

Ultrafast Single Qubit Gates through Multi-Photon Transition Removal

Y. Gao,^{1,2,*} A. Galicia,^{1,2} J. D. Da Costa Jesus,^{3,4} Y. Liu,¹ Y. Haddad,^{1,2}
D. A. Volkov,^{1,2} J. R. Guimarães,^{1,2} H. Bhardwaj,^{1,2} M. Jerger,¹ M. Neis,^{1,2}
B. Li,³ F. A. Cárdenas-López,³ F. Motzoi,^{3,4} P. A. Bushev,¹ and R. Barends^{1,2,†}

¹*Institute for Functional Quantum System (PGI-13), Forschungszentrum Jülich, 52425 Jülich, Germany*

²*Department of Physics, RWTH Aachen University, 52074 Aachen, Germany*

³*Institute for Quantum Control (PGI-8), Forschungszentrum Jülich, 52425 Jülich, Germany*

⁴*Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany*

(Dated: December 1, 2025)

One of the main enablers in quantum computing is having qubit control that is precise and fast [1–4]. However, qubits typically have multilevel structures [5–8] making them prone to unwanted transitions from fast gates. This leakage out of the computational subspace is especially detrimental to algorithms as it has been observed to cause long-lived errors, such as in quantum error correction [9–11]. This forces a choice between either achieving fast gates or having low leakage. Previous works focus on suppressing leakage by mitigating the first to second excited state transition, overlooking multi-photon transitions, and achieving faster gates with further reductions in leakage has remained elusive. Here, we demonstrate single qubit gates with a total leakage error consistently below 2.0×10^{-5} , and obtain fidelities above 99.98% for pulse durations down to 6.8 ns for both X and $X/2$ gates. This is achieved by removing direct transitions beyond nearest-neighbor levels using a double recursive implementation of the Derivative Removal by Adiabatic Gate (DRAG) method, which we name the R2D method. Moreover, we find that at such short gate durations and strong driving strengths the main error source is from these higher order transitions. This is all shown in the widely-used superconducting transmon qubit [7, 12], which has a weakly anharmonic level structure and suffers from higher order transitions significantly. We also introduce an approach for amplifying leakage error that can precisely quantify leakage rates below 10^{-6} . The presented approach can be readily applied to other qubit types as well.

Superconducting qubits are engineered quantum devices that have multiple energy levels, distinct from natural two-level systems. Therefore, overcoming undesired transitions has been a constant challenge in the quantum control of such devices [13–17]. The first works to mitigate this leakage used quadrature modulation to direct the quantum evolution, using the Derivative Removal by Adiabatic Gate (DRAG) method [14], which has since become standard practice. Later works have explored methods ranging from frequency domain optimization to

piecewise constant pulse engineering [18–21]. What has seemingly gone unrecognized, is that when pushing for performance with ever faster gates, the qubit enters the strongly driven regime where the driving strength approaches the anharmonicity, altering gate physics.

Higher order transitions beyond nearest-neighbor energy levels are normally either considered forbidden or can be safely neglected for weak driving. Yet, further leakage reduction is only possible by addressing these transitions, which appear during the evolution and can be revealed through the Schrieffer-Wolff (S-W) transform [22]. We start by describing the system in the rotating frame at the microwave driven frequency ω_d . The Hamiltonian of a capacitively driven qubit can be written as [14],

$$H/\hbar = \sum_j \delta_j |j\rangle\langle j| + \lambda_{j+1} \left(\frac{\Omega_R}{2} \sigma_{j,j+1}^x + \frac{\Omega_I}{2} \sigma_{j,j+1}^y \right), \quad (1)$$

where $\delta_j = \omega_j - j\omega_d$ is the detuning of the j th energy level. $\Omega = \Omega_R + i\Omega_I$ is the control pulse, $\sigma_{j,k}^x = |j\rangle\langle k| + |k\rangle\langle j|$ and $\sigma_{j,k}^y = i|k\rangle\langle j| - i|j\rangle\langle k|$ are the Pauli operators, and λ_j is the relative strength of different transitions, with $\lambda_1 = 1$. When the system is driven at resonance: $\omega_d = \omega_q = \omega_1$, and $\delta_2 = \omega_2 - 2\omega_1 = \Delta$ is the anharmonicity. In this qubit frame, the leakage from $|1\rangle$ to $|2\rangle$ can be straightforwardly identified through the $\sigma_{1,2}^x$ term. However, during the qubit driving, the qubit eigenstates are modified by the driving field, which leads to a superposition of higher qubit states, as well as shifts in eigenfrequencies. Therefore, to fully minimize the leakage out of the computational subspace with faster and stronger drives, we have to consider higher leakage levels and look at the dressed states of the coupled system formed by the qubit and the driving field.

We now use the S-W transform on the system Hamiltonian. Using $U_{S-W} = \exp(-i\Omega_R \sum_{j=0}^2 \lambda_{j+1} \sigma_{j,j+1}^y / 2\Delta)$, and keeping computationally relevant terms to the second order in Ω_R/Δ we get [23]

$$H_{S-W}/\hbar \approx \frac{\Omega_R}{2} \sigma_{0,1}^x + \frac{\Omega_I + \frac{\dot{\Omega}_R}{\Delta}}{2} (\sigma_{0,1}^y + \lambda_2 \sigma_{1,2}^y) + \left(\delta_1 + \frac{4 - \lambda_2^2}{4\Delta} \Omega_R^2 \right) |1\rangle\langle 1| + \frac{\lambda_2}{8\Delta} \Omega_R^2 \sigma_{0,2}^x + \frac{\lambda_2 \lambda_3}{8\Delta} \Omega_R^2 \sigma_{1,3}^x, \quad (2)$$

where the $\sigma_{0,2}^x$ and $\sigma_{1,3}^x$ terms in Eq. (2) reveal the direct transitions between higher levels. See Fig. 1(a) for a schematic representation. Interestingly, those transitions violate selection rules in the qubit frame. The two-photon nature of the transitions is further underlined by the squared term, Ω_R^2 . Moreover, the quadratic dependence of the amplitude emphasizes its importance when the pulse duration is short.

With higher order transitions identified, we propose a pulse shape that suppresses the frequency spectrum of Ω_R^2 at the relevant two-photon transition frequencies. Given a control envelope ansatz Ω_0 , we recursively apply the DRAG technique to remove the multi-photon transitions. To keep Ω_R^2 real and avoid phase errors, we use the second derivative to remove the frequency components. A detailed proof and modeling are given in Ref. [23]. We also include parameters α_{02} and α_{13} which allow for fine-tuning the spectral notches during in-situ optimization,

$$\Omega_R^2 = \left(1 + \frac{\alpha_{02}}{\Delta^2} \frac{d^2}{dt^2} \right) \left(1 + \frac{\alpha_{13}}{9\Delta^2} \frac{d^2}{dt^2} \right) \Omega_0^2. \quad (3)$$

Subsequently, we apply the standard single-photon DRAG correction and include α_{12} ,

$$\Omega = \left(1 - i \frac{\alpha_{12}}{\Delta} \frac{d}{dt} \right) \Omega_R. \quad (4)$$

Similar to Ref. [24], we include a constant detuning to correct the phase error. To ensure that the derivatives exist and the pulse spectrum is finite, the envelope ansatz Ω_0 must be smooth to at least 3rd-order and $\Omega(0) = \Omega(t_p) = 0$, where t_p is the pulse length. In this work, we use $\Omega_0 = A \sin^4(\pi t/t_p)$ and we add 1 ns padding after the pulse to account for possible pulse distortions in the system [25]. Using the Fourier transform, the pulse correction results in,

$$[\Omega_R^2](\omega) = \left(1 - \frac{\alpha_{02}\omega^2}{\Delta^2} \right) \left(1 - \frac{\alpha_{13}\omega^2}{9\Delta^2} \right) [\Omega_0^2](\omega), \quad (5)$$

$$\Omega(\omega) = \left(1 - \frac{\alpha_{12}\omega}{\Delta} \right) \Omega_R(\omega). \quad (6)$$

The effect of the R2D pulse shaping can be intuitively understood in the frequency domain of both Ω and Ω_R^2 [Fig. 1(b)-(c)], where we null the signal spectrum at leakage transition frequencies.

We experimentally validate the R2D pulse shape using an isolated transmon qubit (see Supplement). The qubit

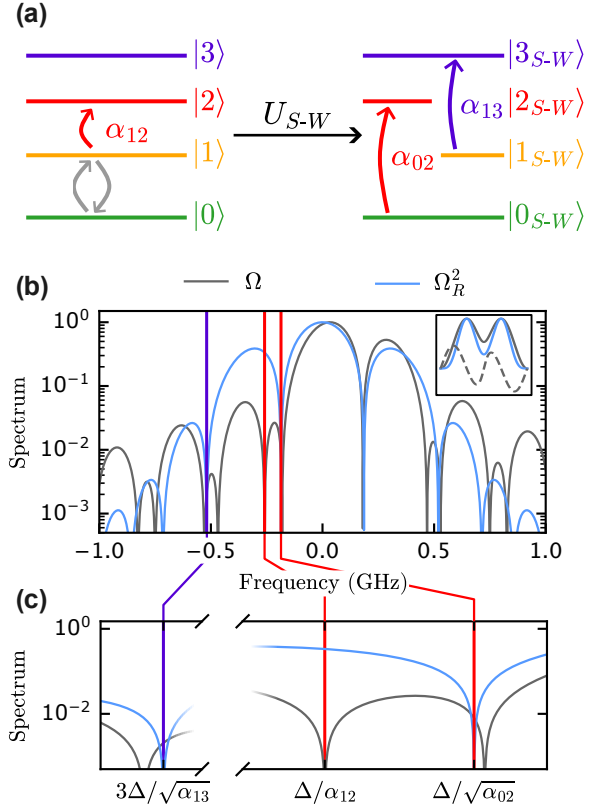


FIG. 1. Schematic of qubit level structures and pulse shapes for fast gates. (a) Schematic representation of qubit energy levels in the qubit eigenbasis (left) and in the dressed eigenbasis under Schrieffer-Wolff transformation (right). The colored arrows show the unwanted leakage transitions. (b) An example R2D pulse for $t_p = 7$ ns. The magnitude spectrum of both $\Omega = \Omega_R + i\Omega_I$ (gray) and Ω_R^2 (blue) are plotted. The solid lines highlight the unwanted transition frequencies to $|2\rangle$ (red) and $|3\rangle$ (purple). The inset shows the normalized time domain pulse for both Ω and Ω_R^2 , and the dashed line shows Ω_I . (c) The same magnitude spectrum of Ω and Ω_R^2 as in (b), highlighting the frequencies of the unwanted single- and multi-photon transitions.

is operated at the flux insensitive spot where it has a frequency f_q of 5.323 GHz and an anharmonicity $\Delta/2\pi$ of -225 MHz.

