UC Berkeley
UC Berkeley Previously Published Works

Title

Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor

Permalink

Journal

ISSN

Authors

Hashim, Akel
Naik, Ravi K
Morvan, Alexis
et al.

Publication Date

DOI

Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor

Akel Hashim[1,2,3,*,†] Ravi K. Naik,[1,3,†] Alexis Morvan[1,3] Jean-Loup Ville,[1] Bradley Mitchell,[1,3]
John Mark Kreikebaum,[1,4,‡] Marc Davis[3] Ethan Smith,[3] Costin Iancu,[3] Kevin P. O'Brien[5]
Ian Hincks,[6,7] Joel J. Wallman[6,7,8] Joseph Emerson,[6,7,8] and Irfan Siddiqi[1,3,4]

[1]*Quantum Nanoelectronics Laboratory, Department of Physics,*
University of California at Berkeley, Berkeley, California 94720, USA
[2]*Graduate Group in Applied Science and Technology,*
University of California at Berkeley, Berkeley, California 94720, USA
[3]*Computational Research Division, Lawrence Berkeley National Lab, Berkeley, California 94720, USA*
[4]*Materials Sciences Division, Lawrence Berkeley National Lab, Berkeley, California 94720, USA*
[5]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,*
Cambridge, Massachusetts 02142, USA
[6]*Quantum Benchmark Inc., Kitchener, Ontario N2H 5G5, Canada*
[7]*Keysight Technologies Canada, Kanata, Ontario K2K 2W5, Canada*
[8]*Institute for Quantum Computing and Department of Applied Mathematics,*
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

The successful implementation of algorithms on quantum processors relies on the accurate control of quantum bits (qubits) to perform logic gate operations. In this era of noisy intermediate-scale quantum (NISQ) computing, systematic miscalibrations, drift, and crosstalk in the control of qubits can lead to a coherent form of error that has no classical analog. Coherent errors severely limit the performance of quantum algorithms in an unpredictable manner, and mitigating their impact is necessary for realizing reliable quantum computations. Moreover, the average error rates measured by randomized benchmarking and related protocols are not sensitive to the full impact of coherent errors and therefore do not reliably predict the global performance of quantum algorithms, leaving us unprepared to validate the accuracy of future large-scale quantum computations. Randomized compiling is a protocol designed to overcome these performance limitations by converting coherent errors into stochastic noise, dramatically reducing unpredictable errors in quantum algorithms and enabling accurate predictions of algorithmic performance from error rates measured via cycle benchmarking. In this work, we demonstrate significant performance gains under randomized compiling for the four-qubit quantum Fourier transform algorithm and for random circuits of variable depth on a superconducting quantum processor. Additionally, we accurately predict algorithm performance using experimentally measured error rates. Our results demonstrate that randomized compiling can be utilized to leverage and predict the capabilities of modern-day noisy quantum processors, paving the way forward for scalable quantum computing.

Subject Areas: Quantum Physics, Quantum Information

---

*To whom correspondence should be addressed.
ahashim@berkeley.edu

†These authors contributed equally to this work.

‡Present address: Google Quantum AI, Mountain View, California, USA.

I. INTRODUCTION

The accuracy of quantum algorithms is limited by different types of errors. Interactions between qubits and the surrounding environment result in incoherent (i.e., nonunitary or irreversible) errors, leading to purity-decreasing processes such as the decoherence of a quantum state. In contrast, systematic imperfections in qubit control (e.g., detuning and calibration errors) and crosstalk on multiqubit processors result in coherent (i.e., unitary or reversible) errors, which are purity-preserving and thus do not result in decoherence. For single qubits, coherent errors manifest as an unwanted unitary rotation by an angle ϵ,

$U(\hat{\mathbf{n}}, \epsilon) = e^{-i\epsilon\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}/2}$, where $\hat{\mathbf{n}}$ is the axis of rotation and $\boldsymbol{\sigma}$ the Pauli vector. Common methods for measuring the average error rate $r(\mathcal{E})$ of quantum gates, such as randomized benchmarking [1–4] (RB), define $r(\mathcal{E})$ as the average gate infidelity,

$$r(\mathcal{E}) = 1 - \mathcal{F} = 1 - \int d\psi \langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle, \quad (1)$$

where the fidelity $\mathcal{F}$ gives the average success probability that preparing an arbitrary pure state $\rho = |\psi\rangle\langle\psi|$ and then evolving the state through a noisy channel $\mathcal{E}(|\psi\rangle\langle\psi|)$ will return the system to the original state. While $r(\mathcal{E})$ captures the average gate infidelity, if coherent errors account for even a small fraction (e.g., ~10%) of the average total error rate [e.g., $r(\mathcal{E}) \sim 10^{-4}$], the worst-case gate infidelity (defined via the diamond norm; see Supplemental Material [5]) can scale as $\sqrt{r(\mathcal{E})}$ (e.g., ~10^{-2}) [11–14]. Thus, the average-case and worst-case infidelities of a single computational gate can differ by orders of magnitude in the presence of coherent errors, as has been explicitly demonstrated for the quantum processor used in this work using simultaneous gate-set tomography [15]. Therefore, the global impact of coherent errors is hard to predict for structured circuits [16] because of both their quadratically worse impact on gate infidelities relative to

average error rates and the potential for interference over the course of an algorithm.

In recent years, there has been growing theoretical interest in randomization methods to mitigate the problem of coherent errors in quantum computations [17–24]. Experimentally, it has been shown that methods such as Pauli-frame randomization [17,18] (PFR) and Pauli twirling can reduce coherent errors in Clifford circuits [25] and the two-qubit CPHASE gate [26], respectively, as measured by gate-set tomography [27–30]. Randomized compiling [20] (RC) is a protocol for reducing coherent error rates in quantum algorithms *in situ*, which is more scalable and generalizable than PFR and simple Pauli twirling, and does not require *a priori* knowledge of the specific error model. In this work, we demonstrate the experimental implementation of RC on a superconducting quantum processor [see Fig. 1(a)]. We show that RC effectively reduces and stabilizes the otherwise unpredictable impact of actual performance-limiting coherent errors in the quantum Fourier transform [31] (QFT) algorithm and in random circuits of variable depth sampled from a universal gate set. Furthermore, we accurately predict algorithm performance under RC from error rates measured in a scalable manner via cycle benchmarking [32] and show how RC performance gains are expected to improve as error rates in
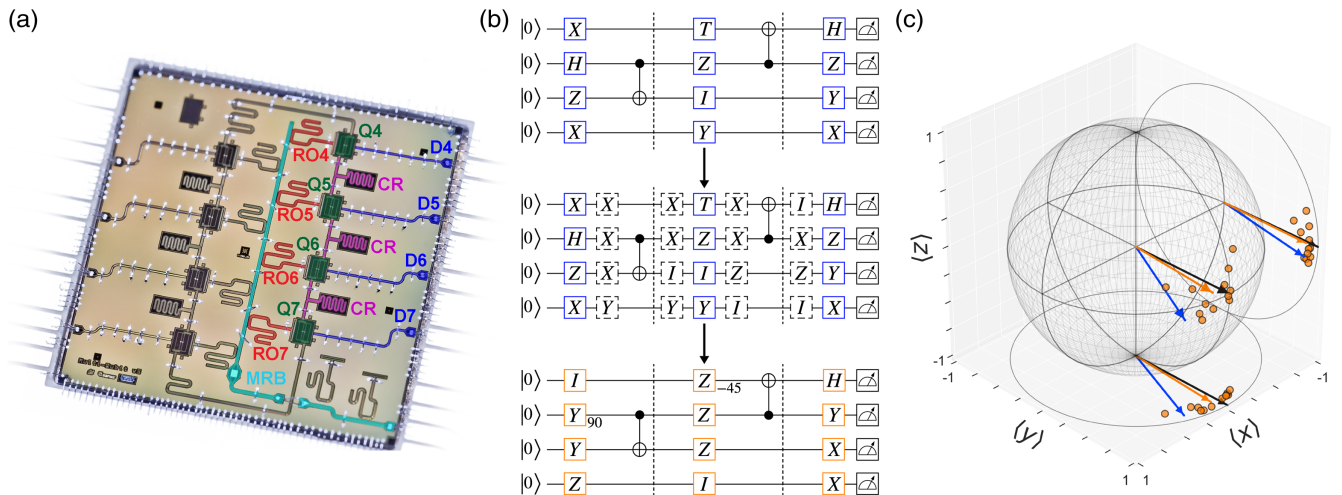

