

## Using Cryogenic CMOS Control Electronics to Enable a Two-Qubit Cross-Resonance Gate

Devin Underwood<sup>1,\*</sup>, Joseph A. Glick<sup>1,†</sup>, Ken Inoue<sup>1</sup>, David J. Frank<sup>1</sup>, John Timmerwilke<sup>1</sup>, Emily Pritchett<sup>1</sup>, Sudipto Chakraborty<sup>1</sup>, Kevin Tien<sup>1</sup>, Mark Yeck<sup>1</sup>, John F. Bulzacchelli<sup>1</sup>, Chris Baks<sup>1</sup>, Raphael Robertazzi<sup>1</sup>, Matthew Beck<sup>1</sup>, Rajiv V. Joshi<sup>1</sup>, Dorothy Wisnieff<sup>1</sup>, Scott Lekuch<sup>1</sup>, Brian P. Gaucher<sup>1</sup>, Daniel J. Friedman<sup>1</sup>, Pat Rosno<sup>2</sup>, Daniel Ramirez<sup>2</sup> and Jeff Ruedinger<sup>2</sup>

<sup>1</sup>IBM Quantum, T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York 10598, USA

<sup>2</sup>IBM Quantum, 2800 37th Street NW, Rochester, Minnesota 55901, USA



(Received 20 August 2023; accepted 19 January 2024; published 14 February 2024)

Qubit control electronics composed of CMOS circuits are of critical interest for next-generation quantum computing systems. A CMOS-based application-specific integrated circuit (ASIC) fabricated in 14-nm fin field-effect transistor (FinFET) technology was used to generate and sequence qubit control wave forms and demonstrate a two-qubit cross-resonance gate between fixed-frequency transmons. The controller was thermally anchored to the  $T = 4$  K stage of a dilution refrigerator and the measured power was 23 mW per qubit under active control. The chip generated single-side banded output frequencies between 4.5 and 5.5 GHz, with a maximum power output of  $-18$  dBm. Randomized-benchmarking (RB) experiments revealed an average number of 1.71 instructions per Clifford (IPC) for single-qubit gates and 17.51 IPC for two-qubit gates. A single-qubit error per gate of  $\epsilon_{1Q} = 8 \times 10^{-4}$  and a two-qubit error per gate of  $\epsilon_{2Q} = 1.4 \times 10^{-2}$  were shown. A drive-induced  $Z$  rotation was observed by way of a rotary-echo experiment; this observation is consistent with the expected qubit behavior given the measured excess local-oscillator (LO) leakage from the CMOS chip. The effect of spurious drive-induced  $Z$  errors was numerically evaluated with a two-qubit model Hamiltonian and shown to be in good agreement with the measured RB data. The modeling results suggest that the  $Z$  error varies linearly with the pulse amplitude.

DOI: [10.1103/PRXQuantum.5.010326](https://doi.org/10.1103/PRXQuantum.5.010326)

### I. INTRODUCTION

Next-generation quantum computers will undergo a paradigm shift whereupon multiqubit devices will predominantly perform fault-tolerant quantum circuits. This new era of quantum computing will require orders of magnitude more qubits than are currently being integrated in today's systems [1]. For large quantum computing systems comprised of solid-state quantum processors (superconducting qubits or quantum dots), cryogenic control electronics is considered a key enabling technology [2–4]. There has been significant development in CMOS electronics for quantum dot processors [5–8], primarily due to the increased input-output demands and the potential for

integrating CMOS electronics with qubits at temperatures  $> 100$  mK. More recently, cryogenic CMOS electronics for superconducting qubits have been developed and have been used for single-qubit-gate demonstrations [9–11].

Research in cryogenic CMOS control electronics has primarily focused on achieving the low-power analog requirements to operate within the thermal-load limitations of a dilution refrigerator (DR). While power dissipation is an important specification for cryogenic control technologies, minimalistic controllers are not sufficient for practical qubit control. For a fully cryogenic control architecture to be useful, a classical processor capable of producing relevant pulse sequences will be required. This classical processor presents an additional source of power dissipation and specialized digital architectures will be needed for cryogenic integration to achieve the required performance within the limited available power budget.

Another important consideration for cryogenic control technologies is the maximum thermal load of the different stages in a dilution refrigerator. For thermalization of the CMOS chip at the  $T = 4$  K plate, the power limitation of a cryogen-free DR is set by the second stage of a cryocooler

\*devin.underwood@ibm.com

†joseph.a.glick@ibm.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

and the maximum thermal load will be determined by the number of cryocoolers in the DR. Notably, the maximum load of cryocoolers is temperature dependent; they yield more cooling power at higher temperatures [12]. If the second stage were allowed to operate at higher temperatures, then the DR would be able to support more cryogenic control channels. Assuming that higher operating temperatures of the second stage do not impact the cooling of other temperature stages, then this strategy will make the integration of cryogenic control technologies more feasible.

Cryocontrolled architectures such as the one shown in Fig. 1(a) may also yield systematic scaling advantages that would warrant using low-power CMOS ASICs. Examples of such advantages include lower communication latency [14], a lower thermal noise floor [15], reduced dispersion of base-band signals [16], reduction in rf loss due to signal delivery [17], wire-count reduction from room temperature

to the  $T = 4$  K stage, a reduction in power per channel, a reduction in cost per channel, and a reduction in total system size.

In this paper, we present measurement results on a transmon-based quantum processor (QP) [18,19] controlled via a custom cryo-CMOS application-specific integrated circuit (ASIC). The ASIC was designed for cryogenic operation and is capable of generating pulse sequences for qubit characterization, verification, and validation (QCVV) experiments common for use with transmon qubits. Here, the CMOS chip is a dual-channel semiautonomous qubit state controller fabricated in 14-nm fin field-effect transistor (FinFET) technology [10,11]. A single-channel block diagram is shown in Fig. 1(b). The on-chip processor facilitates autonomy through its ability to play predefined sequences of qubit control wave forms, a necessary requirement for both quantum error-correction

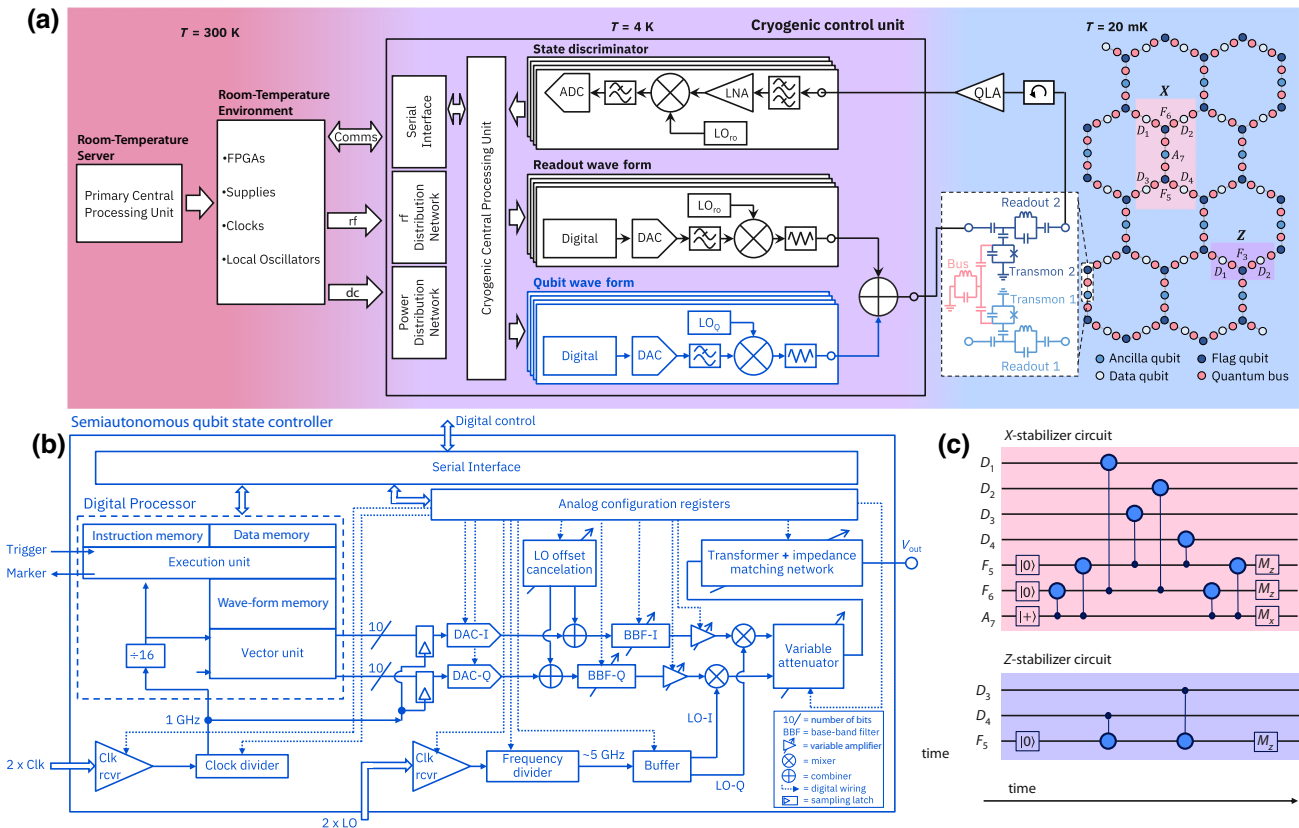


FIG. 1. (a) A block diagram for a fault-tolerant quantum computing architecture in which control signals for the quantum processor (QP) are digitally synthesized at the 4-K stage of the dilution refrigerator. Here, the QP is composed of fixed-frequency transmon qubits coupled together through a fixed-frequency quantum bus and arranged in a heavy hexagonal lattice. The cryogenic control unit (CCU) is composed of a cryogenic central processing unit (CCPU), qubit wave-form generators, readout wave-form generators, and quantum state discriminators. A CCU containing these key elements would be capable of autonomous operation, yielding a shorter-latency loop when running deterministic quantum circuits. A room-temperature processor performs classical computations, orchestrates high-level quantum operations, and interprets the results of quantum algorithms [1]. Room-temperature support electronics are necessary to power, clock, and program active cryogenic electronics. The support electronics interface with a room-temperature server that performs classical computations necessary to run quantum algorithms. (b) An expanded block diagram of the qubit wave-form generator [blue box in (a)], used for this paper. (c) The  $X$  and  $Z$  stabilizer circuits for performing error-correcting protocols on a heavy hexagonal lattice. Stabilizers represent the primary protocol for logical qubit maintenance and the CCU oversees these protocols, which include monitoring physical qubits, decoding errors on physical qubits [13], and generating conditional sequences of pulses.

(QEC) [20–23] and quantum error-mitigation (QEM) workloads [24–26]. In the experiments described in this work, emphasis was placed on understanding the demands of the classical processor (CP), which represents a near-term development challenge for cryogenic control technologies. For example, present-day QPs are primarily used for physics-learning and qubit-characterization experiments [27–29], which can be difficult to support in a low-power processor.

An important qubit-gate-characterization experiment that presents challenges to limited-memory control hardware is randomized benchmarking (RB) [30]. For RB experiments, a random sequence of Clifford gates followed by an inversion pulse are required to be stored in memory. Individual Cliffords correspond to wave forms with independent amplitudes, phases, and durations, while the sequence of Cliffords corresponds to a set of instructions. For complex experiments such as RB, compressing large instruction sets into limited memory is challenging; however, not all experiments are as demanding. In a quantum computing system that primarily performs QEC protocols [Figs. 1(a) and 1(c)], the set of required pulse sequences is simple and repetitive, especially when compared to those needed for QCVV experiments [31]. Understanding the complexity and characteristics of the pulse sequences demanded by these experiments is important for developing optimal qubit control technologies and could lead to special-purpose instruction set architectures (ISAs).

