

Neural-network-based design and implementation of fast and robust quantum gates

Downloaded from: https://research.chalmers.se, 2026-11-03 03:00 UTC

Citation for the original published paper (version of record):
Kuzmanović, M., Moskalenko, I., Chang, Y. et al (2026). Neural-network-based design and implementation of fast and robust quantum gates. Physical Review Applied, 26(3). http://dx.doi.org/10.1103/4zzl-qbsf

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Neural-network-based design and implementation of fast and robust quantum gates

Marko Kuzmanović,[1] Ilya Moskalenko,[1] Yu-Han Chang,[1] Ognjen Stanisavljević,[1] Christopher Warren,[2] Emil Hogedal,[2] Anuj Aggarwal,[2] Irshad Ahmad,[2] Janka Biznárová,[2] Mamta Dahiya,[2] Marcus Rommel,[2] Andreas Nylander,[2] Giovanna Tancredi,[2] and Gheorghe Sorin Paraoanu[1]

[1]*InstituteQ and QTF Centre of Excellence, Department of Applied Physics, School of Science, Aalto University, FI-00076 Aalto, Finland*

[2]*Department of Microtechnology and Nanoscience (MC2), Chalmers University of Technology, SE-41296 Göteborg, Sweden*

(Received 17 July 2025; revised 6 February 2026; accepted 31 August 2026; published 29 September 2026)

We present a continuous-time, neural-network-based approach to optimal control in quantum systems, with a focus on pulse engineering for quantum gates. Leveraging the framework of neural ordinary differential equations (neural ODEs), we construct control fields as outputs of trainable neural networks, thereby eliminating the need for discrete parametrization or predefined bases. This allows for the generation of smooth, control-hardware-agnostic pulses that can be optimized directly using differentiable integrators. As a case study, we design and experimentally implement a short and detuning-robust $\pi/2$ pulse for photon number parity measurements in superconducting transmon circuits. This is achieved through simultaneous optimization for robustness and suppressing the leakage outside of the computational basis. These pulses have very high fidelity ($\approx$99.95%) and realize the parity measurement with an efficiency greater than 99.9% over a detuning range of $\approx \pm 20$ MHz, thereby outperforming traditional techniques while retaining comparable gate durations. This showcases the potential of neural ODEs for high-performance quantum control in experimentally relevant settings.

DOI: 10.1103/4zzl-qbsf

I. INTRODUCTION

High-fidelity control of quantum systems is a central requirement across quantum technologies, especially quantum computing, where quantum advantage and scalability depend on efficient implementations of quantum error-correction algorithms, which in turn rely on robust, error-limited control of the underlying physical qubits [1,2].

While standard approaches such as Rabi pulses or DRAG (derivative removal by adiabatic gate) pulses [3] pulses are often sufficient for idealized two/few-level systems with minimal constraints, they fall short in the presence of noise, detuning, and drive amplitude fluctuations. A solution to these problems was first sought through the use of composite pulse (CP) sequences [4], which do improve robustness but typically suffer from practical limitations, especially when the desired (total) operation is short. CP techniques with phased short-pulse sequences maintain high fidelity in broad ranges of parameters but still require many pulses [5–7]; achieving similar performance with a few pulses remains challenging. Instead, our technique operates at a timescale of single pulses, enabling both fast and robust operations.

More advanced control strategies, such as those based on the GRAPE algorithm [8] or the CRAB method [9], have attempted to address these issues. However, they necessitate either specifying a basis in which the control field is expanded (e.g., [9–13]), or adopting a discretized approach (e.g., [14,15]). Neural networks (NNs) have also been used in a similar way either by discretizing the control field [16–19] or by way of prescribing a basis [20,21] and learning the discrete control parameters. While these approaches generally perform well theoretically, they are associated with certain trade-offs that limit their advantage in experimental implementations. The discretized signal is not generally smooth and requires large bandwidth for faithful representation due to the discontinuities, which are inherently present. This requires *in situ* tuning for each particular combination of quantum and control hardware [15]. Likewise, in addition to the sensitivity to the choice of basis which can be highly problem dependent [10], higher order terms in the expansion typically also increase the required bandwidth.

In contrast, neural networks can be leveraged in a manner that simultaneously addresses both sets of problems: NNs can act as universal function approximators [22], capable of

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

expressing optimal control fields without assuming a fixed basis or discretized representation. The emerging framework of neural ordinary differential equations (neural ODEs) [23], and the tools developed therein, enables a continuous-time approach whereby a neural network *directly* models the control field as a smooth function of time and is optimized through differentiable simulation of the quantum dynamics [24–26]. Additionally, to achieve robustness, most other approaches require specific error modeling for first- or second-order operator cancellation. In contrast, neural networks can leverage existing batch-optimization techniques to achieve high-fidelity operations in the presence of noise or parameter variations.

The primary aim of this work is to introduce and experimentally validate a neural-ODE–based parameterization for quantum control pulses that naturally produces smooth, hardware-compatible waveforms. Within this framework, we design and experimentally verify a high-fidelity, detuning-robust $\pi/2$ pulse for a transmon-based system, motivated by its central role in parity-measurement sequences, its use in Wigner tomography and related cQED experiments. In addition, we numerically construct a robust single-qubit gateset spanning rotation angles $\theta \in [0, \pi]$ to illustrate that the same approach can generate a continuous family of robust control operations.

The NN generated robust $\pi/2$ pulses, owing to their ability to address the qubit with high fidelity even when coupled to a highly populated resonator, not only combat noise in discrete-variable quantum computing but also enable various previously difficult-to-implement operations, such as the creation of large Schrödinger cat states [27], nullifying the need for transmon drive-frequency recalibration. Use of these pulses will also significantly enhance the accuracy of state reconstruction in bosonic QED [28,29] by reducing the coherent dynamical errors during parity mapping—especially for higher dimension states. Such precision is essential for both accurate Wigner tomography [30–32] and reliable error syndrome extraction in bosonic cat-qubit architectures [33]. Besides quantum computing, gates or sequences of gates as considered in this work have applications in other quantum technologies, such as microwave sensing via coherent interaction-free measurements [34–36] and magnetometry [37–39].

Furthermore, the use of a detuning-robust $\pi/2$ pulse can greatly speed up the search for axions and dark photons through efficient microwave photon detection [40]. This is a key motivation behind this work, as well as the context in which the results will be presented. In short, in qubit-based axion detectors a photon populates a resonant cavity, which is coupled to a qubit. The interaction Hamiltonian is $H_{\text{int}} = \chi_{\text{disp}} \sigma_z \hat{n}_{\text{ph}}$, resulting in a shift of the qubit transition frequency by $\delta = 2\chi_{\text{disp}}$ per photon in the cavity. Thus, the following sequence of operations $(\frac{\pi}{2})_X (\beta)_Z (-\frac{\pi}{2})_X$ (where $\beta = 2\chi_{\text{disp}} \langle \hat{n}_{\text{ph}} \rangle \tau$ is the phase accumulated due to the photon associated ac Stark shift during the idle time $\tau = \pi/2\chi_{\text{disp}}$) will result in an identity operation for zero or even number of photons in the detection cavity while it will constitute a $(\pi)_X$ rotation for one, or the odd-numbered, photon state. Thus, reading out the qubit state amounts to performing a photon number parity measurement. For this approach to work, the control pulses used to implement the sequence ought to be true $\frac{\pi}{2}$ pulses, even when detuned by $\delta = 2\chi_{\text{disp}} n_{\text{ph}}$ from the qubit transition frequency.

The problem of engineering *detuning-robust* $\frac{\pi}{2}$ pulses is, therefore, used to demonstrate the utility of the neural network approach. Our method produces smooth, control-hardware-agnostic control fields that outperform standard rectangular and DRAG pulses in robustness while maintaining short durations. We obtain a pulse fidelity of $\approx$99.95% and implement the parity measurement operation with a fidelity greater than 99.9% over a detuning range of $\approx \pm 20$ MHz. This demonstrates precise control over qubit rotations under variable detuning, which, as stated previously, has wide applications in cQED experiments. Furthermore, the method is general, since the design of arbitrary single- and multiqubit gates should be straightforward. While in the present work we restrict experimental and numerical validation to a single qubit and a detuning-robust setting, the approach can also be applied to two-qubit gates. In such multiqubit simulations, Hilbert-space growth increases computational cost, but this can be addressed using reduced effective Hamiltonians that capture the relevant dynamics of the subspace of interest [41,42]. This illustrates that the Neural-ODE framework can, in principle, be extended to multiqubit gates while remaining tractable for simulation, highlighting its practicality. This positions neural ODEs as a useful and easy to use framework for quantum control, capable of addressing challenges that are difficult for traditional approaches, with a minimal increase in technical complexity and without requiring extensive training resources.

The paper is organized as follows: Sec. II gives the details of the neural network approach and describes the experimental setup used for its verification. Sections III A and III B study the problem of generating a detuning-robust $\pi/2$ pulse by considering the qubit and the qutrit subspace, respectively. Section III C evaluates the performance of the obtained pulses in the context of implementing a parity measurement. The main conclusions are delegated to Sec. IV.

II. METHODS

A. Dynamics of a qubit driven by a neural network

Given a quantum system, its time evolution is governed by the propagator U, which satisfies the equation $i\hbar \dot{U}(t) = H(t)U(t)$, with the structure and dimensionality of U and

H set by the specifics of the problem under study. In general, the Hamiltonian contains a number of (fixed) parameters describing the system itself and some number of tunable ones, expressing the drive fields that we can manipulate in order to achieve optimal quantum control. These control fields can be represented as complex time-dependent amplitudes that encapsulate the information about the amplitudes and phases of the externally applied control tones. For the simplest case of a qubit, this reduces to a single control field $\Omega(t)$, with the Hamiltonian (in the frame co-rotating with the qubit):

$$H = \frac{\hbar}{2}[\Omega(t)|1\rangle\langle 0| + \Omega^*(t)|0\rangle\langle 1|]. \tag{1}$$

The control field amplitude $\Omega(t)$ can be identified with the output of a neural network $\Omega(t) = \mathcal{N}(t)$. The equation of motion can now be integrated using a *differentiable* integrator such as "torchdiffeq" [43] This can be done with an adaptive step method (here, we use "dopri5"), without relying on an explicit discretization of time. This allows for the obtained trajectory $U(t)$ to be differentiated with respect to the parameters of the neural network $\mathcal{N}$. Then, one can use gradient descent and backpropagation techniques to tune (train) these parameters until a desired trajectory $U(t)$ or final state $U(t_{\text{end}})$ is obtained.