We start the validation by introducing a method capable of measuring a leakage error down to 10^{-6} . The standard approach uses leakage randomized benchmarking (LRB) [26], where the leakage is measured as it accumulates during Clifford-based randomized benchmarking. However, this approach has drawbacks. First, the random nature causes the accumulation of higher level state population to be incoherent. Second, extracting the leakage error of a specific gate requires interleaving, which can become expensive in terms of time and measurements. Third, since LRB accumulates leakage incoherently, it becomes challenging to distinguish coherent gate leakages from thermal transitions. Moreover,

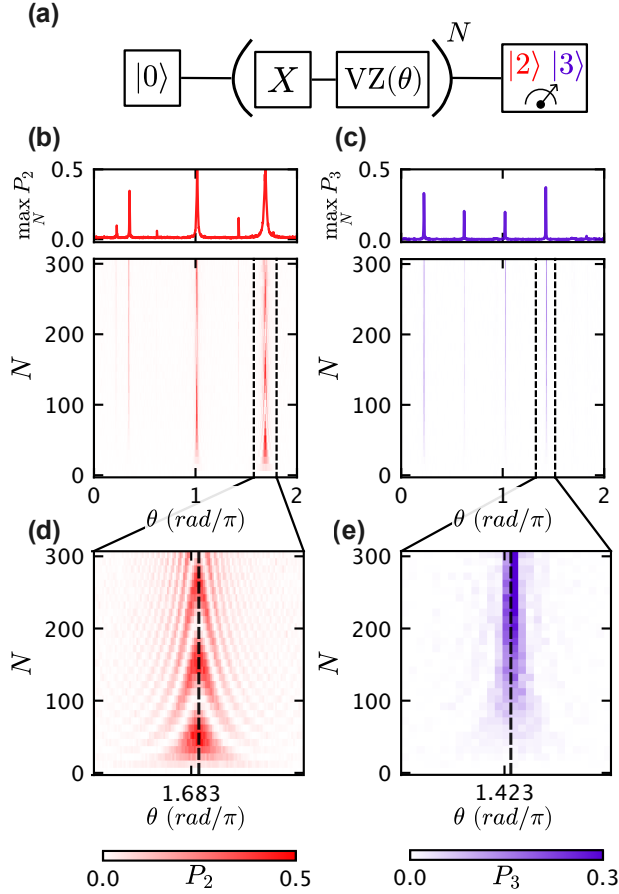


FIG. 2. **Amplified leakage error experiments.** (a) The schematic of the ALE pulse sequence. We repeatedly apply N times a X operation followed by a VZ with angle θ . The sequence ends with measurements of the leakage population. (b) The $|2\rangle$ leakage population as a function of θ and N . The line plot on top shows the maximum leakage population for each θ . (c) Same as (b) but plotted with $|3\rangle$ leakage population. (d)(e) The same data as in (b)(c) but with a magnified window around one of the amplification phases (dashed lines). The clear chevron pattern indicates a coherent oscillation of the leakage population.

we show later that it becomes difficult to obtain precision below 10^{-5} using LRB. Therefore, we introduce an amplified leakage error (ALE) method that coherently accumulates leakage errors for both X and $X/2$ gates, allowing for highly accurate measurements.

The key notion behind the ALE sequence is the phase matching of the leakage population after each operation. Using a Z -rotation with correct angle, the population will be phase matched and coherently accumulated. The ALE sequence consists of N repetitions of an X or composite X operation and a virtual- Z (VZ) gate with angle θ [Fig. 2(a)]. A similar principle was previously employed to amplify errors in two-qubit gates [27, 28]. The unitary of X containing coherent leakages and the VZ

can be written as

$$X = \begin{pmatrix} 0 & 1 & A_1 e^{-i(\phi_1 + \phi_2)} \\ 1 & 0 & A_0 e^{-i(\phi_0 + \phi_2)} \\ -A_0 e^{i(\phi_0 - \phi_2)} & -A_1 e^{i(\phi_1 - \phi_0)} & e^{-i2\phi_2} \end{pmatrix}, \quad (7)$$

$$VZ(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-i\theta} & 0 \\ 0 & 0 & e^{-i2\theta} \end{pmatrix}, \quad (8)$$

where $A_0, A_1 \ll 1$ are magnitudes of the leakage population transferred to $|2\rangle$, ϕ_0 and ϕ_1 are phases of the leakage population, and $2\phi_2$ is the phase acquired in $|2\rangle$ during X . The leakage rate of X is defined as $\epsilon_{leak,|2\rangle} = (A_0^2 + A_1^2)/2$. After we examine the ALE sequence, we find (see Supplement) that leakage is coherently accumulated only when θ satisfies

$$3\theta + 4\phi_2 = 2k\pi, \quad k \in \mathbb{Z}. \quad (9)$$

At the phase $\theta_k = (2k\pi - 4\phi_2)/3$, the qubit population coherently oscillates between the computational subspace and $|2\rangle$, and $P_2 = a_k \sin^2(\omega_k N/2)$, where a_k and ω_k are the amplitude and the frequency of the oscillation. The leakage error of X is proportional to the weighted sum of squared oscillation frequencies at all amplification phases,

$$\epsilon_{leak,|2\rangle} = \frac{1}{6} \sum_{k=1}^3 a_k \omega_k^2. \quad (10)$$

Similar analysis can be applied for the leakage rate to $|3\rangle$, where five distinct phases exist (see Supplement). While the coherent leakage error caused by the control pulse is amplified, the thermal leakage with random phases will remain low. Therefore, the ALE sequence can selectively amplify the coherent leakage error to different states.

We show an example of the experimental implementation of the ALE sequence in Fig. 2(b)-(e). The leakage population is plotted as a function of N and θ . The amplified leakage error to $|2\rangle$ is on the left, and error to $|3\rangle$ is on the right. Despite the existence of state preparation and measurement (SPAM) errors, the amplification of leakages to different states is independent and clearly visible. The line plots in Fig. 2(b)-(c) show the maximum leakage population over N as a function of θ . The target phases which amplify the leakage error are evenly spaced in $[0, 2\pi)$. The small peaks in Fig. 2(b) are caused by SPAM errors. Finally, in Fig. 2(d)-(e), we see clear chevron patterns evidencing a coherent process, with dashed lines at target phases.

The ALE experiment used in the gate calibration procedure can be further simplified. In fact, the target phases satisfying Eq. (9) stay unchanged unless ϕ_2 varies. As ϕ_2 is dominated by the frame frequency of $|2\rangle$, the target phases are stable in a large range of parameters for a constant pulse duration. Thus, only a $2\pi/3$ window is

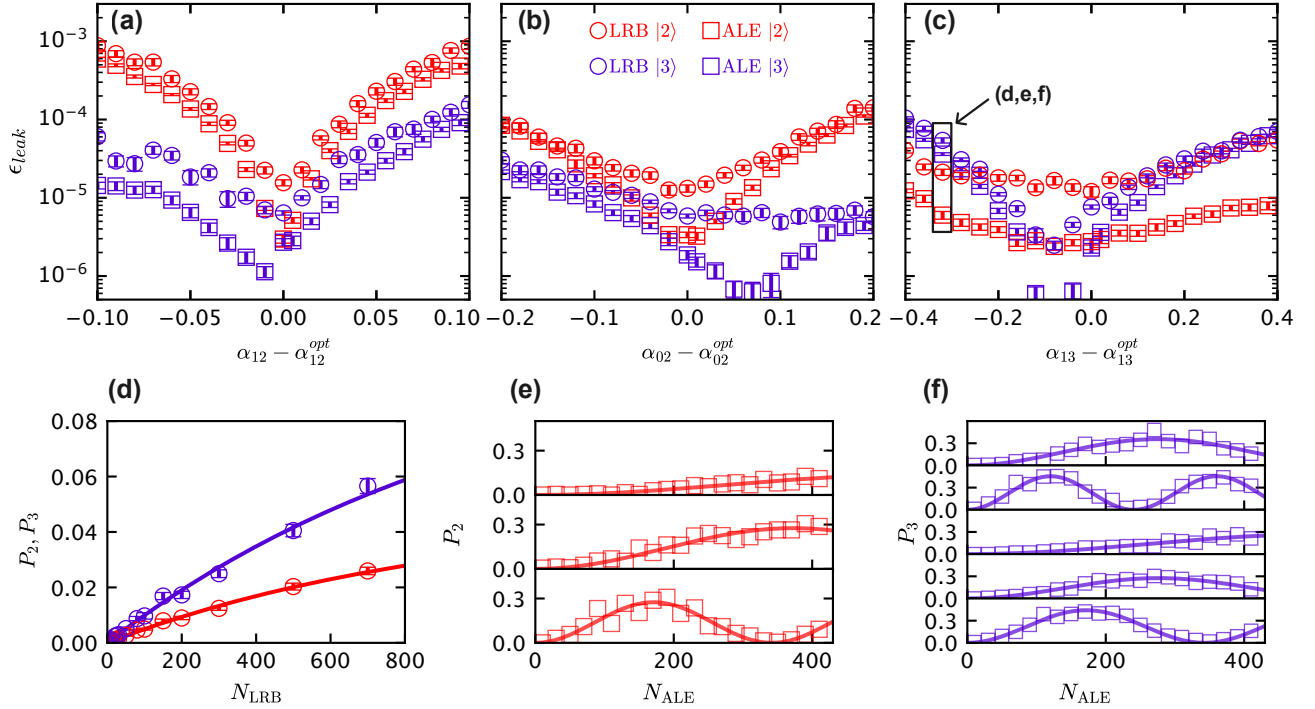


FIG. 3. **Extracting leakage error from leakage randomized benchmarking and amplified leakage error.** (a)(b)(c) Leakage error rates of an X gate at $t_p = 7$ ns extracted from LRB (circle) and ALE (square) for leakage state $|2\rangle$ (red) and $|3\rangle$ (purple) as a function of parameter α_{12} (a), α_{02} (b), and α_{13} (c). The optimal parameter set $(\alpha_{12}^{opt}, \alpha_{02}^{opt}, \alpha_{13}^{opt})$ is obtained through an in-situ optimization. The raw traces of the marked data points in (c) are shown. (d) Raw traces used to extract leakage errors marked in (c) for both states using LRB. Only the interleaved curves are shown. Solid lines are fit curves. (e)(f) Raw traces used to extract leakage errors marked in (c) for $|2\rangle$ (e) and $|3\rangle$ (f) using ALE. Solid lines are fit curves.

