Characterizing Midcircuit Measurements on a Superconducting Qubit Using Gate Set Tomography

Kenneth Rudinger, Guilhem J. Ribeill, Luke C.G. Govia, Matthew Ware, Erik Nielsen, Kevin Young, Thomas A. Ohki, Robin Blume-Kohout, and Timothy Proctor

Phys. Rev. Applied **17**, 014014 — Published 12 January 2022

DOI: 10.1103/PhysRevApplied.17.014014

Characterizing midcircuit measurements on a superconducting qubit using gate set tomography

Kenneth Rudinger,[1,*] Guilhem J. Ribeill,[2,†] Luke C. G. Govia,[2] Matthew Ware,[2]
Erik Nielsen,[1] Kevin Young,[1] Thomas A. Ohki,[2] Robin Blume-Kohout,[1] and Timothy Proctor[1]

[1]*Quantum Performance Laboratory, Sandia National Laboratories, Albuquerque, NM 87185, USA and Livermore, CA 94550, USA*
[2]*Quantum Engineering and Computing, Raytheon BBN Technologies, 10 Moulton St., Cambridge, MA 02138, USA*
(Dated: November 18, 2021)

Measurements that occur within the internal layers of a quantum circuit — *midcircuit measurements* — are an important quantum computing primitive, most notably for quantum error correction. Midcircuit measurements have both classical and quantum outputs, so they can be subject to error modes that do not exist for measurements that terminate quantum circuits. Here we show how to characterize midcircuit measurements, modeled by quantum instruments, using a technique that we call *quantum instrument linear gate set tomography* (QILGST). We then apply this technique to characterize a dispersive measurement on a superconducting transmon qubit within a multiqubit system. By varying the delay time between the measurement pulse and subsequent gates, we explore the impact of residual cavity photon population on measurement error. QILGST can resolve different error modes and quantify the total error from a measurement; in our experiment, for delay times above 1000 ns we measured a total error rate (i.e., half diamond distance) of $\epsilon_\diamond = 8.1 \pm 1.4\%$, a readout fidelity of $97.0 \pm 0.3\%$, and output quantum state fidelities of $96.7 \pm 0.6\%$ and $93.7 \pm 0.7\%$ when measuring 0 and 1, respectively.

I. INTRODUCTION

Gate-model quantum computers perform computations by executing sequences of quantum operations, called quantum circuits. Quantum computations can be performed with circuits containing only qubit initialization, reversible logic gates, and *terminating measurements* [1] — measurements that occur at the circuit's end, converting quantum information stored in the qubits into classical bits. However, circuits can also contain *midcircuit measurements* that extract information from the qubits and alter their state, but don't destroy the qubits nor necessarily collapse their state entirely.

As parity check or stabilizer measurements (Fig. 1a) must extract information about a multiqubit observable, while not disturbing quantum information stored in the logical subspace, high-fidelity midcircuit measurements are essential for quantum error correction (QEC) [2–7]. Midcircuit measurements also have applications to error mitigation and near-term algorithms [8–12].

Midcircuit measurements, however, admit failure modes that don't exist for terminating measurements. Techniques for accurate *characterization* of midcircuit measurements are therefore urgently needed. In this paper, we introduce a protocol (Fig. 1b) for comprehensive, self-consistent characterization of a full set of logic operations that includes midcircuit measurements — which we call *quantum instrument linear gate set tomography* (QILGST). We use QILGST to study single-qubit dispersive measurements on a superconducting transmon processor (Fig. 1d).

Techniques for assessing performance of quantum logic operations can be divided into benchmarking and characterization. Benchmarks quantify overall performance of operations *in situ* on representative tasks, and midcircuit measurements can be benchmarked using QEC (and components thereof) [17–27] or algorithm [9] circuits. However, identifying specific error modes, predicting their impact, and mitigating or eliminating them requires detailed characterization. This is commonly done by tomography, which means estimating a model for the

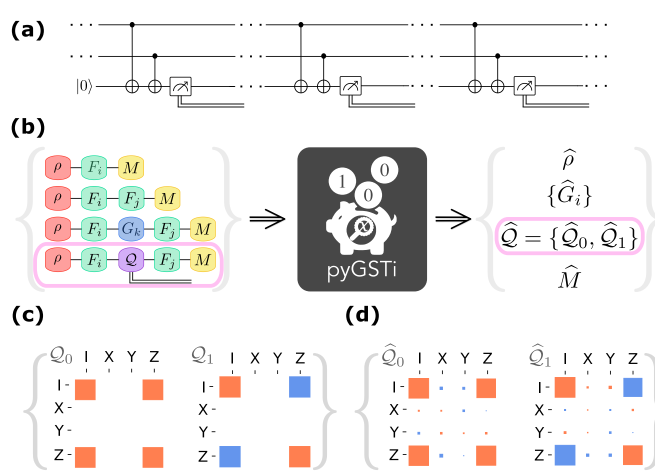

Figure 1. **Characterizing midcircuit measurements.** Many quantum computing primitives require midcircuit measurements, exemplified by **(a)** a repeated parity-check circuit. **(b)** QILGST protocol for characterizing a midcircuit measurement, as part of a complete gate set ($\mathcal{G} = \{\rho, G_i, Q, M\}$). The midcircuit measurement is modeled by a quantum instrument $Q = \{Q_i\}$, which consists of a process matrix for each measurement outcome. QILGST consists of (1) running process tomography circuits on Q, alongside standard GST circuit [13–15]; and (2) matrix inversion or maximum likelihood estimation (using `pyGSTi` [16]) to self-consistently reconstruct the gate set ($\widehat{\mathcal{G}} = \{\widehat{\rho}, \widehat{G}_i, \widehat{Q}, \widehat{M}\}$). Additions to standard GST are circled in pink. We applied QILGST to characterize a dispersive σ_z basis measurement on a transmon qubit. The **(c)** target and **(d)** estimated QI from our experiment, with a readout fidelity of $97.0\% \pm 0.3\%$ and a total error rate of $\epsilon_\diamond = 8.1 \pm 1.4\%$. Each orange (blue) square represents a positive (negative) real number with magnitude proportional to area.

* kmrudin@sandia.gov
† guilhem.ribeill@raytheon.com

operation. Terminating measurements are modeled by *positive operator-valued measures* (POVMs) and can be estimated by *quantum detector tomography* [28–33], but only using precalibrated input states and gates. Gate set tomography (GST) [13–15, 34] removes this requirement, enabling estimation of POVMs self-consistently with initialization and gates. We show how to extend GST to gate sets that include midcircuit measurements, represented as *quantum instruments* [35]. Prior works [36–39] showed how to perform self-testing or tomography of quantum instruments, but, to the best of our knowledge, this is the first protocol for complete *and* self-consistent tomography of midcircuit measurements.

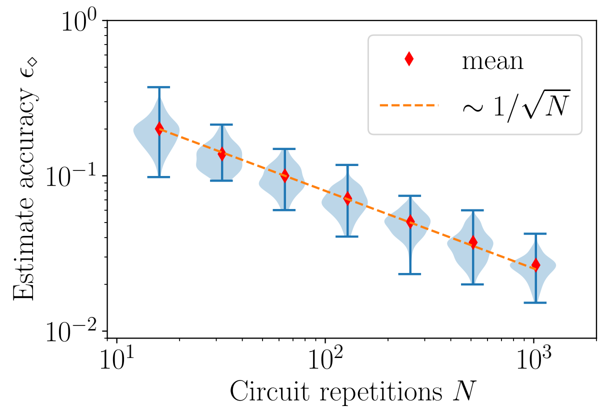

Figure 2. **QILGST accurately characterizes midcircuit measurements.** We simulated single-qubit QILGST under a variety of error models and computed the accuracy of the estimated QI $\widehat{Q}$; see Appendix C. This plot shows the estimation accuracy, measured by half the diamond distance ($\epsilon_\diamond$) [41] between $\widehat{Q}$ and the QI used to generate the data, versus the number of circuit repetitions (N). Each point (violin plot) is the mean (distribution) of the estimation inaccuracy from simulating QILGST under 100 error models. Accuracy scales as $O(1/\sqrt{N})$, the expected shot noise scaling.

