Investigating Microwave-Activated Entangling Gates on Superconducting Quantum Processors

by

Bradley Kenneth Mitchell

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 K. Birgitta Whaley
Professor John Clarke

Spring 2022

Investigating Microwave-Activated Entangling Gates on Superconducting Quantum Processors

Copyright 2022
by
Bradley Kenneth Mitchell

Abstract

Investigating Microwave-Activated Entangling Gates on Superconducting Quantum Processors

by

Bradley Kenneth Mitchell

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Irfan Siddiqi, Chair

Superconducting quantum circuits are a leading technology in the quest toward building a quantum computer, which promises to outperform conventional, or "classical" computers, in solving a variety of tasks. To be useful for computation, a quantum processing unit (QPU) architecture must be scalable and have low error rates. Microwave activated entangling interactions, while often being slower than interactions that use flux-tunable components, are a scalable approach to realizing entanglement in that they are compatible with a minimally complex QPU design: single-junction transmon qubits with fixed qubit-qubit coupling. A prominent microwave-activated gate in this architecture, the cross resonance (CR) gate, is experimentally investigated, including mitigation of leakage errors, and budgeting of coherent and incoherent errors. Importantly, the fidelity of the CR gate is limited in general by the static cross-Kerr, or ZZ, interaction between the qubits. A novel cross-Kerr entangling interaction that commutes with the static cross-Kerr interaction is presented, based on simultaneous off-resonant drives. This cross-Kerr interaction is tunable, enabling the complete cancellation or enhancement of qubit-qubit coupling, a first for this fixed-frequency, fixed-coupling architecture. We show the tunability of the sign of the interaction, and use it to implement a two qubit controlled-phase (CZ) gate. This gate is observed to have lower coherent error than the CR gate. Designing a QPU with this as the native entangling interaction could enable faster microwave-activated gates, improving overall performance while maintaining minimal hardware complexity.

To my wife, family, and friends

Contents

Contents **ii**

List of Figures **iv**

List of Tables **xi**

1 Introduction **1**
1.1 Circuit Quantum Electrodynamics . . . 2
1.1.1 A brief history of superconducting quantum computing . . . 2
1.1.2 The quantum LC circuit . . . 3
1.1.3 The Josephson junction . . . 3
1.1.4 The transmon qubit . . . 4
1.1.5 Qubit-resonator coupling . . . 6
1.1.6 Coupling transmons . . . 7
1.1.7 Coupling to the environment . . . 8
1.2 The Trailblazer QPU . . . 9
1.3 Thesis structure and results summary . . . 12

2 Single-qubit Gates and Crosstalk **14**
2.1 Single Qubit Gate Calibration . . . 14
2.2 Mitigating Classical Crosstalk . . . 17
2.3 Error budgeting single-qubit gates . . . 22
2.3.1 ZZ error budget . . . 22
2.3.2 Sample B error budget . . . 23

3 Cross Resonance Gate **25**
3.1 Principle . . . 25
3.2 Dynamics . . . 28
3.2.1 Computational Space Dynamics . . . 28
3.2.2 Duffing Oscillator Dynamics . . . 28
3.2.3 Duffing Oscillator Dynamics with crosstalk . . . 30
3.2.4 Characterizing dynamics with Hamiltonian tomography . . . 31

3.2.5 Leakage Dynamics . . . 35
3.3 Gate Calibration . . . 37
3.3.1 Direct CNOT calibration . . . 39
3.3.1.1 Leakage reduction with 2D DRAG . . . 40
3.3.1.2 Error budgeting the direct CNOT gate . . . 44
3.3.2 Intrinsic CR gates . . . 47

4 Microwave-Activated Tunable Coupling **56**
4.1 Principle . . . 57
4.2 Perturbation theory derivation . . . 58
4.3 Numerical Simulations . . . 60
4.4 Dynamics . . . 61
4.5 Gate Calibration . . . 64
4.5.1 Entanglement calibration . . . 64
4.5.2 Local Z Gate Calibration . . . 64
4.6 Benchmarking . . . 66
4.6.1 Coherence Dependence on Drive Amplitude . . . 69
4.6.2 Leakage Randomized Benchmarking . . . 69
4.7 Discussion . . . 70

5 Future Directions **73**

A Appendix **75**
A.1 Benchmarking . . . 75
A.1.1 Average gate fidelity and process fidelity . . . 75
A.1.2 Process infidelity and Quantum Error Correction Thresholds . . . 76
A.1.3 Benchmarking protocols for measuring e_F . . . 77
A.1.4 Standard RB, Single and Two-qubit . . . 79
A.1.5 Interleaved RB . . . 79
A.1.6 Extended RB . . . 81
A.1.7 Leakage RB . . . 82
A.1.8 Cycle Benchmarking . . . 83

Bibliography **85**

List of Figures

1.1 **The transmon qubit as an anharmonic oscillator.** (a) Standard LC circuit (b) transmon qubit, replacing the inductor with a Josephson junction (c) comparison of energy spectrum for an LC circuit (dashed line) and a transmon (solid line), with the potential energy in gray and the eigenenergies denoted in the legend. The anharmonicity η is about 5% of the lowest energy transition $\omega_{01} = E_1 - E_0$. for typical transmons in our lab, corresponding to an $E_J/E_c = 70$. 5
1.2 **Two qubit QPU, the 'unit cell' of the Trailblazer quantum processor**. 9
1.3 **Trailblazer quantum processor 8-qubit ring.** Photo credit: John Mark Kreikebaum [31]. 10

2.1 $X_{\pi/2}$ **calibration (a)** Coarse pulse amplitude calibration **(b)** Ramsey frequency calibration with several artificial detunings ν to find the qubit frequency **(c)** Fine amplitude calibration, with pulse repetitions to find increase sensitivity to the optimal drive amplitude. **(d)** Phase error correction, bookends the pulse with virtual-Z_θ gates to correct for phase errors due to Stark shifts of the higher transmon levels. 16
2.2 **Classical crosstalk schematic. (a)** Drive amplitudes are sent to the qubits via their control lines, though crosstalk matrix C mixes the drive signals to realize fields ε_c, ε_t on the transmons. Coupling J can alter the effects that crosstalk has, such as "accidental CR" interactions. **(b)** Image of two coupled qubits on the Trailblazer chip, with control lines on the left. 18
2.3 **Crosstalk calibration for coupled qubits using the CR effect. (a)** Conditional Rabi rate experiment on Q_t by driving the Q_c control line. Here, $Q6$ is Q_t i.e. the targeted or driven qubit, and $Q5$ is Q_c or the spectator qubit. Measuring Rabi rates conditioned on Q_c state gives an estimate of the field present on Q_c. **(b)** Conditional Rabi on Q_t's own drive line, showing some conditionality, indicating crosstalk from L_t to Q_c that can be cancelled with a drive down L_c. **(c)** Calibrating the phase of the crosstalk cancellation tone down L_c. Minimizing the Q_t population difference (color bar) when Q_c is in $|0\rangle$ and $|1\rangle$ versus relative drive phase (y-axis) and pulse duration (x-axis). 19

2.4 **Benchmarking single-qubit gates and crosstalk cancellation.** Without crosstalk cancellation, simultaneous operation has an error rate about five times as large as isolated operation. With crosstalk cancellation, this reduces to simultaneous operation having error rates only twice that of isolated operation. . . . 21
2.5 **Error contribution of ZZ, as estimated by equation 2.7, for a few different pulse times.** Scaling with ZZ is super-linear, but sub-exponential, highlighting benefits of reducing static ZZ and gate time as much as possible. 23

3.1 **Classical picture of Cross Resonance interaction.** The CR effect can be seen between two coupled oscillators, with the control oscillator being nonlinear, i.e. having a state-dependent frequency. **(a)** With the control oscillator in state "0", the drive transmitted to the target is attenuated. **(b)** When the control in "1", the amplitude is attenuated less. **(c)** Depiction of the control oscillator filtering the amplitude transmitted to the target oscillator. The state-dependent frequency of the control oscillator results in different filtered target drive amplitudes. See text for further description. 26
3.2 **CR drive diagram**. Transmons are detuned by $\Delta = \omega_c - \omega_t$ indicated by the black dashed line, and coupled via exchange coupling J. They are driven simultaneously at frequency $\omega_d = \omega_t$ with amplitudes ε_c, ε_t. The bare CR drive only drives the control ($\varepsilon_t = 0$), however a local target drive can be added intentionally, or via classical crosstalk, and is included in the analysis. 29
3.3 **Cross Resonance Dynamics and Hamiltonian tomography.** **(a)** Pulse sequence for Hamiltonian tomography. The control qubit is first prepared in either $|0\rangle$ or $|1\rangle$, and then the CR drive at frequency ω_t is applied to the control qubit. Then, the target qubit Bloch vector is measured via state tomography. **(b)** Example CR dynamics and fit according to the model in equation 3.8. The top three panels are the target qubit Bloch vector components. The second from the bottom plot shows entanglement measure R versus time. At the bottom, the control-qubit $\langle Z \rangle$ as a function of time. Small oscillations are present. **(c)** Target qubit dynamics on the Bloch sphere, conditioned on the control qubit state. . . 32
3.4 **Hamiltonian tomography while varying CR drive ε** **(a)** Hamiltonian terms as a function of $|\varepsilon|$. The presence of higher levels, and microwave crosstalk, results in a nonzero IX term. A local Q_t drive is also applied, offsetting the local terms. **(b)** Varying CR drive phase displays control over the rotational axis of the local and entangling terms. The ZX and IX terms are theoretically opposite in sign, however classical crosstalk in the device modifies this [54]. The IZ and ZZ terms are insensitive to the CR drive phase. The dashed line in the CR amplitude (phase) sweep indicates the fixed CR amplitude (phase) used in the corresponding CR phase (amplitude) sweep. 34

3.5 **Control qubit leakage. (a)** Q_c $|2\rangle$ population dynamics when Q_c is prepared in $|0\rangle$ during the Hamiltonian tomography experiment. Oscillations become clear at larger drive amplitudes $|\varepsilon|$. **(b)** Same as **(a)**, except Q_c is prepared in $|1\rangle$. **(c)** Time-averaged $|2\rangle$ population for varying $|\varepsilon_c|$, showing increasing leakage with drive amplitude. 35
3.6 **CR DRAG Scheme (a)** The CR drive is near-resonant to the control qubit two-photon $|0\rangle \leftrightarrow |2\rangle$ transition (dashed line) and the $|1\rangle \leftrightarrow |2\rangle$ transition. **(b)** Fourier spectrum of CR pulse with carrier frequency at the target qubit frequency (envelope in inset), for a few different DRAG α coefficients. The spectral weight at the leakage transitions (red vertical lines) is modified by α. 36
3.7 **CR leakage DRAG α sweep (a)** Q_c state populations as a function of the CR Rabi pulse duration, with $\alpha = -2.0$ and Q_c initialized in $|0\rangle$. Significant oscillations are seen between $|2\rangle$ and $|0\rangle$, indicating strong driving of the two-photon $|0\rangle \leftrightarrow |2\rangle$ transition. **(b)** Leakage L_{Rabi} as a function of DRAG α, when Q_c is prepared in $|0\rangle$ and $|1\rangle$. Leakage from each transition is not simultaneously minimized when tuning α. 37
3.8 **Minimizing two leakage channels with 2D DRAG.** Left is the leakage from the two-photon $|0\rangle_c |2\rangle_c$ transition. Middle is $|1\rangle_c |2\rangle_c$ leakage, and right is the sum of both. 38
3.9 **CNOT gate time sweep, Hamiltonian tomography (a)** Hamiltonian rates of calibrated CNOT gates as a function of pulse time τ_p without 2D DRAG leakage calibration, and **(b)** with 2D DRAG leakage calibration. **(c)** Target qubit dynamics for $\tau_p = 112\,\text{ns}$ in the $X - Z$ Bloch plane without DRAG and **(d)** with DRAG. 41
3.10 **Direct CNOT gate time sweep, leakage benchmarking (a)** Two-qubit RB survival probability decay curves for $\tau_p = 112\,\text{ns}$, with and without DRAG correction. See the main text for description of survival probability. Decays are comparable, indicating similar gate performance. **(b)** Clifford process fidelity e_F from RB experiments for varying τ_p. **(c)** Leakage RB decay curves at $\tau_p = 112\,\text{ns}$, showing reduced decay and steady-state Q_c $|2\rangle$ population when DRAG is used. **(d)** LRB leakage rate L_1 while varying τ_p. Leakage is robustly reduced with DRAG. 43

3.11 **Benchmarking Direct CNOT Errors (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved intrinsic CR gate. From reference and interleaved decay parameters $p_{RB} = 0.972(1)$ and $p_{IRB} = 963(2)$, respectively, we extract a CR process error $e_g = 7(3) \times 10^{-3}$ (see equation A.7). **(b)** Purity benchmarking shows a unitary (stochastic) $e_U = 11.308(1) \times 10^{-3}$ ($e_S = 15.2925(1) \times 10^{-3}$) error component to the total Clifford process error $e_F^C = 26.600(1) \times 10^{-3}$. **(c)** Leakage Randomized Benchmarking results, obtained by monitoring the $|2\rangle$ state of each transmon when performing RB. The leakage per Clifford L_1 are estimated to be $L_1 = 3.4(4) \times 10^{-3}$ ($0.11(5) \times 10^{-3}$) for Q_c (Q_t). **(d)** Repeated IRB measurements to quantify gate calibration and benchmarking stability. Note that the y-axis is average gate error $r = \frac{d-1}{d} e_F$, where e_F is the process infidelity of the protocol. The statistical average gate process error is $11.3(3) \times 10^{-3}$. 45

3.12 **Cycle Benchmarking the Direct CNOT gate (a)** Exponential decays of the survival probabilities of the Pauli basis preparation states for the direct CNOT gate. **(b)** Process infidelities extracted from the exponential decays, for the case with the direct CNOT cycle (gold) and the identity cycle (blue). See A.1 for more information on Cycle Benchmarking. 47

3.13 **CR amplitude calibration experiment with error-amplification. (a)** Pulse sequence for the experiment. The CR pulse is fixed in time, with the pulse amplitude swept. This is repeated for different numbers of pulse repetitions n. **(b)** Experimental data for several values of n. **(c)** Fitted optimal value, with uncertainties, as a function of pulse repetitions n. 48

3.14 **Closed-loop optimization for calibrating local rotations to realize a CNOT from the CR gate.** The CNOT gate is parameterized by the circuit shown in the top left, where $R_T(\vec{\varphi}_2)$ is an arbitrary single-qubit rotation and the angles for all gates are angles for virtual-Z gates, which have nearly perfect fidelity [41]. The CNOT calibration circuit is inserted into a loss function circuit, such as a Bell state fidelity measurement shown on the right, and a classical optimizer minimizes the loss function in a closed-loop fashion. An example minimization of the loss function using the CMA-ES optimizer [71] is shown at the bottom left, and the parameter convergences are shown at the right. The points in the right plots indicate different parameter trials tested for each optimization step. 50

3.15 **Compiling general circuits with intrinsic CR gates.** General circuits, and particularly randomly-compiled circuits [39, 70],consist of alternating rounds of single- and two-qubit gates. The single-qubit gates R are arbitrary single-qubit rotations, which can be decomposed into three arbitrary-angle Z gates with X_{90} gates in between $Z_{\phi_1}X_{90}Z_{\phi_2}X_{90}Z_{\phi_3}$. Therefore, the single-qubit gates within the CR-based CNOT calibration, shown in the middle step of the compilation, can be folded into the other single-qubit gates without changing the depth of the circuit. An example U_{CR} is shown on the right, with information output from the Keysight/Quantum Benchmark TrueQ software [65]. 51

3.16 **Intrinsic CR Benchmarking (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved intrinsic CR gate. **(b)** Cycle Benchmarking results. The error rate e_i is obtained for the prepared Pauli eigenstate i from the exponential decay of sequence fidelity p_i with increasing sequence length. A larger error rate for a given Pauli eigenstate indicates errors in the cycle that do not commute with the compiled Pauli term. **(c)** Leakage Randomized Benchmarking. By monitoring the $|2\rangle$ state of each transmon when running an IRB experiment, the reference and interleaved $|2\rangle$ state population data are fit to an exponential model to extract the leakage-per-gate for each transmon. **(d)** Purity Benchmarking distinguishes coherent from stochastic errors by measuring the decay of the purity (main plot) of the two-qubit density matrix by performing state tomography for each random Clifford RB sequence, and comparing the purity decay to the RB decay. Inset: breakdown of Clifford process infidelity between coherent and stochastic contributions. 52

3.17 **Comparing Performance of direct CNOT and intrinsic CR gates on sample B.** The points left of the left-most dashed vertical line are from the two-qubit RB experimental suite, comprised of RB, IRB, XRB, and LRB. The middle points between the two dashed vertical lines are from CB, and the right-most points are from single-qubit RB experiments, both isolated and simultaneous [45]. 55

4.1 **Drive scheme for the Stark-induced ZZ interaction**. Transmons are detuned by $\Delta = \omega_c - \omega_t$ indicated by the black dashed line, and coupled via exchange coupling J. They are driven simultaneously with amplitudes ε_c, ε_t between frequencies ω_t and $\omega_c + \eta_c$, indicated with purple dashed lines. The simultaneous driving introduces conditional Stark shifts, i.e., a ZZ interaction. 58

4.2 ZZ **Numerical Simulations (a)** ζ versus ω_d and global drive amplitude $|\varepsilon| = |\varepsilon_t| = |\varepsilon_c|$, with $\varphi_d = \pi$. Driving in the regions $\omega_c^{(12)} < \omega_d < \omega_t$ and $\omega_t < \omega_d < \omega_c$ show consistently larger areas of enhanced ZZ interaction. Driving below $\omega_c^{(12)}$ or above ω_c, the ZZ enhancement is reduced. There are many resonances visible in the simulation. **(b)** ζ versus φ_d and $|\varepsilon|$, with $\omega_d = \omega_t - 40\,\mathrm{MHz}$. **(c)** ζ versus ε_c and ε_t, with $\varphi_d = \pi$, $\omega_d = \omega_t - 40\,\mathrm{MHz}$. 62

4.3 ZZ **as a function of drive parameters, crosstalk (a)** The control and target drive fields ε_c, ε_t are complex combinations of the drive line amplitudes A_c, A_t, mixed via the microwave crosstalk matrix C. **(b)** ZZ versus relative drive phase φ_d, for several overall drive amplitudes $|A| = |A_c| = |A_t|$. Field amplitudes in the color-bar are determined by fitting the data (dots) to numerical simulations including crosstalk (solid lines). With no crosstalk (dashed lines) the predicted ζ is symmetric about $\varphi_d = 0$ and ζ_0. Error bars indicate Ramsey frequency fit uncertainty. **(c)** ZZ versus target drive amplitude A_t for several control drive amplitudes A_c (color-bar). 63

4.4 **CZ Gate Calibration (a)** Pulse sequence for calibrating the amplitude and frequency of the CZ pulse. The target is prepared in superposition, followed by the CZ pulse, and then tomographic pulses T are applied to measure the target qubit Bloch vector r_i for each control qubit state $|i\rangle \in \{0, 1\}$. The global pulse amplitude and frequency are calibrated by selecting parameters that maximize entanglement measure R. **(b)** R as a function of the CZ gate amplitude A and drive detuning from the target $(\omega_d - \omega_t)$. There is a band of frequencies where R is maximal to realize the CZ gate. There is also a pulse amplitude around $A = 0.1$ where R is minimal, corresponding to ZZ cancellation. 65

4.5 **Local Z Gate Calibration (a)** Quantum circuit for compiling a CZ gate using the entangling CZ pulse and local Z corrections. **(b)** Local phase calibration experiment. The control qubit is prepared along the x-axis of the Bloch sphere $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and measured in the x-basis via $Y\left(\frac{\pi}{2}\right), Y\left(-\frac{\pi}{2}\right)$ gates, respectively. **(c)** The local phase ϕ_{ZI} is swept while measuring $\langle ZI\rangle$ for both target preparation states $|0\rangle$ (T0) and $|1\rangle$ (T1). The value of ϕ_{ZI} is calibrated when each input state is mapped to the correct output state, i.e., $|+, 0\rangle \to |0, 0\rangle$ and $|+, 1\rangle \to |11\rangle$. This is analogously done for the target qubit local phase ϕ_{IZ}. . . 66

4.6 **Benchmarking Results (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved CZ gate. **(b)** Cycle Benchmarking results. The error rate e_i is obtained for the prepared Pauli eigenstate i from the exponential decay of sequence fidelity p_i with increasing sequence length. A larger error rate for a given Pauli eigenstate indicates errors in the cycle that do not commute with the compiled Pauli term. The process infidelity is the error averaged across Pauli terms. **(c)** Leakage Randomized Benchmarking. By monitoring the $|2\rangle$ state of each transmon when running an IRB experiment, the reference and interleaved $|2\rangle$ state population data are fit to an exponential model to extract the leakage-per-gate for each transmon. **(d)** Purity Benchmarking distinguishes coherent from stochastic errors by measuring the decay of the purity (main plot) of the two-qubit density matrix by performing state tomography for each random Clifford RB sequence, and comparing the purity decay to the RB decay. Inset: breakdown of Clifford process infidelity between coherent and stochastic contributions. 68

4.7 **Coherence versus Drive Amplitude**. **(a)** Qubit lifetimes T_1 as a function of CZ pulse amplitude. No visible trend is observed. **(b)** Measurements of T_2^{echo} versus CZ drive amplitude for both qubits used in the gate in the main text. Error bars are uncertainties in the fit. The coherence of the sample is reduced at larger CZ pulse amplitudes. 70

4.8 **Comparison of the direct CNOT, intrinsic CR, and CZ gates calibrated on Sample B.** 71

A.1 **Benchmarking circuits (a)** Isolated single-qubit Randomized Benchmarking **(b)** Simultaneous single-qubit Randomized Benchmarking **(c)** Two-qubit Randomized Benchmarking, with possible leakage detection with $|2\rangle$-state readout **(d)** Interleaved Two-qubit RB, with similar option for quantifying leakage as RB**(e)** Extended / Purity RB. Using state tomography to measure decay of purity, i.e. 'unitarity' **(f)** Cycle Benchmarking, a more scalable, precise, and measure for general cycle fidelity as compared to Clifford RB. 80

List of Tables

3.1 Sample parameters for the two qubit pairs studied. 40
3.2 Hamiltonian rates for direct CNOT gates on Sample A (B) (circuit parameters in Table 3.1) with $\tau_p = 144\,\text{ns}$ ($\tau_p = 147\,\text{ns}$), and the resulting Clifford process fidelity e_F^C measured by RB. The e_F^C quoted for sample A corresponds to the results plotted in Figure 3.9 (b). The result for sample B corresponds to an experiment like that shown in Figure 3.11 (a). 44
3.3 Names, symbols, and descriptions of benchmarking metrics used in this thesis. Similar metrics are reported but obtained via CB, however their interpretation is the same. Subsets for isolated and simultaneous single-qubit Clifford RB are reported. See section A.1 for detailed descriptions of these parameters. 53

A.1 Names, symbols, and relationships of various benchmarking parameters used in this thesis. 81

Acknowledgments

It has been a true privilege to pursue my PhD in the Quantum Nanoelectronics Laboratory. I am extremely grateful for the guidance and leadership of my advisor, Professor Irfan Siddiqi, who exemplified the unique blend of scientific inquisitiveness and practical knowledge needed to perform effective research in this field. His teachings on topics ranging from superconducting weak links to refrigerator compressor maintenance provided thorough enrichment in my graduate studies. Irfan's ability to bring together a wonderful team of researchers fostered an amazing research environment in QNL, and I feel fortunate to have been a part of the team.

I would like to thank my committee members, Professors K. Birgitta Whaley and John Clarke, for devoting time to provide feedback on my thesis, and to Professor Hartmut Häffner for serving on my qualifying exam committee.

Over the course of my tenure in the Quantum Nanoelectronics Laboratory, I've had the pleasure to learn from many stellar colleagues. I'd like to thank Vinay Ramasesh for his mentorship in my early graduate school days, when the fabrication equipment was still in Birge. Vinay always brought passion and conviviality to the lab, which eased even the most tedious aspects of research. I also thank Kevin O'Brien and James Colless for their guidance and encouragement to explore research topics when I first joined the lab. I am incredibly thankful to John Mark Kreikebaum for fabricating the devices used in this thesis, and for teaching me fabrication best practices. I thank Will Livingston, for sharing his deep understanding of cryogenics, electronics, and weak measurement. I also thank Dar Dahlen for spending many hours discussing software and calibration routines, which has enabled much of the research produced in this thesis. I give special thanks to Ravi Naik, who, through his exceptional guidance throughout the two-qubit gate projects, helped me immensely to grow as a researcher. I also thank Alexis Morvan for his mentorship, and for exemplifying a thoroughness and enthusiasm for science that was truly contagious.

Beyond the purely research-oriented aspects, I am grateful for all of the fun times we shared as well. Taking Norman Yao's Quantum Optics course together with Vinay, Will, Leigh Martin, and Marie Lu made it one of the most fun and valuable courses in my graduate studies. After many movie nights, game nights, taco Tuesdays, group lunches, gym sessions, and coffee breaks, I am grateful for many colleagues that I can also call friends. I have the fortune to thank many folks for making QNL a friendly and fun place to learn and do science: Andrew Eddins, Sydney Schreppler, John Mark Kreikebaum, Marie Lu, Leigh Martin, Will Livingston, Dar Dahlen, Vinay Ramasesh, Kevin O'Brien, James Colless, Machiel Blok, Emmanuel Flurin, Ravi Naik, Alexis Morvan, Jean-Loupe Ville, Archan Banerjee, Trevor Chistolini, Larry Chen, Akel Hashim, Kasra Nowrouzi, Brian Marinelli, Noah Stevenson, Zahra Pedramrazi, Gerwin Koolstra, Karthik Siva, Long Nguyen, Noah Goss, Linus Kim, Christian Juenger, Wim Lavrijsen, and Costin Iancu.

I'm grateful to everyone who has kept the physics department running smoothly, in particular Anthony Vitan, Anne Takizawa, Joelle Miles, Carlos Bustamante, Stephen Raffel, and Jesse Lopez.

My studies would not have been possible without my support system beyond the lab. I'd like to thank my wife Sumana for her unending support over the years, from near and far. I am grateful for my friends Bryce Kobrin, Nick Dale, Andreas Biekert, and Nikola Maksimovic, who made my time in Berkeley truly an adventure. Finally, I express my deepest gratitude to my parents, Ken and Sharalyn, for always supporting my interests and cheering for my success.

Chapter 1

Introduction

Quantum computation is the idea of encoding data, conventionally represented as bit strings, into a quantum system, typically represented by *qubits*, or quantum bits, and performing logical operations on and measurements of the outcomes, which are based on the laws of quantum mechanics. Richard Feynman and Yuri Manin independently proposed the idea of quantum computers in the 1980s. Since then, progress on the theoretical side showed potentially useful calculations that a quantum computer could perform significantly more efficiently than a conventional, or "classical", computer, such as factoring integers [1], unstructured search [2], and simulating quantum systems [3, 4].

The theoretical possibility of quantum computational advantage spurred intense research toward building a quantum computer, however the challenge of decoherence, where environmental noise dissipates the quantum properties, (i.e. entanglement and superposition) of a quantum state, cast doubt on the possibility of realizing a quantum computer. Fortunately, in 1995, Peter Shor proposed a way around decoherence, by redundantly encoding logical qubits into multiple physical qubits in what's called a quantum error correction (QEC) code [5]. With QEC, the promise of fault-tolerant quantum computation was again seen as realizable, and today researchers are assessing in greater and greater detail the experimental resources needed to bring about quantum computational advantage.

One of the most promising QEC codes being studied for implementation with superconducting circuits is the surface code [6, 7]. This is because it only requires nearest-neighbor connectivity between qubits, and has one of the highest error thresholds of about 1%, meaning that if operations on physical qubits (e.g. initialization, gates, storage, and measurement) have error rates below 1%, corresponding errors on logical qubits will be suppressed as the size of the code (i.e. the number of physical qubits per logical qubit) is increased. The error threshold of a QEC code sets an important minimum performance measure to consider when designing quantum processors and benchmarking their performance.

In superconducting circuits, and also other platforms, two-qubit gates are one of the highest-error operations. Therefore, studying and mitigating the error processes that occur during two-qubit gates is crucial to maximizing quantum processor performance. In this thesis, calibration and error budgeting of entangling gates based on two different microwave-

activated interactions is presented. These interactions are realized on a quantum processor consisting of eight single-junction transmon qubits with fixed coupling in a ring geometry. First, I present a study of the cross-resonance (CR) gate [8], which is one of the most popular entangling gates on single-junction, fixed coupling architectures. Then I present a novel interaction that realizes a controlled-Z gate based on tunable cross-Kerr coupling generated by simultaneous off-resonant driving of the transmons, published in [9]. Besides being useful for generating entanglement, this interaction has the additional benefit that it enables cancellation of static coupling between qubits, a significant error source for architectures with fixed qubit frequencies and fixed coupling.

The structure of this thesis is as follows. In the remainder of this chapter, I will give background of superconducting quantum computing that is sufficient for understanding the main results of this thesis, including an introduction of circuit quantum electrodynamics (cQED) and the experimental quantum processing unit (QPU) used in presented experiments. In Chapter 2, I describe procedures single-qubit gate calibration, microwave crosstalk mitigation on the device, and error budgeting single qubit gates. In Chapter 3, experiments assessing the error budget of cross resonance gates on our hardware are presented, including the implementation of leakage-cancellation techniques on the CR gate, and novel calibration schemes using modern compilation and closed-loop optimization. Chapter 4 details the microwave-activated tunable cross-Kerr (ZZ) coupling, and experiments implementing a CZ gate with this interaction. Future directions are then discussed.

1.1 Circuit Quantum Electrodynamics

1.1.1 A brief history of superconducting quantum computing

While a variety different technologies are being pursued to build quantum computers, including trapped ions, photonic systems, and semiconductor quantum dots, a prominent technology being pursued in academia and industry is superconducting quantum circuits. Superconducting circuits rely on a particular circuit element called a Josephson junction [10], which is simply a thin insulating barrier between two superconducting metals. In 1985, energy-level quantization was first observed in a Josephson junction circuit here at UC Berkeley by Professor John Clarke, Michel Devoret, and John Martinis [11], answering Anthony Leggett's question about whether a degree of freedom comprised of so many atoms can indeed behave quantum-mechanically[1]. This observation showed that superconducting circuits based on Josephson junctions could behave as 'artificial atoms', and subsequently direct evidence of their quantum behavior was shown with the observation of Rabi oscillations in a superconducting charge qubit in 1999 [12], and Ramsey oscillations in a quantronium qubit in 2002 [13]. Milestones continued to be established in superconducting quantum circuits, including strong coupling between a single photon and superconducting charge qubit in 2004 [14], heralding the field of circuit quantum electrodynamics (cQED) [15] (as an analogy

[1]This story is from Steven Girvin's talk, *A Brief History of Superconducting Quantum Computing*

to the cavity QED of optical photons and true atoms) as an architecture for quantum computation. Shortly after came the proposal [16] and realization [17] of the transmon qubit, which is one of the most commonly used superconducting qubits at this time of writing, and is the superconducting qubit used in this thesis.

The framework for cQED is described in the next sections, however see refs. [18, 19] for more details.

1.1.2 The quantum LC circuit

To begin, we consider the quantization of an LC circuit. The Hamiltonian of a circuit with capacitance C and inductance L is

$$H_{LC} = \frac{Q^2}{2C} + \frac{\Phi^2}{2L}, \tag{1.1}$$

where Q is the charge on the capacitor and Φ is the flux threading the inductor. This harmonic Hamiltonian has resonance frequency $\omega_r = 1/\sqrt{LC}$, and impedance $Z_0 = \sqrt{L/C}$. This is analogous to the simple harmonic oscillator (SHO) Hamiltonian, with particle position x and momentum p $H_{\mathrm{SHO}} = p^2/2m + kx^2/2$. While either Q or Φ can be mapped to position x in this analogy, typically Φ is chosen as the 'coordinate' variable, and thus C is the mass and L^{-1} is the spring constant. Quantization then amounts to promoting the conjugate variables to operators, with canonical commutation relation

$$[\Phi, Q] = i\hbar. \tag{1.2}$$

Second quantization then follows with the definition of bosonic creation and annihilation operators $a^\dagger$ and a such that $\Phi = \Phi_{\mathrm{zpf}}(a + a^\dagger)$ and $Q = i_{\mathrm{zpf}}(a - a^\dagger)$, with $\Phi_{\mathrm{zpf}} = \sqrt{\hbar Z_0/2}$,$Q_{\mathrm{zpf}} = \sqrt{\hbar/2Z_0}$ being the flux and charge zero-point fluctuations. Rewriting equation 1.1 in terms of a, $a^\dagger$ we have

$$H_{LC} = \hbar\omega_r \left(a^\dagger a + 1/2\right). \tag{1.3}$$

For an LC circuit to behave in a quantum mechanical manner, two conditions must be met: the circuit must be very low loss, such that the spectral linewidth of the energy states is less than their separation, and the ambient temperature $k_B T \ll \omega_r$. However, since the energy spectrum is linear, resolving or addressing individual transitions is not possible with a quantum LC circuit alone. A nonlinear element is needed to encode a qubit in the Fock states of the oscillator.

1.1.3 The Josephson junction

To encode a qubit in a quantum circuit then, we need a lossless circuit element that is nonlinear, and the Josephson junction meets these requirements. A typical Josephson junction

consists of two superconducting leads, separated by a thin insulating barrier. Remarkably, a dissipationless current flows across the junction according to the DC Josephson effect

$$I = I_c \sin\varphi, \tag{1.4}$$

where I_c is the junction critical current above which lossy current flows, and φ is the phase difference between the superconducting wavefunctions on either side of the junction [20]. The critical current I_c depends on the area of the junction, the thickness of the junction barrier, and the energy gap of the superconductor via the Ambegaokar-Baratoff relations [21]. The voltage across the junction is related to φ in a manner analogous to Faraday's law

$$V = \frac{\Phi_0}{2\pi}\dot{\varphi}, \tag{1.5}$$

where $\Phi_0 = h/2e$ is the magnetic flux quantum. Equation 1.5 describes how a DC voltage across the junction leads to an AC current, creating a voltage-to-frequency converter that has led to using Josephson junction arrays as a tool to establish the SI definition of the volt [22].

From these Josephson relations, we can derive the associated energy of the junction

$$U_J(t) = \int_{-\infty}^{t} IV dt' \tag{1.6}$$

$$= \frac{\Phi_0 I_c}{2\pi}(1 - \cos\varphi) \tag{1.7}$$

$$= E_J(1 - \cos\varphi). \tag{1.8}$$

Since the potential energy is not quadratic, the Josephson junction introduces a nonlinearity to the energy spectrum that is needed for constructing 'artificial atoms' used to encode qubits.

1.1.4 The transmon qubit

The simplest and most common superconducting qubit is the transmon qubit [16], which is realized by replacing the inductor in an LC circuit with a Josephson junction (see Figure 1.1). Defining the generalized 'flux' variable as $\Phi = \int_{-\infty}^{t} V dt'$, The Hamiltonian for the transmon is then given by

$$H = \frac{(Q - Q_g)^2}{2C} - E_J \cos\frac{2\pi\Phi}{\Phi_0} \tag{1.9}$$

$$= 4E_c(n - n_g)^2 - E_J\cos\varphi, \tag{1.10}$$

with $Q = 2en$ being the charge on the capacitor with number of cooper pairs n, and possible offset charge Q_g allowed and defining $E_c = e^2/2C$. Note that C in this case accounts for both the intrinsic capacitance of the junction and the additional shunting capacitance. The

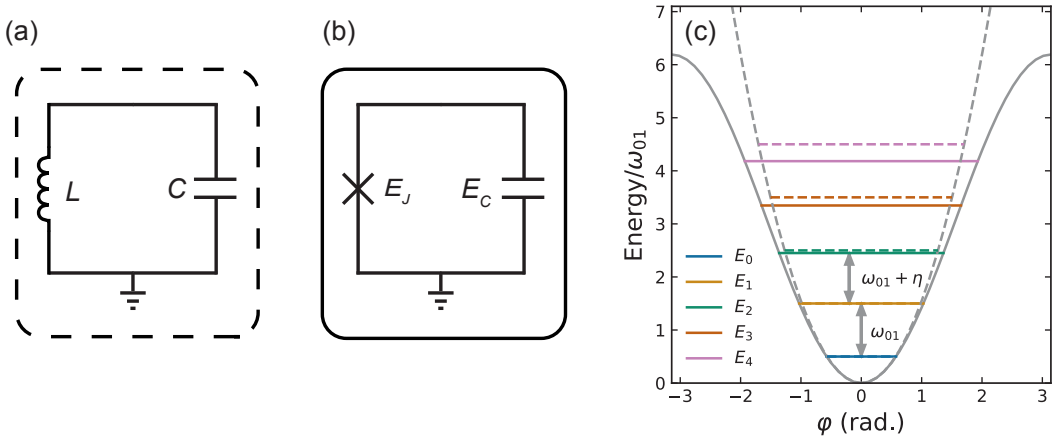


Figure 1.1: **The transmon qubit as an anharmonic oscillator.** (a) Standard LC circuit (b) transmon qubit, replacing the inductor with a Josephson junction (c) comparison of energy spectrum for an LC circuit (dashed line) and a transmon (solid line), with the potential energy in gray and the eigenenergies denoted in the legend. The anharmonicity η is about 5% of the lowest energy transition $\omega_{01} = E_1 - E_0$. for typical transmons in our lab, corresponding to an $E_J/E_c = 70$.

transmon qubit operates in the regime $E_J/E_c \gg 1$ [16], because in this regime charge noise is exponentially suppressed, leading to higher qubit coherence times than for smaller E_J/E_c circuits, known as Cooper Pair Boxes. In the limit $E_J/E_c \gg 1$, the phase difference φ is well localized in the bottom of the cosine potential well around $\varphi = 0$, and from the Hamiltonian, the system behaves as a quantum particle in a cosine-potential well. For $E_J/E_c \gg 1$, the kinetic energy E_c is much lower than the well depth E_J, meaning that the dynamics can be approximated in a perturbative fashion. As such, a typical approximation for the transmon is the Duffing oscillator, which is a harmonic oscillator with an additional quartic term

$$H = \hbar\omega_q a^\dagger a + \hbar\frac{\eta}{2} a^\dagger a(a^\dagger a - 1), \tag{1.11}$$

where $\hbar\omega_q \approx \sqrt{8E_cE_J} - E_c$ is the qubit frequency and $\eta \approx -E_c/\hbar$ is the anharmonicity. In the Duffing oscillator approximation, these values fully determine the transmon Hamiltonian. Note that these approximations give a sense for how to relate circuit parameters to measured transmon frequencies, however higher order or numerical methods should be used when estimating E_J and E_c from measured quantities ω_q and η. Typical values of ω_q and η are $2\pi \times 5\,\text{MHz}$ and $2\pi \times 300\,\text{MHz}$.

1.1.5 Qubit-resonator coupling

The combination of quantum LC circuits and the nonlinear quantum circuits comprised of one or more Josephson junctions establishes a superconducting quantum computing toolbox known as circuit QED [15], with terminology and methods analogous to cavity QED, wherein optical cavities are coupled to true atoms for studying quantum interactions between light and matter. Dispersively coupling a transmon to a resonator enables measurement of the transmon state, and dispersively coupling multiple transmons to a shared resonator enables interactions between the transmons for entangling gates.

