Experimental implementation of non-Clifford interleaved randomized benchmarking with a controlled-S gate

Shelly Garion, Naoki Kanazawa, Haggai Landa, David C. McKay, Sarah Sheldon, Andrew W. Cross, and Christopher J. Wood

1 IBM Quantum, IBM Research Haifa, Haifa University Campus, Mount Carmel, Haifa 31905, Israel
2 IBM Quantum, IBM Research Tokyo, 19-21 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo, 103-8510, Japan
3 IBM Quantum, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
4 IBM Quantum, Almaden Research Center, San Jose, CA 9512, USA

Hardware efficient transpilation of quantum circuits to a quantum devices native gate set is essential for the execution of quantum algorithms on noisy quantum computers. Typical quantum devices utilize a gate set with a single two-qubit Clifford entangling gate per pair of coupled qubits, however, in some applications access to a non-Clifford two-qubit gate can result in more optimal circuit compositions and also allows more flexibility in optimizing over noise. We demonstrate calibration of a low error non-Clifford Controlled- phase (CS) gate on a cloud based IBM Quantum computing using the Qiskit Pulse framework. To measure the gate error of the calibrated CS gate we perform non-Clifford CNOT-Dihedral interleaved randomized benchmarking. We are able to obtain a gate error of $5.9(7) \times 10^{-3}$ at a gate length 263 ns, which is close to the coherence limit of the associated qubits, and lower error than the backends standard calibrated CNOT gate.

I. INTRODUCTION

Quantum computation holds great promise for speeding up certain classes of problems, however near-term applications are heavily restricted by the errors that occur on present day noisy quantum devices [1]. To run a computation on a quantum processor requires first calibrating a universal gate set – a small set of gates which can be used to implement an arbitrary quantum circuit – which has low error rates, and then transpiling the circuit to this set of gates. This transpilation should be done in a hardware-efficient manner to reduce the overall error by minimizing the use of the highest error gates [2]. Two of the most significant error sources on current devices are incoherent errors due to interactions with the environment, quantified by the coherence times of device qubits, and calibration errors in the gates used to implement a quantum computation [3, 4].

If a gate set could be perfectly calibrated the coherence time of the qubits would set the fundamental limit on error rates without active error correction. Thus the goal of gate calibration is to get as close to the coherence limit as possible. Current quantum hardware typically use a gate set consisting of arbitrary single-qubit rotations and a single entangling two-qubit gate [5]. State of the art single-qubit gate error rates in these systems approach $2 \times 10^{-4}$ [6], where two-qubit gate errors are around $10^{-3}$ [7–9], see also Appendix A. In superconducting qubit systems using fixed-frequency transmon qubits a microwave-only two-qubit entangling gate may be implemented using the cross-resonance (CR) interaction [10]. The CR interaction can be used to implement a high fidelity Controlled-NOT (CNOT) gate [11]. Gate sets with a Clifford two-qubit like CNOT are appealing as a variety of averaged errors in a Clifford gate set can be can be robustly measured using various randomized benchmarking (RB) protocols [4, 12–16].

In some cases it may be favourable to introduce an additional two-qubit gate to a gate set if it enables more hardware efficient compilation of relevant circuits, however this adds the overhead of additional calibration and characterization of the gate errors. One such gate is the Controlled-Phase (CS) gate, which is a non-Clifford two-qubit entangling gate that is universal when combined with the Clifford group [17]. The CS gate is particularly attractive to fixed-frequency transmon qubit systems as it can be implemented using the CR interaction, since it is locally equivalent to $\sqrt{\text{CNOT}}$. This means it can be calibrated using the same techniques as the CNOT gate, but with a shorter gate duration or lower power, potentially leading to a higher fidelity two-qubit gate when calibrated close to the coherence limit. Furthermore the CS gate is a member the CNOT-Dihedral group and can be benchmarked using CNOT-Dihedral randomized benchmarking [17]. Recently an optimal decomposition algorithm for two-qubit circuits into the Clifford + CS gates was developed [18]. This method minimizes the number of non-Clifford (CS) gates, which is important in the context of quantum error correction as non-Clifford gates require additional resources such as magic-state distillation to prepare fault-tolerantly [19]. However, in non-fault tolerant near term devices it is often preferable to minimize the total number of two-qubit gates in a decomposition rather than non-Clifford gates. An optimal decomposition for gates generated by the CNOT-Dihedral in terms of the number of CNOT and CS gates has also recently been developed [20]. Another example is the Toffoli gate which can be decomposed into 6 CNOT gates and single qubit gates, but requires only 5 two-qubit gates in its decomposition if the CS and CS$^{-1}$...
gates are also available [21].

In this work we calibrate CS and CS$^{-1}$ gates of varying durations on an IBM Quantum system and benchmark the gate error rates by performing the first experimental demonstration of interleaved CNOT-Dihedral random-
ized benchmarking. For specific gate durations we are able to obtain a high-fidelity CS gate approaching the coherence limit, which due to the shorter CR interaction time results in a lower error rate than can be obtained for a CNOT gate. In addition to RB we also compute the average gate error of the CS gate using two-qubit quantum process tomography (QPT) and compare to the values obtained from RB. Pulse-level calibration was done using Qiskit Pulse [22], and the RB and QPT experiments were implemented using the open source Qiskit computing software stack [23] through the IBM Quantum cloud provider.

