Characterizing quantum gates via randomized benchmarking

Easwar Magesan,[1,2] Jay M. Gambetta,[3] and Joseph Emerson[1,2]

[1]*Department of Applied Mathematics, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[2]*Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[3]*IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA*

We describe and expand upon the scalable randomized benchmarking protocol proposed in Phys. Rev. Lett. **106**, 180504 (2011) which provides a method for benchmarking quantum gates and estimating the gate dependence of the noise. The protocol allows the noise to have weak time and gate dependence, and we provide a sufficient condition for the applicability of the protocol in terms of the average variation of the noise. We discuss how state-preparation and measurement errors are taken into account and provide a complete proof of the scalability of the protocol. We establish a connection in special cases between the error rate provided by this protocol and the error strength measured using the diamond norm distance.

　　　　　　　　　　　　PACS number(s): 03.67.Pp, 03.65.Wj, 03.67.Lx

I. INTRODUCTION

Quantum computers promise an exponential speedup over known classical algorithms for problems such as factoring integers [1], finding solutions to linear systems of equations [2], and simulating physical systems [3,4]. Quantum error-correction methods have been devised for preserving quantum information in the presence of noise [5–7], leading to the theoretical development of a fault-tolerant theory of quantum computing [8–10]. Such a theory promises that quantum computation is possible in the presence of errors, provided the error rate is below a certain threshold value which depends on the particular coding scheme used as well as the error model. This potential has motivated much experimental research dedicated to building a functioning quantum-information processor, with various proposals for possible implementations [11–14].

One of the main challenges in building a quantum-information processor is the nonscalability of *completely* characterizing the noise affecting a quantum system via process tomography [15,16]. A complete characterization of the noise is useful because it allows for the determination of good error-correction schemes, and thus the possibility of reliable transmission of quantum information. Since complete process tomography is infeasible for large systems, there is growing interest in scalable methods for *partially* characterizing the noise affecting a quantum system [17–24].

In Ref. [25] we provided a *scalable* (in the number n of qubits comprising the system) and robust method for benchmarking the full set of Clifford gates by a single parameter using randomization techniques. The concept of using randomization methods for benchmarking quantum gates, commonly called *randomized benchmarking* (RB), was introduced previously in Refs. [18,26]. The simplicity of these protocols has motivated experimental implementations in atomic ions for different types of traps [26–28], NMR [29], superconducting qubits [30,31], and atoms in optical lattices [32]. Unfortunately there are several drawbacks to the methods of Refs. [18,26]. For instance [18] assumes the highly idealized situation of the noise being independent of the chosen gate, in which case the fidelity decay curve averaged over randomly chosen unitaries takes the form of an exponential (in the sequence length). The protocol of Ref. [26] is limited

to the single-qubit case and fits the observed fidelity decay averaged over sequences of single-qubit gates (where each gate consists of a random generator of the Clifford group composed with a random Pauli operator) to an exponential. The decay rate is *assumed* to provide an estimate of the average error probability per Clifford gate. However, conditions for when the assumption of an exponential decay is valid, specifically in the realistic case of gate-dependent and time-dependent noise, were not given. Such a set of conditions would be useful because it is easy to construct pathological examples where the estimated decay rate is not reliable. An unphysical but intuitively simple example is when the error is gate dependent and equal to the exact inverse of the target gate. The error rate given by the protocol is always equal to zero; however, in actuality there is substantial error on each gate (see Sec. IV B). Other important shortcomings of these previous RB protocols are that extensions to multiqubit systems are either not scalable or not well understood, and it is unclear how to explicitly account for state-preparation and measurement errors.

In this paper we give a full analysis of the scalable multiqubit randomized benchmarking protocol for Clifford gates we proposed in Ref. [25], which overcomes the shortcomings described above. We note that since one "gate" in the single-qubit protocol of Ref. [26] consists of a random Clifford generator as well as a random Pauli operator, the cost of implementing a gate in this scheme is 2. In the single-qubit case, our RB scheme can be implemented by explicitly writing down the 24 elements of the Clifford group decomposed into a sequence of the same generators that are randomly applied in Ref. [26]. The average number of generators in such a decomposition is 1.875, which implies that even for the single-qubit case our protocol takes no more time to implement than that of Ref. [26]. Hence, since our protocol is scalable and produces an error estimate which overcomes the various shortcomings listed above, it is reasonable to apply it over other existing schemes regardless of the number of qubits comprising the system.

We provide a detailed proof that our protocol requires at most $O(n^2)$ quantum gates, $O(n^4)$ cost in classical preprocessing (to select each gate sequence), and a number of single-shot repetitions that is independent of n. As well, we give a thorough

explanation of the perturbative expansion of the time- and gate-dependent errors about the average error that leads to the fitting models for the observed fidelity decay. Our zeroth-order model directly shows that for time-independent and gate-independent errors the fidelity decay is indeed modeled by an exponential decay, and the decay rate produces an estimate for the average error rate of the noise.

We derive the first-order fitting model which takes into account the first-order correction terms in the perturbative expansion and provide a detailed explanation of the conditions for when this is a sufficient model of the fidelity decay curve. The fitting formula shows that gate-dependent errors can lead to a deviation from the exponential decay (defining a partial test for such effects in the noise), which was illustrated via numerical examples in Ref. [25]. State-preparation and measurement errors appear as independent fit parameters in the fitting models, and we discuss when the protocol is robust against these errors. In the case of Pauli errors we give some preliminary results regarding the relationship between the benchmarking average error rate and the more common diamond norm error measure [33,34] used in fault-tolerant theory.

The paper is structured as follows: In Sec. II we discuss notation and background material. In Sec. III A we discuss the proposed protocol and then in Sec. III B we present the perturbative expansion and expressions for the zeroth- and first-order fitting models. Section IV provides a sufficient condition for neglecting higher-order terms in the model as well as a simple case for when the benchmarking scheme fails. We also discuss when the protocol is robust against state-preparation and measurement errors. Section V discusses the relationship between the error rate given by the benchmarking scheme and other measures of error commonly used in quantum information. Section VI provides a detailed proof that our protocol is scalable in the number of qubits comprising the system, and a discussion with concluding remarks is contained in Sec. VII.

II. BACKGROUND

Let us first set some notation. Suppose we have an n-qubit quantum system so that the Hilbert space $\mathcal{H}$ representing the system has dimension $d = 2^n$. Thus $\mathcal{H}$ is isomorphic to $\mathbb{C}^d$ and both will generically refer to the Hilbert space of a d-dimensional quantum system throughout the presentation. The set of linear operators on $\mathcal{H}$ will be denoted by $L(\mathcal{H})$. The set of pure states is represented by the complex projective space $\mathbb{CP}^{d-1}$ and the set of all mixed states in $L(\mathcal{H})$, denoted by $\mathcal{D}(\mathcal{H})$, is given by the set of non-negative, trace-1 linear operators on $\mathcal{H}$. Unless otherwise stated, we will be concerned only with quantum operations with the same input and output spaces. The set of linear superoperators mapping $L(\mathcal{H})$ into itself is denoted by $\mathcal{T}(\mathcal{H})$ with the set of quantum channels (completely positive, trace-preserving linear maps) contained in $\mathcal{T}(\mathcal{H})$ denoted by $\mathcal{S}(\mathcal{H})$.

There are various methods for quantifying the distance between quantum operations, we briefly describe those that will be of use to us. Good references for many of the topics in this section are [35–37].

A. Diamond norm, average gate fidelity, and minimum gate fidelity

One method of quantifying the distance between two linear superoperators $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{T}(\mathcal{H})$ is given by the diamond norm distance $\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$. The diamond norm of an arbitrary linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$ is defined as

$$\|\mathcal{R}\|_\diamond = \sup_{k \in \mathbb{N}} \|\mathcal{R} \otimes \mathcal{I}_k\|_1, \qquad (2.1)$$

where $\| \; \|_1$ on superoperators is defined to be the ∞ norm induced by the trace norm $\| \; \|_1$ on $L(\mathbb{C}^m)$ and $L(\mathbb{C}^n)$. It is known that the supremum occurs for $k = m$ and so

$$\|\mathcal{R}\|_\diamond = \|\mathcal{R} \otimes \mathcal{I}_m\|_1 = \max_{A:\|A\|_1 \leqslant 1} \|\mathcal{R} \otimes \mathcal{I}_m(A)\|_1, \quad (2.2)$$

where $A \in L(\mathbb{C}^m \otimes \mathbb{C}^m)$. Hence for $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{T}(\mathcal{H})$,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|(\mathcal{E}_1 - \mathcal{E}_2) \otimes \mathcal{I}_d\|_1. \qquad (2.3)$$

The diamond norm distance is commonly used in quantum information due to its operational meaning of being related to the optimal probability for distinguishing $\mathcal{E}_1$ and $\mathcal{E}_2$ using a binary-outcome positive operator-valued measure (POVM) and single input state (allowing for ancillas) [38].

Another method for quantifying the distance between linear superoperators is given by the $\| \; \|_{1 \to 1}^H$ norm defined for the linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$ as

$$\|\mathcal{R}\|_{1 \to 1}^H = \max_{A:A=A^\dagger, \|A\|_1 \leqslant 1} \|\mathcal{R}(A)\|_1, \qquad (2.4)$$

where $A \in L(\mathbb{C}^m)$. One can see that $\| \; \|_{1 \to 1}^H$ is just $\| \; \|_1$ (which is also denoted $\| \; \|_{1 \to 1}$) restricted to Hermitian inputs. This norm is less common in quantum information due to its lack of operational meaning; however, it is a weaker measure of distance than the diamond norm since for any linear superoperator $\mathcal{R} : L(\mathbb{C}^m) \to L(\mathbb{C}^n)$, $\|\mathcal{R}\|_{1 \to 1}^H \leqslant \|\mathcal{R}\|_\diamond$. This will be of much use to us later when we consider neglecting higher-order effects in the benchmarking scheme.

A commonly used state-dependent measure for comparing quantum operations $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{S}(\mathcal{H})$ is given by the channel fidelity,

$$\mathcal{F}_{\mathcal{E}_1, \mathcal{E}_2}(\rho) = F(\mathcal{E}_1(\rho), \mathcal{E}_2(\rho))$$
$$= \left(\mathrm{tr} \sqrt{\sqrt{\mathcal{E}_1(\rho)} \mathcal{E}_2(\rho) \sqrt{\mathcal{E}_1(\rho)}} \right)^2 \qquad (2.5)$$

where "F" refers to the usual fidelity between quantum states [39]. In the case of a unitary operation $\mathcal{U}$ and quantum operation $\mathcal{E}$, and restricting input states to $\mathbb{CP}^{d-1}$, the channel fidelity is called the gate fidelity. Explicitly, for $\phi \leftrightarrow |\phi\rangle\langle\phi| \in \mathbb{CP}^{d-1}$,

$$\mathcal{F}_{\mathcal{E}, \mathcal{U}}(\phi) = \mathrm{tr}[\mathcal{U}(|\phi\rangle\langle\phi|)\mathcal{E}(|\phi\rangle\langle\phi|)], \qquad (2.6)$$

and defining $\Lambda = \mathcal{U}^\dagger \circ \mathcal{E}$ gives

$$\mathcal{F}_{\mathcal{E}, \mathcal{U}}(\phi) = \mathcal{F}_{\Lambda, \mathcal{I}}(\phi) = \mathrm{tr}[|\phi\rangle\langle\phi|\Lambda(|\phi\rangle\langle\phi|)]. \qquad (2.7)$$

The channel Λ can be thought of as representing how much $\mathcal{E}$ deviates from $\mathcal{U}$ in that if $\mathcal{E} = \mathcal{U}$ then $\Lambda = \mathcal{I}$. The gate fidelity has many nice mathematical properties, including a simple expression for the average over pure states, expressions for the variance in terms of various representations of Λ, and a concentration of measure phenomenon for large systems

[40–42]. The average gate fidelity is obtained by integrating $\mathcal{F}_{\mathcal{E},\mathcal{U}}$ over $\mathbb{CP}^{d-1}$ using the Fubini-Study measure μ_{FS} [43],

$$\overline{\mathcal{F}_{\mathcal{E},\mathcal{U}}} = \overline{\mathcal{F}_{\Lambda,\mathcal{I}}} = \int_{\mathbb{CP}^{d-1}} \mathrm{tr}[|\phi\rangle\langle\phi|\Lambda(|\phi\rangle\langle\phi|)]d\mu_{FS}(\phi). \tag{2.8}$$

Taking the minimum of $\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}$ over all mixed states ρ produces a quantity $\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}^{\min}$ commonly called the minimum channel fidelity,

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}^{\min} = \min_\rho \mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}(\rho).$$

Note that because of concavity of the fidelity, the minimum channel fidelity occurs at a pure state [39]. In the case of the gate fidelity, the minimum is called the minimum gate fidelity.