Coupling a resonator and a transmon is often realized via a coupling capacitance, which can be thought of as electric dipole coupling that is written as

$$H_{int} = -\hbar g(a - a^\dagger)(b - b^\dagger) \tag{1.12}$$

$$\approx -\hbar g(ab^\dagger + a^\dagger b) \tag{1.13}$$

for bosonic operator a of the transmon and b for the cavity. The coupling strength g is given by [18]

$$g = \omega_r \frac{C_g}{C} \left(\frac{E_J}{2E_c}\right)^{1/4} \sqrt{\frac{\pi Z_0}{R_K}}, \tag{1.14}$$

where ω_r is the resonator angular frequency, C_g is the coupling capacitance, C is the transmon capacitance, Z_0 is the resonator impedance, and $R_K = h/e^2$ is the resistance quantum. Equation 1.13 invokes the rotating wave approximation, as terms like $a^\dagger a^\dagger$ that add or remove two photons are perturbatively suppressed by factors of $g/|\omega_q + \omega_r|$, whereas exchange type terms like $ab^\dagger$ are suppressed by a much smaller factor of g/Δ (assuming $|\Delta| \ll |\omega_q + \omega_r|$) and thus terms which remove or add photons to the unperturbed (uncoupled) Hamiltonian have negligible effect on the dynamics compared to the exchange type terms. Thus, the coupled resonator-qubit Hamiltonian (including the uncoupled terms) looks like (letting $\hbar = 1$ from now on)

$$H = \omega_q a^\dagger a + \frac{\eta}{2} a^\dagger a^\dagger a a + \omega_r b^\dagger b + g\left(a^\dagger b + a b^\dagger\right). \tag{1.15}$$

So far, we have not considered the detuning between the resonator and the transmon. In the so-called *dispersive* limit of $\Delta = \omega_q - \omega_r \gg g$, then applying perturbation theory, the energies of the coupled states are shifted and the Hamiltonian in qubit subspace of the transmon is

$$H_D = \omega_q'(1 - \sigma_z)/2 + \omega_r' b^\dagger b - \chi \sigma_z b^\dagger b, \tag{1.16}$$

where the cross-Kerr coupling χ between cavity and resonator is given by

$$\chi = -\frac{g^2 \eta}{\Delta(\Delta + \eta)}, \tag{1.17}$$

and the frequencies are renormalized (i.e. Lamb shifts) as indicated by the primes.

Note that beyond the qubit truncation for the transmon, the dispersive shift χ is different for each higher transmon transition. Additionally, if the transmon were a true qubit, this

corresponds to $\eta \to \infty$, and in this limit $\chi = g^2/\Delta$, whereas for two resonators, this dispersive shift is 0. Finally, if both systems were ideal qubits, then there again would be zero dispersive shift from the coupling g in the dispersive limit.

The dispersive Hamiltonian in equation 1.16 is the basis for non-destructive (or QND) measurement of a qubit in the σ_z basis by measuring the frequency of a readout resonator [15].

1.1.6 Coupling transmons

To realize entangling interactions between transmons, they are typically capacitively coupled either by bringing their shunting capacitor pads directly in proximity to each other, or via a mutual dispersive coupling to a resonator. In the case of the Trailblazer chip that is used in this thesis, a coupling resonator is used. While some approaches generate entanglement between qubits by bringing the $|01\rangle$ and $|10\rangle$ into resonance via flux-tunable qubits or with parametrically driven tunable couplers, on the Trailblazer chip, the qubits are coupled in a 'straddling-regime' configuration, where $|\Delta| = |\omega_c - \omega_t| < \eta_c, \eta_t$ for qubit pair Q_c, Q_t. In this regime, entanglement can be generated using the Cross-resonance interaction [8], as described in Chapter 3, or via a differential Stark shift interaction as described in Chapter 4. One benefit of this approach is that neither qubits nor couplers need to be flux-tunable in order to entangle the qubits. This is beneficial because flux-tunable qubits and couplers introduce additional sources of decoherence for the qubits, and so designs omitting these components tend to produce qubits with longer coherence times.

Similar to the transmon-resonator coupling, transmon-transmon exchange coupling can be described by the Hamiltonian

$$H = H_{q1} + H_{q2} + J(a_1^\dagger a_2 + a_1 a_2^\dagger), \tag{1.18}$$

where H_{qi} is the Duffing oscillator approximation for transmon i, with bosonic annihilation operator a_i. The exchange coupling J for the case of two qubits mutually dispersively coupled to a resonator is approximated by

$$J = \frac{g_1 g_2}{2}\left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2}\right), \tag{1.19}$$

where g_i is the coupling of transmon i to the coupling resonator, and $\Delta_i = \omega_{qi} - \omega_r$ is the detuning of the qubit from the coupling resonator frequency ω_r.

One consequence of equation 1.18 for detuned qubits is, due to higher levels of the transmon, the exchange coupling J results in a cross-Kerr coupling between the transmons in the dispersive limit $\Delta \gg J$. In the qubit subspace of the Hamiltonian, this results in a static $\sigma_z \otimes \sigma_z$ (also often written ZZ) interaction between the qubits, which to leading order is

$$\zeta_0 = 2J^2\left(\frac{1}{\Delta - \eta_t} - \frac{1}{\Delta + \eta_c}\right). \tag{1.20}$$

Crucially, this always-on (but not necessarily desired) interaction between the qubits scales with J^2, whereas the designed the CR interaction scales linearly with J. This sets a design

constraint on the coupling J to be small enough such that it isn't prohibitive for single-qubit gates, but large enough to achieve reasonable two-qubit gate times on this architecture. Typically, for qubit frequencies between $4-6\,\mathrm{GHz}$, and anharmonicities between $250-300\,\mathrm{MHz}$, coupling rates J of order $2-4\,\mathrm{MHz}$ struck a balance between low ζ_0 rates and fast CR rates. However, with the interaction described in Chapter 4, ζ_0 can be cancelled, enabling larger J for two-qubit gates. Additionally, multipath coupler elements employing two coupling paths between qubits allow for reduced ζ_0 for a given effective exchange coupling [23].

1.1.7 Coupling to the environment

So far, we have been discussing quantum circuits as closed quantum systems, and how they couple to each other to enable quantum processing tasks, like qubit measurement by coupling the transmon to a readout resonator, and qubit-qubit entanglement by coupling two transmons via a coupling resonator. However, to communicate to the transmons and to the readout resonators, we need to couple them to the outside world. Important in the design of how strongly to couple components to the outside world, are tradeoffs between speed (e.g. of measurements or qubit gates) and loss channels introduced by coupling the system to the transmission lines used for control and measurement. Additionally, uncontrolled noise sources via materials imperfections, quasiparticle excitations, and even radiative sources introduce noise into quantum systems and cause decoherence. These challenges are summarized in [24], and here I focus on electromagnetic design considerations for balancing control and measurement with the decoherence channels they introduce.

Generally, stochastic noise processes on the qubit can be described as a combination of two processes via the Bloch-Redfield [25–27] picture of two-level system dynamics: longitudinal relaxation Γ_1 and transverse relaxation $\Gamma_2 = \Gamma_1/2 + \Gamma_\phi$ (including pure dephasing rate Γ_ϕ) [19]. The decay rate Γ for a qubit with frequency ω_q coupled to a transmission line with impedance Z_0 with coupling C_c, for instance, can be estimated in the same fashion as an oscillator with external coupling; by modeling the external coupling as a parallel resistance with $R = \mathrm{Re}[Z_0]$. The decay rate then amounts to $\Gamma = Z_0\omega_q^2 C_c^2/C_q$ [28], where C_q is the capacitance of the qubit. Selecting C_c to be too large would cause the qubit lifetime to be limited by emission into the control line. However, choosing a C_c that is too small would cause the practical challenge of delivering the power required to implement sufficiently fast gates without too much heat dissipation on the sample.

In addition to emission through the control line, the qubit can also decay by emitting a photon through the transmission line that the readout resonator is coupled to. To mitigate this, an additional resonator, dubbed a Purcell filter, is placed between the transmission line and the readout resonators, to reduce the density of states in the transmission line at the qubit frequency [29]. While there are various implementations of Purcell filters, a bandpass filter design [30] is used on the Trailblazer chip, enabling multiplexed readout with all readout resonators coupled to the same Purcell filter. Without a Purcell filter, the T_1 limit of the qubit would be $T_1 \leq (\Delta/g)^2/\kappa$, where $\Delta = \omega_q - \omega_r$ is the qubit-resonator detuning, g is the qubit-resonator coupling rate, and κ is the external external coupling rate of the

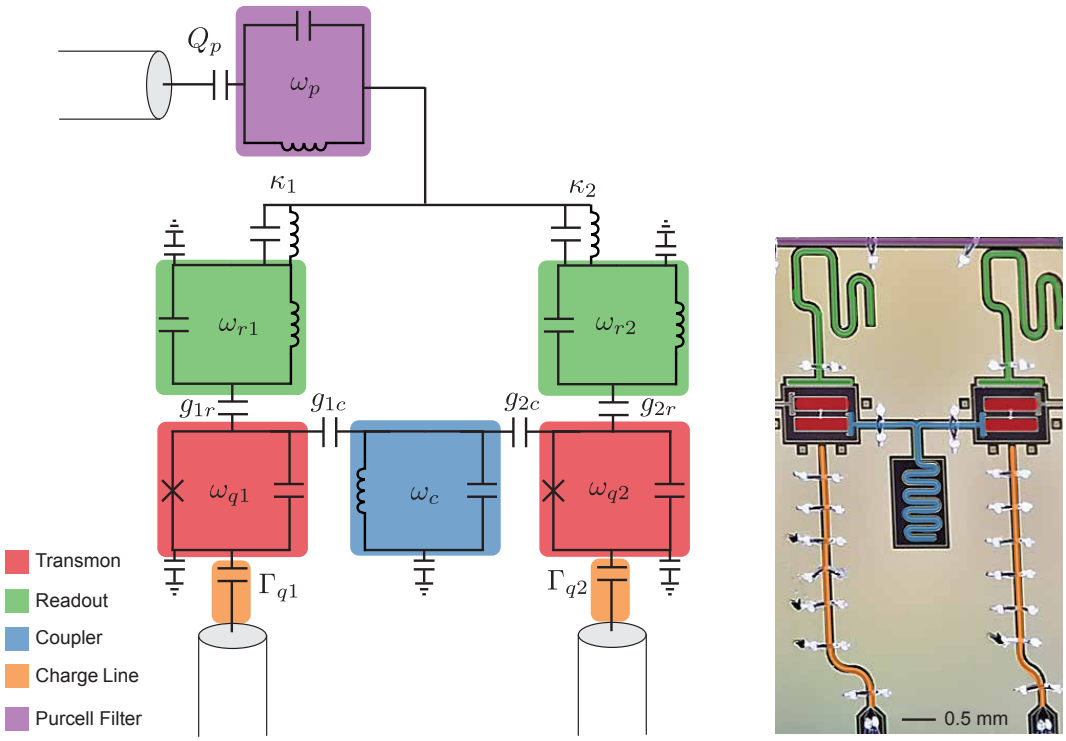


Figure 1.2: **Two qubit QPU, the 'unit cell' of the Trailblazer quantum processor**.

resonator. By including a Purcell filter with quality factor Q_p and frequency ω_p, the T_1 limit becomes [30].

$$T_1 \leq \frac{1}{\kappa}\left(\frac{\Delta}{g}\right)^2\left(\frac{\omega_r}{\omega_q}\right)\left(\frac{2\Delta}{\omega_r/Q_p}\right)^2. \tag{1.21}$$

1.2 The Trailblazer QPU

Thus far, we have summarized the essential elements for the building blocks of a quantum processing unit (QPU): encoding a qubit with a transmon, measuring the qubit by coupling it dispersively to a resonator, coupling qubits by dispersively coupling them both to a resonator, and coupling the qubit to the environment for delivery of control and measurement signals. These elements combined are shown in Figure 1.2, with a circuit diagram and a corresponding micrograph of a portion of the Trailblazer QPU used in this thesis. The full micrograph of the Trailblazer chip is shown in Figure 1.3.

The Trailblazer chip (fabrication details in [31]), first published in [32], was designed by Kevin O'Brien and developed by Kevin O'Brien and John Mark Kreikebaum, and is the beginning multi-qubit planar architecture in the group. It incorporates many standard

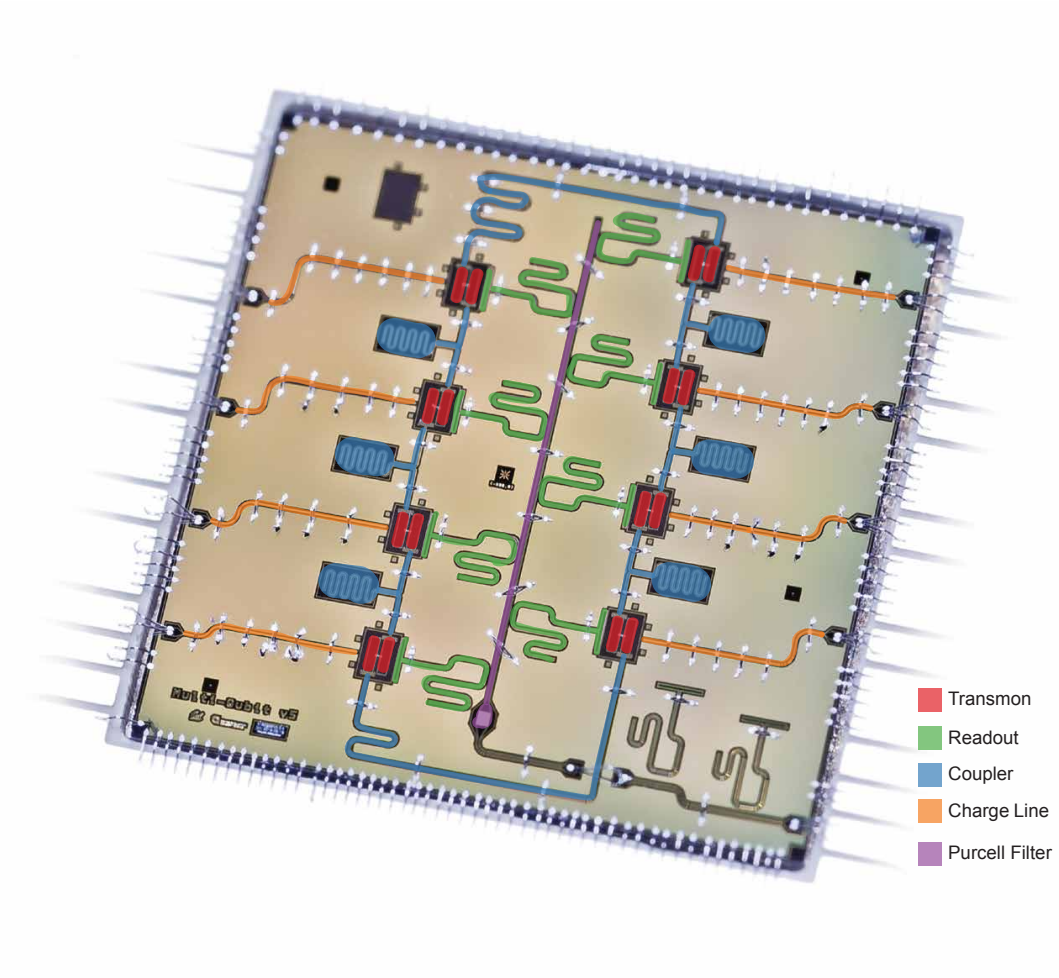


Figure 1.3: **Trailblazer quantum processor 8-qubit ring.** Photo credit: John Mark Kreikebaum [31].

planar cQED QPU design principles, with a few deliberate distinctions. I will describe the design choices as I understand them. There are eight transmons coupled in a ring geometry. The transmons were designed to be high coherence, with the qubit capacitor pads shape chosen to minimize surface participation of dielectric materials in the transmon mode volume [33]. The $E_J/E_c \sim 70$ was chosen to allow for operation of the devices as qutrits (three-level quantum systems) with sufficient coherence of the $|2\rangle$ state. The qubit frequency allocation follows a 'V' pattern, with target neighbor detunings of $\Delta/\eta \sim 0.6$, to maximize the Cross Resonance rate [34]. The average qubit frequency is around 5.5 GHz, which sits comfortably within the frequency band of $4-8$ GHz that most cryogenic microwave component and microwave electronics manufacturers can accommodate, while also being high-enough frequency to satisfy $\hbar\omega_q/k_BT \gg 1$.

The couplers are half-wave transmission line resonators. Transmission line resonators generally have higher internal quality factors over lumped-element circuits because the fields are stored less in the dielectrics and interfaces and are less concentrated. However, transmission-line resonators have higher-order modes that can cause unwanted couplings [35]. The couplers within the two columns of four transmons have coplanar-waveguide (CPW) coupling sections with a coplanar stripline (CPS) resonator at the end. This design was chosen because its higher-order modes are higher in frequency than a standard CPW resonator (determined by finite element simulation), reducing parasitic modes on the chip. The two top and bottom resonators connecting the columns are half-wave CPW resonators. The coupler frequencies are designed to be around 7 GHz to give qubit-qubit couplings $J \sim 3$ MHz. This value was chosen as a balance between fast entangling rates and static ZZ, a trade-off which is discussed often in this thesis.

The readout resonators are quarter-wave CPW resonators, with frequencies evenly spaced by 60 MHz between $\omega_{r0}/2\pi = 6.2$ GHz and $\omega_{r7}/2\pi = 6.7$ GHz. The designed qubit-resonator couplings are $g_{qr}/2\pi \sim 75$ MHz with external couplings $\kappa/2\pi \sim 1$ MHz, targeting the χ/κ ratio for optimal readout $|\chi/\kappa| = 0.5$ [36]. Each resonator is coupled via a mixed-reactance coupling to the Purcell filter that vertically divides the chip, which has frequency $\omega_p/2\pi = 6.5$ GHz in the center of the readout resonator band and external quality factor $Q_p = 10$ to reduce Purcell decay of the qubits into the readout feedline, as discussed earlier. The Purcell filter gives about a factor of 10 improvement in the Purcell limited T_1 through the readout line, from approximately 200 μs to 2 ms.

The control lines are CPW traces with thinner center trace and gap than the other elements, in order to reduce crosstalk caused by dipole radiation from the bondpads. The qubit-control line coupling, set by the separation between the open end of the control line and the qubit capacitor pad, was set (in version 5) to realize a simulated qubit quality factor $Q_q = \omega_q/\Gamma \sim 5 \times 10^6$. For a qubit with $\omega_q/2\pi = 5.5$ GHz, this corresponds to $T_1 \leq Q_q/\omega_q \approx 150\,\mu$s. In version 6 of the Trailblazer chip, these were changed to give $T_1 \leq 1$ ms.

All of these design considerations for the Trailblazer QPU position it for meeting the Divincenzo criteria for a quantum computer [37]:

1. **A scalable physical system with well characterized qubits.** The transmon qubit is the approach taken here.

2. **The ability to initialize the state of the qubits to a simple fiducial state, e.g. $|00...0\rangle$.** Initialization is done by cooling with the dilution refigerator. Residual thermal population can be either actively reset or results processed via post-selection.

3. **Long relevant decoherence times, much longer than the gate operation time.** Measured coherences are $50 \leq T_1 \leq 100\,\mu$s, with 30 ns single-qubit gates and 200 ns two-qubit gates.

4. **A universal set of quantum gates.** Arbitrary single-qubit gates are implemented with individual control lines. Two qubit gates are implemented with microwave-activated interactions described in Chapters 3 and 4. Any CNOT-like two-qubit gate, along with a Hadamard gate and a T gate, is sufficient for universal quantum computation [38].

5. **A qubit-specific measurement capability.** Qubits are measured separately with individual qubit readout resonators that are frequency-multiplexed into a single measurement line.

1.3 Thesis structure and results summary

The remaining components of this thesis focus on the task of calibrating and benchmarking single- and two-qubit gates on the Trailblazer QPU. Experiments in this thesis were performed by myself and Ravi Naik, who measured data presented in Figures 3.11, 3.12, 4.4, 4.5, and 4.6. The two measured samples were mounted to the base plate of a BlueFors XLD400 dilution refrigerator with a base temperature of 10 mK. Cryogenic and room temperature wiring is described in [32, 39]. Chapter 2 describes the single-qubit gate calibration process, classical crosstalk mitigation, and error budgeting of single qubit gates. We find that remaining simultaneous single-qubit gate errors are well budgeted by decoherence and static ZZ interactions.

Chapter 3 details investigations of the cross resonance [8] gate on our hardware, including Hamiltonian dynamics, leakage mitigation using a two-parameter DRAG pulse shape, and two calibration techniques explored to maximize gate performance, as measured with a suite of benchmarking techniques detailed in Appendix A.1. Interestingly, calibrating an 'intrinsic CR' gate, instead of a direct CNOT gate, showed reduced coherent errors, despite being 100 ns slower. Additionally, a sample with higher microwave crosstalk showed better performance than one with lower crosstalk, which is consistent with other works that suggest some microwave crosstalk can reduce overall gate error in the CR gate [34, 40].

In Chapter 4, we discuss a novel tunable cross-Kerr (ZZ) coupling, realized by driving both transmons off-resonantly. The effect is described theoretically and numerically, followed by experimental meeasurement of the tunable ZZ dynamics by varying the relative phase

between the off-resonant drives. The data agree well with the model, when crosstalk is taken into account. This tunable coupling is used to calibrate a CZ gate with fidelity 99.43%, which is found to be dominated by stochastic error sources. Notably, unlike the CR interaction, this interaction commutes with the static ZZ interaction, and therefore the CZ gate fidelity is not limited by this. We further investigate an apparent drive-induced decoherence mechanism observed with this interaction. Finally, we conclude with future directions for leveraging this interaction as the native entangling interaction on future QPUs.

Chapter 2

Single-qubit Gates and Crosstalk

Of the Divincenzo criteria described in the previous chapter, one particular prerequisite for calibrating and benchmarking two-qubit gates, is calibrating and benchmarking single-qubit gates. In this chapter, I will describe the procedure we employed to calibrate single-qubit gates on the Trailblazer chip including an important experimental consideration of classical crosstalk.

2.1 Single Qubit Gate Calibration

In our experiments, we calibrated only one single-qubit gate, the $X_{\pi/2} = e^{-i\sigma_x \pi/4}$ gate, because any single-qubit rotation (i.e. an arbitary matrix in the SU(2) group) can be parameterized via [41]

$$U(\theta, \phi, \lambda) = Z_{\phi-\pi/2} X_{\pi/2} Z_{\theta-\pi} X_{\pi/2} Z_{\lambda-\pi/2}, \tag{2.1}$$

and arbitary Z_θ gates are implemented via virtual Z gates, i.e phase changes in software. Thus, the problem of arbitary single-qubit gate calibration is reduced to simply calibrating the $X_{\pi/2}$ gate.

To understand how to calibrate a gate, it helps to examine the Hamiltonian of a driven transmon in the Duffing oscillator approximation with frequency ω_q and anharmonicity η

$$H_d = \omega_q a^\dagger a + \frac{\eta}{2} a^\dagger a^\dagger a a + \Omega(t) \left(e^{i(\omega_d t+\phi)} + e^{-i(\omega_d t+\phi)}\right) \left(a + a^\dagger\right), \tag{2.2}$$

where $\Omega(t)$ is the drive envelope at carrier frequency ω_d. In the rotating frame of the drive $a \to a e^{-i\omega_d t}$, making the rotating wave approximation, and truncating to the $|2\rangle$ state, H_d is written in matrix form as

$$H_d \approx \begin{pmatrix} 0 & \Omega e^{i\phi_d} & 0 \\ \Omega e^{-i\phi_d} & \Delta & \sqrt{2}\Omega e^{i\phi_d} \\ 0 & \sqrt{2}\Omega e^{-i\phi_d} & 2\Delta + \eta \end{pmatrix}, \tag{2.3}$$

where $\Delta = \omega_q - \omega_d$. Of course, for realizing an $X_{\pi/2}$ gate, we drive on resonance $\Delta = 0$. While we are concerned with single-qubit gates, we include the non-computational $|2\rangle$ state in the

Hamiltonian, because its presence introduces detuning (Z-type) errors in the computational subspace that need to be corrected [42–44]. In these experiments, we used 30 ns Gaussian pulse envelopes with a 5σ cutoff for our $X_{\pi/2}$ pulses. This pulse time was found to give sufficiently high fidelity single-qubit gates, though the gate fidelity as a function of pulse time was not systematically studied, it was found that moving to shorter pulses made crosstalk cancellation, as described in Section 2.2, less effective. The primary two parameters we need to calibrate are the pulse amplitude and frequency, to ensure that the pulse properly reaches the equator from the pole, and that the drive is resonant to the qubit, respectively. Next, phase errors introduced by the Stark shift of the $|1\rangle - |2\rangle$ transition need to be corrected, which is accomplished by calibrating software-defined Z gates, realized by shifting the phase of the qubit drive, called "virtual-Z" gates, on either side of the pulse $Z_\theta X_{\pi/2} Z_\theta$ [41].

After spectroscopic calibration of the qubit frequency, the calibration sequence for the $X_{\pi/2}$ gate then proceeds as follows:

1. Coarse amplitude calibration
2. Ramsey-based frequency calibration
3. Progressive fine amplitude calibration, using error-amplification
4. Phase correction calibration

Coarse amplitude calibration, depicted in Figure 2.1 (a), calibrates a $X_{\pi/2}$ pulse sufficient for Ramsey-based qubit frequency calibration. It consists simply of measuring the qubit $|1\rangle$ population (or equivalently, $\langle Z\rangle$) as a function of pulse amplitude. The point at which $\langle Z\rangle = 0$ corresponds to even superposition of $|0\rangle$ and $|1\rangle$. Once this calibration is done, we are able to measure the qubit frequency by performing Ramsey experiments, depicted in Figure 2.1 (b). Typically, a Ramsey experiment consists of a simple sweep of wait times t between two $X_{\pi/2}$ pulses, however we can improve the sensitivity by adding "artificial detuning" ν using virtual-Z gates during the wait time. This is useful in the case of a small detuning error from the predicted qubit frequency from the actual qubit frequency, where slow oscillations can be difficult to distinguish from decay when fitting. Further, by performing the experiment for multiple artificial detunings ν, the correct qubit frequency can be determined.

After the qubit frequency has been calibrated, the drive amplitude must be recalibrated as well, because the amplitude needed for a detuned drive to put the qubit on the equator of the Bloch sphere is different than a resonant drive. Further, the precision of the optimal drive amplitude can be improved by performing the experiment in Figure 2.1 (c), which involves applying n consecutive $X_{\pi/2}$ pulses. For this sequence, assuming the only error of the pulse is in the amplitude, the qubit $|1\rangle$ population for n $X_{\pi/2}$ can be estimated to be [45]

$$P_1 = -1^{(n+1)} C \cos\left(\frac{2\pi}{4A_{\pi/2}}(n+1)A\right) + y_0, \tag{2.4}$$

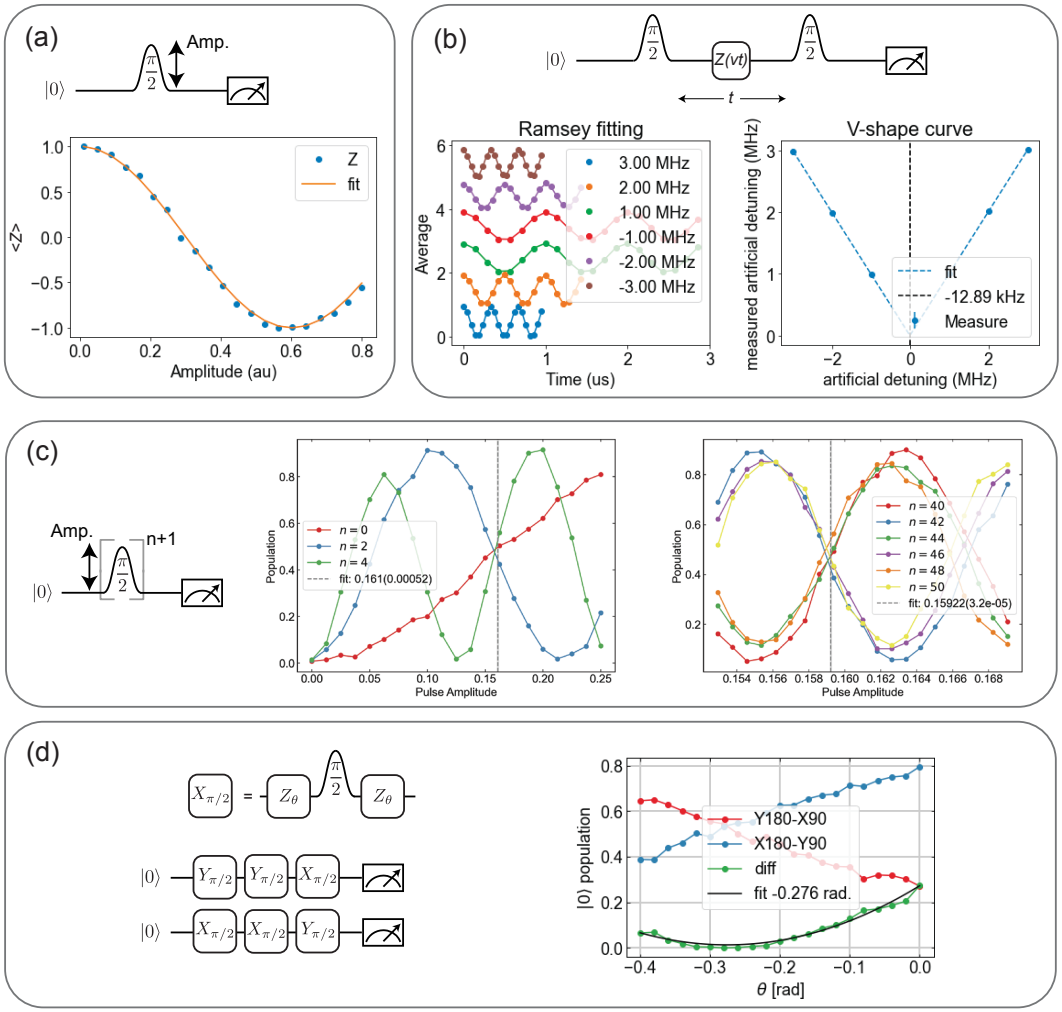


Figure 2.1: $X_{\pi/2}$ **calibration (a)** Coarse pulse amplitude calibration **(b)** Ramsey frequency calibration with several artificial detunings ν to find the qubit frequency **(c)** Fine amplitude calibration, with pulse repetitions to find increase sensitivity to the optimal drive amplitude. **(d)** Phase error correction, bookends the pulse with virtual-Z_θ gates to correct for phase errors due to Stark shifts of the higher transmon levels.

where $y_0 = 0.5$ for perfect $X_{\pi/2}$ pulses, $C = 1$ for perfect measurement contrast, and $A_{\pi/2}$ is the $X_{\pi/2}$ pulse amplitude. With this functional form in mind, the experiments in Fig. 2.1 (c) show amplitude sweeps, first for few n in the left plot, and then for many n in the right plot, illustrating the improved accuracy that can be realized with this approach. Finally, we turn to the errors induced by the presence of higher levels interacting with the pulse. Even though the drive frequency is resonant with the qubit, as calibrated by the Ramsey experiments, the pulse is off-resonantly driving the $|1\rangle \leftrightarrow |2\rangle$ transition. From the form

of H_d given in equation 2.3, this off-resonant drive will shift the energies of the system by $\Delta_{12} \approx 4\Omega^2/\eta$ [46]. This can be thought of as a dynamic detuning of the pulse, that changes as the pulse amplitude changes. In principle, this can be mitigated by detuning the pulse drive frequency [47], however for $X_{\pi/2}$ pulses this can be more simply corrected by calibrating virtual-Z_θ gates on either side of the pulse [41], as shown in Figure 2.1 (d). To calibrate θ, we perform a calibration that compares two sequences that ideally put the qubit on the equator, but are linearly sensitive to detunings and are oppositely-signed [48], as seen in the plot. By fitting to the point where they cross (or the minimum of their squared difference), the optimal θ can be determined.

Note that, in addition to phase errors, leakage errors, where the $|2\rangle$ state is inadvertently populated due to a strong off-resonant drive on the $|1\rangle \leftrightarrow |2\rangle$ transition, can also occur, particularly at when Ω/η approaches the upper end of the perturbative limit. In our case, with single-qubit gate times of 30 ns, this corresponds approximately to $\Omega \sim 20\,\text{MHz} < 0.1\eta$. In principle, minimizing single-qubit gate times can improve fidelity, and further study of minimizing single-qubit gate times while maximizing single-qubit gate fidelities is necessary for maximal algorithm performance. However, in the interest of solving the limiting errors first, reducing two-qubit gate and crosstalk errors is of most concern for improving algorithm performance.

In the next section, I will describe challenges that arise when driving single-qubit gates between qubits simultaneously, specifically due to microwave crosstalk present on our quantum hardware.

2.2 Mitigating Classical Crosstalk

Ideally, single-qubit operations on qubit Q_q are realized by sending control fields $\Omega_q(t)$ to Q_q, and $\Omega_q(t)$ would only interact with Q_q. However, design imperfections of our sample packaging inevitably results in fields being sent to unwanted locations on the chip, resulting in inadvertent driving of other qubits. We present experiments to measure this crosstalk, and describe our approach to mitigating it with additional crosstalk cancellation tones of appropriate amplitudes and phases.

A schematic illustration for classical crosstalk is given in Figure 2.2. Concretely, suppose we want to apply a drive $\varepsilon_t(t)$ that targets qubit Q_t at frequency ω_d. This drive is then applied by sending a microwave pulse $A_t(t)$ down its designated control line L_t. However, due to microwave crosstalk, modeled via a matrix C, A_t can also generate field ε_s on spectating qubit Q_s. Microwave crosstalk originates for example from coupling between control lines and other qubits directly or via common coupling to electromagnetic modes present in the sample box. The crosstalk matrix C can be written as

$$\begin{pmatrix} \varepsilon_\text{s} \\ \varepsilon_\text{t} \end{pmatrix} = \begin{pmatrix} e^{i\theta_\text{s}} & C_\text{st}e^{i\varphi_\text{st}} \\ C_\text{ts}e^{i\varphi_\text{ts}} & 1 \end{pmatrix} \begin{pmatrix} A_\text{s}e^{-i\phi_s} \\ A_\text{t}e^{-i\phi_t} \end{pmatrix}, \tag{2.5}$$

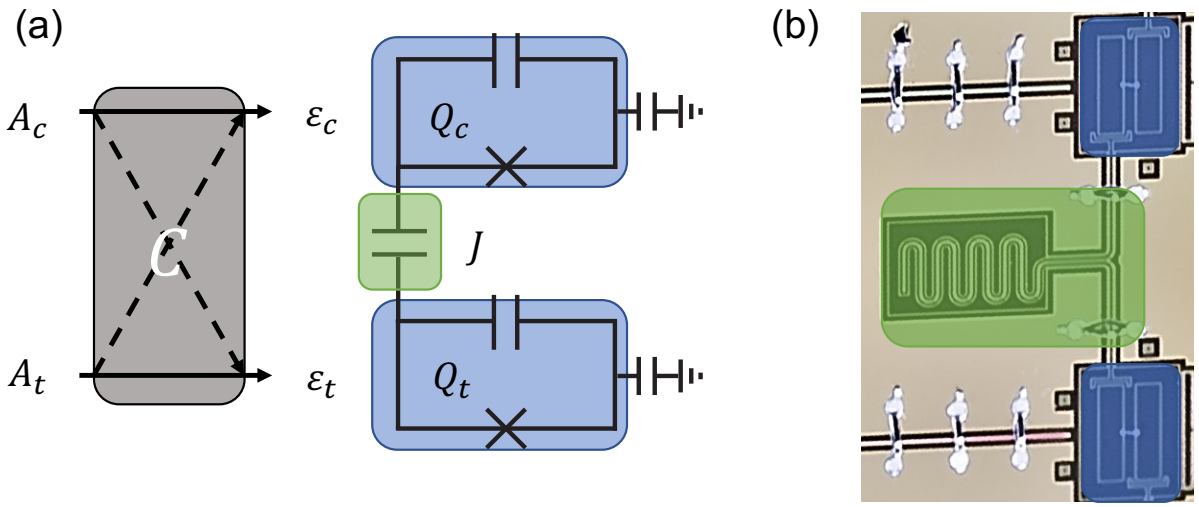


Figure 2.2: **Classical crosstalk schematic. (a)** Drive amplitudes are sent to the qubits via their control lines, though crosstalk matrix C mixes the drive signals to realize fields ε_c, ε_t on the transmons. Coupling J can alter the effects that crosstalk has, such as "accidental CR" interactions. **(b)** Image of two coupled qubits on the Trailblazer chip, with control lines on the left.

where C_{st} (φ_{st}) denotes the crosstalk amplitude (phase) from Q_s to Q_c. The phase θ_{s} results the Q_s line having different electrical delay relative to the Q_t line. If the drive is resonant (off-resonant) with spectating qubit Q_s, we can measure the field strength $|\varepsilon_s(t)|$ incident on Q_s by performing Rabi (Ramsey) type experiments on Q_s when the drive is applied. Then, the field on Q_s can be cancelled by applying a drive down line L_s with amplitude and phase that minimizes the e.g. Stark shift or Rabi rate on Q_s. Measuring and mitigating crosstalk between drive line L_t and spectator qubit Q_s is conceptually simple when Q_t and Q_s are not coupled: measure the Rabi/Ramsey frequency for resonant/detuned drive frequencies, and minimize the Rabi frequency / Stark shift by applying down line L_s the appropriate drive amplitude and phase A_s, ϕ_s. This process is in essence diagonalizing the crosstalk matrix $C(\omega = \omega_t)$ between the Q_t and Q_s.

