Error mitigation for short depth quantum circuits

Kristan Temme, Sergey Bravyi and Jay M. Gambetta
IBM T.J. Watson Research Center, Yorktown Heights NY 10598
(Dated: December 8, 2016)

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 don't require additional qubit resources 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 cancel errors by resampling randomized circuits according to a quasi-probability distribution.

INTRODUCTION

From the time quantum computation generated wide spread interest, the strongest objection to its viability was the sensitivity to errors and noise. In a paper that was released briefly after Peter Shor's seminal factoring algorithm [1], William Unruh [2] found that the coupling to the environment sets an ultimate time limit and size limit for any quantum computation. This limit 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 [3–5]. 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 [6, 7]. 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 [8] seems prohibitively large [9, 10]. 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 lead to the question what computational tasks could be accomplished with only limited, or no error correction. To be clear, the production of entropy in any such circuit will set a limit on the size of any computation that can be performed [11] and it is observed that circuits with only logarithmic depth can be implemented [12]. However, the full computational power of log-depth circuits is not yet fully understood, and it can be argued based on complexity theoretic grounds that even finite depth circuits lie beyond the computational power of a classical computer [13, 14].

The suggestions of near-term applications in such quantum devices mostly center around the natural idea of quantum simulations with short depth circuits [15–17] and approximate optimization algorithms [18]. Furthermore, certain problems in material simulation may be tackled by hybrid quantum-classical algorithms [19]. Such algorithms approximate a many-body quantum system by an effective model with fewer degrees of freedom which can be solved on a relatively small quantum computer [19–21].

In most such applications, the task can be abstracted to applying a short depth quantum circuits 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 or exceeding the one 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 paper we introduce two techniques for *quantum error mitigation* that increases the quality of any such short depth quantum simulations. We find that the quality of the estimate can be increased significantly in the presence of noise with only modest overhead. We are looking for error mitigation techniques that are as simple as possible and don't require additional quantum resources to be applicable in current experiments. We present two different schemes that are applicable in different settings; both schemes however 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. 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 [22]. We consider different noise models, such as the evolutions with Lindblad equation, but also treat the case where the system is coupled to the bath via some weak coupling Hamiltonian. We require that the noise operator time-independent and therefore invariant under time rescaling. It is important to point out that here we

do not need to know the exact form of the noise model. We only require suitable control over the coherent evolution we apply.

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 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

$$K(t) = \sum_\alpha J_\alpha(t) P_\alpha. \tag{1}$$

The total evolution of the open system with initial state ρ_0 will be described by an equation of the following form.

$$\frac{\partial}{\partial t}\rho(t) = -i[K(t), \rho(t)] + \lambda \mathcal{L}(\rho(t)). \tag{2}$$

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 $K(t)$. We ask that there is a parameter $\lambda \ll 1$ that indicates a weak action of the noise and that we can bound $\|\mathcal{L}_{I,t_1} \circ \mathcal{L}_{I,t_2} \circ \ldots \circ \mathcal{L}_{I,t_n}(\rho_0)\|_1 \leq l_n$, where at most $l_n = \mathcal{O}(N^n)$. The map $\mathcal{L}_{I,t}$ is short-hand notation for the transformed map $\mathcal{L}$ into the interaction frame generated by $K(t)$. We consider various forms for the generator $\mathcal{L}$. It could for instance be given in the form of Lindblad operator $\mathcal{L}(\rho) = \sum_\alpha L_\alpha \rho L_\alpha^\dagger - \frac{1}{2}\{L_\alpha^\dagger L_\alpha, \rho\}$ that can be used to model depolarizing noise or dephasing / amplitude damping noise. Alternatively it could correspond to noise of the form $\mathcal{L}(\rho) = -i[V, \rho]$, where V is some Hamiltonian. This of course includes the most general case of open system evolution when $\rho_0 = \rho_S(0) \otimes \rho_B(0)$ and $V = \sum_\alpha S_\alpha \otimes B_\alpha + H_B$ describes the dynamics of out system being coupled to a bath. In this case a small λ indicates a separation of time scales.

The expectation value of the observable A is obtained from the final state $\rho_\lambda(T)$ as

$$E_K(\lambda) = \mathrm{tr}\left(A\rho_\lambda(T)\right). \tag{3}$$

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 from an argument transforming the evolution in the interaction frame w.r.t $K(t)$ and expansion in terms of the Born series (c.f. supp. mat. sec I). Starting from the noise-free expectation value $E^* = \mathrm{tr}\left(H\rho_0(T)\right)$, the expansion in terms of the noise parameter is given by

$$E_K(\lambda) = E^* + \sum_{k=1}^{n} a_k \lambda^k + R_{n+1}(\lambda, \mathcal{L}, T), \tag{4}$$

and the a_k are model-specific constants typically extensive in system size 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. Note that 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 of the noisy evolution. 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 [22, 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 $E_K^n(\lambda)$, which is written as the linear combination

$$\hat{E}_K^n(\lambda) = \sum_{j=0}^{n} \gamma_j \hat{E}_K(c_j\lambda). \tag{5}$$

Here we require the coefficients γ_j to satisfy the linear system of equations [24].

$$\sum_{l=0}^{n} \gamma_j = 1 \quad \text{and} \quad \sum_{j=0}^{n} \gamma_j\, c_j^k = 0 \quad \text{for } k = 1 \ldots n. \tag{6}$$

For different types of error terms $\mathcal{L}$ it may happen that certain powers of λ vanish in the expansion. For some coherent error terms in combination with the right bath states odd contributions in λ can vanish. If this occurs, the Richardson extrapolation method is particularly efficient since a higher order of precision can be obtained at each step. As we will see shortly, the linear combination (5) will be an approximation to E^* up to an error of order $\mathcal{O}(\lambda^{n+1})$ in the noise parameter.

To obtain estimates at different noise rates λ_j, we make use of a simple rescaling trick. We run the same circuit $n+1$ times with rescaled parameters. We follow the **protocol a**:

1. For $j = 0, \ldots, n$
   (a) choose a rescaling coefficient $c_j > 1$ ($c_0 = 1$) and evolve ρ_0 with rescaled Hamiltonian $K^j(t) = \sum_\alpha J_\alpha^j(t) P_\alpha$, where
   $$J_\alpha^j(t) = c_j^{-1} J_\alpha\left(c_j^{-1} t\right), \tag{7}$$
   for time $T_j = c_j T$.
   (b) Estimate observable A to obtain $\hat{E}_K(c_j\lambda)$.
2. Solve equations (6) and compute $\hat{E}_K^n(\lambda)$ as in eqn. (5).

A simple rescaling argument shows that the state $\rho_\lambda^j(T_j)$, which evolves under $\dot{\rho}_\lambda^j = -i[K^j(t), \rho^j] + \lambda\mathcal{L}(\rho^j)$ for time

T_j satisfies $\rho_\lambda^j(T_j) = \rho_{c_j\lambda}(T)$ (c.f. supp. mat. sec II). Hence the estimates $\hat{E}_K(c_j\lambda) = \mathrm{tr}\left(A\rho_\lambda^j(T_j)\right) + \delta_j$ can be obtained from the $n+1$ runs rescaled according to the protocol. Note, that for different experimental circumstances other protocols may be easier to implement. So for example in an optical experiment that is plagued by photon loss, it may be suitable to consider different methods of changing the photon loss rate.

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

$$|E^* - \hat{E}_K^n(\lambda)| \leq \Gamma_n \left(\delta^* + \|A\| \, \frac{l_{n+1}(\lambda T)^{n+1}}{(n+1)!}\right). \quad (8)$$

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 and is presented for the reader's convenience in (supp. mat. sec III). The equations (6) can be solved, and one finds that the coefficients $\gamma_j = \prod_{m\neq j} c_m(c_j - c_m)^{-1}$ are directly related to Lagrange polynomials evaluated at zero so that the constant Γ_n can be evaluated. In the literature [23], several choices for c_j are common. In the Bulirsch - Stoer series the rescalings are chosen so that $c_j = h^j c_0$ constitutes an exponential series, which is typically chosen at base $h = 2$; but harmonic series have also been frequently applied, e.g. for parameters $q > 1, \eta \geq 0$ one can choose $c_j = (j+\eta)^q c_0$. Note that in our experiments we are actually increasing the noise rate starting from the optimal value, wheres it is common in the numerical literature to improve the small parameter. The result here is of course the same; we have just switched the order for notational convenience. For both aforementioned cases, a bound on Γ_n has been derived. Furthermore, observe that the bound in (8) improves exponentially at every order, but can only continue up to a point where $T_j\lambda N < 1$, whenever $l_n \sim N^n$. This sets a bound as to how far the expectation value can be improved.