In certain cases we will be concerned with how close $\mathcal{E}_1$ and $\mathcal{E}_2$ are in terms of the difference between the average fidelity of each channel. To this end we define

$$\Delta F(\mathcal{E}_1,\mathcal{E}_2) := |\overline{\mathcal{F}_{\mathcal{E}_1,\mathcal{I}}} - \overline{\mathcal{F}_{\mathcal{E}_2,\mathcal{I}}}|. \tag{2.9}$$

Last, we note the following relationships between some of the distance measures defined above. First, for $\mathcal{E}_1, \mathcal{E}_2 \in \mathcal{S}(\mathcal{H})$ the following inequalities hold:

$$\Delta F(\mathcal{E}_1,\mathcal{E}_2) \leqslant \|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to1}^H \leqslant \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond, \tag{2.10}$$

where we recall the definition of $\|\ \|_{1\to1}^H$ in Eq. (2.4). The second inequality is clear since,

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to1}^H \leqslant \|\mathcal{E}_1 - \mathcal{E}_2\|_1 \leqslant \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond. \tag{2.11}$$

Now for the first inequality note that

$$\Delta F(\mathcal{E}_1,\mathcal{E}_2) \leqslant \max_{|\phi\rangle}|\mathrm{tr}[(\mathcal{E}_1 - \mathcal{E}_2)(|\phi\rangle\langle\phi|)|\phi\rangle\langle\phi|]|$$
$$\leqslant \max_{|\phi\rangle}\|(\mathcal{E}_1 - \mathcal{E}_2)(|\phi\rangle\langle\phi|)\|_\infty$$

$$= \max_{A:A=A^\dagger,\|A\|_1\leqslant1}\|(\mathcal{E}_1 - \mathcal{E}_2)(A)\|_\infty$$
$$= \|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to\infty}^H, \tag{2.12}$$

where we note that since $\mathcal{E}_1$ and $\mathcal{E}_2$ are completely positive, $\mathcal{E}_1 - \mathcal{E}_2$ is Hermiticity preserving. Hence since $\|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to\infty}^H \leqslant \|\mathcal{E}_1 - \mathcal{E}_2\|_{1\to1}^H$ the inequalities in Eq. (2.10) hold.

Next we show that for any quantum operations $\mathcal{E}_1,\mathcal{E}_2 \in \mathcal{S}(\mathcal{H})$,

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}^{\min} \geqslant 1 - \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond. \tag{2.13}$$

We have that

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \max_{|\psi\rangle\in\mathcal{H}\otimes\mathcal{H}}\|\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)$$
$$- \mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1. \tag{2.14}$$

By the Fuchs–Van de Graaf inequalities [44],

$$\|\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|) - \mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)\|_1$$
$$\geqslant 1 - F(\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|),\mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)), \tag{2.15}$$

so

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$$
$$\geqslant \max_{|\psi\rangle\in\mathcal{H}\otimes\mathcal{H}}[1 - F(\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|),\mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|))]$$
$$= 1 - \min_{|\psi\rangle\in\mathcal{H}\otimes\mathcal{H}}F(\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|),\mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|)). \tag{2.16}$$

Now we have

$$\min_{|\psi\rangle\in\mathcal{H}\otimes\mathcal{H}}F(\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|),\mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|))$$
$$\leqslant \min_{|\phi\rangle\in\mathcal{H}}F(\mathcal{E}_1(|\phi\rangle\langle\phi|),\mathcal{E}_2(|\phi\rangle\langle\phi|)) \tag{2.17}$$

since

$$\min_{|\psi\rangle\in\mathcal{H}\otimes\mathcal{H}}F(\mathcal{E}_1 \otimes \mathcal{I}(|\psi\rangle\langle\psi|),\mathcal{E}_2 \otimes \mathcal{I}(|\psi\rangle\langle\psi|))$$
$$\leqslant \min_{|\phi\rangle\in\mathcal{H}}F(\mathcal{E}_1 \otimes \mathcal{I}(|\phi\rangle\langle\phi| \otimes |\phi\rangle\langle\phi|),\mathcal{E}_2 \otimes \mathcal{I}(|\phi\rangle\langle\phi| \otimes |\phi\rangle\langle\phi|))$$
$$= \min_{|\phi\rangle\in\mathcal{H}}(\mathrm{tr}\sqrt{\sqrt{\mathcal{E}_1(|\phi\rangle\langle\phi|) \otimes |\phi\rangle\langle\phi|}[\mathcal{E}_2(|\phi\rangle\langle\phi|) \otimes |\phi\rangle\langle\phi|]\sqrt{\mathcal{E}_1(|\phi\rangle\langle\phi|) \otimes |\phi\rangle\langle\phi|}})^2$$
$$= \min_{|\phi\rangle\in\mathcal{H}}[\mathrm{tr}(\sqrt{\sqrt{\mathcal{E}_1(|\phi\rangle\langle\phi|)}[\mathcal{E}_2(|\phi\rangle\langle\phi|)]\sqrt{\mathcal{E}_1(|\phi\rangle\langle\phi|)}} \otimes |\phi\rangle\langle\phi|)]^2 = \min_{|\phi\rangle\in\mathcal{H}}F(\mathcal{E}_1(|\phi\rangle\langle\phi|),\mathcal{E}_2(|\phi\rangle\langle\phi|)). \tag{2.18}$$

So

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond \geqslant 1 - \min_{|\phi\rangle\in\mathcal{H}}F(\mathcal{E}_1(|\phi\rangle\langle\phi|),\mathcal{E}_2(|\phi\rangle\langle\phi|)). \tag{2.19}$$

Now by concavity,

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}^{\min} = \min_{|\phi\rangle\in\mathcal{H}}F(\mathcal{E}_1(|\phi\rangle\langle\phi|),\mathcal{E}_2(|\phi\rangle\langle\phi|)) \tag{2.20}$$

and so

$$\mathcal{F}_{\mathcal{E}_1,\mathcal{E}_2}^{\min} \geqslant 1 - \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond. \tag{2.21}$$

B. The Clifford group and t-designs

The Clifford group on n qubits, denoted Clif_n, is defined as the normalizer of the Pauli group $\mathcal{P}_n$ and is generated by the phase (S), Hadamard (H), and controlled-NOT (CNOT)

gates. Clif_n plays an important role in many areas of quantum information such as universality [45], stabilizer code theory and fault tolerance [46], and noise estimation [17].

One extremely useful property of Clif_n, especially for noise estimation, is that the uniform probability distribution over Clif_n comprises a unitary two-design [17]. A unitary t-design is defined as follows.

Definition 1. Unitary t-design.

A unitary t-design is a discrete random variable $\{(q_1,U_1),\dots,(q_K,U_K)\}$, with each $U_i \in U(d)$, such that for every homogeneous complex-valued polynomial p in $2d^2$ indeterminates of degree (s,s) less than or equal to (t,t),

$$\frac{1}{K}\sum_{j=1}^K p(U_j) = \int_{U(d)} p(U)dU. \tag{2.22}$$

The integral is taken with respect to the Haar measure on $U(d)$. Here $p(U)$ is defined to be the evaluation of p at the $2d^2$ values consisting of the d^2 matrix entries of U as well as the d^2 complex conjugates of these matrix entries. In the case $t = 2$ the above reduces to a "twirling" [47] condition

$$\sum_{j=1}^{K} q_j [U_j \Lambda (U_j^\dagger \rho U_j) U_j^\dagger] = \int_{U(d)} [U \Lambda (U^\dagger \rho U) U^\dagger] dU \tag{2.23}$$

being satisfied for any quantum channel Λ and any state ρ [17]. Since a uniform probability distribution on Clif_n forms a two-design, if $\mathrm{Clif}_n = \{C_j : j \in \mathcal{K} = \{1, \ldots, |\mathrm{Clif}_n|\}\}$, then

$$\mathcal{W}(\Lambda)(\rho) := \frac{1}{|\mathrm{Clif}_n|} \sum_{j=1}^{|\mathrm{Clif}_n|} [C_j \Lambda (C_j^\dagger \rho C_j) C_j^\dagger]$$

$$= \int_{U(d)} [U \Lambda (U^\dagger \rho U) U^\dagger] dU. \tag{2.24}$$

As shown in Refs. [18,40], $\int_{U(d)} [U \Lambda (U^\dagger \rho U) U^\dagger] dU$ produces the unique depolarizing channel Λ_d with the same average fidelity as Λ. Hence, if $\overline{\mathcal{F}_{\Lambda, \mathcal{I}}}$ is the average fidelity of Λ, and Λ_d is given by

$$\Lambda_d(\rho) = p\rho + (1 - p)\frac{\mathbb{1}}{d}, \tag{2.25}$$

then

$$\overline{\mathcal{F}_{\Lambda, \mathcal{I}}} = p + \frac{(1 - p)}{d}. \tag{2.26}$$

Thus twirling a quantum operation over the Clifford group produces a depolarizing channel, and the average fidelity is invariant under the twirling operation.

In Sec. III we will be concerned with compositions of both gate-independent and gate-dependent twirls. In the gate-independent case, the sequence of twirls of Λ of length k, $\mathcal{W}(\Lambda)^k$, can be rewritten as the k-fold composition of Λ_d with itself. Using the above representation of Λ_d we get

$$\mathcal{W}(\Lambda)^k(\rho) = p^k \rho + (1 - p^k)\frac{\mathbb{1}}{d}. \tag{2.27}$$

Therefore the average fidelity decreases exponentially to $\frac{1}{d}$ since

$$\overline{\mathcal{F}_{\Lambda_d^k, \mathcal{I}}} = p^k + \frac{(1 - p^k)}{d}. \tag{2.28}$$

We can also write the average fidelity of Λ in terms of its χ matrix [15]. The χ matrix is an important (basis-dependent) object in experimental quantum information as it is directly related to practical methods in process tomography. The χ matrix is obtained by expanding the Kraus operators $\{A_k\}$ of Λ with respect to a particular basis of $L(\mathbb{C}^d)$, which is most often chosen to be the Pauli basis $\{P_j\}_{j=0}^{d^2-1}$ ($P_0 = \mathbb{1}$). This gives

$$\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger = \sum_{i,j} \chi_{i,j} P_i \rho P_j, \tag{2.29}$$

and so a complete description for Λ can be given by estimating the entries of χ. As shown in Ref. [15],

$$\overline{\mathcal{F}_{\Lambda, \mathcal{I}}} = \frac{\chi_{0,0} d + 1}{d + 1}, \tag{2.30}$$

which gives

$$\chi_{0,0} = p\left(1 - \frac{1}{d^2}\right) + \frac{1}{d^2} = \frac{\overline{\mathcal{F}_{\Lambda, \mathcal{I}}}(d + 1) - 1}{d}. \tag{2.31}$$

Therefore the $(0,0)$ entry of the χ matrix for a quantum operation with respect to the Pauli basis is invariant under twirling over a two-design. Moreover, $\chi_{0,0}$ for Λ_d^k decreases to $\frac{1}{d^2}$ exponentially in k.

