purple. For those that are familiar with randomized compiling, when the two qubit layers are just 2 qubit Cliffords, then this is identical to randomized compiling. However, since we also wish to demonstrate our method for non-Clifford 2 qubit gates, the red corrections are there for non-Clifford gates.

   To specify our procedure, we first write C_2 [Eq. (7.1)] in the form

$$C_2 = L_{d+1} L_{\theta_{d+1}} L_d \cdots L_{\theta_{d/2+2}} L_{d/2+1} L_{\theta_{d/2+1}} L_{d/2} L_{\theta_{d/2}} \cdots L_1 L_{\theta_1} L_0,$$

   where $L_{\theta_{d/2+1}}$ is a dummy (empty) 2-qubit gate layer, so that C_2 consists of alternating layers of one- and two-qubit gates. Then:

(a) For each single-qubit gate layer L_i in C_2, sample a uniformly random layer of Pauli gates P_i, that in the following procedure is inserted after and then compiled into L_i.

(b) Replace each two-qubit gate layer L_{θ_i} in C_2 with a new two-qubit gate layer $\mathcal{T}(L_{\theta_i}, P_{i-1})$ that is constructed as follows: For each gate CP_θ in L_{θ_i} with control qubit q_j and target qubit q_k, consider the instructions in P_{i-1} acting on q_j and q_k, denoted by $P_{i-1,j}$ and $P_{i-1,k}$, respectively. If $U(P_{i-1,j}) = I$ or Z, then add CP_ϕ acting on (q_j, q_k) to $\mathcal{T}(L_{\theta_i}, P_{i-1})$ where $\phi = \theta$ if $[U(P), U(P_{i-1,k})] = 0$ and $\phi = -\theta$ otherwise. If $U(P_{i-1,j}) = X$ or Y, then add CP_ϕ acting on (q_j, q_k) to $\mathcal{T}(L_{\theta_i}, P_{i-1})$ where $\phi = -\theta$ if $[U(P), U(P_{i-1,k})] = 0$ and $\phi = \theta$ otherwise.

(c) For each single-qubit gate layer L_i in C_2 with $i > 0$, we define a layer of single-qubit gates P_i^c that undoes the effect of adding P_{i-1} into the circuit—meaning a layer such that $U(P_i^c \mathcal{T}(L_{\theta_i}, P_{i-1}) P_{i-1}) = U(L_{\theta_i})$. Because $\mathbb{G}_2$ is restricted to only controlled Pauli-axis rotations, the correction takes the form $U(P_i^c) = U(P_{i-1} P_{\tilde{\theta}_i})$, where $P_{\tilde{\theta}_i}$ consists of single-qubit Pauli axis rotations. If L_i is not immediately preceded by a two-qubit gate layer, then $P_{\tilde{\theta}_i} = \mathbb{I}$. Otherwise,

$$U(P_{\tilde{\theta}_i}) = U\big(P_{i-1} L_{\theta_i} P_{i-1} \mathcal{T}(L_{\theta_i}, P_{i-1})^{-1}\big). \tag{7.2}$$

(d) Replace each single-qubit gate layer L_i in C_2 with a recompiled layer $\mathcal{R}(P_i L_i P_{i-1}^c)$, defined by

$$U\big(\mathcal{R}(P_i L_i P_{i-1}^c)\big) = U(P_i L_i P_{i-1}^c). \tag{7.3}$$

This randomized compilation step transforms the layer pair $L_i L_{\theta_i}$ into

$$\mathcal{R}(P_i L_i P_{i-1}^c) \mathcal{T}(L_{\theta_i}, P_{i-1}), \tag{7.4}$$

where

$$U\big(\mathcal{R}(P_i L_i P_{i-1}^c) \mathcal{T}(L_{\theta_i}, P_{i-1})\big) = U(P_i L_i L_{\theta_i} P_{i-1}). \tag{7.5}$$

The final circuit produced by this procedure (M) has the property that $U(M) = U(P_{d+1})$, i.e., its overall action is an n-qubit Pauli operator. So, if run perfectly, M returns a single bit string (s_M) that is determined during circuit construction with no additional computation needed.

The final depth-d randomized mirror circuit has the form

$$M = \mathcal{R}(P_{d+1} L_0^{-1} P_d^c)\, \tilde{M}\, \mathcal{R}(P_0 L_0), \tag{7.6}$$

where

$$\tilde{M} = \mathcal{T}(L_{\theta_1}^{-1}, P_d) \mathcal{R}(P_d L_1^{-1} P_{d-1}^c) \cdots \mathcal{R}(P_{d/2+1} L_{d/2}^{-1} P_{d/2}) \\ \mathcal{R}(P_{d/2} L_{d/2} P_{d/2-1}^c) \cdots \mathcal{R}(P_2 L_1 P_1^c) \mathcal{T}(L_{\theta_1}, P_0),$$

is the circuit obtained after applying randomized compilation to the $^{d}/_{2}$ composite layers sampled from Ω and their inverses.

To run these circuits, we chose a range of circuit depths which sets how many layers we have. Then for each depth, we randomly sample our set of available gates K different times. Then we run each circuit many times, typically for us, 1000 times. From each run of these circuits, we get back a string of 0s and 1s representing the measured state of the qubit. Ideally we'd get all 0's back, meaning the qubits returned to their initial state and there were no errors, but usually that's not true. We compare this string to the perfect 'target' string to calculate an effective polarization, which is closely related to the success probability. This step is much like what is done for RB.

7.2 Effective Polarization

We note that the effective polarization here because it's not calculated the way it's normally done for RB. As I showed earlier, our method is different from RB, because we don't calculate the inverse of a bunch of n-qubit unitaries and instead do it layer by layer. However, this means that the effect of twirling our errors in RB is somewhat lost. This okay in RB because errors in Rb are spread amongst the qubits in various directions. But this twirling effect is limited in our randomized mirror circuits. To accommodate for this, we calculate the polarization with an extra correction on top to account for the potential errors we might not be detecting.

We use the same analysis technique as MRB of *Clifford* gate sets [66]. In particular, for each n-qubit circuit C that we run, we estimate its *observed polarization* [66]

$$S = \frac{4^n}{4^n - 1}\left[\sum_{k=0}^{n}\left(-\frac{1}{2}\right)^k h_k\right] - \frac{1}{4^n - 1}, \tag{7.7}$$

where h_k is the probability that the circuit outputs a bit string with Hamming distance k from its target bit string (s_C). As shown in Ref. [66] and discussed further below, the simple additional analysis in computing S simulates an n-qubit 2-design twirl using only local state preparation and measurement. This formula is rather complicated and we leave the details and derivations to [31], but the takeaway here is that the sum, from 0 to n, where n in the maximum number of qubits we have is a sum over the number of possible errors we might be missing. For example for 4 qubits, we might have missed 0 errors, or we might have missed 1, 2, 3, or 4. The weighted sum adds together these possibilities.

A specific MRB experiment is defined by a gate set $\mathbb{G}$, a sampling distribution Ω, and the usual RB sampling parameters (a set of benchmark depths d, the number of circuits K sampled per depth, and the number of times N each circuit is run). Our protocol is the following:

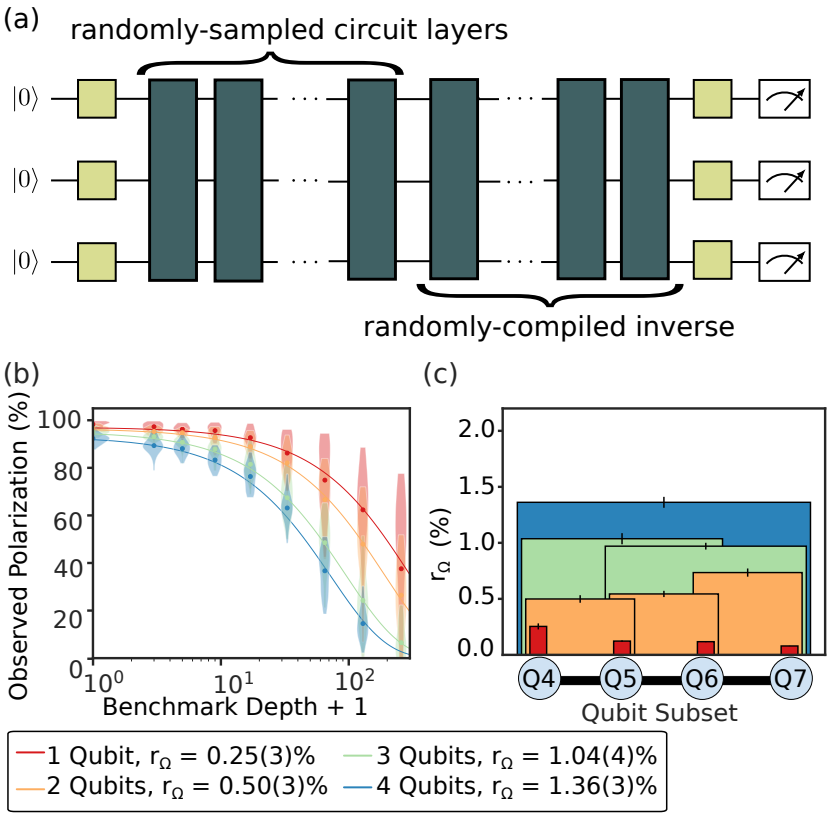

Figure 7.3: Scalable randomized benchmarking of universal gate sets. (a) Randomized mirror circuits combine a simple reflection structure with randomized compiling to enable scalable and robust RB of universal gate sets. (b) Data and fits to an exponential obtained by using our method—MRB of universal gate sets—to benchmark a universal gate set on $n = 1, 2, 3, 4$ qubits of the Advanced Quantum Testbed, and the average error rates of n-qubit layers (r_Ω, where Ω is the layer sampling distribution) extracted from these decays. (c) We benchmarked each connected set of n qubits for $n = 1, 2, 3, 4$, enabling us to map out the average layer error rate (r_Ω) for each subset of qubits. The values of the color bars in figure c) correspond to error rates that are extracted from fitting to the exponential decays shown in figure b).