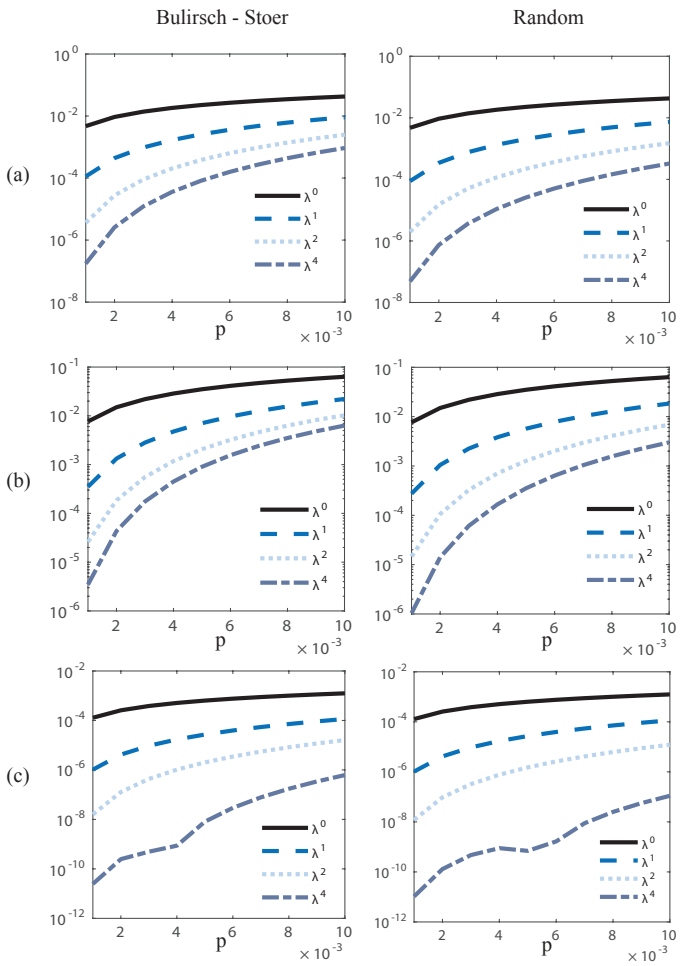

FIG. 1. (color online) The plots show a random Hamiltonian evolution, according to eqn. (9) for $N = 4$ system qubits and $d = 6$ drift steps, each for time $t = 2$. For all systems plot the error $\Delta E = |E^* - \hat{E}_K^n(\lambda)|$ for $n = 0, 1, 2, 3$. Here $\lambda^{n=0}$ always corresponds to the uncorrected error. The noise parameter $\lambda = -1/2\log(1-p)$ is chosen so that all plots have the same perturbation and we only choose the scale the depolarizing strength $p = 10^{-3}\ldots 10^{-2}$ so the different noise models become comparable. The left column shows a choice of $c_j \in \{2^0, 2^1, 2^2, 2^3\}$ for (a) Depolarizing noise (b) Amplitude damping / dephasing noise and (c) the non-Markovian noise model explained in the text. The right column shows the same noise models, where c_j have been chosen at random in the interval $[1, 4]$ including the endpoints.

Examples

To demonstrate this method we will consider three simple numerical examples. In all the examples the time evolution is given by a time-dependent Hamiltonian $K(t)$, that encodes a control problem. For a single *drift step* we evolve with a Hamiltonian $K(t) = U_N(\theta)K_0U_N^\dagger(\theta)$, where the single qubit product Unitary $U_N(\theta) \in SU(2)^{\otimes N}$ is chosen Haar-random and the base Hamiltonian

$$K_0 = \sum_{i,j} J_{i,j} X_i Z_j \quad (9)$$

is chosen with respect to a random graph and Gaussian distributed couplings $J_{i,j}$. We perform d such drift steps to simulate a typical bang-bang control evolution, each for a total dimensionless duration of a time $t = 2$. This evolution is then subjected to three different kinds of noise: first 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 of the N qubits. Second 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 $\mathcal{L}_i = \lambda_1\left(\sigma_i^-\rho\sigma_i^+ - \frac{1}{2}\{\sigma_i^+\sigma_i^-, \rho\}\right) + \lambda_2\left(Z_i\rho Z_i - \rho\right)$ and $\sigma^\pm = 2^{-1}(X \pm iY)$. Third Fig 1 (c), we consider a highly non-Markovian setting, where of the N qubits i is coupled to its own single qubit bath b_i via the Hamiltonian

$V_i = 1/2\ X_i \otimes X_{b_i} + 1/2\ Z_{b_i}$ and the bath is prepared in the initial state $\rho_B = (2\cosh(\beta/2))^{-N} \exp(-\beta \sum_{b_i} \sigma^z_{b_i})$. Then after the full evolution $T = td$ we measure a randomly chosen multi-qubit Pauli operator P_α.

The graphs in Fig 1 show that with modest effort, we consider Richardson extrapolation up to $n = 3$, so the error term enters only up to fourth order, very high precisions can be obtained. The errors in the low noise range, where $p \sim 10^{-3}$ indicate that the relative error in the expectation value can be reduced below $\Delta E \sim 10^{-6} - 10^{-11}$ for these small circuits. The real bound on the precision of how well the observable can be estimated 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 a 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 $\Omega = \{\mathcal{O}_1, \dots, \mathcal{O}_m\}$ 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 non-trivially only on a small subset of qubits, say at most two. This subset is called the support of $\mathcal{O}_\alpha$. 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 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 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_L)$ implements a map $\mathcal{O}_{\boldsymbol{\alpha}} = \mathcal{O}_{\alpha_L} \cdots \mathcal{O}_{\alpha_2} \mathcal{O}_{\alpha_1}$. We shall be interested in the probability of measuring Π_{out} of the circuit $\boldsymbol{\alpha}$ defined as

$$p(\boldsymbol{\alpha}) = \mathrm{Tr}[\Pi_{out}\, \mathcal{O}_{\boldsymbol{\alpha}}(\rho_{in})],$$

where $\rho_{in} = |0\rangle\langle 0|^{\otimes n}$ is the initial state and Π_{out} is some projector diagonal in the Z-basis. For simplicity, here we ignore errors in preparation of the initial state ρ_{in} and in the final measurement. Such errors can be accounted for by adding dummy noisy operations before each measurement and after each qubit initialization.

Below we show that under certain conditions the task of simulating an ideal quantum circuit can be reduced to estimating the output probability $p(\boldsymbol{\alpha})$ 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 length.

Let $\Gamma = \{\mathcal{U}_1, \dots, \mathcal{U}_k\}$ be a fixed basis set of ideal gates. Each gate $\mathcal{U}_\beta$ is described by a unitary TPCP map on n qubits that acts non-trivially on a small subset of qubits, say at most two. An ideal length-L circuit in the basis Γ is a sequence of L gates from Γ. A circuit $\boldsymbol{\beta} = (\beta_1, \dots, \beta_L)$ implements a map $\mathcal{U}_{\boldsymbol{\beta}} = \mathcal{U}_{\beta_L} \cdots \mathcal{U}_{\beta_2} \mathcal{U}_{\beta_1}$. The probability $p_i(\boldsymbol{\beta})$ of measuring Π_{out} in the ideal circuit $\boldsymbol{\beta}$ is defined as

$$p_i(\boldsymbol{\beta}) = \mathrm{Tr}[\Pi_{out}\, \mathcal{U}_{\boldsymbol{\beta}}(\rho_{in})].$$

Here ρ_{in} and Π_{out} are the same as above. We consider a simulation task where the goal is to estimate $p_i(\boldsymbol{\beta})$ with a specified precision δ, say $\delta = 1/10$. It is well known that this problem is BQP-complete [25]. In particular, we do not expect it to be solvable in polynomial time on a classical computer.

The key idea of our scheme is to represent the ideal circuit as a quasi probabilistic 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

$$\mathcal{U}_{\boldsymbol{\beta}} = \gamma_{\boldsymbol{\beta}} \sum_{\boldsymbol{\alpha} \in \Omega_L} P_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \mathcal{O}_{\boldsymbol{\alpha}} \tag{10}$$

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. (10) as a *quasi-probability 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. We note that quasi-probability distributions have been previously used to construct classical algorithms for simulation of quantum circuits [26, 27]. Our work can be viewed as an application of these methods to the problem of simulating ideal quantum circuits by noisy ones.

