

Closed-loop optimization for high-fidelity controlled-Z gates in superconducting qubits

N.J. Glaser^{1,2,*}, F.A. Roy^{1,2,3}, I. Tsitsilin^{1,2}, L. Koch^{1,2}, N. Bruckmoser^{1,2}, J. Schirk^{1,2}, J.H. Romeiro^{1,2}, G.B.P. Huber^{1,2}, F. Wallner^{1,2}, M. Singh^{1,2}, G. Krylov^{1,2}, A. Marx^{1,2}, L. Södergren^{1,2}, C.M.F. Schneider^{1,2}, M. Werninghaus^{1,2} and S. Filipp^{1,2,4,†}

¹ *Technical University of Munich, TUM School of Natural Sciences, Department of Physics, Garching 85748, Germany*

² *Walther-Meißner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany*

³ *Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany*

⁴ *Munich Center for Quantum Science and Technology (MCQST), Schellingstraße 4, 80799 München, Germany*



(Received 9 January 2025; revised 30 June 2025; accepted 15 July 2025; published 20 August 2025)

Achieving fast and high-fidelity qubit operations is crucial for unlocking the potential of quantum computers. In particular, reaching low gate errors in two-qubit gates has been a long-standing challenge in the field of superconducting qubits due to their typically long duration relative to coherence times. To realize fast gates, we use the hybridization between fixed-frequency superconducting qubits with a strongly interacting coupler mode that is tunable in frequency. To reduce population leakage during required adiabatic passages through avoided level crossings, we employ a sensitivity-adaptive closed-loop optimization method to design complex pulse shapes. We compare the performance of Gaussian-square, Fourier-series, and piecewise-constant-slope pulse parametrizations and are able to reach 99.9% controlled-Z-gate fidelity using a 64-ns-long Fourier-series pulse defined by only seven parameters. These high-fidelity values are achieved by analyzing the optimized pulse shapes to identify and systematically mitigate signal-line distortions in the experiment. To improve the convergence speed of the optimization, we implement an adaptive cost function, which continuously maximizes the sensitivity. The demonstrated method can be used for tune-up and recalibration of superconducting quantum processors.

DOI: [10.1103/pckq-2csc](https://doi.org/10.1103/pckq-2csc)

I. INTRODUCTION

In the rapidly developing field of quantum computing, superconducting qubits have emerged as a promising architecture for scalable quantum processors that operate with high speed [1–3]. Superconducting qubits provide a complete gate set by using single-qubit gate rotations of fixed angle, arbitrary Z rotations, and an entangling gate between the individual qubits. Despite significant improvements in coherence times of these qubits [4–8], the fidelity of two-qubit operations such as the controlled- Z (CZ) gates remains a limiting factor.

Many different schemes for realizing two-qubit gates on superconducting qubits have been demonstrated, including

those using parametric coupler interactions [9–11], cross-resonance drives [12,13], microwave-activated controlled phases [14], resonator-induced phase gates (RIP) [15], and the use of higher qubit states [16–18]. To implement CPHASE gates with the latter, the transition between the states $|11\rangle$, where both qubits are excited, and $|20\rangle$, where one qubit is in its second excited state, is used by tuning the frequency of the qubits and potentially also the coupler. In this scheme, high-fidelity gates, with fidelities exceeding 99.8%, have been demonstrated using analytically derived pulse shapes [18].

We investigate an alternative implementation using fixed-frequency qubits that are not affected by flux noise [19] and, therefore, achieve long coherence times.

To realize fast gates, we use the adiabatic hybridization between fixed-frequency superconducting qubits with a strongly interacting coupler mode [20–25]. By varying the frequency of the coupler mode, a ZZ-type interaction between the qubits is activated. This interaction is quantified by the conditional frequency shift $\xi = (\tilde{\omega}_{11} - \tilde{\omega}_{01} - \tilde{\omega}_{10} + \tilde{\omega}_{00})$ of the double-excited state, $|11\rangle \rightarrow e^{i\xi t} |11\rangle$, where $\tilde{\omega}_{n_1 n_2}$ denote the eigenfrequencies of the basis states

*Contact author: niklas.glaser@wmi.badw.de

†Contact author: stefan.filipp@wmi.badw.de

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

$|n_1 n_2\rangle$. To realize a high on-off ratio, ξ must be small during idling operation and as large as possible during gate operation. The requirement for small ξ during idling can be satisfied by exploiting the cancellation of two coupling paths between the qubits: one indirect via the tunable coupler and one through direct capacitive coupling. Due to their opposite signs, a zero ξ configuration can be found with a suitable tunable-coupler frequency in the idling state [23,26,27].

The second condition is realized by the hybridization of higher excited qubit and coupler states. The adiabatic tuning of the coupler frequency then allows adjustment of the conditional frequency shift ξ over several orders of magnitude [21–23,28]. For high-fidelity operation, the coupler trajectory must be shaped such that the correct amount of phase is accumulated and any state transfers caused by energy-level hybridizations are reversed. One option is to use analytically derived pulse-shape descriptions for adiabatic CZ gates [17,23,29]. These, however, typically consider only a single avoided energy-level crossing to derive the pulse description.

Identifying general analytical formulas suitable for the more complex energy structure that arises due to the higher qubit state interactions remains challenging, and one typically resorts to numerical open-loop optimization to efficiently compute pulse shapes [30–32]. In the transfer to experiments, mismatches between models and hardware, such as transfer functions and uncertainties in model parameters may, however, lead to systematic deviations from the targeted system dynamics [33–35]. Although the parameters of the system model and transfer functions can, in principle, all be calibrated individually [36], the large number of potentially correlated parameters leads to a complex optimization problem with low convergence. To overcome these challenges, we instead perform a closed-loop pulse optimization of the CZ gate. Based on a prior pulse shape with low complexity, we optimize pulse shapes with an extended number of parameters on the device, using a global cost function based on randomized benchmarking. This allows us to find optimal parameters even in the presence of strong correlations between parameters. The optimization is able to converge to high-fidelity pulse shapes, correctly considering all avoided crossings and intrinsically correcting for the transfer functions of the control lines.

II. COUPLER-ACTIVATED CONDITIONAL PHASE

The CZ gate is realized in a unit cell of a superconducting qubit architecture with a pair of fixed-frequency transmon-type qubits (Q_1 , Q_2) coupled via a flux-tunable coupling element C . The corresponding circuit diagram shown in

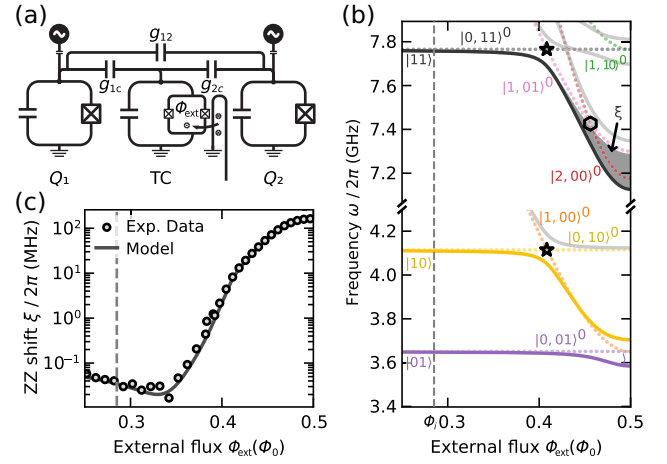


FIG. 1. (a) Lumped-element circuit diagram of the system with two qubits (Q_1 , Q_2) and a frequency-tunable coupler (TC). External drives are applied to each qubit for single-qubit control. The coupler is controlled via a current-driven flux line generating an external flux Φ_{ext} through the SQUID loop. The three elements are capacitively coupled with coupling strength g_{ij} . (b) Energy-level diagram of the two-qubit system in the one-excitation and two-excitation manifolds. The solid lines show the adiabatic computational states $|n_1 n_2\rangle$ as a function of the external flux Φ_{ext} . The dotted lines show the system's bare states $|n_c, n_1 n_2\rangle^0$. The gray shaded area indicates the flux-dependent frequency shift ξ . The dashed line indicates the idle position of the coupler. (c) Frequency shift ξ measured as a function of the external flux Φ_{ext} .

Fig. 1(a) is described by the Hamiltonian

$$\begin{aligned} \hat{H}/\hbar = & \sum_{i=1,2,c} \left(\omega_i \hat{a}_i^\dagger \hat{a}_i + \frac{\alpha_i}{2} \hat{a}_i^\dagger \hat{a}_i^\dagger \hat{a}_i \hat{a}_i \right) \\ & - \sum_{i=1,2} g_{ic} \left(\hat{a}_i^\dagger - \hat{a}_i \right) \left(\hat{a}_c^\dagger - \hat{a}_c \right) \\ & - g_{12} \left(\hat{a}_1^\dagger - \hat{a}_1 \right) \left(\hat{a}_2^\dagger - \hat{a}_2 \right), \end{aligned} \quad (1)$$

with creation and annihilation operators $\hat{a}_i^\dagger$ and $\hat{a}_i$, frequencies ω_i , and anharmonicities α_i , with $i \in \{1, 2\}$ for the qubits and $i = c$ for the tunable coupler. The frequency $\omega_c(\Phi_{\text{ext}})$ of the coupler is tuned by threading an external magnetic flux Φ_{ext} through the superconducting quantum interference device (SQUID) loop of the coupler [37]. The setup is described in Appendix A, and the experimentally realized parameters are summarized in Table I. The elements are capacitively coupled with an asymmetric coupler design [27], shown in Appendix A. The qubits have strong couplings $g_{ic} \sim 60$ MHz to the coupler, while the direct coupling between the qubits $g_{12} = -5.5$ MHz was designed to be small and of opposite sign to reduce the ZZ interaction in the idle state [27].

The conditional frequency shift ξ is tunable from $\xi/2\pi = 20$ kHz to $\xi/2\pi > 100$ MHz via selective energy

TABLE I. Parameters of the device with two qubits Q_1 and Q_2 and a tunable coupler C .

