# Automatic characterization and calibration of a superconducting quantum processor capable of error correction*

V. Negîrneac,[1,2,†] H. Ali,[1,3,†] N. Muthusubramanian,[1,3] F. Battistel,[1]
R. Sagastizabal,[1,3] M. S. Moreira,[1,3] J. F. Marques,[1,3] W. J. Vlothuizen,[1,4]
M. Beekman,[1,4] C. Zachariadis,[1,3] N. Haider,[1,4] A. Bruno,[1,3] and L. DiCarlo[1,3]

[1] *QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands*
[2] *Instituto Superior Técnico, Lisbon, Portugal*
[3] *Kavli Institute of Nanoscience, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands*
[4] *Netherlands Organisation for Applied Scientific Research (TNO), P.O. Box 96864, 2509 JG The Hague, The Netherlands*

(Dated: January 26, 2021)

Simple tuneup of two-qubit gates is essential for the scaling of quantum processors. We introduce the sudden variant (SNZ) of the Net Zero scheme realizing controlled-Z (CZ) two-qubit gates by baseband flux control of transmon frequency. SNZ CZ gates operate at the speed limit of transverse coupling between computational and non-computational states by maximizing intermediate leakage to the latter. The key advantage of SNZ is tuneup simplicity, owing to the regular structure of conditional phase and leakage as a function of two control parameters. We realize SNZ CZ gates in a transmon processor, achieving 99.93 % [99.54(27) % average] fidelity and 0.10 % [0.18(4) % average] leakage on the best pair. SNZ is compatible with scalable schemes for quantum error correction and adaptable to generalized conditional-phase gates useful in intermediate-scale applications.

## CONTENTS



## I. INTRODUCTION

Transmons – a type of superconducting charge qubit – are one of the most promising physical implementations for quantum computing that have witnessed fierce research efforts since its introduction more than a decade ago [2]. Superconducting quantum processors have recently reached important milestones [3], notably the demonstration of quantum supremacy on a 53-transmon processor [4]. Toward quantum error correction (QEC) and fault tolerance [5], recent experiments have used repetitive parity measurements to stabilize two-qubit entanglement [6, 7] and to perform surface-code quantum error detection in a 7-transmon processor [8]. These developments have relied on two-qubit controlled-phase (CPHASE) gates realized by dynamical flux control of transmon frequency, harnessing the transverse coupling $J_2$ between a computational state $|11\rangle$ and a non-computational state such as $|02\rangle$ [9, 10]. Compared to other implementations, e.g., cross-resonance using microwave-frequency pulses [11] and parametric radio-frequency pulsing [12], baseband flux pulses achieve the fastest controlled-Z (CZ) gates (a special case of CPHASE), operating near the speed limit $t_{\rm lim} = \pi/J_2$ [13].

Over the last decade, baseband flux pulsing for two-qubit gating has evolved in an effort to increase gate fidelity and to reduce leakage and residual ZZ coupling. In particular, leakage into non-computational states became a main focus for its negative impact on QEC,

* This article is an extended version of the preprint Ref. 1.
† These authors contributed equally to this work.

adding complexity to error-decoder design [14] and requiring hardware and operational overhead to seep back into the computation space [15–19]. To reduce leakage from linear-dynamical distortion in flux-control lines and limited time resolution in arbitrary waveform generators (AWGs), unipolar square pulses [10, 20] have been superseded by softened counterparts [21, 22] based on fast-adiabatic theory [23]. In parallel, coupling strengths have reduced to $J_2/2\pi \sim 10-20$ MHz in the effort to reduce residual ZZ coupling, which affects single-qubit gates and idling at bias points, and produces crosstalk from spectator qubits [24]. Many groups are actively developing tunable coupling schemes to suppress residual coupling without incurring the slowdown of the gates [25–29].

A main limitation to the fidelity of flux-based CPHASE gates is dephasing from flux noise, as one qubit is displaced 0.5−1 GHz below its flux-symmetry point (i.e., sweetspot [30]) to reach the $|11\rangle$-$|02\rangle$ resonance. To address this limitation, Ref. 31 introduced a bipolar variant [termed Net Zero (NZ)] of the fast-adiabatic scheme, which provides a built-in echo reducing the impact of low-frequency flux noise. The double use of the transverse interaction also reduces leakage by destructive interference, as understood by analogy with a Mach-Zehnder interferometer (MZI). Finally, the zero-average characteristic avoids the buildup of long-timescale distortions in the flux-control lines, significantly improving gate repeatability. NZ pulsing was successfully used in several recent experiments [6, 8, 32], elevating the state of the art for CZ gate fidelity to $99.72\pm0.35\%$ [3]. However, NZ suffers from complicated tuneup, owing to the complex dependence of conditional phase and leakage on fast-adiabatic pulse parameters. This limits the use of NZ for two-qubit gating as quantum processors grow in qubit count.

In this article, we introduce the sudden variant (SNZ) of the NZ scheme implementing CZ gates using baseband flux pulsing. SNZ offers two advantages while preserving the built-in echo, destructive leakage interference, and repeatability characteristic of conventional NZ (CNZ). First, SNZ operates at the speed limit of transverse coupling by maximizing intermediate leakage to the non-computational state. The second and main advantage is greatly simplified tuneup: the landscapes of conditional phase and leakage as a function of two pulse parameters have very regular structure and interrelation, easily understood by exact analogy to the MZI. We realize SNZ CZ gates among four pairs of nearest neighbors in a seven-transmon processor and characterize their performance using two-qubit interleaved randomized benchmarking (2QIRB) with modifications to quantify leakage [31, 33, 34]. The highest performance achieved has $99.93 \pm 0.24\%$ fidelity with $0.10 \pm 0.02\%$ corresponding leakage. SNZ CZ gates are fully compatible with scalable approaches to QEC [35]. The generalization of SNZ to arbitrary CPHASE gates is straightforward and useful for optimization [36], quantum simulation [37], and other noisy intermediate-scale quantum (NISQ) applications [38].

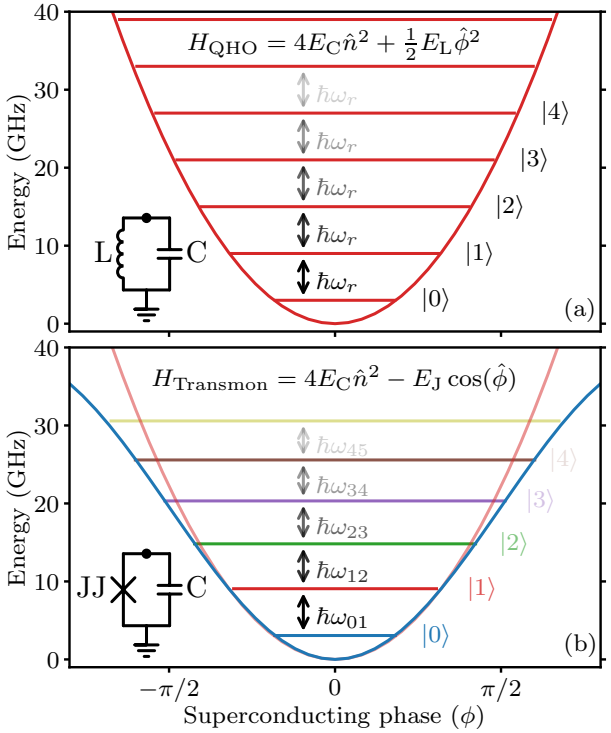


Figure 1. (a) Energy potential of the Quantum Harmonic Oscillator (QHO) and (b) the Transmon. The QHO has a parabolic potential, resulting in equally spaced energy levels with spacing $\hbar\omega_r$. By replacing the inductor with a Josephson junction a nonlinearity is introduced in the Hamiltonian affecting the spacing of the levels which now can be addressed individually by a microwave pulse. The $E_C$ energy is easily set by the capacitor geometry while $E_J$ is very sensible to the fabrication process of the Josephson junction. Adapted from Ref. 39.

## II. TRANSMON AND DEVICE OVERVIEW

This section is a brief overview of the quantum chip, its transmons, single-qubit operations (gates), quantum state readout and the flux-based two-qubit gates.

Despite micrometer scales, simple electrical circuits made of superconducting metals exhibit quantization effects at very low temperatures ($\sim 10$ mK). For example, an LC circuit [Fig. 1(a), inset] behaves as a Quantum Harmonic Oscillator (QHO) with equally spaced energy levels. Although the first two levels resemble a qubit, in practice, the equally spaced ladder does not allow to target specific transitions. However, if the inductor is replaced by a Josephson junction – a nonlinear element – the transitions become anharmonic [Fig. 1(b)]. This enables the individual transitions to be addressed with microwave pulses implementing single-qubit operations (gates) for a practical qubit – the transmon.

If we add one more Josephson junction in parallel and apply a magnetic field (by applying a current through a nearby wire) such that there will be an external magnetic flux $\Phi$ in the closed loop formed by the two junctions (called SQUID loop) as shown in Fig. 2(c), we obtain

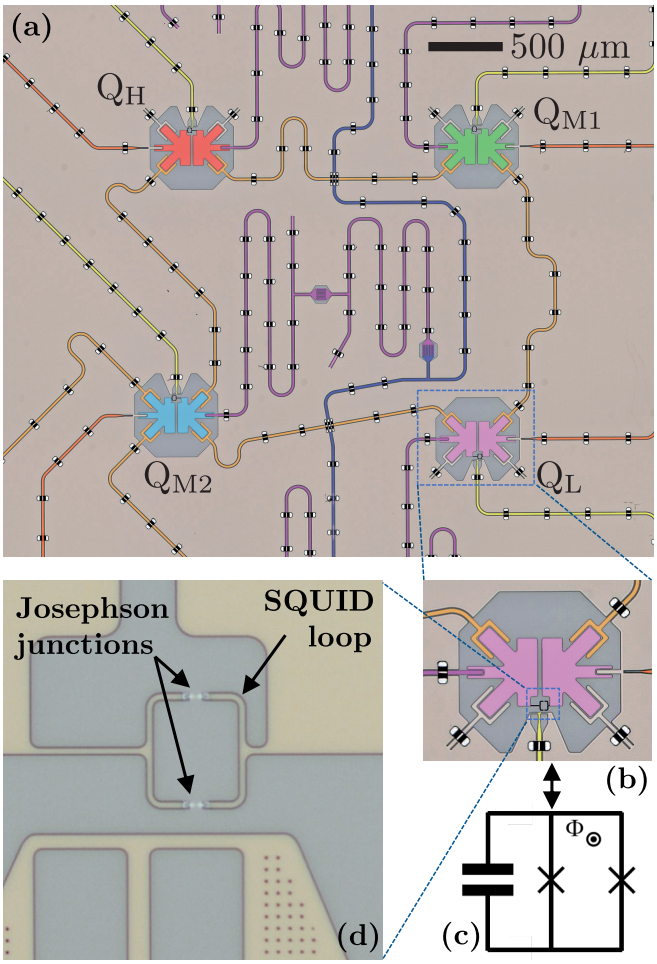


Figure 2. Thin film device and transmon details. (a) False-colored optical image of the device, zoomed in to the four transmons used in this study. Transmons $Q_H$ (red) and $Q_L$ (pink) each connect to $Q_{M1}$ (green) and $Q_{M2}$ (cyan) using dedicated coupling bus resonators for each pair (light orange). Each transmon has a flux-control line for two-qubit gating (yellow), a microwave-drive line for single-qubit gating (dark orange), and dispersively-coupled resonator with Purcell filter for readout (purple) [6, 40]. The readout-resonator/Purcell-filter pair for $Q_{M2}$ is visible at the center of this image. The vertically running common feedline (blue) connects to all Purcell filters, enabling simultaneous readout of the four transmons by frequency multiplexing. Air-bridge crossovers enable the routing of all input and output lines to the edges of the chip, where they connect to a printed circuit board through aluminum wirebonds. (b) Zoom-in at $Q_L$, all transmons on the chip follow a similar design. (c) Transmon circuit diagram with a capacitor representing the two (large) superconducting islands at the center of (b) and the SQUID loop formed by two Josephson junctions. Coupling to the drive-line, to the readout resonator and to other qubits is omitted. The current in the flux-line produces the magnetic flux $\Phi$ in the loop which controls the transmon transition frequency. (d) Zoom-in at the SQUID loop that integrates the two Josephson junctions.