FIG. 1. Experimental realization of noise tailoring via randomized compiling on a superconducting quantum processor. (a) False-colored micrograph of our eight-qubit superconducting quantum processor. In this work, we use four transmon qubits [33] (green) with independent microwave control lines (blue). Two-qubit cross-resonance [34–37] gates are mediated by coupling resonators (CR, purple) between nearest neighbors. The qubits are simultaneously measured via dispersive coupling [38] to independent readout resonators (RO, red) coupled to a multiplexed readout bus (MRB, cyan). (b) Randomization of a quantum circuit. The bare circuit (top), split into K interleaved cycles of easy and hard gates (separated by dashed lines), (middle) is converted into a logically equivalent circuit by inserting random single-qubit twirling gates between each easy and hard cycle, inverting them in the following cycle, and (bottom) then compiling the twirling gates into a new easy-gate cycle. (c) Experimental single-qubit state-tomography results demonstrating noise tailoring. The black vector is the ideal (noiseless) final state of the qubit, but coherent errors cause an over-rotation in the measured state (blue vector). The orange vector represents the final state of the combined distribution of $N = 12$ randomizations (orange data points), which has a lower purity because of the tailored stochastic noise. RC significantly mitigates the impact of coherent errors, as indicated by a reduction in the TVD from $d_{\text{TV,bare}} = 0.170(8)$ to $d_{\text{TV,RC}} = 0.029(2)$ in the $\{|+\rangle, |-\rangle\}$ basis, from $d_{\text{TV,bare}} = 0.069(4)$ to $d_{\text{TV,RC}} = 0.060(1)$ in the $\{|+i\rangle, |-i\rangle\}$ basis, and from $d_{\text{TV,bare}} = 0.073(8)$ to $d_{\text{TV,RC}} = 0.008(2)$ in the $\{|0\rangle, |1\rangle\}$ (computational) basis. This improvement is not captured by the bare ($\mathcal{F} = 0.862$) and RC ($\mathcal{F} = 0.879$) state fidelities, which are approximately equal.

quantum processors continue to decrease, paving the way for more robust large-scale quantum computation.

II. RANDOMIZED COMPILING PROTOCOL

RC tailors coherent errors into a stochastic noise channel by combining the results of many logically equivalent circuits. By inserting and compiling random single-qubit (virtual) twirling gates into a circuit in a way that preserves the overall unitary operation, RC creates a set of "randomized" circuits that are logically equivalent to the original "bare" circuit, without increasing circuit depth. Any bare circuit composed of K cycles of interleaved single-qubit "easy" gates and two-qubit "hard" gates can be randomized using the following method, shown in Fig. 1(b):

  (1) Conjugate each round of easy gates C_k by a twirling gate T_k randomly sampled from a set $\mathcal{T}$ and an inverting operator T_{k-1}^c: $C_k \to T_k C_k T_{k-1}^c$, where T_{k-1}^c is chosen to undo the twirling gate that was inserted in the previous cycle when commuted through the hard-gate cycle G_k: $T_k^c = G_k T_k^\dagger G_k^\dagger$.
  (2) Compile the original single-qubit gates and twirling gates into new easy-gate cycles: $C_k' = T_k C_k T_{k-1}^c$.

In step 1, it is necessary to consider how the tensor product of inverting gates is commuted through two-qubit gates in hard-gate cycles. Typically, $\mathcal{T}$ is chosen to be the set of tensor products of single-qubit Paulis, with the edge terms T_0^c and T_K set to the identity gate; thus, if the two-qubit gates are all Clifford gates (or locally equivalent to Cliffords), then the correction gates will also lie in $\mathcal{T}$, and we need place no restriction on the types of allowed easy gates. Therefore, RC is efficiently compatible with universal quantum computation. However, if the two-qubit gates in hard-gate cycles are not locally equivalent to Clifford gates, then refocusing pulses will be needed, which can result in additional arbitrary two-qubit correction gates at the end of the circuit. No such restrictions are placed on single-qubit gates in hard-gate cycles, which can be arbitrary. In step 2, the new randomized circuit is logically equivalent to the original bare circuit and has the same number of elementary gates. Because the correction gates are computed locally for each cycle, the classical resource requirements for each randomization scale linearly in the number of qubits and circuit depth. Therefore, generating many (N) logically equivalent randomizations of a bare circuit requires very low classical overhead and can be efficiently done before runtime. While certain hardware platforms (e.g., superconducting circuits) may be better equipped to measure large N than others (e.g., trapped ions) because of faster gate times, quantum hardware capable of modifying pulse phases on the fly will enable the utilization of a new randomization per shot [25] on platforms whose single-qubit gates only differ by a change in virtual phases [39].

By measuring each randomization m/N times and computing the union of all N results, we obtain an equivalent statistical distribution for a circuit measured m times in which coherent errors in each computational cycle (except the last) have been averaged into Pauli channels (e.g., random phase and bit flips),

$$\mathcal{E}(\rho) = \sum_{P \in \mathbb{P}^{\otimes n}} c_P P \rho P^\dagger, \qquad (2)$$

where ρ is an n-qubit density matrix, $\mathbb{P}^{\otimes n} = \{I, X, Y, Z\}^{\otimes n}$ the set of 4^n generalized Pauli operators, and c_P the relative probability of an error due to P. Tailoring coherent errors into stochastic Pauli noise has several major advantages: (1) The tailored noise completely suppresses off-diagonal terms in the error process resulting from coherent errors (in the limit of perfectly implemented Pauli twirling; see Supplemental Material [5]), reducing the overall error rate per computational gate cycle. (2) Stochastic Pauli errors have a finite probability of occurring in each gate cycle and only grow linearly with circuit depth (in the small error limit), in contrast to coherent errors which can accumulate up to quadratically with circuit depth in the rare instances in which there is complete constructive interference; thus, RC stabilizes the error rate during algorithms by breaking up the coherent accumulation of unitary errors. (3) Stochastic noise has dramatically lower worst-case error rates than coherent errors occurring at the same average rate, as defined via the diamond norm (see Supplemental Material [5]), and can be directly estimated via randomized benchmarks in order to compare experimental error rates to fault-tolerant error rate thresholds based on Pauli noise. And (4), known fault-tolerant thresholds for stochastic noise [40,41] are orders of magnitude higher than the threshold for generic local errors [42] (for example, those due to coherent errors), potentially enabling fault-tolerant error correction with error rates comparable to those already achieved in modern-day experiments.

To demonstrate noise tailoring via RC, we perform state tomography on a single qubit (Q7) after 50 random gates (see Appendix A), as shown in Fig. 1(c). We find that coherent errors cause a net over-rotation in the measured state compared to the ideal (noiseless) final state. When RC is applied, each randomization results in a different net coherent error; however, the combined result is more aligned with the ideal vector. The state fidelity $\mathcal{F} = 0.862$ and purity $\gamma = 0.938$ for the bare result, and $\mathcal{F} = 0.879$ and $\gamma = 0.881$ for the RC result. While the fidelities are comparable, the rotation error in the bare result has been tailored into stochastic noise under RC, as the fidelity and purity of the RC result are approximately equal in magnitude.