II. CNOT-DIHEDRAL RANDOMIZED BENCHMARKING

We describe the protocol for estimating the average gate error of the CS gate using interleaved CNOT-Dihedral Randomized Benchmarking, which is a natural generalization of the CNOT-Dihedral RB procedure described in [17] with interleaved RB [13] to estimate individual gate fidelities for the CS gate

$$CS = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & i \end{pmatrix}.$$ 

In the following we let $G$ denote the CNOT-Dihedral group on $n$ qubits, $g \in G$ denote a unitary element of $G$, and $I, X, Y, Z$ denote the single-qubit Pauli matrices.

Step 1: Standard CNOT-Dihedral benchmarking.

Randomly sample $l$ elements $g_{j_1}, \ldots, g_{j_l}$ uniformly from $G$, and compute the $(l+1)$th element from the inverse of their composition, $g_{j_{l+1}} = (g_{j_l} \circ \cdots \circ g_{j_1})^{-1}$. Denote by $j_l$ the $l$-tuple $(j_1, \ldots, j_l)$. For each composition, we prepare an input state $\rho$, and apply the composition of the $l+1$ gates that ideally would be

$$S_{j_l} := g_{j_{l+1}} \circ g_{j_l} \circ \cdots \circ g_{j_1},$$

and then measure the expectation value of an observable $E$.

Assuming each gate $g_i$ has an associated error $\Lambda_i(\rho)$, the sequence $S_{j_l}$ is implemented as

$$\tilde{S}_{j_l} := \Lambda_{j_{l+1}} \circ g_{j_{l+1}} \circ (\bigodot_{i=1}^{l} [\Lambda_i \circ g_{j_i}]),$$

The expectation value of $E$ is $(E)_{j_l} := Tr[E \tilde{S}_{j_l}(\rho)]$. Averaging this overlap over $K$ independent sequences of length $l$ gives an estimate of the average sequence fidelity

$$F_{seq}(l, E, \rho) := Tr[E \tilde{S}_{l}(\rho)]$$

where $\tilde{S}_{l}(\rho) := \frac{1}{K} \sum_{j} \tilde{S}_{j}(\rho)$ is the average quantum channel.

We decompose the input state and this final measurement operator in the Pauli basis $P$ (an orthonormal basis of the $n$-qubit Hermitian operators space, constructed of single-qubit Pauli matrices). This gives $\rho = \Sigma_{p \in P} p \rho p$ and $E' = \Sigma_{p \in P} p E p$. Given that the gate errors are close to the average of all errors [17], the average sequence fidelity is

$$F_{seq}(l, E, \rho) = A_Z \alpha_Z^l + A_R \alpha_R^l + \epsilon_l$$

where $A_Z = \Sigma_{p \notin Z} p E p$ and $A_R = \Sigma_{p \notin R} p E p$, with $Z$ being tensor products of $Z$ and $I$ gates.

Each of the two exponential decays $\alpha_Z^l$ and $\alpha_R^l$ can be observed by choosing appropriate input states. For example, if we choose the input state $|0 \ldots 0\rangle$ then $F_{seq} = \epsilon_l + A_R \alpha_R^l$ where $A_R = \Sigma_{p \notin R} p E p$. On the other hand, if we choose $|+\cdots+\rangle$ then $F_{seq} = \epsilon_l + A_Z + \alpha_Z^l$ where $A_Z = \Sigma_{p \notin Z} p E p$. With $X$ tensor products of $X$ and $I$ gates.

The channel parameters $\alpha_Z$ and $\alpha_R$ can be extracted by fitting the average sequence fidelity to an exponential.

$$\alpha = (\alpha_Z + 2^n \alpha_R)/(2^n + 1)$$

and the corresponding average gate error is given by

$$r = (2^n - 1)(1 - \alpha)/2^n.$$ 

Step 2: Interleaved CNOT-Dihedral sequences.

Choose a sequence of unitary gates where the first element $g_{j_1}$ is chosen uniformly at random from $G$, the second is always chosen to be $g$, and alternate between uniformly random elements from $G$ and fixed $g$ up to the $l$-th random gate. The $(l+1)$th element is chosen to be the inverse of the composition of the first $l$ random gates and $l$ interlaced $g$ gates, $g_{j_{l+1}} = (g \circ g_{j_l} \circ \cdots \circ g \circ g_{j_1})^{-1}$. We adopt the convention of defining the length of a sequence by the number of random gates $l$.

For each sequence, we prepare an input state $\rho$, apply

$$\nu_{j_l} := g_{j_{l+1}} \circ g \circ g_{j_l} \circ \cdots \circ g \circ g_{j_1}$$

and measure an operator $E$.

