Arbitrary controlled-phase gate on fluxonium qubits using differential ac Stark shifts

Haonan Xiong,[1,*] Quentin Ficheux,[1,*] Aaron Somoroff,[1] Long B. Nguyen,[1] Ebru Dogan,[2] Dario Rosenstock,[2] Chen Wang,[2] Konstantin N. Nesterov,[3] Maxim G. Vavilov,[3] and Vladimir E. Manucharyan[1]

[1]*Department of Physics, Joint Quantum Institute, and Center for Nanophysics and Advanced Materials, University of Maryland, College Park, Maryland 20742, USA*

[2]*Department of Physics, University of Massachusetts-Amherst, Amherst, Massachusetts 01003, USA*

[3]*Department of Physics and Wisconsin Quantum Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*

(Received 25 March 2021; revised 24 December 2021; accepted 18 February 2022; published 15 April 2022)

Large scale quantum computing motivates the invention of two-qubit gate schemes that not only maximize the gate fidelity but also draw minimal resources. In the case of superconducting qubits, the weak anharmonicity of transmons imposes profound constraints on the gate design, leading to increased complexity of devices and control protocols. Here we demonstrate a resource-efficient control over the interaction of strongly-anharmonic fluxonium qubits. Namely, applying an off-resonant drive to noncomputational transitions in a pair of capacitively-coupled fluxoniums induces a ZZ interaction due to unequal ac Stark shifts of the computational levels. With a continuous choice of frequency and amplitude, the drive can either cancel the static ZZ term or increase it by an order of magnitude to enable a controlled-phase (CP) gate with an arbitrary programmed phase shift. The cross-entropy benchmarking of these non-Clifford operations yields a sub 1% error, limited solely by incoherent processes. Our result demonstrates the advantages of strongly-anharmonic circuits over transmons in designing the next generation of quantum processors.

DOI: 10.1103/PhysRevResearch.4.023040

I. INTRODUCTION

Wiring up a pair of superconducting qubits creates an unintentional ZZ term in the two-qubit Hamiltonian, where Z is the single-qubit Pauli σ_z operator. This ZZ interaction arises from the repulsion between computational and noncomputational energy levels of the coupled circuit; the effect would be absent for purely two-level systems. On one hand, such an interaction realizes a controlled-phase logical operation, inducing a phase shift on one qubit depending on the state of the other one [1,2]. On the other hand, a small but nonzero ZZ term would induce coherent errors during single-qubit operations and lead to quantum cross-talk across the qubit register [3–9]. Therefore, any high-fidelity multiqubit system must either achieve an *in situ* control over the ZZ term or have it permanently eliminated. The former approach invokes flux-tunable qubits and coupler circuits [5,10–14], which comes at the price of limiting the qubit coherence time by the $1/f$ flux noise and introducing a new error channel due to the leakage of quantum information into the coupler degrees of freedom [14,15]. The latter approach critically relies on fine tuning the circuit parameters at the nanofabrication stage, which both reduces the device yield and narrows the parameter space available for performance optimization [16–20].

A technologically attractive control scheme would be to tune the ZZ term in a fixed-parameter circuit using a microwave drive [21–23]. Indeed, as long as the drive does not resonantly excite the circuit transitions, its effect would be reduced to ac Stark shifts, which generally modify the level repulsion structure and hence the magnitude of the ZZ term. Such ideas are hard to implement for transmon qubits, largely due to their weak anharmonicity, responsible for leakage of quantum information outside the computational subspace. Recently, a complete suppression of the static ZZ term was demonstrated in a weakly-anharmonic capacitively-shunted flux qubit thanks to breaking the parity selection rule with an external flux bias [24]. However, the qubit is unavoidably exposed to the first-order $1/f$ flux noise, which limits the qubit coherence and hence the gate error.

In this paper, we unleash the full potential of controlling ZZ interaction with an off-resonant microwave pump [21–24]. This is achieved by using strongly-anharmonic qubits—fluxoniums [25,26]—that are coupled by a fixed capacitance, which is compatible with transmon-based processors, and biased at the half-integer flux quantum, practically eliminating the flux-noise decoherence channel. Our key result is that the ZZ term can be either suppressed to zero or enhanced to about 10 MHz on demand, resulting in a resource-efficient controlled-phase (CP) gate free of appreciable spurious effects, such as coherent state leakage. Previously, we have demonstrated a controlled-Z (CZ) gate in such a system by applying a near-resonant drive to the transition between first and second excited states of one of the qubits [27]. In that

[*]These authors contributed equally to this work.

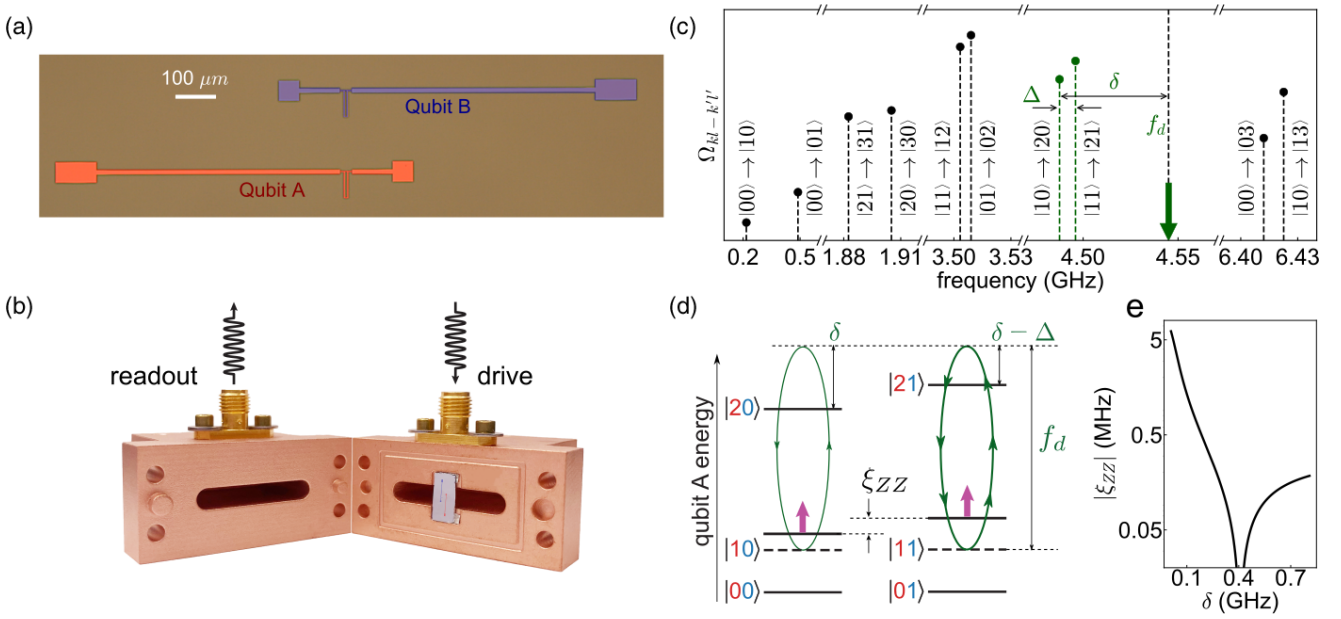

FIG. 1. Differential ac Stark shift in a two-fluxonium circuit. (a) False color picture of the two-fluxonium device. The device is similar to that reported in Ref. [27]. (b) Photograph of the cavity that has the same design as the one used in this experiment. All the microwave drives share the same drive port. The transmitted signal at the cavity frequency is collected from the readout port. The two halves of the cavity are sealed with indium [not shown]. The size of the fluxoniums in this picture is exaggerated for better visibility. (c) Spectrum of allowed transitions of the two-fluxonium circuit extracted from spectroscopy data. The simulated quantity $\Omega_{kl-k'l'}$ defines an on-resonance Rabi frequency that a drive field would induce, assuming the field amplitude is frequency independent. Note that only transitions between states with different parity are allowed. The green arrow indicates the drive frequency f_d used to induce the differential ac Stark shift. (d) Schematic of the energy level diagram for qubit B in the ground state (left) and qubit B in the excited state (right). A drive at frequency f_d pulls the energy level $|11\rangle$ more than the energy level $|10\rangle$, without affecting levels $|00\rangle$ and $|01\rangle$, which is equivalent to a differential ac Stark shift $\xi_{ZZ}^{\text{drive}} > 0$. (e) Calculated total ZZ-interaction rate $\xi_{ZZ} = -|\xi_{ZZ}^{\text{static}}| + \xi_{ZZ}^{\text{drive}}$ for a fixed drive amplitude ($\Omega_{11-21} = 52$ MHz). Note, at an appropriate detuning δ, the total qubit-qubit interaction is switched off, we get $\xi_{ZZ} = 0$.

experiment, the ZZ term was naturally suppressed by the special choice of relatively low qubit frequencies (about 100 MHz). Here we (i) explore a much broader range of qubit frequencies (up to 1.3 GHz), for which the static ZZ term is significant, and (ii) use off-resonant driving, the effect of which can be understood in terms of radiation-dressing and light-shifts.

The practicality of our scheme is illustrated by the performance of the controlled-phase (CP) gate with an arbitrary programmed phase, implemented by rapidly turning the ZZ term on and off. We used the cross-entropy benchmarking technique [5,12,28,29] to measure the CP gate error for a set of 16 equally spaced phase values, obtaining an error of about 3×10^{-3} per radian, while the single qubit error reaches 1×10^{-3}. Notably, both numbers are limited by a manifestly suboptimal coherence in the current experimental setup [30], yet, they are already close to the state-of-the-art for transmons. In fact, our phase-averaged CP gate error of 8×10^{-3} is only few times larger than the best result recently demonstrated on a flux-controlled processor [12,13]. This direct access to the complete family of CP gates is important since they can reduce the required circuit depth in a number of useful algorithms [12,31–35] such as the quantum approximate optimization algorithms (QAOA) and variational quantum eigensolvers (VQE) [36–38]. The minimalism of our qubit-qubit interaction control scheme, enabled in large part by the strong anharmonicity of fluxoniums, can provide a significant advantage for constructing large-scale quantum processors.

II. DIFFERENTIAL AC STARK SHIFT

We describe experiments on two devices with significantly distinct spectral properties. In the main article, we focus on the first device, while the Appendix H supports our conclusions with the data obtained with the second device. The device geometry [shown in Fig. 1(a)] and the measurement setup are similar to those previously reported in Ref. [27]. We conventionally label the coupled energy eigenstates as $|kl\rangle$, where the k and l indices are the uncoupled eigenstates of qubits A and B, respectively. For example, qubit A's (B's) computational transition is labeled $|00\rangle - |10\rangle$ ($|00\rangle - |01\rangle$) and has a frequency $f_A = 217.2$ MHz ($f_B = 488.9$ MHz). While the qubit lifetimes are above 100 μs, the Ramsey coherence times are in the $10 - 14$ μs interval, limited by insufficient thermalization of the measurement lines. The relevant part of the measured two-qubit spectrum is shown in Fig. 1(c). The simulated quantity $\Omega_{kl-k'l'}$ attached to every transition is the Rabi frequency that would be induced by a resonant drive with the same amplitude for all transitions. This frequency scale quantifies the effective coupling of the external drive to the circuit transitions, reflecting both the values of matrix elements of fluxonium charge operators and the asymmetric coupling of the drive field to each qubit [we use only one

port to drive both qubits at shown in Fig. 1(b)]. Details of spectroscopic and time-domain characterization of our device are provided in the Appendix B.

The computational states $|00\rangle$, $|10\rangle$, $|01\rangle$, and $|11\rangle$ are separated in energy from noncomputational states by at least a few GHz, which exceeds the qubit frequencies by almost an order of magnitude. Within the computational subspace, the two-qubit dynamics obeys the Hamiltonian

$$\frac{\hat{H}}{h} = f_A \frac{ZI}{2} + f_B \frac{IZ}{2} + \xi_{ZZ} \frac{ZZ}{4}, \tag{1}$$

where Z is the corresponding Pauli matrix and $\xi_{ZZ} = \xi_{ZZ}^{\text{static}} = -357\,\text{kHz}$ is the ZZ-interaction strength in the present device. The quantity ξ_{ZZ}^{static} has a negative sign because the noncomputational levels push stronger on level $|11\rangle$ than on the other computational levels.