needed to find all phases, and a narrow window around them can be used afterwards. This significantly reduces the calibration time needed in the close-loop optimization.

The amplified leakage error sequence complements the metrology and calibration toolset for characterizing all errors of single qubit gates, together with the amplified phase error (APE) sequence [29] and the amplified amplitude error (AAE) sequence [30]. For each set of parameters $(\alpha_{12}, \alpha_{02}, \alpha_{13})$, APE and AAE experiments can deterministically find out the optimal constant detuning and pulse amplitude (see Supplement). Using the leakage error extracted from ALE experiments as the cost function, the Nelder-Mead optimization can reliably find the optimal gate under the R2D pulse shaping.

We validate the leakage errors obtained from the ALE through a comparison with the standard LRB method. In Fig. 3(a)-(c), leakage errors of an X gate at $t_p = 7$ ns from both methods are plotted as a function of α_{12} , α_{02} , and α_{13} , respectively. The parameters are individually swept around the optimal value obtained from an in-situ gate calibration, $(\alpha_{12}^{opt}, \alpha_{02}^{opt}, \alpha_{13}^{opt}) = (0.864, 1.467, 1.796)$. For each parameter set, APE and AAE experiments are used to find the minimum amplitude and phase error. A set of raw traces for the data points marked in Fig. 3(c) from both methods are shown in Fig. 3(d)-(f). The in-

coherently accumulated leakage of interleaved LRB, averaged over 60 random sequences, as well as the solid fit lines are plotted in Fig. 3(d). The population stays below 0.06 even after 800 Clifford gates. On the other hand, the amplified leakages from ALE experiments for $|2\rangle$ and $|3\rangle$ show clear coherent oscillations at amplification phases within just 400 operations, see the fitted curves in Fig. 3(e)-(f). As shown in Fig. 3(a)-(c), we find that the leakage rates increase by almost two orders of magnitude when the parameters are biased away from the optimal value. While a good correspondence of the leakage rates between two methods is prominently visible, noticeable differences appear for leakages below 10^{-5} . Despite differences in leakage errors between methods, the minimum total leakage is achieved with the same pulse parameters.

The leakage behavior in Fig. 3 provides an intuitive guidance for optimizing pulse parameters. After minimizing the phase and amplitude error, the leakage is solely controlled by $(\alpha_{12}, \alpha_{02}, \alpha_{13})$, and the optimal value can be obtained through the Nelder-Mead optimization. The corrections from α_{02} and α_{13} are nearly orthogonal, which benefits the optimization. Additionally, the ALE sequence requires less operations and shows larger leakage populations, which emphasizes its advantages in terms of time and measurements. Furthermore, a reliable

gate optimization can be achieved through ALE instead of LRB, evidenced by the good correspondence in both leakage errors and the optimal parameter values. We attribute the differences in leakage errors to the fact that the ALE method is insensitive to thermal excitations and $|2\rangle \rightarrow |3\rangle$ transition errors.

In order to evaluate the gate performance from all aspects, we need additional benchmarking tools. Besides LRB characterizing the leakage error ϵ_{leak} , we use standard randomized benchmarking (RB) [31] to characterize the total gate error ϵ_{tot} , and purity randomized benchmarking (PRB) [32] to characterize the decoherence error ϵ_{decoh} (see Supplement). All methods are run in an interleaved fashion to extract errors for the target gate.

To fully demonstrate the performance of the R2D pulse shaping, we perform a complete benchmark of single qubit gates with varying pulse lengths, and construct the error budgets. Using the calibration procedure discussed previously, we bring up both X and $X/2$ gates with pulse length $t_p \in [6.6, 9]$ ns. Pulses shorter than 6.6 ns will result in an unstable gate optimization as negative values appear in Ω_R^2 , and pulses longer than 9 ns will cause benchmarking sequences to exceed the waveform memory limit of the control electronics. To ensure the best results, a constant phase correction is included in the pulse (see Supplement) to compensate for the accumulated phase shift caused by the constant detuning. In Fig. 4, we demonstrate the benchmarking results of X gates on the left, and $X/2$ gates on the right. Error rates from all benchmarks are shown in Fig. 4(a)-(b). We see that the total leakage error stays below 2×10^{-5} down to a pulse length of 6.8 ns. At a pulse duration of 6.6 ns, the leakage error to $|3\rangle$ starts to dominate. At a pulse duration of 6.8 ns, we achieve X and $X/2$ gates with a total gate error of $(1.9 \pm 0.1) \times 10^{-4}$ and $(1.1 \pm 0.1) \times 10^{-4}$ respectively. We note that, with t_π being the π -pulse duration, this is equivalent to a normalized driven strength of $\Omega/\Delta = 1/2t_\pi\Delta = 0.33$, a record value for single qubit gates in superconducting qubits. The solid green line shows the total gate error from a Lindblad master equation simulation. The solid purple line shows the simulated coherent leakage error to $|3\rangle$ in the absence of incoherent processes. The same simulated leakage error to $|2\rangle$ is plotted in red. We use the experiment results to separate the total gate error into constituents: $\epsilon_{tot} = \epsilon_{decoh} + \epsilon_{leak} + \epsilon_{control}$, see Fig. 4(c)-(d). We find that gate errors of both X and $X/2$ are mainly limited by decoherence errors around a level of 1.04×10^{-4} , and coherent control errors increase with decreasing pulse lengths for X gates. We also show the optimal parameters for different pulse lengths in Fig. 4(e)-(f), where symbols and dashed lines represent values from the in-situ calibration and the simulation respectively.

In addition, when comparing the R2D pulse method we find it outperforms the standard single-photon DRAG correction (see Supplement), achieving a leakage reduc-

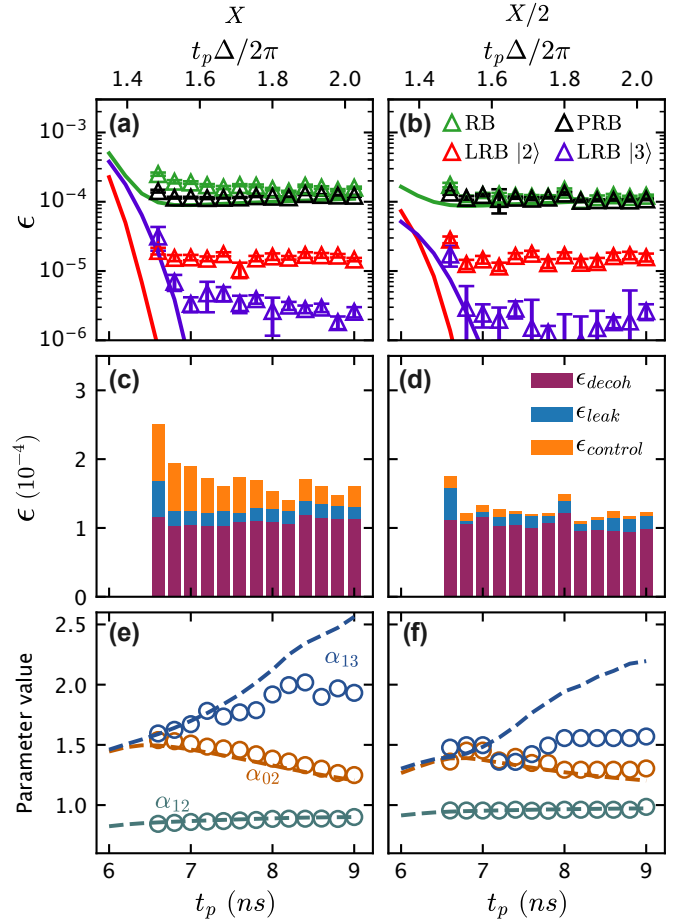


FIG. 4. Gate benchmarks with varying pulse lengths. Gate performance with pulse durations $t_p \in [6.6, 9]$ ns. X gate is on the left, and $X/2$ gate is on the right. (a)(b) Error rates extracted from RB (green), PRB (black), LRB for $|2\rangle$ (red), and LRB for $|3\rangle$ (purple). Solid lines are simulation results. (c)(d) Different contributions to total gate error, including decoherence error (dark magenta), leakage error (blue), and coherent control error (orange). (e)(f) Parameter sets of optimal gates from in-situ calibrations (symbols) and simulations (dashed lines).

tion exceeding a factor of 20.