When the two transmons are coupled and engineered for the Cross Resonance (CR) interaction as they are for the Trailblazer chip, however, measuring crosstalk between coupled qubits needs to be adapted to account for the other Hamiltonian dynamics between them. For example, when targeting qubit Q_t, classical crosstalk between the Q_t drive line L_t and the control qubit Q_c will induce an "accidental CR" interaction between target Q_t and control Q_c. While this field on Q_c is off-resonant (being resonant with Q_t), simply measuring the Stark shift on Q_c as above would not work because the combination of Q_t being driven and the ZZ coupling between Q_c and Q_t results in additional Z-type dynamics on Q_c. To measure the field on Q_c then, we can instead minimize this unwanted CR interaction as

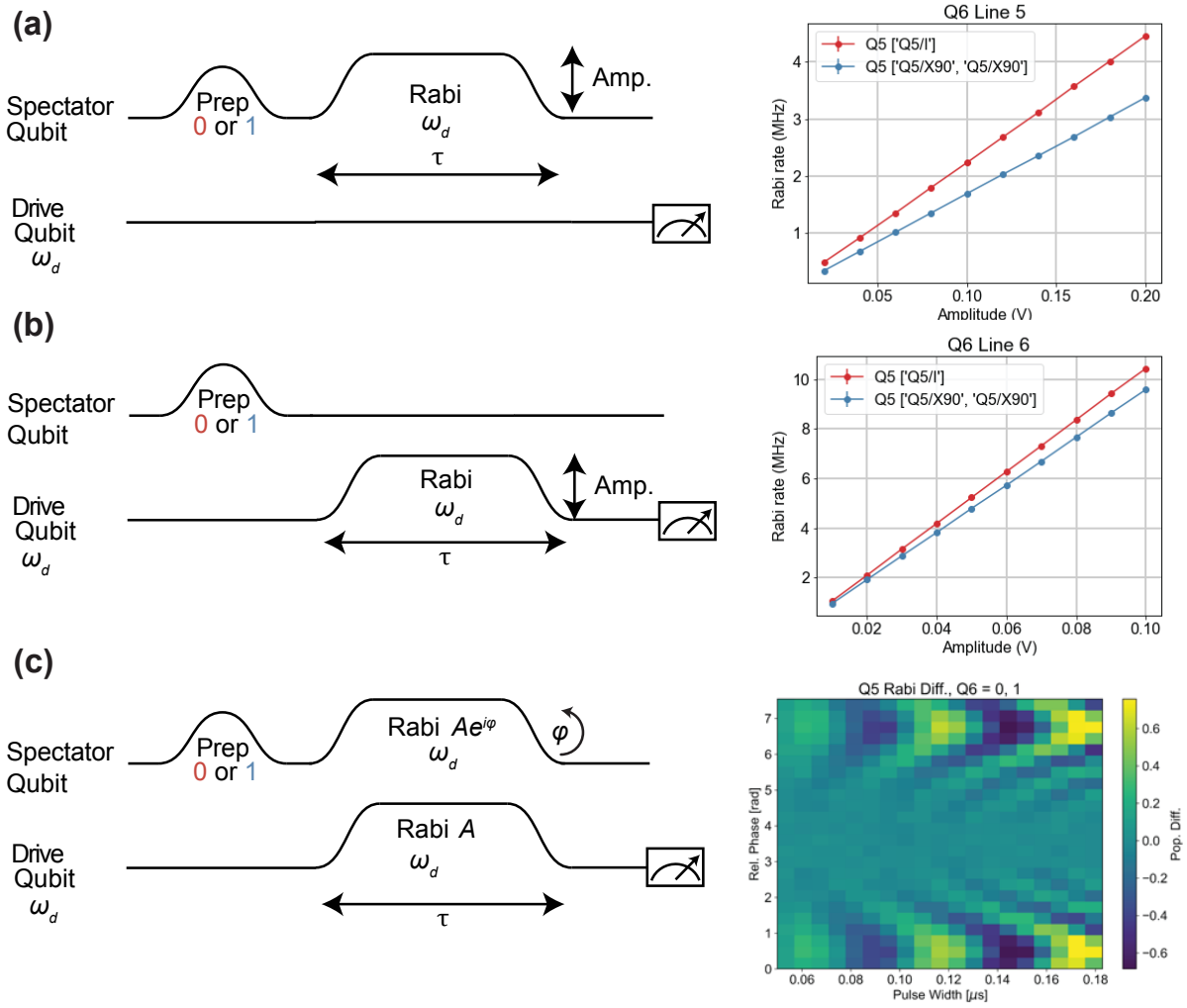


Figure 2.3: **Crosstalk calibration for coupled qubits using the CR effect. (a)** Conditional Rabi rate experiment on Q_t by driving the Q_c control line. Here, $Q6$ is Q_t i.e. the targeted or driven qubit, and $Q5$ is Q_c or the spectator qubit. Measuring Rabi rates conditioned on Q_c state gives an estimate of the field present on Q_c. **(b)** Conditional Rabi on Q_t's own drive line, showing some conditionality, indicating crosstalk from L_t to Q_c that can be cancelled with a drive down L_c. **(c)** Calibrating the phase of the crosstalk cancellation tone down L_c. Minimizing the Q_t population difference (color bar) when Q_c is in $|0\rangle$ and $|1\rangle$ versus relative drive phase (y-axis) and pulse duration (x-axis).

a function of Q_c crosstalk cancellation drive amplitude A_c and phase ϕ_c. Specifically, we minimize the conditional Rabi rate of Q_t as a function of drive amplitudes A_t, A_c, and relative drive phase $\phi_c - \phi_t$. An example of experiments run for this purpose are shown in Figure 2.3. The essential idea is, field at frequency ω_t on Q_c realizes a Q_c-state-dependent Rabi rate on Q_t via the Cross Resonance (CR) effect, so we measure the conditionality of the Rabi rate of Q_t on the state of the control qubit to measure the field strength on Q_c. In Figure 2.3 (a), the Rabi rate as a function of drive amplitude down L_c is measured, when Q_c is in $|0\rangle$ and $|1\rangle$. Fitting the slope of difference in these Rabi rates gives the scaling of the conditional Rabi rate with respect to drive amplitude when driving Q_t from line L_c, denoted $\mu(Q_t, L_c)$. Similarly, we fit the slope of the conditional Rabi rate when driving down Q_t's own line $\mu(Q_t, L_t)$ (this is really the crosstalk we want to cancel). From these experiments, we can match the conditional Rabi rate coming from driving L_t that comes from crosstalk accidentally driving Q_c, by applying the amplitude that gives the equal Rabi rate down L_c. That is, given drive amplitude A_t, from the experiments in Figure 2.3 (b) we predict conditional Rabi rate $\mu(Q_t, L_t)A_t$ down L_t. To match this amplitude down L_c, we apply amplitude A_c such that $\mu(Q_t, L_c)A_c = \mu(Q_t, L_t)A_t$. Importantly though, we need to find the drive phase ϕ_c that cancels the $\mu(Q_t, L_t)$. That is, we apply equal conditional Rabi rates down each line, and sweep the relative phase between the drives to minimize the conditional Rabi rate, thus cancelling the field on Q_c, as shown in Figure 2.3. The phase at which the difference in Q_t populations when Q_c is in $|0\rangle$ and $|1\rangle$ is minimized corresponds to the phase at which the field on Q_c is coherently cancelled. A final measurement of the conditional Rabi rate with the calibrated crosstalk cancellation tones included is performed to confirm crosstalk cancellation.

To benchmark this crosstalk cancellation protocol, we can perform simultaneous and isolated single-qubit Randomized Benchmarking (RB) [45], described in more detail in Section A.1. An example data set is shown in Figure 2.4. Crosstalk should have maximal impact when gates are run simultaneously, and minimal impact when run in isolation. The data in Figure 2.4 shows that without crosstalk cancellation, simultaneous operations have error rates about five times that of isolated operations. With crosstalk cancellation pulses calibrated as described above, the simultaneous error rates are reduced to just under two times the isolated error rate.

While this crosstalk cancellation protocol has shown improvement in simultaneous single-qubit RB, a natural question to ask is what the limitations and drawbacks are of this calibration measurement scheme. First, this approach is continuous-wave field cancellation, and therefore any delays between control lines in pulse delivery are not captured in the conditional Rabi experiments. Additionally, a method that can amplify sensitivity to the optimal crosstalk compensation amplitude and phase is desired. One possible approach is, rather than measuring the Q_t Rabi rate conditioned on the state of Q_c, to instead measure the Stark shift on Q_c when driving at the Q_t frequency down each line, after preparing Q_t along the axis of the drive. This is so that Q_t stays fixed during the drive, avoiding unwanted Z-type dynamics on Q_c resulting from the static ZZ coupling and Q_t dynamics. Note that, because Q_c and Q_t are coupled, with Q_t along the e.g. X axis, ZZ static dynamics will still

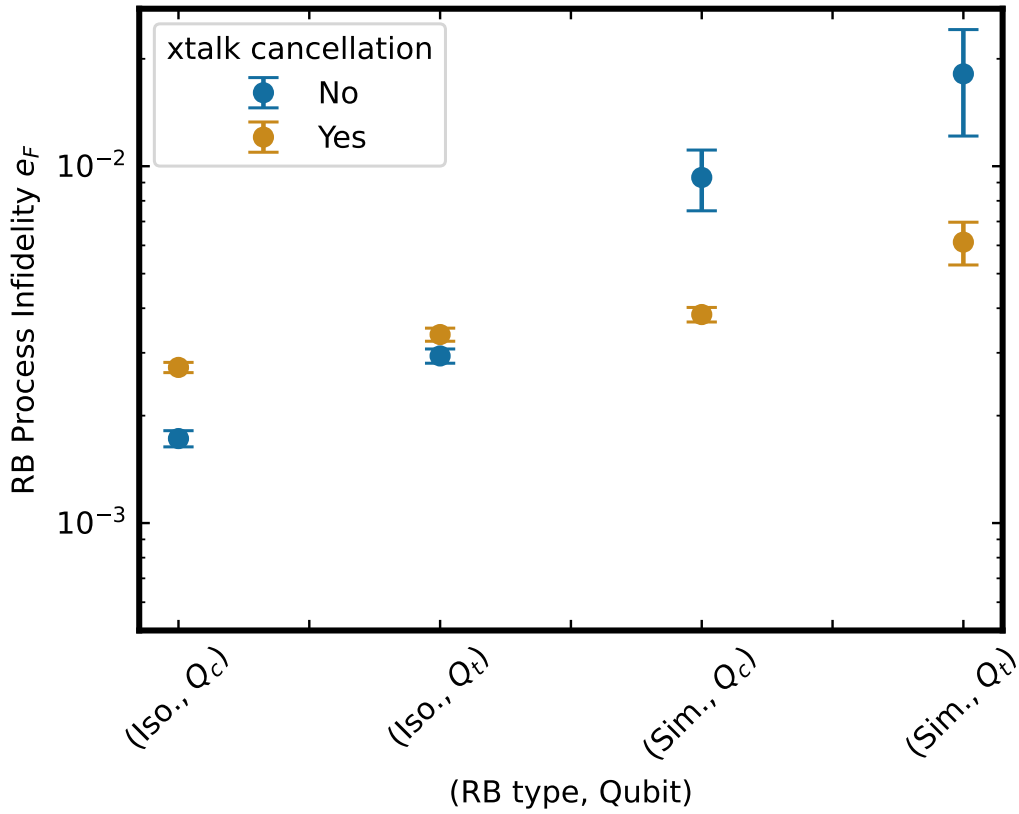


Figure 2.4: **Benchmarking single-qubit gates and crosstalk cancellation.** Without crosstalk cancellation, simultaneous operation has an error rate about five times as large as isolated operation. With crosstalk cancellation, this reduces to simultaneous operation having error rates only twice that of isolated operation.

play a role (since Q_t is in superposition of Z eigenstates). This can be mitigated with an echo pulse on Q_t, to reverse the ZZ dynamics that would affect the Q_c Stark shift measurement. Furthermore, this experiment can be done in a pulsed fashion, replacing the Rabi drive with repeated e.g. $X_{\pi/2}$ pulses. This enables error amplification (going to larger pulse numbers), and it uses the actual pulses that will be employed in circuits, rather than a continuous-wave drive. Similar pulse-level experiments that vary e.g. a delay between the pulse delivery between L_c and L_t could also be useful for measuring delays between control lines, which would complicate this simple cancellation scheme described above.

These drawbacks and potential extensions of this crosstalk cancellation protocol are based on a physical model of crosstalk, and propose physically-motivated experiments to mitigate it. However, a more model-free approach has also been successfully used in the group, where, like the tuning of the intrinsic Cross Resonance gate discussed in Section 3.3.2,

closed-loop optimization is done to tune the crosstalk amplitudes and phases, by minimizing the error in simultaneous RB, similar to previous work using RB [49]. This has the benefit of optimizing the parameters with the very benchmark that is used to verify performance, however, the randomness of the RB circuit also reduces sensitivity of the cost function to the pulse parameters.

2.3 Error budgeting single-qubit gates

Having discussed calibrations of single-qubit gates and crosstalk, it's worth quantifying different error sources of single-qubit gates, including decoherence, leakage, and the static ZZ interaction between qubits. Leakage begins to affect fidelity when gate durations are below $10/\eta$ [42, 50], which for our system is about 5 ns. The reason why shorter gates lead to more leakage can be understood simply by considering the Fourier spectrum of a pulse. A shorter pulse has a broader spectral width, leading to larger spectral weight on neighboring transitions (i.e. the $|1\rangle \leftrightarrow |2\rangle$ transition for transmons), thereby driving those transitions and causing leakage. Since our single-qubit gate times are 30 ns $\gg 10/\eta$, leakage contributions to gate errors are small in comparison to decoherence- and ZZ-induced errors. As mentioned, the gate duration of our single-qubit gates were not made shorter because we found crosstalk cancellation to be less effective below 30 ns. Unfortunately, longer gate times lead to higher error rates due to decoherence. The combined process error for both qubits due to decoherence can be estimated from the Q_t and Q_c lifetimes (coherence times) $T_1^{(t)}$ and $T_1^{(c)}$ ($T_2^{(t)}$ and $T_2^{(c)}$) via

$$e_D \approx \frac{\langle \tau_g \rangle}{4} \left(\frac{1}{T_1^{(c)}} + \frac{1}{T_1^{(t)}} + \frac{2}{T_2^{(c)}} + \frac{2}{T_2^{(t)}} \right). \tag{2.6}$$

For our qubits, with typical parameters ranging between $50\,\mu\text{s} \leq T_1 \leq 100\,\mu\text{s}$ and $70\,\mu\text{s} \leq T_2 \leq 150\,\mu\text{s}$, this corresponds to a decoherence error budget for 60 ns arbitary single qubit gates of $0.7 \times 10^{-3} \leq e_D \leq 1.5 \times 10^{-3}$, in the range of measured simultaneous RB process fidelities.

2.3.1 ZZ error budget

The contribution toward process error due to the static ZZ interaction can be simply estimated by considering the process error of static ZZ on an identity operation of the gate duration τ_g (see for example [51] supplement)

$$e_{ZZ} = 1 - \text{Tr}\left[U_{id}^{\dagger} U_{err}\right]/d^2, \tag{2.7}$$

with $U_{id} = I$ and $U_{err} = e^{-2\pi i \zeta_0 \tau_g |1\rangle\langle 1|}$ with ζ_0 being the static ZZ rate. Figure 2.5 shows the process error from ZZ as a function of the ZZ rate, for a few different single-qubit

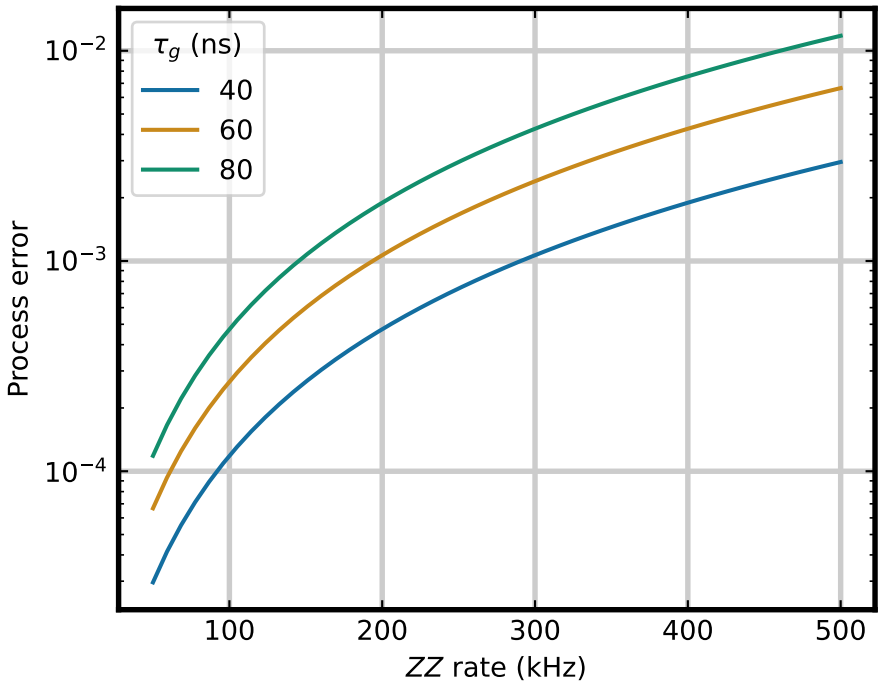


Figure 2.5: **Error contribution of ZZ, as estimated by equation 2.7, for a few different pulse times.** Scaling with ZZ is super-linear, but sub-exponential, highlighting benefits of reducing static ZZ and gate time as much as possible.

gate durations. For the systems studied in this thesis, with e.g. $150\,\text{kHz} \leq \zeta_0 \leq 300\,\text{kHz}$, and with two $X_{\pi/2}$ gates per arbitrary single-qubit gate, the error contribution as estimated above is between $0.1 - 0.25\%$, slightly below the typical single-qubit gate errors measured in our experiments RB experiments. Further, because X-type gates have some decoupling effects on ZZ dynamics, this estimate given above may be an upper bound on the general gate errors.

2.3.2 Sample B error budget

While both ZZ error and decoherence error estimates are comparable, these can in principle be distinguished using Extended Randomized Bencharking (XRB), described in Section A.1. However, in the interest of precisely characterizing one SRB dataset, we can estimate the error contributions from decoherence and ZZ for the simultaneous single-qubit RB data from Sample B (parameters given in Table 3.1). From data shown in Figure 3.17, the average e_F of the qubits is $e_F = 2.25 \times 10^{-3}$, and given the static ZZ $\zeta_0 = 176\,\text{kHz}$, $e_{ZZ} = 0.82 \times 10^{-3}$ and the T_1 and T_2^{echo} of the qubits, the decoherence error estimate is $e_D = 1.17 \times 10^{-3}$. Combined, $e_D + e_{ZZ} = 2.05 \times 10^{-3}$, very close to the measured e_F. This suggests that ZZ

and decoherence are the dominant sources of error for single-qubit gates. The single-qubit errors could thus be improved by reducing the single-qubit gate durations, reducing ZZ, or improving qubit coherences.

Chapter 3

Cross Resonance Gate

The Cross-Resonance gate, first proposed in [8] and demonstrated in [52], is an entangling gate that is widely used on superconducting-circuit-based QPUs because it can be operated on arguably the simplest hardware configuration: fixed-frequency, fixed-coupling transmon qubits. Further, it is predicted to be capable of error rates below 10^{-4} [53], with the lowest published experimental error rate of 1.9×10^{-3} [51]. Because of its established use, compatibility with simple architectures, and predicted low error rates, the CR gate was the first used primary entangling interaction on the Trailblazer multi-qubit chip architecture in QNL.

This chapter begins with the principle of the CR gate and the Hamiltonian dynamics of the CR gate on the Trailblazer chip, including crosstalk and leakage effects. Then, pulse optimization methods used to minimize leakage in CR gates are presented, followed by a comparison of two calibration approaches: the direct CNOT and an *intrinsic* CR gate calibration based on black-box optimization. Each calibration approach is benchmarked with a suite of randomized benchmarking (RB) type protocols, and their performance is compared.

3.1 Principle

The Cross Resonance (CR) interaction between two systems is an interaction where the drive amplitude of one resonant system (the target system) is conditioned on the state of the other (control) system. This is realized by driving the control system at the resonant frequency of the target system, hence the name 'cross-resonance'. The CR interaction between the two systems with frequencies ω_1 and ω_2 needs a few ingredients: (1) an exchange coupling J between the two systems, (2) individual control of each system, and (3) the systems need to be detuned from each other such that $|\Delta| = |\omega_2 - \omega_1| \gg J$. The original CR work [8] and the first demonstration analyzed the interaction assuming each system was a two-level-system, or the "qubit approximation". However, Tripathi *et al.* observed that the interaction is present even between two classical oscillators, provided one of the oscillators is nonlinear [34].

To gain some intuition for this interaction, let us briefly estimate how the Cross Resonance

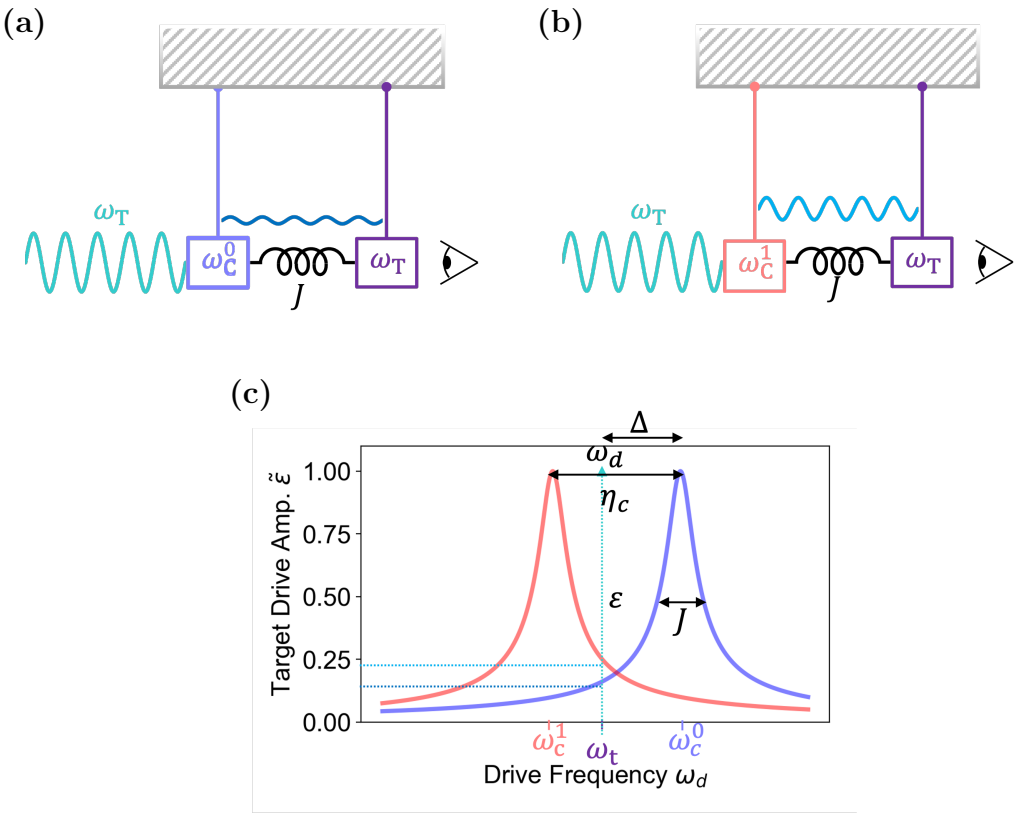


Figure 3.1: **Classical picture of Cross Resonance interaction.** The CR effect can be seen between two coupled oscillators, with the control oscillator being nonlinear, i.e. having a state-dependent frequency. **(a)** With the control oscillator in state "0", the drive transmitted to the target is attenuated. **(b)** When the control in "1", the amplitude is attenuated less. **(c)** Depiction of the control oscillator filtering the amplitude transmitted to the target oscillator. The state-dependent frequency of the control oscillator results in different filtered target drive amplitudes. See text for further description.

interaction scales with the system parameters for the classical coupled oscillators case, as illustrated in 3.1. When driving the control oscillator with amplitude ε at the target oscillator frequency, $\omega_d = \omega_t$, the control oscillator undergoes small forced oscillations at frequency ω_d, with amplitude that scales as ε/Δ. These oscillations in turn drive the target oscillator with amplitude $\tilde{\varepsilon} \sim J\varepsilon/\Delta$. In the case that the control oscillator frequency is nonlinear, its frequency ω_c depends on its state, and therefore the $\tilde{\varepsilon}$ depends on the control qubit state. For instance, if the control oscillator in state 1 detuning doubles $\Delta \to 2\Delta$ as compared to control oscillator state 0, then the difference in target drive amplitudes $\tilde{\varepsilon}_0 - \tilde{\varepsilon}_1 \sim J\varepsilon/(2\Delta)$.

In the case of coupled qubits, the CR interaction can be derived through analysis of the two-qubit Hamiltonian. In absence of the drive, in the uncoupled basis $|c, t\rangle$, the coupled

two-qubit Hamiltonian in the lab frame reads:

$$\hat{H}/\hbar = \omega_c \left|10\right\rangle \left\langle 10\right| + \omega_t \left|01\right\rangle \left\langle 01\right| + (\omega_c + \omega_t) \left|11\right\rangle \left|11\right\rangle + J \left(\left|10\right\rangle \left\langle 01\right| + \text{h.c.}\right). \tag{3.1}$$

The eigenstates $\overline{|c,t\rangle}$ of $\hat{H}$ are superpositions within excitation manifold, where to first order in J/Δ (with $\Delta = \omega_c - \omega_t$), $\overline{|10\rangle} \sim |10\rangle + (J/\Delta)\,|01\rangle$ and $\overline{|01\rangle} \sim |01\rangle - (J/\Delta)\,|10\rangle$ [34], with shifted eigenenergies $\overline{\omega}_{c,t} = \omega_{c,t} \mp J/\Delta$ [52].

The CR drive Hamiltonian in the drive frame is given by $\hat{H}_d = (\varepsilon\,|00\rangle\langle 10| + \text{h.c.})$. The effective target qubit drive rate $\tilde{\varepsilon}_n$ when the control qubit is in state n is determined by the matrix elements of $\hat{H}_d$ in the coupled basis [34]:

$$\tilde{\varepsilon}_n = \overline{\langle n,1|}\hat{H}_d\overline{|n,0\rangle}. \tag{3.2}$$

At lowest order, $\tilde{\varepsilon}_0 = -\frac{J}{\Delta}\varepsilon$ and $\tilde{\varepsilon}_1 = \frac{J}{\Delta}\varepsilon$, and therefore the entangling rate is given by

$$\mu = \frac{\tilde{\varepsilon}_0 - \tilde{\varepsilon}_1}{2} = \frac{-J\varepsilon}{\Delta}, \tag{3.3}$$

to first order. Note that this same CR rate is obtained from the classical analysis, assuming the control-target detuning Δ approaches infinity for one of the oscillator states (i.e., infinite non-linearity).

In terms of the Pauli operators, then, the CR drive induces a Hamiltonian that looks like

$$H_{\text{CR}} = Z \otimes (\delta_{zi} I + |\mu| \cos(\varphi) X + |\mu| \sin(\varphi) Y), \tag{3.4}$$

with Hilbert space convention $\mathcal{H}_c \otimes \mathcal{H}_t$, $\{I, X, Y, Z\} = \{I, \sigma_x, \sigma_y, \sigma_z\}$ is the Pauli single-qubit operator basis, and $\delta_{zi} = \frac{|\varepsilon|^2}{\Delta}$ is the control-qubit Stark shift from the off-resonant drive, and $\varphi = \text{Arg}[\varepsilon]$ is the phase of the CR drive, which maps to the polar angle of the Q_t Bloch rotation axis. Either the ZX or ZY term enable a CNOT gate to be implemented. More concretely, in the Hilbert space defined by $|\text{Q}_\text{c}\rangle \otimes |Q_t\rangle$, a CNOT gate is written as

$$\begin{aligned}
\text{CNOT} &= |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X && (3.5)\\
&= \left(Z_{-\pi/2} \otimes I\right)(Z \otimes X)_{\pi/2}\left(I \otimes X_{-\pi/2}\right) && (3.6)\\
&= \left(Z_{-\pi/2} \otimes I\right)\left(I \otimes Z_{\pi/2}\right)(Z \otimes Y)_{\pi/2}\left(I \otimes Z_{-\pi/2}\right)\left(I \otimes X_{-\pi/2}\right), && (3.7)
\end{aligned}$$

where $(Z \otimes X)_\theta = \exp\left(-\frac{i}{2}\theta Z \otimes X\right)$, and the last line can be understood from the single-qubit Pauli gate relation $Y_{\pi/2} = Z_{\pi/2} X_{\pi/2} Z_{-\pi/2}$, up to a global phase. In other words, a $\pi/2$ rotation about the Y axis can be realized by first rotating about the Z axis by $\pi/2$, performing an $X_\pi/2$ rotation, and then rotating about Z by $-\pi/2$ radians. From this reasoning, either ZY or ZX can be used to realize a CNOT gate.

3.2 Dynamics

While the classical (nonlinear) Harmonic oscillator and the two-level-system limits are illustrative of the CR effect, in reality the systems we use consist of many levels, and consequently the observed dynamics of the Cross Resonance effect are more complex [54]. Several contemporary theoretical analyses have been done to carefully predict the CR dynamics on transmon qubits [34, 53, 55]. In this section, I will outline the experimentally relevant dynamics that we have studied and observed on hardware in QNL.

3.2.1 Computational Space Dynamics

Due to the higher levels of the transmon and microwave crosstalk on the device, the CR drive induces a Hamiltonian of the form [54]

$$\hat{H}_{\mathrm{CR}} = I \otimes \vec{A} + Z \otimes \vec{B}, \tag{3.8}$$

where $\vec{A} = a_x X + a_y Y + a_z Z$ and $\vec{B} = b_I I + b_x X + b_y Y + b_z Z$ are generic sums of Pauli operators. In other words, $\hat{H}_{\mathrm{CR}}$ is a conditional Rabi oscillation of the target qubit, whose Bloch-sphere axis and rotational frequency depends on the control qubit state. Note that in this notation, the CR rate is $\mu = b_x$ and $\delta_{zi} = b_I$. In terms of physical origin, a_y comes from microwave crosstalk, a_x has both intrinsic components and crosstalk-mediated terms, and a_z comes from Stark-shifting of the higher levels of Q_t. For the $\vec{B}$ components, b_I comes from Stark-shifting the Q_c levels, b_y depends on the phase of the CR drive as before, and b_z results from the higher levels of the transmons shifting the $|11\rangle$ state energy.

3.2.2 Duffing Oscillator Dynamics

It is perhaps illustrative to estimate, to lowest order, how each of the Hamiltonian terms in 3.8 arise after including higher transmon levels and microwave crosstalk. To begin, consider two coupled, simultaneously driven transmons, as illustrated in Figure 4.1. The Hamiltonian of the two transmons, in the frame of the drive at frequency ω_d and making the Duffing approximation of the transmon [16], is given by

$$H_{\mathrm{qb}} = \sum_{i=\mathrm{c,t}} \left(\omega_i - \omega_{\mathrm{d}}\right) a_i^\dagger a_i + \frac{\eta_i}{2} a_i^\dagger a_i^\dagger a_i a_i, \tag{3.9}$$

where for transmon i, a_i is the bosonic annihilation operator, ω_i is the transition frequency between $|0\rangle$ and $|1\rangle$, η_i is the anharmonicity, and $\hbar = 1$. Each drive term is given by $H_{\varepsilon_i} = \left(\varepsilon_i a_i + \varepsilon_i^* a_i^\dagger\right)$, where ε_i is the complex drive amplitude, and the coupling term with strength J is $H_J = J\left(a_{\mathrm{c}}^\dagger a_{\mathrm{t}} + a_{\mathrm{c}} a_{\mathrm{t}}^\dagger\right)$, where we denote the higher frequency transmon Q_{c}, and the lower frequency transmon as Q_{t}. The total Hamiltonian is $H = H_{\mathrm{qb}} + H_J + H_{\varepsilon_{\mathrm{c}}} + H_{\varepsilon_{\mathrm{t}}}$.

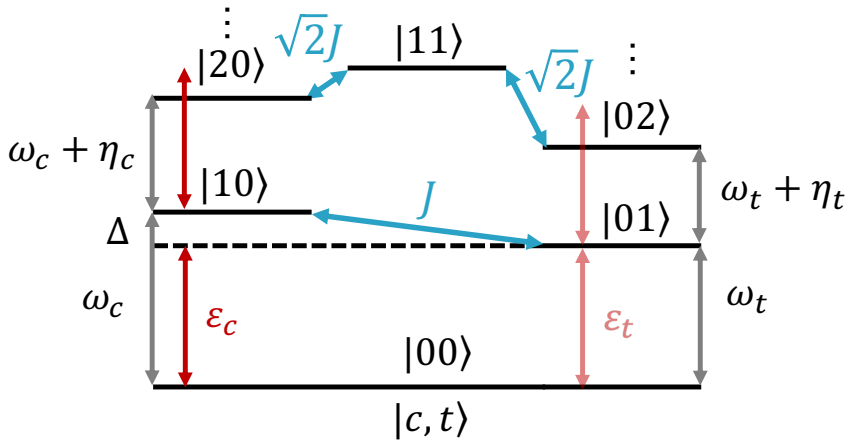


Figure 3.2: **CR drive diagram**. Transmons are detuned by $\Delta = \omega_c - \omega_t$ indicated by the black dashed line, and coupled via exchange coupling J. They are driven simultaneously at frequency $\omega_d = \omega_t$ with amplitudes ε_c, ε_t. The bare CR drive only drives the control ($\varepsilon_t = 0$), however a local target drive can be added intentionally, or via classical crosstalk, and is included in the analysis.

First, we examine the case of no crosstalk, and with $\omega_d = \omega_t$ and $\varepsilon_t = 0$. Following the analysis of [34], the presence of higher levels modifies $\tilde{\varepsilon}_1$ to be $\tilde{\varepsilon}_1 = \frac{J}{\Delta}\left(1 - \frac{2\Delta}{\Delta+\eta_c}\right)\varepsilon_c$. Since $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_1$ are no longer equal and opposite, there is a non-zero drive term ν on Q_t that is independent of Q_c's state

$$\nu = \frac{\tilde{\varepsilon}_0 + \tilde{\varepsilon}_1}{2} = \frac{-J}{\Delta + \eta_c}\varepsilon_c \tag{3.10}$$

and the CR rate becomes

$$\mu = \frac{\tilde{\varepsilon}_0 - \tilde{\varepsilon}_1}{2} = -\frac{J}{\Delta}\frac{\eta_c}{\Delta + \eta_c}\varepsilon_c, \tag{3.11}$$

that is, the higher levels add a correction factor of $\eta_c/(\Delta + \eta_c)$ to μ. Thus, the presence of higher levels with the CR drive, and without crosstalk, the H_{CR} terms have the following

dependence on drive parameters:

$$a_x = \nu \cos(\varphi) \tag{3.12a}$$
$$a_y = \nu \sin(\varphi) \tag{3.12b}$$
$$a_z \sim |\nu|^2/\eta_t \tag{3.12c}$$
$$b_x = \mu \cos(\varphi) \tag{3.12d}$$
$$b_y = \mu \sin(\varphi) \tag{3.12e}$$
$$b_z = 2J^2 \left(\frac{1}{\Delta - \eta_\mathrm{t}} - \frac{1}{\Delta + \eta_\mathrm{c}} \right) \tag{3.12f}$$
$$b_I = \frac{|\varepsilon_c|^2 \eta_c}{2\Delta(\Delta + \eta_c)}, \tag{3.12g}$$

where equation 3.12c is due to off-resonant driving of the $|1\rangle \leftrightarrow |2\rangle$ transition of Q_t, equation 3.12f is the static shift of the $|11\rangle$ state arising from dispersive coupling with $|02\rangle$ and $|20\rangle$ states, and equation 3.12g is the formula for the Stark shift of a Duffing oscillator [56]. Higher order estimates for these parameters can be obtained from time-dependent Schrieffer-Wolff Perturbation Theory (SWPT) [34, 53, 55]. Importantly, the conditional (ZY/ZX) and unconditional (IY/IX) polar drive angles are equal and set by the CR drive phase φ.

3.2.3 Duffing Oscillator Dynamics with crosstalk

When crosstalk is added to the picture, the unconditional terms are increased, and the unconditional and conditional rotation axes can become misaligned [54]. First, it is helpful to distinguish between the fields ε_c and ε_t that are present on Q_c and Q_t, respectively, from the microwave drive amplitudes A_c and A_t and phases ϕ_c and ϕ_t that we apply to the Q_c and Q_t control lines. In the presence of crosstalk, the drive fields at frequency ω_d are mapped to the fields at the qubits via

$$\begin{pmatrix} \varepsilon_\mathrm{c} \\ \varepsilon_\mathrm{t} \end{pmatrix} = \begin{pmatrix} e^{i\theta_\mathrm{c}} & C_\mathrm{ct} e^{i\varphi_\mathrm{ct}} \\ C_\mathrm{tc} e^{i\varphi_\mathrm{tc}} & 1 \end{pmatrix} \begin{pmatrix} A_\mathrm{c} e^{-i\phi_c} \\ A_\mathrm{t} e^{-i\phi_t} \end{pmatrix}, \tag{3.13}$$

where C_ct (φ_ct) denotes the crosstalk amplitude (phase) from Q_t to Q_c. The phase θ_c results the Q_c line having different electrical delay relative to the Q_t line, from which the polar axis for Q_t is defined. With crosstalk present then, driving only down Q_c's line and setting $\phi_c = 0$, the drive fields become

$$\varepsilon_c = A_c e^{i\theta_c} \tag{3.14a}$$
$$\varepsilon_t = C_{tc} A_c e^{i\varphi_{tc}}. \tag{3.14b}$$

Here, we see that the conditional axis has polar angle θ_c. However, because there is now a direct drive amplitude ε_t, the drive ε_t modifies the $I \otimes \vec{A}$ terms to be

$$a_x = \nu \cos(\theta_c) + |\varepsilon_t| \cos(\varphi_{tc}) \tag{3.15a}$$
$$a_y = \nu \sin(\theta_c) + |\varepsilon_t| \sin(\varphi_{tc}) \tag{3.15b}$$
$$a_z \sim |\nu + \varepsilon_t|^2/\eta_t. \tag{3.15c}$$

From these terms, we see that the polar axis for $\vec{A}$ is shifted relative to $\vec{B}$. Because ε_t only drives Q_t, the $Z \otimes \vec{B}$ terms are unaffected by the crosstalk.

3.2.4 Characterizing dynamics with Hamiltonian tomography

Despite the additional complexities that come from the higher transmon levels and classical crosstalk, the unique form of $\hat{H}_{\text{CR}}$–it is only off-diagonal in the Q_t Hilbert space $\mathcal{H}_t$–all the terms (except for b_I) can be determined by measuring the Q_t Bloch sphere dynamics for the case of Q_c in $|0\rangle$ and $|1\rangle$, as depicted in Figure 3.3 (a). In [54], this was dubbed *Hamiltonian tomography*. Example Hamiltonian tomography dynamics are shown in the top three panels of Figure 3.3 (b), demonstrating distinct Bloch sphere dynamics of the target qubit for each control qubit state. These data are plotted on the Bloch sphere in part (c). The absence of population transfer of Q_c is evident in the bottom panel of Figure 3.3 (b). While this experiment does not determine the ZI coefficient of $\hat{H}_{\text{CR}}$, this can be measured using a driven Ramsey experiment outlined in the previous chapter. To extract the Hamiltonian parameters, the Q_t dynamics for each Q_c state from Figure 3.3 (b) are simultaneously fit to a model of a point-particle on a sphere rotating at a fixed frequency $\vec{\Omega}$ about a fixed axis $\hat{n}$ [54]. From this fit, the data in Figure 3.3 give Hamiltonian terms $a_x/2\pi = -4.56\,\text{MHz}$, $a_y/2\pi = -0.57\,\text{MHz}$, $a_z/2\pi = 0.11\,\text{MHz}$, $b_x/2\pi = -3.39\,\text{MHz}$, $b_y/2\pi = -0.46\,\text{MHz}$, and $b_z/2\pi = -0.07\,\text{MHz}$. Note that the frequency resolution of these experiments is limited by the time-span of the data, which here is about 200 ns, corresponding to a Nyquist frequency of about 2.5 MHz. This must be kept in mind when examining low-frequency estimates of parameters with Hamiltonian tomography.

