Two-Fluxonium Cross-Resonance Gate

Ebru Dogan,[1] Dario Rosenstock,[1] Loïck Le Guevel,[1,2] Haonan Xiong,[3] Raymond A. Mencia,[3] Aaron Somoroff,[3] Konstantin N. Nesterov,[4] Maxim G. Vavilov,[4] Vladimir E. Manucharyan,[3] and Chen Wang[1,*]

[1] *Department of Physics, University of Massachusetts-Amherst, Amherst, Massachusetts 01003, USA*
[2] *Department of Electrical and Computer Engineering, University of Massachusetts-Amherst, Massachusetts 01003, USA*
[3] *Department of Physics, Joint Quantum Institute, and Center for Nanophysics and Advanced Materials, University of Maryland, College Park, Maryland 20742, USA*
[4] *Department of Physics and Wisconsin Quantum Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*

(Received 24 April 2023; accepted 7 July 2023; published 4 August 2023)

The superconducting fluxonium qubit has a great potential for high-fidelity quantum gates with its long coherence times and strong anharmonicity at the half-flux-quantum sweet spot. However, current implementations of two-qubit gates compromise fluxonium's coherence properties by requiring either a temporary population of the noncomputational states or tuning the magnetic flux off the sweet spot. Here we realize a fast all-microwave cross-resonance gate between two capacitively coupled fluxoniums with the qubit dynamics well confined to the computational space. We demonstrate a direct controlled NOT (CNOT) gate in 70 ns with fidelity up to $\mathcal{F} = 0.9949(6)$ despite the limitations of a suboptimal measurement setup and device coherence. Our results project a possible pathway towards reducing the two-qubit error rate near to or below 10^{-4} with present-day technologies.

DOI: 10.1103/PhysRevApplied.20.024011

I. INTRODUCTION

The development of superconducting circuits into a leading platform for scaling up quantum computation [1,2] over the past decade has been riding almost exclusively on the optimization of transmons [3]. The transmon has been entrenched as the go-to qubit, as it is easy to build with only one or two Josephson junctions, simple to model in the oscillator basis using perturbation theory, and robust to operate with minimal spurious degrees of freedom. However, it sacrifices anharmonicity, a fundamental quantum resource, for suppression of charge noise, and features a rather restricted parameter space. Recently, the fluxonium qubit [4] has emerged as a serious challenger to the monopoly of the transmon as the building block of a superconducting quantum processor [5,6], becoming the second type of superconducting qubit crossing the 99% fidelity threshold for two-qubit gates [5,7–9].

Moving forward, fluxonium qubits have the potential to outperform transmons in gate fidelity due to their inherent advantages of having both longer coherence times and higher anharmonicity. In particular, the lowest two energy levels of the fluxonium at the half-flux quantum, to be used as the computational states $|0\rangle$ and $|1\rangle$, enjoy substantial protection from dielectric loss due to their low transition frequencies [10] (typically 100 MHz to 1 GHz or lower [11] compared to typical transmons at 4–6 GHz), recently reaching a record 1 ms in coherence time [12]. Despite the low qubit frequency, the noncomputational transitions to higher excited states are in the range of several gigahertz, and this strong anharmonicity provides a large bandwidth and on-demand interactions to enable fast gate operations [13].

However, existing implementations of fluxonium two-qubit gates do not yet fully utilize these core advantages. Controlled Z (CZ) and controlled phase (CPHASE) gate schemes from Refs. [7,8] temporarily populate the higher excited states during the gate operation, and therefore are fundamentally limited by the faster decoherence rates of the plasmon transitions (which are in principle comparable to transmons). The flux-controlled *i*SWAP gates from Refs. [5,9] temporarily bring the fluxonium away from the half-flux "sweet spot" and are therefore susceptible to first-order flux noise in the same way as similar schemes for transmons. To fulfill fluxonium's full potential, several proposals have been put forward to carry out two-qubit gates within the high-coherence computational subspace at fixed frequencies [6,14,15], among which the

*wangc@umass.edu

cross-resonance (CR) controlled NOT (CNOT) gate [15] best leverages the fluxonium anharmonicity.

In the CR scheme, a control qubit is strongly driven at the resonance frequency of a (coupled) target qubit, leading to a CNOT operation or its equivalent. The CR gate has gained substantial popularity in transmon-based systems due to its all-microwave implementation and simplistic experimental requirements [16–22]. After a decade of experimental optimization and theoretical modeling [23–27], the two-transmon CR gate has reached state-of-the-art fidelity of up to 99.8% [28]. Despite this success, transmons are arguably poorly suited for the CR gate: in a nutshell, the CR interaction arises when the control qubit acts as a switchable microwave filter which regulates the amplitude and phase of the effective drive field arriving at the target qubit. The anharmonicity of the qubit is the crucial resource that dictates how strongly and how broadband this filtering effect depends on the control qubit state [23]. With limited anharmonicity, the optimized transmon CR gate has to navigate a fabrication-demanding straddling regime of qubit frequencies, and even then the gate speed is typically well over 100 ns [17,18], limited by leakage error under strong drives in a crowded spectrum. A scheme that relaxes this constraint to some degree is demonstrated in Ref. [29] where a cross-resonance gate is realized between a transmon and a *C*-shunt flux qubit, which offers an improved frequency window for efficient CR interaction.

In this work, we show that the fluxonium qubits are ideally suited for cross-resonance interactions by functioning as broadband quantum switches with high *on*:*off* ratio for external control fields. Following the proposal of the selective-darkening CR scheme [15], we demonstrate a direct CNOT operation between capacitively coupled fluxonium qubits with gate fidelity above 99.4% from interleaved randomized benchmarking ($\mathcal{F} = 99.49(6)\%$ averaged over 7 h, 99.42(6)% averaged over four separate trials spanning two months). The CNOT gate is implemented in a ZZ-canceled two-qubit device with a gate time of 70 ns and minimal pulse shaping. The current gate performance is limited by a severe microwave reflection problem of our drive lines (mitigated by a cancelation technique to be discussed in Appendix A) and lower-than-expected coherence times in our setup, which can be substantially improved in the near future.

II. FLUXONIUM AS A QUANTUM SWITCH

The working principle of the fluxonium CR gate can be conceptually illustrated in the bare single-qubit level diagram in Figs. 1(a) and 1(b). We consider two capacitively coupled fluxonium qubits, where the low-frequency control qubit A is driven at the moderately higher frequency ($\omega_B/2\pi \approx 1$ GHz in our experiment) of the target qubit B. The Hamiltonian is

$$\hat{H} = \hat{H}_A + \hat{H}_B + J_C \hat{n}_A \hat{n}_B + \hat{H}_{\rm cr} + \hat{H}_{\rm dr}, \tag{1}$$

where the individual bare fluxonium Hamiltonian ($\alpha = A, B$),

$$\hat{H}_\alpha = 4E_{C,\alpha}\hat{n}_\alpha^2 + \tfrac{1}{2}E_{L,\alpha}\hat{\varphi}_\alpha^2 - E_{J,\alpha}\cos(\hat{\varphi}_\alpha - \phi_\alpha^{\rm ext}), \tag{2}$$

is written in terms of the dimensionless flux ($\hat{\varphi}_\alpha$) and charge ($\hat{n}_\alpha$) operators and the charging ($E_{C,\alpha}$), inductive ($E_{L,\alpha}$), and Josephson ($E_{J,\alpha}$) energies. The capacitive interaction between the two fluxoniums is described by the coupling energy J_C. The last two terms in Eq. (1) describe the aforementioned cross-resonant drive, $\hat{H}_{\rm cr} = \varepsilon_A \hat{n}_A \cos\omega_B t$, and a small additional drive applied directly to qubit B, $\hat{H}_{\rm dr} = \varepsilon_B \hat{n}_B \cos\omega_B t$.

The highly anharmonic control qubit A functions like a microwave switch due to its very state-dependent response to the off-resonance CR drive. When the control is in $|0\rangle_A$, qubit B receives an effective drive field mediated by the virtually excited $|0\rangle_A$-$|1\rangle_A$ transition [Fig. 1(a)], with a resonant Rabi rate

$$\Omega_0 \approx \varepsilon_A \left[i\frac{J_C}{\hbar}\frac{\langle 0|\hat{n}_A|1\rangle^2}{\omega_B - \omega_A}\right]\langle 0|\hat{n}_B|1\rangle, \tag{3}$$

where the part in square brackets can be understood as a "transmission factor" of the CR drive, while the matrix element $\langle 0|\hat{n}_B|1\rangle$ factors in the response function of B to any electric drive field it receives. The transmission factor flips sign when qubit A is excited to $|1\rangle_A$, giving the original cross-resonance ZX Hamiltonian for the ideal spin-1/2 system [16]. This ZX effect is suppressed by the small charge matrix element of the low-frequency fluxonium transitions.

However, the $|1\rangle_A$ state also opens a stronger pathway of transmission via the virtual noncomputational $|1\rangle_A$-$|2\rangle_A$ transition [Fig. 1(b)]. Even though this transition is several gigahertz detuned, $\omega_{A,12} \gg \omega_A, \omega_B$, its presence is felt strongly since the process doubly benefits from its much larger matrix element $\langle 1|\hat{n}_A|2\rangle$, giving

$$\Omega_1 \approx \varepsilon_A \left(-i\frac{J_C}{\hbar}\right)\left[\frac{\langle 0|\hat{n}_A|1\rangle^2}{\omega_B - \omega_A} + \frac{\langle 1|\hat{n}_A|2\rangle^2}{\omega_{A,12} - \omega_B}\right]\langle 0|\hat{n}_B|1\rangle. \tag{4}$$

This extra term describes an enhanced CR effect beyond the spin-1/2 system, which is known to provide up to a factor of 2 boost for the transmon CR rate in the straddling regime but requires precise frequency placements and may easily cause leakage error [23].

The situation for fluxonium is fundamentally different. This additional transmission factor contains no small parameters related to either qubit frequencies and hence can be the dominant enabler of fast gates for low-frequency qubits. It leads to a conditional rotation of qubit B, and, by applying a small compensation drive directly on B ($\hat{H}_{\rm dr}$) to fully cancel out its dynamics in the *off* state (known as selective darkening [30]), a direct CNOT gate can be

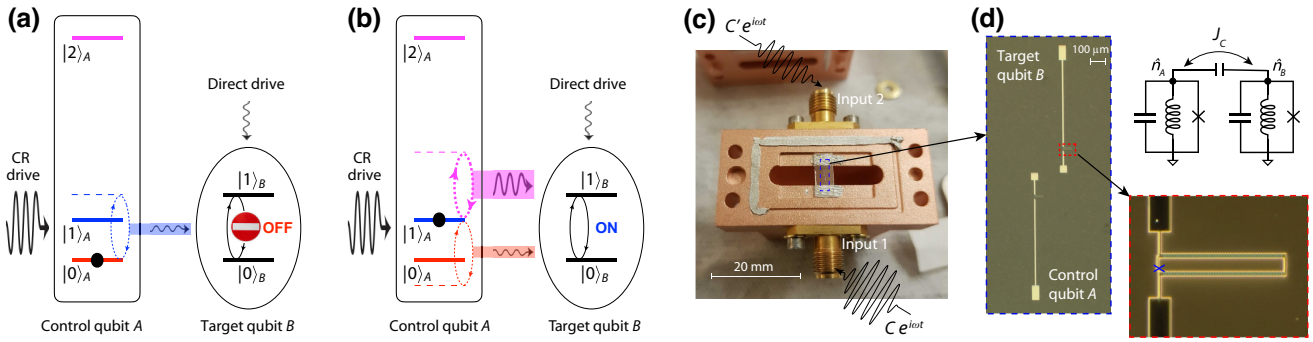

FIG. 1. (a),(b) The mechanism of the fluxonium cross-resonance effect, illustrated in the relevant bare energy-level diagram of two fluxonium qubits. A large CR drive and a small direct drive at ω_B are applied simultaneously to the two qubits. (a) The *off* state of the fluxonium switch. When A is in $|0\rangle$, the relatively weak $|0\rangle$-$|1\rangle$-mediated transmission of the CR drive (blue) is canceled by the direct drive. (b) The *on* state of the fluxonium switch. When A is in $|1\rangle$, the stronger $|1\rangle$-$|2\rangle$-mediated transmission of the CR drive (magenta) is activated. The $|0\rangle$-$|1\rangle$-mediated drive also flips its sign (red) and adds constructively to the total CR effect. Note that the $|0\rangle$-$|2\rangle$ transition is forbidden at half-flux. (c) Photo of half of the cavity with two input ports, where the sapphire chip with two fluxonium qubits is located in the middle. (d) Photo of the two fluxonium qubits, each made of an extended dipole antenna (capacitor), a junction chain superinductor (see the zoomed-in dark-field image), and a small Josephson junction (at the position marked by the blue cross). The dipole of qubit A is extended towards input port 1 and hence more strongly coupled to it, and vice versa for qubit B.

realized. In this discussion, we have omitted the (non-negligible) $|0\rangle$-$|3\rangle$ contribution and non-RWA(rotating wave approximation)-like hybridizations; a more rigorous calculation has been carried out in Ref. [15]. Crucially, in contrast to previous microwave-based two-qubit gates that necessarily populate the noncomputational states [7,8], here the occupation probability of the $|2\rangle_A$ state is minimal since it scales inverse quadratically with the multigigahertz drive detuning (while Ω_1 scales inverse linearly).