III. RANDOMIZED BENCHMARKING

In this section we present both the protocol and a full derivation of the fitting models for randomized benchmarking that were given in Ref. [25]. First, we set some notation and make various definitions that will be used throughout the presentation.

Denote the elements of Clif_n by C_i and the maximum sequence length of applying Clifford gates by M. Suppose that the actual implementation of C_i at time j ($1 \leqslant j \leqslant M$) results in the map $\mathcal{E}_{i,j}$ with $\mathcal{E}_{i,j} = \Lambda_{i,j} \circ C_i$ for some error map $\Lambda_{i,j}$. Hence with each Clifford C_i we associate a sequence $\Lambda_{i,1}, \ldots, \Lambda_{i,M}$ which represents the time-dependent noise operators affecting C_i. We define the average error operator as follows:

Definition 2. Average error operator.

The average error operator affecting the gates in Clif_n is given by

$$\Lambda = \frac{1}{M|\mathrm{Clif}_n|} \sum_j \sum_i \Lambda_{i,j}. \tag{3.1}$$

Consider the twirl of the average error operator over Clif_n. As discussed in Sec. II B, this produces a depolarized channel Λ_d,

$$\Lambda_d(\rho) = \frac{1}{|\mathrm{Clif}_n|} \sum_i C_i^\dagger \circ \Lambda_{\mathrm{ave}} \circ C_i \ (\rho) = p\rho + (1 - p)\frac{\mathbb{1}}{d}. \tag{3.2}$$

Recall from Sec. II B that the average fidelity of Λ, denoted F_{ave}, is invariant under Clifford twirling and so

$$F_{\mathrm{ave}} = p + \frac{1 - p}{d}. \tag{3.3}$$

We now define the average error rate of the set of Clifford gates as follows:

Definition 3. Average error rate.

The average error rate r of the Clifford gates used in a quantum computation is defined to be

$$r = 1 - F_{\mathrm{ave}} = 1 - \left(p + \frac{1 - p}{d}\right) = \frac{(d - 1)(1 - p)}{d}. \tag{3.4}$$

It is important to note that r defined above should not be confused with the "error rate" $r_\mathcal{P}$ of a Pauli channel $\mathcal{P}$. For Pauli channel $\mathcal{P}$, $r_\mathcal{P}$ is defined to be the probability that a nonidentity Pauli operator is applied to the input state. Conditioning on a nonidentity Pauli being applied, there is still a nonzero probability of the input state being unchanged. Subtracting this probability out gives our defined parameter r for $\mathcal{P}$, which is commonly called the "infidelity" of $\mathcal{P}$. One can show that r and $r_\mathcal{P}$ are related via $r_\mathcal{P} = \frac{(d+1)r}{d}$. Following the terminology set in Ref. [26], we will call r the (average) error rate of Λ and note that, in the case where Λ is a Pauli channel, r is equal to the infidelity of Λ.

The parameter r is the figure of merit we want to be able to estimate experimentally. One can estimate p directly using any of standard process tomography [15], ancilla-assisted or entanglement-assisted process tomography [48], or Monte Carlo methods [23,24]. The tomography-based schemes suffer from the unrealistic assumptions of negligible state-preparation and measurement errors, and clean ancillary states and operations. These schemes also require exponential time resources in n, making them infeasible for even relatively small numbers of qubits. The Monte Carlo methods also have the drawback of assuming negligible state-preparation and measurement errors. The advantages of these methods are that the average fidelity of each gate can be estimated and the scheme is efficient in n.

The experimentally relevant challenge therefore is to estimate p while relaxing the assumptions on state preparation, measurement, and ancillary states and processes. Ideally, such a method should also scale efficiently with the number of qubits. As we show below, such an estimate can be obtained through benchmarking the performance of random circuits.

A. Protocol

For a fixed sequence length $m \leqslant M - 1$, the benchmarking protocol consists of choosing K_m sequences of independent and identically distributed uniformly random Clifford elements and calculating the fidelity of the average of the K_m sequences. One repeats this procedure for different values of m and fits the fidelity decay curve to the models we derive below. More precisely, the protocol is as follows.

Fix an initial state $|\psi\rangle$ and perform the following steps:

Step 1. Fix $m \leqslant M - 1$ and generate K_m sequences consisting of $m + 1$ quantum operations. The first m operations are chosen uniformly at random from Clif_n and the $(m + 1)$th operation is uniquely determined as the inverse gate of the composition of the first m. By assumption, each operation $\mathcal{C}_{i_j}$ is allowed to have some error, represented by $\Lambda_{i_j,j}$, and each sequence can be modeled by the operation

$$\mathcal{S}_{\mathbf{i_m}} = \bigcirc_{j=1}^{m+1}\left(\Lambda_{i_j,j} \circ \mathcal{C}_{i_j}\right), \tag{3.5}$$

where $\mathbf{i_m}$ is the m-tuple $(i_1, \ldots, i_m)$ (which we sometimes also denote by $\vec{i_m}$) and i_{m+1} is uniquely determined by $\mathbf{i_m}$.

Step 2. For each of the K_m sequences, measure the survival probability $\text{Tr}[E_\psi \mathcal{S}_{\mathbf{i_m}}(\rho_\psi)]$. Here ρ_ψ is a quantum state that takes into account errors in preparing $|\psi\rangle\langle\psi|$ and E_ψ is the POVM element that takes into account measurement errors. In the ideal (noise-free) case $\rho_\psi = E_\psi = |\psi\rangle\langle\psi|$.

Step 3. Average over the K_m random realizations to find the averaged sequence fidelity,

$$F_{\text{seq}}(m, \psi) = \text{Tr}[E_\psi \mathcal{S}_{K_m}(\rho_\psi)], \tag{3.6}$$

where

$$\mathcal{S}_{K_m} = \frac{1}{K_m} \sum_{\mathbf{i_m}} \mathcal{S}_{\mathbf{i_m}} \tag{3.7}$$

is the average sequence operation.

Step 4. Repeat steps 1 through 3 for different values of m and fit the results for the averaged sequence fidelity [defined in Eq. (3.6)] to the model

$$\mathcal{F}_g^{(1)}(m, |\psi\rangle) = A_1 p^m + B_1 + C_1(m - 1)(q - p^2)p^{m-2} \tag{3.8}$$

derived below. The coefficients A_1, B_1, and C_1 absorb the state-preparation and measurement errors as well as the error on the final gate. The difference $q - p^2$ is a measure of the degree of gate dependence in the errors, and p determines the average error rate r according to the relation given by Eq. (3.4). In the case of gate-independent and time-independent errors the results will fit the simpler model

$$\mathcal{F}_g^{(0)}(m, |\psi\rangle) = A_0 p^m + B_0 \tag{3.9}$$

also derived below, where A_0 and B_0 absorb state-preparation and measurement errors as well as the error on the final gate.

We note that for each m, in the limit of $K_m \to \infty$, $F_{\text{seq}}(m, \psi)$ converges to the exact (uniform) average $\mathcal{F}_g(m, \psi)$, over all sequences,

$$\mathcal{F}_g(m, \psi) = \text{Tr}[E_\psi \mathcal{S}_m(\rho_\psi)], \tag{3.10}$$

where we define the exact average of the sequences to be

$$\mathcal{S}_m = \frac{1}{|\text{Clif}_n|^m} \sum_{(i_1,\ldots,i_m)} \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}. \tag{3.11}$$

Hence the fitting functions by which we model the behavior of $F_{\text{seq}}(m, \psi)$ are derived in terms of $\mathcal{F}_g(m, \psi)$ (see Sec. III B). Note that, since $\mathcal{F}_g(m, \psi)$ is the uniform average over all sequences, we can sum over each index independently,

$$\mathcal{F}_g(m, \psi) = \frac{1}{|\text{Clif}_n|^m} \sum_{i_1,\ldots,i_m} \text{tr}\big[\Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m}$$
$$\circ \mathcal{C}_{i_m} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}(\rho_\psi) E_\psi\big]. \tag{3.12}$$

In order to prepare for the next section where we derive the above fitting models, we write $\mathcal{F}_g(m, \psi)$ in a more intuitive form. We first rewrite $\Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}$ by inductively defining new uniformly random gates from the Clifford group in the following manner:

(1) Define $\mathcal{D}_{i_1} = \mathcal{C}_{i_1}$.

(2) Define $\mathcal{D}_{i_2}$ uniquely by the equation $\mathcal{C}_{i_2} = \mathcal{D}_{i_2} \circ \mathcal{D}_{i_1}^\dagger$, i.e., $\mathcal{D}_{i_2} = \mathcal{C}_{i_2} \circ \mathcal{C}_{i_1} = \bigcirc_{s=1}^{2}\mathcal{C}_{i_s}$.

(3) In general, for $j \in \{2, \ldots, m\}$, if $\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_j}$ and $\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_j}$ have been chosen, define $\mathcal{D}_{i_{j+1}}$ uniquely by the

equation $\mathcal{C}_{i_{j+1}} = \mathcal{D}_{i_{j+1}} \circ \mathcal{D}_{i_j}^\dagger$, i.e.,

$$\mathcal{D}_{i_{j+1}} = \mathcal{C}_{i_{j+1}} \circ \cdots \circ \mathcal{C}_{i_1} = \bigcirc_{s=1}^{j+1} \mathcal{C}_{i_s}. \qquad (3.13)$$

Note that if $j \neq k$, $\mathcal{C}_{i_j}$ and $\mathcal{C}_{i_k}$ are independent and so since the Clifford elements form a group, for each $j = 2, \ldots, m+1$, $\mathcal{D}_{i_j}$ is independent of $\mathcal{D}_{i_{j-1}}$. As well, summing over each i_j index runs over every Clifford element once and only once in $\mathcal{D}_{i_j}$.

We have created a new sequence $(\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_m})$ from $(\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_m})$ uniquely so that

$$\begin{aligned}
\mathcal{S}_{\vec{i_m}} &= \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1} \\
&= \Lambda_{i_{m+1},m+1} \circ \mathcal{D}_{i_{m+1}} \circ \mathcal{D}_{i_m}^\dagger \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \circ \cdots \\
&\quad \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1}.
\end{aligned} \qquad (3.14)$$

Since $\mathcal{C}_{i_{m+1}} = \mathcal{C}_{i_1}^\dagger \circ \cdots \circ \mathcal{C}_{i_m}^\dagger$ and $\mathcal{D}_{i_{m+1}} = \mathcal{C}_{i_{m+1}} \circ \cdots \circ \mathcal{C}_{i_1}$,

$$\mathcal{D}_{i_{m+1}} = \mathbb{1}. \qquad (3.15)$$

Hence the $(m+1)$th gate is decoupled from the rest of the sequence and we have

$$\begin{aligned}
\mathcal{S}_{\vec{i_m}} &= \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \Lambda_{i_m,m} \circ \mathcal{C}_{i_m} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1} \\
&= \Lambda_{i_{m+1},m+1} \circ \mathcal{D}_{i_m}^\dagger \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \circ \cdots \\
&\quad \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1}.
\end{aligned} \qquad (3.16)$$

B. Perturbative expansion and the fitting models

We would like to develop fitting models for $\mathcal{F}_g(m,\psi)$ where the most general noise model allows for the noise to depend upon both the set of gates in Clif_n and time. We can estimate the behavior of $\mathcal{F}_g(m,\psi)$ by considering a perturbative expansion of each $\Lambda_{i,j}$ about the average Λ. We quantify the difference between $\Lambda_{i,j}$ and Λ by defining for all i,j,

$$\delta\Lambda_{i,j} = \Lambda_{i,j} - \Lambda. \qquad (3.17)$$

Our approach will be valid provided $\delta\Lambda_{i,j}$ is a small perturbation from Λ in a sense to be made precise later. Note that each $\delta\Lambda_{i,j}$ is a Hermiticity-preserving, trace-annihilating linear superoperator. Under the above conditions this approach will allow for fitting the experimental fidelity decay sequence to a model with fit parameters that determine not only the average error per gate but also the separate contribution from the combined effects of state-preparation and measurement errors. In the limit of multiple qubits and very precise control, weaker forms of twirling may permit even more detailed modeling of the noise.