Substituting Eq. (10) into the definition of $p_i(\boldsymbol{\beta})$ gives

$$p_i(\boldsymbol{\beta}) = \gamma_{\boldsymbol{\beta}} \sum_{\boldsymbol{\alpha} \in \Omega_L} P_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}). \tag{11}$$

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 string of the noisy circuit $\boldsymbol{\alpha}$ obtained by measuring each qubit of the final state $\mathcal{O}_{\boldsymbol{\alpha}}(\rho_{in})$ in the Z-basis. Note that $\langle x|\Pi_{out}|x\rangle$ is an unbiased estimator of $p(\boldsymbol{\alpha})$ with the variance $O(1)$. Thus from Eq. (11) one infers that the quantity $\gamma_{\boldsymbol{\beta}} \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \langle x|\Pi_{out}|x\rangle$ is an unbiased estimator of the ideal output probability $p_i(\boldsymbol{\beta})$ with the variance $O(\gamma^2_{\boldsymbol{\beta}})$. We can now estimate $p_i(\boldsymbol{\beta})$ with any desired precision δ by the Monte Carlo method. Define

$$M = (\delta^{-1} \gamma_{\boldsymbol{\beta}})^2 \tag{12}$$

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

$$\xi = \frac{\gamma_{\boldsymbol{\beta}}}{M} \sum_{a=1}^{M} \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}^a) \langle x^a|\Pi_{out}|x^a\rangle, \tag{13}$$

where $x^a \in \{0,1\}^n$ is the final readout string of the noisy circuit $\boldsymbol{\alpha}^a$. Computing ξ requires M runs of the noisy circuits with each run producing a single readout string x^a. Note that estimating $p_i(\boldsymbol{\beta})$ with a precision δ in the absence of noise by Monte Carlo method would require approximately δ^{-2} runs of the circuit. Thus the quantity $\gamma^2_{\boldsymbol{\beta}}$ determines the simulation overhead, see Eq. (12). As we discuss in supp. mat. sec IV, the estimate Eq. (13) can be improved by dividing the total budget

of M runs unevenly among different circuits $\boldsymbol{\alpha}^a$ depending on the value of $p(\boldsymbol{\alpha}^a)$ such that circuits $\boldsymbol{\alpha}^a$ with larger (smaller) $p(\boldsymbol{\alpha}^a)$ are given more (fewer) runs.

We shall see that for certain noise models, such as the amplitude damping noise, QPRs can be constructed only if the noisy basis Ω includes some qubit initialization maps. In particular, noisy circuits $\boldsymbol{\alpha}$ that appear in Eq. (10) may apply qubit initializations at intermediate time steps, even though the ideal circuit $\boldsymbol{\beta}$ initializes all the qubits at the very first step. All QPRs constructed below preserve the circuit depth. That is, if the ideal circuit $\boldsymbol{\beta}$ has depth d then all noisy circuits $\boldsymbol{\alpha}$ in Eq. (10) have depth at most d.

Next let us discuss how to construct QPRs with a small overhead. For concreteness, we choose the ideal gate set Γ as the Clifford+T basis. It includes the identity gate I, the Hadamard gate H, phase shift gates $S = \mathrm{diag}[1, i]$ and $T = \mathrm{diag}[1, e^{i\pi/4}]$, and the CNOT. For technical reasons, we shall assume that each CNOT is followed by single-qubit gates (that could be identity gates). We shall consider toy noise models usually studied in the quantum fault-tolerance theory: the depolarizing noise and the amplitude damping noise.

First let us describe product QPRs that can be constructed independently for each gate in the ideal circuit. Consider a fixed ideal gate $\mathcal{U}_\beta \in \Gamma$. Let $\mathcal{O}_1, \ldots, \mathcal{O}_p \in \Omega$ be the list of all noisy operations whose support is contained in the support of $\mathcal{U}_\beta$. Consider the following linear program with $2p$ real variables $\mu_1, \ldots, \mu_p, \eta_1, \ldots, \eta_p$.

$$\textbf{minimize} \quad \sum_{\alpha=1}^{p} \mu_\alpha \tag{14}$$

$$\textbf{subject to} \quad \begin{cases} \eta_\alpha \le \mu_\alpha \\ -\eta_\alpha \le \mu_\alpha \\ \mathcal{U}_\beta = \sum_{\alpha=1}^{p} \eta_\alpha \mathcal{O}_\alpha. \end{cases} \tag{15}$$

Suppose $\{\mu_\alpha, \eta_\alpha\}$ is the optimal solution of the program. Note that $\mu_\alpha = |\eta_\alpha|$ for all α since otherwise the objective function can be decreased. Define $\gamma_\beta = \sum_{\alpha=1}^{p} \mu_p$, $P_\beta(\alpha) = \mu_\alpha/\gamma_\beta$, and $\sigma_\beta(\alpha) = \mathrm{sgn}(\eta_\alpha)$. Then

$$\mathcal{U}_\beta = \gamma_\beta \sum_{\alpha=1}^{p} P_\beta(\alpha)\sigma_\beta(\alpha)\mathcal{O}_\alpha \tag{16}$$

which is a gate-wise version of the QPR Eq. (10). We shall say that a noisy basis Ω simulates a gate $\mathcal{U}_\beta$ with the overhead γ_β if the linear program Eqs. (14,15) has a feasible solution with value γ_β. A product QPR of the ideal circuit $\boldsymbol{\beta}$ is defined as a product of all gate-wise QPRs Eq. (16). It gives $\gamma_{\boldsymbol{\beta}} = \gamma_{\beta_1} \cdots \gamma_{\beta_L}$, $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = P_{\beta_1}(\alpha_1) \cdots P_{\beta_L}(\alpha_L)$ and $\sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \sigma_{\beta_1}(\alpha_1) \cdots \sigma_{\beta_L}(\alpha_L)$. The assumption that all noisy operations $\mathcal{O}_\alpha$ in Eq. (15) act non-trivially only within the support of $\mathcal{U}_\beta$ allows one to restrict Eq. (15) to operations acting on at most two qubits. Such operations can be represented by real matrices of size 16×16 by computing matrix elements of $\mathcal{O}_\alpha$ and $\mathcal{U}_\beta$ in the Pauli basis. Thus the program Eqs. (14,15) can be solved in time $O(1)$. Since the ideal gate set has size $O(n^2)$, product QPRs can be computed in time $O(n^2)$. Furthermore, if two ideal gates have disjoint supports then the gate-wise QPRs defined in Eq. (16) have disjoint supports. Thus product QPRs preserve the circuit depth.

Let us illustrate the construction of product QPRs using the depolarizing noise as an example. Let $\mathcal{D}_k$ be the ϵ-depolarizing channel on k 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 $\mathcal{D}_k\mathcal{U}$. Define a noisy basis Ω 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 [1]. Here $k = 1, 2$. We claim that Ω simulates ideal single-qubit gates $\mathcal{U}_\beta \in \Gamma$ with the overhead $\gamma_\beta = (1 + \epsilon/2)/(1 - \epsilon)$ and simulates CNOTs with the overhead $\gamma_\beta = (1 + 7\epsilon/8)/(1 - \epsilon)$. Indeed, suppose $\mathcal{U}_\beta \in \Gamma$ is a single-qubit gate. Let us look for a solution of Eq. (15) in the form $\mathcal{O}_\alpha = \mathcal{D}_1\mathcal{P}\mathcal{U}_\beta$, where $\mathcal{P} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$. Then Eq. (15) is equivalent to

$$\mathcal{D}_1^{-1} = \eta_1\mathcal{I} + \eta_2\mathcal{X} + \eta_3\mathcal{Y} + \eta_4\mathcal{Z}.$$

One can easily check that the optimal solution minimizing $\sum_\alpha |\eta_\alpha|$ is $\eta_1 = 1 + 3\epsilon/4(1 - \epsilon)$ and $\eta_\alpha = -\epsilon/4(1 - \epsilon)$ for $\alpha = 2, 3, 4$. Therefore $\gamma_\beta = \sum_\alpha |\eta_\alpha| = (1 + \epsilon/2)/(1 - \epsilon)$. The CNOT is simulated in a similar fashion by representing $\mathcal{D}_2^{-1}$ as a linear combination of two-qubit Pauli maps. 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 $p_i(\boldsymbol{\beta})$ from Eq. (13). 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 $\mathcal{U}_{\boldsymbol{\beta}}$ to obtain $\mathcal{O}_{\boldsymbol{\alpha}}$. Numerical implementation of the above method is described in Appendix .