If one wishes to obtain a family of trajectories parameterized by θ (corresponding to, e.g., different rotation angles), the additional parameter can be passed to the neural network $\mathcal{N}(t,\theta)$, such that the entire class of trajectories is optimized for during the training phase, similar to [24,26]. If one wishes to optimize for resilience against deviations in the control parameters, the output of the neural network can be perturbed stochastically in each training iteration, while requiring that the same operation is realized. Unlike previous works [4,25], this technique does not require analytical, propagator-level, cancellation of the error terms, allowing for optimization against complex noise and error models. Namely, if one wishes to optimize for amplitude and/or detuning robustness, the control field can be taken as $\Omega(t) = (1+\alpha)\mathcal{N}(t)e^{i\delta t}$, where the amplitude perturbation α and the detuning perturbation δ are sampled from appropriately chosen distributions. The training procedure, in the presence of perturbations, is represented schematically in Fig. 1.

B. Neural network

In this work, we implement $\mathcal{N}$ as a feedforward deep neural network with $\tan h$ activations. The depth $d = 3$ and width $w = 20$ of the network are chosen as a balance between expressivity and being able to train the network on an ordinary desktop computer in a matter of hours. The network takes two input parameters: the desired rotation angle θ and the time t. The network outputs two real numbers $O_{1,2}$. The first is related to the instantaneous amplitude of the pulse trough $A(t) = \Omega_{\max}\tan h(O_1)$, while the second one is identified as the pulse phase $\varphi = O_2$, resulting in the complex drive field $\Omega(t) = A(t)e^{i\varphi(t)}$. For training purposes, the pulse duration is taken to be $T = 2\pi$, with $t \in [-\pi,\pi]$, leading to a natural amplitude scale $\Omega_{2\pi}$—the amplitude necessary for a 2π rotation during T. In order to limit excessively high pulse amplitudes that might not be feasible experimentally, $\Omega_{\max}$ is set to $\Omega_{\max} = 3\Omega_{2\pi}$.

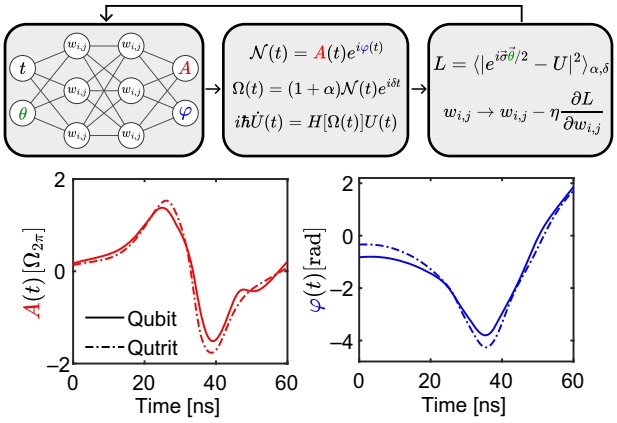


FIG. 1. Top: A schematic representation of the neural network approach: given some target rotation angle θ and time t, the output of the network $\mathcal{N}$ is perturbed by α (amplitude) and δ (detuning) to construct the complex drive field $\Omega(t)$, and the equation of motion is integrated to obtain the propagator U. Based on this, a loss L can be computed, which quantifies the discrepancy between the implemented and the target operation. Lastly, the weights of the neural network $w_{i,j}$ are updated based on the gradients of L w.r.t $w_{i,j}$ and a learning rate η. The cycle is repeated N times or until convergence. Bottom: The resulting pulse amplitude $A(t)$ (left), shown in units of $\Omega_{2\pi} = 2\pi/T_p$ where T_p is the pulse duration, and phase $\varphi(t)$ (right) as a function of time, optimized for a true qubit (solid) and a qutrit (dashed-dotted).

For a qubit, the propagator U can be parameterized via real coefficients c_k as follows:

$$U = c_0\sigma_0 - i\sum_{k\in x,y,z} c_k\sigma_k, \tag{2}$$

with $\sum_{k\in 0,x,y,z} c_k^2 = 1$.

The equation of motion then reads $\partial_t\vec{c} = M\vec{c}$, with:

$$M = \frac{1}{2}\begin{pmatrix} 0 & -\text{Re}(\Omega) & -\text{Im}(\Omega) & 0 \\ \text{Re}(\Omega) & 0 & 0 & \text{Im}(\Omega) \\ \text{Im}(\Omega) & 0 & 0 & -\text{Re}(\Omega) \\ 0 & -\text{Im}(\Omega) & \text{Re}(\Omega) & 0 \end{pmatrix}. \tag{3}$$

We first study the problem of designing a detuning robust gateset, that is a *continuous* family of pulses, which implement rotations by angle $\theta \in [0,\pi]$, which maintain the rotation angle θ as well as the plane of rotation even under

a detuned drive. It is sufficient to obtain a set of rotations generated by σ_x, as the other ones can be realized by applying an additional frame or phase rotation. With these requirements, the loss function is defined as $L = (c_0 - \cos(\theta/2))^2 + (c_x - \sin(\theta/2))^2 + c_y^2 + c_z^2$, the detuning robustness is optimized for in a symmetric interval $\delta \in [-0.4\Omega_{2\pi}, 0.4\Omega_{2\pi}]$, and amplitude robustness is not considered ($\alpha = 0$). As the loss L is only a function of the final operator U, no constraints are placed on the qubit's evolution, allowing the system to take the trajectory that results in a robust operation. For the sake of brevity, the results are relegated to Appendix A, and here, we only note the obtained average fidelity, over the specified θ and δ intervals, of $1 - F \approx 10^{-5}$.

For implementing a parity measurement, we require only a $\pi/2$ rotation and therefore, one of the inputs of the neural network is fixed to $\theta = \pi/2$. Unlike for an arbitrary rotation gate, the exact plane of rotation is not important and need not be fixed. This allows for the constraints to be relaxed, which simultaneously makes the optimization procedure more efficient and leads to improved pulse performance. Since the desired operation is a $\pi/2$ rotation ($c_0 = 1/\sqrt{2}$) in a vertical plane ($c_z = 0$, $c_x^2 + c_y^2 = 1/2$), the loss, or infidelity, of the operation is taken as $L = c_z^2 + (c_0 - 1/\sqrt{2})^2$. Note that, due to the normalization condition, minimizing this quantity simultaneously ensures that $c_x^2 + c_y^2 \approx 1/2$. The form of this loss function is specific to this problem, reflecting the reduced set of constraints, i.e., it is not a special case of the arbitrary rotation one with $\theta = \pi/2$.

Additionally, we focus on detuning robustness, as it is the dominant error source for parity measurements: amplitude errors mainly arise from slow gain drifts, which are largely static over the timescale of typical experiments and can be mitigated by through calibration. In contrast, detuning errors can arise dynamically from photon-number-dependent ac Stark shifts and cannot be addressed in a similar manner. Therefore, α is set to 0, while δ is sampled uniformly from the interval $\delta \in [-0.8\Omega_{2\pi}, 1.1\Omega_{2\pi}]$. This asymmetric interval reflects the experimental circumstances, where the detuning arises from photon-number-dependent ac Stark shifts, $\delta = n_{\mathrm{ph}} 2\chi_{\mathrm{disp}}$, and is therefore positive. While increasing them is feasible and results in higher detuning robustness, we found that it is achieved by an anomalous pulse whose amplitude is nonzero only close to $t = 0$, effectively making it shorter and increasing its bandwidth.

We initialize the network by sampling the weights from a normal distribution with zero mean and unit variance. We evaluate the loss over 100–1000 randomly sampled initial states and start gradient-based optimization from the configuration with the lowest loss, which typically yields a smooth control field. In each training iteration, a batch of 100 to 1000 δ-values is randomly sampled, and the mean loss is computed. The gradient over all of the neural network parameters is computed using automatic differentiation, and their values are updated by using the "Adam" optimizer [44]. On a consumer AMD Zen 2 12-core processor, one such forward pass and automatic differentiation step together take ≈900 ms.

This whole procedure is repeated until convergence (≈10^4 iterations). Although this procedure can still encounter local minima, the training is inexpensive enough to repeat with different seeds until the desired fidelity is achieved. The training procedure is summarized in Fig. 1.

C. Experimental methods

The performance of the obtained pulses is verified on a transmon circuit fabricated using the thick-oxide recipe [45]. The transmon has the transition frequency of $f_{01} = 3.772$ GHz, spectral anharmonicity $\alpha_s = f_{12} - f_{01} = -222.34$ MHz, and relaxation and decoherence times of $T_1 = 131$ μs, $T_2 = 64$ μs, $T_2^{\mathrm{echo}} = 141$ μs. All control signals are generated by the OPX+ arbitrary waveform generator (AWG), which is based on a field-programmable gate array (FPGA), and are up-converted to the microwave domain by the Octave up/down-converter module. Both instruments are manufactured by Quantum Machines. More details about the experimental setup are given in Appendix B.

The neural-ODE optimizer produces a unique waveform for each target gate, which can be stored and played by the OPX+ generator, in the same manner as standard calibrated pulses (e.g., DRAG). A robust gateset corresponds to a finite library of precompiled waveforms. The memory footprint of such a library is modest: for example, a 60 ns pulse sampled at 1 GS/s with I/Q components requires a few hundred bytes. Even a gateset consisting of tens to hundreds of robust pulses therefore occupies well below a megabyte of memory, which is negligible compared to the available waveform storage resources of modern FPGA-based control hardware.

The pulses are characterized through performing standard process tomography [2] at different detunings, δ. The resulting data are used to obtain the experimental process matrix χ_{exp}, which includes contributions from both finite relaxation times and SPAM errors [46]. The unitary generating said operation is estimated in the following way: for a unitary $U_{\mathrm{fit}} = c_0^{\mathrm{fit}}\sigma_0 - i\sum_{k\in x,y,z} c_k^{\mathrm{fit}}\sigma_k$, the corresponding process matrix χ_{fit}, in the basis of $\{\sigma_0, \sigma_x, -i\sigma_y, \sigma_z\}$, is given by $\chi_{\mathrm{fit}}^{j,k} = b_j b_k^*$ where $b_j = c_j$ if $j \in \{0, y\}$ and $b_j = -ic_j$ if $j \in \{x, z\}$. We then look for c_i^{fit}, which minimizes the Frobenius norm $|\chi_{\mathrm{exp}} - \chi_{\mathrm{fit}}|^2$, allowing us to extract the underlying operation.

Additionally, the fidelity of the pulses is estimated using a randomized-benchmark inspired [47] sequence, where an "identity" gate is constructed by 4 successive applications of the pulse. This way deviations from a $\pi/2$ rotation

angle are observed as oscillations, while the fidelity of the pulse can be extracted from the rate of decay toward the maximally mixed state. Lastly, the same analysis is performed on a rectangular and DRAG [3] $\pi/2$ pulse and their performance is compared.

III. $\pi/2$ PULSES FOR PARITY MEASUREMENT

A. Driving a qubit with a neural network

With the decomposition for U presented above, the equations of motion can be presented directly in terms of the real and imaginary components of Ω and the vector of $c_k \in R^4$, allowing for elegant and quick integration. Training the neural network results in a pulse like the one presented in Fig. 1, shown here in terms of its signed amplitude $A(t)$ and phase $\varphi(t)$. For the performance of the pulse to be competitive with a short DRAG pulse, a pulse duration of $T_p = 60$ ns was chosen as a balance between the pulse bandwidth, amplitude, and the width of the robustness region (all of which scale like $\propto T_p^{-1}$). To generate the pulse, the (logical) I and Q quadratures of the AWG are then set to $I(t) = \Omega(t)\cos(\varphi(t))$ and $Q(t) = \Omega(t)\sin(\varphi(t))$.

