Noise tailoring for scalable quantum computation via randomized compiling

Joel J. Wallman
Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo, Waterloo, Canada

Joseph Emerson
Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo, Waterloo, Canada and Canadian Institute for Advanced Research, Toronto, Ontario, Canada M5G 1Z8

(Received 27 June 2016; published 18 November 2016)

Quantum computers are poised to radically outperform their classical counterparts by manipulating coherent quantum systems. A realistic quantum computer will experience errors due to the environment and imperfect control. When these errors are even partially coherent, they present a major obstacle to performing robust computations. Here, we propose a method for introducing independent random single-qubit gates into the logical circuit in such a way that the effective logical circuit remains unchanged. We prove that this randomization tailors the noise into stochastic Pauli errors, which can dramatically reduce error rates while introducing little or no experimental overhead. Moreover, we prove that our technique is robust to the inevitable variation in errors over the randomizing gates and numerically illustrate the dramatic reductions in worst-case error that are achievable. Given such tailored noise, gates with significantly lower fidelity—comparable to fidelities realized in current experiments—are sufficient to achieve fault-tolerant quantum computation. Furthermore, the worst-case error rate of the tailored noise can be directly and efficiently measured through randomized benchmarking protocols, enabling a rigorous certification of the performance of a quantum computer.

DOI: 10.1103/PhysRevA.94.052325

I. INTRODUCTION

The rich complexity of quantum states and processes enables powerful protocols for processing and communicating quantum information, as illustrated by Shor's factoring algorithm [1] and quantum simulation algorithms [2]. However, the same rich complexity of quantum processes that makes them useful also allows a large variety of errors to occur. Errors in a quantum computer arise from a variety of sources, including decoherence and imperfect control, where the latter generally lead to coherent (unitary) errors. It is provably possible to perform a fault-tolerant quantum computation in the presence of such errors provided they occur with at most some maximum threshold probability [3–8]. However, the fault-tolerant threshold probability depends upon the error-correcting code and is notoriously difficult to estimate or bound because of the sheer variety of possible errors. Rigorous lower bounds on the threshold of the order of 10^{-6} [6] for generic local noise and 10^{-4} [9] and 10^{-3} [10] for stochastic Pauli noise have been obtained for a variety of codes. While these bounds are rigorous, they are far below numerical estimates that range from 10^{-2} [11,12] and 10^{-1} [13–15], which are generally obtained assuming stochastic Pauli noise, largely because the effect of other errors is too difficult to simulate [16]. While a threshold for Pauli errors implies a threshold exists for arbitrary errors (e.g., unitary errors), there is currently no known way to rigorously estimate a threshold for general noise from a threshold for Pauli noise.

The "error rate" due to an arbitrary noise map $\mathcal{E}$ can be quantified in a variety of ways. Two particularly important quantities are the average error rate defined via the gate fidelity

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

and the worst-case error rate (also known as the diamond distance from the identity) [17]

$$\epsilon(\mathcal{E}) = \tfrac{1}{2}\|\mathcal{E} - \mathcal{I}\|_\diamond = \sup_\psi \tfrac{1}{2}\|[\mathcal{E}\otimes\mathcal{I}_d - \mathcal{I}_{d^2}](\psi)\|_1, \tag{2}$$

where d is the dimension of the system $\mathcal{E}$ acts on, $\|A\|_1 = \sqrt{\mathrm{Tr}A^\dagger A}$, and the maximization is over all d^2-dimensional pure states (to account for the error introduced when acting on entangled states). The average error rate $r(\mathcal{E})$ is an experimentally convenient characterization of the error rate because it can be efficiently estimated via randomized benchmarking [18–22]. However, the diamond distance is typically the quantity used to prove rigorous fault-tolerance thresholds [6]. The average error rate and the worst-case error rate are related via the bounds [23,24]

$$r(\mathcal{E})d^{-1}(d+1) \leqslant \epsilon(\mathcal{E}) \leqslant \sqrt{r(\mathcal{E})}\sqrt{d(d+1)}. \tag{3}$$

The lower bound is saturated by any stochastic Pauli noise, in which case the worst-case error rate is effectively equivalent to the experimental estimates obtained efficiently via randomized benchmarking [25]. While the upper bound is not known to be tight, there exist unitary channels such that $\epsilon(\mathcal{E}) \approx \sqrt{(d+1)r(\mathcal{E})/4}$, so the scaling with r is optimal [26].

The scaling of the upper bound of Eq. (3) is only saturated by purely unitary noise. However, even a small coherent error relative to stochastic errors can result in a dramatic increase in the worst-case error. For example, consider a single qubit noise channel with $r = 1\times10^{-4}$, where the contribution due to stochastic noise processes is $r = 0.83\times10^{-4}$ and the remaining contribution is from a small unitary (coherent) rotation error. The worst-case error for such noise is $\epsilon \approx 10^{-2}$, essentially two orders of magnitude greater than the infidelity [26].

Here we show that by compiling random single-qubit gates into a logical circuit, noise with arbitrary coherence

and spatial correlations can be converted (or "tailored") into stochastic Pauli noise. We also prove that our technique is robust to gate-dependent errors which arise naturally due to imperfect gate calibration. In particular, our protocol is fully robust against arbitrary gate-dependent errors on the gates that are most difficult to implement, while imperfections in the easier gates introduces an additional error that is essentially proportional to the infidelity.

Our randomized compiling technique requires only a small (classical) overhead in the compilation cost, or, alternatively, can be implemented on the fly with fast classical control. Stochastic Pauli errors with the same average error rate r as a coherent error leads to four major advantages for quantum computation: (i) they have a substantially lower worst-case error rate; (ii) the worst-case error rate can be directly estimated efficiently and robustly via randomized benchmarking experiments, enabling a direct comparison to a threshold estimate to determine if fault-tolerant quantum computation is possible; (iii) the known fault-tolerant thresholds for Pauli errors are substantially higher than for coherent errors; and (iv) the average error rate accumulates linearly with the length of a computation for stochastic Pauli errors, whereas it can accumulate quadratically for coherent errors.

Randomizing quantum circuits has been previously proposed in Refs. [27,28]. However, these proposals have specific limitations that our technique circumvents. The proposal for inserting Pauli gates before and after Clifford gates proposed in Ref. [27] is a special case of our technique when the only gates to be implemented are Clifford gates. However, this technique does not account for non-Clifford gates whereas our generalized technique does. As a large number of non-Clifford gates are required to perform useful quantum computations [29] and are often more difficult to perform fault-tolerantly, our generalized technique should be extremely valuable in practice. Moreover, the proposal in Ref. [27] assumes that the Pauli gates are essentially perfect, whereas we prove that our technique is robust to imperfections in the Pauli gates. Alternatively, Pauli-random-error correction (PAREC) has been shown to eliminate static coherent errors [28]. However, PAREC involves changing the multiqubit gates in each step of the computation. As multiqubit errors are currently the dominant error source in most experimental platforms and typically depend strongly on the gate to be performed, it is unclear how robust PAREC will be against gate-dependent errors on multiqubit gates and consequently against realistic noise. By way of contrast, our technique is completely robust against arbitrary gate-dependent errors on multiqubit gates.