III. DEVICE SETUP AND EXPERIMENTAL METHODS

While the local drives required in the CR scheme may fit more naturally with planar architectures, here we use a convenient three-dimensional (3D) circuit quantum electrodynamics (CQED) design [31] to carry out a proof-of-principle demonstration. Two capacitively coupled fluxoniums are fabricated on a sapphire substrate and enclosed in a copper cavity, and the Hamiltonian parameters extracted from the two-tone spectroscopy are listed in Table I. Compared to a similar device in Ref. [7], our cavity has two drive ports placed on the opposite sides of the sapphire chip. Each port has stronger coupling to one of the qubits due to the asymmetric layout of the fluxoniums on the chip, which gives us spatial selectivity to apply drives to the two-fluxonium circuit [Figs. 1(c) and 1(d)].

A large external superconducting coil is used to apply a static global magnetic field to the circuit. We carry out our experiment at a fixed bias field giving $\phi_A^{\rm ext}/2\pi = 0.5006$ and $\phi_B^{\rm ext}/2\pi = 0.4990$, when both qubits are within 0.7 MHz from their exact half-flux sweet spots. The coherence times for both fluxoniums at this operating point are noted in Table I, which are primarily not limited by flux noise (i.e., insensitive to the exact choice of external flux in the vicinity of this operating point). We attribute the subpar T_1 times to a combination of higher dielectric loss and insufficient IR shielding compared to the current state of the art [12]. Both low-frequency noise and cavity photon shot noise contribute to the low T_2 times.

A. ZZ cancelation, initialization, and readout

The relatively strong capacitive coupling between the two fluxoniums results in a static ZZ interaction $\xi_{ZZ}^0/2\pi = 0.9$ MHz. While ξ_{ZZ}^0 does not limit the fidelity of the direct CNOT gate thanks to its selective-darkening construction [15], it affects other gates and poses crosstalk challenges in scaling up multiqubit systems. Throughout our experiment, we apply an off-resonant continuous microwave tone (at 3.850 GHz, 54 MHz detuned from the $|11\rangle$-$|21\rangle$ transition) at all times to cancel the ZZ interaction by differential ac Stark shift (giving residual $\xi_{ZZ}/2\pi < 20$ kHz) [8]. This ZZ cancelation pump results in an estimate of a total of 2%–3% hybridization between computational states and higher levels, and the impact on qubit coherence is experimentally very minimal. Future studies may further employ multipath or tunable couplers to suppress ξ_{ZZ} [6,32] if needed. For the rest of the paper, we will work exclusively with the dressed eigenstates under the pump tone, which are labeled $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, and form the computational basis of the two-qubit system.

We initialize our qubits in the $|00\rangle$ state with an estimated fidelity of about 94% using a cavity-sideband cooling procedure. This protocol effectively dumps the entropy of the low-frequency qubits to the cold bath of the high-frequency cavity over a duration of 15 μs. Because of the lack of a quantum-limited parametric amplifier in

TABLE I. Device parameters. The qubit frequencies and the (range of daily-averaged) coherence times are measured at the operating flux point of the experiment. The frequency pairs for 1–2 transitions correspond to the cases with the other qubit being in ground (left) and excited (right) state. The circuit Hamiltonian parameters and transition matrix elements are extracted from fitting the qubit spectroscopy data.

Qubit (α)	E_L/h (GHz)	E_C/h (GHz)	E_J/h (GHz)	$\omega_\alpha/2\pi$ (GHz)	$\omega_{\alpha,12}/2\pi$ (GHz)	J_c/h (GHz)	$\lvert\langle 0\rvert\hat{n}_\alpha\lvert 1\rangle\rvert$	$\lvert\langle 1\rvert\hat{n}_\alpha\lvert 2\rangle\rvert$	T_1 (μs)	T_2^* (μs)	T_{2E} (μs)
A	0.78	1.18	4.03	0.5552	3.719/3.796	0.28	0.13	0.55	52.0–60.0	14.0–15.5	22.0–24.0
B	1.42	1.13	4.34	1.0045	3.691/3.610		0.20	0.59	17.0–33.0	5.5–7.0	13.0–16.5

our experiment, the joint two-qubit state is measured by analyzing the averaged readout transmission signal with different combinations of qubit prerotations preceding the measurement [33]. See Appendices C and D for more details of state initialization and readout.

B. Characterization of cross-resonance dynamics

To realize controlled operation of qubit B, we apply microwave drives at ω_B simultaneously to both physical input ports. The two drives are produced by separate in-phase and quadrature modulation of the same local oscillator source. Therefore, they are phase-locked from each other but have independently controllable complex amplitudes C and C'. The CR drive amplitude ε_A and direct drive amplitude ε_B locally incident on the two qubits are linearly related to C and C' by a complex-valued 2×2 "classical crosstalk" matrix, which we do not need to explicitly characterize. We simply find a complex ratio $\eta = C'/C$ experimentally to darken the $|00\rangle$-$|01\rangle$ transition, so that the combined effect of $\hat{H}_{\rm cr}$ and $\hat{H}_{\rm dr}$ gives a rotating-frame effective drive Hamiltonian in a block-diagonal form (written in the form of $A \otimes B$),

$$\hat{H}_{\rm dr} = \frac{\Omega}{2}|1\rangle\langle 1| \otimes \sigma_x + \frac{\Delta_s}{2}\sigma_z \otimes \mathbb{I}, \tag{5}$$

where the first term denotes a conditional X rotation of the target qubit (equivalent to $(\Omega/4)(\mathbb{I} - \sigma_z) \otimes \sigma_x$) and the second term represents an ac Stark shift Δ_s on the control qubit.

To calibrate the complex drive ratio η, we initialize A in $|0\rangle$ and null the Rabi oscillation of B by sweeping the relative amplitude and phase of the drives. After fixing η, we still retain the freedom of choosing C, or the overall amplitude scale and common phase of the two drives, which endows full control of the $|10\rangle$-$|11\rangle$ subspace. When A is initialized in $|1\rangle$, B displays the prototypical Rabi dynamics, as shown in Fig. 2(b) over a range of drive detunings for a given drive power. We can further reconstruct the Bloch-sphere trajectory of qubit B under the

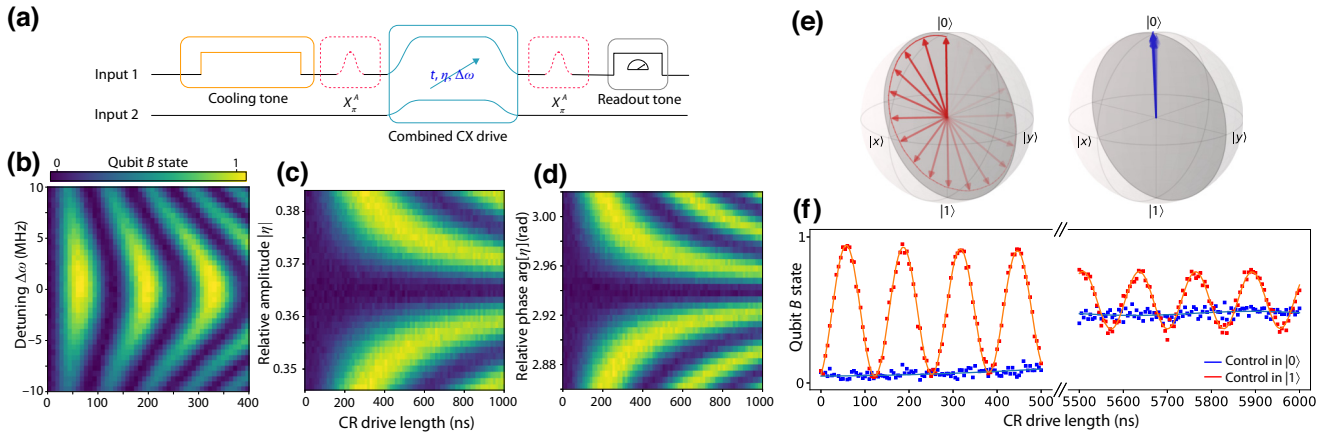

FIG. 2. (a) Pulse sequence for measuring the controlled driven dynamics of qubit B. Here the first X_π^A pulse is applied (skipped) to set the control = *on* (*off*) state, and the second X_π^A pulse is applied (skipped) to reset A to $|0\rangle$ for consistent readout contrast. (b) Rabi oscillation of B at the *on* state as a function of time and drive detuning under the optimized complex drive ratio η for the CR gate. (c),(d) Dynamics of B at the *off* state when η deviates in (c) amplitude and (d) phase from the perfect darkening condition. The ratio η is in arbitrary units, as it includes the difference in attenuation and electrical delay of the two drive lines. (e),(f) Conditional Rabi oscillation of B under resonant CR drive, comparing its dynamics at the *on* (red) and *off* (blue) states. Panel (e) shows the tomographically reconstructed Bloch sphere trajectory of qubit B over the first oscillation period of panel (f). The loss of oscillation contrast over long time is consistent with qubit coherence times. All data are taken with a combined drive amplitude that gives an estimated effective CR drive strength $\epsilon_A\langle 0|\hat{n}_B|1\rangle$ of 90 MHz.

CR drive by performing single-qubit state tomography at different times for the control qubit in $|0\rangle$ and $|1\rangle$, respectively [Fig. 2(e)], which demonstrates that the desired Hamiltonian equation (5) has been realized. We can also vary the conditional Rabi rate Ω as a function of time using a drive envelope $C(t)$, and any envelope that gives an integrated rotation angle $\int \Omega(t)dt = \pi$ would yield a CNOT-equivalent controlled X_π rotation.

C. Single-qubit control and pulse correction

We use a similar calibration procedure to obtain unconditional single-qubit rotations. Since microwaves applied to either input ports contribute to both $\hat{H}_{\rm dr}$ and $\hat{H}_{\rm cr}$, but at different ratios, to realize clean single-qubit rotation of B independent of A, we apply microwave drives at ω_B to both ports simultaneously with a complex drive ratio η' to null the $\sigma_z \otimes \sigma_x$ term of the drive Hamiltonian (see Appendix E). Because the CR effect from the higher-frequency B to the lower-frequency A is weaker for our device parameters, in practice we did not find it necessary to pursue the compensated drive scheme for A and simply used port 1 to drive it.

Fast high-fidelity gates in general require impedance-matched transmission lines free of standing waves. As gate fidelity improves, impedance mismatch can start to cause appreciable harm, the characterization and mitigation of which may require careful studies. Accidentally, we had to carry out the experiment under a challenging condition with extreme reflection problems in our drive lines. Each control pulse bounces off the cavity multiple times, among which the most pronounced secondary impact arrives about 20 ns later than the original pulse and carries $\sim 35\%$ of the original amplitude. We developed an *ad hoc* procedure to calibrate the timing and complex amplitude of the reflected pulses, which extends the method in Ref. [34] to both I and Q quadratures (see Appendix A). By programming waveform predistortion to all control pulses to cancel the reflections throughout our experiment, we improved single-qubit gates from a completely meaningless mess to average fidelity of $\mathcal{F} \gtrsim 99.8\%$ for A and $\mathcal{F} \gtrsim 99.7\%$ for B, as measured by simultaneous randomized benchmarking (see Appendix E). For both qubits, each single-qubit gate uses a 16 ns pulse with a Gaussian envelope ($\sigma = 4$ ns). The fidelity remains lower than the qubit coherence limit, which we attribute to the imperfect waveform correction.

IV. PERFORMANCE OF THE CROSS-RESONANCE GATE

A. Gate tune-up and calibration

Under the CR drive Hamiltonian [Eq. (5)], the dynamics in Fig. 2(e) provides a coarsely tuned controlled X_π rotation (CX_π) after a half Rabi period. To bring up a high-fidelity two-qubit gate, we fine-tune the CR drive parameters and account for extra single-qubit phases (θ_A and θ_B) with rotating frame updates (known as virtual Z rotations [35]) to realize the CX_π gate:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\theta_B} & 0 & 0 \\ 0 & 0 & 0 & -ie^{i\theta_A} \\ 0 & 0 & -ie^{i\theta_A} & 0 \end{pmatrix} \rightarrow \mathrm{CX}_\pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \end{pmatrix}, \tag{6}$$

where the top-left and bottom-right blocks correspond to the $|0\rangle_A$ and $|1\rangle_A$ subspaces, respectively. The CX_π gate is connected to the textbook CNOT gate by an S gate on A, which can be trivially absorbed in θ_A at no additional cost. We targeted CX_π instead of CNOT in our calibration and verification routine to emphasize the underlying continuous dynamics of the gate (which is a conditional rotation). For all purposes, our result can be viewed as applicable for a CNOT gate.