Using the change of variables $\mathcal{D}_{i_j} = \bigcirc_{s=1}^j \mathcal{C}_{i_s}$ described above and expanding to first order we get

$$\begin{aligned}
\mathcal{S}_{\vec{i_m}} &\equiv \Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \cdots \circ \Lambda_{i_j,j} \circ \mathcal{C}_{i_j} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1} = \Lambda_{i_{m+1},m+1} \circ \mathcal{D}_{i_m}^\dagger \circ \Lambda_{i_m,m} \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1} \\
&= \Lambda \circ \mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} + \delta\Lambda_{i_{m+1},m+1} \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + \cdots + \Lambda \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_j}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{D}_{i_j}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + \cdots + \Lambda \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \delta\Lambda_{i_1,1} \circ \mathcal{D}_{i_1}\right) + O\left(\delta\Lambda_{i_j,j}^2\right).
\end{aligned} \qquad (3.18)$$

We define

$$\mathcal{S}_{\vec{i_m}}^{(0)} := \Lambda \circ \mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}, \qquad (3.19)$$

$$\begin{aligned}
\left(\mathcal{S}_{\vec{i_m}}^{(1)}\right) &:= \delta\Lambda_{i_{m+1},m+1} \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + \cdots + \Lambda \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_j}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{D}_{i_j}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}\right) \\
&\quad + \cdots + \Lambda \circ \left(\mathcal{D}_{i_m}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}\right) \circ \cdots \circ \left(\mathcal{D}_{i_1}^\dagger \circ \delta\Lambda_{i_1,1} \circ \mathcal{D}_{i_1}\right),
\end{aligned} \qquad (3.20)$$

and so on for higher-order perturbation terms. As well, recalling the definition of $\mathcal{S}_m$ in Eq. (3.11), we define for each order k,

$$\mathcal{S}_m^{(k)} := \frac{1}{|\mathrm{Clif}_n|^m} \sum_{i_1,\ldots,i_m} \mathcal{S}_{\vec{i_m}}^{(k)} \qquad (3.21)$$

and

$$\mathcal{F}_g^{(k)}(m,\psi) := \mathrm{tr}\left[\left(\sum_{j=0}^k \mathcal{S}_m^{(j)}\right)(\rho_\psi) E_\psi\right], \qquad (3.22)$$

so that

$$\mathcal{S}_m = \sum_{k=0}^{m+1} \mathcal{S}_m^{(k)} \qquad (3.23)$$

and

$$\mathcal{F}_g(m,\psi) = \mathcal{F}_g^{(m+1)}(m,|\psi\rangle) = \mathrm{tr}\left[\left(\sum_{j=0}^{m+1} \mathcal{S}_m^{(j)}\right)(\rho_\psi) E_\psi\right]. \qquad (3.24)$$

1. Zeroth-order model

First, we look at the zeroth-order fitting model $\mathcal{F}_g^{(0)}(m,|\psi\rangle)$ and note that $\mathcal{F}_g^{(0)}(m,|\psi\rangle)$ is exact in the case that the noise is independent of both the gate chosen and time, i.e., $\Lambda_{i_j,j} = \Lambda$. Because of the independence of the $\mathcal{D}_{i_j}$ and the fact that averaging over the ensemble of realizations produces independent twirls which depolarize m factors of Λ (see Sec. II B), we get

$$S_m^{(0)} = \Lambda \circ \Lambda_d \circ \cdots \circ \Lambda_d = \Lambda \circ \left(\bigcirc_{j=1}^m \Lambda_d \right). \quad (3.25)$$

Thus,

$$\mathcal{F}_g^{(0)}(m,|\psi\rangle) = \mathrm{tr}[S_m^{(0)}(\rho_\psi)E_\psi] = \mathrm{tr}[\Lambda(\rho_\psi)E_\psi]p^m$$
$$+ \mathrm{tr}\left[\Lambda\left(\frac{\mathbb{1}}{d}\right)E_\psi\right](1-p^m) = A_0 p^m + B_0,$$
$$(3.26)$$

where

$$A_0 := \mathrm{Tr}\left[E_\psi \Lambda\left(\rho_\psi - \frac{\mathbb{1}}{d} \right) \right] \quad (3.27)$$

and

$$B_0 := \mathrm{Tr}\left[E_\psi \Lambda\left(\frac{\mathbb{1}}{d} \right) \right]. \quad (3.28)$$

Hence, assuming the simplest (ideal) scenario where the noise operator at each step is independent of the applied gate (and is also time invariant), $\mathcal{F}_g(m,\psi) = \mathcal{F}_g^{(0)}(m,|\psi\rangle)$ decays exponentially in p.

2. First-order model

To find $\mathcal{F}_g^{(1)}(m,|\psi\rangle)$ we note that in the definition of $S_{i_m}^{(1)}$ given by Eq. (3.20) there are $\binom{m+1}{1} = m+1$ first-order perturbation terms which contain the gate dependence. First, we consider the $m-1$ terms with $j \in \{2,\ldots,m\}$. For each such j, averaging over the $\{i_1,\ldots,i_m\}$ gives a term of the form

$$\frac{1}{|\mathrm{Clif}_n|^m} \sum_{i_1,\ldots,\,i_m} \Lambda \circ \left(\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \right) \circ \cdots$$

$$\circ \left(\mathcal{D}_{i_j}{}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{D}_{i_j} \right) \circ \left(\mathcal{D}_{i_{j-1}}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_{j-1}} \right) \circ \cdots$$

$$\circ \left(\mathcal{D}_{i_1}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \right). \quad (3.29)$$

For these $m-1$ terms the main trick is to realize that we can reexpand $\mathcal{D}_{i_j} = \mathcal{C}_{i_j} \circ \mathcal{D}_{i_{j-1}}$ in order to depolarize the unitarily rotated perturbation $\mathcal{C}_{i_j}^\dagger \Lambda_{i_j,j} \mathcal{C}_{i_j}$ with the twirling operation $\frac{1}{|\mathrm{Clif}_n|} \sum_{i_{j-1}} \mathcal{D}_{i_{j-1}}^\dagger \cdot \mathcal{D}_{i_{j-1}}$ because the sums are independent. More precisely, the above can be written as

$$\Lambda \circ \Lambda_d^{m-j} \circ \left[\frac{1}{|\mathrm{Clif}_n|^2} \sum_{i_{j-1},i_j} \mathcal{D}_{i_{j-1}}{}^\dagger \circ \mathcal{C}_{i_j}^\dagger \circ \delta\Lambda_{i_j,j} \circ \mathcal{C}_{i_j} \circ \Lambda \circ \mathcal{D}_{i_{j-1}} \right] \circ \left[\sum_{i_{j-2},\ldots,i_1} \left(\mathcal{D}_{i_{j-2}}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_{j-2}} \right) \circ \cdots \circ \left(\mathcal{D}_{i_1}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \right) \right]$$

$$= \Lambda \circ \Lambda_d^{m-j} \circ \left[(\mathcal{Q}_j \circ \Lambda)_d - \Lambda_d^2 \right] \circ \Lambda_d^{j-2}, \quad (3.30)$$

where $\mathcal{Q}_j := \frac{1}{|\mathrm{Clif}_n|} \sum_i \mathcal{C}_i^\dagger \circ \Lambda_{i,j} \circ \mathcal{C}_i$ and the subscript "d" represents the depolarization of the operator within brackets. Using the fact that depolarizing channels commute, we get

$$\Lambda \circ \Lambda_d^{m-j} \circ \left[(\mathcal{Q}_j \circ \Lambda)_d - \Lambda_d^2 \right] \circ \Lambda_d^{j-2} = \Lambda \circ \left[(\mathcal{Q}_j \circ \Lambda)_d - \Lambda_d^2 \right] \circ \Lambda_d^{m-2}. \quad (3.31)$$

For the term with $j=1$, averaging over $i_1,\ldots,i_m$ gives a term of the form

$$\Lambda \circ \Lambda_d^{m-1} \circ \frac{1}{|\mathrm{Clif}_n|} \sum_{i_1} \mathcal{D}_{i_1}{}^\dagger \circ \delta\Lambda_{i_1,1} \circ \mathcal{D}_{i_1} = \Lambda \circ \Lambda_d^{m-1} \circ (\mathcal{Q}_1 - \Lambda_d), \quad (3.32)$$

where

$$\mathcal{Q}_1 := \frac{1}{|\mathrm{Clif}_n|} \sum_{i_1} \left(\mathcal{D}_{i_1}^\dagger \circ \Lambda_{i_1,1} \circ \mathcal{D}_{i_1} \right) = \frac{1}{|\mathrm{Clif}_n|} \sum_i (\mathcal{C}_i^\dagger \circ \Lambda_{i,1} \circ \mathcal{C}_i). \quad (3.33)$$

Last, for the term with $j = m+1$, averaging gives

$$\frac{1}{|\mathrm{Clif}_n|^m} \sum_{i_1,\ldots,\,i_m} \delta\Lambda_{i_{m+1},m+1} \circ \left(\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \right) \circ \cdots \circ \left(\mathcal{D}_{i_1}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \right)$$

$$= \frac{1}{|\mathrm{Clif}_n|^{m-1}} \sum_{i_1,\ldots,\,i_{m-1}} \left(\frac{1}{|\mathrm{Clif}_n|} \sum_{i_m} \delta\Lambda_{i_{m+1},m+1} \circ \left(\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m} \right) \right) \circ \cdots \circ \left(\mathcal{D}_{i_1}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1} \right). \quad (3.34)$$

Since Clif_n is a group, if $i_1, \ldots, i_{m-1}$ is fixed, averaging over the i_m index runs through every Clifford element with equal frequency in the $\mathcal{D}_{i_m}$ random variable. Since $\Lambda_{i_{m+1},m+1}$ is just the error associated with the gate $\mathcal{D}_{i_m}^\dagger$, $\frac{1}{|\mathrm{Clif}_n|} \sum_{i_m} \delta\Lambda_{i_{m+1},m+1} \circ (\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m})$ is independent of the $i_1, \ldots, i_{m-1}$ indices. Hence we can define

$$\mathcal{R}_{m+1} := \frac{1}{|\mathrm{Clif}_n|} \sum_{i_m} \Lambda_{i_{m+1},m+1} \circ (\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m})$$
$$= \frac{1}{|\mathrm{Clif}_n|} \sum_i \Lambda_{i',m+1} \circ (\mathcal{C}_i^\dagger \circ \Lambda \circ \mathcal{C}_i) \quad (3.35)$$

where $\Lambda_{i',m+1}$ denotes the error that arises when the Clifford operation $\mathcal{C}_i^\dagger$ is applied at final time step $m+1$. Again, using the group property of Clif_n, we have

$$\mathcal{R}_{m+1} = \frac{1}{|\mathrm{Clif}_n|} \sum_i \Lambda_{i,m+1} \circ (\mathcal{C}_i \circ \Lambda \circ \mathcal{C}_i^\dagger). \quad (3.36)$$

This decoupling of $\mathcal{R}_{m+1}$ allows us to write

$$\frac{1}{|\mathrm{Clif}_n|^{m-1}}$$
$$\times \sum_{i_1,\ldots,i_{m-1}} \left(\frac{1}{|\mathrm{Clif}_n|} \sum_{i_m} \delta\Lambda_{i_{m+1},m+1} \circ (\mathcal{D}_{i_m}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_m}) \right)$$
$$\circ \cdots \circ (\mathcal{D}_{i_1}{}^\dagger \circ \Lambda \circ \mathcal{D}_{i_1}) = (\mathcal{R}_{m+1} - \Lambda \circ \Lambda_d) \circ \Lambda_d^{m-1}. \quad (3.37)$$