To evaluate the efficacy of RC, we assess algorithmic performance by the total variation distance (TVD), a standard metric for the statistical distance between two probability distributions and a relevant measure in quantum supremacy experiments [43]:

$$d_{\mathrm{TV}}(\mathcal{P}, \mathcal{P}_{\mathrm{ideal}}) = \frac{1}{2} \sum_{x \in X} |\mathcal{P}(x) - \mathcal{P}_{\mathrm{ideal}}(x)|, \qquad (3)$$

where $\mathcal{P}_{\text{ideal}}(x)$ is the ideal probability of measuring a bit string x in a set of possible bit strings X, and $\mathcal{P}(x)$ is the experimentally measured probability. The TVD is a basis-dependent metric that determines the probability of obtaining an incorrect solution, with 0 (1) indicating that the correct distribution of bit strings is always (never) measured. Thus, improvements in algorithmic performance equate to lower TVDs, as exemplified by the observed reduction from $d_{\text{TV,bare}} = 0.073(8)$ to $d_{\text{TV,RC}} = 0.008(2)$ for the single-qubit results in Fig. 1(c), as measured in the computational basis.

As a norm-based error metric, the TVD is generally sensitive to the off-diagonal terms in an error process resulting from coherent errors, in contrast to fidelity-based measures, which are only sensitive to the diagonal terms (see Supplemental Material [5]). Therefore, in the presence of coherent errors, the TVD can be as large as

$$d_{\text{TV}}(\mathcal{P}, \mathcal{P}_{\text{ideal}}) \leq \sqrt{r(\mathcal{E})}\sqrt{d(d+1)}, \qquad (4)$$

but under RC, it is instead upper-bounded directly by the average error rate $r(\mathcal{E})$,

$$d_{\text{TV}}(\mathcal{P}_{\text{RC}}, \mathcal{P}_{\text{ideal}}) \leq r(\mathcal{E})\frac{d+1}{d}, \qquad (5)$$

which is quadratically lower in $r(\mathcal{E})$ and does not scale with the dimension $d = 2^n$ (n qubits). Thus, RC provides a general error reduction from $\sqrt{r(\mathcal{E})} \rightarrow r(\mathcal{E})$. While many NISQ applications [e.g., variational quantum algorithms (VQAs)] depend on measured expectation values as opposed to the TVD, the TVD upper bounds the absolute error of all expectation values measured in the same basis (see Supplemental Material [5]); therefore, a small TVD under RC guarantees a small error in any expectation value estimated from the same probability distribution. Because the TVD depends on simulating the ideal results of a quantum circuit, it is not a scalable error metric. Therefore, it is important to benchmark the TVD performance under RC while quantum circuits can still be efficiently classically simulated, which will inform what TVD improvements can be expected under RC in the post-NISQ era.

III. CYCLE ERROR RECONSTRUCTION

Because RC tailors noise within a quantum circuit, it is useful to benchmark the tailored noise within our multi-qubit system. To do so, we use cycle benchmarking [32] (CB), a scalable protocol that measures errors affecting all qubits during parallel gate cycles. Much like interleaved RB [44], CB interleaves the target cycle of interest between cycles of (tensor products of) random single-qubit Pauli gates and measures an exponential decay as a function of sequence length. A different exponential decay is produced for each state in which the qubit is prepared and measured (i.e., Pauli decay channel). Therefore, the total process

fidelity under CB is the weighted average over all $M \leq 4^n$ measured Pauli decay channels (out of a total of 4^n possible channels for n qubits). The number of Pauli channels that need to be sampled to estimate the process fidelity is constant in the number of qubits and only sets the desired precision in the fidelity estimate [32]. Furthermore, the complexity of fully characterizing an n-qubit system via CB is only polynomially expensive because it depends only on the number of distinct hard-gate cycles one wishes to characterize for any given application.

CB has several advantages over common RB protocols: First, CB enables one to benchmark composite gate cycles, giving more accurate estimates of cycle performance in the context of parallel operations within a quantum circuit; this includes capturing the impact of coherent errors on idling spectator qubits, such as those not explicitly involved in an entangling gate [45]. Second, in contrast to common RB protocols, which perform Clifford twirling, CB performs Pauli twirling, thus producing a stochastic Pauli channel [Eq. (2)], much like RC. Therefore, the effective noise of any cycle under CB is equal to the tailored noise under RC, enabling accurate predictions of algorithmic performance under RC via CB process infidelities.

In this work, we leveraged a cycle error reconstruction (CER) protocol [32,46] based on targeted CB measurements to produce an error map of the Pauli error rates in our four-qubit system. Figure 2(a) outlines the process by which CB can be used to reconstruct k-body gate errors that occur during any hard-gate cycle. The error rates in CER are the coefficients c_P of a Pauli channel in the Kraus representation [Eq. (2)], which can be reconstructed from the Pauli eigenvalues [in the Pauli transfer matrix (PTM) superoperator representation] measured by CB via linear inversion using an inverse Walsh-Hadamard transform [46]. Using CER, we identify the major sources of errors in our system and compensate the most harmful effects with targeted decoupling pulses or virtual phase gates (see Supplemental Material [5] for more details). In Fig. 2(b), we plot one- and two-body Pauli (Kraus) errors reconstructed via CER for the hard-gate cycles in our system after a targeted gate tuneup, showing that the residual error syndromes are broadly distributed, collectively contributing to the process infidelity of each cycle and making further targeted error mitigation less fruitful. By limiting measurements to only one- and two-body errors and making use of marginalized probability distributions, it is not necessary to measure all $M = 4^n$ Pauli decay channels under CER. More generally, it is possible to reconstruct all one-body marginals with $\mathcal{O}(1)$ Pauli decays and all two-body marginals with $\mathcal{O}[\log(n)]$ Pauli decays for n qubits. However, while some k-body errors can efficiently be estimated for $k > 2$, it becomes exponentially expensive to estimate all n-body errors. The results in Fig. 2(b) show that the dominant residual errors in our system are one-body errors. Therefore, even though only one- and two-body errors were measured, the fact that
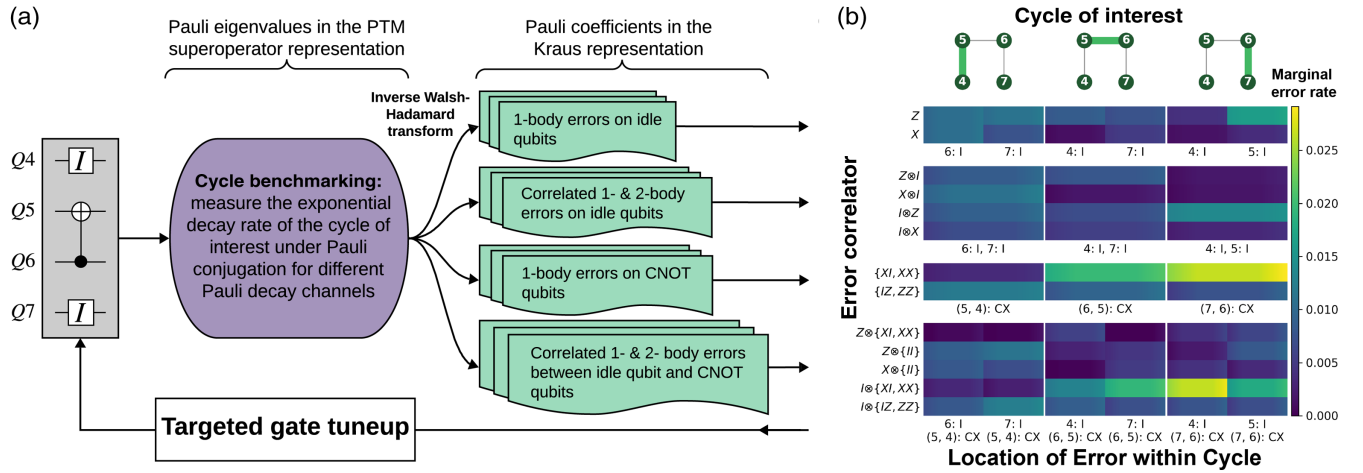
FIG. 2. Cycle error reconstruction of the tailored noise under cycle benchmarking. (a) Schematic of the process by which one- and two-body gate errors can be reconstructed using targeted CB measurements of a parallel gate cycle (e.g., CNOT between Q5 and Q6, with Q4 and Q7 idling), providing detailed information about the Pauli (Kraus) errors occurring during the cycle, as shown in panel (b). (b) Cycle error reconstruction results of four-qubit cycles containing a single CNOT gate and identity gates on the spectator qubits. The y axis (x axis) labels the type of error (where the error occurs in the cycle), and the color (gradient) indicates the marginal error rate from all Pauli contributions (95% confidence interval). The first (third) row of subplots shows one-body errors acting in idling (CNOT) qubits, the second row of subplots shows correlated one- and two-body errors between idling qubits, and the last row of subplots shows correlated one- and two-body errors between an idling spectator qubit and qubits participating in the CNOT. Here, a k-body error is labeled by k nonidentity gates acting on n qubits (e.g., $k = 1$ for CNOT qubits, because these errors occur within a single gate body acting on two qubits). Curly brackets indicate error types that cannot be distinguished due to degeneracies since any local error acting on either qubit in the entangling operation will be transformed by the CNOT. All rows in which all errors are below 30% of the maximum value have been omitted for clarity. This detailed information can be used to perform targeted gate tuneup to address the most harmful errors. The residual errors in our system are broadly distributed among many pathways, so any further targeted tuneup will come with diminishing returns.