Our CX_π gate uses a rounded square pulse envelope typically with 6 ns half Gaussian ($\sigma = 3$ ns) rising and falling edges. At a fixed gate time, we iterate through dedicated subroutines sensitive to specific control errors to calibrate against seven parameters: the relative amplitude and phase (complex η) of the CR drive to ensure selective darkening of the $|00\rangle$-$|01\rangle$ transition; the common amplitude and phase (complex C) and the detuning of the CR drive to ensure a precise X_π rotation in the $|10\rangle$-$|11\rangle$ subspace; the relatively large single-qubit phase θ_A due to the ac Stark shift [$\sigma_Z \otimes \mathbb{I}$ term in Eq. (5)]; and a small phase θ_B possibly due to a spurious $\mathbb{I} \otimes \sigma_Z$ Hamiltonian from a higher-order CR effect [15]. These seven parameters cover all possible control errors of CX_π within its block-diagonal structure. We further monitor possible spurious rotations of A (i.e., leakage drive on A) to ensure the process is block-diagonal. See Appendix F for details of the calibration procedure. A similar routine has been described in a recent report of the CR gate in transmons [28].

B. Gate characterization

We use QPT and IRB to characterize the performance of the CR gate. Figure 3 shows the result for our calibrated CR gate at an optimal gate length of $t_g = 70$ ns. Our process tomography follows the procedure outlined in Ref. [36]. The reconstructed process matrix is in excellent agreement with the ideal CX_π gate, with no outstanding spurious nonzero elements, qualitatively confirming the performance of the gate. The extracted process fidelity is 99.1%, although the QPT fidelity is known to be sensitive to the underlying model of rescaling state preparation and

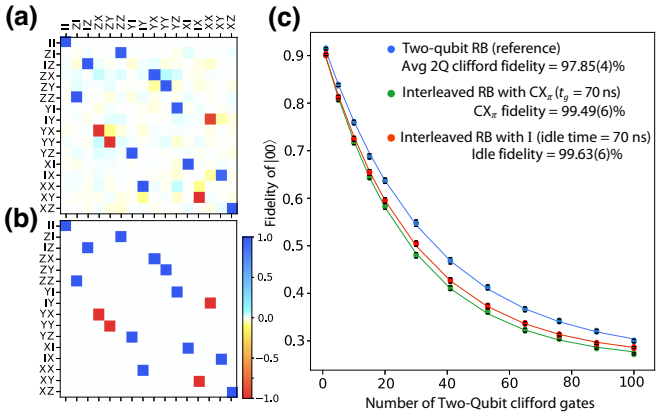

FIG. 3. (a) The reconstructed process matrix of our CR gate calibrated for a total gate length of 70 ns (including two 6-ns Gaussian ramp edges) by quantum process tomography (QPT). (b) The ideal process matrix of CX_π for comparison. (c) Interleaved randomized benchmarking (IRB) of the CR gate. The plot is averaged over 60 randomly generated Clifford sequences per data point over a continuous run of 7 h. The reference error per two-qubit Clifford (2.15%) is consistent with the average number of physical single-qubit gates (6.5) and two-qubit gates (1.5) per Clifford and their fidelity. IRB of a 70-ns idling gate provides an estimate of the decoherence error.

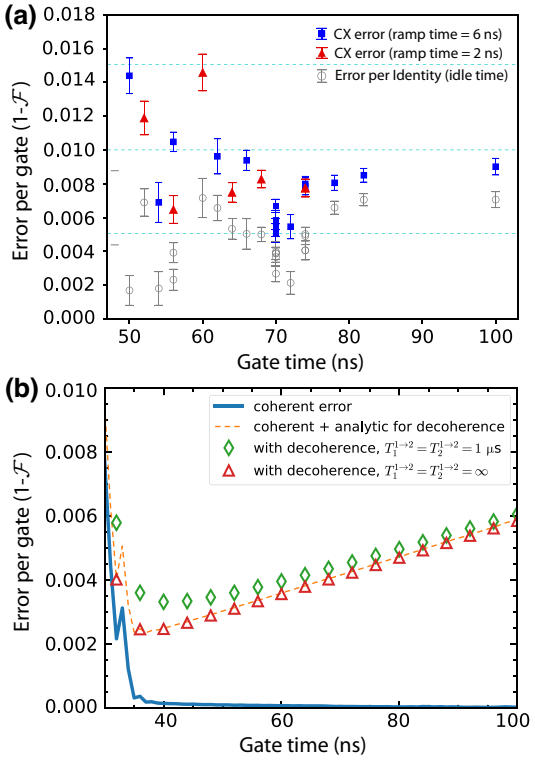

FIG. 4. (a) The two-qubit CX_π gate error extracted from IRB versus gate time for rounded square pulses with ramp times of 6 ns (blue) and 2 ns (red). Also plotted for comparison is idling error from IRB (grey). (b) Numerical simulation results of coherent error per gate (solid blue curve), total error per gate including decoherence of computational states (red triangles), and total error per gate further including decoherence of the noncomputational $|2\rangle$ states (green diamonds). The decoherence rates of the $|2\rangle$ states are exaggerated ($T_1 = T_2 = 1$ μs) to demonstrate that their effect is very minimal. The decoherence rates of the computational states in the simulation use T_1 and T_{2E} values in Table I, whose contribution to gate error is in good agreement with the analytical estimate of t_g/T_{err} (orange dashed line). The simulation follows the numerical procedure in Ref. [15] and is based on the fluxonium parameters from the experiment and considers rounded square pulses with ramp time of 6 ns.

measurement (SPAM) infidelity. The IRB provides a more SPAM-agnostic validation of the CR gate fidelity. Following the IRB procedure [20,37], we compare the sequence fidelity versus sequence length for random two-qubit Clifford gates (reference) and those interleaved with additional CX_π gates, giving the CX_π gate fidelity $\mathcal{F} = 99.49(6)\%$ [Fig. 3(c)]. See Appendix F for more details.

C. Gate speed and pulse shape

We calibrated and benchmarked the CX_π gate at different gate times in the 50–100 ns range, and the IRB fidelity is shown in Fig. 4(a). We achieved CR gate fidelity above 99% over a broad range of gate times, and the fastest well-performing gate takes only 54 ns ($\mathcal{F} \approx 99.3\%$). This is a notable speedup from the transmon CR gates, which have been typically in the range of 150 ns or longer [17,18,22] and only very recently reaching a record of 90 ns [28]. We also benchmarked the fidelity of idling gates with different (idling) time using IRB, which has been used as a sensitive noise spectrometer to probe qubit decoherence on the same time scale as fast gate operations [38]. However, we observed unusual nonmonotonic behavior in idling fidelity which cannot be explained by any Markovian noise models. While part of the fluctuations may be attributed to qubit coherence time fluctuations over the two months of data acquisition, we believe this observation is a clear evidence that the residual pulse reflection problem remains a major contributing factor in our gate error and is responsible for the unexpected peaks and valleys of the CR gate fidelity with respect to gate time.

The optimal choice of the gate time in principle is determined by the tradeoff between the coherent control error at the short-time limit and the incoherent error at the long-time limit. Numerical simulation of the CR gate performance [15] using our device parameters illustrates this tradeoff clearly [Fig. 4(b)]: the unitary error decreases fast with increasing gate time t_g, while the incoherent error increases linearly in time as t_g/T_{err}, where $T_{\text{err}}^{-1} = (T_{1,A}^{-1} + T_{1,B}^{-1} + 2T_{2E,A}^{-1} + 2T_{2E,B}^{-1})/5$ [23]. The simulation suggests further room to speed up the gate to the 40 ns range for optimal performance in a system free of reflection problems.

Another major advantage of highly anharmonic qubits like the fluxonium is the relaxed requirements on pulse shaping. The transmons CR gates often require careful pulse shaping and slow ramp times (at least 15 ns [18,28]) to avoid driving unwanted transitions. The only constraint in pulse shapes in our fluxonium CR gate is to avoid off-resonant excitation of the qubit A (hence limited by qubit-qubit detuning), and once again the small matrix element of the low-frequency qubit A works in our favor. We further tested reducing the pulse ramp from 6 ns to 2 ns, which is faster than the bandwidth of our arbitrary waveform generator (AWG; 350 MHz, with 1 ns digital resolution), and observed no apparent impact in CR gate fidelity (Fig. 4(a)).

V. OUTLOOK

Compared to flux-controlled two-qubit gates, the all-microwave cross-resonance gate has been historically considered a gate that sacrifices speed for simplicity and noise protection. This fundamental tradeoff remains front and center in today's competition to scale with transmon-based architectures. Armed with large anharmonicity and favorable selection rules (i.e., hierarchy of matrix elements), the fluxonium qubits have opened a new dimension for engineering the CR effect and are poised to make the best of both worlds. Among the first generation of two-fluxonium gates that have emerged recently, our CR gate is on par in speed with the flux-controlled gates [5,9] without the need of extra coupler elements or fast flux biasing.

The fluxonium CR gate is expected to well preserve the coherence properties of the computational subspace at half-flux. Although the lingering pulse reflection problem prevented us from calibrating control errors much below the reported gate infidelity, our best gate at 70 ns performs very close to the free-evolution coherence limit [see Fig. 4(b)], showing no hidden obstacles (such as drive-induced instability or decoherence) at least at the level of 10^{-3}. One can be optimistic of rapid fidelity improvement from the current level by unlocking the expected intrinsic fluxonium coherence via better IR shielding and thermalization [10,12]. Assuming the modest capacitive quality factor $Q = 6 \times 10^5$ as in Ref. [12] and that $T_{2E} \approx T_1$, the two qubits at their current working frequency would have coherence times of 400 and 200 μs, respectively, projecting a coherence-limited gate error rate of 2×10^{-4} at 50 ns gate time. We note that modern transmons, with better materials and fabrication processes [39,40], have achieved Q factors 10 times better than this assumption, suggesting substantial room for further improvement in dielectric quality that would reduce the coherence-limited gate error below 10^{-4}. While it remains an open question whether the fluxonium development can follow the same path to close the gap in capacitive Q factor, at least one device in earlier literature hinted at $Q \simeq 3 \times 10^6$ [41]. In addition, the Hamiltonian parameters of our fluxonium qubits are by no means optimal, and there is a large design space to explore optimization strategies.

The selective-darkening CR effect of fluxonium qubits can be more generally viewed as a quantum switch for a broad bandwidth of external control fields. This allows the fluxonium to function as the control qubit for multiple neighbors across a large frequency span when the CR scheme is implemented in multiqubit arrays. Transmon-based architectures with all-microwave control face a severe challenge in frequency crowding that not only demands unprecedented precision in fabrication [42] but also constrains the degree of lattice connectivity (which is a key motivation behind the development of the heavy-hexagonal code [43]). CR gates using fluxonium provide a potential route to combine the simplicity of all-microwave control with favorable connectivity for scalable architectures. For example, recent works [6,44] investigated surface-code lattices based on the fluxonium CR gate, which showed the substantial design freedom afforded by this gate scheme.

As a broadband quantum switch, the fluxonium CR effect provides the means not just for a single and discrete two-qubit gate (CNOT), but also for a broader class of controlled unitary operations. For example, by modifying the phase and detuning of the CR drive, one can efficiently implement controlled rotations of the target qubit around arbitrary axes. Such controlled operations may be further extended to manipulating multimode bosonic or multiqubit systems. Since controlled unitarity is one of the most important building blocks for many quantum algorithms and subroutines (e.g., phase estimation), native parameterized controlled unitary operations will likely bring benefit to noisy intermediate-scale quantum (NISQ) era applications.

ACKNOWLEDGMENTS

We thank Quentin Ficheux, Long Nguyen, and Joseph Bardin for helpful discussions. This research was supported by the ARO-LPS HiPS program (No. W911-NF-18-1-0146). V.E.M. and M.G.V. acknowledge the Faculty Research Award from Google and fruitful conversations with the members of the Google Quantum AI team. L.L.G. acknowledges support from Google.