II. CHARACTERIZING MIDCIRCUIT MEASUREMENTS

A. Quantum instruments

Quantum instruments (QIs) [35] model midcircuit measurements. In tomography, a quantum processor's state is represented by a $d \times d$ density matrix, where d is the (intended) dimension of its Hilbert space. Gates are represented by superoperators acting linearly on density matrices and terminating measurements by POVMs that map states to probability distributions. All these objects are completely positive (CP) and trace-preserving (TP) *quantum processes*, with different input and output spaces. States (density matrices) describe initialization, mapping a trivial space *into* the d^2-dimensional space of mixed states. Gate superoperators map that space to itself. POVMs map quantum states to distributions over outcomes. QIs are processes with a quantum input, and quantum *and* classical outputs. This describes a midcircuit measurement, combining the features (and outputs) of a POVM and a gate. The simplest representation of an m-outcome QI Q is as a set of m CP maps $Q = \{Q_0, \ldots, Q_{m-1}\}$ whose sum $\sum_i Q_i$ is TP. The QI maps ρ to a joint quantum-classical state $\{(p_i, \rho_i)\}_{i=0}^{m-1}$, where

$$p_i = \mathrm{Tr}(Q_i[\rho]) \tag{1}$$

is the probability of observing outcome i and

$$\rho_i = Q_i[\rho]/p_i \tag{2}$$

is the output state *conditional on* observing i. Like gates, each Q_i can be represented using a $d^2 \times d^2$ matrix (see Fig. 1c for an example, with matrix elements defined by $[Q_j]_{kl} = \mathrm{Tr}(\sigma_k Q_j[\sigma_l])$ for $k, l = I, x, y, z$).

Quantum instruments model errors in midcircuit measurements that POVMs cannot. POVMs have strictly classical outputs, but any POVM can be "promoted" to a limited kind of QI called a "measure-and-prepare" process [40], by following it with a conditional re-initialization (i.e., upon observing i, ρ_i is prepared). Measure-and-prepare processes can be characterized with existing methods (e.g., GST), but cannot describe all midcircuit measurements. Measure-and-prepare processes destroy all entanglement with other quantum systems [40], but, e.g., QEC parity-checks should preserve inter-qubit entanglement. Conversely, midcircuit measurements *designed* as measure-and-prepare processes can fail in ways that cannot be modeled without a general QI. QIs can model and describe all Markovian errors in midcircuit measurement. Our goal is to reconstruct (from data) the QI that describes an experimental midcircuit measurement.

B. GST with quantum instruments

GST [13–15] self-consistently reconstructs all elements of a gate set $\mathcal{G}$ — an initialization ρ, two or more logic gates $\{G_i\}$, and a terminating measurement M. It specifies (1) an experiment design (circuits to be performed) and (2) analysis procedures for transforming data into a gate set estimate. Several variants exist [15]; we adapt *linear-inversion* GST (LGST) to gate sets containing midcircuit measurements. LGST resembles process tomography [42–44], with three key innovations: (1) to tomograph each gate G_i, the experiment includes all circuits of the form $F_j^p G_i F_k^m$ where *fiducial circuits* $\{F_j^p\}_{j=1}^{n_p}$ and $\{F_k^m\}_{k=1}^{n_m}$ produce informationally complete ensembles of states and terminating measurements, respectively, using only gates in $\mathcal{G}$; (2) the experiment includes circuits for process tomography on the null operation ($F_j^p F_k^m$ circuits); and (3) systematic errors are removed using the inverse of the tomographed null operation [13, 15]. To extend LGST to a gate set containing a midcircuit measurement, represented by a QI Q [45], we add all circuits of the form $F_j^p Q F_k^m$ to the LGST experiment (Fig. 1b); these circuits output a result from both the midcircuit *and* terminating measurement.

Analyzing QILGST data presents one complication. Whereas each gate G_i is represented by a single CPTP map, a QI defines a *set* of CP maps $\{Q_0, \cdots, Q_{m-1}\}$. Which Q_i appears in each run of the circuit is not controllable; it's determined by the midcircuit measurement's outcome. To reconcile this with LGST analysis, we represent the QI by an $md^2 \times d^2$ process matrix

$$Q = \left(Q_0, \ldots, Q_{m-1} \right)^\top , \quad (3)$$

which is CPTP. The m blocks correspond to copies of the quantum state space, indexed by the measurement's classical outcome. So whereas the LGST linear inversion algorithm for a gate G begins with a $d^2 \times d^2$ matrix of directly measured probabilities $\tilde{G}_{kj}$ [15] — where k labels a final measurement setting and the column j labels a preparation setting — the corresponding algorithm for a QI Q starts with an $md^2 \times d^2$ matrix of probabilities $\tilde{Q}_{kj}$ where k labels a final measurement setting *and* which outcome of Q was observed. With this modification, the LGST algorithm can be directly applied, with the matrix elements of Q estimated to the same absolute precision as those of a gate G.

We call this protocol *quantum instrument linear GST* (QILGST). It requires only 128 circuits to characterize a single-qubit gate set including 3 gates and a QI Q [46]. For all analyses presented here, we used numerical maximum likelihood estimation (implemented in `pyGSTi` [16, 47]), instead of linear inversion. This yields higher accuracy by accounting for heteroskedasticity in data [48]. Data analysis for single-qubit QILGST takes a few seconds on a modern laptop. To verify the correctness of QILGST, we simulated it for a variety of error models (Appendix C); QILGST correctly reconstructs the QIs (Fig. 2).

C. Quantifying errors in a midcircuit measurement

Running QILGST yields estimates of all gates *and* an estimated QI. Like all GST estimates, it has a gauge freedom [15], which we fix by numerically optimizing the gauge to minimize discrepancy between estimated gates and their targets [15]. We denote the gauge-optimized estimate of Q by $\widehat{Q}$. The estimated $\widehat{Q}$ can be compared to the ideal "target" QI (Q_{target}), to quantify errors in the midcircuit measurements. Like gates, midcircuit measurements can display a variety of errors, e.g., measuring the wrong observable, scrambling classical information in the measurement result, or creating wrong post-measurement quantum states. The quality of a logic operation is commonly summarized by metrics such as fidelity or diamond norm. Fidelity between QIs [49] is difficult to interpret because of the joint quantum/classical output, but diamond distance error $\epsilon_\diamond = \frac{1}{2}||\widehat{Q} - Q_{\text{target}}||_\diamond$ [41] is well-defined and a tight upper bound on the change in any experimental probability induced by errors in Q.

III. EXPERIMENTAL RESULTS

We used QILGST to study midcircuit measurements on a transmon qubit within a five-qubit device. We performed dispersive measurements through a microwave cavity coupled to the qubit using standard circuit QED [50] methods. We

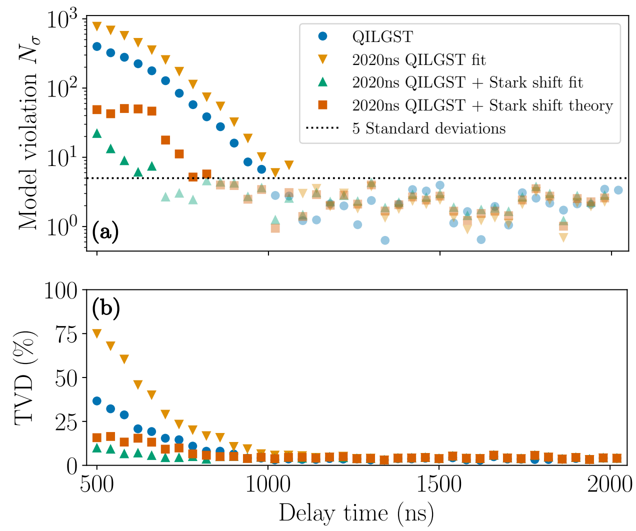

Figure 3. **Characterizing non-Markovian errors in a midcircuit measurement on a superconducting qubit.** Evidence for unmodeled error as a function of the delay time (t_d) between the measurement pulse and subsequent operations, for four different models estimated from the data using QILGST. Evidence for unmodeled error is quantified by **(a)** the number of standard deviations (N_σ) of model violation, and **(b)** the largest total variation distance (TVD) between the model's prediction and the data, for the 36 circuits containing a midcircuit measurement. The gate set estimated by QILGST (circles) does not accurately describe the data for small t_d, indicating non-Markovian errors. This additional error can be modeled by combining QILGST's model estimated from the 2020 ns data — which, alone, is not consistent with the small t_d data (down-triangles) — with a decaying Stark shift error on the gates that follow the measurement (squares and up-triangles).

achieved a high readout fidelity of ~ 96 % using a JTWPA amplifier [51], with a 1 µs measurement pulse resonant with the qubit ground-state shifted cavity frequency, that is subsequently digitized and integrated using a matched-filter kernel [52]. The measurement pulse amplitude, measured through the qubit Stark shift [53], created an average cavity population of $\bar{n} = 122$ for the qubit ground state (and substantially less for the excited state), well below the critical photon number $n_c = \alpha\Delta/[4\chi(\alpha+\Delta)] = 248$. Further device and experimental details can be found in Appendices A and B.

