Error Mitigation for Short-Depth Quantum Circuits

Kristan Temme, Sergey Bravyi, and Jay M. Gambetta
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA
(Received 21 July 2017; published 3 November 2017)

Two schemes are presented that mitigate the effect of errors and decoherence in short-depth quantum circuits. The size of the circuits for which these techniques can be applied is limited by the rate at which the errors in the computation are introduced. Near-term applications of early quantum devices, such as quantum simulations, rely on accurate estimates of expectation values to become relevant. Decoherence and gate errors lead to wrong estimates of the expectation values of observables used to evaluate the noisy circuit. The two schemes we discuss are deliberately simple and do not require additional qubit resources, so to be as practically relevant in current experiments as possible. The first method, extrapolation to the zero noise limit, subsequently cancels powers of the noise perturbations by an application of Richardson's deferred approach to the limit. The second method cancels errors by resampling randomized circuits according to a quasiprobability distribution.

DOI: 10.1103/PhysRevLett.119.180509

Introduction.—From the time quantum computation generated wide spread interest, the strongest objection to its viability was the sensitivity to errors and noise. In an early paper, Unruh [1] found that the coupling to the environment sets an ultimate time and size limit for any quantum computation. This initially curbed the hopes that the full advantage of quantum computing could be harnessed, since it set limits on the scalability of any algorithm. This problem was, at least in theory, remedied with the advent of quantum error correction [2–4]. It was proven that if both the decoherence and the imprecision of gates could be reduced below a finite threshold value, then quantum computation could be performed indefinitely [5,6]. Although it is the ultimate goal to reach this threshold in an experiment that is scalable to larger sizes, the overhead that is needed to implement a fully fault-tolerant gate set with current codes [7] seems prohibitively large [8,9]. In turn, it is expected that in the near term the progress in quantum experiments will lead to devices with dynamics, which are beyond what can be simulated with a conventional computer. This leads to the question: what computational tasks could be accomplished with only limited, or no error correction?.

The suggestions of near-term applications in such quantum devices mostly center around quantum simulations with short-depth circuit [10–12] and approximate optimization algorithms [13]. Furthermore, certain problems in material simulation may be tackled by hybrid quantum-classical algorithms [14]. In most such applications, the task can be abstracted to applying a short-depth quantum circuit to some simple initial state and then estimating the expectation value of some observable after the circuit has been applied. This estimation must be accurate enough to achieve a simulation precision comparable to or exceeding that of classical algorithms. Yet, although the quantum system evolves coherently for the most part of the short-depth circuit, the effects of decoherence already become apparent as an error in the estimate of the observable. For the simulation to be of value, the effect of this error needs to be mitigated.

In this Letter we introduce two techniques for quantum error mitigation that increase the quality of any such short-depth quantum simulations. We find that the accuracy of the expectation value can be increased significantly in the presence of noise. We are looking for error mitigation techniques that are as simple as possible and do not require additional quantum resources. Both techniques require that some noise parameter taken together with system size and circuit depth can be considered a small number. The first scheme does not make any assumption about the noise model other than it being weak and constant in time. In comparison, the second scheme can tolerate stronger noise; however, it requires detailed knowledge of the noise model.

Extrapolation to the zero noise limit.—It is our goal to estimate the expectation value of some quantum observable A with respect to an evolved state $\rho_\lambda(T)$ after time T that is subject to noise characterized by the parameter λ in the limit where $\lambda \to 0$. To achieve this, we apply Richardson's deferred approach to the limit to cancel increasingly higher orders of λ [15].

Although gates are typically used to describe quantum circuits, for our analysis it is more convenient to consider the time-dependent Hamiltonian dynamics implementing the circuit. The time-dependent multi-qubit Hamiltonian is denoted by $K(t)$. It can be expanded into N-qubit Pauli operators $P_\alpha \in \langle \mathbb{1}, X_j, Y_j, Z_j \rangle_{j=1\ldots N}$, where X_j, Y_j, Z_j acts as a single-qubit Pauli matrix on site j and trivially elsewhere. We allow for time-dependent coupling coefficients $J_\alpha(t) \in \mathbb{R}$. The circuit is encoded as $K(t) = \sum_\alpha J_\alpha(t) P_\alpha$. The total evolution of the open system with

initial state $\rho(0)$ will be described by an equation of the following form:

$$\frac{\partial}{\partial t}\rho(t) = -i[K(t), \rho(t)] + \lambda\mathcal{L}[\rho(t)] \tag{1}$$

for time $t \in [0, T]$. We do not specify the exact form of the generator $\mathcal{L}(\rho)$ but only require that it is invariant under time rescaling and independent from the parameters $J_\alpha(t)$ in $K(t)$. The noise term $\mathcal{L}(\rho)$ could be given as a Lindblad operator, or it could correspond to a Hamiltonian that couples to a bath to model non-Markovian dynamics. We ask that there is a parameter $\lambda \ll 1$ that indicates a weak action of the noise and that we can bound $\|\mathcal{L}_{I,t_1} \circ \mathcal{L}_{I,t_2} \circ \ldots \circ \mathcal{L}_{I,t_n}(\rho)\|_1 \leq l_n$, where at most $l_n = \mathcal{O}(N^n)$. The map $\mathcal{L}_{I,t}$ is shorthand notation for the transformation of $\mathcal{L}$ into the interaction frame generated by $K(t)$.

The expectation value of the observable A is obtained from the final state $\rho_\lambda(T)$ as $E_K(\lambda) = \mathrm{tr}[A\rho_\lambda(T)]$. The function $E_K(\lambda)$ can be expressed as a series in λ where the contribution with λ^0 corresponds to the noise-free evolution. This can be seen by transforming the evolution into the interaction frame with respect to $K(t)$ and expanding the Born series; cf. the Supplemental Material, Sec. I [16]. Starting from the noise-free expectation value $E^* = \mathrm{tr}[A\rho_0(T)]$, the expansion is given by

$$E_K(\lambda) = E^* + \sum_{k=1}^{n} a_k \lambda^k + R_{n+1}(\lambda, \mathcal{L}, T). \tag{2}$$

The a_k are model-specific constants typically growing like $a_k \sim N^k T^k$. Here $R_{n+1}(\lambda, \mathcal{L}, T)$ is the remainder of the expansion and can be bounded by $|R_{n+1}(\lambda, \mathcal{L}, T)| \leq \|A\| l_{n+1} (\lambda T)^{n+1}/(n+1)!$ by standard arguments. Since we assumed an extensive scaling of l_n, such an expansion is only meaningful whenever $NT\lambda$ is small. We are of course interested in $\lim_{\lambda \to 0} E_K(\lambda) = E^*$; however, we are faced with a small but finite parameter λ. Since we only have access to $E_K(\lambda)$, our estimate of E^* will be off by $\mathcal{O}(\lambda)$.

This estimate can be improved by Richardson's deferred approach to the limit [15,23]. To explain the idea, let us assume we can run the quantum circuit at different noise rates λ_j, with $j = 0, \ldots, n$ and obtain experimental estimates $\hat{E}_K(\lambda_j) = E_K(\lambda_j) + \delta_j$. Here the $\lambda_j = c_j\lambda$ are appropriately rescaled values of the experimental noise rate λ. The estimate deviates from the actual expectation value due to experimental inaccuracies and finite sampling errors by an error δ_j. The estimate of E^* can be significantly improved by considering the approximation $\hat{E}^n_K(\lambda)$, which is written as the linear combination

$$\hat{E}^n_K(\lambda) = \sum_{j=0}^{n} \gamma_j \hat{E}_K(c_j\lambda). \tag{3}$$

Here we require the coefficients γ_j to satisfy the linear system of equations [23].

$$\sum_{l=0}^{n} \gamma_j = 1 \quad \text{and} \quad \sum_{j=0}^{n} \gamma_j c_j^k = 0 \quad \text{for } k = 1 \ldots n. \tag{4}$$

The linear combination Eq. (3) will be an approximation to E^* up to an error of order $\mathcal{O}(\lambda^{n+1})$.

To obtain estimates at different noise rates λ_j, we use a rescaling trick. We run the same circuit $n+1$ times with rescaled parameters in $K(t)$. We follow the protocol: 1. For $j = 0, \ldots, n$ (a) choose a rescaling coefficient $c_j > 1$ ($c_0 = 1$) and evolve $\rho(0)$ with rescaled Hamiltonian $K^j(t) = \sum_\alpha J^j_\alpha(t) P_\alpha$, where

$$J^j_\alpha(t) = c_j^{-1} J_\alpha(c_j^{-1} t), \tag{5}$$

for time $T_j = c_j T$. (b) Estimate observable A to obtain $\hat{E}_K(c_j\lambda)$. 2. Solve Eq. (4) and compute $\hat{E}^n_K(\lambda)$ as in Eq. (3).