APPENDIX A: ACTIVE REFLECTION CANCELATION VIA PULSE PREDISTORTION

Within our current fridge setup, an impedance mismatch in the signal transmission chain severely impedes the integrity of the microwave pulses at the two qubit frequencies. Clean microwave pulses played into our qubit input lines generate several reflections hitting the qubit at a later time (on the order of tens of nanoseconds) with different amplitude (up to 35% of the original

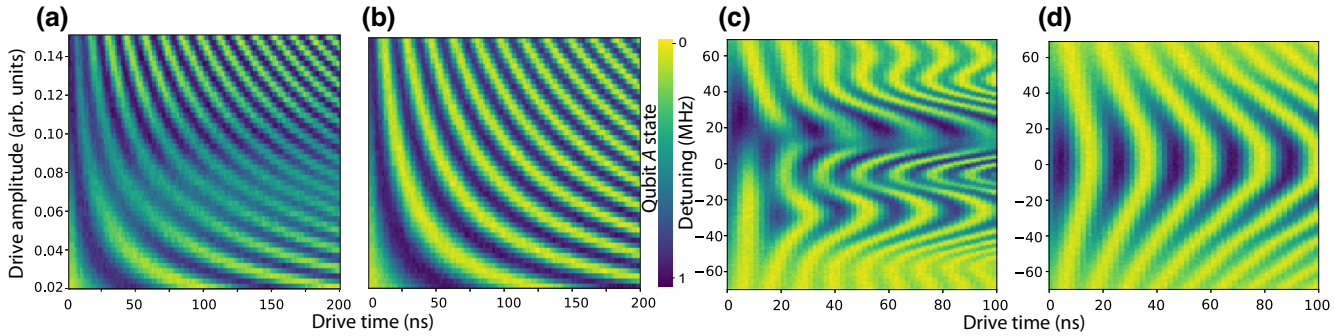

FIG. 5. (a),(b) Rabi oscillation of qubit A as a function of drive time and amplitude under a single-port drive (a) without and (b) with the predistortion. Problematic contrast loss is resolved and the correct oscillation frequency is recovered for (b). (c),(d) Rabi oscillation of qubit A as a function of drive time and drive detuning (the "chevron pattern") for a single-port drive (c) without and (d) with the predistortion, driven at an amplitude corresponding to the π pulse used throughout the experiment. The time axis does not include the ramp time of the pulse, which is 6 ns for each edge. The reflection problem applies to both qubits and both drive lines: similar traces are observed for all four combinations.

amplitude) and phase, leading to severe deviation from the intended rotation. More critically, the pulse reflections overlap with subsequent pulses, resulting in unpredictable behavior of gate sequences such as contrast loss in power or time Rabi and strongly distorted chevron patterns [see Figs. 5(a) and 5(c)]. We attribute the higher-than-usual reflections to a noncryogenic directional coupler placed at the mixing chamber with its coupled port connected to the cavity through a few Eccosorb filters [see Fig. 16(b)]. The unmatched nature of the cavity inputs combined with the coupled port load deviating from the ideal 50-Ω value at cryogenic temperature and the absence of attenuation on the reflection path result in this critical situation.

We mitigated this issue by applying a simple predistortion model to the qubit pulses. Our model consists of duplicating the ideal pulse into a train of delayed pulses whose timing, amplitudes, and phases are chosen to overlap and cancel the mismatch-induced reflections at the qubit port [Fig. 6(a)]. All corrections are applied in the background and do not affect the effective qubit gate times.

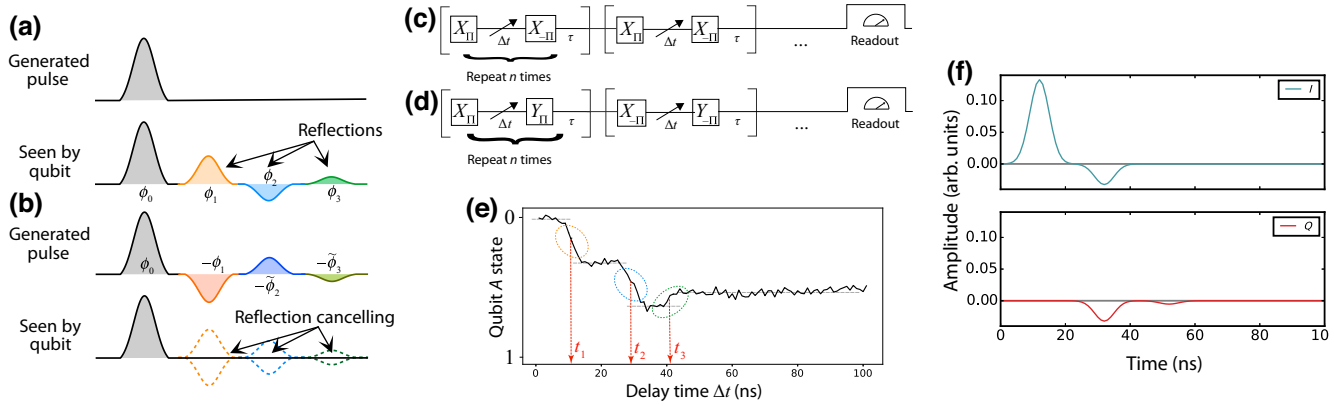

FIG. 6. (a),(b) Illustration of the generated pulse envelope and resulting envelope at the qubit port (a) without any correction and (b) with predistortion. The phases for reflections in (b) are not the same as in (a), as each correction implemented creates its own reflections that change the phase of the ones following it. (c),(d) The pulse sequences for characterizing the (c) quadrature-phase and (d) in-phase components of reflections. We sweep over the delay time Δt for a given number of repetitions n. Here, τ is a time long enough (80 ns) for reflections from each set to not overlap with the next set or the readout. We start with $n = 1$ and proceed by increasing up to $n = 7$ for further amplification of the features. For the sequence in (d), the sign of the π pulse changes for each iteration. (e) An example trace from the sequence from (c) ($n = 1$) for characterizing the quadrature-phase reflection. Each jump that makes the qubit deviate from its previous state can be attributed to an additional reflection. The delay times with which these reflections follow the original pulse can be pinned down by summing the X_π pulse length with times t_i pointed out with the red arrows shown. For both in-phase and quadrature-phase reflection characterizations, we use X_π pulses as fast as 8 ns. (f) The set of envelopes that would need to be played from AWG I and Q channels in order to have a clean X_π rotation for qubit A.

1. Reflection characterization

Pulse distortions due to impedance mismatch issues are not unknown to the superconducting qubit community but mainly arise in fast flux lines driving a low-impedance inductor, and the suggested pulse predistortion schemes mostly target the flux drive lines [45,46]. Our case of microwave pulse reflection in our qubit drive lines is more relatable to an earlier work [34] from which we make use of a pulse sequence characterizing the Q-quadrature reflection and extend it to our own model.

We start by defining a reflection with respect to the main pulse by three parameters: its delay time, relative amplitude, and relative phase. More conveniently in the qubit context, the reflections in the I-Q plane are defined with the relative in-phase and relative quadrature-phase components. Equivalently, we can characterize these reflection parameters independently with relative amplitudes for in-phase and quadrature-phase components at their relevant delay times.

The first step of the characterization is pinning down the delays of these reflections. An intuitive way of extracting them is by varying the delay between the following two rotations: X_π and $X_{-\pi}$. This is a sequence that isolates and accumulates the quadrature-phase component for the reflections while cancelling the in-phase component [34]. For short delay times where the reflections for X_π arrive after the pulse $X_{-\pi}$ with no overlaps, all reflections due to X_π should be canceled out by the reflections of $X_{-\pi}$: the measurement should result in the ground state. As the delay increases, reflections of X_π start overlapping with the pulse $X_{-\pi}$. For sufficiently short pulses, the result is the steplike evolution of the qubit state where we can clearly see the effects of individual reflections [see Fig. 6(e)]. It is possible to identify each reflection from such a trace, as each of the "jumps" can be attributed to an additional reflection hitting the qubit. The points with the highest derivatives on these traces can be pinned down to determine the delay times for these reflections.

We then use the same experiment with our targeted gate length and adjust the quadrature reflection component associated with each delay until the qubit state remains in ground for all delays, a sign of a good predistortion canceling the reflections at the qubit ports. The X_π and $X_{-\pi}$ pairs are repeated up to seven times, amplifying the deviation from ground, in a single measurement to finely tune the quadrature components [Fig. 6(c)].

The in-phase component is tuned in a similar measurement consisting of varying delay between X_π and Y_π rotations [Fig. 6(d)]. The pulse sequence for the in-phase component characterization is sensitive to both in-phase and quadrature components. It is beneficial to have a reasonably well-tuned cancelation of the quadrature component before moving on with the in-phase component. We then need to proceed with an iterative modification of both

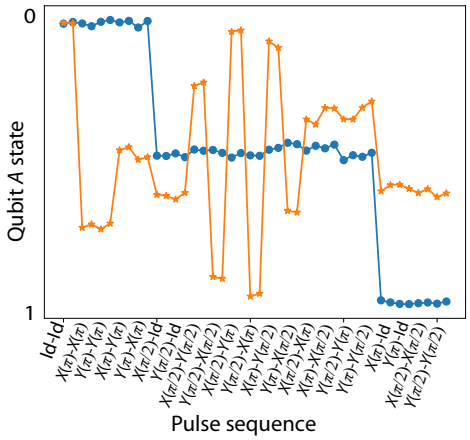

FIG. 7. AllXY experiment [47] for the qubit A without (orange) and with (blue) the reflection cancelation scheme after the same set of tune-up procedures. The AllXY experiment consists of three sets of sequences: the first set of points land in the ground state, the second in the equator, and the third in the excited state. Pulse errors will show up as deviations from this configuration.

I and Q of each reflection, until both pulse sequences result in the designated final states with no significant deviations.

Both qubits' drives require this protocol to be implemented, with the case of qubit B being more challenging than qubit A for being a combination of two individual drives. A sample envelope played from the AWG for an X_π for the qubit A can be seen from Fig. 6(f). Rather than a single Gaussian on the I channel for a π rotation around the X axis, a set of two envelopes modified for reflection cancelations are played on both I and Q channels. We depend on this predistortion scheme for having meaningful and reliable operations in our system. An "AllXY" experiment [47] as a visual indicator of the level of improvement can be seen in Fig. 7.

APPENDIX B: CALIBRATION OF INITIAL STATE POPULATION WITH AN ENTANGLING GATE

Without a single-shot readout, we take extra steps to calibrate the initial ground-state population of the qubits, i.e., to measure our state initialization fidelity. This may be accomplished by making use of higher transitions. Comparing the amplitudes of $|1\rangle$-$|2\rangle$ oscillations with and without population inversion [48] would provide the information about the initial state populations. However, this process involves extra complexity for a two-fluxonium system whose $|1\rangle$-$|2\rangle$ transitions have frequency and readout visibility strongly dependent on the state of the other qubit. In order to conveniently measure and track our qubit initialization fidelity over the long course of our experiment without requiring extra control resources, we make use of the entangling gate available to us, the CX_π gate, to calibrate qubit populations. Intuitively, if the conditional

Rabi drive generates the same contrast as the unconditional Rabi drive, it proves that the control qubit is in a pure $|1\rangle$ state. A quantitative procedure can be developed around this principle to calibrate any deviation of the control qubit from the pure state.

We stress that the lack of knowledge about the initial state purity does not represent an impediment to benchmarking the two-qubit gate. The fact that the two-qubit gate fidelity is obtained via the decay rate of the randomized benchmarking (RB) measurements allows extraction of the CX_π gate fidelity as long as we can measure the relative qubit populations renormalized to the initial states. Our joint readout scheme, to be elaborated in Appendix D, provides such measurements to certify the CX_π gate fidelity regardless of the initialization fidelity. Nevertheless, for definiteness we report actual qubit populations as calibrated in this section throughout the paper.

The initial state populations of qubit A (control) and qubit B (target) are calculated with two complementary sets of measurements, each yielding the population information for one of the two qubits. The way to design such a scheme would not be unique. As long as we have a number of independent measurements sufficient to set up a system of equations with robust solutions, any of them would be applicable. We describe our procedure under three approximations and in the end we will discuss the caveats of these approximations. These assumptions can be stated as follows: (1) the whole population lies within the computational space ($p_{00}+p_{01}+p_{10}+p_{11}=1$); (2) the initial residual populations for the two qubits are independent from each other; and (3) the CX gate is perfect.

1. Measurement protocol for individual qubit populations

To measure qubit A's population, we apply four Ramsey-like measurements on qubit B (Fig. 8) to obtain a set of four equations. In the presence of an entangling gate, the conditional flips result in different oscillation amplitudes for qubit B, depending on the residual population of qubit A. In the ideal case of no residual population for qubits A and B, these four measurements yield oscillations with the following contrasts in readout voltages:

$$\begin{aligned}(a) &\rightarrow M_{01}-M_{00},\\ (b) &\rightarrow M_{01}-M_{00},\\ (c) &\rightarrow M_{11}-M_{10},\\ (d) &\rightarrow M_{11}-M_{10}.\end{aligned}$$

Here M_{ij} is the demodulated complex readout voltage for the pure $|ij\rangle$ state. For this ideal case, qubit A has all the population in the ground state so the contrasts $(a)-(b)$ and $(c)-(d)$ are the same. There will be no

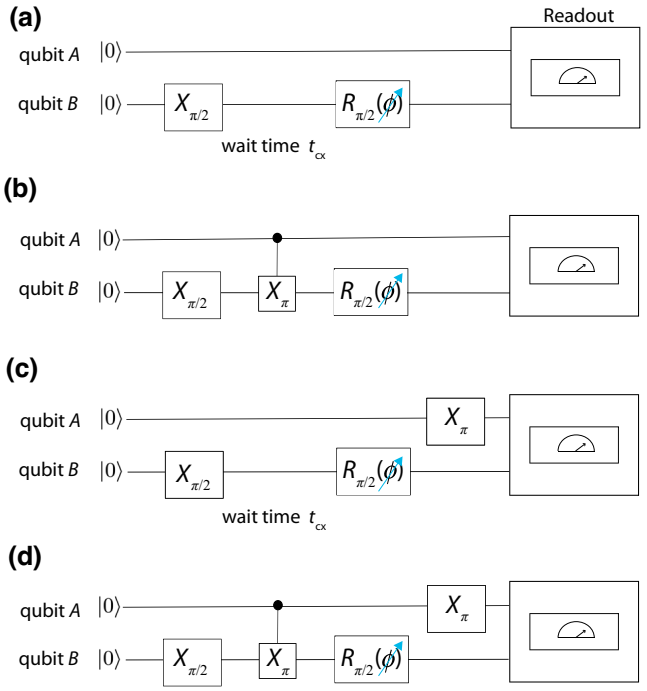

FIG. 8. Pulse sequences for the set of measurements to determine the residual population of qubit A. All four measurements result in oscillations as we sweep the phase of the $\pi/2$ pulse for qubit B.

difference between the unconditional and conditional oscillations since CX_π has no effect on qubit B (target) with qubit A (control) being in ground.