Tuning the magnitude of ξ_{ZZ} by an externally applied microwave drive can be understood as follows. First, let us note that the capacitive coupling leads to much stronger interaction between the noncomputational states, splitting the otherwise degenerate transitions $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ by $\Delta = 8.47\,\text{MHz} \gg |\xi_{ZZ}^{\text{static}}|$, which provides us the resource to dynamically modify ξ_{ZZ}. The same is true for other transition pairs, e.g., $|11\rangle - |12\rangle$, $|01\rangle - |02\rangle$, or $|00\rangle - |03\rangle$, $|10\rangle - |13\rangle$, although the splitting may vary. Consider driving the circuit at the frequency f_d, blue detuned from the $|10\rangle - |20\rangle$ resonance by an amount $\delta \gg \Delta$ [see Fig. 1(d)]. Except for the paired transition $|11\rangle - |21\rangle$, there are no other circuit transitions in the GHz vicinity [see Fig. 1(c)] due to the large anharmonicity. Therefore, the effect of the drive reduces to the attraction between levels in pairs $\{|10\rangle, |20\rangle\}$ and $\{|11\rangle, |21\rangle\}$, while other transitions are not driven. Specifically, if qubit B is in the ground state, there is a positive ac Stark shift $\delta f^{\text{Stark}}(\Omega, \delta) = (\sqrt{\Omega^2 + \delta^2} - \delta)/2$ on the qubit A frequency, which can be thought of as pulling level $|10\rangle$ towards level $|20\rangle$ by the δ-detuned drive. However, this shift is larger when qubit B is in the excited state, because the detuning $(\delta - \Delta)$ is smaller and the effective drive amplitude Ω is larger. Therefore, each qubit acquires a differential ac Stark shift, which is equivalent to modifying the ZZ term in Eq. (1) as $\xi_{ZZ} = \xi_{ZZ}^{\text{static}} + \xi_{ZZ}^{\text{drive}}$, where $\xi_{ZZ}^{\text{drive}} = \delta f^{\text{Stark}}(\Omega_{11-21}, \delta - \Delta) - \delta f^{\text{Stark}}(\Omega_{10-20}, \delta)$. Because $\xi_{ZZ}^{\text{static}} < 0$ and $\xi_{ZZ}^{\text{drive}} > 0$ for typical fluxonium parameters, the total qubit-qubit interaction ξ_{ZZ} in Eq. (1) can be tuned through zero or increased by about an order of magnitude compared to the static value by adjusting the drive frequency and amplitude [see Fig. 1(e) and Fig. 2(b), right panel].

III. TOMOGRAPHY OF DRIVE-TUNED ZZ INTERACTION

We verify that the qubit-qubit interaction indeed takes the form of Eq. (1) using a tomography protocol depicted in Fig. 2(a) [39]. The pulse sequence shown results in the observation of Ramsey-type fringes oscillating at the frequency ξ_{ZZ}. Figure 2(b) shows the measured oscillations by sweeping the driving frequency around $f_d \approx 4.5\,\text{GHz}$ and fixing the amplitude such that $\Omega_{11-21} = 52\,\text{MHz}$. As the drive frequency approaches either the $|10\rangle - |20\rangle$ or $|11\rangle - |21\rangle$ transitions (marked by the black dashed lines at 4.488 GHz and 4.496 GHz, respectively), the ZZ-interaction rate reaches

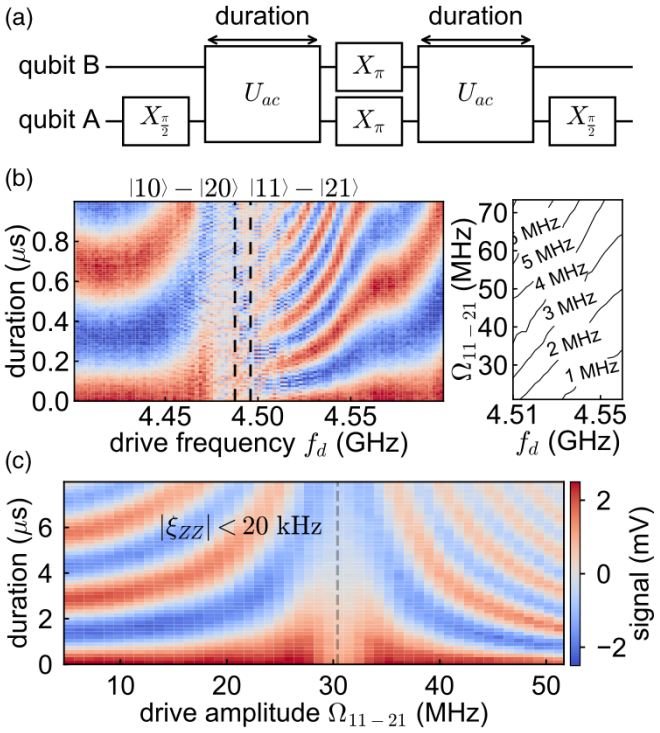

FIG. 2. Tuning the ZZ interaction. (a) Pulse sequence used to measure the interaction rate ξ_{ZZ}. The qubits evolve under U_{ac} that corresponds to the ZZ term in Eq. (1). We use a refocusing pulse on each qubit in the middle of the sequence to cancel single-qubit Z rotations. (b) Induced ZZ interactions (a two-qubit Ramsey-type fringe) as a function of drive frequency and amplitude around the $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ transitions. The color scale is proportional to $\langle ZI \rangle$. The oscillation rate expectedly increases on approaching a resonance condition (left panel). The induced interaction rate reaches about $\xi_{ZZ} = 5\,\text{MHz}$ (right panel). (c) Cancellation of the qubit-qubit interaction. When applying a drive at 4.65 GHz, the ZZ oscillations slow down and speed back up as the drive amplitude passes though the value $\Omega_{11-21} = 30.4\,\text{MHz}$. At this point, we can extract that $\xi_{ZZ} < 20\,\text{kHz}$.

about 6 MHz before the off-resonant ac Stark shift picture breaks down [Fig. 2(b), right inset]. Beyond that point, the drive field and the two noncomputational transitions undergo a coherent energy exchange, witnessed by the detection of rapidly oscillating ripple features in Fig. 2(b). The ZZ interaction can also be controlled by off-resonantly driving the pair of doublets of transitions $|00\rangle - |03\rangle$, $|10\rangle - |13\rangle$, $|01\rangle - |31\rangle$, $|00\rangle - |30\rangle$, near the frequency $f_d' \approx 6.5\,\text{GHz}$, where we obtained $\xi_{ZZ} > 10\,\text{MHz}$ (Appendix G and Fig. 8). Furthermore, with a more optimal structure of the noncomputational transitions, we observed $\xi_{ZZ} > 23\,\text{MHz}$ in the second device (Appendix H).

The immediate application of our differential ac Stark shift phenomenon is to permanently cancel the static ZZ interaction between qubits. This is achieved by applying a microwave tone at the frequency $f_d = 4.65\,\text{GHz}$ and the amplitude such that $\Omega_{11-21} = 30.4\,\text{MHz}$ [see Fig. 2(c)]. As we fine tune the drive amplitude, the two-qubit Ramsey fringe slows down from the static value of 357 kHz to the value $\xi_{ZZ} < 20\,\text{kHz}$, beyond which the oscillations cannot be resolved due to the

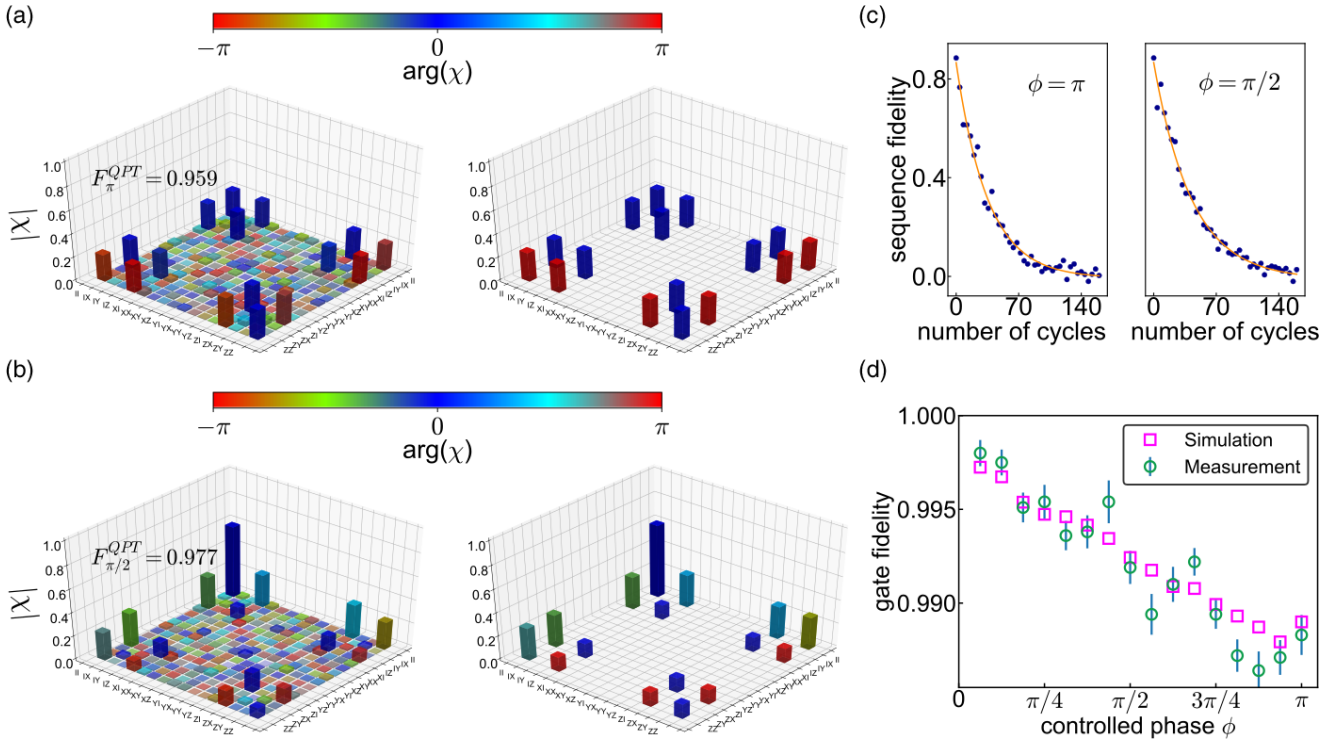

FIG. 3. Benchmarking of arbitrary CP gates. [(a),(b)] Examples of the quantum process tomography (QPT) for the two values of the controlled phase $\phi = \pi$ (top) and $\phi = \pi/2$ (bottom). Left (right) panels indicate experimental (theoretical) values. The process tomography χ matrix reproduces the ideal process with a fidelity $F_{\phi=\pi}^{\mathrm{QPT}} = 0.959$ and $F_{\phi=\pi/2}^{\mathrm{QPT}} = 0.977$. (c) Cross-entropy benchmarking (XEB) for the phases $\phi = \pi/2$ and π corresponding to the process tomography shown in [(a),(b)]. We extract the Pauli and gate errors from the decay of the XEB sequence fidelity with the number of cycles. The gate fidelity at $\phi = \pi/2$ reaches $(99.2 \pm 0.1)\%$. (d) Fidelity of the controlled-phase gate family. The green circles are obtained by cross-entropy benchmarking and the magenta squares are obtained by master equation simulations using the average T_1 and T_2^E shown in Table I (see Appendix I). The gate fidelity averaged over the complete family exceeds 99.2%.

qubit's finite coherence time. We characterized single-qubit operations, in the presence of the ZZ-canceling drive, using conventional randomized benchmarking sequences [40]. In the case of individual benchmarking, the average single-qubit gate fidelity is 0.9969 (45-ns-long pulses) for qubit A and 0.9991 for qubit B (26-ns-long pulses). The simultaneous qubit benchmarking requires longer pulse sequences, which reduces the fidelity to 0.9963 ± 0.0001 for qubit A and 0.9957 ± 0.0001 for qubit B (see Appendix D for details).

IV. ARBITRARY CONTROLLED-PHASE GATE

The ZZ term in Eq. (1) can be switched on and off on a time scale of about 10 ns using Gaussian-edge pulses supplemented with the commonly used derivative removal (DRAG) distortion [41,42]. During the time interval of $\xi_{ZZ}(t) \neq 0$, the states $|10\rangle$ and $|11\rangle$ accumulate different phases ϕ_{10} and ϕ_{11}, which is equivalent to the action of a unitary evolution operator $U = \mathrm{diag}(1, e^{-i\phi_{10}}, 1, e^{-i\phi_{11}})$ in the computational subspace. Using virtual Z rotations [40], the accumulated phase can be entirely assigned to any state such as $|11\rangle$ to implement a controlled-phase operation $U_{\mathrm{CP}}(\phi) = \mathrm{diag}(1, 1, 1, e^{-i\phi})$ where $\phi = \phi_{11} - \phi_{10}$. In principle, one can modulate both the drive frequency and amplitude during the gate pulse, but for simplicity we fixed the drive frequency and tried two different values, $f_d = 4.545$ GHz (detuned by 49 MHz from transition $|11\rangle - |21\rangle$) and $f_d' = 6.665$ GHz (detuned by 55 MHz from $|00\rangle - |30\rangle$, see Appendix G).

We start the CP gate characterization by performing quantum process tomography (QPT). The process tomography matrix (the χ matrix) [43,44] is obtained by preparing 16 independent input states, applying the CP gate, and performing the state tomography of the final quantum state (see Appendix E). By comparing the measured χ-matrix to the theoretical one, we perform the initial tune up of the pulse parameters required to implement the CP gate with the given phase ϕ: gate duration, drive amplitude, DRAG coefficient, and the phases of the virtual Z rotations. Optimized QPT examples for $\phi = \pi$ and $\phi = \pi/2$, with the single-qubit Z rotations adjusted to exhibit only the ZZ evolution, are shown in Figs. 3(a) and 3(b). The χ-matrix fidelity [45] reached 0.959 for $\phi = \pi$ and 0.977 for $\phi = \pi/2$, likely limited by the state preparation and measurement (SPAM) errors.