The data in Fig. 4 show that the R2D pulse shape enables gates with nearly 99.99% fidelity at the normalized driven strength Ω/Δ up to 0.33, limited by qubit decoherence. First, the leakage errors remain below 2.0×10^{-5} over a large domain, and the increase in leakage errors below a pulse length of 6.8 ns is well predicted by the simulated coherent leakage. This evidently demonstrates the importance of accounting for multi-photon leakages in single qubit gates. We attribute the lower bound of leakages to excitations dominated by the voltage noise of room temperature electronics for the microwave control. Second, both X and $X/2$ gates show small coherent errors, with the error being substantially below 1.0×10^{-4} for $X/2$ gates. For X gates, we attribute the coherent

error increase with decreasing duration to microwave distortions, as it becomes more noticeable for shorter gates and higher amplitudes. This can be mitigated by introducing a small padding after the pulse. Third, the optimal pulse parameters from simulation and in-situ calibration compare well. Parameter α_{13} differs more for longer pulses because the change in $|3\rangle$ leakage is below the measurement threshold. The simulated and experimental results being consistent in all aspects validates the R2D pulse shape.

In conclusion, we demonstrate fast and precise X and $X/2$ gates through suppression of multi-photon transitions in the strongly driven regime using only three parameters. We also introduce the ALE sequence which allows precision near 10^{-6} in coherent leakage characterization, and complements the metrology toolset for the single qubit gate calibration. The presented approaches can be straightforwardly adopted by other platforms that use microwave qubit control. In addition, the ability to suppress multiple transitions is also beneficial to driving qubits simultaneously. Moreover, the same approaches can be used to mitigate the leakage in fast two-qubit gates. These results set a new benchmark for single qubit gates in terms of both fidelities and speed, and with enhanced qubit coherence and precise distortion characterization, the proposed methodology opens a path to achieving gates with a fidelity of 99.999%.

ACKNOWLEDGMENTS

The authors are grateful for TWPA calibration support from Marcello Guardascione, data visualization support from Niklas Oertel, software support from Juan Cereijo, and theory support from Dimitrios Georgiadis. The device is made in the Helmholtz Nano Facility. We acknowledge the support by the German Federal Ministry of Research, Technology and Space (BMFTR), funding program “Quantum technologies - from basic research to market”, project QSolid (Grant No. 13N16149). We acknowledge funding from the Horizon Europe program (HORIZON-CL4-2021-DIGITAL-EMERGING-02-10) via the project 101080085 (QCFD) and by HORIZON-CL4-2022-QUANTUM-01-SGA Project under Grant 101113946 OpenSuperQPlus100.

CONTRIBUTIONS

Y.G. and R.B. designed the experiments. Y.G. performed the experiments and analyzed the data. Y.G. conducted the simulation. J.D.D.C.J. and F.M. provided theoretical framework and models. Y.G. and A.G. implemented key components of the data-taking infrastructure. Y.G. designed the device. Y.L., Y.H., D.A.V., J.R.G., H.B., and M.N. contributed to device fabrication.

Y.G. and R.B. prepared the manuscript. All authors contributed to the experimental infrastructures or the theoretical studies, and to the manuscript revision.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

* y.gao@fz-juelich.de

† r.barends@fz-juelich.de

- [1] J. Eisert and J. Preskill, [Mind the gaps: The fraught road to quantum advantage](#) (2025), [arXiv:2510.19928 \[quant-ph\]](#).
- [2] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Physical Review A* **86**, [10.1103/physreva.86.032324](#) (2012).
- [3] M. Mohseni, A. Scherer, K. G. Johnson, O. Wertheim, M. Otten, N. A. Aadit, Y. Alexeev, K. M. Bresniker, K. Y. Camsari, B. Chapman, *et al.*, [How to build a quantum supercomputer: Scaling from hundreds to millions of qubits](#) (2025), [arXiv:2411.10406 \[quant-ph\]](#).
- [4] R. Barends and F. K. Wilhelm, [Performance-centric roadmap for building a superconducting quantum computer](#) (2025), [arXiv:2506.23178 \[quant-ph\]](#).
- [5] L. B. Nguyen, Y.-H. Lin, A. Somoroff, R. Mencia, N. Grabon, and V. E. Manucharyan, High-coherence fluxonium qubit, *Phys. Rev. X* **9**, [041041](#) (2019).
- [6] K. Dodge, Y. Liu, A. R. Klots, B. Cole, A. Shearow, M. Senatore, S. Zhu, L. B. Ioffe, R. McDermott, and B. L. T. Plourde, Hardware implementation of quantum stabilizers in superconducting circuits, *Phys. Rev. Lett.* **131**, [150602](#) (2023).
- [7] 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, *Physical Review A* **76**, [10.1103/physreva.76.042319](#) (2007).
- [8] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, Noise spectroscopy through dynamical decoupling with a superconducting flux qubit, *Nature Physics* **7**, [565–570](#) (2011).
- [9] M. Mathews, L. Pahl, D. Pahl, V. L. Addala, C. Tang, W. D. Oliver, and J. A. Grover, [Placing and routing quantum ldpc codes in multilayer superconducting hardware](#) (2025), [arXiv:2507.23011 \[quant-ph\]](#).
- [10] M. McEwen, D. Kafri, Z. Chen, J. Atalaya, K. J. Satzinger, C. Quintana, P. V. Klimov, D. Sank, C. Gidney, A. G. Fowler, *et al.*, Removing leakage-induced correlated errors in superconducting quantum error correction, *Nature Communications* **12**, [10.1038/s41467-021-21982-y](#) (2021).
- [11] R. Acharya, D. A. Abanin, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya, *et al.*,

- Quantum error correction below the surface code threshold, *Nature* **638**, 920–926 (2024).
- [12] R. Barends, J. Kelly, A. Megrant, D. Sank, E. Jeffrey, Y. Chen, Y. Yin, B. Chiaro, J. Mutus, C. Neill, P. O’Malley, P. Roushan, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Coherent josephson qubit suitable for scalable quantum integrated circuits, *Phys Rev Lett* **111**, 080502 (2013).
 - [13] E. Knill, Quantum computing with realistically noisy devices, *Nature* **434**, 39–44 (2005).
 - [14] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, *Physical Review Letters* **103**, 10.1103/physrevlett.103.110501 (2009).
 - [15] R. McDermott and M. G. Vavilov, Accurate qubit control with single flux quantum pulses, *Phys. Rev. Appl.* **2**, 014007 (2014).
 - [16] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer’s guide to superconducting qubits, *Applied Physics Reviews* **6**, 10.1063/1.5089550 (2019).
 - [17] E. Leonard, M. A. Beck, J. Nelson, B. Christensen, T. Thorbeck, C. Howington, A. Opremcak, I. Pechenezhskiy, K. Dodge, N. Dupuis, M. Hutchings, J. Ku, F. Schlenker, R. Suttle, C. Wilen, S. Zhu, M. Vavilov, B. L. T. Plourde, and R. McDermott, Digital coherent control of a superconducting qubit, *Phys. Rev. Appl.* **11**, 014009 (2019).
 - [18] E. Hyyppä, A. Vepsäläinen, M. Papič, C. F. Chan, S. Inel, A. Landra, W. Liu, J. Luus, F. Marxer, C. Ockeloen-Korppi, S. Orbell, B. Tarasinski, and J. Heinsoo, Reducing leakage of single-qubit gates for superconducting quantum processors using analytical control pulse envelopes, *PRX Quantum* **5**, 10.1103/prxquantum.5.030353 (2024).
 - [19] V. M. Frey, S. Mavadia, L. M. Norris, W. de Ferranti, D. Lucarelli, L. Viola, and M. J. Biercuk, Application of optimal band-limited control protocols to quantum noise sensing, *Nature Communications* **8**, 10.1038/s41467-017-02298-2 (2017).
 - [20] B. Chiaro and Y. X. Zhang, Active leakage cancellation in single qubit gates, *Physical Review Letters* **135**, ARTN 130601 10.1103/4kz9-w97h (2025).
 - [21] 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 Information* **7**, 10.1038/s41534-020-00346-2 (2021).
 - [22] F. Motzoi and F. K. Wilhelm, Improving frequency selection of driven pulses using derivative-based transition suppression, *Physical Review A* **88**, 10.1103/PhysRevA.88.062318 (2013).
 - [23] J. D. Da Costa Jesus *et al.*, In preparation (2025).
 - [24] Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. O’Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. White, A. Korotkov, and J. M. Martinis, Measuring and suppressing quantum state leakage in a superconducting qubit, *Physical Review Letters* **116**, 10.1103/physrevlett.116.020501 (2016).
 - [25] S. Gustavsson, O. Zwiernik, J. Bylander, F. Yan, F. Yoshihara, Y. Nakamura, T. P. Orlando, and W. D. Oliver, Improving quantum gate fidelities by using a qubit to measure microwave pulse distortions, *Phys. Rev. Lett.* **110**, 040502 (2013).
 - [26] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, *Physical Review A* **97**, 10.1103/physreva.97.032306 (2018).
 - [27] K. X. Wei, E. Pritchett, D. M. Zajac, D. C. McKay, and S. Merkel, Characterizing non-markovian off-resonant errors in quantum gates, *Physical Review Applied* **21**, 10.1103/physrevapplied.21.024018 (2024).
 - [28] B. Li, T. Calarco, and F. Motzoi, Experimental error suppression in cross-resonance gates via multi-derivative pulse shaping, *npj Quantum Information* **10**, 10.1038/s41534-024-00863-4 (2024).
 - [29] E. Lucero, J. Kelly, R. C. Bialczak, M. Lenander, M. Mariantoni, M. Neeley, A. D. O’Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis, Reduced phase error through optimized control of a superconducting qubit, *Physical Review A* **82**, 10.1103/PhysRevA.82.042339 (2010).
 - [30] 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, *Physical Review A* **93**, 10.1103/physreva.93.012301 (2016).
 - [31] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Physical Review Letters* **106**, 10.1103/physrevlett.106.180504 (2011).
 - [32] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, *New Journal of Physics* **17**, 113020 (2015).
 - [33] D. Zeuch *et al.*, Exact rotating wave approximation, *Annals of Physics* **423**, 168327 (2020).
 - [34] N. Lambert, E. Giguère, P. Menczel, B. Li, P. Hopf, G. Suárez, M. Gali, J. Lishman, R. Gadhvi, R. Agarwal, A. Galicia, N. Shammah, P. Nation, J. Johansson, S. Ahmed, S. Cross, A. Pitchford, and F. Nori, Qutip 5: The quantum toolbox in python, *Physics Reports* **1153**, 1 (2026).
 - [35] R. Barends, J. Kelly, A. Veitia, A. Megrant, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, E. Jeffrey, C. Neill, P. J. J. O’Malley, J. Mutus, C. Quintana, P. Roushan, D. Sank, J. Wenner, T. C. White, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Rolling quantum dice with a superconducting qubit, *Phys. Rev. A* **90**, 030303 (2014).