1. For a range of integers $d \geq 0$, sample K randomized mirror circuits that have a benchmark depth of d, using the sampling distribution Ω, and run each one $N \geq 1$ times.

2. Estimate each circuit's observed polarization S.

3. Fit $\bar{S}_d$, the mean of S at benchmark depth d, to

$$\bar{S}_d = Ap^d, \tag{7.8}$$

where A and p are fit parameters, and then compute

$$r_\Omega = (4^n - 1)(1 - p)/4^n \tag{7.9}$$

which is the average error per layer.

In Fig. 7.3b, we show an example of the exponential decay that results from measuring the effective polarization at various circuit depths. For each circuit depth we generate k = 30 randomly sampled circuits and we measure each circuit 1000 times so that gives us 30,000 bit strings back that we calculate the effective polarization for. We plot the effective polarization we calculate from every run of the circuits. Each violin plot represents the spread of the results that come back from the 1000 runs of each circuit for a given depth. The exponential fit is done to the average point in this spread. From the fits we extract an r_Ω which is an averaged error per layer.

7.3 Simulations

We have made modifications to traditional RB in both the circuit structure and polarization calculation in order to make RB more scalable. It is important to verify that these changes still allow us to capture the error accurately. In particular, the theory for MRB suggests that MRB is particularly robust when the two-qubit gates are Clifford gates and when all errors are stochastic Pauli errors.

We use numerical simulations to investigate the robustness of MRB, studying whether the MRB error rate (r_Ω) closely approximates the error rate of Ω-distributed layers (ϵ_Ω). We choose an artificial amount of error for each gate during the simulations. We use the simulated bit string to calculate the effective polarization. The effective polarization at various depths is fitted to an exponential decay, allowing us to extract the "measured" error, r_Ω, which we compare to the amount of injected error. We simulated MRB for n-qubit layer sets constructed from the gate set $\mathbb{G}_1 = \mathbb{SU}(2)$ and $\mathbb{G}_2 = \{\mathsf{cs}, \mathsf{cs}^\dagger\}$ and $n = 1, 2, 4$, with all-to-all connectivity. We used a sampling distribution Ω_2 for which the two-qubit gate density is $\xi = 1/2$. In these simulations (and our experiments) each single-qubit gate is decomposed into the following sequence of $\mathsf{x}_{\pi/2}$ and z_θ gates:

$$\mathsf{u}(\theta, \phi, \lambda) = \mathsf{z}_{-\phi-\pi/2}\, \mathsf{x}_{\pi/2}\, \mathsf{z}_{\pi-2\theta}\, \mathsf{x}_{\pi/2}\, \mathsf{z}_{-\lambda+\pi/2}. \tag{7.10}$$

Here $\mathsf{x}_{\pi/2}$ is a $\pi/2$ rotation around the X axis and z_θ is a rotation around the Z axis by $\theta \in [0, 2\pi)$. Note that even when a shorter sequence of gates can implement the required unitary (e.g., $\mathsf{u}(0,0,0)$ implements the identity so it could be implemented with no gates) we always use this sequence of five gates. Therefore, the only difference between any two single-qubit gates is the angles of the z_θ gates.

We simulated three different families of error model: stochastic Pauli errors, Hamiltonian errors, and stochastic and Hamiltonian errors. Hamiltonian errors refer to coherent errors, such as over-rotations. These error models are specified using the error generator framework of Ref. [7], and they consist of gate-dependent errors specified by randomly sampling error rates for each type of error and each gate. We simulated error models that are crosstalk free (note that our theory encompasses crosstalk errors) so each error model is specified by the rates of each type of local error on each gate. In particular, for an m-qubit gate we randomly sample $4^m - 1$ stochastic error generators, or $4^m - 1$ Hamiltonian error generators, or both, depending on the error model family. We sampled the error rates so that the infidelity of each two-qubit gate was approximately q, and the infidelity of each one-qubit gate was approximately $0.1q$, where q is a parameter swept over a range of values.

Figure 7.4 shows the results of these simulations. It compares the true average layer error rate per qubit

$$\epsilon_{\Omega,\,\mathrm{perQ}} = 1 - (1 - \epsilon_\Omega)^{1/n} \approx \epsilon_\Omega/n \tag{7.11}$$

to the observed MRB error rate per qubit

$$r_{\Omega,\,\mathrm{perQ}} = 1 - (1 - r_\Omega)^{1/n} \approx r_\Omega/n \tag{7.12}$$

in each simulation, separated into the three families of error model (1σ error bars are shown, computed using a standard bootstrap).

We see a 1:1 correspondence, meaning that whatever error we put in, we are indeed detecting. This is an intuitive check that our method is working. Here, stochastic errors refer to incoherent errors, relating to things like t2 decays. Whereas Hamiltonian errors are coherent errors, like over rotations. We recognize that realistic devices will have a combination of stochastic and Hamiltonian errors the last simulation gives equal weight to each type of error in the simulation. Here plot the same data but with the y axis normalized and we actually see that 1 qubit simulations show a LOT more error spread than simulations for higher numbers of qubits. This is because often for 2 and more qubits, we have 2 qubit gates that spread the error out among the qubits. Creating a kind of averaging effect. For single qubits subject to over-rotation errors, depending on what exactly the circuit you have is, it could show up as a huge error or be totally canceled out. So the spread is very large. We will see this in our data later too.

7.4 Experimental Implementation

In our experiments we investigated three different choices for $(\mathbb{G}_1, \mathbb{G}_2)$: $(\mathbb{SU}(2), \{\backslash, \mathsf{cs}^\dagger\})$, $(\mathbb{SU}(2), \{\mathsf{cz}\})$, and $(\mathbb{C}_1, \{\mathsf{cz}\})$, where $\mathbb{C}_1$ is the set of all 24 single-qubit Clifford gates. Thus