A rescaling of the equations shows that the state $\rho^j_\lambda(T_j)$, which evolves under $\dot{\rho}^j_\lambda = -i[K^j(t), \rho^j] + \lambda\mathcal{L}(\rho^j)$ for time T_j, satisfies $\rho^j_\lambda(T_j) = \rho_{c_j\lambda}(T)$; cf. Supplemental Material, Sec. II [16]. Hence the estimates $\hat{E}_K(c_j\lambda) = \mathrm{tr}[A\rho^j_\lambda(T_j)] + \delta_j$ can be obtained from the $n+1$ runs rescaled according to the protocol.

If the protocol is performed for $n+1$ steps, the error between the exact expectation value E^* and the estimator $\hat{E}^n_K(\lambda)$ can be bounded by

$$|E^* - \hat{E}^n_K(\lambda)| \leq \Gamma_n \left(\delta^* + \|A\| \frac{l_{n+1} (\lambda T)^{n+1}}{(n+1)!} \right). \tag{6}$$

Here $\Gamma_n = \sum_{j=0}^{n} |\gamma_j| c_j^{n+1}$ and $\delta^* = \max_j |\delta_j|$ is the largest experimental error.

This follows from repeated application of the triangle inequality; cf. Supplemental Material, Sec. III [16]. The Eq. (4) can be solved, and one finds that the coefficients $\gamma_j = \prod_{m \neq j} c_m (c_j - c_m)^{-1}$, so that the constant Γ_n can be evaluated. In the literature [23], several choices for progression of c_j are common. The two most frequent series are exponential decrease (Bulirsch-Stoer) and harmonic decay. In our experiments we are actually increasing the noise rate starting from the optimal value, whereas it is common in the numerical literature to improve the small parameter. The result is, of course, the same.

Examples.—To demonstrate this method we will consider three numerical examples. In all the examples the time evolution is given by a Hamiltonian $K(t)$ that encodes a control problem. For a single *drift step* we evolve with a Hamiltonian $K_R(t) = U_N(\theta) K_0 U_N^\dagger(\theta)$, where the single qubit product unitary $U_N(\theta) \in SU(2)^{\otimes N}$ is chosen Haar-random, and the drift Hamiltonian $K_0 = \sum_{i,j} J_{i,j} X_i Z_j$ is

chosen with respect to a random graph and Gaussian distributed couplings $J_{i,j}$. The evolution is subject to three different noise models: First, in Fig. 1(a), we evolve in the presence of depolarizing noise described by the sum of single qubit generators $\mathcal{L}_i = -\lambda(2^{-1}\mathrm{tr}_i(\rho) - \rho)$ acting on all N qubits. Second, in Fig. 1(b), we consider dephasing and amplitude damping noise on every qubit, where we have chosen a ratio of $\lambda_1/\lambda_2 = 1.5$ with a generator $\mathcal{L}_i = \lambda_1(\sigma_i^- \rho \sigma_i^+ - \frac{1}{2}\{\sigma_i^+\sigma_i^-, \rho\}) + \lambda_2(Z_i\rho Z_i - \rho)$ and $\sigma_i^{\pm} = 2^{-1}(X_i \pm iY_i)$. Third, in Fig. 1(c), we consider a highly non-Markovian setting, where each of the N qubits i is coupled to its own single-qubit bath b_i via the Hamiltonian $V_i = 1/2X_i \otimes X_{b_i} + 1/2Z_{b_i}$ and the bath is prepared in the initial state $\rho_B = [2\cosh(\beta/2)]^{-N}\exp(-\beta\sum_{b_i}\sigma_{b_i}^z)$. Then, after the evolution of each noisy circuit $T = td$ we measure a randomly chosen multiqubit Pauli operator P_α.

The graphs in Fig. 1 show that with modest effort very high precisions can be obtained. In the low noise range $\epsilon \sim 10^{-3}$ the relative error can be reduced to $\Delta E \sim 10^{-6}$–10^{-11}. The precision is then essentially determined by the sampling error δ^*, which we have neglected in the plots.