Hence combining Eqs. (3.25), (3.31), (3.32), and (3.37) gives

$$S_m^{(0)} + S_m^{(1)} = \Lambda \circ \Lambda_d^m + (\mathcal{R}_{m+1} - \Lambda \circ \Lambda_d) \circ \Lambda_d^{m-1}$$
$$+ \sum_{j=2}^m \Lambda \circ [(\mathcal{Q}_j \circ \Lambda)_d - \Lambda_d^2] \circ \Lambda_d^{m-2}$$
$$+ \Lambda \circ \Lambda_d^{m-1} \circ (\mathcal{Q}_1 - \Lambda_d)$$
$$= \mathcal{R}_{m+1} \circ \Lambda_d^{m-1} + \sum_{j=2}^m [\Lambda \circ (\mathcal{Q}_j \circ \Lambda)_d \circ \Lambda_d^{m-2}]$$
$$+ \Lambda \circ \Lambda_d^{m-1} \circ \mathcal{Q}_1 - m(\Lambda \circ \Lambda_d^m). \quad (3.38)$$

To calculate $\mathcal{F}_g^{(1)}(m,|\psi\rangle) := \mathrm{tr}[(S_m^{(0)} + S_m^{(1)})(\rho_\psi)E_\psi]$ we have

$$\mathrm{tr}[\mathcal{R}_{m+1} \circ \Lambda_d^{m-1}(\rho_\psi)E_\psi] = G_{1,m+1}p^{m-1} + H_{1,m+1}, \quad (3.39)$$

$$\mathrm{tr}[\Lambda \circ (\mathcal{Q}_j \circ \Lambda)_d \circ \Lambda_d^{m-2}(\rho_\psi)E_\psi] = A_0 q_j p^{m-2} + B_0, \quad (3.40)$$

$$\mathrm{tr}[\Lambda \circ \Lambda_d^{m-1} \circ \mathcal{Q}_1(\rho_\psi)E_\psi] = A_{1,1}p^{m-1} + B_0, \quad (3.41)$$

$$\mathrm{tr}[\Lambda \circ \Lambda_d^m(\rho_\psi)E_\psi] = A_0 p^m + B_0, \quad (3.42)$$

where $G_{1,m+1} := \mathrm{tr}[\mathcal{R}_{m+1}(\rho_\psi - \frac{1}{d})E_\psi]$, $H_{1,m+1} := \mathrm{tr}[\mathcal{R}_{m+1}(\frac{1}{d})E_\psi]$, $A_{1,1} := \mathrm{tr}\{\Lambda[\mathcal{Q}_1(\rho_\psi) - \frac{1}{d}]E_\psi\}$, A_0 and B_0 are as given in Eqs. (3.27) and (3.28), and q_j is the

depolarization parameter for $(\mathcal{Q}_j \circ \Lambda)_d$. Thus

$$\mathcal{F}_g^{(1)}(m,|\psi\rangle) = G_{1,m+1}p^{m-1} + H_{1,m+1} + \sum_{j=2}^m (A_0 q_j p^{m-2} + B_0)$$
$$+ A_{1,1}p^{m-1} + B_0 - m(A_0 p^m + B_0)$$
$$= p^{m-1}(G_{1,m+1} + A_{1,1} - A_0 p)$$
$$+ (m-1)A_0 p^{m-2}\left(\frac{\sum_{j=2}^m q_j}{m-1} - p^2 \right) + H_{1,m+1}. \quad (3.43)$$

Finally, we can also rewrite Eq. (3.43) as

$$\mathcal{F}_g^{(1)}(m,|\psi\rangle) = A_1(m)p^m + B_1(m) + C_1(m-1)$$
$$\times [q(m) - p^2]p^{m-2}, \quad (3.44)$$

where

$$A_1(m) = \mathrm{Tr}\left[E_\psi \Lambda\left(\frac{\mathcal{Q}_1(\rho_\psi)}{p} - \rho_\psi + \frac{(p-1)\mathbb{1}}{pd} \right) \right]$$
$$+ \mathrm{Tr}\left[E_\psi \mathcal{R}_{m+1}\left(\frac{\rho_\psi}{p} - \frac{\mathbb{1}}{pd} \right) \right],$$
$$B_1(m) = \mathrm{Tr}\left[E_\psi \mathcal{R}_{m+1}\left(\frac{\mathbb{1}}{d} \right) \right], \quad C_1 = \mathrm{Tr}\left[E_\psi \Lambda\left(\rho_\psi - \frac{\mathbb{1}}{d} \right) \right],$$
$$q(m) = \sum_{j=2}^m q_j/(m-1), \quad (3.45)$$

and q_j is the depolarizing parameter defined by

$$(\mathcal{Q}_j \circ \Lambda)_d(\rho) = q_j\rho + (1-q_j)\frac{\mathbb{1}}{d}. \quad (3.46)$$

We write the first-order model in the form of Eq. (3.44) because of its similarity to that of the zeroth-order model given by Eq. (3.26). The difference between Eqs. (3.44) and (3.26) is the $C_1(m-1)[q(m) - p^2]p^{m-2}$ term contained in Eq. (3.44), which can be thought of as a measure of the gate dependence of the noise.

Again, we see that the edge effects and state-preparation and measurement errors are embedded in the three coefficients $A_1(m)$, $B_1(m)$, and C_1. Note that the m dependence in $q(m)$ and the $A_1(m)$, and $B_1(m)$ coefficients due to the last gate disappears if the errors do not change as a function of time.

IV. NEGLECT OF HIGHER ORDERS

A. Bounding higher-order perturbation terms

We would like to give conditions showing when one is justified in stopping the expansion at some order k. The main idea, as expressed in Eq. (4.1) below, is to bound the "size" of the terms in $S_m^{(k+1)}$ and we use the "$1 \to 1$" norm on linear superoperators maximized over Hermitian inputs, denoted $\| \|_{1\to1}^H$, to make this precise (see Sec. II). Note that $\| \|_{1\to1}^H$ has the following useful properties:

(a) submultiplicativity for Hermiticity-preserving superoperators,

(b) unitary invariance, and

(c)$\|\mathcal{E}\|^H_{1\to1} \leqslant 1$ for any quantum operation $\mathcal{E}$.

Later we will discuss the motivation for using $\|\ \|^H_{1\to1}$ as opposed to more familiar norms used in quantum-information theory such as the diamond norm $\|\ \|_\diamond$.

From Sec. II A we have that

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| = \left|\text{tr}\left[\left(\sum_{j=0}^{k+1} S_m^{(j)}\right)(\rho_\psi)E_\psi\right] - \text{tr}\left[\left(\sum_{j=0}^{k} S_m^{(j)}\right)(\rho_\psi)E_\psi\right]\right| = \left|\text{tr}\left[S_m^{(k+1)}(\rho_\psi)E_\psi\right]\right| \leqslant \left\|S_m^{(k+1)}\right\|^H_{1\to1}$$

(4.1)

and so bounding $S_m^{(k+1)}$ provides a bound for how much the k- and $(k+1)$-order fidelities will differ. We first look at the case of stopping at first order, i.e., $k=1$. There are $\binom{m+1}{2} = \frac{(m+1)m}{2}$ second-order perturbation terms in Eq. (3.18). Let us look at at a term with perturbations at j_1 and j_2 where without loss of generality we assume $j_2 > j_1$. Using the properties listed above, along with the triangle inequality, we have

$$\left\|\frac{1}{|\text{Clif}_n|^m} \sum_{\vec{i}_m} \Lambda \circ \mathcal{D}^\dagger_{i_m} \circ \Lambda \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_2}} \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}} \circ \cdots \circ \mathcal{D}^\dagger_{i_1} \circ \Lambda \circ \mathcal{D}_{i_1}\right\|^H_{1\to1}$$

$$\leqslant \frac{1}{|\text{Clif}_n|^m} \sum_{\vec{i}_m} \left\|\Lambda\right\|^H_{1\to1} \left\|\mathcal{D}^\dagger_{i_m} \circ \Lambda \circ \mathcal{D}_{i_m}\right\|^H_{1\to1} \cdots \left\|\mathcal{D}^\dagger_{i_{j_2}} \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|^H_{1\to1} \cdots \left\|\mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|_{1\to1} \cdots \left\|\mathcal{D}^\dagger_{i_1} \circ \Lambda \circ \mathcal{D}_{i_1}\right\|^H_{1\to1}$$

$$= \left(\left\|\Lambda\right\|^H_{1\to1}\right)^{m-1} \frac{1}{|\text{Clif}_n|} \sum_{i_{j_2}} \left\|\mathcal{D}^\dagger_{i_{j_2}} \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|^H_{1\to1} \frac{1}{|\text{Clif}_n|} \sum_{i_{j_1}} \left\|\mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|^H_{1\to1}$$

$$\leqslant \frac{1}{|\text{Clif}_n|} \sum_{i_{j_2}} \left\|\mathcal{D}^\dagger_{i_{j_2}} \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}}\right\|^H_{1\to1} \frac{1}{|\text{Clif}_n|} \sum_{i_{j_1}} \left\|\mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}}\right\|^H_{1\to1} = \gamma_{j_2}\gamma_{j_1},$$

(4.2)

where we define the time-dependent variation in the noise,

$$\gamma_j := \frac{1}{|\text{Clif}_n|} \sum_i \left\|\Lambda_{i,j} - \Lambda\right\|^H_{1\to1}.$$

(4.3)

Summing over all j_1, j_2 with $j_2 > j_1$ gives

$$\left\|S_m^{(2)}\right\|^H_{1\to1} = \left\|\frac{1}{|\text{Clif}_n|^m} \sum_{\vec{i}_m} S_{\vec{i}_m}^{(2)}\right\|^H_{1\to1}$$

$$= \left\|\frac{1}{|\text{Clif}_n|^m} \sum_{\vec{i}_m} \sum_{j_2 > j_1} \Lambda \circ \mathcal{D}^\dagger_{i_m} \circ \delta\Lambda \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_2}} \circ \delta\Lambda_{i_{j_2}}\right.$$

$$\left. \circ \mathcal{D}_{i_{j_2}} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}} \circ \cdots \circ \mathcal{D}^\dagger_{i_1} \circ \Lambda \circ \mathcal{D}_{i_1}\right\|^H_{1\to1}$$

$$\leqslant \sum_{j_2 > j_1} \left\|\frac{1}{|\text{Clif}_n|^m} \sum_{\vec{i}_m} \Lambda \circ \mathcal{D}^\dagger_{i_m} \circ \delta\Lambda \circ \mathcal{D}_{i_m} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_2}}\right.$$

$$\left. \circ \delta\Lambda_{i_{j_2}} \circ \mathcal{D}_{i_{j_2}} \circ \cdots \circ \mathcal{D}^\dagger_{i_{j_1}} \circ \delta\Lambda_{i_{j_1}} \circ \mathcal{D}_{i_{j_1}} \circ \cdots \circ \mathcal{D}^\dagger_{i_1} \circ \Lambda \circ \mathcal{D}_{i_1}\right\|^H_{1\to1} \leqslant \sum_{j_2 > j_1} \gamma_{j_2}\gamma_{j_1}.$$

(4.4)

In terms of the fidelity we thus have from Eqs. (4.1) and (4.4),

$$\left|\mathcal{F}_g^{(2)}(m,|\psi\rangle) - \mathcal{F}_g^{(1)}(m,|\psi\rangle)\right| \leqslant \sum_{j_2 > j_1} \gamma_{j_2}\gamma_{j_1}.$$

(4.5)

Note that if the noise is time independent then we have

$$\sum_{j_2 > j_1} \gamma^2 = \frac{(m+1)m}{2}\gamma^2,$$

(4.6)

which gives

$$\left|\mathcal{F}_g^{(2)}(m,|\psi\rangle) - \mathcal{F}_g^{(1)}(m,|\psi\rangle)\right| \leqslant \frac{(m+1)m}{2}\gamma^2. \quad (4.7)$$

It is straightforward to show that bounds on higher-order terms go as

$$\left\|S_m^{(k)}\right\|_{1\to1}^H \leqslant \sum_{j_k>\cdots>j_1} \gamma_{j_k}\cdots\gamma_{j_1}, \quad (4.8)$$

so that the difference between the k- and $(k+1)$-order fidelities is bounded by

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| \leqslant \sum_{j_k>\cdots>j_1} \gamma_{j_k}\cdots\gamma_{j_1}. \quad (4.9)$$

Again, if the noise is time independent,

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| \leqslant \binom{m+1}{k}\gamma^k. \quad (4.10)$$

We now discuss our motivation for using $\|\ \|_{1\to1}^H$ as opposed to more familiar norms for distinguishing superoperators, such as the diamond norm. For any superoperator norm $\|\ \|$ that satisfies the properties listed above, the following inequality holds:

$$\left|\mathcal{F}_g^{(k+1)}(m,\psi) - \mathcal{F}_g^{(k)}(m,\psi)\right| \leqslant \binom{m+1}{k}\gamma^k, \quad (4.11)$$

where

$$\gamma := \frac{1}{|\mathrm{Clif}_n|}\sum_i \|\Lambda_i - \Lambda\| \quad (4.12)$$

and for simplicity we have assumed time-independent noise.