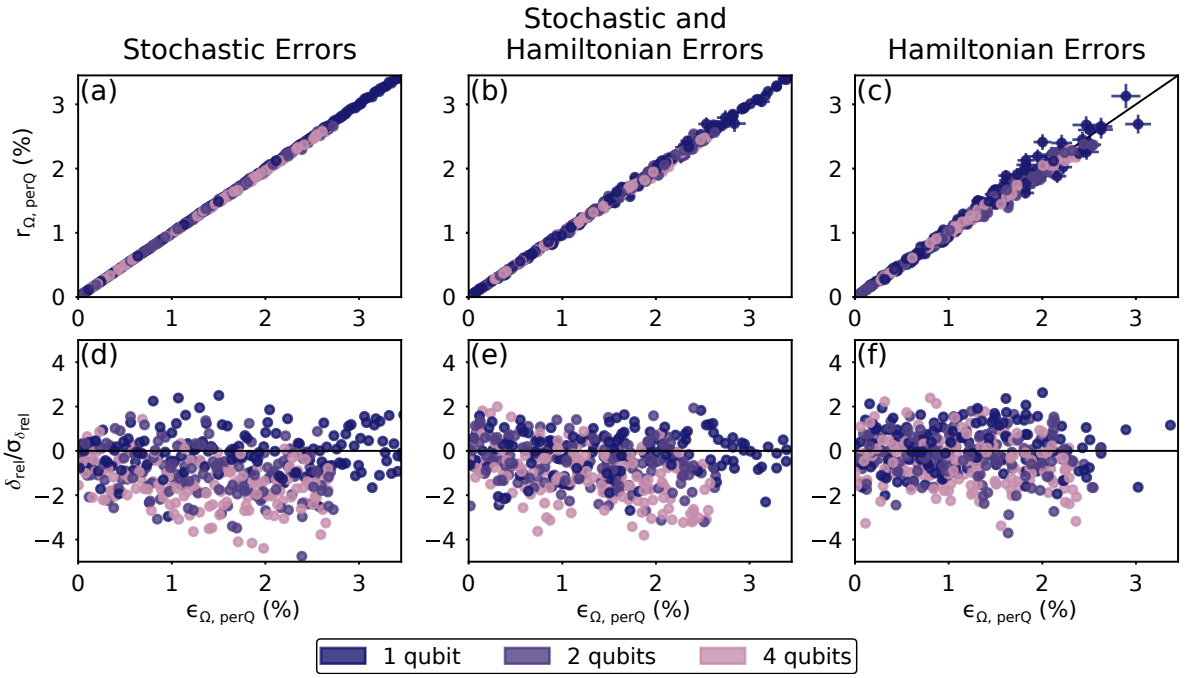

Figure 7.4: **Investigating the reliability of MRB using simulations.** We simulated MRB on n all-to-all-connected qubits for $n = 1, 2, 4$ on the gate set $(\mathbb{G}_1, \mathbb{G}_2) = (\mathbb{SU}(2), \{\mathsf{cs}, \mathsf{cs}^\dagger\})$ with randomly-sampled gate-dependent errors. From left to right, the columns show results from simulations with crosstalk-free error models consisting of only stochastic errors (a,c), a combination of stochastic and Hamiltonian errors (b,d), and only Hamiltonian errors (e,f). (a-c): The MRB error rate per qubit $[r_{\Omega,\,\mathrm{perQ}} = 1-(1-r_\Omega)^{1/n}]$ versus the average composite layer error rate per qubit $[\epsilon_{\Omega,\,\mathrm{perQ}} = 1-(1-\epsilon_\Omega)^{1/n}]$ for each randomly sampled error model. The MRB error rate r_Ω closely approximates ϵ_Ω, and the agreement is closest under purely stochastic errors. (d-f): The relative error $\delta_{\mathrm{rel}} = {}^{(r_{\Omega,\,\mathrm{perQ}}-\epsilon_{\Omega,\,\mathrm{perQ}})}/_{\epsilon_{\Omega,\,\mathrm{perQ}}}$ divided by its uncertainty $\sigma_{\delta_{\mathrm{rel}}}$ for each randomly sampled error model ($\sigma_{\delta_{\mathrm{rel}}}$ is calculated via a standard non-parametric bootstrap). The MRB error rate r_Ω is biased towards very slightly underestimating ϵ_Ω for $n > 2$ qubits, which is expected from our theory (see main text).

we have a non-Clifford only, mixed, and Clifford only gate set, respectively.

MRB enables benchmarking each layer set on any connected set of qubits, and the error rates on subsets of a device can be used to learn about the location and type of errors. We benchmarked n-qubit layer sets for every possible connected set $\mathbb{Q} \subseteq \{\mathtt{Q4}, \mathtt{Q5}, \mathtt{Q6}, \mathtt{Q7}\}$ of n qubits with $n = 1, 2, 3, 4$, resulting in 10 different qubit subsets. Independently benchmarking every connected subset of qubits allows us to study the spatial variation in gate performance in detail and determine the size of crosstalk error on circuits with 3 and 4 qubits (see Section **??**). For each RB experiment, we sampled $K = 30$ circuits at a set of exponentially-spaced benchmarking depths ($d = 0, 2, 4, 8 \ldots$).

For each of the three gate sets $(\mathbb{G}_1, \mathbb{G}_2)$, and each qubit subset $\mathbb{Q}$, we ran experiments with a two-qubit gate density of $\xi = 1/2$. To investigate the effect of varying ξ, we also ran experiments with $\xi = 1/8$ for one of the gate sets—$(\mathbb{SU}(2), \{\backslash, \mathsf{cs}^\dagger\})$—and every $\mathbb{Q}$. For each qubit subset we therefore ran 4 MRB experiments, defined by [1]:

1. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{\backslash, \mathsf{cs}^\dagger\}$, and $\xi = 1/8$.
2. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{\backslash, \mathsf{cs}^\dagger\}$, and $\xi = 1/2$.
3. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{\mathsf{cz}\}$, and $\xi = 1/2$.
4. $\mathbb{G}_1 = \mathbb{C}_1$, $\mathbb{G}_2 = \{\mathsf{cz}\}$, and $\xi = 1/2$.

We implement these circuits on the AQT 8 qubit device. devices shown in Fig. 7.5. We use qubits 4 thru 7. These are fixed frequency qubits highlighted in green on the image. They are linearly connected with coupling resonators highlighted in pink. Each qubit has its own drive line in blue and readout resonator in red. The line down the middle in bright blue is a multiplexed readout bus. Our results are shown in Fig. 7.6. Each row corresponds to different gate set. We have the universal set with with the CS/CSD gates on top, the universal set with Clifford CZs, and the Clifford only group on the bottom. The left column shows the same exponential fits of the effective polarizations, which are used to extract R_Ω, the average error per layer. The middle column shows the errors from all 10 combinations of qubit subsets. And the height of each bar indicates r_Ω.

By comparing (e) and (f), we find that the average error rate of a layer set is approximately independent of whether single-qubit gates are sampled from $\mathbb{SU}(2)$ or from $\mathbb{C}_1$ (the single-qubit Clifford group)—that is, $r(\mathbb{SU}(2), \{\mathsf{cz}\}, \mathbb{Q}, 1/2) \approx r(\mathbb{C}_1, \{\mathsf{cz}\}, \mathbb{Q}, 1/2)$ for all ten subsets of qubits $\mathbb{Q}$. All single-qubit gates in our experiments are implemented using a composite $\mathsf{u}(\theta, \phi, \lambda)$ gate [see Eq. (7.10)] that contains two $\mathsf{x}_{\pi/2}$ gates and three z_θ gates. This is the case even for unitaries that do not require two $\mathsf{x}_{\pi/2}$ pulses, such as the identity. The difference between any two single-qubit gates is therefore only in the angles of the three z_θ gates within $\mathsf{u}(\theta, \phi, \lambda)$. These gates are implemented by in-software phase updates on later pulses [50], so it is expected that these "virtual gates" cause negligible errors.

[1] For the four one-qubit subsets, three of the cases coincide—as they differ only by the two-qubit gate set or the two-qubit gate density, which are unused parameters in one-qubit circuits. In that case we only sample and run only of the three identical MRB designs.

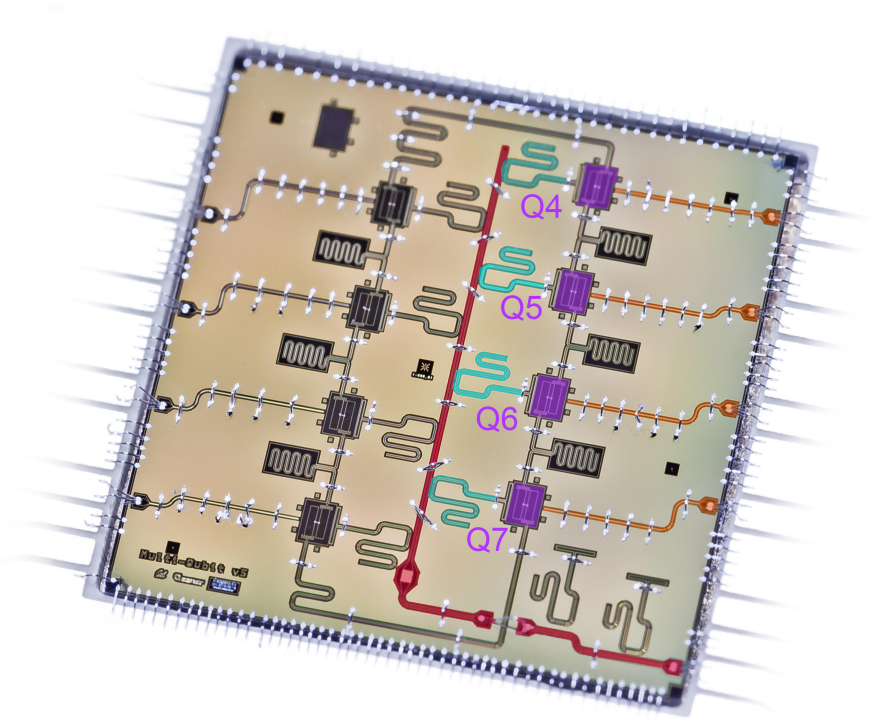

Figure 7.5: **The Advanced Quantum Testbed.** We performed MRB experiments on four qubits (Q4-Q7) of AQT's eight-qubit superconducting transmon processor (AQT@LBNL Trailblazer8-v5.c2). The processor includes 8 fixed frequency transmons coupled in a ring geometry. Each qubit (purple) has its own control line (orange) and readout resonator (cyan) coupled to a shared readout bus (red) for multiplexed readout.