Here, we report the use of cryo-CMOS to generate qubit control wave forms for a suite of characterization experiments including  $T_1$ ,  $T_2$ ,  $T_2^*$ , Carr-Purcell-Meiboom-Gill sequences (CPMG), rotary echo, Hamiltonian tomography, and RB of single-qubit and two-qubit gates. The classical processor was characterized during qubit measurements, highlighting the efficiency of the ISA. Transmon calibration routines were performed in order to realize the aforementioned characterization experiments. These measurements serve as a promising demonstration of cryo-CMOS based control technology, with single-qubit and two-qubit error per gate (EPG) observed to be  $\epsilon_{1Q} = 8 \times 10^{-4}$  and  $\epsilon_{2Q} = 1.4 \times 10^{-2}$ , respectively. We show through Lindblad-master-equation simulations that the observed error is set by control noise. The primary error source is a pulse-induced qubit  $Z$ -axis rotation that arises due to spurious spectral content observed in the CMOS controller output. Additionally, a Gaussian-distributed pulse-amplitude noise was observed on the rf pulses but simulations showed that this noise did not significantly impact gate errors.

## II. CRYO-CMOS CONTROL ELECTRONICS

An ideal quantum control architecture will have the capability to autonomously generate wave-form sequences conditioned on the measurement of physical qubits [32]. A

control unit able to satisfy this requirement will be composed of circuit blocks for qubit wave-form generation, entanglement wave-form generation, readout wave-form generation, and qubit state discrimination. Furthermore, the unit will require a central processor that manages these blocks and conditionally asserts logic for running the appropriate quantum circuits [Fig. 1(a)]. CMOS is an ideal technology for realizing such a control unit because of existing industrial fabrication capabilities and ease of integration of the different circuit blocks required.

In a quantum computing system optimized to execute specific quantum algorithms, the fault-tolerant operations are in principle deterministic, based on the decoding of physical qubit measurements [13,33,34], implying that full autonomy is achievable. If the above-mentioned capabilities are performed autonomously and within the dilution refrigerator, this approach will yield a reduced-latency control configuration [Fig. 1(a)], thereby increasing the achievable number of circuit layer operations per second (CLOPS) [35]. The primary latency concerns addressed here are a reduction in the round-trip transient time and the response time for conditional wave-form generation.

The proposed autonomous control unit is composed of distinguishable circuit blocks, which can be developed either as stand-alone chips in a multichip configuration or as a single larger integrated circuit. The work detailed in this paper focuses on a demonstration with a stand-alone circuit block consisting of two distinct rf channels for qubit wave-form generation. Here, the same rf generator is used for both single-qubit control and generating entanglement between qubit pairs; this approach leverages the inherent wiring advantage associated with the cross-resonance-based architecture [19]. As shown in Fig. 1(b), this semiautonomous qubit state controller consists of an analog block with a low-power digital-to-analog converter (DAC), a special-purpose classical processor, and a serial interface for communicating with room-temperature electronics [10,11].

### A. Analog block

The qubit state controller was designed and fabricated in 14-nm FinFET technology, a technology choice that was desirable due to its high switching efficiency, large transistor on-off ratio, and lower threshold voltages that lead to reduced power dissipation [36]. The choice of a highly scaled transistor technology also reduces the cost of adding a high degree of digital programmability. For example, the qubit controller is capable of being configured to the following modes of conversion operation: double sideband with suppressed carrier (DSB SC), single-sideband direct-conversion lower sideband (SSB LSB), and single-sideband direct-conversion upper sideband (SSB USB). The reported experiments utilized SSB LSB as a mode of operation. In this configuration, the in-line low-pass filters

provided additional attenuation of the LO when it was placed higher in frequency than the transmon qubits.

For experimental versatility, the analog control block of the qubit state controller was made configurable, utilizing over 200 bits of digital control. The analog control block was composed of two 10-bit DACs (in-phase and quadrature, or IQ), two base-band filters (IQ), a complex mixer to provide, e.g., SSB-LSB output, and a tunable output stage. The complex mixer received quadrature clock signals for up-conversion of the base-band signal of the DAC to the  $|0\rangle$ -to- $|1\rangle$  transition frequency  $\omega_{01}$  of the qubit. The SSB mode was chosen as the primary mode of operation in order to reduce circuit complexity, while also reducing analog power consumption. In order to maximize dynamic range while simultaneously minimizing noise generation, a fully differential-current-mode design was implemented. Notable advantages of the design include current reuse among multiple functional blocks, high-bandwidth interfaces between circuit elements, convenient implementation of the variable gain stages (using current scaling and current steering [37]), and low switching noise at the output [10,11].

The differential wiring extended from the DAC output to the output stage of the chip. The output stage consisted of a balun, which converted the differential signal to a single-ended one; the balun resonance frequency was tunable to support the range of desired SSB frequencies. The balun resonance and shape tuning were controlled using 4 bits of center frequency adjustment and 2 bits of quality-factor adjustment. The output impedance was adjustable to match the fridge wiring, which was connected to the balun output. A variable attenuator provided 20 dB of programmable attenuation for noise reduction, plus an additional 25 dB to be switched in for blanking the AWG during readout. To satisfy dynamic range requirements, two variable gain stages were used in the analog control path. The first gain stage was placed between the base-band filter (BBF) and the SSB up-converter, while the second gain stage was placed at the output of the SSB up-converter. Both gain stages were unidirectional (to provide reverse isolation) and yielded a total of 34 dB of gain control, with an average step size close to 2 dB. The BBF bandwidth (as reflected in the 3 dB cutoff frequency) was configurable over a range of 100–800 MHz using a 5-bit bandwidth-control configuration.

The DAC could be programmed to produce an IF offset within approximately 400 MHz of the LO and the in-band spurious tones at the output were suppressed to a spurious-free dynamic range of 40 dB out to 500 MHz [10,11]. A microwave source with a frequency between 8 and 12 GHz was delivered from room temperature to the cryo-CMOS chip, on which LO signals in a range of 4–6 GHz were generated with a 2:1 frequency divider. Leveraging the differential-current-mode architecture of the analog control path, programmable dc currents were added to the output

currents of the DACs in order to compensate differential offsets, which helped to reduce LO leakage in the rf output.

## B. Digital block

The digital architecture featured a processor implementing 32 bit fixed-point instructions for programming flexibility, including special instructions for wave-form generation and phase rotations and was designed to minimize power consumption for cryogenic operation. The ISA of the processor ISA defined 32 general-purpose instructions (eight branch or flow control, ten data movement, and 14 arithmetic) to enable trigger-controlled loops, subroutines, and computation, as well as five special instructions for the generation of wave forms and digital output signals. The processor core used three SRAM banks, with 32 kbyte dedicated for instructions, 20 kbyte dedicated for wave forms, and 32 kbyte dedicated for data, respectively. To minimize power consumption, the processor had a fast clock domain operating at the sampling frequency  $f_s$  of the DACs for providing wave-form data to the DACs and a slow clock domain operating at  $f_{\text{CLK}} = f_s/16$  for program control. The microarchitecture implemented fetch, decode, branch resolution, and scalar arithmetic execution instructions within one slow clock cycle.

Wave-form data were stored as an envelope modulated by intermediate frequency IQ sinusoids with an initial phase of zero. This approach reduced the wave-form memory footprint and avoided the power overhead associated with the sine or cosine evaluations of a numerically controlled oscillator (NCO) [38,39]. The compute waveform coefficients (CWC) instruction prepared the IQ coefficients used by the play waveform (PW) instruction to set the phase and amplitude of the output wave form. These coefficients were calculated relative to the frame phase, which could be modified by special instructions such as add frame phase (ADDFP) to effect a virtual Z rotation of the qubit phase [40]. The coefficients were applied to the stored wave-form data through 16-way single-instruction multidata vector arithmetic logic in the slow clock domain. One PW instruction could play up to 4096 wave-form samples. The samples were serialized into the fast clock domain and sent to the IQ DACs in the analog section. The wave-form retrieval and processing functions progressed independently of the program flow and control facilities of the processor.

TABLE I. The cryo-CMOS parameters used during the qubit control experiments.

Channel	LO (GHz)	IF (MHz)	Digital clock (GHz)
CH <sub>CTRL</sub>	5.6	261.77	2.25
CH <sub>TRGT</sub>	5.6	348.81	2.25



TABLE II. The average qubit parameters measured using the cryo-CMOS chip. The reported coherence measurements were interleaved between 2Q RB measurements shown in Fig. 8(f).

Qubit	$T_1$ ( $\mu$ s)	$T_2$ ( $\mu$ s)	$T_2^*$ ( $\mu$ s)	$ZZ$ (kHz)
$Q_{\text{CTRL}}$	$57.57 \pm 4.46$	$68.99 \pm 2.04$	$23.13 \pm 1.91$	$103.0 \pm 67$
$Q_{\text{TRGT}}$	$61.58 \pm 5.99$	$69.16 \pm 4.80$	$24.21 \pm 3.22$	$103.0 \pm 67$

### III. CMOS PROCESSOR FOR QUBIT CONTROL

Even with a specialized ISA, some experiments were challenging to accommodate with this low-power processor, primarily due to its limited memory. Note that all experiments performed with the cryo-CMOS processor were originally developed using room-temperature electronics that offer significantly more memory and higher performance as compared to the custom processor design. This discrepancy created issues for the porting of experiments to the low-power processor. Details regarding issues encountered are illustrated in Fig. 2 and are further discussed in Sec. IV. Here, Fig. 2 shows the memory usage of the processor for the different experiments performed. Many experiments required no change from routines developed with room-temperature electronics but in some cases additional effort was required in order to fit pulse sequences into processor memory. The default approach to reducing memory demands was to decrease the point density by customizing parameters, while ensuring that enough data points were collected to extract accurate fit results. In cases when the default approach was not sufficient, it was necessary to rewrite experiment routines or substitute new pulse types. One experiment for which new pulse definitions were required was Hamiltonian tomography [41] and details for how this experiment was made to work are reviewed in Sec. IV.

With respect to how they were operated using room-temperature control electronics, most experiments fell into the category of not requiring change or only needing custom parameters to achieve successful execution. RB is an example of an experiment made to work through parameter adjustment: in this case, the instruction memory was the limitation to be overcome. The strategy used to address this challenge was first to reduce the number of Clifford sequences and then to use logarithmic spacing between the different sequence lengths, both of which helped to reduce the number of instructions required. However, as shown in Figs. 6(a) and 7(a), the last data point in an RB experiment consumes the most memory, implying that simple point-reduction methods will not scale as error rates improve. Lower error rates will require more Clifford gates for the exponential decay to converge, which will require more instruction memory. As shown in Fig. 2, the instruction-memory requirements for an RB experiment are predicted to increase from approximately 32 kbyte to approximately