II. STANDARDIZED FORM FOR COMPILED QUANTUM CIRCUITS

We begin by proposing a standardized form for compiled quantum circuits based on an operational distinction between "easy" and "hard" gates, that is, gates that can be implemented in a given experimental platform with relatively small and large amounts of noise respectively. We also propose a specific choice of easy and hard gates that is well suited to many architectures for fault-tolerant quantum computation.

In order to experimentally implement a quantum algorithm, a quantum circuit is compiled into a sequence of elementary gates that can be directly implemented or have been specifically optimized. Typically, these elementary gates can be divided into easy and hard gate sets based either on how many physical qubits they act on or how they are implemented within a fault-tolerant architecture. In near-term applications of universal quantum computation without quantum error correction, such as quantum simulation, the physical error model and error rate associated with multiqubit gates will generally be distinct from, and much worse than, those associated with single qubit gates. In the long term, fault-tolerant quantum computers will implement some operations either transversally (that is, by applying independent operations to a set of physical qubits) or locally in order to prevent errors from cascading. However, recent "no-go" theorems establish that for any fault-tolerant scheme, there exist some operations that cannot be performed in such a manner [30,31] and so must be implemented via other means, such as magic-state injection [32] or gauge fixing [33].

The canonical division that we consider is to set the easy gates to be the group generated by Pauli gates and the phase gate $R = |0\rangle\langle 0| + i|1\rangle\langle 1|$, and the hard gate set to be the Hardamard gate H, the $\pi/8$ gate $\sqrt{R}$, and the two-qubit controlled-Z gate $\Delta(Z) = |0\rangle\langle 0| \otimes I + |0\rangle\langle 0| \otimes Z$. Such circuits are universal for quantum computation and naturally suit many fault-tolerant settings, including Calderbank-Shor-Steane (CSS) codes with a transversal T gate (such as the 15-qubit Reed-Muller code), color codes, and the surface code. While some of the hard gates may be easier than others in a given implementation, it is beneficial to make the set of easy gates as small as possible since our scheme is completely robust to arbitrary variations in errors over the hard gates.

With such a division of the gates, the target circuit can be reorganized into a circuit (the "bare" circuit) consisting of K clock cycles, wherein each cycle consists of a round of easy gates followed by a round of hard gates applied to disjoint qubits as in Fig. 1(a). To concisely represent the composite operations performed in individual rounds, we use the notational shorthand $\vec{A} = A_1 \otimes \cdots \otimes A_n$ and define G_k to be the product of all the hard gates applied in the kth cycle. We also set $G_K = I$ without loss of generality, so that the circuit begins and ends with a round of easy gates.

III. RANDOMIZED COMPILING

We now specify how standardized circuits in the above form can be randomized in order to average errors in the implementations of the elementary gates into an effective stochastic channel, that is, into a channel $\mathcal{E}$ that maps any n-qudit state ρ to

$$\mathcal{E}(\rho) = \sum_{P \in \mathbf{P}_d^{\otimes n}} c_P P \rho P^\dagger, \tag{4}$$

where $\mathbf{P}_d^{\otimes n}$ is the set of d^{2n} generalized Pauli operators and the coefficients c_P are a probability distribution over $\mathbf{P}_d^{\otimes n}$. For qubits ($d = 2$), $\mathbf{P}_2$ is the familiar set of four Hermitian and unitary Pauli operators $\{I, X, Y, Z\}$.

Let $\mathbf{C}$ denote the group generated by the easy gates and assume that it contains a subset $\mathbf{T}$ such that

$$\mathcal{E}^{\mathbf{T}} = \mathbb{E}_T T^\dagger \mathcal{E} T \tag{5}$$

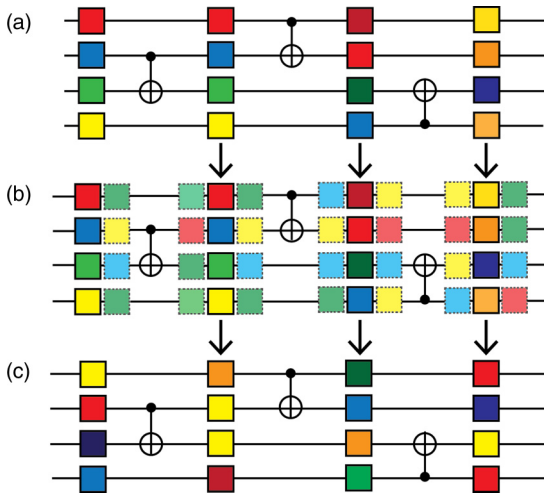

FIG. 1. (a) Example of a bare circuit that is arranged into cycles wherein each cycle consists of a round of easy single-qubit gates and a round of hard gates (here, the hard gates are controlled-NOT gates). (b) A randomized circuit wherein twirling gates have been inserted before and after every easy gate. (c) A randomized circuit wherein the twirling gates have been compiled into the easy gates, resulting in a new circuit that is logically equivalent to the bare circuit and has the same number of elementary gates.

is a stochastic channel for any channel $\mathcal{E}$, where $\mathbb{E}_x f(x) = |X|^{-1} \sum_{x\in X} f(x)$ denotes the uniform average over a set X (typically a gate set implicit from the context). The canonical example of such a set is $\mathbf{P}_d^{\otimes n}$ or any group containing $\mathbf{P}_d^{\otimes n}$.

We propose the following randomized compiling technique, where the randomization should ideally be performed independently for each run of a given bare circuit. Each round of easy gates $\vec{C}_k$ in the bare circuit of Fig. 1(a) is replaced with a round of randomized dressed gates

$$\tilde{C}_k = \vec{T}_k \vec{C}_k \vec{T}_{k-1}^c \tag{6}$$

as in Fig. 1(b), where the $T_{j,k}$ are chosen uniformly at random from the twirling set $\mathbf{T}$ and the correction operators are set to $\vec{T}_k^c = G_k \vec{T}_k^\dagger G_k^\dagger$ to undo the randomization from the previous round. The edge terms $\vec{T}_0^c$ and $\vec{T}_K$ can either be set to the identity or also randomized depending on the choice of the twirling set and the states and measurements.