The above equations show that in order to give the tightest bound on the fidelity difference we would like to find the norm $\|\ \|$ that provides the smallest value of γ. The diamond norm $\|\ \|_\diamond$ is a candidate; however, by Eq. (2.11) $\|\ \|_{1\to1}^H$ is much weaker than $\|\ \|_\diamond$. Therefore γ associated with $\|\ \|_{1\to1}^H$ will be much smaller than γ associated with $\|\ \|_\diamond$, providing a tighter bound on the fidelity difference.

B. Case where benchmarking fails

There is a simple (and highly unphysical) case in which benchmarking fails. Suppose the noise is time independent and for each i, $\Lambda_i = \mathcal{C}_i^\dagger$. Then $F_g(m,\psi) = 1$ for every m even though there is substantial error on each $\mathcal{C}_i$ and so benchmarking clearly fails. The key point to note here is that the noise is highly dependent on the gate chosen, and so we expect that the sufficient condition derived above for ignoring higher-order terms will not be satisfied (i.e., γ in this example will be far from 0). To see that this is the case, note that since Clif_n is a unitary two-design it is also a unitary one-design. Hence, since Clif_n is † closed,

$$\frac{1}{|\mathrm{Clif}_n|}\sum_{i=1}^{|\mathrm{Clif}_n|}\Lambda_i = \frac{1}{|\mathrm{Clif}_n|}\sum_{i=1}^{|\mathrm{Clif}_n|}\mathcal{C}_i^\dagger = \frac{1}{|\mathrm{Clif}_n|}\sum_{i=1}^{|\mathrm{Clif}_n|}\mathcal{C}_i = \Omega, \quad (4.13)$$

where Ω is the totally depolarizing channel mapping every input state to the maximally mixed state $\frac{\mathbb{1}}{d}$. Therefore,

$$\|\Lambda_i - \Lambda\|_{1\to1}^H = \|\mathcal{C}_i^\dagger - \Omega\|_{1\to1}^H. \quad (4.14)$$

Now $\|\Lambda_i - \Lambda\|_{1\to1}^H$ is achieved at a pure state, and for any pure state $|\psi\rangle$

$$(\Lambda_i - \Lambda)(|\psi\rangle\langle\psi|) = C_i^\dagger|\psi\rangle\langle\psi|C_i - \frac{\mathbb{1}}{d}. \quad (4.15)$$

Hence, if $|\phi\rangle$ is a pure state at which $\|\Lambda_i - \Lambda\|_{1\to1}^H$ is achieved,

$$\begin{aligned}\|\Lambda_i - \Lambda\|_{1\to1}^H &= \left\|C_i^\dagger|\phi\rangle\langle\phi|C_i - \frac{\mathbb{1}}{d}\right\|_1 \\ &= 1 - \frac{1}{d} + (d-1)\frac{1}{d} = \frac{2(d-1)}{d}. \quad (4.16)\end{aligned}$$

Therefore in this case,

$$\gamma = \frac{1}{|\mathrm{Clif}_n|}\sum_i \|\Lambda_i - \Lambda\|_{1\to1}^H = \frac{2(d-1)}{d} \geqslant 1, \quad (4.17)$$

and so our sufficient condition is not satisfied, as expected.

It is important to note that one can devise tests for when such a pathological case is occurring. One simple test is as follows: If the input state is $|\psi\rangle$ then choose Clifford elements $\mathcal{C}_i$ that map $|\psi\rangle$ to an orthogonal state in the measurement basis containing $|\psi\rangle$. For each i, apply $\mathcal{C}_i$ to $|\psi\rangle$ and perform the measurement. For small noise strength the output of the measurement should almost never be ψ; however, if the noise is something close to the inverse of the gate the measurement result will be ψ with high probability.

C. State preparation and measurement errors

In this section we analyze the effect of state-preparation and measurement errors on the benchmarking protocol. The main result is that these errors can be ignored in situations of practical relevance. For simplicity of the discussion, let us assume the gate dependence of the noise is weak enough so that the zeroth-order expression given in Eq. (3.26) is a valid model for the fidelity decay curve. One can obtain an estimate for p as long as the fidelity curve is not constant. As state-preparation and measurement errors are accounted for in A_0 and B_0, we can obtain an estimate for p regardless of the form of the state-preparation and measurement errors whenever the curve is not constant. Thus the protocol is robust against any state-preparation or measurement errors unless these errors create a constant fidelity curve. It is straightforward to characterize exactly when the fidelity curve is constant.

From Eq. (3.26) an exponential decay occurs if and only if A_0 is nonzero and p lies in $(0,1)$. Hence no decay occurs if and only if one of $p = 0$, $p = 1$, or $A_0 = 0$ occurs. We look at each case separately.

$p = 0$. This occurs if and only if Λ is the totally depolarizing channel and in this case the fidelity is constant at $B_0 = \frac{\mathrm{tr}(E_\psi)}{d} \leqslant \frac{1}{d}$. Since we have assumed small gate dependence, this case is possible only if most of the errors are approximately centered around the totally depolarizing channel with little variation. This situation is of little practical relevance since the gate operations being characterized are usually reasonably precise.

$p = 1$. This case corresponds to Λ being the identity channel, which means all gates are perfect. Again, in practice this situation is unlikely as the implementation of any gate will have some associated error. Note that in this case the fidelity is equal to $A_0 + B_0$ which is just $\mathrm{tr}[\Lambda(\rho_\psi)E_\psi] = \mathrm{tr}(\rho_\psi E_\psi)$. Hence the constant decay curve is a measure of the overlap between the imperfect input state and the imperfect POVM element.

$A_0 = 0$. The case $A_0 = 0$ occurs if and only if

$$\mathrm{tr}[E_\psi \Lambda(\rho_\psi)] = \mathrm{tr}\left[E_\psi \Lambda\left(\frac{\mathbb{1}}{d}\right)\right]. \tag{4.18}$$

Thus $\Lambda(\rho_\psi)$ and $\Lambda(\frac{\mathbb{1}}{d})$ have the same probability of producing the output "ψ" from the measurement. Since gates are reasonably precise in practice, this situation occurs when at least one of the state preparation or measurement has substantial error. Note that the fidelity will be equal to B_0 in this case and so can take any value in $[0,1]$.

Of the above three cases, the only one that depends upon state-preparation or measurement errors is the case $A_0 = 0$. Since this case occurs when at least one of the state preparation or measurement has substantial error, it is unlikely to arise in practice. This discussion shows that a constant fidelity decay curve can occur only in extreme cases, and so it is safe to assume the protocol is independent of state-preparation and measurement errors.

V. AVERAGE ERROR RATE AND THE DIAMOND NORM

In terms of connections between the average error rate r and relevant fault-tolerant measures of error, it is natural to ask how the error rate r between Λ and $\mathcal{I}$ is related to the diamond norm between Λ and $\mathcal{I}$. In general an explicit relationship will be impossible to obtain; however, we show that in certain cases that are relevant in various fault-tolerant noise models we can obtain such a relationship. First we give an additional proof of a previously established result [38] for calculating the diamond norm distance between generalized Pauli channels. The proof we present here illustrates how one can apply a semidefinite program to calculate the diamond norm distance between quantum channels [49]. Ideally, this proof technique could be used either to explicitly calculate or to place bounds on the diamond norm distance between more general classes of quantum channels. This could allow for obtaining further relationships between r and the diamond norm distance which hold in more general cases.

A. Calculating the diamond norm distance between generalized Pauli channels

Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are Pauli channels, or more generally any channels with Kraus operators given by an orthogonal (normalized to d) basis of unitary operators $\{P_i\}_{i=0}^{d^2}$ (which we call generalized Pauli channels),

$$\mathcal{E}_1(\rho) = \sum_{i=0}^{d^2-1} q_i P_i \rho P_i^\dagger, \tag{5.1}$$

$$\mathcal{E}_2(\rho) = \sum_{i=0}^{d^2-1} r_i P_i \rho P_i^\dagger. \tag{5.2}$$

Define the vector $\vec{v}$ of length d^2 by

$$v_i = q_i - r_i \tag{5.3}$$

for all $i \in \{0, \ldots, d^2 - 1\}$. Then

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|\vec{v}\|_1 = \sum_{i=0}^{d^2-1} |v_i|. \tag{5.4}$$

To prove Eq. (5.4) using the semidefinite program in Ref. [49], first note that $\Phi = \mathcal{E}_1 - \mathcal{E}_2$ has action

$$\Phi(\rho) = \sum_{i=0}^{d^2-1} (q_i - r_i) P_i \rho P_i^\dagger. \tag{5.5}$$

The semidefinite program has the following primal and dual problems:

Primal problem. Maximize $\langle J(\Phi), W \rangle$ subject to $W \leqslant \mathbb{1}_d \otimes \rho$, $W \in \mathrm{Pos}[L(\mathbb{C}^d \otimes \mathbb{C}^d)]$, $\rho \in D(L(\mathbb{C}^d))$,

Dual problem. Minimize $\|\mathrm{tr}_1(Z)\|_\infty$ subject to $Z \geqslant J(\Phi)$, $Z \in \mathrm{Pos}[L(\mathbb{C}^d \otimes \mathbb{C}^d)]$, where $J(\Phi)$ is the Choi matrix [50] of Φ. If α and β are the solutions to the primal and dual problems then the case that $\alpha = \beta$ is called *strong duality*. It is shown in Ref. [49] that the above semidefinite program always has the property of strong duality and the solution to the program is $\alpha = \frac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond$. Note also that it is always the case that $\alpha \leqslant \beta$.

By definition,

$$\begin{aligned} J(\Phi) &= d\,\Phi \otimes \mathcal{I}(|\psi_0\rangle\langle\psi_0|) \\ &= d\sum_{i=0}^{d^2-1} (q_i - r_i) P_i \otimes \mathbb{1}|\psi_0\rangle\langle\psi_0| P_i^\dagger \otimes \mathbb{1}. \end{aligned} \tag{5.6}$$

Noting that $\{|\psi_i\rangle := P_i \otimes \mathbb{1}|\psi_0\rangle\}_{i=0}^{d^2-1}$ forms an orthonormal basis of maximally entangled states for $\mathbb{C}^d \otimes \mathbb{C}^d$, which we call the generalized Bell basis (GBB), we have that $J(\Phi)$ is diagonal when written in the GBB with diagonal elements (eigenvalues) $d(q_i - r_i)$. Let Π_+ denote the projector onto the eigenspace with non-negative eigenvalues and Π_- denote the projector onto the eigenspace with negative eigenvalues.