Assuming that the gate $g$ has an associated error $\Lambda_g(\rho)$ and that each gate $g_i$ has an associated error $\Lambda_i(\rho)$, the sequence $\nu_{j_l}$ is implemented as

$$\tilde{\nu}_{j_l} := \Lambda_{j_{l+1}} \circ g_{j_{l+1}} \circ (\bigodot_{i=1}^{l} [\Lambda_g \circ g \circ \Lambda_i \circ g_{j_i}]).$$
The overlap with $E$ is $\text{Tr}[E\hat{\nu}_3(\rho)]$. Averaging this overlap over $K$ independent sequences of length $l$ gives an estimate of the new sequence fidelity

$$F_{\text{seq}}(l, E, \rho) := \text{Tr}[E\hat{\nu}(\rho)]$$

where $\hat{\nu}(\rho) := \frac{1}{K} \sum_k \hat{\nu}_3(\rho)$ is the average quantum channel.

Similarly to Step 1, we fit $F_{\text{seq}}(l, E, \rho)$ and obtain the depolarizing parameter $\alpha$, according to Eq. (3). Using the values obtained for $\alpha$ and $\alpha_g$, the gate error of $\Lambda_g$, which is given by

$$r_g^{rb} = \frac{(2^n - 1)(1 - \alpha_g/\alpha)}{2^n},$$

and must lie in the range $[r_g^{rb} - \epsilon, r_g^{rb} + \epsilon]$, where $\epsilon$ is estimated in [13] Eq. (5). Note that one has to be careful in interpreting the results of an interleaved experiment, as in some cases $\epsilon$ might be large compared to $r_g^{rb}$.

III. IMPLEMENTING THE CONTROLLED-S GATE

We calibrate CS gates of varying gate durations using Qiskit Pulse and measure the average gate error using the interleaved CNOT-Dihedral RB protocol in II. We use the CR pulse sequence as a generator of two-qubit entanglement [10, 24]. The CR pulse is realized by irradiating one (control) qubit with a microwave pulse at the transition frequency of another (target) qubit. The stimulus drives the quantum state of the target qubit with the direction of rotation depending on the quantum state of the control qubit. This controlled rotation is used to create two-qubit entangling gates such as CNOT and CS.

The two-qubit system driven by the CR pulse with amplitude $A$ and phase $\phi$ can be approximated by an effective block-diagonal time-independent Hamiltonian [25, 26]

$$\overline{H}_{\text{CR}}(A, \phi) = \sum_{P=I,X,Y,Z} \frac{\omega_{ZP}(A, \phi)}{2} Z \otimes P$$

$$+ \sum_{Q=X,Y,Z} \frac{\omega_{IQ}(A, \phi)}{2} I \otimes Q,$$

where the qubit ordering is control $\otimes$ target, and $\omega_{ZP}$ and $\omega_{IQ}$ represent the interaction strength of the corresponding Pauli Hamiltonian terms. In the absence of noise, the ideal CR evolution for a constant-amplitude pulse is written as an unitary operator

$$U_{\text{CR}}(A, \phi) = \exp \left\{ -it_{\text{CR}} \overline{H}_{\text{CR}}(A, \phi) \right\},$$

where $t_{\text{CR}}$ is the length of the CR pulse. We also define the unitary operator created by an arbitrary two-qubit generator as

$$[BC]_\theta = \exp \left\{ -i\theta \frac{1}{2} (B \otimes C) \right\}$$

where $B, C$ are arbitrary single qubit operators, and we use $[BC] \equiv [BC]_\pi$.

As can be seen by examining Eq. (7), the CR pulse induces three entangling interaction terms ($ZX, ZY$, and $ZZ$), in addition to potentially many unwanted local rotations with different amplitudes. By appropriately calibrating the phase of the CR drive $\phi$, the $ZX$ term is the dominant term among the interactions and is the key term for executing two-qubit gates in this system. As with the standard CNOT gate, we can compose a CS gate by isolating the $ZX$ interaction with a refocusing sequence and single qubit pre- and post-rotations:

$$CS = [IH] \circ [IX]_{\frac{\pi}{2}} \circ [ZI]_{\frac{\pi}{2}} \circ [ZX]_{-\frac{\pi}{2}} \circ [IH],$$

where $H$ is the Hadamard operator. As shown in Eq. (10), we need to develop the calibration procedure to find an amplitude $A$ and a phase $\phi$ where $|\omega_{ZX}|_{\text{CR}} = \pi/4$ and the other terms become zero. The CR Hamiltonian includes a large $ZI$ term as a result of the off-resonant driving of the control qubit; $IX, ZX$ and $IZ$ can also be large for transmon qubits [25]. However, the strengths of $ZZ$ and $IZ$ terms are expected to be negligible weak in our device. We note that both $ZI$ and $IX$ terms commute with the $ZX$ term of interest, while $ZI$ and $ZX$ terms anti-commute with the inversion of the control qubit $XI$. In addition, the $ZI$ term is the even function and both $IX$ and $ZX$ terms are odd functions of the drive amplitude $A$. Accordingly, we can effectively eliminate the impact of those unwanted terms with the two-pulse echoed CR sequence [27] expressed as

$$U_{\text{echo}}(A, \phi) = [XI] \circ U_{\text{CR}}(-A, \phi) \circ [XI] \circ U_{\text{CR}}(A, \phi).$$

This sequence consists of two CR pulses with opposite drive amplitude, each one followed by a $\pi$-rotation refocus pulse $XI$ on the control qubit. Here we also assume the negligible impact of the $IY$ term which is generally introduced by the physical crosstalk between the control and the target qubit [11].