Now we can take into account that qubit A has some nonzero population e in its excited state:

$$\begin{aligned}(a) &\rightarrow (1-e)(M_{01}-M_{00})+e(M_{11}-M_{10}),\\ (b) &\rightarrow (1-e)(M_{01}-M_{00})-e(M_{11}-M_{10}),\\ (c) &\rightarrow (1-e)(M_{11}-M_{10})+e(M_{01}-M_{00}),\\ (d) &\rightarrow (1-e)(M_{11}-M_{10})-e(M_{01}-M_{00}).\end{aligned}$$

The voltages M_{ij} can be eliminated by a correct manipulation of these four equations, i.e.,

$$e=\frac{(c)-(d)}{(a)+(b)+(c)-(d)}\quad\text{or}$$
$$e=\frac{(a)-(b)}{(a)-(b)+(c)+(d)},\tag{B1}$$

to give the residual population for the control qubit.

Qubit B's initial excited-state population ϵ should also be taken into account to complete this approach. However, it merely brings an overall $(1-2\epsilon)$ factor to all four equations that gets canceled out during the elimination [see (B1)].

The population ϵ of qubit B can be measured in a completely analogous manner with these four sets of sequences

if we have a CX_π gate in the reverse direction, with B as control and A as target. While this $CX_{B\to A}$ gate is not native to our system, it can be constructed by the native $CX_{A\to B}$ gate and single-qubit Hadamard gates. Given that $CX_{A\to B}$ and $CNOT_{A\to B}$ only vary by an S gate on qubit A (which is accounted for with a virtual Z in the software), we apply the decomposition $(CNOT_{B\to A}) = H_A H_B (CNOT_{A\to B}) H_A H_B$ for reversing the direction of our gate.

2. Initial state populations

Using the procedure above, we have measured the residual excited-state populations of both qubits after applying the reset protocol detailed in Appendix C. Over the course of multiple weeks, we obtain the average residual population e of qubit A to be $(1.0 \pm 0.3)\%$, and the average residual population ϵ of qubit B to be $(5.0 \pm 1.3)\%$. If we assume the initial qubit states are uncorrelated, we can report the average initial state population after the reset protocol to be $\vec{P}_i = [P_{00}, P_{01}, P_{10}, P_{11}] \approx [0.94, 0.05, 0.01, 0.00]$. It is worth considering our assumptions and possible caveats to this initial state population measurement.

First, since the CX_π gate is not perfect, the measurement outcome of (b) and (d) (from Fig. 8) may be subject to a relative error bounded by the infidelity of the CX gate, $1 - \mathcal{F}$. From Eq. (B1), recall that $(a) \approx (b)$ and $(c) \approx (d)$ (and (a) and (c) are similar in magnitude), and so the possible absolute error of the calculated populations e and ϵ is bounded by $\frac{1}{2}(1 - \mathcal{F})$, or 0.005 for a 99% gate fidelity.

Second, it is possible that the initial qubit states are correlated after our sideband cooling mechanism. However, the most correlated scenario of our initial state given the data would be $\vec{P}_i \approx [0.95, 0.04, 0.00, 0.01]$, which corresponds to a change comparable to the measurement noise level. Essentially, when one of the qubit states is proven to be sufficiently pure, it strongly limits the impact of possible correlations.

Third, the initial states may have residual population in noncomputational states, e.g., in $|2\rangle$. This population is expected to be very small considering that we managed to deplete the $|1\rangle$ state efficiently in our initialization and that the $|1\rangle$-$|2\rangle$ transition has high frequency and short relaxation time. In addition, if qubit A is in $|2\rangle$, the Rabi oscillation of B under CX drive would register a distinct signature (i.e., oscillations of a different period) that is entirely absent in the experiment (to sub-1% level). We cannot place as stringent of a bound to the possibility of qubit B starting in its $|2\rangle$ state, but this small population would not affect the experiment in any way except for the overall scale of readout voltages. For all purposes, we may simply consider $\vec{P}$ as the population vector normalized within the computation subspace.

APPENDIX C: STATE PREPARATION WITH MULTIPHOTON COOLING

Our initial state preparation protocol aims for the reset of qubit A. Assuming a subsystem composed of A and the cavity R, we drive a two-photon sideband transition from the state $|1\rangle_A|0\rangle_R$ to $|0\rangle_A|1\rangle_R$, which then quickly decays into $|0\rangle_A|0\rangle_R$ [Fig. 9(a)]. Initially, without any reset protocol, qubit A is close to a 60%–40% mixture state.

To enact this multiphoton reset protocol, we apply a 15-μs-long square pulse at a frequency of 3.524 GHz, which is equal to half the difference between the cavity and qubit A frequencies.

To calibrate the optimal cooling tone parameters, we sweep over the frequency of the cooling tone for various powers, comparing the *on*:*off* contrast of qubit A. Using the qubit population measurement presented in Appendix B, we can observe the effect of the cooling tone with respect to different cooling tone lengths, reducing the residual excited-state population of A to $\sim 1\%$ and saturating after about 10 μs. For reasons yet to be understood, this reset protocol simultaneously also resets qubit B to $\sim 5\%$ level. We briefly attempted further cooling of qubit B with a $|1\rangle_B|0\rangle_R \leftrightarrow |0\rangle_B|1\rangle_R$ two-photon drive, though it did not bring additional improvement to the initialization.

APPENDIX D: JOINT READOUT WITH PRE-PULSES

We implement a joint readout method that makes use of a number of averaged measurements to extract the population distribution of the two-qubit system [33,49]. Despite increasing the time and number of measurements, this simple scheme (with no additional experimental complexity to the readout setup) makes it possible to implement two-qubit experiments with no single-shot readout. This used to be a common practice before Josephson parametric amplifiers became widely available. The characteristics of our readout cavity and dispersive shifts, taken with respect to the $|00\rangle$ state, are as follows: $\omega_{RO} = 7.598$ GHz, $\kappa_{RO}/2\pi = 5.6$ MHz, $\chi_{|10\rangle}/2\pi = -6.0$ MHz, $\chi_{|01\rangle}/2\pi = 2.6$ MHz, and $\chi_{|11\rangle}/2\pi = 0.8$ MHz. The dispersive shifts were calculated using a weaker measurement than was typically used throughout the experiment to mitigate relaxation effects during the readout.

At the end of any measurement, the readout gives us a complex demodulated voltage of the cavity transmission that involves contributions from all states weighted with their corresponding populations:

$$V = p_{00}M_{00} + p_{01}M_{01} + p_{10}M_{10} + p_{11}M_{11}.$$

Here $\vec{p} = \mathrm{diag}[\rho] = [p_{00}, p_{01}, p_{10}, p_{11}]^T$ is the population distribution for any state of the two-qubit system. This population four-vector is particularly essential for benchmarking purposes where we need to know how much of

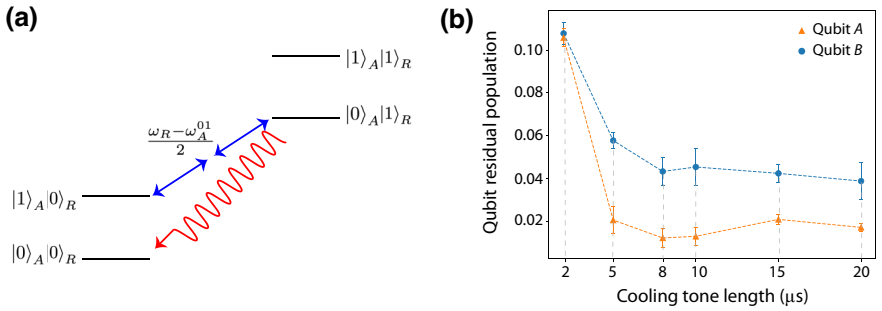

FIG. 9. (a) Multiphoton cooling scheme used for initial state preparation. (b) Effect of cooling tones with various lengths on qubit residual populations for the tone power and frequency specified in Appendix C. Each point corresponds to the average of 10 measurements (with 5000 averages each) done with the same cooling configuration, and the error bars are the standard error for the distribution of 10 measurements.

the final state is in $|00\rangle$. This single complex voltage value by itself is not sufficient to pin down the population vector $\vec{p}$. In order to obtain $\vec{p}$ at any moment, we first need to know about the four-vector $\vec{M}$ where the components are complex demodulated voltages of the cavity transmission for the computational states $[M_{00}, M_{01}, M_{10}, M_{11}]$.

We also need to have a sufficient number of independent measurements that will let us solve for the unknowns p_{00}, p_{01}, p_{10}, and p_{11}. For that, any prepared state will be measured with four different readout configurations, varied by pre-readout single-qubit rotations:

(1) no rotation (II),
(2) π pulse on qubit B (IX),
(3) π pulse on qubit A (XI),
(4) π pulse on both qubits (XX).

With this scheme, each measurement will eventually yield a set of four measured voltages: $\vec{V} = [V_{\mathrm{II}}, V_{\mathrm{IX}}, V_{\mathrm{XI}}, V_{\mathrm{XX}}]^T$. The correct mapping relates the four populations to the four measured voltages:

$$\begin{pmatrix} V_{\mathrm{II}} \\ V_{\mathrm{IX}} \\ V_{\mathrm{XI}} \\ V_{\mathrm{XX}} \end{pmatrix} = \begin{pmatrix} M_{00} & M_{01} & M_{10} & M_{11} \\ M_{01} & M_{00} & M_{11} & M_{10} \\ M_{10} & M_{11} & M_{00} & M_{01} \\ M_{11} & M_{10} & M_{01} & M_{00} \end{pmatrix} \begin{pmatrix} p_{00} \\ p_{01} \\ p_{10} \\ p_{11} \end{pmatrix}.$$

The matrix M is nothing but a permutation of the components of $\vec{M}$. The vector $\vec{M}$ will allow us to reconstruct the inverse of the mapping relation that will give any state's population distribution, given the measured voltages:

$$\vec{p} = M^{-1}\vec{V}. \tag{D1}$$

In the case of perfect state preparation, obtaining components of the $\vec{M}$ would be straightforward. Although it is not a realistic assumption, it is intuitive to look at the case with $\vec{P}_i = [1, 0, 0, 0]^T$, where $\vec{P}_i$ is the four-vector for the initial states of the system:

$$\begin{pmatrix} V_{\mathrm{II}} \\ V_{\mathrm{IX}} \\ V_{\mathrm{XI}} \\ V_{\mathrm{XX}} \end{pmatrix} = \begin{pmatrix} M_{00} & M_{01} & M_{10} & M_{11} \\ M_{01} & M_{00} & M_{11} & M_{10} \\ M_{10} & M_{11} & M_{00} & M_{01} \\ M_{11} & M_{10} & M_{01} & M_{00} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} M_{00} \\ M_{01} \\ M_{10} \\ M_{11} \end{pmatrix}.$$

For a more realistic case, we know that the initial population vector should have nonzero populations for the remaining computational states as well. With the initial state populations $\vec{P}_i$ (obtained using the method from Appendix B), the mapping can be written in the following way:

$$\begin{pmatrix} M_{00} \\ M_{01} \\ M_{10} \\ M_{11} \end{pmatrix} = \begin{pmatrix} P_{00} & P_{01} & P_{10} & P_{11} \\ P_{01} & P_{00} & P_{11} & P_{10} \\ P_{10} & P_{11} & P_{00} & P_{01} \\ P_{11} & P_{10} & P_{01} & P_{00} \end{pmatrix}^{-1} \begin{pmatrix} V_{\mathrm{II}} \\ V_{\mathrm{IX}} \\ V_{\mathrm{XI}} \\ V_{\mathrm{XX}} \end{pmatrix},$$

which gives us the required complex demodulated voltages. With the components of $\vec{M}$ at hand, we can now construct the full mapping from Eq. (D1) that gives us the population distribution for the computational states at any point of a measurement.

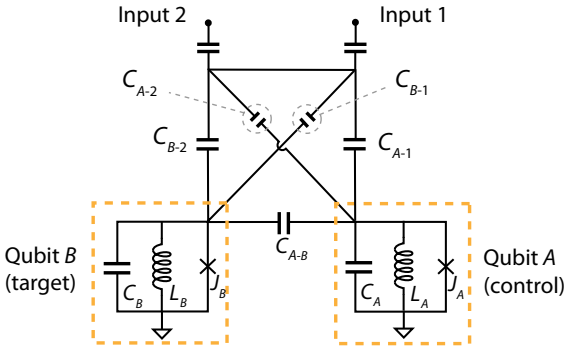

FIG. 10. Effective circuit diagram of the capacitively coupled two-fluxonium system with asymmetric coupling to two input ports. The cavity mode is omitted for simplicity.