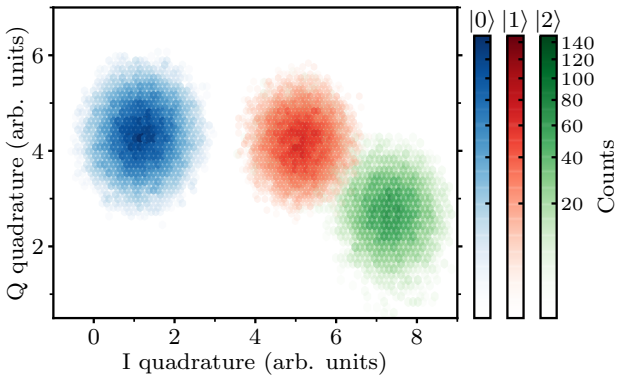


Figure 3. Exemplary read-out histograms for a transmon, prepared in $|0\rangle, |1\rangle$ and $|2\rangle$ states, represented on the (I, Q)-plane. For the $|2\rangle$ state the transmon is excited with a driving pulse calibrated for the $|1\rangle \rightarrow |2\rangle$ transition after preparing $|1\rangle$. Dataset generated with routines from Ref. 41. The relative positions of the centroids can vary between transmons.

flux-dependent energy levels – a key mechanism exploited in this work. We return to it after introducing the physical device and its architecture.

In this work we employ a transmon device designed to execute quantum circuits with the long-term goal of a universal Fault-Tolerant Quantum Computer. The holy grail, in overcoming the inherent errors of the physical qubits, are logical qubits encoded redundantly by physical qubits in an architecture such as the Surface Code [5, 35], allowing for Quantum Error Detection and Quantum Error Correction. The 7-transmons chip used is designed in a configuration representing a sub-patch of the Surface Code [35]. For more detailed chip design considerations and transmons background we point the reader to chapter 2 of Ref. 42 and Ref. 39.

The superconducting transmon chip is placed inside a dilution refrigerator that maintains it under vacuum at the coldest stage ($\sim$ 10 mK, mixing chamber temperature). The terminals of the chip are routed up to the room-temperature control electronics through the attenuation/amplification chain using coaxial cables.

In this study we make use of four transmons. A zoomed in optical image of the device is shown in Fig. 2. In Figs. 2(b) and 2(c) we detail the structure of one of these transmon while in Fig. 2(d) we present the SQUID loop. High- and low-frequency transmons ($Q_H$ and $Q_L$, respectively) connect to two mid-frequency transmons ($Q_{M1}$ and $Q_{M2}$) using bus resonators dedicated to each pair. Every transmon has a dedicated microwave-drive line for single-qubit gating, a flux-control line used for two-qubit gating, and a dispersively-coupled readout resonator with dedicated Purcell filter [6, 40] for quantum state readout. Qubits initialization is based on the energy relaxation to the ground state $|0\rangle$ (a few hundred microseconds).

The drive-live implements arbitrary single qubit rotations around axes on the $\hat{x}\hat{y}$-plane (of the Bloch sphere).

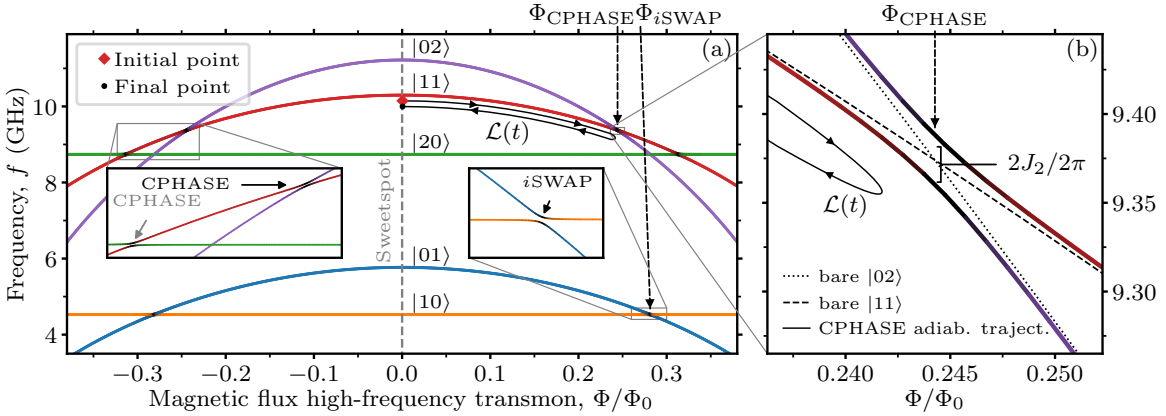


Figure 4. (a) Approximate eigenenergies spectrum of two coupled transmons with typical parameters as a function of the (local) magnetic flux in the SQUID loop of the high-frequency transmon, normalized to the flux quantum $\Phi_0$. Only the first five excited levels are included. All branches are symmetrical with respect to the sweetspot. Single qubit excitations $|10\rangle$ and $|01\rangle$ implement an $i$SWAP gate at $\Phi_{i\mathrm{SWAP}}$ (inset on the right). Interactions of $|11\rangle$ with non-computational levels $|20\rangle$ and $|02\rangle$ give rise to CPHASE-type avoided crossings (inset on the left). Black path between red diamond (initial point) and black dot (final point) indicates typical (unipolar) CPHASE adiabatic trajectory [$\mathcal{L}(t)$]. The zero point of energy is chosen to be $f_{00}$. (b) Zoom-in at the avoided crossing between $|11\rangle$ and $|02\rangle$ at $\Phi_{\mathrm{CPHASE}}$ employed for CZ gating.

The driving signal is composed of two sinusoids resonant with the qubit frequency, one in phase (I) and one in quadrature (Q), shaped by Gaussian and Gaussian-derivative envelopes according to the DRAG scheme [43]. An IQ up-converting mixer is employed for this modulation. The DRAG scheme is essential to avoid (residually) driving the $|1\rangle \rightarrow |2\rangle$ transition which otherwise results in *leakage* ($L_1$) into states outside the computation space – a severe gate error.

Readout is performed by frequency multiplexed measurement of a common feedline capacitively connected to all four transmons through their respective readout resonator and Purcell filters. The scenario of a transmon coupled to a resonator (effectively a QHO) is known as circuit Quantum Electrodynamics (cQED) and it is analogous to an atom in an optical cavity. The important here is that the resonator-qubit interaction leads to a resonator whose frequency depends on the qubit state. In practice this makes possible to probe the transmission line at a fixed frequency and based on the transmission/reflection infer the qubit state. An IQ modulation/demodulation chain is employed for this task [44]. To build some intuition, in Fig. 3 we illustrate an exemplary histogram of the (demodulated and averaged) read-out voltages for a transmon prepared in each of its first 3 levels and measured immediately after.

Despite being possible to operate transmon at other frequencies, due to flux-noise impact each transmon is statically flux-biased at its sweetspot (maximum transition frequency point) countering residual offsets (this procedure is detailed in chapter 3 of Ref. 42). Flux pulsing is performed using a Zurich Instruments HDAWG-8 ($t_s$ = 1/2.4 ns). Following prior work [31, 45], linear-dynamical distortions in the flux-control lines are corrected using real-time filters.

When considering a resonator coupling capacitively two transmons, and assuming no photons in the resonator, it is possible to show [39, 42] that this system can be regarded as two directly coupled transmons. Combined with the frequency tuneability of the higher-frequency qubit, the pair can be used to implement two-qubit gates, such as the CPHASE-type gates and the CZ in articular [10].

The CPHASE unitary can be expressed in the computational space $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as:

$$U_{\mathrm{CPHASE}}(\phi_{2\mathrm{Q}}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi_{2\mathrm{Q}}} \end{pmatrix},$$

which has the effect of applying a relative quantum phase conditional on both qubits being in the excited state $|11\rangle$, hence the name controlled-phase gate. A CZ gate is achieved when

$$\phi_{2\mathrm{Q}} = \pi \;\Rightarrow\; U_{\mathrm{CPHASE}}(\phi_{2\mathrm{Q}}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \equiv U_{\mathrm{CZ}}.$$

To elucidate the implementation principle behind the CPHASE, it is useful to inspect the energy spectrum of two coupled flux-tunable transmons with typical parameters. In Fig. 4 we plot the (approximate) eigenenergies of the two-transmon system as a function of the magnetic flux in the SQUID loop of the high-frequency qubit. We note that, as we increase the absolute flux (linearly dependent on the flux-line current), the non computational state $|02\rangle$ decreases in energy much more than $|01\rangle$. As a consequence, it "crosses" the $|11\rangle$. Furthermore, due to the negative anharmonicity of this type of transmon, the $|02\rangle$ crosses $|11\rangle$ before the crossing of the computational states $|01\rangle$ and $|10\rangle$.

When zooming in at the point where $|11\rangle$ and $|02\rangle$ would become degenerate [Fig. 4(b)] we instead find an *avoided crossing*. The insets of Fig. 4(a) depict 3 such avoided crossings (on each sides of the sweetspot): two of the CPHASE-type and one of the $i$SWAP-type, same physics, only distinct interacting levels. In practice, due to the increased sensitivity to flux noise at higher flux, only the first avoided crossing is suitable for high fidelity gates [Fig. 4(a), left inset, arrow in black].