Probabilistic error cancellation.—Here we discuss a noise reduction scheme for quantum circuits subject to Markovian noise. First let us state our assumptions on the noise model. A noisy N-qubit device will be described by a basis set of noisy operations $\Omega = \{\mathcal{O}_1, \ldots, \mathcal{O}_m\}$ that can be implemented on this device. Each operation $\mathcal{O}_\alpha$ is a trace-preserving completely positive (TPCP) map on N qubits that acts nontrivially only on a small subset of qubits, say at most two. For example, $\mathcal{O}_\alpha$ could be a noisy unitary gate applied to a specified pair of qubits or a noisy qubit initialization. We assume that noise in the system can be

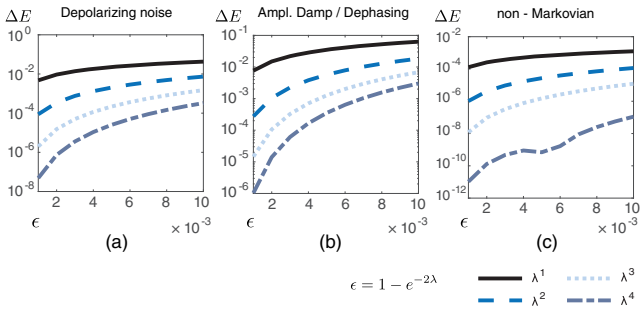

FIG. 1. The plots show a random Hamiltonian evolution for $N = 4$ system qubits and $d = 6$ drift steps, each for time $t = 2$. For all systems plot the error $\Delta E = |E^* - \hat{E}_K^n(\lambda)|$ for $n = 0, 1, 2, 3$. Here λ^1, $n = 0$ corresponds to the uncorrected error. The noise parameter $\lambda = -1/2\log(1-\epsilon)$ is chosen so that all plots have the same perturbation measured in the depolarizing strength $\epsilon = 10^{-3}\ldots10^{-2}$. The plot shows the mitigation of (a) depolarizing noise, (b) amplitude damping/dephasing noise, and (c) non-Markovian noise, for $\{c_j\}$ chosen as random partition of in the interval [1,4].

fully characterized such that the map $\mathcal{O}_\alpha$ is known for each α. A circuit of length L in the basis Ω is a sequence of L operations from Ω. Let Ω_L be a set of all length-L circuits in the basis Ω. A circuit $\alpha = (\alpha_1, \ldots, \alpha_L)$ implements a map $\mathcal{O}_{\boldsymbol{\alpha}} = \mathcal{O}_{\alpha_L}\ldots\mathcal{O}_{\alpha_2}\mathcal{O}_{\alpha_1}$. The expectation value of an observable A on the final state produced by a noisy circuit α is

$$E(\boldsymbol{\alpha}) = \mathrm{tr}[A\mathcal{O}_{\boldsymbol{\alpha}}(|0\rangle\langle 0|^{\otimes n})].$$

For simplicity, we ignore errors in the initial state preparation and in the final measurement. Such errors can be accounted for by adding dummy noisy operations before each measurement and after each qubit initialization. Furthermore, we shall assume that A is diagonal in the Z basis and $\|A\| \leq 1$.

Below we show that under certain conditions the task of simulating an ideal quantum circuit can be reduced to estimating the expectation value $E(\boldsymbol{\alpha})$ for a suitable random ensemble of noisy quantum circuits $\boldsymbol{\alpha}$. Moreover, the ideal and the noisy circuits act on the same number of qubits and have the same depth.

Let $\Gamma = \{\mathcal{U}_1, \ldots, \mathcal{U}_k\}$ be a fixed basis set of ideal gates. Each gate $\mathcal{U}_\beta(\rho) = U_\beta \rho U_\beta^\dagger$ is described by a unitary TPCP map on N qubits that acts nontrivially on a small subset of qubits. An ideal length-L circuit in the basis Γ is a sequence of L gates from Γ. A circuit $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ implements a map $\mathcal{U}_{\boldsymbol{\beta}} = \mathcal{U}_{\beta_L}\ldots\mathcal{U}_{\beta_2}\mathcal{U}_{\beta_1}$. Define an ideal expectation value

$$E^*(\boldsymbol{\beta}) = \mathrm{tr}[A\mathcal{U}_{\boldsymbol{\beta}}(|0\rangle\langle 0|^{\otimes n})].$$

We consider a simulation task where the goal is to estimate $E^*(\boldsymbol{\beta})$ with a specified precision δ.