The dressed gates should then be compiled into and implemented as a single round of elementary gates as in Fig. 1(c) rather than being implemented as three separate rounds of elementary gates. In order to allow the dressed gates to be compiled into a single easy gate, we require $\vec{T}_k^c \in \mathbf{C}^{\otimes n}$ for all $\vec{T}_k \in \mathbf{T}^{\otimes n}$. The example with $\mathbf{T} = \mathbf{P}_d$ that has been implicitly appealed to and described as "toggling the Pauli frame" previously [11] is a special case of the above technique when the hard gates are Clifford gates (which are defined to be the gates that map Pauli operators to Pauli operators under conjugation), but breaks down when the hard gates include non-Clifford gates such as the single-qubit $\pi/8$ gate. For the canonical division into easy and hard gates from the previous section, we set $\mathbf{T} = \mathbf{P}_2$, $\mathbf{C}$ to be the group generated by R and $\mathbf{P}_2$ (which is isomorphic to the dihedral group of order 8) and the hard gates to be rounds of H, $\sqrt{R}$, and $\Delta(Z)$ gates. Conjugating a Pauli gate by H or $\Delta(Z)$ maps it to another Pauli gate, while conjugating by $\sqrt{R}$ maps $X^x Z^z$ to $R^x X^x Z^z$ (up to a global phase). Therefore the correction gates, and hence the dressed gates, are all elements of the easy gate set.

The tailored noise is not realized in any individual choice of sequences. Rather, it is the average over independent random sequences. However, while each term $T^\dagger \mathcal{E} T$ in the tailored noise can have a different effect on an input state ρ, if the twirling gates are independently chosen on each run, then the expected noise over multiple runs is exactly the tailored noise. Independently sampling the twirling gates each time the circuit is run introduces some additional overhead, since the dressed gates (which are physically implemented) depend on the twirling gates and so need to be recompiled for each experimental run of a logical circuit. However, this recompilation can be performed in advance efficiently on a classical computer or else applied efficiently "on the fly" with fast classical control. Moreover, this fast classical control is exactly equivalent to the control required in quantum error correction so imposes no additional experimental burden.

We will prove below that our technique tailors noise satisfying various technical assumptions into stochastic Pauli noise. We expect the technique will also tailor more general noise into approximately stochastic noise, though leave a fully general proof as an open problem.

Robustness to arbitrary independent errors on the hard gates

We now prove that our randomized compiling scheme results in an average stochastic noise channel for Markovian noise that depends arbitrarily upon the hard gates but is independent of the easy gate. Under this assumption, the noisy implementations of the kth round of easy gates $\vec{C}_k$ and hard gates G_k can be written as $\mathcal{E}_e \vec{C}$ and $G_k \mathcal{E}(G_k)$ respectively, where $\mathcal{E}_e$ and $\mathcal{E}(G_k)$ are n-qubit channels that can include multiqubit correlations and $\mathcal{E}(*)$ can depend arbitrarily on the argument, that is, on which hard gates are implemented.

Theorem 1. Randomly sampling the twirling gates $\vec{T}_k$ independently in each round tailors the noise at each time step (except the last) into stochastic Pauli noise when the noise on the easy gates is gate independent.

Proof. The key observation is that if the noise in rounds of easy gates is some fixed noise channel $\mathcal{E}_e$, then the dressed gates in Eq. (6) have the same noise as the bare gates and so compiling in the extra twirling gates in Fig. 2(c) does not change the noise at each time step, as illustrated in Figs. 2(a) and 2(d). Furthermore, the correction gates $T_{k,j}^c$ are chosen so that they are the inverse of the randomizing gates when they are commuted through the hard gates, as illustrated in Figs. 2(b) and 2(c). Consequently, uniformly averaging over the twirling gates in every cycle reduces the noise in the kth cycle to the tailored noise

$$\mathcal{T}_k = \mathbb{E}_{\vec{T}} \vec{T}^\dagger \mathcal{E}(G_k) \mathcal{E}_e \vec{T}, \tag{7}$$

where for channels $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{AB}$ denotes the channel whose action on a matrix M is $\mathcal{A}[\mathcal{B}(M)]$. When $\mathbf{T} = \mathbf{P}$, the above channel is a Pauli channel [19]. Moreover, by the definition of a unitary one-design [21], the above sum is independent of the choice of $\mathbf{T}$ and so is a Pauli channel for any unitary one-design. ■

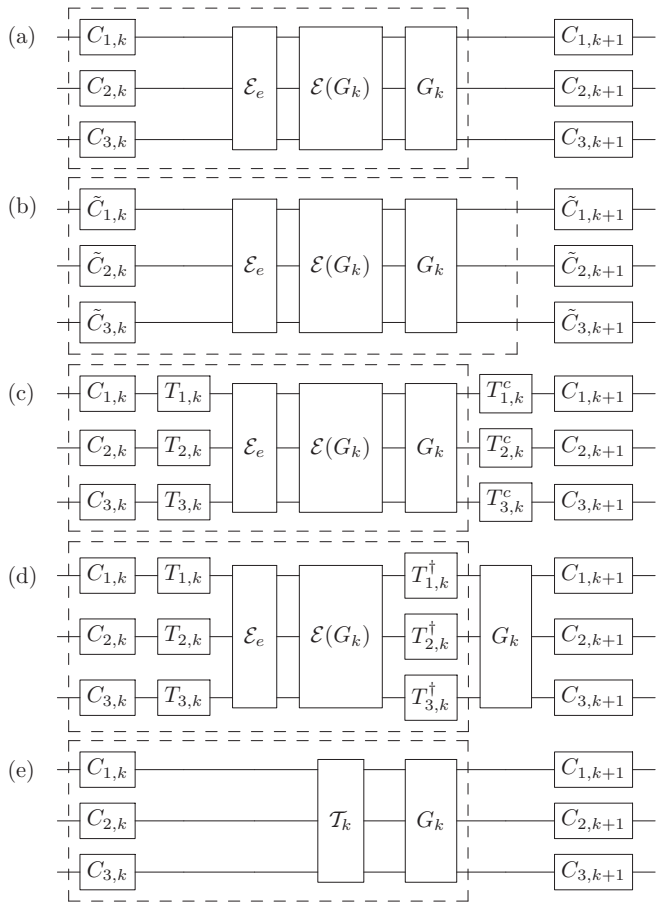

FIG. 2. (a) Fragment of a noisy bare circuit with the kth cycle indicated by the dashed box, where the gate-independent noise from subsequent cycles are omitted for brevity. (b) An equivalent fragment where the easy gates have been replaced by the dressed gates in Eq. (6), which leaves the noise unchanged. (c) The equivalent fragment where we have expanded out the dressed gates. (d) The equivalent fragment where we commute the correction gates through G_k, the round of hard gates. (e) The tailored circuit obtained by averaging over randomized circuits where $\mathcal{T}_k$ is the tailored stochastic Pauli channel in Eq. (7).