At $\phi = \pi$ the CP gate belongs to the Clifford group, i.e. it becomes the controlled-Z gate. Hence, it can be characterized using randomized benchmarking (RB), which evades SPAM limitations. Using procedures similar to those described in Ref. [27], we optimized the gate pulses and obtained a CZ gate fidelity of 0.989 ± 0.001 at $f_d = 4.545$ GHz and 0.991 ± 0.001 at $f_d' = 6.665$ GHz. Next, we characterize the CP gate at $\phi = \pi$ using the cross-entropy (XEB) benchmarking technique, which is applicable to non-Clifford operations, and also evades the SPAM errors [5,12,29]. The XEB procedure

consists of a succession of cycles, each composed of one randomly chosen single-qubit Clifford gate on each qubit and a given CP gate. The sequence fidelity is calculated from the cross-entropy between the measured and expected qubit state distribution [29], which decays exponentially with the number of cycles (see Appendix F). The optimal gate parameters are found by optimizing the sequence fidelity at a fixed number of cycles with the Nelder-Mead algorithm. The XEB procedure applied to the CZ gate at $f_d = 4.545$ GHz yields a gate fidelity of 0.988 ± 0.001, which agrees with the results of randomized benchmarking and hence validates the use of XEB for other values ϕ of the controlled phase.

Finally, we apply the XEB procedure to a family of CP gates with the value of ϕ equally spaced by $\pi/16$. The extracted gate error grows approximately linearly in ϕ with a slope of about 3×10^{-3} per radian [Fig. 3(d), circular markers]. For $\phi = \pi/16$, the CP gate error reaches 2×10^{-3}, which is close to the experimental resolution limit. In order to understand the origin of the gate error, we performed detailed master-equation simulations of the driven two-fluxonium system (see Appendix I). Our numerical model closely reproduces the data while relying only on experimentally measured parameters [Fig. 3(d), square markers]. According to the model, the error is entirely due to incoherent processes, while the coherent error is absent down to the 10^{-4} level thanks to the strong anharmonicity of fluxonium's noncomputational transitions.

V. DISCUSSION

The off-resonant drive in our experiment essentially replaces the original (undriven) qubit states with the "dressed" ones, e.g., $|10\rangle \to |10\rangle + (\lambda/2)|20\rangle$ and $|11\rangle \to |11\rangle + (\lambda/2)|21\rangle$ in the case shown in Fig. 1(d), where $\lambda \approx \Omega/\delta$. It is this dressing that leads to the tunable interaction rate ξ_{ZZ} in Eq. (1). One can further show that in the presence of energy relaxation between the undressed single-qubit states $|20\rangle$ and $|10\rangle$ at a rate $1/T_1^{(2\to1)}$, the dressed qubit transition $|00\rangle - |10\rangle$ inherits a pure dephasing rate $\lambda^2/8T_1^{(2\to1)}$. Remarkably, dressing due to the ZZ-cancellation drive is so weak, $\lambda^2 \approx 0.04$, that even for a low relaxation time $T_1^{(2\to1)} \approx 5\,\mu$s in our present experiment, the extra dephasing becomes significant only at the 1 ms level. Thus, the highest measured single-dressed-qubit gate fidelity reached 0.999, and it is limited by the undressed coherence time of about $10 - 15\,\mu$s. In other words, our scheme for eliminating the ZZ interaction has practically no adverse effects, and hence can be generally used to cancel the quantum cross-talk in fluxonium-based processors.

During the CP gate, both ξ_{ZZ} and λ temporarily grow stronger, and hence the energy relaxation of the noncomputational transitions becomes an important contributor to the incoherent gate error. Let us note that the $|1\rangle - |2\rangle$ transition of fluxoniums has the frequency and charge matrix element of typical transmons, whose relaxation time usually belongs to the $10 - 100\,\mu$s range. Therefore, we expect the same range for $T_1^{(2\to1)}$ in fluxoniums once the fabrication and thermalization procedures are properly optimized [30]. In this case, a ZZ-interaction rate in the range $\xi_{ZZ} = 5 - 10$ MHz can be induced with no decoherence exposure at the level of a few hundred microseconds. In fact, the numerical model projects a CP gate error (for $\phi = \pi$) well in the 10^{-4} range for the next generation of fluxonium devices.

VI. CONCLUSIONS

The demonstrated all-microwave control over the ZZ interaction is technologically attractive for scaling-up superconducting quantum processors. Our scheme has the following advantages. First, it does not require close arrangement of qubit frequencies, in contrast to the case of the cross-resonance (CR) gate [17,18] for transmons, which would mitigate the spectral crowding issues [7,17], allows tolerance to fabrication variability, and enables multiplexing of the qubit control. For example, here we used only one input port for the entire experiment. The qubit frequency could be as low as 100 MHz, as in our earlier demonstration of the CZ gate [27], but it can also be around $200 - 500$ MHz (the main device here) or even $700 - 1300$ MHz (the second device here). Second, fluxonium's highly anharmonic spectra allow great flexibility of choosing drive frequencies and to operate the gate in the far-detuned regime, in which higher states are almost unoccupied during the gate operation, reducing leakage error and error due to decoherence of higher states. The required drive power is also comparable to single-qubit rotations [23]. This makes our method highly scalable. Third, our scheme does not require any additional circuitry as in schemes to cancel ZZ coupling using tunable couplers.

Among the previously explored two-qubit gates with a high degree of flexibility, our scheme is most reminiscent of the resonator-induced phase (RIP) gate [22,46,47] for transmons. The RIP gate populates an auxiliary bus mode with off-resonant photons to induce a differential Stark shift. In comparison, our CP gate requires no auxiliary modes or complex pulse-shaping, has closely-confined driven dynamics, and already enables a higher fidelity in devices with very suboptimal coherence, all thanks to the strong anharmonicity of fluxoniums.

Dressing qubits with microwave drives to control their interaction opens up a new route to implement on-demand interaction in a qubit register. We note that similar methods for transmon qubits have been very recently reported by two groups [48,49]. In both cases, the ZZ coupling is controlled by driving off-resonantly computational transitions of superconducting circuits. Moreover, Ref. [49] successfully applied this technique to a multiqubit system, demonstrating the scalability of this approach.

ACKNOWLEDGMENTS

We acknowledge the support from DOE - SC0019449 and the ARO-LPS HiPS Program (W911NF1810146). V.E.M. and M.G.V acknowledge the Faculty Research Award from Google. We acknowledge Lincoln Labs and IARPA for providing a Josephson Traveling Wave Parametric Amplifiers used here to enhance the readout.

APPENDIX A: EXPERIMENTAL SETUP

We use the same readout and initialization procedures described in Ref. [27]. The device is embedded into a rect-

angular copper cavity resonator with a resonance frequency $f_C = 7.538$ GHz and linewidth $\kappa/2\pi = 5$ MHz, thermally anchored to the base plate of a dilution refrigerator at 14 mK. External driving is provided through a single input port to the cavity and the transmission signal is monitored using a stronger coupled output port. Spectroscopy versus flux data was used to accurately extract circuit parameters and calculate transition matrix elements. The table of relevant coherence times is provided below.

We perform a single-shot joint readout of the two qubit states [50] by preamplifying the readout signal with the Josephson traveling parametric-wave amplifier (JTWPA) [51]. The equilibrium population of the two qubit states are obtained by fitting the single-shot histograms with 4 Gaussian distributions, and we compensate the readout error with an empirical model. Prior to each experiment, the qubits are initialized by populating the cavity with a large number of photons, which conveniently prepares the two qubits in a mixed state with the excited state populations of 69% and 82%, respectively. Such a degree of state initialization is sufficient to perform accurate gate error measurement, including the quantum process tomography. We did not focus on improving initialization in this paper. There are some high-fidelity initialization procedures already implemented on single fluxoniums using sideband transitions [25], using a fast feedback loop [52] or on a "heavy" fluxonium [53] by using two driving tone to convert qubit excitations into cavity photons. The details of experimental procedures are provided in the Appendix E.

A schematic of the experimental setup is depicted in Fig. 4. We use standard modulation and demodulation techniques of microwave signals. Intermediate frequency pulses are generated by two M3202A PXIe arbitrary waveform generators (not represented) and the readout signal is digitized by a M3102A PXIe Digitizer (not represented) in the same M9010A PXIe Chassis. The signal departing the cavity is amplified by a JTWPA provided by Lincoln labs [51].

APPENDIX B: SPECTROSCOPY AND TIME-DOMAIN MEASUREMENTS

The two-fluxonium spectrum is obtained with the standard two-tone experiment technique where we readout the cavity after exciting the system with a probe tone. The qubit parameters are extracted by fitting the spectrum at different external flux points with the numerical diagonalization of the Hamiltonian given by Eq. (I2) as shown in Figs. 5(a) and 5(b). The $|1\rangle - |2\rangle$ transition of qubit A splits into two transitions $|11\rangle - |21\rangle$ and $|10\rangle - |20\rangle$ due to its capacitive coupling to qubit B. Off-resonantly driving these two transitions induces the ZZ interaction required to cancel the static ZZ interaction and to perform for CP gates as described in the main text.

TABLE I. Energy relaxation time T_1, Ramsey coherence time T_2^R, and spin echo coherence time T_2^E. The ranges corresponds to time fluctuations during the experiment.

	T_1 (μs)	T_2^R (μs)	T_2^E (μs)
$\|00\rangle - \|10\rangle$	158 – 207	10 – 12	14 – 15
$\|00\rangle - \|01\rangle$	116 – 141	13	20 – 25
$\|11\rangle - \|21\rangle$	4.9 – 6.2	2.6 – 2.8	3.3

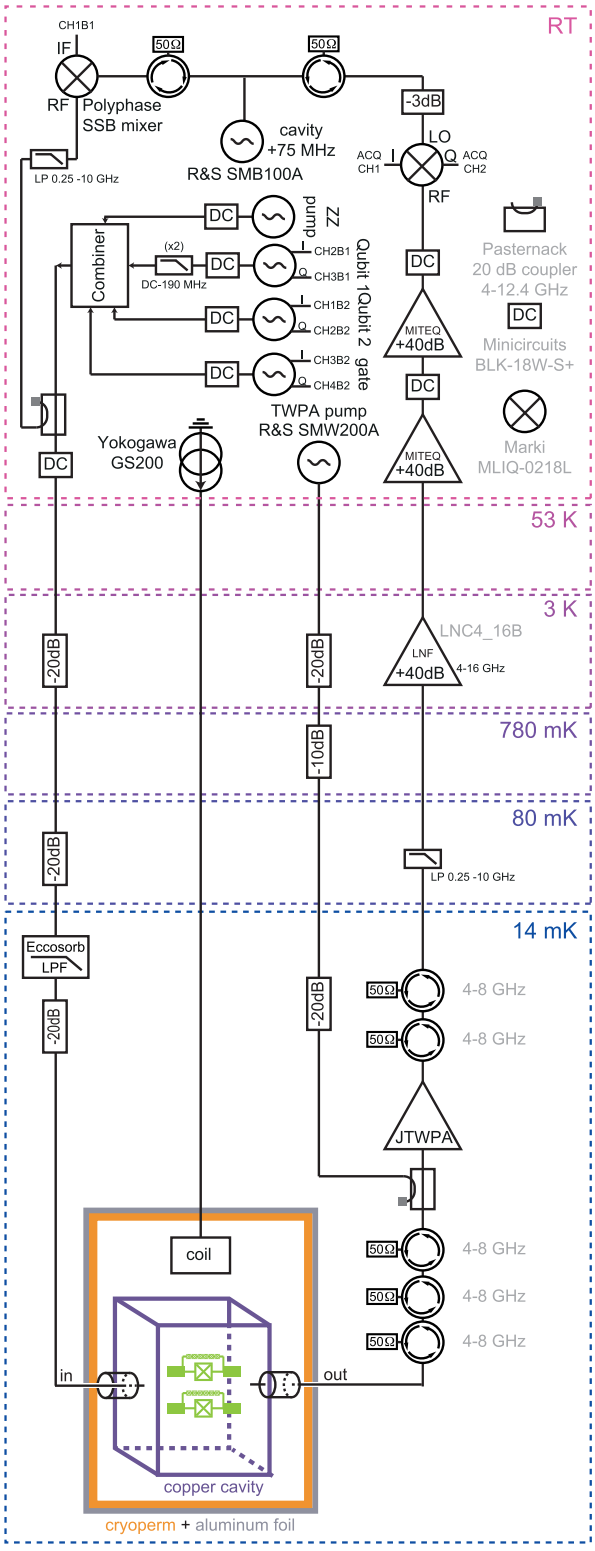

FIG. 4. Schematics of the experimental setup. See labeling on the plot for details.