The gate set $\mathcal{G}$ consisted of $\pi/2$ rotations around the σ_x and σ_y axes, an idle operation, midcircuit and terminating measurements in the σ_z basis, and state preparation in $|0\rangle$ (implemented by a 500 µs idle reset). The midcircuit measurement's target QI (Fig. 1c) is $Q_{\text{target}} = \{Q_{\text{target},0}, Q_{\text{target},1}\}$, where

$$Q_{\text{target},k}[\rho] = \text{Tr}\left[\frac{1}{2}\left(\sigma_I + (-1)^k\sigma_z\right)\rho\right]\left(\sigma_I + (-1)^k\sigma_z\right). \quad (4)$$

For this gate set, there are 128 QILGST circuits [54], 36 of which contain a midcircuit measurement. We ran the QILGST experiment (with $N = 1024$ circuit repetitions) multiple times; each run used a different time delay t_d between the midcircuit measurement pulse and subsequent operations, with 500 ns ≤

$t_d \leq 2020\,$ns [55]. This produced a QILGST dataset $D(t_d)$ for each t_d. We applied the QILGST analysis to each $D(t_d)$ independently, producing an estimated gate set $\widehat{\mathcal{G}}(t_d)$ for each t_d.

A. Testing for non-Markovian errors

The gate set estimated by QILGST will accurately describe the data if errors on all operations are Markovian. Non-Markovian errors are common however [14, 56–60], so we check whether $\widehat{\mathcal{G}}(t_d)$ is consistent with $D(t_d)$ using the log-likelihood ratio test statistic λ_{LLR} [14, 58, 61]. This λ_{LLR} is $N_\sigma \sim 400$ standard deviations above its expected value (if the QILGST model is valid) when $t_d = 500\,$ns, but $N_\sigma \lesssim 5$ if $t_d \geq 1020\,$ns (Fig. 3a, circles). Therefore, short delay times are causing non-Markovian errors. We quantify the size of the unmodeled effect by the total variation distance (TVD) between the probabilities predicted by $\widehat{\mathcal{G}}(t_d)$ and the observed frequencies $D(t_d)$ [58]. The maximum TVD for the 36 QI-containing circuits is large at the shortest delay times (37% at $t_d = 500\,$ns), but is small ($< 6\%$) for $t_d \geq 1020\,$ns (Fig. 3c, circles). We attribute this large non-Markovian error at short delay times to residual photon population in the measurement cavity (we measure the relaxation rate of the cavity to be $\kappa^{-1} = 242\,$ns), as we didn't perform active reset of the cavity state [53, 62].

B. Midcircuit measurements with long delay times

Before investigating the non-Markovian effects observed for $t_d \leq 900\,$ns, we present the results of QILGST at long delay times ($t_d \geq 1020\,$ns), where the estimated gate sets *do* accurately describe the data. When $t_d \geq 1020\,$ns $\approx 4.2/\kappa$ the cavity photon population is negligible, so we expect that the only variation across those t_d values will be a small increase in relaxation errors, in the midcircuit measurement's preparation of $|1\rangle\langle 1|$, for longer delay times. As $\widehat{\mathcal{G}}(2020\,\mathrm{ns})$ accurately models the data for all $t_d \geq 1120\,$ns (Fig. 3, down-triangles), we focus on $\widehat{\mathcal{G}} \equiv \widehat{\mathcal{G}}(2020\,\mathrm{ns})$ and $\widehat{Q} \equiv \widehat{Q}(2020\,\mathrm{ns})$, our estimate of the midcircuit measurement's QI. Figs. 1d and 1c show process matrices for $\widehat{Q}$ and the target Q_{target}, respectively. We find a total error in $\widehat{Q}$ is $\epsilon_\diamond = 8.1 \pm 1.4\%$ (error bars are 2σ). This metric quantifies *all* errors in the measurement, including readout errors *and* errors in the post-measurement state.

To verify the estimate's consistency with standard metrics, we calculate the readout fidelity $F = \frac{1}{2}[P_{0|0} + P_{1|1}]$ of the midcircuit measurement, where $P_{0|0}$ (resp., $P_{1|1}$) is the (marginal) probability of reading out 0 (resp., 1) in the midcircuit measurement of the prepare-measure-measure (resp., prepare-π-pulse-measure-measure) circuit. Both circuits are part of the QILGST experiment, so we can predict F from $\widehat{\mathcal{G}}$ and compare this to observed frequencies in $D(2020\,\mathrm{ns})$. The predicted and observed values are $F = 97.0 \pm 0.3\%$ and $F = 97.3 \pm 0.4\%$, respectively, which are consistent with each other *and* with independent readout fidelity measurements (see Appendix A).

The readout fidelity does not quantify all error in the midcircuit measurement ($F = 97\%$ whereas $\epsilon_\diamond = 8\%$). From $\widehat{Q}$'s two process matrices $\{\widehat{Q}_0, \widehat{Q}_1\}$ (Fig. 1d) we can ascertain and quantify the types of errors that are occurring. As it ideally should, the measurement destroys all coherence between $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. This is because, to within statistical uncertainty, $\widehat{Q}_i[\sigma_x] = \widehat{Q}_i[\sigma_y] = 0$ for both $i = 0$ and $i = 1$ (i.e., only the corner elements of the matrices in Fig. 1d are inconsistent with zero). $\widehat{Q}$ is therefore entirely described by the probabilities $p_{i|j} = \mathrm{Tr}(\widehat{Q}_i[|j\rangle\langle j|])$ and output states $\rho_{i|j} = \widehat{Q}_i[|j\rangle\langle j|]/p_{i|j}$. We find that $p_{0|0} = 99.7 \pm 0.6\%$ and $p_{1|1} = 99.0 \pm 0.6\%$ (these probabilities imply a readout fidelity of $\tilde{F} = 99.3 \pm 0.4\%$, which differs from F above — but it's not inconsistent, as F includes contributions from errors in the state input into the midcircuit measurement, whereas $\tilde{F}$ does not [63]). We find that $\rho_{i|i} = \sigma_I + z_i\sigma_z$ where $z_0 = 0.93$ and $z_1 = -0.86$, implying state fidelities between $\rho_{i|i}$ and the ideal (post-midcircuit-measurement) preparations $|i\rangle\langle i|$ of $96.8 \pm 0.6\%$ and $93.7 \pm 0.8\%$, for $i = 0$ and $i = 1$, respectively. Error in the output quantum state therefore dominates error in the midcircuit measurement. This error is not quantified by readout fidelity, and cannot be measured by detector tomography. The probability for the excited state to decay during the full $3.02\,\mu$s measurement and delay time is $\sim 4.2\%$, so there is an additional source of 2-3% error in the measurement operation, which we conjecture is due to effects beyond the dispersive model [64–69].

C. Modelling non-Markovian errors in midcircuit measurements with short delay times

We now return to consider the source of the non-Markovian errors observed for short post-measurement delay times. A likely source of observed non-Markovian error when $t_d < 1020\,$ns is residual cavity photons, which induce an AC Stark shift the qubit frequency. This causes a $\delta\sigma_z$ Hamiltonian error in all post-measurement gates where δ (1) decays over time, and (2) depends on the midcircuit measurement outcome. In the context of tomography, this is a non-Markovian effect — it cannot be modeled by a single CPTP map per gate. To test whether the Stark shift explains the data, we constructed a model $\widehat{G}_{\mathrm{stark}}$ that is identical to $\widehat{G}(2020\,\mathrm{ns})$ except for added errors that model the Stark shift. We replaced $\widehat{G}_k$ with

$$\widehat{G}_k(\alpha, r, i, m) = \exp\Big[\log\Big(\widehat{G}_k\Big) + \alpha_i(t_d)\exp(-mr_i)\mathcal{Z}\Big], \qquad (5)$$

for $k = x, y$, where $\mathcal{Z}[\rho] = -i\sigma_z\rho + i\rho\sigma_z$ generate σ_z rotations, $m = 0, 1, \ldots$ indexes the number of gates since the midcircuit measurement, i is the midcircuit measurement's outcome, r_i is the Stark shift's decay rate, and $\alpha_i(t_d)$ is the initial phase error for delay time t_d. Both r_i and $\alpha_i(t_d)$ can be fully described by dispersive theory and independent device characterizations. This model explains most of the discrepancy between the QILGST fits and the data for $t_d < 1100\,$ns (Fig. 3, squares). With *zero* fit parameters, at $t_d = 500\,$ns N_σ is decreased by almost an order of magnitude, and the maximum TVD from 80% to 15%. This is strong evidence that the main source of the non-Markovianity is this decaying Stark shift. As this shift adds a coherent Z component to post-measurement gates, this kind of error could be mitigated "for free" via virtual Z rotations [70].