A. Gate Calibration and Benchmarks

To experimentally implement the CS and $\text{CS}^I$ gates we use the 27 qubit IBM Quantum system $\text{ibmq-paris}$ with fixed-frequency and dispersively coupled transmon qubits. Qubit 0 and the qubit 1 of this system are assigned as the control and the target qubit, respectively. The resonance frequency and anharmonicity of the control (target) qubit are 5.072 (5.020) GHz and -336.0 (-321.0) MHz.

The pulses realized in practice are not constant-amplitude pulses, rather the amplitude is increased and decreased smoothly. We implement the CR pulse as a flat top Gaussian, with flat-top length $\tau_{\text{sq}}$, and Gaussian rising and falling edges each with length $\tau_{\text{edge}}$ ($t_{\text{CR}} = \tau_{\text{sq}} + 2\tau_{\text{edge}}$). We use a constant Gaussian edge with $\tau_{\text{edge}} = 28.16$ ns with 14.08 ns standard deviation and
vary the length of the duration of the square flat-top pulse \( \tau_{sq} \). The minimum pulse duration is \( \tau_{sq} = 0 \) ns, yielding a pure Gaussian shape. The overhead of single-qubit gates in the echoed CS sequence in Eq. (10) for the ibmq_paris backend is 106.7 ns, giving a total echoed CS gate time of \( \tau_{CS} = 2\tau_{CR} + 106.7 \) ns. The single-qubit gates are optimized by merging consecutive rotations using the Qiskit circuit transpiler with optimization_level = 1 followed by conversion to a pulse schedule [22].

We performed calibration to a CR rotation angle \( \omega_{2Z}(A, \phi)\tau_{CR} \approx \pi/4 \) for different values of \( \tau_{sq} \). This was done by first performing a rough calibration of \( (A, \phi) \) by scanning those parameters, followed by the closed-loop fine calibration with standard error amplification sequences (see Appendix B for details). The calibrated pulse schedule of the CS gate with \( \tau_{sq} = 21.3 \) ns (\( \tau_{CS} = 263.1 \) ns) is shown in Fig. 1(a).

The average gate error of the calibrated CS gate is evaluated by using the interleaved CNOT-Dihedral RB with 10 sequence lengths \( l \in (1, 5, 10, 20, 30, 50, 75, 100, 125, 150) \), and 10 samples for each \( l \). Each experiment is executed 1024 times for both input states \( \ket{00} \) and \( \ket{++} \) both with and without interleaving the CS gate. An example of measured RB decay curves for \( \tau_{sq} = 21.3 \) ns are shown in Fig. 1(b). The exponential fit of the decay curves yields \( \alpha = 9.78(1) \times 10^{-1} \) and \( \alpha_g = 9.73(1) \times 10^{-1} \), giving an estimated average gate error of the CS gate of \( r_{gb} = 5.2(7) \times 10^{-3} \). According to [13], the theoretical bound of the error is calculated to be \([0, 2.8 \times 10^{-2}]\). In addition to RB we also perform quantum process tomography (QPT) [29] and compute the average gate fidelity from the reconstructed process. QPT was done using maximum likelihood estimation with the tomography module of Qiskit Ignis [30], using a preparation basis of \( \{0\}, \{1\}, \{\pm\}, \{\mp\} \) and measurement basis of \( \{X, Y, Z\} \) for each qubit. We performed 1024 repetitions (shots) per basis configuration and correct for measurement errors using the readout error mitigation technique [31].

The average gate error calculated from the tomographic fit for \( \tau_{sq} = 21.3 \) ns was \( r_{gb}^{\text{exp}} = 8.8 \times 10^{-3} \) which is comparable to the value estimated from the interleaved CNOT-Dihedral RB experiment.

### B. Gate Duration Dependence

We perform the same calibration and benchmarking procedures for different flat-top width \( \tau_{sq} \) from 0 ns to 355.6 ns (\( \tau_{CS} \) from 219.3 ns to 930.5 ns) and measure the average gate errors by both the interleaved CNOT-Dihedral RB experiment and QPT. In this experiment, we use a reduced set of RB sequence lengths \( l \in (1, 10, 25, 50, 100, 150) \) to reduce the total number of experiments while keeping the accuracy of the estimated gate error high.