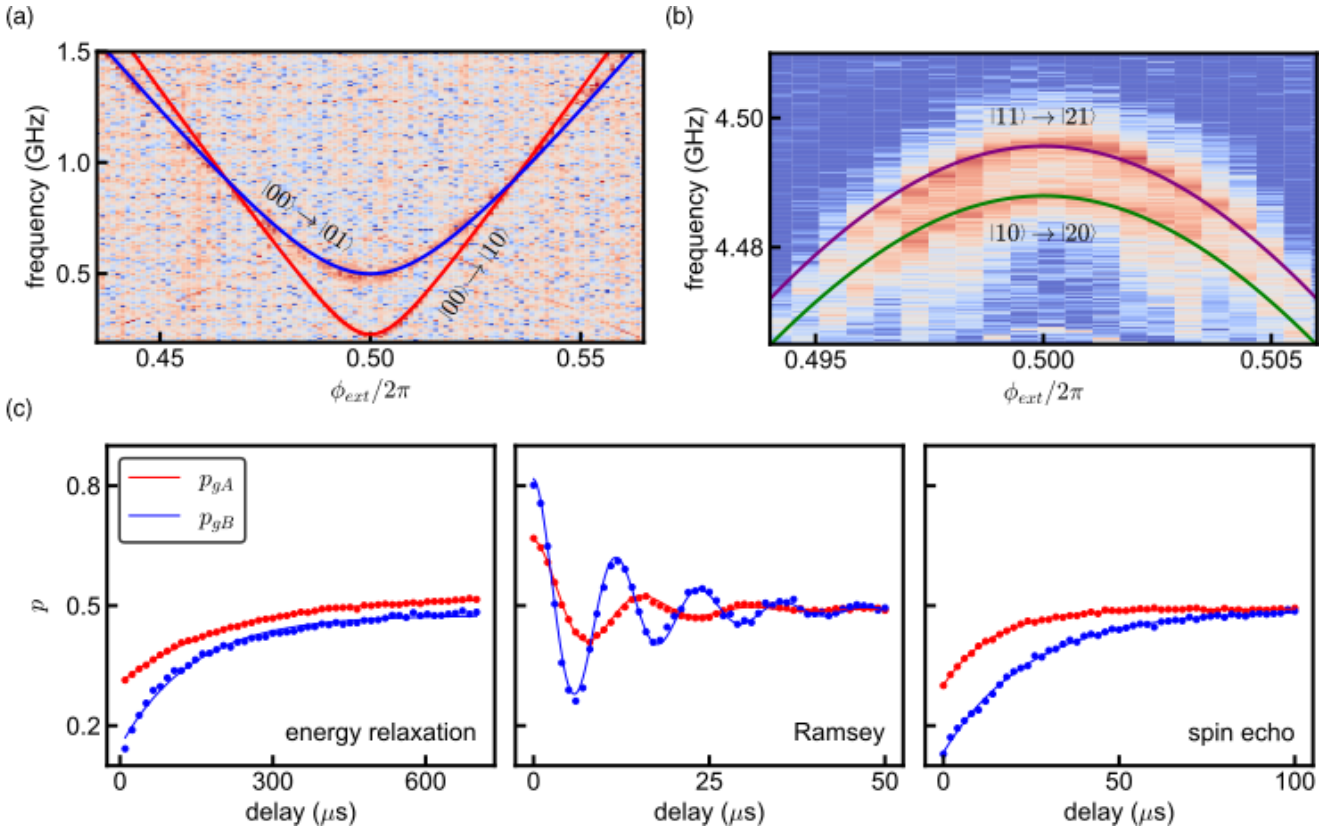

FIG. 5. Two-tone spectroscopy and time-domain characterization. (a) Two-tone spectroscopy of the fluxonium transitions as a function of the external flux threading both fluxonium loops. The $|00\rangle - |10\rangle$ and $|00\rangle - |01\rangle$ transitions show the qubit frequencies at the sweet spot of 217.2 MHz and 488.9 MHz, respectively. (b) Two-tone spectroscopy of the $|11\rangle - |21\rangle$ and $|10\rangle - |20\rangle$ transitions used to induce the ZZ interaction in the main text. (c) Energy relaxation time T_1, Ramsey coherence time T_2^{Ramsey}, and spin-echo coherence time T_2^{Echo} characterization of the qubit transitions.

The coherence time of the fluxonium transitions are extracted using time-domain measurements preceded by an initialization pulse at the cavity frequency. Measuring the energy decay of the qubits from the initialized state yields $T_1^A = (207 \pm 3.7)\,\mu$s and $T_1^B = (141 \pm 5.2)\,\mu$s, while standard Ramsey and spin-echo sequences give the coherence times $T_{2,R}^A = (9.69 \pm 0.33)\,\mu$s, $T_{2,R}^B = (13 \pm 0.4)\,\mu$s, $T_{2,E}^A = (14.3 \pm 0.27)\,\mu$s, $T_{2,E}^B = (24.6 \pm 0.47)\,\mu$s [see Fig. 5(c)]. The fluctuation range of these numbers are reported in Table I.

APPENDIX C: READOUT ERRORS

We adopt an empirical method to compensate for readout errors similar to the one in Ref. [27]. We assume the measured populations p'_{ij} are linked to the populations p_{ij} at the beginning of the readout by a linear transformation $\hat{M}$,

$$\begin{pmatrix} p'_{gg} \\ p'_{ge} \\ p'_{eg} \\ p'_{ee} \end{pmatrix} = \hat{M} \begin{pmatrix} p_{gg} \\ p_{ge} \\ p_{eg} \\ p_{ee} \end{pmatrix}. \tag{C1}$$

The error matrix $\hat{M}$ has 6 parameters that account for the incorrect mapping $|0\rangle \rightarrow |1\rangle$ and $|1\rangle \rightarrow |0\rangle$ for both qubits and transitions between the two qubits during the readout. The error matrix reads

$$M = \begin{pmatrix} 1-a_1-b_1 & b_2 & a_2 & 0 \\ b_1 & 1-a_1-b_2-c_1 & c_2 & a_2 \\ a_1 & c_1 & 1-a_2-b_1-c_2 & b_2 \\ 0 & a_1 & b_1 & 1-a_2-b_2 \end{pmatrix}. \tag{C2}$$

Among the 6 parameters, a_1 and a_2 represent incorrect mappings $|0\rangle_A \rightarrow |1\rangle_A$ and $|1\rangle_A \rightarrow |0\rangle_A$. b_1 and b_2 represent incorrect mappings $|0\rangle_A \rightarrow |1\rangle_B$ and $|1\rangle_B \rightarrow |0\rangle_A$. c_1 and c_2 describe the readout cross-talk where the two qubits swap excitations. In addition to these six parameters, the initial populations for the two qubits p_{gA}^0, p_{gB}^0 are also two

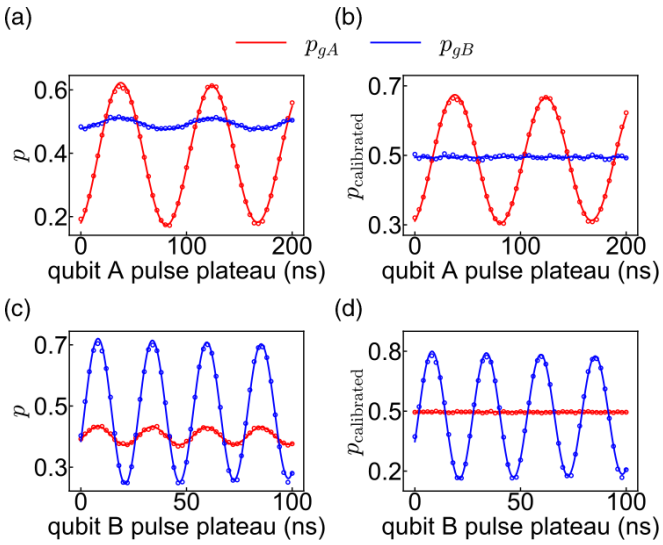

FIG. 6. Rabi oscillations with and without readout errors correction. Oscillations of qubit A [(a),(b)] and qubit B [(c),(d)]. Before correction (left), the Rabi oscillations are not centered around 0.5 because of qubit transitions occurring during the readout. We also correct for readout cross-talk: qubit B readout signal oscillates when qubit A is rotated before correction, and vice versa.

unknown variables that need to be determined. To calibrate all the parameters, we perform Rabi measurements on each qubit while the other qubit is left in the $|+\rangle$ state, as shown in Fig. 6. Assuming that the system starts with the populations

$$\begin{pmatrix} p_{gg} \\ p_{ge} \\ p_{eg} \\ p_{ee} \end{pmatrix} = \begin{pmatrix} p^0_{gA} p^0_{gB} \\ p^0_{gA}\left(1-p^0_{gB}\right) \\ \left(1-p^0_{gA}\right)p^0_{gB} \\ \left(1-p^0_{gA}\right)\left(1-p^0_{gB}\right) \end{pmatrix}, \tag{C3}$$

one can predict the Rabi oscillation amplitudes and offsets in the two Rabi measurements of Fig. 6 as a function of the 6 parameters of the error matrix—and then deduce $\hat{M}$. The corrected populations are obtained by applying $\hat{M}^{-1}$ to the measured populations $(p'_{gg}, p'_{ge}, p'_{eg}, p'_{ee})^T$. With this calibration technique, the populations for both qubits oscillate around 0.5 and only the rotated qubit displays oscillations during the Rabi experiment.

APPENDIX D: SINGLE-QUBIT RANDOMIZED BENCHMARKING

Single-qubit randomized benchmarking (RB) is performed on each qubit—individually and simultaneously—by applying a series of randomly chosen Clifford gates listed in Table II. Z rotations are performed with virtual Z gates [40]—which simply change the phase of X, Y pulses in subsequent gates. Therefore, our Z rotations and identity gates have a zero duration. In the individual single-qubit RB sequence, one Clifford gate contains on average 0.83 physical pulses. As shown in Fig. 7, the average individual single-qubit fidelity is $(99.69 \pm 0.01)\%$ for qubit A with 45-ns-long pulses and $(99.91 \pm 0.003)\%$ for qubit B with 26-ns-long pulses.

TABLE II. Single-qubit Clifford gates used in the RB sequences. Z rotations are realized with virtual Z gates. Only X and Y pulses have a nonzero duration.

Single qubit Clifford group
I, X_π, Y_π, Z_π, $X_{\pi/2}$, $X_{-\pi/2}$, $Y_{\pi/2}$, $Y_{-\pi/2}$, $Z_{\pi/2}$, $Z_{-\pi/2}$,
$Y_{\pi/2}Z_{\pi/2}$, $Y_{-\pi/2}Z_{-\pi/2}$, $Y_{\pi/2}Z_{-\pi/2}$, $Y_{-\pi/2}Z_{\pi/2}$, $X_{\pi/2}Z_{-\pi/2}$, $X_{-\pi/2}Z_{\pi/2}$,
$X_{\pi/2}Z_{\pi/2}$, $X_{-\pi/2}Z_{-\pi/2}$,
$Z_{\pi/2}X_{\pi/2}Z_{\pi/2}$, $Z_{\pi/2}X_{-\pi/2}Z_{\pi/2}$, $Z_{-\pi/2}Y_{\pi/2}Z_{-\pi/2}$, $Z_{-\pi/2}Y_{-\pi/2}Z_{-\pi/2}$,
$Z_{-\pi/2}X_\pi Z_{-\pi}$, $Z_{\pi/2}X_\pi Z_{-\pi}$

In our simultaneous RB sequence, there is a possible idle time for one qubit when the other qubit is rotated on the Bloch sphere. So the simultaneous RB sequence is longer than the individual RB sequences with the same number of Clifford gates. The change of fidelity when operating the qubits simultaneously cannot be directly associated to cross-talks. With such sequences, the simultaneous single qubit fidelity is $(99.63 \pm 0.01)\%$ for qubit A and $(99.57 \pm 0.01)\%$ for qubit B. These numbers are used to extract the CP gate fidelity in the XEB measurements.

APPENDIX E: QUANTUM PROCESS TOMOGRAPHY

We perform a process tomography of the gate operation using a standard procedure described in [43,44]. The quantum process to be characterized is interleaved between one preparation pulse and one tomography pulse chosen in Table III. The process matrix χ characterizes fully the quantum process $\mathcal{E}$. We prepare 16 input states $\{\rho_j\}$ and measure the

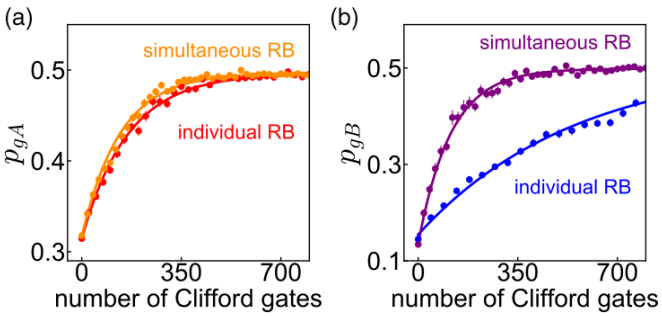

FIG. 7. Randomized benchmarking of single-qubit gates. (a) Population of qubit A and (b) population of qubit B as a function of the sequence length. A list of 51 randomly chosen sequence of Clifford gates are applied to the individually or simultaneously to the qubits before applying a recovery gate. The average single qubit fidelity for qubit A [resp. qubit B] for a gate duration of 45 ns [resp. 26 ns] is $(99.69 \pm 0.01)\%$ [resp. $(99.91 \pm 0.003)\%$] in the individual case and $(99.63 \pm 0.01)\%$ [resp. $(99.57 \pm 0.01)\%$] in the simultaneous case. The duration of a simultaneous operation is given by the longest gate time (gate time of qubit A or gate time of qubit B if a zero-duration pulse is applied to A). The fidelities extracted from the simultaneous RB are used to calculate the CP gate fidelity in Fig. 3 of the main text.