Theorem 1 establishes that the noise in all but the final cycle can be exactly tailored into stochastic noise (albeit under somewhat idealized conditions which will be relaxed below). To account for noise in the final round, we can write the effect corresponding to a measurement outcome $|\vec{z}\rangle$ as $\mathcal{A}(|\vec{z}\rangle\langle\vec{z}|)$ for some fixed noise map $\mathcal{A}$. If $\mathbf{P} \subset \mathbf{C}$, we can choose $\vec{T}_K$ uniformly at random from $\mathbf{P}^{\otimes n}$. A virtual Pauli gate can then be inserted between the noise map $\mathcal{A}$ and the idealized measurement effect $|\vec{z}\rangle\langle\vec{z}|$ by classically relabeling the measurement outcomes to map $\vec{z} \to \vec{z} \oplus \vec{x}$, where $\oplus$ denotes entrywise addition modulo 2. Averaging over $\vec{T}_K$ with this relabeling reduces the noise in the final round of single-qubit Clifford gates and the measurement to

$$\overline{\mathcal{A}} = \mathbb{E}_{\vec{P}} \vec{P} \mathcal{A} \mathcal{E}_e \vec{P}. \tag{8}$$

This technique can also be applied to quantum nondemolition measurements on a subset of the qubits (as in, for example, error-correcting circuits), where the unmeasured qubits have randomizing twirling gates applied before and after the measurement.

IV. ROBUSTNESS TO INDEPENDENT ERRORS ON THE EASY GATES

By Theorem 1, our technique is fully robust to the most important form of gate dependence, namely, gate-dependent errors on the hard gates. However, Theorem 1 still requires that the noise on the easy gates is effectively gate independent. Because residual control errors in the easy gates will generally produce small gate-dependent (coherent) errors, we will show that the benefits of noise tailoring can still be achieved in this physically realistic setting.

When the noise depends on the easy gates, the tailored noise in the kth cycle from equation (7) becomes

$$\mathcal{T}_k^{\mathrm{GD}} = \mathbb{E}_{\vec{T}_1,\ldots,\vec{T}_k} \vec{T}_k^\dagger \mathcal{E}(G_k) \mathcal{E}(\vec{\tilde{C}}_k) \vec{T}_k, \tag{9}$$

which depends on the previous twirling gates through $\vec{\tilde{C}}_k$ by Eq. (6). This dependence means that we cannot assign independent noise to each cycle in the tailored circuit.

However, in Theorem 2 we show that implementing a circuit with gate-dependent noise $\mathcal{E}(\vec{\tilde{C}}_k)$ instead of the corresponding gate-independent noise

$$\mathcal{E}_k^{\mathbf{T}} = \mathbb{E}_{\vec{\tilde{C}}_k} \mathcal{E}(\vec{\tilde{C}}_k) = \mathbb{E}_{\vec{T}_k,\vec{T}_{k-1}} \mathcal{E}\left(\vec{T}_k \vec{C}_k \vec{T}_{k-1}^c\right) \tag{10}$$

introduces a relatively small additional error. We show that the additional error is especially small when $\mathbf{T}$ is a group normalized by $\mathbf{C}$, that is, $CTC^\dagger \in \mathbf{T}$ for all $C \in \mathbf{C}$, $T \in \mathbf{T}$. This condition is satisfied in many practical cases, including the scenario where $\mathbf{T}$ is the Pauli group and $\mathbf{C}$ is the group generated by Pauli and R gates. The stronger bound reduces the contributions from every cycle by orders of magnitude in parameter regimes of interest (i.e., $\epsilon[\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}}] \leqslant 10^{-2}$, comparable to current experiments), so that the bound on the additional error grows very slowly with the circuit length.

Theorem 2. Let $\mathcal{C}_{\mathrm{GD}}$ and $\mathcal{C}_{\mathrm{GI}}$ be tailored circuits with gate-dependent and gate-independent noise on the easy gates respectively. Then

$$\|\mathcal{C}_{\mathrm{GD}} - \mathcal{C}_{\mathrm{GI}}\|_\diamond \leqslant \sum_{k=1}^{K} \mathbb{E}_{\vec{T}_1,\ldots,\vec{T}_K} \left\|\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\right\|_\diamond. \tag{11}$$

When $\mathbf{T}$ is a group normalized by $\mathbf{C}$, this can be improved to

$$\begin{aligned}\|\mathcal{C}_{\mathrm{GD}} - \mathcal{C}_{\mathrm{GI}}\|_\diamond \leqslant \sum_{k=2}^{K} 2\mathbb{E}_{\vec{\tilde{C}}_k} \left\|\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\right\|_\diamond \epsilon\left[\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}}\right] \\ + \mathbb{E}_{\vec{\tilde{C}}_1} \left\|\mathcal{E}(\vec{\tilde{C}}_1) - \mathcal{E}_1^{\mathbf{T}}\right\|_\diamond.\end{aligned} \tag{12}$$

Proof. Let

$$\begin{aligned}\mathcal{A}_k &= G_k \mathcal{E}(G_k) \mathcal{E}(\vec{\tilde{C}}_k) \vec{\tilde{C}}_k, \\ \mathcal{B}_k &= G_k \mathcal{E}(G_k) \mathcal{E}_k^{\mathbf{T}} \vec{\tilde{C}}_k,\end{aligned} \tag{13}$$

where $\mathcal{A}_k$ and $\mathcal{B}_k$ implicitly depend on the choice of twirling gates. Then the tailored circuits under gate-dependent and

gate-independent noise are

$$\begin{aligned}\mathcal{C}_{\mathrm{GD}} &= \mathbb{E}_a \mathcal{A}_{K:1},\\ \mathcal{C}_{\mathrm{GI}} &= \mathbb{E}_a \mathcal{B}_{K:1},\end{aligned} \tag{14}$$

respectively, where $\mathcal{X}_{a:b} = \mathcal{X}_a \ldots \mathcal{X}_b$ (note that this product is noncommutative) with $\mathcal{A}_{K:K+1} = \mathcal{B}_{0:1} = \mathcal{I}$ and the expectation is over all $\vec{T}_a$ for $a = 1, \ldots, K$. Then by a straightforward induction argument,

$$\mathcal{A}_{K:1} - \mathcal{B}_{K:1} = \sum_{k=1}^{K} \mathcal{A}_{K:k+1}(\mathcal{A}_k - \mathcal{B}_k)\mathcal{B}_{k-1:1} \tag{15}$$

for any fixed choice of the twirling gates. By the triangle inequality,

$$\begin{aligned}\|\mathcal{C}_{\mathrm{GD}} - \mathcal{C}_{\mathrm{GI}}\|_\diamond &= \left\| \mathbb{E}_a \sum_{k=1}^{K} \mathcal{A}_{K:k+1}(\mathcal{A}_k - \mathcal{B}_k)\mathcal{B}_{k-1:1} \right\|_\diamond\\ &\leqslant \mathbb{E}_a \sum_{k=1}^{K} \|\mathcal{A}_{K:k+1}(\mathcal{A}_k - \mathcal{B}_k)\mathcal{B}_{k-1:1}\|_\diamond\\ &\leqslant \mathbb{E}_a \sum_{k=1}^{K} \left\|\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\right\|_\diamond,\end{aligned} \tag{16}$$

where the second inequality follows from the submultiplicativity

$$\|\mathcal{A}\mathcal{B}\|_\diamond \leqslant \|\mathcal{A}\|_\diamond \|\mathcal{B}\|_\diamond \tag{17}$$

of the diamond norm and the normalization $\|\mathcal{A}\|_\diamond = 1$ which holds for all quantum channels $\mathcal{A}$.