A more interesting example is the noise described by the amplitude damping channel $\mathcal{A}(\rho) = A_0\rho A_0^\dagger + A_1\rho A_1^\dagger$, where

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & (1-\epsilon)^{1/2} \end{bmatrix} \quad \text{and} \quad A_1 = \epsilon^{1/2} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

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

[1] A Pauli map $\mathcal{P}$ corresponding to a Pauli operator $P \in \{I, X, Y, Z\}$ is defined by $\mathcal{P}(\rho) = P\rho P$.

gates $\mathcal{A}^{\otimes k}\mathcal{U}$ alone cannot simulate any ideal unitary gate. Indeed, assume the contrary. Suppose $\mathcal{U}_\beta$ is a single-qubit gate that has a QPR Eq. (16) with $\mathcal{O}_\alpha = \mathcal{A}\mathcal{V}_\alpha$ for some unitary maps $\mathcal{V}_\alpha$. Rewrite Eq. (16) as

$$\mathcal{A}^{-1} = \gamma_\beta \sum_{\alpha=1}^{p} P_\beta(\alpha)\sigma_\beta(\alpha)\mathcal{V}_\alpha\mathcal{U}_\beta^{-1}.$$

Since the maps $\mathcal{V}_\alpha\mathcal{U}_\beta^{-1}$ are unital, we infer that $\mathcal{A}^{-1}$ and $\mathcal{A}$ are unital which is false. Thus Eq. (16) has no solutions.

To overcome this problem we shall extend the noisy basis by adding state preparations. Also we shall employ non-product QPRs. Given a single-qubit state $|\psi\rangle$, define a state preparation map

$$\mathcal{P}_{|\psi\rangle}(\rho) = \mathrm{Tr}(\rho)\cdot|\psi\rangle\langle\psi| \tag{17}$$

Let $\mathcal{S}(\rho) = S\rho S^{-1}$ be the S-gate considered as a TPCP map. Define a noisy basis Ω that includes noisy state preparations $\mathcal{A}\mathcal{P}_{|\psi\rangle}$ with $|\psi\rangle = |+\rangle, |-\rangle, |0\rangle, |1\rangle$, noisy single-qubit gates $\mathcal{A}\mathcal{U}_\beta$, $\mathcal{A}\mathcal{S}^{\pm 1}\mathcal{U}_\beta$ for each ideal single-qubit gate $\mathcal{U}_\beta \in \Gamma$, and noisy two-qubit gates

$$\mathcal{A}_c\mathcal{A}_t\mathcal{S}_c^y\mathcal{S}_t^z\mathcal{U}_{\rm cnot}, \quad y,z \in \{0,\pm 1\}$$

where c,t are the control and the target qubits of a CNOT gate $\mathcal{U}_{\rm cnot} \in \Gamma$. Here the subscripts indicate qubits acted upon by each map. We claim that this noisy basis Ω simulates any ideal Clifford+T circuit $\boldsymbol{\beta}$ with the overhead

$$\gamma_{\boldsymbol{\beta}} \le \gamma^{L_1+2L_2}, \quad \gamma \equiv \frac{1+\epsilon}{1-\epsilon}, \tag{18}$$

where L_k is the number of k-qubit gates in $\boldsymbol{\beta}$. The corresponding QPR Eq. (10) preserves the circuit depth, although it does not have a simple product form as above.

Indeed, consider a single-qubit gate $\mathcal{U}_\beta \in \Gamma$. Let us look for a solution of Eq. (15) with $p = 4$ and

$$\mathcal{O}_1 = \mathcal{A}\mathcal{U}_\beta,\ \mathcal{O}_2 = \mathcal{A}\mathcal{S}\mathcal{U}_\beta,\ \mathcal{O}_3 = \mathcal{A}\mathcal{S}^{-1}\mathcal{U}_\beta,\ \mathcal{O}_4 = \mathcal{P}_{|0\rangle}.$$

Note that $\mathcal{P}_{|0\rangle} \in \Omega$ since $\mathcal{A}\mathcal{P}_{|0\rangle} = \mathcal{P}_{|0\rangle}$. Furthermore, since $\mathcal{P}_{|0\rangle} = \mathcal{A}\mathcal{P}_{|0\rangle}\mathcal{U}_\beta$, one can rewrite Eq. (15) as

$$\mathcal{A}^{-1} = \eta_1\mathcal{I} + \eta_2\mathcal{S} + \eta_3\mathcal{S}^{-1} + \eta_4\mathcal{P}_{|0\rangle}. \tag{19}$$

One can easily check that the optimal solution minimizing $\sum_\alpha |\eta_\alpha|$ is

$$\eta_1 = \frac{1}{\sqrt{1-\epsilon}},\ \eta_2 = \eta_3 = \frac{1-\sqrt{1-\epsilon}}{2(1-\epsilon)},\ \eta_4 = -\frac{\epsilon}{1-\epsilon}.$$

Therefore, Ω simulates $\mathcal{U}_\beta$ with the overhead $\sum_\alpha |\eta_\alpha| = \gamma$, where γ is defined in Eq. (18).

Next consider the CNOT gate $\mathcal{U}_{\rm cnot} \in \Gamma$. Consider a decomposition of $\mathcal{A}_c^{-1}\mathcal{A}_t^{-1}$ obtained by applying Eq. (19) twice. Multiplying this decomposition on the right by $\mathcal{U}_{\rm cnot}$ and on the left by $\mathcal{A}_c\mathcal{A}_t$ one obtains

$$\mathcal{U}_{\rm cnot} = \sum_\alpha \eta_\alpha \mathcal{O}'_\alpha\mathcal{O}_\alpha, \quad \sum_\alpha |\eta_\alpha| = \gamma^2 \tag{20}$$

where $\mathcal{O}_\alpha = \mathcal{A}_c\mathcal{A}_t\mathcal{S}_c^y\mathcal{S}_t^z\mathcal{U}_{\rm cnot} \in \Omega$ is a valid noisy operation and $\mathcal{O}'_\alpha$ is either identity or a state preparation map $\mathcal{P}_{|0\rangle}$ applied to the control and/or target qubits. Here we noted that $\mathcal{A}\mathcal{P}_{|0\rangle} = \mathcal{P}_{|0\rangle}\mathcal{A}$. Although $\mathcal{O}'_\alpha\mathcal{O}_\alpha$ might not be a valid noisy operation from Ω, we may merge $\mathcal{O}'_\alpha$ with the next gate applied after the CNOT. Indeed, by assumption, each CNOT in the ideal circuit is followed by some single-qubit gates $\mathcal{U}_c$ and $\mathcal{U}_t$ applied to the control and the target qubits. The gates I, S, T can be absorbed into $\mathcal{P}_{|0\rangle}$ since they act trivially on the state $|0\rangle$. The only non-trivial case is when $\mathcal{P}_{|0\rangle}$ is merged with the Hadamard gate. In this case the latter is replaced by the state preparation $\mathcal{P}_{|+\rangle}$.

Since $\mathcal{P}_{|+\rangle}$ can now appear in the ideal circuit, we must be able to use noisy operations from Ω to simulate $\mathcal{P}_{|+\rangle}$. Let us look for a solution of Eq. (15) with $\mathcal{U}_\beta \equiv \mathcal{P}_{|+\rangle}$ in the form

$$\mathcal{P}_{|+\rangle} = \eta_1\mathcal{A}\mathcal{P}_{|+\rangle} + \eta_2\mathcal{A}\mathcal{P}_{|-\rangle} + \eta_3\mathcal{A}\mathcal{P}_{|1\rangle}. \tag{21}$$

Note that the righthand side of Eq. (21) contains only noisy operations from Ω. One can rewrite Eq. (21) as

$$|+\rangle\langle+| = \eta_1\mathcal{A}(|+\rangle\langle+|) + \eta_2\mathcal{A}(|-\rangle\langle-|) + \eta_3\mathcal{A}(|1\rangle\langle 1).$$

The optimal solution minimizing $\sum_\alpha |\eta_\alpha|$ is

$$\eta_{1,2} = \pm\frac{1}{2}\left(\frac{1}{\sqrt{1-\epsilon}} \pm \frac{1-2\epsilon}{1-\epsilon}\right), \quad \eta_3 = \frac{\epsilon}{1-\epsilon}.$$

Therefore Ω simulates the ideal state preparation $\mathcal{P}_{|+\rangle}$ with the overhead $\gamma' = \sum_\alpha |\eta_\alpha| \le \gamma$, where γ is defined in Eq. (18).