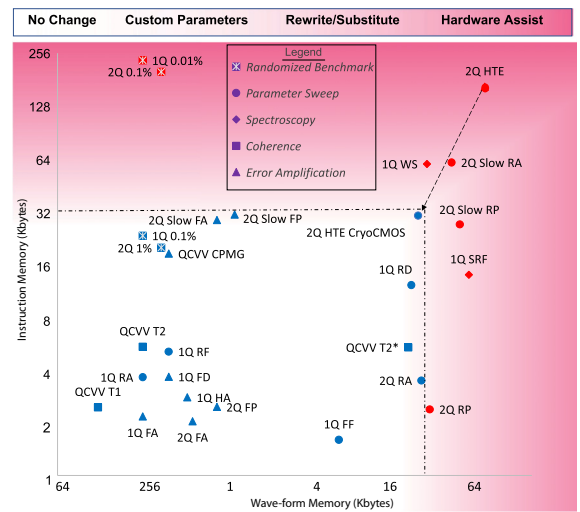


FIG. 2. The instruction and wave-form memory requirements for each of the calibration sequences, QCVV experiments, and RB experiments, using nominal pulse widths of 42.67 ns for single-qubit gates, 71.1 ns for the shortest cross-resonance (CR) pulse width, and 711.1 ns for longest CR pulse width (marked “Slow” in the data labels). The unit on the  $X$ - and  $Y$ -axes is the memory size in bytes, in logarithmic scale. The cryo-CMOS memory limits of 32 kbyte for instructions and 20 kbyte for wave forms are denoted by the dotted square. “1Q” in the data labels identifies single-qubit calibrations and RB experiments, “2Q” or “2Q Slow” identifies two-qubit calibrations and RB experiments, and “QCVV” identifies the characterization experiments ( $T_1$ ,  $T_2$ ,  $T_2^*$ , and CPMG). For calibrations, the last string identifies the type of calibration, encoded as follows: R, rough; F, fine; A, amplitude; F, frequency; D, drag; P, phase. For RBs, the last string specifies the errors per Clifford and the corresponding marker identifies the memory size needed to extract an error rate from an exponential decay converging to  $P(1) = 0.49$ . WS, wideband spectroscopy; SRF, super-rough frequency calibration; HTE, Hamiltonian tomography with echoed CR. Vast differences in the memory requirements are evident for these different experiments. For instance, CR amplitude, CR phase, spectroscopy, and HTE require large wave-form memory, while RB, CPMG, spectroscopy, and HTE require very large instruction memory. The colors indicate the degree of difficulty to accommodate the cryo-CMOS memory limitations. Experiments that fall in the “No Change” zone (white) run without any modifications. “Parameter Change” requires reducing and/or carefully recalculating the parameters of the experiment, such as reducing the number of steps and/or reducing the sweep range. “Rewrite/Substitute” requires rewriting the experiment itself and/or substituting pulse types, to devise an equivalent experiment that fits better in memory by exploiting cryo-CMOS core features. 2Q HTE is in this category. “Hardware Assist” (purple) identifies experiments that may require changes to the current cryo-CMOS hardware and/or instruction set. 1Q and 2Q RB with lower error rates fall in this category, requiring further innovation going forward.

256 kbyte when the 1Q error and 2Q error are reduced from 0.1% to 0.01% and 1% to 0.1%, respectively.

To accommodate longer sequences, ASICs may require more effective instruction memory, a more specialized

ISA, or a custom compiler designed to specifically for the ISA [42]. Alternatively, new measurement techniques with lower memory overhead could also be adapted to extract error rates; e.g., using simple characterization measurements along with error models to infer errors [43–46], interpreting the error through an under-sampled RB experiment [47], or performing experiments such as quantum process tomography [48,49] that place lesser demands on memory. It is also worth noting that longer sequences may not be needed in future quantum computers. Experiments such as RB are useful for performance evaluation during hardware and software development but are not the intended use case for quantum computers. Experiments such as QEC [20–23] and QEM [24–26] will be the primary use cases for useful applications of quantum computers.

#### IV. MEMORY REDUCTION IN A LOW-POWER PROCESSOR

The on-chip processor supported 32 kbyte of SRAM for instructions and 20 kbyte of SRAM for wave forms. For simplicity, these memory banks had a single-purpose designation: wave-form memory was used store wave-form data and instruction memory was used to store programs that played sequences of qubit control pulses. Resource bottlenecks were observed when the wave forms became too long, and/or sequences became too long. When this happened, memory-reduction techniques were necessary. One example technique was to reduce wave-form memory by partitioning wave forms.

A key example is the Hamiltonian-tomography experiment from Fig. 5(d), which required many long square-Gaussian wave forms of different lengths that did not fit in wave-form memory. To overcome this problem, the flat section of the square-Gaussian wave form was partitioned into small equal-sized segments and the wave form was constructed by stepping through instruction memory to add segments together, as shown in Fig. 3. This technique made a fundamental physics experiment feasible with limited wave-form memory but at the cost of more instruction memory. The width of the segment was used as a sweep parameter in order to optimize the trade space between wave-form size and the number of instructions.

This wave-form memory-reduction technique was also applied to other 2Q calibrations for consistency. In these cases, the width of the smallest segment was chosen to align with the fastest 2Q cross-resonance (CR) pulse. To increase the pulse width, the number of instructions increased linearly with the number of segments, as was observed in Fig. 2 for 2Q calibrations marked with the “Slow” prefix. In this specific case, the smallest segment was chosen to be 71.1 ns and the slowest CR pulse used

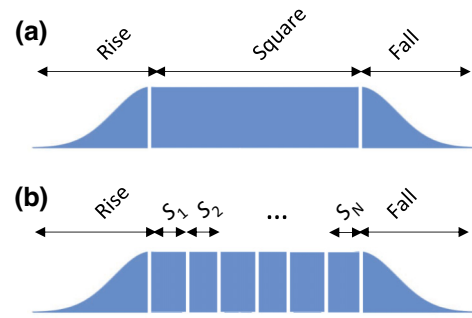


FIG. 3. (a) The standard wave form used for cross-resonance Hamiltonian-tomography experiments. (b) An illustration of the wave-form partitioning used for memory reduction of pulses that were too long to be stored in wave-form memory.

was 711.1 ns; resulting in an approximately  $10\times$  increase in instruction-memory usage. It is conceivable that the arithmetic and branch facilities of the processor could be exploited to reduce the  $10\times$  increase but this is not trivial because of phase-alignment requirements between adjacent segments. As mentioned in Sec. II B, the processor supports phase-manipulation instructions, so in theory this can be overcome, but in practice it is difficult to implement correctly and accurately.

Another challenge associated with this simplistic ISA model is resource starvation, because there are often no free clock cycles to allow execution of arithmetic or branch instructions. For example, in these experiments the shortest 1Q gate instrumented in Fig. 8(e) is just 28.44 ns wide, which corresponds to two processor clock cycles. In order to generate the shortest pulse, two instructions are required: one to compute the wave-form coefficients (CWCs) and another to play a wave form (PW). Shorter gates are desirable because they are shown to reduce errors but a dependency exists between the processor clock frequency and the 1Q gate length. The processor clock frequency sets a lower bound on the 1Q gate speed. Increasing the processor clock frequency would speed up the execution of the CWC and PW instructions but would result in more power dissipation.

The difficulty in optimizing instruction sequences on the ISA was compounded by the layered structure of the existing software stack, in which high-level pulse definitions were generated in a manner that supported a variety of room-temperature control hardware. Since off-the-shelf AWGs do not provide quantum specific instructions, such as frame phase adjustments or parametric looping; the existing pulse-generation software did not support the efficiency yielded by a quantum specific processor. Ideally, a special-purpose compiler would be used to take in high-level pulse definitions and convert them into instructions for special-purpose processors such as the one presented in this paper [42].

## V. EXPERIMENTAL SETUP

The experimental setup in a closed-cycle dilution refrigerator, detailed in Fig. 4(a), consisted primarily of a cryo-CMOS AWG payload (labeled CP) mounted to the 4-K plate and a qubit payload (labeled QP) mounted to the mixing-chamber (MXC) plate at 10 mK. The CP was composed of an Au-plated Cu-machined mount that mechanically and thermally anchored a printed circuit board (PCB). The PCB housed the cryo-CMOS AWG chip, decoupling

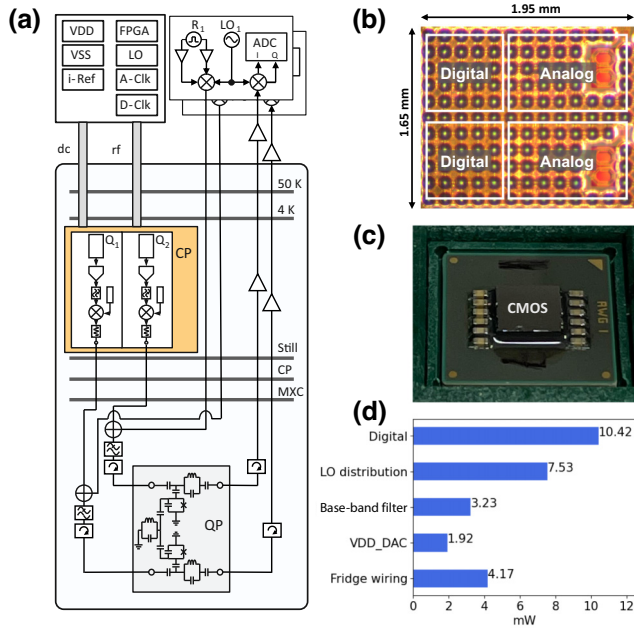


FIG. 4. (a) A block diagram of the experimental setup. The cryo-CMOS payload (CP) is mounted to the  $T = 4$  K plate of a dilution refrigerator and drives control signals down to a pair of transmons in the qubit payload (QP) on mixing chamber (MXC) plate at 10 mK to perform single-qubit and two-qubit cross-resonance gates. Between the CP and the QP, there is 22 dB of cold attenuation, a Mini Circuits VLF 5500 low-pass filter, a ferrite isolator, and a directional coupler that combines the control and the readout signals together. The dc bias supplies, reference currents, clock frequencies, local oscillator (LO), readout electronics, and the field-programmable gate array (FPGA) are all located at room temperature. (b) A micrograph of the CMOS chip, highlighting the digital and analog sections of the two AWG channels. (c) The CMOS chip, bonded to a laminate with NPO ceramic decoupling caps, that sits inside in a pogo-pin socket. The chip is thermally anchored to  $T = 4$  K through a Cu backing plate on the lid of the socket. (d) The power dissipation of the cryo-CMOS chip measured while under active control, for each subcomponent of the chip, and the passive heat load due to wiring from 50 K to 4 K. Including wiring, the total power dissipation per control channel is 27.27 mW. The reported powers were extracted from the on-chip supply voltage measured by a sense line and the current sourced by the power supply. The power dissipation due to fridge wiring was calculated using cryogenic material models [50].

capacitors, and routed traces that connected the various pins of the chip to the wiring connectors. The cryo-CMOS AWG chip, shown in Fig. 4(b), was bump bonded to a laminate with NPO ceramic decoupling capacitors and was inserted into a pogo-pin socket on the PCB, shown in Fig. 4(c). The cryo-CMOS chip was thermally anchored to the 4 K stage via a Cu block on the lid of the socket and a Cu strap that was connected directly to the Cu back plane of the mount. The dc power supplies, reference currents, clock and local-oscillator (LO) sources, and a field-programmable gate array (FPGA), all used to power and control the chip, were located at room temperature near the cryostat. The output of the two cryo-CMOS AWG channels, set to the resonant frequencies of the qubits (approximately 5 GHz), was sent via coaxial cables down to the MXC plate. At the MXC plate, the signals passed through a total of  $-22$  dB of cold attenuation, a Mini Circuits VLF 5500 low-pass filter, and a ferrite cryogenic isolator before connecting to the qubits.

The QP consisted of a pair of transmon qubits that were connected by an  $LC$  cancellation bus that helped to reduce the effect of constant  $ZZ$ -type errors. The transmons were operated and measured in transmission, each with their own designated Purcell filters and readout resonators. The readout pulses, supplied by room-temperature mixer-based signal generators, were set to the frequency of the respective qubit readout resonators (approximately 7 GHz) and driven into the fridge. The readout control signals were combined at the MXC plate with the cryo-CMOS AWG signals into a single line using a directional coupler. After passing through the QP, the output signals from each of the two qubits were sent through another cryogenic isolator, a HEMT amplifier, and a room-temperature amplifier, before being sent to an ADC to be digitized. The maximum power level of the control signals from the cryo-CMOS AWG that could be delivered to the qubits (after passing through the  $-22$  dB of cold attenuation between them) was approximately  $-40$  dBm. The net gain in the readout path was approximately 16 dB, defined to be the total amplification of the HEMT and room-temperature amplifiers, minus the losses associated with the coaxial cables. The readout chain did not include use of a quantum limited amplifier (such as a Josephson-based traveling-wave parametric amplifier) but if desired one could be added to improve the signal-to-noise ratio (SNR) of the output signal. Even without the use of such amplifiers, state-discrimination measurements showed that fidelities of approximately 97% could be achieved.