The key idea of our scheme is to represent the ideal circuit as a quasiprobabilistic mixture of noisy ones. Let us say that a noisy basis Ω simulates an ideal circuit $\boldsymbol{\beta}$ with the overhead $\gamma_{\boldsymbol{\beta}} \geq 1$ if there exists a probability distribution $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ on the set of noisy circuits $\boldsymbol{\alpha} \in \Omega_L$ such that

$$\mathcal{U}_{\boldsymbol{\beta}} = \gamma_{\boldsymbol{\beta}} \sum_{\boldsymbol{\alpha}\in\Omega_L} P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha})\mathcal{O}_{\boldsymbol{\alpha}} \tag{7}$$

for some coefficients $\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \pm 1$. We also require that the distribution $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is sufficiently simple so that one can efficiently sample $\boldsymbol{\alpha}$ from $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$. The coefficients $\gamma_{\boldsymbol{\beta}}$, $\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ must be efficiently computable. We shall refer to Eq. (7) as a *quasiprobability representation* (QPR) of the ideal circuit $\boldsymbol{\beta}$. Note that $\gamma_{\boldsymbol{\beta}} \geq 1$ because $\mathcal{U}_{\boldsymbol{\beta}}$ and $\mathcal{O}_{\boldsymbol{\alpha}}$ are trace preserving. Quasiprobability distributions have been previously used to construct classical algorithms for simulation of quantum circuits [20,21]. Our work can be viewed as an application of these methods to the problem of simulating ideal quantum circuits by noisy ones.

Substituting Eq. (7) into the definition of $E^*(\boldsymbol{\beta})$ gives

$$E^*(\boldsymbol{\beta}) = \gamma_{\boldsymbol{\beta}} \sum_{\boldsymbol{\alpha}\in\Omega_L} P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha})E(\boldsymbol{\alpha}). \tag{8}$$

Let $\boldsymbol{\alpha} \in \Omega_L$ be a random variable drawn from $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and $x \in \{0,1\}^n$ be the final readout of the noisy circuit α obtained by measuring each qubit of the final state $\mathcal{O}_{\boldsymbol{\alpha}}(|0\rangle\langle 0|^{\otimes n})$ in the Z basis. Note that $\langle x|A|x\rangle$ is an unbiased estimator of $E(\boldsymbol{\alpha})$ with the variance $O(1)$. Thus from Eq. (8) one infers that $\gamma_{\boldsymbol{\beta}}\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha})\langle x|A|x\rangle$ is an unbiased estimator of the ideal expectation value $E^*(\boldsymbol{\beta})$ with the variance $O(\gamma_{\boldsymbol{\beta}}^2)$. We can now estimate $E^*(\boldsymbol{\beta})$ with any desired precision δ by the Monte Carlo method. Define

$$M = (\delta^{-1}\gamma_{\boldsymbol{\beta}})^2 \tag{9}$$

and generate M samples $\alpha^1, \ldots, \boldsymbol{\alpha}^M \in \Omega_L$ drawn from $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$. By Hoeffding's inequality, $E^*(\boldsymbol{\beta})$ is approximated within error $O(\delta)$ w.h.p. by a random variable

$$\hat{E}(\boldsymbol{\beta}) = \frac{\gamma_{\boldsymbol{\beta}}}{M}\sum_{a=1}^{M} \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}^a)\langle x^a|A|x^a\rangle, \tag{10}$$

where $x^a \in \{0,1\}^n$ is the final string of the noisy circuit α^a. Computing the estimator $\hat{E}(\boldsymbol{\beta})$ requires M runs of the noisy circuits, with each run producing a single readout string x^a. Estimating $E^*(\boldsymbol{\beta})$ with a precision δ in the absence of noise by the Monte Carlo method would require approximately δ^{-2} runs. Thus the quantity $\gamma_{\boldsymbol{\beta}}^2$ determines the simulation overhead [see Eq. (9)].

A systematic method of constructing QPRs with a small overhead is given in the Supplemental Material, Sec. IV [16]. Here we illustrate the method using toy noise models usually studied in the quantum fault-tolerance theory: the depolarizing noise and the amplitude damping noise. For concreteness, we choose the ideal gate set Γ as the standard Clifford+T basis.

