Context-aware fidelity estimation

Dripto M. Debroy,[1,*,†] Élie Genois ◉,[1,2,*,‡] Jonathan A. Gross,[1,*,§] Wojciech Mruczkiewicz ◉,[1] Kenny Lee,[1]
Sabrina Hong,[1] Zijun Chen,[1] Vadim Smelyanskiy,[1] and Zhang Jiang[1]

[1]*Google Quantum AI, Venice, California 90291, USA*

[2]*Institut quantique & Département de Physique, Université de Sherbrooke, Québec, Canada J1K 2R1*

We present context-aware fidelity estimation (CAFE), a framework for benchmarking quantum operations that offers several practical advantages over existing methods such as randomized benchmarking (RB) and cross-entropy benchmarking. In CAFE, a gate or a subcircuit from some target experiment is repeated n times before being measured. By using a subcircuit, we account for effects from the spatial and temporal circuit context. Since coherent errors accumulate quadratically while incoherent errors grow linearly, we can separate them by fitting the measured fidelity as a function of n. One can additionally interleave the subcircuit with dynamical decoupling sequences to remove certain coherent error sources from the characterization when desired. We have used CAFE to experimentally validate our single- and two-qubit unitary characterizations by measuring fidelity against estimated unitaries. In numerical simulations, we find that CAFE produces fidelity estimates at least as accurate as interleaved RB while using significantly fewer resources. We also introduce a compact formulation for preparing an arbitrary two-qubit state with a single entangling operation and use it to present a concrete example using CAFE to study controlled-Z gates in parallel on a Sycamore processor.

I. INTRODUCTION

Reliably understanding the structure of noise in quantum processors is vital for advancing quantum computation. There are a variety of characterization methods available to quantify and validate the performance of individual quantum operations such as state preparation, single- and two-qubit gates, measurement, and reset [1]. Amongst these techniques, some use randomness to estimate gate fidelities efficiently, such as randomized benchmarking (RB) [2–7], cross-entropy benchmarking (XEB) [8,9], channel spectrum benchmarking [10], and direct fidelity estimation [11–13], while others use complete sets of input states, such as unitary tomography [14,15], quantum process tomography [16,17], and gate set tomography [18,19]. These methods all have upsides and downsides in terms of speed, scalability, and the amount of information provided [1].

An important problem that many existing techniques face is that as quantum computers scale, complex intercomponent interactions arise, such as control cross talk or temporal correlations caused by residual pulse tails. For example, ex-periments may use amplifiers with temperature-dependent gain profiles, which could be linked to pulse duty cycle and cause circuit-dependent errors. These effects can make the performance of a quantum operation highly context dependent [9,20–22], and there is a growing need for characterization methods which account for them.

In this paper, we describe a characterization method called *context-aware fidelity estimation* (CAFE) which measures the fidelity between an experimentally implemented quantum operation and a reference unitary. In CAFE, we repeat a gate or a subcircuit from a target experiment n times and measure the average gate fidelity against the reference unitary raised to the nth power. For example, the subcircuit can be part of a stabilizer extraction circuit in a quantum error correction experiment, as demonstrated experimentally in Sec. VI. This procedure allows us to capture context-related errors which isolated characterizations do not capture [23,24]. Since coherent errors accumulate quadratically as a function of n while incoherent errors grow linearly, we can separate their contributions to the infidelity. This is in contrast with RB and XEB, where random compiling is used to twirl coherent errors into incoherent ones.

Throughout the main text of this paper, we focus on using CAFE to study the performance of controlled-Z (CZ) gates implemented in parallel, as two-qubit operations have been shown to be a dominant error source across recent large-scale experiments on a number of experimental platforms, especially in the presence of stray interactions [25–27]. We also discuss CAFE for a subcircuit of a stabilizer extraction circuit. In the Appendixes, we present results from CAFE experiments characterizing parallel single-qubit operations.

---

*These authors contributed equally to this work.

†dripto@google.com

‡elie.genois@usherbrooke.ca
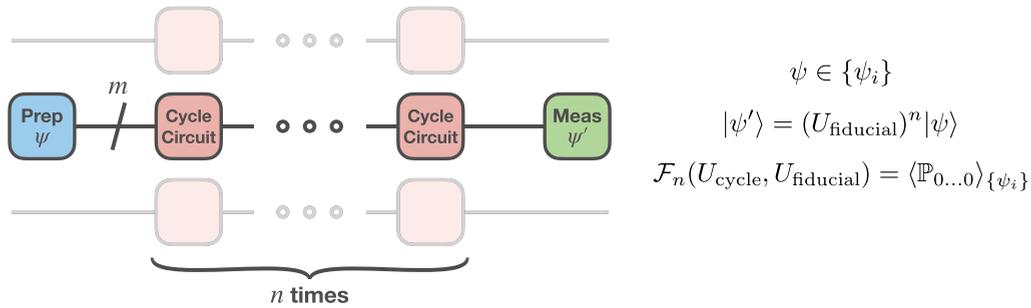
§jarthurgross@google.com

FIG. 1. A schematic of the circuits used to measure the fidelity of *n* repetitions of the cycle circuit, which is selected to be a subcircuit from some target experiment. First, a state pulled from an *m*-qubit 2-design, $\{\psi_i\}$, is prepared (blue). Second, the cycle circuit being characterized is applied *n* times (pink). Third, we undo the combined action of the previous circuit in a single step, similar to the inversion step in randomized benchmarking, and measure the resulting state in the computational basis (green). The fidelity between *n* repetitions of the applied operation and the *n*th power of the reference unitary can be found by averaging the probability of getting $|0\rangle^{\otimes m}$ over all 4^m initial states. The faded operations represent the spatial context of the cycle circuit being characterized. Meas, measurement; Prep, preparation.

II. CONTEXT-AWARE FIDELITY ESTIMATION

A typical CAFE experiment is performed in three steps, schematically shown in Fig. 1. First, a state $|\psi\rangle$ is prepared from an *m*-qubit 2-design $\{\psi_i\}$ [28–31]. These ensembles match the Haar random distribution up to second moments, allowing us to measure fidelity. In Appendix B, we present a method to construct shallow circuits that prepare arbitrary two-qubit entangled states for this purpose. The method uses a single entangling operation to prepare a two-qubit state with the desired entanglement signature, followed by single-qubit operations to reach the desired state.

Next, the *m*-qubit circuit of interest, referred to as the *cycle circuit*, is repeated *n* times, along with operations on neighboring uninvolved qubits. The cycle circuit can include multiple operations, allowing close matching to the circuit context found in the target experiment. As an example, an experiment which includes dynamical decoupling (DD) would be robust in the presence of certain coherent errors, and including these DD gates in the cycle circuit allows CAFE to neglect contributions from these coherent errors, focusing on the errors which impact performance. Inserting DD gates is one of several ways for CAFE to separate coherent and incoherent contributions to gate errors, which we discuss further in Sec. IV.

Lastly, we apply a circuit that maps the state back to $|0\rangle^{\otimes m}$, assuming that the cycle circuit implements the desired "fiducial unitary," before measuring the qubits in the computational basis. This final step is analogous to the inversion step of RB, used to undo the action of the total applied circuit before measurement. Moreover in CAFE, this final step allows us to validate the performance of different unitary characterization methods, as discussed further in Sec. V. The average gate fidelity of the operation, which we refer to as *fidelity* in the following, can be found by averaging over the experiments for all 2-design states $\{\psi_i\}$ [14]:

$$\mathcal{F}(U_{\text{cycle}}, U_{\text{fiducial}}) = \langle \mathbb{P}_{0\cdots0} \rangle_{\{\psi_i\}}. \quad (1)$$

We note that in the $m = 2$ case, the preparation and measurement stages only use a single entangling gate, so their contribution to the total error rate is far smaller than the circuit being characterized for most values of *n*. In general, since the information we extract from CAFE comes from fitting the

fidelity over different values of *n*, reasonable state preparation and measurement (SPAM) errors do not significantly impact the characterization results, as they only cause a constant offset in the fidelity curve. While the number of 2-design states scales exponentially with qubit count, random circuits can approximate these states in polynomial depth [29], which could be used for performing CAFE on $m > 2$ qubits.

III. COMPARING CAFE WITH RANDOMIZED BENCHMARKING

In this section, we compare our CAFE approach to the widely used interleaved randomized benchmarking (IRB) protocol for estimating the fidelity of noisy CZ gates. The error
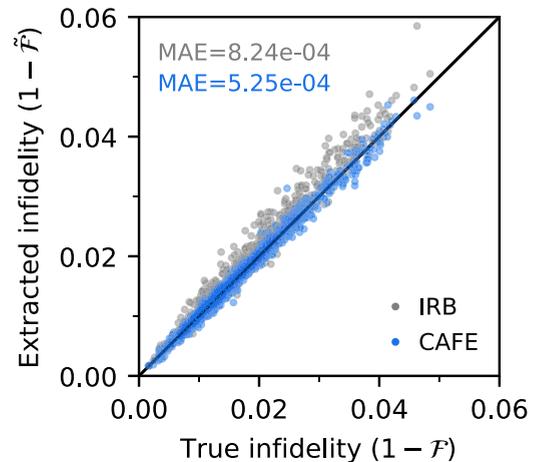

FIG. 2. Simulations comparing the CAFE and interleaved randomized benchmarking (IRB) techniques for characterizing the average gate infidelity of noisy CZ gates with amplitude and phase damping noise in addition to coherent errors. For CAFE (blue), we use depths $n \in [0, 2, 4, 6, 8]$, whereas for IRB (gray) we use depths $n \in [5, 10, 15, 20, 25, 30, 35]$ for obtaining both the reference RB curve and the acquisition with interleaved CZ gates. Both approaches use 2000 shots per circuit. Labels presents the median absolute errors (MAEs) of IRB and CAFE over $N = 1000$ different noisy CZ gates. MAE = median($|\tilde{\mathcal{F}}_1 - \mathcal{F}_1|, \ldots, |\tilde{\mathcal{F}}_N - \mathcal{F}_N|$). We note that MAE scales with the true error of the gates being considered.

model for these gates, which is fully described in Appendix E, contains coherent errors together with amplitude and phase damping as incoherent noise. In Fig. 2, we show that for this model, and using realistic error rates, CAFE yields a more accurate average gate fidelity estimation than IRB, while requiring significantly fewer experimental resources.