	Q_1	Q_2	C
$\omega_i/2\pi$	4.115 GHz	3.651 GHz	3.7–6.3 GHz
$\alpha_i/2\pi$	−261 MHz	−275 MHz	−124 MHz
$g_{ic}/2\pi$	67 MHz	61 MHz	
$g_{12}/2\pi$	−5.5 MHz		
T_1	(51±6) μ s	(125±25) μ s	(16±7) μ s
T_2^E	(90±30) μ s	(30±4) μ s	
T_2^R	(23±3) μ s	(6.6±1.4) μ s	
$\mathcal{F}^{\text{SQ}}$	(99.93±0.01)%	(99.7±0.1)%	

shifts caused by avoided crossings with tunable coupler states, as shown in Fig. 1(b). At the idle position ($\Phi_i = \Phi_{\text{ext}} = 0.285 \Phi_0$), far detuned from any avoided crossing, the eigenstates $|n_1 n_2\rangle$ of the Hamiltonian in Eq. (1) are approximately equal to the corresponding bare states labeled $|n_c, n_1 n_2\rangle^0$, where n_1 and n_2 are the excitation numbers for the qubits and n_c is the excitation number of the tunable coupler. At this point, coherence-limited single-qubit gate fidelities $\mathcal{F}^{\text{SQ}}$ above 99.7% are measured in simultaneous randomized benchmarking experiments (Table I).

Adiabatically lowering the coupler frequency ω_c by a change in Φ_{ext} leads first to an adiabatic excitation transfer between qubit Q_1 and the coupler C via the interaction of the $|0, 1 n_2\rangle^0$ and $|1, 0 n_2\rangle^0$ states at $\Phi_\star = \Phi_{\text{ext}} = 0.408 \Phi_0$ [indicated by $\star$ in Fig. 1(b)]. At this point, the energy shift is largely independent of the state of qubit Q_2 . Only when tuning the frequency ω_c lower such that the second excited coupler state $|2, 00\rangle^0$ hybridizes with the $|0, 11\rangle^0$ and $|1, 01\rangle^0$ states, the conditional frequency shift becomes large. The interaction between these states, indicated by $\hexagon$ in Fig. 1(b), is present only in the double-excitation manifold, i.e., it is conditional on Q_2 being in the excited state. At the point $\Phi_f = \Phi_{\text{ext}} = 0.456 \Phi_0$, we observe a conditional frequency shift $\xi/2\pi > 75$ MHz, which enables the implementation of conditional phase rotations within nanosecond timescales.

To measure ξ as a function of external flux, we perform Ramsey measurements interleaved by adiabatic flux pulses [see black circles in Fig. 1(c)]. The resulting values agree well with a numerical diagonalization of the Hamiltonian in Eq. (1) using the parameters in Table I. In the numerical simulations, the Hamiltonian is truncated to three qubit levels and four coupler levels. To implement a controlled-Z operation, the flux pulse must be chosen such that the accumulated conditional phase is $\phi_{ZZ} = \int \xi(t) dt = \pi$. Note that the single-qubit phases φ_1 and φ_2 accumulated during the gate operation are corrected using virtual-Z rotations [38,39].

To realize short gate times, the coupler needs to be tuned past the avoided crossings $\Phi_{\text{ext}} > \Phi_\star$ to an interaction

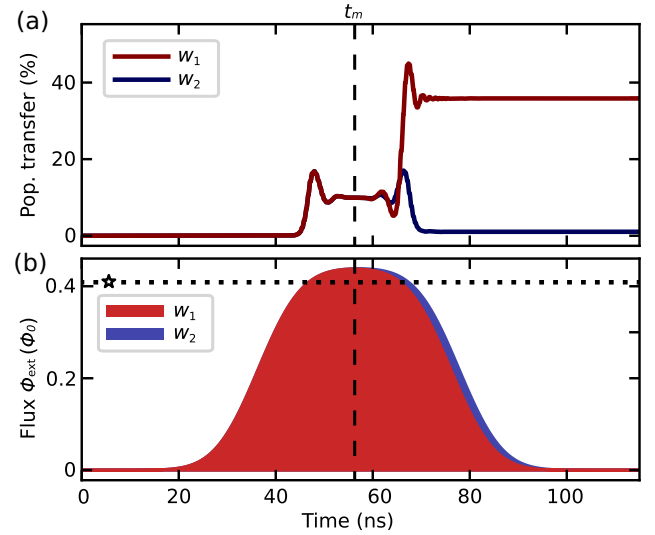


FIG. 2. Population leakage and retrieval. (a) Simulated population transfer to other states for two Gaussian-square pulses with pulse widths w_1 and w_2 . The curves show the population transferred from the $|11\rangle$ state to the $|1, 01\rangle$ and $|2, 00\rangle$ states. (b) Pulse waveforms of the Gaussian-square pulses are shown. The dotted line indicates the flux amplitude at the avoided crossing $\Phi_\star$.

region with a large conditional frequency shift ξ ; however, to remain in the adiabatic regime, the pulse cannot be arbitrarily short [21], as the adiabatic condition limits the slope of the pulse to $d\omega/dt \ll g^2$ with coupling g between two interacting states. Relaxing the adiabaticity condition results in a population transfer between the interacting eigenstates [17,21,40]; however, population loss can be minimized by choosing a suitable pulse width, ensuring that any population lost while passing the avoided crossing is recovered when returning to the idle position Φ_i . This coherent effect is known as Landau-Zener-Stückelberg (LZS) interference [41–43] and depends in particular on the relative accumulated phase differences $\Lambda = \int (\omega_i(t) - \omega_j(t)) dt$ of the populated eigenstates $|i\rangle$ and $|j\rangle$. Using the Landau-Zener model, the remaining population $P(t_e)$ after the return passage is periodic in Λ , with $P_e = 1 - L_m \sin^2(\Lambda)$. The maximal population loss $1 - P_m = 4P_m(1 - P_m)$ is given by the population transferred following a single passage of the avoided crossings $(1 - P_m)$. In Fig. 2, two pulses with pulse width $w_1 = 39.67$ ns and $w_2 = 41.39$ ns are shown, along with a simulation of the population loss from the $|11\rangle$ state to the $|1, 01\rangle$ and $|2, 00\rangle$ states. This simulation highlights how the evolved phases, influenced by the pulse width, determine whether population loss occurs or is mitigated.

To realize a high-fidelity gate, the errors due to leakage must be minimized while at the same time collecting correct phases. For a calibration using a Gaussian-square pulse shape with three pulse-shape parameters (Φ_f, w, τ_r) , we achieve a gate fidelity of 95%, as discussed

in Appendix B. To further increase the fidelity, more control parameters are needed to simultaneously reduce both leakage errors and phase-accumulation errors.

III. PULSE-SHAPE PARAMETRIZATIONS

To realize a gate with a minimal error rate, we explore the following pulse parametrizations, which we calibrate using a closed-loop optimization algorithm [35, 44]: a five-parameter Gaussian-square pulse for a basic implementation of the CZ gate, a piecewise-constant-slope (PiCoS) parametrization with 21 parameters, and a Fourier-component decomposition with eight parameters.

The Gaussian-square pulse serves as a benchmarking baseline, with a minimal parameter set $\mathcal{K}_{\text{GS}} = \{A, w, \tau_R, \varphi_1, \varphi_2\}$, where A denotes the amplitude scaling, w is its pulse width, and τ_R is the rise-time. $\varphi_{1,2}$ are single-qubit Z rotations for the qubits Q_1 and Q_2 .

The PiCoS parametrization is a simplification of the widely used piecewise-constant (PWC) parameterization for pulses digitized at the device sampling rate [30, 35]. It is defined by N_Y amplitudes Y_i with equal spacing w/N_Y across the pulse width w . The N_Y nodes are interpolated via a constant slope, as shown in the inset of Fig. 3(c). This approach significantly reduces the $\mathcal{O}(100)$ parameters of the PWC parameterization, which are determined by the sampling rate of the arbitrary-waveform generator (in our experiment 2 GS/s [45]) and the pulse duration (typically around 50 ns). Additional parameters included in the optimization are a global amplitude scaling A , the pulse width w , and Z rotations φ_1 and φ_2 applied on the respective qubits. This results in a parameter set $\mathcal{K}_{\text{PiCoS}} = \{Y_i, A, w, \varphi_1, \varphi_2\}$. Although parameters such as A and w are redundant with the PiCoS approach, they allow for global adaptations to rescale the pulse, adding significant robustness to the calibration.

A less complex yet similarly versatile pulse-shape parametrization is provided by a decomposition of the envelope into a sum of Fourier components

$$\Phi(t) = A \sum_{n=1}^N \lambda_n \left(1 - \cos \left(2\pi n \frac{t - t_p/2}{w} \right) \right), \quad (2)$$

with the maximum amplitude A , the width of the pulse shape w , within a total sampled pulse duration t_p , and the normalized Fourier coefficients λ_n . This parametrization results in a parameter set $\mathcal{K}_{\text{Fourier}} = \{A, w, \lambda_n, \varphi_1, \varphi_2\}$. Using only low-frequency Fourier coefficients, smooth pulse shapes can be ensured. Numerical open-loop optimizations indicate that a good balance between the number of control parameters and sufficient controllability can be achieved with five Fourier coefficients [28, 46].