This model does not, however, entirely explain observations at the shortest delay times. This could be due to inaccuracies in the device parameter characterization, or effects beyond dispersive theory [64–69]. To test the first hypothesis, we *fit* the four parameters $\alpha_0(t_d)$, $\alpha_1(t_d)$, r_0, and r_1 to the data at each t_d [71]. This final model is almost consistent with the data (Fig. 3, up-triangles), and its optimized parameter values predict behavior close to that predicted by the independently measured device parameters at long delay times. (See Appendix E for more details on how we incorporate the decaying Stark shift into the QILGST estimate.) This demonstrates how QILGST can be combined with device physics to develop and validate microscopic models of device dynamics, while also suggesting that additional physics is needed to fully describe dispersive measurements on superconducting qubits.

IV. DISCUSSION

Quantum computing experiments that rely on well-calibrated midcircuit measurements are becoming increasingly prominent [17–26]. Techniques like QILGST will be essential for characterizing them. The most striking features of our experimental results are the non-Markovianity of the midcircuit measurement at short delay times, and the large error in the post-measurement state even with the longest post-measurement delay. These effects could not have been discovered and quantified using quantum detector tomography, randomized benchmarking, or readout fidelity measurements, and they suggest that active cavity and qubit reset [53] will be critical for low-error midcircuit measurements on superconducting qubits. As with standard tomographic methods, the number of circuits required for QILGST scales exponentially with the number of qubits. However, QILGST could potentially be combined with recent advances in many-qubit GST [72, 73] to obtain polynomial resource scaling. By enabling complete characterizations of, e.g., many-qubit syndrome extraction cycles, this would provide invaluable insight into experimental QEC.

DATA AND CODE AVAILABILITY

All data presented herein and all analysis code are available online [74].

ACKNOWLEDGMENTS

The authors would like to thank G. E. Rowlands for assistance with experimental software infrastructure, A. Wagner for assistance with device fabrication, B. Hassick for experimental assistance, M. da Silva, J. Gamble, and C. Granade for useful discussions, and W. D. Oliver for providing the JTWPA. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research through the Quantum Testbed Program; the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA); and by the U.S. Army Research Office under Contract No: W911NF-14-C-0048. Sandia National Laboratories is a multi-program laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. Fabrication of the devices presented in this paper was partially conducted at the Harvard Center for Nanoscale Systems, a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. 1541959. All statements of fact, opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views or policies of the U.S. Army Research Office, the U.S. Department of Energy, IARPA, the ODNI, or the U.S. Government. This article does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

Appendix A: Device Parameters

The superconducting transmon device was fabricated by BBN in collaboration with Raytheon RF Components. The device ground plane, resonators and qubit capacitors are 200 nm niobium sputtered on high-resistivity intrinsic silicon, cleaned with an HF-last RCA clean [75] before sputtering. The niobium metallization was optically patterned and etched with an SF_6+O_2 RIE-ICP plasma etch. Post-etch residues were removed using an oxygen ash and a HF etch. The qubits' single Josephson junction was patterned using a Dolan bridge [76] technique using a PMMA-MMA bilayer resist and electron beam lithography. The junction was fabricated using aluminum electron beam evaporation after an Ar^+ ion mill etch to remove surface oxides. The sample was mounted in and wirebonded to a custom copper sample holder, with additional aluminum wirebonds across on-chip resonators to short parasitic resonances. This package was in turn mounted to the cold stage of a dilution refrigerator inside a light-tight, magnetically shielded sample can.

The qubit chip consists of five fixed-frequency transmon qubits, designed to be similar to those described in [77], connected by bus resonators in two pairs of three. A micrograph of the device is shown in Figure 4. For the experiments described in this paper, only one qubit (Q_3) is measured, while the other transmons are detuned by at least 140 MHz (with coupling only through bus resonators) and so have no impact on its operation and can thus be safely ignored. Q_3 is dispersively coupled to a readout resonator through which control drives resonant with the qubit are also applied. A detailed description of the control wiring, electronics and software stack can be found in the Control Electronics section. Relevant device parameters are listed in Table I. In particular, the photon number popula-

Parameter	Symbol	Value	Measurement
Qubit frequency	$\omega_{01}/2\pi$	4.764 18 GHz	Low-power qubit spectroscopy
Qubit anharmonicity	α	−310 MHz	Two-tone qubit spectroscopy
Resonator dressed frequency	$\omega_r/2\pi$	6.734 64 GHz	Low power resonator spectroscopy
Resonator-qubit coupling	$g/2\pi$	62.5 MHz	Calculated [80]
Qubit dispersive shift	$\chi/2\pi$	−0.270 MHz	Resonator spectroscopy with qubit in $\lvert 0\rangle$ and $\lvert 1\rangle$
Resonator photon decay rate	$1/\kappa$	242 ns	Cavity photon number decay [53]

Table I. Device parameters for transmon Q_3.

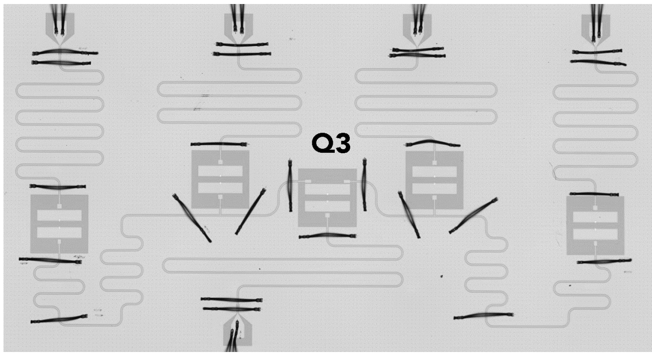

Figure 4. Micrograph showing the device studied in this paper. Q_3 is the central qubit.

tion evolution in the qubit cavity and its relaxation time $1/\kappa$ were measured using the Stark shift [53]. Qubit coherences were measured using standard inversion recovery, Ramsey and Hahn echo sequences, and are listed in Table II. The $X_{\frac{\pi}{2}}$ and $Y_{\frac{\pi}{2}}$ qubit rotation gates were implemented as Gaussian pulses with a 60 ns length. Single-qubit error per Clifford gate was measured using randomized benchmarking [78] and found to be $r = 1.1 \times 10^{-3}$ (Fig. 5a), consistent with the results of GST (Fig. 6). Qubit measurement fidelity, here defined as $F = (P_{0|0} + P_{1|1})/2$, where $P_{1|1}$ is the probability of correctly identifying the qubit state as $\lvert 1\rangle$ when prepared in $\lvert 1\rangle$, was determined from calibration data taken simultaneously with the QILGST sequences. To calibrate the measurement fidelity, we used 1.3×10^5 preparations each of the qubit in its ground and excited states. The reflected cavity signal was downconverted and integrated using a matched kernel filter [52], and binned resulting in the well-separated readout histograms shown in Figure 5b. Integrating and taking the difference of these histograms yields a fidelity $F = 96.35\,\%$, while an approach using logistic regression [79] yields a fidelity $F = 96.43\,\% \pm 0.7\,\%$.

Appendix B: Control electronics

QILGST control sequences are generated in the `pyGSTi` software package then compiled and time-ordered using BBN's Quantum Gate Language (QGL) [81]. QGL ouputs a hardware efficient representation of the experiments which are sent to the control hardware over an ethernet interface. The physical control and readout pulses are sequenced using BBN's custom Arbitrary Pulse Sequencer II (APS-II). The sequencing capabilities of the APS-II allow for continuous playback of the QILGST experiments in a interleaved fashion collecting 1024 shots for each t_d without interruption for waveform or data loading.

Our superconducting device is measured in a Bluefors LD-400 dilution refrigerator. Fig. 9 outlines the complete measurement system. The amplifier pump and qubit control and readout microwave tones are generated using Holzworth9000A microwave synthesizers. To correct for any residual phase instability in the measurement tone, we use an 'autodyne' measurement technique [52]. Control and readout pulses are mixed with the microwave tones using Marki IQ-4509 mixers. Control pulses are generated by BBN custom Arbitrary Pulse Sequencer-II (APS-II) [82] units. The readout and control channels are combined at room temperature, and the qubit cavity is measured in reflection through a Krytar directional coupler at the cold stage. A K&L micro machined 6L250 low-pass filter provides the qubit with protection from high frequency noise above 12 GHz, and a Quinstar QCI cryogenic isolator provides further isolation from the rest of the readout chain. The cavity signal and a pump tone are then combined using a second directional coupler and sent through a Josephson Traveling-Wave Parametric Amplifier (JTWPA). The JTWPA provides roughly 25 dB of gain at the cavity frequency. Additional isolation is provided by a second Krytar QCI isolator and QCY circulator. The readout signal is then amplified at the 4 K stage using an LNF LNC4_8C HEMT amplifier.

T_1 (µs)	T_2^* (µs)	T_2 (µs)
70.2	43.8	82.5

Table II. Transmon average coherence times, measured continuously over 8 h.