Comparing Figs. 7.6 (d) and (f), we observe that the error rates for layers containing cs and $\mathsf{cs}^\dagger$ gates are all almost equal to, but slightly larger than, the error rates for layers containing cz gates. The largest relative difference is in the experiments on the 3-qubit set $\{\mathsf{Q4}, \mathsf{Q5}, \mathsf{Q6}\}$:

$$r(\mathbb{SU}(2), \{\mathsf{cs}, \mathsf{cs}^\dagger\}, \{\mathsf{Q4}, \mathsf{Q5}, \mathsf{Q6}\}, 1/2) = 1.64(5)\% \tag{7.13}$$

and

$$r(\mathbb{SU}(2), \{\mathsf{cz}\}, \{\mathsf{Q4}, \mathsf{Q5}, \mathsf{Q6}\}, 1/2) = 1.48(4)\% \tag{7.14}$$

. The three different two-qubit gates (cs, $\mathsf{cs}^\dagger$, and cz) on each qubit pair were *a priori* expected to have similar error rates, due to their similar calibration procedures. The slightly larger error rates for cs and $\mathsf{cs}^\dagger$ were cross-validated using cycle benchmarking [20]. Therefore, these results are experimental evidence for the robustness of MRB with non-Clifford two-qubit gates.

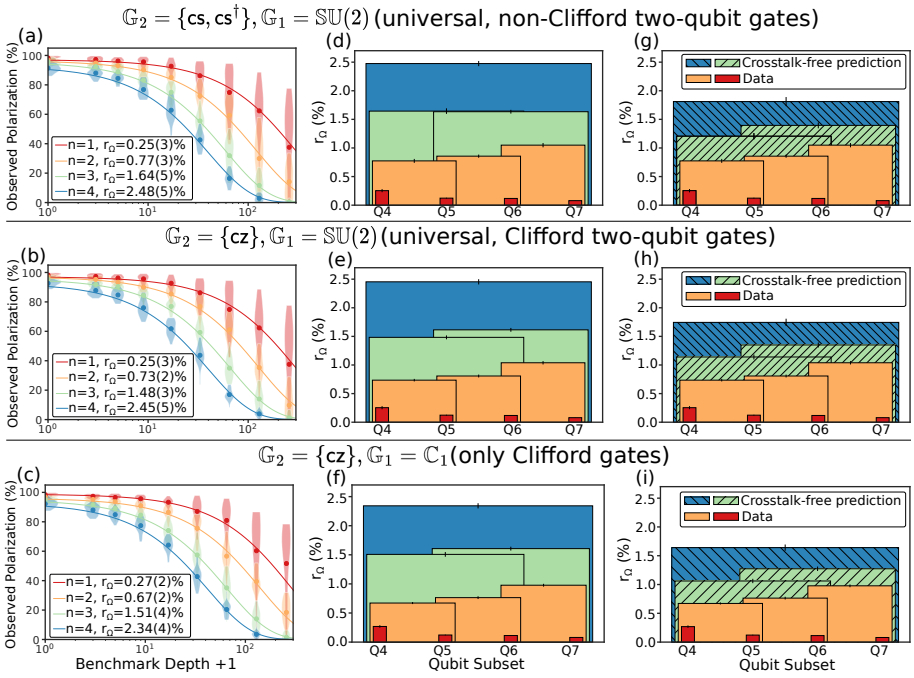

Figure 7.6: **Randomized benchmarking of universal gate sets on four qubits of the Advanced Quantum Testbed.** We used MRB to benchmark n-qubit layers constructed from three different gate sets, on each connected n-qubit subset of a linearly-connected set of four qubits $\{\mathtt{Q4}, \mathtt{Q5}, \mathtt{Q6}, \mathtt{Q7}\}$ in an eight-qubit superconducting transmon processor (`AQT@LBNL Trailblazer8-v5.c2`). The rows correspond to results from three different choices of gate set, each consisting of a two-qubit gate set $\mathbb{G}_2$ and a single-qubit gate set $\mathbb{G}_1$. From top to bottom, the rows correspond to: a universal gate set containing two non-Clifford entangling gates and the set of all single-qubit gates [$\mathbb{G}_2 = \{\mathsf{cs}, \mathsf{cs}^\dagger\}$, $\mathbb{G}_1 = \mathbb{SU}(2)$]; a universal gate set containing a Clifford entangling gate and the set of all single-qubit gates [$\mathbb{G}_2 = \{\mathsf{cz}\}$, $\mathbb{G}_1 = \mathbb{SU}(2)$]; and a non-universal, Clifford gate set [$\mathbb{G}_2 = \{\mathsf{cz}\}$, $\mathbb{G}_1 = \mathbb{C}_1$ where $\mathbb{C}_1$ is the one-qubit Clifford group]. (a-c): MRB decays for the qubit subsets $\{\mathtt{Q4}\}$, $\{\mathtt{Q4}, \mathtt{Q5}\}$, $\{\mathtt{Q4}, \mathtt{Q5}, \mathtt{Q6}\}$, and $\{\mathtt{Q4}, \mathtt{Q5}, \mathtt{Q6}, \mathtt{Q7}\}$. Violin plots and points show the distribution and mean, respectively, of the MRB circuit's observed polarization (S_d) versus benchmark depth (d). The curve is a fit of the mean of S_d ($\bar{S}_d$) to $\bar{S}_d = Ap^d$. The average error rate of an n-qubit layer (r_Ω) is given by $r_\Omega = (4^n - 1)(1 - p)/4^n$. The observed $\bar{S}_d$ decays exponentially, as predicted by our theory for MRB. (d-f): The estimated error rate r_Ω for each qubit subset that we benchmarked. (g-i): Predictions for the average layer error rate of 3- and 4-qubit subsets (hatched) based on the experimental 1- and 2-qubit error rates (un-hatched) and the assumption of no crosstalk errors. The difference between (d-f) and (g-i) quantifies the contribution of crosstalk errors to the average error rate of an n-qubit layer, for $n = 3, 4$. For all three gate sets and $n = 4$, we see that crosstalk errors are contributing approximately 0.7% error to r_Ω, which is approximately $1/3$ of r_Ω.

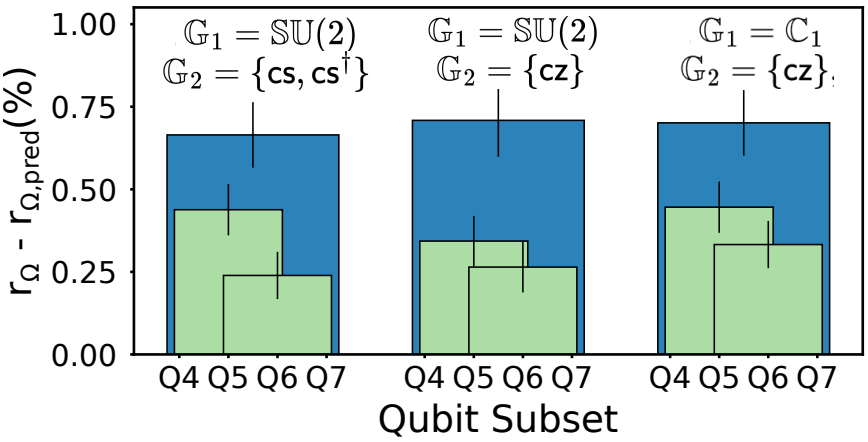

Figure 7.7: tbd

In Fig. 7.6g-i, we begin to use our data to extract information about crosstalk on our system. We take the error values for 1 and 2 qubit circuits from the middle plot and put them back into simulations for 3 and 4 qubit circuits. Then we run the simulations and data analysis to see what we predict should be the average error per layer given we are limited to JUST single and 2qubit errors. We plot those in the striped bars on the third column. We see that these stripped predicted errors are much lower than what we actually get in the middle column for 3 and 4 qubit circuits . Indicating that there are errors that only appear when more qubits participate in the circuits. We plot the difference between the expected crosstalk free and measured values for 3 and 4 qubit circuits in Fig. 7.7.

7.5 Error Models

In addition to using RB data to extra the average error per layer, as we have done, we can also use this data to look at the error for a specific layer using error models. Fitting to error models is a known technique can also be a computationally expensive step that scales exponentially with the number of qubits. We take our simulations from before and generate circuits with a certain amount of error injected in. We tweak the quantities of these errors until the resulting simulation result is as close to our own data as possible. Figure Fig. 7.8 shows the two models we use. The first kind of model is a depolarizing model. We give each kind of 1 and 2 qubit gate an error. (click) We then use this error to calculate an overall error rate for an n-qubit layer, represented in the circuit here as $D_{\gamma L}$. In other words for each dressed layer of gates we insert a probability of failure. In this model the probability of any n qubit error (XXXX, XIXX etc etc) is equal. We give it a probability that the state at the end of this layer becomes maximally mixed. We verify the results of our first model with a second more complex model that is less efficient to run. The second model is the Pauli Stochastic model, where more parameters vary. The probability of error is allowed to

