Lawrence Berkeley National Laboratory
LBL Publications

Title
Demonstrating scalable randomized benchmarking of universal gate sets

Permalink
https://escholarship.org/uc/item/99r2c42w

Authors
Hines, Jordan
Lu, Marie
Naik, Ravi K
et al.

Publication Date
2022-07-14

Copyright Information

Demonstrating scalable randomized benchmarking of universal gate sets

Jordan Hines,[1,2,*] Marie Lu,[3] Ravi K. Naik,[3,4] Akel Hashim,[3,4] Jean-Loup Ville,[3,†] Brad
Mitchell,[3,4] John Mark Kriekebaum,[3,5,‡] David I. Santiago,[3,4] Stefan Seritan,[2] Erik Nielsen,[2]
Robin Blume-Kohout,[2] Kevin Young,[2] Irfan Siddiqi,[3,4,5] Birgitta Whaley,[6] and Timothy Proctor[2,§]

[1]*Department of Physics, University of California, Berkeley, CA 94720*
[2]*Quantum Performance Laboratory, Sandia National Laboratories, Albuquerque, NM 87185 and Livermore, CA 94550*
[3]*Quantum Nanoelectronics Laboratory, Department of Physics, University of California, Berkeley, CA 94720*
[4]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
[5]*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
[6]*Department of Chemistry, University of California, Berkeley, CA 94720*

Randomized benchmarking (RB) protocols are the most widely used methods for assessing the performance of quantum gates. However, the existing RB methods either do not scale to many qubits or cannot benchmark a universal gate set. Here, we introduce and demonstrate a technique for scalable RB of many universal and continuously parameterized gate sets, using a class of circuits called randomized mirror circuits. Our technique can be applied to a gate set containing an entangling Clifford gate and the set of arbitrary single-qubit gates, as well as gate sets containing controlled rotations about the Pauli axes. We use our technique to benchmark universal gate sets on four qubits of the Advanced Quantum Testbed, including a gate set containing a controlled-S gate and its inverse, and we investigate how the observed error rate is impacted by the inclusion of non-Clifford gates. Finally, we demonstrate that our technique scales to many qubits with experiments on a 27-qubit IBM Q processor. We use our technique to quantify the impact of crosstalk on this 27-qubit device, and we find that it contributes approximately ⅔ of the total error per gate in random many-qubit circuit layers.

I. INTRODUCTION

Quantum computers suffer from a diverse range of errors that must be quantified if their performance is to be understood and improved. Errors that are localized to single qubits or pairs of qubits can be studied in detail using tomographic techniques [1, 2]. However, many-qubit circuits are often subject to large additional errors, such as crosstalk [3–8], that are not apparent in isolated one- or two-qubit experiments. There are now techniques for partial tomography on individual many-qubit circuit layers (also called "cycles"), including cycle benchmarking [9] and Pauli noise learning [10–12]. But quantum computers can typically implement exponentially many different circuit layers, and it is only feasible to characterize a small subset of them.

Randomized benchmarks [5–8, 13–31] make it possible to quantify the rate of errors in an average n-qubit layer, by probing a quantum computer's performance on random n-qubit circuits. However, established randomized benchmarks cannot measure the performance of *universal* layer sets in the many-qubit regime, where quantum computational advantage may be possible. Those randomized benchmarks that can be applied to universal layer sets, such as standard randomized benchmarking (RB) [15, 16] and cross-entropy benchmarking (XEB) [27–29], require classical computations that scale exponentially in the number of qubits (n). XEB requires classical simulation of random circuits that are famously infeasible to simulate for more than approximately 50 qubits [28]. This is because XEB requires estimating the (linear) cross-entropy

---

* jordanh@berkeley.edu
† Now at Alice & Bob, 53 Bd du Général Martial Valin, 75015 Paris
‡ Now at Google Quantum AI, Mountain View, CA, USA.
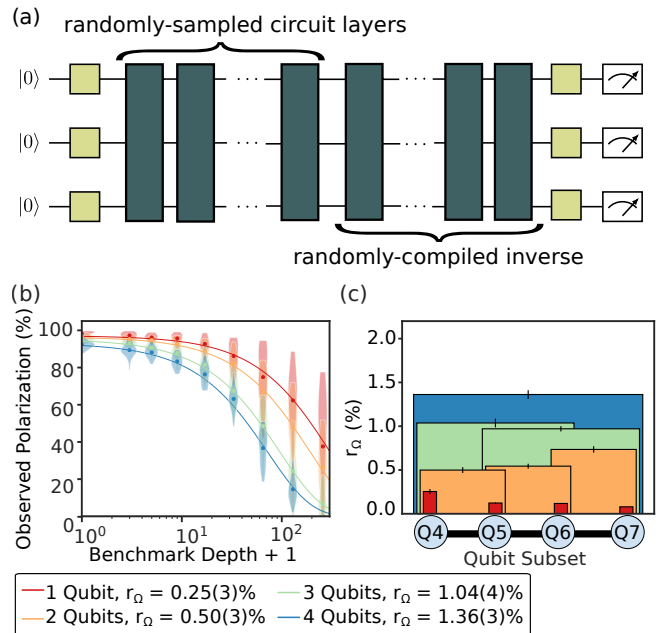§ tjproct@sandia.gov

FIGURE 1. **Scalable randomized benchmarking of universal gate sets.** (a) Randomized mirror circuits combine a simple reflection structure with randomized compiling to enable scalable and robust RB of universal gate sets. (b) Data and fits to an exponential obtained by using our method—MRB of universal gate sets—to benchmark a universal gate set on $n = 1, 2, 3, 4$ qubits of the Advanced Quantum Testbed, and the average error rates of n-qubit layers (r_Ω, where Ω is the layer sampling distribution) extracted from these decays. (c) We benchmarked each connected set of n qubits for $n = 1, 2, 3, 4$, enabling us to map out the average layer error rate (r_Ω) for each subset of qubits.

between each circuit's actual and ideal output distributions.