We can substantially improve the above bound by evaluating some of the averages over twirling gates before applying the triangle inequality. In particular, leaving the average over $\vec{T}_{k-1}$ inside the diamond norm in Eq. (16) for every term except $k = 1$ gives

$$\begin{aligned}\|\mathcal{C}_{\mathrm{GD}} - \mathcal{C}_{\mathrm{GI}}\|_\diamond \leqslant \sum_{k=1}^{K} \mathbb{E}_{a\neq k-1} \|\mathbb{E}_{k-1}\delta_k\gamma_k\|_\diamond\\ + \mathbb{E}_a \left\|\mathcal{E}(\vec{\tilde{C}}_1) - \mathcal{E}_1^{\mathbf{T}}\right\|_\diamond,\end{aligned} \tag{18}$$

where

$$\begin{aligned}\delta_k &= \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}},\\ \gamma_k &= \vec{\tilde{C}}_k G_{k-1} \mathcal{E}(G_{k-1}) \mathcal{E}_{k-1}^{\mathbf{T}} \vec{T}_{k-1},\end{aligned} \tag{19}$$

and $\delta_k\gamma_k$ is the only factor of $\mathcal{A}_{K:k+1}(\mathcal{A}_k - \mathcal{B}_k)\mathcal{B}_{k-1:1}$ that depends on $\vec{T}_{k-1}$. Substituting $\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}} = [\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}} - \mathcal{I}] + \mathcal{I}$ in γ_k gives

$$\begin{aligned}\mathbb{E}_{k-1}\delta_k = \mathbb{E}_{k-1}\delta_k \vec{\tilde{C}}_k G_{k-1}\left[\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}} - \mathcal{I}\right]\vec{T}_{k-1}\\ + \mathbb{E}_{k-1}\delta_k \vec{T}_k \vec{C}_k G_{k-1},\end{aligned} \tag{20}$$

where the only factor in the second term that depends on $\vec{T}_{k-1}$ is δ_k. When $\mathbf{T}$ is a group normalized by $\mathbf{C}$,

$$\begin{aligned}\mathbb{E}_{k-1}\delta_k &= \mathbb{E}_{k-1}\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\\ &= \mathbb{E}_{k-1}\mathcal{E}\left(\vec{C}_k[\vec{C}_k^\dagger \vec{T}_k \vec{C}_k]\vec{T}_{k-1}^c\right) - \mathcal{E}_k^{\mathbf{T}}\\ &= \mathbb{E}_{\vec{T}'}\mathcal{E}(\vec{C}_k\vec{T}') - \mathcal{E}_k^{\mathbf{T}} = 0\end{aligned} \tag{21}$$

for any fixed value of $\vec{T}_k$, using the fact that $\{hg : g \in \mathbf{G}\} = \mathbf{G}$ for any group $\mathbf{G}$ and $h \in \mathbf{G}$ and that $\mathbb{E}_{\vec{T}'}\mathcal{E}(\vec{C}_k\vec{T}')$ is independent of $\vec{T}_k$. Therefore

$$\begin{aligned}\|\mathcal{C}_{\mathrm{GD}} - \mathcal{C}_{\mathrm{GI}}\|_\diamond &= \|\mathbb{E}_j(\mathcal{A}_{K:1} - \mathcal{B}_{K:1})\|_\diamond\\ &\leqslant \|\delta_1\|_\diamond + \sum_{k=2}^{K} \mathbb{E}_{j\neq k-1}\|\mathbb{E}_{k-1}\delta_k\gamma_k\|_\diamond\\ &\leqslant \|\delta_1\|_\diamond + \sum_{k=2}^{K} \mathbb{E}_{j\neq k-1}\|\mathbb{E}_{\vec{T}_{k-1}}\delta_k \vec{\tilde{C}}_k\\ &\quad \times G_{k-1}\left[\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}} - \mathcal{I}\right]\vec{T}_{k-1}\|_\diamond\\ &\leqslant \sum_{k=2}^{K} \mathbb{E}_j \|\delta_k\|_\diamond \left\|\mathcal{E}(G_{k-1})\mathcal{E}_{k-1}^{\mathbf{T}} - \mathcal{I}\right\|_\diamond + \|\delta_1\|_\diamond,\end{aligned} \tag{22}$$

where we have had to split the sum over k as $\vec{T}_0$ is fixed to the identity. ■

There are two particularly important scenarios in which the effect of gate-dependent contributions need to be considered and which determine the physically relevant value of K. In near-term applications such as quantum simulators, the following theorem would be applied to the entire circuit, while in long-term applications with quantum error correction, the following theorem would be applied to fragments corresponding to rounds of error correction. Hence under our randomized compiling technique, the noise on the easy gates imposes a limit either on the useful length of a circuit without error correction or on the distance between rounds of error correction. It is important to note that a practical limit on K is already imposed, in the absence of our technique, by the simple fact that even Pauli noise accumulates linearly in time, so $r(\mathcal{T}_k) \ll 1/K$ is already required to ensure that the output of any realistic circuit remains close to the ideal circuit.

While Theorem 2 provides a very promising bound, it is unclear how to estimate the quantities $\frac{1}{2}\mathbb{E}_{\vec{\tilde{C}}_k}\|\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\|_\diamond$ without performing full process tomography. To remedy this, we now provide the following bound in terms of the infidelity, which can be efficiently estimated via randomized benchmarking. We expect the following bound is not tight as we use the triangle inequality to turn the deviation from the average noise into deviations from no noise, which should be substantially larger. However, even the following loose bound is sufficient to rigorously guarantee that our technique significantly reduces the worst-case error, as illustrated in Fig. 3 for a two-qubit gate in the bulk of a circuit (i.e., with $k > 1$).

The following bound could also be substantially improved if the noise on the easy gates is known to be close to depolarizing (even if the hard gates have strongly coherent errors), as quantified by the unitarity [34–36]. However, rigorously determining an improved bound would require analyzing the protocol for estimating the unitarity under gate-dependent noise, which is currently an open problem.