For the calibration of the pulse shapes, we employ a closed-loop optimization based on the covariance matrix adaptation evolution strategy (CMA-ES) algorithm, which

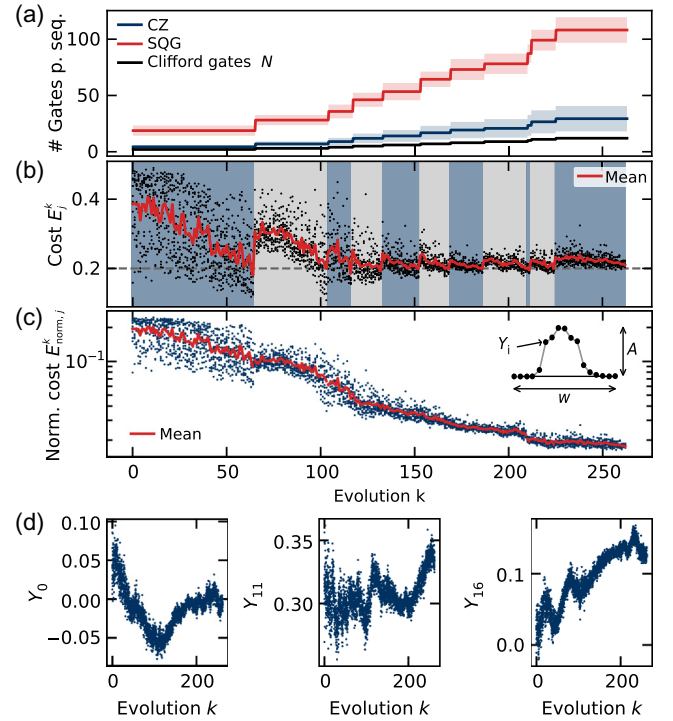


FIG. 3. Exemplary optimization curves for a PiCoS pulse-shape parametrization. (a) Number of CZ gates and single-qubit gates (SQG) per sequence, averaged over the 80 measured randomized sequences. (b) Measured sequence infidelity E_j^k during the optimization. Blue points show the sequence infidelities of each candidate. The red line shows the mean value over the candidate solutions of each evolution. The change in shading indicates increases in cost function sensitivity S . (c) Convergence of normalized sequence infidelity $E_{\text{norm},j}^k = E_j^k/N$ on a logarithmic axis. The inset illustrates a schematic of the PiCoS pulse shape with the nodes Y_i , the amplitude A , and the width w . (d) Evolution of selected nodes Y_i during the optimization of the PiCoS parametrization.

converges efficiently even in the presence of noise [47]. Because, in contrast to simple one-parameter sweeps, it simultaneously optimizes all pulse parameters, it can efficiently cope with parameter correlations, as detailed above and in Appendix B. We take the sequence fidelity of randomized benchmarking sequences as the cost function and enhance the optimization procedure by adapting the number of Clifford gates during the optimization. Starting from a roughly calibrated gate, the gate fidelity can then be continuously optimized without requiring additional restarts.

In each optimization iteration (evolution) k the optimizer chooses P parameter-combinations $\{\mathcal{K}_j^k, j \in [1, P]\}$ to be probed. The parameters are sampled from Gaussian distributions, defined via a mean $m^{(k)}$ for each parameter and a covariance matrix. We use the optimized randomized benchmarking for immediate tune-up (ORBIT) protocol to determine a cost E_j^k for each candidate solution $\mathcal{K}_j^k$, which

reflects the performance of the resulting gate [35,44]. The cost measures the qubit population after randomized benchmarking sequences comprising N Clifford gates. We measure $M = 80$ different randomizations that are each averaged over 128 shots. For every parameter combination $\mathcal{K}_j^k$ of the CZ gate, the cost E_j^k is measured. The weighted average and the covariance of $\mathcal{K}_j^k$ with respect to E_j^k are then used to update the sampling distribution's mean and covariance matrix. This process is iterated until the distribution has converged to a minimum. This makes the optimizer generally well suited to experimental optimization, as it provides robustness to slow drifts in the cost function [47]. If possible, all j evaluations should be performed within a single experiment to distribute system fluctuations evenly across all cost-function values E_j^k . Furthermore, the classical processing overhead is reduced by using a single experiment per evolution [35].

The measured cost functions E_j^k are influenced by experimental noise, making it essential to maximize the sensitivity S_k during each evolution. The sensitivity of the cost function with respect to changes in the CZ gate error ϵ is defined as $S_k(N) = (dE_j^k/d\epsilon) \mathcal{N}^{-1}$, where $\mathcal{N}$ is a normalization constant. The maximum sensitivity depends on the number of Clifford gates N and the average gate errors ϵ for a given parameter combination $\{\mathcal{K}_j^k\}$. Although the exact error of the gate is not known during the optimization, the ideal sensitivity S_k can be approximated based on the value of the cost function E_j^k , as discussed in more detail in Appendix C 1. Generally, the sensitivity $S_k(N)$ remains above 90% of its maximum value for cost-function values E_j^k ranging between 0.2 and 0.4. In contrast to other works [22,44], we employ a dynamic update of the number of Clifford gates N during the optimization process. In particular, the number of Clifford gates is increased if the mean cost-function value across the evolution $\bar{E}^k$ falls below 0.2.

As this changes the actual value of the cost function, an optimization procedure capable of handling changes to the investigated cost function is needed. This is in contrast to an update rule where only the number of repetitions is adjusted [48], which, however, increases the run time for both low- and high-fidelity evaluations.

In the CMA-ES algorithm, the cost function can be rescaled during the optimization process as it compares values within each set of parameter combinations $\mathcal{K}_j^k$ of a single evolution. This contrasts with commonly used optimizers, such as Nelder-Mead [49] or L-BFGS [50], where the cost E_j^k of parameter combinations is added individually and compared with previous evaluations. For additional details on the algorithm, see Appendix C.

An exemplary optimization run for the PiCoS parametrization is shown in Fig. 3. The optimization begins with low sensitivity S and $N = 2$ Clifford gates for an initial parameter combination $\{\mathcal{K}_j^0\}$, where individual candidate solutions

with two-qubit gate errors up to $\epsilon \sim 50\%$ are measured. As the optimizer identifies combinations with higher fidelities, the sensitivity S and, therefore, the number of gates, is incremented in steps of $N \rightarrow N + 1$ to $N = 12$ Clifford gates [Fig. 3(a)]. This stepwise increase allows for the separation of small variations between the high-fidelity candidate parameters, as the mean cost per evolution $\bar{E}^k$ remains between 0.2 and 0.4 [Fig. 3(b)].

The optimization of the PiCoS pulse with 21 parameters converges after 265 evolutions and a total time of 12 h to a gate error of $\epsilon^{\text{PiCoS}} = 0.25 \pm 0.09\%$. The optimization is terminated if $\bar{E}^k$ remains steady over multiple evolutions k . The final result of the optimization is chosen as the parametrization $\mathcal{K}_j^k$ with the lowest E_j^k within all evolutions since the last sensitivity update. To illustrate the convergence of the protocol, we compute the normalized Clifford infidelity $E_{\text{norm},j}^k = E_j^k/N$ by dividing the measured sequence infidelity E_j^k by the number of Cliffords per sequence. After 200 evolutions, $E_{\text{norm},j}^k$ is reduced by an order of magnitude, as shown in Fig. 3(c). A select set of amplitude values Y_i is shown in Fig. 3(d).

The required number of evolutions k_{max} and parameter combinations per evolution P depend on the number of pulse-shape parameters, as further discussed in Appendix C 4.

IV. HIGH-FIDELITY CZ GATE OPTIMIZATION

We apply the optimization algorithm described above to different pulse parametrizations. Initially, the pulse shapes are coarsely calibrated in amplitude, pulse width, and single-qubit phases φ_1 and φ_2 , ensuring an error of $\lesssim 10\%$. The gates are then optimized using the optimization procedure described above. Following the optimization, we evaluate the gate error ϵ using interleaved randomized benchmarking [51]. The performance of the different pulse implementations, along with their respective pulse durations, is summarized in Table II.

The calibrated Gaussian-square pulse parametrization with a pulse width $w = 60$ ns results in a gate error of $\epsilon^{\text{G-Sq}} = 4.79(19)\%$. This high error indicates that the gate is strongly limited by coherent processes, suggesting

TABLE II. Characteristics of optimized gates. Performance of the discussed gates—Gaussian-square (GS), Fourier-series, and PiCoS pulses—with number of optimized parameters n_p , fixed pulse length t_p , CZ gate error ϵ , and applied precompensation for (Fourier, C) and (PiCoS, C).

	n_p	t_p	t_b	ϵ
GS	5	50 ns	50 ns	$(4.79 \pm 0.19)\%$
Fourier	8	60 ns	0 ns	$(0.52 \pm 0.15)\%$
PiCoS	21	60 ns	0 ns	$(0.25 \pm 0.09)\%$
Fourier, C	8	64 ns	0 ns	$(0.09 \pm 0.10)\%$
PiCoS, C	21	20 ns	12 ns	$(0.21 \pm 0.09)\%$

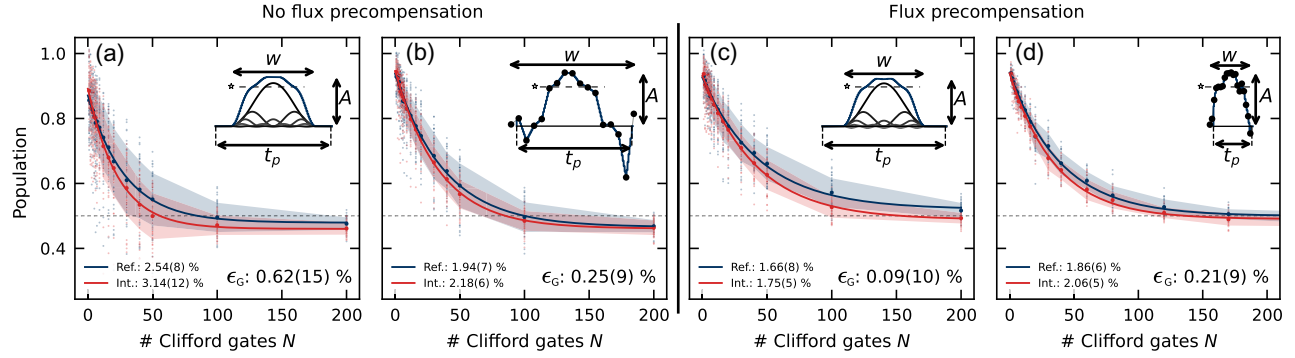


FIG. 4. Optimized pulse-shape parametrizations and interleaved randomized benchmarking results comparing (a),(c) the Fourier-series parametrization and (b),(d) the PiCoS parametrization, with gate errors denoted as ϵ . Panels (a),(b) display the results before applying pulse precompensation, while panels (c),(d) show the results after the implementation of precompensation.

that this parametrization is not well suited to achieving high-fidelity coherence-limited gates. In contrast, gates with the Fourier-series parametrization with five Fourier components and a pulse width $w = 60$ ns achieve an optimized gate infidelity of $\epsilon^{\text{Fourier}} = 0.62 \pm 0.15\%$ [see Fig. 4(a)]. This is a significant improvement in comparison to the Gaussian-square pulse and the coarse calibration. This optimization converges after 80 evolutions, requiring a calibration time of 142 min. With the PiCoS pulse parametrization, the gate error is further reduced by a factor of 3 to $\epsilon^{\text{PiCoS}} = 0.25 \pm 0.09\%$ [see Fig. 4(b)]. Moreover, interleaved randomized benchmarking measurements for the PiCoS pulses show a lower spread between randomizations compared to the Fourier-series pulse, indicating a reduction of unitary errors [52].