A QPR of the ideal circuit $\boldsymbol{\beta}$ with the overhead Eq. (18) is constructed in two steps. First, one applies the decomposition Eq. (20) to each CNOT of $\boldsymbol{\beta}$ and merges state preparation maps $\mathcal{P}_{|0\rangle}$ that appear in $\mathcal{O}'_\alpha$ (if any) with the single-qubit gates of $\boldsymbol{\beta}$ following the CNOT. Now all CNOT gates are replaced by noisy gates from Ω. The rest of the circuit consists of single-qubit gates $\mathcal{U}_\beta \in \Gamma$ and state preparations $\mathcal{P}_{|+\rangle}$. At the second step each of these ideal operations is replaced by its QPR constructed above. Note that each CNOT contributes γ^2 to the total overhead $\gamma_{\boldsymbol{\beta}}$, see Eq. (20). Each single-qubit gate $\mathcal{U}_\beta \in \Gamma$ or a state preparation $\mathcal{P}_{|+\rangle}$ contributes at most γ to the total overhead. This proves Eq. (18).

The examples considered above suggest that noisy circuits in which the noise is well characterized can simulate ideal ones with the 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 for a given device by performing quantum process tomography [28] and solving the linear program Eqs. (14,15) for each ideal gate. Using Eq. (12) one can estimate the number of noisy circuit runs M required to simulate an ideal length-L circuit as $M \sim \exp(2c\epsilon L)$. Assuming that near-term quantum devices can achieve error rates in the range $\epsilon \sim 10^{-3}$, it may be possible to use such devices to simulate ideal circuits with $O(1000)$ gates without resorting to the conventional quantum error correction. It appears likely that such simulations cannot be accomplished using only classical computers.

Construction of QPRs with small overhead for more general operation-dependent noise is an interesting open problem. A preliminary analysis shows that a noisy basis Ω that includes noisy versions of all single-qubit and two-qubit Clifford gates, T-gates, and noisy qubit initializations in the X, Y, Z basis can simulate any ideal gate $\mathcal{U}_\beta$ from the Clifford+T gate set with the overhead $\gamma_\beta \leq 1+O(\epsilon)$ provided that each noisy operation is ϵ-close to its ideal analogue. Unfortunately, the constant coefficient in this upper bound is far too large to have any practical implications. Furthermore, the full Clifford group on two-qubits contains 11520 gates. It may not be feasible to perform process tomography for each of those gates.

A full characterization of noise for each computational operation with the precision in the range 10^{-3} may be impossible in practice. However, if one is willing to sacrifice optimality, a Pauli twirling [29] or a Clifford twirling [30] can be applied that converts any noise channel into a simple probabilistic mixture of Pauli errors or the depolarizing noise. The latter depends only on a few parameters making the noise characterization task more manageable.

CONCLUSIONS

In this paper we have presented two different schemes for mitigating the effects of decoherence and noise in short depth quantum circuits. The first scheme is able to minimize the effect of errors due to decoherence and noise on the estimation of expectation values, if the initial noise rate is sufficiently small. For this scheme the explicit knowledge of the noise model is not necessary and the method works for both phenomenological local Markovian noise models as well as noise derived from first principle by jointly evolving system and bath unitarily. The second scheme works for errors which are larger in magnitude and is able to deal well with typical gate errors if the underlying noise model and the gate error can be sufficiently well characterized. Most importantly, both approaches do not require any additional hardware and overhead in resources. This means that these error mitigation techniques can be applied immediately in current and near-term experiments to increase the quality of results.

A very recent independent paper by Li and Benjamin [31] discusses similar issues to those addressed here.

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

[1] P. W. Shor, SIAM review **41**, 303 (1999).
[2] W. G. Unruh, Physical Review A **51**, 992 (1995).
[3] P. W. Shor, Physical review A **52**, R2493 (1995).
[4] A. M. Steane, Physical Review Letters **77**, 793 (1996).
[5] A. R. Calderbank and P. W. Shor, Physical Review A **54**, 1098 (1996).
[6] D. Aharonov and M. Ben-Or, in *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing* (ACM, 1997) pp. 176–188.
[7] A. Y. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[8] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Physical Review A **86**, 032324 (2012).
[9] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Physical Review X **2**, 031007 (2012).
[10] S. J. Devitt, A. M. Stephens, W. J. Munro, and K. Nemoto, Nature communications **4** (2013).
[11] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, arXiv preprint quant-ph/9611028 (1996).
[12] F. Pastawski, A. Kay, N. Schuch, and I. Cirac, Physical review letters **103**, 080501 (2009).
[13] B. M. Terhal and D. P. DiVincenzo, Quant. Inf. Comp. **4**, 134 (2004).
[14] E. Farhi and A. W. Harrow, arXiv preprint arXiv:1602.07674 (2016).
[15] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O?Brien, Nature communications **5** (2014).
[16] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, arXiv preprint arXiv:1509.04279 (2015).
[17] D. Wecker, M. B. Hastings, and M. Troyer, Physical Review A **92**, 042303 (2015).
[18] E. Farhi, J. Goldstone, and S. Gutmann, arXiv preprint arXiv:1411.4028 (2014).
[19] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, arXiv preprint arXiv:1510.03859 (2015).
[20] P.-L. Dallaire-Demers and F. K. Wilhelm, Physical Review A **93**, 032303 (2016).
[21] N. C. Rubin, arXiv preprint arXiv:1610.06910 (2016).
[22] L. F. Richardson and J. A. Gaunt, Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character **226**, 299 (1927).
[23] A. Sidi, "Practical extrapolation methods: Theory and applications, volume 10 of cambridge monographs on applied and computational mathematics," (2003).
[24] C. Schneider, Numerische Mathematik **24**, 177 (1975).
[25] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and quantum computation*, Vol. 47 (American Mathematical Society Providence, 2002).
[26] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Physical review letters **115**, 070501 (2015).
[27] N. Delfosse, P. A. Guerin, J. Bian, and R. Raussendorf, Physical Review X **5**, 021003 (2015).
[28] M. Mohseni, A. Rezakhani, and D. Lidar, Phys. Rev. A **77**, 032322 (2008).
[29] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Physical Review A **78**, 012347 (2008).
[30] E. Magesan, J. M. Gambetta, and J. Emerson, Physical Review A **85**, 042311 (2012).
[31] Y. Li and S. C. Benjamin, arXiv preprint arXiv:1611.09301 (2016).

SUPPLEMENTAL MATERIAL

Reducing noise by Richardson extrapolation

In numerical analysis, Richardson extrapolation is a sequence acceleration method, used to improve the rate of convergence of a sequence. We use the same technique to extrapolate to the zero noise limit in short depth quantum circuits in the presence of noise. We assume that the noise process is constant in time and does not depend on the rescaling of the system Hamiltonian parameters. We consider various noise models in continuous time.

It is our goal to estimate the expectation value of some observable A with respect to the evolved state $\rho_\lambda(T)$. The actual computation is now encoded in the time-dependent Hamiltonian $K(t)$, and the full evolution is given by the equation

$$\frac{\partial}{\partial t}\rho = -i[H(t),\rho] + \lambda\mathcal{L}(\rho), \tag{22}$$

where we have that

$$K_\alpha(t) = \sum_\alpha J_\alpha(t) P_\alpha \tag{23}$$

is some multi-qubit Pauli Hamiltonian with time-dependent coupling constant $J_\alpha(t)$. Starting from some initial state ρ_0 the system is evolved for some time T. We consider different forms of noise $\mathcal{L}(\rho)$. As the simplest form of noise, we assume a time-independent Lindblad operator of the form

$$\mathcal{L}(\rho) = \sum_\beta L_\beta \rho L_\beta^\dagger - \frac{1}{2}\{L_\beta^\dagger L_\beta, \rho\}. \tag{24}$$

However, we can also imagine other forms of errors such as,

$$\mathcal{L}(\rho) = -i[V,\rho], \tag{25}$$

where V is some Hamiltonian. This setting in useful when we want to consider more general, possibly non-Markovian noise models, or a noisy evolution derived from first principle. One can make the assumption that the initial state is given by $\rho_0 = \rho_S(0)\otimes\rho_B(0)$ and give the most general form of an interaction Hamiltonian between system and bath, such as

$$V = \sum_\alpha S_\alpha \otimes B_\alpha + H_B \tag{26}$$

Here we take the point of view that a small λ indicates a separation of time scales and $\rho_S(0) = |\psi_0\rangle\langle\psi_0|$ may be the initial state of the computation. We assume that $\rho(0) = \rho_S(0)\otimes\rho_B$, where the bath state is a steady state w.r.t the bath Hamiltonian $[H_B,\rho_B] = 0$. The observable we want to estimate $A = A_S\otimes\mathbb{1}$ is then only supported on the system degrees of freedom.