For the primal problem let $W = \frac{\Pi_+}{d}$ and $\rho = \frac{\mathbb{1}}{d}$. Then

$$\langle J(\Phi), W \rangle = \sum_{k:q_k-r_k\geqslant 0} q_k - r_k = \frac{1}{2}\sum_k |q_k - r_k| = \frac{1}{2}\|\vec{v}\|_1. \tag{5.7}$$

Thus $\alpha \geqslant \frac{1}{2}\|\vec{v}\|_1$.

For the dual problem take $Z = d\Pi_+ J(\Phi)\Pi_+$ which is just $\sum_{k:q_k-r_k\geqslant 0}(q_k - r_k)|\psi_k\rangle\langle\psi_k|$ and note that $Z \geqslant J(\Phi)$. Moreover, $\mathrm{tr}_1(Z) = d(\sum_{k:q_k-r_k\geqslant 0} q_k - r_k)\frac{\mathbb{1}}{d}$ and so

$$\|\mathrm{tr}_1(Z)\|_\infty = \left(\sum_{k:q_k-r_k\geqslant 0} q_k - r_k\right) = \frac{1}{2}\|\vec{v}\|_1. \tag{5.8}$$

Thus $\alpha \leqslant \frac{1}{2}\|\vec{v}\|_1$, which implies $\alpha = \frac{1}{2}\|\vec{v}\|_1$ and $\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|\vec{v}\|_1$ as desired.

As a simple corollary to Eq. (5.4) note that if $\mathcal{E}_1$ and $\mathcal{E}_2$ are depolarizing channels with fidelity parameters p_1 and p_2, respectively, then

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \frac{2|p_1 - p_2|(d^2 - 1)}{d^2}. \tag{5.9}$$

To see this note that

$$q_0 = \frac{(d+1)\overline{F_{\mathcal{E}_1,\mathcal{I}}} - 1}{d} = \frac{(d+1)\left(p_1 + \frac{1-p_1}{d}\right) - 1}{d}$$

$$= \frac{(d^2-1)p_1 + 1}{d^2} \tag{5.10}$$

and similarly

$$r_0 = \frac{(d^2-1)p_2 + 1}{d^2}. \tag{5.11}$$

Thus, for every $1 \leqslant i \leqslant d^2 - 1$,

$$q_i = \frac{1 - q_0}{d^2 - 1} = \frac{1 - p_1}{d^2} \tag{5.12}$$

and

$$r_i = \frac{1 - r_0}{d^2 - 1} = \frac{1 - p_2}{d^2}. \tag{5.13}$$

So

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \|v\|_1 = |q_0 - r_0| + \sum_{i=1}^{d^2-1} |q_i - r_i|$$

$$= \left| \frac{(d^2-1)p_1 + 1}{d^2} - \left(\frac{(d^2-1)p_2 + 1}{d^2} \right) \right|$$

$$+ (d^2 - 1)\left| \frac{1 - p_1}{d^2} - \left(\frac{1 - p_2}{d^2} \right) \right|$$

$$= 2\frac{(d^2-1)|p_1 - p_2|}{d^2}. \tag{5.14}$$

B. Relating the diamond norm and error rate in benchmarking

Now suppose that $\mathcal{E}_2 = \mathcal{I}$ in Eq. (5.4). Then, $r_0 = 1$ and for every $1 \leqslant i \leqslant d^2 - 1$, $r_i = 0$. Hence in this case

$$\|\mathcal{E}_1 - \mathcal{I}\|_\diamond = \|\vec{v}\|_1 = |q_0 - 1| + 1 - q_0 = 2(1 - q_0). \tag{5.15}$$

We know that q_0 is related to the average fidelity of $\mathcal{E}_1$, $\overline{F_{\mathcal{E}_1,\mathcal{I}}}$, by

$$\overline{F_{\mathcal{E}_1,\mathcal{I}}} = \frac{q_0 d + 1}{d + 1} \tag{5.16}$$

and so

$$\|\mathcal{E}_1 - \mathcal{I}\|_\diamond = \frac{2(d+1)(1 - \overline{F_{\mathcal{E}_1,\mathcal{I}}})}{d}. \tag{5.17}$$

Therefore in the case of randomized benchmarking (where we define the error rate $r = 1 - \overline{F_{\Lambda,\mathcal{I}}}$), if Λ is a generalized Pauli channel, r and $\|\Lambda - \mathcal{I}\|_\diamond$ are related by

$$\|\Lambda - \mathcal{I}\|_\diamond = 2\frac{(d+1)r}{d}. \tag{5.18}$$

VI. SCALABILITY OF THE PROTOCOL

In this section we fill in the details of the scalability proof of our RB protocol that was briefly outlined in Ref. [25]. First, we note that the size of the Clifford group scales as $2^{O(n^2)}$ and so the number of sequences of length m scales as $2^{mO(n^2)}$. Hence, if full averaging over the Clifford group is required for each sequence length, our protocol does not scale well in either of n or m. As mentioned in Ref. [25], there are three obstacles to overcome in order for the above protocol to be scalable:

(1) Sequence length: Since the number of sequences of length m scales as $2^{mO(n^2)}$, averaging over *all* sequences for each m is clearly inefficient.

(2) Uniform sampling: Since the size of the Clifford group scales as $2^{O(n^2)}$, sampling directly from a list of all Clifford elements becomes impossible for large n (writing down every element is inefficient in n).

(3) Implementing Clifford operations: In practice, one can only implement a generating set for the Clifford group. Hence even if random sampling can be accomplished there must be a scalable method for implementing each Clifford using only this generating set.

We now describe how to overcome each of the above obstacles.

Solution to (1). From Eq. (3.12), $\mathcal{F}_g(m, \psi)$ is the uniform average of the random variable

$$\mathcal{F}_g^{\vec{i}_m}(m, |\psi\rangle) := \mathrm{tr}[S_{\vec{i}_m}(\rho_\psi) E_\psi]$$

$$= \mathrm{tr}[\Lambda_{i_{m+1},m+1} \circ \mathcal{C}_{i_{m+1}} \circ \cdots \circ \Lambda_{i_1,1} \circ \mathcal{C}_{i_1}(\rho_\psi) E_\psi] \tag{6.1}$$

over $|\mathrm{Clif}_n|^m$ sequences $(i_1, \ldots, i_m)$. The benchmarking protocol requires choosing a sequence at random, evaluating the above fidelity, repeating for many sequences, and taking the average of the results.

Let $S_k(m, |\psi\rangle) = \frac{\mathcal{F}_g^{\vec{i}_m}(m,|\psi\rangle) + \cdots + \mathcal{F}_g^{\vec{i}_m}(m,|\psi\rangle)}{k}$ be the normalized k-fold sum of the random variable $\mathcal{F}_g^{\vec{i}_m}(m, |\psi\rangle)$ and note that $\mathbb{E}[S_k(m, |\psi\rangle)] = \mathcal{F}_g(m, \psi)$. A probabilistic bound on $|S_k(m, |\psi\rangle) - \mathcal{F}_g(m, \psi)|$ is given by Höeffding's inequality,

$$\mathbb{P}[|S_k(m, |\psi\rangle) - \mathcal{F}_g(m, \psi)| \geqslant \epsilon]$$

$$\leqslant 2e^{-2(k\epsilon)^2/k(b-a)^2} = 2e^{-2k\epsilon^2/(b-a)^2}, \tag{6.2}$$

where $[a, b]$ is the range of $\mathcal{F}_g^{\vec{i}_m}(m, |\psi\rangle)$. Since $\mathcal{F}_g^{\vec{i}_m}(m, |\psi\rangle)$ is a fidelity it must lie in $[0, 1]$ (in reality it will lie in a much smaller interval; for now we continue to assume it lies in $[a, b] \subseteq [0, 1]$). Suppose we want

$$\mathbb{P}[(|S_k(m, |\psi\rangle) - \mathcal{F}_g(m, \psi)| \geqslant \epsilon] \leqslant \delta, \tag{6.3}$$

where ϵ represents the accuracy of the estimate and $1 - \delta$ represents the desired confidence level. We can find how many trials one needs to perform to obtain this accuracy by setting $\delta = 2e^{-2k\epsilon^2/(b-a)^2}$ and solving for k,

$$k = \frac{\ln\left(\frac{2}{\delta}\right)(b-a)^2}{2\epsilon^2}. \tag{6.4}$$

Note that k is explicitly independent of m and n which provides a solution to (1).

It is instructive to obtain an estimate of the size of k for realistic parameter values of δ and ϵ. Since $1 - \delta$ represents our desired confidence level we set $\delta = 0.05$. Fault tolerance provides a wide range for the error tolerance of a physical (0-level) gate in the fault-tolerant construction. The value of the error tolerance depends on both the coding scheme and the noise model and typical values lie somewhere between 10^{-6} and 10^{-2}. Let us assume that the physical gates have errors on the order of 10^{-4}. Intuitively, since the fidelity curve decays in sequence length it is reasonable to assume that ϵ can be relaxed as m grows large. Similarly, $b - a$ can be assumed to be relatively small for small values of m but will converge to $1 - \frac{1}{d}$ as m grows large. As a result both $b - a$ and ϵ have an *implicit* dependence on m, and this implicit dependence is advantageous when choosing ϵ for large values of m. Let us assume $m = 100$ and a fidelity decay curve that is well approximated by an exponential. Then we expect fidelity values on the order of 0.99 at this value of m and so we take $\epsilon = 10^{-3}$, $b - a = 0.2$. With these values for ϵ, δ, and $b - a$ we get

$$k = \frac{\ln\left(\frac{2}{0.05}\right)(0.2)^2}{2(10^{-3})^2} \sim 7 \times 10^4. \tag{6.5}$$

While this number is large it is independent of n and thus compares favorably with quantum process tomography which scales as 16^n. As a direct comparison, performing process tomography on a four-qubit system already requires 65 536 measurements.

Solution to (2). For the second problem we present a method to scalably sample uniformly from the full Clifford group that utilizes the symplectic representation of the Clifford group (see Refs. [51,52]). Since the Clifford group is the normalizer of the Pauli group, every Clifford element is completely determined by its action under conjugation on the Pauli group. In particular, since the Pauli group is generated by the set of all X_i and Z_i (the label i refers to X or Z being in the ith position with identity operators elsewhere), an element of the Clifford group is completely determined by its action on this set. In the symplectic representation this corresponds to each Clifford element $\mathcal{Q}$ being associated uniquely with a $2n \times 2n$ binary symplectic matrix C and length-$2n$ binary vector h which records negative signs in the images of X_i and Z_i. The only constraints on $\mathcal{Q}$ are that commutation relations and Hermiticity of the generating set must be preserved under $\mathcal{Q}$. Hence we can construct a random Clifford element $\mathcal{Q}$ by inductively constructing a random symplectic matrix C and vector h.

Since h corresponds to keeping track of negative signs, the binary entries of h can be chosen uniformly at random. C is inductively constructed column by column, where the first n columns correspond to the images of X_1 through X_n, and the last n columns correspond to the images of Z_1 through Z_n (all of which are written in binary notation as in Ref. [52]). Preservation of commutation relations is phrased through the symplectic inner product and so at each step one chooses the new column by finding a random solution to a system of linear equations which represents the inner-product conditions. Since randomly choosing $2n$ elements of the Pauli group that satisfy the required commutation relations

is equivalent to inductively choosing random solutions to $2n$ sets of linear equations [which requires $O(n^3)$ operations], we can produce a random Clifford element in $O(n^4)$ (classical) operations.

Solution to (3). Any Clifford element can be decomposed into a sequence of $O(n^2)$ one- and two-qubit generators in $O(n^2)$ time [52] [alternatively, there are slower methods which produce a "canonical" decomposition into $O(n^2/\log n)$ generators [53]]. We describe this method which again utilizes the symplectic representation of the Clifford group. As mentioned above, every Clifford element $\mathcal{Q}$ is represented up to a phase by a binary, symplectic matrix C and a binary vector h. The main goal is to decompose C into generators as the negative signs represented by h can be accounted for via multiplication by single-qubit Pauli operators. The main theorem used in the decomposition of Clifford elements is Theorem 4 of Ref. [52] which states that if C is a binary symplectic matrix then C can be decomposed as a product of five binary symplectic matrices, which we denote by T_1 through T_5.

