Calibration of Drive Nonlinearity for Arbitrary-Angle Single-Qubit Gates Using Error Amplification

Stefania Lazăr, Quentin Ficheux, Johannes Herrmann, Ants Remm, Nathan Lacroix, Christoph Hellings, Francois Swiadek, Dante Colao Zanuz, Graham J. Norris, Mohsen Bahrami Panah, Alexander Flasby, Michael Kerschbaum, Jean-Claude Besse, Christoph Eichler, and Andreas Wallraff

1 Department of Physics, ETH Zurich, Zurich CH-8093, Switzerland
2 ETH Zurich—PSI Quantum Computing Hub, Paul Scherrer Institute, Villigen CH-5232, Switzerland
3 Quantum Center, ETH Zurich, Zurich 8093, Switzerland

(Received 6 December 2022; revised 26 June 2023; accepted 31 July 2023; published 15 August 2023)

The ability to execute high-fidelity operations is crucial to scaling up quantum devices to large numbers of qubits. However, signal distortions originating from nonlinear components in the control lines can limit the performance of single-qubit gates. In this work, we use a measurement based on error amplification to characterize and correct the small single-qubit rotation errors originating from the nonlinear scaling of the qubit drive rate with the amplitude of the programmed pulse. With our hardware, and for a 15-ns pulse, the rotation angles deviate by up to several degrees from a linear model. Using purity benchmarking, we find that control errors reach $2 \times 10^{-4}$, which accounts for half of the total gate error. Using cross-entropy benchmarking, we demonstrate arbitrary-angle single-qubit gates with coherence-limited errors of $2 \times 10^{-4}$ and leakage below $6 \times 10^{-5}$. While the exact magnitude of these errors is specific to our setup, the presented method is applicable to most sources of nonlinearity. Our work shows that the nonlinearity of qubit drive line components imposes a limit on the fidelity of single-qubit gates, independent of improvements in coherence times, circuit design, or leakage mitigation when not corrected for.

DOI: 10.1103/PhysRevApplied.20.024036

1. INTRODUCTION

As quantum processors accommodate an ever larger number of qubits, understanding the errors that limit their performance remains an important task to enable further progress [1–3]. In particular, errors in single- and two-qubit gates limit the fidelities of quantum error-correction codes [4–8], variational quantum algorithms [9,10], or digital quantum simulations [11–13]. These errors can be reduced by increasing the coherence times, but also by mitigating control errors [14–19] or by reducing the circuit depth when compiling algorithms with a larger gate set [10,20,21].

For superconducting quantum processors, standard gate calibration techniques [22–25] assume that the control parameters respond linearly to the control fields. However, commonly used electronic components can introduce nonlinear effects, which result in control errors if unaccounted for. For example, in-phase and quadrature (IQ)-mixer imperfections produce spurious frequency components [26–28], gain compression in amplifiers saturates the power of control signals [29], and saturation of pulse-generation devices leads to waveform distortions [30].

For the past decade, two-qubit gates have imposed the main limitation on superconducting quantum processors [2,31]. However, considerable efforts in recent years have reduced the errors of two-qubit gates into the range of $10^{-3}$ [19,32–36], making their contribution to computation errors comparable to that of single-qubit gates [1]. Moreover, current algorithms often require a larger number of single-qubit gates than two-qubit gates, especially for dynamical decoupling schemes [37,38]. Thus, further improvements in the performance of single-qubit gates are needed to enhance the overall fidelity of quantum algorithms.

Arbitrary single-qubit rotation angles are routinely calibrated by assuming that the rotation of the state vector scales linearly with the pulse amplitude [22–25]. However, nonlinear electronic components in the drive line can produce a deviation from this linear model that leads to rotation errors. So far, few techniques have been developed to compensate for similar imperfections, such as waveform predistortion [32,39,40], pulse-shape optimization methods [41,42], and correction of drive-pulse amplitudes affected by imperfect signal generation [30].
Here, we present an in situ amplification scheme for small rotation errors, which allows us to calibrate coherence-limited arbitrary-angle single-qubit gates on a transmon qubit [43] even in the presence of nonlinearity. With this technique, we correct for residual rotation errors when using pulses with amplitudes approaching the 1-dB compression point of the frequency conversion device in our microwave line, which is nonlinear at high input powers [29]. Using randomized benchmarking [44], purity benchmarking [45], and cross-entropy benchmarking [46], we demonstrate arbitrary-angle 15-ns-long single-qubit gates with coherence-limited performance and with leakage outside the computational subspace below $6 \times 10^{-5}$. Importantly, we find that, for short gate duration, coherent errors in the computational subspace caused by nonlinear pulse amplitude distortions are one order of magnitude larger than leakage, imposing a lower bound on the gate error for a given gate duration, independent of improvements in coherence times.