Further, the form of $\hat{H}_{\text{CR}}$ in equation 3.8 admits a simple measure for entanglement between Q_c and Q_t, which we term the conditionality distance R, given by

$$R = \frac{1}{2}||r_0 - r_1||^2, \tag{3.16}$$

where r_i is the target qubit Bloch vector when the control qubit is in state $|i\rangle$. When R is maximal, it indicates that the target qubit is located at maximally orthogonal locations on the Bloch sphere, conditioned on the control qubit state. Unlike more general two-qubit entanglement measures such as concurrence [57], estimating R does not require full two-qubit tomography. This is because the action on Q_c is only along Z. Therefore, it requires only tomography of the target qubit Q_t, for both conditions of Q_c. The catch here is, R measures only the entanglement produced by conditional-unitary type gates. Indeed, R is maximized

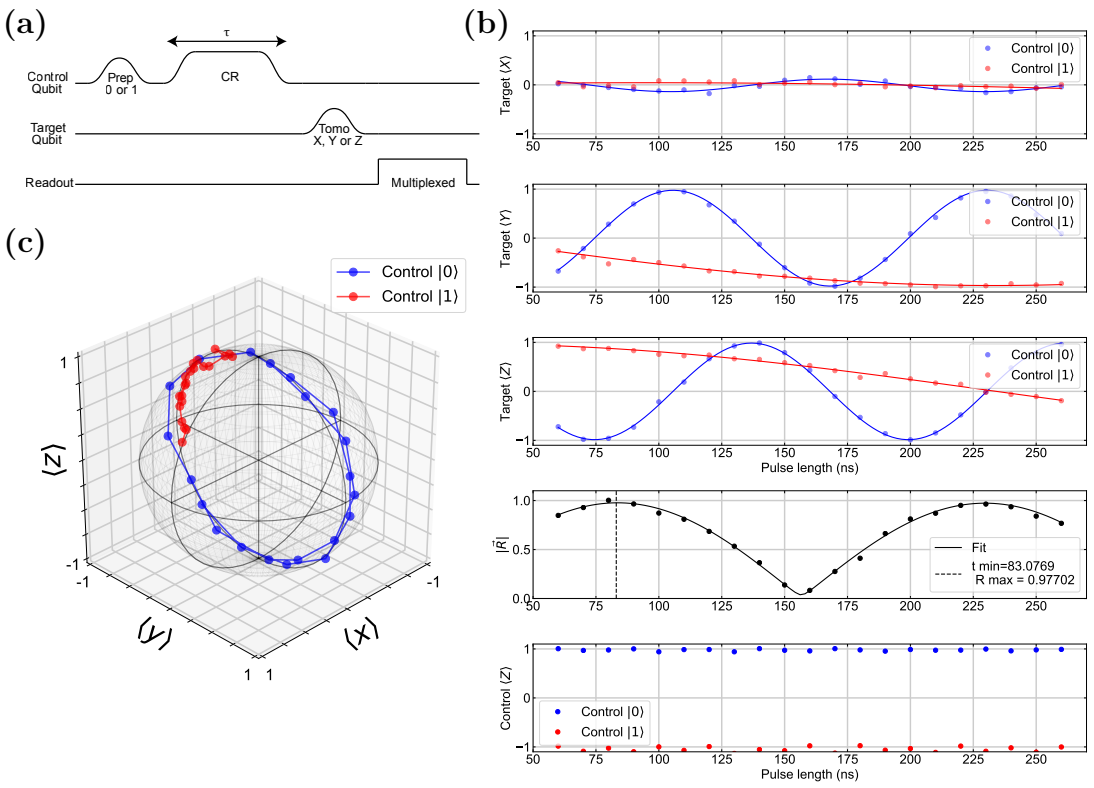


Figure 3.3: **Cross Resonance Dynamics and Hamiltonian tomography.** **(a)** Pulse sequence for Hamiltonian tomography. The control qubit is first prepared in either $|0\rangle$ or $|1\rangle$, and then the CR drive at frequency ω_t is applied to the control qubit. Then, the target qubit Bloch vector is measured via state tomography. **(b)** Example CR dynamics and fit according to the model in equation 3.8. The top three panels are the target qubit Bloch vector components. The second from the bottom plot shows entanglement measure R versus time. At the bottom, the control-qubit $\langle Z \rangle$ as a function of time. Small oscillations are present. **(c)** Target qubit dynamics on the Bloch sphere, conditioned on the control qubit state.

only when the unitary acting on Q_t when Q_c is in $|0\rangle$, is orthogonal to the unitary acting on Q_t when Q_c is in $|1\rangle$ (i.e. $\mathrm{Tr}[U_{0,t}^{\dagger}U_{1,t}] = 0$, where $U_{i,t}$ denotes the unitary acting on Q_t when Q_c is in state $|i\rangle$). For other types of entangling interactions, such as SWAP-like or iSWAP-like, R is not a good measure, because the interactions cannot be characterized by only performing tomography on one qubit. The fourth panel in Figure 3.3 (b) shows R as a function of time, with a fit indicating the pulse duration where the CR interaction has maximally entangled the qubits. Notably, this measure is insensitive to local rotations. The precise local rotations to realize a CNOT gate can be subsequently calibrated with target qubit rotations, and a control-qubit Z-gate.

With the experimental toolkit of Hamiltonian tomography, and a model that captures the CR dynamics with realistic experimental condition, we can study how the CR effect scales with our experimental parameters, with the ultimate goal of calibrating a high-fidelity CNOT gate. Recall that the parameters of interest for the CR interaction are the qubit-qubit detuning Δ, the Q_c anharmonicity η_c, the exchange coupling J, and the (complex) CR drive ε_c. On the fixed-frequency, fixed-coupling 8-qubit ring platform primarily employed in QNL, the only parameter we can control after fabrication is the CR drive. While previous works have explored the CR effect while varying Δ [58], theoretical prediction has shown two regions of frequency detuning [34, 53] to target when designing fixed-frequency samples, namely, when $\Delta \sim 0.6|\eta_c|$ is predicted to yield the largest CR rates, and $\Delta \sim 0.3|\eta_c|$ yields a larger μ/b_z ratio [53].

Figure 3.4 shows an example dataset for illustrating the interplay of CR dynamics with classical crosstalk. Figure 3.4 (a) and (b) show Hamiltonian tomography data as a function of CR drive amplitude and phase, respectively. The dashed vertical lines in the plots indicate the amplitude (phase) at which the phase (amplitude) sweep experiments were performed, and allow comparison between the two datasets. In Figure 3.4 (b), $A_t = 0$, whereas in Figure 3.4 (a) there is an applied target drive A_t. This is evident by the differences in the local terms between the two experiments. Since it has only the CR drive, to understand this data we begin with Figure 3.4 (b). First, we see that the IX, IY, ZX, and ZY terms are related according to the expected sinusoidal relation, and the IZ and ZZ rates are approximately independent of the drive phase. This agrees with our understanding from equation 3.12 and 3.15, that IZ and ZZ terms originate from Stark shift of the $|1\rangle \leftrightarrow |2\rangle$ transition of Q_t, and the static coupling between the qubits, respectively.

Next, we can estimate the crosstalk and complex drive parameter ε_c by fitting it to our model of the system with crosstalk that was described earlier. The qubits (Sample A of Table 3.1) used in this experiment at this time had parameters $\omega_c/2\pi = 5.841\,\mathrm{GHz}$, $\omega_t/2\pi = 5.687\,\mathrm{GHz}$, $\eta_c/2\pi = -243.6\,\mathrm{MHz}$, $\eta_t/2\pi = -247.3\,\mathrm{MHz}$, with static ZZ rate $\zeta/2\pi = 300\,\mathrm{kHz}$, corresponding to an inferred coupling strength $J/2\pi = 3.31\,\mathrm{MHz}$. From this, the theory predicts that the $b_x/a_x = \mu/\nu = \eta_c/\Delta \approx -1.57$, for the above parameters. However, notably, in the data in Figure 3.4, $|a_x| > |b_x|$, and their signs are aligned, rather than opposite. This is a manifestation of microwave crosstalk. From these data, $|\varepsilon_c|/2\pi = 60\,\mathrm{MHz}$, corresponding to the arbitrary-valued CR pulse amplitude in the dashed line of part (a). The relative drive line phase is $\theta_c = 1.38\,\mathrm{rad}$, the crosstalk amplitude is $C_{tc} = 0.112$, and crosstalk phase is

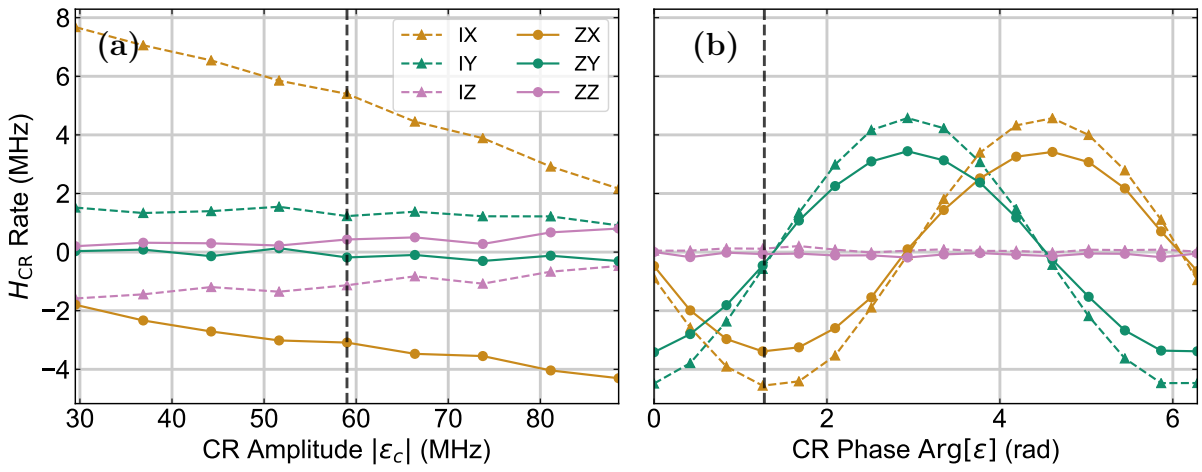


Figure 3.4: **Hamiltonian tomography while varying CR drive ε (a)** Hamiltonian terms as a function of $|\varepsilon|$. The presence of higher levels, and microwave crosstalk, results in a nonzero IX term. A local Q_t drive is also applied, offsetting the local terms. **(b)** Varying CR drive phase displays control over the rotational axis of the local and entangling terms. The ZX and IX terms are theoretically opposite in sign, however classical crosstalk in the device modifies this [54]. The IZ and ZZ terms are insensitive to the CR drive phase. The dashed line in the CR amplitude (phase) sweep indicates the fixed CR amplitude (phase) used in the corresponding CR phase (amplitude) sweep.

$\varphi_{tc} = -1.76\,\text{rad}$.

With the information gleaned from Figure 3.4 (b), we can understand the effect that an additional drive term $A_t e^{-i\phi_t}$ has on the dataset in part (a), which has fixed CR phase indicated by the dashed line in Figure 3.4 (b). First, the IX and IY values are offset by approximately 10 MHz and 1.7 MHz, respectively, corresponding to $A_t = 10.1\,\text{MHz}$ and $\phi_t = -2.97\,\text{rad}$. This nonzero A_t also adds to the field ε_c, but because $C_{ct} = 0.44$, $\varphi_{ct} = -0.09\,\text{rad}$ (as independently determined with crosstalk calibrations described in Chapter 2), the magnitude of additional field on Q_c is only$\sim 4\,\text{MHz}$, which is a small contribution to the entangling dynamics compared to CR drive $A_c \sim 60\,\text{MHz}$ as noted earlier, and is evidenced by the similar ZX and ZY rates at the dashed lines between parts (a) and (b). In terms of how the Hamiltonian terms scale with $|\varepsilon_c|$, the IX and IY terms scale linearly, and the IZ term reduces as the local drive amplitude decreases, as expected via the AC Stark effect. The ZZ and ZY terms appear to increase, however marginally, due to possible higher order effects [55] on the former and imperfect CR phase alignment on the latter. The entangling $Z \otimes \vec{B}$ terms begin to show a saturation behavior, as predicted theoretically [34, 53, 55].

In general, the crosstalk on each device can vary, with older versions of microwave packaging having much larger crosstalk, resulting in larger local target qubit dynamics. Further,

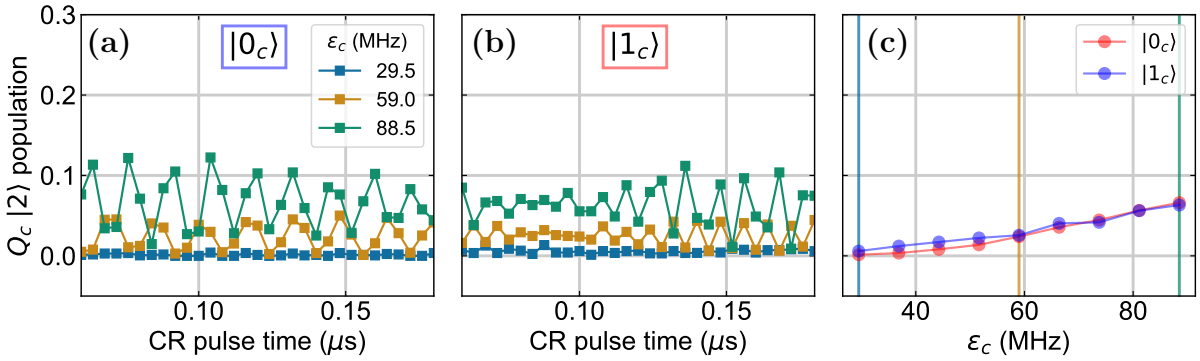


Figure 3.5: **Control qubit leakage. (a)** Q_c $|2\rangle$ population dynamics when Q_c is prepared in $|0\rangle$ during the Hamiltonian tomography experiment. Oscillations become clear at larger drive amplitudes $|\varepsilon|$. **(b)** Same as **(a)**, except Q_c is prepared in $|1\rangle$. **(c)** Time-averaged $|2\rangle$ population for varying $|\varepsilon_c|$, showing increasing leakage with drive amplitude.

the crosstalk matrix is generally frequency-dependent. Nevertheless, with these experiments, we are able to determine the Hamiltonian and crosstalk parameters, engineer the desired dynamics, and distinguish entangling from local terms as measured by R.

3.2.5 Leakage Dynamics

While the model originally provided in [54] and presented in equation 3.8 accounts for unitary dynamics of the CR effect within the computational subspace, such as microwave crosstalk and the non-resonant interactions involving higher transmon levels, at higher drive powers additional physics comes into play; notably, leakage out of the computational subspace can occur if the CR tone is near-resonant to the Q_c $|0\rangle \leftrightarrow |2\rangle$ or $|1\rangle \leftrightarrow |2\rangle$ transitions.

The physical origin of leakage dynamics in the CR drive can be understood in the same way as with single-qubit gates [42, 47]. While it has been proposed as a source of error theoretically [34], leakage dynamics of the CR gate have only been recently studied experimentally [23]. In short, because the CR drive frequency is near-resonant to the Q_c $|0\rangle \leftrightarrow |2\rangle$ and $|1\rangle \leftrightarrow |2\rangle$ transitions, there is sufficient spectral weight of the CR pulse at the frequency of these transitions, driving leakage out of the computational subspace. In the Hamiltonian tomography experiment of Figure 3.4, we monitored the $|2\rangle$ state population [59] for Q_c. In Figure 3.5 (a) and (b), we see oscillations of increasing amplitude.

To mitigate unwanted driving of a near-resonant transition, simple modifications to the CR pulse can be made, using what's called Derivative Removal by Adiabatic Gate (DRAG) [42]. The idea of DRAG is to modify the pulse envelope $b(t) \to b'(t)$ in a simple way

$$b'(t) = b(t) - i\frac{\alpha}{\Delta}\frac{d}{dt}\left(b(t)\right), \tag{3.17}$$

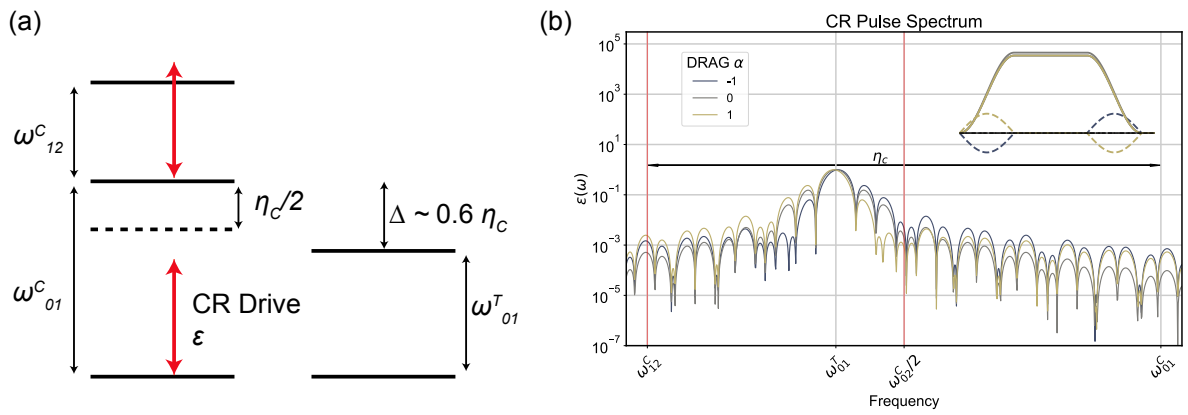


Figure 3.6: **CR DRAG Scheme (a)** The CR drive is near-resonant to the control qubit two-photon $|0\rangle \leftrightarrow |2\rangle$ transition (dashed line) and the $|1\rangle \leftrightarrow |2\rangle$ transition. **(b)** Fourier spectrum of CR pulse with carrier frequency at the target qubit frequency (envelope in inset), for a few different DRAG α coefficients. The spectral weight at the leakage transitions (red vertical lines) is modified by α.

to eliminate the spectral weight of the pulse at the frequency of interest $\omega_l = \omega_d + \Delta$. Fourier analysis shows that when the weighting parameter $\alpha = 1$, the pulse spectral weight at detuning Δ, $|b(\omega = \Delta)| = 0$, however in practice the optimal α may be different due to waveform distortions between the generator and the qubit [60]. The idea of DRAG is illustrated in Figure 3.6. The CR drive frequency ω_t for the qubit pair under study in this chapter is detuned just below the two-photon $\omega_c + \eta_c/2$ transition by $0.13\eta_c \approx -32\,\text{MHz}$, and also just above the $\omega_c + \eta_c$ transition by 89 MHz. The Fourier spectrum of the CR pulse, shown in Figure 3.6 (b), thus has non-zero spectral weight on these transitions. Adding the DRAG term to the pulse, illustrated in the inset showing the pulse shape, modifies the spectrum of the pulse, and thus the spectral weight of the pulse on each transition. By tuning the DRAG α parameter, the driving rate of these transitions can be tuned. In Figure 3.7, we perform an experiment to quantify the leakage to the Q_c $|2\rangle$ state. The left plot shows the Q_c state dynamics when applying the CR pulse for varying lengths of time, with $\alpha = -2.0$. The Rabi oscillations between $|0\rangle$ and $|2\rangle$ are evident. To quantify leakage, we calculate the mean-squared error of the dynamics of the $|2\rangle$ state

$$L_{\text{Rabi}} = \frac{1}{T}\sum_{t=0}^{T}\left(p_2(t) - \langle p_2\rangle_t\right)^2, \tag{3.18}$$

where p_2 is the qubit $|2\rangle$-state population. To calibrate α, we can thus minimize L_{Rabi}. In Figure 3.7 (b), we measure L_{Rabi} as a function of α, and we see that the optimal α is different depending on whether Q_c is prepared in $|0\rangle$ or $|1\rangle$. This is because the frequency detuning of

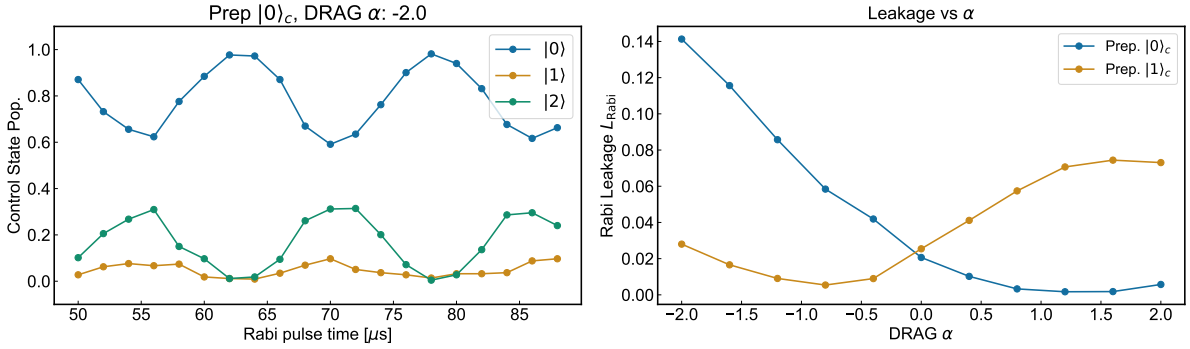


Figure 3.7: **CR leakage DRAG α sweep (a)** Q_c state populations as a function of the CR Rabi pulse duration, with $\alpha = -2.0$ and Q_c initialized in $|0\rangle$. Significant oscillations are seen between $|2\rangle$ and $|0\rangle$, indicating strong driving of the two-photon $|0\rangle \leftrightarrow |2\rangle$ transition. **(b)** Leakage L_{Rabi} as a function of DRAG α, when Q_c is prepared in $|0\rangle$ and $|1\rangle$. Leakage from each transition is not simultaneously minimized when tuning α.

the $|0\rangle \leftrightarrow |2\rangle$ and $|1\rangle \leftrightarrow |2\rangle$ transitions are different, and thus a single α parameter cannot minimize the CR pulse spectral weight on both simultaneously.

To simultaneously reduce leakage on both transitions, the DRAG formalism can be extended with additional pulse shaping [47, 61]. With a base envelope $b(t)$ of a flat-topped pulse with a cosine ramp, we adopt a 2D DRAG envelope

$$A_c(t) = b(t) + i\alpha_0 \frac{d}{dt}\left(b(t)\right) + \alpha_1 \left(i\frac{d}{dt}\left(b(t)\right)\right)^2, \tag{3.19}$$

and calibrate the DRAG parameters α_0 and α_1 to simultaneously minimize leakage from both transitions. A sweep of α_0,α_1 is shown in Figure 3.8. With 2D DRAG pulse shaping, we can thus reduce unwanted leakage dynamics that arise during the CR gate, which is crucial for high-fidelity quantum logic [62]. This technique is shown in section 3.3.1.1 to substantially reduce leakage while maintaining gate fidelity.

3.3 Gate Calibration

The Hamiltonian dynamics give physical understanding of the interactions between the qubits, however leveraging these interactions into quantum logical elements requires developing gate calibration and benchmark protocols. In this section, I will discuss two main approaches to calibrating entangling gates with the CR effect: direct CNOT calibration, and intrinsic CR calibration. The direct CNOT calibration realizes a CNOT gate by combining the CR interaction with additional local drives of the target qubit, realizing a CNOT gate

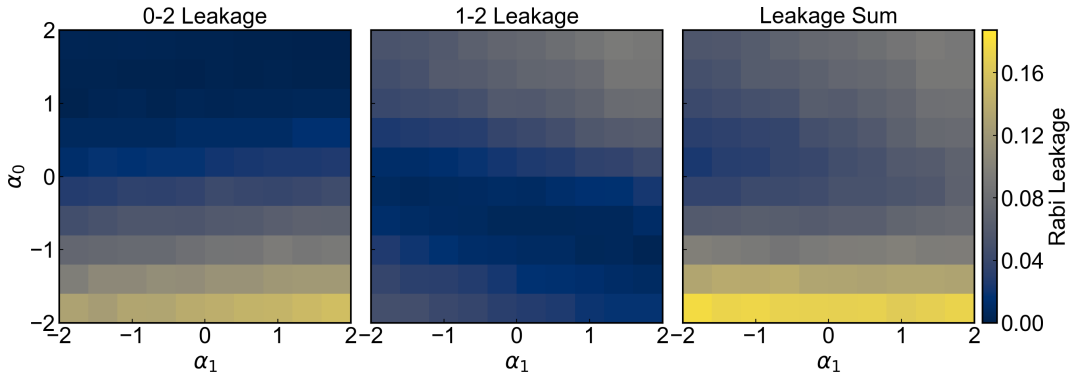


Figure 3.8: **Minimizing two leakage channels with 2D DRAG.** Left is the leakage from the two-photon $|0\rangle_c |2\rangle_c$ transition. Middle is $|1\rangle_c |2\rangle_c$ leakage, and right is the sum of both.

without any subsequent single-qubit gates (aside from instantaneous virtual Z gates [41]). We first show that leakage of the CR gate can be mitigated by shaping the CR pulse spectrum with the 2D DRAG protocol discussed previously. However, reducing leakage did not appear to reduce the overall gate error. On a separate pair of qubits that had higher anharmonicity and therefore were less susceptible to leakage, a more detailed set of error benchmarking was performed to assess the dominant error contributions, revealing that coherent (i.e. calibration or ZZ) errors formed nearly half of the error contribution from this gate.

To reduce the coherent error contribution, an alternative calibration approach was explored, leaving out the additional local target qubit drives during the CR gate and relegating local rotations to subsequent single qubit gates. Dubbed the *intrinsic* CR gate calibration approach, this separated the CNOT gate calibration into (i) precisely calibrating the CR gate amplitude to maximize entanglement with error-amplification circuits, and (ii) calibrating the single-qubit gates needed to map the entangling CR unitary U_{CR} to a CNOT gate. While this approach on the surface results in longer CNOT gates due to the additional single-qubit gates needed after the CR pulse, this issue is mitigated using by compiling quantum circuits into a gate set that replaces the CNOT gate with the intrinsic CR gate U_{CR}.

The benchmarking results of the direct CNOT and intrinsic CR gates are compared, and while both approaches yielded gate fidelities above 99%, the intrinsic CR gate had notably lower coherent error, suggesting that additional pulses used in the direct CNOT gate suffered from calibration errors. The lower coherent error of the intrinsic CR gate suggests that a simpler approach to gate calibration, focused on maximizing entanglement and delegating the local rotations to compilation-based solutions, may enable higher fidelity quantum algorithms.

3.3.1 Direct CNOT calibration

The dynamics discussed in the previous section provide guidance in leveraging the CR interaction to realize a CNOT gate. The unitary operation for a CR drive with envelope $\varepsilon(t)$ $H_{\mathrm{CR}}(t) = I \otimes \vec{A}(\varepsilon(t)) + Z \otimes \vec{B}(\varepsilon(t))$ with duration τ_p is given by

$$U_{\mathrm{CR}} = e^{-i \int_0^{\tau_p} H_{\mathrm{CR}}(t) dt}, \tag{3.20}$$

assuming H_{CR} commutes with itself at all times. Since a CNOT gate is locally equivalent to $ZX\left(\frac{\pi}{2}\right) = \exp\left\{-\frac{i}{2}\frac{\pi}{2}ZX\right\}$ operation, via [8]

$$\mathrm{CNOT} = ZI\left(-\frac{\pi}{2}\right) ZX\left(\frac{\pi}{2}\right) IX\left(-\frac{\pi}{2}\right), \tag{3.21}$$

a natural approach would be to calibrate the CR drive parameters such that the H_{CR} terms are $b_x = b_I = -a_x$, with all other terms being zero. Then, the interaction can be driven for time τ_p such that $\int_0^{\tau_p} \varepsilon(t) dt = \frac{\pi}{2}$. From the dynamics shown in Figure 3.4, we see that b_y can be calibrated to zero by adjusting the CR drive phase Arg[ε]. Further, any remaining a_y can be adjusted by applying drives to Q_t [54].

However, a_z , b_I, and b_z do not depend on drive phase. Since b_I commutes with the b_x interaction, this can be corrected later via a virtual-Z gate [41], but because b_z and a_z do not commute with b_x, they can reduce the overall entanglement achievable by the drive. While an echo on the control qubit can reduce the overall ZZ error in the unitary produced by the CR drive [54], as b_z/b_x becomes larger, the contrast of the conditional Rabi oscillations on Q_c is reduced, and thus this is an important measure of how entangling the interaction can be. Minimizing b_z/b_x has been explored using engineered ZZ suppression with a multipath coupler [23], indicating that this poses a fundamental minimal coherent error bound on CR gate error. Additionally, a strong IX drive on the target qubit can reduce the effect of b_z terms on the dynamics [40], however this increases a_z. The local Stark shift on Q_t a_z can finally be adjusted by adding a detuning the CR drive frequency ω_d, or by adding DRAG pulse shaping to the drive [42, 47].

To summarize, we can engineer the CR Hamiltonian to directly implement a CNOT gate, with the following procedure:

1. Calibrate the entangling rate by measuring R versus $|\varepsilon_c|$. This involves a full Hamiltonian tomography measurement set, meaning full Q_t tomography and preparing Q_c in $|0\rangle$ and $|1\rangle$, totaling to six experiments for each $|\varepsilon_c|$.

2. Minimize ZY by tuning the CR phase Arg[ε]. Since ZY only causes non-zero oscillations of r_x the x-component of Q_t Bloch vector r, measuring only r_x is sufficient, only four experiments per CR drive phase are needed.

3. Minimize IY with an IY drive on Q_t. Similar to ZY, only r_x needs to be measured, and thus there are four experiments per IY drive amplitude.

name	Sample A	Sample B
ω_c (GHz)	5.84769	5.45868
ω_t (GHz)	5.69371	5.29787
η_c (MHz)	-244.21	-271.35
η_t (MHz)	-246.87	-275
Δ (MHz)	153.98	160.81
Δ/η_c	0.631	0.593
ζ_0 (MHz)	0.3	0.176
J (MHz)	3.361	2.819
Q_c T_1 (μs)	60-80	65
Q_t T_1 (μs)	160	58
Q_c T_2^{echo} (μs)	120-160	86
Q_t T_2^{echo} (μs)	190	77

Table 3.1: Sample parameters for the two qubit pairs studied.

4. Calibrate IX drive amplitude to put Q_t in $|0\rangle$ when Q_c is prepared in $|0\rangle$. Here, r_z is all that need be measured, and only one preparation state, corresponding to just two experiments per IX drive amplitude.

5. ZZ can also be reduced with the IX drive, i.e. a 'rotary. echo' [40]

6. IZ can be reduced by using DRAG [42, 47] or pulse detuning.

7. ZI is corrected using a virtual-Z gate [41] on Q_c.

8. Leakage is calibrated with DRAG, with experiments described in section 3.2.5.

3.3.1.1 Leakage reduction with 2D DRAG

With this above approach, we calibrated CNOT gates on Sample A in Table 3.1 and varied the CR pulse time τ_p, with the aim of investigating whether there is a gate speed at which leakage is a dominant error source, and how robustly we can reduce leakage as gate speed is varied. Note that we did not calibrate the drive to minimize ZZ or IZ as described above in these experiments, as they depend more weakly on the drive parameters, and the focus of this study is the impact of leakage on gate fidelity. To quantify the effect of leakage mitigation with 2D DRAG pulse shaping, the gates were calibrated with and without the 2D DRAG calibration step of section 3.2.5. Figure 3.9 shows Hamiltonian tomography data for each gate after the calibration procedure, with the right (left) plot showing calibration with (without) 2D DRAG. The Hamiltonian tomography experiments show nearly the desired dynamics of a pure CNOT interaction, as seen in the lower two plots, with the Q_t approximately stationary

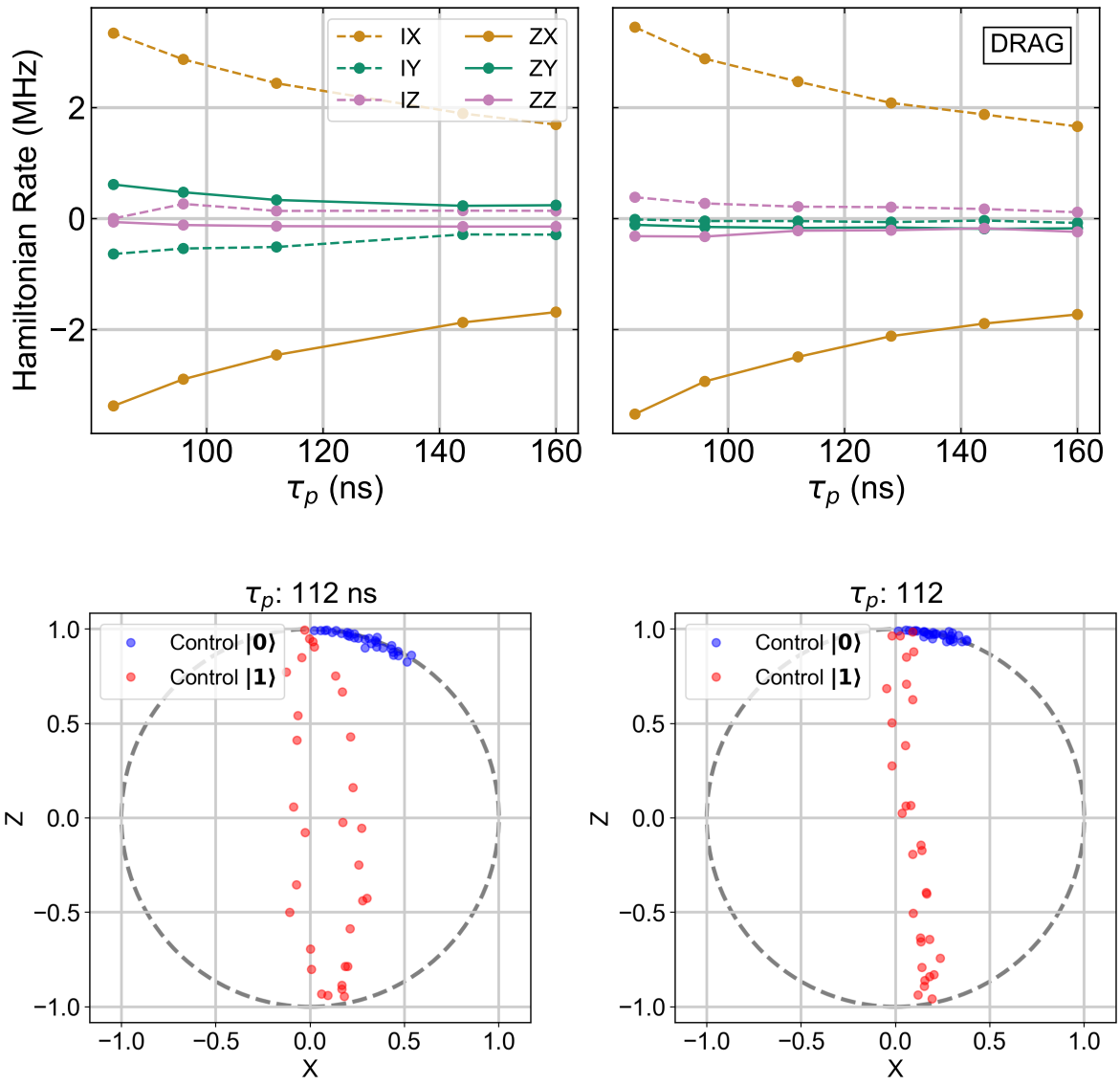


Figure 3.9: **CNOT gate time sweep, Hamiltonian tomography (a)** Hamiltonian rates of calibrated CNOT gates as a function of pulse time τ_p without 2D DRAG leakage calibration, and **(b)** with 2D DRAG leakage calibration. **(c)** Target qubit dynamics for $\tau_p = 112\,\text{ns}$ in the $X - Z$ Bloch plane without DRAG and **(d)** with DRAG.

when Q_c is prepared in $|0\rangle$, and undergoing rotations approximately about the x-axis with Q_c in $|1\rangle$. Some IY and ZY dynamics are seen in the data without DRAG, and IZ and ZZ in the data with DRAG. However, slower oscillations are difficult to characterize precisely

due to the limited frequency bandwidth of Hamiltonian tomography experiments.

While analyzing Hamiltonian tomography data gives insight into the dynamics, we next seek to quantify the performance of the CNOT gates according to standardized measures of gate fidelity, namely Randomized Benchmarking (RB) experiments [63, 64]. The RB protocol involves running randomly sampled gate sequences of increasing length that simplify to the identity operation. These random paths through Hilbert space have the effect of averaging general gate errors, which can be coherent or incoherent, into a depolarizing channel, a process called "twirling". The RB protocol measures the "survival probability" (plotted on the y-axis of Figure 3.10 (a)), or how well the qubits returned to their initial state, at the end of the random net-identity path through Hilbert space, as a function of sequence length. As the paths get longer, errors in the gates accumulate further, reducing the probability that the qubits will return to their initial state. When the paths are sufficiently random (as is accomplished by using random sequences of Clifford gates [64]), the survival probability undergoes an exponential decay. Appendix A.1 gives more detailed information about RB and other benchmarks. Using Quantum Benchmark's software TrueQ [65], we measured the average fidelity and, by monitoring the $|2\rangle$ state, the leakage rate [47, 66] of the Direct CNOT gate for a variety of gate times τ_p. Figure 3.10 (a) shows example RB decay curves for $\tau_p = 112\,\text{ns}$ for CNOT gates with (blue) and without (gold) DRAG correction. The decay rates are comparable with and without DRAG, and this is consistent across all gate times studied, as shown in Figure 3.10 (b). Here $e_F = \frac{d^2-1}{d^2}\left(1 - p_{RB}\right)$ is the process infidelity of the average Clifford gate, is between 3.5% and 7%. This is substantially above the decoherence limit indicated with the dashed gray line, suggesting the presence of substantial coherent error in the gates. The process infidelity from decoherence for two qubits is approximately $e_D = \frac{1}{4}\tau_p\left(1/T_1^{(c)} + 1/T_1^{(t)} + 2/T_2^{(c)} + 2/T_2^{(t)}\right)$. For the qubits studied in this chapter, the average T_1 and T_2^{echo} times of around $70\,\mu$s and $120\,\mu$s, respectively, however even using a low estimate of $T_1 = T_2 = 50\,\mu$s predicts e_D about one order of magnitude lower, between 0.5% and 0.8%. Setting aside the high RB process infidelity, the 2D DRAG calibration does successfully reduce leakage, as shown in Figure 3.10 (c) and (d). In part (c), the $|2\rangle$ state population of Q_c is plotted as a function of RB sequence length, for the CNOT gate again $\tau_p = 112\,\text{ns}$, as in part (a). The gate with DRAG shows substantially slower decay rate and reduced steady-state population. This is further reflected in part (d), where the leakage rate L_1 is plotted for all pulse times studied and leakage is shown to be robustly reduced when using 2D DRAG. This experiment highlights the value of pulse spectrum management on managing error sources in CR gates, particularly on qubits with smaller anharmonicity, such as those optimized for qutrit processors [32].

These experiments showed a robust reduction of leakage with 2D DRAG calibration, without sacrificing gate fidelity. However, the absolute RB error rates are well above the decoherence limit. This points us to sources of coherent error, including the static ZZ and calibration errors.