Such modulated pulses can have higher bandwidth and amplitudes resulting in implementation challenges: the finite bandwidth and nonzero nonlinearity of the generation scheme might lead to discrepancies between the desired and the generated pulse. These can be accounted for by measuring the output of the generator with a high-frequency oscilloscope. Furthermore, the time domain trace can be integrated to estimate the evolution of the qubit under the *generated* drive tone. The results of this procedure, presented in Appendix C, demonstrate the ability to synthesize these complex pulses.

Figure 2 shows the results of the process tomography as a function of the detuning frequency. A detailed description of the quantum process tomography procedure can be found in Appendix D. The experimental traces (colored traces shown on the left panel) resemble the theory (black), though with some discrepancies. Most notable is an offset of the c_z component w.r.t the expected value. Naively, one might interpret this as a frequency offset between the laboratory frame and the qubit. However, the discrepancy remains after careful calibration, and it approximately corresponds to a detuning of ≈400 kHz, well above the level of experimental uncertainty.

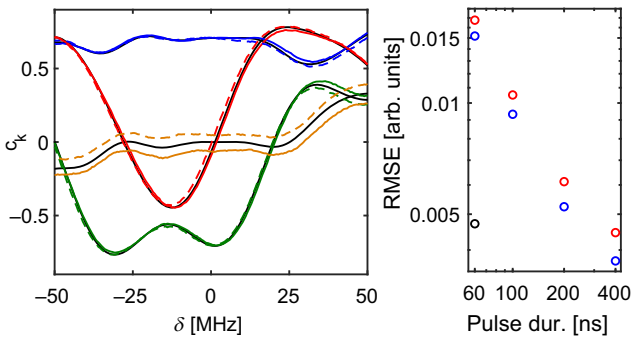


FIG. 2. Left: The experimentally obtained coefficients c_0 (blue), c_x (red), c_y (green), and c_z (yellow) for the pulse (solid) and its conjugate form (dashed) along with the theoretically expected value (solid black). Right: the RMSE between the theory and the experiment as a function of the pulse duration for the direct (red) and conjugated (blue) pulses, as well as the pulse optimized to suppress the interactions with the $|2\rangle$ state (black).

Additionally, if one plays the *conjugated* pulse $\Omega(t) = A(t)e^{-i\varphi(t)}$, we see that the c_z offset changes sign (cf. left panel of Fig. 2), indicating an alternative mechanism. Conjugating the pulse has the same effect as flipping its frequency, or rather detuning with respect to the qubit frequency. Therefore, it also comes with an additional change of sign of δ, which was taken into account for reasons of presentation in this figure.

The $\varphi(t) \nrightarrow -\varphi(t)$ and $\delta \nrightarrow -\delta$ symmetry breaking could have several causes. *The first* is the fact that the signal generation scheme used in this experiment relies on upconversion by mixing the low frequency signal with a local oscillator (LO) close to the qubit frequency. For systems with negative anharmonicity, the LO is typically placed above the qubit transition frequency (here, we have $f_{\mathrm{LO}} = 3.9$ GHz) to avoid driving higher order transitions. However, we find that the LO leakage and image are suppressed by at least 40–50 dBc typically and their contributions are not clearly present in the scope traces. Together with repeating the measurements with different f_{LO}, this makes the upconversion scheme an unlikely culprit for the observed effect. *Another* possible explanation is that the qubit approximation is not valid, and the symmetry is naturally broken by the presence of higher excited states. To this end, we perform the same type of measurements as shown in Fig. 2 as a function of pulse duration and measure the root-mean-square error (RMSE) between the theory and experiment, for both the non-conjugated and conjugated pulses. Figure 2 shows that the RMSE follows a power law, similar to what is observed when the finite anharmonicity becomes the limiting factor at short pulse durations [3]. The fact that the loss for conjugated pulses is systematically lower than for non-conjugated ones indicates that the two-level approximation is not adequate even for long pulse durations, with low RMSE. It is worthy to note that if the c_z offset was generated by a detuned qubit frequency, or LO leakage, its effect would be exaggerated by increasing the pulse duration.

B. Driving a qutrit with a neural network

With the results obtained in Sec. III A, we turn to optimizing the pulse in the presence of the second excited state, for our specific device: the Hamiltonian of Eq. (1) is

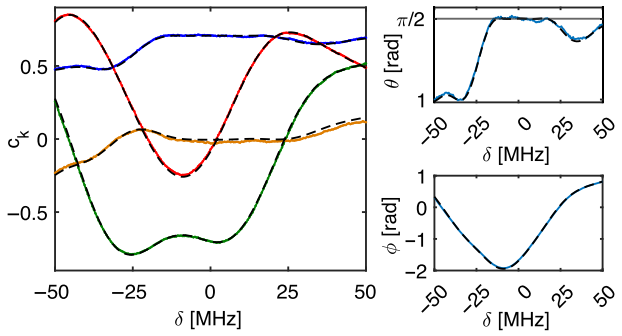


FIG. 3. Left: The experimentally obtained coefficients c_0 (blue), c_x (red), c_y (green), and c_z (yellow) for the pulse optimized to suppress the interactions with the $|2\rangle$ state. Right: the azimuthal and polar rotation angles, θ and ϕ, top and bottom, respectively. The dashed black traces on all panels correspond to the theoretically expected behavior.

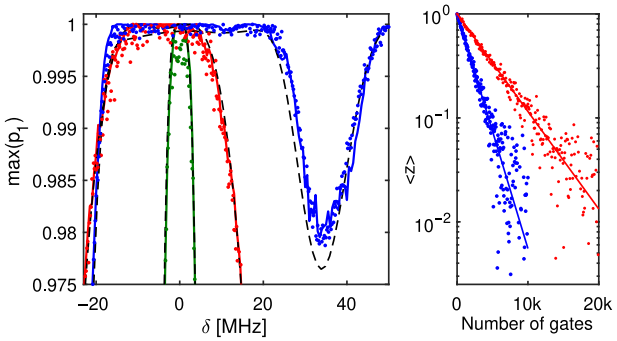


FIG. 4. Left: The maximum $|0\rangle \rightarrow |1\rangle$ transfer efficiency as a function of detuning δ for the robust (blue), DRAG (red) and rectangular (green) pulses. The dots correspond to directly measured values, the solid lines to the QPT inferred ones and the dashed black to the theoretical expectation. Right: The z component of the density as a function of the number of applied robust (blue) and DRAG (red) pulses.

expanded with $\Delta|2\rangle\langle 2|$ and $\lambda\Omega(t)|2\rangle\langle 1| + \text{H.c.}$ terms. As the numerical integration is performed over $t \in [-\pi, \pi]$, here $\Delta = 2\pi\alpha_s T_p$ is set by the pulse duration T_p and the anharmonicity α_s, while λ is the ratio of the $\langle 2|\hat{q}|1\rangle$ and $\langle 1|\hat{q}|0\rangle$ matrix elements of the electrical dipole moments. By fitting the transition frequencies of our system, we extract the Josephson and charging energies and evaluate lambda to be $\lambda \approx 1.37$. The SU(3) group does not permit a linear and real parametrization of the propagator U [48], forcing the parameter space to be $\mathbb{C}^{3\times 3} \cong \mathbb{R}^{18}$, therefore significantly increasing the computational complexity: using the same hardware one iteration takes ≈4.6 s, a fivefold increase compared to the qubit case. Thus, we opt to *refine* the pulse, i.e., resume training the neural network from the previous state, thereby reducing the resources needed for the optimization. The resulting pulse is also shown in Fig. 1: seemingly minor differences, such as a smoother envelope $A(t)$, lead to a significantly different behavior. This can be understood by performing QPT as a function of detuning δ, as shown in Fig. 3 (left panel). Indeed, we find that $c_0 \approx 1/\sqrt{2}$ and $c_z \approx 0$ for a wide range of detuning frequencies, and thus implicitly $c_x^2 + c_y^2 = 1/2$. When these conditions are met, the resulting rotation angle θ is $\theta \approx \pi/2$, while the plane of rotation ϕ is frequency dependent, see Fig. 3 (top and bottom right panel, respectively). The latter is not an issue for the parity measurement experiment, as the sequence is not sensitive to an overall phase. To quantify the agreement with the theoretically expected behavior, the RMSE loss is calculated and shown in Fig. 2 by the black circle: the agreement is significantly improved and is on the level of the error observed for the longest pulse durations.

C. Pulse sequence for a parity measurement

In the context of the parity measurement experiment, the figure of merit can be constructed as the efficiency with which two consequentially applied pulses, with an appropriately chosen delay time in between them, transfer the population of the qubit from the ground $|0\rangle$ state to the excited one $|1\rangle$. With this benchmark in mind, we compare our designer pulse to a rectangular pulse of the same duration $T_p = 60$ ns and a $T_p = 20$ ns DRAG pulse: we initialize the qubit in the ground state, apply the sequence, and measure the probability to find it in the excited one. The result of this procedure, as a function of the detuning δ is shown in Fig. 4 (left panel). Here, we show the theoretically expected trace, along with the experimentally determined ones based on both a direct measurement and a QPT reconstruction, for all three pulses. For more details on this procedure, see Appendix E. As expected, due to its comparatively small bandwidth, the rectangular pulse performs the worst. The DRAG pulse fares much better; however, its efficiency falls off rather quickly for positive detunings $\delta > 0$, while it stays high for negative ones: $\max(p_1) > 0.99$ for $-13.5 < \delta < 5$ MHz. This behavior is the result of the *inherent* spectral asymmetry present in the DRAG pulse: the nonzero out-of-phase component $Q(t) \propto \sin(2\pi t/T)$ introduces some phase modulation, thus skewing the spectrum of the pulse. This constitutes an effective detuning, which counteracts the ac Stark shift of the second excited state $|2\rangle$ and cannot be eliminated without degrading the performance of the pulse.

In contrast, the efficiency of the robust scheme remains high (above ≈0.98) for detunings as large as $\delta = 50$ MHz, while matching the performance of the DRAG pulse for negative detunings: $\max(p_1) > 0.99$ for $-16.5 < \delta < 22$ MHz. In Appendix E 2, we demonstrate that our NN approach outperforms GRAPE pulses optimized for the same task. Additionally, to achieve a comparable level of detuning robustness using composite pulses requires more than 5 pulses, significantly lengthening the operation—c.f. Appendix E 3. It is worth noting that the sequence realized with DRAG and rectangular pulses

does not strictly implement a parity measurement even when the transfer efficiency is high: as the resulting unitary has a nonzero σ_z component (cf. Appendix E), the corresponding rotation does not lie in a vertical plane. As a result, the qubit must acquire an additional phase to maximize the transfer efficiency. With a single delay time τ, this condition cannot be met simultaneously for all detunings δ, associated with the number of photons as $\delta = 2\chi_{\mathrm{disp}} n_{\mathrm{ph}}$. In contrast, the robust pulse preserves the desired property as $c_z \approx 0$, which is detailed in Appendix E as well.