We study the effect described above on two flux-tunable superconducting transmon qubits from two different 17-qubit devices labeled A and B, with similar designs to that discussed in Ref. [7]. The devices are controlled by nominally the same electronics. The measurements presented in Secs. II and V were done on device A, while those discussed in Secs. III and IV were performed on device B (see Appendix A, where we also summarize the device parameters).

II. PULSE AMPLITUDE CALIBRATION USING ERROR AMPLIFICATION

Nonlinear components in qubit control lines distort coherent microwave drive pulses, effectively changing the amplitude of the driving field at the qubit frequency in an unintended way [26]. We illustrate this effect in Figs. 1(a) and 1(b), where a pulse produced by the waveform generator with amplitude $A$ (purple) reaches the qubit with an amplitude that shows a reduction from a linear behavior (black and gray lines, respectively). This reduction becomes more pronounced with larger input amplitudes and leads to a rotation angle of the qubit state that is smaller than the targeted one. The pulse distortions introduced by the nonlinearity could be corrected with a sample-by-sample nonlinear transformation. However, we show that a simple scaling of the pulse amplitude by the inverse of the response function of the drive line (orange pulses) is enough to mitigate the rotation errors produced by the nonlinearity.

In order to characterize the impact of nonlinearity on the pulse amplitudes, we measure the rotation errors that are introduced by reducing the pulse amplitude proportionally to the amplitude of a $\pi$ rotation; see the dashed black line in Fig. 1(b). To this end, we use the pulse calibration method based on error amplification described in Refs. [47–49], which we extend to arbitrary angles. We refer to this technique as an $N$-pulse calibration method.

To calibrate any rotation errors left over from an initial Rabi calibration of a $\pi$ pulse, we start by initializing the qubit in the superposition state $|+i\rangle = (|g\rangle + i|e\rangle)/\sqrt{2}$, and then we apply $N$ repetitions of $X_{\pi+i\epsilon}$ gates, which have a small error in the rotation angle, $\epsilon$. The pulse sequence is shown above Fig. 1(c). In the ideal case ($\epsilon = 0$), the qubit remains in the equatorial plane of the Bloch sphere, where the ground and excited-state populations are $p_g = p_e = 0.5$. For imperfect gates ($\epsilon \neq 0$), the rotation errors accumulate to produce a deviation of the excited-state population from 0.5 [see the purple points in Fig. 1(c)]. We ignore any small errors in the initialization pulse as they will not be amplified during the pulse sequence.

We extend the methods in Refs. [47–49] by extracting the rotation error per gate from a fit of the excited-state population of the qubit to a master-equation-based model (see Appendix B). After applying the method to a 32-ns pulse with the amplitude calibrated with a single-pulse Rabi experiment, we find around $0.9^\circ$ of rotation error. This corresponds to around 1.7 mV of amplitude error for a pulse amplitude of 335 mV at the output of the arbitrary waveform generator. We shape the pulse using a Gaussian envelope with standard deviation $\sigma = 6.4$ ns, which is truncated at $\pm 2.5 \sigma$. We use a value of $0.68 \sigma$ for the quadrature scaling parameter of the derivative removal by adiabatic gate (DRAG) method [50,51] to cancel the phase errors arising from the ac-Stark shift of the $e_f$ transition during the drive pulse [52,53].

Repeating the measurement after scaling the pulse amplitude based on the fit result, we are able to reduce the observed rotation errors [see the orange points in Fig. 1(c)]. The small residual oscillations exhibited by the orange points are not caused by rotation errors captured by our model, and could originate, for example, from off-resonant driving by higher harmonic frequency components, or from uncompensated nonlinear distortions of the pulse. The latter effects are beyond the scope of the current work.

To determine the impact of nonlinearity for rotation angles smaller than $\pi$, we generalize the technique by using groups of pulses implementing rotation angles that add up to $\pi$. The initial pulse amplitude for these gates is obtained from a proportional reduction with respect to the previously calibrated $\pi$-pulse amplitude.