Standard RB of a universal layer set is restricted to even smaller n, because it requires compiling and running Haar random n-qubit unitaries [15]. This compilation requires classical computations that are exponentially expensive in n, and results in circuits containing $O(2^n)$ two-qubit gates [32]. Due to the large overhead, even standard RB on Clifford gates—which has lower overheads and non-exponential scaling—has only been implemented on up to 5 qubits [5, 6, 8].

In this paper we introduce and demonstrate a simple and scalable technique for RB of a broad class of universal gate sets. Our technique uses a novel kind of *randomized mirror circuits*, shown in Fig. 1 (a), and advances on a recently introduced method—*mirror* RB (MRB)—that enables scalable RB of Clifford gates [6]. Our randomized mirror circuits use a layer-by-layer inversion structure that enables classically efficient circuit construction and prediction of that circuit's error-free output. Furthermore, randomized mirror circuits do not contain the exponentially large subroutines used in other RB methods, which enables MRB of even hundreds or thousands of qubits. To perform MRB on universal gate sets, we run randomized mirror circuits of varied depths (d) and compute their mean *observed polarization* [6], a quantity that is closely related to success probability. The mean observed polarization versus circuit depth is fit to an exponential decay, as shown in Fig. 1 (b). As in standard RB, the estimated decay rate is then simply rescaled to estimate the average error rate of n-qubit layers. MRB therefore preserves the core strengths and simplicity of standard RB and XEB, while avoiding the classical simulation and compilation roadblocks that have prevented scalable and efficient RB of universal layer sets.

We use MRB to study errors in two different quantum computing systems, based on superconducting qubits. Our experiments are on 4 qubits of the Advanced Quantum Testbed (AQT) [33] and on all of the qubits of a 27-qubit IBM Q quantum computer (`ibmq_montreal`) [34]. In our experiments on AQT we use MRB to quantify and compare the performance of three different layer sets on each subset of n qubits (for $n = 1, 2, 3, 4$), including a layer set containing non-Clifford two-qubit gates [see Fig. 1 (b-c)]. In our experiments on `ibmq_montreal` we show that our method scales to many qubits by performing MRB on a universal gate set on up to 27 qubits.

Multi-qubit MRB enables probing and quantifying crosstalk, which is an important source of error in contemporary many-qubit processors [3–5, 7] that cannot be quantified by only testing one or two qubits in isolation. We quantify the contribution of crosstalk errors to the observed error rates in our experiments on AQT and further divide the error into contributions from individual layers and gates. The techniques we introduce for these analyses complement other established RB-like methods for estimating the error rates of individual gates—such as interleaved RB [35–37] and cycle benchmarking [9]. In our experiments on `ibmq_montreal`, we use MRB to study how crosstalk errors vary on this device as n increases, with n ranging from $n = 1$ up to $n = 27$. We find that crosstalk errors dominate in circuit layers on $n \gg 1$ qubits.

This paper is structured as follows: In Section II we introduce our notation and define the error rate that our method measures. In Section III we define the MRB protocol. In Section IV we present theory and simulations that show that MRB is reliable. In Sections V and VI we present the results of our experiments on AQT and IBM Q's quantum processors, respectively.

II. DEFINITIONS AND PRELIMINARIES

In this section, we introduce our notation and background information related to our method. In Section II A we introduce the notation and definitions used throughout this paper. In Section II B we define the error rate that MRB is designed to measure. In Section II C we describe the gate sets that our method can be used to benchmark, i.e., we state the conditions that a gate set must satisfy if it is to be benchmarked with MRB.

A. Definitions

We begin by introducing our notation and definitions. A k-qubit *gate* g is an instruction to perform a particular unitary operation $U(g) \in \mathbb{SU}(2^k)$ on k qubits. We will only consider $k = 1, 2$, and we use $\mathbb{G}_1$ and $\mathbb{G}_2$ to denote a set of one- and two-qubit gates, respectively. In this work $\mathbb{G}_2$ will only contain controlled rotations about the X, Y, or Z axis, denoted CP_θ and defined by

$$U(CP_\theta) = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes e^{-i\frac{\theta}{2}P}, \qquad (1)$$

where θ is the angle of rotation and P is the axis of rotation. Our experiments use four of these gates, which we write as $\mathsf{cs} = CZ_{\pi/2}$, $\mathsf{cs}^\dagger = CZ_{-\pi/2}$, $\mathsf{cphase} = CZ_\pi$, and $\mathsf{cnot} = CX_\pi$. We denote the single-qubit gate that is a rotation by θ about P by P_θ. An n-qubit, depth-d *circuit* is a length-d sequence of n-qubit layers $C = L_d L_{d-1} \cdots L_2 L_1$. An n-qubit *layer L* is an instruction to perform a particular unitary operation $U(L) \in \mathbb{SU}(2^n)$ on those n qubits. In this work, we use layers that consist of parallel applications of only one-qubit gates or only two-qubit gates. We use $\mathbb{L}(\mathbb{G})$ to denote the set of all layers constructed by parallel applications of gates from the gate set $\mathbb{G}$. Often it will be convenient to think of random circuits and layers as random variables, and when we do so we use the L font, e.g., we often use L to denote a layer-valued random variable, meaning that $\mathsf{L} = L$ with probability $\Omega(L)$ for some distribution Ω over $\mathbb{L}(\mathbb{G})$. We use L^{-1} to denote an instruction to perform the operation $U(L)^{-1}$.