Interestingly, this direct CNOT gate calibration approach yielded a lower error CR gate on a separate qubit pair, denoted Sample B in Table 3.1, which had much higher intrinsic

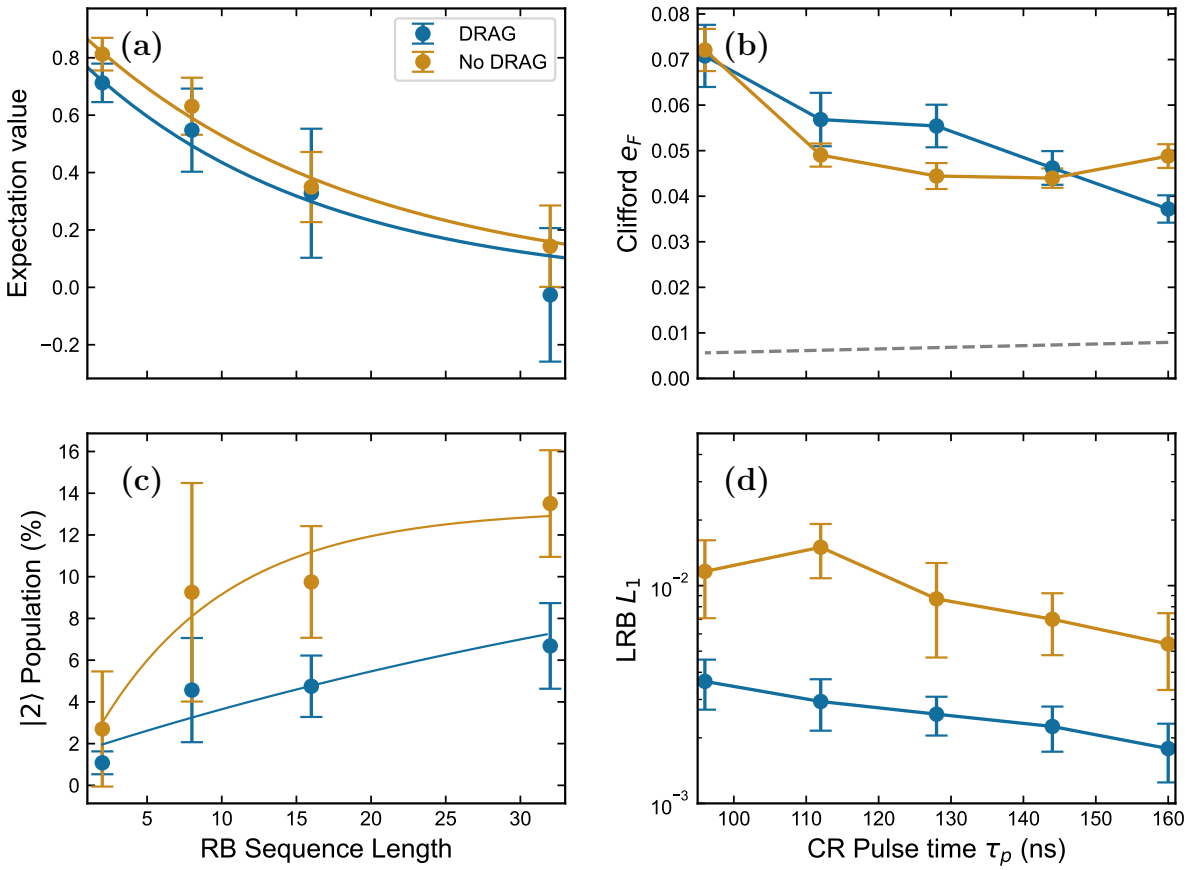


Figure 3.10: **Direct CNOT gate time sweep, leakage benchmarking (a)** Two-qubit RB survival probability decay curves for $\tau_p = 112\,\text{ns}$, with and without DRAG correction. See the main text for description of survival probability. Decays are comparable, indicating similar gate performance. **(b)** Clifford process fidelity e_F from RB experiments for varying τ_p. **(c)** Leakage RB decay curves at $\tau_p = 112\,\text{ns}$, showing reduced decay and steady-state Q_c $|2\rangle$ population when DRAG is used. **(d)** LRB leakage rate L_1 while varying τ_p. Leakage is robustly reduced with DRAG.

local drive terms (i.e. classical crosstalk) present. Table 3.3.1.1 shows a comparison of the Hamiltonian rates extracted with Hamiltonian tomography between Samples A and B, for gates with approximately equal durations of $\tau_p = 144\,\text{ns}$ and $\tau_p = 147\,\text{ns}$, respectively. The ZZ and ZX rates are similar, but the IX and IZ terms on sample B are about 15 times larger than Sample A. This results in the same polar angle tilt $\sin^{-1}\left(\frac{IZ}{IX}\right) \approx 4.5°$ of the unconditional rotations. However, as noted earlier, the error introduced by the ZZ term may be partially mitigated by the larger natural IX drive on Sample B, as predicted [34] and observed [40]. While we were not able to achieve comparable gate fidelities on sample A,

Rate (MHz)	Sample A	Sample B
IX	1.894890	26.771392
IY	-0.285226	0.045184
IZ	0.143450	-2.041381
ZX	-1.871643	2.024811
ZY	0.231025	-0.490854
ZZ	-0.143615	-0.132003
$e_F^C \times 10^2$	4.6(4)	2.0(2)

Table 3.2: Hamiltonian rates for direct CNOT gates on Sample A (B) (circuit parameters in Table 3.1) with $\tau_p = 144\,\text{ns}$ ($\tau_p = 147\,\text{ns}$), and the resulting Clifford process fidelity e_F^C measured by RB. The e_F^C quoted for sample A corresponds to the results plotted in Figure 3.9 (b). The result for sample B corresponds to an experiment like that shown in Figure 3.11 (a).

given their comparable circuit parameters and single-qubit gate fidelities, adding additional drive on the target qubit may have improved the gate performance [40]. Notably, because the leakage was lower on Sample B owing to the larger anharmonicity of the transmons, we were able to mitigate the large IZ error using DRAG, reducing the polar angle tilt of the unconditional drive, however this did not improve the direct CNOT gate RB process fidelity. In the next section, I will describe approaches to further distinguish and reduce error mechanisms of the direct CNOT gate on Sample B.

3.3.1.2 Error budgeting the direct CNOT gate

With our demonstrated ability to reduce leakage errors, we next describe a broader suite of benchmarking experiments, used to determine the dominant noise sources of the direct CNOT gate. The calibration procedure here is the same as was described earlier, except leakage was lower for this gate owing to the larger anharmonicity of the qubits. Because of this, a single CR DRAG parameter was instead calibrated to reduce IZ error of the CR gate, as mentioned in the direct CNOT calibration procedure. Rather than sweeping pulse time for this pair, we focused on thoroughly benchmarking a single gate time of $\tau_p = 147\,\text{ns}$, with a 30 ns cosine ramp. The benchmarking protocols we used were interleaved randomized benchmarking (IRB), extended randomzied benchmarking (XRB), also known as purity benchmarking, and Cycle Benchmarking (CB). The first protocol is a commonly-used benchmark for estimating process error for a gate of interest, rather than process error of the average Clifford gate, as was reported in the previous section. Further, XRB distinguishes coherent and incoherent error sources [67]. Finally CB enables a tighter bound on the process error, since the randomizing gate set is the Pauli group, rather than the Clifford group [68]. For more information about the benchmark protocols, see section A.1.

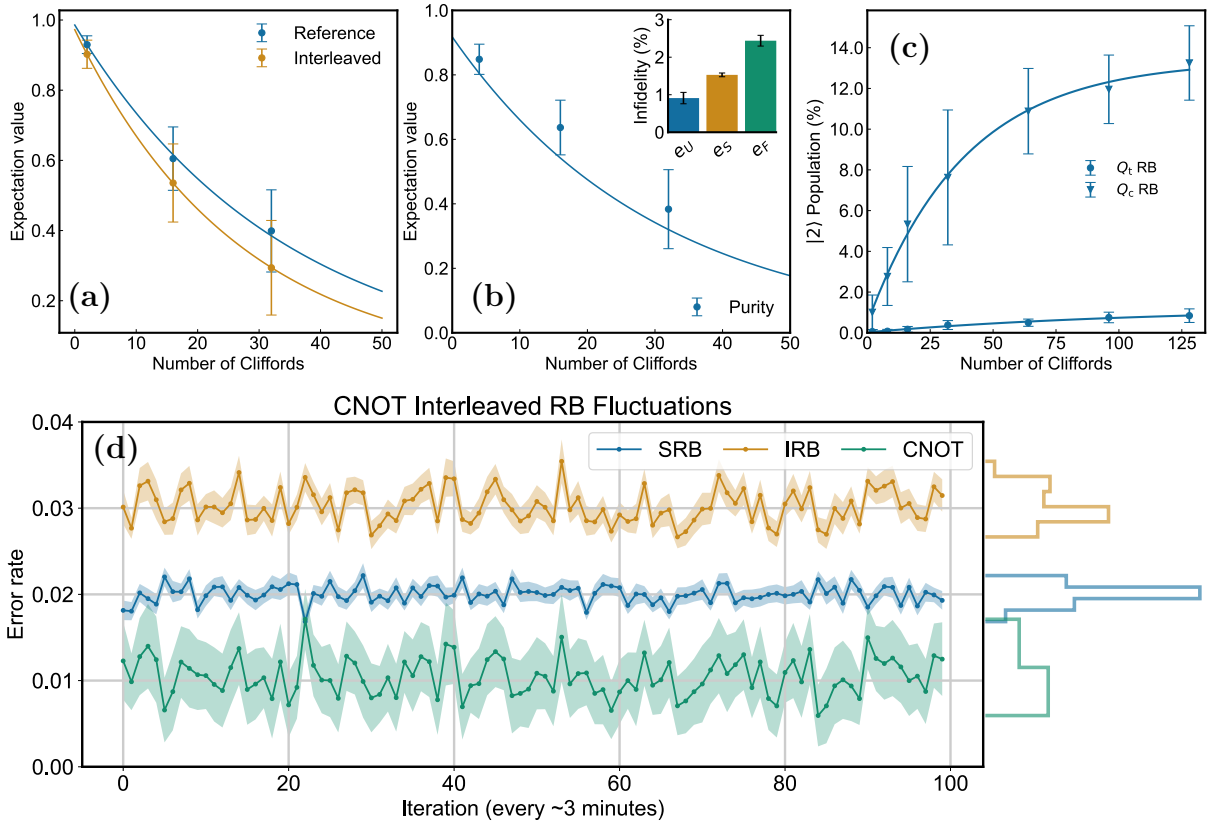


Figure 3.11: **Benchmarking Direct CNOT Errors (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved intrinsic CR gate. From reference and interleaved decay parameters $p_{RB} = 0.972(1)$ and $p_{IRB} = 963(2)$, respectively, we extract a CR process error $e_g = 7(3) \times 10^{-3}$ (see equation A.7). **(b)** Purity benchmarking shows a unitary (stochastic) $e_U = 11.308(1) \times 10^{-3}$ ($e_S = 15.2925(1) \times 10^{-3}$) error component to the total Clifford process error $e_F^C = 26.600(1) \times 10^{-3}$. **(c)** Leakage Randomized Benchmarking results, obtained by monitoring the $|2\rangle$ state of each transmon when performing RB. The leakage per Clifford L_1 are estimated to be $L_1 = 3.4(4) \times 10^{-3}$ ($0.11(5) \times 10^{-3}$) for Q_c (Q_t). **(d)** Repeated IRB measurements to quantify gate calibration and benchmarking stability. Note that the y-axis is average gate error $r = \frac{d-1}{d} e_F$, where e_F is the process infidelity of the protocol. The statistical average gate process error is $11.3(3) \times 10^{-3}$.

Benchmarking results for this gate are shown in Figure 3.11. The highest gate fidelity, as measured by IRB and shown in Figure 3.11 (a), was 99.5(2)%, with worst/best case fidelity bounds of 90.8(3)%/99.97(2)% [68]. However, measuring IRB experiments repeatedly for several hours gave an average gate fidelity of 99.1%.

To understand the dominant error sources, we first examine the XRB data, shown in Figure 3.11 (b). The inset shows the breakdown of the total Clifford process infidelity e_F, in terms of unitary or coherent, and stochastic or incoherent, error sources, denoted e_U and e_S, respectively. First, it's worth comparing e_S to a naive estimate of the decoherence error as calculated from the T_1 and T_2^{echo} times of the sample. This is calculated as,

$$e_D \approx \frac{\langle \tau_C \rangle}{4} \left(\frac{1}{T_1^{(c)}} + \frac{1}{T_1^{(t)}} + \frac{2}{T_2^{(c)}} + \frac{2}{T_2^{(t)}} \right), \tag{3.22}$$

which we estimate to be $e_D \approx 7.5 \times 10^{-3}$ given an average clifford time $\tau_{\text{cliff}} = 383\,\text{ns}$, which is a bit under half of the measured e_S. The source of this discrepancy is a topic of further study, though it has been predicted that the Stark shifting of the control qubit during the CR drive will introduce additional decoherence channels via the (less coherent) higher levels of Q_c [34].

While the XRB, IRB, and LRB experiments gave some insight into error sources, the high coherent error of the gate means that there is high uncertainty in the worst-case error of the gate [69]. To get an improved estimate, we perform Cycle Benchmarking (CB) on the gate as well, which is described in the Appendix A.1. Similar to RB, in CB, the probability of the qubits returning to the expected state is measured after a sequence of random gates. This is the y axis of Figure 3.12 (a). Cycle Benchmarking is distinct from Randomzied Benchmarking in that the randomizing gates are n-qubit Pauli gates, rather than Clifford gates. This has the benefit that n-qubit Pauli gates do not need any entangling gates, and therefore the twirling operation has higher fidelity. However, unlike Clifford gates used in RB, Pauli gates do not twirl general errors into a depolarizing channel. Instead, they twirl errors into a Pauli channel, and so, in order to characterize the errors of this channel, each Pauli decay rate of this Pauli channel must be measured. This is done by preparing all eigenstates of the Pauli basis, running the CB circuits, and measuring the decay of the probability that the qubits are in the expected state, given the prepared input state and the random circuit instance. See Appendix A.1 for further information. Results for CB on the direct CNOT gate are shown in Figure 3.12, give an estimated process fidelity for the CNOT gate $e_g = 7.2(5) \times 10^{-3}$, with lower/upper bounds $e_{g,L}/e_{g,U}$ of $1.3(2)/38.3(9) \times 10^{-3}$, assuming fully stochastic/coherent noise, respectively [69]. These bounds are much tighter than the estimate from IRB/XRB experiments, which estimated $e_g = 7.5 \times 10^{-3}$, and $e_{g,L}/e_{g,U}$ of $0.925(1)/59.39(1) \times 10^{-3}$. Furthermore, CB gives an estimate of the cycle fidelity during a randomly-compiled algorithm, and therefore CB metrics are useful in the context of applications [70].

While the majority of the noise for this direct CNOT gate is incoherent, nearly 1% of the error is from coherent error sources. Coherent error sources are essentially limited to

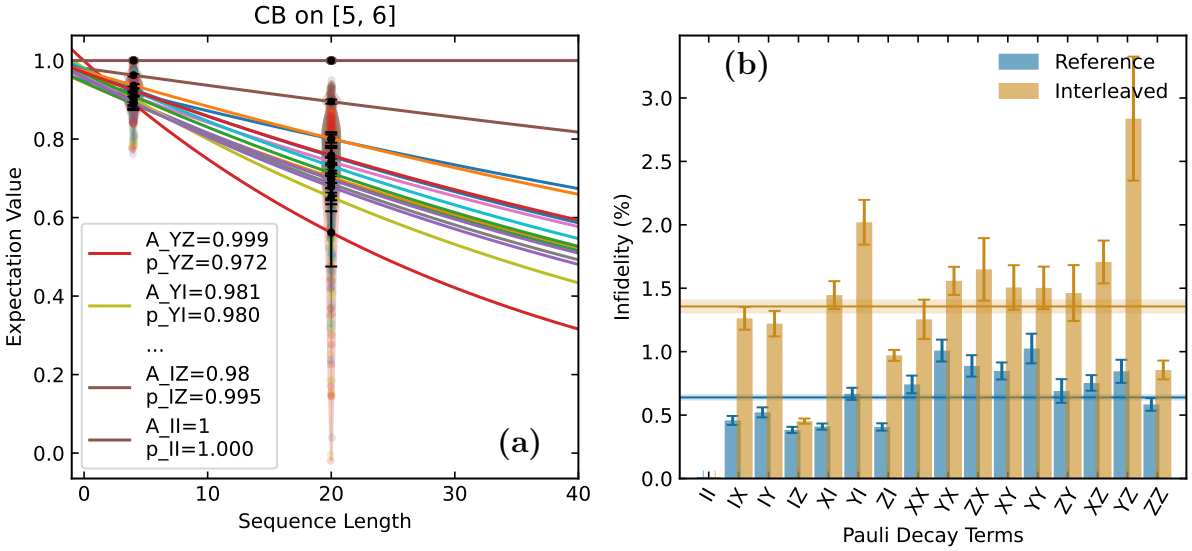


Figure 3.12: **Cycle Benchmarking the Direct CNOT gate (a)** Exponential decays of the survival probabilities of the Pauli basis preparation states for the direct CNOT gate. **(b)** Process infidelities extracted from the exponential decays, for the case with the direct CNOT cycle (gold) and the identity cycle (blue). See A.1 for more information on Cycle Benchmarking.

calibration errors and the static ZZ interaction. In effort to reduce coherent error in the gate, we next explore a different calibration approach, where the unitary of the fully entangling CR gate, which is locally equivalent to a CNOT gate, is measured and used as a native gate.

3.3.2 Intrinsic CR gates

With the demonstrated ability to reduce leakage in our toolkit, and our identification of substantial coherent error in the direct CNOT gate, we next explored ways to reduce coherent errors. The approaches are two-fold. First, we develop an error-amplification calibration procedure, to more precisely calibrate the entangling portion of the CR gate, namely the CR pulse amplitude that maximizes R. Then, we use closed-loop optimization to find the single-qubit gates needed to map the *intrinsic* CR unitary U_{CR} to a CNOT, by maximizing a loss function, e.g. Bell state fidelity. Then, since the CR gate is locally-equivalent to a CNOT, we use the intrinsic CR gate as the native gate when compiling circuits, rather than using a CNOT.

The error-amplification procedure that we employ to maximize entanglement is one similar to other experiments used in pulse calibration: it leverages pulse repetition to accumulate coherent errors of interest. The protocol, depicted in Figure 3.13 (a), maximizes R with re-

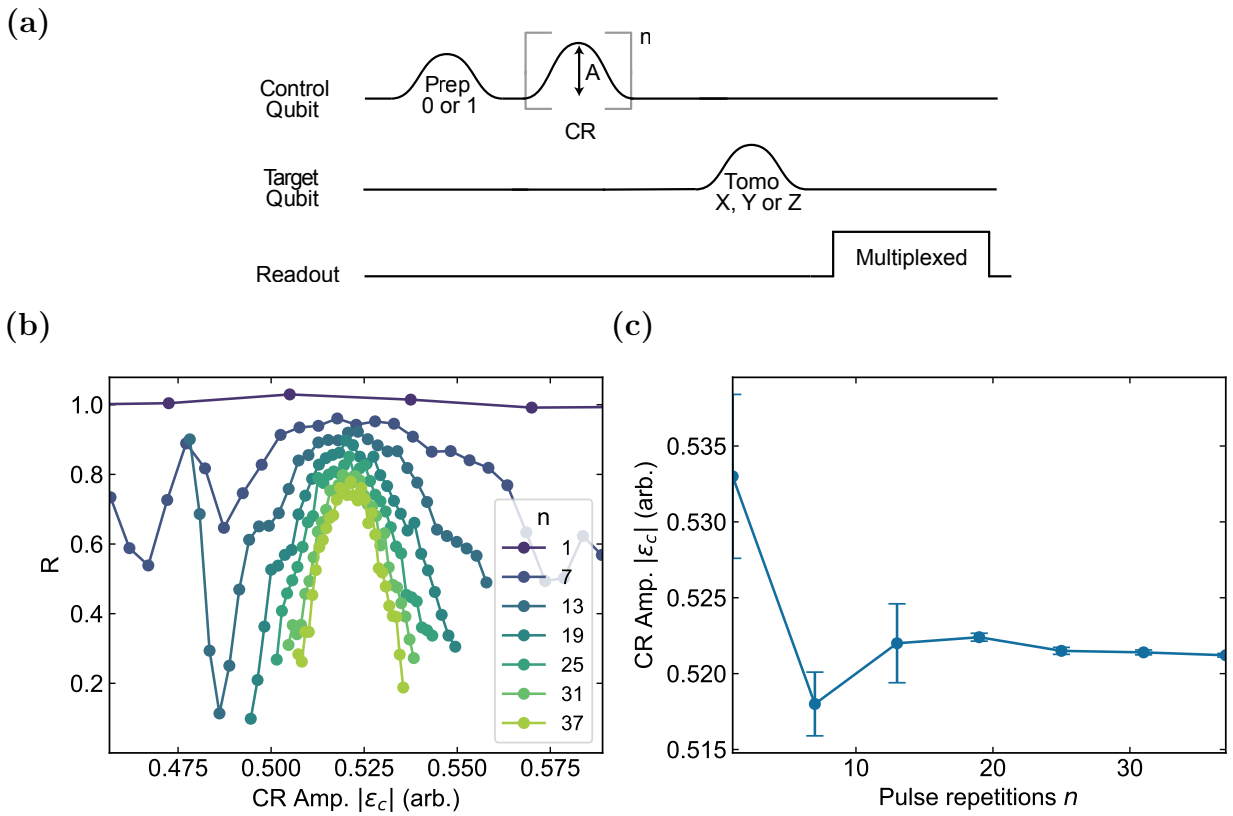


Figure 3.13: **CR amplitude calibration experiment with error-amplification. (a)** Pulse sequence for the experiment. The CR pulse is fixed in time, with the pulse amplitude swept. This is repeated for different numbers of pulse repetitions n. **(b)** Experimental data for several values of n. **(c)** Fitted optimal value, with uncertainties, as a function of pulse repetitions n.

spect to $|\varepsilon_c|$. We increase the sensitivity by repeating the pulse an odd number of times, as shown in Figure 3.13 (b). Interestingly, there are dips observed in the behavior or R vs ε_c for some $n > 1$. This behavior, which is also observed in simulation, may be due to the static ZZ interaction, which is not echoed away in this protocol, interfering with the CR dynamics. Additionally, the maximal value of R is reduced with increasing n due to decoherence. Figure 3.13 (c) shows the parabola fit estimate as a function of n. At few n, the precision is lower, and around $n = 19$, the error in the fit estimate is reduced substantially.

With the improved calibration of the entangling portion of the CR gate, we know that we have a fully entangling, CNOT-like gate. As the direct CNOT calibration described previously suffered from limited precision in calibrating the local rotations with the additional drive pulses, and also introduced coherent errors during the CR pulse, another approach to

calibrate a CNOT is to simply append single-qubit rotations after the CR pulse, to realize the desired unitary. If we used exclusively CNOT gates as our 'native' entangling gate, this approach could be costly, as the CNOT gate time is longer due to the overhead of additional single-qubit gates after the CR gate (this is what we were trying to avoid with the direct CNOT gate). However, with modern compilation software that supports more general two-qubit unitaries as native gates, the intrinsic CR gate can itself be used as the native entangling gate, avoiding this added cost. In order to use the compiler effectively, however, the intrinsic CR unitary needs to be accurately determined. To determine the CR unitary, we employ closed-loop optimization. In particular, we optimize the parameterized calibration circuit shown in Figure 3.14 to realize a CNOT gate. The circuit parameters ϕ_1 and θ_c are present to modify the ZI and IZ type errors, and the $R_T\left(\vec{\varphi}_2\right)$ gate is an arbitrary Bloch sphere rotation in the $ZXZXZ$ decomposition [41]. In terms of cost function, we have had most success by minimizing infidelity of Bell states $|\Phi\rangle = \left(|00\rangle + |11\rangle\right)/\sqrt{2}$ and $|i\Phi\rangle = \left(|00\rangle + i\,|11\rangle\right)/\sqrt{2}$. Other experimental works have used RB cost functions with a fixed sequence length [49]. We have explored this as well, however we found that *ab initio* calibration is most successful with the Bell state infidelity. We note that RB sequences, due to their error-randomizing nature, tend to reduce sensitivity to the parameters, but may be useful when the gate has already a sufficiently high fidelity.

To optimize over the parameter space, we use the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [71]. CMA-ES is a gradient-free optimizer that is similar to a particle-swarm scheme in that multiple trial solutions are run for each optimization step, and the selection of the next region where parameters will be taken is determined by fitting the covariance of the loss function along the parameter space, and moving the parameter sampling region in the direction of greatest covariance. The lower plots in Figure 3.14 show the loss function convergence of the Bell state fidelity preparation circuit on the left, and the parameter convergence in the right plots. The points in the right plots indicate different samples within the parameter sampling region of the optimizer, at a given optimization step.

After the intrinsic CR unitary U_{CR} is determined via closed-loop optimization, we specify U_{CR} as a native two-qubit gate in the Keysight/Quantum Benchmark TrueQ [65] circuit transpiler, where the CR gate can then be substituted for the CNOT gate in arbitary circuits before running on our hardware, as indicated in Figure 3.15. The top circuit indicates a subset of an arbitrary quantum circuit, with rounds of arbitary single-qubit gates R in between rounds of entangling gates. The transpilation process begins by inserting the calibrated quantum circuit for realizing a CNOT from a CR gate. Then, because any arbitrary single-qubit gate can be expressed via three arbitrary-angle Z gates with X_{90} gates in between $Z_{\phi_1}X_{90}Z_{\phi_2}X_{90}Z_{\phi_3}$ [41], the parameters of neighboring single-qubit gates R are simply modified to include the single-qubit rotations from the CNOT calibration circuit, $R \rightarrow R'$.

Note that, in general, quantum circuits may not have the form as shown in the top circuit; however, any circuit that is randomly-compiled [70], a compilation strategy that has been shown to improve the reliability of quantum algorithms by tailoring coherent noise into stochastic noise [39], will have this form. Since any quantum algorithm benefits from tailoring coherent noise into stochastic noise [70], this compilation strategy will not result in

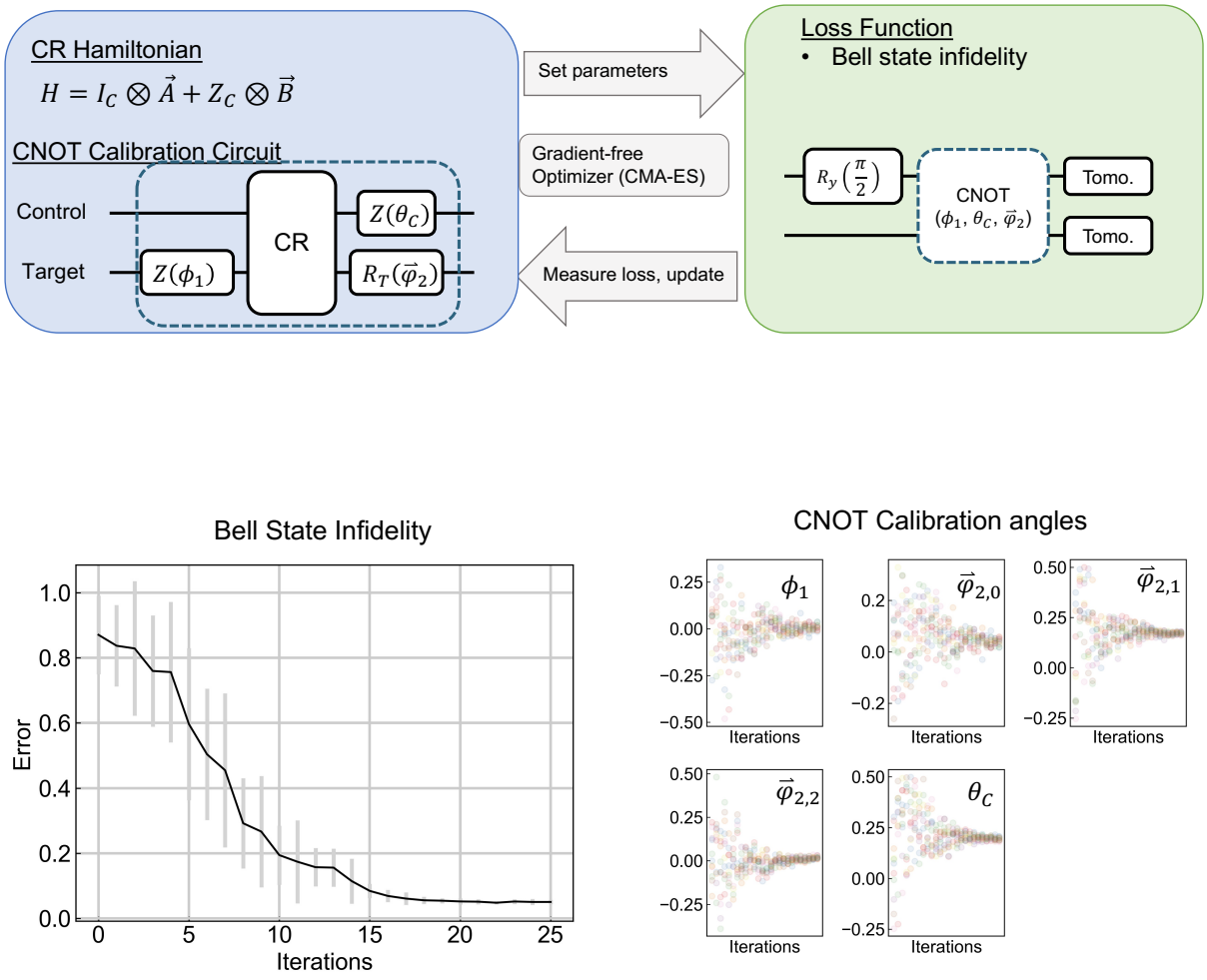


Figure 3.14: **Closed-loop optimization for calibrating local rotations to realize a CNOT from the CR gate.** The CNOT gate is parameterized by the circuit shown in the top left, where $R_T(\vec{\varphi}_2)$ is an arbitrary single-qubit rotation and the angles for all gates are angles for virtual-Z gates, which have nearly perfect fidelity [41]. The CNOT calibration circuit is inserted into a loss function circuit, such as a Bell state fidelity measurement shown on the right, and a classical optimizer minimizes the loss function in a closed-loop fashion. An example minimization of the loss function using the CMA-ES optimizer [71] is shown at the bottom left, and the parameter convergences are shown at the right. The points in the right plots indicate different parameter trials tested for each optimization step.

increased gate depths for any randomly-compiled circuit.

In Figure 3.16, benchmarking results are shown for an intrinsic CR gate calibrated on the same qubit pair as described in the previous section. The gate duration and pulse shape were modified from the direct CNOT gate, from $\tau_p = 147\,\text{ns} \rightarrow 250\,\text{ns}$, and from a ramp

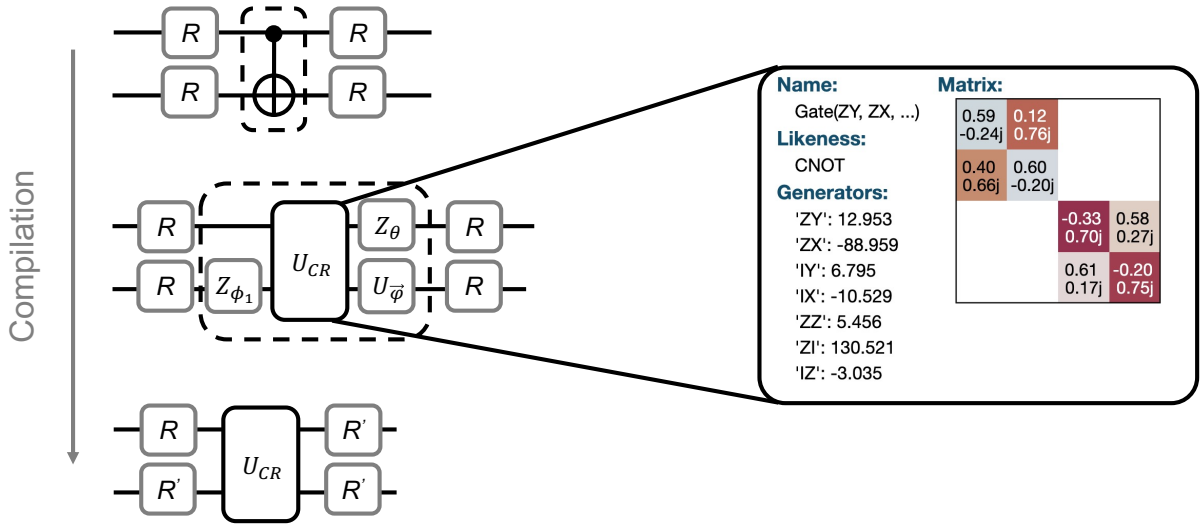


Figure 3.15: **Compiling general circuits with intrinsic CR gates.** General circuits, and particularly randomly-compiled circuits [39, 70],consist of alternating rounds of single- and two-qubit gates. The single-qubit gates R are arbitrary single-qubit rotations, which can be decomposed into three arbitrary-angle Z gates with X_{90} gates in between $Z_{\phi_1}X_{90}Z_{\phi_2}X_{90}Z_{\phi_3}$. Therefore, the single-qubit gates within the CR-based CNOT calibration, shown in the middle step of the compilation, can be folded into the other single-qubit gates without changing the depth of the circuit. An example U_{CR} is shown on the right, with information output from the Keysight/Quantum Benchmark TrueQ software [65].

fraction $\tau_r = 30\,\text{ns}$ to a "full ramp" cosine pulse shape. These changes were made in effort to reduce leakage on this pair, which was indeed a successful strategy. The CR gate process error, as estimated by Cycle Benchmarking, is $e_g = 11.52(3) \times 10^{-3}$, with worst/best case errors $e_{g,U}/e_{g,L}$ of $29.27(4) \times 10^{-3}/4.5(1) \times 10^{-3}$. This value is comparable to the average estimate over repeated IRB measurements for the direct CNOT gate, but is higher than the best reported direct CNOT gate error, and that which was measured with CB for the direct CNOT gate.

A full comparison of the suite of benchmarks that were measured for both gates is shown in Figure 3.17, with descriptions of quantities in Table 3.3.2. Both top and bottom plots are the same data, but the top is log-scale on the y-axis, facilitating differentiation of the smaller parameters, and the bottom plot is in linear scale for comparing larger errors. From these data, we can make a few remarks:

- Cycle Benchmarking reliably finds narrower bounds on the best/worst case process errors of the gate $e_{g,L}$, $e_{g,U}$.
- As measured by XRB, the intrinsic CR gate, despite being 100 ns longer, has a smaller

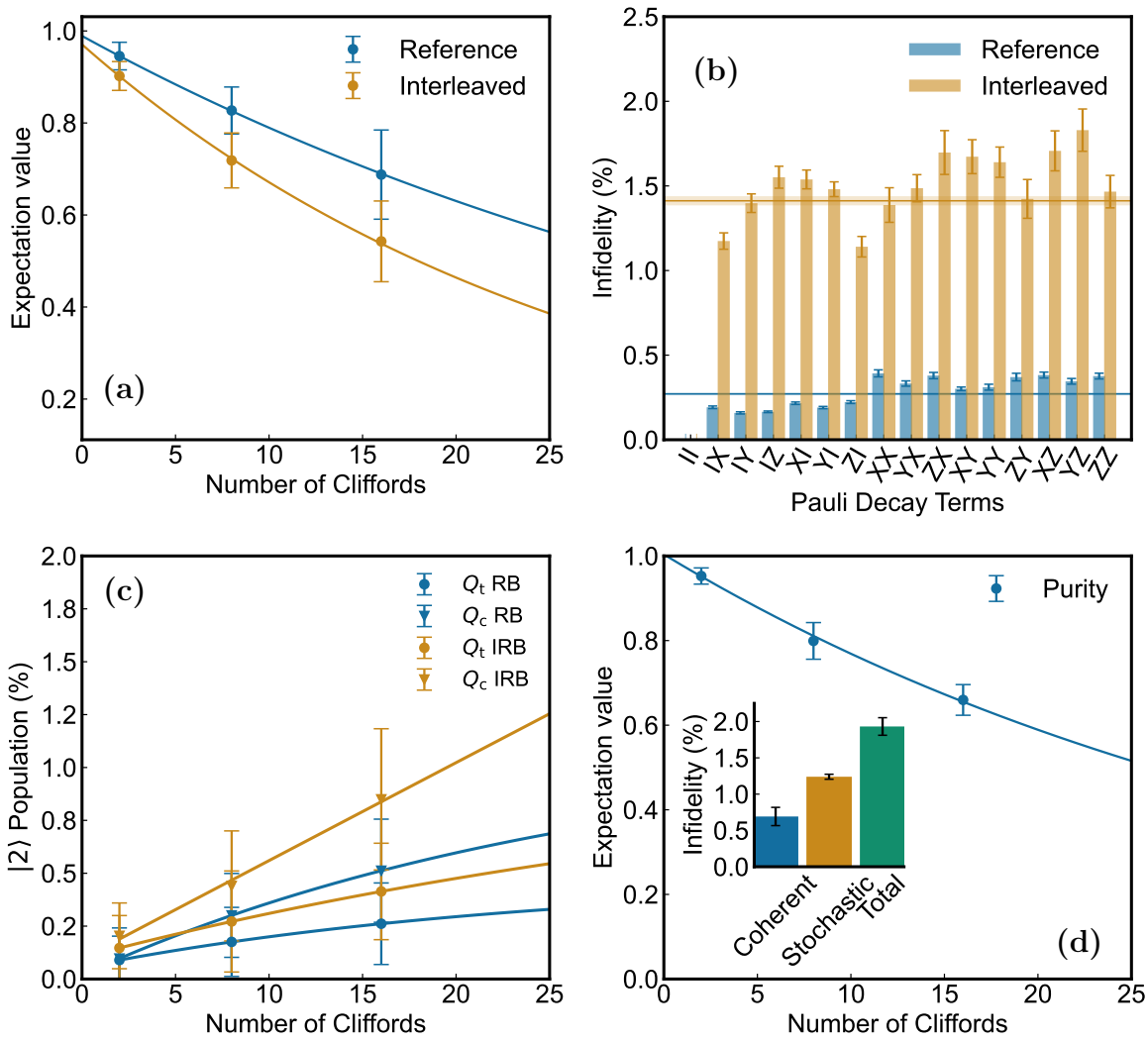


Figure 3.16: **Intrinsic CR Benchmarking (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved intrinsic CR gate. **(b)** Cycle Benchmarking results. The error rate e_i is obtained for the prepared Pauli eigenstate i from the exponential decay of sequence fidelity p_i with increasing sequence length. A larger error rate for a given Pauli eigenstate indicates errors in the cycle that do not commute with the compiled Pauli term. **(c)** Leakage Randomized Benchmarking. By monitoring the $|2\rangle$ state of each transmon when running an IRB experiment, the reference and interleaved $|2\rangle$ state population data are fit to an exponential model to extract the leakage-per-gate for each transmon. **(d)** Purity Benchmarking distinguishes coherent from stochastic errors by measuring the decay of the purity (main plot) of the two-qubit density matrix by performing state tomography for each random Clifford RB sequence, and comparing the purity decay to the RB decay. Inset: breakdown of Clifford process infidelity between coherent and stochastic contributions.

Metric Name	Symbol	Description	Protocol
Clifford Process error	$e_F^{(C)}$	Clifford process infidelity	RB
Clifford unitary error	e_U	Coherent error contribution to $e_F^{(C)}$	XRB
Clifford stochastic error	e_S	Stochastic error contribution to $e_F^{(C)}$	XRB
Gate process error	e_g	Process error measured via reference, interleaved RB	IRB, RB
Lower bound of e_g	$e_{g,L}$	Best case estimate given unitarity	IRB, RB, XRB
Upper bound of e_g	$e_{g,U}$	Worst case estimate given unitarity	IRB, RB, XRB
Leakage rate	L_1	Gate leakage out of computational subspace	LRB

Table 3.3: Names, symbols, and descriptions of benchmarking metrics used in this thesis. Similar metrics are reported but obtained via CB, however their interpretation is the same. Subsets for isolated and simultaneous single-qubit Clifford RB are reported. See section A.1 for detailed descriptions of these parameters.

stochastic error e_S than the direct CNOT gate. It even has a smaller unitary error e_U. The overall improvement in two-qubit SRB with the intrinsic CR gate is notable, given the comparable single-qubit SRB data for each experiment. Interestingly, for the average clifford time with the intrinsice CR gate of $\tau_{\text{Cliff}} = 542\,\text{ns}$, the predicted decoherence error $e_D = 1.06\%$ is close to the measured $e_S = 1.2647(1)\%$ via XRB, while the predicted e_D differed substantially from e_S for the direct CNOT gate.

- Using a full pulse ramp for the intrinsic CR gate lowered the leakage rate L_1 substantially. However, even for the direct CNOT gate, L_1 here is comparable to the DRAG-mitigated CR gate on sample A (see Figure 3.10 (d)).