Comparing the Fourier-series and PiCoS pulses, the pulse shape is similar, with a distinct slowdown of the trajectories upon entering the interaction region near $\star$. This aligns with the expectation that reducing the slope at an avoided crossing helps to minimize leakage. The Fourier-series pulse exhibits signs of leakage out of the qubit subspace, as indicated by the population deviating from the fully depolarized value in the limit of large N_C [35,53,54]. In contrast, the PiCoS pulse mitigates leakage due to its additional degrees of freedom. Most leakage will occur into the coupler states, as confirmed by numerical simulations; however, since the coupler states are not directly measurable on the current device, an experimental verification of the leakage channels remains outside the scope of this work.

The optimization of the PiCoS pulse also results in asymmetric shoulders and introduces negative flux values in place of the zero-amplitude regions [see Fig. 4(b)]. This suggests that the optimization routine seeks to compensate for distortion effects in the control line; however, due to the slow response time, of the order of microseconds, the routine cannot fully eliminate all distortion effects within the pulse duration. Therefore, memory effects influencing subsequent pulses in the experimental sequence remain

uncorrected [55]. To address these limitations and further enhance gate quality, we characterize the line-transfer function using an extension of the cryoscope protocol described in Ref. [56]. The extension is adapted to our architecture, which does not allow direct readout of the coupling element. Instead, we use the interaction between the qubit and the tunable coupling element, as described in more detail in Appendix D. The characterization reveals significant low-pass filtering effects that limit the fidelity of consecutive CZ gates. To counteract the signal distortion, we systematically correct the transfer function by employing on-device signal predistortion using an HDAWG from Zurich Instruments [57]. With this, we suppress low-pass filtering, reducing the characteristic timescales from 2.5 μ s to 10 ns. The optimizer can then adapt the pulse shapes more effectively, fully compensating for remaining distortions within the pulse duration.

Reoptimization of Fourier-series pulses with 64 ns duration, now incorporating predistortions, achieves significantly reduced gate errors of $\epsilon^{\text{Fourier, C}} = 0.09 \pm 0.1\%$ [see Fig. 4(c)]. Based on the experimental data, the gate fidelity is predominantly limited by the dephasing of the qubit states. Repeating the procedure for the PiCoS pulse reveals that, while the gate fidelity has reached its optimal level, there remains potential to reduce the gate duration. In fact, for pulses with a length of 20 ns followed by a 12-ns buffer period of zero signal amplitude, we achieve a gate error as low as $\epsilon^{\text{PiCoS, C}} = 0.21 \pm 0.09\%$, as shown in Fig 4(d). Further reducing the pulse length, however, leads to increased gate errors, which we attribute primarily to line distortions occurring on similar timescales. For the presented pulse, we observe minimal indications of leakage from the interleaved randomized benchmarking experiment, particularly when compared to the Fourier-series pulse.

V. DISCUSSION AND CONCLUSION

We have demonstrated a closed-loop optimization of parametrized pulses in combination with a signal

predistortion to correct on timescales exceeding the gate duration. This approach achieves two-qubit gate fidelities around 99.9% in less than 64 ns, a significant improvement compared to Gaussian-square pulses, which yield error rates of around 5% but can be optimized using conventional tune-up methods based on individual parameter calibrations. Our approach of a closed-loop CMA-ES-based simultaneous parameter optimization in conjunction with increased-complexity pulse shapes, i.e., Fourier-series and PiCoS pulses, enables gate errors to be reduced below 0.3%. Notably, we have demonstrated that the optimized PiCoS parameterization allows us to reduce leakage and intrinsically correct flux-line distortions.

In addition, optimization results have been used to identify systematic errors. In particular, flux-line distortions are reflected in the pulse shapes after the optimization routine has converged to a high-fidelity gate. This inference of system properties through black-box optimal control, followed by the subsequent compensation of systematic deviations, enables the realization of control pulses with reduced complexity, a cornerstone of optimal control theory [33–35]. System identification, as demonstrated in this work, will facilitate more accurate open-loop optimal control and provide an improved understanding of the system parameters.

As part of the systematic distortion mitigation, we extended the cryoscope protocol [56] to indirectly measure flux-line distortions via nontunable elements. By inverting the system transfer functions directly on the classical control hardware [57], we compensate for these distortions. This approach suppresses long-term memory effects on the scale of the experimental sequences, which allows us to reach average gate fidelities above 99.9%.

Furthermore, the closed-loop optimal control scheme shows promise for a quantum processor tune-up application. For this, the required time per optimization has to be further reduced. We have shown that this can be achieved through system identification, reducing the optimization time from 12 to 4.5 h by using a pulse shape with less complexity. We have analyzed the primary factors contributing to optimization run time, including classical processing time and passive reset, as detailed in Appendix C 3. There, we outline strategies to decrease the run time by more than a factor of 100.

This optimization method is also well suited to recalibration tasks, where metaparameters such as previous parameter spreads and covariance matrices can be retained. This allows the recalibration process to converge rapidly, especially when only minor adjustments to the optimal parameters are required. This approach could ensure high system uptime by interleaving optimization evolutions with algorithmic tasks, thereby tracking and compensating for small parameter drifts in real time. Such a strategy would preserve high-fidelity operations throughout the algorithm's run time.

ACKNOWLEDGMENTS

We acknowledge financial support from the German Federal Ministry of Education and Research via the funding program *Quantum Technologies – From Basic Research to the Market* under Contracts No. 13N15680 *GeQCoS* and No. 13N16188 *MUNIQC-SC*; the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via Project No. FI2549/1-1 and Germany's Excellence Strategy EXC-2111-390814868 *MCQST*; and the European Union's Horizon research and innovation program through the projects *OpenSuperQPlus100* (Grant No. 955479) and *MOlecular Quantum Simulations (MOQS)* (Grant No. 955479). We also acknowledge support from the EU MSCA Cofund *International, Interdisciplinary, and Intersectoral Doctoral Program in Quantum Science and Technologies (QUSTEC)* (Grant No. 847471). This research is part of the *Munich Quantum Valley*, supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: EXPERIMENTAL SETUP

The qubits are controlled and measured using a *Super High Frequency Qubit Controller* (SHFQC) from Zurich Instruments. A schematic of the experimental setup is shown in Fig. 5. We use charge-coupled direct drive lines on the chip for qubit control. The readout resonators are coupled to a common feedline. The feedline is attenuated by -80 dB distributed over the temperature stages and filtered with a 5.5-GHz high-pass filter (VHF-5050+). The drive lines are attenuated by -60 dB distributed over the different temperature stages and filtered with a 3.5- to 4.5-GHz bandpass filter (VBFZ-4000-S+).

We use two 4–12 GHz isolators on the output line. The output signal is amplified by a 4–8 GHz 40-dB HEMT cryogenic low-noise amplifier (LNF-LNC4_8G) thermalized at the 3-K stage. The signal is amplified by a 40-dB low-noise room-temperature amplifier (DBLNA104000800A) before being digitized at the SHFQC input. The flux offset and drive to the coupler are generated using one channel of a Zurich Instruments HDAWG. The flux line is attenuated at room temperature by -10 dB and by -20 dB at the 3 K stage and filtered using a 1.35 GHz low-pass filter (VLFX-1350+).

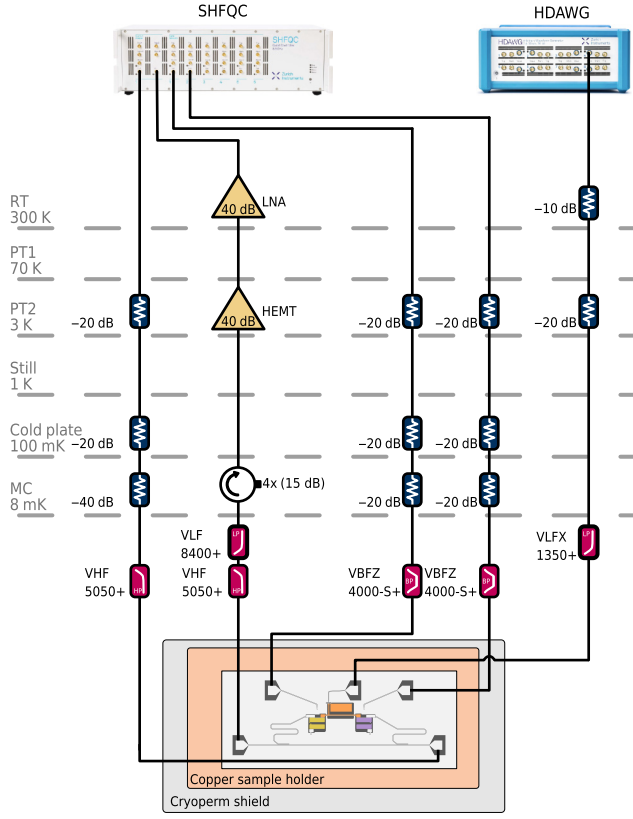


FIG. 5. Experimental setup and schematic of the wiring for the superconducting qubit experiment.

APPENDIX B: ZZ PHASE AND LEAKAGE

We illustrate the effect of coherent errors emerging from nonideal pulse shapes and their correlated dependence on pulse-shape parameters using simple pulse parametrizations. As discussed in the main text, the hybridization with the avoided crossings may lead to population losses. We tune the flux of the coupler, and thereby the coupler frequency, using a Gaussian-square pulse defined as

$$\Phi(t) = \Phi_f \prod_{\pm} \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{\pm t + w/2}{\tau_R} \right) \right), \quad (\text{B1})$$

where w is the pulse width, Φ_f is the maximal flux value, τ_R is the rise time, and erf is the Gauss error function. An illustration of the parametrization is shown in the inset of Fig. 6(c). We measure the population $P_{n_1 n_2}$ after applying a flux pulse to the coupler, with the system initialized in one of the computational basis states $|\psi\rangle_i = |P_{n_1 n_2}\rangle$. Figures 6(a) and 6(b) depict the final populations of $|\psi\rangle_2 = |10\rangle$ and $|\psi\rangle_3 = |11\rangle$ as functions of the pulse width w and the flux amplitude Φ_f . For these states, we observe a population transfer indicating stronger interactions between the coupler and the qubit states for increasing flux pulse amplitude Φ_f . A significant increase in population transfer occurs when the system enters the regime beyond the

avoided crossing at $\Phi_f > 0.43\Phi_0$ [see Fig. 1(b)]. The population transfer exhibits coherent oscillations with respect to the pulse width w , consistent with LZS oscillations. For short gate times $w < 20$ ns, no leakage is observed, as the maximum pulse amplitude is constrained by the pulse rise time. For the states $|\psi\rangle_0 = |00\rangle$ and $|\psi\rangle_1 = |01\rangle$, the population transfer is within the measurement errors and is therefore not shown.