For a layer or circuit L, we use $\mathcal{U}(L)$ and $\phi(L)$ to denote the superoperator for its perfect and imperfect implementations, respectively, so $\mathcal{U}(L)[\rho] = U(L)\rho U^\dagger(L)$. We often represent superoperators as matrices, acting on states represented as vectors in Hilbert-Schmidt space (denoted by $|\rho\rangle\rangle$). A layer L's error map is defined by $\mathcal{E}(L) = \phi(L)\mathcal{U}^\dagger(L)$. The *entanglement fidelity* of $\phi(L)$ to $\mathcal{U}(L)$ is defined by

$$F(\phi(L), \mathcal{U}(L)) = F(\mathcal{E}) = \langle\langle\varphi|(\mathbb{I} \otimes \mathcal{E}(L))[|\varphi\rangle\langle\varphi|]|\varphi\rangle\rangle, \quad (2)$$

where φ is any maximally entangled state of $2n$ qubits [38]. This fidelity is linearly related to another widely-used fidelity variant—the *average gate fidelity* [38]

$$\bar{F}(\phi(L), \mathcal{U}(L)) = \int d\psi \, \langle \psi | \mathcal{E}(L)[\psi] | \psi \rangle \tag{3}$$

$$= \frac{2^n F(\phi(L), \mathcal{U}(L)) + 1}{2^n + 1}, \tag{4}$$

where $d\psi$ is the Haar measure on pure states. Throughout, we use the term "(in)fidelity" to refer to the entanglement (in)fidelity.

Our theory will make use of the *polarization* of a channel $\mathcal{E}$, which is a rescaling of $\mathcal{E}$'s fidelity given by

$$\gamma(\mathcal{E}) = \frac{4^n}{4^n - 1} F(\mathcal{E}) - \frac{1}{4^n - 1}, \tag{5}$$

as well as stochastic Pauli channels. An n-qubit stochastic Pauli channel $\mathcal{E}_{\mathrm{pauli},\{\gamma_P\}}$ is parameterized by a probability distribution $\{\varepsilon_P\}$ over the 4^n Pauli operators ($\mathbb{P}_n$), and it has the action

$$\mathcal{E}_{\mathrm{pauli},\{\varepsilon_P\}}[\rho] = \sum_{P \in \mathbb{P}_n} \varepsilon_P P \rho P, \tag{6}$$

with $\sum_{P \in \mathbb{P}_n} \varepsilon_P = 1$. For a stochastic Pauli channel, the total probability of a fault, i.e., the probability it applies a non-identity Pauli operator, is $1 - \varepsilon_{\mathbb{I}_n} = 1 - F(\mathcal{E}_{\mathrm{pauli},\{\varepsilon_Q\}})$.

B. Fidelity decay and Ω-distributed random circuits

In this work we introduce a method for RB of a family of universal n-qubit layer sets. By its definition (see Section III B), our method estimates *an* error rate (r_Ω) for these n-qubit layers. To confirm that our method works correctly, we must show that it reliably estimates an independently-defined and well-motivated notion for the average error rate for a set of n-qubit layers. Surprisingly, defining the error rate that our method (or any other RB method) should aim to estimate is challenging. RB protocols are often formulated as methods for measuring the mean infidelity of a set of n-qubit gates or layers, but this is subtly flawed: the mean infidelity is not an observable property of a set of physical gates—it is not "gauge-invariant" [39]. One solution to this problem, which we adopt herein, was introduced in Ref. [40]: the rate of decay of the mean fidelity of a family of random circuits, as a function of increasing circuit depth, is (approximately) gauge-invariant. This decay rate can therefore be what an RB protocol aims to measure.

Defining the error rate that an RB protocol aims to measure in terms of the decay rate of fidelity of random circuits raises an interesting question: *which* random circuits? A "random circuit" is not uniquely defined: there are many possible distributions from which random circuits can be sampled, and the rate that the fidelity of those circuits decays will vary from family to family. For example, the degree to which coherent errors can systematically add across circuit layers depends on the speed at which those layers scramble errors. In this work,

we introduce a natural family of circuits—which we call Ω-distributed random circuits—and we aim to estimate the decay rate of the mean fidelity of these circuits.