We assume that $\rho_\lambda(T)$ is the state which we obtain after the noisy evolution for time T. From this state we can estimate the expectation value of the observable A, by various methods. Typically we will sample the expectation value

$$E_K(\lambda) = \mathrm{tr}\left(A\rho_\lambda(T)\right). \tag{27}$$

so that an additional sampling error δ is introduced, and we obtain from out measurement the statistic $\hat{E}_K(\lambda) = E_K(\lambda) + \delta$. The error can for be asymptotically Gaussian $\delta = \mathcal{O}\left(M^{-1/2}\sqrt{\mathrm{tr}\left(\rho_\lambda(T)(A - E_K(\lambda))^2\right)}\right)$ since one typically repeats the experiment $M \gg 1$ times so that the i.i.d hypothesis holds.

I Series expansion in the noise parameter

We now show that 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. We also provide a bound on the error term. To this end, we transform into the interaction picture of $K(t)$. We define $U_K(t) = \mathcal{T}\left\{\exp(-i\int_0^t K(t')dt')\right\}$, where $\mathcal{T}\{\cdot\}$ defines the time order expansion. We define the interaction picture through

$$\rho_I(t) = U_K(t)\rho(t)U_K^\dagger(t) \quad \text{and} \quad \mathcal{L}_{I,t}(\circ) = U_K(t)\mathcal{L}\left(U_K^\dagger(t)\circ U_K(t)\right)U_K^\dagger(t), \tag{28}$$

where now the generator $\mathcal{L}_{I,t}$ has become time-dependent The evolution equation in the interaction picture now reads

$$\partial_t \rho_I(t) = \lambda \mathcal{L}_{I,t}\left(\rho_I(t)\right), \tag{29}$$

Recall that every first order differential equation can be reformulated as an integral equation

$$\rho_I(T) = \rho_I(0) + \lambda \int_0^T \mathcal{L}_{I,t}\left(\rho_I(t)\right) dt. \tag{30}$$

This equation can be recursively solved to increasing order in λ so that

$$\begin{aligned}\rho_I(T) = \rho_I(0) + \lambda \int_0^T \mathcal{L}_{I,t}\left(\rho_I(0)\right) dt + \lambda^2 \int_0^T \int_0^t \mathcal{L}_{I,t} \circ \mathcal{L}_{I,t'}\left(\rho_I(0)\right) dt dt' \\ + \lambda^3 \int_0^T \int_0^t \int_0^{t'} \mathcal{L}_{I,t} \circ \mathcal{L}_{I,t'} \circ \mathcal{L}_{I,t''}\left(\rho_I(0)\right) dt dt' dt'' \ldots .\end{aligned} \tag{31}$$

Recall that $\rho_I(0) = \rho(0)$. Furthermore, we can conjugate the full expression on both sides with the unitary $U_K(T)$. Let us for notational convenience define $\rho_\lambda(T)$ as the resulting state after evolution with noise rate λ. We observe that $U_K(T)^* \rho(0) U_K(T) = \rho_0(T)$, whereas $U_K(T)^* \rho_I(T) U_K(T) = \rho_\lambda(T)$, so that we obtain the expression in the Schrödinger picture as

$$\begin{aligned}\rho_\lambda(T) = \rho_0(T) + \sum_{k=1}^{n} \lambda^k \int_0^T \int_0^{t_1} \ldots \int_0^{t_{k-1}} U_K^\dagger(T) \mathcal{L}_{I,t_1} \circ \mathcal{L}_{I,t_2} \circ \ldots \circ \mathcal{L}_{I,t_k}\left(\rho(0)\right) U_K(T) dt_1 dt_2 \ldots dt_k \\ + \lambda^{n+1} \int_0^T \int_0^{t_1} \ldots \int_0^{t_n} U_K^\dagger(T) \mathcal{L}_{I,t_1} \circ \mathcal{L}_{I,t_2} \circ \ldots \circ \mathcal{L}_{I,t_n+1}\left(\rho_I(t_{n+1})\right) U_K(T) dt_1 dt_2 \ldots dt_{n+1}.\end{aligned} \tag{32}$$

The expectation value $E_K(\lambda) = \mathrm{tr}\left(A \rho_\lambda(T)\right)$ for the observable A can immediately be expanded in a series with parameter λ of the form

$$E_K(\lambda) = \mathrm{tr}\left(A \rho_0(T)\right) + \sum_{k=1}^{n} a_k \lambda^k + R_{n+1}(\lambda, \mathcal{L}, T). \tag{33}$$

where the constants a_k and the remainder $R_{n+1}(\lambda, \mathcal{L}, T)$ are obtained by pairing the integrals with the trace $\mathrm{tr}\left(A \cdot\right)$ and $\mathrm{tr}\left(A \rho_0(T)\right) = E^*$ corresponds to the noise-free evolution we seek to extrapolate to. We read off that

$$a_k = \int_0^T \int_0^{t_1} \ldots \int_0^{t_{k-1}} \mathrm{tr}\left(U_K(T) A U_K^\dagger(T) \mathcal{L}_{I,t_1} \circ \mathcal{L}_{I,t_2} \circ \ldots \circ \mathcal{L}_{I,t_k}\left(\rho(0)\right)\right) dt_1 dt_2 \ldots dt_k, \tag{34}$$

as well as

$$R_{n+1}(\lambda, \mathcal{L}, T) = \lambda^{n+1} \int_0^T \int_0^{t_1} \ldots \int_0^{t_n} \mathrm{tr}\left(U_K(T) A U_K^\dagger(T) \mathcal{L}_{I,t_1} \circ \ldots \circ \mathcal{L}_{I,t_n+1}\left(\rho_I(t_{n+1})\right)\right) dt_1 \ldots dt_{n+1} \tag{35}$$

We can bound $|R_{n+1}(\lambda, \mathcal{L}, T)|$ by a simple application of Cauchy's mean value theorem and Hölder's inequality. We observe by first applying the midpoint Theorem that there exist $\xi_1, \ldots, \xi_{n+1}$ so that

$$R_{n+1}(\lambda, \mathcal{L}, T) = \frac{\lambda^{n+1} T^{n+1}}{(n+1)!} \mathrm{tr}\left(U_K(T) A U_K^\dagger(T) \mathcal{L}_{I,\xi_1} \circ \ldots \circ \mathcal{L}_{I,\xi_n+1}\left(\rho_I(\xi_{n+1})\right)\right). \tag{36}$$

We can then of course immediately bound the inner product

$$|\mathrm{tr}\left(U_K(T) A U_K^\dagger(T) \mathcal{L}_{I,t_1} \circ \ldots \circ \mathcal{L}_{I,t_n+1}\left(\rho_I(t_{n+1})\right)\right)| \leq \|A\| \| \mathcal{L}_{I,t\xi_1} \circ \ldots \circ \mathcal{L}_{I,\xi_n+1}\left(\rho_I(\xi_{n+1})\right) \|_1. \tag{37}$$

by a direct application of Hölder's inequality. Note that all Schatten norms are unitarily invariant, so when the map $\mathcal{L}$ is bounded, we can apply the subsequent operator norm inequalities

$$\| \mathcal{L}_{I,t\xi_1} \circ \ldots \circ \mathcal{L}_{I,\xi_n+1}\left(\rho_I(\xi_{n+1})\right) \|_1 \leq \|\mathcal{L}\|_{1\to 1}^{n+1}. \tag{38}$$

Alternatively, if $\mathcal{L}$ is unbounded and models the unitary evolution between a system and bath, we can transform into the Heisenberg picture for the observable $A = A_S \otimes \mathbb{1}$, and bound

$$|\mathrm{tr}\left(U_K(T)AU_K^\dagger(T)\mathcal{L}_{I,t_1}\circ\ldots\circ\mathcal{L}_{I,t_n+1}\left(\rho_I(t_{n+1})\right)\right)| \leq \|A(T)\|\|\mathcal{L}_{I,t\xi_1}\circ\ldots\circ\mathcal{L}_{I,\xi_n+1}\left(\rho_I(0)\right)\|_1. \tag{39}$$