Supplementary Materials for “Ultrafast Single Qubit Gates through Multi-Photon Transition Removal”

Y. Gao,^{1, 2} A. Galicia,^{1, 2} J. D. Da Costa Jesus,^{3, 4} Y. Liu,¹ Y. Haddad,^{1, 2}
D. A. Volkov,^{1, 2} J. R. Guimarães,^{1, 2} H. Bhardwaj,^{1, 2} M. Jerger,¹ M. Neis,^{1, 2}
B. Li,³ F. A. Cárdenas-López,³ F. Motzoi,^{3, 4} P. A. Bushev,¹ and R. Barends,^{1, 2}

¹*Institute for Functional Quantum System (PGI-13), Forschungszentrum Jülich, 52425 Jülich, Germany*

²*Department of Physics, RWTH Aachen University, 52074 Aachen, Germany*

³*Institute for Quantum Control (PGI-8), Forschungszentrum Jülich, 52425 Jülich, Germany*

⁴*Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany*

(Dated: December 1, 2025)

QUBIT PARAMETERS AND STATE DISCRIMINATION

In this work, we use an isolated tunable transmon qubit that does not couple to any other qubits or couplers. The qubit chip is fabricated using the E-beam lithography. The Al-Al₂O₃-Al junctions are defined using the standard Manhattan process on a Si substrate.

The qubit has $E_C/h = 204.8$ MHz and $E_J/h = 18.7$ GHz. The operating point is near the flux insensitive point, where the qubit is at the frequency $f_q = 5.323$ GHz and has anharmonicity $\Delta/2\pi = -225.0$ MHz. The qubit XY control line is capacitively coupled to the qubit, and the coupling capacitance is $C_c = 95$ aF. The qubit is dispersively readout from a quarter-wave resonator. The resonator is at the frequency $f_r = 6.799$ GHz, and has a qubit-resonator coupling of $g_{qr}/2\pi = 45$ MHz. The average qubit coherence properties are $T_1 = 40$ μ s from T_1 experiments, $T_2^E = 56$ μ s from Echo experiments, and $T_2^R = 45$ μ s from Ramsey experiments.

To better quantify the leakage rate to different states, we use four-state discrimination. We reset the qubit to $|0\rangle$ by relaxation, and measure $2^{17} = 131072$ shots in each state. The acquisition length is 2.048 μ s. The raw IQ points and the fitted 2D Gaussian distributions are shown in Fig. S1. To discriminate the readout data, the nearest center method is used. We obtain the assignment matrix as

$$\Lambda = \begin{pmatrix} 89.64 & 9.49 & 0.75 & 0.12 \\ 13.10 & 86.03 & 0.80 & 0.07 \\ 13.67 & 11.43 & 69.93 & 4.97 \\ 10.46 & 2.76 & 15.78 & 71.00 \end{pmatrix} \times 10^{-2}, \quad (\text{S1})$$

where Λ_{ij} , the element at i -th row and j -th column, indicates the probability of measuring $|j\rangle$ when qubit is prepared at $|i\rangle$. In this assignment matrix, a large number of thermal population $\bar{n} \approx 0.1$ is observed. We use postselection to reduce the initialization error, and the new assignment matrix is

$$\Lambda^{ps} = \begin{pmatrix} 97.793 & 2.129 & 0.073 & 0.007 \\ 7.873 & 91.604 & 0.519 & 0.004 \\ 6.726 & 11.173 & 77.602 & 4.499 \\ 3.116 & 2.082 & 15.958 & 78.843 \end{pmatrix} \times 10^{-2}. \quad (\text{S2})$$

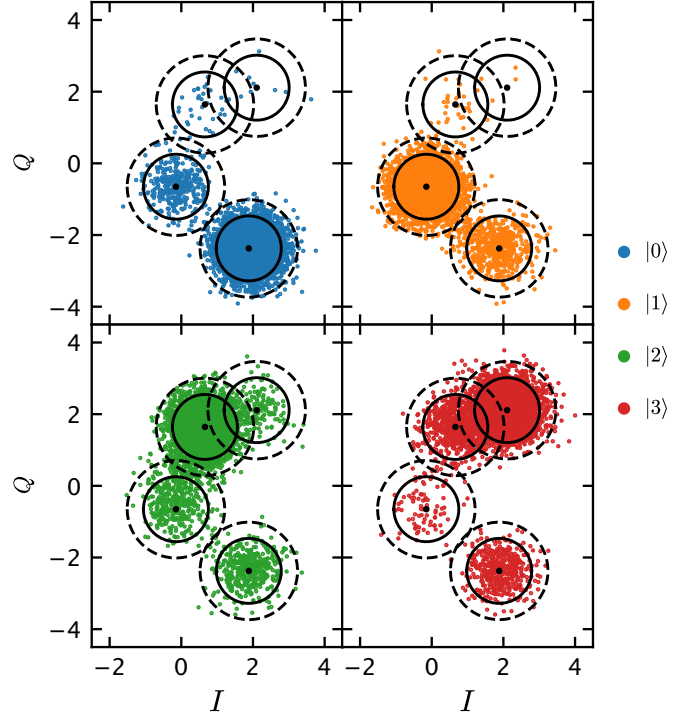


FIG. S1. Raw IQ points from the single shot experiment where qubit is prepared in the $|0\rangle$ (blue), $|1\rangle$ (orange), $|2\rangle$ (green), and $|3\rangle$ (red) states. We show 5000 out of 2^{17} preparations for each state. The black points are the fitted state centers, the circles have the radii corresponding to two (solid line) and three (dashed line) standard deviations of the data.

In all experiments performed in this work, we use post-selection to reset the qubit to $|0\rangle$, and we do not correct for SPAM errors.

CONSTANT DETUNING AND PHASE CORRECTION

In this work, we use a constant detuning to correct the phase error. The theoretical value for the constant detuning δ_ω satisfies

$$\delta_\omega = 0.712 \frac{(4\alpha_{12} - 2) \pi^2}{\Delta t_g^2}, \quad (\text{S3})$$

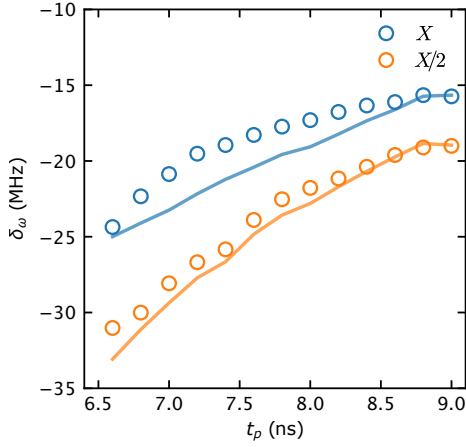


FIG. S2. Constant detuning corrections for different pulse durations from experiments (symbols) and predictions using the experimentally calibrated α_{12} (lines).

see Ref. [23] for detailed derivations. It matches well with experimental results [Fig. S2], and can be used as initial values for the pulse. Additional to the constant detuning correction, an extra phase correction is included to compensate the accumulated phase shift,

$$\Omega'(t) = \Omega(t)e^{-i\delta_\omega(t-t_p/2)}, \quad (\text{S4})$$

where $\Omega(t)$ is the pulse envelope as in Eq. (4) in the main text. The phase correction $\delta_\omega t_p/2$ is particularly important when other single qubit gates of the same qubit are at different lengths, or their phase errors are corrected by other approaches. In the absence of this phase correction, the effective rotation axis will be away from x axis in the $x-y$ plane. This explains the issue reported in Ref. [24] that the best results are only achievable by using the same detuning and the same length for X and $X/2$ gates.

SINGLE QUBIT GATE SIMULATION

We simulate the qubit evolution without RWA, as the rotating wave approximation (RWA) is only valid under weak coupling [33]. The bare transmon is modeled in the qubit charge basis as

$$H_q = 4E_C \hat{n}^2 + E_J \cos \hat{\phi}, \quad (\text{S5})$$

where $\hat{n}$ is the charge operator, $\hat{\phi}$ is the phase operator, and

$$\cos \hat{\phi} = \frac{1}{2} \sum_n |n\rangle \langle n+1| + |n+1\rangle \langle n|. \quad (\text{S6})$$

Transmon parameters E_C and E_J are set such that the lowest four energy levels compare well with experiment

results. There are in total 19 charge states in the simulation. A capacitively coupled microwave drive is modeled as

$$H_d = \hbar \Omega_{RF}(t) \hat{n}. \quad (\text{S7})$$

Microwave distortion is not considered. We also include 1 ns padding after the pulse as in the experiment. The pulse is modulated by the qubit frequency obtained from Eq. (S5). The QuTiP [34] package is used to simulate the qubit evolution. In Fig. S3, simulation results of the optimal pulses at $t_p = 7$ ns using R2D and DRAG pulse shapes are shown.