We measure the qubit coherence times \( T_1 \) and \( T_2 \) with relaxation and Hahn echo sequences [32], respectively, to monitor the stability of physical properties of qubits. These experiments are inserted immediately before each calibration experiment and yield coherence times of \( T_1 = 59.6 \pm 15.6 \) (77.1 \( \pm \) 7.3) \( \mu \)s and \( T_2 = 92.5 \pm 22.1 \) (69.1 \( \pm \) 4.8) \( \mu \)s for the control (target) qubit during the experiment. Here, the error bars correspond to the standard deviation over the duration of the whole set of calibration and benchmarking experiments, presented in Fig. 2. The device was accessed via the cloud through a fair-share queuing model used in IBM Quantum systems. The time in between experiments was about 168 minutes on average, thus the experiment could be subject to some parameter fluctuations due to noise.

![FIG. 1. The CS gate realized with a closed-loop calibration. (a) Pulse schedule with the flat-top width \( \tau_{sq} = 21.3 \) ns. The schedule consists of two CR pulses \( CR_1 \) and \( CR_2 \) on the ControlChannel \( u_0 \) with echo pulses \( X(\pi) \) applied on DriveChannel \( d_0 \) of the control qubit. Local gates in Eq. (10) are also applied to the DriveChannel \( d_1 \) of the target qubit. Modulation frequencies of \( d_0 \) and \( d_1 \) are locked at the resonance frequencies of the control and the target qubits, respectively. The \( u_0 \) channel is associated with the control qubit with the frame of the target qubit to drive CR pulses at appropriate phase and frequency. A Circular arrow of \( VZ(\theta) \) represents the virtual-Z rotations with rotation angle \( \theta \). (b) CNOT-Dihedral interleaved RB. Dotted lines show fit curves of the ground state population measured by standard RBs in \( \ket{00} \) and \( \ket{++} \) basis, while solid lines show fits of interleaved RB. Triangle and cross symbols show raw experiment data of 10 different random circuits.](image-url)
FIG. 2. Average gate errors as a function of the flat-top width of the CR pulse \( \tau_{sq} \) estimated by different benchmark techniques. The corresponding total gate time \( \tau_{CS} \) is shown in the top axis. Blue circles and red triangles represent \( r_{q}^{rb} \) and \( r_{q}^{opt} \), respectively. The Green dotted line shows the theoretical lower bound of the average gate error calculated by the total gate time \( \tau_{CS} \) and the average \( T_{1} \) and \( T_{2} \) values of the qubits during the experiment. The filled area represents the coherence limit with \( T_{1} \) and \( T_{2} \) values with variance of 1 \( \sigma \). See text for a detailed discussion.

with a long characteristic time [33].

Nevertheless, as Fig. 2 shows, our calibration method provides highly accurate results and allows to approach the coherence limit for appropriately chosen gate times. It can also be seen that \( r_{q}^{rb} \) and \( r_{q}^{opt} \) show a similar trend as a function of \( \tau_{sq} \). This dependence on \( \tau_{sq} \) agrees well with the slope predicted by the coherence limit for \( \tau_{sq} \gtrsim 21.3 \) ns. We emphasize that \( r_{q}^{rb} \) gives a more robust estimate of the gate error than \( r_{q}^{opt} \) as it is not as sensitive to state preparation and measurement errors. The interleaved CNOT-Dihedral RB experiment also requires only 24 circuit executions per single error measurement, while the two-qubit QPT requires 144 circuit executions with the readout error mitigation. The smaller experimental cost to measure \( r_{q}^{rb} \) enables us to average the result over 10 different random circuits, which is empirically sufficient to obtain a reproducible outcome, at a practical queuing time with \textit{ibmq-paris}.

The nearly stable offset of \( r_{q}^{rb} \) from the coherence limit possibly indicates the presence of coherent errors due to imperfection of calibration. The measured \( r_{q}^{opt} \) obtained consistantly smaller error values than \( r_{q}^{rb} \).

In the region \( \tau_{sq} \lesssim 21.3 \) ns, both gate errors show a significant increase from the coherence limit. In this regime the drive amplitude of the CR pulse rapidly increases in order to guarantee that the total accumulated rotation angle is \( \pi/4 \) for shorter \( \tau_{CR} \). The amplitude of crosstalk \( \sqrt{\omega_{IX}^{2} + \omega_{IY}^{2}} \) measured at \( \tau_{sq} = 0 \) ns is 176.2 kHz, while one at \( \tau_{sq} = 355.6 \) ns is 19.4 kHz. Although the \( IX \) term is refocused and has negligible contribution, the remained \( IY \) term can still impact on the measured gate errors. Thus, at \( \tau_{sq} = 0 \) we calibrate a CS gate with a compensation tone on the target qubit to suppress the physical crosstalk between qubits (see Appendix C for details). The calibrated pulse sequences with and without the compensation tone yield \( r_{q}^{rb} \) of \( 2.1(3) \times 10^{-2} \) and \( 2.2(2) \times 10^{-2} \), respectively. These comparable results indicate the physical crosstalk is relatively suppressed in this quantum device and other noise sources are dominant for \( \tau_{sq} \lesssim 21.3 \) ns. For example, at high power the perturbation theory used to obtain the average CR Hamiltonian may break down, and hence also calibration scheme based on this decomposition.