The accrual of a relative conditional phase happens when we excite both qubits into $|11\rangle$ [Fig. 4(a), red diamond] and apply a flux trajectory $\mathcal{L}(t)$ towards the avoided crossing with $|02\rangle$ at $\Phi_{\text{CPHASE}}$ for a total time $\tau$ in a slow enough manner compared to the timescale set by $J_2$ and then returning back following the same excursion [Fig. 4(a), black trajectory]. Ideally, this would ensure following an instantaneous eigenstate of the system and never populating the $|02\rangle$. It is common to label such trajectory as being *adiabatic*. Effectively, the avoided crossing with the non-computational state gives rise to an effective ZZ-coupling within the computational state yielding a CPHASE.

However, if the transmons are too close in frequency, the avoided crossing would be much closer to the sweetspot with a non-negligible always-on residual ZZ-coupling that gives rise to detrimental phase accrual [46]. Same applies for too strong couplings which motivated their reductions over the last decade.

Contrasting with the adiabatic approach, the ideal extreme of a sudden pulse into the avoided crossing, could also implement a CZ gate. When jumping suddenly to $\Phi_{\text{CPHASE}}$, after preparing the transmons in $|11\rangle$, and letting the system evolve for a time $\tau$, the system performs a Larmor-type rotation [44] – a precession-like phenomenon – acquiring the relative conditional phase. The $J_2$ rate establishes the fastest CPHASE gates possible, in particular a $t_{\text{lim}} = 2\pi/2J_2$ for the CZ.

The disadvantage of interacting with states outside the computational space is that, for small offsets from the optimal parameters, the system ends up in superposition with those states (leakage) being very detrimental in a two-level-based quantum computation model [14]. Many efforts went into designing adiabatic trajectories compatible with the quantum error correction threshold ($\sim 1\,\%$), culminating recently in the Net-Zero scheme [31].

Unfortunately, its calibration is laborious because, besides feeding the simulations with an accurate characterization of the device, it requires to first heuristically optimize the pulse duration in simulation. This increases the complexity of the problem even in simulation. Note that in experiment it is almost unfeasible owing to the $\sim 30\,\text{s}$ necessary to evaluate a single point on the 2D calibration landscape. Furthermore, despite its current adoption, the complex structure and interdependence of the calibration landscapes (see chapter 4 in Ref. 42 or Ref. 31 for some examples) limits the use of the NZ as quantum processors grow in qubit number. It is worth emphasizing that the qualitative features of these calibration landscapes vary significantly between pulse durations and among transmon-pairs.

Motivated by these practical limitations we introduce the SNZ variant described in the next section.

## III. THE SNZ CONCEPT

A flux pulse harnessing the $|11\rangle$-$|02\rangle$ interaction implements the unitary

$$U_{\text{CPHASE}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{i\phi_{01}} & 0 & 0 & 0 \\ 0 & 0 & e^{i\phi_{10}} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{1-4L_1}e^{i\phi_{11}} & \sqrt{4L_1}e^{i\phi_{02,11}} \\ 0 & 0 & 0 & \sqrt{4L_1}e^{i\phi_{11,02}} & \sqrt{1-4L_1}e^{i\phi_{02}} \end{pmatrix}$$

in the $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle, |02\rangle\}$ subspace, neglecting decoherence and residual interaction between far off-resonant levels. Here, $\phi_{01}$ and $\phi_{10}$ are the single-qubit phases, and $\phi_{11} = \phi_{01} + \phi_{10} + \phi_{2\text{Q}}$, where $\phi_{2\text{Q}}$ is the conditional phase. Finally, $L_1$ is the leakage parameter. Even though we follow the leakage definition where it accounts for any computational state to leak into any other non-computation state [34], for this illustration we simplify and attribute all leakage to the main $|11\rangle \to |02\rangle$ channel, hence a factor of 4 under the square roots.

In a rotating frame which absorbs the single-qubit phases, the system Hamiltonian is

$$H = \Delta_{02}(t)\,|02\rangle\langle 02| + J_2\,(|02\rangle\langle 11| + |11\rangle\langle 02|)\,,$$

where $\Delta_{02}(t)$ is the dynamical detuning between $|02\rangle$ and $|11\rangle$. Each half of the bipolar NZ pulse implements the unitary

$$U_{\text{A}} = U_{-\text{A}} = \begin{pmatrix} \gamma e^{i\phi_{\text{a}}} & \beta e^{i\phi_{\text{b}}} \\ \beta e^{i\phi_{\text{c}}} & \gamma e^{i\phi_{\text{d}}} \end{pmatrix},$$

in the $\{|11\rangle, |02\rangle\}$ subspace, where $\gamma, \beta \in [0,1]$ satisfy $\gamma^2 + \beta^2 = 1$ and $\phi_{\text{a}} + \phi_{\text{d}} = \phi_{\text{b}} + \phi_{\text{c}} + \pi \pmod{2\pi}$. In the MZI analogy, this unitary is the action of each beamsplitter, ideally identical BS1 = BS2.

In SNZ [Fig. 5(a)], each half pulse is a square pulse with amplitude $\pm A$ and duration $t_{\text{p}}/2 = t_{\text{lim}}/2$. SNZ intentionally adds an idling period $t_\phi$ between the half pulses to perfect the analogy to the MZI [Fig. 10(a) inset], allowing accrual of relative phase $\phi$ in between the beamsplitters BS1 and BS2. The unitary action of this idling is

$$U_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}.$$

An ideal CZ gate, our target here, achieves $\phi_{01} = \phi_{10} = 0 \pmod{2\pi}$, $\phi_{2\text{Q}} = \pi \pmod{2\pi}$ (phase condition PC), and $L_1 = 0$ (leakage condition LC), with arbitrary $\phi_{02}$. Accomplishing both conditions with $U_{-\text{A}} U_\phi U_{\text{A}}$ requires

$$\gamma^2 e^{i2\phi_{\text{a}}} + \beta^2 e^{i(\phi_{\text{b}}+\phi_{\text{c}}+\phi)} = -1 \text{ (PC)}$$

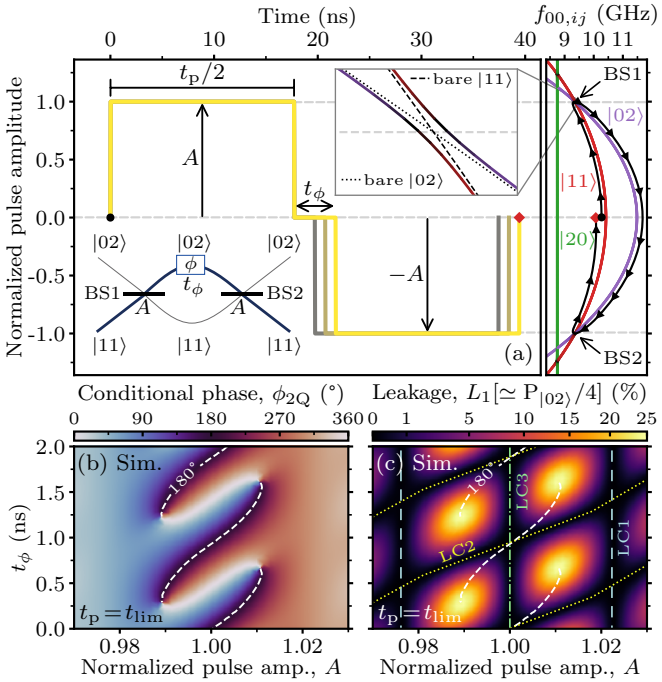


Figure 5. Numerical simulation of a SNZ pulse with infinite time resolution for $Q_L$-$Q_{M2}$, with $\Delta_{02}/2\pi = 1.063$ GHz (at bias point) and $t_{lim} = 35.40$ ns for the $|11\rangle$-$|02\rangle$ interaction. (a) Schematic of the SNZ flux pulse, with $t_p = t_{lim}$ and variable $A$ and $t_\phi$. Amplitude $A$ is normalized to the $|11\rangle$-$|02\rangle$ resonance. (b, c) Landscapes of conditional phase $\phi_{2Q}$ (b) and leakage $L_1$ (c) as a function of $A$ and $t_\phi$.

simultaneously with either one of three conditions: (LC1) $\beta = 0$; (LC2) $\phi_a - \phi_d - \phi = \pi \pmod{2\pi}$; or (LC3) $\gamma = 0$. LC1 (LC3) corresponds to perfect reflection (transmission) at each beamsplitter. LC2 corresponds to destructive interference of the leakage produced by the first beamsplitter.

The key advantage of SNZ over CNZ is the very straightforward procedure to simultaneously meet PC and low leakage. To appreciate this, consider first the ideal scenario where the pulses has infinite time resolution. For, $t_p = t_{lim}$, $t_\phi = 0$ ($\phi = 0$), and $A = 1$ (the $|\Delta_{02}|$ minimum) each half pulse implements an $i$SWAP gate between $|11\rangle$ and $|02\rangle$. Thus, $\gamma = 0$ (meeting LC3) and $\phi_b = \phi_c = -\pi/2$ (meeting PC). In the MZI analogy, the first beamsplitter fully transmits $|11\rangle$ to $-i|02\rangle$ (producing maximal intermediate leakage), and the second fully transmits $-i|02\rangle$ to $-|11\rangle$.

Consider now the effect of $t_\phi$. The two-qutrit numerical simulation, i.e. including three energy levels ($|0\rangle$, $|1\rangle$ and $|2\rangle$), with infinite time resolution in Figs. 5(b) and 5(c) show that the landscapes of $\phi_{2Q}$ and $L_1$ as a function of $A$ and $t_\phi$ have a clear structure and link to each other. Evidently, $U_{-A}U_\phi U_A$ is $2\pi$-periodic in $\phi$, so both landscapes are vertically periodic. The $L_1$ landscape shows a vertical leakage valley at $A = 1$, due to LC3, and two other vertical valleys due to LC1 [see Fig. 5(c)]. LC2 gives rise to additional diagonal valleys. Crucially, juxtaposing the $\phi_{2Q} = 180°$ contour shows that PC is met at the crossing of LC3 and LC2 valleys. This regular leakage landscape therefore provides on its own useful crosshairs for simultaneously achieving PC and low leakage. We note that along the LC3 valley, $\phi_{2Q}(t_\phi)$ changes monotonically, allowing for CPHASE gates with arbitrary $\phi_{2Q}$. We leave this generalization for future work, focusing here on CZ gates.

There are practical reasons to include $t_\phi$ in experiment: any flux-pulse distortion remaining from the first half pulse during the second (e.g., due to finite pulse rise time) will break the symmetry $U_{-A} = U_A$. Due to the fixed time resolution $t_s$ of the AWG used for flux control, $\phi$ can only increment in steps of $-\Delta_{02}t_s$, where $\Delta_{02}$ is the detuning at the bias point. As typically $\Delta_{02}/2\pi = 0.5$–$1$ GHz and $t_s \sim 1$ ns, one may only use the number of intermediate sampling points in $t_\phi$ for coarse control. For fine control, we propose the (simultaneous) tunning of the amplitude $\pm B$ of two sampling points: the first and last sampling points during $t_\phi$ [see Fig. 6(b)].