- Based on a simple model for the coherent error from static ZZ during the gate presented in Section 2.3.1, the static ZZ error for the 147 ns direct CNOT gate and the 250 ns CR gate are $e_{ZZ} \sim 4.5 \times 10^{-3}$ and $e_{ZZ} \sim 15 \times 10^{-3}$, respectively. These errors, while being over-estimates, convey that the static non-commuting ZZ interaction is a substantial source of error for these gates. The approach in Chapter 4 addresses this challenge.

- The simultaneous single-qubit gate process errors are a few parts in 10^{-3}, suggesting that measuring two-qubit gate errors below that is an uncertain task with these as the

twirling operations.

In summary, we have addressed several challenges towards maximizing CR and CNOT gate fidelities. We have demonstrated leakage reduction using 2D DRAG pulses on low-anharmonicity transmons. We explored a faster scheme for implementing a CNOT, called the direct CNOT, and identified ways to reduce coherent errors, via local driving of Q_t [40] as evidenced by the lower errors for direct CNOT gates on Sample B, which had more local driving than Sample A due to crosstalk. We also developed a novel calibration scheme with intrinsic CR gates as the native entangling gate, which is operationally simpler.

When considering scaling the system calibrations beyond two qubits, the optimizer-based approach to calibration may hold even more promise, due to its general, black-box approach. Still, knowledge of the physical interactions present on a quantum processor, like spectator errors from static ZZ [72], crosstalk, and frequency collisions are important for realizing high fidelity quantum operations on larger systems. Interestingly, the physical interaction discussed in Chapter 4 was discovered during our calibrations of CR gates on the four-qubit processor of which Sample B is a part.

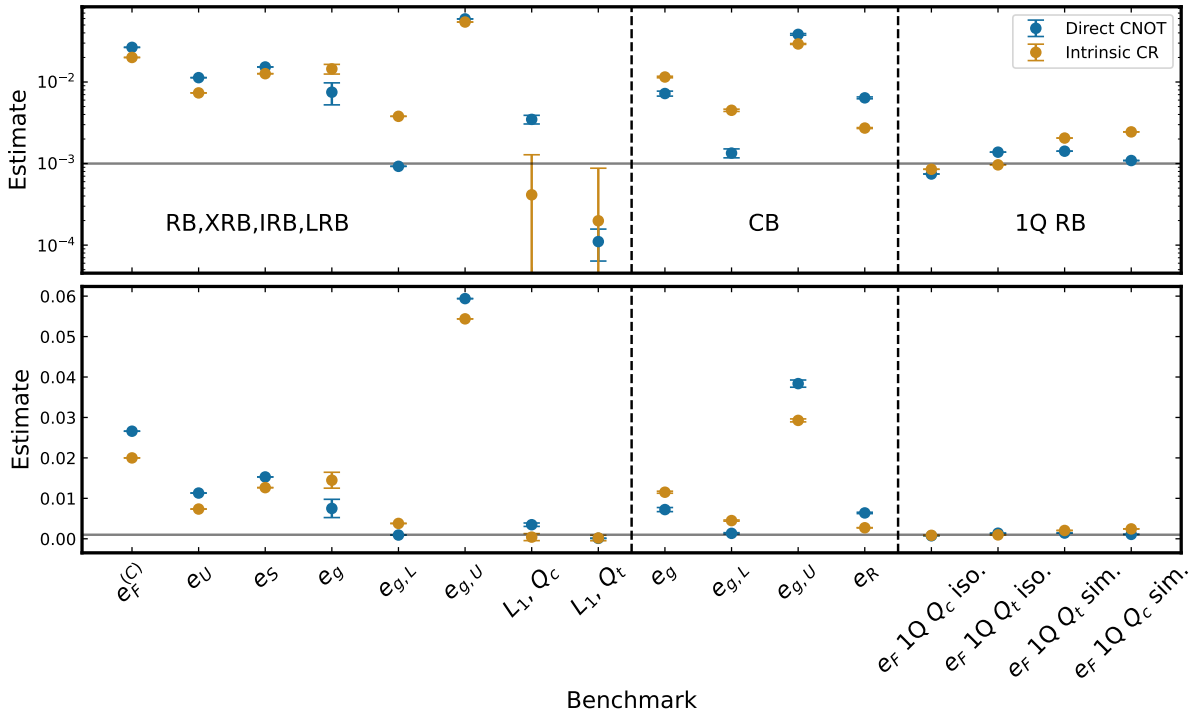


Figure 3.17: **Comparing Performance of direct CNOT and intrinsic CR gates on sample B.** The points left of the left-most dashed vertical line are from the two-qubit RB experimental suite, comprised of RB, IRB, XRB, and LRB. The middle points between the two dashed vertical lines are from CB, and the right-most points are from single-qubit RB experiments, both isolated and simultaneous [45].

Chapter 4

Microwave-Activated Tunable Coupling

During our studies in Chapter 3 on the Cross Resonance gate, we observed another entangling interaction on our hardware. While applying a crosstalk-cancellation drive during a CR gate on a spectator qubit neighboring the control qubit, an enhanced spectator-control ZZ interaction was observed that was generated by this simultaneous, off-resonant drive. This interaction has the notable benefit of commuting with the static ZZ coupling, suggesting that with this, a ZZ-error-free two-qubit gate could be implemented, giving a performance boost over the CR gate. Furthermore, this interaction offers the possibility for decoupling qubits, by cancelling the static ZZ interaction with minimal hardware resources. Approaches utilizing qubits and/or couplers with flux-tunable transition frequencies have shown high tunability [73, 74], full decoupling of qubits [75], and gate fidelities above 99.8% [76–78]. However, these approaches require additional control circuitry, and introduce decoherence channels.

A scheme with reduced hardware complexity while retaining the above capabilities would simplify the scaling of high-fidelity two-qubit gates to many-qubit processors. Charge-activated two-qubit gate schemes [8, 79–84] do not require additional control circuitry, and they are compatible with fixed-frequency qubits. Particularly, the cross resonance (CR) gate [8] has demonstrated a gate fidelity as high as 99.7% [85]. However, this scheme has residual idle ZZ coupling between qubits [86] due to interactions between their noncomputational transitions. This idle coupling causes correlated errors, dephasing, and spectator errors [72, 87–89]. Mitigating idle ZZ coupling without a flux-tunable coupler was done using dynamical decoupling [40, 54, 83] and using opposite-anharmonicity qubits [90], with added overhead in circuit depth and in hardware complexity, respectively. Recently, cancellation of the ZZ coupling between capacitively-coupled fluxonium qubits was also reported [91].

In this chapter, we describe experiments (first appearing in [9]) demonstrating a tunable, charge-activated ZZ interaction between two fixed-frequency transmon qubits [16] with fixed coupling. We show full cancellation of ZZ coupling between the qubits, and realize a CZ gate with a fidelity of 99.43(1)%. The interaction is realized by driving the transmons

simultaneously at a frequency between the $|0\rangle \to |1\rangle$ and $|1\rangle \to |2\rangle$ transitions, generating a conditional Stark shift. In contrast to the CR interaction, here the drive frequency can be tuned to avoid driving unwanted transitions, making this interaction suitable for use on large devices with crowded frequency spectra [92, 93]. Further, *in situ* control over the coupling enables controlled phase gates with arbitrary phase angles $\mathrm{CZ}(\phi)$, which are useful for noisy intermediate-scale quantum (NISQ) algorithms [94–96].

4.1 Principle

Consider two coupled, simultaneously driven transmons, as illustrated in Figure 4.1. The Hamiltonian of the two transmons, in the frame of the drive at frequency ω_d and making the Duffing approximation of the transmon [16], is given by

$$H_\mathrm{qb} = \sum_{i=\mathrm{c,t}} (\omega_i - \omega_\mathrm{d})\, a_i^\dagger a_i + \frac{\eta_i}{2} a_i^\dagger a_i^\dagger a_i a_i, \tag{4.1}$$

where for transmon i, a_i is the bosonic annihilation operator, ω_i is the transition frequency between $|0\rangle$ and $|1\rangle$, η_i is the anharmonicity, and $\hbar = 1$. Each drive term is given by $H_{\varepsilon_i} = \left(\varepsilon_i a_i + \varepsilon_i^* a_i^\dagger\right)$, where ε_i is the complex drive amplitude, and the coupling term with strength J is $H_J = J\left(a_\mathrm{c}^\dagger a_\mathrm{t} + a_\mathrm{c} a_\mathrm{t}^\dagger\right)$, where we denote the higher frequency transmon Q_c, and the lower frequency transmon as Q_t. The total Hamiltonian is

$$H = H_\mathrm{qb} + H_J + H_{\varepsilon_\mathrm{c}} + H_{\varepsilon_\mathrm{t}}. \tag{4.2}$$

The Stark-induced ZZ interaction can be understood through the lens of the CR effect: when driving the control qubit with amplitude ε_c at the target qubit frequency ω_t, the target experiences a drive amplitude $\tilde{\varepsilon}_n$ that depends on the control state $|n\rangle$ [8, 34]. In the qubit subspace, the CR drive realizes an entangling ZX interaction with rate $\mu = (\tilde{\varepsilon}_0 - \tilde{\varepsilon}_1)/2$. In the limit $\varepsilon_\mathrm{c}/\Delta_\mathrm{t} \ll 1$, detuning the drive frequency from the target qubit frequency $\Delta_\mathrm{t} = \omega_\mathrm{t} - \omega_\mathrm{d}$ results in a conditional Stark shift of the target qubit frequency $\tilde{\delta}_n$, where

$$\tilde{\delta}_n = \frac{\tilde{\varepsilon}_n^2}{\Delta_\mathrm{t}}. \tag{4.3}$$

The drive-induced ZZ interaction ζ then is given by $\zeta = \tilde{\delta}_0 - \tilde{\delta}_1$, which can be expressed in terms of μ as

$$\zeta = 2\mu\,(\tilde{\varepsilon}_0 + \tilde{\varepsilon}_1)/\Delta_\mathrm{t}. \tag{4.4}$$

This conditional Stark shift is much smaller than the CR rate $\zeta \ll \mu$. However, applying a drive simultaneously to the target qubit at amplitude ε_t modifies the total drive amplitude on the target to $\tilde{\varepsilon}_n + \varepsilon_\mathrm{t}$. Replacing $\tilde{\varepsilon}_n \to \tilde{\varepsilon}_n + \varepsilon_\mathrm{t}$ above, ζ then scales linearly to first order with ε_t

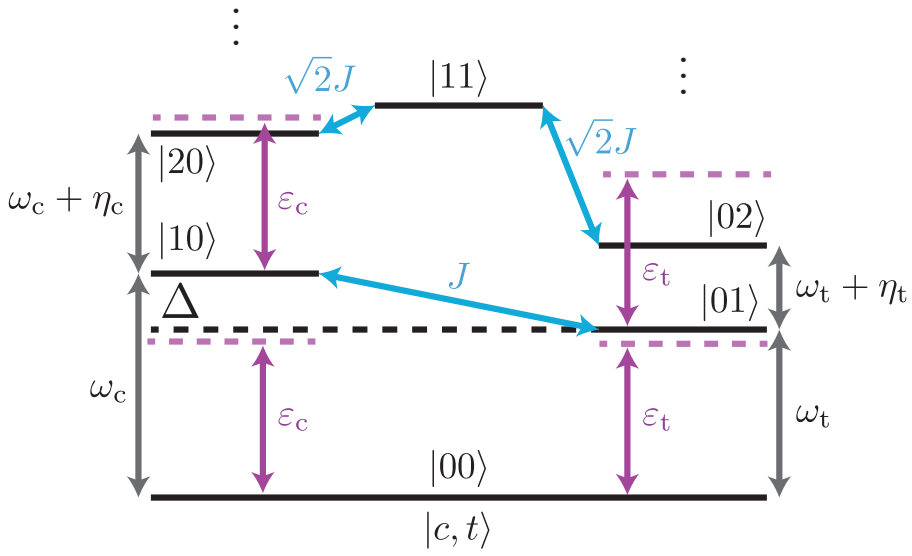


Figure 4.1: **Drive scheme for the Stark-induced ZZ interaction**. Transmons are detuned by $\Delta = \omega_c - \omega_t$ indicated by the black dashed line, and coupled via exchange coupling J. They are driven simultaneously with amplitudes ε_c, ε_t between frequencies ω_t and $\omega_c + \eta_c$, indicated with purple dashed lines. The simultaneous driving introduces conditional Stark shifts, i.e., a ZZ interaction.

$$\zeta = \frac{2\mu}{\Delta_t}\left(\tilde{\varepsilon}_0 + \tilde{\varepsilon}_1 + 2\varepsilon_t\right) + O(|\varepsilon_t|^2). \tag{4.5}$$

By driving both transmons simultaneously, one can generate a ZZ coupling between transmons with detuning Δ suitable for the CR gate, with interaction rates comparable to the CR effect. Note that the qubits are interchangeable and this can be generalized to systems that are not transmons [91].

4.2 Perturbation theory derivation

Here, we present a derivation of the effect using perturbation theory. The Hamiltonian of two (uncoupled) off-resonantly driven transmons in the drive frame is given by:

$$H_{\text{qb}} = \sum_{i=\text{c,t}} \left(\omega_i - \omega_d\right) a_i^\dagger a_i + \frac{\eta_i}{2} a_i^\dagger a_i^\dagger a_i a_i + \left(\varepsilon_i e^{i\phi} a + \text{h.c.}\right). \tag{4.6}$$

For a single transmon, truncated to the $|2\rangle$ level, the matrix looks like

$$H = \begin{pmatrix} 0 & \varepsilon e^{i\phi} & 0 \\ \varepsilon e^{-i\phi} & \Delta & \sqrt{2}\varepsilon e^{i\phi} \\ 0 & \sqrt{2}\varepsilon e^{-i\phi} & 2\Delta + \eta \end{pmatrix} \tag{4.7}$$

The dressed states of H to first order become:

$$|\overline{0}\rangle \approx |0\rangle + \frac{\varepsilon e^{-i\phi}}{-\Delta}|1\rangle \tag{4.8}$$

$$|\overline{1}\rangle \approx |1\rangle + \frac{\varepsilon e^{i\phi}}{-\Delta}|0\rangle + \frac{\sqrt{2}\varepsilon e^{-i\phi}}{-(\Delta+\eta)}|2\rangle \tag{4.9}$$

$$|\overline{2}\rangle \approx |2\rangle + \frac{\sqrt{2}\varepsilon e^{i\phi}}{\Delta+\eta}|1\rangle \tag{4.10}$$

$$\tag{4.11}$$

To estimate ζ, we first calculate the joint computational states dressed by the drives (without any coupling between the transmons). Then, we calculate the energy shift to first order by the coupling H_J.

The joint computational levels $|c,t\rangle$ dressed by respective drives $\varepsilon_c e^{i\phi_c}$ and $\varepsilon_t e^{i\phi_t}$ are given by:

$$|\overline{00}\rangle \approx |00\rangle + \frac{\varepsilon_t e^{-i\phi_t}}{-\Delta_t}|01\rangle + \frac{\varepsilon_c e^{-i\phi_c}}{-\Delta_c}|10\rangle + \frac{\varepsilon_c\varepsilon_t e^{-i(\phi_c+\phi_t)}}{\Delta_c\Delta_t}|11\rangle \tag{4.12}$$

$$\begin{aligned} |\overline{01}\rangle \approx |01\rangle &+ \frac{\varepsilon_c e^{-i\phi_c}}{-\Delta_c}|11\rangle + \frac{\varepsilon_t e^{i\phi_t}}{\Delta_t}|00\rangle + \frac{\varepsilon_c\varepsilon_t e^{-i(\phi_c-\phi_t)}}{-\Delta_c\Delta_t}|10\rangle \\ &+ \frac{\sqrt{2}\varepsilon_t e^{-i\phi_t}}{-(\Delta_t+\eta_t)}|02\rangle + \frac{\varepsilon_c e^{-i\phi_c}}{-\Delta_c}\frac{\sqrt{2}\varepsilon_t e^{-i\phi_t}}{-(\Delta_t+\eta_t)}|12\rangle \end{aligned} \tag{4.13}$$

$$\begin{aligned} |\overline{10}\rangle \approx |10\rangle &+ \frac{\varepsilon_t e^{-i\phi_t}}{-\Delta_t}|11\rangle + \frac{\varepsilon_c e^{\phi_c}}{\Delta_c}|0\rangle \otimes \left(|0\rangle + \frac{\varepsilon_t e^{-i\phi_t}}{-\Delta_t}|1\rangle\right) \\ &+ \frac{\sqrt{2}\varepsilon_c e^{-i\phi_c}}{-(\Delta_c+\eta_c)}|2\rangle \otimes \left(|0\rangle + \frac{\varepsilon_t e^{-i\phi_t}}{-\Delta_t}|1\rangle\right) \end{aligned} \tag{4.14}$$

$$\begin{aligned} |\overline{11}\rangle \approx |1\rangle &\otimes \left(|1\rangle + \frac{\varepsilon_t e^{i\phi_t}}{\Delta_t}|0\rangle + \frac{\sqrt{2}\varepsilon_t e^{-i\phi_t}}{-(\Delta_t+\eta_t)}|2\rangle\right) \\ &+ \frac{\varepsilon_c e^{i\phi_c}}{\Delta_c}|0\rangle \otimes \left(|1\rangle + \frac{\varepsilon_t e^{i\phi_t}}{\Delta_t}|0\rangle + \frac{\sqrt{2}\varepsilon_t e^{-i\phi_t}}{-(\Delta_t+\eta_t)}|2\rangle\right) \\ &+ \frac{\sqrt{2}\varepsilon_c e^{-i\phi_c}}{-(\Delta_c+\eta_c)}|2\rangle \otimes \left(|1\rangle + \frac{\varepsilon_t e^{i\phi_t}}{\Delta_t}|0\rangle + \frac{\sqrt{2}\varepsilon_t e^{-i\phi_t}}{-(\Delta_t+\eta_t)}|2\rangle\right) \end{aligned} \tag{4.15}$$

The coupling Hamiltonian $H_J = J\left(a_c^\dagger a_t + \text{h.c.}\right)$ can be written in the subspace truncated to the $|2\rangle$ state as

$$\begin{aligned} H_J \approx & J\left(|01\rangle\langle 10| + \text{h.c.}\right) + \\ & \sqrt{2}J\left(|11\rangle\langle 20| + |11\rangle\langle 02| + \text{h.c.}\right) + \\ & 2J\left(|12\rangle\langle 21| + \text{h.c.}\right). \end{aligned} \tag{4.16}$$

With this, we calculate the energy shifts of the computational states via $E_{ij} \approx \langle\overline{i,j}|H_J|\overline{i,j}\rangle$ for $i, j \in \{0, 1\}$:

$$\langle\overline{00}|H_J|\overline{00}\rangle = \frac{2J\varepsilon_c\varepsilon_t}{\Delta_c\Delta_t}\cos\varphi \tag{4.17}$$

$$\langle\overline{01}|H_J|\overline{01}\rangle = \frac{2J\varepsilon_c\varepsilon_t}{\Delta_c}\left(\frac{-1}{\Delta_t} + \frac{2}{\Delta_t + \eta_t}\right)\cos\varphi \tag{4.18}$$

$$\langle\overline{10}|H_J|\overline{10}\rangle = \frac{2J\varepsilon_c\varepsilon_t}{\Delta_t}\left(\frac{-1}{\Delta_c} + \frac{2}{\Delta_c + \eta_c}\right)\cos\varphi \tag{4.19}$$

$$\begin{aligned} \langle\overline{11}|H_J|\overline{11}\rangle = 2J\varepsilon_c\varepsilon_t\cos\varphi\Bigg(&\frac{1}{\Delta_c\Delta_t} + \frac{-2}{(\Delta_c + \eta_c)\Delta_t} + \\ &\frac{-2}{(\Delta_t + \eta_t)\Delta_c} + \frac{4}{(\Delta_c + \eta_c)(\Delta_t + \eta_t)}\Bigg). \end{aligned} \tag{4.20}$$

Then finally, we use these to calculate the ZZ interaction rate ζ:

$$\zeta = \langle\overline{00}|H_J|\overline{00}\rangle + \langle\overline{11}|H_J|\overline{11}\rangle - \left(\langle\overline{01}|H_J|\overline{01}\rangle + \langle\overline{10}|H_J|\overline{10}\rangle\right) \tag{4.21}$$

$$\begin{aligned} = 2J\varepsilon_c\varepsilon_t\cos\varphi\Bigg(&\frac{2}{\Delta_c\Delta_t} + \frac{-2}{(\Delta_c + \eta_c)\Delta_t} + \frac{-2}{(\Delta_t + \eta_t)\Delta_c} + \frac{4}{(\Delta_c + \eta_c)(\Delta_t + \eta_t)} \\ &- \left(\frac{-2}{\Delta_c\Delta_t} + \frac{2}{(\Delta_c + \eta_c)\Delta_t} + \frac{2}{(\Delta_t + \eta_t)\Delta_c}\right)\Bigg) \end{aligned} \tag{4.22}$$

$$= 8J\varepsilon_c\varepsilon_t\cos\varphi\left(\frac{1}{\Delta_c\Delta_t} + \frac{-1}{(\Delta_c + \eta_c)\Delta_t} + \frac{-1}{(\Delta_t + \eta_t)\Delta_c} + \frac{1}{(\Delta_c + \eta_c)(\Delta_t + \eta_t)}\right) \tag{4.23}$$

$$= \frac{8J\varepsilon_c\varepsilon_t\eta_c\eta_t}{\Delta_c\Delta_t(\Delta_c + \eta_c)(\Delta_t + \eta_t)}\cos\varphi. \tag{4.24}$$

We see that the coupling is linear in J, ε_c, ε_t, and the magnitude can be tuned simply by changing the relative phase φ between the drives.

4.3 Numerical Simulations

In addition to perturbation theory, we solve for the ZZ rates by numerically diagonalizing the Hamiltonian in equation 4.2 using QuTiP [97]. Here, we present simulation results

in Figure 4.2 showing how ζ behaves as a function of ω_d, φ_d, ε_c, and ε_t. We include 7 levels for each transmon in the calculation and use the parameters for the first presented pair: $\omega_\mathrm{c}/2\pi = 5.845\,\mathrm{GHz}$, $\omega_\mathrm{t}/2\pi = 5.690\,\mathrm{GHz}$, $\eta_\mathrm{c}/2\pi = -244.1\,\mathrm{MHz}$, $\eta_\mathrm{t}/2\pi = -247.1\,\mathrm{MHz}$, and $J = 3.45\,\mathrm{MHz}$. In Figure 4.2 (a), ζ is calculated for a range of drive frequencies and amplitudes. The regions of largest ζ are when $\omega_\mathrm{c}^{(12)} < \omega_\mathrm{d} < \omega_\mathrm{t}$, where $\omega_\mathrm{c}^{(12)} = \omega_\mathrm{c} + \eta_\mathrm{c}$, and $\omega_\mathrm{t} < \omega_\mathrm{d} < \omega_\mathrm{c}$. There are also many sharp features associated with the drive interacting resonantly with higher levels of the transmons. Figure 4.2 (b) shows the dependence of ζ on φ_d and $|\varepsilon|$ with $\omega_\mathrm{d} = \omega_\mathrm{t} - 40\,\mathrm{MHz}$. The sinusoidal dependence is visible, as well as a resonance around $|\varepsilon| = 37\,\mathrm{MHz}$. In Figure 4.2 (c) we sweep the amplitudes independently at the same drive frequency as before, keeping constant $\varphi_\mathrm{d} = \pi$. The dependence on each amplitude is symmetric, except for the resonance observed for $\varepsilon_\mathrm{t} > 12.5\,\mathrm{MHz}$.

4.4 Dynamics

We next experimentally investigate how the Stark-induced ZZ interaction depends on the field amplitudes on each qubit ε_c, ε_t and the frequency of the drive field ω_d. We measure ζ using Ramsey interferometry [98] and extract the frequency shift of Q_t conditioned on the state Q_c being $|0\rangle$ or $|1\rangle$ when applying the drive. We find agreement between our measurements and numerical simulations when microwave crosstalk is included.

We show in Figure 4.3 experiments measuring ZZ as a function of drive parameters. The experiments were performed on a pair of fixed-frequency, fixed-coupling transmons on a device of the same design as in [32], with parameters $\omega_\mathrm{c}/2\pi = 5.845\,\mathrm{GHz}$ ($\omega_\mathrm{t}/2\pi = 5.690\,\mathrm{GHz}$), $\eta_\mathrm{c}/2\pi = -244.1\,\mathrm{MHz}$ ($\eta_\mathrm{t}/2\pi = -247.1\,\mathrm{MHz}$) with static ZZ coupling $\zeta_0/2\pi = 307\,\mathrm{kHz}$, corresponding to an inferred exchange coupling strength $J = 3.45\,\mathrm{MHz}$.

In the presence of microwave crosstalk, ε_c and ε_t are complex linear combinations of CZ drive amplitudes A_c and A_t applied to the control and target transmon drive lines, as shown in Figure 4.3 (a). This is expressed via the crosstalk matrix

$$\begin{pmatrix} \varepsilon_\mathrm{c} \\ \varepsilon_\mathrm{t} \end{pmatrix} = \begin{pmatrix} e^{i\theta_\mathrm{c}} & C_\mathrm{ct} e^{i\varphi_\mathrm{ct}} \\ C_\mathrm{tc} e^{i\varphi_\mathrm{tc}} & 1 \end{pmatrix} \begin{pmatrix} A_\mathrm{c} \\ A_\mathrm{t} e^{-i\varphi_\mathrm{d}} \end{pmatrix}, \tag{4.25}$$

where C_ct (φ_ct) denotes the crosstalk amplitude (phase). The phase θ_c results from electrical delay between the drive lines. For these experiments in Figure 4.3 (b), we set $\Delta_\mathrm{t} = 40\,\mathrm{MHz}$ and measured the ZZ interaction while varying drive phase φ_d and global drive amplitude A, where $A_\mathrm{c} = A_\mathrm{t} = A$. The experimental data deviates from the crosstalk-free simulations, which diagonalize the full system Hamiltonian (see supplementary material). This is remedied by including the crosstalk matrix parameters in the model fit. With no crosstalk, ζ is symmetric about ζ_0 between in-phase and out-of-phase driving. Additionally, in Figure 4.3 (c) we varied A_t and A_c independently while keeping φ_d fixed to $\varphi_\mathrm{d} = 1.31\,\mathrm{rad}$. The linear dependence of ζ on the drive amplitude is observed for non-zero A_c amplitudes, as predicted in the theoretical description. Note that the predicted linear dependence is valid only when the drive amplitude is weak compared to the drive detuning from the transmon transition

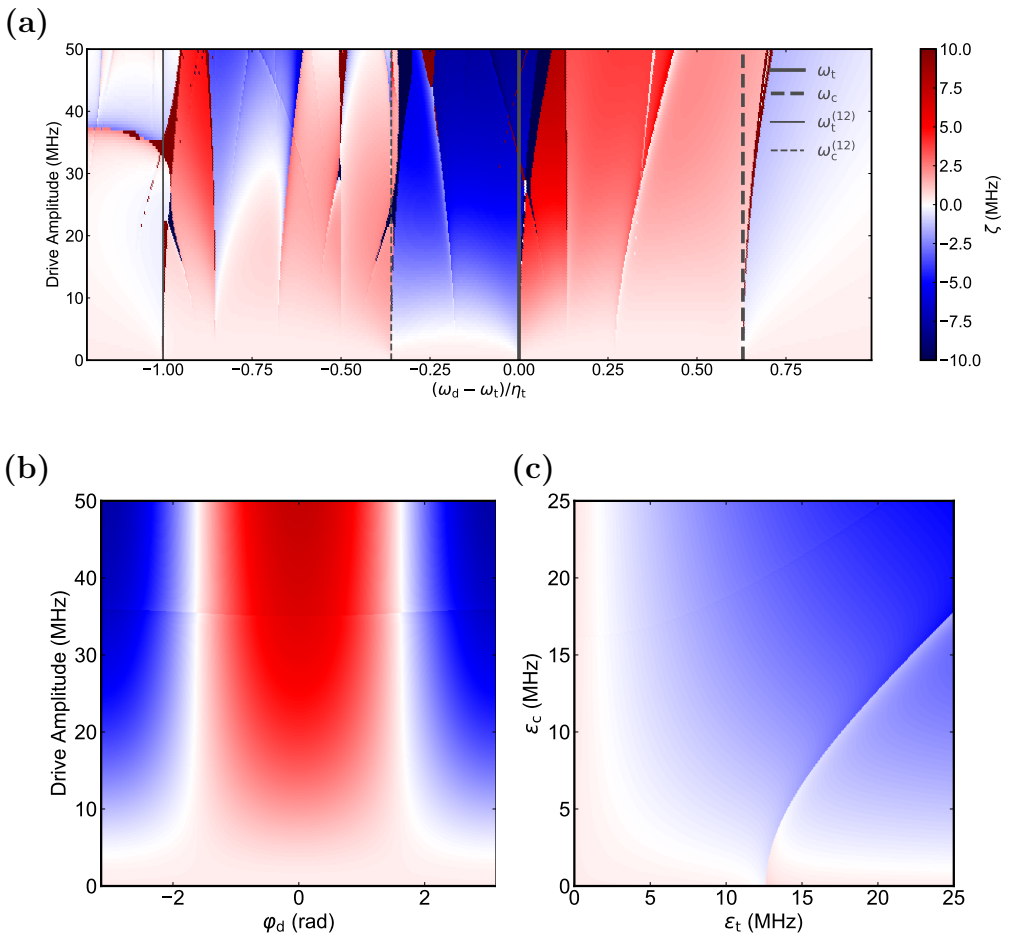


Figure 4.2: ZZ **Numerical Simulations (a)** ζ versus ω_d and global drive amplitude $|\varepsilon| = |\varepsilon_t| = |\varepsilon_c|$, with $\varphi_d = \pi$. Driving in the regions $\omega_c^{(12)} < \omega_d < \omega_t$ and $\omega_t < \omega_d < \omega_c$ show consistently larger areas of enhanced ZZ interaction. Driving below $\omega_c^{(12)}$ or above ω_c, the ZZ enhancement is reduced. There are many resonances visible in the simulation. **(b)** ζ versus φ_d and $|\varepsilon|$, with $\omega_d = \omega_t - 40\,\text{MHz}$. **(c)** ζ versus ε_c and ε_t, with $\varphi_d = \pi$, $\omega_d = \omega_t - 40\,\text{MHz}$.

frequencies. These experiments demonstrate tunability of the sign of the ZZ coupling, with magnitude $\zeta \approx 0$, to one order of magnitude larger than ζ_0, by adjusting the relative phase between the drives. This flexibility is well-captured by numerical simulation, and is resilient to microwave crosstalk,

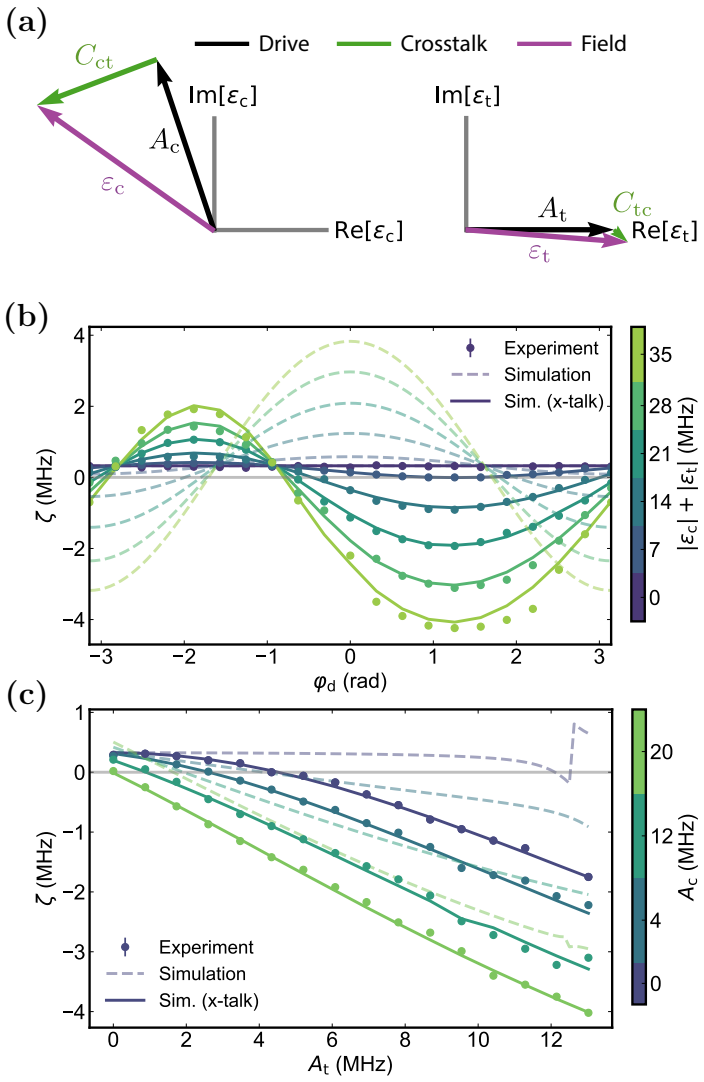


Figure 4.3: ZZ **as a function of drive parameters, crosstalk (a)** The control and target drive fields ε_c, ε_t are complex combinations of the drive line amplitudes A_c, A_t, mixed via the microwave crosstalk matrix C. **(b)** ZZ versus relative drive phase φ_d, for several overall drive amplitudes $|A| = |A_c| = |A_t|$. Field amplitudes in the color-bar are determined by fitting the data (dots) to numerical simulations including crosstalk (solid lines). With no crosstalk (dashed lines) the predicted ζ is symmetric about $\varphi_d = 0$ and ζ_0. Error bars indicate Ramsey frequency fit uncertainty. **(c)** ZZ versus target drive amplitude A_t for several control drive amplitudes A_c (color-bar).

4.5 Gate Calibration

4.5.1 Entanglement calibration

Using the enhanced ZZ interaction, we next calibrate a CZ gate, which is realized from Hamiltonian terms IZ, ZI, and ZZ: $\mathrm{CZ} = \exp\left(-\frac{i}{2}\frac{\pi}{2}\left(-ZI - IZ + ZZ\right)\right)$. We first calibrate the entangling term ZZ, and then correct local phase errors on the each qubit using virtual Z gates [41]. To calibrate the ZZ term, we prepare the target qubit in superposition, apply the CZ pulse, and measure the target qubit Bloch vector r_0 (r_1) when the control qubit is in the $|0\rangle$ ($|1\rangle$) state (see Figure 4.4 (a)). To maximize entanglement, we maximize the quantity

$$R = \frac{1}{2}||r_0 - r_1||^2, \tag{4.26}$$

which measures the normalized vector distance between target Bloch vectors conditioned on the control qubit state [54]. For the drive pulse, we use a cosine ramp with a flat top, with the flat top of the pulse set to 40% of the total pulse time. The parameters to calibrate include pulse time τ_p, drive frequency ω_d, drive amplitudes A_c and A_t, and relative phase φ_d. Pulse time is determined by experiments sweeping drive amplitudes and relative phase, like those shown in Figure 4.3.

We calibrate the gate on a separate qubit pair that was used for the preceding experiments, with parameters $\omega_\mathrm{c}/2\pi = 5.4696\,\mathrm{GHz}$ ($\omega_\mathrm{t}/2\pi = 5.315\,\mathrm{GHz}$), $\eta_\mathrm{c}/2\pi = -270.5\,\mathrm{MHz}$ ($\eta_\mathrm{t}/2\pi = -273.0\,\mathrm{MHz}$) with static ZZ coupling $\zeta_0/2\pi = 170\,\mathrm{kHz}$, corresponding to an inferred exchange coupling strength $J = 2.79\,\mathrm{MHz}$. We set $\tau_p = 201\,\mathrm{ns}$, and pulse amplitudes A_c, A_t and relative phase φ_d are selected to maximize $|\zeta|$.

To calibrate the drive frequency ω_d and overall pulse amplitude A, we find the values that maximize R. In Figure 4.4 (b), one observes a bandwidth of 40 MHz where maximal R is achievable. There is also a region around $A = 0.1$ where R is approximately zero, corresponding to ZZ interaction cancellation. To map the pulse to a CZ gate, local phase corrections are calibrated by measuring the individual qubit Pauli Z error using Ramsey-type experiments.

4.5.2 Local Z Gate Calibration

Here we outline how the local terms of the CZ gate are calibrated. After calibrating the ZZ term of the Hamiltonian to the maximally entangling angle of $\frac{\pi}{2}$, the operation of the pulse on the qubits is of the form $\exp\left(-\frac{i}{2}\left(\alpha IZ + \beta ZI + \frac{\pi}{2}ZZ\right)\right)$. To correct for phases α and β such that the phases on each qubit are aligned to realize a CZ gate, we apply virtual Z (VZ) gates [41] with angles ϕ_{ZI} and ϕ_{IZ} after the entangling pulse, as illustrated in Fig 4.5 (a). With the VZ gates, the combined unitary applied to the qubits is $\exp\left(-\frac{i}{2}\left(\left(\alpha + \phi_{IZ}\right)IZ + \left(\beta + \phi_{ZI}\right)ZI + \frac{\pi}{2}ZZ\right)\right)$. To calibrate ϕ_{ZI}, we prepare the control qubit along the x-axis of the Bloch sphere $|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$, apply the CZ gate circuit, and measure along the x-axis, as shown in Figure 4.5 (b). We sweep ϕ_{ZI} and measure $\langle ZI\rangle$

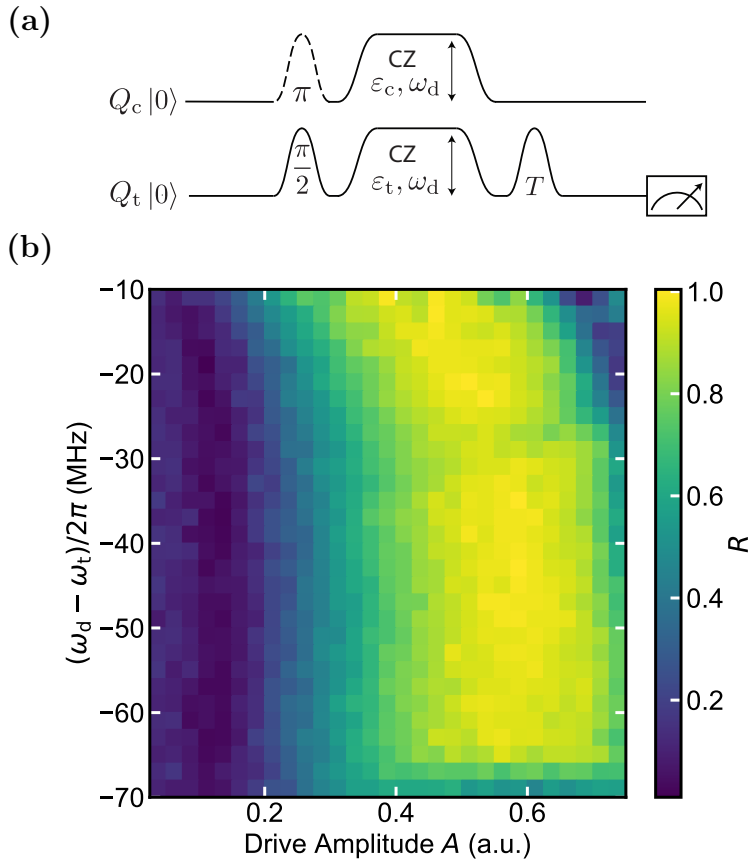


Figure 4.4: **CZ Gate Calibration (a)** Pulse sequence for calibrating the amplitude and frequency of the CZ pulse. The target is prepared in superposition, followed by the CZ pulse, and then tomographic pulses T are applied to measure the target qubit Bloch vector r_i for each control qubit state $|i\rangle \in \{0, 1\}$. The global pulse amplitude and frequency are calibrated by selecting parameters that maximize entanglement measure R. **(b)** R as a function of the CZ gate amplitude A and drive detuning from the target $(\omega_d - \omega_t)$. There is a band of frequencies where R is maximal to realize the CZ gate. There is also a pulse amplitude around $A = 0.1$ where R is minimal, corresponding to ZZ cancellation.