The reasons for imperfection of two-qubit gates in superconducting qubits have been investigated and associated with various mechanisms such as nonideal signal generation, residual ZZ coupling, CR-induced ZZ interaction [34–36], and leakage to the higher energy levels [37, 38]. Although a further analysis of the error mechanisms in this regime of high-power pulses is beyond the scope of this study, initial results indicate that coherent population transfer out of the two-qubit manifold into the higher levels, and ZZ interaction terms, are not the relevant mechanisms [39]. At the same time, the coherence limit can be further lowered by reducing the time spent on single-qubit gates. At \( \tau_{sq} = 21.3 \) ns with the minimum \( r_{q}^{rb} \) of \( 5.9(7) \times 10^{-3} \), the refocusing pulse and local rotations occupy 40% of the total gate time \( \tau_{CS} \), yielding a non-negligible impact on the gate error.

IV. CONCLUSION

We have demonstrated calibration of a high fidelity non-Clifford CS gate on 27 qubit IBM Quantum system \textit{ibmq-paris}. This gate is not currently included in the standard basis gates of IBM Quantum systems, and it was calibrated and benchmarked entirely using open source software available in Qiskit. Since the CS gate is non-Clifford, robust characterization of the average gate error cannot be done using standard RB. To benchmark performance of the non-Clifford gate we performed the first experimental demonstration of two-qubit interleaved CNOT-Dihedral RB, which allow efficient and robust characterization of a universal gateset containing the CS gate.

We obtained a minimal gate error of \( 5.9(7) \times 10^{-3} \) with appropriately shaped echoes and a total gate time of 263.1 ns. The gate error reported for the standard two-qubit CNOT gate provided by \textit{ibmq-paris} is \( 1.3 \times 10^{-2} \). Thus the presented CS gate error is comparable with half the CNOT error. By performing RB and QPT for a variety of gate lengths we were also able to study the performance of the CS gate in different regimes and observed a break down in performance if gate lengths were reduced below the best value obtained for 263.1 ns. This is consistent with previous literature on CNOT calibration using the cross-resonance interaction in the high power regime.
The expansion of the native two-qubit gateset of a Cloud quantum device with additional low error calibrated gates allows for improved hardware efficient translation of quantum circuits. This is important for executing quantum algorithms on noisy quantum devices without error correction, and for reducing the error correction overhead when fault-tolerant devices with active error correction are available.

ACKNOWLEDGEMENTS

We thank Ken Xuan Wei for discussion about the CS gate calibration and providing us with the pulse sequence to investigate local coherent errors. DCM and SS acknowledge partial support from the ARO under Contract No. W911NF-14-1-0124.