We obtain almost the same type of bound from Hölder's inequality since $\|A(T)\| = \|A\|$ where the only difference is now that $\|\mathcal{L}_{I,t\xi_1}\circ\ldots\circ\mathcal{L}_{I,\xi_n+1}\left(\rho_I(0)\right)\|_1 \leq l_{n+1}$ only depends on the inital state. In either case, we will now write for the bound on $|R_{n+1}(\lambda,\mathcal{L},T)|$ from now on

$$|R_{n+1}(\lambda,\mathcal{L},T)| \leq \|A\|\ l_{n+1}\ \frac{\lambda^{n+1}T^{n+1}}{(n+1)!}. \tag{40}$$

Furthermore, the coefficients a_k can be bounded in a similar fashion. Note, that, if we assume that noise acts locally on each qubit, such as for instance, when the dissipator $\mathcal{L}$ corresponds to single-qubit depolarizing noise, so that $\mathcal{L}(\rho) = \sum_{i=1}^{N}(\frac{1}{2}\mathrm{tr}_{[i]}(\rho) - \rho)$. We have that $||\mathcal{L}||_{1\to 1} = \mathcal{O}(N)$ is extensive in the system size. A similar argument holds for the case when the individual qubits couple to a bath. From this we can deduce that for local noise we typically find $l_k = \mathcal{O}(N^k)$ as mentioned in the main text, and that $|a_k| \leq \mathcal{O}((N\ T)^k)$.

II Experimental rescaling of the noise parameter

In order to apply Richardson extrapolation, we have to be able to evaluate $E_K(\lambda)$ for different values of λ. In an actual experiment, we can't directly control the parameter λ, however, we may control the evolution $K(t)$. To this end, we introduce a rescaling. We redefine

$$T \to T' = cT \quad \text{as well as} \quad J(t) \to J'(t) = c^{-1}J(c^{-1}t) \quad \text{from which also} \quad \rho(t) \to \rho'(t) = \rho(h^{-1}t) \tag{41}$$

We claim that this rescaling maps $\rho'_\lambda(T') = \rho_{h\lambda}(T)$ if the noise operator does not also depend on the Hamiltonian couplings $J(t)$ and is constant in time. This rescaled density matrix then leads to a new evaluation $E'_K(\lambda) \to E_K(c\lambda)$ of the expectation value. To see that the rescaling has the desired effect, we again make use of the integral representation of $\rho_\lambda(T)$. For which we can write now in the Schrödinger picture

$$\rho_\lambda(T) = \rho(0) - i\int_0^T [K(t),\rho(t)]dt + \lambda\int_0^T \mathcal{L}(\rho(t))dt. \tag{42}$$

We can now choose a re-parametrization of the evolution $c^{-1}J(c^{-1}t)$ and an increased runtime cT, and write

$$\rho'_\lambda(T') = \rho(0) - i\int_0^{cT} [K'(t),\rho'(t)]dt + \lambda\int_0^{cT} \mathcal{L}(\rho'(t))dt \tag{43}$$

with $K'(t) = \sum_\alpha c^{-1}J_\alpha(c^{-1}t)P_\alpha$. If we now re-parametrize the integral according to $t = ct'$ we have that $dt = cdt'$, which leads to

$$\begin{aligned}
\rho'_\lambda(T') &= \rho(0) - i\int_0^T \sum_\alpha c^{-1}J_\alpha(t')[P_\alpha,\rho(t')]cdt' + \lambda\int_0^T \mathcal{L}(\rho'(t))cdt' \\
&= \rho(0) - i\int_0^T \sum_\alpha [K(t'),\rho(t')]dt' + \lambda c\int_0^T \mathcal{L}(\rho(t'))dt' \\
&= \rho_{c\lambda}(T).
\end{aligned} \tag{44}$$

Hence, rescaling the evolution according to equation (41) leads to an effective rescaling of the dissipative rate λ. This can be done for any dissipator $\mathcal{L}$ and allows the experimenter to evaluate $E_K(\lambda)$ for different values of $c\lambda$ so that we can apply the Richardson extrapolation procedure.

III Error bounds on the noise-free estimator

Let us now show that the **protocol a** leads to the desired error bound on the estimated expectation value as claimed in eqn. (8). Recall that we first choose a set of $n+1$ rescaling parameters $c_0 = 1 < c_1 < \ldots < c_n$, to evolve with respect to the rescaled

Hamiltonian $K^j(t)$ for time $T_j = c_j T$. As discussed in the previous section, this evolution leads to a state $\rho^j_\lambda(T_j) = \rho_{c_j\lambda}(T)$, c.f. eqn (7) as was discussed in section II. If we now measure the observable A on these states we obtain for $j = 0 \dots n+1$ the estimates $\hat{E}_K(c_j\lambda) = E_K(c_j\lambda) + \delta_j$. Recall the set of equations for γ_j defined in eqn. (6), which requires for the $\{c_j\}$ that

$$\begin{aligned} &\sum_{l=0}^{n} \gamma_j = 1 \\ &\sum_{j=0}^{n} \gamma_j \; c_j^k = 0 \quad \text{for } k = 1 \dots n. \end{aligned} \tag{45}$$

Now we observe that estimators $\hat{E}_K(c_j\lambda)$ can be expressed as

$$\hat{E}_K(c_j\lambda) = E^* + \sum_{k=1}^{n} a_k c_j^k \lambda^k + R(c_j\lambda, \mathcal{L}, T) + \delta_j. \tag{46}$$

due to the expansion (33) discussed in section I. Recall now the definition of our improved estimator $\hat{E}^n_K(\lambda)$ as given in eqn. (5) $\hat{E}^n_K(\lambda) = \sum_{j=0}^n \gamma_j \hat{E}_K(c_j\lambda)$, for which then

$$\begin{aligned} \hat{E}^n_K(\lambda) &= \sum_{j=0}^{n} \left(\gamma_j E^* + \sum_{k=1}^{n} a_k c_j^k \lambda^k + R(c_j\lambda, \mathcal{L}, T) + \delta_j \right) \\ &= E^* \left(\sum_{j=0}^{n} \gamma_j \right) + \sum_{k=1}^{n} a_k \lambda^k \left(\sum_{j=0}^{n} \gamma_j c_j^k \right) + \left(\sum_{j=0}^{n} \gamma_j R(c_j\lambda, \mathcal{L}, T) + \delta_j \right). \end{aligned} \tag{47}$$

recall the equations for γ_j from which we can then infer after the application of the triangle inequality

$$|E^* - \hat{E}^n_K(\lambda)| \leq \sum_{j=0}^{n} |\gamma_j| \left(|R(c_j\lambda, \mathcal{L}, T)| + |\delta_j| \right). \tag{48}$$

After the application of the bound $|R(c_j\lambda, \mathcal{L}, T)| \leq \|A\| \; l_{n+1} \; c_j^{n+1} \lambda^{n+1} T^{n+1} ((n+1)!)^{-1}$ and the observation that $c_j \geq 1$, we can bound the difference with $\Gamma_n = \sum_{j=0}^n |\gamma_j| c_j^{n+1}$ as stated in eqn. (8).

IV Depolarizing noise cancellation: numerical results

The error cancellation method was tested numerically for small Clifford+T circuits subject to the depolarizing noise. We choose the ideal circuit $\mathcal{U}_{\boldsymbol{\beta}}$ as a composition of d alternating layers of gates with each layer being either a tensor product of n single-qubit gates I, H, S, T (for odd layers) or a tensor product of $n/2$ CNOTs (for even layers). The resulting circuit $\mathcal{U}_{\boldsymbol{\beta}}$ has depth d. Simulations were performed for 500 random circuits $\mathcal{U}_{\boldsymbol{\beta}}$ as above with the initial state $|+\rangle^{\otimes n}$. Each single-qubit gate was picked randomly from the set $\{I, H, S, T\}$. Control and target qubits for each CNOT were picked at random.