Proper synchronization, timing, and triggering between the cryo-CMOS AWGs and the readout control electronics is critical to performing all of the qubit calibrations and measurements. All the supplies (clock, LO, and those for the readout electronics) were connected to a common 10-MHz clock reference. When an experiment was to be performed, the wave-form data and the instruction

sequences were first loaded onto the cryo-CMOS AWG via the FPGA and the serial communication interface; then, a signal was relayed back to the FPGA to confirm that the program was successfully loaded into memory. At the beginning of each cryo-CMOS AWG program, there was a WAIT instruction. A trigger signal, marking the start of the experiment, was sent from the readout electronics at room temperature to the cryo-CMOS AWG, satisfying the WAIT condition and commencing the programmed pulse sequence. At the end of a sequence, when a readout pulse was being played, the cryo-CMOS AWG used its on-chip programmable attenuation to blank the output of the AWG signal by approximately 45 dB. By adjusting the pulse timing within the sequence itself and by using a buffer delay on the room-temperature readout electronics, it was ensured that the readout pulses began soon after the control pulses ended, approximately 10 ns after the blanker feature on the cryo-CMOS AWG was engaged. The coordination and timing between the control pulses, the blanking window, and the readout pulses was confirmed using a high-speed oscilloscope. If necessary for a given experiment, the cryo-CMOS AWGs could output their own marker pulses to trigger one another or to indicate that a sequence was complete. At the end of a sequence, the processor then looped back to the beginning of the program and waited for another trigger. Through this procedure, the cryo-CMOS AWG could initialize the qubits into any desired Clifford state, kick off a pulse sequence—such as for calibration, QEC, or to perform a specific qubit experiment—and ensure that at the end, they were followed up by well-aligned readout pulses to complete the measurement.

## VI. QUBIT CALIBRATIONS

Performing calibration routines with room-temperature control electronics is common practice for transmon-based devices and detailed discussions can be found in Refs. [51, 52]. Executing these routines using a novel low-power cryo-CMOS ASIC represents an important demonstration of functionality. As shown in Fig. 5, experiments to optimize the amplitude, frequency, phase, and width of the control pulses were performed, from which a set of optimized parameters were found and stored in wave-form memory. The successful execution of these calibrations is

necessary to facilitate the high-fidelity two-qubit echoed CR gate demonstrated in this paper.

### A. Single-qubit wave forms

The DACs produced in-phase and quadrature signals of the form  $V_I(t) = \Omega(t) \cos(\omega_{SSB}t - \phi)$  and  $V_Q(t) = \Omega(t) \sin(\omega_{SSB}t - \phi)$ , respectively, where  $\omega_{SSB}$  is the single-sideband frequency and  $\phi$  is the phase. For single-qubit gates, the signals were shaped with a Gaussian envelope of the form

$$\Omega_G(t) = \begin{cases} \Omega_0 \frac{e^{t^2/2\sigma^2} - e^{t_g^2/2\sigma^2}}{1 - e^{t_g^2/2\sigma^2}}, & t \leq t_g, \\ 0, & \text{else,} \end{cases} \quad (1)$$

where  $\Omega_0$  is the amplitude, the width of the pulse is defined by  $t_g$ , and  $\sigma$  is the standard deviation. The functional form of  $\Omega_G$  was chosen to enforce that the pulse started and ends with zero amplitude [40,53]. The pulse amplitudes for  $X(\pi)$  and  $X(\pi/2)$  were calibrated by driving the qubit at the  $|0\rangle$  to  $|1\rangle$  transition frequency. During the Gaussian pulse, the higher-energy levels of the transmon experienced a Stark shift, resulting in a drive-induced phase shift about the  $Z$  axis of the Bloch sphere. This phase shift was amplitude dependent and was mitigated with a derivative removal via adiabatic gate (DRAG) calibration [40,53–55]. The DRAG pulse was implemented by applying  $\Omega_G(t)$  and the derivative  $\beta\dot{\Omega}_G(t)$  to the in-phase and quadrature channels, respectively. As shown in Fig. 5(c), the DRAG pulse was repeated  $N$  times while sweeping the DRAG scale parameter  $\beta$ ; the collective minima corresponded to the optimal parameter value. Subsequent single-qubit calibrations consisted of error-amplification measurements for fine tuning all pulse parameters [56,57].

### B. Two-qubit wave forms

In order to drive the cross-resonance interaction, a square-Gaussian wave form was applied to the control qubit at the resonant frequency of the target qubit [58]. The pulse shape was defined by a Gaussian rise and fall of length  $\tau_r$  and standard deviation  $\sigma_r$ , had a flat top of length  $\tau_p - 2\tau_r$  and amplitude  $\Omega_0$ , and had a total pulse length of  $\tau_p$ . The expression describing the pulse shape is given by

$$\Omega_{GS}(t) = \begin{cases} \Omega_0 \left( e^{\frac{(t-\tau_r)^2}{2\sigma_r^2}} - e^{\frac{\tau_r^2}{2\sigma_r^2}} \right) / \left( 1 - e^{\frac{\tau_r^2}{2\sigma_r^2}} \right), & 0 < t < \tau_r, \\ \Omega_0, & \tau_r < t < \tau_p - \tau_r, \\ \Omega_0 \left( e^{\frac{[t-(\tau_p-\tau_r)]^2}{2\sigma_r^2}} - e^{\frac{\tau_r^2}{2\sigma_r^2}} \right) / \left( 1 - e^{\frac{\tau_r^2}{2\sigma_r^2}} \right), & \tau_p - \tau_r < t < \tau_p. \end{cases} \quad (2)$$



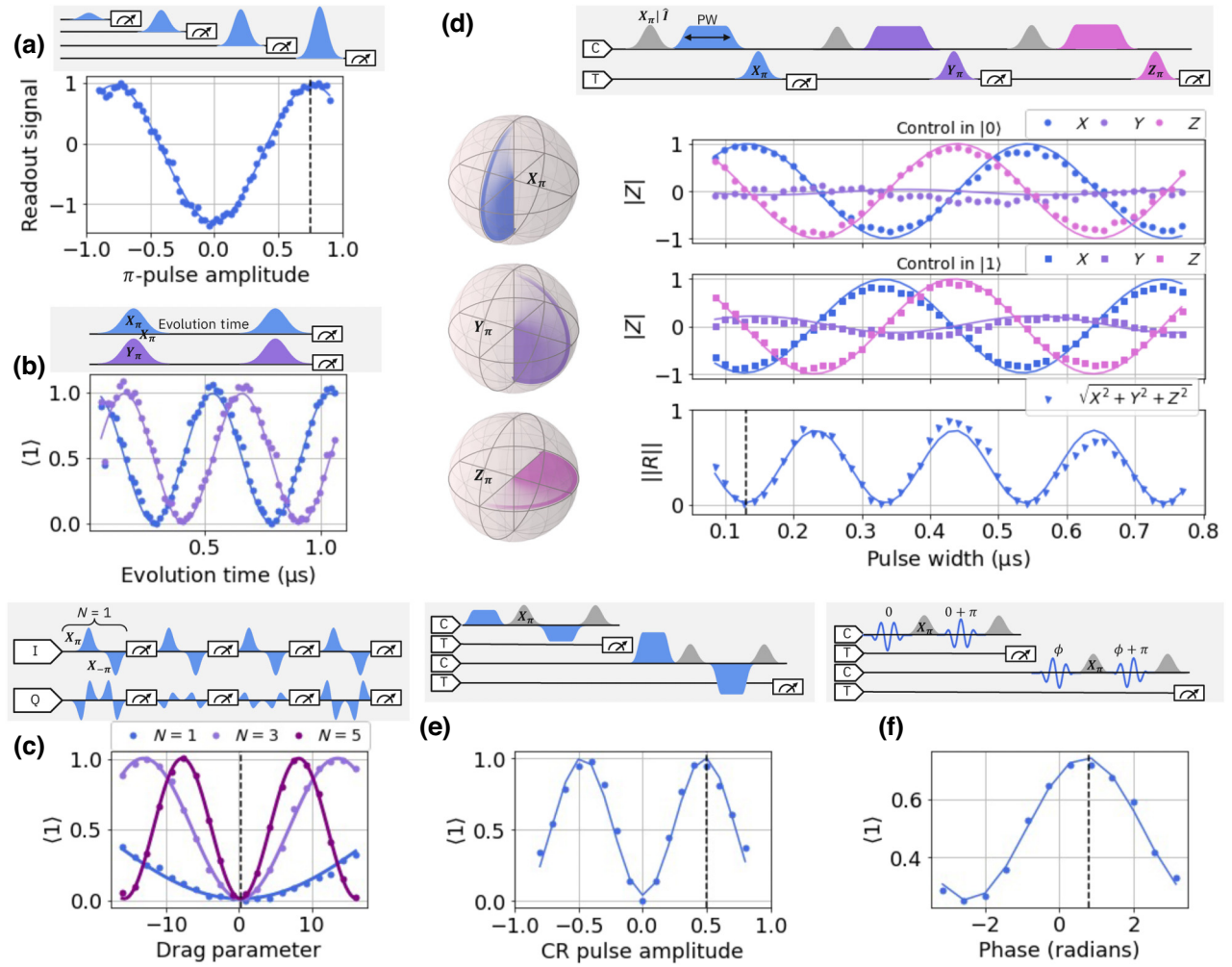


FIG. 5. Calibration data for single-qubit (1Q) and two-qubit (2Q) wave forms and the corresponding pulse sequences. Where relevant, dashed vertical lines indicate optimal values extracted from calibration. (a) The 1Q Rabi measurement to tune the  $\pi$ -pulse amplitude, optimally at the maxima in the curve. (b) The 1Q Ramsey measurement to tune the qubit frequency. Playing either a  $X(\pi/2)$  (blue) or  $Y(\pi/2)$  (purple) pulse as the second pulse in the sequence yields two curves with a relative phase of  $\pi/2$ . The period of the curve(s) determines how offset the driven frequency is from the transition frequency of the qubit. (c) Derivative removal by adiabatic gate (DRAG) calibration to add a derivative of the Gaussian quadrature component to the pulse shape. The  $\pi$  pulses are repeated  $N$  times within a pulse sequence for different sizes of the DRAG parameter. Each sequence yields a curve with a different period. The optimal DRAG parameter is the collective minima of the different curves. (d) Hamiltonian tomography measured as a function of the CR pulse width. The  $|Z\rangle$  state of the target qubit is measured after projecting into  $\langle X\rangle$ ,  $\langle Y\rangle$ , and  $\langle Z\rangle$ . These measurements are performed with the control qubit prepared in either the  $|0\rangle$  or the  $|1\rangle$  state. The oscillations on the target are fitted to a Hamiltonian, which can be used to extract device parameters and/or provide information about the optimal pulse width and the extent of errors such as  $IY$ . By computing the Bloch vector  $|\vec{R}||$ , one can extract the optimal CR pulse length at the first minima of the curve. The qubit-to-qubit coupling was extracted to be  $J = 2.7$  MHz. (e) CR amplitude calibration, optimally at the first maxima. (f) CR phase calibration, optimally at the first maxima.

The interactions between control and target qubits could be driven by sweeping the pulse width for a fixed amplitude or, conversely, by sweeping the amplitude for a fixed pulse width. In Fig. 5(d), full-state Hamiltonian tomography is performed on the target qubit. This calibration is a method to identify the coherent error terms relevant to the microwave drive [41].