[1] J. Preskill, Quantum 2, 79 (2018).
[2] A. D. Córcoles, A. Kandala, A. Javadi-Abhari, D. T. McClure, A. W. Cross, K. Temme, P. D. Nation, M. Steffen, and J. Gambetta, arXiv preprint arXiv:1910.02894 (2019).
[3] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Phys. Rev. A 93, 012301 (2016).
[4] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, New Journal of Physics 17, 113020 (2015).
[5] D. C. McKay, T. Alexander, L. Bello, M. J. Biercuk, L. Bishop, J. Chen, J. M. Chow, A. D. Crócoles, D. J. Egger, S. Filipp, J. Gomez, M. Hush, A. Javadi-Abhari, D. Moreda, P. Nation, B. Paulovics, E. Winston, C. J. Wood, J. Wootton, and J. M. Gambetta, “Qiskit backend specifications for OpenQASM and OpenPulse experiments,” (2018), arXiv:1809.03452 [quant-ph].
[6] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Phys. Rev. A 96, 022320 (2017).
[7] J. M. Gambetta, J. M. Chow, and M. Steffen, npj Quantum Information 3 (2017), 10.1038/s41534-016-0004-0.
[8] Y. Xu, J. Chu, J. Yuan, J. Qiu, Y. Zhou, L. Zhang, X. Tan, Y. Yu, S. Liu, J. Li, F. Yan, and D. Yu, “High-fidelity, high-scalability two-qubit gate scheme for superconducting qubits,” (2020), arXiv:2006.11860 [quant-ph].
[9] B. Foxen, C.Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, S. Boixo, D. Buell, B. Burkett, Y. Chen, R. Collins, E. Farhi, A. Fowler, C. Giudici, M. Giustina, R. Graff, M. Harrigan, T. Huang, S. V. Isakov, E. Jefrey, Z. Jiang, D. Kafri, K. Kechedzhi, P. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, E. Lucero, J. McClean, M. McEwen, X. Mi, M. Mohseni, J. Y. Mutus, O. Naaman, M. Neeley, M. Niu, A. Petukhov, C. Quintana, N. Rubin, D. Sank, V. Smelyanskiy, A. Vainsencher, T. C. White, Z. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, “Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms,” (2020), arXiv:2001.08543 [quant-ph].
[10] J. M. Chow, A. Córcoles, J. M. Gambetta, C. Rigetti, B. Johnson, J. A. Smolin, J. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, et al., Physical review letters 107, 080502 (2011).
[11] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Phys. Rev. A 93, 060302 (2016).
[12] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. 106, 180504 (2011).
[13] 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, Phys. Rev. Lett. 109, 080505 (2012).
[14] C. J. Wood and J. M. Gambetta, Phys. Rev. A 97, 032306 (2018).
[15] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, Phys. Rev. Applied 6, 064007 (2016).
[16] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, arXiv preprint arXiv:2003.02354 (2020).
[17] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, npj Quantum Information 2 (2016), 10.1038/npjqi.2016.12.
[18] A. N. Glaudell, N. J. Ross, and J. M. Taylor, “Optimal two-qubit circuits for universal fault-tolerant quantum computation,” (2020), arXiv:2001.05997 [quant-ph].
[19] S. Bravyi and A. Kitaev, Physical Review A 71 (2005), 10.1103/physreva.71.022316.
[20] S. Garion and A. W. Cross, “On the structure of the cnot-dihedral group,” (2020), arXiv:2006.12042 [quant-ph].
[21] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A 52, 3457 (1995).
[22] T. Alexander, N. Kanazawa, D. J. Egger, L. Capelluto, C. J. Wood, A. Javadi-Abhari, and D. McKay, arXiv preprint arXiv:2004.06755 (2020).
[23] H. Abraham and et al, “Qiskit: An open-source framework for quantum computing,” (2019).
[24] C. Rigetti and M. Devoret, Physical Review B 81, 134507 (2010).
[25] E. Magesan and J. M. Gambetta, “Effective hamiltonian models of the cross-resonance gate,” (2018), arXiv:1804.04073 [quant-ph].
[26] M. Malekakhlagh, E. Magesan, and D. C. McKay, “First-principles analysis of cross-resonance gate operation,” (2020), arXiv:2005.00133 [quant-ph].
[27] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. Plourde, and M. Steffen, Physical Review A 87, 030301 (2013).
[28] A. Patterson, J. Rahamim, T. Tsunoda, P. Spring, S. Jебari, K. Ratter, M. Mergenthaler, G. Tancredi, B. Vlastakis, M. Esposito, et al., Physical Review Applied 12, 064013 (2019).
[29] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A 77, 032322 (2008).
[30] IBM, “Qiskit Ignis,” https://github.com/Qiskit/qiskit-ignis (2019), version 0.3.
[31] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, “Mitigating measurement errors in multi-qubit experiments,” (2020), arXiv:2006.14044 [quant-ph].
The two-qubit CNOT gate error between these qubits is shown in graph edges. Error values are represented by color maps shown in the bottom.

**Appendix B: Calibrating CS Gate**

The single qubit gates used for the echo sequence and local rotations are provided by ibmq.paris. We calibrate the CR pulse amplitude $A$ and its phase $\phi$ by the rough parameter scan followed by the closed-loop calibration. These parameters are determined based on the two-pulse echoed CR sequence $U_{echo}$ shown in Eq. (11). This approach simplifies the calibration, namely, we don’t need to take non-negligible $ZI$ and $IX$ terms into account when we fit the experimental results for calibration parameters. Calibrated sequence $U_{echo} \sim [ZX]_{t/2}$ is used to realize the $CS$ with local rotations shown in Eq. (10).

### 1. Rough Parameter Scan

We initialized both qubits in the ground state and perform a rough scan of the CR pulse amplitude with the pulse schedule:

$$S^{scan}_{A}(A) \equiv U_{echo}(A, 0).$$

The schedule is followed by the measurement of the target qubit in the Z-basis. The sinusoidal fit for the measured population of the target qubit with $S^{scan}_{A}$ with different $A$ gives an estimate of the CR amplitude $A_0$ where the angle of controlled rotation is approximately $\pi/4$. A typical experimental result for $\tau_{sq} = 21.3$ ns is shown in Fig. 4(a).

By using this $A_0$, we scan the CR phase with two pulse schedules $S^{scan}_{\phi_g}$ and $S^{scan}_{\phi_e}$:

$$S^{scan}_{\phi_g}(\phi) \equiv [IZ]_{\frac{t}{2}} \circ [IX]_{\frac{t}{2}} \circ U_{echo}(A_0, \phi)^2,$$

$$S^{scan}_{\phi_e}(\phi) \equiv [IZ]_{\frac{t}{2}} \circ [IX]_{\frac{t}{2}} \circ U_{echo}(A_0, \phi)^2 \circ [XI].$$