Achieving a high-fidelity gate requires carefully selecting the amplitude Φ_f and pulse width w to minimize leakage while ensuring the correct acquired ZZ phase ϕ_{ZZ} . The ZZ phase ϕ_{ZZ} is measured via Ramsey-type experiments, following a protocol similar to that in Ref. [39]. The control qubit is initialized in either $|0\rangle$ or $|1\rangle$, while the target qubit is set to the $| -Y \rangle$ state using an $R_{\pi/2}^X$ rotation. The CZ gate, consisting of the flux pulse on the coupler and Z rotations on the qubits, is then applied. A final $R_{\pi/2}^{\theta}$ pulse maps the evolved state back to the z axis for measurement. The rotation axis angle θ is swept over ten points from -1.1π to 1.1π , and the resulting population is fitted to a cosine function to extract the acquired phase. The ZZ phase ϕ_{ZZ} is determined as the difference between the acquired phases with the control qubit in $|1\rangle$ and $|0\rangle$.

The errors in leakage and ZZ phase ϕ_{ZZ} are estimated using the average gate fidelity [58]. The phase-mismatch error was derived to be

$$E_{\phi} = 1 - \left(\frac{2(d-1)}{d^2 + d} \cos \delta\phi + \frac{d(d-1) + 2}{d^2 + d} \right), \quad (\text{B2})$$

and the population-loss error is

$$E_P = 1 - \frac{1}{(d^2 + d)} \left(d + \left| d - \sum 1 - P_{n_1 n_2} \right|^2 \right), \quad (\text{B3})$$

where d is the Hilbert space dimension. Simultaneously minimizing both errors is essential for a high-fidelity pulse. As shown in Fig. 6(d), for a slice at $\Phi_f = 0.44\Phi_0$, the estimated errors exhibit unsynchronized oscillations, making it difficult to identify parameters for a common minimum error. As discussed in the main text, employing more complex pulse shapes with additional parameters helps to mitigate both errors simultaneously.

APPENDIX C: OPTIMIZATION

This appendix outlines the optimization process, detailing the ORBIT cost function and its sensitivity, the optimization algorithm and its parameters, a run-time analysis, and the convergence time as a function of the parameter count.

1. ORBIT cost function

The cost function for the optimization is derived from randomized benchmarking sequences of fixed length [44].

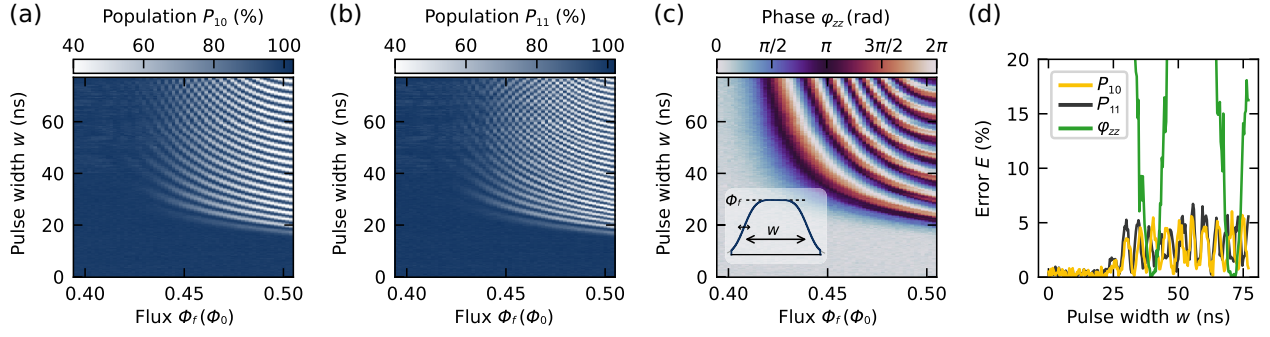


FIG. 6. Leakage induced by the CPHASE gate pulse measured from the (a) $|10\rangle$ and (b) $|11\rangle$ states, alongside the (c) accumulated ZZ phase φ_{zz} . Measurements are shown as functions of pulse width w and amplitude Φ_f . (d) Error rates E for population loss from the $|11\rangle$ ($1 - P_{11}$) and $|10\rangle$ ($1 - P_{10}$) states, along with the accumulated ZZ phase φ_{zz} , at a fixed amplitude of $\Phi_f = 0.44 \Phi_0$. Note that the error is truncated at 20%.

The sequence fidelity of the Clifford-based cost function depends on the number of Clifford gates N and follows the form $\mathcal{F}(N) = A(\mathcal{F}_C)^N + B$ [51]. The fidelity per Clifford gate, $\mathcal{F}_C$, is approximated as $\mathcal{F}_C = (\mathcal{F}_{SQ})^{\bar{n}_{SQ}} \times (\mathcal{F}_{CZ})^{\bar{n}_{CZ}}$, where $\mathcal{F}_{SQ}$ and $\mathcal{F}_{CZ}$ are the single- and two-qubit gate fidelities, respectively, and $\bar{n}_{SQ} = 8.5$ and $\bar{n}_{CZ} = 1.5$ represent the average number of gates per Clifford. This approximation is based on a Clifford-gate decomposition that includes X_π , Y_π , $X_{\pi/2}$, $Y_{\pi/2}$, $X_{-\pi}$, $Y_{-\pi}$, $X_{-\pi/2}$, $Y_{-\pi/2}$, and CZ gates, under the assumption of uncorrelated errors.

The sensitivity S of the sequence fidelity $\mathcal{F}(N)$ to parameter changes depends on the number of Clifford gates and the fidelity of the individual gates in the sequence. The sensitivity S to changes in the CZ gate fidelity, normalized by the CZ gate error, is evaluated with

$$S = \frac{1}{1 - \mathcal{F}_{CZ}} \frac{d\mathcal{F}}{d\mathcal{F}_{CZ}} \quad (C1)$$

and is shown in Fig. 7 for various CZ and single-qubit gate fidelities. For CZ gates with $\mathcal{F}_{CZ} \approx 50\%$, optimal sensitivity S is achieved using short sequences of $N = 2$ Clifford gates. As gate fidelities improve, achieving maximal sensitivity requires longer sequences, with N determined by $N = -1/\log(\mathcal{F}_C)$.

Because single-qubit gates are part of the Clifford decomposition, their fidelity $\mathcal{F}_{SQ}$ also influences the optimal choice of N . As shown in Figs. 7(b) and 7(c), errors in the single-qubit gates reduce the optimal N and weaken the sensitivity to changes in $\mathcal{F}_{CZ}$.

2. Optimization algorithm

The pulse-optimization procedure is carried out using the CMA-ES optimizer [47], as outlined in Algorithm 1. First, the CMA-ES optimizer is initialized with the parameter set $\mathcal{K}_j^k$ to be optimized, along with the respective hyperparameters, such as their initial values and search range. The parameters are normalized based on

the expected optimization range. The choice of the initial parameter spread σ strikes a balance between global exploration and local exploitation, with its exact value depending on the particular pulse optimization.

A batch of candidate solutions $\mathcal{K}_j^k$ with population size P is sampled by the optimizer based on a Gaussian distribution. The cost value E_j^k is computed for each candidate solution by applying the gate parametrized according to $\mathcal{K}_j^k$ in the decomposition of the Clifford sequences (Fig. 8). We prepare $M = 80$ Clifford sequences, each consisting of N Clifford gates. The last Clifford gate is chosen to invert the operation of the previous gates, so that the entire

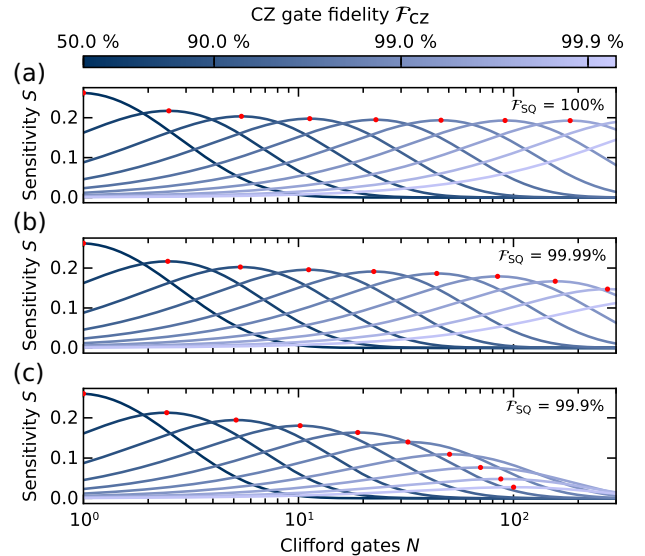


FIG. 7. Sensitivity S of the Clifford sequence to fidelity changes plotted as a function of the Clifford gate count N using Eq. (C1) for CZ gate errors $\epsilon_{CZ} \in [0.5, 0.001]$. Different CZ gate errors are indicated by the color shading. For single-qubit gate fidelities we choose (a) $F_{SQ} = 100\%$, (b) $F_{SQ} = 99.99\%$, and (c) $F_{SQ} = 99.9\%$. Red dots indicate the points of maximal sensitivity.