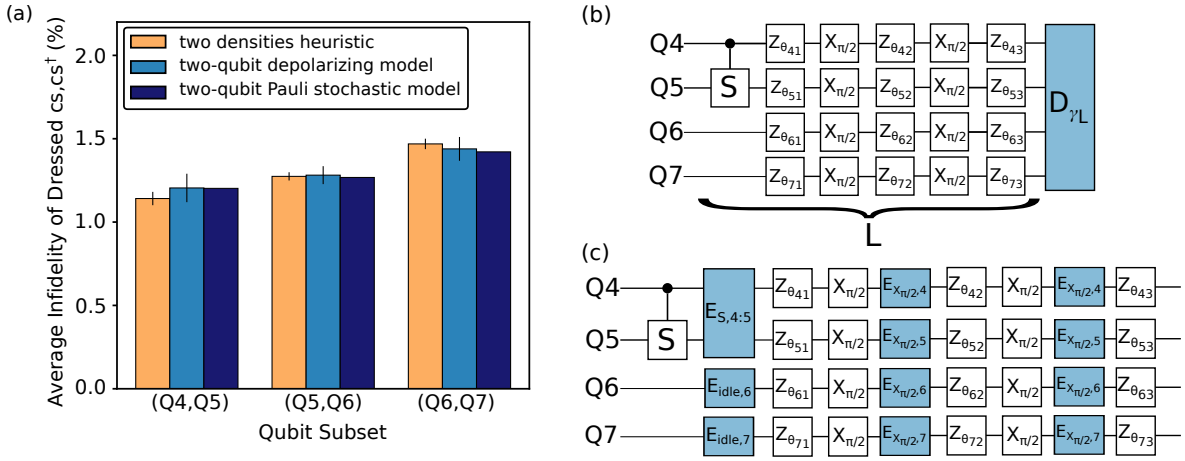

Figure 7.8: **Fitting error models to MRB data and estimating gate error rates.** We fit two types of error models to MRB data to estimate the infidelity of individual circuit layers. (a) By running two MRB experiments with two different two-qubit gate densities ξ, we can estimate the mean infidelity of a set of one or more two-qubit gates—here cs and $\mathsf{cs}^\dagger$—using basic linear algebra. We call this procedure the *two densities heuristic*. The estimates of the average gate error obtained from the two densities heuristic (orange) are compared to independent estimates obtained from two more rigorous but more complex and computationally intensive procedures: fitting each set of two-qubit MRB data to (1) a depolarizing model (light blue), and (2) a stochastic Pauli errors model (dark blue). (b) To fit a depolarizing model, we assign an error rate to each dressed layer and an error rate to each qubit's readout. (c) To fit a Pauli stochastic model, we assign a Pauli stochastic channel to each possible gate except the virtual z_θ gates.

be different for each qubit. For instance after our two qubit gate on Q4 and Q5 on the left, we insert the purple $E_{s4,5}$. Which represents the probability of that gate failing. During that time, we also let the two idle qubits (6-7) have a probability of error. In addition, we assign a probability of error to the remaining single qubit gates.

Using these models we fit the data and extract process infidelities for each 2 qubit gate over the three pairs of connected qubits (45, 56, and 67), displayed in Fig. 7.9. The process infidelities are between 2-3 % for the models in bright blue and dark blue. The takeaway from these plots is that the two qubit model in orange is underpredicting error by 56%. Where as the 4 qubit models capture the error much better. We attribute this difference between the 2 qubit models and the 4 qubit models to the presence of crosstalk. There doesn't seem to be much obvious different between the depolarizing model (which is much less computationally expensive) and the Pauli stochastic one, so this was a nice verification that in the future we can rely more on the easier depolarizing model.

In addition, we verify our models against cycle benchmarking data (shown in light blue) in Fig. 7.9. Our results are broadly similar to the CB results. We get at most 22% disagreement

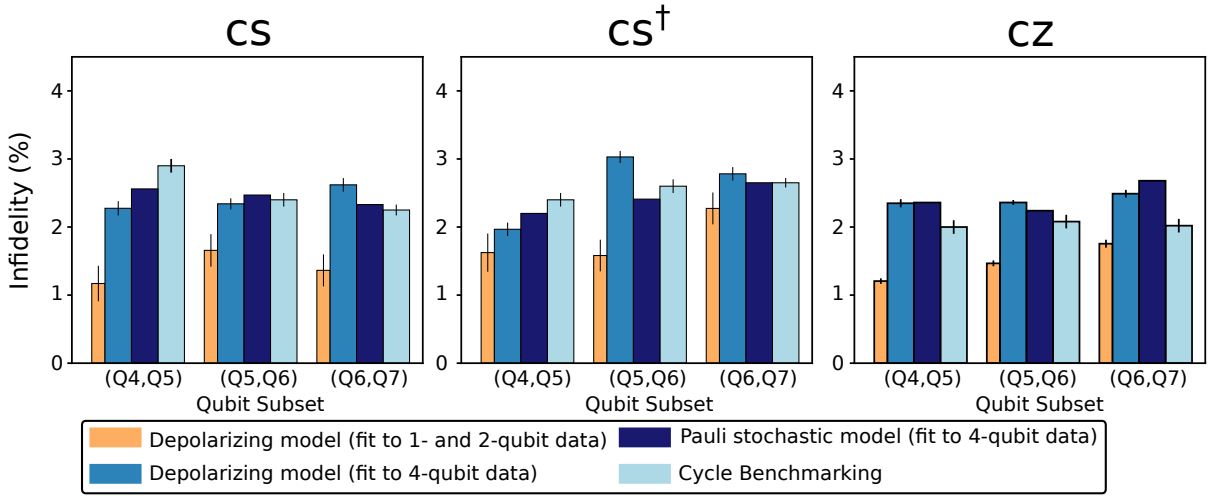

Figure 7.9: **Estimating the infidelity of dressed 4-qubit layers.** By fitting error models to MRB data, we can estimate the infidelity of each $\mathbb{G}_1$-dressed layer used in the MRB circuits. Here we show four different estimates of the infidelities of 4-qubit layers containing a single CS, CS† or CZ gate on one of the three connected pairs of qubits. We fit a simple n-qubit depolarizing model to (1) the 4-qubit data, and (2) the 1- and 2-qubit data, and use both models to estimate the infidelity of 4-qubit $\mathbb{G}_1$-dressed layers. The estimates from fitting to the 1- and 2-qubit data do not account for any additional crosstalk errors that occur in 4-qubit layers, so the additional error estimated when fitting to the 4-qubit data is a quantification of crosstalk. We also fit a more sophisticated stochastic Pauli error model to the 4-qubit circuit data, resulting in comparable estimates to those obtained from the simple depolarizing model (which uses a scalable, less computationally intensive analysis). To validate our results against an established technique, we compare to infidelities independently estimated using cycle benchmarking [20]. We observe qualitative agreement. The cycle benchmarking experiments measure the infidelities of layers dressed with one-qubit gates sampled from a different gate set (the Pauli group) to that used in our MRB experiments [$\mathbb{SU}(2)$ or $\mathbb{C}_1$, the single-qubit Clifford group], and these experiments were implemented on a different day than the MRB circuits, so exact agreement is not expected.

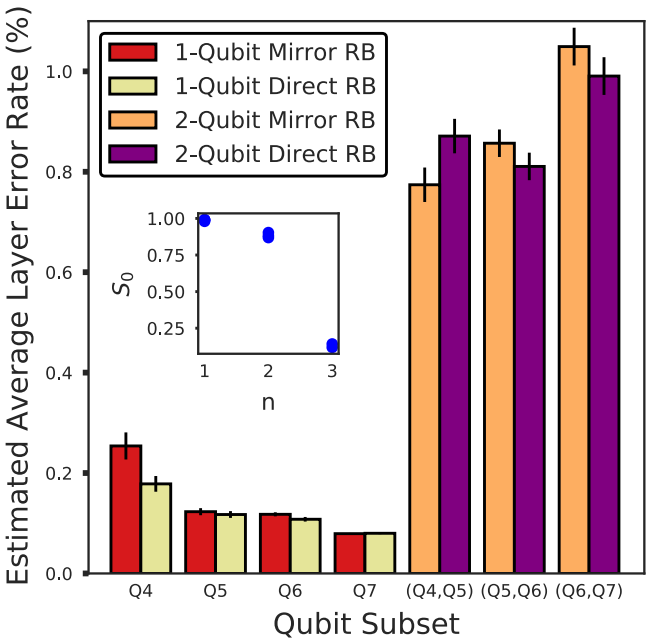

Figure 7.10: Comparing to DRB to mirror circuits

between the models and cycle benchmarking. This difference is expected because of two reasons: 1) the CB data was taken on a day earlier than the other data. And 2) CB estimates layers that are dressed with random Paulis where as our circuits have layers dressed by single qubit Clifford gates. Overall though, this plot suggests crosstalk in our system.