Ω-distributed random circuits are similar to the circuits used in XEB and other benchmarking routines. They are defined in terms of a customizable gate set $\mathbb{G}$ and sampling distribution Ω over that gate set. This gate set consists of one- and two-qubit gate sets $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$, and Ω is determined by two probability distributions Ω_1 and Ω_2 over n-qubit layer sets $\mathbb{L}(\mathbb{G}_1)$ and $\mathbb{L}(\mathbb{G}_2)$, respectively. An Ω-distributed random circuit with a *benchmark depth* of d is a circuit-valued random variable $\mathsf{C}_d = \mathsf{L}_{2d} \cdots \mathsf{L}_2 \mathsf{L}_1$ where the d odd-indexed layers are Ω_1-distributed and the d even-indexed layers are Ω_2-distributed. These circuits consist of interleaved layers of one and two-qubit gates, so it is useful to define a *composite layer* to be a pair of layers of the form $\mathsf{L} = \mathsf{L}_2 \mathsf{L}_1$ where $\mathsf{L}_1 \in \mathbb{L}(\mathbb{G}_1)$ is a layer of one-qubit gates and $\mathsf{L}_2 \in \mathbb{L}(\mathbb{G}_2)$ a layer of one-qubit gates. We denote the set of all composite layers by $\mathbb{L}(\mathbb{G})$. An Ω-distributed random circuit of benchmark depth d then consists of d composite layers that are Ω-distributed over $\mathbb{L}(\mathbb{G})$ with $\Omega(L_2 L_1) = \Omega_1(L_1)\Omega_2(L_2)$. We require that our Ω-distributed random circuits are highly scrambling [41] (meaning that, in expectation, they quickly delocalize errors) which is not true for every possible $\mathbb{G}$ and Ω [5]. In Section II C we present the conditions we require of $\mathbb{G}$ and Ω, which guarantee this scrambling.

Our protocol aims to estimate the rate at which the fidelity of Ω-distributed random circuits decays with depth. The average fidelity of Ω-distributed random circuits with benchmark depth d ($\bar{F}_d$) is given by

$$\bar{F}_d = \mathop{\mathbb{E}}_{\mathsf{C}_d} F(\mathcal{U}(\mathsf{C}_d), \phi(\mathsf{C}_d)). \tag{7}$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation value (here over the circuit-valued random variable C_d). The requirement that our Ω-distributed circuits are highly scrambling ensures that $\bar{F}_d$ decays exponentially, and therefore has a well-defined rate of decay. In Section IV we show that $\bar{F}_d$ decays exponentially in depth for our circuits, i.e., $\bar{F}_d \approx A p_{\mathrm{rc}}^d + B$, for constants A and B. We then define

$$\epsilon_\Omega = (4^n - 1)(1 - p_{\mathrm{rc}})/4^n. \tag{8}$$

We choose ϵ_Ω to be this particular rescaling of p_{rc} because p_{rc} corresponds to the effective polarization of a random composite layer in an Ω-distributed random circuit—i.e., the polarization in a depolarizing channel that would give the same fidelity decay—and so ϵ_Ω is the effective average fidelity of a layer sampled from Ω. When stochastic Pauli errors are the dominant source of error, ϵ_Ω is approximately equal to the commonly-used metric of average layer infidelity (see Appendix A 3).

C. The gate set and sampling distributions

Our technique requires certain conditions of the gate set $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$ and the sampling distributions Ω_1 and Ω_2. The gate set $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$ must have the following properties:

1. $\mathbb{G}_1$ is closed under inverses and conjugation by Pauli operators, and is a unitary 2-design over $\mathbb{SU}(2)$. Examples of valid $\mathbb{G}_1$ are the set of all single-qubit gates $\mathbb{SU}(2)$ and the set of all 24 single-qubit Clifford gates $\mathbb{C}_1$.

2. $\mathbb{G}_2$ contains at least one entangling gate, it is closed under inverses, and each gate in $\mathbb{G}_2$ is a CP_θ gate (defined in Section II A). Examples of valid $\mathbb{G}_2$ are {cnot} and {cs, cs†}.

3. For all $CP_\theta \in \mathbb{G}_2$, $\mathbb{G}_1$ is closed under multiplication by the single-qubit Pauli axis rotation P_θ. This is guaranteed to hold if $\mathbb{G}_1$ is the set of all single-qubit gates $\mathbb{SU}(2)$.

We require the following properties of the sampling distributions Ω_1 and Ω_2:

1. Ω_1 is the uniform distribution over $\mathbb{G}_1$.

2. Ω_2 is invariant under exchanging any subset of the gates in a two-qubit gate layer L with their inverses.

3. Ω_2-distributed layers quickly delocalize errors.

Our method can be generalized to single-qubit gate sets $\mathbb{G}_1$ that only *generate* a unitary 2-design and to distributions Ω_1 other than the uniform distribution. However, this complicates the analysis, so we do not consider this more general case herein.

III. SCALABLE RANDOMIZED BENCHMARKING OF UNIVERSAL GATE SETS

In this section we introduce our method for MRB of universal gate sets. In Section III A we introduce the family of randomized mirror circuits used in MRB. In Section III B we explain the MRB data analysis and define the complete MRB protocol.

A. Randomized mirror circuits for universal gate sets

Our protocol uses a novel family of *randomized mirror circuits* [6, 7, 31] that we now introduce. The structure of these randomized mirror circuits allows our protocol to measure ϵ_Ω, the average error rate of n-qubit layers sampled from Ω (see Section II B), without expensive classical computation. One approach to estimating ϵ_Ω is to run Ω-distributed random circuits of varied depths, and then estimate the decay in the (linear) cross entropy between these circuits' ideal and actual output probability distributions [27, 29]. This is because the decay rate of this cross entropy is known to be approximately equal to ϵ_Ω [27, 29]. The problem with this method is that the classical computation cost of computing the ideal output probability distribution scales exponentially in the number of qubits (n) when the gate set is universal [28], limiting it to $n \lesssim 50$. To estimate ϵ_Ω without expensive classical computation our protocol runs Ω-*distributed randomized mirror*

circuits, which use an inversion structure to transform an Ω-distributed random circuit into a circuit with an efficiently-computable outcome.