Outside the cryostat, microwaves are amplified further using a L3Harris Narda-MITEQ AMF-4F-04001200-15-10P before downconversion to the 13 MHz intermediate frequency with a Marki doubly balanced mixer. A Stanford Research Systems SR445A preamplifier with a voltage gain of 25 is the last stage of amplification before the signal is captured using a X6-1000M Innovative Integration digitizer card running custom firmware [83] which further decimates, digitally downconverts and integrates the data using a matched filter [52]. Data collection and pipelining is orchestrated by a the Auspex software package [84]. All sources, digitizers and sequencers share a

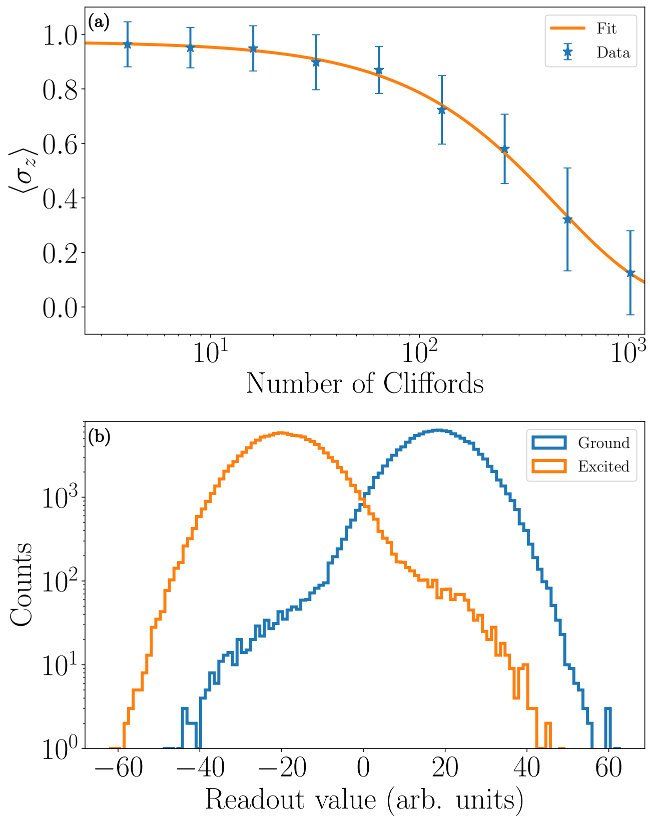

Figure 5. (a) Randomized benchmarking of single-qubit Clifford gates on Q_3. Cliffords are generated from $\{I, X(\pm\pi/2), Y(\pm\pi/2), X(\pi), Y(\pi)\}$ gates with an average of 1.71 gates per Clifford. Points are averages of 32 independent randomized sequences of Clifford gates for each length, while the solid curve is an exponential fit to the data used to extract the error per Clifford $r = 1.1 \times 10^{-3}$. (b) Histogram of measurement results after matched filter integration for 1.3×10^5 ground and excited state preparations, corresponding to $F = 96.4\,\%$.

global 10 MHz clock provided by an SRS SF725 Rubidium frequency standard.

Appendix C: QILGST Simulations

In the main text, we demonstrated that QILGST worked correctly using simulated data. Here we provide the details of these simulations. We simulated single-qubit QILGST on a gate set consisting of $\pi/2$ rotations around the σ_x and σ_y axes, midcircuit and terminating measurements in the σ_z basis, and state preparation in $|0\rangle$. The target QI is $Q_{\text{target}} = \{Q_{\text{target},0}, Q_{\text{target},1}\}$ where

$$Q_{\text{target},k}[\rho] = \text{Tr}\left[\frac{1}{2}\left(\sigma_I + (-1)^k \sigma_z\right)\rho\right]\left(\sigma_I + (-1)^k \sigma_z\right). \quad \text{(C1)}$$

We generated 100 different error models. Errors on the midcircuit measurement were randomly sampled, and errors on all other operations were held constant. The X and Y operations each had over-rotations of 10^{-3} radians along both X and Y axes and were subject to 10^{-2} depolarization, while SPAM was subject to 10^{-3} depolarization. The midcircuit measurement was subject to randomly chosen errors, both on the classical and quantum portions of the channel. With a probability chosen uniformly from 0 to 10^{-2} the $|0\rangle$ was misidentified as $|1\rangle$ (and vice-versa, with another independently chosen probability). Additionally, X and Y coherences were chosen to persist post-midcircuit measurement, both strengths equal but chosen uniformly at random between 0 and 10^{-2}.

For each error model, we simulated drawing N samples from each of the QILGST circuits, with N varying logarithmically from 16 to 1024. For the data with each value of N, we applied the QILGST analysis to obtain an estimate of the gate set. We then computed half the diamond distance between the estimated QI $\widehat{Q}$ and the *true* QI Q_{true} used in the simulation (*not* Q_{target}), as a measure of the estimation inaccuracy. Fig. 2 in the main text shows estimation inaccuracy versus N. It scales as $1/\sqrt{N}$ (standard quantum-limited scaling), indicating that QILGST is correctly reconstructing the QI up to the expected statistical fluctuations.

Appendix D: QILGST Experimental Results

Here we include some additional analysis of the QILGST experimental results. Fig 6 shows the the half diamond distance error ($\epsilon_\diamond$) for each gate for $t_d \geq 1020$ ns (when the datasets are Markovian). We examine the estimates obtained from both QILGST and LGST. (The latter does not incorporate circuits containing any midcircuit measurements, and therefore does not reconstruct an estimate for the quantum instrument.). There is good agreement between the LGST and QILGST reconstructions of the non-QI operations, and the error rates are reasonably stable across the examined delay times.

We do note that qubit relaxation does increase with post-measurement delay time, as expected. However, this effect is only on the order of a couple of percent and is not currently a dominant error source for large post-measurement delay times. The strength of this particular error mechanism is illustrated in Fig. 7, where we consider the fidelity of the post-midcircuit measurement state with respect to the $|1\rangle\langle 1|$ state, post-selected on having prepared a $|1\rangle$ state *and* having received a classical "1" from the midcircuit measurement.

Finally, we also provide the QILGST reconstruction of $\widehat{\mathcal{G}}(2020\,\text{ns})$ in Table III.

Appendix E: Stark Shift Model for Non-Markovian Error

The QILGST fits in the main text show evidence for a considerable amount of non-Markovian error. A possible model to explain this error is the AC-Stark shift of the qubit frequency due to residual photons in the measurement cavity leftover from the midcircuit measurement. The AC-Stark shift will be time-dependent as the photons leak out of the cavity, and thus induce a non-Markovian error on subsequent gates following the midcircuit measurement. A qubit-only Hamiltonian

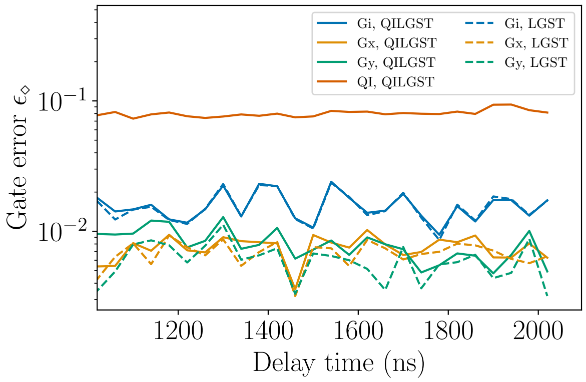

Figure 6. Errors as measured by half diamond distance for QILGST (solid lines) and LGST (dashed lines) reconstructions for $t_d > 1020ns$. G_i denotes the idle operation, G_x the $\pi/2$ X rotation, G_y the $\pi/2$ Y rotation, and *QI* the quantum instrument.

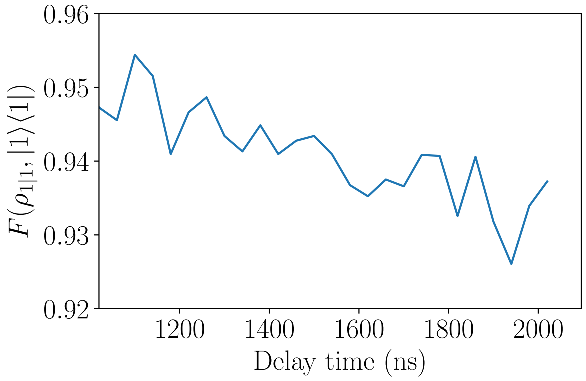

Figure 7. Fidelity of $\rho_{1|1}$ with respect to the ideal state $|1\rangle\langle 1|$ as a function of post-midcircuit measurement delay time. $\rho_{1|1}$ is the state output by the midcircuit measurement, conditioned on the input state $|1\rangle$ being fed into it *and* the classical bit "1" being read out from the midcircuit measurement.

describing this error model is given by

$$H(t) = H_k + \delta(t)\sigma_z, \tag{E1}$$

where H_k for $k \in \{x, y\}$ is the Hamiltonian describing the intended gate, and δ is a parameter describing the Stark-shift. Based on the lowest order dispersive theory, we would expect that $\delta(t) = \chi n(t)$, where $n(t) = \left\langle \hat{a}^\dagger \hat{a} \right\rangle(t) = n_i e^{-\kappa t}$ is the time-dependent expectation value of the cavity photon population, with n_i the initial photon population that depends on the outcome of the midcircuit measurement, labeled by $i \in \{0, 1\}$, as we drive on one of the shifted cavity lines for measurement. All parameters in this model have been measured by independent calibration experiments, see Table I.