To calibrate rotation angles up to $\pi/2$, we use a pulse sequence in which each $\pi$ pulse is split into $k$ repetitions of $X_{\pi/k+i\epsilon}$ gates; see the pulse sequence above Fig. 1(d). For an $X_{\pi/2}$ rotation, we find an over-rotation error of around $0.4^\circ$ when setting the pulse amplitude to half of the amplitude of the $\pi$ pulse that was already corrected with the $N$-pulse calibration [purple points and line in Fig. 1(d)]. This value corresponds to an amplitude error of around 0.7 mV.

Similarly, to measure the nonlinearity for rotation angles in the range $\pi/2$ to $\pi$, we apply groups of two $X$ rotations,
where the first is a previously calibrated $\pi/k$ gate, and the second is a pulse implementing a rotation of $\pi - \pi/k + \varepsilon$, which is to be calibrated [indicated by the colored box in the pulse sequence above Fig. 1(e)]. We perform the calibration for an $X_{3\pi/6}$ gate with the pulse amplitude set to 5/6 of the $\pi$-pulse amplitude that was already corrected with the $N$-pulse calibration. We find an under-rotation error of around $-1.0^\circ$, corresponding to an amplitude error of around $1.9 \text{ mV}$.

For both $X_{\pi/2}$ and $X_{3\pi/6}$, the accumulation of rotation errors is prevented by adjusting the pulse amplitude with the value extracted from the corresponding fit (see the orange points). We note that the amplitude of the oscillations shown by the purple points in Figs. 1(d) and 1(e) is captured by our model (see Appendix B).

We reconstruct the nonlinear response curve of the qubit drive line by performing the calibration for 11 distinct rotation angles with a gate duration of 15 ns; see Figs. 1(f) and 1(g). This gate duration corresponds to a $\pi$-pulse amplitude of 730 mV at the output of the arbitrary-waveform generator (AWG), which is around 100 mV below the 1-dB compression point of our up-conversion device. Figures 1(h) and 1(i) shows that a simple linear downscaling of the amplitude with respect to the calibrated $\pi$-pulse amplitude introduces a systematic over-rotation error; see the filled purple region in Fig. 1(i). For rotation angles between $0^\circ$ and $180^\circ$, we model this response with a polynomial correction that preserves the antisymmetry between positive and negative amplitudes:

$$\theta(\tilde{A}) = 180^\circ (1 + b(\tilde{A}^2 - 1) + a(\tilde{A}^4 - 1))\tilde{A}, \quad (1)$$

Here, $\tilde{A} = A/A_\pi$ is the pulse amplitude scaled with respect to the amplitude of a $\pi$ rotation, and $a$ and $b$ are fitting parameters capturing the nonlinearity of the up-conversion circuitry. The degree of the polynomial is chosen heuristically based on the shape of the deviation from the linear scaling shown in Fig. 1(i). Parameters $a$ and $b$ are determined from a fit to the data in Fig. 1(h) (orange line), for which we obtain residuals of the order of $0.2^\circ$. In Sec. V, we use the model in Eq. (1) to implement arbitrary rotation angles.

In the specific case of our setup, we identify the frequency up-conversion device as the dominant source of the nonlinear response by comparing the results measured with the $N$-pulse calibration to those obtained by rerouting
the up-converted drive signal directly to the detection electronics at room temperature (see Appendix C). However, not all the observed nonlinearity is explained by this measurement, indicating some further source of nonlinearity in other parts of the drive line, which could not be avoided by only using this type of calibration. On the other hand, the $N$-pulse calibration method is a calibration tool for the entire microwave drive line that can also be used when the nonlinear distortions have a different origin than that we have identified in our setup.

III. IMPACT OF NONLINEARITY ON SHORT SINGLE-QUBIT GATES

A lower bound on single-qubit gate errors is imposed by the ratio between the gate duration and the coherence time of the system. However, decreasing the gate duration requires an increase in the pulse amplitude. As discussed above, this leads to a larger nonlinear distortion [see Figs. 1(f) and 1(g)], and, hence, more rotation errors if not calibrated out, but also to possible leakage out of the computational subspace.