We construct a specific randomized mirror circuit on n qubits with benchmark depth d via the three-step procedure shown in Fig. 2. This procedure consists of first sampling a circuit C_1 consisting of an Ω-distributed random circuit preceded by an initial layer of random single-qubit gates that randomizes the state input into the circuit (enabling estimation of the circuit's fidelity using the method of Ref. [42]). We then append the inverse of C_1 to obtain C_2, a simple form of mirror (or motion-reversal) circuit whose error-free output is definite and easy to predict. Finally, C_2 is randomly compiled, to prevent systematic coherent addition or cancellation of errors between the Ω-distributed random circuit and its inverse—which is essential for reliable estimation of ϵ_Ω. The exact procedure is as follows:

1. (Sample a random circuit) Construct a circuit $C_1 = L_{d/2}L_{\theta_{d/2}}\cdots L_1 L_{\theta_1}L_0$ consisting of:

   (a) A layer L_0 sampled from Ω_1, which consists of a single-qubit gate on each qubit.

   (b) $d/2$ composite layers $L_i L_{\theta_i}$, where L_i is sampled from Ω_1, and L_{θ_i} is sampled from Ω_2.

2. (Construct simple mirror circuit) Add to the circuit C_1 the layers in step 1 in reverse order, with each layer replaced with its inverse. The result is a circuit

$$C_2 = L_0^{-1}L_{\theta_1}^{-1}L_1^{-1}\cdots L_{\theta_{d/2}}^{-1}L_{d/2}^{-1}L_{d/2}L_{\theta_{d/2}}\cdots L_1 L_{\theta_1}L_0, \quad (9)$$

such that $U(C_2) = \mathbb{I}$.

3. (Randomized compiling) Construct a new circuit M by starting with C_2 and replacing layers using the following randomized compilation procedure, which reduces to standard Pauli frame randomization [43] when the two-qubit gates are all Clifford gates. To specify our procedure, we first write C_2 [Eq. (9)] in the form

$$C_2 = L_{d+1}L_{\theta_{d+1}}L_d\cdots L_{\theta_{d/2+2}}L_{d/2+1}L_{\theta_{d/2+1}}L_{d/2}L_{\theta_{d/2}}\cdots L_1 L_{\theta_1}L_0,$$

where $L_{\theta_{d/2+1}}$ is a dummy (empty) 2-qubit gate layer, so that C_2 consists of alternating layers of one- and two-qubit gates. Then:

   (a) For each single-qubit gate layer L_i in C_2, sample a uniformly random layer of Pauli gates P_i, that in the following procedure is inserted after and then compiled into L_i.

   (b) Replace each two-qubit gate layer L_{θ_i} in C_2 with a new two-qubit gate layer $\mathcal{T}(L_{\theta_i}, P_{i-1})$ that is constructed as follows: For each gate CP_θ in L_{θ_i} with control qubit q_j and target qubit q_k, consider the instructions in P_{i-1} acting on q_j and q_k, denoted by $P_{i-1,j}$ and $P_{i-1,k}$, respectively. If $U(P_{i-1,j}) = I$ or Z, then add CP_ϕ acting on (q_j, q_k) to $\mathcal{T}(L_{\theta_i}, P_{i-1})$ where $\phi = \theta$ if $[U(P), U(P_{i-1,k})] = 0$ and $\phi = -\theta$ otherwise.

If $U(P_{i-1,j}) = X$ or Y, then add CP_ϕ acting on (q_j, q_k) to $\mathcal{T}(L_{\theta_i}, P_{i-1})$ where $\phi = -\theta$ if $[U(P), U(P_{i-1,k})] = 0$ and $\phi = \theta$ otherwise.

(c) For each single-qubit gate layer L_i in C_2 with $i > 0$, we define a layer of single-qubit gates P_i^c that undoes the effect of adding P_{i-1} into the circuit—meaning a layer such that $U(P_i^c \mathcal{T}(L_{\theta_i}, P_{i-1}) P_{i-1}) = U(L_{\theta_i})$. Because $\mathbb{G}_2$ is restricted to only controlled Pauli-axis rotations, the correction takes the form $U(P_i^c) = U(P_{i-1} P_{\tilde{\theta}_i})$, where $P_{\tilde{\theta}_i}$ consists of single-qubit Pauli axis rotations. If L_i is not immediately preceded by a two-qubit gate layer, then $P_{\tilde{\theta}_i} = \mathbb{I}$. Otherwise,

$$U(P_{\tilde{\theta}_i}) = U(P_{i-1} L_{\theta_i} P_{i-1} \mathcal{T}(L_{\theta_i}, P_{i-1})^{-1}). \tag{10}$$

(d) Replace each single-qubit gate layer L_i in C_2 with a recompiled layer $\mathcal{R}(P_i L_i P_{i-1}^c)$, defined by

$$U(\mathcal{R}(P_i L_i P_{i-1}^c)) = U(P_i L_i P_{i-1}^c). \tag{11}$$

This randomized compilation step transforms the layer pair $L_i L_{\theta_i}$ into $\mathcal{R}(P_i L_i P_{i-1}^c)\mathcal{T}(L_{\theta_i}, P_{i-1})$, where

$$U(\mathcal{R}(P_i L_i P_{i-1}^c)\mathcal{T}(L_{\theta_i}, P_{i-1})) = U(P_i L_i L_{\theta_i} P_{i-1}). \tag{12}$$

The final circuit produced by this procedure (M) has the property that $U(M) = U(P_{d+1})$, i.e., its overall action is an n-qubit Pauli operator. So, if run perfectly, M returns a single bit string (s_M) that is determined during circuit construction with no additional computation needed.