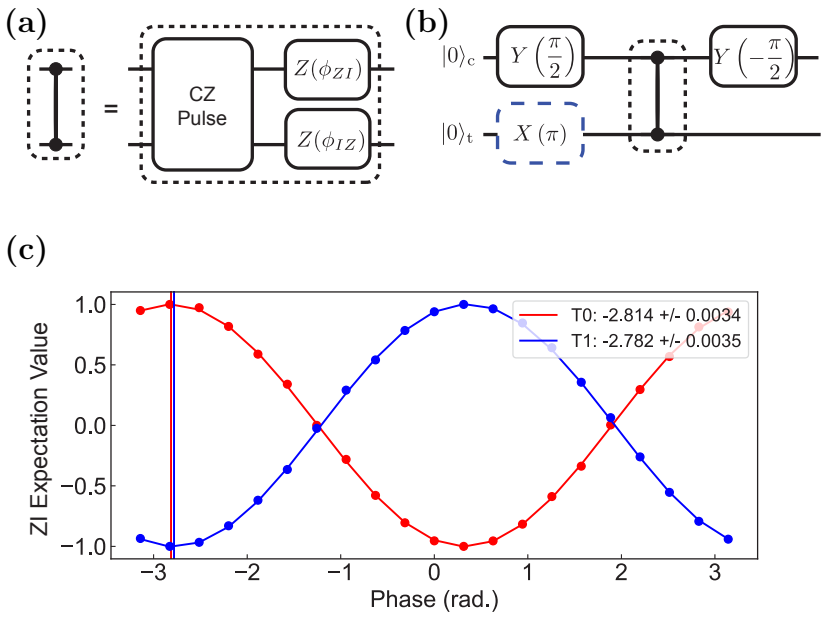


Figure 4.5: **Local Z Gate Calibration (a)** Quantum circuit for compiling a CZ gate using the entangling CZ pulse and local Z corrections. **(b)** Local phase calibration experiment. The control qubit is prepared along the x-axis of the Bloch sphere $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and measured in the x-basis via $Y\left(\frac{\pi}{2}\right)$, $Y\left(-\frac{\pi}{2}\right)$ gates, respectively. **(c)** The local phase ϕ_{ZI} is swept while measuring $\langle ZI\rangle$ for both target preparation states $|0\rangle$ (T0) and $|1\rangle$ (T1). The value of ϕ_{ZI} is calibrated when each input state is mapped to the correct output state, i.e., $|+,0\rangle \rightarrow |0,0\rangle$ and $|+,1\rangle \rightarrow |11\rangle$. This is analogously done for the target qubit local phase ϕ_{IZ}.

when the target qubit is in $|0\rangle$ and $|1\rangle$, as shown in Figure 4.5 (c). The value of ϕ_{ZI} is calibrated when each input state is mapped to the correct output state, i.e., $|+,0\rangle \rightarrow |0,0\rangle$ and $|+,1\rangle \rightarrow |11\rangle$. We fit the expectation values for both target preparations to and set the calibrated value to the average of the two fit results. This experiment is analogously applied for ϕ_{IZ}, exchanging the roles of the control and target qubit in the experiment.

4.6 Benchmarking

To assess gate performance, we perform randomized benchmarking (RB) [63, 64] experiments. We first perform Interleaved Randomized Benchmarking (IRB) [99], shown in Figure 4.6 (a). The IRB protocol interleaves a gate of interest between a sequence of randomly chosen Clifford gates, also called Clifford twirling, which randomizes gate errors to a depolarizing channel. The sequence fidelity is measured with increasing sequence length m and

the decay parameter p is extracted by fitting to the exponential model $P(m) = A \cdot p^m$ [100]. Because IRB estimates F_{CZ} using exponential decays, it is insensitive to state preparation and measurement (SPAM) errors. The decay parameter for the interleaved experiment $p_{IRB} = 0.9672(7)$ and the reference experiment $p_{RB} = 0.9744(9)$ give an estimate of the CZ gate fidelity $F_{CZ} = 1 - \frac{d-1}{d}(1 - p_{IRB}/p_{RB}) = 99.44(9)\%$, where $d = 2^n$ for n qubits (here $d = 4$). The upper- and lower-bounds on gate fidelity estimates from IRB have been shown to span orders of magnitude [69]. When p_{RB} is comparable to p_{IRB}, the IRB F_{CZ} estimate uncertainty increases. From these IRB results, the upper- and lower-bounds on F_{CZ} are between 91.9(2)% and 99.96(1)%, spanning nearly 2 orders of magnitude in gate error (1-F_{CZ}).

To reduce gate fidelity estimate uncertainty, we run the cycle benchmarking (CB) protocol [68], shown in Figure 4.6 (b). The CB protocol is similar to IRB, in that the gate, or cycle, of interest is interleaved between randomly chosen gates. Instead of Clifford gates, in CB the cycle is twirled with multi-qubit Pauli gates, which are tensor products of single-qubit Pauli gates. Pauli twirling maps gate errors into stochastic Pauli errors, which are measured by preparing each eigenstate of the multi-qubit Pauli basis, e.g. XX or YZ for two-qubit CB, and fitting the sequence fidelity to an exponential decay parameter (e.g., p_{XX}, p_{YZ}) as a function of CB sequence length. The error rate $e_i = 1 - p_i$ for Pauli eigenstate i measures errors in the cycle that do not commute with that Pauli operator. We performed CB with cycle lengths of $m \in \{2, 16, 32\}$ for the CZ cycle, and for the empty cycle to estimate the fidelity of the Pauli twirling gates, and extracted the error rate of each Pauli term, which is plotted in Figure 4.6 (b). Averaging over all Pauli preparations, we extract average Pauli decay parameters for both cycles $p_{CZ} = 0.98937(8)$ and $p_I = 0.99702(3)$. We then estimate the CZ gate fidelity as in the IRB protocol, to be $F_{CZ} = 1 - \frac{d-1}{d}(1 - p_{CZ}/p_I) = 99.43(1)\%$, with a worst-case (best-case) fidelity bound of 97.52(2)% (99.764(5)%). Note that the interval between these bounds from CB are narrower than that of IRB, because the fidelity of the Pauli twirling operation is higher than Clifford twirling.

To understand how to reduce the CZ gate error, it is important to distinguish the different error sources. Different error types include coherent errors, such as mis-calibration, stochastic errors, such as dephasing errors, and leakage errors, involving population transfer to non-computational states of the system. We measure the leakage-per-gate using leakage randomized benchmarking (LRB) [47, 66], realized by extracting $|2\rangle$-state outcomes for Q_c and Q_t in IRB experiments. We fit the qubit $|2\rangle$-state population data shown in Figure 4.6 (c) to an exponential model [47, 66] for both reference RB and interleaved RB experiments. We resolve the leakage-per-gate for each transmon to be 0.014% and 0.007% for Q_c and Q_t respectively, indicating leakage is not a dominant source of errors for this gate.

To distinguish coherent and stochastic error sources, we perform purity benchmarking [65, 67], which measures the decay of the purity of the two-qubit density matrix by performing state tomography after Clifford RB sequences. Figure 4.6 (d) shows the purity decay curve, and the inset shows the breakdown of the Clifford process infidelity $e_F = (1 - p_{RB})(1 - 1/d^2) = 1.78(3) \cdot 10^{-2}$ between coherent ($e_U = 0.37(3) \cdot 10^{-2}$) and stochastic ($e_S = 1.41(1) \cdot 10^{-2}$) error types. The dominant source of error in the gate is stochastic

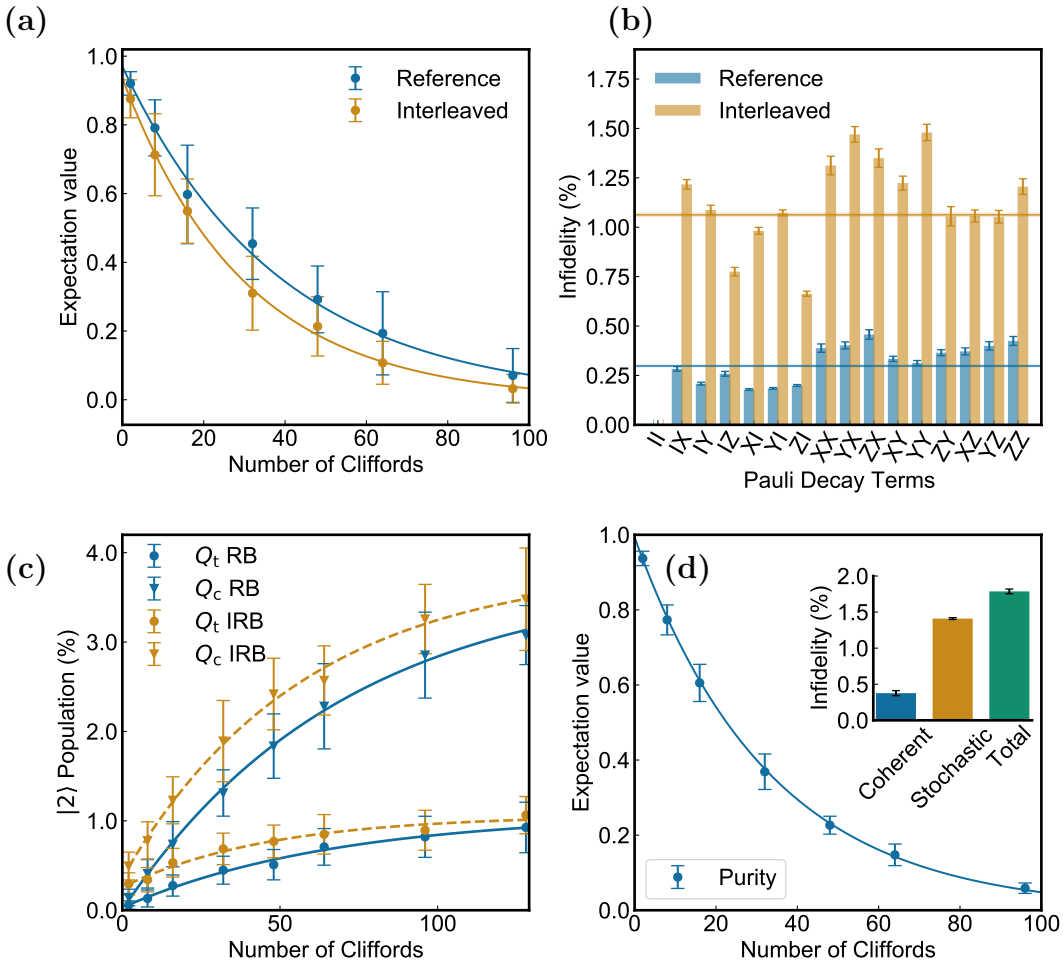


Figure 4.6: **Benchmarking Results (a)** Interleaved RB. Exponential decay of Clifford sequence fidelity (y-axis) for different Clifford sequence lengths with (gold) and without (blue) the interleaved CZ gate. **(b)** Cycle Benchmarking results. The error rate e_i is obtained for the prepared Pauli eigenstate i from the exponential decay of sequence fidelity p_i with increasing sequence length. A larger error rate for a given Pauli eigenstate indicates errors in the cycle that do not commute with the compiled Pauli term. The process infidelity is the error averaged across Pauli terms. **(c)** Leakage Randomized Benchmarking. By monitoring the $|2\rangle$ state of each transmon when running an IRB experiment, the reference and interleaved $|2\rangle$ state population data are fit to an exponential model to extract the leakage-per-gate for each transmon. **(d)** Purity Benchmarking distinguishes coherent from stochastic errors by measuring the decay of the purity (main plot) of the two-qubit density matrix by performing state tomography for each random Clifford RB sequence, and comparing the purity decay to the RB decay. Inset: breakdown of Clifford process infidelity between coherent and stochastic contributions.

error. We estimate the decoherence-limited Clifford process infidelity $e_{\text{decoh.}}$ using the measured T_1 and T_2^{echo} for Q_c (Q_t) of $T_1 = 65(5)\,\mu$s ($T_1 = 58(9)\,\mu$s) and $T_2^{\text{echo}} = 86(6)\,\mu$s, ($T_2^{\text{echo}} = 77(8)\,\mu$s), and the average Clifford gate length of 389 ns, which gives $e_{\text{decoh.}} = 0.76 \cdot 10^{-2}$. The larger observed e_S than the $e_{\text{decoh.}}$ suggests that other forms of stochastic error are present beyond those introduced by relaxation and dephasing of the qubit transition levels. This could be due to decoherence channels of higher transmon levels participating in the interaction via state hybridization between computational and noncomputational levels during the drive [91]. We measured reduced coherence when applying the CZ drive, described in the supplementary material. To summarize, we measure the CZ gate fidelity to be 99.43(1)%, with low leakage and the dominant source of remaining errors being stochastic error. The discrepancy between stochastic error and predicted error from decoherence suggests additional sources of stochastic error are present during the gate.

4.6.1 Coherence Dependence on Drive Amplitude

To investigate whether the CZ drive generated additional sources of incoherent noise, we measured coherences and liftimes of the qubits when the CZ drive is on. In Figure 4.7 (a) we show T_1 versus CZ pulse amplitude, where little trend is observed, as expected. The first 10 data points (about 5 μs) were omitted from the time series data when fitting, due to off-resonant driving of the transmons at high amplitudes which caused deviation from exponential decay of the qubit populations. Further, in Figure 4.7 (b) T_2^{echo} is measured as a function of CZ drive amplitude. We see that the echo time decreases with increasing CZ pulse amplitude, consistent with the discrepancy between measured stochastic errors of the gate and what is predicted by coherence estimates, described earlier. We also monitored the average $|2\rangle$-state population during in these experiments (not shown), and the largest $|2\rangle$ population of about 8% is observed on Q_c when performing the echo experiment on Q_c, indicating off-resonant driving of the $|1\rangle \rightarrow |2\rangle$ transition of Q_c.

4.6.2 Leakage Randomized Benchmarking

To estimate leakage-per-gate, we perform interleaved randomized benchmarking while resolving the $|2\rangle$-state for both transmons (see Appendix A.1 for further details). Imperfect $|2\rangle$-state readout fidelity results in an underestimate of the leakage rate in a way approximately proportional to the $|2\rangle$-state readout infidelity (see [47] supplement). For our LRB experiment, the readout fidelities for the three states are $p_c^{|0\rangle} = 0.996$, $p_c^{|1\rangle} = 0.984$, $p_c^{|2\rangle} = 0.975$ for the control transmon, and $p_t^{|0\rangle} = 0.996$, $p_t^{|1\rangle} = 0.984$, $p_t^{|2\rangle} = 0.962$ for the target transmon, suggesting that the leakage rates we measure are a few percent smaller than the actual leakage rates. We fit the observed $|2\rangle$ state populations as a function of Clifford sequence length

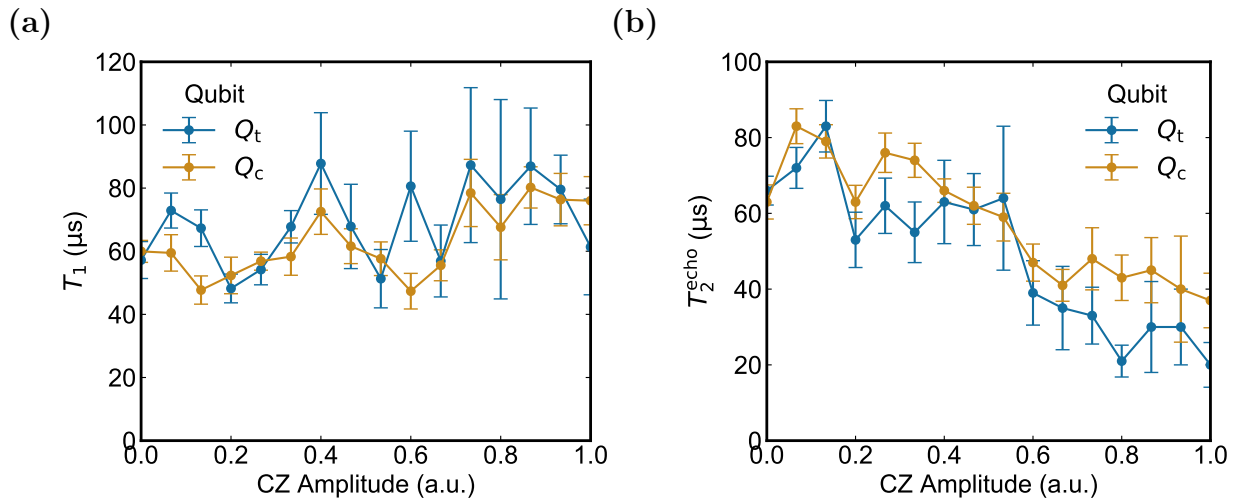


Figure 4.7: **Coherence versus Drive Amplitude**. **(a)** Qubit lifetimes T_1 as a function of CZ pulse amplitude. No visible trend is observed. **(b)** Measurements of T_2^{echo} versus CZ drive amplitude for both qubits used in the gate in the main text. Error bars are uncertainties in the fit. The coherence of the sample is reduced at larger CZ pulse amplitudes.

m from both the referenced and interleaved IRB experiments to the exponential model

$$P_{|2\rangle} = B - Ae^{-\Gamma m} \tag{4.27}$$
$$\Gamma = \gamma_\uparrow + \gamma_\downarrow \tag{4.28}$$
$$B = \gamma_\uparrow/\Gamma, \tag{4.29}$$

where $\gamma_\uparrow$ is the leakage rate and $\gamma_\downarrow$ is the seepage rate [47, 66]. Fitting both interleaved and reference data-sets to this model, we extract the leakage-per-gate as $\gamma_\uparrow^{\text{CZ}} = \gamma_\uparrow^{\text{IRB}} - \gamma_\uparrow^{\text{RB}}$ [84].

4.7 Discussion

We have demonstrated a tunable ZZ interaction between fixed frequency, fixed coupling transmons using off-resonant, simultaneous charge drives. This tunable ZZ coupling enables both cancellation and enhancement of the static ZZ interaction. We implemented a high fidelity CZ gate that is resilient against drive crosstalk and static ZZ interactions during the gate. We expect with this interaction one can leverage higher exchange coupling J between qubits to further reduce CZ gate times, and multi-path couplers that suppress static ZZ [85] can be combined with this drive to eliminate the unwanted ZZ during idling [75] or during operation of other gates. Further, the off-resonant character of the interaction provides drive frequency flexibility, reducing frequency crowding constraints with scaling up fixed

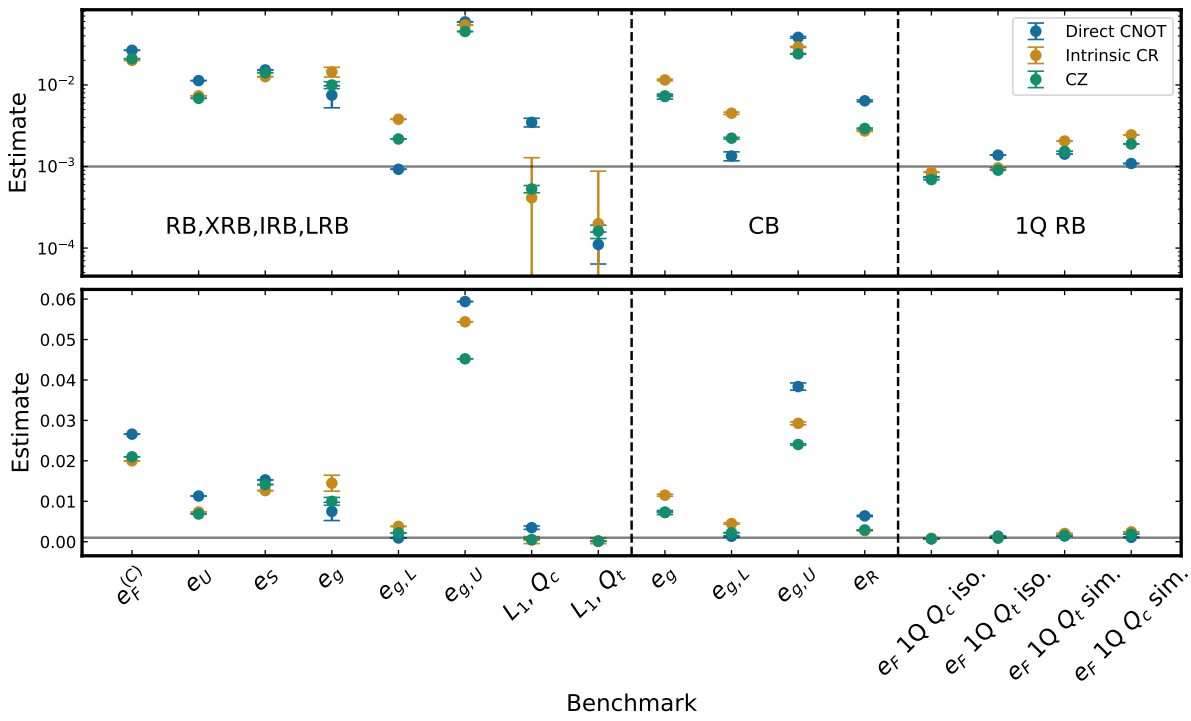


Figure 4.8: **Comparison of the direct CNOT, intrinsic CR, and CZ gates calibrated on Sample B.**

frequency, fixed coupling quantum processors. Quantifying multi-qubit errors of this gate on a larger quantum processor, including spectator errors and simultaneous gate operation with spectator ZZ cancellation, is a subject of future work. While we utilize this charge-activated tunable ZZ interaction to implement a CZ gate, it can also be applied to simulation of exotic quantum many-body physics, including the extended Bose-Hubbard model [101–104] and nonreciprocal interacting photonic systems [105, 106], as well as realizing native quantum stabilizer measurements [107], without the need for additional flux-based tunable components.

As a final comparison, Figure 4.8 compares the benchmarking results for each entangling gate calibrated on Sample B: the direct CNOT, intrinsic CR, and the CZ gate. Notably, the estimated gate process errors are quite close for the CZ and direct CNOT gate, as measured by CB. However, the leakage rate of the CZ gate was substantially lower than the direct CNOT, more similar to the intrinsic CR gate, which may be explained by the difference their respective gate times (direct CNOT was 147 ns, CZ was 200 ns, and intrinsic CR was 250 ns). Both the intrinsic CR and the CZ also had similar unitary error e_U, with the CZ error being slightly smaller. This small difference may be due to the driven ZZ Hamiltonian commuting with the static ZZ, however further study is necessary to determine these differences, as well

as to determine the dominant coherent error sources in the CZ gate. The higher e_U of the direct CNOT gate may be explained by coherent errors from the additional drives.

Chapter 5

Future Directions

In this thesis, we investigated microwave-activated entangling gates on the Trailblazer superconducting quantum processor. As a prerequisite, we calibrated and benchmarked single-qubit gates, finding that the static ZZ interaction and decoherence account for the measured error rates. For the cross resonance gate, we explored two calibration techniques, and made several improvements and discoveries. First, we mitigated leakage of the CR gate on a low-anharmonicity sample using two-dimensional DRAG pulse shaping. On a separate sample with higher anharmonicity, we compared the direct CNOT gate calibration approach with an intrinsic CR gate calibration, and found lower coherent errors on the intrinsic CR gate, despite it being substantially longer. For both cases, the remaining process error was consistent with errors from the static ZZ coupling. Following an observed increase in ZZ interaction when applying a crosstalk-cancellation drive on a spectator qubit when operating the CR gate, we discovered a microwave-activated tunable coupling. We studied this tunable coupling theoretically, numerically, and experimentally, finding good agreement with experiments when crosstalk is accounted for. We then implemented a high-fidelity CZ gate with the tunable coupling, and found the gate error was dominated by stochastic errors.

One overarching goal of the work in this thesis is to maximize performance of fixed frequency, fixed coupling quantum processors. To this end, future QPU designs could incorporate the tunable coupling interaction as the native interaction, rather than the CR interaction. The potential benefits for this are three fold: improved yield due to fewer frequency collision constraints when compared to the CR gate [108], static ZZ cancellation [51], and, as a consequence of the second point, higher exchange coupling J for faster gates. For a qubit with $100\,\mu$s T_1, achieving a process error of 0.1% requires a two-qubit gate time of 80 ns. Therefore, a path toward two-qubit gate performance of 10^{-3} would be to approximately double the coupling J. Importantly, this also increases the static ZZ coupling, which scales with J^2. This larger scaling of ζ_0 with J is why J is generally kept small for architectures based on the CR interaction, however with the tunable ZZ coupling presented in Chapter 4, the static coupling can be cancelled.

While static ZZ cancellation is possible with the driven ZZ coupling, larger static ZZ couplings require larger drive amplitudes to cancel, which worsens the drive-induced deco-

herence presented in Figure 4.7. One additional way to temper the increased static ZZ rates with J is to use multiple coupling paths between the qubits [23]. This allows a larger J/ζ_0. By incorporating multipath coupling elements, I estimate that we can maintain the same static ZZ couplings, while nearly doubling the exchange coupling J. This suggests that two-qubit gate times could be nearly halved, while maintaining similar single-qubit gate fidelities without requiring active ZZ cancellation.

Finally, experiments utilizing this drive-induced interaction for implementing a two-*qutrit* entangling gate are underway. Utilizing the $|2\rangle$ state for computation has been shown to enable hardware-efficient algorithms in this group [32], however this demonstration was based on the static ZZ interaction between the qutrits and the CR interaction. Amplifying the ZZ interaction via off-resonant drives could enable higher-fidelity quantum algorithms to be implemented on qutrit systems.

Appendix A

Appendix

A.1 Benchmarking

Benchmarking quantum circuits is the process of running standardized protocols and measuring the fidelity, or error, of the outcomes. Benchmarking is crucial for understanding the limitations of a given gate set in faithful execution of quantum circuits. In this section, I will summarize the key protocols used in this thesis to benchmark the quantum gates studied, namely isolated and simultaneous single-qubit gates, the direct CNOT gate, the intrinsic CR gate, and the CZ gate.

A.1.1 Average gate fidelity and process fidelity

Starting from first principles, the simplest question to ask is, how to fully determine the action of a quantum circuit? In general, any quantum circuit is a quantum channel, that is, a completely positive (CP), trace-preserving linear operator that maps density matrices to density matrices [38]. In particular, a quantum process Φ is termed a *superoperator*, because it acts on density matrices, in contrast to standard quantum mechanical operators, which act on wavefunctions.

Given an experimentally determined quantum process Φ, there are a couple of ways to evaluate how well we have implemented the desired unitary U. Ideally, the unitary U acts on a state $|\psi\rangle$, and we can define the *average gate fidelity* $\overline{F}$ as simply the average state fidelity between states generated by $\Phi(\rho_\psi)$ and states generated by the ideal $\rho_\psi^\dagger$, averaged over input states ρ_ψ

$$\overline{F}(\Phi, U) = \int d\psi \, \langle\psi| \, U^\dagger \Phi(\rho_\psi) U \, |\psi\rangle \, . \tag{A.1}$$

In other words, $\overline{F}$ measures how well Φ approximates the U [109], with $\rho_{psi} = |\psi\rangle \langle\psi|$ and $d\psi$ is the uniform Haar measure over state space [109]. Essentially, this is the average state fidelity over all possible (pure) input states. The average gate fidelity, or infidelity $r = 1 - \overline{F}$ is a common measure of gate performance.

However, there is an important observation about r that is relevant for benchmarking quantum gates. Native gates on a quantum computer that we want to benchmark in almost all cases act on a *subset* of qubits, and the average gate fidelity does not account for effects that Φ might have on the entanglement between the subsystem its acting on $|\psi\rangle$ and the rest of the quantum computer. A measure that is more relevant to quantum error correction [110] is the *process* fidelity, or sometimes called *entanglement* fidelity.

The process fidelity can be understood by invoking the channel-state duality, known as the Choi-Jamiolkowski isomorphism [111, 112], which represents the quantum process Φ as a density matrix ρ_Φ, often called the Choi matrix, of the form $A \otimes B$ for input density matrices A and output density matrices B

$$\rho_\Phi = \frac{1}{d} \sum_{i,j} E_{ij} \otimes \Phi(E_{ij}), \tag{A.2}$$

where E_{ij} is the $d \times d$ matrix with value 1 at row i and column j, but is 0 elsewhere. Note that, the trace-preserving property of Φ is equivalent to $\mathrm{Tr}_B[\rho_\Phi] = I/d$, and the unital property (normally presented via the operator-sum representation of Φ as $\sum_i K_i^\dagger K_i = 1$ [38] for Kraus operators K_i) given by $\mathrm{Tr}_A[\rho_\Phi] = I/d$. Now, the process fidelity F for a quantum operation Φ is simply the *state fidelity* of the experimentally measured Choi matrix, with the ideal Choi matrix

$$F = \mathrm{Tr}[\rho_\Phi^{(\mathrm{ideal})} \rho_\Phi^{(\mathrm{exp.})}]. \tag{A.3}$$

Note that the equation for state fidelity used here assumes that $\rho_\Phi^{(\mathrm{ideal})}$ is a pure state, since generally ideal quantum circuits are unitary. If this is not the case, the more general formula for state fidelity must be used.

Importantly, the average gate fidelity $\overline{F}$ and process fidelity F are related via [109]

$$\overline{F} = \frac{dF + 1}{d + 1}. \tag{A.4}$$

Note that the average gate fidelity is always larger than the process fidelity. For this reason, and because process fidelity considers entanglement preservation with other subsystems as described in [110], process fidelity should be the metric used to benchmark quantum processes, as opposed to average gate fidelity. To give a sense for how much these quantities differ, For two-qubit gates, $d = 4$, so $\overline{F} = \frac{1}{5} + \frac{4}{5}F$. A measurement reporting $\overline{F} = 99\%$ gives an $F = 98.75\%$. Thus, choosing whether to report gate fidelity or process fidelity, can impact whether a gate is perceived to be "below threshold" for quantum error correction. We again stress that gate fidelity is inappropriate to use for this comparison.

A.1.2 Process infidelity and Quantum Error Correction Thresholds

In the context of quantum error correction (QEC), there is a critical *error threshold* for a QEC code where, if the physical operations are below this threshold, the *logical* error rate will

reduce with growing system size, enabling the implementation of logical qubit storage, logical gates, and eventually fault-tolerant quantum computation, ushering in a new paradigm of computing power. The error thresholds reported in QEC codes typically are modeled as a general depolarizing error rate per QEC cycle p, and each operation (e.g. preparation, gate, storage, measurement) are assumed to have the same error rate.

So, how do we compare measures like process error e_F to the QEC error thresholds? In the case of more general and physically realistic errors, the QEC threshold error rate for quantum channel $\mathcal{E}$ can be characterized by the diamond-distance from the identity [113] $\epsilon(\mathcal{E})$

$$\epsilon(\mathcal{E}) = \frac{1}{2}||\mathcal{E} - \mathcal{I}||_\diamond = \sup_\psi \frac{1}{2}||\left[\mathcal{E} \otimes \mathcal{I}_d - \mathcal{I}_{d^2}\right](\psi)||_1, \tag{A.5}$$

which measures the worst-case error between two quantum channels. Here, $||\cdot||_1$ is the trace norm. While RB cannot measure the diamond-norm directly, measuring the process error e_F can be used to bound it [70, 114]:

$$e_F \leq \epsilon(\mathcal{E}) \leq d\sqrt{e_F}. \tag{A.6}$$

Essentially, if gate errors are coherent, this means that the worst case error given a particular input state, could be much larger than the overall process error, since coherent errors can affect some input states much more than others. So, if coherent errors are substantial, then the worst-case error rate may be much higher than the average case. Stochastic Pauli errors on the other hand saturate the lower bound of equation A.6. Note that, while the upper bound listed in equation A.6 is not known to be tight, the scaling with e_F is optimal [115]. The implication of the right-hand inequality in equation A.6 is that, an RB estimate of e_F could indicate a ϵ error orders of magnitude larger than what is reported by RB. However, utilizing a software protocol called Randomized Compiling, which converts coherent errors to incoherent errors on average, when running quantum algorithms is one way to bring predictions of algorithm performance by RB metrics and actual algorithmic performance into agreement [39].

A.1.3 Benchmarking protocols for measuring e_F

Because Φ is a linear operator, it can be represented as a matrix in the vector space of density matrices, and therefore determining Φ amounts to measuring all of the matrix elements of Φ, by preparing a spanning set of input density matrices, and measuring a spanning set of the output density matrices; this is called quantum process tomography (QPT). QPT is a natural approach is to measure Φ.

As QPT amounts to determining the elements of a matrix that maps density matrices to density matrices, the naive complexity of QPT is $\mathcal{O}(d^4)$, where $d = 2^n$ is the Hilbert space dimension of n qubits [64]. Therefore, while in principle giving complete information of the quantum process, it does not scale well with increasing system size. Furthermore, state preparation and measurement (SPAM) errors limit the precision with which Φ can

be determined [116]. An extension of QPT, called Gate Set Tomography (GST) [117], is a SPAM-independent way to determine Φ, however GST also scales exponentially with system size.

Because of scalability and precision challenges posed by measuring Φ precisely, an alternative approach called Randomized Benchmarking (RB) [63, 64] has become a standard procedure for benchmarking quantum gates. Rather than determining a specific process matrix Φ, RB protocols and their extensions measure the average gate fidelity $\overline{F}$, and via equation A.4, the process fidelity F can be determined.

The two main varieties of benchmarking protocols used in this thesis are Randomized Benchmarking (RB) [63, 64] protocols, and Cycle Benchmarking [68]. The RB protocol is an industry-standard technique to measure average error rates of random sequences of *Clifford* circuits (see definition of Clifford gates in the next paragraph), which can be thought of as running net-identity circuits that take random paths through Hilbert space [118]. While coherent errors and incoherent errors in the gates will affect each individual path differently, the errors of the average trajectory behave as if all errors are *depolarization* errors. That is, the error channel becomes a depolarizing channel Λ acting on density matrix ρ_0

$$\Lambda(\rho_0) = pU\rho_0 U^{-1} + (1-p)\frac{\mathbb{1}}{d}, \tag{A.7}$$

where p is the depolarization constant and U is the target unitary to be implemented (for the identity operation, $U = \mathbb{1}$). Intuition for how Clifford twirling tailors generic noise into depolarizing noise comes from the notion the definition of *twirling* a quantum channel. Twirling a quantum channel Φ is simply the operation of conjugating Φ with random unitaries. A given quantum channel Φ may have coherent and incoherent error sources, but because the Clifford group forms a *unitary 2-design* [118], performing the twirled operation $\Phi_T \equiv \int dU U^\dagger \Phi(U\rho U^\dagger)U$, where the integral is over the uniform (Haar) measure dU on the space of $d \times d$ unitary matrices [109], is well approximated by twirling Φ over random Clifford gates. The randomness over Hilbert space converts errors, in a statistical sense, onto a depolarizing channel, which is a purely random error channel on the state, having no preferred direction in Hilbert space.

Clifford RB measures the depolarization constant p by performing random identity-equivalent Clifford circuits of increasing depth m, measuring the decay of the *survival probability* $P(m)$, and fitting $P(m) = A_0 p^m + B_0$, where A_0 and B_0 contain state preparation and measurement (SPAM) error information. Specifically, this is done efficiently by selecting N randomly-sampled, length-m sequences of n-qubit Clifford operations, drawn from the Clifford group, which is defined as the group that normalizes the n-qubit Pauli group $\mathbf{P}_n$

$$\mathbf{C}_n = \{V \in U_{2^n} \mid V\mathbf{P}_n V^\dagger = \mathbf{P}_n\}. \tag{A.8}$$

The n-qubit Clifford group is generated by the single-qubit Hadamard (H) and $S = \sqrt{Z}$

gates, and the two-qubit CNOT gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{A.9}$$

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \tag{A.10}$$

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{A.11}$$

Note that while RB protocols don't tell the type of errors present in a gate, the error rates reported by RB are insensitive to SPAM errors, unlike QPT. Furthermore, it requires fewer experiments per qubit than GST, and is therefore more scalable.

Once the depolarization parameter p is estimated, modifications to the standard RB protocol allow estimation of fidelity of a particular gate (via interleaved RB [99]), of the coherent and incoherent relative error contributions (via purity benchmarking [67, 69]), and of the leakage rate (via leakage RB [47, 66]). Crosstalk errors can be characterized by comparing single-qubit RB decay rates when performing single-qubit gates simultaneously, and in isolation [87]. In the next sections, I will describe in further detail the different forms of RB protocols, as well as Cycle Benchmarking.

A.1.4 Standard RB, Single and Two-qubit

Standard, or Clifford, RB gives an estimate of the depolarization constant p corresponding to the random circuits run. From these decay parameters, the process infidelity of the average Clifford gate can be estimated as $e_F^C = \frac{d^2-1}{d}(1-p)$, where $d = 2^n$ is the Hilbert space dimension for n qubits. Another commonly reported measure is the average error per cycle (EPC) $r = \frac{d-1}{d}(1-p) = \frac{d}{d+1}e_F^C$, which is notably always lower than the per-cycle process infidelity.

Single-qubit RB experiments are generally done simultaneously, as plotted in Figure A.1 (b), from which a simultaneous single-qubit Clifford process fidelity is extracted for each qubit. This can be compared to single-qubit RB performed in isolation (Figure A.1 (a)) to obtain a measure of crosstalk [87]. Simultaneous single-qubit RB for two qubits tends to be in the range of $e_F^{1\text{Q,sim}} \sim 3 \times 10^{-3}$, with precise measurements reported in Figure 3.15.