Lastly, we set out to measure the fidelity of the robust $\frac{\pi}{2}$ pulse and compare it to that of the DRAG. To this end, we construct a pseudoidentity operation out of 4 pulses $(\frac{\pi}{2})^{4n} = (-1)^n \hat{I}$, initialize the system in the ground state and measure the z component of the density matrix as a function of the number of repetitions n. We perform this procedure only on resonance ($\delta = 0$)—see Fig. 4 (right panel). By fitting the results with an exponential decay, we are able to extract the fidelity per pulse: $F_{\mathrm{rob}} = 0.999481 \pm 0.000017$ and $F_{\mathrm{DRAG}} = 0.999784 \pm 0.000008$. While higher single-qubit fidelities have been reported for isolated transmons (e.g., [49,50]), our results are close to the record-breaking ones for a transmon embedded in a circuit (e.g., [51]).

The estimation for the single-qubit incoherent errors corresponding to a gate length T_p with relaxation T_1 and pure Markovian dephasing T_ϕ is given by [52,53]: $F^{\max} = 1 - T_p/3(1/T_1 + 1/T_\phi)$, where T_ϕ is estimated as $1/T_\phi = 1/T_2 - 1/2T_1$. For our experimental parameters, this yields $F^{\max}_{60\ \mathrm{ns}} = 0.999611$ and $F^{\max}_{20\ \mathrm{ns}} = 0.999870$, placing the results within $\approx 10^{-4}$ from the theoretical limit. The total incoherent error constitutes approximately 75% of the total robust pulse error and 60% of the total DRAG pulse error. These percentages are obtained by comparing the experimentally measured gate infidelity to the theoretical incoherent error limit $1 - F^{\max}$ computed from the independently measured T_1 and T_2 times. The remaining 25% and 40% we assign to the gate implementation and leakage errors, demonstrating a better implementation fidelity of the robust pulse compared to the DRAG. As a result, we can conclude that the performance of the pulses is limited by finite relaxation time of the transmon and not by the implementation.

IV. DISCUSSION AND CONCLUSIONS

We presented a neural ordinary differential equation (neural ODE) framework for quantum control, enabling continuous-time, basis-free synthesis of high-fidelity control pulses. By representing drive fields as outputs of trainable neural networks, and optimizing them through automatic differentiation, our approach provides universal, control-hardware-agnostic control strategies without relying on predefined bases or discretized control points.

We validated this method by generating, and benchmarking experimentally, detuning-robust $\pi/2$ pulses for use with superconducting qubits, and achieved superior detuning resilience compared to standard rectangular and DRAG pulses. Furthermore, we demonstrate that our pulses outperform both GRAPE-optimized and composite sequences. The achieved robustness is particularly relevant for bosonic QED experiments, where the effective qubit detuning arises from photon-number-dependent ac Stark shifts due to dispersive coupling to a resonator. In such a setting, including Wigner tomography and parity-based measurements, the qubit cannot be maintained on resonance across different photon occupations by a single frequency calibration. The demonstrated detuning-robust $\pi/2$ pulses therefore directly address a central experimental limitation, enabling high-fidelity operations even in the presence of large and state-dependent frequency shifts.

This framework was also demonstrated to generate robust arbitrary rotation gates and can be generalized to arbitrary multiqubit gates. It outperforms conventional techniques in applications such as error-mitigation for quantum computing, bosonic state reconstruction, Wigner tomography, magnetometry, and parity measurements for axion detection. This work demonstrates that neural-ODE-based continuous-time control parameterizations can serve as an effective enabling tool for the experimental realization of detuning-robust quantum gates in superconducting circuits.

ACKNOWLEDGMENTS

We thank Miroslav Bogdanović for assistance with the machine-learning portion of the work, and Jonas Bylander for comments on the manuscript. This project has received funding from the European Union through Horizon Europe and the Quantum Flagship, with OpenSuperQPlus100, Grant Agreement No. 101113946, under HORIZON-CL4-2022-QUANTUM-01-SGA. We also acknowledge support from Research Council of Finland Proof-of-Concept program (Project No. 359397) and from the Centre of Excellence program (Project No. 352925). The Chalmers team acknowledges financial support by the Knut and Alice Wallenberg Foundation through the Wallenberg Center for Quantum Technology (WACQT). The fabrication of the quantum processor was performed at Myfab Chalmers. The Aalto team acknowledges the provision of facilities and technical support by the Aalto University at the national research infrastructure OtaNano—Low Temperature Laboratory and Micronova.

M. K. carried out the modeling, numerical optimization and together with I. M., Y.-H. C., and O. S. performed the experiments. The Chalmers team designed and fabricated the transmon device. All members of the Aalto team contributed to the writing of the manuscript. G. S. P. supervised the project.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: DETUNING ROBUST GATESET

This appendix presents a numerical study of a robust single-qubit gateset, for a range of rotation angles $\theta \in [0, \pi]$, optimized independently from the $\pi/2$ gate presented in the main text. The optimization is carried out as described in Sec. II: Minimizing the loss function defined as $L = (c_0 - \cos(\theta/2))^2 + (c_x - \sin(\theta/2))^2 + c_y^2 + c_z^2$, over $\theta \in [0, \pi]$ and $\delta \in [-0.4\Omega_{2\pi}, 0.4\Omega_{2\pi}]$ ($\Omega_{2\pi} = 2\pi/T_p$), results in a detuning robust gateset, generated by σ_x. Amplitude robustness is not considered ($\alpha = 0$). Employing this optimization strategy yields a complete θ gate set within a single training execution, at no extra computational expense. The whole procedure is carried out within the qubit approximation, i.e., the presence of the second excited state is neglected.

Figure 5 shows the c_k coefficients of the resulting unitary operation U as a function of detuning δ for several θ values, chosen such that the transition probability $|\langle 1|U|0\rangle|^2 = \sin(\theta/2)^2$ is linearly spaced from 0 to 1 (indicated by the dashed horizontal lines). It is important to note that this is only done for sake of presentation, the training is done on a dense set of θ values spanning the entire $\theta \in [0, \pi]$ interval. We observe that the target values for c_k are indeed obtained in the entire δ training interval, with an average (in)fidelity of $1 - F = 1 - \mathrm{tr}(U_\theta^{-1} U)/2 \approx 10^{-5}$, as shown in Fig. 6. The same figure also shows the time-dependent pulse amplitudes $A(t)$ and phases $\varphi(t)$.

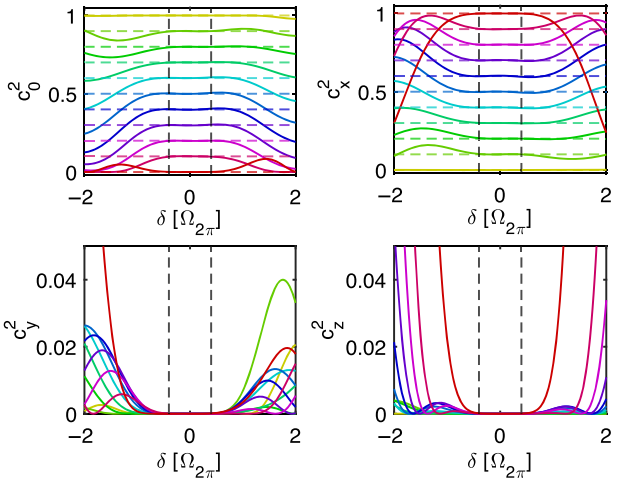


FIG. 5. The squared coefficients c_0^2, c_x^2, c_y^2, and c_z^2 as a function of δ for a detuning robust gateset. The rotation angles θ satisfy $\sin(\theta_k/2)^2 = k/10$ from $k = 0$ (yellow) to $k = 10$ (red), with their target c_0^2 and c_x^2 indicated by the dashed horizontal lines. The dashed vertical lines indicate the training region.

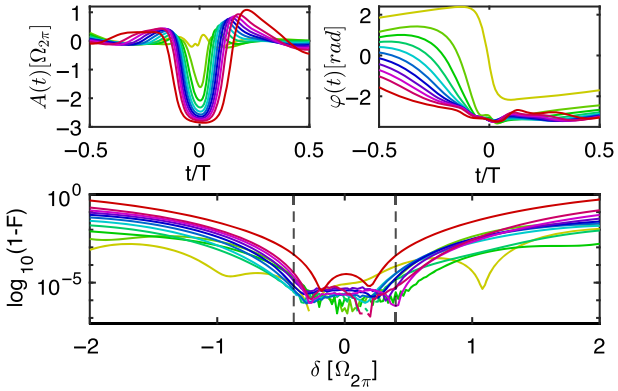


FIG. 6. The time-dependent amplitude $A(t)$ (top left) and phase $\varphi(t)$ (top right) required to realize the unitary operations presented in Fig. 5. The bottom panel shows the fidelity F of the operation as a function of detuning δ (the dashed vertical lines indicate the training region). All traces are color-matched to the ones in Fig. 5.

Although in the training the rotation angle θ is discretely sampled, our large training volume (on the order of 10^4 iterations and a batch size on the order of 1000) yields a highly dense sampling distribution with an effective angular spacing on the order of 100 μ°. Translating this theoretical precision to experimental Rabi pulses would require an amplitude stability of approximately 1 ppm. This corresponds to a 20-bit DAC resolution, exceeding the 16-bit limit typical of modern arbitrary waveform generators. Additionally, given the relatively low computational cost, it is feasible to increase the theoretical resolution of our gateset by more than an order of magnitude, without the need for high-performance computing resources, far surpassing current instrumentation limits.

APPENDIX B: EXPERIMENTAL SETUP

The transmon qubit used in the experiment is a subset of a two-qubit device similar to the quantum processor studied in Ref. [54], with two fixed-frequency transmons capacitively coupled via a frequency-tunable coupler. The chip is packaged, protected from the spurious magnetic noise by a μ-metal shield and thermally anchored to the mixing chamber of the refrigerator. During the experiments, only one qubit (Q_1) is used, and the coupler is detuned to its sweet spot at the maximum frequency to minimize its impact on the qubit.