These symplectic matrices can be decomposed into symplectic matrices representing one- and two-qubit Clifford operations that correspond to Hadamard, single-qubit $\frac{\pi}{2}$ rotations about σ_Z, two-qubit $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$, two-qubit permutation operations, and CNOT operations. The overall discussion can be condensed into the following main result:

Main result. Every Clifford operation $\mathcal{Q}$ can be realized by a sequence of one- and two-qubit Clifford operations which consists of the following six rounds of operations:

(1) an initial round of single-qubit Pauli operators,

(2) application of a sequence of CNOT and two-qubit permutation operations,

(3) application of a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z,

(4) application of Hadamard operations,

(5) application of a sequence of $\frac{\pi}{2}$ rotations about $\sigma_Z \otimes \sigma_Z$ followed by a sequence of $\frac{\pi}{2}$ rotations about σ_Z, and

(6) application of a final round of CNOT and two-qubit permutation operations.

Note that the operations within each of the rounds 3, 4, and 5 all commute and can be performed in any order.

The time complexity in decomposing a symplectic matrix into the sequence of one- and two-qubit Clifford operations given above is $O(n^3)$ since one needs to solve linear systems of equations to obtain T_1 through T_5. In many cases one would like to have a decomposition of a Clifford element into a particular generating set for the Clifford group, such as $G_n :=$ {H,S,CNOT} which consists of Hadamard's (H) and phase gates (S) on each qubit, as well as CNOT gates on all pairs of qubits. There are $n^2 + n$ elements in G_n, and it is a straightforward process to decompose the operations in rounds 1 through 6 above into H, S, and CNOT gates.

In total, for an n-qubit system, we can efficiently choose Clifford gates uniformly at random and decompose each gate into a canonical subsequence of elements from the generating set G_n. The total time complexity of these two procedures is $O(n^4) + O(n^3) = O(n^4)$. The number of trials k one needs to perform to estimate $\mathcal{F}_g(m, \psi)$ to an accuracy ϵ with probability at least $1 - \delta$ is given by Eq. (6.4), which is independent of m and n. Thus if we perform the protocol for R different values

of m, the total time complexity is

$$\frac{O(n^4)R\ln(2/\delta)}{2\epsilon^2} \qquad (6.6)$$

which implies that the protocol is scalable in n.

VII. DISCUSSION

We have shown that randomized benchmarking provides a scalable method for benchmarking the set of Clifford gates. The protocol allows for time- and gate-dependent noise, and the fitting models for the fidelity function take into account state-preparation and measurement errors. In addition to providing an estimate of the average fidelity across all Clifford gates, the first-order model provides a measure of the gate dependence of the noise.

We have provided here rigorous proofs of the conditions for the validity of the protocol, as well as the scalability of the protocol in the number of qubits n comprising the system. We have also established an exact relationship between the average fidelity estimate provided by the protocol and a stronger characterization of the average error operator strength given by the diamond norm for the case of random Pauli errors. The proof of this relationship utilizes a semidefinite program for computing the diamond norm [49] which has the potential to establish further connections between these two notions of error strength.

While benchmarking the full unitary group would be ideal, this is a provably inefficient task since just generating a Haar-random unitary operator is inefficient in n. On the other hand as we have shown here, benchmarking the Clifford group is an efficient task. It is not difficult to see that benchmarking the Clifford group provides significant information for both fault-tolerant quantum computation as well as obtaining a benchmark for a generating set of the full unitary group. First, any realistic implementation of a quantum computer will have to take advantage of error-correction codes in order to perform fault-tolerant quantum computation. The fact that most of the codes used in fault-tolerant theory are stabilizer codes implies that the encoding and decoding operations that have to be performed can be chosen to be Clifford operations. Hence a benchmark of Clifford operations provides direct information regarding the robustness of these encoding and decoding schemes.

Second, the unitary group can be generated by adding just one single-qubit rotation not in the Clifford group (for instance the $\frac{\pi}{8}$ gate). Hence a benchmark for the Clifford group can actually provide useful information regarding a benchmark for a generating set of the full unitary group. In addition, it has been shown that any unitary operation can be implemented using Clifford gates, a single-qubit ancilla state called a magic state [54] and measurements in the computational basis. Hence in this model of quantum computation the only gates that need to be benchmarked for universal quantum computation are Clifford gates.

Various interesting questions and comments arise from the benchmarking analysis presented here. First, there is a key point to emphasize regarding the zeroth- and first-order fitting models. As depicted in Ref. [25], there exist physically relevant noise models for which when the true value of the depolarization fidelity parameter p is used, the first-order model fits the experimental data much better than the zeroth-order model. However, it may be the case that a least-squares fitting procedure using the functional form of the zeroth-order model produces a very good fit to the experimental data, albeit producing an incorrect value for p. Therefore in order to obtain a more accurate value for p one should always use the first-order fitting model unless prior knowledge of the noise indicates that it is effectively gate independent.

It will be useful to obtain a better understanding of when a least-squares fitting procedure using the zeroth-order model produces a value for p that is close to its true value. Clearly, in the gate-independent case the zeroth-order model fits the fidelity decay curve exactly. Moreover, for weakly gate-dependent noise one can see from our continuity argument that the zeroth-order model is still a sufficient fitting function for the fidelity decay curve. Hence the most interesting case to analyze is when there is a non-negligible amount of gate dependence in the noise and the condition for using the first-order model to fit the decay curve is satisfied. A useful test that would indicate gate dependence in the noise, and thus the validity of the value of p obtained from fitting to the zeroth-order model, is to perform the least-squares fitting procedure using both the zeroth- and first-order fitting models. If the estimates of p obtained in each case differ significantly then the zeroth-order model must be a poor choice of fitting function even though it may fit the data well. In this case the noise must have a strong gate dependence because otherwise $q - p^2$ would be small which implies that the two fitting functions would produce similar estimates for p.

An interesting question is how to extract a meaningful average error rate over a generating set of the Clifford group, for instance G_n defined previously, from the average error rate r over the entire Clifford group. One might argue that benchmarking a generating set for the Clifford group is sufficient for benchmarking the full Clifford group; however, it is entirely plausible that noise correlations between the n physical qubits create large errors on elements of Clif_n, even when the errors on the generating set can be controlled [55]. In fact an assumption that is often made in fault-tolerant estimates is that the correlation in noise between qubits either is small or can be ignored.

With regard to scalability, while we have shown that the protocol itself is scalable in n, a useful direction for further research would be an analysis of how the sufficient condition of weak average variation of the noise depends on n. As previously noted, the noise associated with a multiqubit Clifford element is given by the noise associated with the sequence of generators comprising the Clifford. A determination of whether these noise operators continue to satisfy the sufficient condition when it is met for small numbers of qubits will be useful for understanding the applicability of the protocol.

Rigorous fault-tolerant analyses sometimes invoke the diamond norm as a measure of the error strength rather than the weaker characterization provided by the average fidelity. Hence it is desirable to find relationships between these two quantities that is more general than the special case of random Pauli errors presented here. As mentioned above, the semidefinite program we have used to deduce the relationship

appears to be a promising tool for further research in this area. From the expression given in Eq. (2.2), one can see that the diamond norm is essentially a "worst-case" maximization over input (entangled) states. In quantum computation it is the case that the measure of accessible states (states that can be reached in polynomial time using a generating set for the unitary group) is equal to 0. Hence there is a high probability that the maximization criterion demanded by the diamond norm is a much stronger condition than necessary for understanding the strength of the errors affecting the computation. This point becomes even more relevant for an algorithm-specific (i.e., nonuniversal) quantum computer. An interesting direction of further research is to provide precise conditions for when

the average fidelity provides an indication or bound on the error strength in terms of stronger characterizations such as the diamond norm.

Additionally, if one were able to obtain an estimate of the minimum gate fidelity from knowledge of the average fidelity they could use the direct relationship between the minimum gate fidelity and diamond norm given by Eq. (2.21) to obtain information about the error strength in terms of the diamond norm. A result that may be useful in this direction of research is the "concentration-of-measure effect" of the gate fidelity, which implies that as n increases, the measure of the set of states which produce a fidelity close to the minimum yet far from the average is exponentially small in n [41,42].

[1] P. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society Press, Los Alamitos, CA, 1994).

[2] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).

[3] R. Feynman, Int. J. Theor. Phys. **21**, 467 (1982).

[4] S. Lloyd, Science **273**, 1073 (1996).

[5] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).

[6] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).

[7] A. Steane, Proc. R. Soc. London, Ser. A **452**, 2551 (1996).

[8] D. Aharonov and M. Ben-Or, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, edited by F. T. Leighton and P. W. Shor (ACM, New York, 1997).

[9] E. Knill, R. Laflamme, and W. Zurek, Proc. R. Soc. London, Ser. A **454**, 365 (1997).

[10] J. Preskill, e-print arXiv:quant-ph/9712048.

[11] J. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).

[12] D. G. Cory, M. D. Price, and T. F. Havel, Physica D: Nonlin. Phen. **120**, 82 (1998).

[13] D. Loss and D. P. Divincenzo, Phys. Rev. A **57**, 120 (1998).

[14] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai, Nature (London) **398**, 786 (1999).

[15] I. Chuang and M. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[16] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).

[17] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009).

[18] J. Emerson, R. Alicki, and K. Zyczkowski, J. Opt. B: Quantum Semiclass. Opt. **7**, S347 (2005).

[19] B. Levi, C. C. Lopez, J. Emerson, and D. G. Cory, Phys. Rev. A **75**, 022314 (2007).

[20] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. Cory, and R. Laflamme, Science **317**, 1893 (2007).

[21] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Phys. Rev. A **78**, 012347 (2008).

[22] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).

[23] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

[24] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).

[25] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).

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

[27] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Quantum Inf. Comput. **9**, 0920 (2009).

[28] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. A **84**, 030303 (2011).

[29] C. Ryan, M. Laforest, and R. Laflamme, New J. Phys. **11**, 013034 (2009).

[30] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009).

[31] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **82**, 040305 (2010).

[32] S. Olmschenk, R. Chicireanu, K. D. Nelson, and J. V. Porto, New J. Phys. **12**, 113007 (2010).

[33] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing, Dallas, TX, 1998*, edited by J. S. Vitter (ACM, New York, 1998).

[34] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Boston, 2002).

[35] R. Horn and C. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1990).

[36] J. Watrous, Quantum Inf. Comput. **5**, 058 (2005).

[37] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, Cambridge, UK, 2002).

[38] M. Sacchi, J. Opt. B: Quantum Semiclassical Opt. **7**, S333 (2005).

[39] M. Nielsen and I. Chuang, *Quantum Computation and Information* (Cambridge University Press, Cambridge, UK, 2000).

[40] M. Nielsen, Phys. Lett. A **303**, 249 (2002).

[41] E. Magesan, R. Blume-Kohout, and J. Emerson, Phys. Rev. A **84**, 012309 (2011).

[42] E. Magesan, Quantum Inf. Comput. **11**, 0466 (2011).

[43] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, UK, 2006).

[44] C. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[45] P. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Computer Society Press, Los Alamitos, CA, 1999).

[46] D. Gottesman, Ph.D. thesis, Caltech, 1997, e-print arXiv:quant-ph/9705052.

[47] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[48] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White, Phys. Rev. Lett. **90**, 193601 (2003).

[49] J. Watrous, Theory Comput. **5**, 217 (2009).

[50] M. Choi, Lin. Alg. Appl. **10**, 285 (1975).

[51] C. Dankert, Ph.D. thesis, University of Waterloo, 2005, e-print arXiv:quant-ph/0512217v2.

[52] J. Dehaene and B. De Moor, Phys. Rev. A **68**, 042318 (2003).

[53] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).

[54] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).

[55] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **106**, 130506 (2011).