The final depth-d randomized mirror circuit has the form

$$M = \mathcal{R}(P_{d+1} L_0^{-1} P_d^c) \, \tilde{M} \, \mathcal{R}(P_0 L_0), \tag{13}$$

where

$$\tilde{M} = \mathcal{T}(L_{\theta_1}^{-1}, P_d)\mathcal{R}(P_d L_1^{-1} P_{d-1}^c)\cdots\mathcal{R}(P_{d/2+1} L_{d/2}^{-1} P_{d/2})$$
$$\mathcal{R}(P_{d/2} L_{d/2} P_{d/2-1}^c)\cdots\mathcal{R}(P_2 L_1 P_1^c)\mathcal{T}(L_{\theta_1}, P_0),$$

is the circuit obtained after applying randomized compilation to the $d/2$ composite layers sampled from Ω and their inverses.

B. RB with non-Clifford randomized mirror circuits

We now introduce our protocol—MRB for universal gate sets. Our protocol has the same general structure as standard RB [15] and many of its variants: an exponential decay is fit to data from random circuits. However, our data analysis method is different from standard RB. We use the same analysis technique as MRB of *Clifford* gate sets [6]. In particular, for each n-qubit circuit C that we run, we estimate its *observed polarization* [6]

$$S = \frac{4^n}{4^n - 1}\left[\sum_{k=0}^{n}\left(-\frac{1}{2}\right)^k h_k\right] - \frac{1}{4^n - 1}, \tag{14}$$

where h_k is the probability that the circuit outputs a bit string with Hamming distance k from its target bit string (s_C). As
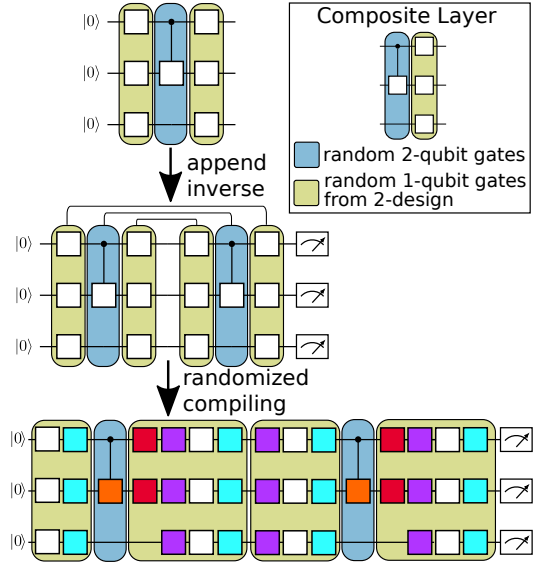

FIGURE 2. **Randomized mirror circuits over universal gate sets.** To construct a randomized mirror circuit of benchmark depth d (and total depth $2d + 2$) we first sample a random depth $d + 1$ circuit. This circuit alternates between layers of randomly sampled one-qubit gates and layers of randomly sampled two-qubit gates. It can be thought of as consisting of a single initial layer of random one-qubit gates followed by $d/2$ composite layers (see inset). We then append to this circuit its inverse, i.e., the circuit in reverse with each layer replaced with its inverse. This creates a depth $2d + 2$ circuit that will, if run perfectly, always return the all zeros bit string. This circuit is susceptible to systematic addition or cancellation of errors between the two halves of the circuit. To prevent this unwanted effect we then apply randomized compiling to the circuit. We insert a layer of random single-qubit Pauli gates (cyan) after each one-qubit gate layer. In order to guarantee that this randomly compiled circuit still always, if run perfectly, returns a single bit string s, our procedure (1) changes the rotation angles in the two-qubit gates (orange) if these gates are not Clifford gates, (2) adds in single-qubit Pauli axis rotations following the two-qubit gates (red) and, (3) adds in correction Pauli gates (purple) prior to each single-qubit gate layer. The yellow boxes show gates that are compiled together to create the final circuit of depth $2d + 2$. This circuit contains d composite layers, which we call its benchmark depth.

shown in Ref. [6] and discussed further below, the simple additional analysis in computing S simulates an n-qubit 2-design twirl using only local state preparation and measurement.

A specific MRB experiment is defined by a gate set $\mathbb{G}$, a sampling distribution Ω, and the usual RB sampling parameters (a set of benchmark depths d, the number of circuits K sampled per depth, and the number of times N each circuit is run). Our protocol is the following:

1. For a range of integers $d \geq 0$, sample K randomized mirror circuits that have a benchmark depth of d, using the sampling distribution Ω, and run each one $N \geq 1$ times.

2. Estimate each circuit's observed polarization S.

3. Fit $\bar{S}_d$, the mean of S at benchmark depth d, to

$$\bar{S}_d = Ap^d, \tag{15}$$

where A and p are fit parameters, and then compute

$$r_\Omega = (4^n - 1)(1 - p)/4^n. \tag{16}$$

Our claim is that r_Ω is a reliable estimate of the average error rate of Ω-distributed n-qubit circuit layers, i.e., $r_\Omega \approx \epsilon_\Omega$ under broad circumstances.