The experimental setup used is presented in Fig. 7. The pulses are generated at room temperature by using an ensemble OPX+/Octave from Quantum Machines, capable of generating microwave pulses within a 350 MHz bandwidth. To achieve low enough temperatures, we employ a BlueFors XLD400 dilution fridge with base temperature of ≈10 mK. Single-shot measurements are enabled by the traveling wave parametric amplifier (TWPA) from VTT Technical Research Center of Finland, while at 4 K the

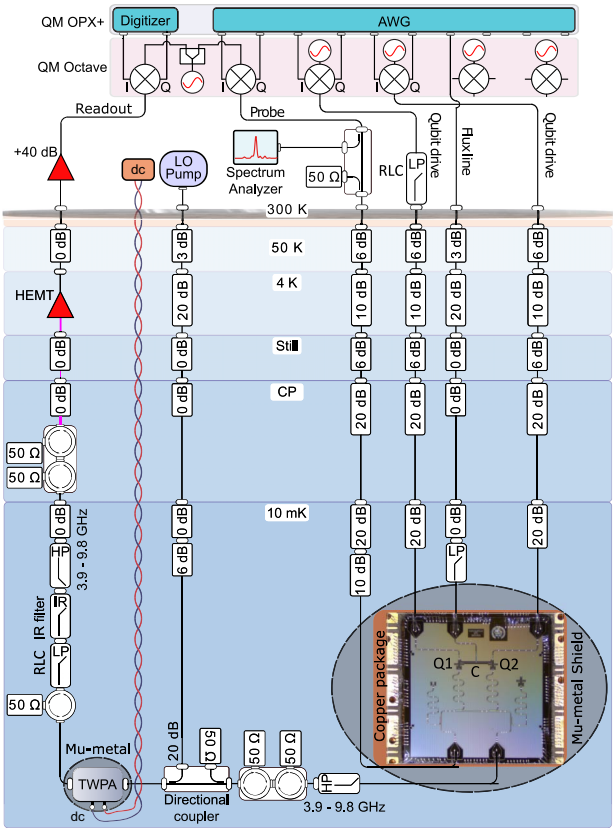


FIG. 7. A diagram of the experimental setup. Microscope image of the quantum processor with two transmon qubits. Only the left qubit (Q_1) is used for the experiment. The sample is mounted inside a μ-metal shield to protect it from stray magnetic fields.

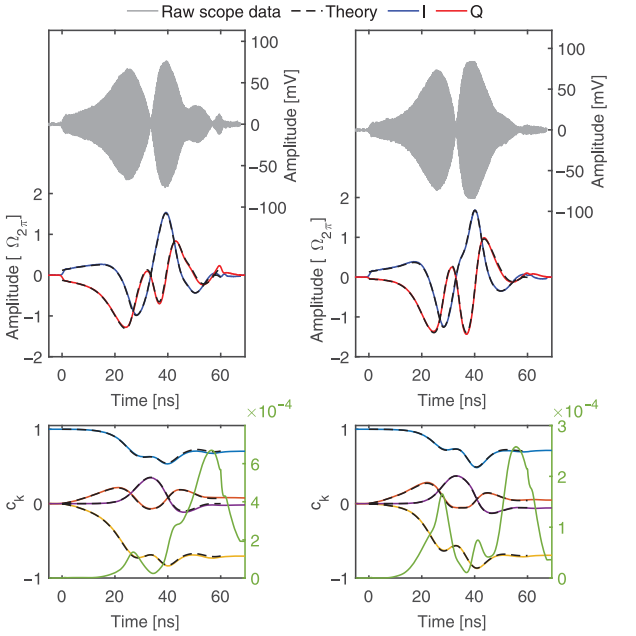


FIG. 8. Top: The real and imaginary part of the pulse $\Omega(t)e^{i\varphi(t)}$ (black dashed), along with the measured pulse (gray) and its demodulated real (blue) and imaginary (red) components. Bottom: The c_0 (blue), c_x (red), c_y (yellow), and c_z (purple) coefficients as a function of time for the scope data, along with the theory (black dashed). The right-hand scale, along with the solid green trace, shows the infidelity (1-F) of the experimental trajectory with respect to the theoretical one. The left and right columns correspond to the qubit and qutrit optimized pulses, respectively.

readout signal is further amplified by a high-electron-mobility transistor (HEMT) fabricated by Low Noise Factory. After further room-temperature amplification (by Narda Miteq LNA-40-04000800-07-10P), the readout signal is fed back into the OPX+/Octave ensemble, where it is downconverted and digitized.

The pulses used for state preparation and quantum process tomography (cf. Appendix D) use a cosine envelope with a 40 ns duration combined with the DRAG technique and calibrated according to "Google" method described in Ref. [55]. The average gate fidelity characterized with Clifford RB [47] exceeds 99.96%.

The single-shot readout uses a 2 μs pulse, with an average two-state discrimination fidelity of $F_{\text{Assign}(|0\rangle,|1\rangle)} = 95.65\%$, and $F_{\text{Assign}(|0\rangle,|1\rangle,|2\rangle)} = 94.13\%$ for three states.

APPENDIX C: IMPLEMENTATION VERIFICATION

The pulses generated by the procedure described in the main text can then be digitized using a sufficiently fast oscilloscope directly (Keysight MSOS804A was used here). With the captured trace (cf. Fig. 8), an *analytic*, complex valued, signal representation can be obtained by applying a Hilbert transform [56]. The analytic form can be easily demodulated by multiplication by $e^{-i\omega_{01}t}$, with $\omega_{01} = 2\pi f_{01}$ being the qubit/signal center frequency. Afterward, it lends itself to comparison to the theory directly, through the real and the imaginary part.

Based on these reconstructed drive amplitudes, the corresponding trajectory can be calculated, which are also shown in Fig. 8. The fidelity of the reconstructed trajectory, defined as $F = \text{tr}\{U_{\text{exp}}(t)U^{\dagger}_{\text{theory}}(t)\}/2$, remains high ($F > 0.999$) throughout the evolution (also in Fig. 8).

APPENDIX D: QUANTUM PROCESS TOMOGRAPHY

Quantum process tomography is performed in the manner described in Ref. [2]: 4 input states are prepared by initializing the system in the ground state and then applying an identity (no-op) gate, a $(-\frac{\pi}{2})_X$, $(-\frac{\pi}{2})_Y$, or a $(\pi)_X$ gate, resulting in the following initial states $|0\rangle\langle 0|$, $|x\rangle\langle x|$, $|y\rangle\langle y|$, and $|1\rangle\langle 1|$, respectively. Then, the operation-under-test is applied, followed up by state tomography where the z component is read out directly and for the y and x ones, the readout is preceded by applying $(\frac{\pi}{2})_X$, $(\frac{\pi}{2})_Y$.

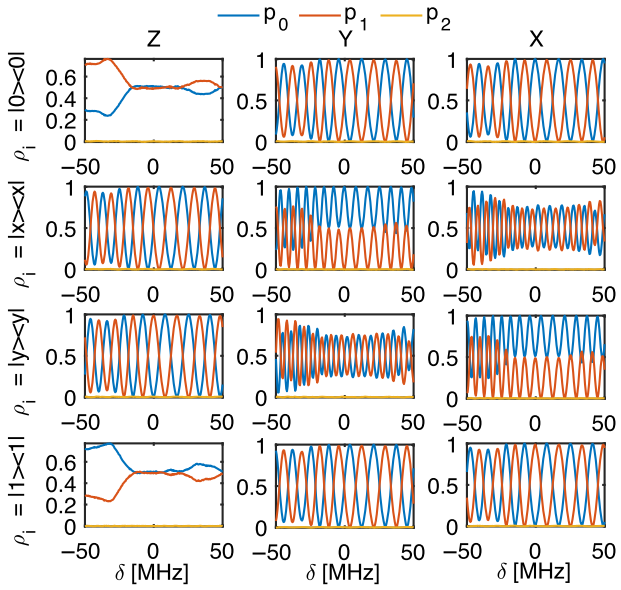


FIG. 9. The raw results of the frequency-dependent process tomography of the robust $(\frac{\pi}{2})$ pulse. Different rows correspond to different initial states, while the columns correspond to reading out the x, y, and z components of the density matrix. See the text for more details.

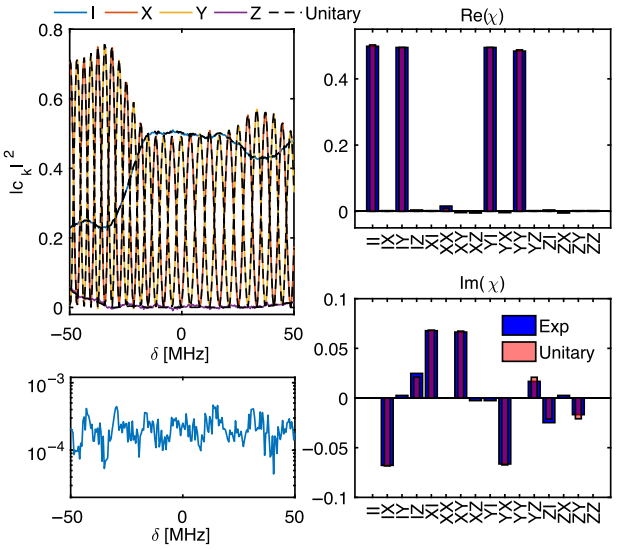


FIG. 10. Left: The diagonal elements of the experimental process matrix (solid, colored), along with that of the unitary fit (black, dashed), as a function of detuning δ, along with their Frobenius distance (bottom). Right: The real (top) and imaginary (bottom) parts of the experimental and unitary process matrices at $\delta = 0$.

This is then performed as a function of the test-pulse frequency; the results for the robust, neural network based, $(\frac{\pi}{2})$ are shown in Fig. 9.

From there, we see that the population of the $|2\rangle$ state is negligible, allowing us to carry out the process tomography only in the qubit subspace. The detuned pulses are generated by briefly changing the frequency of the IF oscillator in the Quantum Machines OPX+ controller, and the phase difference accumulated during this frequency excursion leads to the rapidly oscillating behavior observed in most panels of Fig. 9, which can later be accounted for.

The data can be processed to extract the process matrix [2], resulting in a complex $4 \times 4 \chi_{\text{exp}}$ matrix, as shown in Fig. 10. This experimentally obtained process matrix contains contributions not only from the unitary evolution, but also from relaxation, decoherence, and SPAM errors [46]. For estimating the underlying unitary operation, we fit the experimental process matrix with a unitary one χ_{fit}, by minimizing their Frobenius distance. To gauge the errors introduced by this process, we note the following: for a strictly unitary process, $U = c_0\sigma_0 - i\sum_{k\in x,y,z} c_k\sigma_k$, the diagonal elements of χ are just $\chi_{k,k} = |c_k|^2$. In Fig. 10, we compare the diagonal elements of χ_{exp} and that of χ_{fit}, along with their Frobenius distance. As their agreement is quite good, and the distance small, we assume that this leads to a reliable estimate of the implemented unitary evolution, which might not be true for higher SPAM or decoherence errors.

The c_k coefficients shown in Fig. 10 are identical to the ones presented in Fig. 3 in the main text up to an additional rotation in the X/Y subspace. As mentioned before, this rotation angle is $\beta = 2\pi\delta(T_{\text{pulse}} + T_{\text{dead}})$, where $T_{\text{dead}} \approx 10$ ns is the time needed for the IF oscillator frequency change. Applying the inverse rotation in the X/Y subspace $c'_x + ic'_y = (c_x + ic_y)e^{-i\beta}$ yields the results presented in Fig. 3.