To incorporate incoherent processes during the evolution, we use Lindblad master equation,

$$\dot{\rho} = \frac{i}{\hbar} [\rho, H_q + H_d] + \sum_i L_i \rho L_i^\dagger - \frac{1}{2} \left\{ \rho, L_i^\dagger L_i \right\} \quad (\text{S8})$$

where ρ is the density operator, and we use collapse operators $L_- = \sqrt{\Gamma_-} a$, $L_+ = \sqrt{\Gamma_+} a^\dagger$, and $L_\phi = \sqrt{\Gamma_\phi} a^\dagger a$. $\Gamma_- = (1 - \bar{n})/T_1$, $\Gamma_+ = \bar{n}/T_1$, and $\Gamma_\phi = 1/T_2^E - 1/2T_1$ can be measured in experiments.

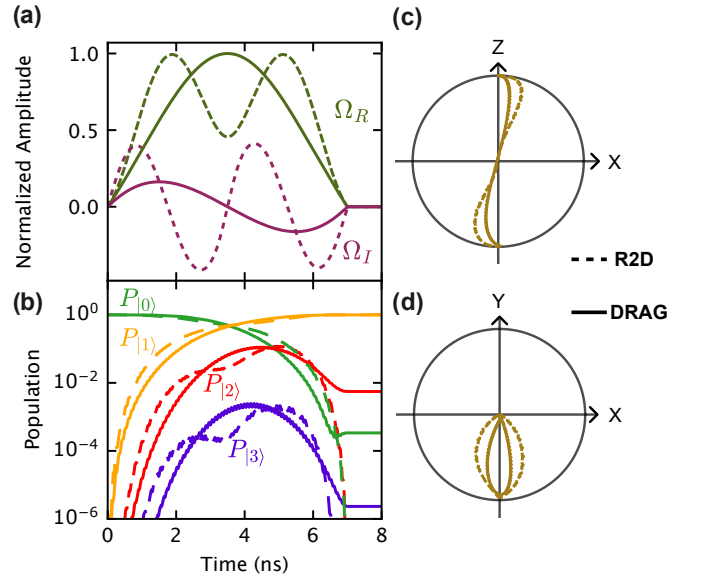


FIG. S3. Simulation results of the optimal pulses using R2D (dashed lines) and DRAG (solid lines) pulse shapes. (a) The time domain signals of the real part and the imaginary part. (b) Simulated population of lowest four states. (c) Qubit state trajectories projected in X - Z plane. (d) Qubit state trajectories projected in X - Y plane.

PULSE SHAPE IN TIME AND FREQUENCY DOMAIN

It is important to understand in what way the corrections are affecting the pulse shape. In Fig. S4, we show how the time and frequency domain pulses change when

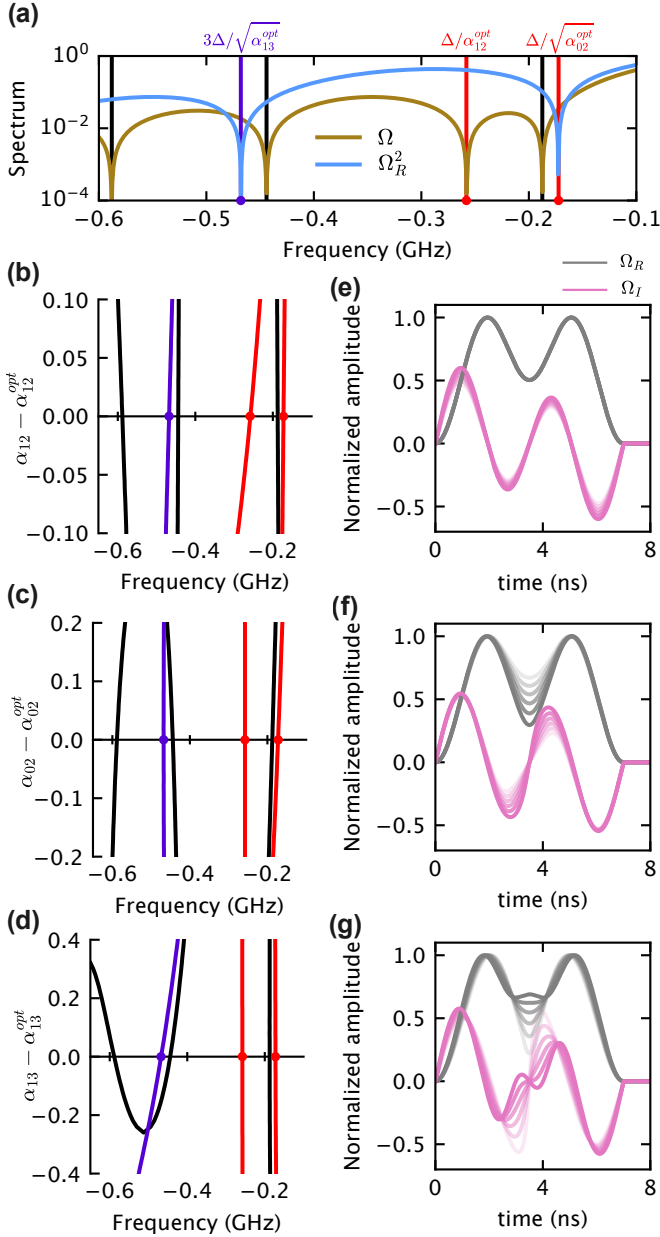


FIG. S4. R2D pulse shape in the frequency domain and the time domain. (a) The optimal pulse at $t_p = 7$ ns with pulse parameter $(\alpha_{12}^{opt}, \alpha_{02}^{opt}, \alpha_{13}^{opt}) = (0.864, 1.467, 1.796)$. (b) The frequency domain pulses when parameter α_{12} is swept around α_{12}^{opt} . The colored lines are the notches shown in (a), and the corresponding points are marked with dots. (e) The time domain pulses when parameter α_{12} is swept around α_{12}^{opt} . The lighter color indicates a smaller value of the parameter. (c)(f) Same plots as in (b)(e) but sweeping α_{02} . (d)(g) Same plots as in (b)(e) but sweeping α_{13} .

the pulse parameters are swept. In Fig. S4(a), the spectrum of an optimal pulse at $t_p = 7$ ns is plotted, and the parameters $(\alpha_{12}^{opt}, \alpha_{02}^{opt}, \alpha_{13}^{opt}) = (0.864, 1.467, 1.796)$ are obtained from in-situ optimization. We mark the notches that we actively optimize with colors and the

rest of notches black. The pulse parameters are swept individually around the optimal value and the pulses are plotted in both time and frequency domain. For time domain pulse, we use a lighter color to indicate a smaller value of the parameter. We sweep α_{12} in Fig. S4(b)(e), α_{02} in Fig. S4(c)(f), and α_{13} in Fig. S4(d)(g). From the frequency domain notches, the corrections from α_{02} and α_{13} are nearly orthogonal, which agrees with the leakage rates scan in Fig. 3 in the main text. Note the opposite effects on $\Omega_R(t)$ as α_{02} and α_{13} increase, this enables a short pulse since larger α_{02} corrections in fast pulses tends to make Ω_R^2 negative.

EXPERIMENTAL SETUP AND SIGNAL DISTORTION

The experimental setup and an optical picture of the transmon are shown in Fig. S5. The transmon qubit chip is packaged, and protected by a μ -metal shield. The device is mounted at the mixing chamber of a Bluefors XLD fridge. The qubit flux is directly biased by a Zürich Instruments HDAWG using the voltage offset of the channel. A Zürich Instruments SHFQA is used for the qubit readout. The readout channel has $f_{LO} = 7.0$ GHz, output range is 10 dBm, and input range is -10 dBm. The qubit XY control signal is generated by a Zürich Instruments SHFSG. We use $f_{LO} = 5.4$ GHz and output range of 10 dBm.

To mitigate the microwave signal distortion caused by the electronics and control wiring, we measure the frequency response of the system through an oscilloscope. The inverse transfer function of the system is applied to the pulse samples before the samples are sent to SHFSG. This approach is significantly limited by the waveform memory of the electronics, which prohibits us to benchmark gates at longer pulse duration.

DERIVATION OF AMPLIFIED LEAKAGE ERROR

Here, we derive the amplification phases of amplified leakage error (ALE) sequence, and extract the leakage error.

In the absence of phase and amplitude errors, the unitary of an X operation with only leakage to $|2\rangle$ can be written as

$$X = \begin{pmatrix} 0 & 1 & -A_1 e^{-i(\phi_1 + \phi_2)} \\ 1 & 0 & -A_0 e^{-i(\phi_0 + \phi_2)} \\ A_0 e^{i(\phi_0 - \phi_2)} & A_1 e^{i(\phi_1 - \phi_2)} & e^{-i2\phi_2} \end{pmatrix}, \quad (\text{S9})$$

where $A_0, A_1 \ll 1$ are the leakage population to $|2\rangle$, ϕ_0, ϕ_1 are phases of the leakage population, and $2\phi_2$ is the phase acquired in $|2\rangle$ during X . The leakage error

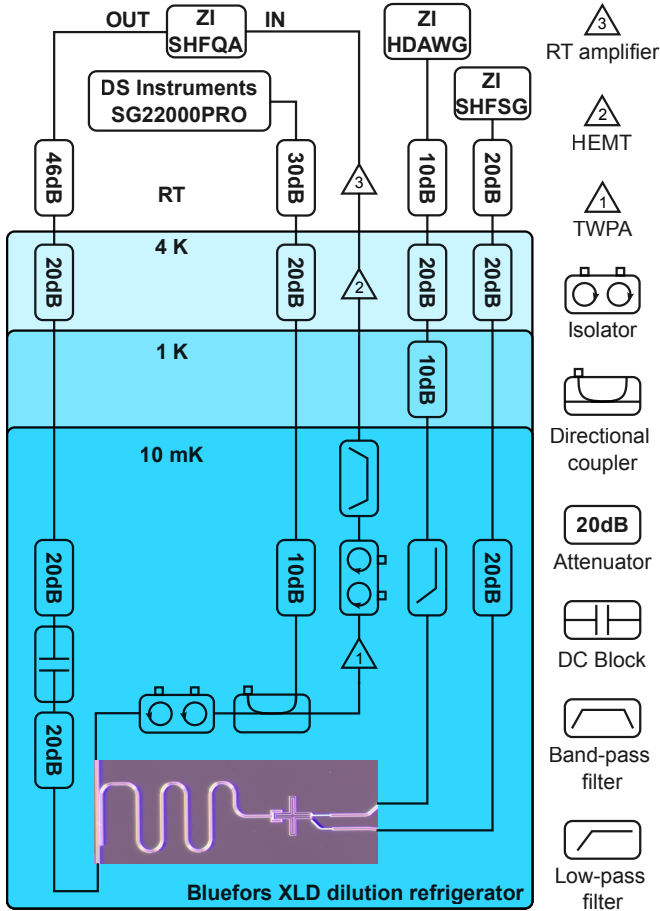


FIG. S5. The electronics and control wiring schematic.