### A. Comparison between CNZ and SNZ pulses

The conventional NZ (CNZ) strong pulse [Fig. 6(a)] consists of two back-to-back half pulses of duration $t_p/2$ each, applied on the higher-frequency transmon. Typically, $t_p/t_{lim} \sim 1.1$–$1.6$. The strong half pulses are formally parametrized as in Ref. 23. For the purposes of illustration, here we can loosely lump this parametrization as affecting the amplitude ($\pm A$) and curvature ($A'$) of the strong half pulses. Immediately following the strong pulse, weak bipolar pulses of duration $t_{1Q}$ are applied on both the higher- and lower-frequency transmons with amplitudes $\pm C$ and $\pm D$, respectively, in order to null the single-qubit phases acquired by each. Typically, $t_{1Q} = 10$ ns. In conventional NZ there is no intermediate idling period between the strong half pulses, so the analogy to the MZI is not exact [Fig. 6(c)]. During tuneup, one searches the $(A, A')$ space to achieve $U_{-A}U_A = U_{CPHASE}(\phi_{2Q} = \pi)$ by only affecting the unitary action of the two beamsplitters. Because for typical $t_p$ conventional NZ produces significant leakage at the first strong pulse, achieving minimal leakage relies on meeting LC2. The structure of the $\phi_{2Q}(A, A')$ and $L_1(A, A')$ landscapes and especially their interrelation are not straightforward, so the search for an $(A, A')$ setting satisfying both PC and LC2 is not easily guided [see figures in Ref. 31].

The SNZ pulses introduced here [Fig. 6(b)] differ in two key ways. First, the strong half pulses are replaced by square half pulses each with duration $t_p/2$ as close as possible to $t_{lim}/2$ (as allowed by the AWG sampling period) but not shorter. Second, an intermediate idling period $t_\phi$ is added to accrue relative phase $\phi$ between $|02\rangle$ and $|11\rangle$, perfecting the analogy to the MZI [Fig. 6(d)]. We use the amplitude $\pm B$ of the first and last sampling points in $t_\phi$ and the number of intermediate zero-amplitude points

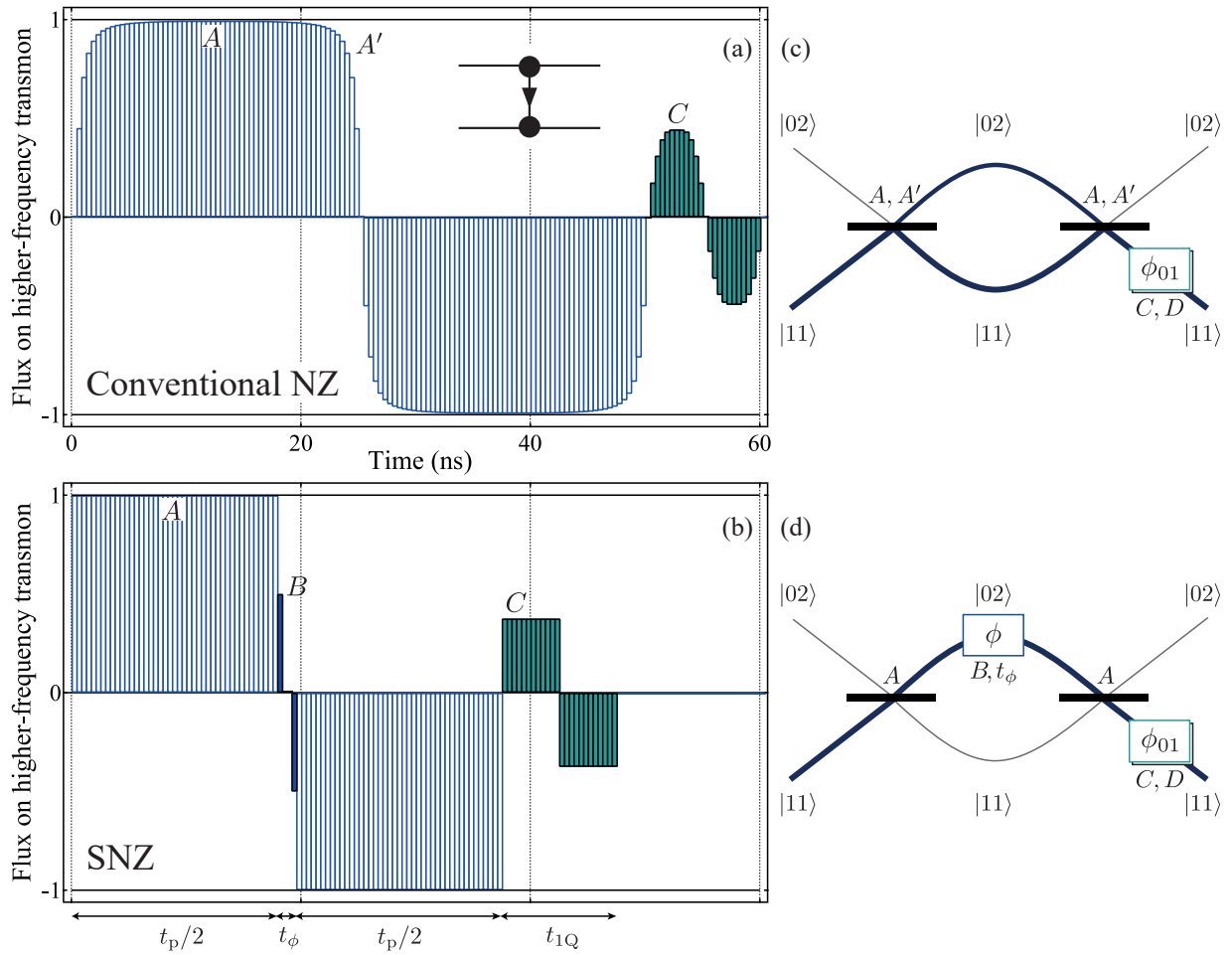


Figure 6. Comparison of conventional NZ (CNZ) and SNZ pulses for CZ gates. (a) Conventional NZ CZ pulses consist of two back-to-back strong half pulses of duration $t_p/2$ each, followed by two weak back-to-back half pulses of duration $t_{1Q}/2$ each on the higher-frequency qubit. The amplitude ($\pm A$) and curvature ($A'$) of the strong pulses are jointly tuned to set the conditional phase $\phi_{2Q}$ at minimal leakage $L_1$, while the amplitude $\pm C$ of the weak pulses is used to null the single-qubit phase on the higher-frequency transmon. Weak pulses (amplitude $\pm D$) on the lower-frequency qubit (not shown here) are also used to null its single-qubit phase. (b) In SNZ, the strong pulses are replaced by square pulses with $t_p$ as close as possible to $t_{lim}$ but not shorter. Also, an intermediate idling period $t_\phi$ is added to accrue relative phase $\phi$ between $|02\rangle$ and $|11\rangle$. The amplitude $\pm B$ of the first and last sampling points in $t_\phi$ and the number of intermediate zero-amplitude points provide fine and coarse control of this relative phase, respectively. SNZ CZ gates also use weak bipolar pulses (now square) of total duration $t_{1Q}$ to null single-qubit phases. (c) The MZI analogy for CNZ pulses is incomplete. Each strong half pulse implements a beamsplitter (ideally identical) with scattering parameters affected by $A$ and $A'$. However, there is no possibility to independently control the relative phase in the two arms between the beamsplitters. (d) The MZI analogy is exact for SNZ pulse. The scattering at the beamsplitters is controlled by $A$ and the relative phase $\phi$ is controlled finely using $B$ and coarsely using $t_\phi$.

to achieve fine and coarse control of $\phi$, respectively. As in CNZ, we use weak bipolar pulses on both transmons (also with $t_{1Q} = 10\,\text{ns}$) to null the single-qubit phases. During tuneup, we search the $(A, B)$ space to achieve $U_{-A}U_\phi U_A = U_{CPHASE}(\phi_{2Q} = \pi)$. In Section IV A we will show that the SNZ pulse design gives a very simple structure to the $\phi_{2Q}(A, B)$ and $L_1(A, B)$ landscapes. Crucially, the crossing point of leakage valleys satisfying LC2 and LC3 matches $\phi_{2Q} = 180°$. This simplicity of tuneup is the key advantage of the SNZ over the conventional NZ.

Another advantage of SNZ over conventional CZ is the reduced total time $t_{total} = t_p + t_\phi + t_{1Q}$ required for a CZ gate. However, due to the 20 ns timing grid of the control electronics and the transverse coupling strengths in this device, this speedup is insufficient to reduce the total time allocated per CZ gate from 60 to 40 ns. Nonetheless, in SNZ, the fluxed transmon spends more time at its sweetspot, which reduces the dephasing due to flux noise.

Additionally, in Fig. 7 we compare schematically the trajectories implemented by the strong CNZ and SNZ pulses on a system initialized in the $|11\rangle$ state. A CNZ pulse attempts to adiabatically approach twice the interaction point while remaining in the $|11\rangle$ state [Fig. 7(a)]. Because it is hard to meet such condition, during the positive branch of the pulse, some population is lost through

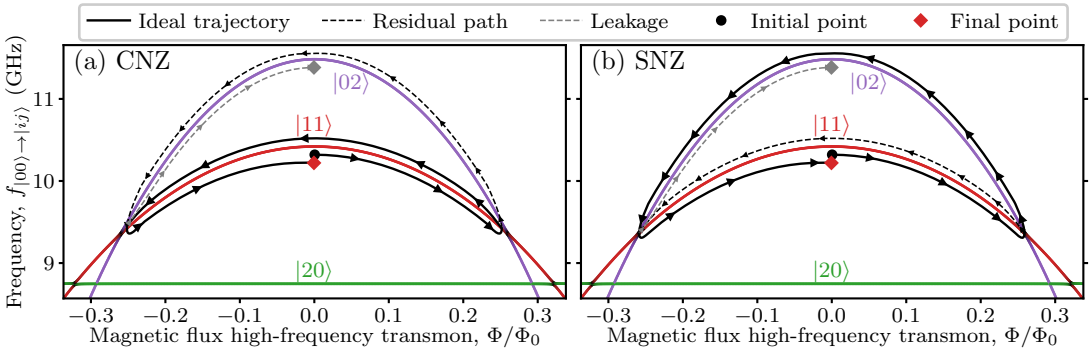


Figure 7. Schematic comparison between realistic trajectories for CNZ and SNZ pulses when applied to the double-excited state $|11\rangle$. The trajectory for single-qubit phase corrections are omitted. Note that in both cases, the most of the trajectory is spent close to (CNZ) or at the center (SNZ) of the interaction point. (a) Trajectory for a CNZ pulse. The system remains in the $|11\rangle$ state during the full trajectory for a sufficiently long pulse such that the adiabatic condition holds. Imperfections and/or too short pulses can lead to part of the population leaking into (or seeping back from) the $|02\rangle$ state. (b) Trajectory for an SNZ pulse. Ideally, maximal leakage is produced during the first half-pulse and the full population of the $|11\rangle$ state is recovered during the second half-pulse.

the residual path of $|02\rangle$. Imperfect destructive interference of the $|02\rangle$ state during the negative branch of the pulse leads to leakage at the end of the gate. In case of an SNZ pulse [Fig. 7(b)], after suddenly moving into the avoided crossing, during the first half-pulse, we maximize the transfer of population to the $|02\rangle$ state by fixing the amplitude $A = 1$ for a duration of $t_p/2 \sim t_{lim}/2$. As we switch to the second half-pulse approaching the avoided crossing for the second time, the single-sample amplitudes $\pm B$, together with $t_\phi$, provide robust tunning knobs to maximize the recovery of the $|11\rangle$ population with the desired conditional phase.