APPENDIX E: COMPARISON WITH OTHER PULSES

Here, we first provide further details about the comparison with the DRAG and rectangular drives. In Appendices E 2 and E 3, we discuss the alternative approach of using GRAPE and composite pulses, respectively.

1. DRAG and rectangular drive

As stated in the main text, the same QPT procedure was carried out for the $T_p = 60$ ns rectangular and $T_p = 20$ ns pulses, the results of which are shown in Fig. 11. Figure 11 also shows the discrepancy of the implemented operation with respect to a $(\frac{\pi}{2})$ rotation in an *arbitrary*, but vertical plane. It is quantified as $F_\chi = \text{tr}(\tilde{\chi}_{\text{exp}}\chi_{\pi/2})$, where $\tilde{\chi}_{\text{exp}}$ is the *purified* process matrix $\tilde{\chi}_{\text{exp}} = \chi_{\text{exp}}/\sqrt{\text{tr}(\chi^2_{\text{exp}})}$ and $\chi_{\pi/2}$ is the process matrix corresponding to a unitary with the following expansion $c_0 = 1/\sqrt{2}$, $c_x = \cos(\Theta)/\sqrt{2}$,

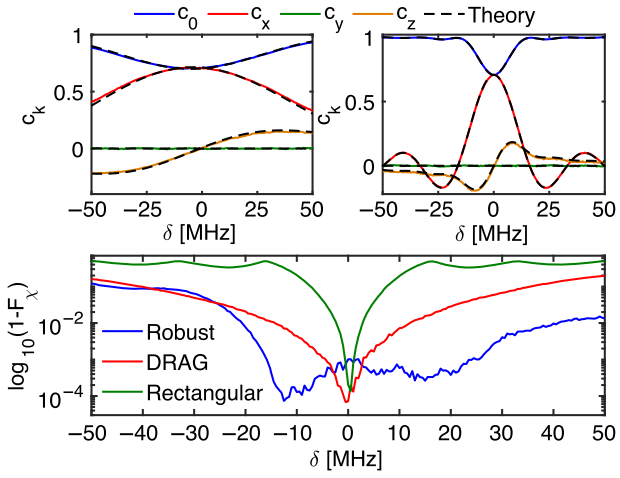


FIG. 11. Top: The c_k decomposition of the DRAG (left) and rectangular (right) pulses as a function of detuning. Bottom: the process matrix fidelity F_χ between the experimentally obtained ones and a vertical $\pi/2$ rotation.

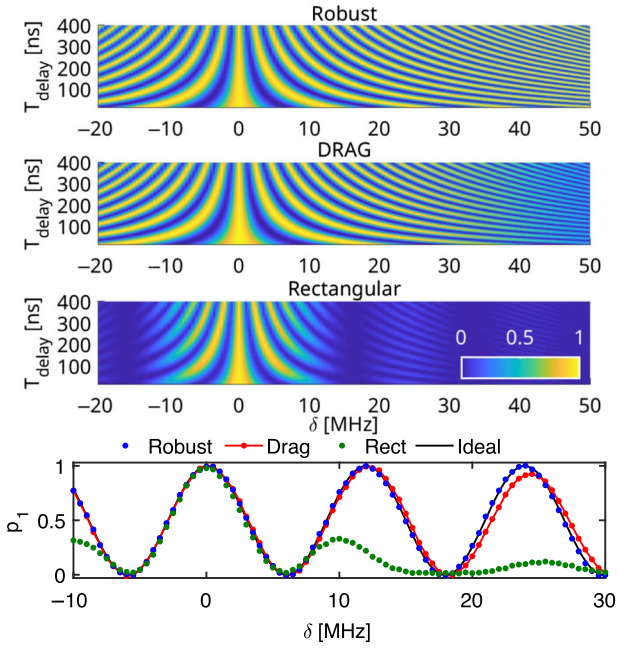


FIG. 12. Top: The Ramsey interference pattern measured for the robust, DRAG and rectangular pulses. Bottom: slices of the former as a function of detuning δ for a fixed T_{delay} (colored symbols), along with the ideal response (solid black).

$c_y = \sin(\Theta)/\sqrt{2}$, and $c_z = 0$, where Θ is chosen such to maximize F_χ. The wide range of detuning frequencies for which $F_\chi \approx 1$ demonstrates yet again the robustness of the neural network based pulse.

Much like in Fig. 4 of the main text, from Fig. 11 we see that the response of the DRAG pulse is shifted to the left. As stated in the main text, this shift is inherent to a DRAG pulse: the phase modulation and frequency offset introduced by the nonzero out-of-phase Q quadrature counteracts the ac Stark effect from the $|2\rangle$ state and is necessary for a high fidelity operation. Introducing an additional detuning δ cannot center the response of the pulse without sacrificing performance or changing the effective value of the DRAG coefficient, making the frequency response skewed towards $\delta < 0$ once again.

Unlike for the robust pulse, where the σ_z component $c_z \approx 0$ (see Fig. 3 of the main text), for here that is not the case. These nonzero c_z components limit the usefulness of DRAG pulses for the parity measurement protocol: Ramsey interference patterns are shown in Fig. 12 for all three pulses. By taking a slice at a constant protocol time $T_{\text{protocol}} = T_{\text{delay}} + T_{\text{pulse}}$ and plotting the excited state population as a function of frequency δ, also shown in Fig. 12, we can see the effect of the z rotation. While the envelope of the oscillation corresponds to the one obtained from measuring $\max(p_1)$ and shown in Fig. 4 of the main text, we see that the positions of the maxima/minima start deviating from an ideal sinusoidal pattern for $\delta \gg 0$. As a result, the parity measurement implemented with DRAG pulses is limited to lower photon numbers compared to the robust pulse.

To understand this further, we can study the following: the idle time T_{delay} is equivalent to a z frame rotation $e^{i\beta\sigma_z/2}$ by the angle $\beta = 2\pi\delta T_{\text{delay}}$. Sandwiching this between two operations, given as $U = c_0\sigma_0 - i\sum_{k\in x,y,z} c_k\sigma_k$, results in the total operation $U_{\text{tot}} = Ue^{i\beta\sigma_z/2}U$. The squared norm of the matrix element $U_{\text{tot},1,2} = \langle 1|U_{\text{tot}}|0\rangle$ is the probability for the transition to occur from the ground to the excited state, which is precisely what we measure experimentally. After some straightforward algebra, it evaluates to $|U_{\text{tot},1,2}|^2 = 4(c_x^2 + c_y^2)(c_0\cos(\beta/2) + c_z\sin(\beta/2))^2 = 4(c_x^2 + c_y^2)(c_z^2 + c_0^2)\cos(\frac{\beta-\beta_0}{2})^2$, where $\beta_0 = 2\arctan(c_z/c_0)$. The transition probability reaches unity only when $(c_x^2 + c_y^2) = (c_z^2 + c_0^2) = \frac{1}{2}$ and $\beta = \beta_0$. However, as β_0 depends nonlinearly on δ through $c_{0,z}(\delta)$, and β is a linear function of δ, the last condition cannot be satisfied for all δ by an appropriate choice of T_{delay}.

Now, we turn to analyzing the performance of a parity measurement experiment using these pulses. As described in Appendix E 1, the transition $|0\rangle \to |1\rangle$ probability in a Ramsey-like experiment can be estimated from the QPT data, or theoretical model, as $\max(p_1) = 4(c_x^2 + c_y^2)(c_z^2 + c_0^2)$, which is obtained for an accumulated phase $\beta = 2\pi k + \beta_0$. This implies that the delay time in the Ramsey sequence cannot be chosen solely based on the detuning; it requires prior knowledge of the $c_{0,z}(\delta)$ dependence. Hence, we go about measuring $\max(p_1)$ by applying a *virtual* z rotation in-between the two drive pulses. A 2D map, which measures the transition probability, as a function of detuning δ and the frame rotation angle β, is shown in Fig. 13. As the β dependence is

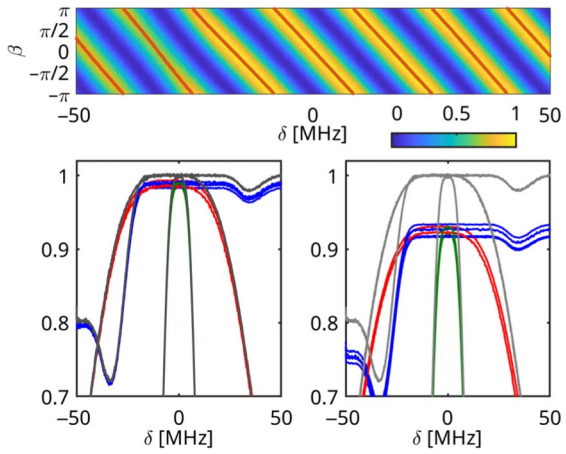


FIG. 13. Top: A 2D map of the excited state population as a function of the detuning δ and relative pulse phase β. The red dots indicate the optimal β values for each δ. Bottom: The measured excited state population as a function of δ, using the optimal β values, for the robust (blue), DRAG (red), and rectangular (green) pulses before, and after correcting for the readout fidelity (gray). The left and right panels correspond to the averaged and single-shot readout schemes, respectively.

sinusoidal, the optimal value can easily be determined by performing a Fourier-like fit.

Using these optimal values, we proceed to measure the excited state population as a function of δ. The state of the transmon is read out by probing a dispersively coupled resonator, either in a single-shot fashion or in the averaged regime. The first is foremost limited by the SNR of the readout and the system relaxation times relative to the readout duration, while the latter is sensitive mostly to the postprocessing procedure used to extract the populations, as described in detail in Ref. [57]. As we need to resolve minute variations in p_1, a lot of measurement shots are required to discern them; we used $N = 5 \times 10^5$ shots. While this results in very low shot noise for the single-shot readout and good SNR for the averaged readout scheme, the long measurement time makes the experiment highly susceptible to errors induced by drifts in the amplification/readout chain. Figure 13 shows several datasets, for each of the studied pulses, where single-shot and averaged readout were performed simultaneously. Both readout schemes suffer from variations and drifts in their efficiency, which amounts to a systematic error present in the readout procedure. However, when one accounts for the variations, all of the datasets collapse to the same trace. This, along with the excellent agreement with the QPT-based estimate, gives validity to the correction procedure.

2. GRAPE pulses

In order to make a comparison to the neural-network approach, in the GRAPE approach the pulse first needs to be discretized. With the sampling rate of our AWG (1 Gs/s) and the chosen pulse duration $T_p = 60$ ns, we parameterize the pulse in terms of $N = 60I_k$ and Q_k quadrature samples each. This defines a piecewise constant set of drives, each $\Delta t = 1$ ns long, corresponding to a propagator $U_k = e^{-iH(I_k,Q_k)\Delta t}$. The total propagator can then be obtained by $U = \prod_k U_k$. The propagators, evaluated over a batch of detuning values δ, can then be used to calculate the loss function L_U in the same way as described in the main text—designed to create a detuning-robust parity-measurement $\pi/2$ pulse. In addition, the loss function should be penalized for discontinuities in the drive fields through $L_{\mathrm{sm}} = \sum_k (I_k - I_{k+1})^2 + (Q_k - Q_{k+1})^2$. The total loss is then the sum of L_U and L_{sm}: $L = L_u + sL_{\mathrm{sm}}$, with a Lagrange multiplier s, which can be tuned to obtain sufficiently smooth pulses.