To quantify the impact of rotation errors produced by drive-line nonlinearity on the performance of single-qubit gates, we measure the total average gate error with randomized benchmarking (RB) [44,54,55]. In addition, we distinguish between coherent and incoherent errors using purity benchmarking (PB) [45,56]. Both methods rely on the application of random sequences sampled from the 24 single-qubit Clifford operations, which we decompose into $X$ and virtual $Z$ gates [57]. However, for purity benchmarking, we apply the tomography pulses $I$ (identity), $X_{\pi/2}$, and $Y_{\pi/2}$ at the end of each sequence to measure the state purity (see the pulse sequence at the top of Fig. 2). We record 4096 acquisitions of each sequence using single-shot readout, and discriminate between the evolution confined to the qubit computational subspace and leakage into the $f$ state with an average readout assignment error over the three qubit states of around 4%. The results are then corrected for readout errors [58,59].

We calculate the average $f$-state population for each sequence length and extract the leakage error per gate $L$ [Fig. 2(d) and Appendix D]. To estimate the gate errors within the qubit subspace, we then renormalize the populations of the qubit subspace so that they add up to one. The total gate error $E$ [Fig. 2(a)] and the incoherent error per gate $E_{\text{inc}}$ [Fig. 2(b)] are extracted independently from the decay of sequence fidelity and state purity, respectively (see Appendix D). Thus, the difference between $E$ and $E_{\text{inc}}$ provides an estimate of the coherent control errors per gate $E_{\text{coh}}$ in the computational subspace [Fig. 2(c)]. We stress that, by renormalizing the qubit subspace before estimating the gate errors, $L$ does not contribute to the total gate error $E$.

In Sec. II we found that not correcting for the effects of nonlinearity results in around 1° of rotation error for a 32-nsec gate, and 3.6° when the length is decreased to 15 ns. Here we use a different device (device B) with better coherence and smaller $\pi$-pulse amplitude (see Appendix A) to measure the coherent errors introduced when the $\pi/2$-pulse amplitude is scaled linearly relative to the $\pi$-pulse amplitude (purple points in Fig. 2). By calibrating the $\pi/2$-pulse amplitude with the $N$-pulse method (orange), we show that these errors can be mitigated. In both cases, the amplitude of the $\pi$ pulse is calibrated with the $N$-pulse method. (a) Total gate error $E$, (b) incoherent error per gate $E_{\text{inc}}$, (c) coherent error per gate $E_{\text{coh}}$, and (d) leakage error per gate $L$ of single-qubit gates as a function of gate duration $\tau$ (bottom axes) and $\pi$-pulse amplitude $A_{\pi}$ (top axes). The black line in (b) is a linear fit to the orange data points. Colors have the same meaning as in the sequence schematic.

For both implementations of the $\pi/2$ pulse, we find an expected linear increase of $E_{\text{inc}}$ with gate duration; see the black line in Fig. 2(b). At 20 ns, we observe an increase in the coherent errors when using a linear scaling of the
amplitude [Fig. 2(c)]. This increase becomes more pronounced at 15 ns, where the coherent gate errors account for about half of the total gate error [Fig. 2(a)], while the leakage remains comparatively low at around $6 \times 10^{-5}$. With the $N$-pulse calibration we obtain coherence-limited gate errors for all gate lengths; see the orange points. This indicates that the measured coherent errors originate primarily from the nonlinearity identified in Sec. II, which becomes more pronounced at shorter gate lengths where larger pulse amplitudes are needed to achieve the same rotation (see the top axes in Fig. 2). In our experiment, the coherent errors produced by this effect are up to one order of magnitude larger than leakage into the noncomputational states.

**IV. IMPACT OF NONLINEARITY ON ARBITRARY-ANGLE SINGLE-QUBIT GATES**

In this section, we show that the $N$-pulse calibration method can be used to implement arbitrary-angle single-qubit gates with coherence-limited performance. Purity benchmarking cannot be used to benchmark arbitrary-rotation single-qubit gates, since these gates are not in the Clifford group. Hence, we use cross-entropy benchmarking (XEB) [20,46,60] with the pulse sequence shown at the top of Fig. 3. Each cycle consists of an $X$ rotation of fixed angle $\theta$ followed by a virtual $Z_{\phi_i}$ gate, with $\phi_i$ sampled randomly from a uniform distribution between 0 and $2\pi$. We generate $K = 200$ random sequences up to 4096 cycles and measure in the $Z$ basis using single-shot readout with 4096 acquisitions.