---

**Appendix A: Basis Gate Information**

In this paper all experiments are performed via cloud access to IBM Quantum system ibmq.paris. The backend provider calibrates single-qubit and two-qubit basis gates on a regular basis and provides pulse schedules and gate errors to users. The gate error distribution at the time of experiment (2020-05-20 05:48 UTC) is shown in Fig. 3. The averaged single-qubit gate error is $5.0 \times 10^{-4}$, while that of two-qubit gates is $1.4 \times 10^{-2}$. The single-qubit gate error of the qubit 0 and 1, which are use in the CS gate, are $4.0 \times 10^{-4}$ and $3.7 \times 10^{-4}$, respectively. The two-qubit CNOT gate error between these qubits is $1.3 \times 10^{-2}$.
The schedule $S_{\text{scan}}^{\phi}$ ($S_{\text{scan}}^{\phi}$) drives the echo sequence $U_{\text{echo}}(A_0, \phi)$ twice with the control qubit of the ground (excited) state. Note that the last two operations correspond to the projection into $Y$-basis for the following measurement. The flip of the state of the control qubit leads to the controlled rotation of the target qubit state with opposite direction as illustrated in Fig. 4(b). This opposite rotation of $\pi/2$ around an azimuthal angle $\theta = \theta_0 - \phi$ of the target qubit Bloch sphere yields measured outcome of $\mp 1$ for $S_{\phi g}^{\text{scan}}$ and $S_{\phi e}^{\text{scan}}$, respectively, at the optimal phase $\phi = \phi_0$ where $\theta = 0$. Here $\theta_0$ is the phase offset from the unknown transfer function of the coaxial cable assembly [42]. The phase $\phi_0$ gives a rough estimate of the CR phase where the $ZX$ term of interest is maximized while the unwanted $ZY$ term is eliminated.

2. Closed-loop Fine Calibration

We use the roughly estimated parameters $(A_0, \phi_0)$ as an initial guess of closed-loop calibrations. We first optimize the CR pulse amplitude with following experiment:

$$S_{\text{fine}}^{\phi}(A) \equiv U_{\text{echo}}(A, \phi_0)^4 N \circ [I X]_{\vec{z}},$$

where $N$ is number of repeated sequences. This schedule prepares the target qubit in the superposition state and repeats the echo sequence $4N$ times to apply a controlled rotation of $N \pi$. Because the initial guess of $A_0$ is estimated by the parameter scan in the coarse precision with a finite error $\delta_A$, repeating $S_{\text{fine}}^{\phi}$ for different $N$ can accumulate $\delta_A$ and this error appears as over rotation from the superposition state, as shown in Fig. 4(c). The fit for the over rotation as a function of $N$ yields precise estimate of $\delta_A$, and we iteratively update the initial guess to optimize the CR pulse amplitude to $A_1$ where $\delta_A \sim 0$. Here we use $N = 0, 1, 2, ..., 9$ and repeat updating the CR amplitude until the over rotation error reaches below the threshold value of $10^{-3} \pi$ rad.

With the optimized amplitude $A_1$, we tune the CR phase with following experiment:

$$S_{\phi}^{\text{fine}}(\phi) \equiv [I Y]_{\vec{z}} \circ (U_{\text{echo}}(A_1, \phi_0)^4 N \circ [I X]_{\vec{z}}).$$

This sequence also accumulates the small phase error $\delta_\phi$ as function of $N$. We iteratively update the CR phase until the same threshold value with the amplitude calibration to obtain the optimal phase $\phi_1$ where $\delta_\phi \sim 0$.

Appendix C: Crosstalk Estimation

The unwanted local rotation terms $IX$ and $IY$ can be simultaneously amplified with the following sequence combined with the XY-4 dynamical decoupling [43] on the control qubit:

$$S_{\text{XY4}}^{\phi} \equiv ([I Y] \circ U_{CR} \circ [X I] \circ U_{CR})^{2N},$$
where $U_{CR} = U_{CR}(A_1, \phi_1)$. Here, the CR pulse with the same sign is repeatedly applied while changing the state of control qubit. This pulse sequence refocuses (and hence eliminates) controlled rotation terms such as $ZX$ and $ZY$, allowing us to precisely estimate the strength of weak local rotation terms $\sqrt{\omega_{IX}^2 + \omega_{IY}^2}$, amplified in the absence of strong two-qubit interactions.

This technique can be used to calibrate a compensation tone that eliminates the $IY$ term caused by the physical crosstalk between qubits [11]. The compensation tone is applied to the drive channel of the target qubit $d_1$, in parallel with $U_{CR}$. This single-qubit pulse is shaped as a flat-top pulse with Gaussian edges of identical duration as the $U_{CR}$ pulse, with its own calibrated amplitude and phase $(A', \phi')$. First, we repeat $S^{xy4}$ for $N = 0, 2, 4, ..., 32$ without the compensation tone and measure the Pauli $Z$ expectation value of the target qubit. The fit for the oscillation over the total CR gate time $8\tau_{CR}N$ yields the strength of the total unwanted local rotation terms. At $\tau_{sq} = 0$ ns, the unwanted local rotation strength of 176.2 kHz was observed. This strength was reduced to 6.7 kHz with the calibrated compensation tone with $A' = 0.00102$ and $\phi' = -0.962$ rad. The experimental result is shown in Fig. 4(d).