More concretely, the randomized benchmarking protocol uses 20 different circuits for seven depths $5 \leqslant n \leqslant 35$ of random two-qubit Clifford gates (which often require two CZ's together with single-qubit gates once compiled), in addition to repeating these same circuits interleaved with a CZ at every depth in order to obtain the CZ gate fidelity estimate. In contrast, CAFE uses only 16 different circuits and five depths $0 \leqslant n \leqslant 8$ of CZ's (with at most two additional CZ layers for state preparation and measurement). Moreover, the single-qubit gates used by IRB to create the random Clifford gates introduce unwanted error sources from decoherence and systematic errors, a problem that CAFE alleviates altogether by only repeating the cycle circuit of interest. Note that, unlike CAFE, RB additionally provides an estimate of the average gate fidelity of all two-qubit Clifford gates, which can be of independent interest. We also want to point out that using additional depths n and sampling shots can improve both approaches, and that considering a different range of noise parameters can change the resulting median absolute errors significantly. However, we have found that the conclusion remains that CAFE allows one to do a gate characterization at least as accurate as IRB, while requiring significantly less run time in experiment.

$$\tilde{U}(\Delta\theta, \Delta\gamma, \Delta\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\Delta\gamma}\cos(\Delta\theta) & -ie^{-i\Delta\gamma}\sin(\Delta\theta) & 0 \\ 0 & -ie^{-i\Delta\gamma}\sin(\Delta\theta) & e^{-i\Delta\gamma}\cos(\Delta\theta) & 0 \\ 0 & 0 & 0 & -e^{-i(\Delta\phi+2\Delta\gamma)} \end{pmatrix}, \tag{2}$$

where $\tilde{U}(0,0,0) = $ CZ, and the swap, single-qubit phase, and controlled-phased miscalibration angles are assumed to be small, such that $\Delta\theta, \Delta\gamma, \Delta\phi \ll 1$. To simplify further the expressions here, we also assume that the noisy quantum channel implementing the cycle circuit can be described by a two-qubit depolarizing channel

$$\mathcal{E}(\rho) = (1 - p_{\text{depol}})\tilde{U}\rho\tilde{U}^\dagger + p_{\text{depol}} I_d/d, \tag{3}$$

which outputs a totally mixed state with probability p_{depol} and otherwise applies the cycle unitary $\tilde{U}$. With such a channel, the average gate fidelity for n repetitions of the cycle circuit is given by

$$\begin{aligned} \mathcal{F}_n = \tfrac{1}{4} - \epsilon_{\text{SPAM}} - \tfrac{1}{20}(1 - p_{\text{depol}})^n \\ \times (1 - |1 + 2e^{-in\Delta\gamma}\cos(n\Delta\theta) + e^{-in(2\Delta\gamma+\Delta\phi)}|^2), \end{aligned} \tag{4}$$

where we have explicitly included the SPAM errors ϵ_{SPAM}, which are assumed to vary slowly over the timescale of an experiment. We note that in this model we made a steady-state assumption about the gates in our system, but modeling transient behavior could be achieved with a more advanced fit. Using the expression in Eq. (4), we can model the

IV. BUDGETING COHERENT AND INCOHERENT ERRORS

Using CAFE, one can accurately estimate the fidelity of any unitary operation and report this single number as a performance metric, which is useful for validation and for estimating the performance of different quantum algorithms [1]. However, such a metric alone provides very little information as to how to improve the gate fidelity in practice. A central feature of CAFE is that one can extract actionable information about the origins of gate error by budgeting the coherent and incoherent contributions. To do so, one can fit the fidelity decay curve to a physical model or modify the cycle circuit to echo out different parts of the gate errors, for example, by leveraging dynamical decoupling (DD) pulses. We note that such budgeting is helpful for directing the research focus, as the interventions required to mitigate coherent and incoherent errors tend to be different.

A. Separating coherent and incoherent errors through fitting to a model

One way to obtain an error budget from CAFE is to fit the experimental data to a model. Although the CAFE experiment and resulting error budget are valid for any m-qubit unitary, we will again focus on the two-qubit CZ case to simplify the discussion in the main text. A more general derivation is presented in Appendix C. We assume that the cycle unitary is an excitation-preserving two-qubit gate close to a CZ

experimental data to obtain the gate fidelity $\mathcal{F}$, in addition to the incoherent errors ϵ_{incoh} and coherent errors ϵ_{coh} of the quantum operation. To get these parameters in a way that is robust in the presence of SPAM errors, we use

$$1 - \mathcal{F} = 1 - \frac{\mathcal{F}_1}{(1 - \epsilon_{\text{SPAM}})}, \tag{5}$$

$$\epsilon_{\text{incoh}} = 1 - \frac{\mathcal{F}_1(p_{\text{depol}} = 0)}{(1 - \epsilon_{\text{SPAM}})}, \tag{6}$$

$$\epsilon_{\text{coh}} = 1 - \frac{\mathcal{F}_1(\Delta\theta = \Delta\gamma = \Delta\phi = 0)}{(1 - \epsilon_{\text{SPAM}})}. \tag{7}$$

In numerical simulations, which are presented in Appendix E, this analysis procedure was shown to be valid and robust in the presence of different unitary errors, as well as amplitude and phase damping channels. We use this method to budget errors in Figs. 2–5. A similar budgeting approach for single-qubit X(π) gates, together with experimental results acquired on a Sycamore chip, are presented in Appendix D.

As shown in Fig. 3, the simple model of Eq. (4) allows us to accurately fit the CAFE curves spanning a wide range
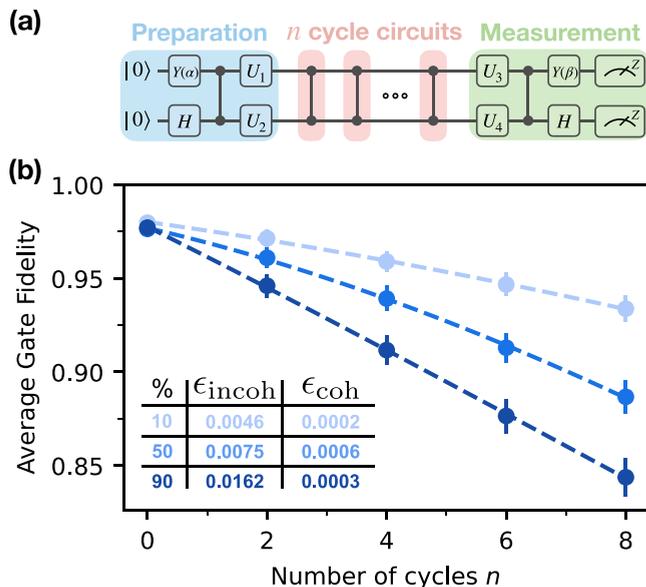
(a)

(b)

FIG. 3. Experimental characterization of CZ gates performed in parallel using CAFE. (a) Structure of the 16 circuits used to characterize the fidelity of a CZ gate for a given depth. The angles $\{\alpha, \beta\}$ and single-qubit unitaries U_j are found using the method described in Appendix B. Here the cycle circuit is a single CZ gate. (b) Data representing the cycle circuit fidelity compared with the ideal CZ unitary for different cycle repetitions, along with fits using the model presented in Eq. (4) and resulting gate budgets in the inset table. We plot the data for three specific gates—at the 10th, 50th, and 90th percentiles on a Sycamore device in terms of average gate infidelity—to illustrate the accuracy of the modeling over a wide range of data. Error bars represent the standard deviation of binomial distributions scaled by a factor of 5 to be visible.
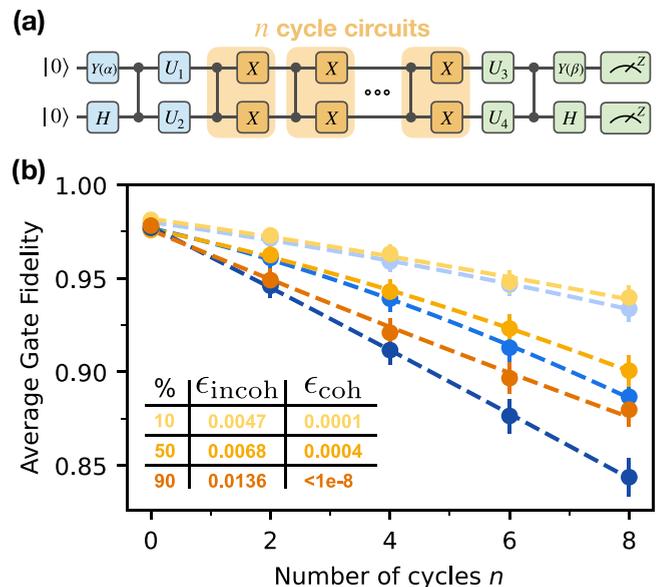
(a)

(b)

FIG. 4. CAFE combined with dynamical decoupling (DECAF). (a) Structure of the 16 DECAF circuits for characterizing a CZ gate at a given depth. Only the cycle circuit highlighted in orange differs from the CAFE circuits of Fig. 3(a). (b) Experimental results on characterizing CZ gates in parallel when dynamical decoupling gates are part of the cycle circuit (orange), alongside the standard CAFE data presented in Fig. 3 (blue). Interleaving the CZ gates with X gates on both qubits echoes out low-frequency Z noise, which contributes to ϵ_{incoh}, and removes the sensitivity to single-qubit phase unitary errors, which contribute to ϵ_{coh}. Performing the simple DECAF experiment thus provides valuable information about the error mechanisms in CZ gates. Error bars are scaled by a factor of 5 to be visible.

of CZ gate fidelities executed in parallel on a Sycamore chip. Additional data showing the consistency of the resulting error budget with XEB are presented in Appendix H.