TABLE III. Preparation and tomography pulses used for quantum process tomography (QPT). The first letter refers to the rotation axis and the subscript is the rotation angle. I is the identity. We prepare $d^2 = 16$ input states (where $d = 4$ is the dimension of the Hilbert space). An overcomplete set of 36 pulses is used for the state tomography – combined with a maximum likelihood estimation – to reduce the sensitivity of our tomography to pulse imperfections.

Preparation pulses	Tomography Pulses
$II, IX_{\pi/2},$ $IY_{\pi/2}, IX_{-\pi/2},$ $X_\pi I, X_\pi X_\pi,$ $X_\pi Y_{\pi/2}, X_\pi X_{-\pi/2},$ $Y_{\pi/2} I, Y_{\pi/2} X_\pi,$ $Y_{\pi/2} Y_{\pi/2}, Y_{\pi/2} X_{-\pi/2},$ $X_{-\pi/2} I, X_{-\pi/2} X_\pi,$ $X_{-\pi/2} Y_{\pi/2}, X_{-\pi/2} X_{-\pi/2}$	$II, X_\pi I, X_{\pi/2} I, X_{-\pi/2} I,$ $Y_{\pi/2} I, Y_{-\pi/2} I, IX_\pi, X_\pi X_\pi,$ $X_{\pi/2} X_\pi, X_{-\pi/2} X_\pi, Y_{\pi/2} X_\pi, Y_{-\pi/2} X_\pi,$ $IX_{\pi/2}, X_\pi X_{\pi/2}, X_{\pi/2} X_{\pi/2}, X_{-\pi/2} X_{\pi/2},$ $Y_{\pi/2} Y_{\pi/2}, Y_{-\pi/2} Y_{\pi/2}, IX_{-\pi/2}, X_\pi X_{-\pi/2},$ $X_{\pi/2} X_{-\pi/2}, X_{-\pi/2} X_{-\pi/2},$ $Y_{\pi/2} X_{-\pi/2}, Y_{-\pi/2} X_{-\pi/2},$ $IY_{\pi/2}, X_\pi Y_{\pi/2}, X_{\pi/2} Y_{\pi/2}, X_{-\pi/2} Y_{\pi/2},$ $Y_{\pi/2} Y_{\pi/2}, Y_{-\pi/2} Y_{\pi/2}, IY_{-\pi/2}, X_\pi Y_{-\pi/2},$ $X_{\pi/2} Y_{-\pi/2}, X_{-\pi/2} Y_{-\pi/2},$ $Y_{\pi/2} Y_{-\pi/2}, Y_{-\pi/2} Y_{-\pi/2}$

output states $\mathcal{E}(\rho_j) = \sum_{m,n} \chi_{mn} P_m \rho_j P_n^\dagger$ where P_i are the 16 two-qubit Pauli operators.

1. State tomography

The density matrix of each output state is reconstructed by maximizing the likelihood of a list of 36 measurements obtained by applying the tomography pulses given in Table III. The maximum likelihood estimation (MLE) technique [54] searches the space of density matrices to find the one that is the most likely to reproduce the observations. To calculate the expected output signal for an arbitrary density matrix, we first calibrate the measurement operator

$$\hat{M} = \beta_{II} \mathbb{I} \otimes \mathbb{I} + \beta_{IZ} \mathbb{I} \otimes \sigma_Z + \beta_{ZI} \sigma_Z \otimes \mathbb{I} + \beta_{ZZ} \sigma_Z \otimes \sigma_Z \tag{E1}$$

where β_{ij} are complex coefficients and σ_i are Pauli matrices. β_{ij} are measured by applying the pulses $\{II, IX_\pi, X_\pi I, X_\pi X_\pi\}$ to the initial state before measuring the output signal. After calibrating the measurement operator, we can predict the output signal for a state ρ after any of the 36 tomography pulses. We parametrize the density matrix with 16 real parameters as $\rho = \mathcal{T}^\dagger \mathcal{T} / \mathrm{Tr}(\mathcal{T}^\dagger \mathcal{T})$, where $\mathcal{T}$ is a lower triangular matrix given by

$$\mathcal{T} = \begin{pmatrix} t_1 & 0 & 0 & 0 \\ t_5 + it_6 & t_2 & 0 & 0 \\ t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\ t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4 \end{pmatrix}. \tag{E2}$$

This Cholesky decomposition ensures that ρ corresponds to a physical quantum state (see Sec. V of the supplementary material of Ref. [55]). The state ρ is then obtained by maximizing the likelihood of the 36 output measurements over the 16 parameters of $\mathcal{T}$.

2. Process matrix

The process matrix χ is then calculated by linear inversion following the procedure described in Ref. [43]. This procedure is valid in our case even if the input states ρ_j are not pure states because they are still linearly independent. The gate fidelity [45] compared to the ideal process matrix χ_{ideal} is $F(\chi, \chi_{\mathrm{ideal}}) = (\mathrm{Tr}(\chi^\dagger \chi_{\mathrm{ideal}}) d + 1)/(d + 1)$, where d is the dimension of the Hilbert space.

APPENDIX F: CROSS-ENTROPY BENCHMARKING (XEB)

The corrected populations are used to calculate the cross-entropy between the experimental populations and the expected populations starting from the state after the initialization pulse. The cross entropy yields the XEB sequence fidelity. By fitting the sequence fidelity versus number of cycles with $Ap^m + B$, the Pauli error per cycle r^P_{cycle} is given by

$$r_{\mathrm{cycle}} = \frac{N-1}{N}(1-p) \tag{F1}$$

$$r^P_{\mathrm{cycle}} = \frac{N+1}{N} r_{\mathrm{cycle}} \tag{F2}$$

where N is the dimension of the system. Then we extract the CP gate Pauli error r^P_{CP} from the equation $(1 - r^P_{\mathrm{cycle}}) = (1 - r^P_A)(1 - r^P_B)(1 - r^P_{\mathrm{CP}})$, where r^P_A, r^P_B are the average single-qubit Pauli error measured from the simultaneous RB experiment described in Appendix D. The arbitrary CP gate fidelity is calculated by converting Pauli error back to the gate error r_{CP} as shown in Figs. 3(c) and 3(d).

APPENDIX G: ZZ INTERACTIONS INDUCED WHEN DRIVING AROUND 6.5 GHz

As stated in the main text, our gate scheme is applicable by driving near any transition from the computational subspace to higher states with a large transition dipole. Indeed, ZZ interactions can be induced using the four transitions $|00\rangle - |03\rangle$, $|10\rangle - |13\rangle$, $|01\rangle - |31\rangle$, and $|00\rangle - |30\rangle$. Figure 8 shows ZZ oscillations similar to the one of Fig. 2 of the main text by driving around 6.5 GHz. We find that the induced interaction rate in this frequency range can exceed the one achievable in the vicinity of the $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ transitions—and reach values greater than 10 MHz.

Our optimal CZ gate in this frequency window is obtained at $f'_d = 6.665$ GHz (blue line in Fig. 8) with a drive amplitude $\Omega_{00-30} = 91$ MHz. Because the detuning to the four transitions are on the same order of magnitude compared to the drive amplitude, all four transitions contribute to the induced ZZ interaction during the gate. Even though more transitions are involved when performing a CZ gate at f'_d compared to the gate at $f_d = 4.545$ GHz, the gate fidelity is still higher (99.1% at f'_d compared to 98.9% at f_d) due to the larger ZZ-interaction rate.

APPENDIX H: OVER 20 MHZ ZZ INTERACTIONS

We realize the same type of CP gate at a faster speed (40 ns for a CZ gate) in the second two-fluxonium device measured in a different laboratory. This device has a similar

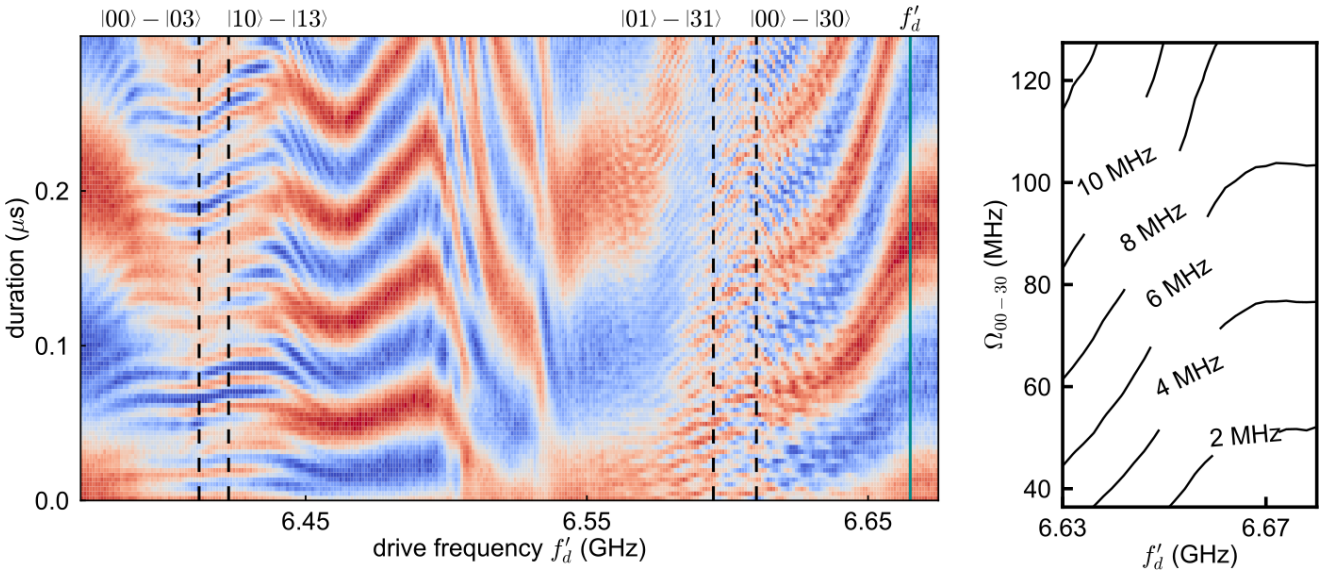

FIG. 8. ZZ interactions induced by driving around 6.5 GHz. We use the pulse protocol depicted in Fig. 2(a) of the main text to measure ZZ oscillations (left panel). At this drive frequency, the four transitions $|00\rangle - |03\rangle$, $|10\rangle - |13\rangle$, $|01\rangle - |31\rangle$, and $|00\rangle - |30\rangle$ contribute to the ac Stark shift in the computational subspace. We achieve interaction rates exceeding 10 MHz (right panel).

design but with different parameters shown in Table IV. The coupling constant J_C is comparable with that in the main text, which does not require large coupling capacitors. Despite abnormally low coherence times in this particular device, we measured a CZ gate fidelity of $(97.6 \pm 0.3)\%$ using RB.

This device has a stronger static ZZ interaction $\xi_{ZZ}^{\text{static}} = -2.1$ MHz. We use an always-on tone (detuned from the $|11\rangle - |21\rangle$ transition by $\delta = 80$ MHz, Rabi rate $\Omega = 43$ MHz) to cancel the ZZ interaction at all times. We perform simultaneous randomized benchmarking to characterize our single-qubit gates and obtain gate fidelities of $(99.35 \pm 0.03)\%$ and $(99.16 \pm 0.03)\%$ for qubits A and B, respectively.

The single-qubit gate fidelity is predominantly limited by the low coherence times of the computational transitions compared with the single-qubit gate time of 24 ns. Qubit A (707 MHz) has $T_1 = 11\,\mu$s, $T_2^R = 4.2\,\mu$s, $T_2^E = 5.3\,\mu$s, Qubit B (1310 MHz) has $T_1 = 8.1\,\mu$s, $T_2^R = 1.6\,\mu$s, $T_2^E = 2.4\,\mu$s. The $|1\rangle - |2\rangle$ transitions for both qubits have $T_1 = 0.7\,\mu$s, $T_2^R = 1.0\,\mu$s, $T_2^E = 1.3\,\mu$s on average. The low coherence time is suspected to be associated with dielectric loss deterioration during the mailing or packaging.

By increasing the amplitude of the same tone above $\Omega = 190$ MHz, we can obtain an induced ZZ coupling strength $\xi_{ZZ}^{\text{ind}} = 23$ MHz to implement a fast CP gate. Using a flat-topped Gaussian microwave pulse, we achieve a CZ gate time of 40 ns. We characterize this two-qubit gate with interleaved RB and show the results in Fig. 9. The data is produced from 45 randomized gate sequences and the two-qubit Clifford group is constructed with on average 5.583 physical single-qubit gates (incorporating virtual Z gates for the single-qubit Z rotations) and 1.5 CZ gates per Clifford gate. The CP gate fidelity extracted from the interleaved RB measurement is $(97.6 \pm 0.3)\%$. This value is slightly lower than the simulated error 98.7% from Appendix I, which should come from the insufficient optimization of the pulse parameters or the temporal fluctuation of the relaxation and dephasing time.

TABLE IV. Charging, inductive, Josephson energy, and coupling constant of the device discussed in Appendix H.