APPENDIX E: CALIBRATION AND BENCHMARKING OF SINGLE-QUBIT GATES

The fact that we are working in a 3D geometry with no local drives and with a system that has a strong CR component brings up the challenge of addressing the qubits individually. Given that both input ports are coupled to both qubits (see Fig. 10), any qubit B drive from one of the individual ports (at frequency ω_B) will also be seen by qubit A, leading to conditional oscillations of B [Figs. 11(a) and 11(b)]. In order to realize high-fidelity single-qubit gates, we need to be able to address each qubit independently from the state of the other one. As mentioned in the main text, the frequency of the qubit A oscillations driven from input 1 does not have a significant dependence on the qubit B state and can be considered as a "clean" single-qubit drive by itself. On the other hand, in order to realize a "clean" single-qubit gate for B, the two individual drives need to be combined with a scheme similar to the CR gate, only this time to fulfill the requirement: $\langle 00|\hat{H}_{\rm dr}|01\rangle = \langle 10|\hat{H}_{\rm dr}|11\rangle$ with the drive Hamiltonian from Eq. (5) of the main text. With a similar procedure, one can find the right relative amplitude and phase between the two drives to satisfy this condition.

This vector compensation scheme is illustrated in more detail in Fig. 12. We can consider the drive from each port as having two components: a direct drive, which is independent of the state of qubit A, and a CR drive, which is turned on whenever qubit A is in $|1\rangle$. We tune the complex ratio η' of the drives on each port such that the CR components are of equal amplitude and opposite phase, thereby leaving only the direct drives. With the proper ratio, qubit B's Rabi frequency no longer depends on the state of qubit A [Fig. 11(c)].

We benchmark our single-qubit gates using simultaneous single-qubit randomized benchmarking (insensitive to SPAM errors) [50]. We generate random sequences of N pairs of Clifford gates, one for each qubit, with N ranging from 1 to 200, and apply them to each qubit simultaneously in time. At the conclusion of each sequence, we measure two different quantities: the fidelity of the $|0\rangle$ state of qubit A and the fidelity of the $|0\rangle$ state of qubit B. This

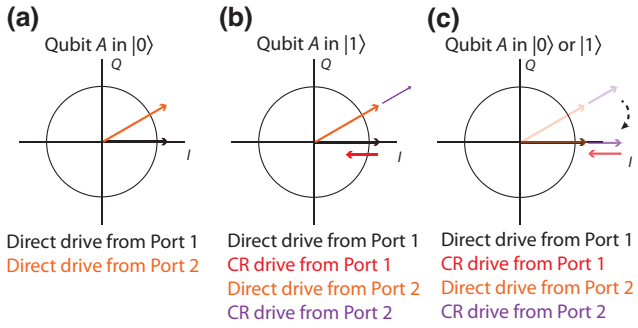

FIG. 12. (a) Effective drive on qubit B, shown in the I-Q plane, when qubit A is in $|0\rangle$. The qubit is driven with a different amplitude and phase depending on which driving port is used. (b) Effective drive on qubit B, shown in the I-Q plane, when qubit A is in $|1\rangle$. The qubit is driven with a different effective amplitude and phase due to the CR component of the drive now being turned on. (c) Relative amplitude and phase are applied to the two driving ports to exactly cancel the CR drives and achieve unconditional single-qubit rotations.

measurement was repeated 456 times over the course of several weeks, with different random sequences of N Clifford pairs being generated each time. The results were averaged together, separately for each qubit, and each fit to a single exponential. The resulting traces are shown in Fig. 13. From these traces, we can extract the average error per Clifford on each individual qubit. Taking into account the construction of our single-qubit Clifford gate set, which utilizes virtual Z gates and contains 1.167 physical single-qubit gates per Clifford, we report the following average fidelities per physical gate: $\mathcal{F}_A = 0.99835(4)$ and $\mathcal{F}_B = 0.99734(6)$.

APPENDIX F: CALIBRATION AND BENCHMARKING OF THE CR GATE

The preliminary CX_π gate tune up involves the calibration of the relative phase, the relative amplitude, and the overall amplitude, as elaborated in the main text. However, a more sophisticated calibration procedure is needed to realize a high-fidelity CX gate. Assuming that we already have a gate with the above three parameters calibrated

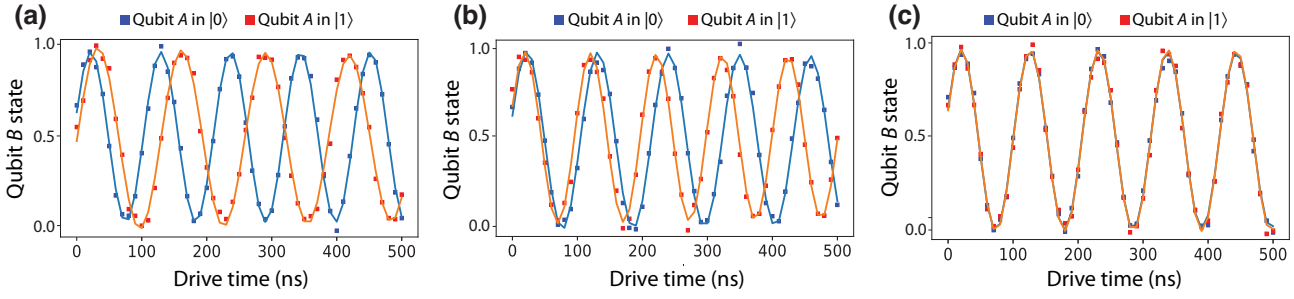

FIG. 11. (a) Rabi oscillations of qubit B driven from port 1. (b) Rabi oscillations of qubit B driven from port 2. (c) Rabi oscillations of qubit B after applying the complex ratio η' between the two driving ports.

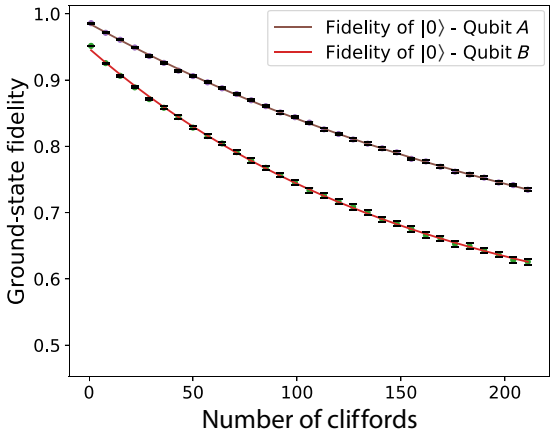

FIG. 13. Simultaneous single-qubit randomized benchmarking. The two traces shown are measured fidelities to the $|0\rangle$ state of qubit A and the $|0\rangle$ state of qubit B at the conclusion of N random pairs of single-qubit Clifford gates. Each trace is the averaged result of 456 randomized benchmarking experiments performed over the course of several weeks. The error bars shown are the standard errors taken across these 456 experiments.

reasonably well, we can now proceed with the next level of calibration. We should note that we do not apply any brute-force or black-box tools (such as Nelder-Mead or machine learning algorithms) to numerically optimize the gate parameters against the two-qubit gate fidelity. We rely on the following completely intuitive calibration procedure targeting individual parameters for realizing high-fidelity gates. It has the benefit of speed and transparency.

From the observed control Rabi dynamics (Fig. 2), we know that the process has a block-diagonal structure. The remaining degrees of freedom that would not be distinguishable for the drives from Fig. 2 are the imprinted phases for the two qubits. Taking them into account, we can write our experimental CR matrix to have the following form:

$$\mathrm{CR}_{\mathrm{exp}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\theta_B} & 0 & 0 \\ 0 & 0 & 0 & -ie^{i\theta_A} \\ 0 & 0 & -ie^{i\theta_A} & 0 \end{pmatrix}, \tag{F1}$$

where θ_A and θ_B are phases picked by qubit A and B during our cross-resonance gate operation. Brought into a superposition state, the phases θ_A (θ_B) would be what we would observe to be the additional phase accumulated on the qubits A (B) under the CR drive. It is possible to measure these phases using the same pulse sequences from Fig. 8. The phase difference between the oscillations resulting from the measurements of (a) and (b) on qubit A (qubit B) will give us θ_A (θ_B).

1. Integrating single-qubit Z rotations

Given that our experimental cross-resonance gate realization brings along extra phase accumulations on both qubits A and B, we first need to show the equivalence of our experimental gate to a CX_π gate.

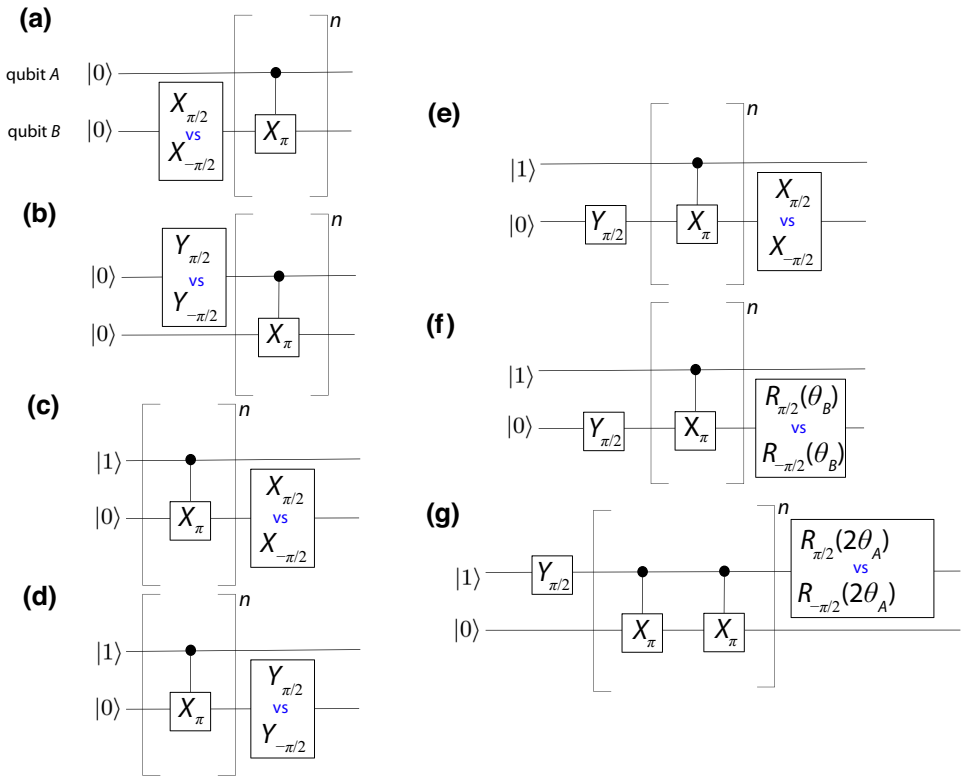

FIG. 14. CR tune-up sequences. These seven experiments check the following criteria. (1) No drive on qubit B when qubit A is *off* : the calibration experiment (a) makes sure that qubit B does not rotate with respect to the X axis when qubit A is in $|0\rangle_A$; and experiment (b) does a similar check for the Y axis. (2) Resonant-π condition when qubit A is *on*: experiment (c) checks whether qubit B overrotates or not, given qubit A is in state $|1\rangle_A$; and experiment (d) does the complementary check by making sure that qubit B is driven along the big circle on the Bloch sphere. (3) Alignment check: experiment (e) checks if the CR rotation aligns with the X axis of qubit B. (4) Correction of single-qubit frames: experiments (f) and (g) make sure that the extra phases on qubit A and qubit B are compensated in the correct way.

The relevant Z-rotation operations that qubit A and B undergo are expressed as

$$Z^A(\theta_A) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & e^{i\theta_A} \\ 0 & 0 & e^{i\theta_A} & 0 \end{pmatrix}$$

and

$$Z^B(\theta_B) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\theta_B} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\theta_B} \end{pmatrix}.$$

It can be shown that our experimental CR gate is equivalent to a CX gate up to a number of Z rotations:

$$[\mathrm{CX}_\pi] = \left[Z^B\left(-\frac{\theta_B}{2}\right)\right][\mathrm{CR}_{\mathrm{exp}}]\left[Z^B\left(-\frac{\theta_B}{2}\right)\right] \times \left[Z^A\left(\frac{\theta_B}{2} - \theta_A\right)\right]. \tag{F2}$$

Since these Z rotations can simply be applied via virtual Z's [35] in the software, there is a complete equivalence between our experimental cross-resonance gate and a CX_π gate as long the correct virtual Z operations are applied. We can also trivially change θ_A by $\pi/2$ to convert the gate to a CNOT. For the rest of the section and for all our benchmarking purposes, our gate is composed of the form as in Eq. (F2).