A useful and intuitive picture to analyze the CAFE data is to consider the incoherent and coherent errors as linear and quadratic contributions to the gate infidelity, respectively. This can be seen directly by expanding Eq. (4) up to terms $O((\Delta\theta)^4, (\Delta\gamma)^4, (\Delta\phi)^4, p_{\text{depol}}^2)$:

$$\mathcal{F}_n \approx 1 - \epsilon_{\text{SPAM}} - \epsilon_{\text{lin}} \, n - \epsilon_{\text{quad}} \, n^2, \quad (8)$$

$$\epsilon_{\text{lin}} = \frac{3 p_{\text{depol}}}{4} \, n, \quad (9)$$

$$\epsilon_{\text{quad}} = \frac{8[(\Delta\theta)^2 + (\Delta\gamma)^2 + \Delta\gamma\,\Delta\phi] + 3(\Delta\phi)^2}{20} \, n^2. \quad (10)$$

This approximate quadratic form can be found for different cycle unitaries under similar noise channels and could be used directly in the budgeting procedure given low gate infidelities and shallow cycle repetitions n or in cases lacking an accurate analytical model for the quantum channel. Note that the fit model and resulting CAFE error budgeting can be straightforwardly extended to another platform or experiment. For example, the quantum channel describing the cycle circuit can be parametrized by a set of Kraus operators and computed numerically to fit the CAFE fidelities at different depths n. However, since CAFE performs only the minimal set of cir-

cuits needed to extract the average gate fidelity, the data will not contain enough information to fully characterize generic quantum channels, and additional experiments or other characterization tools should be considered if that is the goal.

We also note that the quantity ϵ_{incoh} estimates the average gate infidelity when no coherent control errors are present. This useful characterization metric is defined as $R(\mathcal{E})$ in Ref. [32], where the authors show that it bounds the *unitarity* of the channel $u(\mathcal{E})$:

$$\frac{d-1}{d}(1 - \sqrt{u(\mathcal{E})}) \leqslant R(\mathcal{E}). \quad (11)$$

As such, we can also relate our incoherent error estimate to unitarity. For instance, in the case of depolarizing noise, the unitarity is

$$u(\mathcal{E}_{\text{depol}}) = (1 - p_{\text{depol}})^2, \quad (12)$$

and the incoherent error obtained from CAFE is, as derived in Appendix C,

$$1 - \epsilon_{\text{incoh}} = \frac{1}{d} + \frac{d-1}{d}(1 - p_{\text{depol}}) \quad (13)$$

$$= \frac{1}{d} + \frac{d-1}{d}\sqrt{u(\mathcal{E}_{\text{depol}})}. \quad (14)$$

B. Isolating coherent channels using dynamical decoupling

Exploiting the versatility of CAFE, a second method for error budgeting is to modify the cycle circuit in order to isolate
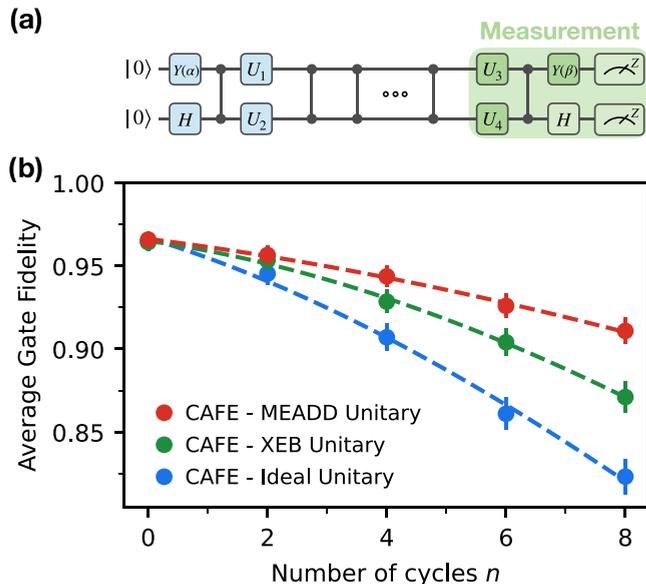
(a)

(b)

FIG. 5. Using CAFE to validate different unitary characterizations of a CZ gate with 50th percentile infidelity. (a) Only the three single-qubit gates $\{U_3, U_4, Y(\beta)\}$ highlighted in green need to be modified to compare the fidelity of the experimental CZ gate with an arbitrary unitary. (b) The different curves represent the fidelity between the experimental gate and an ideal CZ unitary (blue), a unitary extracted from an XEB experiment (green), and a unitary characterized with a newly introduced method called MEADD [35]. Since only the premeasurement circuit changes in the three different CAFE experiments, the improvement in fidelity can entirely be attributed to considering a gate unitary that is closer to the experimental CZ implementation. Error bars are scaled by a factor of 5 to be visible.

specific error channels. In particular, this is viable when the impact of the modifications is insignificant relative to the error channels being isolated. As a relevant example, we can leverage the fact that the cycle circuit is repeated n times by inserting dynamical decoupling gates in between repetitions. This approach of combining DD and CAFE, which we refer to as *DECAF*, is particularly useful to characterize the amount of certain coherent error present in the cycle circuit without necessitating full unitary tomography. We note that this is somewhat similar to the work in Ref. [5], with a single repetition of the cycle circuit, and the randomized unitaries replaced with specific DD pulses.

In the case of a CZ gate, adding an X gate to both qubits in the CAFE cycle circuit echoes out the single-qubit phase errors, in addition to mitigating low-frequency noise, as demonstrated in Appendix F [33,34]. As shown in Fig. 4, the DECAF data have higher fidelities and decrease more linearly than the CAFE data in practice. Looking at the resulting fits over all characterized CZ gates, we see a significant decrease in the median coherent error from 5.5×10^{-4} to 9.4×10^{-5}, which highlights single-qubit phase miscalibrations, alongside a decrease of the median incoherent error from 7.2×10^{-3} to 6.7×10^{-3}, which indicates the level of low-frequency noise in the device.

V. VALIDATING UNITARY CHARACTERIZATIONS

One key difference between CAFE and other methods for extracting error rates is the final unentangling step before measurement. By doing all of the inversion in a single step, similar to RB, but using an arbitrary characterized reference unitary for the inversion, we can validate different unitary characterizations while respecting the non-Clifford nature of most coherent error models. As such, we can use CAFE to benchmark unitary characterizations, simply by changing the final measurement step and seeing which predictions most accurately map the final state back to $|0\rangle^{\otimes m}$, in conjunction with the methods presented in Sec. IV A. As illustrated in Fig. 5(a), for a two-qubit cycle circuit, this change corresponds to modifying three single-qubit gates in the last step of the CAFE protocol.

In Fig. 5(b), we compare the CZ unitary extracted by XEB to the one extracted using a unitary characterization method called matrix-element amplification by dynamical decoupling (MEADD) [35], which is a characterization technique that we have developed based on Floquet characterization [36,37]. MEADD allows us to isolate and precisely measure the different unitary parameters in any phased fermionic simulation (FSIM) gate (a general excitation-number-preserving two-qubit gate). We can see very clearly that the unitary predicted by MEADD is significantly better at predicting the coherent error than the unitary extracted by XEB. By increasing the amount of context around the gate, for example, including some microwave operations or measurements on the surrounding qubits to include cross talk or measurement-induced dephasing effects, we can see which characterizations break down in other contexts (see Sec. VI) and build trust that the structures seen in our characterizations are the dominant effects impacting algorithm performance on the processor.