	$E_{C,i}$ (GHz)	$E_{L,i}$ (GHz)	$E_{J,i}$ (GHz)	J_C (GHz)
Qubit A	1.1	0.84	3.5	0.33
Qubit B	1.0	1.7	4.0	

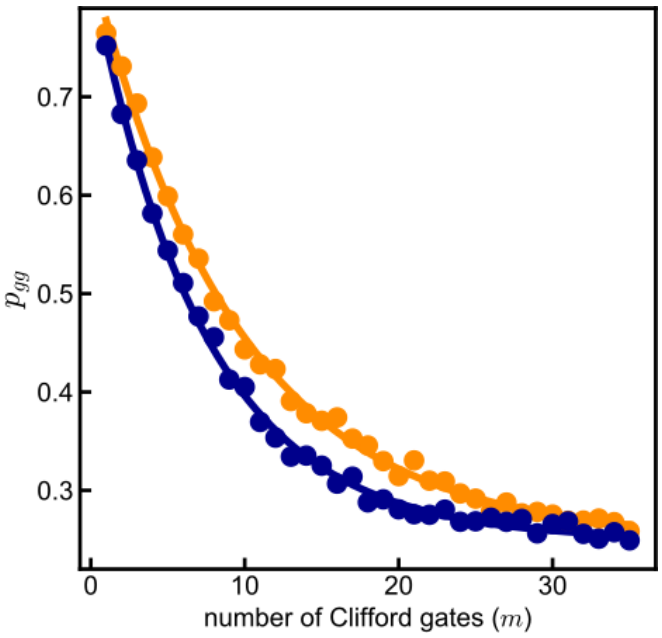

FIG. 9. Interleaved randomized benchmarking of the CZ gate for the second device. The blue and yellow circles depict the interleaved and reference sequence. We extract the CZ gate fidelity $(97.6 \pm 0.3)\%$.

TABLE V. Parameters of the Hamiltonian (I2) extracted from the two-tone spectroscopy measurements and predicted qubit transition frequencies and charge matrix elements.

	$E_{C,\alpha}$ (GHz)	$E_{L,\alpha}$ (GHz)	$E_{J,\alpha}$ (GHz)	J_C (GHz)	f_{01}^{α} (GHz)	f_{12}^{α} (GHz)	$\lvert\langle 0\rvert\hat{n}_\alpha\lvert 1\rangle_\alpha\rvert$	$\lvert\langle 1\rvert\hat{n}_\alpha\lvert 2\rangle_\alpha\rvert$
Qubit A	1.051	0.753	5.263	0.248	0.217	4.489	0.066	0.576
Qubit B	1.069	0.771	3.870		0.489	3.510	0.131	0.559

In all measurements of this device, we did not optimize our readout parameters sufficiently to perform single-shot joint readout of the two-qubit states. As an alternative, to measure the population distribution of any two-qubit state $\vec{P} = [P_{gg}, P_{ge}, P_{eg}, P_{ee}]^T$, we repeatedly prepare the same state and measure the cavity transmission following four different single-qubit rotations: no rotation, π rotation on qubit B, π rotation on qubit A, and π rotation on both qubits. These measurements yield a 4-vector of complex transmission coefficients $\vec{V} = [V_{II}, V_{I\pi}, V_{\pi I}, V_{\pi\pi}]^T$. This measurement signal can be converted to the population vector via a calibrated measurement matrix M: $\vec{P} = M\vec{V}$,

$$\begin{pmatrix} p_{gg} \\ p_{ge} \\ p_{eg} \\ p_{ee} \end{pmatrix} = \begin{pmatrix} M_{gg} & M_{ge} & M_{eg} & M_{ee} \\ M_{ge} & M_{gg} & M_{ee} & M_{eg} \\ M_{eg} & M_{ee} & M_{gg} & M_{ge} \\ M_{ee} & M_{eg} & M_{ge} & M_{gg} \end{pmatrix}^{-1} \begin{pmatrix} V_{II} \\ V_{I\pi} \\ V_{\pi I} \\ V_{\pi\pi} \end{pmatrix}. \quad \text{(H1)}$$

The M matrix is calibrated and updated throughout all experiments by measuring $\vec{V}$ of the initial state of the system $\vec{P}_i$ (which typically is about $[0.87, 0.04, 0.09, 0.00]^T$ following a driven reset protocol). $\vec{P}_i$ is determined self-consistently with additional calibration routines.

APPENDIX I: THEORETICAL ANALYSIS

In this Appendix, we analyze the controlled-phase (CP) gate with fluxonium qubits theoretically. First, we simulate the unitary dynamics during the gate operation, calculate the unitary error, and discuss coherent leakage to higher noncomputational levels. Second, we estimate the incoherent error caused by relaxation and dephasing, which are the dominant sources of error in our devices. Our estimate of this error is in good agreement with experimental gate infidelities in Fig. 3 of the main text. Finally, we discuss effects of the ZZ-cancellation tone.

1. Unitary dynamics and coherent leakage to higher levels

We describe the unitary dynamics of the driven system of two capacitively coupled fluxonium qubits by the Hamiltonian

$$\hat{H} = \hat{H}_{\text{static}} + \hat{H}_{\text{drive}}(t). \quad \text{(I1)}$$

Here the static part of the Hamiltonian reads (α labels the two circuits A and B)

$$\frac{\hat{H}_{\text{static}}}{h} = \sum_{\alpha=A,B}\left[4E_{C,\alpha}\hat{n}_\alpha^2 + \frac{1}{2}E_{L,\alpha}\hat{\varphi}_\alpha^2 - E_{J,\alpha}\cos(\hat{\varphi}_\alpha - \phi_{\text{ext},\alpha})\right] + J_C\hat{n}_A\hat{n}_B, \quad \text{(I2)}$$

where $E_{C,\alpha}$ is the charging energy (antenna) of fluxonium α, $E_{L,\alpha}$ is its inductive energy (junction array), $E_{J,\alpha}$ is Josephson energy (small junction), and $\phi_{\text{ext},\alpha}$ is proportional to the external magnetic flux threading the loop formed by the junction array and the small junction of qubit α. Here we operate at the flux sweet spot where $\phi_{\text{ext},\alpha} = \pi$. The operators $\hat{\varphi}_\alpha$ and $\hat{n}_\alpha$ are the generalized flux and Cooper-pair number operators of fluxonium α, respectively, which satisfy the commutation relation $[\hat{\varphi}_\alpha, \hat{n}_{\alpha'}] = i\delta_{\alpha\alpha'}$. Table V gives the experimental values of the parameters in the Hamiltonian (I2) obtained by fitting two-tone spectroscopy measurements to the numerical diagonalization of the Hamiltonian and calculated values of single-qubit transition frequencies and charge matrix elements. The notation f_{kl}^{α} stands for the frequency of the $|k\rangle_\alpha - |l\rangle_\alpha$ transition. Our qubits are coupled by an interaction term $J_C\hat{n}_A\hat{n}_B$ with a coupling constant $J_C = 0.248$ GHz. This affects two-qubit transition frequencies, which are shown in Fig. 1(c) of the main text.

We model the time-dependent drive term in the Hamiltonian (I1) by

$$\frac{\hat{H}_{\text{drive}}(t)}{h} = (\epsilon_A\hat{n}_A + \epsilon_B\hat{n}_B)[g_x(t)\cos(2\pi f_d t) + g_y(t)\sin(2\pi f_d t)]. \quad \text{(I3)}$$

Here the envelope functions $g_x(t)$ and $g_y(t)$ describe two independent quadrature controls. Assuming that they are normalized as $\sqrt{g_x^2 + g_y^2} = 1$ for a continuous microwave tone, we find that the on-resonance Rabi frequency for the two-qubit transition $|kl\rangle - |k'l'\rangle$ is given by

$$\Omega_{kl-k'l'} = |\langle kl|(\epsilon_A\hat{n}_A + \epsilon_B\hat{n}_B)|k'l'\rangle|. \quad \text{(I4)}$$

Here $|kl\rangle$ is the eigenstate of the interacting Hamiltonian (I2) that is connected adiabatically to the noninteracting state $|k\rangle_A \otimes |l\rangle_B$. To match the experimentally measured induced ZZ rate of $\xi_{ZZ}^{\text{ind}} = 2.9$ MHz at $f_d = 4.545$ GHz at the drive power corresponding to $\Omega_{11-21} = 52.4$ MHz, we choose $\epsilon_B/\epsilon_A = 1.3$ in the drive term (I3). Equation (I4) has been used in simulating Rabi frequencies shown in Fig. 1(c) in the main text.

For the control $g_x(t)$, we use a Gaussian flat-topped pulse with the length of each of the rising and lowering edges t_{rise} and the duration of the flat part t_{flat}. Up to a normalization constant, the rising edge of the pulse at $0 < t < t_{\text{rise}}$ is given by

$$g_x(t) \propto \exp\left[-\frac{(t - t_{\text{rise}})^2}{2\sigma^2}\right] - \exp\left[-\frac{t_{\text{rise}}^2}{2\sigma^2}\right], \quad \text{(I5)}$$

where $\sigma = t_{\text{rise}}/\sqrt{2\pi}$, and the lowering edge is given by a similar expression at $t_{\text{gate}} - t_{\text{rise}} < t < t_{\text{gate}}$, where $t_{\text{gate}} = 2t_{\text{rise}} + t_{\text{flat}}$. For the orthogonal quadrature, we use the DRAG

approach [41], which gives

$$g_y(t) = \alpha_{\rm DRAG}\frac{dg_x}{dt}, \tag{I6}$$

where $\alpha_{\rm DRAG}$ is the DRAG coefficient adjusted to avoid leakage errors during the gate.

To calculate the unitary gate errors, we first simulate the operator of evolution between $t = 0$ and $t = t_{\rm gate}$ for a system described by the Hamiltonian (I1) and project it into the computational subspace to obtain $\hat{U}$. To properly compare $\hat{U}$ with the target CP gate operator $\hat{U}_{\rm CP}(\phi)$, where

$$\hat{U}_{\rm CP}(\phi) = {\rm diag}(1, 1, 1, e^{-i\phi}), \tag{I7}$$

we adjust $\hat{U}$ with single-qubit Z rotations as follows. We define the phase accumulation for operator $\hat{U}$ as

$$\phi_U = \phi_{00} + \phi_{11} - \phi_{10} - \phi_{01}, \tag{I8}$$

where $\phi_{kl} = -{\rm arg}\langle kl|\hat{U}|kl\rangle$ is the opposite of the phase of the corresponding diagonal matrix element. The combination (I8) is invariant under single-qubit Z rotations. It results in the phase mismatch

$$\delta\phi = \phi_U - \phi \tag{I9}$$

between the accumulated and target phases. We then calculate $\hat{U}' = \hat{U}_Z\hat{U}$, where

$$\hat{U}_Z = {\rm diag}[e^{i(\phi_{00}-\delta\phi/4)}, e^{i(\phi_{10}+\delta\phi/4)}, e^{i(\phi_{01}+\delta\phi/4)}, e^{i(\phi_{11}-\delta\phi/4-\phi)}] \tag{I10}$$

is the product of two single-qubit Z rotations and an overall phase factor. Using the operator $\hat{U}'$, we use the standard expression for the two-qubit gate fidelity [56]:

$$F = \frac{{\rm Tr}[(\hat{U}')^\dagger\hat{U}'] + |{\rm Tr}[\hat{U}_{\rm CP}^\dagger(\phi)\hat{U}']|^2}{20}. \tag{I11}$$

There are two major sources of coherent errors: the phase error and the error due to leakage to noncomputational levels. To calculate the phase error, we take $\hat{U} = {\rm diag}(e^{-i\phi_{00}}, e^{-i\phi_{01}}, e^{-i\phi_{10}}, e^{-i\phi_{11}})$ and find using Eq. (I10) that $\hat{U}' = {\rm diag}(e^{-i\delta\phi/4}, e^{i\delta\phi/4}, e^{i\delta\phi/4}, e^{-i\delta\phi/4-i\phi})$. The infidelity $1 - F$ calculated using Eq. (I11) then gives the phase error

$$\mathcal{E}_\phi = \frac{4}{5}\sin^2\left(\frac{\delta\phi}{4}\right). \tag{I12}$$

The coherent leakage error is the average leakage probability, which is given by

$$P_{\rm leak} = 1 - \tfrac{1}{4}{\rm Tr}(\hat{U}^\dagger\hat{U}). \tag{I13}$$

Since $\hat{U}$ is obtained following the projection into the computational subspace, it is generally nonunitary, so $P_{\rm leak}$ can be nonzero.

Using this procedure, we calculated coherent errors for 16 target phases $\phi = \pi/16, \pi/8, \ldots, \pi$, discussed in the main text. For each target phase, we used experimental pulse duration parameters $t_{\rm rise}$ and $t_{\rm flat}$ and optimized $1 - F$ numerically over the DRAG parameter $\alpha_{\rm DRAG}$, the overall drive amplitude, and the drive frequency, which we allowed to vary within a ± 5 MHz window around its experimental value $f_d = 4.545$ GHz. For each of the 16 CP gates optimized this way, we found that $1 - F < 10^{-4}$, $P_{\rm leak} < 10^{-4}$, and $\mathcal{E}_\phi < 10^{-5}$. We show these errors as a function of $t_{\rm flat}$ in Fig. 10(a) and 10(b) for $\phi = \pi/2$ and $\phi = \pi$, respectively, where we keep other parameters fixed at their optimal values. For $\phi = \pi/2$, Fig. 10(a), we notice that $P_{\rm leak} < 10^{-4}$ for any value of $t_{\rm flat}$, which is explained by long edges of the pulse with $t_{\rm rise} = 50$ ns. Therefore, for this set of gate parameters, we can obtain a complete family of CP gates with the total unitary error $1 - F < 10^{-4}$ for any target phase ϕ by simply varying $t_{\rm flat}$. We have checked that in this case, the accumulated phase (I8) is a linear function of $t_{\rm flat}$ and is given by $\phi_U = 2\pi\xi_{ZZ}t_{\rm flat}$ up to a constant with $\xi_{ZZ} = 4.3$ MHz. In comparison, a shorter $t_{\rm rise} = 10$ ns has been used in simulations and experiment for the target phase $\phi = \pi$, see Fig. 10(b), which resulted in maxima of $P_{\rm leak}(t_{\rm flat})$ around 10^{-2}. In this case, the optimized gate requires simultaneous elimination of leakage in both $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ transitions at the end of the pulse, which happens at sharp minima of $P_{\rm leak}(t_{\rm flat})$ in Fig. 10(b).