2. Fine tuning the gate for high fidelity

The fine-tune-up procedure consists of the calibration of the following parameters:

(1) relative amplitude of the two drives,
(2) relative phase of the two drives,
(3) overall amplitude of the drive,
(4) detuning of the CR drive with respect to the qubit B drive (CR detuning $\Delta\omega_{\mathrm{CR}}$),
(5) misalignment of the CR drive with respect to the qubit B drive (CR angle $\Delta\phi_{\mathrm{CR}}$),
(6) phase picked up by qubit B during the CR drive, and
(7) phase picked up by qubit A during the CR drive.

Of these parameters, the requirements for CR detuning (4) and CR angle (5) have been determined heuristically. The optimal values for these two parameters are small but nonzero ($\Delta\omega_{CR} \sim$ order of 100 kHz, $\Delta\phi_{\mathrm{CR}} \sim$ order of 10^{-2} rad) and we suspect that their presence is related to the remaining underlying reflection issue. Our procedure is similar to the calibration procedure from Ref. [28] and has been tailored to fine-tune the seven degrees of freedom of our CX_π gate. The detailed procedure and the pulse sequences can be seen from Fig. 14. Sweeping over the calibration parameter, the two measurements that differ from each other by an inversion of one of their pulses result in a crossing point that gives the correct value for each of these seven parameters. In order to increase the sensitivity of these calibrations, we can increase the number of repetitions of the CR gate in the sequence (an even number of repetitions is required for the θ_A calibration.)

This calibration procedure is run at regular intervals during gate benchmarking and is completed with a visual syndrome check that includes all the parameter checks simultaneously.

3. Quantum process tomography with joint readout

The quantum process matrix of the CX_π gate is estimated over a set of 1044 measurements with joint state readout. We first apply the two-qubit gate on 36 initial states prepared by applying all combinations of simultaneous target and control single-qubit gates: I, $\mathrm{X}_{\pi/2}$, $\mathrm{X}_{-\pi/2}$, X_π, $\mathrm{Y}_{\pi/2}$, and $\mathrm{Y}_{-\pi/2}$. State tomography is then applied on the 36 new states following Ref. [36], with 29 pre-readout pulses and measurements from which the final density matrix is found via least-squares optimization. The CX_π process matrix is estimated via a simultaneous least-squares optimization over the 36 final density matrices.

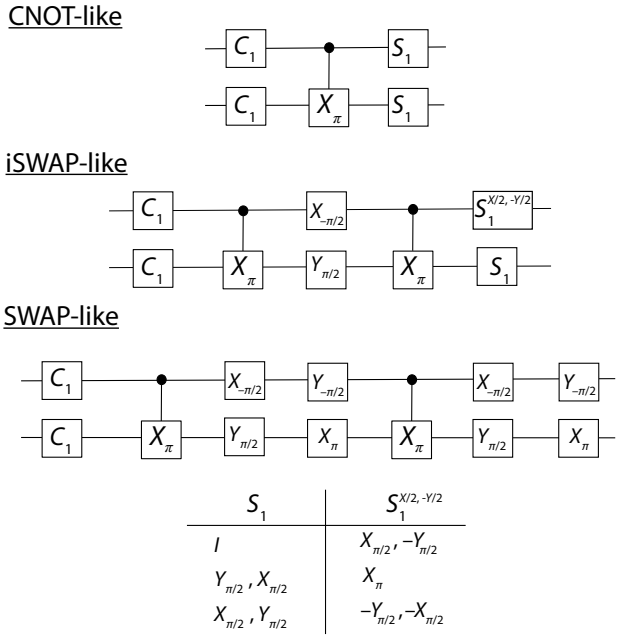

FIG. 15. Two-qubit Clifford group composition with CX_π: C_1 refers to the single-qubit Clifford group, while S_1 and $\mathrm{S}_1^{X/2,-Y/2}$ are each three-element subsets of C_1.

4. CX_π randomized benchmarking

To measure the fidelity of our CX_π gate, we perform interleaved two-qubit randomized benchmarking. To construct our two-qubit Clifford gate set, we follow the procedures in Refs. [20,37], modifying it to create the two-qubit Clifford group using our CX_π gate as the generator. The gate decomposition can be seen in Fig. 15. Each two-qubit Clifford contains on average 6.483 physical single-qubit gates and 1.5 CX gates.

We extract fidelity by measuring the sequence fidelity of at least 12 different N values with N typically going from 1 to 100. To measure sequence fidelity at each N, we generate at least 40 different random Clifford sequences, and for each random sequence we average 5000 times to reduce noise. The average sequence fidelity is fit according to the model $F = Ap^m + B$ (with no constraints) to obtain the decay constants, and the error per Clifford is $\mathrm{EPC} = \frac{3}{4}(1-p)$. The error bars on Fig. 3(c) are standard errors ($\mathrm{stdev}/\sqrt{N_t}$), where "stdev" is the standard deviation from the mean and N_t is the number of runs. The uncertainties for the EPCs and fidelity numbers are calculated from the fit errors. The three different kinds of EPCs ($\mathrm{EPC}_{\mathrm{ref}}$ for reference two-qubit RB; $\mathrm{EPC}_{\mathrm{CX}}$ for interleaved RB with CX_π; and $\mathrm{EPC}_{\mathrm{idle}}$ for interleaved RB with I) are calculated separately, and average fidelities for these three are obtained via the following relations:

average 2Q Clifford fidelity $= (1 - \mathrm{EPC}_{\mathrm{ref}})$,
average CX_π fidelity $= (1 - \mathrm{EPC}_{\mathrm{CX}})/(1 - \mathrm{EPC}_{\mathrm{ref}})$,
average idle fidelity $= (1 - \mathrm{EPC}_{\mathrm{idle}})/(1 - \mathrm{EPC}_{\mathrm{ref}})$.

APPENDIX G: EXPERIMENTAL SETUP

The most fundamental microwave setup requirement for our experimental scheme is the phase locking between the two drive tones played into the two input ports. We realize this by using a single microwave generator that resources all individual qubit drives. The same local oscillator feeds the two "home-made" I-Q mixers (one per control line) to maintain phase coherence during combined drives. The LO frequency is set to 840 MHz based on hardware constraints and to minimize spurious tones in the qubit drive spectrum. Each "home-made" I-Q mixer is composed of two discrete three-port mixers and power

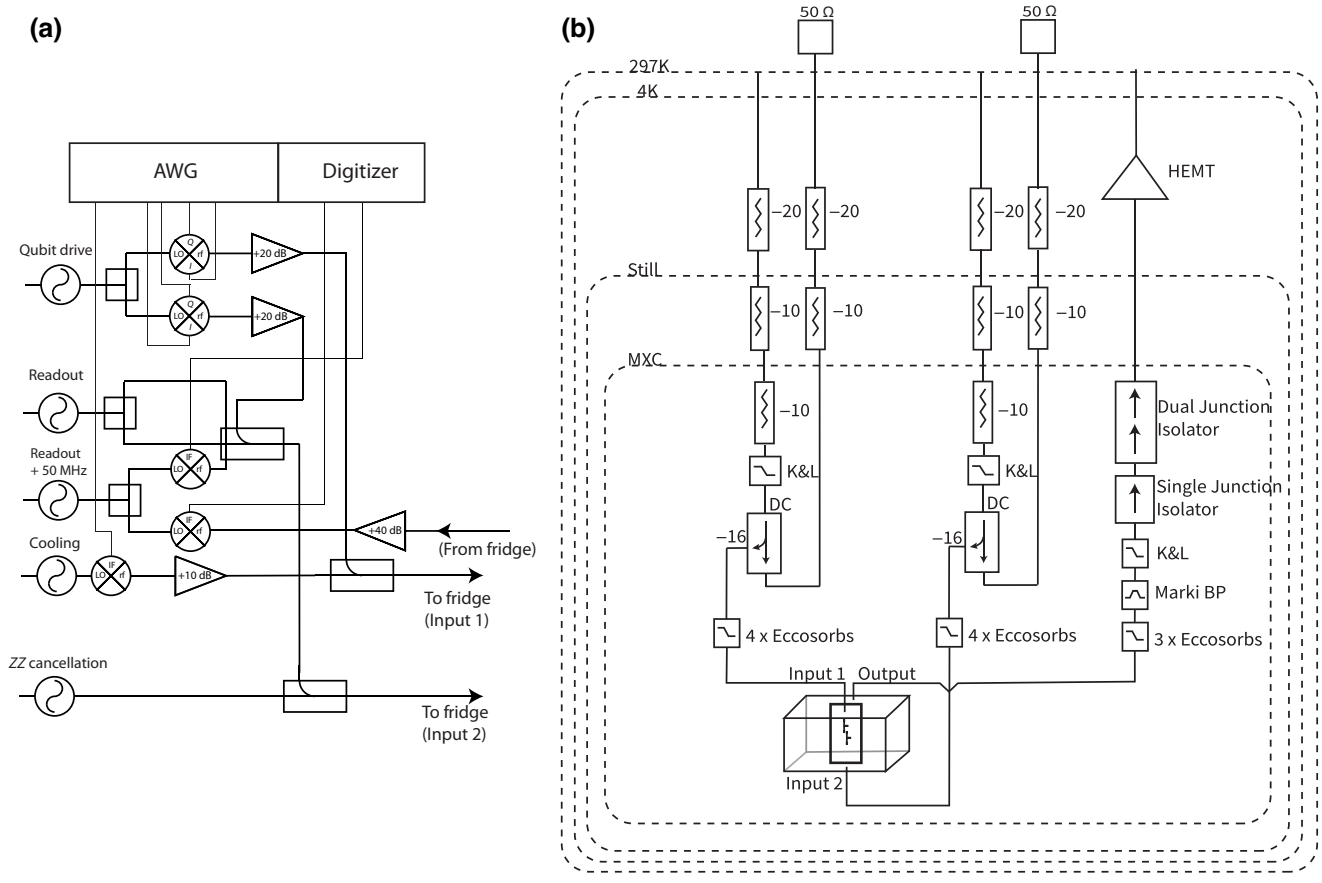

FIG. 16. (a) Diagram of the room-temperature microwave setup. Note that what is depicted as a standard, four-port I-Q mixer is the "home-made" I-Q mixer described in Appendix G. The lines are also heavily filtered with low-pass and bandpass filters to eliminate spurious tones. These details are omitted from the diagram for simplicity. (b) Diagram of the cryogenic setup.

combiners (splitters) carefully chosen to cover both qubit frequencies after up-mixing (corresponding to intermediate frequencies of −285 MHz for qubit A and 164 MHz for qubit B). The entire room-temperature microwave setup is shown in Fig. 16(a).

The fridge setup is shown in Fig. 16(b): The noncryogenic directional coupler to which we attribute the impedance mismatch can be seen below the mixing chamber. All the Eccosorb filters are made in-house and have been chained up (4× for each input line and 3× for the output line) between the cavity ports and the closest microwave component. The experiment is carried out in an Oxford Triton 500 Dilution Refrigerator with a base temperature below 20 mK.

APPENDIX H: DEVICE FABRICATION

The fabrication procedure for the two-fluxonium device is identical to that in Ref. [12]. The device is fabricated on a 430-μm-thick sapphire chip with dimensions 9 × 4 mm^2. Following the cleaning of the chip, the double-layer resist's first layer is prepared by spinning methyl methacrylate EL-13 at 5000 revolutions per minute (rpm) and baking at 180 °C for 1 min. This is followed by the second layer of resist: 950 PMMA A3 spun at 4000 rpm for 1 min and baked at the same temperature for half an hour. As the anticharging layer, 11 nm of Al is deposited using a Plassys deposition system. The electron-beam lithography is realized using a current of 1 nA on a 100 kV Elionix EBL system. After the electron-beam lithography, the anticharging layer is removed by keeping the chip in a 0.1 M potassium hydroxide (KOH) aqueous solution. The development is done in a 3:1 isopropyl alcohol:deionized water (IPA:DI) solution at a temperature of 6 °C for 2 min. A Plassys deposition system is used for the double-angle Al deposition (with a Al deposition thickness of 20 nm for the first and 40 nm for the second layer) with the same process steps as detailed in Ref. [12]. Liftoff is realized by keeping the chip in acetone at a temperature of 60 °C for 3 h, followed by brief sonications in acetone and IPA (5 and 10 s). As the last step, the chip is dried with N_2.

APPENDIX H: SIMULATION AND DISCUSSION OF GATE FIDELITY

We carry out numerical simulations of the CR gate using the Hamiltonian and coherence parameters extracted from the experiment (Table I). The simulations [Fig. 4(b) and Fig. 17] are based on the procedure from Ref. [15], which searches for the lowest coherent gate error with respect to the CNOT operation over the seven parameters of the CR gate (similar to the actual experiment). Decoherences of both the computational and noncomputational states are then added to the simulation based on the optimized parameter set.