VI. CHARACTERIZING MULTILAYER CIRCUITS USING CAFE

Here, we provide an example of a CAFE experiment that includes both spatial and temporal context by characterizing a cycle circuit containing multiple layers. As illustrated in Fig. 6(a), we consider a portion of the stabilizer extraction circuit for a distance-3 surface code experiment used in Ref. [26]. In Fig. 6(b), we show how this section of the circuit can be considered on its own and broken down into qubit groups. We can then run a parallel CAFE experiment on these qubit groups which uses the highlighted section as the cycle circuit. The resulting experimental data are shown in Fig. 6(c). We can see that one of the $m = 1$ qubit groups experiences significant error, despite its individual single-qubit X and Hadamard (H) gates showing good fidelities when characterized in isolation. This highlights the importance of context-aware characterization, as the error mode being experienced was only visible when the gates were run in conjunction. This experiment also highlights the flexibility of CAFE in being able to characterize multioperation circuits simultaneously on qubit groups of varying size.
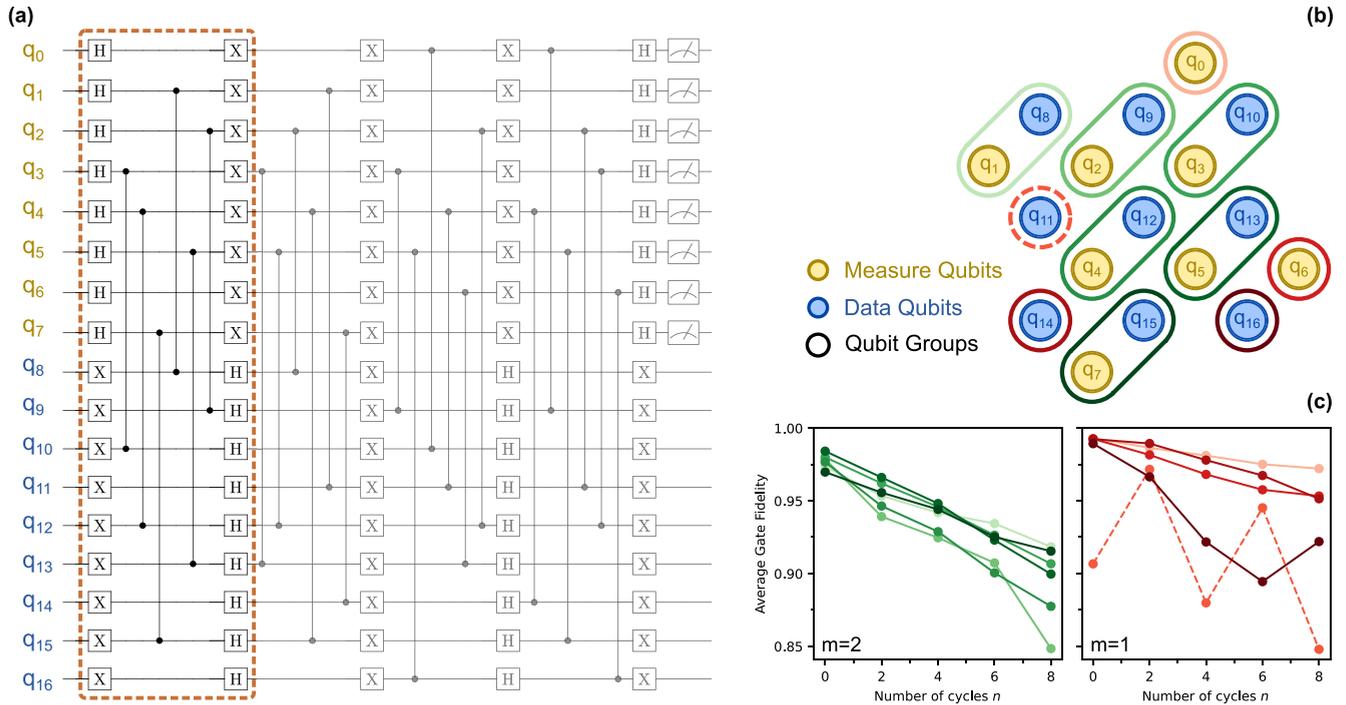
FIG. 6. (a) The stabilizer extraction circuit for a distance-3 surface code, with the section to be characterized boxed in orange. (b) A layout for the CAFE experiment along with the qubit groups characterized. (c) CAFE data for repetitions of the cycle circuit from (a) on the qubit groups in (b), with the $m = 2$ groups in green and the $m = 1$ groups in red. CAFE allows us to see that q_{11} (red dashed curve) is experiencing a significant context-dependent error.

VII. CONCLUSION

CAFE separates itself from other gate characterization methods due to its simplicity, efficiency, and flexibility, most notably as a complementary tool to other gate characterizations that provide more granular output. We have found its ability to split coherent and incoherent errors to be more reliable than that of other methods, and the ability to experimentally test unitary characterizations has allowed us to design and evaluate novel characterization methods more effectively.

Overall, our results demonstrate that CAFE yields accurate fidelity and error budget estimates using significantly less time in practice than complementary characterization protocols. This approach is thus particularly useful to quickly identify the worst performing individual quantum operations, typically single- and two-qubit gates, directly in the context of the larger quantum circuit in which they are found. Moreover, CAFE yields actionable information about the origins of the gate errors, namely the contributions of incoherent and coherent noise, which allows one to calibrate the quantum device to optimize performance on the experiment of interest. This usefulness of CAFE is demonstrated in Sec. VI for characterizing and recalibrating a portion of the stabilizer extraction circuit of a distance-3 surface code experiment.

The flexibility of CAFE allows for many other as-of-yet unexplored variations, from creating fits which include the impact of leakage, to changing the cycle circuit round-by-round to echo out components in different ways, to analyzing the different input bases separately to extract details about the error structure. Our hope is that other researchers will be able to create their own modifications of the CAFE framework to study the errors facing their own systems.

APPENDIX A: DEVICE SPECIFICS

All experimental data presented here were collected on a subset of a Sycamore device with 72 transmon qubits and 121 tunable couplers, where each qubit is coupled to up to four neighbors. In Fig. 7 we present median error rates for microwave single-qubit gates, CZ gates, and measurements over the time frame during which data were acquired for this paper, and in Fig. 8 we show the cumulative density functions for these errors. All errors reported are measured simultaneously. For more information on the architecture, see Refs. [9,38], and for more information on the implementation of the gates, see Ref. [39].
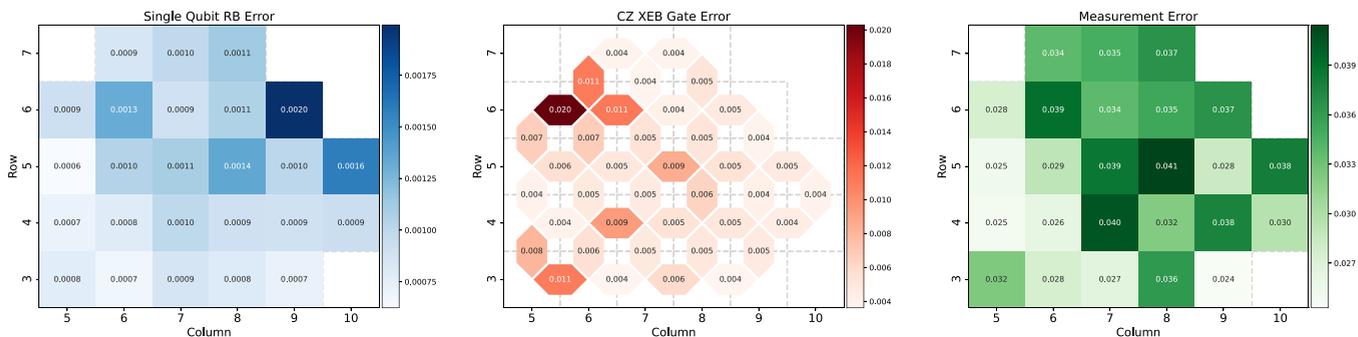
FIG. 7. Heat maps for per-qubit and per-pair median error rates for single-qubit randomized benchmarking, CZ XEB, and average readout identification error.

APPENDIX B: PREPARING AND MEASURING TWO-QUBIT STATES

We describe a method to prepare or measure an arbitrary two-qubit state with a single maximally entangling operation. We start by considering a general two-qubit target state

$$|\psi\rangle = A|00\rangle + B|01\rangle + C|10\rangle + D|11\rangle,$$
$$|A|^2 + |B|^2 + |C|^2 + |D|^2 = 1. \quad \text{(B1)}$$

We can then write a matrix with the amplitudes of this state,

$$M_\psi = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad \text{(B2)}$$

and perform the singular value decomposition (SVD) $M_\psi = U_\psi S_\psi V_\psi^\dagger$. Considering the initial singular value, S_ψ^{00}, we can quantify the level of entanglement in the targeted state. The first step to generate this state is to prepare a state with a matching entanglement signature.

In the case where we intend to use a CZ gate, this can be done by putting one qubit in $|+\rangle$, and applying a Y rotation to the other qubit with an angle of $\alpha = 2\arccos(S_\psi^{00})$. This state, labeled as $|$CZ$\rangle$ in Fig. 9, is equivalent to the final state up to single-qubit rotations. To find these rotations, we take the SVD of the 2×2 matrix corresponding to this intermediate state, $U_{CZ}S_{CZ}V_{CZ}^\dagger$. The single-qubit unitary required for the first qubit is given by $U_1 = U_\psi U_{CZ}^{-1}$, and the unitary for the second qubit is $U_2 = V_\psi V_{CZ}^{-1}$. The resulting circuit is shown in Fig. 9. Constructions for CZ and $\sqrt{\text{ISWAP}}$ are implemented as the methods `prepare_two_qubit_state_with_cz` and `prepare_two_qubit_state_with_iswap` in the open-source software CIRQ [40].

APPENDIX C: DERIVATION OF EQUATION (4)

In this Appendix, we derive Eq. (4) as an example of how one could do the same for other unitaries and error models of interest. Given a general quantum channel described by a set
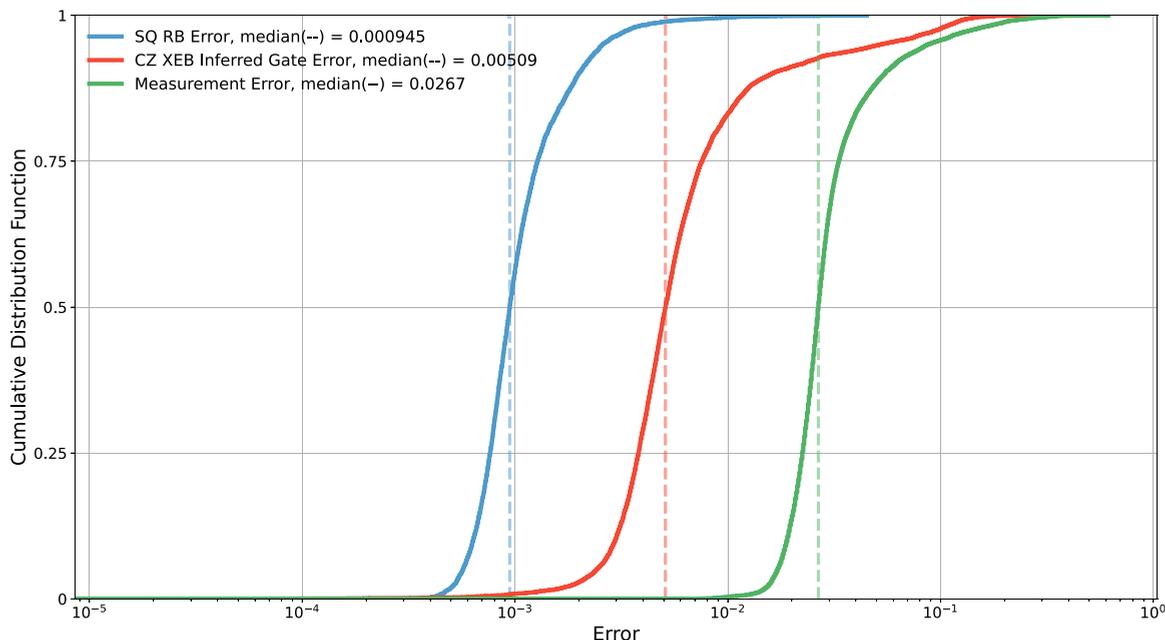