The gate generated by this Hamiltonian is given by

$$G_k(t) = \mathcal{T}_{\leftarrow} \exp\left(-i \int_{t_0}^{t_0+t_{\text{gate}}} H(t)dt\right) \approx \tag{E2}$$

$$\exp\left(-i \int_{t_0}^{t_0+t_{\text{gate}}} H(t)dt\right) = \exp\left(-i\left[H_k t_{\text{gate}} + \varphi_{i,m}\sigma_z\right]\right),$$

where the approximation sign is an indication that we have approximated the full time-ordered integral with the first order term of the Magnus expansion. We have verified that the second order term of the Magnus expansion results in a phase error that is at least an order of magnitude smaller than the first order phase error $\varphi_{i,m}$. From the experimental calibration and dispersive theory, the first order phase error is given by

$$\varphi_{i,m}(t_d) = \frac{\chi n_i}{\kappa}\left(1 - e^{-\kappa t_{\text{gate}}}\right) e^{-\kappa\left(m t_{\text{gate}} + t_d\right)}, \tag{E3}$$

where $m \in \{0, 1, 2\}$ labels the gates following the midcircuit measurement.

For the modelling results presented in the main text, we approximate the implemented gate more accurately by replacing H_k with $\log\left(\widehat{G}_k\right)$, the generator [85] of the superoperator representation of the gate $\widehat{G}_k$ characterized by QILGST at 2020ns delay. We model each gate following the midcircuit measurement as

$$\widehat{G}_k(\alpha, r, i, m) = \exp\left[\log\left(\widehat{G}_k\right) + \alpha_i(t_d)\exp(-m r_i)\mathcal{Z}\right], \tag{E4}$$

where

$$\alpha_i(t_d) = \frac{\chi n_i}{\kappa}\left(1 - e^{-\kappa t_{\text{gate}}}\right) e^{-\kappa t_d}, \tag{E5}$$

and $r_i = \kappa t_{\text{gate}}$ under the dispersive model.

In addition to the model given above with $\varphi_{i,m}$ of Eq. (E3) determined entirely by independently characterized parameters, we also fit a model of the form of Eq. (E4) with all $\alpha_i(t_d)$ and r_i as free parameters. For this model fitting, these free parameters are independently fit at each delay time, and for each midcircuit measurement result. The model fit results in a phase error

$$\phi_{i,m}(t_d) = \alpha_i(t_d) e^{-m r_i}. \tag{E6}$$

Fig. 8 shows the calculated value of $\varphi_{i,m}$, and the fit estimate of $\phi_{i,m}$ as a function of delay time for each gate following the midcircuit measurement. The two models agree reasonably well above $t_d \approx 900ns$. We note that a) the effect is much more pronounced (as expected) when 0 is read out than 1, and b) the fit model does not quite follow an exponential decay at low t_d. This latter point indicates that while the Stark shift inspired model of Eq. (E4) with free fit parameters is a good effective model for the data, it does not agree with simple microscopic dispersive theory. This indicates that the discrepancy is likely not due to mis-characterization of the system parameters, but of qualitatively distinct physics arising from effects outside of dispersive theory.

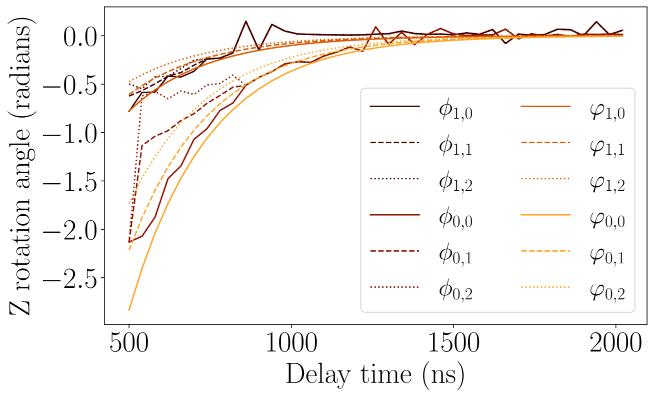

Figure 8. The total amount of σ_z rotation angle induced by the post-measurement Stark shift in each post-measurement gate. $\varphi_{i,m}$ indicates the theoretical prediction for the m^{th} post-measurement gate when i is read out; $\phi_{i,m}$ is the same quantity, but where we numerically optimized the fit parameters of the model (independently at each t_d).

Operation label	$\mathcal{G}_{\text{target}}$	$\widehat{\mathcal{G}}(2020\,\text{ns})$	2σ error bars
$\|\rho\rangle\rangle$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}$	$\begin{pmatrix}1\\-0.016\\-0.008\\0.953\end{pmatrix}$	$\begin{pmatrix}0\\0.008\\0.009\\0.006\end{pmatrix}$
$\langle\langle M\|$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1&0&0&1\end{pmatrix}$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1.002&-0.002&-0.01&0.997\end{pmatrix}$	$\begin{pmatrix}0.002&0.006&0.008&0.003\end{pmatrix}$
G_i	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$	$\begin{pmatrix}1.0&0.0&0.0&0.0\\-0.004&0.993&-0.001&0.021\\0.01&0.008&0.989&-0.008\\0.005&-0.022&0.003&0.99\end{pmatrix}$	$\begin{pmatrix}0.0&0.0&0.0&0.0\\0.008&0.009&0.024&0.027\\0.008&0.024&0.009&0.03\\0.009&0.027&0.03&0.01\end{pmatrix}$
G_x	$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&-1\\0&0&1&0\end{pmatrix}$	$\begin{pmatrix}1.0&0.0&0.0&0.0\\-0.001&0.999&0.003&-0.004\\0.0&-0.004&0.011&-0.999\\0.0&-0.003&0.999&0.011\end{pmatrix}$	$\begin{pmatrix}0.0&0.0&0.0&0.0\\0.004&0.004&0.012&0.012\\0.002&0.012&0.007&0.003\\0.002&0.012&0.003&0.006\end{pmatrix}$
G_y	$\begin{pmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&-1&0&0\end{pmatrix}$	$\begin{pmatrix}1.0&0.0&0.0&0.0\\-0.001&0.005&0.004&0.999\\-0.001&-0.005&0.999&-0.004\\0.001&-0.999&-0.005&0.006\end{pmatrix}$	$\begin{pmatrix}0.0&0.0&0.0&0.0\\0.003&0.006&0.013&0.003\\0.004&0.013&0.003&0.013\\0.003&0.003&0.013&0.006\end{pmatrix}$
Q_0	$\begin{pmatrix}0.5&0&0&0.5\\0&0&0&0\\0&0&0&0\\0.5&0&0&0.5\end{pmatrix}$	$\begin{pmatrix}0.504&0.003&-0.006&0.493\\-0.01&0.002&0.005&-0.014\\-0.007&-0.005&0.002&-0.0\\0.454&0.0&0.005&0.478\end{pmatrix}$	$\begin{pmatrix}0.003&0.011&0.011&0.005\\0.013&0.023&0.023&0.016\\0.013&0.023&0.022&0.016\\0.006&0.014&0.014&0.009\end{pmatrix}$
Q_1	$\begin{pmatrix}0.5&0&0&-0.5\\0&0&0&0\\0&0&0&0\\-0.5&0&0&0.5\end{pmatrix}$	$\begin{pmatrix}0.496&-0.003&0.006&-0.493\\0.004&0.001&0.001&-0.009\\0.009&-0.003&-0.005&-0.009\\-0.418&0.004&0.0&0.448\end{pmatrix}$	$\begin{pmatrix}0.003&0.011&0.011&0.005\\0.013&0.023&0.023&0.015\\0.013&0.023&0.023&0.015\\0.007&0.015&0.016&0.01\end{pmatrix}$

Table III. QILGST-reconstructed estimates for all operations at $t_d = 2020\,\text{ns}$.

[1] D. P. DiVincenzo, The physical implementation of quantum computation, Fortschritte der Physik: Progress of Physics **48**, 771 (2000).