For each circuit $\mathcal{U}_{\boldsymbol{\beta}}$ we defined the output projector Π_{out} by selecting a subset of 2^{n-1} basis vectors whose probability in the final state is relatively high due to the constructive interference. Intuitively, this makes the output probability more sensitive to the depolarizing noise that tends to destroy interference. More formally, we set

$$\Pi_{out} = \arg\max_{\Pi} \mathrm{Tr}[\Pi \, \mathcal{U}_{\boldsymbol{\beta}}(\rho_{in})]$$

where the maximum is over all projectors Π that are diagonal in the Z-basis and have rank 2^{n-1}. Equivalently, Π_{out} projects onto the subspace spanned by 2^{n-1} basis vectors that have the largest probability in the final state $\mathcal{U}_{\boldsymbol{\beta}}(\rho_{in})$. Thus, by construction, $p_i(\boldsymbol{\beta}) \geq 1/2$. We observed that $p_i(\boldsymbol{\beta})$ is well separated from $1/2$ for most of the circuits see Fig. 3. Furthermore, the output probability is $1/2$ in the limit of strong noise.

Recall that we define a noisy version of a k-qubit unitary gate $\mathcal{U}$ as $\mathcal{D}_k\mathcal{U}$, where

$$\mathcal{D}_k(\rho) = (1-\epsilon)\rho + \frac{\epsilon I}{2^k} \mathrm{Tr}(\rho)$$

is the depolarizing channel on k qubits. Noise was added after all gates including the identity gates. In this case the total simulation overhead $\gamma_{\boldsymbol{\beta}}$ depends only on the number of qubits and the circuit depth, namely

$$\gamma_{\boldsymbol{\beta}} = \left[\frac{1+\epsilon/2}{1-\epsilon}\right]^{nd/2} \cdot \left[\frac{1+7\epsilon/8}{1-\epsilon}\right]^{nd/4}.$$

Consider a fixed ideal circuit $\mathcal{U}_{\boldsymbol{\beta}}$ and let $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$, $\mathcal{O}_{\boldsymbol{\alpha}}$ be the random ensemble of noisy circuits obtained from $\mathcal{U}_{\boldsymbol{\beta}}$ by inserting random Pauli operators and adding noise as described in the main text. Instead of using the estimate Eq. (13) for the ideal output probability $p_i(\boldsymbol{\beta})$ we opted for a slightly optimized estimate. It is defined by dividing the total budget of M runs into K groups such that the j-th group contains M_j runs,

$$M = \sum_{j=1}^{K} M_j.$$

Define a random variable

$$\xi \equiv \gamma_{\boldsymbol{\beta}} K^{-1} \sum_{j=1}^{K} \sigma_{\boldsymbol{\beta}}(\boldsymbol{\alpha}^j) \frac{1}{M_j} \sum_{a=1}^{M_j} \langle x_j^a | \Pi_{out} | x_j^a \rangle, \tag{49}$$

where $\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^K$ are independent samples drawn from the distribution $P_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ and $x_j^a \in \{0,1\}^n$ are readout strings obtained by measuring each qubit of the final state $\mathcal{O}_{\boldsymbol{\alpha}^j}(\rho_{in})$ in the Z-basis. We prepare a fresh copy of the final state to generate each string x_j^a. Thus computing ξ requires M runs of the noisy circuits with each run producing a single readout string. One can easily check that ξ is an unbiased estimator of $p_i(\boldsymbol{\beta})$ for any choice of $\{M_j\}$. Our goal is to choose $\{M_j\}$ that minimize the variance of ξ for a fixed M. One can easily check that the optimal choice is

$$M_j \approx \frac{M\sigma_j}{\sum_{i=1}^{K} \sigma_i}$$

where $\sigma_j^2 = p(\boldsymbol{\alpha}^j) - p(\boldsymbol{\alpha}^j)^2$ and $p(\boldsymbol{\alpha}^j)$ is the output probability of the noisy circuit $\mathcal{O}_{\boldsymbol{\alpha}^j}$. In order to choose optimal values of M_j one has to run each circuit $\boldsymbol{\alpha}^j$ at least a few times which gives a rough estimate of $p(\boldsymbol{\alpha}^j)$ and thus σ_j. Numerical simulations were performed for the following parameters:

number of qubits	$n = 6$
circuit depth	$d = 20$
error rate	$\epsilon = 0.01$
total number of runs	$M = 4,000$
simulation overhead	$\gamma_{\boldsymbol{\beta}} \approx 4.3$

Our results are presented on the left panel of Figure 2. For each of ≈ 500 ideal circuits $\boldsymbol{\beta}$ generated at random we computed a simulation precision $\delta(\boldsymbol{\beta}) \equiv |\xi - p_i(\boldsymbol{\beta})|$ where ξ is the estimate defined in Eq. (49). The plot on Figure 2, left, shows distribution of the ideal circuits $\boldsymbol{\beta}$ according to their simulation precision $\delta(\boldsymbol{\beta})$. The median value of $\delta(\boldsymbol{\beta})$ is approximately 0.05. This is consistent with the estimate

$$\delta(\boldsymbol{\beta}) \approx \frac{\gamma_{\boldsymbol{\beta}}}{\sqrt{M}} = \frac{4.3}{\sqrt{4000}} \approx 0.07.$$

We also computed a simulation precision $\delta_0(\boldsymbol{\beta})$ that one would obtain by running the circuit $\boldsymbol{\beta}$ directly on a noisy device without error cancellation, see the right panel of Figure 2. It is defined as $\delta_0(\boldsymbol{\beta}) = |p(\boldsymbol{\beta}) - p_i(\boldsymbol{\beta})|$. For each circuit the output probability $p(\boldsymbol{\beta})$ was estimated using $M = 4,000$ circuit runs. Thus the simulations presented on the left and the right panels of Figure 2 have access to exactly the same resources. The median value of $\delta_0(\boldsymbol{\beta})$ is approximately 0.15. We conclude that error cancellation significantly improves the simulation precision.

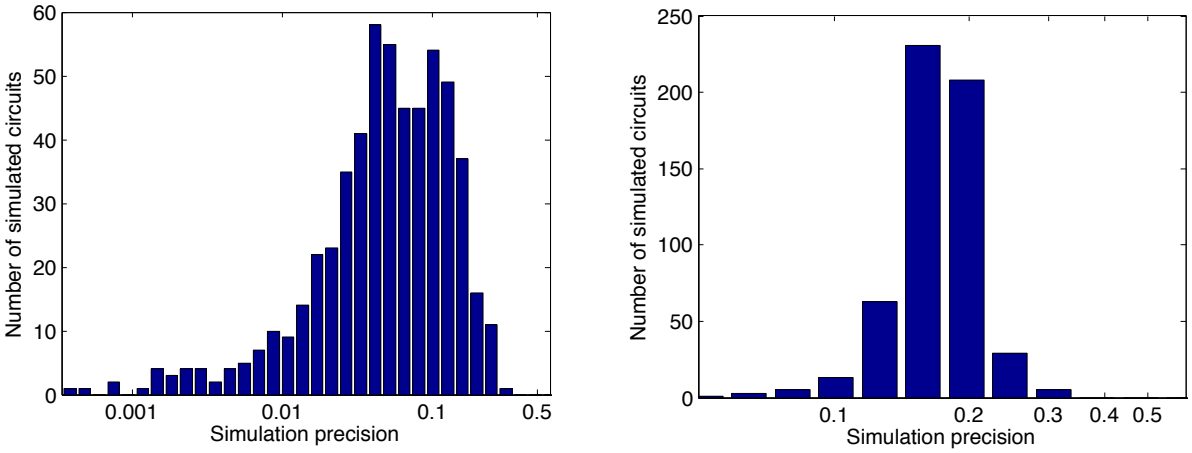

FIG. 2. Simulation precision for ≈ 500 randomly generated ideal Clifford+T circuits on $n = 6$ qubits with depth $d = 20$. The left and the right panels show results for simulations with and without error cancellation. In both cases each ideal circuit was simulated by $M = 4000$ runs of the noisy circuit.

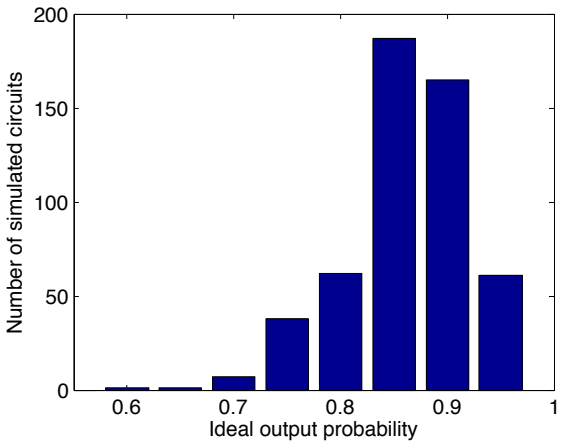

FIG. 3. Distribution of the ideal circuits according to their output probability.