FIG. 8. Cumulative density functions for single-qubit (SQ) RB, CZ XEB, and readout errors.
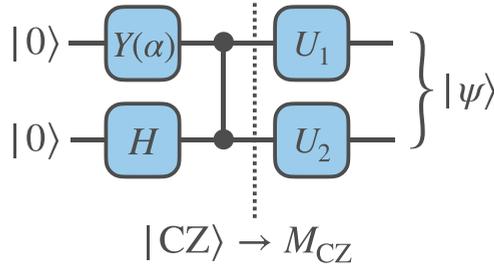
FIG. 9. A simple circuit that maps $|00\rangle$ to an arbitrary two-qubit state $|\psi\rangle$ using only one CZ operation. The overlap of any two-qubit state with the state $|\psi\rangle$ can be obtained by executing the inverse of this circuit and reporting the probability of measuring $|00\rangle$ afterwards.

of Kraus operators K_α,

$$\mathcal{E}(\rho) = \sum_\alpha K_\alpha \rho K_\alpha^\dagger, \tag{C1}$$

the average fidelity of the quantum operation $\mathcal{E}$ with respect to a unitary operation U is

$$\mathcal{F}(\mathcal{E}, U) = \int \langle\psi(\boldsymbol{x})| U^\dagger \mathcal{E}(|\psi(\boldsymbol{x})\rangle \langle\psi(\boldsymbol{x})|) U |\psi(\boldsymbol{x})\rangle \, d\mu(\boldsymbol{x}) \tag{C2}$$

$$= \sum_\alpha \int |\langle\psi(\boldsymbol{x})| U^\dagger K_\alpha |\psi(\boldsymbol{x})\rangle|^2 \, d\mu(\boldsymbol{x}), \tag{C3}$$

where $\boldsymbol{x}$ is a parametrization of pure quantum states and $\mu(\boldsymbol{x})$ is the Haar measure. To evaluate $\mathcal{F}$, we introduce the projector on the symmetric subspace of the system and its replica, which is expressed as

$$P_S = \frac{d(d+1)}{2} \int |\psi(\boldsymbol{x})\rangle \langle\psi(\boldsymbol{x})| \otimes |\psi(\boldsymbol{x})\rangle \langle\psi(\boldsymbol{x})| \, d\mu(\boldsymbol{x}), \tag{C4}$$

where $d = 2^m$ is the dimension of the m-qubit Hilbert space. Using this projector, the fidelity takes the form

$$\mathcal{F}(\mathcal{E}, U) = \frac{2}{d(d+1)} \mathrm{tr}\left(P_S \sum_\alpha K_\alpha^\dagger U \otimes U^\dagger K_\alpha \right) \tag{C5}$$

$$= \frac{1}{d(d+1)} \sum_\alpha \mathrm{tr} K_\alpha^\dagger K_\alpha + |\mathrm{tr}(U^\dagger K_\alpha)|^2 \tag{C6}$$

$$= \frac{1}{d+1} + \frac{1}{d(d+1)} \sum_\alpha |\mathrm{tr}(U^\dagger K_\alpha)|^2. \tag{C7}$$

In order to compute the fidelity expression in Eq. (C7), one computes the eigenvalues of $U^\dagger K_\alpha$ and sums them up directly to get $\mathrm{tr}(U^\dagger K_\alpha)$. For example, we consider the case where the operation of interest can be described by a quantum channel that either applies the unitary $\tilde{U}$ or totally depolarizes the qubits with probability p_{depol}

$$\mathcal{E}(\rho) = (1 - p_{\mathrm{depol}}) \tilde{U} \rho \tilde{U}^\dagger + p_{\mathrm{depol}} I_d/d, \tag{C8}$$

where I_d is the d-dimensional identity operator. The fidelity of this channel after n consecutive applications is given by

$$\mathcal{F}_n = (1 - p_{\mathrm{depol}})^n \frac{d + |\mathrm{tr}[(U^\dagger)^n \tilde{U}^n]|^2}{d(d+1)} + [1 - (1 - p_{\mathrm{depol}})^n] \frac{1}{d}. \tag{C9}$$

After subtracting ϵ_{SPAM} from the right-hand side to account for experimental SPAM errors, the expression in Eq. (C9) can be used to fit the CAFE data of any m-qubit unitary $\tilde{U}$, subject to symmetric depolarizing noise, with respect to a reference unitary U. Note that in the case where the quantum operation under characterization, $\tilde{U}$, corresponds exactly with the reference unitary U, the CAFE data should obey the following single-exponential decay:

$$\mathcal{F}_n = \frac{1}{d} + \frac{d-1}{d} (1 - p_{\mathrm{depol}})^n. \tag{C10}$$

Otherwise, when $\tilde{U} \neq U$, coherent errors are present, and the first term of (C9) will introduce errors that scale quadratically with n to first order.

In the case we are interested in, we want to use CAFE to characterize a two-qubit gate, where $d = 4$, $U = \text{CZ}$, and $\tilde{U}$ is a number-preserving FSIM gate close to a CZ gate:

$$\tilde{U}(\Delta\theta, \Delta\gamma, \Delta\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\Delta\gamma}\cos(\Delta\theta) & -ie^{-i\Delta\gamma}\sin(\Delta\theta) & 0 \\ 0 & -ie^{-i\Delta\gamma}\sin(\Delta\theta) & e^{-i\Delta\gamma}\cos(\Delta\theta) & 0 \\ 0 & 0 & 0 & -e^{-i(2\Delta\gamma+\Delta\phi)} \end{pmatrix}, \tag{C11}$$

with some residual SWAP-like error captured by the angle $\Delta\theta$, single-qubit phase error captured by $\Delta\gamma$, and controlled-phase (CPHASE)-like error captured by $\Delta\phi$, with $\Delta\theta, \Delta\gamma, \Delta\phi \ll 1$. Since CZ is diagonal, we can show that the eigenvalues of $(U^\dagger)^n \tilde{U}^n$ are

$$\boldsymbol{\lambda} = \begin{pmatrix} 1 & e^{-in(\Delta\gamma-\Delta\theta)} & e^{-in(\Delta\gamma+\Delta\theta)} & e^{-in(2\Delta\gamma+\Delta\phi)} \end{pmatrix}, \tag{C12}$$

for $n \in \mathbb{N}$. As such, $\mathrm{tr}[(U^\dagger)^n \tilde{U}^n] = \mathrm{sum}(\boldsymbol{\lambda})$, and substituting into Eq. (C9) gives

$$\mathcal{F}_n = \frac{1}{4} - (1 - p_{\mathrm{depol}})^n \left(\frac{1 - |1 + 2e^{-in\Delta\gamma}\cos(n\Delta\theta) + e^{-in(2\Delta\gamma+\Delta\phi)}|^2}{20} \right) \tag{C13}$$

$$= 1 - \frac{3p_{\mathrm{depol}}}{4} n - \frac{8(\Delta\theta)^2 + 8(\Delta\gamma)^2 + 8(\Delta\gamma \Delta\phi) + 3(\Delta\phi)^2}{20} n^2 \tag{C14}$$

$$+ O((\Delta\theta)^4, (\Delta\gamma)^4, (\Delta\phi)^4, p_{\mathrm{depol}}^2), \tag{C15}$$
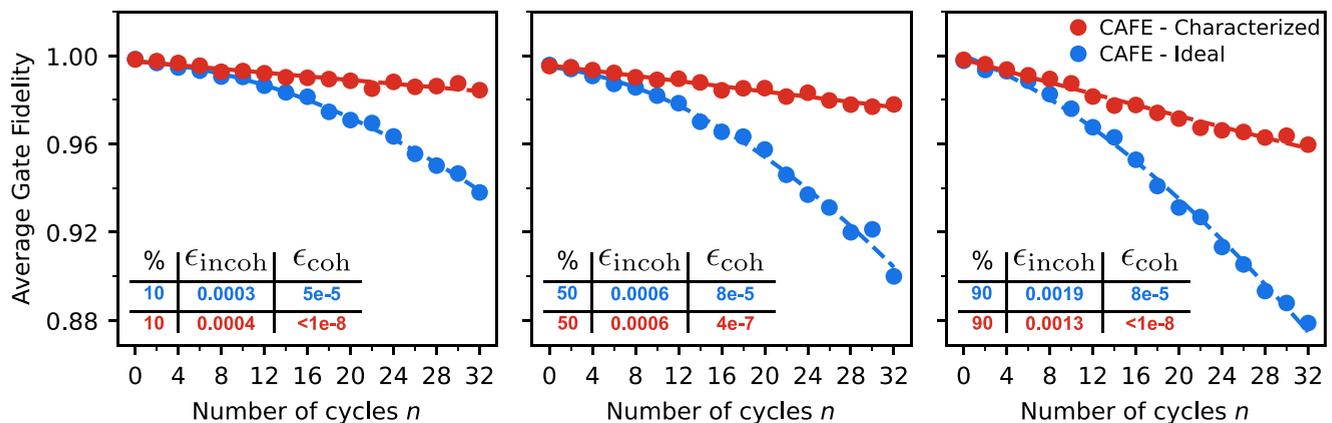
FIG. 10. Single-qubit parallel CAFE results for three X gates close to the 10th (left), 50th (center), and 90th (right) percentiles of infidelity, from a total of 68 parallel X gates on a Sycamore device. The data in blue compute the fidelity with an ideal $U = X$ unitary, whereas the CAFE data in red use a characterized unitary. Note that in these data, the characterized single-qubit unitaries were not used to optimize the gate parameters, thus the imperfect calibration.

which is the expression in Eq. (4) of the main text, after subtracting the SPAM errors ϵ_{SPAM} from the right-hand side to account for experimental imperfections.