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

[3] E. Knill, Quantum computing with realistically noisy devices,

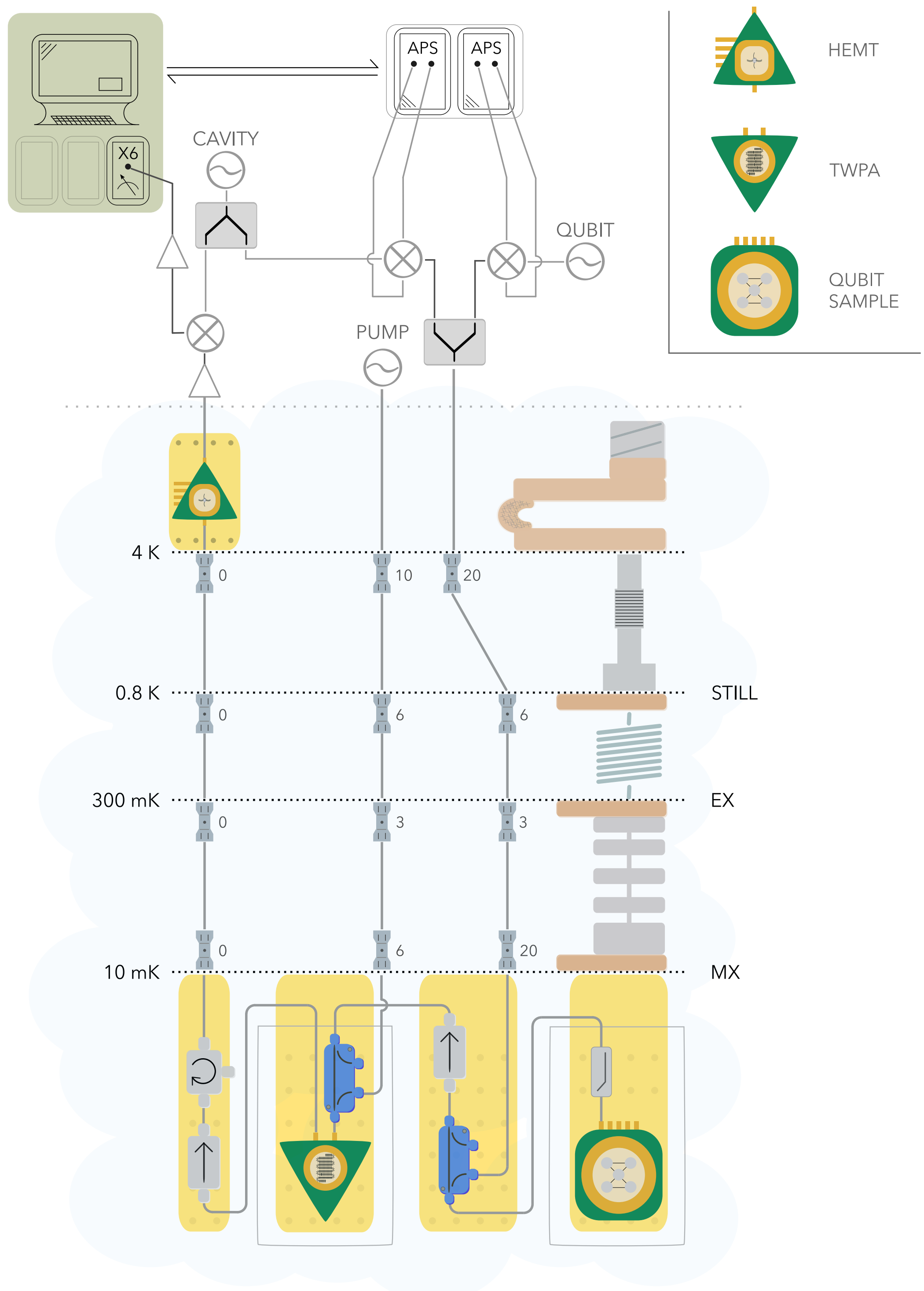

Figure 9. Experimental diagram. Microwave control signals are synthesized at room temperature and mixed up to the qubit and cavity frequencies. These signals are routed into a dilution refrigerator and bounce off the qubit sample. The microwaves then pass through both a JTWPA and HEMT amplifier before being down-mixed and further amplified again at room temperature. Microwave attenuation levels are listed for each temperature stage of the dilution refrigerator. The measurement signal from the cavity is converted to an intermediate frequency using an autodyne technique [52] and digitized using an commercially available digitzer running custom firmware [83].

Nature **434**, 39 (2005).
[4] D. P. DiVincenzo and P. Aliferis, Effective fault-tolerant quantum computation with slow measurements, Phys. Rev. Lett. **98**, 020501 (2007).
[5] A. J. Landahl, J. T. Anderson, and P. R. Rice, Fault-tolerant quantum computing with color codes, arXiv preprint arXiv:1108.5738 (2011).
[6] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).
[7] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Reports on Progress in Physics **76**, 076001 (2013).
[8] D. Ristè, L. C. G. Govia, B. Donovan, S. D. Fallek, W. D. Kalfus, M. Brink, N. T. Bronn, and T. A. Ohki, Real-time processing of stabilizer measurements in a bit-flip code, npj Quantum Information **6**, 71 (2020).
[9] A. D. Corcoles, M. Takita, K. Inoue, S. Lekuch, Z. K. Minev, J. M. Chow, and J. M. Gambetta, Exploiting dynamic quantum circuits in a quantum algorithm with superconducting qubits, arxiv (2021), arXiv:2102.01682 [quant-ph].
[10] M. Urbanek, B. Nachman, and W. A. de Jong, Error detection on quantum computers improving the accuracy of chemical calculations, Phys. Rev. A **102**, 022427 (2020).
[11] A. Holmes, M. R. Jokar, G. Pasandi, Y. Ding, M. Pedram, and F. T. Chong, NISQ+: Boosting quantum computing power by approximating quantum error correction, in *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)* (2020) pp. 556–569.
[12] A. Paetznick and K. M. Svore, Repeat-until-success: Non-deterministic decomposition of single-qubit unitaries, Quantum Info. Comput. **14**, 1277–1301 (2014).
[13] D. Greenbaum, Introduction to quantum gate set tomography, arxiv (2015), arXiv:1509.02921 [quant-ph].
[14] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 14485 (2017).
[15] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate Set Tomography, Quantum **5**, 557 (2021).
[16] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, Probing quantum processor performance with pyGSTi, Quantum Sci. Technol. **5**, 044002 (2020).
[17] G. Q. AI, Exponential suppression of bit or phase errors with cyclic error correction, Nature **595**, 383 (2021).
[18] 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 (2020).
[19] A. Erhard, H. P. Nautrup, M. Meth, L. Postler, R. Stricker, M. Stadler, V. Negnevitsky, M. Ringbauer, P. Schindler, H. J. Briegel, *et al.*, Entangling logical qubits with lattice surgery, Nature **589**, 220 (2021).
[20] V. Negnevitsky, M. Marinelli, K. K. Mehta, H.-Y. Lo, C. Flühmann, and J. P. Home, Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register, Nature **563**, 527 (2018).
[21] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. J. Schoelkopf, Fault-tolerant detection of a quantum error, Science **361**, 266 (2018).
[22] S. S. Elder, C. S. Wang, P. Reinhold, C. T. Hann, K. S. Chou, B. J. Lester, S. Rosenblum, L. Frunzio, L. Jiang, and R. J. Schoelkopf, High-fidelity measurement of qubits encoded in multilevel superconducting circuits, Phys. Rev. X **10**, 011001 (2020).
[23] C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo, Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements, Science Advances **6**, eaay3050 (2020).
[24] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, Scalable extraction of error models from the output of error detection circuits, arXiv preprint arXiv:1405.1454 (2014).
[25] M. Takita, A. D. Córcoles, E. Magesan, B. Abdo, M. Brink, A. Cross, J. M. Chow, and J. M. Gambetta, Demonstration of weight-four parity measurements in the surface code architecture, Phys. Rev. Lett. **117**, 210505 (2016).
[26] D. Willsch, M. Willsch, F. Jin, H. De Raedt, and K. Michielsen, Testing quantum fault tolerance on small systems, Phys. Rev. A **98**, 052348 (2018).
[27] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, Logical randomized benchmarking, arXiv preprint arXiv:1702.03688 (2017).
[28] A. Luis and L. L. Sánchez-Soto, Complete characterization of arbitrary quantum measurement processes, Phys. Rev. Lett. **83**, 3573 (1999).
[29] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Tomography of quantum detectors, Nat. Phys. **5**, 27 (2009).
[30] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, Phys. Rev. A **64**, 024102 (2001).
[31] G. M. D'Ariano, L. Maccone, and P. Lo Presti, Quantum calibration of measurement instrumentation, Phys. Rev. Lett. **93**, 250407 (2004).
[32] S. Izumi, J. S. Neergaard-Nielsen, and U. L. Andersen, Tomography of a feedback measurement with photon detection, Phys. Rev. Lett. **124**, 070502 (2020).
[33] Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, Detector tomography on ibm quantum computers and mitigation of an imperfect measurement, Phys. Rev. A **100**, 052315 (2019).
[34] 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).
[35] E. B. Davies and J. T. Lewis, An operational approach to quantum probability, Commun. Math. Phys. **17**, 239 (1970).
[36] N. Miklin, J. J. Borkała, and M. Pawłowski, Semi-device-independent self-testing of unsharp measurements, Phys. Rev. Research **2**, 033014 (2020).
[37] K. Mohan, A. Tavakoli, and N. Brunner, Sequential random access codes and self-testing of quantum measurement instruments, New J. Phys. **21**, 083034 (2019).
[38] S. Wagner, J.-D. Bancal, N. Sangouard, and P. Sekatski, Device-independent characterization of quantum instruments, Quantum **4**, 243 (2020).
[39] J. Z. Blumoff, K. Chou, C. Shen, M. Reagor, C. Axline, R. T. Brierley, M. P. Silveri, C. Wang, B. Vlastakis, S. E. Nigg, L. Frunzio, M. H. Devoret, L. Jiang, S. M. Girvin, and R. J. Schoelkopf, Implementing and characterizing precise multiqubit measurements, Phys. Rev. X **6**, 031041 (2016).
[40] M. Horodecki, P. W. Shor, and M. B. Ruskai, Entanglement breaking channels, Rev. Math. Phys. **15**, 629 (2003).
[41] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, in *Proceedings of the thirtieth annual ACM symposium on Theory of computing - STOC '98* (ACM Press, New York, New York, USA, 1998) pp. 20–30.
[42] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum process to-