In cross-entropy benchmarking, the cross-entropy is used as a metric to compare the output qubit-state distribution after $K$ random sequences to that calculated assuming ideal gates [20]. By then fitting the exponential decay of this cross-entropy fidelity versus the number of cycles in the sequence, we extract the average error $E$ of the $X_\theta$ gate (see Appendix E). In addition, the ratio of the variance of the measured distribution and that of the Porter-Thomas distribution allows us to extract the average purity of the final qubit state over the distribution of random sequences [60]. This quantity also decays exponentially with the number of cycles, from which we extract the average incoherent errors per gate, $E_{\text{inc}}$. Then we calculate the coherent errors per gate $E_{\text{coh}}$ as the difference between $E$ and $E_{\text{inc}}$. Finally, as also done for purity benchmarking, we use the measured $f$-state population to estimate the average leakage error per gate $L$, and renormalize the qubit subspace before estimating the gate errors. Thus, $L$ does not contribute to the total gate error $E$.

To minimize the incoherent errors, we choose the shortest, 15-ns gate in Fig. 2, which is also maximally sensitive to errors caused by the nonlinearity of the drive-line components. We benchmark the performance of $X$ rotations with angles $\pi/4$, $\pi/3$, $\pi/2$, $3\pi/4$, $5\pi/6$ using the $N$-pulse calibration method, and obtain coherence-limited gate errors in the computational subspace of around $2.0(1) \times 10^{-4}$ for all rotation angles [see the orange points in Fig. 3(a)]. Thus, we observe no coherent errors within error bars.

For reference, we repeat the measurements for the case where the amplitude is scaled linearly with respect to the calibrated $X$-pulse amplitude (see the purple points), and we find a dependence of the control errors on the rotation angle [Fig. 3(c)], which is similar to the curve in Fig. 1(i). From the $N$-pulse calibration of each rotation angle performed before the corresponding XEB measurement, we calculate the expected control errors per gate, and find good agreement with the data in Fig. 3(c); see the open black squares.

With both implementations, the incoherent errors per gate remain constant across all measured angles [Fig. 3(b)], as expected for a fixed gate duration. However, the leakage shows a small increase with rotation angle from around $2 \times 10^{-5}$ for $\pi/4$ to around $6 \times 10^{-5}$ at $5\pi/6$ [Fig. 3(d)], which is expected as the drive amplitude becomes larger. Nevertheless, the leakage is smaller by one order of magnitude than the coherent errors introduced by the nonlinearity, signaling the latter as the main limitation on overall gate performance.

FIG. 3. Characterization of arbitrary-angle single-qubit gates using cross-entropy benchmarking (XEB). Top: illustration of an XEB pulse sequence. For each angle $\theta$, the amplitudes of the $X_\theta$ gates are obtained either from a linear scaling with respect to the calibrated amplitude of the $X_{5\pi/6}$ gate (purple), or from the $N$-pulse calibration (orange). (a) Total gate error $E$, (b) incoherent error per gate $E_{\text{inc}}$, (c) coherent error per gate $E_{\text{coh}}$, and (d) leakage error per gate $L$ for single-qubit $X_\theta$ gates of various rotation angles $\theta$. The black squares in (c) are calculated from the amplitude corrections obtained with the $N$-pulse calibration.
CONTINUOUSLY PARAMETERIZED SINGLE-QUBIT GATES

Finally, we show that an individual calibration of the amplitude of every arbitrary-rotation single-qubit gate used in a sequence is not required to achieve coherence-limited gates in our experimental setup. To this end, we implement a continuously parameterized gate set using a polynomial scaling of the pulse amplitude obtained from the fit of the data in Fig. 1(h) to the model in Eq. (1).

To benchmark continuously parameterized single-qubit gates, we replace the fixed-angle $X_\theta$ gates in the protocol described in Sec. IV by $X_\theta$ gates, where in each cycle, the angles $\theta$ are sampled fully randomly from a uniform distribution between $0$ and $\pi$. We use 400 random sequences of up to 2048 cycles, and measure each sequence in the $Z$ basis using single-shot readout with 8192 acquisitions. We discriminate between the computational subspace and leakage, and process the data as described in Sec. IV.