Theorem 3. For arbitrary noise,

$$\mathbb{E}_{\vec{\tilde{C}}_k}\left\|\mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}}\right\|_\diamond \leqslant 2\epsilon\left(\mathcal{E}_k^{\mathbf{T}}\right) + 2\sqrt{\mathbb{E}_{\vec{\tilde{C}}_k}\epsilon[\mathcal{E}(\vec{\tilde{C}}_k)]^2}. \tag{23}$$

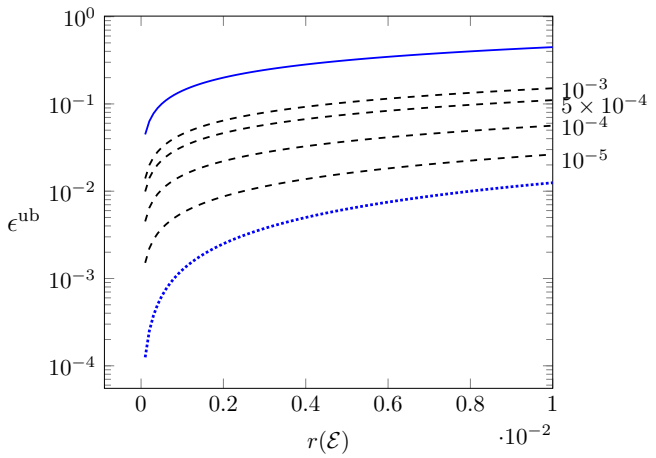

FIG. 3. Upper bounds ϵ^{ub} on the worst-case error for a two-qubit hard gate in the bulk of a circuit (e.g., a controlled-NOT gate with $k > 1$) as a function of its infidelity $r(\mathcal{E})$ with [dashed black, from theorems 2 and 3] and without [solid blue, from Eq. (3)] our tailoring technique under gate-dependent local noise on the single-qubit gates with infidelity $r(\overline{\mathcal{E}}^{\mathbf{T}}_{j,k}) = 10^{-5}, 10^{-4}, 5 \times 10^{-3}, 10^{-3}$ respectively. The worst-case error rate achieved by our technique for gate-independent noise (over the dressed gates) is plotted for comparison [dotted blue, from Eq. (3)].

For n-qubit circuits with local noise on the easy gates,

$$\mathbb{E}_{\vec{\tilde{C}}_k} \left\| \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}} \right\|_\diamond \leqslant \sum_{j=1}^{n} 4\sqrt{6r\left(\mathcal{E}_{j,k}^{\mathbf{T}}\right)} \tag{24}$$

for $k = 2, \ldots, K$, where $\mathcal{E}_{j.k}^{\mathbf{T}} = \mathbb{E}_{\tilde{C}_{j,k}} \mathcal{E}_j(\tilde{C}_{j,k})$ is the local noise on the jth qubit averaged over the dressed gates in the kth cycle.

Proof. First note that, by the triangle inequality,

$$\begin{aligned} \mathbb{E}_{\vec{\tilde{C}}_k} \left\| \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}} \right\|_\diamond &= \mathbb{E}_{\vec{\tilde{C}}_k} \left\| \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{I} + \mathcal{I} - \mathcal{E}_k^{\mathbf{T}} \right\|_\diamond \\ &\leqslant \mathbb{E}_{\vec{\tilde{C}}_k} \| \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{I} \|_\diamond + \left\| \mathcal{I} - \mathcal{E}_k^{\mathbf{T}} \right\|_\diamond \\ &\leqslant \mathbb{E}_{\vec{\tilde{C}}_k} 2\epsilon[\mathcal{E}(\vec{\tilde{C}}_k)] + 2\epsilon\left(\mathcal{E}_k^{\mathbf{T}}\right). \end{aligned} \tag{25}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} \left(\mathbb{E}_{\vec{\tilde{C}}_k} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)]\right)^2 &= \left(\sum_{\vec{\tilde{C}}_k} |\#\vec{\tilde{C}}_k|^{-1} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)] \right)^2 \\ &\leqslant \left(\sum_{\vec{\tilde{C}}_k} |\#\vec{\tilde{C}}_k|^{-2} \right) \left(\sum_{\vec{\tilde{C}}_k} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)]^2 \right) \\ &\leqslant |\#\vec{\tilde{C}}_k|^{-1} \left(\sum_{\vec{\tilde{C}}_k} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)]^2 \right) \\ &= \mathbb{E}_{\vec{\tilde{C}}_k} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)]^2, \end{aligned} \tag{26}$$

where $\#\vec{\tilde{C}}_k$ is the number of different dressed gates in the kth round.

For local noise, that is, noise of the form $\mathcal{E}_1 \otimes \ldots \mathcal{E}_n$ where $\mathcal{E}_j$ is the noise on the jth qubit,

$$\begin{aligned} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)] &= \tfrac{1}{2} \left\| \bigotimes_{j=1}^{n} \mathcal{E}_j(\tilde{C}_{j,k}) - \mathcal{I} \right\|_\diamond \\ &\leqslant \sum_{j=1}^{n} \tfrac{1}{2} \| \mathcal{E}_j(\tilde{C}_{j,k}) - \mathcal{I} \|_\diamond \\ &\leqslant \sum_{j=1}^{n} \epsilon[\mathcal{E}(\tilde{C}_{j,k})], \end{aligned} \tag{27}$$

where we have used the analog of Eq. (15) for the tensor product and

$$\|A \otimes B\|_\diamond \leqslant \|A\|_\diamond \|B\|_\diamond, \tag{28}$$

which holds for all A and B due to the submultiplicativity of the diamond norm, and the equality $\|A \otimes I\|_\diamond = \|A\|_\diamond$. Similarly,

$$\epsilon\left(\mathcal{E}_k^{\mathbf{T}}\right) = \sum_{j=1}^{n} \epsilon\left(\mathcal{E}_{j,k}^{\mathbf{T}}\right), \tag{29}$$

where $\mathcal{E}_{j,k}^{\mathbf{T}} = \mathbb{E}_{T_{j,k-1}, T_{j,k}} \mathcal{E}(\tilde{C}_{j,k})$. We then have

$$\begin{aligned} \mathbb{E}_{\vec{\tilde{C}}_k} \epsilon[\mathcal{E}(\vec{\tilde{C}}_k)] &\leqslant \sum_{j=1}^{n} \mathbb{E}_{\tilde{C}_{j,k}} \epsilon[\mathcal{E}(\tilde{C}_{j,k})] \\ &\leqslant \sum_{j=1}^{n} \sqrt{\mathbb{E}_{\tilde{C}_{j,k}} \epsilon[\mathcal{E}(\tilde{C}_{j,k})]^2}, \end{aligned} \tag{30}$$

where the second inequality is due to the Cauchy-Schwarz inequality as in Eq. (26). Returning to Eq. (25), we have

$$\begin{aligned} \mathbb{E}_{\vec{\tilde{C}}_k} \left\| \mathcal{E}(\vec{\tilde{C}}_k) - \mathcal{E}_k^{\mathbf{T}} \right\|_\diamond &\leqslant \sum_{j=1}^{n} 2\epsilon\left(\mathcal{E}_{j,k}^{\mathbf{T}}\right) + 2\sqrt{\mathbb{E}_{\tilde{C}_{j,k}} \epsilon[\mathcal{E}(\tilde{C}_{j,k})]^2} \\ &\leqslant \sum_{j=1}^{n} 2\sqrt{6r\left(\mathcal{E}_{j,k}^{\mathbf{T}}\right)} + 2\sqrt{\mathbb{E}_{\tilde{C}_{j,k}} 6r[\mathcal{E}(\tilde{C}_{j,k})]} \\ &\leqslant \sum_{j=1}^{n} 4\sqrt{6r\left(\mathcal{E}_{j,k}^{\mathbf{T}}\right)} \end{aligned} \tag{31}$$

for local noise, where the second inequality follows from Eq. (3) with $d = 2$ and the third from the linearity of the infidelity. ∎