two-body errors are negligible shows that three- or more body errors are also negligible. These error rates can be further utilized to simulate and predict algorithm performance under RC, as shown in Fig. 3(c).

IV. QUANTUM FOURIER TRANSFORM

RC can be applied to any gate-based quantum algorithm, including those at the heart of many quantum applications, like the QFT. Here, we utilize a synthesis algorithm [47] to numerically approximate the four-qubit QFT circuit unitary in order to reduce the CNOT count to $K = 13$ for our linear connectivity (see circuit diagram in Supplemental Material [5]). Much like the classical discrete Fourier transform, the QFT maps singular inputs (e.g., $|0000\rangle$) into uniform distributions, and it maps superposition states (e.g., $|+ + ++\rangle$) into singular distributions. To measure the performance of RC for different resultant probability distributions, we apply the QFT to various single-qubit product states involving permutations of Pauli basis states $\{|0\rangle, |1\rangle, |+\rangle\}$, as well as random separable input states $[SU(2)_{\text{rand}}^{\otimes 4}|0000\rangle]$; see Fig. 3(a) for several examples of the measured distributions.

In Figs. 3(a) and 3(b), we show that the relative RC performance is best (equivalent) when the algorithm generates a uniform (singular) distribution across all measurement basis states. This is due to the basis dependence of the TVD: Given a small angle error of θ relative to the ideal final state of a system, the TVD scales as $d_{\text{TV}}(\mathcal{P}, \mathcal{P}_{\text{ideal}}) \simeq \theta^2 \simeq r(\mathcal{E})$ if the target state is in an eigenstate of the measurement basis (i.e., deterministic algorithms), whereas the TVD scales as $d_{\text{TV}}(\mathcal{P}, \mathcal{P}_{\text{ideal}}) \simeq \theta$ [i.e., the $\sqrt{r(\mathcal{E})}$ dependence in Eq. (4)] if the target state is coherently spread across the measurement basis (i.e., nondeterministic algorithms; see Supplemental Material [5]). Therefore, if the (ideal) target state is in an eigenstate of the measurement basis, the raw probabilities will not be sensitive to off-diagonal terms in the error process resulting from coherent errors [which scale as $\theta \simeq \sqrt{r(\mathcal{E})}$], so the TVD will not benefit from the general $\sqrt{r(\mathcal{E})} \to r(\mathcal{E})$ improvement provided by RC. As such, distribution uniformity is a good proxy for the susceptibility of the target state to coherent errors with respect to the measurement basis, and it is thus correlated with improvement under RC. More generally, error assessments that are only sensitive to the diagonal terms in the error process (e.g., process or average gate infidelity, or the TVD of deterministic algorithms) cannot benefit from the suppression of the off-diagonal terms via RC when averaging over many randomizations. In contrast, norm-based error metrics, such as the TVD of nondeterministic algorithms and the diamond
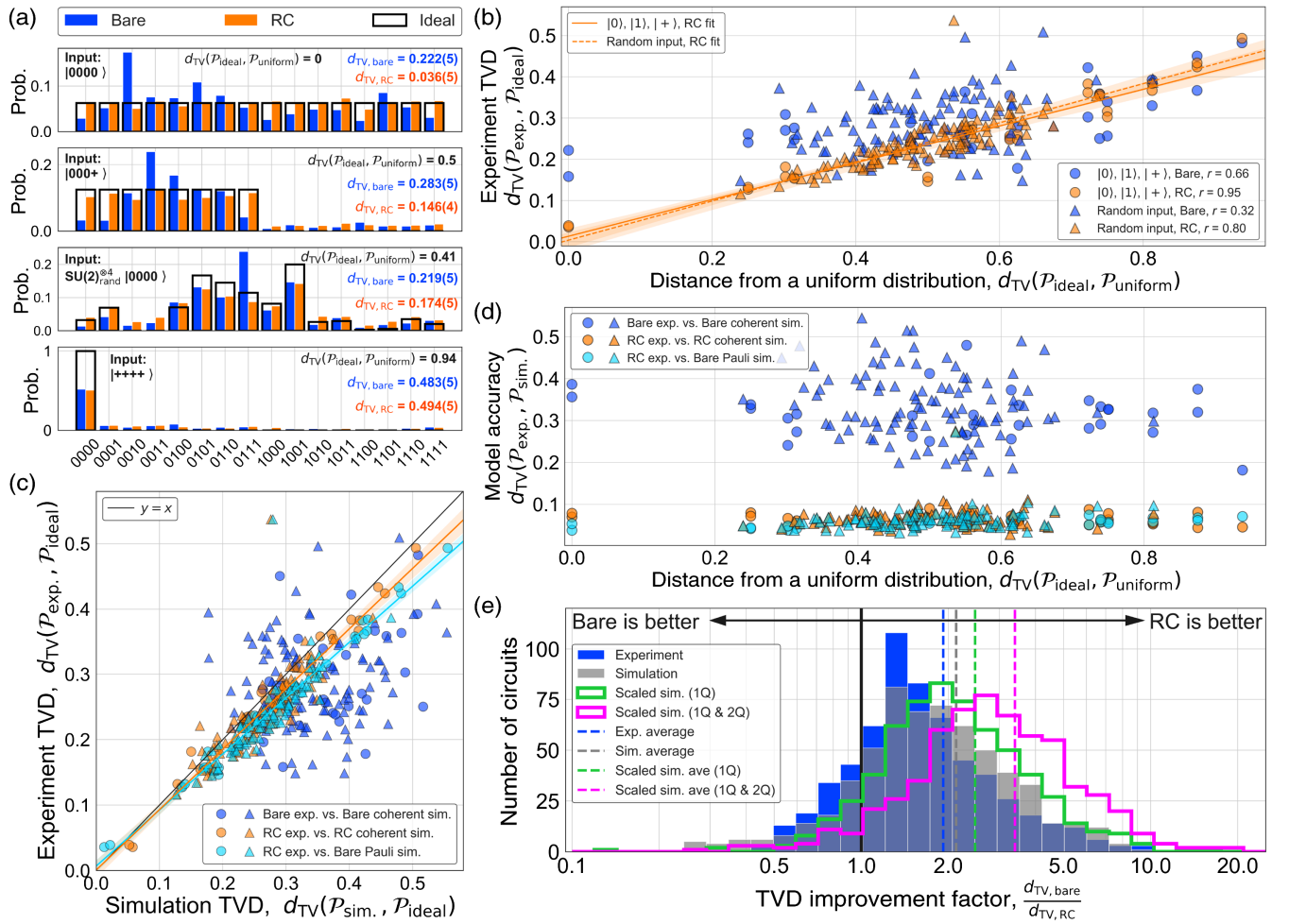
FIG. 3. Improving the quantum Fourier transform with randomized compiling. (a) Measured probability distributions for the QFT applied to $|0000\rangle$, $|000+\rangle$, $|++++\rangle$, and a random input state $[SU(2)_{\text{rand}}^{\otimes 4}|0000\rangle]$. (b) Bare and RC TVDs for all four-qubit QFT results, as a function of distribution uniformity of the ideal results. RC provides more improvement as the resultant distribution spans more basis states $[d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}}) \to 0]$. Circles indicate the QFT applied to Pauli basis states $(\{|0\rangle, |1\rangle, |+\rangle\})$, and triangles indicate random inputs states $[SU(2)_{\text{rand}}^{\otimes 4}|0000\rangle]$; we use the same distinction in panels (c) and (d), but omit the description in the legend for clarity. Pearson r values listed in the legend quantify the correlation strength of each data set, justifying linear fits for the RC data (transparent bands indicate the 95% confidence intervals). (c) Experimental vs simulated TVDs from two models based on the Pauli error rates in Fig. 2(b). The blue (orange) markers denote the bare (RC) circuits simulated with the coherent error model, and the cyan markers denote the bare circuits simulated with the Pauli model. (d) Accuracy of the two models compared to experimental results. The bare circuits simulated with the Pauli model are plotted against the experimental RC results in panel (c), which are also used to compute the model accuracy in panel (d). (e) Summary of the improvement under RC for all two-, three-, and four-qubit random input QFT results, showing good agreement between experiment (blue) and theory (gray). Simulations in which single-qubit (green) and two-qubit (pink) error rates have been scaled down by a factor of 10 suggest that RC performance increases as error rates decrease. (Error bars on the TVD $[\mathcal{O}(10^{-3})]$ for panels (b)–(d) are smaller than the markers.)