Let $\mathcal{D}_k$ be the depolarizing noise on $k = 1, 2$ qubits that returns the maximally mixed state with probability ϵ and does nothing with probability $1-\epsilon$. Define a noisy version of a k-qubit unitary gate $\mathcal{U}$ as $\mathcal{D}_k\mathcal{U}$. The noisy basis Ω is obtained by multiplying ideal gates on the left by arbitrary Pauli operators and adding the depolarizing noise. Thus Ω is a set of operations $\mathcal{O}_\alpha = \mathcal{D}_k\mathcal{P}\mathcal{U}$, where $\mathcal{U} \in \Gamma$ is a k-qubit ideal gate and $\mathcal{P} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes k}$ is a Pauli TPCP map. The random ensemble of noisy circuits $\mathcal{O}_{\boldsymbol{\alpha}}$ that simulates an ideal circuit $\mathcal{U}_{\boldsymbol{\beta}}$ is constructed in three steps: (1) Start from the ideal circuit, $\mathcal{O}_{\boldsymbol{\alpha}} = \mathcal{U}_{\boldsymbol{\beta}}$. (2) Modify $\mathcal{O}_{\boldsymbol{\alpha}}$ by adding a Pauli X, Y, Z after each single-qubit gate with probability $p_1 = \epsilon/(4+2\epsilon)$. The gate is unchanged with probability $1-3p_1$. (3) Modify $\mathcal{O}_{\boldsymbol{\alpha}}$ by adding a Pauli $IX, IY, \ldots, ZZ$ after each CNOT with probability $p_2 = \epsilon/(16+14\epsilon)$. The CNOT is unchanged with probability $1-15p_2$. The resulting circuit is then implemented on a noisy device (which adds the depolarizing noise after each gate) and the final readout string x is recorded. By generating M samples of x one can estimate $E^*(\boldsymbol{\beta})$ from Eq. (10). The sign function $\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is equal to $(-1)^r$, where r is the number of Pauli operators added to the ideal circuit $\mathcal{U}_{\boldsymbol{\beta}}$. As shown in in the Supplemental Material, Sec. IV [16], the above defines a QPR of the ideal circuit $\mathcal{U}_{\boldsymbol{\beta}}$ with the overhead $\gamma_{\boldsymbol{\beta}} \approx 1 + \epsilon(3L_1/2 + 15L_2/8)$, where L_1 is the number of single-qubit gates and L_2 is the number of CNOTS in the ideal circuit. The method has been tested numerically for random noisy Clifford + T circuits; see Fig. 2.

A more interesting example is the noise described by the amplitude-damping channel $\mathcal{A}$ that resets every qubit to its ground state with probability ϵ. A noisy version of a k-qubit unitary gate $\mathcal{U}$ is defined as $\mathcal{A}^{\otimes k}\mathcal{U}$. In contrast to the previous example, noisy unitary gates $\mathcal{A}^{\otimes k}\mathcal{U}$ alone cannot simulate any ideal unitary gate since $\mathcal{A}$ is not a unital map. To overcome this, we extend the noisy basis Ω by adding noisy versions of single-qubit state preparations $\mathcal{A}\mathcal{P}_{|\psi\rangle}$, where $\mathcal{P}_{|\psi\rangle}$ maps any input state to $|\psi\rangle\langle\psi|$. Our scheme requires state preparations for single qubit states $|\psi\rangle = |+\rangle, |-\rangle, |0\rangle, |1\rangle$ that can be performed at any time step (not only at the beginning). In the Supplemental Material Sec. V [16] we show how to construct a QPR of the ideal Clifford + T circuit $\mathcal{U}_{\boldsymbol{\beta}}$ with the overhead $\gamma_{\boldsymbol{\beta}} \approx 1 + \epsilon(2L_1 + 4L_2)$. The examples considered above suggest that well-characterized noisy circuits can simulate ideal ones with overhead $\gamma \approx (1+c\epsilon)^L$, where ϵ is the typical error rate and c is a small constant. The value of c can be determined by performing quantum process tomography