## IV. EXPERIMENTAL REALIZATION OF SNZ CZ GATES

### A. Calibration procedure

We now turn to the experimental realization of SNZ CZ gates between the nearest-neighbor pairs among four transmons. We exemplify the tuneup of SNZ using pair $Q_L$-$Q_{M2}$ (Fig. 8). We first identify $t_{lim}$ for the $|11\rangle$-$|02\rangle$ interaction and amplitude $A$ bringing the two levels on resonance. Both are extracted from the characteristic chevron pattern of $|2\rangle$-population in $Q_{M2}$ as a function of the amplitude and duration of a unipolar square flux pulse acting on $|11\rangle$ [Fig. 8(a)]. The symmetry axis corresponds to $A = 1$ and the oscillation period along this axis gives $t_{lim}$. We set $t_p = t^+_{lim} \equiv 2nt_s$, where $n$ is the smallest integer satisfying $2nt_s \geq t_{lim}$. Next, we measure the landscapes of $\phi_{2Q}$ and leakage estimate $\tilde{L}_1$ in the range $A \in [0.9, 1.1]$, $B \in [0, A]$, using conditional-oscillation experiments [31]. As expected, the landscape of $\tilde{L}_1$ [Fig. 8(c)] reveals a vertical valley at $A = 1$ and a diagonal valley. Juxtaposing the $\phi_{2Q} = 180°$ contour from Fig. 8(b), we observe the matching of PC at the crossing of these valleys. These observations are in ex-

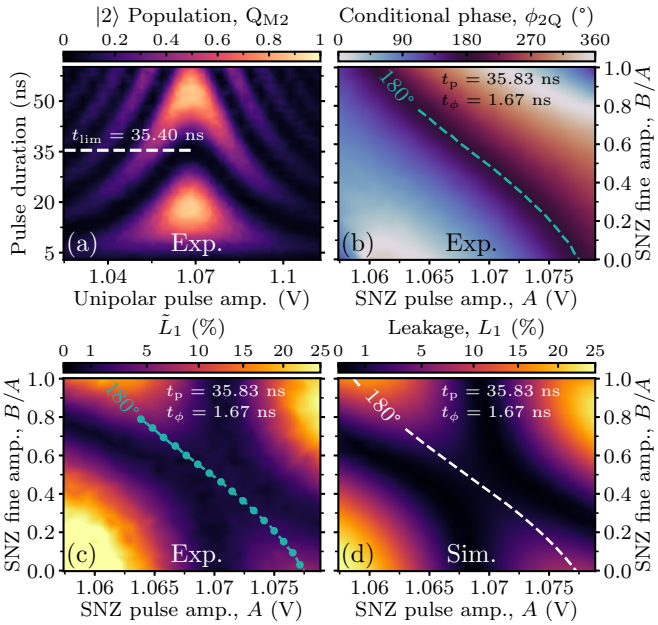


Figure 8. Calibration of the SNZ pulse for pair $Q_L$-$Q_{M2}$ and comparison to simulation. (a) $|2\rangle$-state population of $Q_{M2}$ as a function of the amplitude and duration of a unipolar square pulse making $|11\rangle$ interact with $|02\rangle$. (b, c) Landscapes of conditional phase $\phi_{2Q}$ and leakage estimate $\tilde{L}_1$ as a function of SNZ pulse amplitudes $A$ and $B$, with $t_p = t^+_{lim}$ and $t_\phi = 1.67$ ns. The juxtaposed $\phi_{2Q} = 180°$ contour runs along the opposite diagonal compared to Figs. 5(b) and 5(c) because, contrasting with $t_\phi$, increasing $B$ (which decreases $\Delta_{02}$) changes $\phi$ in the opposite direction from $t_\phi$. Data points marked with dots are measured with extra averaging for examination in Fig. 9. (d) Numerical simulation of leakage $L_1$ landscape and $\phi_{2Q} = 180°$ contour with parameters and flux-pulse distortions from experiment. All landscapes (also in Fig. 9) are sampled using an adaptive algorithm based on Ref. 47.

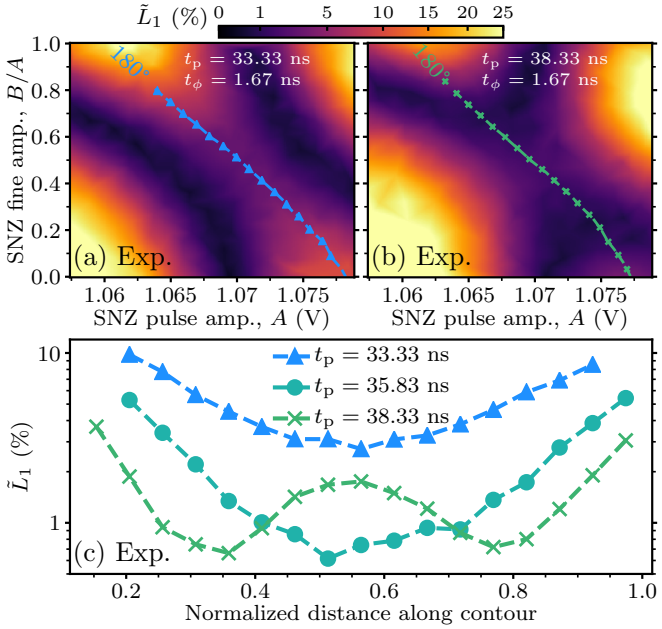


Figure 9. (a, b) Landscapes of the leakage estimate $\tilde{L}_1$ for intentionally short and long ($t_p = t_{lim}^{+} \pm 6t_s$) flux pulses on $Q_{M2}$. (c) Extracted $\tilde{L}_1$ along the $\phi_{2Q} = 180°$ contours from (a), (b), and Fig. 8(c).

cellent agreement with a numerical two-qutrit simulation [Fig. 8(d)].

### 1. *SNZ robustness to hardware limitations*

Experimentally, it is nearly impossible to precisely match $t_p = t_{lim}$ due to the discreteness of $t_s$. To understand the consequences, we examine the $\phi_{2Q}$ and $\tilde{L}_1$ landscapes for SNZ pulses with intentionally set $t_p = t_{lim}^{+} \pm 6t_s$ (Fig. 9). We find that the PC contour remains roughly unchanged in both cases. However, there are significant effects on $\tilde{L}_1$. In both cases, we observe that $\tilde{L}_1$ lifts at the prior crossing of LC2 and LC3 valleys where $\phi_{2Q} = 180°$. For too-short pulses [Fig. 9(a)], there remain two valleys of minimal $\tilde{L}_1$, but these are now curved and do not cross $\phi_{2Q} = 180°$. For too-long pulses [Fig. 9(b)], there are also two curved valleys. Crucially, these cross the $\phi_{2Q} = 180°$ contour, and it remains possible to achieve PC and minimize leakage at two $(A, B)$ settings. Extracting $\tilde{L}_1$ along the $\phi_{2Q} = 180°$ contours [Fig. 9(c)] confirms that the minimal leakage obtainable for $t_p = t_{lim}^{+} + 6t_s$ matches that for $t_p = t_{lim}^{+}$. The impossibility to achieve minimal leakage at $\phi_{2Q} = 180°$ for $t_p < t_{lim}$ is a clear manifestation of the speed limit set by $J_2$. In turn, the demonstrated possibility to do so for $t_p > t_{lim}$ (even when overshooting the duration significantly) is an important proof of the viability of the SNZ pulse in practice.

| Parameter | $Q_{M1}$-$Q_H$ | $Q_{M2}$-$Q_H$ | $Q_L$-$Q_{M1}$ | $Q_L$-$Q_{M2}$ |
|---|---|---|---|---|
| $t_{lim}$ (ns) | 32.20 | 29.00 | 40.60 | 35.40 |
| $t_p$, $t_\phi$ (ns) | 32.50, 2.92 | 29.10, 3.75 | 40.83, 1.25 | 35.83, 1.67 |
| $t_{total}$ (ns) | 45.42 | 42.91 | 52.08 | 47.50 |
| Interaction | $\|11\rangle$-$\|02\rangle$ | $\|11\rangle$-$\|02\rangle$ | $\|11\rangle$-$\|20\rangle$ | $\|11\rangle$-$\|02\rangle$ |
| Parked qubit | $Q_{M2}$ | $Q_{M1}$ | – | – |
| Avg. $F$ (%) | $98.89 \pm 0.35$ | $99.54 \pm 0.27$ | $93.72 \pm 2.10$ | $97.14 \pm 0.72$ |
| Avg. $L_1$ (%) | $0.13 \pm 0.02$ | $0.18 \pm 0.04$ | $0.78 \pm 0.32$ | $0.63 \pm 0.11$ |
| Max. $F$ (%) | $99.77 \pm 0.23$ | $99.93 \pm 0.24$ | $99.15 \pm 1.20$ | $98.56 \pm 0.70$ |
| Min. $L_1$ (%) | $0.07 \pm 0.04$ | $0.10 \pm 0.02$ | $0.04 \pm 0.08$ | $0.41 \pm 0.10$ |

Table 1. Summary of SNZ CZ pulse parameters and achieved performance for the four transmon pairs. Single-qubit phase corrections are included in $t_{total}$. Gate fidelities and leakage are obtained from 2QIRB keeping the other two qubits in $|0\rangle$. Statistics (average and standard deviation) are taken from repeated 2QIRB runs [see Ref. 46]. The maximum $F$ and minimum $L_1$ quoted are not necessarily from the same run.