IV. THEORY AND SIMULATIONS OF MRB ON UNIVERSAL GATE SETS

In this section we present a theory for MRB of universal gate sets that shows that our method is reliable. We show that the average observed polarization ($\bar{S}_d$) decays exponentially, and that the MRB error rate (r_Ω) approximately equals the average error rate of Ω-distributed layers [ϵ_Ω, Eq. (8)]. In Section IV A we show that $r_\Omega \approx \epsilon_\Omega$ assuming Pauli stochastic error on each circuit layer. In Sections IV B and IV C we present theory and simulations of the performance of MRB under general Markovian errors to further validate our method. In particular, we show that the randomized compilation step of our circuit construction guarantees that all errors in the circuit are twirled into Pauli stochastic error (implying that $r_\Omega \approx \epsilon_\Omega$) under the assumption that all two-qubit gates are Clifford gates.

A. MRB with stochastic Pauli errors

We now show that $r_\Omega \approx \epsilon_\Omega$ under the assumption of stochastic Pauli error on each circuit layer. A more detailed derivation can be found in Appendix A. Throughout this section, we will treat circuits and circuit layers as random variables. We assume each circuit layer has gate-dependent Markovian error, $\phi(\mathsf{L}) = \mathcal{E}(\mathsf{L})\mathcal{U}(\mathsf{L})$. However, we assume that the error on the single-qubit gates is independent of the Pauli gates and Pauli-axis rotations inserted by randomized compiling, i.e., $\phi(\mathcal{R}(\mathsf{PLP}^c)) = \mathcal{E}(\mathsf{L})\mathcal{U}(\mathcal{R}(\mathsf{PLP}^c))$. This is implied by the stronger condition of gate-independent errors on single-qubit gates—which is an assumption commonly made in the theory of RB and related methods. We will model the error on state preparation and measurement (SPAM) and the first and last circuit layers of a randomized mirror circuit [$\mathcal{R}(P_0\mathsf{L}_0)$ and $\mathcal{R}(P_d\mathsf{L}_0^{-1}P_{d-1}^c)$, respectively] as a single global depolarizing error channel $\mathcal{E}_{\mathrm{SPAM}}[\rho] = \gamma_{\mathrm{SPAM}}\rho + (1 - \gamma_{\mathrm{SPAM}})\frac{\mathbb{I}}{2^n}$ occurring immediately before the final circuit layer. We assume $\mathcal{E}_{\mathrm{SPAM}}$ is independent of L_0 and the target bit string of the circuit.

We start by showing that the mean observed polarization [Eq. (14)] of randomized mirror circuits, which is measured in the MRB protocol, equals the mean polarization of the overall error map of a randomized mirror circuit. An implementation of the depth-d randomized mirror circuit M_d [whose structure is given in Eq. (13)] can be expressed in terms of its error and its target evolution $\mathcal{U}(\mathsf{P}_{d+1})$ as

$$\phi(\mathsf{M}_d) = \mathcal{U}(\mathsf{P}_{d+1})\mathcal{U}(\mathsf{L}_0^{-1})\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)\mathcal{U}(\mathsf{L}_0), \tag{17}$$

where

$$\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d) = \mathcal{E}_{\mathrm{SPAM}}\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_d) \tag{18}$$
$$= \mathcal{E}_{\mathrm{SPAM}}\mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1}^{-1}, \mathsf{P}_d)} \cdots \mathcal{E}'_{\mathsf{L}_{d/2}^{-1}}\mathcal{E}'_{\mathsf{L}_{d/2}} \cdots \mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1}, \mathsf{P}_0)} \tag{19}$$

and

$$\mathcal{E}'_{\mathsf{L}_i} = \mathcal{U}(\mathsf{L}_1)^{-1} \cdots \mathcal{U}(\mathsf{L}_i)^{-1}\mathcal{E}(\mathsf{L}_i)\mathcal{U}(\mathsf{L}_i) \cdots \mathcal{U}(\mathsf{L}_1). \tag{20}$$

Eq. (18) defines an overall error map for M_d, which includes the error from the $d/2$ Ω-distributed circuit layers and their inverses (after randomized compilation). To extract the polarization [Eq. (5)] of this error map, we average over the initial circuit layer L_0, making use of a fidelity estimation technique that requires only single-qubit gates: the fidelity of any error channel $\mathcal{E}$ can be found by averaging over a tensor product of single-qubit 2-designs [42]. In particular, for any bit string $y \in \{0,1\}^n$,

$$\gamma(\mathcal{E}) = \frac{4^n}{4^n - 1}\sum_{x \in \{0,1\}^n}(-1/2)^{h(x,y)}\langle\!\langle x + y|\bar{\mathcal{E}}|0\rangle\!\rangle - \frac{1}{4^n - 1}, \tag{21}$$

where $\bar{\mathcal{E}} = \mathbb{E}_{\mathsf{L}}[\mathcal{U}(\mathsf{L})^\dagger\mathcal{E}\mathcal{U}(\mathsf{L})]$ and $\mathsf{L} = \otimes_{i=1}^n \mathsf{L}_i$, where each L_i is a independent, single-qubit 2-design [42]. Applying Eq. (21) to Eq. (17), we find that

$$\gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)) = \mathbb{E}_{\mathsf{L}_0} S(\mathsf{M}_d) \tag{22}$$

where $S(\mathsf{M}_d)$ denotes the observed polarization [Eq. (14)] of M_d. Therefore, the mean observed polarization over all depth-d randomized mirror circuits is

$$\bar{S}_d = \mathbb{E}_{\mathsf{M}_d} \gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)). \tag{23}$$

Equation (23) says that the average observed polarization $\bar{S}_d$, which is estimated in the MRB protocol, is equal to the expected polarization of the error channel of a depth-d randomized mirror circuit.