ALGORITHM 1. Optimization.

```

1: function OPTIMIZATION:
2:   optimizer  $\leftarrow$  CMAES
3:    $N \leftarrow 1$ 
4:   seqs  $\leftarrow$  GET_SEQUENCES(length=N)
5:    $k \leftarrow 0$ 
6:   repeat
7:      $\mathcal{K}_j^k \leftarrow$  optimizer.ASK(P)
8:     experiments  $\leftarrow []$ 
9:     for  $s_j^k \in_j \mathcal{K}_j^k$  do
10:      experiments  $\leftarrow$  COMPILE(seqs,  $s_j^k$ )
11:    end for
12:    results  $\leftarrow$  EXECUTE(experiments)
13:     $E_j^k \leftarrow$  PROCESS(results)
14:    optimizer.TELL( $E_j^k$ )
15:    if  $\text{MEAN}_j(E_j^k) < 0.2$  then
16:       $N \leftarrow N + 1$ 
17:      sequence  $\leftarrow$  GET_SEQUENCE(length=N)
18:    end if
19:     $k \leftarrow k + 1$ 
20:  until stop
21: end function

```

sequence effectively implements a pseudoidentity operation. For interleaved randomized benchmarking [51], half of the M sequences are interleaved with two CZ gates, placed between each decomposed Clifford gate. The set of Clifford sequences remains unchanged for all candidate solutions probed with the same N .

The cost value E_j^k is calculated as the average final ground-state population of Q_1 over M sequences, each repeated 128 times (shots). We use LabOne Q [59] to compile the instruction set for the control devices based on the candidate solution parameters $\mathcal{K}_j^k$ and the Clifford sequences before execution on the control hardware. If the mean cost $\bar{E}^k = \text{mean}_j(E_j^k)$ of the evolution k falls below 0.2, the number of Clifford gates per sequence is increased by one to enhance the sensitivity S of the cost function. The execution is terminated when the mean cost

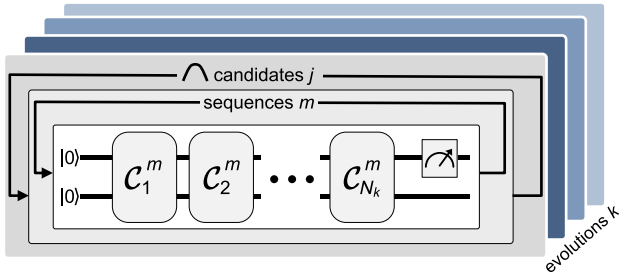


FIG. 8. Illustration of the execution flow of the quantum circuits. The process involves N Clifford gates C_i^m followed by a measurement, iterated over multiple shots. This procedure is repeated for M randomized sequences, indexed by m . Additionally, the process is repeated for J candidate solutions, indexed by j , to evaluate the parameter set $\mathcal{K}_j^k$ in each evolution k .

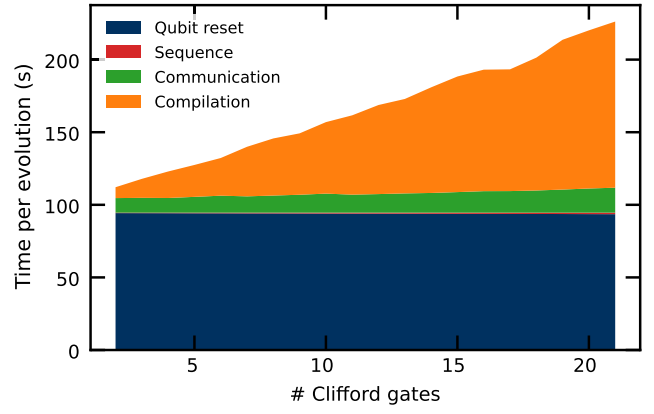


FIG. 9. Breakdown of the time required per evolution during the optimization process, categorized into qubit reset, sequence execution, device communication, and compilation duration, as a function of the number of Clifford gates in the executed sequences.

$\bar{E}^k$ remains steady over multiple evolutions k . Alternatively, to improve the signal-to-noise ratio, the number of Clifford sequences M itself or the number of experimental repetitions can be increased; however, this comes at the expense of a longer run time per evolution due to the increased number of measurements, while maintaining a constant trigger rate.

3. Optimization run time

The optimization duration is primarily determined by the number of evolutions $k_{\max}$ and the time required per optimization step, i.e., per evolution k . The duration is limited from below by the time needed to execute the quantum circuits, including single- and two-qubit gates, measurements, and reset operations, multiplied by the number of experimental repetitions and sequence randomizations. To evaluate the overhead caused by device-instruction preparation and communication, we consider an optimization with a repetition rate of 1.4 kHz, $M = 80$ sequence randomizations, 128 single-shot repetitions, and 13 candidate solutions, resulting in a measurement time of 93 s for one evolution k (blue and red bar in Fig. 9). In our experiment, the repetition rate is primarily bound by the time required for thermal qubit reset (blue bar in Fig. 9), while the actual sequence duration, including measurements and pulses, is almost negligible (red bar in Fig. 9).

The communication time required for control-device programming and data transfers adds a constant offset to the time spent per evolution. This time scales only minimally with sequence complexity and, therefore, the number of Clifford gates per sequence (green area in Fig. 9); however, the compilation of waveforms and instructions in LabOne Q [59] for device control scales linearly with the number of Clifford gates per sequence (orange area in Fig. 9).

The run time can in the future be reduced by employing active reset mechanisms [60–63]. Assuming measurement-based active reset, the on-device measurement time, including reset pulses, corresponds to 0.85 s per evolution for the given number of repetitions, meaning that the total measurement time of the presented PiCoS pulse optimization would be under 4 min; however, due to the limited readout fidelities, which can be improved by an increased resonator coupling, this was beyond the capabilities of the device under investigation.

Furthermore, a full compilation of instructions and waveforms is used. If only minimal updates (atomic waveforms and virtual-Z rotation values) were required, the classical processing time could be drastically reduced. By using efficient processing methods, the classical processing time could in the future be reduced from 120 s to (0.72 ± 0.01) s. This has been validated in a proxy setup with reduced experimental complexity and now needs to be implemented at the full algorithmic scale.

4. Convergence speed

Increasing the complexity of the pulse shape can yield a higher-fidelity solution; however, the increased number of parameters to be optimized also affects the evolutions $k_{\max}$ required for convergence, thereby impacting the overall optimization duration. To guide hyperparameter selection, we examine the scaling behavior of the algorithm's required iterations using several standard optimization problems. For the CMA-ES optimizer, we numerically optimize over test functions with added normally distributed noise. We evaluate the number of evolutions $k_{\max}$ required until the variance of the cost function results E_j^k across the candidate solutions j becomes smaller than the variance of the added noise. For the evaluation of the performance of optimization algorithms, a variety of test functions are commonly used as benchmarks [64,65]. Using such test functions, we observe that the number of evolutions follows a power-law dependence $k_{\max} \sim D^\nu + k_1$, where D is the dimension of the parameter space and k_1 is an offset according to the single parameter cases [see Fig. 10]. For the Spherical, Styblinski-Tang, and Rastrigin test functions, the increase is sublinear with $\nu \approx 0.65$. In contrast, for the Rosenbrock function and a polynomial function, the scaling factors are $\nu = 1.7$ and $\nu = 1.5$, respectively. These scaling factors align with the experimental observations. We observe an approximately linear relationship between the required evolutions $k_{\max}$ and the number of parameters in the pulse optimization, using different parametrizations.

The population size P of the candidate solutions evaluated per evolution k is also influenced by the number of parameters $N_p = |\mathcal{K}|$ and is determined as $P(N_p) = 4 + 4 \log(N_p)$ [47], resulting in a logarithmic scaling of the required candidate points, as shown in Fig. 10(b).

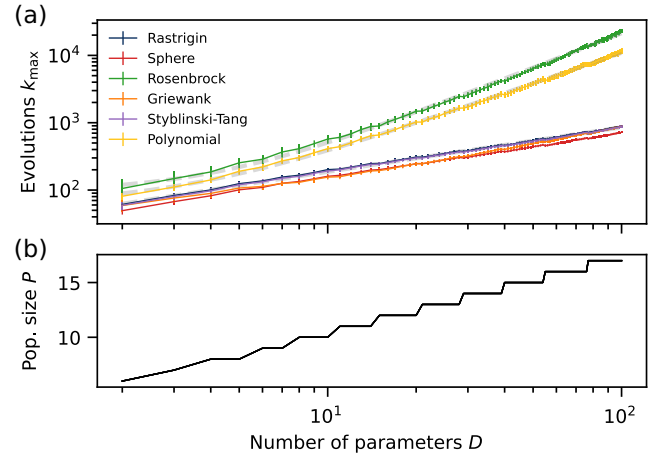


FIG. 10. (a) Scaling of the required number of optimization evolutions $k_{\max}$ with the number of parameters D for various test functions, including Rastrigin, Sphere, Rosenbrock, Griewank, Styblinski-Tang, and Polynomial. (b) Population size P_k per evolution as a function of D , following the logarithmic scaling $P_k(N_p) = 4 + 4 \log(N_p)$.

APPENDIX D: FLUX DISTORTION

1. Signal-line characterization

The transients of the flux pulses are characterized by an extension of the *cryoscope* protocol described in Ref. [56]. The original protocol determines the response of the flux line by measuring the change in the frequency of the qubit connected to it.

In this work, however, the response of the coupler instead of the qubit has to be measured. Since the coupler is not directly connected to a readout line, the protocol has been extended, and the response is measured via the coupled fixed-frequency qubit Q_1 . The eigenstate of the fixed-frequency qubit Q_1 experiences a frequency shift due to the hybridization with the coupler state (see Fig. 1 in the main text).

The phase shift of the qubit state is measured using a Ramsey-type experiment. Qubit Q_1 is initialized in the ground state and then placed into a superposition state by an $R_{\pi/2}^X$ pulse. A nominally rectangular flux pulse is applied via the arbitrary-waveform generator (Zurich Instruments HDAWG), tuning the coupler's frequency close to the fixed-frequency qubit's frequency. The qubit frequency $\tilde{\omega}_1$ is shifted by the interaction with $\omega_c(\Phi)$. A final $R_{\pi/2}^{\theta}$ pulse with a variable rotation-axis angle θ is applied after a fixed interval following the initial $R_{\pi/2}^X$ gate.