In Fig. 10(c), we show simulations for the target phase $\phi = \pi$ for the second device, which is discussed in detail in Appendix H. This device has a stronger repulsion between $|12\rangle$ and $|21\rangle$, which results in a larger detuning between frequencies of the $|10\rangle - |20\rangle$ and $|11\rangle$ - $|21\rangle$ transitions and, therefore, in a faster CP gate. In simulations, we have chosen $\epsilon_A = \epsilon_B$ in the drive term (I3) to match the experimental value of $\Omega_{11-21}/\Omega_{10-20}$. To match experimental time parameters, we have used $t_{\rm rise} = 18$ ns and $\sigma = 30/\sqrt{2\pi}$ ns in the Gaussian edge (I5). As is evident from Fig. 10(c), this device allows a fast CP, while maintaining a small leakage $P_{\rm leak} < 10^{-3}$. Here, it is a specific device spectrum rather than a short gate duration that prevents us from easily achieving leakage error below 10^{-4} as in Figs. 10(a) and 10(b). For this device, the $|01\rangle - |02\rangle$ and $|11\rangle - |12\rangle$ transitions are additional important leakage channels as their frequencies are only 100–200 MHz below those of the $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ transitions. This makes it harder to match off-resonant Rabi oscillations in all the leaking transitions to achieve sharp minima as in Fig. 10(b). Thus, a longer Gaussian edge as in Fig. 10(a) or more advanced pulse shaping is required to obtain $1 - F < 10^{-4}$.

2. Incoherent error

In the present experiment, the actual gate fidelity is limited by decoherence rather than coherent errors. To estimate this limit, we use a simple model in the rotating-wave approximation with only six levels in the Hilbert space: four computational states as well as states $|20\rangle$ and $|21\rangle$. Thus, we use

$$\frac{\hat{H}_{\rm RWA}}{h} = -\delta|20\rangle\langle 20| - (\delta - \Delta)|21\rangle\langle 21| + \frac{g_x(t)}{2}[\Omega_{10-20}(|10\rangle\langle 20| + |20\rangle\langle 10|) + \Omega_{11-21}(|11\rangle\langle 21| + |21\rangle\langle 11|)]$$
$$- \frac{ig_y(t)}{2}[\Omega_{10-20}(|10\rangle\langle 20| - |20\rangle\langle 10|) + \Omega_{11-21}(|11\rangle\langle 21| - |21\rangle\langle 11|)], \tag{I14}$$

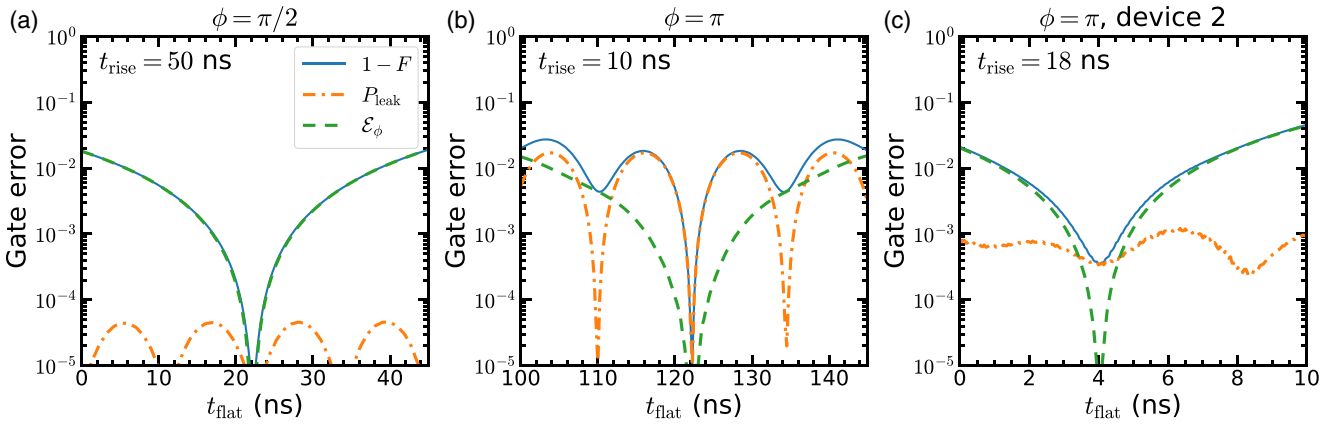

FIG. 10. Simulated coherent gate errors. The total unitary error $1-F$ (solid lines), leakage error $P_{\rm leak}$ (dash-dotted lines), and the phase error $\mathcal{E}_\phi$ (dashed lines) are shown vs $t_{\rm flat}$, the duration of the flat part of the pulse, for the target phases $\phi = \pi/2$ (a) and π (b) and for the second device that is described in Appendix H with the target phase of π (c). Each gate is first optimized for a fixed Gaussian flat-topped envelope $g_x(t)$ with experimental values of $t_{\rm flat}$, $t_{\rm rise}$, and σ, see Eq. (I5), and then simulated at other values of $t_{\rm flat}$.

where $\delta = f_d - f_{10-20} \approx 57$ MHz and $\Delta = f_{11-21} - f_{10-20} \approx 8$ MHz, see Fig. 1(c) in the main text. In addition, here $\Omega_{11-21}/\Omega_{10-20} = 1.114$ to match the ratio ϵ_B/ϵ_A in Eq. (I3) and we use the same pulse-shaping envelopes as in Eqs. (I5) and (I6). We find that contributions due to decoherence of both computational and noncomputational transitions are important in our device. The relevant time parameter for the first contribution is the total gate duration $2t_{\rm rise} + t_{\rm flat}$, while the relevant time parameter for the second contribution is the effective duration $t_{\rm rise} + t_{\rm flat}$. Similarly to the calculation of unitary errors, we use the experimental values of $t_{\rm rise}$ and $t_{\rm flat}$ to better match our estimates with measurements.

For given target phase accumulation ϕ and experimental $t_{\rm rise}$ and $t_{\rm flat}$, we optimize the unitary gate error over the drive amplitude and the DRAG coefficient $\alpha_{\rm DRAG}$. We calculate this error following the same procedure as in Appendix I 1, but for the Hamiltonian (I14). Once the optimal parameters are found, we perform simulations of nonunitary dynamics using Lindblad master equation. For the 6×6 density matrix ρ, it has the form

$$\frac{d\hat{\rho}}{dt} = -\frac{i}{\hbar}[\hat{H}_{\rm RWA}, \hat{\rho}] + \sum_k \left[\hat{L}_k \hat{\rho} \hat{L}_k^\dagger - \frac{1}{2}(\hat{L}_k^\dagger \hat{L}_k \hat{\rho} + \hat{\rho} \hat{L}_k^\dagger \hat{L}_k)\right], \quad \text{(I15)}$$

where we use six collapse operators $\hat{L}_k$. We form them by continuing tensor products of single-qubit collapse and identity operators into interacting states. For example, we take $|0\rangle_A\langle 1|_A \otimes (|0\rangle_B\langle 0|_B + |1\rangle_B\langle 1|_B)$ to obtain $|00\rangle\langle 10| + |01\rangle\langle 11|$, which is exact for an uncoupled system. Thus, for the relaxation in the main qubit transitions, we use the following collapse operators:

$$\hat{L}_1^A = \sqrt{\Gamma_1^A}(|00\rangle\langle 10| + |01\rangle\langle 11|), \quad \text{(I16a)}$$

$$\hat{L}_1^B = \sqrt{\Gamma_1^B}(|00\rangle\langle 01| + |10\rangle\langle 11| + |20\rangle\langle 21|). \quad \text{(I16b)}$$

We notice that while the unitary dynamics of states $|00\rangle$ and $|01\rangle$ is independent of that of other levels in our simplified model (I14), these states are no longer uncoupled from the rest of the Hilbert space once incoherent channels are accounted for. For the pure dephasing of the main qubit transitions, we use

$$\hat{L}_\varphi^A = \sqrt{2\Gamma_\varphi^A}(|00\rangle\langle 00| + |01\rangle\langle 01|), \quad \text{(I17a)}$$

$$\hat{L}_\varphi^B = \sqrt{2\Gamma_\varphi^B}(|00\rangle\langle 00| + |10\rangle\langle 10| + |20\rangle\langle 20|). \quad \text{(I17b)}$$

For relaxation and pure dephasing of the $|1\rangle_A \to |2\rangle_A$ transition of qubit A, we use

$$\hat{L}_1^{1-2,A} = \sqrt{\Gamma_1^{1-2,A}}(|10\rangle\langle 20| + |11\rangle\langle 21|), \quad \text{(I18a)}$$

$$\hat{L}_\varphi^{1-2,A} = \sqrt{2\Gamma_\varphi^{1-2,A}}(|20\rangle\langle 20| + |21\rangle\langle 21|). \quad \text{(I18b)}$$

In these equations, the relaxation and dephasing rates are given by $\Gamma_1 = 1/T_1$ and $\Gamma_\varphi = 1/T_2^E - 1/2T_1$, where T_1 and T_2^E are the relaxation and T_2 echo times of the corresponding transition. In simulations, we use average values of T_1 and T_2^E for the ranges shown in Table I.

After solving the master equation (I15) for an initial state $|\psi_0\rangle$ at $t = 0$, we find the density matrix $\hat{\rho}$ describing the system at time $t = t_{\rm gate}$. We calculate the state fidelity as $F_\rho = {\rm Tr}(\hat{\rho}\hat{\rho}_{\rm ideal})$, where $\rho_{\rm ideal} = |\psi_{\rm ideal}\rangle\langle\psi_{\rm ideal}|$ with $|\psi_{\rm ideal}\rangle = \hat{U}_Z^\dagger \hat{U}_{\rm CP}(\phi)|\psi_0\rangle$ and $\hat{U}_Z$ describing virtual Z rotations used to calculate unitary gate error, see Eq. (I10). We then estimate the gate error $1-F$ by averaging $1-F_\rho$ over 36 initial two-qubit states generated from the set of six initial single-qubit states $\{|0\rangle, |1\rangle, (|0\rangle \pm |1\rangle)/\sqrt{2}, (|0\rangle \pm i|1\rangle)/\sqrt{2}\}$. Gate errors calculated this way are shown by squares in Fig. 3(d) of the main text.

The simulated gate error for the accumulated phase $\phi = \pi$ is $1 - F \approx 1.1 \times 10^{-2}$, which agrees well with the experimental value. A natural question to ask is how small this error can be in devices with longer coherence times. For $T_1 =$

$T_2 = 500\ \mu$s of single-qubit transitions and $T_1 = T_2 = 50\ \mu$s of the $|1\rangle_A \to |2\rangle_A$ transition, we find that this error reduces to 7×10^{-4}. When these relaxation and coherence times are further increased by a factor of two to 1000 and 100 μs, we estimate $1 - F \approx 4 \times 10^{-4}$. To obtain $1 - F < 10^{-4}$, we additionally need to change the device parameters of Table V to allow for a shorter gate. For example, this can be achieved by increasing J_C by a factor of 2, which increases Δ by a factor of 4 and thus reduces the shortest possible gate duration by the same factor.

A similar approach to estimate gate error due to relaxation and dephasing applied to the second device, which is discussed in Fig. 10(c) and in Appendix H, gives the error of about 1.3%. Since in that device, the frequencies of the $|01\rangle - |02\rangle$ and $|11\rangle - |12\rangle$ transitions are only 100–200 MHz away from those of the $|10\rangle - |20\rangle$ and $|11\rangle - |21\rangle$ transitions, we extended the model (I14) to add levels $|02\rangle$ and $|12\rangle$ and modified the set of collapse operators accordingly with the extra relaxation and dephasing channels.