distance, will generally be sensitive to the off-diagonal terms in the error process and will thus generally benefit from RC.

We quantify the "distance from a uniform distribution" by computing the TVD of each ideal probability distribution with the uniform distribution in $d = 2^n$ dimensions for n qubits, $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}})$, which is 0 (maximized) when $\mathcal{P}_{\text{ideal}}$ is uniform (singular). In Fig. 3(b), the bare and RC TVDs are plotted as a function of $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}})$

for all four-qubit QFT results. For singular input states ($|0000\rangle$ or $|1111\rangle$), RC significantly reduces the TVD, but for a superposition input state ($|++++\rangle$), the bare and RC TVDs are approximately equal. We compute the Pearson correlation coefficient r to quantify the correlation strength between the experiment TVD and $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}})$, where $+1$ (-1) indicates exact positive (negative) correlation and 0 implies no linear correlation. The RC results are strongly correlated [$r = 0.95(0.80)$ for basis (random)

inputs] compared to the bare results [$r = 0.66(0.32)$ for basis (random) inputs], underscoring the stability and predictability of RC compared to nonrandomized circuits.

In Fig. 3(c), we predict the TVD performance of the QFT using two models: (1) a Pauli model of our system consisting of the Pauli error rates extracted from the CER results in Fig. 2(b); (2) a coherent error model that, under simulated CER, produces approximately equal error rates as the experimental results in Fig. 2(b). The coherent error model is generated by finding a desired completely positive and trace-preserving (CPTP) map for a fixed unitarity [13] (i.e., fixed fraction of the total error rate due to coherent errors), which has been minimized with respect to the experimental CER results (see Supplemental Material [5] for more details). Even though such a model is based on experimental error rates, it does not accurately capture detailed information about coherent errors in our system (e.g., context-dependent rotation axes and angles [15]). Therefore, we do not expect the coherent error model to accurately simulate the performance of any individual bare circuit; rather, we expect it to only capture the average performance of the bare circuits, which is possible as long as the magnitude of the unitarity in the model is correct. We note that, in general, measuring an accurate coherent error model for multiqubit systems with a continuous single-qubit gate set is not experimentally tractable.

Using these models, we simulate the QFT circuits and compare the results to the experimental TVDs in three different ways:

(1) Coherent error model applied to the bare circuits vs experimental bare results.
(2) Coherent error model applied to the RC circuits vs experimental RC results.
(3) Pauli model applied to the bare circuits vs experimental RC results.

For (3), it is unnecessary to simulate the RC circuits with the Pauli model since the Pauli model already contains the tailored error rates. In Fig. 3(c), we see excellent agreement between experiment and simulation for (2) and (3), but unreliable predictability for (1). This indicates that we can accurately predict the results of an algorithm *a priori* with CER error rates using (a) a Pauli model, as long as the experimental circuit is performed using RC, and (b) a coherent error model, as long as both the simulated and experimental circuits are performed using RC. In Fig. 3(d), we validate the model accuracy by computing the TVD of the experimental results with the simulated results. Almost all of the simulated results for (2) and (3) are accurate to within 10% with respect to the experimental RC results, but the accuracy of the simulated results for (1) are much worse. By utilizing RC in conjunction with CB/CER, we avoid the difficulty in capturing and modeling the complex interplay of coherent errors, and thus close the gap between the circuit performance predicted from benchmarking diagnostics and experimental results.

In practice, the input states to quantum algorithms will not be known *a priori*, such as when the QFT is used in

Shor's algorithm [48]. While unknown inputs are not guaranteed to be random, we use random inputs as a proxy for when the QFT is used as a subroutine in other algorithms. Figure 3(b) shows that when the QFT is applied to random input states, most of the results are improved under RC. A histogram of the TVD improvement for two-, three-, and four-qubit random input QFT results can be seen in Fig. 3(e) (two- and three-qubit QFT results are provided in the Supplemental Material [5]), showing that the vast majority of circuits (more than 81%) are improved under RC by an average of $d_{\mathrm{TV,bare}}/d_{\mathrm{TV,RC}} \approx 1.9$. Here, we include two- and three-qubit results in order to summarize the RC QFT performance using a larger sample size drawn from systems that include differing error rates. In the rare instances in which coherent errors in a circuit benignly cancel, RC can hurt performance ($d_{\mathrm{TV,bare}}/d_{\mathrm{TV,RC}} < 1$); however, in general, this becomes vanishingly unlikely for longer depth circuits.

While the coherent error model may not accurately predict the individual result of any given bare circuit, it does predict the average performance relative to RC. To demonstrate this, we compute the TVD improvement under RC ($d_{\mathrm{TV,bare}}/d_{\mathrm{TV,RC}}$) for all of the two-, three-, and four-qubit random input QFT circuits simulated using the coherent error model. As seen in Fig. 3(e), the distribution of improvement under RC predicted by simulation agrees well with experiment, with an overlapping index of 0.94 (out of a maximum of 1), which quantifies the percentage that one normal distribution overlaps with another. The good agreement between experiment and theory in Fig. 3(e) suggests that we can predict the average improvement under RC as error rates decrease. Included in Fig. 3(e) are simulated results in which single-qubit error rates are reduced by a factor of 10, resulting in a modest improvement, and when both single- and two-qubit error rates are reduced by a factor of 10, in which case RC improves more than 94% of the simulated circuits by an average of $d_{\mathrm{TV,bare}}/d_{\mathrm{TV,RC}} \approx 3.4$. In agreement with the predictions made in Ref. [20], these results demonstrate that RC is expected to provide a larger relative improvement as gate infidelities decrease (for a fixed fraction of the total error rate due to coherent errors). Therefore, as quantum processors improve and error rates decrease, we can expect RC to outperform nonrandomized circuits as long as coherent errors persist.