Our last piece of verification is to compare our results with other established forms of benchmarking. We do this for up to two qubits because randomized benchmarking is still practical to implement for a couple of qubits. Specifically, we choose Direct Randomized Benchmarking (DRB), a variant of standard RB developed by Tim Proctor's team [68] because standard RB measures an error rate the is not directly comparable the error rate derived from Mirror randomized benchmarking. In contrast, DRB is designed to do so. Furthermore, DRB is now an established technique that is known to be reliable. We run DRB on the gate set that contains non-Cliffords gates: cs and $\mathsf{cs}^\dagger$. In Fig. 7.10, the red and orange bars are for Mirror RB (single and two qubit circuits, respectively) And then the yellow and purple bars are for Direct RB results. We see good agreement between the error rates from MRB and Direct RB. The inset of this plot is the effective polarization obtained

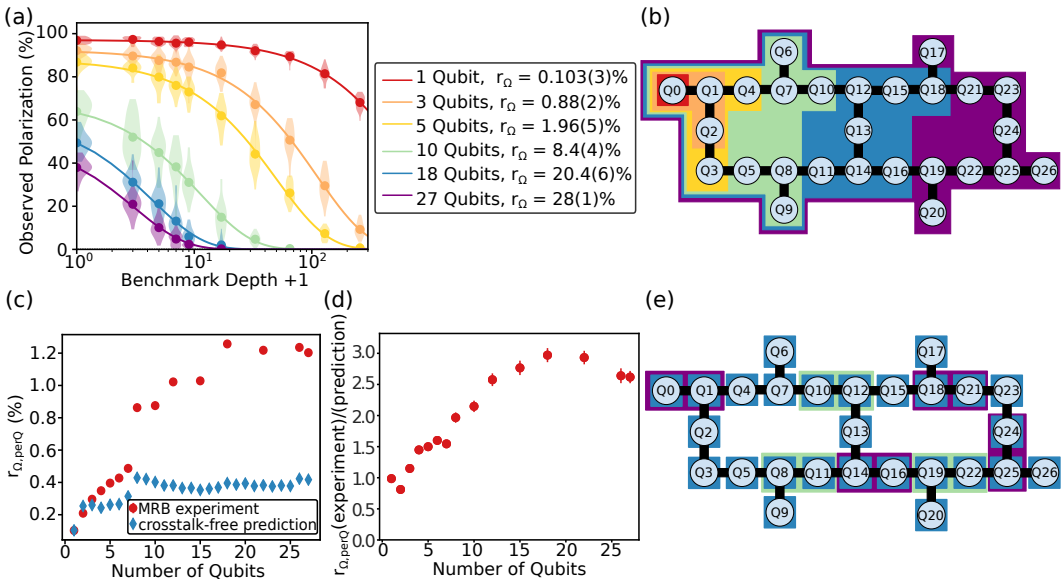

Figure 7.11: Comparing to IBM

from Direct RB for a fixed depth as a function of the qubit number. The sharp decrease in polarization happened because of the overwhelming overhead required to implement a Haar-random unitary for DRB. The circuits were so large that we were not able to measure anything for 3 or more qubits. This again demonstrates that DRB cannot be scalable. But it's encouraging that the MRB results are comparable and trustworthy, so that we can use it for larger numbers of qubits in situations out of reach for DRB.

7.6 Demonstrating Scalability

One important feature of our benchmarking is its scalability. We demonstrate this by repeating our measurements on the IBMQ-Montreal device, which has 27 qubits, and characterizing the crosstalk. The universal gate set consists of the CNOT gate and single qubit rotations. We first run circuits for different qubit numbers, ranging from 1 to 27 qubits. Fig. 7.11b shows the connectivity of the chip and highlights which qubits were used for each type of circuit. For instance Q0 highlighted in red was used for the 1 qubit circuit and corresponds to the familiar exponential plot in red on the left. Similarly a 3 qubit circuit was run on qubits 0, 1, and 2, highlighted in orange on the chip and the data is in orange on the left. And so on. Even for the 27 qubit circuits, we can still get effective polarizations that are much greater than 0 and extract an average layer error rate of 28%, showing that our MRB protocol is indeed scalable.

Next we would like to understand the crosstalk present on the IBM chip. We take a second round of data where we run 27 single-qubit circuits simultaneously and also several

2 qubit circuits simultaneously. Ideally we'd run them all separately, but that would take a good bit of time on the IBM processor and there was a long queue of submitted user projects so we ended up running them simultaneously.

In the Fig. 7.11e, the blue shaded squares around each qubit indicate that a single qubit circuit was run on each qubit. In addition, the purple and green boxes indicate the qubits that were paired up to use for 2 qubit circuits. The pairs of qubits for the 2 qubit circuits were selected based on frequency constraints

Much like before, we use these single and 2 qubit circuit results to extract an average error per layer. Then we use these error rates in our simulations to predict what error rates we would expect for circuits using more qubits. We compare these predictions to our data in the middle plot.

I will point out here that we are plotting r_Ω divided by the number of qubits, which is the observed error rate per qubit. The predictions are in blue. There is a small rise in error between 1 and 2 qubits, which is pretty typical and expected. As two qubit gates generally have more error than 1 qubit gates. Then the plot remains flat for the most part for higher qubit numbers. This complies with our intuition for 2 reasons. First without crosstalk, it is expected that the error per qubit to be unaffected by the presence of other qubits participating. Second, we also take care to keep the density of 2 qubit gates constant throughout all the circuits. We note that there is a slight kink upwards at around $n = 8$. We believe this is due to the fact that not all of the qubits on the IBM chip have the same error. Some are a bit higher than others and contribute a little more error[7]. But for the most part the error per qubit is constant. This is very different from the points in red, which are the results of actual circuits for qubit numbers ranging from 1 through 27. Here, the error per qubit continues to climb far past $n = 2$ and is about 0.5% higher than the prediction. This is further emphasized in the last plot. The Y axis has been adjusted to be the ratio between the error obtained from data and the predicted error. We divided the red points by the blue points. There is a clear 250-300% increase that saturates at around $n = 15$. This indicates crosstalk with a finite spatial radius.

7.7 Conclusion and Outlook

Scalable benchmarking methods are needed to quantify the integrated performance of medium- and large-scale quantum processors. We have introduced Mirror randomized circuits as a new method to benchmark processors. We have showed that it is scalable for both Clifford and also universal gate sets. The method reliably measures the error rate of a random n-qubit circuit layer sampled from a user-specified distribution Ω. We verified our results using simulations and other established techniques like Cycle benchmarking and direct randomized benchmarking for low numbers of qubits. Additionally, we showed that Mirror RB data contains more information than just the average error per layer. We quantified the

[7]Kpop fun fact 7/7: BTS was the first group to reach No. 1 on the US Albums chart and also No. 1 on the US Artists 100. They are also the most streamed group on Spotify (16,300,000,000 streams total).

amount of crosstalk on two different processors and obtain two qubit gate fidelities through fitting error models, such as the scalable depolarizing model. Which gave us results comparable to the more computationally expensive Pauli stochastic model. Our results indicate that MRB on many qubits reveals and quantifies errors not present in one- and two-qubit circuits, highlighting the importance of scalable benchmarks.

Our method can be viewed as both an adaptation of standard RB and its variants, to enable efficient and scalable benchmarking of universal gate sets, and as an adaptation of XEB that removes XEB's inefficient circuit simulation step. It therefore provides a link between two widely used benchmarking methodologies, and so we anticipate that the ideas introduced here will lead to further advances in randomized benchmarking.

We expect that a variety of interesting benchmarking methods can be constructed using MRB and extensions or adaptations of this method. First, while our simulations and data indicate that our method works well for non-Clifford gates, we believe that we can make this method even more sensitive to coherent errors in non-Clifford gates by changing the way the circuits are sampled. For example we can change the layer structure and include consecutive 2 qubit gate layers. Adjusting the circuit structures will also improve the predictive power of our error model fitting. In addition, many algorithms have more structure. Our method is still based on randomized gates and it would be interesting to look at how we can apply our mirror technique to make scalable benchmarking methods with more structured circuits. Furthermore, we anticipate that MRB can form the foundation of methods for estimating the error rates of individual gates and layers, within the context of many-qubit circuits. In this work we demonstrated a simple example of such a technique—fitting MRB data to a depolarizing model—and we expect that a variety of robust methods could be developed, that would complement or advance on existing methods for this task [45, 20, 21] such as interleaved RB. Alternatively, we anticipate that MRB can be adapted to construct scalable "full-stack" benchmarks based on random circuits, such as a scalable variant of the widely-used quantum volume benchmark [15].

Bibliography

[1] M. Virginia P. Altoé et al. *Localization and reduction of superconducting quantum coherent circuit losses*. 2020. arXiv: 2012.07604 [quant-ph].

[2] *AQT@LBL - SC Qubit Testbed*. https://aqt.lbl.gov/. Accessed: 2020-03-01.

[3] Camille Aron, Manas Kulkarni, and Hakan E. Türeci. "Steady-state entanglement of spatially separated qubits via quantum bath engineering". In: *Phys. Rev. A* 90 (6 Dec. 2014), p. 062305. DOI: 10.1103/PhysRevA.90.062305. URL: https://link.aps.org/doi/10.1103/PhysRevA.90.062305.

[4] Frank Arute et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (2019), pp. 505–510. URL: https://www.nature.com/articles/s41586-019-1666-5.