The calibration is performed by sweeping the width of the CR wave form applied to the control and then applying a  $X_\pi$ ,  $Y_\pi$ , or  $Z_\pi$  pulse in order to project the state of the target onto the  $X$ ,  $Y$ , or  $Z$  axes of the Bloch sphere. This process is carried out for the control in  $|0\rangle$  and  $|1\rangle$ . The data are then fitted to a block-diagonal Hamiltonian and the six interaction terms  $IX$ ,  $IY$ ,  $IZ$ ,

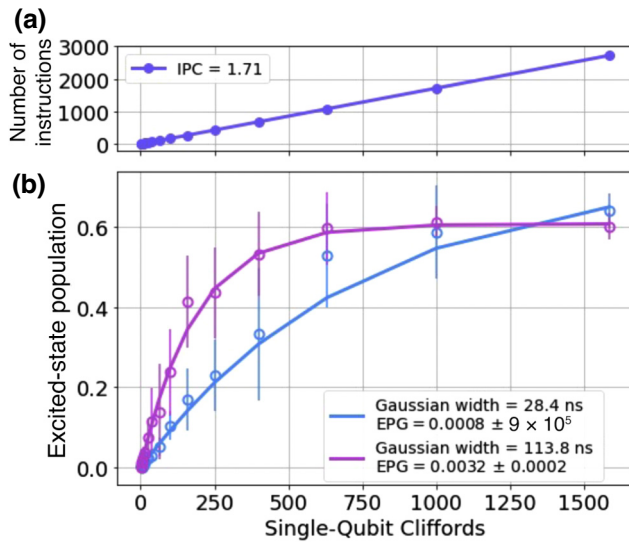


FIG. 6. The RB data for the single-qubit experiments. (a) The number of instructions increases linearly with the Clifford count, with 36.94% of the instructions coming from the last RB data point. A single-qubit RB experiment requires 1.71 IPC, which is extracted from the slope of the plotted line. (b) The single-qubit RB data for Gaussian widths of 28.4 ns and 113.8 ns, respectively. The longer gate has more error  $\epsilon_{1Q} = 0.0032$  but converges more quickly, requiring fewer instructions. The shorter gate has less error  $\epsilon_{1Q} = 0.0008$  but requires longer Clifford sequences to measure and thus more instructions. The observed decay does not converge to 0.5, indicating leakage outside of the computational basis [40]. For example, leakage into the  $|2\rangle$  state gives rise to IQ counts that are different from  $|0\rangle$  and  $|1\rangle$ . This yields a measurement result of the form  $V_0p_0 + V_1p_1 + V_2p_2$ , which converges above 0.5, without proper binning of the higher excited states. We believe this effect to be caused by spurious spectral content such as LO leakage.

$ZX$ ,  $ZY$  and  $ZZ$  are parametrized. The quantity  $\|R\| = \sqrt{(\langle X_0 \rangle + \langle X_1 \rangle)^2 + (\langle Y_0 \rangle + \langle Y_1 \rangle)^2 + (\langle Z_0 \rangle + \langle Z_1 \rangle)^2}$  in Fig. 5(d) is the two-norm distance between Bloch vectors of the target for the two states of the control. When  $\|R\| = 0$ , the two qubits are maximally entangled and this indicates an optimal pulse width  $\tau_p$ .

## VII. BENCHMARKING A CROSS-RESONANCE GATE

The cryo-CMOS chip was specially designed to generate wave forms for transmon qubits in a CR-based architecture, as shown in Fig. 1(a). A detailed description of the experimental setup with the CMOS chip thermalized to the  $T = 4$  K stage in a dilution refrigerator is provided in Sec. V. In a cross-resonance-based qubit device, entanglement between neighboring qubits was generated via the CR interaction [19,41,56,59–62]. The CR interaction gives rise to a  $ZX$  term in the Hamiltonian, which describes a state-dependent Rabi oscillation on the target

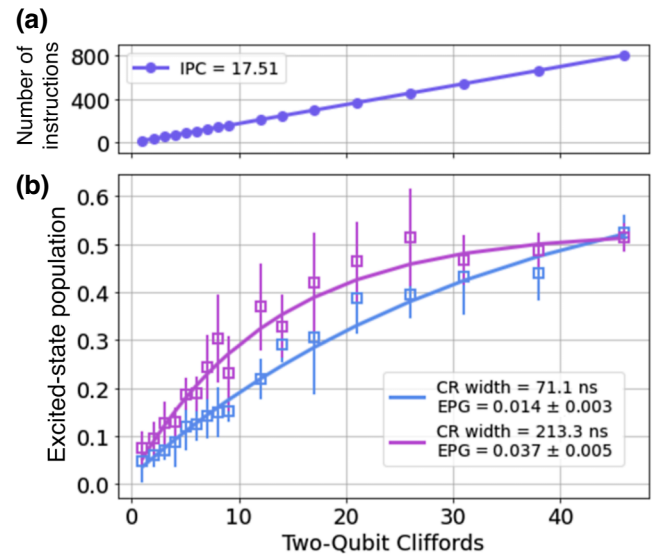


FIG. 7. The RB data for the two-qubit experiments. (a) Two-qubit RB is more demanding on the processor, requiring 17.51 IPC, with 18.41% of the instructions coming from the last data point. For both RB experiments, a small fraction of the total number of instructions come from presequence calibration, buffering, and pulse idle times, and contribute to the IPC. (b) Two-qubit RB measurements for CR pulse widths of 71.1 ns and 213.3 ns, respectively. The shorter gate lengths have less error— $\epsilon_{2Q} = 0.014$  compared to  $\epsilon_{2Q} = 0.037$ —but require more computational overhead in order to measure with precision using traditional RB. The error bars for the RB experiments are averaged over ten rounds of the same random seed.

qubit, which in turn depends on the state of the control qubit. The entanglement is generated by applying an rf drive to the control qubit at the  $|0\rangle$ -to- $|1\rangle$  transition frequency of the target qubit. The always-on coupling in CR devices gives rise to parasitic terms in the Hamiltonian, such as  $ZZ$ ,  $ZI$ , and  $IZ$ . These undesirable terms can be mitigated in hardware through cancellation buses [62] and digitally through echoed gate sequences [41,61]. Here, an echoed gate was realized on a device with a cancellation bus. This gate sequence used Gaussian wave forms for single-qubit rotations and square-Gaussian wave forms for generating entanglement.

### A. Single-qubit randomized benchmarking

Single-qubit RB experiments were performed both individually and simultaneously on the control and target qubits and characterized as a function of the gate length  $t_g$ . For each  $t_g$ , the wave forms were calibrated and measurements of  $T_1$  and  $T_2$  were interleaved between RB experiments. The results for individual RB experiments are displayed in Fig. 8(e). The data show an error reduction as a function of the gate length  $t_g$ . However, the observed errors are not solely explained by decoherence. For example, the errors do not track with the first-order single-qubit

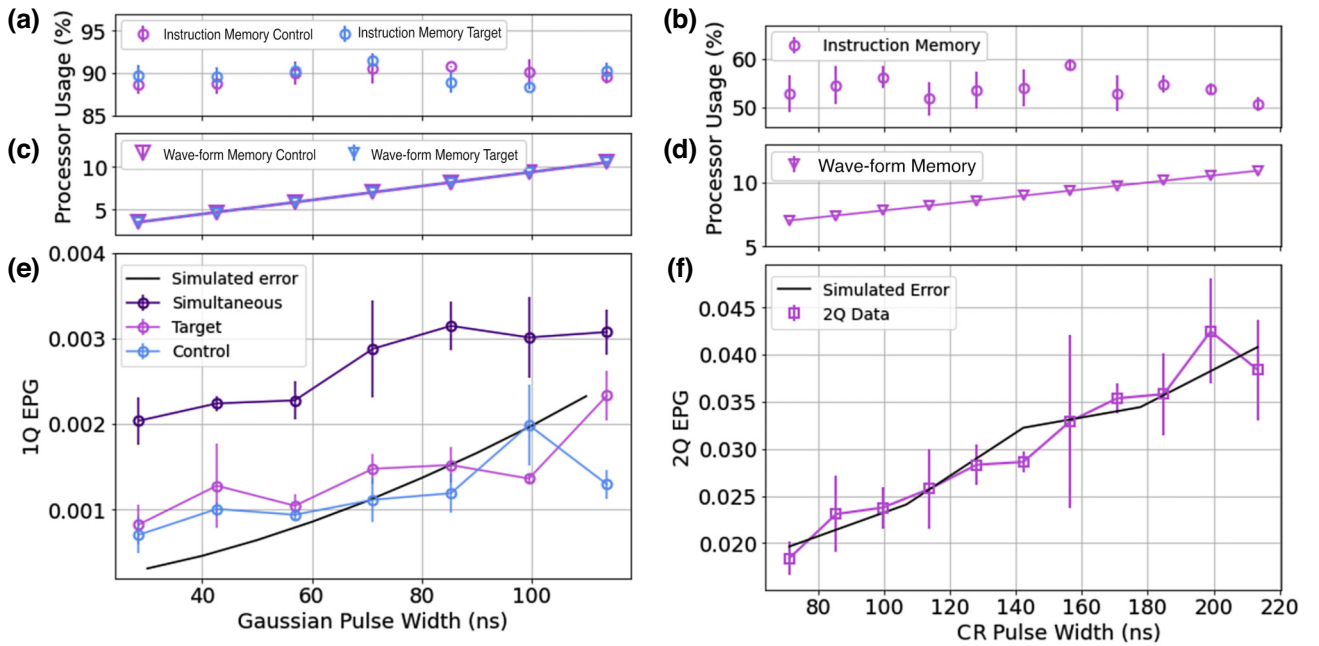


FIG. 8. Two-qubit RB data for different single-qubit and two-qubit gate lengths. All data and error bars are averaged over three different RB experiments performed at each gate length. The random Clifford-gate sequences are stored in instruction memory and played out sequentially during and experiment. The amount of instruction memory used is plotted as a percentage of fullness for (a) single qubit gate pulse widths, and (b) two qubit gate pulse widths. The number of instructions is independent of gate width, and the usage fluctuates because the number of gates in a random sequence varies. The amount of waveform memory used increases linearly with (c) single qubit gate pulse widths, and (d) two qubit gate pulse widths. (e) Single-qubit RB measured on each qubit individually and both qubits simultaneously, while sweeping the Gaussian pulse width. The individual RB error is modeled with a Hamiltonian simulation assuming 83.1 kHz of Z rotation, consistent with oscillations observed in the rotary-echo experiment. For simultaneous RB, each qubit is measured and the average error is reported. The simultaneous error is believed to be due to an increase in quantum crosstalk from ZZ and classical crosstalk from the CMOS chip. (f) Two-qubit RB of an echoed cross-resonance gate as a function of the width of the cross-resonance pulse. Faster gates are observed to have reduced error, consistent with a reduction in decoherence errors; however, simulations reveal that qubit coherence is not the leading source of error. Using a parametrized two-qubit Hamiltonian, the additional error was modeled by assuming an amplitude-dependent Z error on the target qubit and a constant Z error on the control qubit.

error model  $\epsilon_{1Q}$  [46,63]. This discrepancy implies that the control electronics are contributing to the excess error measured in the RB experiments. Potential control-related error sources include the spectral content, as shown in Fig. 10, and pulse-amplitude noise, detailed in Sec. VIII. The spurious spectral content and quasistatic amplitude noise give rise to over- or under-rotation errors that vary quadratically with the noise source amplitude.

During simultaneous operation, each control channel plays unique random Clifford sequences, which gives rise to coherent quantum crosstalk errors of the form  $\epsilon_{1Q}^{ZZ} = 1/6 (2\pi ZZ t_g)^2$  that contribute to the total observed error [44]. The increase in simultaneous error observed in Fig. 8(e) is not explained by an error analysis that assumes only coherent quantum crosstalk, a result that implies an external noise source. It is suspected that classical crosstalk on the CMOS chip arises during simultaneous wave-form generation, which gives rise to the observed error.