can be calculated as $\epsilon_{\text{leak},|2\rangle} = (A_0^2 + A_1^2)/2$. A virtual-Z rotation with angle θ can be written as

$$\text{VZ}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-i\theta} & 0 \\ 0 & 0 & e^{-i2\theta} \end{pmatrix}. \quad (\text{S10})$$

Using Eq. (S9) and Eq. (S10), every double applications of X and VZ will result in,

$$U = (\text{VZ}(\theta) \cdot X)^2 = \begin{pmatrix} 1 & 0 & u_{00} \\ 0 & 1 & u_{01} \\ u_{10} & u_{11} & e^{-i(4\phi_2 + 3\theta)} \end{pmatrix}. \quad (\text{S11})$$

with u_{ij} being complex elements of U . In order to coherently accumulate the leakage population, the amplification phases θ needs to satisfy

$$4\phi_2 + 3\theta_k = 2k\pi. \quad (\text{S12})$$

At the amplification phases θ_k , the leakage population can only be amplified if it has the same phase after each operation. This indicates the ability to filter out the thermal excitation.

To extract the leakage rate, we first obtain the effective Hamiltonian of the system,

$$U^{N/2} = e^{iH_{\text{eff}}N/\hbar} \approx (I + i\frac{2}{\hbar}H_{\text{eff}})^{N/2}. \quad (\text{S13})$$

Then we solve the the time dependent Schrödinger equation using H_{eff} ,

$$-\frac{i}{\hbar}H_{\text{eff}}|\Psi(N)\rangle = \frac{d}{dN}|\Psi(N)\rangle, \quad (\text{S14})$$

with $|\Psi(0)\rangle = |0\rangle$, the $|2\rangle$ population after each operation can be expressed as,

$$P_2 = |\langle 2|\Psi(N)\rangle|^2 = a_k \sin^2(\omega_k N/2), \quad (\text{S15})$$

$$a_k \omega_k^2 = A_0^2 + A_1^2 - 2A_0A_1 \cos(\Phi + \theta_k), \quad (\text{S16})$$

where N is the number of applications, $a_k = 1/2$ for $|2\rangle$, and Φ is a function of ϕ_0, ϕ_1, ϕ_2 . Using Eq. (S16) and the identity $\sum_{k=1}^n \cos(a + 2k\pi/n) = 0$,

$$\epsilon_{\text{leak},|2\rangle} = \frac{1}{6} \sum_{k=1}^3 a_k \omega_k^2. \quad (\text{S17})$$

In the case where $\omega_k N \ll 2\pi$, the Eq. (S15) reduces to $P_2 = a_k \omega_k^2 N^2/4$. This can be used as the fit model when the leakage is small.

The same analysis can be applied to $|3\rangle$, and the leakage error satisfies,

$$P_3 = a_k \sin^2(\omega_k N), \quad (\text{S18})$$

$$\epsilon_{\text{leak},|3\rangle} = \frac{1}{10} \sum_{k=1}^5 a_k \omega_k^2 = \frac{1}{2} (B_0^2 + B_1^2). \quad (\text{S19})$$

where B_i are the leakage population from $|i\rangle$ to $|3\rangle$. We would like to point out that the a_k for $|3\rangle$ depends on leakage rates from different states, including $|2\rangle$, whereas the final leakage error for the X operation does not. This indicates that ALE sequence is insensitive to transitions between $|2\rangle$ and $|3\rangle$.

SINGLE QUBIT GATE CALIBRATION PROCEDURE

In this section, we give more details about the single qubit gate calibration procedure. We use the amplified amplitude error method (AAE) [30], also called ping-pong experiment, to minimize the amplitude error. We also use the amplified phase error approach (APE) [29] to minimize the phase error. In Fig. S6, we show the schematic of the single qubit gate calibration procedure. We assume the qubit frequency f_q and anharmonicity Δ are well calibrated by Ramsey experiments before the Nelder-Mead (N-M) optimization. In Fig. S6(a), we show the flow chart of the optimization procedure. For each

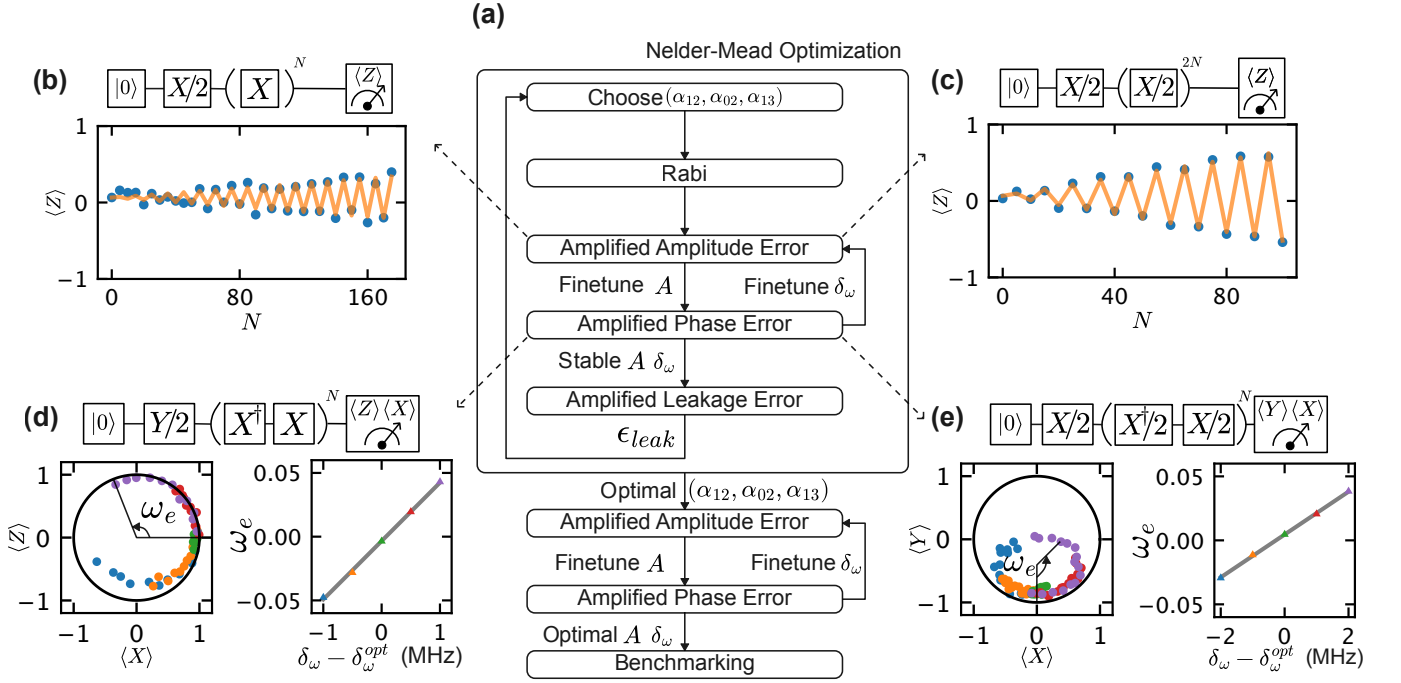


FIG. S6. Schematic of the single qubit gate calibration procedure. (a) Flow chart of the Nelder-Mead optimization. We assume the qubit frequency f_q and the anharmonicity Δ is well calibrated before the Nelder-Mead optimization. (b)(c) The amplified amplitude error experiments for X gate (b) and for $X/2$ gate (c). The pulse sequence, the raw result (blue dots), and the fit (orange line) are shown. (d)(e) The amplified phase error experiments for X gate (d) and for $X/2$ gate (e). The pulse sequence, the raw result (colored symbols) and the fit (gray line) are shown.

iteration in N-M optimization, a set of $(\alpha_{12}, \alpha_{02}, \alpha_{13})$ is chosen by the optimizer. We first run a Rabi experiment to coarsely calibrate the pulse amplitude. Then, we use iterative AAE and APE experiments to finetune the pulse amplitude A and the constant detuning δ_ω . After both parameters are stable (typically we need two iterations), an ALE experiment is performed to evaluate the leakage error ϵ_{leak} . This ϵ_{leak} will be used as the loss function for optimizer to choose the next set of parameters until the algorithm converges. After the N-M optimization, extra iterations of AAE and APE are used to perform the final adjustments on the pulse amplitude A and the constant detuning δ_ω .

Fig. S6(b)-(c) show examples of AAE experiments for both X and $X/2$ gates. The results are fitted to the function

$$f = a \sin(N(\pi + \phi_e)), \quad (\text{S20})$$

where ϕ_e is the over-/under-rotation angle.

Fig. S6(d)-(e) show examples of APE experiments for both X and $X/2$ gates. Note the differences in the initial rotation and the final measurements. The error in the constant detuning can be extracted by fitting to the model

$$f = r_I \cos(\omega_e N + \phi_0) + i r_Q \sin(\omega_e N + \phi_0) + c_I + i c_Q, \quad (\text{S21})$$

where the rotation frequency ω_e is proportional to the errors in the constant detuning, and $\omega_e = 0$ when the constant detuning is optimal. Theoretically, the rotation radii $(r_I, r_Q) = (1, 1)$ for X gate and $(r_I, r_Q) = (1, 0.5)$ for $X/2$ gate, the initial angle $\phi_0 = 0$ for X gate and $\phi_0 = -\pi/2$ for $X/2$ gate, the rotation centers $(c_I, c_Q) = (0, 0)$ for X gate and $(c_I, c_Q) = (0, -0.5)$ for $X/2$ gate.