3. Effect of the ZZ-cancellation tone

We have seen in our simulations that the relaxation and dephasing of the $|1\rangle_A - |2\rangle_A$ transition of qubit A is responsible for about half of the CP gate error because the population of state $|2\rangle_A$ can become as large as 20% during gate operation. Similarly, the microwave tone that is used to cancel static ZZ-interaction results in modified relaxation and dephasing channels within the computational subspace as expected from the hybridization of $|10\rangle$ with $|20\rangle$ and $|11\rangle$ with $|21\rangle$. However, in comparison to a fast CP gate, where $\Omega_{11\to21} \sim \delta - \Delta$, the ZZ-cancellation tone has a lower power of $\Omega_{11-21} \approx 30$ MHz and a larger detuning $|f_d - f_{11-21}| \approx 150$ MHz. In the rotating frame [the frame of the Hamiltonian (I14)], this leads to a weak hybridization between computational and higher levels. Keeping only terms that are linear in $\lambda = \Omega_{11-21}/|f_d - f_{11-21}| \approx \Omega_{10-20}/|f_d - f_{10-20}| \approx 0.2$, we find the following dressed states [eigenstates of the Hamiltonian (I14)]:

$$|11\rangle_\lambda = |11\rangle + \frac{\lambda}{2}|21\rangle, \tag{I19a}$$

$$|21\rangle_\lambda = -\frac{\lambda}{2}|11\rangle + |21\rangle, \tag{I19b}$$

and similarly for $|10\rangle_\lambda$ and $|20\rangle_\lambda$. The set of states $\{|00\rangle, |01\rangle, |10\rangle_\lambda, |11\rangle_\lambda\}$ is the computational subspace in the presence of the cancellation tone (the dressed computational subspace). When the system is in one of the dressed states $|10\rangle_\lambda$ or $|11\rangle_\lambda$, the population of the bare state $|20\rangle$ or $|21\rangle$ is only $\lambda^2/4 \approx 0.01$, so we anticipate only a weak reduction of the relaxation and dephasing rates of qubit transitions in comparison to their values at $\lambda = 0$.

To calculate these rates accurately, one needs to apply the transformation given by Eqs. (I19a) and (I19b) to the Hamiltonian describing system-bath interaction, trace out bath degrees of freedom, and obtain a master equation in the new basis [57]. Here we only estimate the effect of the interplay of the cancellation tone and relaxation and dephasing of the $|1\rangle_A - |2\rangle_A$ transition to demonstrate that such an effect is small. To this end, we apply the transformation given by Eqs. (I19a) and (I19b) to the collapse operators (I18a) and (I18b). To the zeroth order in λ, we find the collapse operators having the same forms as Eqs. (I18a) and (I18b) but written in terms of the dressed states. The corrections to them are given by

$$\delta\hat{L}_1^{1-2,A} = \frac{\lambda}{2}\sqrt{\Gamma_1^{1-2,A}}(|10\rangle_\lambda\langle 10|_\lambda - |20\rangle_\lambda\langle 20|_\lambda + |11\rangle_\lambda\langle 11|_\lambda - |21\rangle_\lambda\langle 21|_\lambda) + O(\lambda^2), \tag{I20a}$$

$$\delta\hat{L}_\varphi^{1-2,A} = \frac{\lambda}{2}\sqrt{2\Gamma_\varphi^{1-2,A}}(|10\rangle_\lambda\langle 20|_\lambda + |20\rangle_\lambda\langle 10|_\lambda + |11\rangle_\lambda\langle 21|_\lambda + |21\rangle_\lambda\langle 11|_\lambda) + O(\lambda^2). \tag{I20b}$$

We observe that the relaxation channel introduces a small dephasing of the $|1\rangle_A - |2\rangle_A$ transition and vice versa. In addition, the terms with $|10\rangle\langle 10|$ and $|11\rangle\langle 11|$ in the right-hand side of Eq. (I20a) yield additional dephasing of computational transitions. This additional dephasing can be characterized by the rate Γ'_ϕ defined as $\sqrt{2\Gamma'_\phi} \approx (\lambda/2)\sqrt{\Gamma_1^{1-2,A}}$. For our parameters, we find $1/\Gamma'_\phi \approx 8/(\lambda^2\Gamma_1^{1-2,A}) \approx 1$ ms, which is much longer than qubit dephasing times in our device. This number is comparable with longest fluxonium coherence times reported so far [30,58]. As we anticipate a significant improvement in T_1 of the $|1\rangle \to |2\rangle$ transitions of future devices, $1/\Gamma'_\phi$ is expected to become much longer. Therefore, we do not predict any significant increase in the dephasing rate caused by the cancellation tone in both current and future devices.

We notice that Eqs. (I20a) and (I20b) do not predict directly any relaxation within the computational subspace. Cancellation-tone-induced relaxation is possible in a more detailed model (I2), which properly accounts for the capacitive interaction. With the interaction-induced hybridization of states $|10\rangle$ and $|01\rangle$ taken into account, an additional relaxation channel is possible with the rate Γ'_1, where $\sqrt{\Gamma'_1} \sim \lambda\nu\sqrt{\Gamma_1^{1\to2,A}}$, $\nu = J_C n_{01}^A n_{01}^B/|f_{01}^A - f_{01}^B| \approx 0.008$ describes the hybridization strength, and $n_{kl}^\alpha = |\langle k|\hat{n}_\alpha|l\rangle|$. Alternatively, qubit relaxation can be facilitated by a stronger hybridization of states $|21\rangle$ and $|12\rangle$ with the strength $J_C n_{12}^A n_{12}^B/|f_{12}^A - f_{12}^B| \approx 0.08$. For both of these cases, we find that the theory predicts a correction to the qubit relaxation rate negligible compared to the energy relaxation rates of the qubit transitions of our devices.

[1] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Quantum Logic Gates for Coupled Superconducting Phase Qubits, Phys. Rev. Lett. **91**, 167005 (2003).

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

[3] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Characterization of Addressability by Simultaneous Randomized Benchmarking, Phys. Rev. Lett. **109**, 240504 (2012).

[4] 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).

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[6] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing Context-Dependent Errors in Quantum Processors, Phys. Rev. X **9**, 021045 (2019).

[7] S. Krinner, S. Lazar, A. Remm, C. K. Andersen, N. Lacroix, G. J. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, Benchmarking Coherent Errors in Controlled-Phase Gates Due to Spectator Qubits, Phys. Rev. Applied **14**, 024042 (2020).

[8] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, Correlated randomized benchmarking, arXiv:2003.02354.

[9] A. Morvan, V. V. Ramasesh, M. S. Blok, J. M. Kreikebaum, K. O'Brien, L. Chen, B. K. Mitchell, R. K. Naik, D. I. Santiago, and I. Siddiqi, Qutrit Randomized Benchmarking, Phys. Rev. Lett. **126**, 210504 (2021).

[10] 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).

[11] P. Zhao, P. Xu, D. Lan, J. Chu, X. Tan, H. Yu, and Y. Yu, High-Contrast ZZ Interaction Using Superconducting Qubits with Opposite-Sign Anharmonicity, Phys. Rev. Lett. **125**, 200503 (2020).

[12] B. Foxen, C. Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, S. Boixo, D. Buell, B. Burkett, Y. Chen, R. Collins *et al.*, Demonstrating a Continuous Set of Two-qubit Gates for Near-term Quantum Algorithms, Phys. Rev. Lett. **125**, 120504 (2020).

[13] V. Negîrneac, H. Ali, N. Muthusubramanian, F. Battistel, R. Sagastizabal, M. S. Moreira, J. F. Marques, W. Vlothuizen, M. Beekman, N. Haider, A. Bruno, and L. DiCarlo, High-Fidelity Controlled-Z Gate with Maximal Intermediate Leakage Operating at the Speed lLimit in a Superconducting Quantum Processor, Phys. Rev. Lett. **126**, 220502 (2021).

[14] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim, A. Melville, B. M. Niedzielski, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Realization of High-Fidelity CZ and ZZ-free iSWAP Gates with a Tunable Coupler, Phys. Rev. X **11**, 021058 (2021).

[15] M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Lazar, J. C. Besse, T. Walter, A. Wallraff, and C. Eichler, Implementation of Conditional Phase Gates Based on Tunable ZZ Interactions, Phys. Rev. Lett. **125**, 240502 (2020).

[16] 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).

[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(R) (2016).

[18] 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).

[19] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa *et al.*, Demonstration of quantum volume 64 on a superconducting quantum computing system, arXiv:2008.08571.

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

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

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

[23] S. Krinner, P. Kurpiers, B. Royer, P. Magnard, I. Tsitsilin, J. C. Besse, A. Remm, A. Blais, and A. Wallraff, Demonstration of an All-Microwave Controlled-Phase Gate between Far-Detuned Qubits, Phys. Rev. Applied **14**, 044039 (2020).

[24] A. Noguchi, A. Osada, S. Masuda, S. Kono, K. Heya, S. P. Wolski, H. Takahashi, T. Sugiyama, D. Lachance-Quirion, and Y. Nakamura, Fast parametric two-qubit gates with suppressed residual interaction using the second-order nonlinearity of a cubic transmon, Phys. Rev. A **102**, 062408 (2020).

[25] 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).

[26] V. E. Manucharyan, N. A. Masluk, A. Kamal, J. Koch, L. I. Glazman, and M. H. Devoret, Evidence for coherent quantum phase slips across a Josephson junction array, Phys. Rev. B **85**, 024521 (2012).

[27] 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).

[28] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya *et al.*, A blueprint for demonstrating quantum supremacy with superconducting qubits, Science **360**, 195 (2018).

[29] R. Barends, C. M. Quintana, A. G. Petukhov, Y. Chen, D. Kafri, K. Kechedzhi, R. Collins, O. Naaman, S. Boixo, F. Arute, K. Arya, D. Buell, B. Burkett, Z. Chen, B. Chiaro, A. Dunsworth, B. Foxen, A. Fowler, C. Gidney, M. Giustina *et al.*, Diabatic Gates for Frequency-Tunable Superconducting Qubits, Phys. Rev. Lett. **123**, 210501 (2019).

[30] A. Somoroff, Q. Ficheux, R. A. Mencia, H. Xiong, R. V. Kuzmin, and V. E. Manucharyan, Millisecond coherence in a superconducting qubit, arXiv:2103.08578.

[31] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. L. Heras, R. Babbush, A. G. Fowler, B. Campbell, Y. Chen *et al.*, Digitized adiabatic quantum computing with a superconducting circuit, Nature (London) **534**, 222 (2016).

[32] P. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus *et al.*, Scalable Quantum Simulation of Molecular Energies, Phys. Rev. X **6**, 031007 (2016).

[33] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[34] M. Ganzhorn, D. J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer, Phys. Rev. Applied **11**, 044092 (2019).

[35] I. Aleiner, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, A. Bourassa *et al.*, Accurately computing electronic properties of materials using eigenenergies, arXiv:2012.00921.

[36] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[37] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Simulation of electronic structure Hamiltonians using quantum computers, Mol. Phys. **109**, 735 (2011).

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

[39] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, and M. Steffen, Implementing a strand of a scalable fault-tolerant quantum computing fabric, Nat. Commun. **5**, 4015 (2014).

[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] 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).

[42] 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).

[43] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[44] 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(R) (2013).

[45] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit, Phys. Rev. Lett. **102**, 090502 (2009).

[46] A. W. Cross and J. M. Gambetta, Optimized pulse shapes for a resonator-induced phase gate, Phys. Rev. A **91**, 032325 (2015).

[47] S. Puri and A. Blais, High-Fidelity Resonator-Induced Phase Gate with Single-Mode Squeezing, Phys. Rev. Lett. **116**, 180501 (2016).

[48] B. K. Mitchell, R. K. Naik, A. Morvan, A. Hashim, J. M. Kreikebaum, B. Marinelli, W. Lavrijsen, K. Nowrouzi, D. I. Santiago, and I. Siddiqi, Hardware-Efficient Microwave-Activated Tunable Coupling Between Superconducting Qubits, Phys. Rev. Lett. **127**, 200502 (2021).

[49] K. X. Wei, E. Magesan, I. Lauer, S. Srinivasan, D. F. Bogorin, S. Carnevale, G. A. Keefe, Y. Kim, D. Klaus, W. Landers, *et al.*, Quantum crosstalk cancellation for fast entangling gates and improved multi-qubit performance, arXiv:2106.00675.

[50] 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).

[51] C. Macklin, D. Hover, M. E. Schwartz, X. Zhang, W. D. Oliver, and I. Siddiqi, A near-quantum-limited Josephson traveling-wave parametric amplifier, Science **350**, 307 (2015).

[52] R. Gebauer, N. Karcher, D. Gusenkova, M. Spiecker, L. Grünhaupt, I. Takmakov, P. Winkel, L. Planat, N. Roch, W. Wernsdorfer, A. V. Ustinov, M. Weber, M. Weides, I. M. Pop, and O. Sander, State preparation of a fluxonium qubit with feedback from a custom FPGA-based platform, AIP Conf. Proc. **2241**, 020015 (2020).

[53] H. Zhang, S. Chakram, T. Roy, N. Earnest, Y. Lu, Z. Huang, D. Weiss, J. Koch, and D. I. Schuster, Universal Fast-Flux Control of a Coherent, Low-Frequency Qubit, Phys. Rev. X **11**, 011010 (2020).

[54] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).

[55] T. Roy, S. Kundu, M. Chand, S. Hazra, N. Nehra, R. Cosmic, A. Ranadive, M. P. Patankar, K. Damle, and R. Vijay, Implementation of Pairwise Longitudinal Coupling in a Three-Qubit Superconducting Circuit, Phys. Rev. Applied **7**, 054025 (2017).

[56] L. H. Pedersen, N. M. Møller, and K. Mølmer, Fidelity of quantum operations, Phys. Lett. A **367**, 47 (2007).

[57] M. Boissonneault, J. M. Gambetta, and A. Blais, Dispersive regime of circuit QED: Photon-dependent qubit dephasing and relaxation rates, Phys. Rev. A **79**, 013819 (2009).

[58] 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).