V. RANDOM CIRCUITS OF VARIABLE DEPTH

To illustrate the broad applicability and generic benefits of RC for universal circuits, we demonstrate achievable performance gains for RC applied to four-qubit circuits of variable depth composed of K interleaved cycles of easy and hard gates randomly sampled from a universal (Clifford + T) gate set (see Appendix C). As shown in Fig. 4(a), RC reduces the average TVD at all circuit depths tested (with $N = 20$ randomizations for each bare circuit),
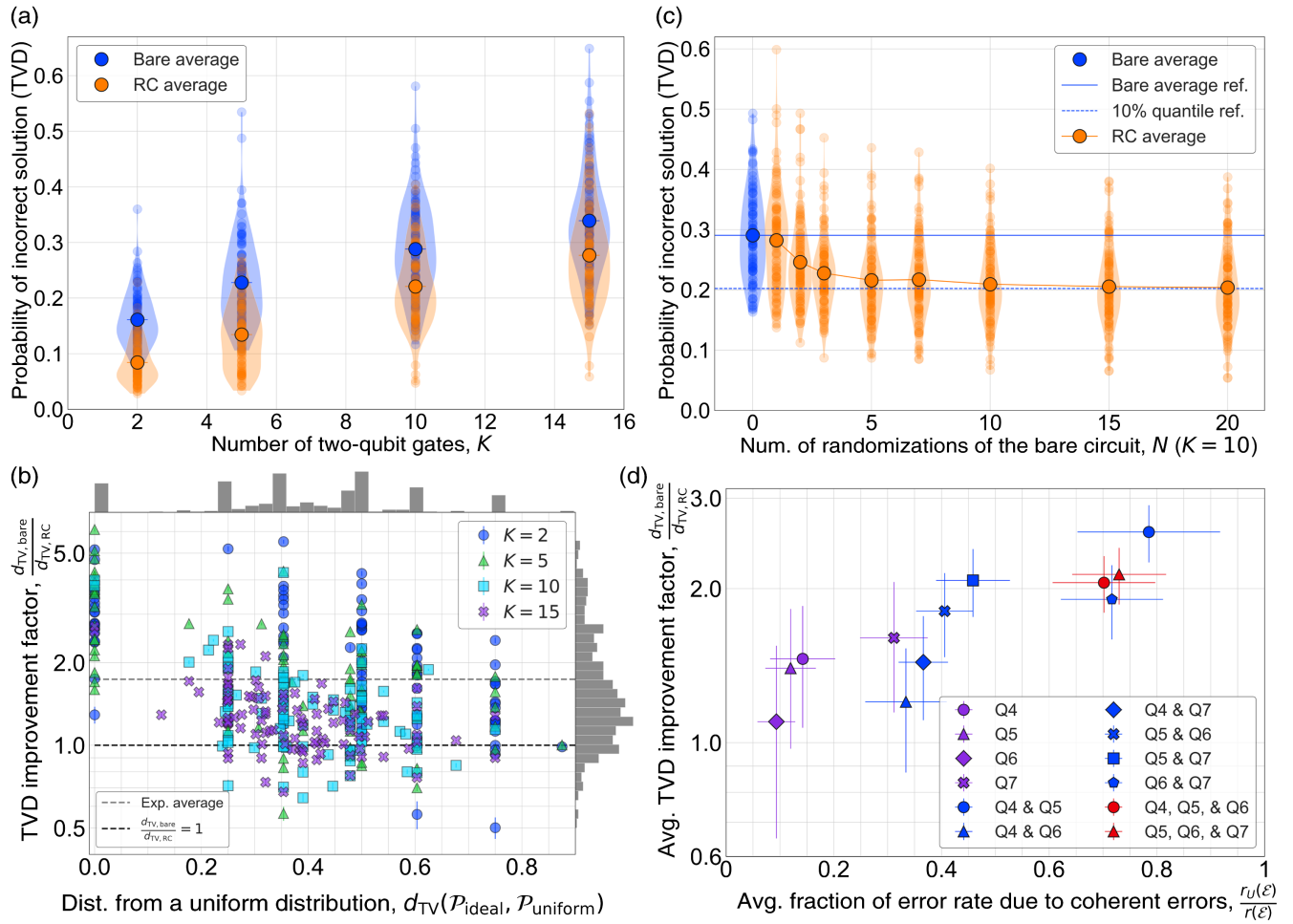
FIG. 4. Randomized compiling extends the computational reach with respect to circuit depth. (a) Bare and RC TVDs as a function of circuit depth K. RC reduces the TVD on average for all circuit depths tested, allowing one to perform longer gate sequences under a fixed TVD error budget. The semitransparent blue (orange) points indicate the TVDs of the individual random circuits (the unioned data over all $N = 20$ randomizations of the corresponding bare circuits). Violin plots depict the distribution of results. The TVD error grows approximately linearly with K for both bare and RC results, suggesting that the dominant reason for an improvement under RC is an overall reduction in the error for each gate cycle K, rather than a suppression of the adversarial accumulation of coherent errors (although RC can additionally provide this benefit). (b) RC TVD improvement factor for the random circuits in panel (a) as a function of distribution uniformity. The average improvement is $d_{\text{TV,bare}}/d_{\text{TV,RC}} \approx 1.7$. (c) TVD as a function of number of randomizations, with $K = 10$ fixed. The average TVD under RC converges to a value close to the 10% quantile level (dashed line) of the nonrandomized circuits for $N = 20$. However, only $N = 10$ randomizations are needed to converge to within 2.7% of the $N = 20$ level. (d) For a fixed total error rate, RC provides a larger TVD improvement for systems with a higher fraction of coherent errors. The colored subsets listed in the legend highlight random single-qubit circuits that were performed in isolation (purple) or in parallel (blue and red). The average fraction of the total error rate due to coherent errors was quantified using measurements of RB and unitary RB under isolated or simultaneous operation.

demonstrating how longer-depth quantum circuits can be performed under RC given a fixed error budget in the TVD. These results highlight an important distinction in how RC can improve algorithm performance: Because random circuits are already robust to the coherent accumulation of unitary errors, we observe a linear (not quadratic) growth in the average TVD as a function of circuit depth for both the bare and RC circuits. Therefore, the reduction in the TVD under RC is not due to the suppression of the adversarial accumulation of coherent errors (although each

randomization under RC can additionally provide this benefit for structured circuits that are susceptible to such errors). Rather, it is due to a reduction in the off-diagonal terms in the error process due to coherent errors per computational gate cycle K when averaging over many randomizations (see Supplemental Material [5]). As previously noted, this improvement is not observed for fidelity-based error metrics [see, for example, the state fidelity results in Fig. 1(c)], which under-represent the global impact of coherent errors on quantum algorithms, as

they are only sensitive to the diagonal terms in the error process. The relative improvement under RC is reduced at longer circuit depths since both the bare and RC results will converge to a uniform distribution (i.e., statistical mixture) due to decoherence in the limit of large K.

The TVD improvement under RC for all of the random circuits in Fig. 4(a) is plotted in Fig. 4(b) as a function of $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}})$, with an average improvement of $d_{\text{TV,bare}}/d_{\text{TV,RC}} \approx 1.7$. Because randomly sampled circuits approach an approximate unitary-2 design as a function of circuit depth [49], we observe that the results are not uniformly spread across $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}})$ for small K; rather, they are highly concentrated at several uniformities and only begin spreading out for larger K. Given the concentration of results in the range between $d_{\text{TV}}(\mathcal{P}_{\text{ideal}}, \mathcal{P}_{\text{uniform}}) \in [0.2, 0.8]$ in both Figs. 3(b) and 4(b), we would expect typical algorithms to fall within this range.