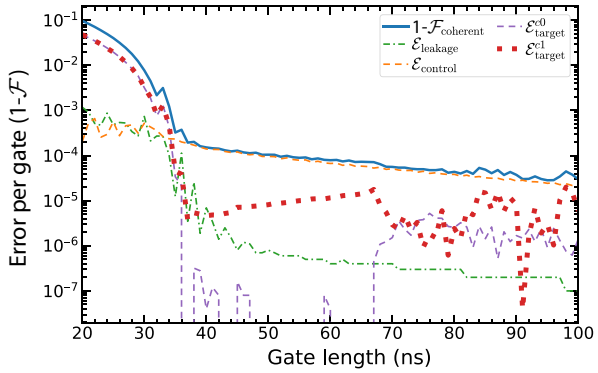

FIG. 17. Numerical simulation for detailed breakdown of coherent control errors in logarithmic scale, showing error due to leakage to noncomputational states ($\mathcal{E}_{\text{leakage}}$), unintended rotation of control qubit ($\mathcal{E}_{\text{control}}$), and misrotation of qubit B in $|0\rangle_A$ and $|1\rangle_A$ state ($\mathcal{E}^{0/1}_{\text{target}}$). The simulation considers rounded square pulses with ramp time of 6 ns. The dominant control errors at gate time above 40 ns, $\mathcal{E}_{\text{control}}$, can be substantially reduced by using slower pulse ramps.

An additional observation about the simulation in Fig. 4(b) is that the contribution from decoherence of the noncomputational transitions is very small, as expected for our CR scheme where qubit dynamics is well confined in the computational subspace. Experimentally we found $T_1^{1\to2}$ and $T_2^{1\to2}$ times for the $|1\rangle$-$|2\rangle$ transitions in the range of 3–7 μs. A more precise characterization is nontrivial due to the strong hybridization of noncomputational states, the need for calibrating the joint readout for many combinations of qubit states, and the need to actively cancel pulse reflection at multiple frequencies. Nevertheless, for the range of gate length explored in our experiment, we found that even a very short coherence time of $T_1^{1\to2} = T_2^{1\to2} =$ 1 μs for both qubits would only contribute at 10% level to the total decoherence error. Therefore, decoherence of noncomputational states is expected to have negligible contribution in our experiment (less than the day-to-day fluctuations of the computational state decoherence).

The simulation in Fig. 17 shows that the coherent control error is easily below 10^{-4} at gate length above 50 ns. Moreover, the leading error is unintended rotation of control qubit A by the CR drive. This is purely driven by the fast ramp time of the control pulse, which contains a spectral tail that marginally excite the control qubit with a detuning of ∼ 450 MHz. The error can be suppressed substantially by choosing a slightly slower ramp time as needed, and is therefore considered unimportant in Ref. [15].

[1] F. Arute, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[2] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[3] 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).

[4] V. E. Manucharyan, J. Koch, L. I. Glazman, and M. H. Devoret, Fluxonium: Single Cooper-pair circuit free of charge offsets, Science **326**, 113 (2009).

[5] F. Bao, *et al.*, Fluxonium: An Alternative Qubit Platform for High-Fidelity Operations, Phys. Rev. Lett. **129**, 010502 (2022).

[6] L. B. Nguyen, G. Koolstra, Y. Kim, A. Morvan, T. Chistolini, S. Singh, K. N. Nesterov, C. Jünger, L. Chen, Z. Pedramrazi, B. K. Mitchell, J. M. Kreikebaum, S. Puri, D. I. Santiago, and I. Siddiqi, Blueprint for a High-Performance Fluxonium Quantum Processor, PRX Quantum **3**, 037001 (2022).

[7] Q. Ficheux, L. B. Nguyen, A. Somoroff, H. Xiong, K. N. Nesterov, M. G. Vavilov, and V. E. Manucharyan, Fast Logic with Slow Qubits: Microwave-Activated Controlled-Z Gate on Low-Frequency Fluxoniums, Phys. Rev. X **11**, 021026 (2021).

[8] H. Xiong, Q. Ficheux, A. Somoroff, L. B. Nguyen, E. Dogan, D. Rosenstock, C. Wang, K. N. Nesterov, M. G. Vavilov, and V. E. Manucharyan, Arbitrary controlled-phase gate on fluxonium qubits using differential ac Stark shifts, Phys. Rev. Res. **4**, 023040 (2022).

[9] I. N. Moskalenko, I. A. Simakov, N. N. Abramov, A. A. Grigorev, D. O. Moskalev, A. A. Pishchimova, N. S. Smirnov, E. V. Zikiy, I. A. Rodionov, and I. S. Besedin, High fidelity two-qubit gates on fluxoniums using a tunable coupler, npj Quantum Inf. **8**, 130 (2022).

[10] L. B. Nguyen, Y.-H. Lin, A. Somoroff, R. Mencia, N. Grabon, and V. E. Manucharyan, High-Coherence Fluxonium Qubit, Phys. Rev. X **9**, 041041 (2019).

[11] H. Zhang, S. Chakram, T. Roy, N. Earnest, Y. Lu, Z. Huang, D. K. Weiss, J. Koch, and D. I. Schuster, Universal Fast-Flux Control of a Coherent, Low-Frequency Qubit, Phys. Rev. X **11**, 011010 (2021).

[12] A. Somoroff, Q. Ficheux, R. A. Mencia, H. Xiong, R. V. Kuzmin, and V. E. Manucharyan, Millisecond Coherence in a Superconducting Qubit, Phys. Rev. Lett. **130**, 267001 (2023).

[13] K. N. Nesterov, I. V. Pechenezhskiy, C. Wang, V. E. Manucharyan, and M. G. Vavilov, Microwave-activated controlled-*Z* gate for fixed-frequency fluxonium qubits, Phys. Rev. A **98**, 030301 (2018).

[14] K. N. Nesterov, Q. Ficheux, V. E. Manucharyan, and M. G. Vavilov, Proposal for Entangling Gates on Fluxonium Qubits via a Two-Photon Transition, PRX Quantum **2**, 020345 (2021).

[15] K. N. Nesterov, C. Wang, V. E. Manucharyan, and M. G. Vavilov, Controlled-NOT Gates for Fluxonium Qubits via Selective Darkening of Transitions, Phys. Rev. Appl. **18**, 034063 (2022).

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

[17] 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 (2016).

[18] A. Kandala, K. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. Keefe, D. Klaus, O. Dial, and D. McKay, Demonstration of a High-Fidelity CNOT Gate for Fixed-Frequency Transmons with Engineered ZZ Suppression, Phys. Rev. Lett. **127**, 130501 (2021).

[19] K. Heya and N. Kanazawa, Cross-Cross Resonance Gate, PRX Quantum **2**, 040336 (2021).

[20] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, Process verification of two-qubit quantum gates by randomized benchmarking, Phys. Rev. A **87**, 030301 (2013).

[21] S. Hazra, K. V. Salunkhe, A. Bhattacharjee, G. Bothara, S. Kundu, T. Roy, M. P. Patankar, and R. Vijay, Engineering cross resonance interaction in multi-modal quantum circuits, Appl. Phys. Lett. **116**, 152601 (2020).

[22] A. Patterson, J. Rahamim, T. Tsunoda, P. Spring, S. Jebari, K. Ratter, M. Mergenthaler, G. Tancredi, B. Vlastakis, M. Esposito, and P. Leek, Calibration of a Cross-Resonance Two-Qubit Gate Between Directly Coupled Transmons, Phys. Rev. Appl. **12**, 064013 (2019).

[23] V. Tripathi, M. Khezri, and A. N. Korotkov, Operation and intrinsic error budget of a two-qubit cross-resonance gate, Phys. Rev. A **100**, 012301 (2019).

[24] A. Petrescu, C. L. Calonnec, C. Leroux, A. Di Paolo, P. Mundada, S. Sussman, A. Vrajitoarea, A. A. Houck, and A. Blais, Accurate Methods for the Analysis of Strong-Drive Effects in Parametric Gates, Phys. Rev. Appl. **19**, 044003 (2023).

[25] E. Magesan and J. M. Gambetta, Effective Hamiltonian models of the cross-resonance gate, Phys. Rev. A **101**, 052308 (2020).

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

[27] S. Ashhab, F. Yoshihara, T. Fuse, N. Yamamoto, A. Lupascu, and K. Semba, Speed limits for two-qubit gates with weakly anharmonic qubits, Phys. Rev. A **105**, 042614 (2022).

[28] K. X. Wei, E. Magesan, I. Lauer, S. Srinivasan, D. F. Bogorin, S. Carnevale, G. A. Keefe, Y. Kim, D. Klaus, W. Landers, N. Sundaresan, C. Wang, E. J. Zhang, M. Steffen, O. E. Dial, D. C. McKay, and A. Kandala, Hamiltonian Engineering with Multicolor Drives for Fast Entangling Gates and Quantum Crosstalk Cancellation, Phys. Rev. Lett. **129**, 060501 (2022).

[29] J. Ku, X. Xu, M. Brink, D. C. McKay, J. B. Hertzberg, M. H. Ansari, and B. Plourde, Suppression of Unwanted ZZ Interactions in a Hybrid Two-Qubit System, Phys. Rev. Lett. **125**, 200504 (2020).

[30] P. C. de Groot, J. Lisenfeld, R. N. Schouten, S. Ashhab, A. Lupaşcu, C. J. P. M. Harmans, and J. E. Mooij, Selective darkening of degenerate transitions demonstrated with two superconducting quantum bits, Nat. Phys. **6**, 763 (2010).

[31] H. Paik, D. I. Schuster, L. S. Bishop, G. Kirchmair, G. Catelani, A. P. Sears, B. R. Johnson, M. J. Reagor, L. Frunzio, L. I. Glazman, S. M. Girvin, M. H. Devoret, and R. J.

Schoelkopf, Observation of High Coherence in Josephson Junction Qubits Measured in a Three-Dimensional Circuit QED Architecture, Phys. Rev. Lett. **107**, 240501 (2011).

[32] I. N. Moskalenko, I. S. Besedin, I. A. Simakov, and A. V. Ustinov, Tunable coupling scheme for implementing two-qubit gates on fluxonium qubits, Appl. Phys. Lett. **119**, 194001 (2021).

[33] S. Filipp, P. Maurer, P. J. Leek, M. Baur, R. Bianchetti, J. M. Fink, M. Göppl, L. Steffen, J. M. Gambetta, A. Blais, and A. Wallraff, Two-Qubit State Tomography Using a Joint Dispersive Readout, Phys. Rev. Lett. **102**, 200402 (2009).

[34] S. Gustavsson, O. Zwier, J. Bylander, F. Yan, F. Yoshihara, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Improving Quantum Gate Fidelities by Using a Qubit to Measure Microwave Pulse Distortions, Phys. Rev. Lett. **110**, 040502 (2013).

[35] 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).

[36] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Universal Quantum Gate Set Approaching Fault-Tolerant Thresholds with Superconducting Qubits, Phys. Rev. Lett. **109**, 060501 (2012).

[37] R. Barends, *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature **508**, 500 (2014).

[38] P. O'Malley, *et al.*, Qubit Metrology of Ultralow Phase Noise Using Randomized Benchmarking, Phys. Rev. Appl. **3**, 044009 (2015).

[39] A. P. M. Place, *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, Nat. Commun. **12**, 1779 (2021).

[40] C. Wang, *et al.*, Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds, npj Quantum Inf. **8**, 1 (2022).

[41] I. M. Pop, K. Geerlings, G. Catelani, R. J. Schoelkopf, L. I. Glazman, and M. H. Devoret, Coherent suppression of electromagnetic dissipation due to superconducting quasiparticles, Nature **508**, 369 (2014).

[42] J. B. Hertzberg, E. J. Zhang, S. Rosenblatt, E. Magesan, J. A. Smolin, J.-B. Yau, V. P. Adiga, M. Sandberg, M. Brink, J. M. Chow, and J. S. Orcutt, Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors, npj Quantum Inf. **7**, 129 (2021).

[43] C. Chamberland, G. Zhu, T. J. Yoder, J. B. Hertzberg, and A. W. Cross, Topological and Subsystem Codes on Low-Degree Graphs with Flag Qubits, Phys. Rev. X **10**, 011022 (2020).

[44] A. Ciani, B. M. Varbanov, N. Jolly, C. K. Andersen, and B. M. Terhal, Microwave-activated gates between a fluxonium and a transmon qubit, Phys. Rev. Res. **4**, 043127 (2022).

[45] M. Jerger, A. Kulikov, Z. Vasselin, and A. Fedorov, In Situ Characterization of Qubit Control Lines: A Qubit as a Vector Network Analyzer, Phys. Rev. Lett. **123**, 150501 (2019).

[46] 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).

[47] M. Reed, Entanglement and quantum error correction with superconducting qubits, ArXiv:1311.6759 (2013).

[48] K. Geerlings, Z. Leghtas, I. M. Pop, S. Shankar, L. Frunzio, R. J. Schoelkopf, M. Mirrahimi, and M. H. Devoret, Demonstrating a Driven Reset Protocol for a Superconducting Qubit, Phys. Rev. Lett. **110**, 120501 (2013).

[49] J. M. Chow, L. DiCarlo, J. M. Gambetta, A. Nunnenkamp, L. S. Bishop, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Detecting highly entangled states with a joint qubit readout, Phys. Rev. A **81**, 062325 (2010).

[50] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 180504 (2011).