When the pulse amplitudes are scaled with the polynomial function, we measure coherence-limited random-angle $X$ gates with an average error of $8.9(4) \times 10^{-4}$ on device A; see the orange data points in Fig. 4 and the exponential fit (line). With a linear amplitude scaling (purple points and line), we find around $3.8 \times 10^{-4}$ of coherent control errors. This number matches the mean value of the gate errors calculated from the rotation errors in Figs. 1(h) and 1(i).

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under the U.S. Army Research Office Grant No. W911NF-16-1-0071, from the EU Flagship on Quantum Technology H2020-FETFLAG-2018-03 project 820363 OpenSuperQ, from the National Centre of Competence in Research Quantum Science and Technology (NCCR QSIT), a research instrument of the Swiss National Science Foundation (SNSF), from the SNFS R’equip grant
We estimate the final qubit excited-state population. We apply Eq. (B2) to the qubit ground state \((y(0) = 0, z(0) = -1)\), and then continue to apply this equation to the qubit state after each pulse in the \(N\) pulse calibration measurement. We estimate the final qubit excited-state population as \(p_e = (1 + z(t))/2\). Denoting the gate duration by \(\tau\), we express \(\Omega = \alpha \tau / \tau\), and extract the dimensionless fraction of a \(\pi\) rotation \(\alpha\) from a fit of the measured data to Eq. (B2).

\[
\begin{align*}
\tilde{y}(t) &= e^{-(3\Gamma_1 + 2\Gamma_\phi) t/4} \left( \cos(\nu t) + \sin(\nu t) (\Gamma_1 - 2\Gamma_\phi) / 4\nu \right) \\
\tilde{z}(t) &= z(t) + \frac{2\Omega \Gamma_1}{\Gamma_1 (\Gamma_1 + 2\Gamma_\phi) + 2\Omega^2} \\
\nu &= \sqrt{\Omega^2 - \frac{(\Gamma_1 - 2\Gamma_\phi)^2}{16}},
\end{align*}
\]

and \(\tilde{y}(0), \tilde{z}(0)\) are the coordinates at the start of the evolution.

We apply Eq. (B2) to the qubit ground state \((y(0) = 0, z(0) = -1)\), and then continue to apply this equation to the qubit state after each pulse in the \(N\) pulse calibration measurement. We estimate the final qubit excited-state population as \(p_e = (1 + z(t))/2\). Denoting the gate duration by \(\tau\), we express \(\Omega = \alpha \tau / \tau\), and extract the dimensionless fraction of a \(\pi\) rotation \(\alpha\) from a fit of the measured data to Eq. (B2).
FIG. 5. (a) Simplified schematic of the room-temperature and measurement setups used to operate the qubits. Intermediate frequency pulses are generated with an AWG and up-converted to the gigahertz range using the conversion device indicated in red. This device contains an IQ mixer and an amplifier, and is the main source of nonlinearity in our drive line. We configure a switch to either send the pulses into the dilution refrigerator via the qubit drive line (path 1) or directly into the detection chain consisting of another IQ mixer, which down-converts the pulses into the frequency band of the detection device (path 2). (b) Measured qubit rotation angle as a function of the applied pulse amplitude relative to the $\pi$-pulse amplitude. The blue points are measured with the $N$-pulse calibration via path 1, and the black points are obtained from the amplitudes of the demodulated pulses measured via path 2 and converted to the same units. (c) Difference between the rotation angles in (b) and a linear interpolation between the last point and the origin. The lines in (b) and (c) are fits to the fifth-order polynomial in Eq. (1) of the main text.

Using the found value of $\alpha$, we calculate the pulse amplitude as $A = \alpha A_{\pi}$, where $A_{\pi}$ is the amplitude of a $\pi$ rotation.

APPENDIX C: SIGNAL NONLINEARITY DUE TO ELECTRONIC COMPONENTS

The response curves shown schematically in Fig. 1(b) and measured in Figs. 1(f)–1(i) are typical for electronic components such as mixers or amplifiers, which become nonlinear above a certain input power [29].