Additionally, we show that a small number of randomizations is sufficient to saturate the lowest-possible TVD under RC for a fixed circuit depth ($K = 10$), plotted in Fig. 4(c). After $N = 20$ randomizations, the average TVD under RC converges to a value that is better than approximately 90% of the nonrandomized circuits. However, after only $N = 10$ randomizations, the average RC TVD is already within 2.7% of the $N = 20$ level, highlighting the resource efficiency of this protocol.

Finally, in Fig. 4(d), we plot the average TVD improvement factor for random single-qubit circuits (at a fixed circuit depth $K = 5$; see Supplemental Material [5]) performed in isolation or in parallel as a function of the average fraction of the total error rate due to coherent errors. From these data, we see that RC provides a larger relative TVD improvement as the fraction of the total error rate due to coherent errors increases (for a fixed total error rate). We measure the average total error rate $r(\mathcal{E})$ using RB and the average error rate due to coherent errors $r_U(\mathcal{E})$ using unitary RB [13,50] (see Supplemental Material [5]). Even though the average total error rates of any two results are not exactly equal, we group the data by the number of qubits performed in parallel (differentiated by color) since $r_U(\mathcal{E})/r(\mathcal{E})$ can be more directly compared across these subsets independently. While RC performance decreases as $r_U(\mathcal{E})/r(\mathcal{E}) \to 0$, we note that even for single-qubit systems that are close to coherence limited $[r_U(\mathcal{E})/r(\mathcal{E}) \lesssim 0.1]$, RC still provides an average improvement. Therefore, while the trade-off between decreased RC improvement as $r_U(\mathcal{E})/r(\mathcal{E}) \to 0$ [for $r(\mathcal{E})$ fixed] and increased RC improvement as $r(\mathcal{E}) \to 0$ [for $r_U(\mathcal{E})/r(\mathcal{E})$ fixed] will depend on each system individually, our results suggests that any system with coherent errors can benefit from RC on average, even those which are nearly coherence limited.

VI. OUTLOOK

In this work, we have demonstrated the promising capabilities of randomized compiling, a universal protocol in gate-based quantum computing for suppressing coherent errors that is agnostic to specific error models and hardware platforms. RC provides a strategy for mitigating complex and intractable crosstalk dynamics, extending the computational reach of noisy quantum processors. Additionally, novel error reconstruction methods using CB are well suited to characterize the new and emergent forms of crosstalk errors seen on multiqubit processors, and they offer a method for accurately predicting error rates under RC. This improved predictability is essential for scalable quantum computing and is necessary for comparing experimental error rates to fault-tolerance thresholds.

We believe that our methods and results have broad relevance across many experimental and theoretical efforts exploring gate-based quantum computing applications, including NISQ algorithms such as VQAs, which depend on the accurate measurement of expectation values. Additionally, while VQAs aimed at finding ground-state energies of quantum systems can converge even in the presence of coherent errors, the true parametrization of the ground-state wave function may be incorrect; we suspect that RC can facilitate faster convergence to the ground-state energy in the presence of coherent errors while also finding the best estimate of the true ground-state wave function of the system. Finally, RC may continue to be useful in the fault-tolerant era since it is expected that, under certain conditions, coherent errors will continue to persist and remain a problem even with quantum error correction [51,52]; utilizing RC will ensure that fault-tolerant error thresholds are set by stochastic errors, not coherent errors. To this end, we expect that RC is not just a stopgap measure in the NISQ era, but will continue to be a powerful technique beyond NISQ.

Supplemental Material [5] is available for this paper. All data are available in the manuscript, supplemental materials, or from the corresponding author upon reasonable request.

designed and J. M. K. fabricated the device. M.D, E. S., and C. I. developed the software to generate the optimized circuits [47]. I. H. provided True-Q [53,54] software support, helped analyze the cycle error reconstruction data, and generated the model simulator. I. S. supervised all work. A. H., R. K. N., J. E., and I. S. wrote the manuscript, with input from all coauthors.

APPENDIX A: SINGLE-QUBIT STATE TOMOGRAPHY

The random circuit used for the single-qubit state tomography results in Fig. 1(c) was generated by randomly sampling $K = 25$ interleaved cycles of easy and hard single-qubit gates, as defined by the following gatesets: the single-qubit Clifford set, $C_{\text{easy}} = \{\mathbf{C_1}\}$, and common non-Clifford gates, $G_{\text{hard}} = \{X45, Y45, T = Z45\}$. State tomography results are reconstructed by performing ensemble measurements of the same final state in the X, Y, and Z bases. For the experimental results presented in Fig. 1(c), 6000 shots were taken in each measurement basis for the bare circuit. Since 12 randomizations of the bare circuit were utilized in the RC result, in order to normalize shot statistics between the bare and RC results, 500 shots were taken for each randomization in each basis.

APPENDIX B: QUANTUM FOURIER TRANSFORM

Each bare QFT circuit was measured 10 000 times. Note that $N = 50$ randomizations were generated for each bare circuit, and each randomization was measured 200 times. All "random" input states were generated by applying random $SU(2)$ unitaries to each qubit independently before applying the QFT algorithm. For the data presented in Fig. 3(b), 100 random inputs were generated.

APPENDIX C: RANDOM CIRCUITS OF VARIABLE DEPTH

Random bare circuits were generated by randomly sampling K interleaved cycles of easy and hard gates from the following gatesets: For the four-qubit circuits in Figs. 4(a)–4(c), $C_{\text{easy}} = \{\mathbf{C_1}, X45, Y45, T\}$ and $G_{\text{hard}} = \{CX = CNOT, CY, CZ\}$, where $\mathbf{C_1}$ is the single-qubit Clifford set. For the single-qubit circuits in Fig. 4(d), $C_{\text{easy}} = \{\mathbf{C_1}\}$ and $G_{\text{hard}} = \{X45, Y45, T\}$. For each circuit depth K, which we define in terms of the number of two-qubit (non-Clifford) gates for the multiqubit (single-qubit) circuits, 100 random bare circuits were generated, and each was measured 4000 times. Note that $N = 20$ randomizations were generated for each random bare circuit, and each randomization was measured 200 times. All $N = 20$

randomizations of the corresponding bare circuits were combined to obtain an equivalent statistical distribution for a circuit measured 4000 times. The circuit depth was fixed at $K = 5$ for the single-qubit results in Fig. 4(d).

For the results presented in Fig. 4(c), a total of $\mathcal{N} = 20$ randomizations were generated for each bare circuit, and each was measured 4000 times. For each N along the x axis, the union over only $N \in \mathcal{N}$ randomizations was computed for each bare circuit, and then $m = 4000$ shots were randomly chosen from a total of $m_{\text{total}} = \sum_i^N m_i$ shots (with the exception of $N = 1$). This is equivalent to throwing out $m_{\text{total}} - 4000$ shots at random from the unioned RC result in order to sample a smaller distribution from the full distribution.

---

[1] J. Emerson, R. Alicki, and K. Życzkowski, *Scalable Noise Estimation with Random Unitary Operators*, J. Opt. B **7**, S347 (2005).

[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] 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).

[4] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and Robust Randomized Benchmarking of Quantum Processes*, Phys. Rev. Lett. **106**, 180504 (2011).

[5] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.11.041039 for a description of the experimental setup, qubit properties, gate error rates, and analysis of cycle benchmarking and cycle error reconstruction, which includes Refs. [6–10]. Also included is an expanded discussion of the impact of coherent errors on various error metrics, as well as supplemental data for the experimental results presented in the main text.

[6] M. Blok, V. Ramasesh, T. Schuster, K. O'Brien, J. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Yao, and I. Siddiqi, *Quantum Information Scrambling on a Superconducting Qutrit Processor*, Phys. Rev. X **11**, 021010 (2021).

[7] J. Kreikebaum, K. O'Brien, A. Morvan, and I. Siddiqi, *Improving Wafer-Scale Josephson Junction Resistance Variation in Superconducting Quantum Coherent Circuits*, Supercond. Sci. Technol. **33**, 06LT02 (2020).