The training is carried out through gradient descent. A naive approach would evaluate the gradients by varying I_k, Q_k and thus would calculate the propagator (at least) $2N + 1$ times. Instead, here we employ the framework of automatic differentiation, as done with the neural network. For a δ batch of size 500, one such iteration takes $\approx$100 ms in total, with the forward pass taking $t_{\mathrm{fwd}} \approx 50$ ms and calculating the gradient and updating the parameters another $t_{\mathrm{bck}} \approx 50$ ms (using 4 threads on the same CPU as stated in the main text). Therefore, it provides a speedup of a factor of $\approx$60 compared to the naive approach.

The same technique can be used in conjunction with the neural network, instead of the differentiable integration: the fixed-step discretized integration can be used for training, while the variable-step integration can be used for validation (and refining the discretized grid if the two deviate), thus speeding up the training by a factor of $\approx$10. For the neural network designed pulses presented in this section, we adopt this approach.

For both, we sample the initial configuration 50 times and train the best performing one for 50 k iterations within the qubit approximation and another 50 k for the qutrit refinement. Therefore, the total training time on our hardware is expected to be approximately 3 h. In reality, we find it to be closer to 4 h, with slight run to run variation. In addition, we vary the initial random number generator seed and repeat the entire procedure many times. The total dataset we obtained in this way consists of 20 neural-network pulses and 10 GRAPE pulses for $s \in \{0.01, 0.1, 1, 10, 100\}$ each. A selected set of representative pulses is verified experimentally on a different sample with the following qubit parameters: $f_{01} = 4.172$ GHz, $f_{12} = 3.974$ GHz (fitted to obtain $\lambda = 1.379$), and relaxation and decoherence times of $T_1 = 66$ µs, $T_2 = 80$ µs, $T_2^{\mathrm{echo}} = 91$ µs. The original sample (described in Appendix B) could not be utilized anymore due to degraded performance after several thermal cycles.

In order to compare these pulses, we first plot the $|0\rangle \rightarrow |1\rangle$ transition probability p_1 after the application of

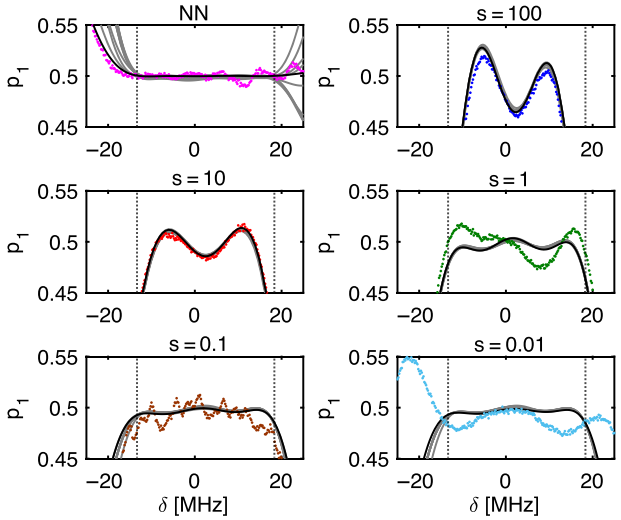


FIG. 14. The $|0\rangle$ to $|1\rangle$ transfer probability p_1 as a function of the detuning δ for the NN and GRAPE pulses: Solid gray lines represent the full dataset. Colored dots indicate the experimental validation for a representative pulse, with the corresponding theoretical prediction denoted by the solid black trace.

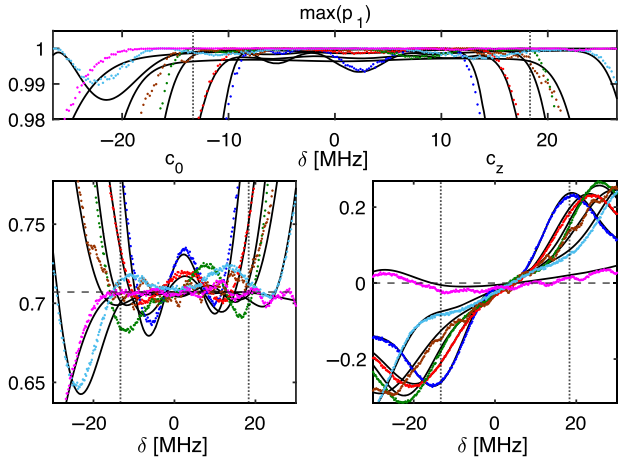


FIG. 15. Top: The QPT predicted $\max(p_1)$ vs δ for the NN pulse (magenta) and the GRAPE ones for $s \in \{0.01, 0.1, 1, 10, 100\}$ (teal, brown, green, red, and blue, respectively). Bottom: the QPT extracted c_0 (left) and c_z (right) coefficients of the same set of pulses. The solid black traces correspond to the theoretically expected values, while the vertical dashed lines indicate the training region, and the horizontal ones show the target $c_{0/z}$ values.

a single pulse as a function of detuning δ, see Fig. 14. Each of the panels shows one dataset: the NN one, and the GRAPE ones grouped by the smoothing factor s. We highlight that while the NN pulses show optimal behavior in the training domain (indicated by the vertical dashed lines), the response of the GRAPE pulses is not flat. Moreover, the individual training runs are remarkably close to one another, but pulses with different smoothing factors s differ substantially from one another. This is indicative of the optimization getting trapped in a local minimum, highly dependent on s.

We perform quantum process tomography and observe that the c_z component of the GRAPE pulses deviates significantly more from zero (the target value) than the NN one, see Fig. 15. The same is true for the c_0 component, but to a lesser extent. Consequently, the GRAPE pulses $\max(p_1)$ also exhibit higher deviations from the ideal value of 1 and are robust in a smaller range of detuning frequencies δ in comparison with the NN one. Additionally, the performance of the pulses is highly sensitive to the smoothing factor s.

Overall, the experimental data align well with the theoretical models. Although both the QPT (Fig. 15) and the direct p_1 (Fig. 14) capture the qualitative behavior of the simulated system, quantitative agreement degrades at lower values of s. This discrepancy is most pronounced for the $s = 1$ and $s = 0.01$ datasets, where measuring multiple pulses yielded varying experimental traces despite virtually indistinguishable theoretical predictions. We tentatively attribute this to signal distortion: the higher bandwidth requirements of the less-smoothed pulses may exceed the finite bandwidth capabilities of the AWG and IQ-mixer upconversion stage. Further investigation is required to confirm this hardware limitation.

In summary, the NN approach offers better performance without the necessity of fine tuning the smoothing Lagrange multiplier. At the same time, with the aforementioned hybrid fixed/variable time-step training approach, the overall training duration is the same as that of the GRAPE approach.

3. Composite pulses

Here, we compare our scheme with a short composite pulse sequence. The rationale for this is the following: the robust pulse is comparatively long $T_p = 60$ ns to the typical durations of DRAG pulses $T_p = 10$–20 ns. Thus, a sequence of 3–6 pulses can be implemented in the same duration.

Famously, the CORPSE [4] sequence even provides an analytical recipe how to implement a 3-pulse long detuning-robust drive. These results were later generalized by, e.g., [58], and longer sequences have also been studied [59]. However, they rely on first- or second-order error cancellation around $\delta = 0$, which does not yield good performance for larger detuning values, which are of relevance here. Thus, instead of relying on analytical models we numerically searched for the optimal sequences, 3–5 pulses long, using a combination of a genetic algorithm (for the sake of parameter-space exploration) and refining the result using a gradient-based optimizer. The loss function to be optimized was made identical to the one used for the NN-based approach. For the sake of visual clarity, in Fig. 16 we report only the c_0 and c_z components

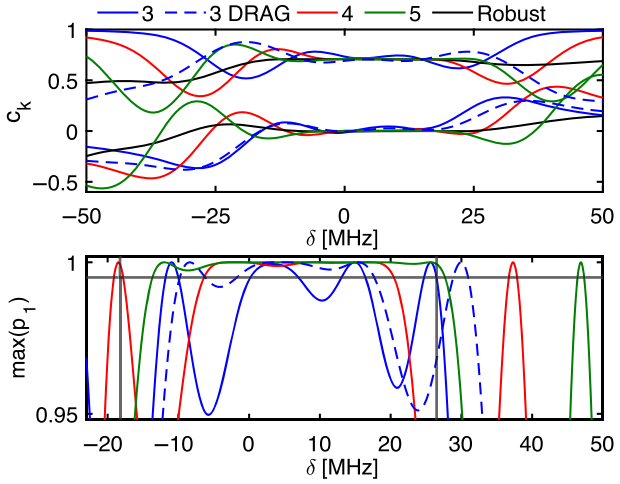


FIG. 16. Top: The c_0 and c_z traces for various composite pulse sequence lengths. The solid-colored lines correspond to the rectangular pulses, the black ones to the robust pulse, and the dashed-blue to a 3-pulse long sequence implemented with DRAG pulses. Bottom: The corresponding transition probabilities in a Ramsey-like experiment; the solid gray lines indicate the region in which the robust pulse has an efficiency greater than 0.995.

of the optimal composite pulses, as well as the corresponding $\max(p_1)$ traces. We find that for both metrics the composite sequences exhibit larger variations than the NN pulse, which is to be expected due to the low number of parameters. On the other hand, these numerically obtained sequences vastly outperform the analytical solutions (not shown).

The claim, made in Ref. [4], that the spectral content of the pulses does not affect the performance of the composite sequence is true near $\delta = 0$: close to resonance, the DRAG and the rectangular pulse look the same to first order, if made the same duration—see Fig. 11. However, for larger $|\delta|$ the argument breaks down. When attempting to implement a 3-pulse long sequence using DRAG pulses, the spectral response changes substantially leading to a different result, even after optimization—see Fig. 16. This finding further complicates implementing a composite pulse sequence based on DRAG: the DRAG coefficient is not rotation-angle independent, and the spectral response is sensitive to its value. Thus, to properly fine-tune the sequence one must perform the optimization in a closed loop, where the sequence and the pulses themselves are tuned simultaneously, making it costly both numerically and experimentally.

[1] 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).

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).

[3] 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).

[4] H. K. Cummins and J. A. Jones, Use of composite rotations to correct systematic errors in NMR quantum computation, New J. Phys. **2**, 6 (2000).

[5] B. T. Torosov and N. V. Vitanov, Arbitrarily accurate variable rotations on the bloch sphere by composite pulse sequences, Phys. Rev. A **99**, 013402 (2019).