Here we verify that the nonlinearity observed with the $N$-pulse calibration originates mostly from the control electronics used for driving the qubits. We perform the measurements on a different setup than those presented in the main text, but with nominally identical control electronics. We use slightly longer 50-ns DRAG pulses, with a $\sigma = 10$-ns standard deviation of the Gaussian envelope truncated at $\pm 2.5 \sigma$. The pulses are generated at an intermediate frequency of $-100$ MHz before up-conversion to the qubit frequency using a circuit that consists of an IQ mixer and amplifiers [see the red outline in the schematic in Fig. 5(a)]. We choose to either apply the pulses to the qubit and perform the $N$-pulse calibration [path 1 in Fig. 5(a)] or we directly down-convert the pulses to an intermediate frequency of $200$ MHz before detecting the quadratures of the signal (path 2). The signal for path 2 is attenuated by $-20$ dB to ensure that the down-conversion mixer and subsequent amplifiers are operated in their linear regime. We then demodulate the signal to dc, filter out spurious frequency components larger than the spectral width of the pulse, and extract the amplitudes of the resulting pulses. These data are then converted to a rotation angle [see the black squares in Fig. 5(b)] for better comparison to the results obtained with the $N$-pulse calibration via path 1 (blue circles).

We find that the nonlinear scaling of the Rabi angle with the pulse amplitude identified with the $N$-pulse calibration technique originates mostly from our frequency conversion device; see Fig. 5(c), where we plot the difference between the measured data points in Fig. 5(b) and a linear interpolation between the origin and the last data point. However, the results obtained via path 2 systematically overestimate those measured using the qubit by about $0.2^\circ$ at an amplitude ratio of 0.5, showing that the room-temperature calibration presented here cannot fully replace the in situ calibration presented in the main text.

APPENDIX D: PURITY BENCHMARKING

In order to distinguish between incoherent and control errors, we use a simple extension of the randomized benchmarking protocol known as purity benchmarking. Purity benchmarking consists of performing tomography of the qubit state vector after each random sequence of Clifford gates [45,56]. This is illustrated by the schematic in Fig. 6(a), where the orange block consists of the single-qubit tomography pulses $I$ (identity), $X_{\pi/2}$, and $Y_{\pi/2}$. We...
estimate the purity of the state after each random sequence as

\[ \mathcal{P} = \langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2, \]  

(D1)

and compute the mean of this quantity over the random sequences for a fixed number of Clifford gates.

In Fig. 6(b) we show the exponential decay of the \( \langle \sigma_x \rangle \) operator (blue) from the RB measurement of the 40-ns gate in Fig. 2 of the main text calibrated with the \( N \)-pulse technique. We compare this decay to the decay of \( \langle \sqrt{\mathcal{P}} \rangle \) (orange) from the corresponding PB experiment. In the absence of control errors, the two curves should coincide.

For RB and PB, respectively, we fit to

\[ \langle \sigma_x \rangle(m) = A \alpha^m + B, \]

\[ \langle \mathcal{P} \rangle(m) = A' u^m + B', \]  

(D2)

where \( m \) is the number of Clifford gates, \( \alpha \) is the depolarization parameter [44], \( u \) is the unitarity [45], and \( A, A', B, B' \) are fit parameters capturing state preparation and measurement errors. The total average error per gate and the average incoherent error per gate are then respectively calculated as

\[ E = \frac{1 - \alpha}{2N}, \]

\[ E_{\text{inc}} = \frac{1 - \sqrt{u}}{2N}, \]  

(D3)

where \( N = 1.125 \) is the average number of physical gates per Clifford in the HZ decomposition [57]. The virtual Z gates are implemented in software by a qubit phase update, and hence it has perfect fidelity. Subtracting these two quantities gives an estimate of the coherent control error per gate, \( E_{\text{coh}} = E - E_{\text{inc}} \).

The analysis described above is performed after discriminating between the qubit subspace and leakage, and normalizing the qubit populations so that they add up to one. In Fig. 6(c), we show the average \( f \)-state population as a function of the RB sequence length (points). We fit these data to the rate equation [22]

\[ p_f(m) = \frac{l}{l+s}(1 - e^{-(l+s)m}) + p_0 e^{-(l+s)m}, \]  

(D4)

from which we find the average leakage per gate, \( L = l/N \).

Here, \( s \) is the decay rate into the computational subspace and \( p_0 \) is the initial \( f \)-state population.

From the data shown in Fig. 6, we extract a leakage rate per gate of around \( 1.1(1) \times 10^{-5} \), and a total average gate error of \( 4.4(2) \times 10^{-4} \), which is limited by decoherence at around \( 4.1(1) \times 10^{-4} \).