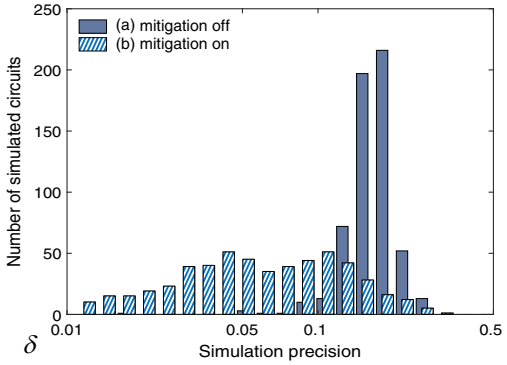

FIG. 2. Simulation precision $\delta(\beta) = |\hat{E}(\boldsymbol{\beta}) - E^*(\boldsymbol{\beta})|$ for 500 randomly generated ideal Clifford+T circuits on $N = 6$ qubits with depth $d = 20$. The gates are subject to single- and two-qubit depolarizing noise $\epsilon = 10^{-2}$. The figure shows results for simulations without (a) and with (b) error cancellation. In both cases each ideal circuit was simulated by $M = 4000$ runs of the noisy circuit. For each circuit $\mathcal{U}_{\boldsymbol{\beta}}$ we defined the observable A as a projector Π_{out} onto the subset of 2^{N-1} basis vectors with the largest weight in the final state. The results are consistent with $\gamma_{\boldsymbol{\beta}} \approx 4.3$ so that $\gamma_{\boldsymbol{\beta}} M^{-1/2} \approx 0.07$.

[22] and finding the QPR for each ideal gate. Using Eq. (9), one can estimate the number of noisy circuit runs of length L as $M \sim \exp(2c\epsilon L)$. Assuming error rates in the range $\epsilon \sim 10^{-3}$, it may be possible to simulate ideal circuits with $O(10^3)$ gates.

Conclusions.—Both error mitigation schemes require no additional quantum hardware such as ancilla or code qubits and work directly with the physical qubits. The zero-noise extrapolation requires sufficient control of the time evolution to implement the rescaled dynamics and hinges on the assumption of a large time-scale separation between the dominant noise and the controlled dynamics. For the probabilistic error cancellation a full characterization of the noisy computational operations is necessary. To obtain this to a precision of $\sim 10^{-3}$ is challenging in practice. However, if one is willing to sacrifice optimality, a Pauli- or Clifford-twirling [24,25] can be applied that converts any noise channel into a simple mixture of Pauli errors or depolarizing noise, making the characterization task much more manageable. A very recent independent paper by Li and Benjamin [26] discusses similar issues to those addressed here.

We thank Antonio Mezzacapo for insightful discussions, and we acknowledge support from the IBM Research Frontiers Institute.

[1] W. G. Unruh, Phys. Rev. A **51**, 992 (1995).
[2] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).
[3] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).
[4] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).
[5] D. Aharonov and M. Ben-Or, in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing* (ACM, 1997), pp. 176–188.
[6] A. Y. Kitaev, Russ. Math. Surv. **52**, 1191 (1997).
[7] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Phys. Rev. A **86**, 032324 (2012).
[8] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Phys. Rev. X **2**, 031007 (2012).
[9] S. J. Devitt, A. M. Stephens, W. J. Munro, and K. Nemoto, Nat. Commun. **4** (2013).
[10] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nat. Commun. **5** (2014).
[11] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New J. Phys. **18**, 023023 (2016).
[12] D. Wecker, M. B. Hastings, and M. Troyer, Phys. Rev. A **92**, 042303 (2015).
[13] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1411.4028.
[14] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Phys. Rev. X **6**, 031045 (2016).
[15] L. F. Richardson and J. A. Gaunt, Phil. Trans. R. Soc. A **226**, 299 (1927).
[16] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.119.180509, which includes Refs. [17–22], for detailed calculation.
[17] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).
[18] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press on Demand, 2002).
[19] C. Schneider, Numer. Math. **24**, 177 (1975).
[20] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Phys. Rev. Lett. **115**, 070501 (2015).
[21] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf, Phys. Rev. X **5**, 021003 (2015).
[22] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
[23] A. Sidi, *Practical extrapolation methods: Theory and applications*, Cambridge Monographs on Applied and Computational Mathematics (2003), Vol. 10.
[24] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Phys. Rev. A **78**, 012347 (2008).
[25] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[26] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).