APPENDIX D: SINGLE-QUBIT CAFE

In order to give an additional example of how to deploy the CAFE characterization framework, we present an experiment characterizing 68 single-qubit X gates in parallel on a Sycamore processor. The results are presented in Fig. 10, showing the fidelity of the experimental gate with both an ideal unitary (in blue) and a characterized one (in red), in addition to fits using an analytical model derived below, for three example gates. In the single-qubit case, the minimal 2-design contains only four states which can all be prepared with a single gate, which significantly speeds up the characterization. We have used cycle repetitions n up to 32 here, showing how one can modify the experiment when studying especially high fidelity operations to maintain good fit robustness.

Assuming that the incoherent noise of the experimental gate can be described by a totally depolarizing channel, our model to fit the single-qubit CAFE data can again be derived from Eq. (C9). For this case where $U = R_x(\pi)$ and $d = 2$, we can parametrize a single-qubit unitary

$$\tilde{U}(\Delta\mu) = \begin{pmatrix} -\sin(\Delta\mu/2) & -i\cos(\Delta\mu/2) \\ -i\cos(\Delta\mu/2) & -\sin(\Delta\mu/2) \end{pmatrix}, \quad \text{(D1)}$$

where $\tilde{U}(0) = R_x(\pi)$. One can easily show that the eigenvalues of $(U^\dagger)^n \tilde{U}^n$ are

$$\boldsymbol{\lambda}_n = \begin{cases} (i^n e^{in\Delta\mu/2} & i^n e^{-in\Delta\mu/2}), & \text{for } n \text{ even} \\ (e^{in\Delta\mu/2} & e^{-in\Delta\mu/2}), & \text{for } n \text{ odd}, \end{cases} \quad \text{(D2)}$$

where $n \in \mathbb{N}$. As such,

$$|\mathrm{tr}[(U^\dagger)^n \tilde{U}^n]|^2 = |\mathrm{sum}(\boldsymbol{\lambda}_n)|^2 = 4\cos^2(n\Delta\mu/2), \quad \text{(D3)}$$

and Eq. (C9) becomes

$$\mathcal{F}_n = \frac{1}{2} - \epsilon_{\mathrm{SPAM}} + (1 - p_{\mathrm{depol}})^n \left(\frac{2}{3}\cos^2(n\Delta\mu/2) - \frac{1}{6} \right), \quad \text{(D4)}$$

where, as before, we have explicitly included the SPAM errors ϵ_{SPAM} to account for experimental imperfections. The error budget in terms of the fidelity $\mathcal{F}$, incoherent error contribution $\epsilon_{\mathrm{incoh}}$, and coherent error contribution ϵ_{coh} is obtained the same way as in Eqs. (5)–(7) of the main text, i.e., by fitting the data with this expression, evaluating it at $n = 1$ with different noise parameters set to zero, and normalizing by the SPAM errors. Figure 10 shows that such a model allows us to accurately reproduce the CAFE data obtained in the experiment.

APPENDIX E: NUMERICAL CAFE SIMULATIONS

In this Appendix, we present numerical simulations where we characterize a CZ gate using the CAFE approach presented in the main text. We simulate the same $2^4 = 16$ circuits executed on the Sycamore device with noisy two-qubit unitaries, together with either depolarizing noise in Appendix E 1 or amplitude and phase damping noise in Appendix E 2. We also used the same low numbers of cycle repetitions $n \in [0, 2, 4, 6, 8]$ which make the CAFE experiment have especially low execution time relative to other two-qubit benchmarking techniques. In Sec. III we showed that using CAFE to characterize a CZ gate can allow for a more accurate fidelity estimation than the widely used randomized benchmarking (RB) protocol while using significantly fewer experimental resources. Note that we also simulate the sampling of 2000 shots used experimentally, which places an upper bound on the standard deviation of all the CAFE data points presented in Figs. 3–5 of

$$\sigma_{\mathcal{F}} \leqslant \frac{1}{8\sqrt{N_{\mathrm{shots}}}} \approx 0.0028, \quad \text{(E1)}$$

which is smaller than all the markers used in the plots of the main text.

Importantly, these simulations allow us to confirm the validity of the CAFE experiment and the following fitting procedure to obtain accurate estimates of the following: the average gate fidelity of the quantum operation with regard to any reference unitary, the incoherent error contribution to the infidelity, and the coherent error contribution.
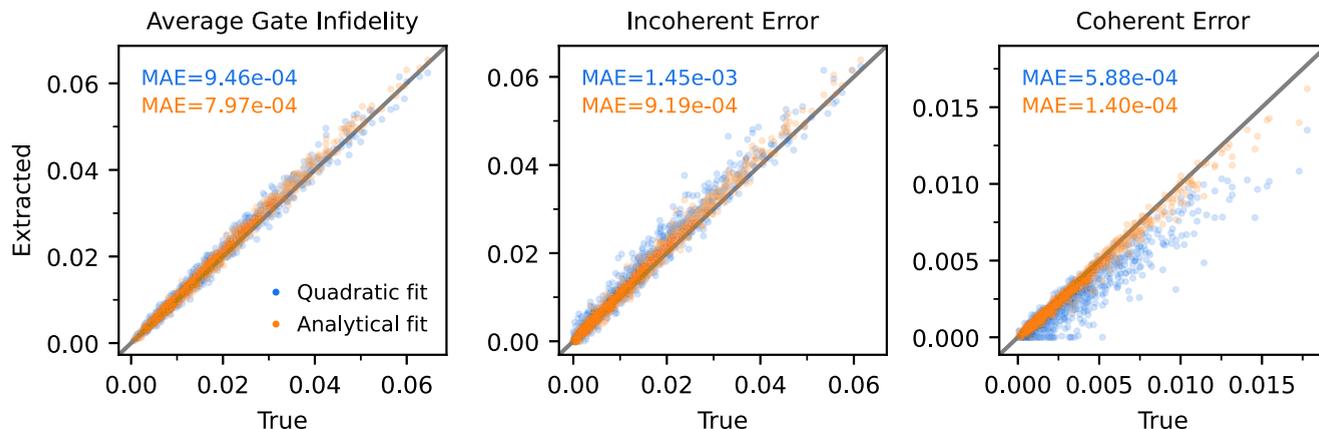
FIG. 11. Simulating the CAFE experiment for characterizing 1000 different CZ gates with realistically small coherent errors and depolarizing incoherent noise. Two error budgeting techniques obtained by fitting these CAFE data are presented: in blue, using a quadratic fit for the fidelity at depths $n \in [0, 2, 4]$, and in orange using the analytical expression of Eq. (4) for depths $n \in [0, 2, 4, 6, 8]$. In each panel we show the median absolute errors (MAEs) of the two error budgeting techniques from the true values inserted into the simulation. We note that MAE scales with the true error of the gates being considered. The true incoherent (coherent) errors are obtained from the channel's average gate infidelity when only depolarizing errors (unitary errors) are present in the simulation.

The simulations were performed using the open-source software CIRQ [40].

1. Depolarizing noise

In a first case, we consider that the noisy CZ gate we are trying to characterize is described by the quantum channel

$$\mathcal{E}(\rho) = (1 - p_{\text{depol}}) \tilde{V} \rho \tilde{V}^\dagger + p_{\text{depol}} I_d / d, \tag{E2}$$

which outputs a totally depolarized state with probability p_{depol}, and otherwise applies the excitation-preserving unitary

$$\tilde{V}(\theta, \zeta, \chi, \gamma, \phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i(\gamma+\zeta)}\cos\theta & -i\,e^{-i(\gamma-\chi)}\sin\theta & 0 \\ 0 & -i\,e^{-i(\gamma+\chi)}\sin\theta & e^{-i(\gamma-\zeta)}\cos\theta & 0 \\ 0 & 0 & 0 & -e^{-i(2\gamma+\phi)} \end{pmatrix}, \tag{E3}$$

where we allow for miscalibrations in all of the five angles that parametrize this gate, where $\tilde{V}(0, 0, 0, 0, 0) = $ CZ. For the simulation, we then sample the depolarizing probability uniformly in the range from 0 to 0.05 and sample the miscalibrated angles from a normal distribution with zero mean and standard deviation of 0.05 rad such that

$$p_{\text{depol}} \propto \mathcal{U}(0, 0.05), \tag{E4}$$

$$\{\theta, \zeta, \chi, \gamma, \phi\} \propto \mathcal{N}(0, 0.05^2). \tag{E5}$$

Using this model, we can simulate the noisy CAFE experiment for different cycle repetitions n and then fit these data points to obtain an error budgeting of the noisy CZ gate, which we can compare directly with the true parameter values that were drawn for the simulation.

In Fig. 11, we present the average gate infidelity $(1 - \mathcal{F})$, incoherent error contribution ϵ_{incoh}, and coherent error contribution ϵ_{coh} to this infidelity for 1000 independently sampled unitaries and depolarizing probabilities. We present two valid gate error budgeting approaches: one which fits the CAFE data with the analytical expression in Eq. (4) (orange points) and one which uses the simple quadratic form of Eq. (8) (blue points). These scatterplots demonstrate the validity of CAFE

to characterize the fidelity of a quantum operation and budget its coherent and incoherent contributions with an imprecision smaller than or equal to 0.001 on median for realistic gate fidelities. Note that the reported median absolute errors depend on the specific range of gate fidelities considered. For example, the MAE values decrease significantly when considering only gate infidelities <1% in the ensemble.