**APPENDIX E: CROSS-ENTROPY BENCHMARKING**

To benchmark the performance of gates that do not belong to the Clifford group, such as arbitrary-rotation single-qubit gates, we use the XEB technique [20,46,60]. In the XEB protocol, sequences of random cycles of gates are applied to a system prepared in a chosen state (typically the ground state). The only requirement to take into consideration when choosing the cycle design is that sequences with enough cycles should fully explore the Hilbert space of the system [46].

The average sequence fidelity is estimated by comparing the distribution of measured outcomes \( Q_{\text{meas}} \) to the calculated, ideal distribution \( Q_{\text{ideal}} \) using the cross-entropy as a metric of overlap. The cross-entropy between two distributions \( Q \) and \( R \) is given by

\[ S(Q, R) = -\sum_i q_i \ln(r_i), \]  

(E1)

where \( q_i, r_i \) are samples from distributions \( Q, R \). For a fixed number of cycles, we estimate the average XEB sequence fidelity of a distribution of random sequence outcomes as [20]

\[ F_{\text{XEB}} = \frac{S(Q_{\text{incoh}}, Q_{\text{ideal}}) - S(Q_{\text{meas}}, Q_{\text{ideal}})}{S(Q_{\text{incoh}}, Q_{\text{ideal}}) - S(Q_{\text{meas}}, Q_{\text{ideal}})}. \]  

(E2)

Here, \( Q_{\text{incoh}} \) is the incoherent distribution where all the bit strings have an equal probability of \( 1/2^n \) (with \( n = 1 \) for one qubit), and \( Q_{\text{meas}} \) and \( Q_{\text{ideal}} \) are the measured and calculated distributions of probabilities \( (p_g, p_e)_i \) describing the qubit state at the end of each random sequence \( i \).
To constrain $F_{\text{XEB}}$ between 0 and 1, we normalize $Q_{\text{meas}}$, $Q_{\text{inc}}$, and $Q_{\text{deal}}$ by the total number of random sequences for a fixed number of cycles.

From the variance of the measured probability distribution, we also estimate the average purity of the final qubit state as [60]

$$P = \frac{\text{Var}(Q_{\text{meas}})}{\text{Var}(Q_{\text{PT}})},$$

(E3)

where $\text{Var}(Q_{\text{PT}}) = (2^n - 1)/(2^{2n}(2^n + 1))$ is the variance of the Porter-Thomas distribution, with $n = 1$ for a single qubit.

The points in Fig. 7(a) show the exponential decays of the XEB fidelity (blue) and the square root of the sequence purity (orange) for the purple point in Fig. 3 of the main text, corresponding to an $X$ gate of angle $5\pi/6$. The lines are fits to

$$F_{\text{XEB}}(m) = A \left(1 - \frac{2^n}{2^n - 1} E\right)^m + B,$$

$$\sqrt{P}(m) = A' \left(1 - \frac{2^n}{2^n - 1} E_{\text{inc}}\right)^m + B',$$

(E4)

from which we estimate the total average errors per cycle $E$, and the average incoherent errors per cycle $E_{\text{inc}}$. Here $m$ is the number of cycles, and $A, A', B, B'$ are fit parameters capturing the state preparation and measurement errors. We then calculate the amount of coherent control errors per cycle as $E_{\text{coh}} = E - E_{\text{inc}}$. Since both cycle designs chosen for this work (see Figs. 3 and 4) contain only one physical gate, $E$, $E_{\text{inc}}$, and $E_{\text{coh}}$ quantify the average errors per gate. From the data set in Fig. 7(a), we find a total average gate error of $6.3(3) \times 10^{-4}$, with a contribution from decoherence errors of $5.6(2) \times 10^{-4}$, giving around $8(3) \times 10^{-5}$ control errors.

The analysis described above is performed after discriminating between the qubit subspace and leakage, and normalizing the qubit populations so that they add up to one. In Fig. 7(b), we show the average $f$-state population (points) as a function of the XEB sequence length, and a fit to the rate equation in Eq. (D4) (line), from which we extract an average leakage rate per gate of around $9.5(7) \times 10^{-5}$.


[38] H. Y. Carr and E. M. Purcell, Effects of diffusion on free precession in nuclear magnetic resonance experiments, Phys. Rev. 94, 630 (1954).