mography of a controlled-not gate, Phys. Rev. Lett. **93**, 080502 (2004).

[43] J. F. Poyatos, J. I. Cirac, and P. Zoller, Complete characterization of a quantum process: The Two-Bit quantum gate, Phys. Rev. Lett. **78**, 390 (1997).

[44] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[45] We only consider a single QI here; extending to multiple QIs in the same gate set is trivial.

[46] Only 92 circuits are needed if idle is omitted. The experiment presented here uses the gates $\{X_{\frac{\pi}{2}}, Y_{\frac{\pi}{2}}, \mathbb{1}\}$; the simulations omit idle.

[47] E. Nielsen, R. J. Blume-Kohout, K. M. Rudinger, T. J. Proctor, L. Saldyt, and Others, *Python GST Implementation (PyGSTi) v. 0.9*, Tech. Rep. (Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2019).

[48] It also allows us to constrain the gates to be CPTP. We constrain the QI to be TP but not CP.

[49] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).

[50] 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, Phys. Rev. A **69**, 062320 (2004).

[51] K. O'Brien, C. Macklin, I. Siddiqi, and X. Zhang, Resonant phase matching of josephson junction traveling wave parametric amplifiers, Phys. Rev. Lett. **113**, 157001 (2014).

[52] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, Phys. Rev. A **91**, 022118 (2015).

[53] D. T. McClure, H. Paik, L. S. Bishop, M. Steffen, J. M. Chow, and J. M. Gambetta, Rapid Driven Reset of a Qubit Readout Resonator, Phys. Rev. Appl. **5**, 011001 (2016), arXiv:1503.01456.

[54] The preparation and measurement fiducials were the same. They were G_x, G_y, G_xG_x, $G_xG_xG_x$, and $G_yG_yG_y$, where G_k denotes a σ_k rotation by $\pi/2$.

[55] The 500 ns minimum time between a measurement pulse and subsequent gate operation is a limitation of our custom digitizer firmware.

[56] Y. Wan, D. Kienzler, S. D. Erickson, K. H. Mayer, T. R. Tan, J. J. Wu, H. M. Vasconcelos, S. Glancy, E. Knill, D. J. Wineland, A. C. Wilson, and D. Leibfried, Quantum gate teleportation between separated qubits in a trapped-ion processor, Science **364**, 875 (2019).

[57] R. Blume-Kohout, K. Rudinger, E. Nielsen, T. Proctor, and K. Young, Wildcard error: Quantifying unmodeled errors in quantum processors, arXiv preprint arXiv:2012.12231 (2020).

[58] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing context-dependent errors in quantum processors, Phys. Rev. X **9**, 021045 (2019).

[59] T. Proctor, M. Revelle, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, Detecting and tracking drift in quantum information processors, Nat. Commun. **11**, 5396 (2020).

[60] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, Noise spectroscopy through dynamical decoupling with a superconducting flux qubit, Nat. Phys. **7**, 565 (2011).

[61] E. Nielsen, K. Rudinger, T. Proctor, K. Young, and R. Blume-Kohout, Efficient flexible characterization of quantum processors with nested error models, arXiv preprint arXiv:2103.02188 (2021).

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

[63] $\tilde{F}$ is also not directly measurable; it is not gauge-invariant.

[64] M. Boissonneault, J. M. Gambetta, and A. Blais, Nonlinear dispersive regime of cavity qed: The dressed dephasing model, Phys. Rev. A **77**, 060305 (2008).

[65] M. Boissonneault, J. M. Gambetta, and A. Blais, Improved superconducting qubit readout by qubit-induced nonlinearities, Phys. Rev. Lett. **105**, 100504 (2010).

[66] L. C. G. Govia and F. K. Wilhelm, Unitary-feedback-improved qubit initialization in the dispersive regime, Phys. Rev. Applied **4**, 054001 (2015).

[67] L. C. G. Govia and F. K. Wilhelm, Entanglement generated by the dispersive interaction: The dressed coherent state, Phys. Rev. A **93**, 012316 (2016).

[68] M. Khezri, E. Mlinar, J. Dressel, and A. N. Korotkov, Measuring a transmon qubit in circuit qed: Dressed squeezed states, Phys. Rev. A **94**, 012347 (2016).

[69] P. Huembeli and S. E. Nigg, Towards a heralded eigenstate-preserving measurement of multi-qubit parity in circuit qed, Phys. Rev. A **96**, 012313 (2017).

[70] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient z gates for quantum computing, Phys. Rev. A **96**, 022330 (2017).

[71] This fitting procedure still uses QILGST formalism and code; it just fits to a different model.

[72] E. Nielsen, *Efficient Scalable Tomography of Many-Qubit Quantum Processors*, Tech. Rep. SAND2020-10518 (Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 2020).

[73] L. C. G. Govia, G. J. Ribeill, D. Ristè, M. Ware, and H. Krovi, Bootstrapping quantum process tomography via a perturbative ansatz, Nature Communications **11**, 1084 (2020).

[74] K. Rudinger, G. Ribeill, L. Govia, M. Ware, E. Nielsen, K. Young, T. Ohki, R. Blume-Kohout, and T. Proctor, Data and code for "Characterizing mid-circuit measurements on a superconducting qubit using gate set tomography" v1.0.2, 10.5281/zenodo.5146074 (2021).

[75] W. Kern and J. E. Soc, The Evolution of Silicon Wafer Cleaning Technology, J. Electrochem. Soc. **137**, 1887 (1990).

[76] G. J. Dolan, Offset masks for lift-off photoprocessing, Appl. Phys. Lett. **31**, 337 (1977).

[77] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, and M. Steffen, Implementing a strand of a scalable fault-tolerant quantum computing fabric, Nat. Commun. **5**, 1 (2014), arXiv:1311.6330.

[78] C. Ryan, M. Laforest, and R. Laflamme, Randomized benchmarking of single-and multi-qubit control in liquid-state nmr quantum information processing, New J. Phys. **11**, 013034 (2009).

[79] Z. H. Ding, J. M. Cui, Y. F. Huang, C. F. Li, T. Tu, and G. C. Guo, Fast High-Fidelity Readout of a Single Trapped-Ion Qubit via Machine-Learning Methods, Phys. Rev. Appl. **12**, 014038 (2019), arXiv:1810.07997.

[80] M. F. Gely, G. A. Steele, and D. Bothner, Nature of the lamb shift in weakly anharmonic atoms: From normal-mode splitting to quantum fluctuations, Phys. Rev. A **98**, 053808 (2018).

[81] B. Johnson and et al., Quantum Gate Language, `https://github.com/BBN-Q/QGL` (2021).

[82] C. A. Ryan, B. R. Johnson, D. Ristè, B. Donovan, and T. A. Ohki, Hardware for dynamic quantum computing, Review of Scientific Instruments **88**, 104703 (2017).

[83] C. Ryan and et al., BBN-QDSP-X6, https://github.com/BBN-Q/BBN-QDSP-X6 (2016).
[84] G. Rowlands and et al., Auspex, https://github.com/BBN-Q/Auspex (2021).
[85] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, A taxonomy of small markovian errors, arXiv preprint arXiv:2103.01928 (2021).