The quadratic fit results shown in blue in Fig. 11 are an indication of how the CAFE framework can be deployed in order to estimate the fidelity of an operation in context. However, this simplest approach gives slightly biased results for ϵ_{incoh} and ϵ_{coh}, which is understood from the breakdown of the $n p_{\text{depol}} \ll 1$ approximation and/or the $n\alpha \ll 1$ approximation, where α here stands for any of the five miscalibrated CZ angles. Consequently, a quadratic fit should be used carefully, for operations with high fidelity (>99%) and using shallow cycle repetitions n. In fact, for these simulations, we have found that using larger depths than $n = 4$ for the quadratic fit decreased the accuracy of the resulting error budgets. However, if the fidelity of the operation of interest is very high, for example, with single-qubit gates, using a quadratic fit becomes an attractive strategy when lacking a proper analytical model of the error origins.
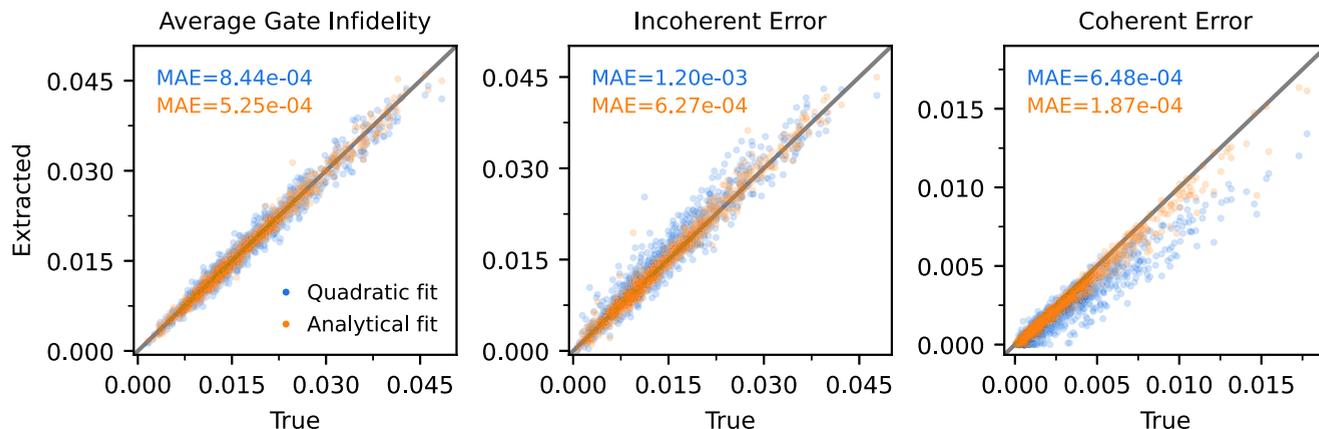
FIG. 12. Similar simulations as presented in Fig. 11, now implementing realistic amplitude and phase damping incoherent noise instead of a depolarizing channel. The analytical fit, which still uses Eq. (4) that effectively lumps the incoherent noise into a p_{depol} probability, works remarkably well even under this different noise channel.

On the other hand, the analytical fit performs well for larger depths n since it does not rely on any approximation. Here, we have only used the depths $n \in [0, 2, 4, 6, 8]$ to be consistent with the experiment results of the main text. These results are an important validation for CAFE but not necessarily surprising since we are using the same error model to simulate the experiment and to fit the resulting data. In the next section, we show that this model also holds very well for different noise models such as amplitude and phase damping, provided they cause realistic gate infidelities of a few percent or less.

2. Amplitude and phase damping noise

To verify the robustness of the CAFE framework in the presence of different incoherent noise processes, we have also simulated a two-qubit CAFE experiment where the quantum channel is described by the same unitary $\tilde{V}$ with randomly sampled angles as before, but now followed by a single-qubit amplitude and phase damping channel on both qubits. The total decay and phase-flip probabilities for both qubits were

sampled from a normal distribution with a standard deviation of 0.03, before taking the absolute value

$$\{p_{\text{decay,total}}, p_{\text{phaseflip,total}}\} \propto |\mathcal{N}(0, 0.03^2)|, \qquad (E6)$$

and the individual decay and phase-flip probabilities of the two qubits were splitting these total probabilities with a ratio drawn from a normal distribution with mean of 0.5 and standard deviation of 0.1, such that the qubits have similar but distinct coherence properties.

In Fig. 12, we show that the CAFE framework we developed again gives very accurate CZ characterizations and error budgets. In particular, using the analytical expression of Eq. (4) to fit these CAFE data works remarkably well given that the noise present in the gate cannot be fully described by the assumed depolarizing channel. Since the incoherent noise is small, but in the realistic range of creating about 0.1–4% gate infidelity, the fit is able to approximate well the incoherent error contribution to the single exponential of the model. This gives us confidence that we can leverage this
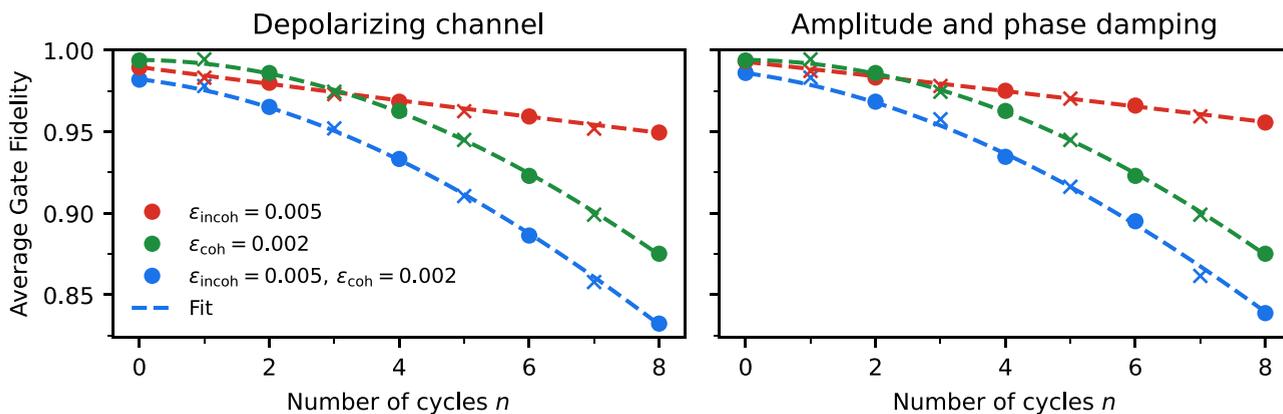

FIG. 13. Example of fitting the CAFE curve obtained in simulations with the depolarizing noise channel used in Appendix E 1 (left) and the amplitude and phase damping noise channel used in Appendix E 2 (right). In blue, we show simulations including both incoherent errors of strength $\epsilon_{\text{incoh}} = 0.005$ and coherent errors of $\epsilon_{\text{coh}} = 0.002$, and the dashed curves show the fits of these data using Eq. (4). As in the main text, only the even depths are used in the fit; the odd-depth points are presented with crosses simply to show their consistency with the model. Note that the green data are the same in both panels as we fix the incoherent errors to be zero.

model in realistic gate characterization scenarios. Moreover, the simulations show that the accuracy of using this model increases with increasing gate fidelities, which is a great indication for the continued usefulness of this method as gates and qubit coherences keep improving.

3. Fitting CAFE data

In Fig. 13, we present examples of the CAFE data together with fits to the model of Eq. (4) realized on the simulation data to obtain the CZ error budgets presented in Figs. 11 and 12. We see that the model of Eq. (4) fits the CAFE data very well for realistic noise of strength $\epsilon_{\text{incoh}} = 0.005$ and $\epsilon_{\text{coh}} = 0.002$, even when the incoherent noise comes from an amplitude and phase damping channel. In Fig. 13, we also show simulation data using the same noise channel but an ideal unitary ($\epsilon_{\text{coh}} = 0$) in red, and using the same unitary without any incoherent noise ($\epsilon_{\text{incoh}} = 0$) in green. The linear and purely quadratic behaviors of these curves, respectively, can be easily understood from the quadratic form of Eq. (8) where the incoherent errors build up linearly with n, whereas the coherent errors build up quadratically.

Since we simulate the actual 16 circuits used to obtain the average gate fidelity at different depths, we capture some

depth-dependent behaviors that are due to coherent effects in the single- and two-qubit unitaries (see, for example, the green cross at $n = 1$ in Fig. 13, which is actually higher than the point at $n = 0$). This is not an artifact of the finite sampling, but is due to the fact that the preparation and measurement circuits both require an imperfect CZ gate, which can coherently map the state closer to or further away from the desired state $|00\rangle$, depending on the specific imperfect unitaries. Similarly, part of the incoherent noise channel can anticommute with the CZ unitary and produce back-and-forth behaviors between the odd and even depths n. We have found in simulation that using only the even depths avoids this issue, while performing similarly or better in terms of accuracy in error budgeting, compared with using all the even and odd depths.