V. NUMERICAL SIMULATIONS

Tailoring experimental noise into stochastic noise via our technique provides several dramatic advantages, which we now illustrate via numerical simulations. Our simulations are all of six-qubit circuits with the canonical division into easy and hard gates. That is, the easy gates are composed of Pauli gates and the phase gate $R = |0\rangle\langle 0| + i|1\rangle\langle 1|$, while the hard gates are the Hadamard, $\pi/8$ gate $T = \sqrt{R}$, and the two-qubit controlled-Z gate $\Delta(Z) = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes Z$. Such circuits are universal for quantum computation and naturally suit many fault-tolerant settings, including CSS codes with a

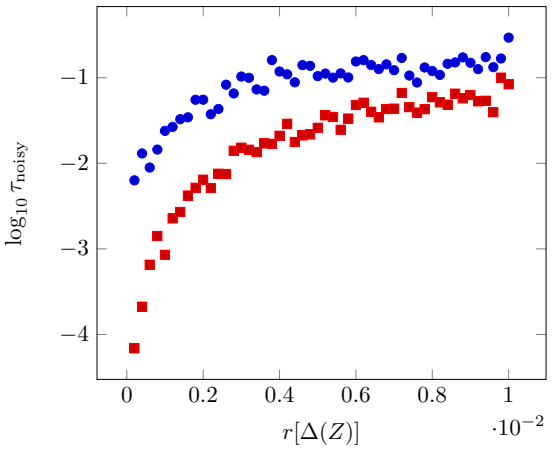

FIG. 4. Semilog plots of the error τ_{noise} from Eq. (32) with respect to computational basis measurement outcomes as a function of the average gate error $r[\Delta(Z)]$ of the noise on the $\Delta(Z)$ gate in six-qubit bare (blue circles) and tailored (red squares) circuits. Each data point corresponds to an independent random circuit with 100 cycles, where the (gate-dependent) noise on each gates is an over-rotation about the relevant eigenbasis with infidelity $r[\Delta(Z)]$ for the $\Delta(Z)$ gates and $r[\Delta(Z)]/10$ for all single-qubit gates. The data points for the tailored noise correspond to an average over 10^3 independent randomizations of the corresponding bare circuit via Eq. (6). The total error for the bare and tailored circuits differs by a factor of approximately 2 on a log scale, mirroring the separation between the worst-case errors for stochastic and unitary channels from Eq. (3) (although here the error is not maximized over preparations and measurements).

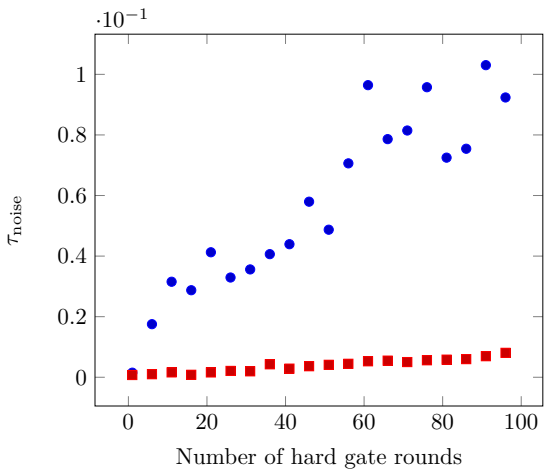

FIG. 5. Plots of the error τ_{noise} from Eq. (32) with respect to computational basis measurement outcomes as a function of the average gate error $r[\Delta(Z)]$ of the noise on the $\Delta(Z)$ gate in six-qubit bare (blue circles) and tailored (red squares) circuits. Each data point corresponds to an independent random circuit with K cycles, where the (gate-dependent) noise on each gates is an over-rotation about the relevant eigenbasis with infidelity 10^{-3} for the $\Delta(Z)$ gates and 10^{-5} for all single-qubit gates. The data points for the tailored noise correspond to an average over 10^3 independent randomizations of the corresponding bare circuit via Eq. (6). The error rate grows approximately linearly with the number of gate cycles, suggesting that the dominant reason for the error suppression is that the error at each location is suppressed (where there are a linear number of total locations), rather than the suppression of possible quadratic accumulation of coherent errors between locations.

transversal T gate (such as the 15-qubit Reed-Muller code), color codes, and the surface code.

We quantify the total noise in a noisy implementation $\mathcal{C}_{\text{noisy}}$ of an ideal circuit $\mathcal{C}_{\text{id}}$ by the variational distance

$$\tau_{\text{noisy}} = \sum_j \tfrac{1}{2} |\Pr(j|\mathcal{C}_{\text{noisy}}) - \Pr(j|\mathcal{C}_{\text{id}})| \tag{32}$$

between the probability distributions for ideal computational basis measurements after applying $\mathcal{C}_{\text{noisy}}$ and $\mathcal{C}_{\text{id}}$ to a system initialized in the $|0\rangle^{\otimes n}$ state. We do not maximize over states and measurements, rather, our results indicate the effect of noise under practical choices of preparations and measurements.

For our numerical simulations, we add gate-dependent over-rotations to each gate, that is, we perturb one of the eigenvectors of each gate U by $e^{i\delta_U}$. For single-qubit gates, the choice of eigenvector is irrelevant (up to a global phase), while for the two-qubit $\Delta(Z)$ gate, we add the phase to the $|11\rangle$ state.

We perform two sets of numerical simulations to illustrate two particular properties. First, Fig. 4 shows that our technique introduces a larger relative improvement as the infidelity decreases, that is, approximately a factor of 2 on a log scale, directly analogous to the $r/\sqrt{r}$ scaling for the worst case error (although recall that our simulations are for computational basis states and measurements and do not maximize the error over preparations and measurements). For these simulations, we set δ_U so that the $\Delta(Z)$ gate has an infidelity of $r[\Delta(Z)]$ and so that all single-qubit gates have an infidelity of $r[\Delta(Z)]/10$ (regardless of whether they are included in the easy or the hard set). For the bare circuits (blue circles), each data point is the variational distance of a randomly chosen six-qubit circuit with a hundred alternating rounds of easy and hard gates, each sampled uniformly from the sets of all possible easy and hard gate rounds respectively. For the tailored circuits (red squares), each data point is the variational distance between $\Pr(j|\mathrm{C}_{\text{id}})$ and the probability distribution $\Pr(j|\mathrm{C}_{\text{noisy}})$ averaged over 10^3 randomizations of the bare circuit obtained by replacing the easy gates with (compiled) dressed gates as in Eq. (6).