In our experiment, we use nine linearly spaced axis angles θ in the range $[0, 2\pi]$, which allows us to extract the phase shift via a cosine-function fit. With the individual phase fits per time step, the data is robust against fluctuations of the coherence time and population loss of the qubit. The flux pulse length t_p is varied from 0 to

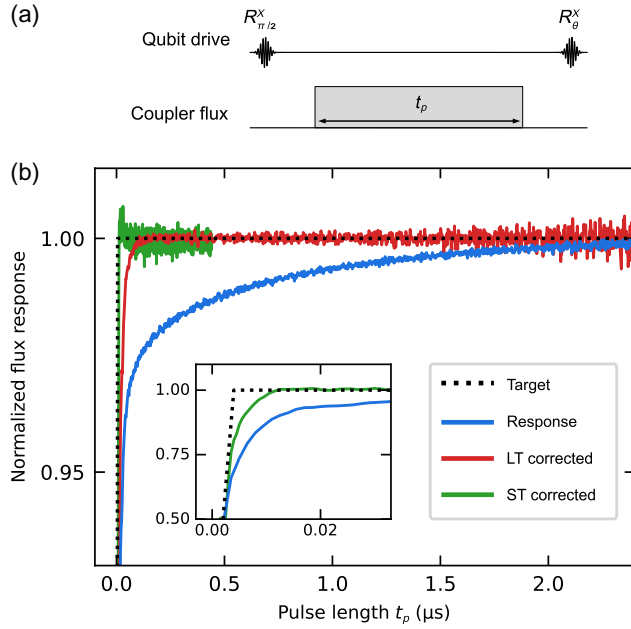


FIG. 11. Measurement of the coupler flux response. (a) Schematic of the sequence used to measure the response of the flux to the applied pulse. The qubit drive sequence is depicted, including the initial $R_{\pi/2}^X$ rotation, the coupler flux pulse of duration t_p , and the final $R_{\pi/2}^X$ rotation. (b) Measured flux responses. The response is plotted as a function of time, normalized to the calibrated nominal amplitude. The uncorrected flux response is shown in blue, showing significant deviation from the target pulse due to distortions over various timescales. The flux response after applying long-timescale corrections (red curve) using two exponential infinite impulse response (IIR) filters improves the alignment with the target pulse. The fully corrected flux-response, incorporating four exponential IIR and finite impulse response (FIR) filter corrections, is depicted by the green curve. Due to the finer temporal resolution and less adiabatic hybridization, the extracted step-response function is noisier. We present only the measurement up to 500 ns due to the increased number of sampling points and the good agreement of the long-term correction for longer times. The black dashed line shows the nominal pulse sent to the AWG. The inset provides an enlarged view of the initial 30 ns of the response.

2.5 μ s, as shown in Fig. 11(a). The effective qubit frequency change as a function of the flux pulse length t_p is estimated by the numerical derivative of the phases with respect to the flux pulse length t_p . Suddenly tuning the coupler close to the avoided crossing induces nonadiabatic transitions, leading to population exchange between interacting states. These transitions introduce additional amplitude and phase oscillations during the interaction on timescales of nanoseconds. We employ a Savitzky-Golay filter with a window length of 10 ns to smooth out these oscillations, thereby limiting the resolution of the flux-response measurement. To calibrate the applied flux amplitude to the resulting qubit frequency change, we

TABLE III. Fitted parameters for the exponential IIR filter corrections.

i	1	2	3	4
A_i	-0.021	-0.012	-0.393	+0.595
τ_i (ns)	846	151	36.0	21.6

measure the frequency shift induced by various nominal flux amplitudes for pulse lengths >300 ns. The nonlinear relation between the applied flux and the qubit frequency accurately resolves the coupler's response near the qubit while suppressing the contribution of turn-off transients when the coupler is detuned from the qubit.

The measured flux-pulse step response [blue curve in Fig. 11(b)] exhibits low-pass filtering characteristics, influenced by various factors. These include the skin effect in the cabling [66,67], which induces distortions on longer timescales, as well as filters, packaging, and on-chip distortions that predominantly impact shorter timescales. After 60 ns, the step response reaches 97% of the final flux value, while reaching 99% requires over 650 ns, with residual deviations persisting for up to 2.5 μ s.

The measured flux response is modeled using exponential infinite impulse response (IIR) filters of the form

$$y(t) = y_{\text{in}}(t) \prod_i (1 + A_i e^{-t/\tau_i}), \quad (\text{D1})$$

where A_i are the amplitudes and τ_i are the decay constants. Long-time-scale distortions are corrected using two exponential IIR filters ($i \in [1, 2]$), resulting in the red curve in Fig. 11(b). Based on this corrected signal, the short-time-scale behavior is addressed using two additional exponential IIR filters ($i \in [3, 4]$). The parameters of the exponential filters are provided in Table III. These field-programmable gate array (FPGA)-based real-time distortion compensations of the HDAWGs [57] are constrained to a granularity of 4 ns by the FPGA clock rate. The initial samples are further corrected using a finite impulse response (FIR) filter, which convolves the signal as follows:

$$y[n] = \sum_{m=0}^{M-1} l[m] y_{\text{in}}[n - m], \quad (\text{D2})$$

with $M = 72$ calibrated coefficients $l[m]$. This process results in the fully corrected signal, shown as the green curve in Fig. 11(b), achieved using the full set of four IIR filters ($i \in [1, 2, 3, 4]$) and the FIR filter. The corrected response aligns with the target response function within $\pm 0.5\%$, with no resolvable distortions after 60 ns.

2. Gate-fidelity effects

Since flux pulses are exclusively used within CZ gates, we evaluate the impact of flux distortions by interleaving

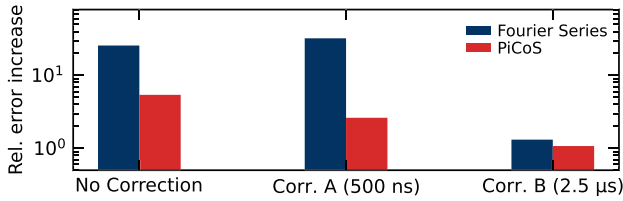


FIG. 12. Relative increase in gate errors when using $n = 2$ CZ gates in comparison to $n = 1$ CZ gates in interleaved randomized benchmarking, separated for Fourier Series and PiCoS pulses. The following cases with no correction, and flux distortion corrections measured and corrected for up to (A) 500 ns, and (B) 2.5 μ s.

two CZ gates—instead of a single CZ gate—between the Clifford gates in the interleaved randomized benchmarking sequence, such that the flux pulses directly follow each other. With no correction, the total errors of the multigate measurement are $(16.6 \pm 2.4)\%$ (up from the individual error $(0.62 \pm 0.15)\%$) for the Fourier-series pulses, respectively. The relative increase in errors is shown in Fig. 12. In contrast, the PiCoS pulse exhibits only a moderately increased error of $(1.6 \pm 0.4)\%$ (up from the individual error $(0.52 \pm 0.19)\%$). This indicates that the PiCoS parameterization effectively reduces the impact of signal distortions; however, it cannot mitigate distortion effects that occur outside the pulse duration.

With the additional on-device flux-distortion correction, the errors of the optimized pulses are reduced, particularly mitigating the error increase observed after the second gate. Our analysis demonstrates the necessity of integrating the complete system response into the evaluation. During our experiments, the flux-pulse response was measured and corrected twice: first with a 500-ns flux-pulse length, and later with a 2.5- μ s flux-pulse length. The initial correction reduced flux distortions, leading to high-fidelity CZ gates, assessed through interleaved randomized benchmarking; however, errors accumulated when probing two successive interleaved CZ gates, suggesting that immediately adjacent flux pulses still degrade the average gate performance, and corrections on longer timescales are needed (see Corr. A in Fig. 12). Later, the response function was measured over a longer time window of 2.5 μ s, close to the maximum time that could be reliably resolved in the presence of decoherence and population transfer into coupler states. This correction effectively eliminated the error increase for successive CZ gates, as shown by Corr. B in Fig. 12. The results also show that the PiCoS pulse minimizes additional errors across all cases.

The error due to flux-pulse bleeding is most pronounced in the accumulated phase φ_1 of Q_1 , which is highly sensitive to the exact coupler-frequency trajectory during the pulse. Even minor changes caused by pulse bleeding can result in significant variations in the acquired phases due to

the strong dependence of the state frequency on the coupler frequency, as shown in Fig. 1(b).

Note that the immediate concatenation of two CZ gates on the same control line is rarely encountered in practical applications. The entangling gates are typically alternated with single-qubit operations or operations on other qubit pairs, which naturally mitigates long-term effects [68,69].

- [1] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, *Annu. Rev. Condens. Matter Phys.* **11**, 369 (2020).
- [2] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann *et al.*, Realizing repeated quantum error correction in a distance-three surface code, *Nature* **605**, 669 (2022).
- [3] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
- [4] A. P. Place, L. V. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, A. Vrajitoarea, S. Sussman *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, *Nat. Commun.* **12**, 1779 (2021).
- [5] C. Wang, X. Li, H. Xu, Z. Li, J. Wang, Z. Yang, Z. Mi, X. Liang, T. Su, C. Yang *et al.*, Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds, *Npj Quantum Inf.* **8**, 3 (2022).
- [6] R. T. Gordon, C. E. Murray, C. Kurter, M. Sandberg, S. A. Hall, K. Balakrishnan, R. Shelby, B. Wacaser, A. A. Stabile, J. W. Sleight, M. Brink, M. B. Rothwell, K. P. Rodbell, O. Dial, and M. Steffen, Environmental radiation impact on lifetimes and quasiparticle tunneling rates of fixed-frequency transmon qubits, *Appl. Phys. Lett.* **120**, 074002 (2022).
- [7] H. Deng, Z. Song, R. Gao, T. Xia, F. Bao, X. Jiang, H.-S. Ku, Z. Li, X. Ma, J. Qin *et al.*, Titanium nitride film on sapphire substrate with low dielectric loss for superconducting qubits, *Phys. Rev. Appl.* **19**, 024013 (2023).
- [8] M. Bal, A. A. Murthy, S. Zhu, F. Crisa, X. You, Z. Huang, T. Roy, J. Lee, D. v. Zanten, R. Pilipenko *et al.*, Systematic improvements in transmon qubit coherence enabled by niobium surface encapsulation, *Npj Quantum Inf.* **10**, 43 (2024).
- [9] R. Bialczak, M. Ansmann, M. Hofheinz, M. Lenander, E. Lucero, M. Neeley, A. O’Connell, D. Sank, H. Wang, M. Weides *et al.*, Fast tunable coupler for superconducting qubits, *Phys. Rev. Lett.* **106**, 060501 (2011).
- [10] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, A universal gate for fixed-frequency qubits via a tunable bus, *Phys. Rev. Appl.* **6**, 064007 (2016).
- [11] G. B. P. Huber, F. A. Roy, L. Koch, I. Tsitsilin, J. Schirck, N. J. Glaser, N. Bruckmoser, C. Schweizer, J. Romeiro, G. Krylov *et al.*, Parametric multielement coupling architecture for coherent and dissipative control of superconducting qubits, *PRX Quantum* **6**, 030313 (2025).
- [12] C. Rigetti, and M. Devoret, Fully microwave-tunable universal gates in superconducting qubits with linear couplings