APPENDIX F: DYNAMICAL DECOUPLING IN CAFE

In this Appendix, we show how applying an X gate to both qubits after the CZ in the cycle circuit decouples the FSIM unitary from both of its single-qubit phases. First, we can show that interleaving a pair of FSIM gates with parallel X gates gives a cycle unitary that also corresponds to an FSIM gate. To see this, consider what happens when conjugating an FSIM unitary by parallel X's:

$$(X \otimes X)\text{FSIM}(X \otimes X) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\gamma}u_{11} & e^{-i\gamma}u_{12} & 0 \\ 0 & e^{-i\gamma}u_{21} & e^{-i\gamma}u_{22} & 0 \\ 0 & 0 & 0 & e^{-i(2\gamma+\phi)} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{F1}$$

$$= \begin{pmatrix} e^{-i(2\gamma+\phi)} & 0 & 0 & 0 \\ 0 & e^{-i\gamma}u_{22} & e^{-i\gamma}u_{21} & 0 \\ 0 & e^{-i\gamma}u_{12} & e^{-i\gamma}u_{11} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{F2}$$

where we have factored the FSIM unitary into phases between the particle-number subspaces and a special unitary U in the single-excitation subspace

$$U = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \begin{pmatrix} e^{-i\zeta}\cos\theta & -ie^{i\chi}\sin\theta \\ -ie^{-i\chi}\sin\theta & e^{i\zeta}\cos\theta \end{pmatrix}. \tag{F3}$$

Following this with a second FSIM gate gives

$$\text{FSIM}(X \otimes X)\text{FSIM}(X \otimes X) = e^{-i2\gamma} \begin{pmatrix} e^{-i\phi} & 0 & 0 & 0 \\ 0 & \tilde{u}_{11} & \tilde{u}_{12} & 0 \\ 0 & \tilde{u}_{21} & \tilde{u}_{22} & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix}, \tag{F4}$$

$$UXUX = \tilde{U} = \begin{pmatrix} \tilde{u}_{11} & \tilde{u}_{12} \\ \tilde{u}_{21} & \tilde{u}_{22} \end{pmatrix} = \begin{pmatrix} u_{11}u_{22} + u_{12}^2 & u_{11}(u_{21} + u_{12}) \\ u_{22}(u_{21} + u_{12}) & u_{11}u_{22} + u_{21}^2 \end{pmatrix}. \tag{F5}$$

Here, the common phase γ has been eliminated (only appearing as a global phase), and the controlled phase ϕ is a relative phase between the even- and odd-parity subspaces. For the case of something close to a CZ gate (small swap angle θ and small
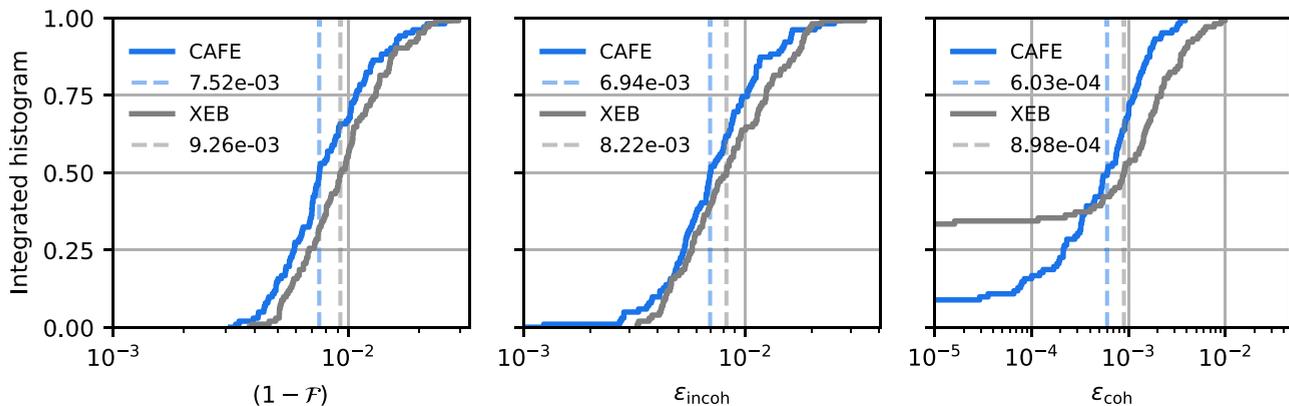
FIG. 14. Error budgets of parallel CZ gates on a Sycamore processor. Data in blue are obtained from the CAFE experiment using the procedure detailed in the main text, where three samples of these data are shown in Fig. 3. For reference, we present similar gate error budget information as obtained from standard XEB in gray, where the incoherent error is extracted from the speckle purity and the coherent error is obtained from the difference with the total gate error. The data show 103 different two-qubit gates.

differential phase ζ), the single-excitation unitary simplifies to

$$U = I - i(\theta \cos \chi X - \theta \sin \chi Y + \zeta Z) + O(\zeta^2) + O(\theta^2),$$
(F6)

$$XUX = I - i(\theta \cos \chi X + \theta \sin \chi Y - \zeta Z) + O(\zeta^2) + O(\theta^2),$$
(F7)

$$UXUX = I - i2\theta \cos \chi X + O(\zeta^2) + O(\theta^2) + O(\zeta\theta).$$
(F8)

As such, the effects of the differential phase ζ, to first order, are also removed by this echoing.

APPENDIX G: SWAP ERRORS WITHOUT PHASE MATCHING

When running CAFE with characterized unitaries in the experiments performed in this work, we do not attempt to correct the swap errors. This is for two reasons. First, the swap errors are the smallest errors in the system and tend to be overshadowed by the other sources of error. Second, in our system, the relative phases accumulated between qubits sitting at different idle frequencies are accounted for by changing the microwave phases. This is possible because the two-qubit gate we are using, the CZ gate, ideally commutes with such differential phases. The effect of not "phase matching" (that is, removing this differential phase by shifting qubit frequencies) on swap errors is that the swap phase χ is shifted from cycle to cycle, preventing the swaps from coherently adding and further diminishing their effect relative to the other error sources. Should one be concerned with relatively large swap angles θ, one could incorporate the application-dependent value of χ in the determination of the circuit to map the predicted state back to $|00\rangle$.

APPENDIX H: ADDITIONAL CZ DATA

For completeness, we present in Fig. 14 the entire error budget results from the parallel CZ CAFE data set acquired on a Sycamore processor. These data illustrate that CAFE can be straightforwardly deployed on large-scale quantum processors to characterize the fidelity of quantum operations in parallel, while budgeting the incoherent and coherent error contributions. For reference, we also present an error budgeting of the same CZ gates obtained from XEB. Note that in this latter case, the coherent error contribution is obtained by subtracting the incoherent speckle purity error from the total XEB error, which gives an unphysical negative ϵ_{incoh} value for 33 out of the 103 gates. In our tests, this problem did not arise for CAFE.

[1] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).

[2] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

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

[4] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, New J. Phys. **16**, 103032 (2014).

[5] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Characterizing errors on qubit operations via iterative randomized benchmarking, Phys. Rev. A **93**, 012301 (2016).

[6] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Scalable randomized benchmarking of quantum computers using mirror circuits, Phys. Rev. Lett. **129**, 150502 (2022).

[7] A. M. Polloreno, A. Carignan-Dugas, J. Hines, R. Blume-Kohout, K. Young, and T. Proctor, A theory of direct randomized benchmarking, arXiv:2302.13853.

[8] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, Nat. Phys. **14**, 595 (2018).

[9] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[10] Y. Gu, W.-F. Zhuang, X. Chai, and D. E. Liu, Benchmarking universal quantum gates via channel spectrum, Nat. Commun. **14**, 5880 (2023).

[11] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical characterization of quantum devices without tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[12] S. T. Flammia and Y.-K. Liu, Direct fidelity estimation from few Pauli measurements, Phys. Rev. Lett. **106**, 230501 (2011).

[13] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

[14] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2002).

[15] C. H. Baldwin, A. Kalev, and I. H. Deutsch, Quantum process tomography of unitary and near-unitary maps, Phys. Rev. A **90**, 012110 (2014).

[16] G. M. D'Ariano, M. G. Paris, and M. F. Sacchi, Quantum tomography, in *Advances in Imaging and Electron Physics* (Elsevier, New York, 2003), Vol. 128, pp. 205–308.

[17] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 062119 (2013).

[18] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921.

[19] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[20] J. Kelly, P. O'Malley, M. Neeley, H. Neven, and J. M. Martinis, Physical qubit calibration on a directed acyclic graph, arXiv:1803.03226.

[21] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing context-dependent errors in quantum processors, Phys. Rev. X **9**, 021045 (2019).

[22] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li *et al.*, Strong quantum computational advantage using a superconducting quantum processor, Phys. Rev. Lett. **127**, 180501 (2021).

[23] P. Murali, D. C. Mckay, M. Martonosi, and A. Javadi-Abhari, Software mitigation of crosstalk on noisy intermediate-scale quantum computers, in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems: ASPLOS'20* (ACM Press, New York, 2020), pp. 1001–1016.

[24] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, Quantum **4**, 321 (2020).

[25] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G. J. Norris, C. K. Andersen, M. Müller, A. Blais, C. Eichler, and A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, Nature (London) **605**, 669 (2022).

[26] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).

[27] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, Phys. Rev. X **11**, 041058 (2021).

[28] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[29] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[30] H. Zhu, Y. S. Teo, and B.-G. Englert, Two-qubit symmetric informationally complete positive-operator-valued measures, Phys. Rev. A **82**, 042308 (2010).

[31] D. Lu, H. Li, D.-A. Trottier, J. Li, A. Brodutch, A. P. Krismanich, A. Ghavami, G. I. Dmitrienko, G. Long, J. Baugh, and R. Laflamme, Experimental estimation of average fidelity of a Clifford gate on a 7-qubit quantum processor, Phys. Rev. Lett. **114**, 140505 (2015).

[32] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, New J. Phys. **17**, 113020 (2015).

[33] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, Phys. Rev. A **58**, 2733 (1998).

[34] L. Viola and E. Knill, Robust dynamical decoupling of quantum systems with bounded controls, Phys. Rev. Lett. **90**, 037901 (2003).

[35] E. Genois, J. A. Gross, Z.-P. Cian, D. M. Debroy, and Z. Jiang, Matrix element amplification using dynamical decoupling (unpublished).

[36] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, A. Bengtsson, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, Y.-A. Chen, B. Chiaro, R. Collins, S. J. Cotton, W. Courtney *et al.*, Observation of separated dynamics of charge and spin in the Fermi-Hubbard model, arXiv:2010.07965.

[37] C. Neill, T. McCourt, X. Mi, Z. Jiang, M. Y. Niu, W. Mruczkiewicz, I. Aleiner, F. Arute, K. Arya, J. Atalaya, R. Babbush, J. C. Bardin, R. Barends, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell *et al.*, Accurately computing the electronic properties of a quantum ring, Nature (London) **594**, 508 (2021).

[38] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Tunable coupling scheme for implementing high-fidelity two-qubit gates, Phys. Rev. Appl. **10**, 054062 (2018).

[39] Google Quantum AI, Exponential suppression of bit or phase errors with cyclic error correction, Nature (London) **595**, 383 (2021).

[40] Cirq Developers, Cirq, 2022, https://github.com/quantumlib/Cirq/graphs/contributors.