## B. Experimental performance of SNZ CZ gates

With these insights, we proceed to tuning the four SNZ CZ gates, following similar procedures. We use final weak bipolar pulses of total duration $t_{1Q} = 10$ ns to null the single-qubit phases in the frame of microwave drives. Since our codeword-based control electronics has a 20 ns timing grid, and 40 ns $< t_{total} = t_p + t_\phi + t_{1Q} <$ 60 ns for all pairs, we allocate 60 ns to every CZ gate. Some pair-specific details must be noted (see inset in Fig. 10 for intuition on qubits connectivity and the detuning involved during the CZ gates). Owing to the transitions overlap of mid-frequency qubits, implementing CZ between $Q_H$ and $Q_{M1}$ ($Q_{M2}$) requires a parking (bipolar) flux pulse on $Q_{M2}$ ($Q_{M1}$) during the SNZ pulse on $Q_H$ [8, 35]. The parking flux pulse on the other mid-frequency transmon (the one not participating in the CZ) downshifts its frequency (during the CZ gate) in order to avoid the undesired interaction with this transmon. For most pairs, we employ the $|11\rangle$-$|02\rangle$ interaction, which requires the smallest flux amplitude (reducing the impact of dephasing from flux noise) and does not require crossing any other interaction. However, for $Q_L$-$Q_{M1}$, we cannot reliably use this interaction as there is a flickering two-level system (TLS) overlapping $Q_{M1}$ transition [46]. For this pair, we therefore employ the $|11\rangle$-$|20\rangle$ interaction. Here, the SNZ brigs a side benefit: it minimizes exchange between $|01\rangle$ and the TLS, $|11\rangle$ and $|20\rangle$, and $|01\rangle$ and $|10\rangle$ as their resonances are crossed as suddenly as possible.

Table 1 summarizes the timing parameters and performance attained for the four SNZ CZ gates. The CZ gate fidelity $F$ and leakage $L_1$ are extracted using a 2QIRB protocol [31, 34]. For each pair, we report the best, average and standard deviation of both based on at least 10 repetitions of the protocol spanning more than 8 h [46]. Several observations can be drawn. First, CZ gates involving $Q_H$ perform better on average than those involving $Q_L$. This is likely due to the shorter $t_{lim}$ and correspondingly longer time 60 ns $- t_p$ spent near the

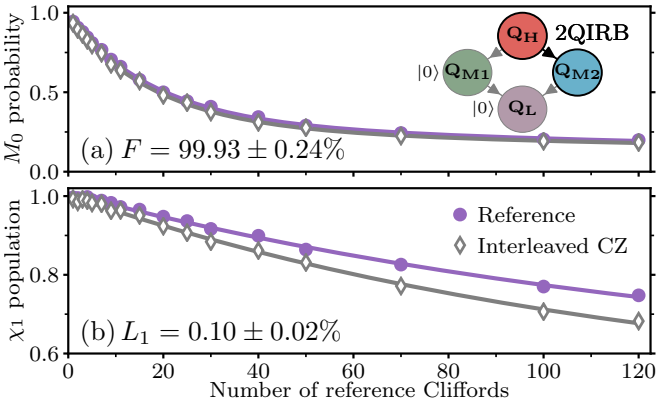


Figure 10. Best SNZ CZ gate performance achieved from a single run of 2QIRB. (a) Reference and CZ-interleaved return probability $M_0$ to $|00\rangle$ and (b) population in the computational space $\chi_1$ as a function of the number of two-qubit Cliffords in the reference curve. Errors bars in $F$ and $L_1$ are obtained from the uncertainty of exponential-decay fits.

sweetspot. Additionally, the frequency downshifting required of $Q_H$ to interact with $Q_{M1}$ and $Q_{M2}$ is roughly half that required of the latter to interact with $Q_L$. This reduces the impact of dephasing from flux noise during the pulse. Not surprisingly, performance is worst for $Q_L$-$Q_{M1}$. Here, the pulse must downshift $Q_{M1}$ the most to reach the distant $|11\rangle$-$|20\rangle$ interaction, increasing dephasing from flux noise. Also, there may be residual exchange with the TLS and the crossed resonances. Overall, there is significant temporal variation in performance as gleaned by repeated 2QIRB characterizations. We believe this reflects the underlying variability of qubit relaxation and dephasing times, which however were not tracked simultaneously. In addition to having the best average performance, pair $Q_{M2}$-$Q_H$ displays the maximum $F$ of $99.93 \pm 0.24\%$ (Fig. 10). To the best of our knowledge, this is the highest CZ fidelity extracted from 2QIRB characterization in a transmon processor.

To understand the dominant sources of infidelity $\varepsilon = 1 - F$ and leakage, we run numerical simulations [31], for both SNZ and CNZ, with input parameters from experiment for pair $Q_{M2}$-$Q_H$. We dissect an error budget versus various models finding similar contributions for both gates (see Ref. 46 for details). Nevertheless, the results suggest that SNZ slightly outperforms CNZ, likely due to the shorter time spent away from the sweetspot.

## V. CONCLUSION

In summary, we have proposed and implemented the sudden variant of the NZ pulsing scheme [31]. SNZ operates at the speed limit of transverse coupling by maximizing intermediate leakage to the non-computational state. The key advantage of SNZ over CNZ is ease of tuneup, owing to the simple structure of error landscapes as a function of pulse parameters. We have demonstrated high-fidelity CZ gates between four transmon pairs in a patch of a 7-transmon processor. Control architectures without a timing grid will further benefit from the increased speed of SNZ over CNZ by reducing the total gate time and thus the impact of decoherence. Harnessing the tuning simplicity, we already employ SNZ CZ gates in the Starmon-5 processor publicly available via the QuTech Quantum Inspire platform [48]. Moving forward, the compatibility of SNZ with our proposed [35] scalable scheme for surface coding makes SNZ our choice for CZ gates for quantum error correction. Finally, the straightforward extension of SNZ to arbitrary conditional-phase gates can find immediate use in NISQ applications.

## ACKNOWLEDGMENTS

We thank L. Janssen, M. Rol, M. Sarsby, T. Stavenga, B. Tarasinski, and M. Venkatesh for experimental assistance, C. Eichler and B. Terhal for discussions, and G. Calusine and W. Oliver for providing the traveling-wave parametric amplifier used in the readout amplification chain. This research is supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the U.S. Army Research Office Grant No. W911NF-16-1-0071, and by Intel Corporation. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. F. B. is supported by ERC Grant EQEC No. 682726.

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

[2] 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).

[3] 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, Annual Review of Condensed Matter Physics **11**, 369 (2020).

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

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

[6] C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo, Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements, Science Advances **6**, 10.1126/sciadv.aay3050 (2020).

[7] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, J. Heinsoo, J.-C. Besse, M. Gabureac, A. Wallraff, and C. Eichler, Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits, npj Quantum Information **5**, 1 (2019).

[8] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, Repeated quantum error detection in a surface code, Nat. Phys. **16**, 875.

[9] 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).

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

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

[12] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. da Silva, and B. R. Johnson, Demonstration of a parametrically activated entangling gate protected from flux noise, Phys. Rev. A **101**, 012302 (2020).

[13] 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, R. Graff, T. Huang, E. Jeffrey, J. Kelly, P. V. Klimov, F. Kostritsa, D. Landhuis, E. Lucero, M. McEwen, A. Megrant, X. Mi, J. Mutus, M. Neeley, C. Neill, E. Ostby, P. Roushan, D. Sank, K. J. Satzinger, A. Vainsencher, T. White, J. Yao, P. Yeh, A. Zalcman, H. Neven, V. N. Smelyanskiy, and J. M. Martinis, Diabatic gates for frequency-tunable superconducting qubits, Phys. Rev. Lett. **123**, 210501 (2019).

[14] B. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, B. M. Terhal, and L. DiCarlo, Characterization and detection of leakage in simulations of a transmon-qubit surface-17 architecture, arXiv:2002.07119 (in prepartion (2020)).

[15] P. Aliferis and B. M. Terhal, Fault-tolerant quantum computation for local leakage faults, Quantum Info. Comput. **7**, 139 (2007).

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

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

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

[19] J. Ghosh and A. G. Fowler, Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements, Phys. Rev. A **91**, 020302 (2015).

[20] L. DiCarlo, M. D. Reed, L. Sun, B. R. Johnson, J. M. Chow, J. M. Gambetta, L. Frunzio, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf, Preparation and measurement of three-qubit entanglement in a superconducting circuit, Nature **467**, 574 (2010).

[21] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance., Nature **508**, 500 (2014).

[22] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, A. N. Cleland, J. Wenner, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, Nature **519**, 66 (2015).

[23] J. M. Martinis and M. R. Geller, Fast adiabatic qubit gates using only $\sigma_z$ control, Phys. Rev. A **90**, 022307 (2014).

[24] S. Krinner, S. Lazar, A. Remm, C. Andersen, N. Lacroix, G. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, Benchmarking coherent errors in controlled-phase gates due to spectator qubits, arXiv preprint arXiv:2005.05914 (2020).

[25] Y. Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, A. Megrant, J. Y. Mutus, P. J. J. O'Malley, C. M. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, M. R. Geller, A. N. Cleland, and J. M. Martinis, Qubit architecture with high coherence and fast tunable coupling, Phys. Rev. Lett. **113**, 220502 (2014).

[26] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Tunable coupling scheme for implementing high-fidelity two-qubit gates, Physical Review Applied **10**, 054062