## B. Two-qubit randomized benchmarking

Two-qubit gates are more difficult in practice than single-qubit gates. When compared to single-qubit gates, two-qubit gates are longer and there are more ways in which errors can arise; furthermore, the coupled qubits are more sensitive than single qubits to error sources. For example, in the decoherence error model  $\epsilon_{2Q}$  [46,63], the error coefficients are larger and there are more terms that factor into the error (see Sec. X). In addition to decoherence, there are other potential sources of error, including phase noise [64], amplitude noise, pulse-induced decay [65,66], spurious spectral tones, coherent quantum crosstalk [67], microwave crosstalk [68,69], and leakage outside of the computational basis [40,70]. Additionally, the two-qubit wave-form generation is more complex, requiring on average 17.51 IPC (Fig. 7), compared to 1.71 IPC for single-qubit gates (Fig. 6).

The echoed two-qubit cross-resonance gate was calibrated and benchmarked for different two-qubit-gate

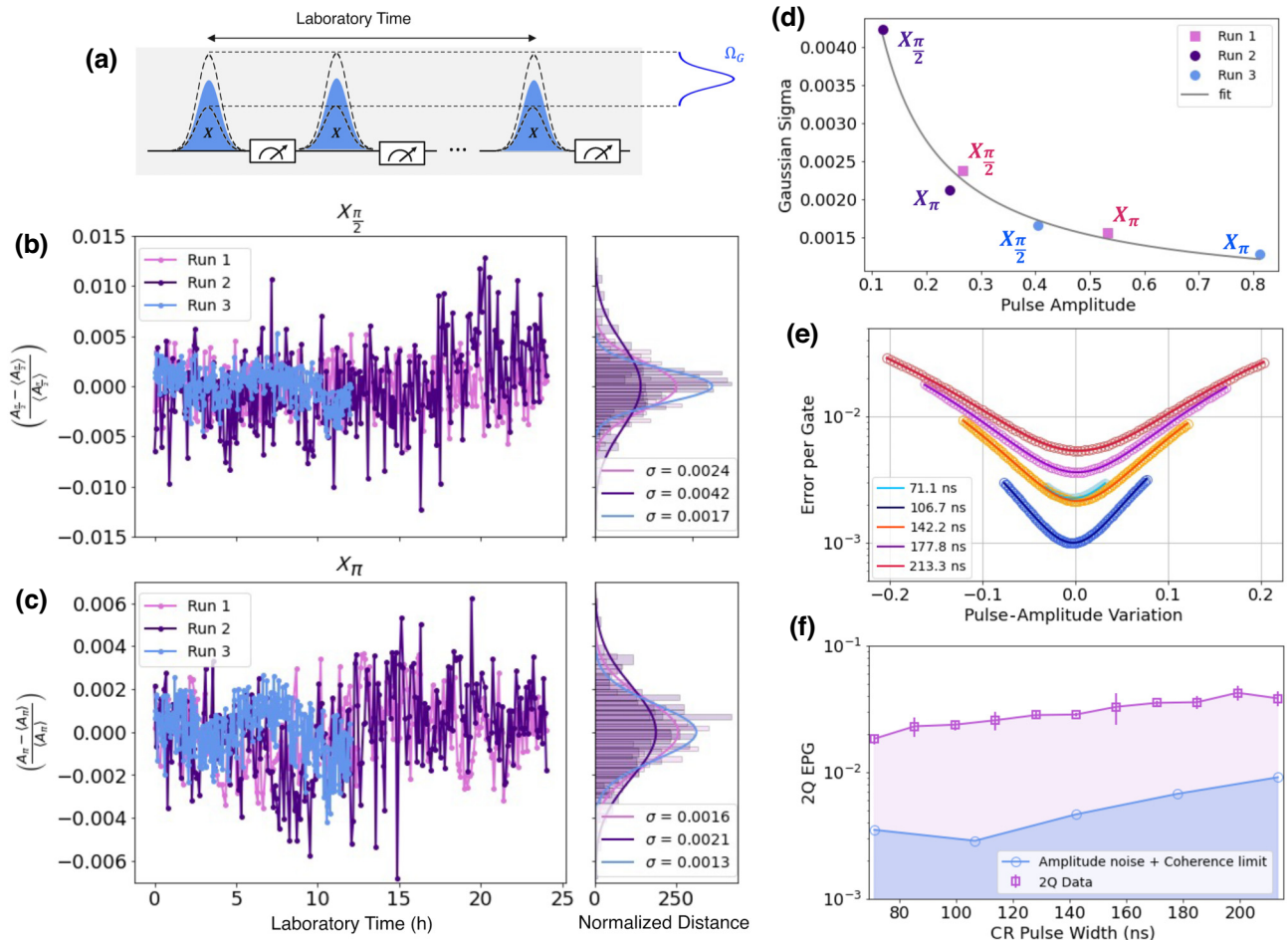


FIG. 9. (a) The measurement sequence performed to extract quasistatic amplitude noise. The pulse-amplitude calibrations were repeated consecutively and the  $X_{\pi/2}$  and  $X_{\pi}$  calibrations were interleaved. The  $t_g$  value was the same for  $X_{\pi/2}$  and  $X_{\pi}$  and the amplitude was varied. For each calibration, the amplitude coefficient was observed to fluctuate, consistent with a normal noise distribution. (b),(c) The percent difference of the pulse-amplitude coefficients plotted for (b)  $X_{\pi/2}$  and (c)  $X_{\pi}$ , respectively. The time series pulse-amplitude coefficient data are binned, fitted to a normal distribution, and are projected to the panel to the right. The noise was measured for three different experimental configurations: run 1 was the standard experiment, run 2 had additional filtering on the room-temperature supply lines and a shorter  $t_g$ , and run 3 had increased on-chip attenuation and a longer  $t_g$ . The pulse amplitude varied with  $t_g$ . (d) The widths of the normal distributions plotted as a function of the amplitude coefficient. The relationship between the quasistatic noise and the pulse amplitude is shown to follow a  $1/x$  dependence of the form  $f(x) = 0.00041/x + 0.0007$ . This result indicates that the SNR improves for larger DAC amplitudes. (e) The simulated gate error as a function of the quasistatic amplitude noise  $\sigma$  and for different CR pulse widths. The error is fitted to a quadratic heuristic model of the form  $c_0 + c_1 \Delta_{amp} + c_2 \Delta_{amp}^2$ . The coefficients are listed in Table III. (f) The measured 2Q gate error as a function of the CR pulse width, along with the simulated error due to observed amplitude noise. The modeling implies that amplitude noise is not the leading source of error.

lengths, as shown in Fig. 8(d). The pulses driving the single-qubit rotations were held constant and measurements of  $T_1$  and  $T_2$  were interleaved with RB experiments. The error rates were simulated using a two-qubit model Hamiltonian parametrized with experimental data and the measured RB data do not track with decoherence or coherent quantum crosstalk.

The simulations suggest that the error is due to an always-on  $Z$  rotation combined with an amplitude-dependent  $Z$  error on the target qubit. The simulated

$Z$  error is consistent with rotary-echo experiments performed on the target qubit, which are observed to have an average  $Z$  rotation of 83.1 kHz, and modeled using Lindblad master equations [Fig. 10(a)]. The target  $Z$  rotation is further consistent with spectrum-analyzer measurements of the CMOS-chip output that reveal excess LO leakage when measured  $T = 5$  K in a closed-cycle  $^4\text{He}$  cryostat. The presence of off-resonance spectral content will give rise to an ac Stark effect [71] that shifts the qubit frequency, resulting in a  $Z$  rotation. As shown



TABLE III. The fit parameters for the heuristic gate-error model that assumes quasistatic amplitude noise. The fitting routines were applied to numerical results from Hamiltonian simulations using measured device parameters. The quadratic behavior is consistent with “theta squared” errors that arise from over- or under-angle rotations.

Coefficient	71 ns	106.7 ns	142.2 ns	177.8 ns	213.3 ns	Mean
$c_2$	0.593127	0.353895	0.471876	0.523495	0.548932	0.498265
$c_1$	0.003120	0.001647	-0.001617	-0.002103	-0.004213	-0.000633
$c_0$	0.002252	0.001001	0.002150	0.003619	0.005373	0.002879

in Fig. 10(c) the spurious content was observed to be channel dependent, with the LO leakage on the control-qubit channel being the worst of the two. This random

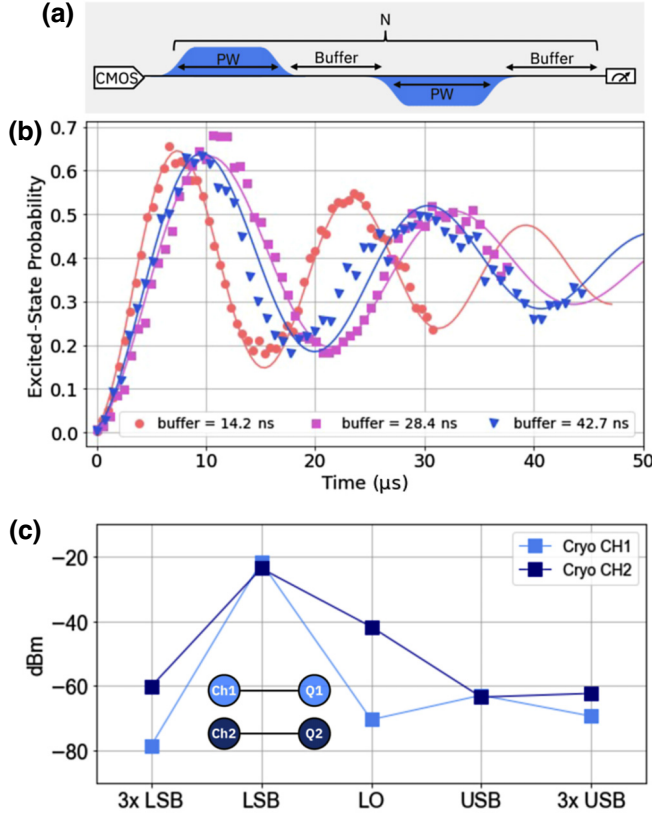


FIG. 10. (a) A pulse sequence for a rotary-echo experiment. (b) A rotary-echo measurement on the control qubit shows oscillatory behavior rather than the expected exponential decay. The data are fitted to a master-equation simulation that includes  $T_1$ ,  $T_2$ , and a constant  $Z$ . (c) The measured amplitudes of the most prominent spurious peaks observed in a spectrum analyzer. The LO was set to 5 GHz, with a sideband frequency of 250 MHz. The cryogenic measurements were performed at 5 K in a  $^4\text{He}$  closed-cycle cryostat prior to loading into a dilution refrigerator for qubit testing. The same bias conditions as for the CMOS chip were used for the qubit measurements. Channel 1 was connected to the control qubit and channel 2 was connected to the target qubit. Here, channel 2 is observed to have more spurious content, which is consistent with observed behavior on the control qubit. The additional spurious content Stark shifts the qubit, which shifts the energy levels and gives rise to an always-on  $Z$  error on the target qubit.

variation is in excess of what was expected in the chip design phase, so the designed tuning range did not cover the distribution observed in hardware samples; the data show an example where there was sufficient range for one channel and insufficient for the second.

## VIII. AMPLITUDE NOISE

In Fig. 9, a measurement was performed to observe quasistatic amplitude noise. The measurement sequence consisted of repeated amplitude calibrations for  $X_\pi$  followed by  $X_{\pi/2}$ . For each calibration, the DAC amplitude coefficient was stored and the percent difference was plotted as a function of the laboratory time. The data were binned and then fitted to a normal distribution. Three different measurements were performed, each with slightly different experimental configurations: run 1 was performed with the nominal device configuration and gate length, run 2 was performed with a shorter gate length after adding additional low-pass filtering to the supply voltages, and for run 3 the on-chip attenuation was set to the maximum value of 20 dB, requiring a long low-amplitude Gaussian pulse to generate a  $\pi$  rotation.