[6] B. T. Torosov and N. V. Vitanov, Experimental demonstration of composite pulses on IBM's quantum computer, Phys. Rev. Appl. **18**, 034062 (2022).

[7] M. Harutyunyan, F. Holweck, D. Sugny, and S. Guérin, Digital optimal robust control, Phys. Rev. Lett. **131**, 200801 (2023).

[8] 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. Res. **172**, 296 (2005).

[9] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Phys. Rev. A **84**, 022326 (2011).

[10] A. Pagano, M. M. Müller, T. Calarco, S. Montangero, and P. Rembold, Role of bases in quantum optimal control, Phys. Rev. A **110**, 062608 (2024).

[11] P. Rembold, Quantum optimal control of spin systems and trapped atoms, Ph.D. thesis, Universität zu Köln, 2022.

[12] N. Oshnik, P. Rembold, T. Calarco, S. Montangero, E. Neu, and M. M. Müller, Robust magnetometry with single nitrogen-vacancy centers via two-step optimization, Phys. Rev. A **106**, 013107 (2022).

[13] P. Doria, T. Calarco, and S. Montangero, Optimal control technique for many-body quantum dynamics, Phys. Rev. Lett. **106**, 190501 (2011).

[14] P. M. Poggi, G. De Chiara, S. Campbell, and A. Kiely, Universally robust quantum control, Phys. Rev. Lett. **132**, 193801 (2024).

[15] 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).

[16] Y. Ding, Y. Ban, J. D. Martín-Guerrero, E. Solano, J. Casanova, and X. Chen, Breaking adiabatic quantum control with deep learning, Phys. Rev. A **103**, L040401 (2021).

[17] M.-Z. Ai, Y. Ding, Y. Ban, J. D. Martín-Guerrero, J. Casanova, J.-M. Cui, Y.-F. Huang, X. Chen, C.-F. Li, and G.-C. Guo, Experimentally realizing efficient quantum control with reinforcement learning, Sci. China Phys. Mech. Astron. **65**, 250312 (2022).

[18] L. Giannelli, S. Sgroi, J. Brown, G. S. Paraoanu, M. Paternostro, E. Paladino, and G. Falci, A tutorial on optimal control and reinforcement learning methods for quantum technologies, Phys. Lett. A **434**, 128054 (2022).

[19] J. Brown, S. Sgroi, L. Giannelli, G. S. Paraoanu, E. Paladino, G. Falci, M. Paternostro, and A. Ferraro,

Reinforcement learning-enhanced protocols for coherent population-transfer in three-level quantum systems, New J. Phys. **23**, 093035 (2021).

[20] F. Preti, T. Calarco, and F. Motzoi, Continuous quantum gate sets and pulse-class meta-optimization, PRX Quantum **3**, 040311 (2022).

[21] H. Hutin, P. Bilous, C. Ye, S. Abdollahi, L. Cros, T. Dvir, T. Shah, Y. Cohen, A. Bienfait, F. Marquardt *et al.*, Preparing Schrödinger cat states in a microwave cavity using a neural network, PRX Quantum **6**, 010321 (2025).

[22] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Netw. **2**, 359 (1989).

[23] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, Neural ordinary differential equations, Adv. Neural Inf. Process. Syst. **31 (NeurIPS 2018)** (2018).

[24] F. Sauvage and F. Mintert, Optimal control of families of quantum gates, Phys. Rev. Lett. **129**, 050507 (2022).

[25] U. Güngördü and J. P. Kestner, Robust quantum gates using smooth pulses and physics-informed neural networks, Phys. Rev. Res. **4**, 023155 (2022).

[26] B. Bhattacharyya, F. An, D. Kozbiel, A. J. Goldschmidt, and F. T. Chong, Using optimal control to guide neural-network interpolation of continuously-parameterized gates, in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, New York, 2024), Vol. 1, pp. 1336–1346.

[27] O. Milul, B. Guttel, U. Goldblatt, S. Hazanov, L. M. Joshi, D. Chausovsky, N. Kahn, E. Çiftyürek, F. Lafont, and S. Rosenblum, Superconducting cavity qubit with tens of milliseconds single-photon coherence time, PRX Quantum **4**, 030336 (2023).

[28] T. Krisnanda, P. Song, A. Copetudo, C. Y. Fontaine, T. Paterek, T. C. H. Liew, and Y. Y. Gao, Experimental demonstration of enhanced quantum tomography via quantum reservoir processing, Quantum Sci. Technol. **10**, 035041 (2025).

[29] T. Krisnanda, C. Y. Fontaine, A. Copetudo, P. Song, K. X. Lee, N.-N. Huang, F. Valadares, T. C. H. Liew, and Y. Y. Gao, Demonstrating efficient and robust bosonic state reconstruction via optimized excitation counting, PRX Quantum **6**, 010303 (2025).

[30] B. Vlastakis, G. Kirchmair, Z. Leghtas, S. E. Nigg, L. Frunzio, S. M. Girvin, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Deterministically encoding quantum information using 100-photon Schrödinger cat states, Science **342**, 607 (2013).

[31] P. Bertet, A. Auffeves, P. Maioli, S. Osnaghi, T. Meunier, M. Brune, J. M. Raimond, and S. Haroche, Direct measurement of the Wigner function of a one-photon Fock state in a cavity, Phys. Rev. Lett. **89**, 200402 (2002).

[32] L. G. Lutterbach and L. Davidovich, Method for direct measurement of the Wigner function in cavity QED and ion traps, Phys. Rev. Lett. **78**, 2547 (1997).

[33] L. Sun, A. Petrenko, Z. Leghtas, B. Vlastakis, G. Kirchmair, K. M. Sliwa, A. Narla, M. Hatridge, S. Shankar, J. Blumoff, L. Frunzio, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Tracking photon jumps with repeated quantum non-demolition parity measurements, Nature (London) **511**, 444 (2014).

[34] J. J. McCord, S. Dogra, and G. S. Paraoanu, Coherent interaction-free detection of noise, Phys. Rev. A **110**, 032404 (2024).

[35] J. J. McCord, S. Dogra, and G. S. Paraoanu, Theory of coherent interaction-free detection of pulses, Phys. Rev. Res. **5**, 033012 (2023).

[36] S. Dogra, J. J. McCord, and G. S. Paraoanu, Coherent interaction-free detection of microwave pulses with a superconducting circuit, Nat. Commun. **13**, 7528 (2022).

[37] N. N. Gusarov, M. R. Perelshtein, P. J. Hakonen, and G. S. Paraoanu, Optimized emulation of quantum magnetometry via superconducting qubits, Phys. Rev. A **107**, 052609 (2023).

[38] A. R. Shlyakhov, V. V. Zemlyanov, M. V. Suslov, A. V. Lebedev, G. S. Paraoanu, G. B. Lesovik, and G. Blatter, Quantum metrology with a transmon qutrit, Phys. Rev. A **97**, 022115 (2018).

[39] S. Danilin, A. V. Lebedev, A. Vepsäläinen, G. B. Lesovik, G. Blatter, and G. S. Paraoanu, Quantum-enhanced magnetometry by phase estimation algorithms with a single artificial atom, npj Quantum Inf. **4**, 29 (2018).

[40] A. V. Dixit, S. Chakram, K. He, A. Agrawal, R. K. Naik, D. I. Schuster, and A. Chou, Searching for dark matter with a superconducting qubit, Phys. Rev. Lett. **126**, 141302 (2021).

[41] S. A. Caldwell *et al.*, Parametrically activated entangling gates using transmon qubits, Phys. Rev. Appl. **10**, 034050 (2018).

[42] M. Roth, M. Ganzhorn, N. Moll, S. Filipp, G. Salis, and S. Schmidt, Analysis of a parametrically driven exchange-type gate and a two-photon excitation gate between superconducting qubits, Phys. Rev. A **96**, 062323 (2017).

[43] R. T. Q. Chen, torchdiffeq (2018), https://github.com/rtqichen/torchdiffeq.

[44] K. Diederik and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[45] A. Osman, J. Fernández-Pendás, C. Warren, S. Kosen, M. Scigliuzzo, A. Frisk Kockum, G. Tancredi, A. Fadavi Roudsari, and J. Bylander, Mitigation of frequency collisions in superconducting quantum processors, Phys. Rev. Res. **5**, 043001 (2023).

[46] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[47] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).

[48] J. B. Bronzan, Parametrization of su (3), Phys. Rev. D **38**, 1994 (1988).

[49] Z. Li, P. Liu, P. Zhao, Z. Mi, H. Xu, X. Liang, T. Su, W. Sun, G. Xue, J.-N. Zhang *et al.*, Error per single-qubit gate below 1e-4 in a superconducting qubit, npj Quantum Inf. **9**, 111 (2023).

[50] M. P. Bland, F. Bahrami, J. G. Martinez, P. H. Prestegaard, B. M. Smitham, A. Joshi, E. Hedrick, A. Pakpour-Tabrizi, S. Kumar, A. Jindal *et al.*, 2d transmons with lifetimes

and coherence times exceeding 1 millisecond, arXiv:2503.14798.

[51] 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).

[52] L. Ding, M. Hays, Y. Sung, B. Kannan, J. An, A. Di Paolo, A. H. Karamlou, T. M. Hazard, K. Azar, D. K. Kim, B. M. Niedzielski, A. Melville, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, J. A. Grover, K. Serniak, and W. D. Oliver, High-fidelity, frequency-flexible two-qubit fluxonium gates with a transmon coupler, Phys. Rev. X **13**, 031035 (2023).

[53] T. Abad, J. Fernández-Pendás, A. Frisk Kockum, and G. Johansson, Universal fidelity reduction of quantum operations from weak dissipation, Phys. Rev. Lett. **129**, 150504 (2022).

[54] A. Bengtsson, P. Vikstål, C. Warren, M. Svensson, X. Gu, A. F. Kockum, P. Krantz, C. Križan, D. Shiri, I.-M. Svensson, G. Tancredi, G. Johansson, P. Delsing, G. Ferrini, and J. Bylander, Improved success probability with greater circuit depth for the quantum approximate optimization algorithm, Phys. Rev. Appl. **14**, 034010 (2020).

[55] Z. Chen *et al.*, Measuring and suppressing quantum state leakage in a superconducting qubit, Phys. Rev. Lett. **116**, 020501 (2016).

[56] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals* (Cambridge University Press, Cambridge, England, 2010).

[57] I. Björkman, M. Kuzmanović, and G. S. Paraoanu, Observation of the two-photon Landau-Zener-Stückelberg-Majorana effect, Phys. Rev. Lett. **134**, 060602 (2025).

[58] S. Kukita, H. Kiya, and Y. Kondo, General off-resonance-error-robust symmetric composite pulses with three elementary operations, Phys. Rev. A **106**, 042613 (2022).

[59] S. Kukita, H. Kiya, and Y. Kondo, Short composite quantum gate robust against two common systematic errors, J. Phys. Soc. Jpn. **91**, 104001 (2022).