[8] B. K. Mitchell, R. K. Naik, A. Morvan, A. Hashim, J. M. Kreikebaum, B. Marinelli, W. Lavrijsen, K. Nowrouzi, D. I. Santiago, and I. Siddiqi, *Hardware-Efficient Microwave-Activated Tunable Coupling between Superconducting Qubits*, arXiv:2105.05384.

[9] A. Y. Kitaev, *Quantum Computations: Algorithms and Error Correction*, Usp. Mat. Nauk **52**, 53 (1997).

[10] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing Quantum Gates via Randomized Benchmarking*, Phys. Rev. A **85**, 042311 (2012).

[11] J. J. Wallman and S. T. Flammia, *Randomized Benchmarking with Confidence*, New J. Phys. **16,** 103032 (2014).

[12] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, *Bounding Quantum Gate Error Rate Based on Reported Average Fidelity*, New J. Phys. **18,** 012002 (2015).

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

[14] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, *Comparing Experiments to the Fault-Tolerance Threshold*, Phys. Rev. Lett. **117,** 170502 (2016).

[15] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan *et al.*, *Experimental Characterization of Crosstalk Errors with Simultaneous Gate Set Tomography*, arXiv:2103.09890.

[16] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Measuring the Capabilities of Quantum Computers*, arXiv:2008.11294.

[17] E. Knill, *Fault-Tolerant Postselected Quantum Computation: Threshold Analysis*, arXiv:quant-ph/0404104.

[18] O. Kern, G. Alber, and D. L. Shepelyansky, *Quantum Error Correction of Coherent Errors by Randomization*, Eur. Phys. J. D **32,** 153 (2005).

[19] M. R. Geller and Z. Zhou, *Efficient Error Models for Fault-Tolerant Architectures and the Pauli Twirling Approximation*, Phys. Rev. A **88,** 012314 (2013).

[20] J. J. Wallman and J. Emerson, *Noise Tailoring for Scalable Quantum Computation via Randomized Compiling*, Phys. Rev. A **94,** 052325 (2016).

[21] M. B. Hastings, *Turning Gate Synthesis Errors into Incoherent Errors*, arXiv:1612.01011.

[22] E. Campbell, *Shorter Gate Sequences for Quantum Computing by Mixing Unitaries*, Phys. Rev. A **95,** 042306 (2017).

[23] E. Campbell, *Random Compiler for Fast Hamiltonian Simulation*, Phys. Rev. Lett. **123,** 070503 (2019).

[24] Z. Cai, X. Xu, and S. C. Benjamin, *Mitigating Coherent Noise Using Pauli Conjugation*, npj Quantum Inf. **6,** 17 (2020).

[25] M. Ware, G. Ribeill, D. Ristè, C. A. Ryan, B. Johnson, and M. P. da Silva, *Experimental Pauli-Frame Randomization on a Superconducting Qubit*, Phys. Rev. A **103,** 042604 (2021).

[26] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, *Quantum Computation with Universal Error Mitigation on a Superconducting Quantum Processor*, Sci. Adv. **5,** eaaw5686 (2019).

[27] 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).

[28] C. Stark, *Self-Consistent Tomography of the State-Measurement Gram Matrix*, Phys. Rev. A **89,** 052109 (2014).

[29] R. J. Blume-Kohout, J. K. Gamble, E. Nielsen, P. L. W. Maunz, T. L. Scholten, and K. M. Rudinger, *Turbocharging Quantum Tomography*, OSTI https://doi.org/10.2172/1168946 (2015).

[30] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, *Demonstration of Qubit Operations below a Rigorous Fault Tolerance Threshold with Gate Set Tomography*, Nat. Commun. **8,** 14485 (2017).

[31] D. Coppersmith, *An Approximate Fourier Transform Useful in Quantum Factoring*, arXiv:quant-ph/0201067.

[32] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, *Characterizing Large-Scale Quantum Computers via Cycle Benchmarking*, Nat. Commun. **10,** 5347 (2019).

[33] J. Koch, M. Y. Terri, J. Gambetta, A. A. Houck, D. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, *Charge-Insensitive Qubit Design Derived from the Cooper Pair Box*, Phys. Rev. A **76,** 042319 (2007).

[34] G. Paraoanu, *Microwave-Induced Coupling of Superconducting Qubits*, Phys. Rev. B **74,** 140504(R) (2006).

[35] C. Rigetti and M. Devoret, *Fully Microwave-Tunable Universal Gates in Superconducting Qubits with Linear Couplings and Fixed Transition Frequencies*, Phys. Rev. B **81,** 134507 (2010).

[36] 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.*, *Simple All-Microwave Entangling Gate for Fixed-Frequency Superconducting Qubits*, Phys. Rev. Lett. **107,** 080502 (2011).

[37] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, *Procedure for Systematically Tuning up Cross-Talk in the Cross-Resonance Gate*, Phys. Rev. A **93,** 060302(R) (2016).

[38] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, *Strong Coupling of a Single Photon to a Superconducting Qubit Using Circuit Quantum Electrodynamics*, Nature (London) **431,** 162 (2004).

[39] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, *Efficient Z Gates for Quantum Computing*, Phys. Rev. A **96,** 022330 (2017).

[40] P. Aliferis and A. W. Cross, *Subsystem Fault Tolerance with the Bacon-Shor Code*, Phys. Rev. Lett. **98,** 220502 (2007).

[41] P. Aliferis, D. Gottesman, and J. Preskill, *Accuracy Threshold for Postselected Quantum Computation*, Quantum Inf. Comput. **8,** 181 (2008).

[42] D. Aharonov and M. Ben-Or, *Fault-Tolerant Quantum Computation with Constant Error Rate*, SIAM J. Computing **39,** 1207 (2008).

[43] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, *Quantum Computational Advantage Using Photons*, Science **370,** 1460 (2020).

[44] 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 *et al.*, *Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking*, Phys. Rev. Lett. **109,** 080505 (2012).

[45] S. Krinner, S. Lazar, A. Remm, C. Andersen, N. Lacroix, G. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, *Benchmarking Coherent Errors in*

Controlled-Phase Gates Due to Spectator Qubits, Phys. Rev. Applied **14**, 024042 (2020).

[46] S. T. Flammia and J. J. Wallman, *Efficient Estimation of Pauli Channels*, ACM Trans. Quantum Comput. **1**, 1 (2020).

[47] M. G. Davis, E. Smith, A. Tudor, K. Sen, I. Siddiqi, and C. Iancu, *Heuristics for Quantum Compiling with a Continuous Gate Set*, arXiv:1912.02727.

[48] P. W. Shor, *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*, SIAM Rev. **41**, 303 (1999).

[49] A. W. Harrow and R. A. Low, *Random Quantum Circuits are Approximate 2-Designs*, Commun. Math. Phys. **291**, 257 (2009).

[50] B. Dirkse, J. Helsen, and S. Wehner, *Efficient Unitarity Randomized Benchmarking of Few-Qubit Clifford Gates*, Phys. Rev. A **99**, 012315 (2019).

[51] C. Chamberland, J. Wallman, S. Beale, and R. Laflamme, *Hard Decoding Algorithm for Optimizing Thresholds under General Markovian Noise*, Phys. Rev. A **95**, 042332 (2017).

[52] D. Greenbaum and Z. Dutton, *Modeling Coherent Errors in Quantum Error Correction*, Quantum Sci. Technol. **3**, 015007 (2017).

[53] The True-Q software was used to perform randomized compiling, generate cycle benchmarking sequences, analyze the cycle error reconstruction results, and simulate the circuits described in this work.

[54] S. J. Beale, A. Carignan-Dugas, D. Dahlen, J. Emerson, I. Hincks, A. Jain, D. Hufnagel, E. Ospadov, J. Saunders, J. J. Wallman, and A. Winick, *True-Q* https://doi.org/10.5281/zenodo.3945250 (2020).