For each experimental run, the amplitude required to drive a  $X_\pi$  and a  $X_{\pi/2}$  changed due to different amounts of in-line attenuation. The relative fluctuations in the pulse amplitude were evaluated and shown to increase for larger pulse amplitudes. In Fig. 9, the distribution width  $\sigma$  is plotted versus the pulse amplitude, which yields a  $1/x$

TABLE IV. The fit parameters extracted from the master-equation simulation. A Nelder-Mead optimization routine was performed on the data set from Fig. 10 to determine the best fit. The  $Z$  rate was a free parameter in each fit. The  $X$  rate,  $T_1$ , and  $T_2^*$  were allowed small bounds in order to achieve an optimal fit. We note that the optimal  $T_1$ , and  $T_2^*$  fit parameters are different from the observed values reported in Tables II. The deviations in  $T_1$  are attributed to two-level system (TLS) fluctuations that occurred before the rotary echo rotary-echo data was collected [74,75].

Buffer (ns)	$X$ rate (MHz)	$Z$ rate (kHz)	$T_1$ ( $\mu\text{s}$ )	$T_2^*$ ( $\mu\text{s}$ )
14.2	1.897	94.3	82.2	33.5
28.4	1.927	72.5	82.2	32.6
42.7	1.927	82.6	79.9	33.1

dependency. Since the different experimental configurations did not deviate from the  $1/x$  dependency, these results imply that the source of the noise is on-chip and could be related to an increase in low-frequency noise at cryogenic temperatures [72].

Error analysis was performed using the  $1/x$  fit of the noise as an input into the error model. We have found that the amplitude noise is larger than the decoherence errors but is not significant enough to explain the observed 2Q gate error. For small low-frequency fluctuations, the amplitude noise will result in either an over- or under-rotation during the gate. The error per gate from an over- or under-rotation can be fitted to the form  $c_0 + c_1 \Delta_{amp} + c_2 \Delta_{amp}^2$ . The error due to quasistatic amplitude noise was simulated using modeling techniques described in Sec. X. The simulated error was fitted to the quadratic error model and the coefficients for different sample lengths as shown in Table III, resulting in a simple heuristic error model.

### IX. ROTARY-ECHO EXPERIMENT

A rotary-echo experiment [73] was performed to measure driven decay, as shown in Fig. 10(b). The qubit was pulsed in the  $+X$  direction, followed by an idle gate, and then pulsed in the  $-X$  direction. The pulse width was 200 ns and the sequence was repeated for  $N = 31$  times. Perfectly symmetric and noiseless pulses will yield an exponential decay consistent with the qubit lifetime. Here, a time-dependent oscillation is observed and it is believed to be due to an ac Stark shift that arises during the pulse. The oscillations are fitted using a two-level qubit model where an additional  $Z$  rotation occurs during the  $+X$  and  $-X$  rotations. This  $Z$  rotation arises from spurious peaks that Stark shift the qubit while it is being driven, which is consistent with the spurious peaks shown in Fig. 10(c).

An average  $Z$ -rotation strength of 83.1 kHz was extracted from fitting to the data with the three different buffer lengths: 14.2 ns, 28.4 ns, and 42.7 ns. The simulation consisted of solving the Lindblad master equation for a single qubit with a time-dependent  $X$  pulse with an additional time-dependent  $Z$  pulse. All pulses were square-Gaussian [58] shaped with a 16-ns sigma. The  $X$ -rotation rate,  $Z$ -rotation rate,  $T1$ , and  $T2^*$  were all input variables used in the master-equation simulations to fit the experimental rotary-echo data. The best fit parameters are listed in Table IV. The time dependent oscillations were most sensitive to variations the  $Z$ -rotation strength.

### X. MODELING CROSS-RESONANCE GATE ERRORS

To model the effect of spurious  $Z$  errors on our qubits, we numerically calibrated the CR gate using a two-qubit

model Hamiltonian,

$$H = \sum_{c \in \{a,b\}} \left[ \omega_c c^\dagger c + \frac{\delta_c}{2} c^\dagger c (c^\dagger c - \mathbb{I}) \right] + J(a + a^\dagger)(b + b^\dagger) + \Omega_x(t) \cos(\omega_b t)(a + a^\dagger), \quad (3)$$

which describes two transmon qubits with lowering operators  $a$  and  $b$  within a Duffing model with anharmonicity  $\delta_{a/b}$  and coupling strength  $J$ . The results from this modeling are shown in Fig. 11 and indicate that error-inducing  $Z$  noise is not constant with the pulse amplitude and that the  $Z$  noise varies linearly with the pulse amplitude.

To generate the cross-resonance entangling interaction, the control-qubit term  $(a + a^\dagger)$  was driven at the frequency ( $\omega_b$ ) of the target qubit.  $\Omega_x(t)$  describes the pulse envelope,

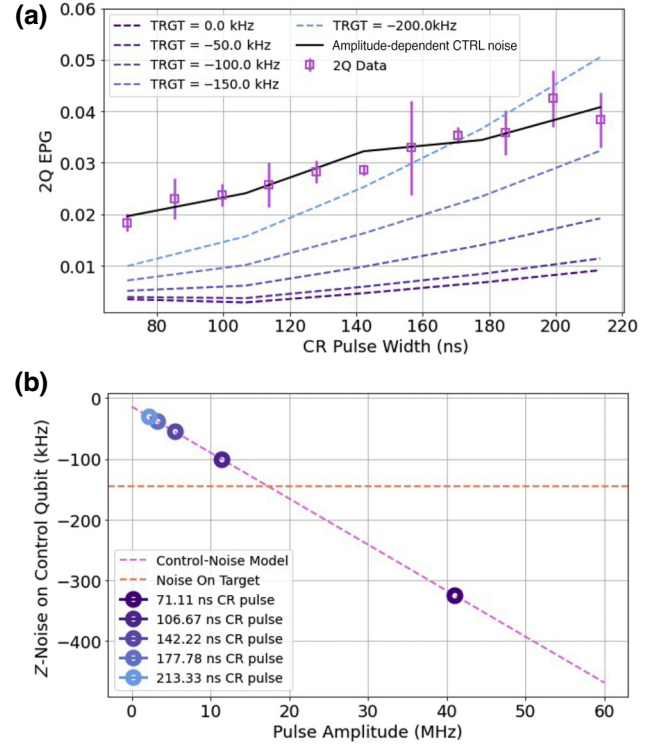


FIG. 11. (a) The measured and simulated two-qubit EPG as a function of the CR pulse width. The simulated error is shown for different amounts of  $Z$  noise on the target qubit, while the  $Z$  noise on the control qubit is held constant at the measured 85 kHz. The difference in slopes between the measured error and the simulated error implies that a constant- $Z$ -noise model is not sufficient for describing the observed errors. A best fit is obtained by introducing an amplitude-dependent  $Z$  noise on the control qubit, with the  $Z$  noise on the target qubit fixed to 145 kHz. We note that the amplitude-dependent  $Z$  noise on the control and the constant  $Z$  noise on the target are not explained by the experimental data that were collected. (b) The simulated  $Z$  noise as a function of the CR pulse amplitude. The observed EPG was best captured by assuming  $Z$  noise that varies linearly with the pulse amplitude.

where for CR gates we used the Gaussian-square shape  $\Omega_{\text{GS}}(t)$  described in Eq. 2. The pulse was allowed to rise (and fall) over twice  $\sigma_{\text{GS}}$ , the Gaussian width, before (and after) a square pulse of duration  $\tau_p - 4\sigma_{\text{GS}}$ . As shown in Fig. 11(c), for each pulse width considered, the amplitude of the pulse in Eq. (3) was calibrated to minimize the total error of a  $U_{\text{ideal}} = e^{-i\pi ZX/4}$  rotation (the native entangler produced by the CR drive). This was done by performing time-domain simulations of the Hamiltonian [Eq. (3)] to estimate  $U_{\text{sim}}$  and then minimizing the two-qubit-gate error, as defined by

$$\epsilon_{2Q} = 1 - \left( \left| \text{Tr}[U_{\text{sim}}^\dagger U_{\text{ideal}}] \right|^2 / 4 + 1 \right) / 5, \quad (4)$$

which yielded a linear dependence of the pulse amplitude and the error.

## XI. CONCLUSIONS

We have developed a low-power CMOS ASIC designed to operate at  $T = 4$  K that is able to generate sequences of rf wave forms for controlling, calibrating, and benchmarking a universal set of quantum gates between a pair of transmon qubits. The cryogenic control electronics were used to demonstrate high-fidelity two-qubit cross-resonance gates. A two-qubit Hamiltonian model has provided an insight into the behavior of spurious  $Z$  errors, which indicates that the control-electronics noise has an amplitude dependence. The modeling and analysis suggest that the observed drive-dependent  $Z$  rotation during rotary-echo experiments and of LO leakage in the output of the ASIC are connected, implying that spurious content from the CMOS chip is the primary source of gate error.

The CMOS processor was characterized across a wide variety of qubit experiments, demonstrating its viability for providing control pulses to next-generation quantum computers. Furthermore, these results highlight challenges with low-power cryogenic control electronics related to the instruction and memory requirements for standard qubit experiments. These results underscore the need for further innovation of digital architectures as gate error rates approach fault-tolerant thresholds.

## XII. OUTLOOK

The primary concern with realizing cryo-CMOS electronics for controlling qubits is the cooling limitation set by the dilution refrigerator. The CMOS ASIC used in these experiments was comprised of two low-power rf analog front ends that dissipated approximately 8.92 mW of power per channel and a quantum specific processor that dissipated approximately 10.42 mW per channel (see Fig. 4). Assuming the cooling powers in a standard dilution refrigerator, it is believed that up to 100 cryo-CMOS channels could be integrated into a quantum computing

system. The outlook becomes more optimistic when taking into account circuit innovations and advances in cooling infrastructure.

A key insight from this work is that the processor requirements for common calibrations and quantum information experiments will need to increase as the performance of the quantum processor improves but that increasing the processor capabilities will lead to more power dissipation from the CMOS chip. To mitigate the demand for more processing power, new ISAs may be required or new low-overhead experiments for calibrating and characterizing qubits will be needed.

The operation of CMOS control electronics within the dilution refrigerator has the potential for improved gate performance due to lower noise, reduced loss, and less dispersion; however, the technical nuances of cryogenic operation and system integration make it challenging to achieve this potential. For example, most foundry circuit models are not reliable below 50 K, surface-mount components do not meet the necessary specification at cryogenic temperatures, and the electrical path length between the support electronics and ASICs is long and lossy when compared to CMOS-based servers. Consequently, this technology will take time to achieve its full potential.

This paper has described the first demonstration of two-qubit RB with a cryogenic CMOS controller, with an observed error per gate of  $\epsilon_{2Q} = 1.4 \times 10^{-2}$ . The leading source of error has been shown to arise from the electronics; however, an advantage of custom CMOS electronics is that new circuits can be designed to mitigate error sources after they have been identified. The primary engineering challenge arises from being able to distinguish errors arising from devices physics versus errors arising from control electronics. The identification of error sources is a nontrivial task but these efforts are becoming simpler due to innovative approaches being developed in the field of QCVV.

## ACKNOWLEDGMENTS

We thank Oliver Dial, Muir Kumph, David McKay, and George Zettles for technical advice and support. We acknowledge Donald Bethune for his development of the cryogenic infrastructure and Thomas Fox for his contribution to the ISA and we thank Jiri Stehlik, David Zajac, and Seth Merkle for helpful technical discussions.

- 
- [1] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, The future of quantum computing with superconducting qubits, *J. Appl. Phys.* **132**, 160902 (2022).
  - [2] J. M. Gambetta, J. M. Chow, and M. Steffen, Building logical qubits in a superconducting quantum computing system, *npj Quantum Inf.* **3**, 2 (2017).