[5] Alexandre Blais et al. "Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation". In: *Phys. Rev. A* 69 (6 June 2004), p. 062320. DOI: 10.1103/PhysRevA.69.062320. URL: https://link.aps.org/doi/10.1103/PhysRevA.69.062320.

[6] M. G. Blamire. "The Potential of the Josephson Energy". In: *Journal of Superconductivity and Novel Magnetism* 34.6 (June 2021), pp. 1629–1632. DOI: 10.1007/s10948-020-05614-y. URL: https://doi.org/10.1007/s10948-020-05614-y.

[7] Robin Blume-Kohout et al. "A Taxonomy of Small Markovian Errors". In: *PRX Quantum* 3 (2 May 2022), p. 020335. DOI: 10.1103/PRXQuantum.3.020335. URL: https://link.aps.org/doi/10.1103/PRXQuantum.3.020335.

[8] Sergio Boixo et al. "Characterizing quantum supremacy in near-term devices". In: *Nat. Phys.* 14.6 (2018), p. 595. URL: https://www.nature.com/articles/s41567-018-0124-x.

[9] Winton G. Brown and Bryan Eastin. "Randomized benchmarking with restricted gate sets". In: *Phys. Rev. A* 97 (2018), p. 062323. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.97.062323.

[10] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson. "Characterizing universal gate sets via dihedral benchmarking". In: *Phys. Rev. A* 92.6 (2015), p. 060302. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.92.060302.

[11] Isaac L. Chuang and M. A. Nielsen. "Prescription for experimental determination of the dynamics of a quantum black box". In: *Journal of Modern Optics* 44.11-12 (1997), pp. 2455–2467. DOI: 10.1080/09500349708231894. eprint: https://www.tandfonline.com/doi/pdf/10.1080/09500349708231894.

[12] Jahan Claes, Eleanor Rieffel, and Zhihui Wang. "Character Randomized Benchmarking for Non-Multiplicity-Free Groups With Applications to Subspace, Leakage, and Matchgate Randomized Benchmarking". In: *PRX Quantum* 2 (1 Mar. 2021), p. 010351. DOI: 10.1103/PRXQuantum.2.010351. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.010351.

[13] Antonio D. Córcoles et al. "Protecting superconducting qubits from radiation". In: *Applied Physics Letters* 99.18 (2011), p. 181906. DOI: 10.1063/1.3658630. eprint: https://doi.org/10.1063/1.3658630. URL: https://doi.org/10.1063/1.3658630.

[14] Andrew W Cross et al. "Scalable randomised benchmarking of non-Clifford gates". In: *NPJ Quantum Inf.* 2 (2016), p. 16012. URL: http://www.nature.com/articles/npjqi201612.

[15] Andrew W Cross et al. "Validating quantum computers using randomized model circuits". In: *Phys. Rev. A* 100 (2019), p. 032328. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.100.032328.

[16] Daoyi Dong et al. "Learning robust pulses for generating universal quantum gates". In: *Scientific Reports* 6.1 (Oct. 2016), p. 36090. DOI: 10.1038/srep36090. URL: https://doi.org/10.1038/srep36090.

[17] A. Eddins et al. "Stroboscopic Qubit Measurement with Squeezed Illumination". In: *Phys. Rev. Lett.* 120 (4 Jan. 2018), p. 040505. DOI: 10.1103/PhysRevLett.120.040505. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.040505.

[18] Joseph Emerson, Robert Alicki, and Karol Życzkowski. "Scalable noise estimation with random unitary operators". In: *J. Opt. B Quantum Semiclass. Opt.* 7.10 (2005), S347. URL: http://iopscience.iop.org/article/10.1088/1464-4266/7/10/021/meta.

[19] Joseph Emerson et al. "Symmetrized characterization of noisy quantum processes". In: *Science* 317.5846 (2007), pp. 1893–1896. URL: http://science.sciencemag.org/content/317/5846/1893.

[20] Alexander Erhard et al. "Characterizing large-scale quantum computers via cycle benchmarking". In: *Nat. Commun.* 10 (2019), p. 5347. URL: https://www.nature.com/articles/s41467-019-13068-7.

[21] Steven T Flammia. "Averaged circuit eigenvalue sampling". In: (Aug. 2021). arXiv: 2108.05803 [quant-ph]. URL: http://arxiv.org/abs/2108.05803.

[22] Steven T. Flammia and Joel J. Wallman. "Efficient Estimation of Pauli Channels". In: *ACM Trans. Quant. Comp.* 1 (2020), p. 3. URL: https://doi.org/10.1145/3408039.

[23] Jay M Gambetta et al. "Characterization of addressability by simultaneous randomized benchmarking". In: *Phys. Rev. Lett.* 109.24 (2012), p. 240504. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.109.240504.

[24] Shelly Garion et al. "Experimental implementation of non-Clifford interleaved randomized benchmarking with a controlled-S gate". In: *Phys. Rev. Research* 3 (1 Mar. 2021), p. 013204. DOI: 10.1103/PhysRevResearch.3.013204. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.3.013204.

[25] Robin Harper and Steven T Flammia. "Estimating the fidelity of T gates using standard interleaved randomized benchmarking". In: *Quantum Sci. Technol.* 2.1 (2017), p. 015008. URL: http://iopscience.iop.org/article/10.1088/2058-9565/aa5f8d/meta.

[26] Robin Harper, Steven T. Flammia, and Joel J. Wallman. "Efficient Learning of Quantum Noise". In: *Nat. Phys.* 16 (2020), pp. 1–5. URL: https://www.nature.com/articles/s41567-020-0992-8.

[27] A. K. Hashagen et al. "Real Randomized Benchmarking". In: *Quantum* 2 (2018), p. 85. URL: https://doi.org/10.22331/q-2018-08-22-85.

[28] Jonas Helsen et al. "A general framework for randomized benchmarking". In: (Oct. 2020). arXiv: 2010.07974 [quant-ph]. URL: http://arxiv.org/abs/2010.07974.

[29] Jonas Helsen et al. "A new class of efficient randomized benchmarking protocols". In: *arXiv preprint arXiv:1806.02048* (2018). URL: https://arxiv.org/abs/1806.02048.

[30] Jonas Helsen et al. "Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates". In: (Nov. 2020). arXiv: 2011.13048 [quant-ph]. URL: http://arxiv.org/abs/2011.13048.

[31] Jordan Hines et al. *Demonstrating scalable randomized benchmarking of universal gate sets*. 2022. DOI: 10.48550/ARXIV.2207.07272. URL: https://arxiv.org/abs/2207.07272.

[32] IBM Quantum https://quantum-computing.ibm.com. 2021.

[33] X. Y. Jin et al. "Thermal and Residual Excited-State Population in a 3D Transmon Qubit". In: *Phys. Rev. Lett.* 114 (24 June 2015), p. 240501. DOI: 10.1103/PhysRevLett.114.240501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.240501.

[34] J. E. Johnson et al. "Heralded State Preparation in a Superconducting Qubit". In: *Phys. Rev. Lett.* 109 (5 Aug. 2012), p. 050506. DOI: 10.1103/PhysRevLett.109.050506. URL: https://link.aps.org/doi/10.1103/PhysRevLett.109.050506.

[35] A Kandala et al. "Demonstration of a High-Fidelity cnot Gate for Fixed-Frequency Transmons with Engineered Z Z Suppression". In: *Physical Review Letters* 127.13 (2021), p. 130501.

[36] M. E. Kimchi-Schwartz et al. "Stabilizing Entanglement via Symmetry-Selective Bath Engineering in Superconducting Qubits". In: *Phys. Rev. Lett.* 116 (24 June 2016), p. 240503. DOI: 10.1103/PhysRevLett.116.240503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.240503.

[37] Emanuel Knill et al. "Randomized benchmarking of quantum gates". In: *Phys. Rev. A* 77.1 (2008), p. 012307. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.77.012307.

[38] Jens Koch et al. "Charge-insensitive qubit design derived from the Cooper pair box". In: *Phys. Rev. A* 76 (4 Oct. 2007), p. 042319. DOI: 10.1103/PhysRevA.76.042319. URL: https://link.aps.org/doi/10.1103/PhysRevA.76.042319.

[39] J M Kreikebaum et al. "Optimization of infrared and magnetic shielding of superconducting TiN and Al coplanar microwave resonators". In: *Superconductor Science and Technology* 29.10 (Aug. 2016), p. 104002. DOI: 10.1088/0953-2048/29/10/104002. URL: https://iopscience.iop.org/article/10.1088/0953-2048/29/10/104002.

[40] S. Krinner et al. "Engineering cryogenic setups for 100-qubit scale superconducting circuit systems". In: *EPJ Quantum Technology* 6.1 (2019), p. 2. DOI: 10.1140/epjqt/s40507-019-0072-0. URL: https://doi.org/10.1140/epjqt/s40507-019-0072-0.

[41] Yunchao Liu et al. *Benchmarking near-term quantum computers via random circuit sampling*. 2021. arXiv: 2105.05232 [quant-ph].