CLIFFORD-BASED RANDOMIZED BENCHMARKING

The comprehensive benchmarking of the single qubits gates is done through standard randomized benchmarking (RB) [31], leakage randomized benchmarking (LRB) [26], and purity benchmarking (PRB) [32]. We use RB to characterise the total gate error ϵ_{tot} , LRB to characterize leakage error ϵ_{leak} , and PRB to characterize the decoherence error ϵ_{decoh} . The benchmarkings are run in an interleaved fashion to obtain the error for the target gate. See Fig. S7 for schematics of all benchmarking experiments.

All three benchmarking protocols include running multiple random sequences that consist of random Clifford gates C_i in different length (N) and a final recovering gate C_r .

RB measures the averaged population (P) of $|0\rangle$ at the

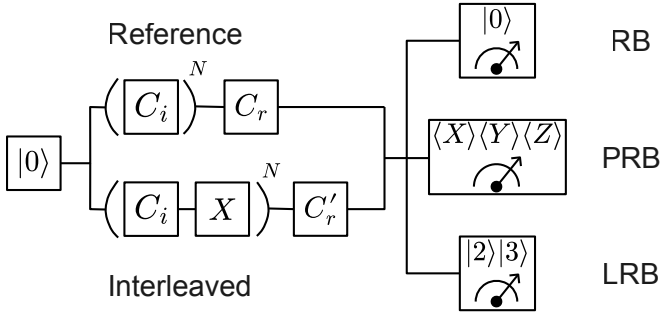


FIG. S7. Schematics of all benchmarking experiments. Both reference and interleaved sequences are shown.

end of each sequence, and fits it to model

$$P = A + Bp^N. \quad (\text{S22})$$

The total gate error of the target gate operation is calculated by

$$\epsilon_{tot} = (1 - p_{int}/p_{ref})/2, \quad (\text{S23})$$

where p_{ref} and p_{int} are fitted results from reference and interleaved sequences.

LRB measures the averaged population of leakage states, and fits it to model in Eq. (S22). The average leakage error of Clifford gates is $\epsilon = A(1 - p)$, and the leakage error of the target gate operation is

$$\epsilon_{leak} = \epsilon_{int} - \epsilon_{ref}, \quad (\text{S24})$$

where ϵ_{ref} and ϵ_{int} are leakage results calculated from reference and interleaved sequences.

PRB measures the averaged purity $U = \langle X \rangle^2 + \langle Y \rangle^2 + \langle Z \rangle^2$ at the end of each sequence, fits it to model

$$U = A + Bu^N. \quad (\text{S25})$$

We use $X/2$ and $Y/2$ gate after each sequence to project the qubit state to different axes, and $\langle Z \rangle = P_0 - P_1$ without the normalization of $P_0 + P_1 = 1$. The purity error of the target gate operation is calculated by

$$\epsilon_{pur} = \left(1 - \sqrt{u_{ref}/u_{int}}\right)/2, \quad (\text{S26})$$

where u_{ref} and u_{int} are the fitted results from reference and interleaved sequences. As mentioned in [32], the decoherence error is obtained by $\epsilon_{decoh} \approx \epsilon_{pur} - \epsilon_{leak}/2$ in the presence of leakage channels.

For the Clifford gates decomposition, we follow the mapping stated in Ref. [35]. The $t_p = 8$ ns X gate is used when the pulse lengths of $X/2$ gate are swept, and vice versa. As the waveform memory of the control electronics is limited, the benchmarkings cannot reach the saturation values. We use $A = (\Lambda_{01}^{ps} + \Lambda_{10}^{ps})/2$ for RB fitting, where Λ_{ij}^{ps} are the elements in the assignment matrix. We use $A = 0$ for PRB fitting. In the case of low

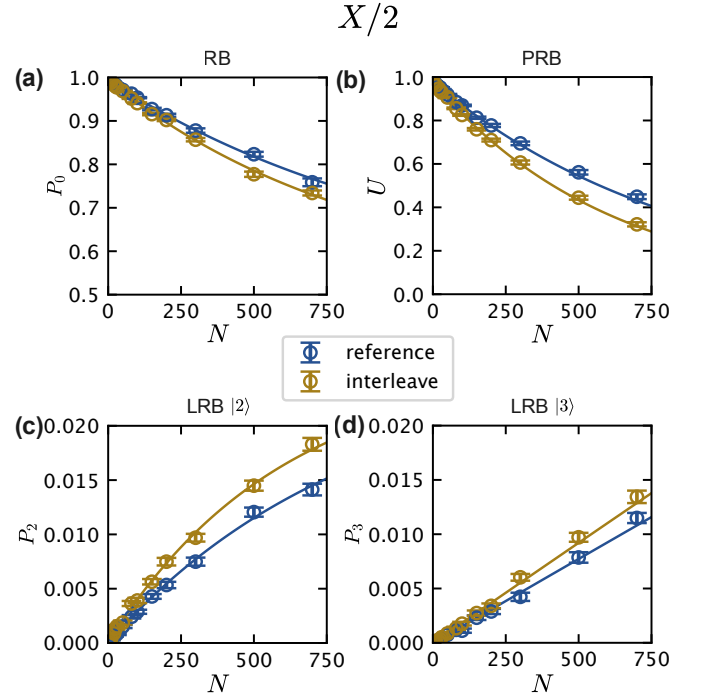


FIG. S8. Benchmarkings of the optimal $X/2$ gate at $t_p = 6.8$ ns, averaged over 60 random sequences. The error bar indicates the standard error of the mean. The solid lines are fit curves. (a) RB. (b) PRB. (c) LRB for $|2\rangle$. (d) LRB for $|3\rangle$.

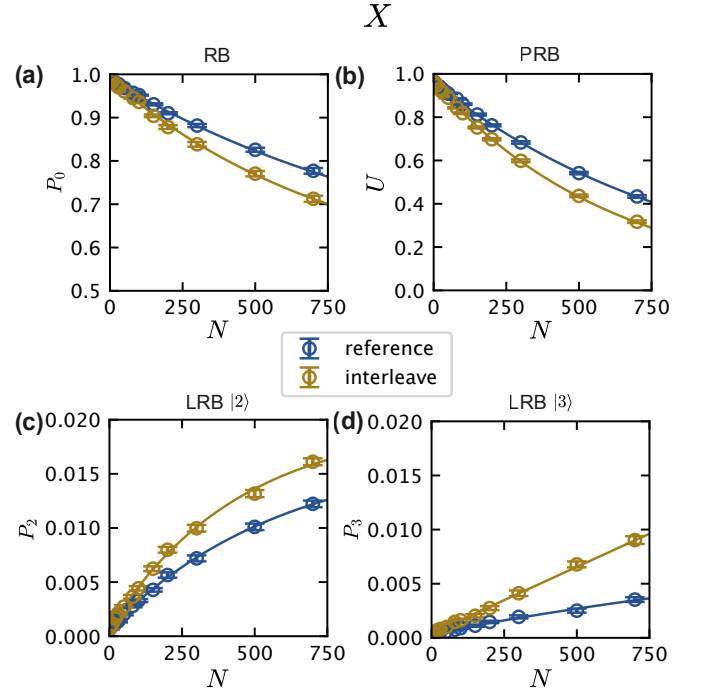


FIG. S9. Benchmarkings of the optimal X gate at $t_p = 6.8$ ns, averaged over 60 random sequences. The error bar indicates the standard error of the mean. The solid lines are fit curves. (a) RB. (b) PRB. (c) LRB for $|2\rangle$. (d) LRB for $|3\rangle$.

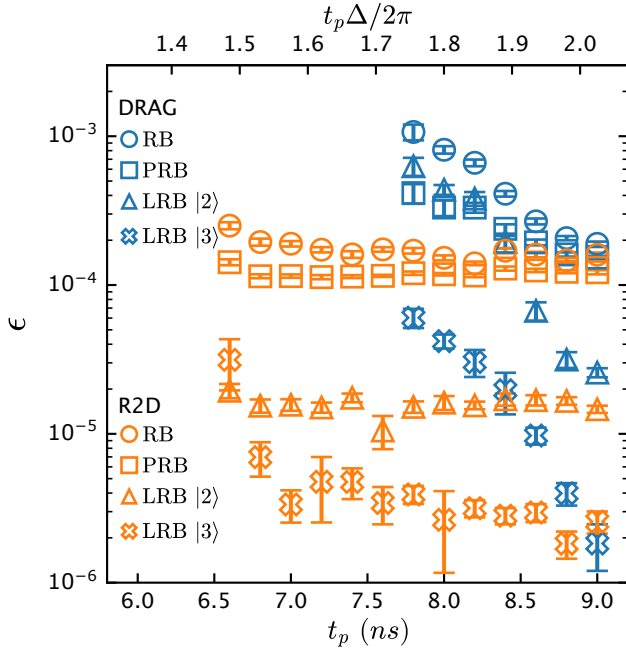


FIG. S10. Benchmarking results for X gates using the standard DRAG and the R2D methods. R2D outperforms standard DRAG.

leakage population in LRB, we use $P = (1-p)AN$ as the fit model and assume $A = -B$. Fig. S8 and Fig. S9, we show the benchmark curves for the optimal $X/2$ and X gates at $t_p = 6.8$ ns, respectively. All benchmark curves are averaged over 60 random sequences.

We directly compare the benchmarking results for X gates using standard DRAG and R2D corrections in Fig. S10. For the standard DRAG correction, the α_{12} is optimized to minimize the total leakage error, and α_{02} and α_{13} are set to 0.