(2018).
[27] P. Mundada, G. Zhang, T. Hazard, and A. Houck, Suppression of qubit crosstalk in a tunable coupling superconducting circuit, Physical Review Applied **12**, 054023 (2019).
[28] 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, arXiv preprint arXiv:2005.08863 (2020).
[29] Y. Xu, J. Chu, J. Yuan, J. Qiu, Y. Zhou, L. Zhang, X. Tan, Y. Yu, S. Liu, J. Li, F. Yan, and D. Yu, High-fidelity, high-scalability two-qubit gate scheme for superconducting qubits, arXiv preprint arXiv:2006.11860 (2020).
[30] J. A. Schreier, A. A. Houck, J. Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Suppressing charge noise decoherence in superconducting charge qubits, Phys. Rev. B **77**, 180502 (2008).
[31] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo, Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits, Phys. Rev. Lett. **123**, 120502 (2019).
[32] M. Kjaergaard, M. E. Schwartz, A. Greene, G. O. Samach, M. O. A. Bengtsson, C. M. McNally, J. Braumüller, D. K. Kim, P. Krantz, M. Marvian, A. Melville, B. M. Niedzielski, Y. Sung, R. Winik, J. Yoder, D. Rosenberg, S. L. K. Obenland, . T. P. Orlando, I. Marvian, S. Gustavsson, and W. D. Oliver, A quantum instruction set implemented on a superconducting quantum processor, ArXiv:2001.08838 (2020).
[33] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).
[34] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, Phys. Rev. A **97**, 032306 (2018).
[35] R. Versluis, S. Poletto, N. Khammassi, B. M. Tarasinski, N. Haider, D. J. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, Scalable quantum circuit and control for a superconducting surface code, Phys. Rev. Appl. **8**, 034021 (2017).
[36] N. Lacroix, C. Hellings, C. Kraglund Andersen, A. Di Paolo, A. Remm, S. Lazar, S. Krinner, G. J. Norris, M. G. Gabureac, A. B. Blais, C. Eichler, and A. Wallraff, Improving the performance of deep quantum optimization algorithms with continuous gate sets, arXiv preprint arXiv:2005.05275 (2020).
[37] R. Barends, L. Lamata, J. Kelly, L. García-Álvarez, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, E. Solano, and J. Martinis, Digital quantum simulation of fermionic models with a superconducting circuit, Nat. Commun. **6**, 7654 (2015).
[38] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[39] M. A. Rol, *Control for Programmable Superconducting Quantum Systems*, Ph.D. thesis, Delft University of Technology (2020).
[40] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler, Rapid high-fidelity multiplexed readout of superconducting qubits, Phys. Rev. Appl. **10**, 034040 (2018).
[41] M. A. Rol, C. Dickel, S. Asaad, C. C. Bultink, R. Sagastizabal, N. K. L. Langford, G. de Lange, B. C. S. Dikken, X. Fu, S. R. de Jong, and F. Luthi, PycQED (2016).
[42] V. Negîrneac, *Automatic characterization and calibration of a superconducting quantum processor capable of error correction*, Master's thesis, Instituto Superior Técnico, Lisbon, Portugal (2021).
[43] 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).
[44] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Applied Physics Reviews **6**, 021318 (2019).
[45] 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, Applied Physics Letters **116**, 054001 (2020).
[46] See supplemental material.
[47] B. Nijholt, J. Weston, J. Hoofwijk, and A. Akhmerov, *Adaptive*: parallel active learning of mathematical functions (2019).
[48] Qutech Quantum Inspire, `https://www.quantum-inspire.com/`.
[49] C. C. Bultink, B. Tarasinski, N. Haandbaek, S. Poletto, N. Haider, D. J. Michalak, A. Bruno, and L. DiCarlo, General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed, Appl. Phys. Lett. **112**, 092601 (2018).
[50] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Optimized driving of superconducting artificial atoms for improved single-qubit gates, Phys. Rev. A **82**, 040305 (2010).

## SUPPLEMENTAL MATERIAL

This supplement provides additional information in support of statements and claims made in the main text. Appendix A provides further details on the measured transmon parameters. Appendix B presents the characterization of single-qubit gate performance. Appendix C provides evidence for the two-level system affecting the realization of SNZ CZ gates in pair $Q_L$-$Q_{M1}$ using the $|11\rangle$-$|02\rangle$ interaction. Appendix D presents a characterization of the residual ZZ coupling between qubits at the bias point. Appendix E summarizes the technical details of the CZ characterization by repeated 2QIRB. Appendix F presents the numerical simulation of the error budget for SNZ and CNZ CZ gates on pair $Q_{M2}$-$Q_H$.

### Appendix A: Transmon parameters

Our experimental study focuses on four transmons in a patch of our 7-qubit processor. Table S1 provides a summary of measured parameters for the four transmons.

### Appendix B: Single-qubit gate performance

All single-qubit gates are implemented as DRAG-type [43, 50] microwave pulses with a total duration of $4\sigma = 20$ ns, where $\sigma$ is the Gaussian width of the main-quadrature Gaussian pulse envelope. We perform two sets of experiments to jointly quantify the infidelity $\varepsilon$ and leakage $L_1$ of these gates. First, we perform individual single-qubit randomized benchmarking (1QRB) keeping the other three qubits in $|0\rangle$. Second, we perform simultaneous single-qubit randomized benchmarking (S1QRB) on pairs of qubits, keeping the other two qubits in $|0\rangle$. The results obtained from both types of experiment are reported as diagonal and off-diagonal elements in the matrices presented in Fig. S1.

### Appendix C: Flickering two-level system

As mentioned in the main text, we were unable to realize the SNZ CZ gate between pair $Q_L$-$Q_{M1}$ using the $|11\rangle$-$|02\rangle$ interaction due to the presence of a two-level system (TLS) interacting intermittently with $Q_{M1}$ at the flux amplitude placing $|11\rangle$ and $|02\rangle$ on resonance. Figures S2(a) and S2(b) show the negative impact of this TLS when attempting to characterize the $|11\rangle$-$|02\rangle$ interaction by the standard time-domain chevron measurement. While experience shows that it is probable that such a TLS could be displaced or eliminated by thermal cycling at least above the critical temperature of aluminum, we chose instead to use the more flux distant $|11\rangle$-$|20\rangle$ interaction to realize the SNZ CZ gate for this pair. For this interaction, a standard, stable chevron pattern is observed [Figs. S2(c) and S2(d)].

| | $Q_H$ | $Q_{M1}$ | $Q_{M2}$ | $Q_L$ |
|---|---|---|---|---|
| Qubit transition frequency at sweetspot, $\omega_Q/2\pi$ (GHz) | 6.4329 | 5.7707 | 5.8864 | 4.5338 |
| Transmon anharmonicity, $\alpha/2\pi$ (MHz) | -280 | -290 | -285 | -320 |
| Readout frequency, $\omega_{RO}/2\pi$ (GHz) | 7.4925 | 7.2248 | 7.0584 | 6.9132 |
| Relaxation time, $T_1$ (µs) | $37 \pm 1$ | $40 \pm 1$ | $47 \pm 1$ | $66 \pm 1$ |
| Ramsey dephasing time, $T_2^*$ (µs) | $38 \pm 1$ | $49 \pm 1$ | $47 \pm 1$ | $64 \pm 1$ |
| Echo dephasing time, $T_2$ (µs) | $54 \pm 2$ | $68 \pm 1$ | $77 \pm 1$ | $94 \pm 2$ |
| Residual qubit excitation, (%) | 1.4 | 1.2 | 4.3 | 1.7 |
| Best readout fidelity, $F_{RO}$ (%) | 99.1 | 98.5 | 99.4 | 97.8 |

Table S1. Summary of frequency, coherence, residual excitation, and readout parameters of the four transmons. The statistics of coherence times for each transmon are obtained from 30 repetitions of standard time-domain measurements [44] taken over $\sim 4$ h. The residual excitation is extracted from double-Gaussian fits of single-shot readout histograms with the qubit nominally prepared in $|0\rangle$. The readout fidelity quoted is the average assignment fidelity [49], extracted from single-shot readout histograms after mitigating residual excitation by post-selection on a pre-measurement.

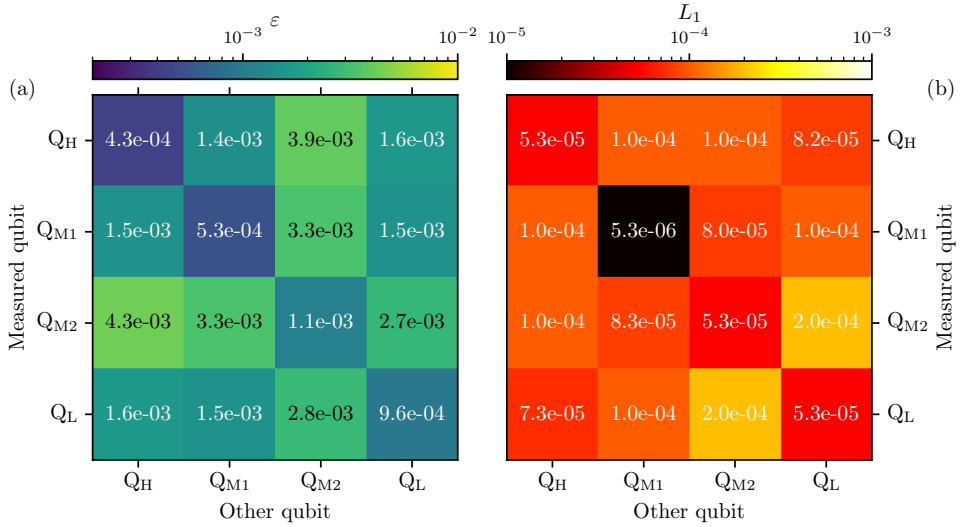


Figure S1. Characterization of single-qubit gate infidelity $\varepsilon$ (a) and leakage $L_1$ (b) using randomized benchmarking (100 randomization seeds). Diagonal elements are extracted from individual single-qubit randomized benchmarking keeping the other 3 qubits in $|0\rangle$. Off-diagonal elements are extracted from simultaneous one-qubit randomized benchmarking on pairs of qubits, keeping the other two qubits in $|0\rangle$.

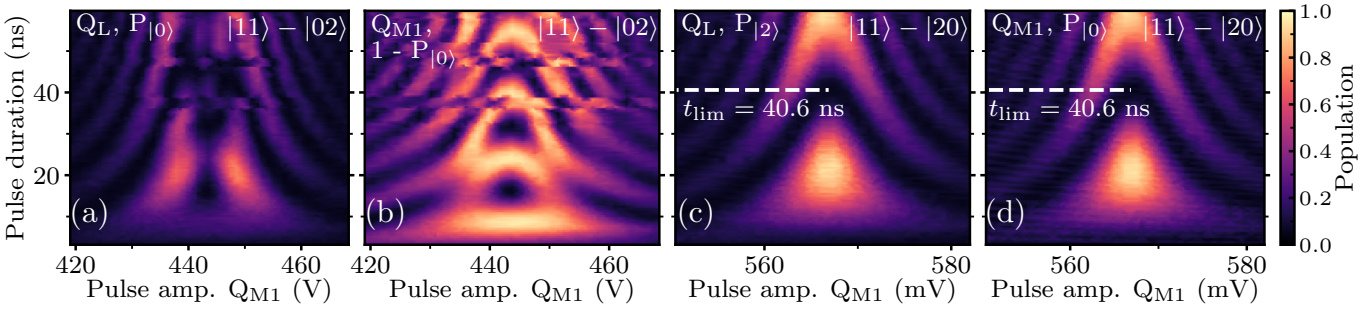


Figure S2. Time-domain characterization of the $|11\rangle$-$|02\rangle$ and (c, d) $|11\rangle$-$|20\rangle$ interactions for pair $Q_L$-$Q_{M1}$. (a, b) Landscapes of (a) ground-state population $P_{|0\rangle}$ of $Q_L$ and (b) total excited-state population $1 - P_{|0\rangle}$ of $Q_{M1}$ as a function of the amplitude and duration of a unipolar square pulse near the $|11\rangle$-$|02\rangle$ resonance. The absence of the expected chevron pattern in these landscapes reflects a flickering TLS resonant with the qubit transition of $Q_{M1}$ at this pulse amplitude. Horizontally shifting fringes in (a) and (b) are due to flickering of the TLS on the scale of a few minutes. These observations preclude the use of the $|11\rangle$-$|02\rangle$ interaction to realize the CZ gate. In contrast, the landscapes of (c) two-state population $P_{|2\rangle}$ of $Q_L$ and (d) $P_{|0\rangle}$ of $Q_{M1}$ and as a function of unipolar square pulse parameters near the $|11\rangle$-$|20\rangle$ resonance reveal a standard, stable chevron pattern. All landscapes were sampled using an adaptive algorithm based on Ref. 47.