Second, Fig. 5 shows that the typical error for both the bare and tailored circuits grows approximately linearly with the length of the circuit. This suggests that, for typical circuits, the primary reason that the total error is reduced by our technique is not because it prevents the worst-case quadratic accumulation of fidelity with the circuit length (although it does achieve this). Rather, the total error is reduced because the contribution from each error location is reduced, where the number of error locations grows linearly with the circuit size. For these simulations, we set δ_U so that the $\Delta(Z)$ gate has an infidelity of 10^{-3} and the easy gates have infidelities of 10^{-5}. For the bare circuits (blue circles), each data point is the variational distance of a randomly chosen six-qubit circuit as above with K alternating rounds of easy and hard gates, where K varies from 5 to 100. The tailored circuits (red squares) again give the variational distance between the ideal distribution and the

probability distribution averaged over 10^3 randomizations of the bare circuit.

VI. DISCUSSION

We have shown that arbitrary Markovian noise processes can be reduced to effective Pauli processes by compiling different sets of uniformly random gates into sequential operations. This randomized compiling technique can reduce the worst-case error rate by orders of magnitude and enables threshold estimates for general noise models to be obtained directly from threshold estimates for Pauli noise. Physical implementations can then be evaluated by directly comparing these threshold estimates to the average error rate r estimated via efficient experimental techniques, such as randomized benchmarking, to determine whether the experimental implementation has reached the fault-tolerant regime. More specifically, the average error rate r is that of the tailored channel for the composite noise on a round of easy and hard gates and this can be directly estimated using interleaved randomized benchmarking with the relevant choice of group [37,38].

Our technique can be applied directly to gate sets that are universal for quantum computation, including all elements in a large class of fault-tolerant proposals. Moreover, our technique only requires *local* gates to tailor general noise on multiqubit gates into Pauli noise. Our numerical simulations in Figs. 4 and 5 demonstrate that our technique can reduce worst-case errors by orders of magnitude. Furthermore, our scheme should generally produce an even greater effect as fault-tolerant protocols are scaled up, since fault-tolerant protocols are designed to suppress errors, for example, $\epsilon \to \epsilon^k$ for some scale factor k (e.g., number of levels of concatenation), so any reduction at the physical level is improved exponentially with k. However, while the technique can be directly integrated into fault-tolerant implementations, analyzing the impact on encoded error rates remains an open problem and will generally depend on the specific noise and error-correcting code.

A particularly significant open problem is the robustness of our technique to noise that remains non-Markovian on a timescale longer than a typical gate time. Non-Markovian noise can be mitigated by techniques such as randomized dynamic decoupling [39,40], which correspond to applying random sequences of Pauli operators to echo out non-Markovian contributions. Due to the random gates compiled in at each time step, we expect that our technique may also suppress non-Markovian noise in a similar manner.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with A. Carignan-Dugas, D. Cory, S. Flammia, D. Gottesman, T. Jochym-O'Connor, and R. Laflamme. This research was supported by the U.S. Army Research Office through Grant No. W911NF-14-1-0103, CIFAR, the Government of Ontario, and the Government of Canada through NSERC and Industry Canada.

[1] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM Rev. **41**, 303 (1999).

[2] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[3] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[4] D. Gottesman, Class of quantum error-correcting codes saturating the quantum Hamming bound, Phys. Rev. A **54**, 1862 (1996).

[5] E. Knill, Resilient quantum computation, Science **279**, 342 (1998).

[6] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (1999).

[7] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, Quantum error correction via codes over GF(4), IEEE Trans. Inf. Theory **44**, 1369 (1998).

[8] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (NY) **303**, 2 (2003).

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

[10] P. Aliferis, D. Gottesman, and J. Preskill, Accuracy threshold for postselected quantum computation, Quant. Inf. Comput. **8**, 181 (2008).

[11] E. Knill, Quantum computing with realistically noisy devices, Nature (London) **434**, 39 (2005).

[12] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Surface code quantum computing with error rates over 1%, Phys. Rev. A **83**, 020302 (2011).

[13] G. Duclos-Cianci and D. Poulin, Fast Decoders for Topological Quantum Codes, Phys. Rev. Lett. **104**, 050504 (2010).

[14] J. R. Wootton and D. Loss, High Threshold Error Correction for the Surface Code, Phys. Rev. Lett. **109**, 160503 (2012).

[15] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Strong Resilience of Topological Codes to Depolarization, Phys. Rev. X **2**, 021004 (2012).

[16] D. Puzzuoli, C. Granade, H. Haas, B. Criger, E. Magesan, and D. G. Cory, Tractable simulation of error correction with honest approximations to realistic fault models, Phys. Rev. A **89**, 022306 (2014).

[17] A. Kitaev, Quantum computations: algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).

[18] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

[19] J. Emerson, M. Silva, O. Moussa, C. A. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Symmetrized characterization of noisy quantum processes, Science **317**, 1893 (2007).

[20] 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).

[21] 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).

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

[23] S. Beigi and R. König, Simplified instantaneous non-local quantum computation with applications to position-based cryptography, New J. Phys. **13**, 093036 (2011).

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

[25] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[26] 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).

[27] E. Knill, Fault-tolerant postselected quantum computation: Threshold analysis, arXiv:quant-ph/0404104.

[28] O. Kern, G. Alber, and D. L. Shepelyansky, Quantum error correction of coherent errors by randomization, Eur. Phys. J. D **32**, 153 (2005).

[29] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

[30] B. Eastin and E. Knill, Restrictions on Transversal Encoded Quantum Gate Sets, Phys. Rev. Lett. **102**, 110502 (2009).

[31] M. E. Beverland, O. Buerschaper, R. Koenig, F. Pastawski, J. Preskill, and S. Sijher, Protected gates for topological quantum field theories, J. Math. Phys. **57**, 022201 (2016).

[32] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[33] H. Bombín, Gauge color codes: optimal transversal gates and gauge fixing in topological stabilizer codes, New J. Phys. **17**, 083002 (2015).

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

[35] 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).

[36] J. J. Wallman, Bounding experimental quantum error rates relative to fault-tolerant thresholds, arXiv:1511.00727.

[37] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[38] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Characterizing universal gate sets via dihedral benchmarking, Phys. Rev. A **92**, 060302(R) (2015).

[39] L. Viola and E. Knill, Random Decoupling Schemes for Quantum Dynamical Control and Error Suppression, Phys. Rev. Lett. **94**, 060502 (2005).

[40] L. F. Santos and L. Viola, Enhanced Convergence and Robust Performance of Randomized Dynamical Decoupling, Phys. Rev. Lett. **97**, 150501 (2006).