A.1.5 Interleaved RB

Interleaved RB (IRB), shown in Figure A.1 (d) is a way to estimate the fidelity of a particular gate G, by interleaving G between the random Clifford gates in a standard RB sequence [99]. Note that for Clifford RB, G must itself be in the Clifford group. Interleaving a gate of interest with randomly sampled gates is defined as "twirling" the gate G. From these two

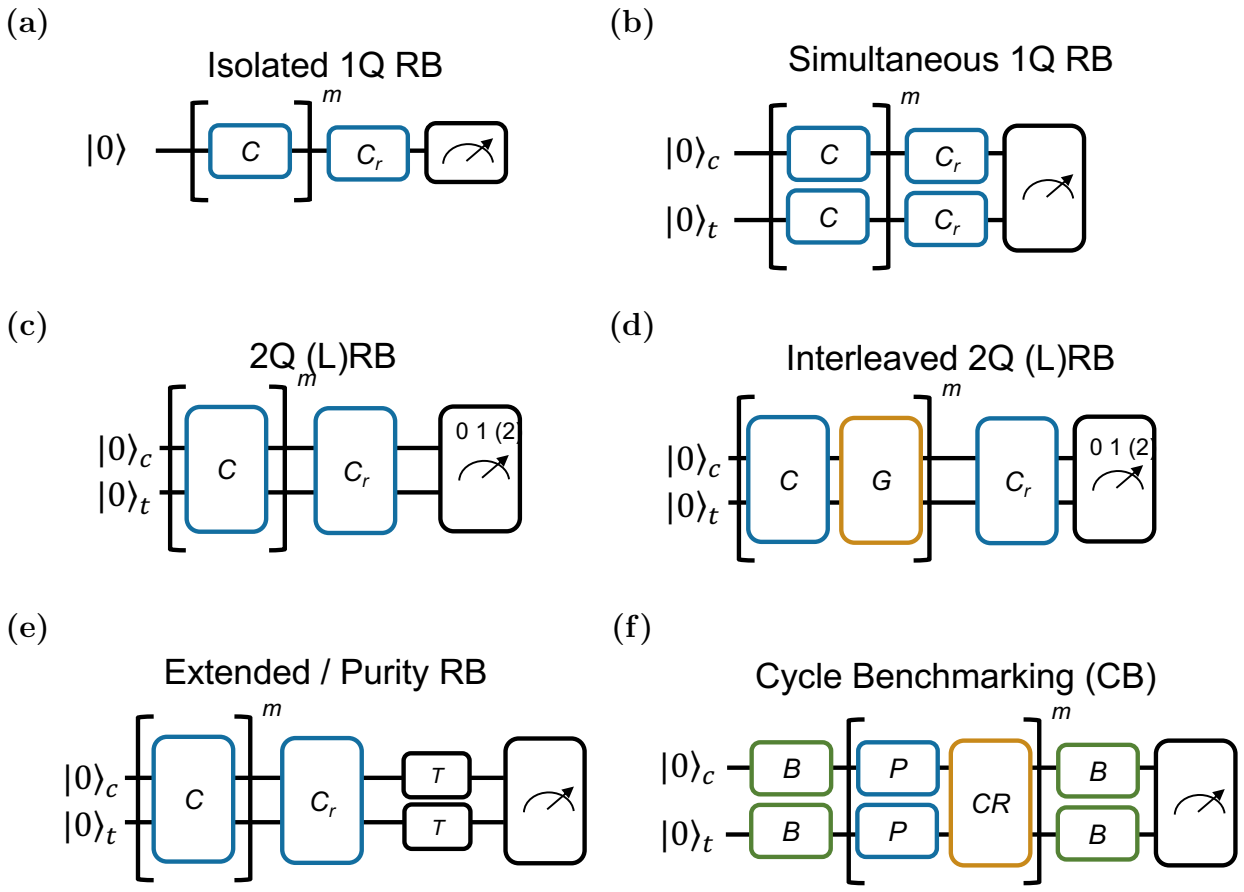


Figure A.1: **Benchmarking circuits (a)** Isolated single-qubit Randomized Benchmarking **(b)** Simultaneous single-qubit Randomized Benchmarking **(c)** Two-qubit Randomized Benchmarking, with possible leakage detection with $|2\rangle$-state readout **(d)** Interleaved Two-qubit RB, with similar option for quantifying leakage as RB**(e)** Extended / Purity RB. Using state tomography to measure decay of purity, i.e. 'unitarity' **(f)** Cycle Benchmarking, a more scalable, precise, and measure for general cycle fidelity as compared to Clifford RB.

Metric Name	Symbol	Relation to p or other	Comments
Depolarization constant	p	1	
Depolarization error	e_d	$1-p$	
Process/Entanglement error	e_F	$\frac{d^2-1}{d^2}(1-p)$	Independent of d
Process/Entanglement fidelity	F	$1-e_F$	
Error per Cycle (EPC)	r	$\frac{d-1}{d}(1-p)$	$=\frac{d}{d+1}e_F$
Error per Gate (EPG)	EPG	r/n_g	Upper bounds gate error from IRB [77]
Interleaved gate/cycle error	r_c	$\frac{d-1}{d}(1-p_c/p)$	via IRB; large uncertainty [69]
Unitarity	u		via XRB[67]
Stochastic error	e_S	$1-\sqrt{u\,(1-1/d^2)+1/d^2}$	via XRB[67]
Unitary error	e_U	e_F-e_S	via XRB[67]
Leakage rate	L_1		via LRB[47, 66]

Table A.1: Names, symbols, and relationships of various benchmarking parameters used in this thesis.

experiments, the decay parameters for the circuits with the interleaved gates and for standard are measured p_{IRB} and p_{RB}. The standard way to estimate the process infidelity for the gate of interest e_g from IRB is given in Table 3.3.2, however this is subject to a large systematic uncertainty when gate errors have coherent components [69]. The lower- and upper-bounds on e_g, denoted $e_{g,L}$ and $e_{g,U}$ are given in Equation A.12, with error bars saturated from below in the case of stochastic errors (meaning state purity decays as p^2_{RB}), and from above by fully coherent errors, corresponding no decay of state purity. These bounds can be more precisely estimated using Extended, or Purity RB (XRB), described next.

A.1.6 Extended RB

Extended RB (XRB) or Purity RB, shown in Figure A.1 (e) allows estimation of the unitary e_U and incoherent or stochastic (e_S) contributions to the overall process infidelity e_F, by

performing full state tomography after each RB sequence. This is because stochastic errors contribute to the decay of state purity, while unitary errors do not contribute. Additionally, measuring the unitarity with this method enables improved bounds of the worst/best case error bounds on interleaved cycle fidelity [69],. To estimate the worst/best case infidelity bounds e_{IU}, e_{IL} of a gate, the decay parameters p_{IRB}, p_{RB}, and purity decay (or 'unitarity') u are estimated via IRB, RB, and XRB experiments respectively. From these parameters, e_{IU}, e_{IL} are estimated via

$$e_{g,U} = 1 - \left(\frac{p_C p_R}{u} - \sqrt{\left(1 - \frac{p_R^2}{u}\right)\left(1 - \frac{p_C^2}{u}\right)} \right) \tag{A.12}$$

$$e_{g,L} = 1 - \left(\frac{p_C p_R}{u} + \sqrt{\left(1 - \frac{p_R^2}{u}\right)\left(1 - \frac{p_C^2}{u}\right)} \right). \tag{A.13}$$

Note that, if the unitarity is not known or measured, using $u = 1$ (i.e. purely coherent error) gives the worst case bounds of the upper and lower bounds of the gate process infidelity, while setting $u = p_{\mathrm{RB}}^2$ (i.e. purely stochastic error) gives the best case upper and lower bounds.

A.1.7 Leakage RB

Leakage RB is an extension of an RB experiment where populations of the non-computational (e.g. $|2\rangle$) states of the transmons are also measured [47, 66]. This enables the estimation of the leakage rate of the average Clifford gate, and also of the interleaved gate, in the case of an IRB experiment with $|2\rangle$-state measurement enabled. The leakage rate of the gate set is important because leakage is often not accounted for in most QEC threshold calculations, and methods used to address them often increase the resource overhead of QEC cycles [62, 119]. To benchmark the leakage-per-cycle of a RB experiment, we perform the same analysis as was done in [47], where the average $|2\rangle$-state population after m e.g. Clifford cycles is modeled via the rate equation

$$p_{|2\rangle}(m+1) = p_{|2\rangle}(m) + L_1\left(1 - p_{|2\rangle}(m)\right) - L_2\left(p_{|2\rangle}(m)\right), \tag{A.14}$$

where L_1 (L_2) is the leakage (seepage) rate per Clifford. From the rate equation above for $p_{|2\rangle}$, which is only valid when averaged over all Clifford circuit instances, the solution is

$$p_{|2\rangle} = C(1-\Gamma)^m + p_\infty, p_\infty = \frac{L_1}{\Gamma}, \Gamma = L_1 + L_2. \tag{A.15}$$

When $\Gamma \ll 1$, the solution can be replaced with

$$p_{|2\rangle} = (p_{|2\rangle}(0) - p_\infty)e^{-\Gamma m} + p_\infty, \tag{A.16}$$

and estimating L_1 and L_2 amounts to fitting the $p_{|2\rangle}(m)$ data to this model.

In terms of leakage effects in quantum error correction, it generally reduces the error rate threshold of a QEC code by requiring additional circuitry like Leakage-Reduction-Units to mitigate [120]. In one notable work [121], it was found that the error threshold was reduced by only a factor of 2 when leakage and depolarizing noise rates are comparable, assuming three-outcome measurements. From this, a rule of thumb seems to be to reduce leakage rate L_1 below the depolarizing noise rate.

To give a sense of leakage rates among two-qubit gates in superconducting qubit systems, ref. [96] found leakage rates per cycle (or per gate) $L_1 \sim 5 \times 10^{-4}$. With gates of this form, leakage errors have been studied and mitigated in the bit-flip error correction code [122], showing an enhanced error suppression factor when unconditional reset is implemented. This suggests that leakage rates of this magnitude are manageable with reset schemes. The leakage-per-gate measured by the intrinsic CR gate (Figure 3.15) and the CZ gates (Figure 4.6 in this thesis) are comparable to that measured in [96]. The leakage of the direct CNOT on sample A (Figure 3.10), however, is higher, around 3×10^{-3}, suggesting that the anharmonicity for this sample is too low for low-leakage gates, or alternative methods to mitigate leakage on samples with lower anharmonicities are needed.

A.1.8 Cycle Benchmarking

Figure A.1 (f) shows the quantum circuit for the Cycle Benchmarking (CB) protocol [68]. While RB typically uses the Clifford group as the randomizing/twirling group, CB uses the Pauli group. Importantly, Clifford twirling converts general errors into a depolarization error as in equation A.7, while Pauli twirling converts general errors into Pauli errors e_{P}, realizing the quantum operation Λ

$$\Lambda(\rho) = \sum_{P \in \{I,X,Y,Z\}^{\otimes n}} \mu(P) P \rho P^{\dagger}, \tag{A.17}$$

where $\mu(P)$ is a probability distribution. In CB, $\mu(P)$ of the dressed cycle is measured by preparing each Pauli eigenstate with the corresponding basis gates B in Figure A.1 (f). The Pauli twirling gates P ensure that the cycle (indicated in the brackets) is a Pauli depolarization channel with exponential decays p_{P} for each preparation, measured via the exponential decay of the prepared eigenstate as a function of cycle length m. Note that the cycle lengths m are typically chosen such that the net operation of the cycle $\mathcal{G}$ after m repetitions is the identity, $\mathcal{G}^m = \mathbb{1}$. Also, the basis measurement gates B at the end are in general different than the preparation basis gates at the beginning, and are calculated by computing the action of the cycle on the input Pauli state.

To estimate the bare cycle fidelity, similar to IRB, the dressed cycle fidelity is estimated by performing CB with the cycle of interest interleaved with the Pauli gates, while the fidelity of the twirling gates themselves (the 'reference' circuit) are separately measured. Estimating the gate fidelity and worst/best case fidelity bounds are then estimated in the same way as in IRB.

More explicitly, any quantum channel Λ can be expressed via the Kraus representation

$$\Lambda(\rho) = \sum_{k=0}^{d^2} A_k \rho A_k^\dagger, \quad \sum_k A_k^\dagger A_k = \mathbb{1}. \tag{A.18}$$

If the quantum channel is a noisy gate with ideal unitary U, the noise can be on average converted to Pauli errors by inserting random Pauli gates between the gate of interest U [68, 123]. With Pauli noise, the Kraus operators A_k are of the form

$$A_k = \sqrt{\frac{e_k}{d^2-1}} P_k U, k \neq 0 \tag{A.19}$$

$$A_0 = \sqrt{a_0} P_0 U \tag{A.20}$$

$$a_0 = 1 - \overline{e_P} \tag{A.21}$$

$$\overline{e}_P = \frac{1}{d^2-1} \sum_{k=1}^{d^2} e_k, \tag{A.22}$$

where e_k is the error probability for n-qubit Pauli operator P_k, and P_0,a_0 are the identity operator and probability of no error. By preparing input states that are n-qubit Pauli operators and measuring their expectation value by choosing the corresponding measurement basis, the exponential decays with sequence length Ap_k^m can be isolated and fitted, giving $p_k = 1 - e_k$ and avoiding the issue of fitting to multiple decays, similar to character benchmarking [124]. With these decays, the overall process fidelity a_0 can be computed.

Note that while CB is used here as a way to determine the process fidelity of a gate such as IRB, CB can be used on general, multi-qubit cycles. The dressed fidelity of a given cycle corresponds to the estimated process fidelity of that cycle when running a randomly-compiled algorithm and thus the goal of CB is to predict performance of general cycles in an algorithm.

While general benchmarking protocols like those presented here offer a general way to evaluate a quantum computer, and to try to render their performance more predictable, application-specific benchmarking protocols also provide valuable insight into how general quantum computer performance actually is. Indeed, it has been shown that performance of current superconducting quantum hardware can vary substantially, depending on the type of circuits being run. One notable work is the application of structured mirror-type circuits, which were demonstrated to exhibit very different performance on commercial quantum systems than randomized methods [125]. Nevertheless, general benchmarking methods remain valuable as a way to estimate future quantum computer performance, to provide guidance into which sources of error are most important to reduce in pursuit of bringing physical error rates below the quantum error correction threshold [126].

Bibliography

[1]P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring", 124–134 (1994).

[2]L. K. Grover, "A fast quantum mechanical algorithm for database search", STOC '96, 212–219 (1996).

[3]R. P. Feynman, "Simulating physics with computers", International Journal of Theoretical Physics **21**, 22 (1982).

[4]S. Lloyd, "Universal quantum simulators", Science **273**, 1073–1078 (1996).

[5]P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", Phys. Rev. A **52**, R2493–R2496 (1995).

[6]S. B. Bravyi and A. Y. Kitaev, "Quantum codes on a lattice with boundary", arXiv preprint quant-ph/9811052 (1998).

[7]A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation", Physical Review A **86**, 032324 (2012).

[8]C. Rigetti and M. Devoret, "Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies", Physical Review B **81**, 134507 (2010).

[9]B. K. Mitchell, R. K. Naik, A. Morvan, A. Hashim, J. M. Kreikebaum, B. Marinelli, W. Lavrijsen, K. Nowrouzi, D. I. Santiago, and I. Siddiqi, "Hardware-efficient microwave-activated tunable coupling between superconducting qubits", Physical Review Letters **127**, 200502 (2021).

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

[11]J. M. Martinis, M. H. Devoret, and J. Clarke, "Energy-level quantization in the zero-voltage state of a current-biased josephson junction", Phys. Rev. Lett. **55**, 1543–1546 (1985).

[12]Y. Nakamura, Y. A. Pashkin, and J. S. Tsai, "Coherent control of macroscopic quantum states in a single-cooper-pair box", Nature **398**, 786–788 (1999).

[13]D. Vion, A. Aassime, A. Cottet, P. Joyez, H. Pothier, C. Urbina, D. Esteve, and M. H. Devoret, "Manipulating the quantum state of an electrical circuit", Science **296**, 886–889 (2002).

[14]A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R. S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, "Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics", Nature **431**, 162–167 (2004).

[15]A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, "Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation", Physical Review A, 14 (2004).

[16]J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Charge-insensitive qubit design derived from the cooper pair box", Physical Review A **76**, 042319 (2007).

[17]J. A. Schreier, A. A. Houck, J. Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Suppressing charge noise decoherence in superconducting charge qubits", Phys. Rev. B **77**, 180502 (2008).

[18]A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, "Circuit quantum electrodynamics", Rev. Mod. Phys. **93**, 025005 (2021).

[19]P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "A quantum engineer's guide to superconducting qubits", Applied Physics Reviews **6**, 021318 (2019).

[20]M. Tinkham, *Introduction to superconductivity*, 2nd ed. (Dover Publications, June 2004).

[21]V. Ambegaokar and A. Baratoff, "Tunneling between superconductors", Phys. Rev. Lett. **10**, 486–489 (1963).

[22]F. Muller, R. Popel, J. Kohlmann, J. Niemeyer, W. Meier, T. Weimann, L. Grimm, F.-W. Dunschede, and P. Gutmann, "Optimized 1 v and 10 v josephson series arrays", IEEE Transactions on Instrumentation and Measurement **46**, 229–232 (1997).

[23]A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, "Demonstration of a high-fidelity cnot gate for fixed-frequency transmons with engineered z z suppression", Physical Review Letters **127**, 130501 (2021).

[24]I. Siddiqi, "Engineering high-coherence superconducting qubits", Nature Reviews Materials **6**, 875–891 (2021).

[25]R. K. Wangsness, "Sublattice effects in magnetic resonance", Phys. Rev. **91**, 1085–1091 (1953).

[26]F. Bloch, "Generalized theory of relaxation", Phys. Rev. **105**, 1206–1222 (1957).

[27]A. G. Redfield, "On the theory of relaxation processes", IBM Journal of Research and Development **1**, 19–31 (1957).

[28]A. A. Houck, J. A. Schreier, B. R. Johnson, J. M. Chow, J. Koch, J. M. Gambetta, D. I. Schuster, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Controlling the spontaneous emission of a superconducting transmon qubit", Phys. Rev. Lett. **101**, 080502 (2008).

[29]M. D. Reed, B. R. Johnson, A. A. Houck, L. DiCarlo, J. M. Chow, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, "Fast reset and suppressing spontaneous emission of a superconducting qubit", Applied Physics Letters **96**, 203110 (2010).

[30]E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O'Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, "Fast accurate state measurement with superconducting qubits", Physical Review Letters **112**, 190504 (2014).

[31]J. M. Kreikebaum, K. P. O'Brien, A. Morvan, and I. Siddiqi, "Improving wafer-scale Josephson junction resistance variation in superconducting quantum coherent circuits", Superconductor Science and Technology **33**, 06LT02 (2020).

[32]M. S. Blok, V. V. Ramasesh, T. Schuster, K. O'Brien, J. M. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Y. Yao, and I. Siddiqi, "Quantum information scrambling on a superconducting qutrit processor", Physical Review X **11**, 021010 (2021).

[33]J. M. Gambetta, C. E. Murray, Y.-K.-K. Fung, D. T. McClure, O. Dial, W. Shanks, J. W. Sleight, and M. Steffen, "Investigating surface loss effects in superconducting transmon qubits", IEEE Transactions on Applied Superconductivity **27**, Conference Name: IEEE Transactions on Applied Superconductivity, 1–5 (2017).

[34]V. Tripathi, M. Khezri, and A. N. Korotkov, "Operation and intrinsic error budget of a two-qubit cross-resonance gate", Physical Review A **100**, 10.1103/PhysRevA.100.012301 (2019).

[35]M. Göppl, A. Fragner, M. Baur, R. Bianchetti, S. Filipp, J. M. Fink, P. J. Leek, G. Puebla, L. Steffen, and A. Wallraff, "Coplanar waveguide resonators for circuit quantum electrodynamics", Journal of Applied Physics **104**, 113904 (2008).

[36]J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, "Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement", Phys. Rev. A **76**, 012325 (2007).

[37]D. P. DiVincenzo, "The physical implementation of quantum computation", Fortschritte der Physik **48**, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1521-3978%28200009%2948%3A PROP771%3E3.0.CO%3B2-E, 771–783 (2000).

[38]M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th anniversary edition* (Cambridge University Press, 2010).

[39]A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, "Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor", Physical Review X **11**, 041039 (2021).

[40]N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta, "Reducing unitary and spectator errors in cross resonance with optimized rotary echoes", PRX Quantum **1**, 020318 (2020).

[41]D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, "Efficient Z gates for quantum computing", Physical Review A **96**, 022330 (2017).

[42]F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple pulses for elimination of leakage in weakly nonlinear qubits", Physical Review Letters **103**, 110501 (2009).

[43]J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Optimized driving of superconducting artificial atoms for improved single-qubit gates", Physical Review A **82**, 040305 (2010).

[44]E. Lucero, J. Kelly, R. C. Bialczak, M. Lenander, M. Mariantoni, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis, "Reduced phase error through optimized control of a superconducting qubit", Physical Review A **82**, 042339 (2010).

[45]S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, "Characterizing errors on qubit operations via iterative randomized benchmarking", Physical Review A **93**, 012301 (2016).

[46]Z. Chen, "Metrology of quantum control and measurement in superconducting qubits", 241.

[47]Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and J. M. Martinis, "Measuring and suppressing quantum state leakage in a superconducting qubit", Physical Review Letters **116**, 020501 (2016).

[48]M. D. Reed, "Entanglement and quantum error correction with superconducting qubits", 382.

[49]J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, "Optimal quantum control using randomized benchmarking", Physical Review Letters **112**, 240504 (2014).

[50]M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, "Leakage reduction in fast superconducting qubit gates via optimal control", npj Quantum Information **7**, 14 (2021).

[51]K. X. Wei, E. Magesan, I. Lauer, S. Srinivasan, D. F. Bogorin, S. Carnevale, G. A. Keefe, Y. Kim, D. Klaus, W. Landers, N. Sundaresan, C. Wang, E. J. Zhang, M. Steffen, O. E. Dial, D. C. McKay, and A. Kandala, "Quantum crosstalk cancellation for fast entangling gates and improved multi-qubit performance", arXiv:2106.00675 [quant-ph] (2021).

[52]J. M. Chow, A. D. Córcoles, J. M. Gambetta, C. Rigetti, B. R. Johnson, J. A. Smolin, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, "Simple all-microwave entangling gate for fixed-frequency superconducting qubits", Physical Review Letters **107**, 080502 (2011).

[53]M. Malekakhlagh, E. Magesan, and D. C. McKay, "First-principles analysis of cross-resonance gate operation", Physical Review A **102**, 042605 (2020).

[54]S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, "Procedure for systematically tuning up cross-talk in the cross-resonance gate", Physical Review A **93**, 060302 (2016).

[55]E. Magesan and J. M. Gambetta, "Effective hamiltonian models of the cross-resonance gate", Physical Review A **101**, 052308 (2020).

[56]M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, "Dynamics of superconducting qubit relaxation times", arXiv:2105.15201 [cond-mat, physics:quant-ph] (2021).

[57]S. A. Hill and W. K. Wootters, "Entanglement of a pair of quantum bits", Phys. Rev. Lett. **78**, 5022–5025 (1997).

[58]M. Ware, B. R. Johnson, J. M. Gambetta, T. A. Ohki, J. M. Chow, and B. L. T. Plourde, "Cross-resonance interactions between superconducting qubits with variable detuning", arXiv:1905.11480 [quant-ph] (2019).

[59]R. Bianchetti, S. Filipp, M. Baur, J. M. Fink, C. Lang, L. Steffen, M. Boissonneault, A. Blais, and A. Wallraff, "Control and tomography of a three level superconducting artificial atom", Physical Review Letters **105**, 223601 (2010).

[60]S. Gustavsson, O. Zwier, J. Bylander, F. Yan, F. Yoshihara, Y. Nakamura, T. P. Orlando, and W. D. Oliver, "Improving quantum gate fidelities by using a qubit to measure microwave pulse distortions", Physical Review Letters **110**, 040502 (2013).

[61]L. S. Theis, F. Motzoi, S. Machnes, and F. K. Wilhelm, "Counteracting systems of diabaticities using DRAG controls: the status after 10 years [(a)]", EPL (Europhysics Letters) **123**, 60001 (2018).

[62]J. Ghosh, A. G. Fowler, J. M. Martinis, and M. R. Geller, "Understanding the effects of leakage in superconducting quantum-error-detection circuits", Physical Review A **88**, 062329 (2013).

[63]E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, "Randomized benchmarking of quantum gates", Physical Review A, 7 (2008).

[64]J. Emerson, R. Alicki, and K. Życzkowski, "Scalable noise estimation with random unitary operators", Journal of Optics B: Quantum and Semiclassical Optics **7**, S347–S352 (2005).

[65]S. J. Beale, A. Carignan-Dugas, D. Dahlen, J. Emerson, I. Hincks, P. Iyer, A. Jain, D. Hufnagel, E. Ospadov, J. Saunders, A. Stasiuk, J. J. Wallman, and A. Winick, *True-Q*, June 2020.

[66]C. J. Wood and J. M. Gambetta, "Quantification and characterization of leakage errors", Physical Review A **97**, 032306 (2018).

[67]J. Wallman, C. Granade, R. Harper, and S. T. Flammia, "Estimating the coherence of noise", New Journal of Physics **17**, 113020 (2015).

[68]A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, "Characterizing large-scale quantum computers via cycle benchmarking", Nature Communications **10**, 5347 (2019).

[69]A. Carignan-Dugas, J. J. Wallman, and J. Emerson, "Bounding the average gate fidelity of composite channels using the unitarity", New Journal of Physics **21**, 053016 (2019).

[70]J. J. Wallman and J. Emerson, "Noise tailoring for scalable quantum computation via randomized compiling", Physical Review A **94**, 052325 (2016).

[71]N. Hansen, yoshihikoueno, ARF1, K. Nozawa, L. Rolshoven, M. Chan, Y. Akimoto, briegl-hostis, and D. Brockhoff, *Cma-es/pycma: r3.2.1*, version r3.2.1, Mar. 2022.

[72]S. Krinner, S. Lazar, A. Remm, C. Andersen, N. Lacroix, G. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, "Benchmarking Coherent Errors in Controlled-Phase Gates due to Spectator Qubits", Physical Review Applied **14**, 024042 (2020).

[73]M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Lazar, J.-C. Besse, T. Walter, A. Wallraff, and C. Eichler, "Implementation of Conditional Phase Gates Based on Tunable Z Z Interactions", Physical Review Letters **125**, 240502 (2020).

[74]Y. Xu, J. Chu, J. Yuan, J. Qiu, Y. Zhou, L. Zhang, X. Tan, Y. Yu, S. Liu, J. Li, F. Yan, and D. Yu, "High-Fidelity, High-Scalability Two-Qubit Gate Scheme for Superconducting Qubits", Physical Review Letters **125**, 240503 (2020).

[75]P. Mundada, G. Zhang, T. Hazard, and A. Houck, "Suppression of Qubit Crosstalk in a Tunable Coupling Superconducting Circuit", Physical Review Applied **12**, 054023 (2019).

[76]V. Negîrneac, H. Ali, N. Muthusubramanian, F. Battistel, R. Sagastizabal, M. S. Moreira, J. F. Marques, W. Vlothuizen, M. Beekman, N. Haider, A. Bruno, and L. DiCarlo, "High-fidelity controlled-Z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor", arXiv:2008.07411 [quant-ph], arXiv: 2008.07411 (2020).

[77]J. Stehlik, D. M. Zajac, D. L. Underwood, T. Phung, J. Blair, S. Carnevale, D. Klaus, G. A. Keefe, A. Carniol, M. Kumph, M. Steffen, and O. E. Dial, "Tunable Coupling Architecture for Fixed-frequency Transmons", arXiv:2101.07746 [quant-ph], arXiv: 2101.07746 (2021).

[78]Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim, A. Melville, B. M. Niedzielski, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler", arXiv:2011.01261 [quant-ph] (2020).

[79]G. S. Paraoanu, "Microwave-induced coupling of superconducting qubits", Physical Review B **74**, 140504 (2006).

[80]J. M. Chow, J. M. Gambetta, A. W. Cross, S. T. Merkel, C. Rigetti, and M. Steffen, "Microwave-activated conditional-phase gate for superconducting qubits", New Journal of Physics **15**, 115012 (2013).

[81]S. Puri and A. Blais, "High-Fidelity Resonator-Induced Phase Gate with Single-Mode Squeezing", Physical Review Letters **116**, 180501 (2016).

[82]H. Paik, A. Mezzacapo, M. Sandberg, D. T. McClure, B. Abdo, A. D. Córcoles, O. Dial, D. F. Bogorin, B. L. T. Plourde, M. Steffen, A. W. Cross, J. M. Gambetta, and J. M. Chow, "Experimental Demonstration of a Resonator-Induced Phase Gate in a Multiqubit Circuit-QED System", Physical Review Letters **117**, 250502 (2016).

[83]J. Long, T. Zhao, M. Bal, R. Zhao, G. S. Barron, H.-s. Ku, J. A. Howard, X. Wu, C. R. H. McRae, X.-H. Deng, G. J. Ribeill, M. Singh, T. A. Ohki, E. Barnes, S. E. Economou, and D. P. Pappas, "A universal quantum gate set for transmon qubits with strong ZZ interactions", arXiv:2103.12305 [quant-ph], arXiv: 2103.12305 (2021).

[84]S. Krinner, P. Kurpiers, B. Royer, P. Magnard, I. Tsitsilin, A. Blais, and A. Wallraff, "Demonstration of an All-Microwave Controlled-Phase Gate between Far-Detuned Qubits", Physical Review Applied, 11 (2020).

[85]A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, "Demonstration of a High-Fidelity CNOT for Fixed-Frequency Transmons with Engineered ZZ Suppression", arXiv:2011.07050 [quant-ph], arXiv: 2011.07050 (2020).

[86]L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Demonstration of two-qubit algorithms with a superconducting quantum processor", Nature **460**, 240–244 (2009).

[87]J. M. Gambetta, A. D. Corcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, "Characterization of addressability by simultaneous randomized benchmarking", Physical Review Letters **109**, arXiv: 1204.6308, 240504 (2012).

[88]D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, "Three-Qubit Randomized Benchmarking", Physical Review Letters **122**, 200502 (2019).

[89]A. Morvan, V. V. Ramasesh, M. S. Blok, J. M. Kreikebaum, K. O'Brien, L. Chen, B. K. Mitchell, R. K. Naik, D. I. Santiago, and I. Siddiqi, "Qutrit randomized benchmarking", arXiv:2008.09134 [quant-ph], arXiv: 2008.09134 (2020).

[90]J. Ku, X. Xu, M. Brink, D. C. McKay, J. B. Hertzberg, M. H. Ansari, and B. L. T. Plourde, "Suppression of Unwanted Z Z Interactions in a Hybrid Two-Qubit System", Physical Review Letters **125**, 200504 (2020).

[91]H. Xiong, Q. Ficheux, A. Somoroff, L. B. Nguyen, E. Dogan, D. Rosenstock, C. Wang, K. N. Nesterov, M. G. Vavilov, and V. E. Manucharyan, "Arbitrary controlled-phase gate on fluxonium qubits using differential ac-Stark shifts", arXiv:2103.04491 [cond-mat, physics:quant-ph], arXiv: 2103.04491 (2021).

[92]J. B. Hertzberg, E. J. Zhang, S. Rosenblatt, E. Magesan, J. A. Smolin, J.-B. Yau, V. P. Adiga, M. Sandberg, M. Brink, J. M. Chow, and J. S. Orcutt, "Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors", arXiv:2009.00781 [cond-mat, physics:quant-ph], arXiv: 2009.00781 (2020).

[93]E. J. Zhang, S. Srinivasan, N. Sundaresan, D. F. Bogorin, Y. Martin, J. B. Hertzberg, J. Timmerwilke, E. J. Pritchett, J.-B. Yau, C. Wang, W. Landers, E. P. Lewandowski, A. Narasgond, S. Rosenblatt, G. A. Keefe, I. Lauer, M. B. Rothwell, D. T. McClure, O. E. Dial, J. S. Orcutt, M. Brink, and J. M. Chow, "High-fidelity superconducting quantum processors via laser-annealing of transmon qubits", arXiv:2012.08475 [quant-ph], arXiv: 2012.08475 (2020).

[94]J. Preskill, "Quantum Computing in the NISQ era and beyond", Quantum **2**, arXiv: 1801.00862, 79 (2018).

[95]N. Lacroix, C. Hellings, C. K. Andersen, A. Di Paolo, A. Remm, S. Lazar, S. Krinner, G. J. Norris, M. Gabureac, J. Heinsoo, A. Blais, C. Eichler, and A. Wallraff, "Improving the Performance of Deep Quantum Optimization Algorithms with Continuous Gate Sets", PRX Quantum **1**, 110304 (2020).

[96]B. Foxen, C. Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, S. Boixo, D. Buell, B. Burkett, Y. Chen, R. Collins, E. Farhi, A. Fowler, C. Gidney, M. Giustina, R. Graff, M. Harrigan, T. Huang, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, P. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, E. Lucero, J. McClean, M. McEwen, X. Mi, M. Mohseni, J. Y. Mutus, O. Naaman, M. Neeley, M. Niu, A. Petukhov, C. Quintana, N. Rubin, D. Sank, V. Smelyanskiy, A. Vainsencher, T. C. White, Z. Yao, P. Yeh, A. Zalcman, H. Neven, J. M. Martinis, and Google AI Quantum, "Demonstrating a Continuous Set of Two-qubit Gates for Near-term Quantum Algorithms", Physical Review Letters **125**, 120504 (2020).

[97]R. Johansson, P. Nation, A. Vardhan, A. Pitchford, J. Lishman, C. Granade, E. Giguère, B. Li, S. Saraogi, N. Shammah, A. L. Grimsmo, M. Baden, P. Migdał, D. Vasilyev, K. Fischer, T. Raheja, S. Ahmed, sbisw002, A. Galicia, M. S. Costa, S. Cross, B. Criger, W. Rzadkowski, J. Feist, A. Upadhyaya, Canoming, Zomtir, alexbrc, N. Quesada, and S. Krastanov, *Qutip/qutip: QuTiP 4.6.1*, May 2021.

[98]N. F. Ramsey, "A Molecular Beam Resonance Method with Separated Oscillating Fields", Physical Review **78**, 695–699 (1950).

[99]E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, "Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking", Physical Review Letters **109**, 10.1103/PhysRevLett.109.080505 (2012).

[100]R. Harper, "Statistical analysis of randomized benchmarking", Physical Review A, 7 (2019).

[101]M. Kounalakis, C. Dickel, A. Bruno, N. K. Langford, and G. A. Steele, "Tuneable hopping and nonlinear cross-Kerr interactions in a high-coherence superconducting circuit", npj Quantum Information **4**, 38 (2018).

[102]P. Roushan, C. Neill, J. Tangpanitanon, V. M. Bastidas, A. Megrant, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, B. Foxen, M. Giustina, E. Jeffrey, J. Kelly, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. White, H. Neven, D. G. Angelakis, and J. Martinis, "Spectroscopic signatures of localization with interacting photons in superconducting qubits", Science **358**, 1175–1179 (2017).

[103]Y. Ye, Z.-Y. Ge, Y. Wu, S. Wang, M. Gong, Y.-R. Zhang, Q. Zhu, R. Yang, S. Li, F. Liang, J. Lin, Y. Xu, C. Guo, L. Sun, C. Cheng, N. Ma, Z. Y. Meng, H. Deng, H. Rong, C.-Y. Lu, C.-Z. Peng, H. Fan, X. Zhu, and J.-W. Pan, "Propagation and Localization of Collective Excitations on a 24-Qubit Superconducting Processor", Physical Review Letters **123**, 050502 (2019).

[104]J. Jin, D. Rossini, R. Fazio, M. Leib, and M. J. Hartmann, "Photon Solid Phases in Driven Arrays of Nonlinearly Coupled Cavities", Physical Review Letters **110**, 163605 (2013).

[105]P. Roushan, C. Neill, A. Megrant, Y. Chen, R. Babbush, R. Barends, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, E. Jeffrey, J. Kelly, E. Lucero, J. Mutus, P. J. J. O'Malley, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. White, E. Kapit, H. Neven, and J. Martinis, "Chiral ground-state currents of interacting photons in a synthetic magnetic field", Nature Physics **13**, 146–151 (2017).

[106]X. Guan, Y. Feng, Z.-Y. Xue, G. Chen, and S. Jia, "Synthetic gauge field and chiral physics on two-leg superconducting circuits", Physical Review A **102**, 032610 (2020).

[107]C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, "Repeated quantum error detection in a surface code", Nature Physics **16**, 875–880 (2020).

[108]A. Morvan, L. Chen, J. M. Larson, D. I. Santiago, and I. Siddiqi, "Optimizing frequency allocation for fixed-frequency superconducting quantum processors", arXiv:2112.01634 [quant-ph] (2022).

[109]M. A. Nielsen, "A simple formula for the average gate fidelity of a quantum dynamical operation", Physics Letters A **303**, 249–252 (2002).

[110]M. A. Nielsen, "The entanglement fidelity and quantum error correction", arXiv:quant-ph/9606012 (1996).

[111]A. Jamiołkowski, "Linear transformations which preserve trace and positive semidefiniteness of operators", Reports on Mathematical Physics **3**, 275–278 (1972).

[112]M.-D. Choi, "Completely positive linear maps on complex matrices", Linear Algebra and its Applications **10**, 285–290 (1975).

[113]P. Aliferis, F. Brito, D. P. DiVincenzo, J. Preskill, M. Steffen, and B. M. Terhal, "Fault-tolerant computing with biased-noise superconducting qubits: a case study", New Journal of Physics **11**, 013061 (2009).

[114]J. J. Wallman and S. T. Flammia, "Randomized benchmarking with confidence", New Journal of Physics **16**, 103032 (2014).

[115]Y. R. Sanders, J. J. Wallman, and B. C. Sanders, "Bounding quantum gate error rate based on reported average fidelity", New Journal of Physics **18**, 012002 (2015).

[116]S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, "Self-consistent quantum process tomography", Phys. Rev. A **87**, 062119 (2013).

[117]D. Greenbaum, "Introduction to quantum gate set tomography", arXiv:1509.02921 [quant-ph] (2015).

[118]C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation", Phys. Rev. A **80**, 012304 (2009).

[119]A. G. Fowler, "Coping with qubit leakage in topological codes", Phys. Rev. A **88**, 042308 (2013).

[120]P. Aliferis and B. M. Terhal, "Fault-tolerant quantum computation for local leakage faults", arXiv:quant-ph/0511065 (2006).

[121]M. Suchara, A. W. Cross, and J. M. Gambetta, "Leakage suppression in the toric code", Quantum Info. Comput. **15**, 997–1016 (2015).

[122]M. McEwen, D. Kafri, Z. Chen, J. Atalaya, K. J. Satzinger, C. Quintana, P. V. Klimov, D. Sank, C. Gidney, A. G. Fowler, F. Arute, K. Arya, B. Buckley, B. Burkett, N. Bushnell, B. Chiaro, R. Collins, S. Demura, A. Dunsworth, C. Erickson, B. Foxen, M. Giustina, T. Huang, S. Hong, E. Jeffrey, S. Kim, K. Kechedzhi, F. Kostritsa, P. Laptev, A. Megrant, X. Mi, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Niu, A. Paler, N. Redd, P. Roushan, T. C. White, J. Yao, P. Yeh, A. Zalcman, Y. Chen, V. N. Smelyanskiy, J. M. Martinis, H. Neven, J. Kelly, A. N. Korotkov, A. G. Petukhov, and R. Barends, "Removing leakage-induced correlated errors in superconducting quantum error correction", Nature Communications **12**, 1761 (2021).

[123]E. Knill, "Quantum computing with realistically noisy devices", Nature **434**, 39–44 (2005).

[124]A. Carignan-Dugas, J. J. Wallman, and J. Emerson, "Characterizing universal gate sets via dihedral benchmarking", Phys. Rev. A **92**, 060302 (2015).

[125]T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, "Measuring the capabilities of quantum computers", Nature Physics **18**, 75–79 (2022).

[126]Google Quantum AI, Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, S. Hong, C. Jones, A. Petukhov, D. Kafri, S. Demura, B. Burkett, C. Gidney, A. G. Fowler, A. Paler, H. Putterman, I. Aleiner, F. Arute, K. Arya, R. Babbush, J. C. Bardin, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, N. Bushnell, B. Chiaro, R. Collins, W. Courtney, A. R. Derk, D. Eppens, C. Erickson, E. Farhi, B. Foxen, M. Giustina, A. Greene, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, T. Huang, W. J. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, K. Kechedzhi, S. Kim, A. Kitaev, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, T. McCourt, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Newman, M. Y. Niu, T. E. O'Brien, A. Opremcak, E. Ostby, B. Pató, N. Redd, P. Roushan, N. C. Rubin, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. White, Z. J. Yao, P. Yeh, J. Yoo, A. Zalcman, H. Neven, S. Boixo, V. Smelyanskiy, Y. Chen, A. Megrant, and J. Kelly, "Exponential suppression of bit or phase errors with cyclic error correction", Nature **595**, 383–387 (2021).