## Appendix D: Residual ZZ coupling at bias point

Coupling between nearest-neighbor transmons in our device is realized using dedicated coupling bus resonators. The non-tunability of said couplers leads to residual ZZ coupling between the transmons at the bias point. We quantify the residual ZZ coupling between every pair of qubits as the shift in frequency of one qubit when the state of the other changes from $|0\rangle$ to $|1\rangle$. We extract this frequency shift using a simple time-domain measurement: we perform a standard echo experiment on one qubit (the echo qubit), but add a $\pi$ pulse on the other qubit (control qubit) halfway through the free-evolution period simultaneous with the refocusing $\pi$ pulse on the echo qubit. The results are presented as a matrix in Fig. S3. We observe that the residual ZZ coupling is highest between $Q_H$ and the mid-frequency qubits $Q_{M1}$ and $Q_{M2}$. This is consistent with the higher (lower) absolute detuning between $Q_H$ ($Q_L$) and the mid-frequency transmons, and the higher (lower) transverse coupling $J_2 = \pi/t_{\text{lim}}$ for the upper (lower) pairs.

An alternative way to evidence this residual ZZ coupling is to extract the fidelity of idling using 2QIRB and to compare this fidelity to that of CZ. To this end, we perform 2QIRB of idling (for 60 ns) on pairs $Q_{M2}$-$Q_H$ and $Q_L$-$Q_{M2}$. The results, shown in Fig. S4, show striking differences for the two pairs. For $Q_{M2}$-$Q_H$, the pair with strongest residual coupling, the idling fidelity is significantly lower than the CZ fidelity. This is because the residual ZZ coupling is a

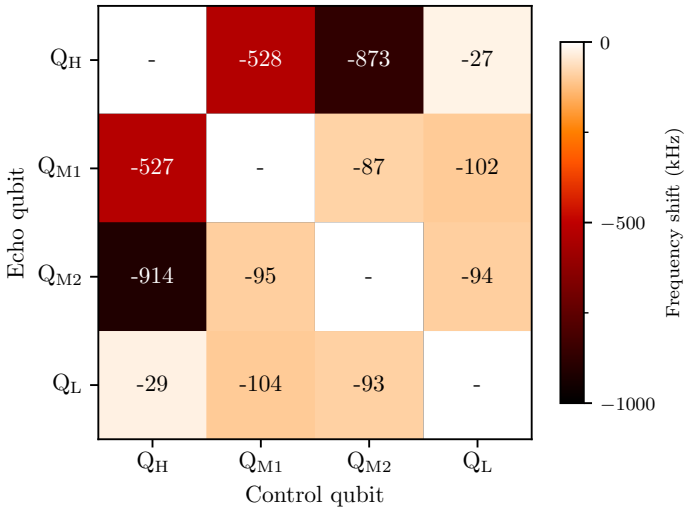


Figure S3. Extracted residual ZZ coupling between all pairs of qubits at their bias points. We report the frequency shift in one qubit (named echo qubit) when the computational state of another qubit (named control qubit) is shifted from $|0\rangle$ to $|1\rangle$.

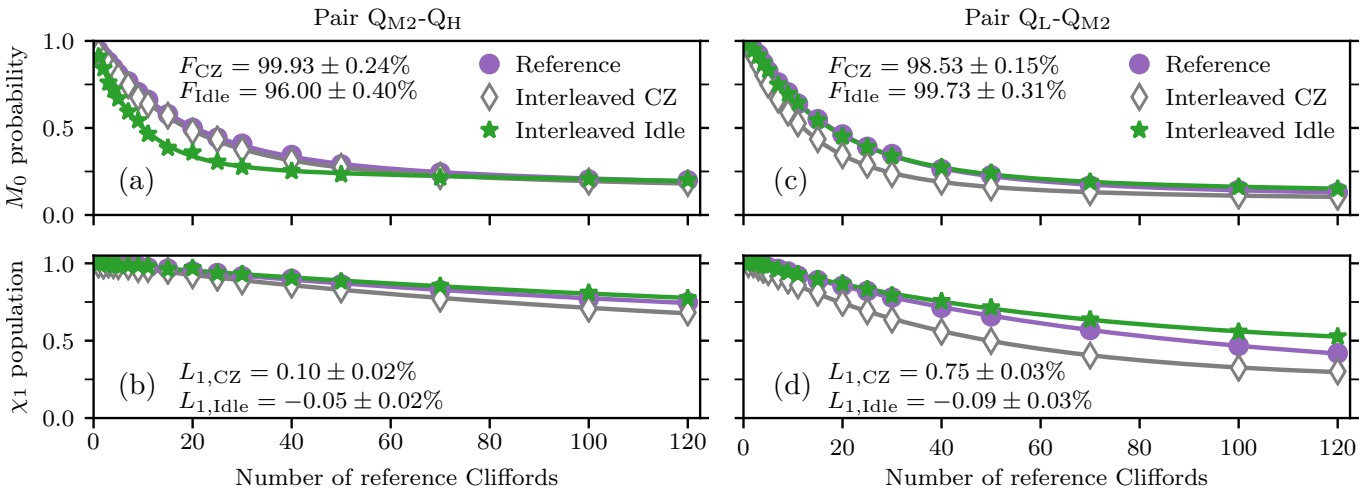


Figure S4. Comparison by 2QIRB of the fidelity and leakage of SNZ CZ versus idling (for an equivalent 60 ns) for pairs $Q_{M2}$-$Q_H$ and $Q_L$-$Q_{M2}$. SNZ gate parameters are provided in Table 1 of the main text. (a, c) Return probability to $|00\rangle$ as a function of the number $N_{2Q}$ of two-qubit Clifford operations in the reference curve. For $Q_{M2}$-$Q_H$, the extracted idling fidelity is significantly lower than the SNZ CZ fidelity. This is due to the high residual ZZ coupling between these two qubits as reported in Fig. S3, which is not refocused during idling but absorbed into the tuneup of the SNZ CZ gate. For $Q_L$-$Q_{M2}$, idling fidelity exceeds SNZ CZ fidelity as the residual coupling is one order of magnitude weaker. (b, d) Population in the computational subspace as a function of $N_{2Q}$. Leakage as a function of $N_{2Q}$ is weakest when interleaving idling steps, leading to negative $L_{1,Idle}$. This is due to seepage back into the computational space (during idling) of the leakage produced by the reference two-qubit Cliffords.

source of error during idling but is absorbed into the tuneup of SNZ. For $Q_L$-$Q_{M2}$, for which the residual coupling is one order of magnitude lower, this trend is not observed.

## Appendix E: Technical details on 2QIRB

Table S2 details technical aspects of the characterization of CZ gates by repeated 2QIRB runs.

| Parameter | $Q_{M1}$-$Q_H$ | $Q_{M2}$-$Q_H$ | $Q_L$-$Q_{M1}$ | $Q_L$-$Q_{M2}$ |
|---|---|---|---|---|
| Number of 2QIRB runs | 39 | 10 | 88 | 35 |
| Number of randomization seeds | 100 | 300 | 100 | 100 |
| Same randomization seeds | No | No | Yes | No |
| Avg. time per 2QIRB run (min) | 17 | 50 | 9 | 17 |
| Total wall-clock time (h) | 28.8 | 16.9 | 16.7 | 14.8 |

Table S2. Technical details of the characterization of CZ gates by repeated 2QIRB. The average time per 2QIRB run is the time required to perform back-to-back measurements of the reference and the CZ-interleaved curves. The total wall-clock time includes the overhead from compilation of RB sequences and other measurements performed in between the CZ 2QIRB runs, e.g., idling 2QIRB (Fig. S4).

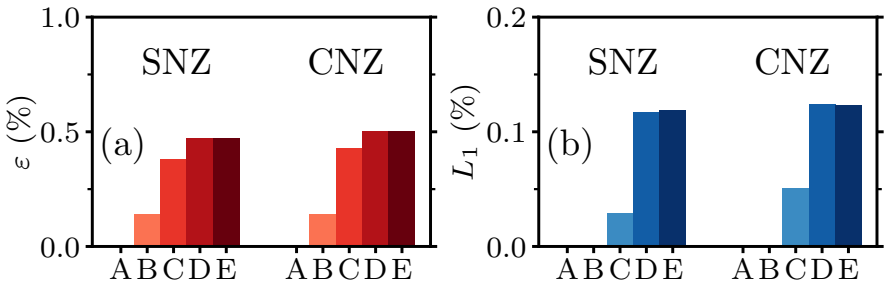


Figure S5. Error budgets for infidelity $\varepsilon$ (a) and leakage $L_1$ (b) obtained by a numerical simulation (as in Ref. 31) of the $Q_{M2}$−$Q_H$ SNZ CZ gate with parameters in Fig. 10 and for a conventional NZ gate with optimized parameters (see text for details). The simulation incrementally adds errors using experimental input parameters for this pair: (A) no noise; (B) relaxation; (C) Markovian dephasing; (D) dephasing from quasistatic flux noise; and (E) flux-pulse distortion.

### Appendix F: Simulation results for SNZ and conventional NZ CZ gates versus different error models

To identify the dominant sources of infidelity $\varepsilon = 1 - F$ and leakage for SNZ CZ gates, we perform a two-qutrit numerical simulation for pair $Q_{M2}$-$Q_H$ with incremental addition of measured error sources (Fig. S5), as in our previous work on conventional NZ [31]. The simulation cumulatively adds: (A) no noise; (B) relaxation; (C) Markovian dephasing; (D) dephasing from quasistatic flux noise; and (E) the remaining measured flux-pulse distortion. The experimental inputs for models B, C and D combine measured qubit relaxation time $T_1$ at the bias point, and measured echo and Ramsey dephasing times ($T_2$ and $T_2^*$) as a function of qubit frequency. The input to E consists of a final Cryoscope measurement of the flux step response using all real-time filters. The simulation suggests that the main source of $\varepsilon$ is Markovian dephasing (as in Ref. 31), while the dominant contribution to $L_1$ is low-frequency flux noise. The latter contrasts with Ref. 31, where simulation identified flux-pulse distortion as the dominant leakage source. We identify two possible reasons for this difference: in the current experiment, the $1/f$ low-frequency flux noise is $\sim$4 times larger (in units of $\Phi_0/\sqrt{\text{Hz}}$) and the achieved flux step response is noticeably sharper. Finally, we use the simulation to compare performance of SNZ to conventional NZ CZ. For the latter, we fix $t_\phi = 0$, $t_{1Q} = 60$ ns $- t_p$, and use the fast-adiabatic pulse shape and $t_p = 45.83$ ns optimized by simulation. Overall, the error sources contribute very similarly to the error budget for both cases. The marginally higher overall performance found for SNZ is likely due to the increased time spent at the sweetspot during the gate time.