- [3] R. Van Meter and D. Horsman, A blueprint for building a quantum computer, *Communications of the ACM* **56**, 84–93 (2013).
- [4] J. M. Hornibrook, J. I. Colless, I. D. Conway Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J. D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly, Cryogenic control architecture for large-scale quantum computing, *Phys. Rev. Appl.* **3**, 024010 (2015).
- [5] X. Xue *et al.*, CMOS-based cryogenic control of silicon quantum circuits, *Nature* **593**, 205 (2021).
- [6] S. J. Pauka, K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, C. Cantaloube, N. Dick, G. C. Gardner, M. J. Manfra, and D. J. Reilly, A cryogenic CMOS chip for generating control signals for multiple qubits, *Nat. Electron.* **4**, 64 (2021).
- [7] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, in *2016 IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA, USA, 2016), p. 13.5.1.
- [8] D. J. Reilly, in *2019 IEEE International Electron Devices Meeting (IEDM)* (San Francisco, CA, USA, 2019), p. 31.7.1.
- [9] J. C. Bardin *et al.*, in *IEEE Journal of Solid-State Circuits* (2019), Vol. 54, p. 3043.
- [10] D. J. Frank *et al.*, in *2022 IEEE International Solid-State Circuits Conference (ISSCC)* (San Francisco, CA, USA, 2022), Vol. 65, p. 360.
- [11] S. Chakraborty *et al.*, A cryo-CMOS low-power semi-autonomous transmon qubit state controller in 14-nm Fin-FET technology, *IEEE J. Solid-State Circuits* **57**, 3258 (2022).
- [12] G. E. Miller *et al.* in *International Cryocooler Conference* (Boulder, CO, 2007), p. 655.
- [13] P. Das, A. Locharla, and C. Jones, LILLIPUT: A lightweight low-latency lookup-table based decoder for near-term quantum error correction, *arXiv:2108.06569* (2021).
- [14] C. Guo, J. Lin, L.-C. Han, N. Li, L.-H. Sun, F.-T. Liang, D.-D. Li, Y.-H. Li, M. Gong, Y. Xu, S.-K. Liao, and C.-Z. Peng, Low-latency readout electronics for dynamic superconducting quantum computing, *AIP Adv.* **12**, 045024 (2022).
- [15] S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinsoo, R. Keller, J. Lütolf, C. Eichler, and A. Wallraff, Engineering cryogenic setups for 100-qubit scale superconducting circuit systems, *EPJ Quantum Technol.* **6**, 2 (2019).
- [16] M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, C. C. Bultink, Y. Salathe, N. Haandbaek, J. Sedivy, and L. DiCarlo, Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor, *Appl. Phys. Lett.* **116**, 054001 (2020), publisher: American Institute of Physics.
- [17] S. Simbierowicz, V. Y. Monarkha, S. Singh, N. Mesasoudi, P. Krantz, and R. E. Lake, Microwave calibration of qubit drive line components at millikelvin temperatures, *Appl. Phys. Lett.* **120**, 054004 (2022), publisher: American Institute of Physics.
- [18] 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, *Phys. Rev. A* **76**, 042319 (2007).
- [19] 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, *Phys. Rev. Lett.* **107**, 080502 (2011).
- [20] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [21] D. Bacon, Operator quantum error-correcting subsystems for self-correcting quantum memories, *Phys. Rev. A* **73**, 012340 (2006).
- [22] 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).
- [23] P. Aliferis and A. W. Cross, Subsystem fault tolerance with the Bacon-Shor code, *Phys. Rev. Lett.* **98**, 220502 (2007).
- [24] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* **567**, 491 (2019).
- [25] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature* **567**, 209 (2019).
- [26] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, *PRX Quantum* **3**, 010345 (2022).
- [27] 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).
- [28] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
- [29] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, Three-qubit randomized benchmarking, *Phys. Rev. Lett.* **122**, 200502 (2019).
- [30] 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, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [31] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, *Nat. Rev. Phys.* **2**, 382 (2020).
- [32] A. D. Córcoles, M. Takita, K. Inoue, S. Lekuch, Z. K. Mineev, J. M. Chow, and J. M. Gambetta, Exploiting dynamic quantum circuits in a quantum algorithm with superconducting qubits, *Phys. Rev. Lett.* **127**, 100501 (2021).
- [33] R. J. Overwater, M. Babaie, and F. Sebastiano, Neural-network decoders for quantum error correction using surface codes: A space exploration of the hardware cost-performance tradeoffs, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [34] E. H. Chen, T. J. Yoder, Y. Kim, N. Sundaresan, S. Srinivasan, M. Li, A. D. Córcoles, A. W. Cross, and M. Takita,



- Calibrated decoders for experimental quantum error correction, *Phys. Rev. Lett.* **128**, 110504 (2022).
- [35] A. Wack, H. Paik, A. Javadi-Abhari, P. Jurcevic, I. Faro, J. M. Gambetta, and B. R. Johnson, Quality, speed, and scale: Three key attributes to measure the performance of near-term quantum computers, [arXiv:2110.14108](https://arxiv.org/abs/2110.14108) (2021).
- [36] D. Bhattacharya and N. K. Jha, FinFETs: From devices to architectures, *Adv. Electron.* **2014**, 365689 (2014).
- [37] K. O’Sullivan, C. Gorman, M. Hennessy, and V. Callaghan, A 12-bit 320-MSample/s current-steering CMOS D/A converter in 0.44 mm<sup>2</sup>, *IEEE J. Solid-State Circuits* **39**, 1064 (2004).
- [38] J.-S. Park, S. Subramanian, L. Lampert, T. Mladenov, I. V. Klotchkov, D. Kurian, E. Juárez-Hernández, B. P. Esparza, S. R. Kale, K. T. A. Beevi, S. P. Premaratne, T. Watson, S. Suzuki, M. Rahman, J. Timbadiya, S. Soni, and S. Pellerano, A fully integrated cryo-CMOS SoC for qubit control in quantum computers capable of state manipulation, readout and high-speed gate pulsing of spin qubits in Intel 22nm FFL FinFET technology, *2021 IEEE International Solid-State Circuits Conference (ISSCC)* **64**, 208 (2021).
- [39] J. P. G. Van Dijk *et al.* in *IEEE International Solid-State Circuits Conference (ISSCC)* (2020), Vol. 55, p. 304.
- [40] 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).
- [41] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Procedure for systematically tuning up cross-talk in the cross-resonance gate, *Phys. Rev. A* **93**, 060302(R) (2016).
- [42] L. Moro, M. G. A. Paris, M. Restelli, and E. Prati, Quantum compiling by deep reinforcement learning, *Commun. Phys.* **4**, 178 (2021).
- [43] L. H. Pedersen, N. M. Møller, and K. Mølmer, Fidelity of quantum operations, *Phys. Lett. A* **367**, 47 (2007).
- [44] P. J. J. O’Malley *et al.*, Qubit metrology of ultralow phase noise using randomized benchmarking, *Phys. Rev. Appl.* **3**, 044009 (2015).
- [45] D. Willsch, M. Nocon, F. Jin, H. De Raedt, and K. Michielsen, Gate-error analysis in simulations of quantum computers with transmon qubits, *Phys. Rev. A* **96**, 062302 (2017).
- [46] T. Abad, J. Fernández-Pendás, A. Frisk Kockum, and G. Johansson, Universal fidelity reduction of quantum operations from weak dissipation, *Phys. Rev. Lett.* **129**, 150504 (2022).
- [47] J. Kelly *et al.*, Optimal quantum control using randomized benchmarking, *Phys. Rev. Lett.* **112**, 240504 (2014).
- [48] A. V. Rodionov, A. Veitia, R. Barends, J. Kelly, D. Sank, J. Wenner, J. M. Martinis, R. L. Kosut, and A. N. Korotkov, Compressed sensing quantum process tomography for superconducting quantum gates, *Phys. Rev. B* **90**, 144504 (2014).
- [49] A. Gaikwad, K. Shende, Arvind, and K. Dorai, Implementing efficient selective quantum process tomography of superconducting quantum gates on IBM quantum experience, *Sci. Rep.* **12**, 3688 (2022).
- [50] Cryogenic properties of commonly used metals, <https://trc.nist.gov/cryogenics/materials/materialproperties.htm>.
- [51] I. Quantum, Calibrating qubits with QISKIT pulse, *Learn Quantum Computation Using Qiskit*, <https://qiskit.org/textbook/ch-quantum-hardware/index-pulses.html>.
- [52] A. D. Patterson, J. Rahamim, T. Tsunoda, P. A. Spring, S. Jebari, K. Ratter, M. Mergenthaler, G. Tancredi, B. Vlastakis, M. Esposito, and P. J. Leek, Calibration of a cross-resonance two-qubit gate between directly coupled transmons, *Phys. Rev. Appl.* **12**, 064013 (2019).
- [53] J. M. Gambetta, F. Motzoi, S. T. Merkel, and F. K. Wilhelm, Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator, *Phys. Rev. A* **83**, 012308 (2011).
- [54] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, *Phys. Rev. Lett.* **103**, 110501 (2009).
- [55] 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, *Phys. Rev. A* **82**, 040305(R) (2010).
- [56] 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, *Phys. Rev. A* **93**, 012301 (2016).
- [57] N. V. Vitanov, Relations between single and repeated qubit gates: Coherent error amplification for high-fidelity quantum-gate tomography, *New J. Phys.* **22**, 023015 (2020).
- [58] M. Malekakhlagh and E. Magesan, Mitigating off-resonant error in the cross-resonance gate, *Phys. Rev. A* **105**, 012602 (2022).
- [59] E. Magesan and J. M. Gambetta, Effective Hamiltonian models of the cross-resonance gate, *Phys. Rev. A* **101**, 052308 (2020).
- [60] M. Malekakhlagh, E. Magesan, and D. C. McKay, First-principles analysis of cross-resonance gate operation, *Phys. Rev. A* **102**, 042605 (2020).
- [61] 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).
- [62] 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 ZZ suppression, *Phys. Rev. Lett.* **127**, 130501 (2021).
- [63] K. X. Wei, E. Pritchett, D. Zajac, D. McKay, and S. Merkel, Characterizing non-Markovian off-resonant errors in quantum gates, [arxiv:2302.10881](https://arxiv.org/abs/2302.10881) (2023).
- [64] H. Ball, W. D. Oliver, and M. J. Biercuk, The role of master clock stability in quantum information processing, *npj Quantum Inf.* **2**, 16033 (2016).
- [65] F. Yan, S. Gustavsson, J. Bylander, X. Jin, F. Yoshihara, D. G. Cory, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Rotating-frame relaxation as a noise spectrum analyser of a superconducting qubit undergoing driven evolution, *Nat. Commun.* **4**, 2337 (2013).
- [66] 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).

- [67] J. Ku, X. Xu, M. Brink, D. C. McKay, J. B. Hertzberg, M. H. Ansari, and B. L. T. Plourde, Suppression of unwanted ZZ interactions in a hybrid two-qubit system, *Phys. Rev. Lett.* **125**, 200504 (2020).
- [68] Y. Ding, P. Gokhale, S. Lin, R. Rines, T. Propson, and F. T. Chong, in *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)* (IEEE Computer Society, Los Alamitos, California, 2020), p. 201.
- [69] R. Wang, P. Zhaob, Y. Jin, and H. Yu, Control and mitigation of microwave crosstalk effect with superconducting qubits, *Appl. Phys. Lett.* **121**, 15 (2022).
- [70] 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 Inf.* **7**, 14 (2021).
- [71] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf, ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field, *Phys. Rev. Lett.* **94**, 123602 (2005).
- [72] S. Sekiguchi, M.-J. Ahn, T. Mizutani, T. Saraya, M. Kobayashi, and T. Hiramoto, in *IEEE Journal of the Electron Devices Society* (2021), Vol. 9, p. 1151.
- [73] S. Gustavsson *et al.*, Driven dynamics and rotary echo of a qubit tunably coupled to a harmonic oscillator, *Phys. Rev. Lett.* **108**, 170503 (2012).
- [74] M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, Dynamics of superconducting qubit relaxation times, *npj Quantum Inf.* **8**, 132 (2022).
- [75] T. Thorbeck, A. Eddins, I. Lauer, D. T. McClure, and M. Carroll, Two-level-system dynamics in a superconducting qubit due to background ionizing radiation, *PRX Quantum* **4**, 020356 (2023).