- and fixed transition frequencies, *Phys. Rev. B* **81**, 134507 (2010).
- [13] 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 *et al.*, Simple all-microwave entangling gate for fixed-frequency superconducting qubits, *Phys. Rev. Lett.* **107**, 080502 (2011).
 - [14] 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).
 - [15] H. Paik, A. Mezzacapo, M. Sandberg, D. T. McClure, B. Abdo, A. D. Córcoles, O. Dial, D. F. Bogorin, B. Plourde, M. Steffen *et al.*, Experimental demonstration of a resonator-induced phase gate in a multiqubit circuit-QED system, *Phys. Rev. Lett.* **117**, 250502 (2016).
 - [16] B. Foxen, C. Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends *et al.*, Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms, *Phys. Rev. Lett.* **125**, 120504 (2020).
 - [17] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim *et al.*, Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler, *Phys. Rev. X* **11**, 021058 (2021).
 - [18] F. Marxer, A. Vepsäläinen, S. W. Jolin, J. Tuorila, A. Landra, C. Ockeloen-Korppi, W. Liu, O. Aho, A. Auer, L. Belzane *et al.*, Long-distance transmon coupler with CZ-gate fidelity above 99.8%, *PRX Quantum* **4**, 010314 (2023).
 - [19] 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).
 - [20] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. Campbell, J. I. J. Wang, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A tunable coupling scheme for implementing high-fidelity two-qubit gates, *Phys. Rev. Appl.* **10**, 054062 (2018).
 - [21] 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).
 - [22] Y. Xu, J. Chu, J. Yuan, J. Qiu, Y. Zhou, L. Zhang, X. Tan, Y. Yu, S. Liu, J. Li *et al.*, High-fidelity, high-scalability two-qubit gate scheme for superconducting qubits, *Phys. Rev. Lett.* **125**, 240503 (2020).
 - [23] J. Stehlik, D. M. Zajac, D. L. Underwood, T. Phung, J. Blair, S. Carnevale, D. Klaus, G. A. Keefe, A. Carniol, M. Kumph *et al.*, Tunable coupling architecture for fixed-frequency transmon superconducting qubits, *Phys. Rev. Lett.* **127**, 080505 (2021).
 - [24] N. J. Glaser, F. Roy, and S. Filipp, Controlled-controlled-phase gates for superconducting qubits mediated by a shared tunable coupler, *Phys. Rev. Appl.* **19**, 044001 (2023).
 - [25] R. Li, K. Kubo, Y. Ho, Z. Yan, Y. Nakamura, and H. Goto, Realization of high-fidelity CZ gate based on a double-transmon coupler, *Phys. Rev. X* **14**, 041050 (2024).
 - [26] P. Mundada, G. Zhang, T. Hazard, and A. Houck, Suppression of qubit crosstalk in a tunable coupling superconducting circuit, *Phys. Rev. Appl.* **12**, 054023 (2019).
 - [27] E. A. Sete, A. Q. Chen, R. Manenti, S. Kulshreshtha, and S. Poletto, Floating tunable coupler for scalable quantum computing architectures, *Phys. Rev. Appl.* **15**, 064063 (2021).
 - [28] J. Chu, and F. Yan, Coupler-assisted controlled-phase gate with enhanced adiabaticity, *Phys. Rev. Appl.* **16**, 054020 (2021).
 - [29] J. M. Martinis, and M. R. Geller, Fast adiabatic qubit gates using only σ_z control, *Phys. Rev. A* **90**, 022307 (2014).
 - [30] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms, *J. Magn. Reson.* **172**, 296 (2005).
 - [31] D. M. Reich, M. Ndong, and C. P. Koch, Monotonically convergent optimization in quantum control using Krotov's method, *J. Chem. Phys.* **136**, 104103 (2012).
 - [32] P. Doria, T. Calarco, and S. Montangero, Optimal control technique for many-body quantum dynamics, *Phys. Rev. Lett.* **106**, 190501 (2011).
 - [33] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Tunable, flexible, and efficient optimization of control pulses for practical qubits, *Phys. Rev. Lett.* **120**, 150401 (2018).
 - [34] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes, Integrated tool set for control, calibration, and characterization of quantum devices applied to superconducting qubits, *Phys. Rev. Appl.* **15**, 034080 (2021).
 - [35] M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, Leakage reduction in fast superconducting qubit gates via optimal control, *Npj Quantum Inf.* **7**, 14 (2021).
 - [36] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Characterizing errors on qubit operations via iterative randomized benchmarking, *Phys. Rev. A* **93**, 012301 (2016).
 - [37] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, *Appl. Phys. Rev.* **6**, 021318 (2019).
 - [38] 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).
 - [39] M. Ganzhorn, D. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner *et al.*, Gate-efficient simulation of molecular eigenstates on a quantum computer, *Phys. Rev. Appl.* **11**, 044092 (2019).
 - [40] Q. Ding, A. V. Oppenheim, P. T. Boufounos, S. Gustavsson, J. A. Grover, T. A. Baran, and W. D. Oliver, Pulse design of baseband flux control for adiabatic controlled-phase gates in superconducting circuits, *ArXiv:2407.02722*.
 - [41] S. N. Shevchenko, S. Ashhab, and F. Nori, Landau-Zener-Stückelberg interferometry, *Phys. Rep.* **492**, 1 (2010).
 - [42] C. Zener, Non-adiabatic crossing of energy levels, *Proc. R. Soc. Lond. A* **137**, 696 (1932).
 - [43] Y. Gefen, and D. Thouless, Zener transitions and energy dissipation in small driven systems, *Phys. Rev. Lett.* **59**, 1752 (1987).

- [44] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey *et al.*, Optimal quantum control using randomized benchmarking, *Phys. Rev. Lett.* **112**, 240504 (2014).
- [45] We employ the Zurich Instruments QCCS for the generation of control signals.
- [46] N. J. Glaser, Optimal control of entangling gates in superconducting tunable-coupler architectures, (2021).
- [47] N. Hansen, The CMA evolution strategy: A tutorial, *ArXiv:1604.00772*.
- [48] F. Sauvage, and F. Mintert, Optimal quantum control with poor statistics, *PRX Quantum* **1**, 020322 (2020).
- [49] J. A. Nelder, and R. Mead, A simplex method for function minimization, *Comput. J.* **7**, 308 (1965).
- [50] D. C. Liu, and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Math. Program.* **45**, 503 (1989).
- [51] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. Da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki *et al.*, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [52] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Effect of noise correlations on randomized benchmarking, *Phys. Rev. A* **93**, 022303 (2016).
- [53] C. J. Wood, and J. M. Gambetta, Quantification and characterization of leakage errors, *Phys. Rev. A* **97**, 032306 (2018).
- [54] J. J. Wallman, M. Barnhill, and J. Emerson, Robust characterization of leakage errors, *New J. Phys.* **18**, 043021 (2016).
- [55] M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, C. C. Bultink, Y. Salathe, N. Haandbæk, J. Sedivy, and L. DiCarlo, Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor, *Appl. Phys. Lett.* **116**, 054001 (2020).
- [56] M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, C. C. Bultink, Y. Salathe, N. Haandbæk, J. Sedivy *et al.*, Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor, *Appl. Phys. Lett.* **116** (2020).
- [57] Zurich Instruments AG, HDAWG User Manual, 750 MHz Arbitrary Waveform Generator, 2025, https://docs.zhinst.com/pdf/ziHDAWG_UserManual.pdf
- [58] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
- [59] Zurich Instruments AG, zhinst/laboneq, 2024, <https://github.com/zhinst/laboneq>
- [60] D. Ristè, J. G. van Leeuwen, H.-S. Ku, K. W. Lehnert, and L. DiCarlo, Initialization by measurement of a superconducting quantum bit circuit, *Phys. Rev. Lett.* **109**, 050507 (2012).
- [61] P. Magnard, P. Kurpiers, B. Royer, T. Walter, J.-C. Besse, S. Gasparinetti, M. Pechal, J. Heinsoo, S. Storz, A. Blais *et al.*, Fast and unconditional all-microwave reset of a superconducting qubit, *Phys. Rev. Lett.* **121**, 060502 (2018).
- [62] M. D. Reed, B. R. Johnson, A. A. Houck, L. DiCarlo, J. M. Chow, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, Fast reset and suppressing spontaneous emission of a superconducting qubit, *Appl. Phys. Lett.* **96** (2010).
- [63] K. Geerlings, Z. Leghtas, I. M. Pop, S. Shankar, L. Frunzio, R. J. Schoelkopf, M. Mirrahimi, and M. H. Devoret, Demonstrating a driven reset protocol for a superconducting qubit, *Phys. Rev. Lett.* **110**, 120501 (2013).
- [64] R. Haupt, and S. Haupt, *Practical Genetic Algorithms, Appendix I: Test Functions* (John Wiley & Sons, Ltd, Hoboken, New Jersey, 2003), p. 205
- [65] T. Back, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms* (Oxford University Press, New York, New York, 1996).
- [66] R. Wigington, and N. Nahman, Transient analysis of coaxial cables considering skin effect, *Proc. IRE* **45**, 166 (1957).
- [67] N. Nahman, A discussion on the transient analysis of coaxial cables considering high-frequency losses, *IRE Trans. Circuit Theory* **9**, 144 (1962).
- [68] S. H. Sack, and D. J. Egger, Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation, *Phys. Rev. Res.* **6**, 013223 (2024).
- [69] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth *et al.*, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).