[42] C. Macklin et al. "A near–quantum-limited Josephson traveling-wave parametric amplifier". In: *Science* 350.6258 (2015), pp. 307–310. DOI: 10.1126/science.aaa8525. URL: https://www.science.org/doi/abs/10.1126/science.aaa8525.

[43] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. "Characterizing quantum gates via randomized benchmarking". In: *Phys. Rev. A* 85.4 (2012), p. 042311. URL: https://journals.aps.org/pra/pdf/10.1103/PhysRevA.85.042311.

[44] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. "Scalable and robust randomized benchmarking of quantum processes". In: *Phys. Rev. Lett.* 106.18 (2011), p. 180504. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.106.180504.

[45] Easwar Magesan et al. "Efficient measurement of quantum gate error by interleaved randomized benchmarking". In: *Phys. Rev. Lett.* 109.8 (2012), p. 080505. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.109.080505.

[46] Esteban A Martinez et al. "Compiling quantum algorithms for architectures with multi-qubit gates". In: *New Journal of Physics* 18.6 (June 2016), p. 063029. DOI: 10.1088/1367-2630/18/6/063029. URL: https://doi.org/10.1088%2F1367-2630%2F18%2F6%2F063029.

[47] Dmitri Maslov and Yunseong Nam. "Use of global interactions in efficient quantum circuit constructions". In: *New Journal of Physics* 20.3 (Mar. 2018), p. 033018. ISSN: 1367-2630. DOI: 10.1088/1367-2630/aaa398. URL: http://dx.doi.org/10.1088/1367-2630/aaa398.

[48] Karl Mayer et al. "Theory of mirror benchmarking and demonstration on a quantum computer". In: *arXiv [quant-ph]* (2021). arXiv: 2108.10431 [quant-ph]. URL: https://arxiv.org/abs/2108.10431.

[49] David C McKay et al. "Correlated Randomized Benchmarking". In: (Mar. 2020). arXiv: 2003.02354 [quant-ph]. URL: http://arxiv.org/abs/2003.02354.

[50] David C McKay et al. "Efficient Z gates for quantum computing". In: *Phys. Rev. A* 96.2 (Aug. 2017), p. 022330. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.96.022330. URL: https://link.aps.org/doi/10.1103/PhysRevA.96.022330.

[51] Bradley K Mitchell et al. "Hardware-Efficient Microwave-Activated Tunable Coupling Between Superconducting Qubits". In: *arXiv preprint arXiv:2105.05384* (2021).

[52] Klaus Mølmer and Anders Sørensen. "Multiparticle Entanglement of Hot Trapped Ions". In: *Phys. Rev. Lett.* 82 (9 Mar. 1999), pp. 1835–1838. DOI: 10.1103/PhysRevLett.82.1835. URL: https://link.aps.org/doi/10.1103/PhysRevLett.82.1835.

[53] Thomas Monz et al. "14-Qubit Entanglement: Creation and Coherence". In: *Phys. Rev. Lett.* 106 (13 Mar. 2011), p. 130506. DOI: 10.1103/PhysRevLett.106.130506. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.130506.

[54] Gary J Mooney, Charles D Hill, and Lloyd CL Hollenberg. "Entanglement in a 20-qubit superconducting quantum computer". In: *Scientific reports* 9.1 (2019), pp. 1–8.

[55] A. Morvan et al. "Qutrit Randomized Benchmarking". In: *Phys. Rev. Lett.* 126 (21 May 2021), p. 210504. DOI: 10.1103/PhysRevLett.126.210504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.210504.

[56] K. W. Murch et al. "Cavity-Assisted Quantum Bath Engineering". In: *Physical Review Letters* 109.18 (Oct. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.183602. URL: http://dx.doi.org/10.1103/PhysRevLett.109.183602.

[57] V Negirneac et al. "High-fidelity controlled-z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor". In: *Physical Review Letters* 126.22 (2021), p. 220502.

[58] Erik Nielsen et al. "Gate Set Tomography". In: *Quantum* 5 (Oct. 2021), p. 557. ISSN: 2521-327X. DOI: 10.22331/q-2021-10-05-557. URL: https://doi.org/10.22331/q-2021-10-05-557.

[59] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. 2nd ed. Massachusetts: Cambridge University Press, 2000.

[60] Simon E. Nigg et al. "Black-Box Superconducting Circuit Quantization". In: *Phys. Rev. Lett.* 108 (24 June 2012), p. 240502. DOI: 10.1103/PhysRevLett.108.240502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.108.240502.

[61] Kevin O'Brien et al. "Resonant Phase Matching of Josephson Junction Traveling Wave Parametric Amplifiers". In: *Phys. Rev. Lett.* 113 (15 Oct. 2014), p. 157001. DOI: 10.1103/PhysRevLett.113.157001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.157001.

[62] Alexander P. M. Place et al. "New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds". In: *Nature Communications* 12.1 (Mar. 2021), p. 1779. DOI: 10.1038/s41467-021-22030-5. URL: https://doi.org/10.1038/s41467-021-22030-5.

[63] J. F. Poyatos, J. I. Cirac, and P. Zoller. "Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate". In: *Phys. Rev. Lett.* 78 (2 Jan. 1997), pp. 390–393. DOI: 10.1103/PhysRevLett.78.390. URL: https://link.aps.org/doi/10.1103/PhysRevLett.78.390.

[64] Timothy Proctor et al. *Establishing trust in quantum computations*. 2022. arXiv: 2204.07568 [quant-ph].

[65] Timothy Proctor et al. "Measuring the capabilities of quantum computers". In: *Nature Physics* 18.1 (Jan. 2022), pp. 75–79. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01409-7. URL: https://doi.org/10.1038/s41567-021-01409-7.

[66] Timothy Proctor et al. "Scalable randomized benchmarking of quantum computers using mirror circuits". In: (Sept. 2020). arXiv: 2112.09853 [quant-ph]. URL: https://arxiv.org/abs/2112.09853.

[67] Timothy J Proctor et al. "Direct randomized benchmarking for multiqubit devices". In: *Phys. Rev. Lett.* 123.030503 (2019). URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.123.030503.

[68] Timothy J Proctor et al. "Direct randomized benchmarking: robust and flexible benchmarking of quantum gates". In: *In preparation* 030503 (2021).

[69] Kenneth Rudinger et al. "Experimental Characterization of Crosstalk Errors with Simultaneous Gate Set Tomography". In: *PRX Quantum* 2 (4 Nov. 2021), p. 040338. DOI: 10.1103/PRXQuantum.2.040338. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.040338.

[70] Mohan Sarovar et al. "Detecting crosstalk errors in quantum information processors". In: *Quantum* 4 (2020), p. 321. URL: https://quantum-journal.org/papers/q-2020-09-11-321/.

[71] Vivek V. Shende, Igor L. Markov, and Stephen S. Bullock. "Minimal universal two-qubit controlled-NOT-based circuits". In: *Physical Review A* 69.6 (June 2004). ISSN: 1094-1622. DOI: 10.1103/physreva.69.062321. URL: http://dx.doi.org/10.1103/PhysRevA.69.062321.

[72] Anders Sørensen and Klaus Mølmer. "Entanglement and quantum computation with ions in thermal motion". In: *Phys. Rev. A* 62 (2 July 2000), p. 022311. DOI: 10.1103/PhysRevA.62.022311. URL: https://link.aps.org/doi/10.1103/PhysRevA.62.022311.

[73] Anders Sørensen and Klaus Mølmer. "Quantum Computation with Ions in Thermal Motion". In: *Phys. Rev. Lett.* 82 (9 Mar. 1999), pp. 1971–1974. DOI: 10.1103/PhysRevLett.82.1971. URL: https://link.aps.org/doi/10.1103/PhysRevLett.82.1971.

[74] J Stehlik et al. "Tunable Coupling Architecture for Fixed-Frequency Transmon Superconducting Qubits". In: *Physical Review Letters* 127.8 (2021), p. 080505.

[75] Joel J Wallman and Joseph Emerson. "Noise tailoring for scalable quantum computation via randomized compiling". In: *Phys. Rev. A* 94 (2016), p. 052325. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.94.052325.

[76] Xi-Lin Wang et al. "18-Qubit Entanglement with Six Photons' Three Degrees of Freedom". In: *Phys. Rev. Lett.* 120 (26 June 2018), p. 260502. DOI: 10.1103/PhysRevLett.120.260502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.260502.

[77] M. Werninghaus et al. "Leakage reduction in fast superconducting qubit gates via optimal control". In: *npj Quantum Information* 7.1 (Jan. 2021), p. 14. DOI: 10.1038/s41534-020-00346-2. URL: https://doi.org/10.1038/s41534-020-00346-2.

[78] Fei Yan et al. "Rotating-frame relaxation as a noise spectrum analyser of a superconducting qubit undergoing driven evolution". In: *Nature Communications* 4.1 (Aug. 2013). ISSN: 2041-1723. DOI: 10.1038/ncomms3337. URL: http://dx.doi.org/10.1038/ncomms3337.