We now show how $\bar{S}_d$ depends on the error rate of layers sampled from Ω (ϵ_Ω). To do so, we use the fact that a depth-d randomized mirror circuit consists of randomized compilation of a circuit consisting of a depth-$d/2$ Ω-distributed random circuit $\mathsf{C}_{d/2}$ followed by its inverse. These two depth-$d/2$ circuits are both Ω-distributed (even after randomized compilation), but they are correlated. In particular,

$$\bar{S}_d = \gamma(\mathcal{E}_{\mathrm{SPAM}}) \mathbb{E}_{\mathsf{C}_{d/2}} \gamma(\mathcal{U}(\mathsf{C}_{d/2})\bar{\mathcal{E}}_{\mathrm{eff}}(\mathsf{C}_{d/2}^{-1})\mathcal{U}(\mathsf{C}_{d/2})^{-1}\mathcal{E}_{\mathrm{eff}}(\mathsf{C}_{d/2})), \tag{24}$$

where $\mathcal{E}_{\mathrm{eff}}(\mathsf{C}_{d/2})$ is the overall error map for $\mathsf{C}_{d/2}$ [i.e., $\phi(\mathsf{C}_{d/2}) = \mathcal{U}(\mathsf{C}_{d/2})\mathcal{E}_{\mathrm{eff}}(\mathsf{C}_{d/2})$] and $\bar{\mathcal{E}}_{\mathrm{eff}}(\mathsf{C}_{d/2}^{-1})$ denotes the average error map over all possible circuits C' resulting from applying randomized compilation to $\mathsf{C}_{d/2}^{-1}$. Expressing Eq. (24) in terms of the mean observed polarization of the overall error map on an Ω-distributed random circuit, we have

$$\bar{S}_d = \gamma(\mathcal{E}_{\mathrm{SPAM}})\left(\bar{\Gamma}_{d/2}^2 - \Delta_\Omega\right), \tag{25}$$

Stochastic Errors **Stochastic and Hamiltonian Errors** **Hamiltonian Errors**
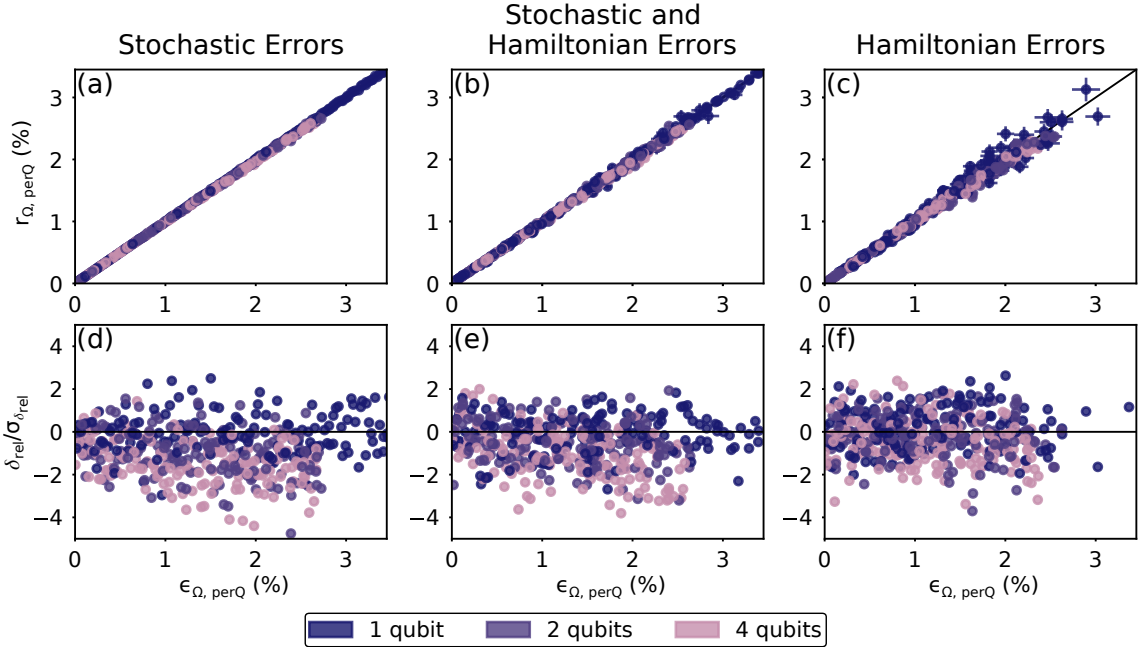
FIGURE 3. **Investigating the reliability of MRB using simulations.** We simulated MRB on n all-to-all-connected qubits for $n = 1, 2, 4$ on the gate set $(\mathbb{G}_1, \mathbb{G}_2) = (\mathbb{SU}(2), \{\mathsf{cs}, \mathsf{cs}^\dagger\})$ with randomly-sampled gate-dependent errors. From left to right, the columns show results from simulations with crosstalk-free error models consisting of only stochastic errors (a,c), a combination of stochastic and Hamiltonian errors (b,d), and only Hamiltonian errors (e,f) (see Section IV C for details). (a-c): The MRB error rate per qubit $[r_{\Omega, \mathrm{perQ}} = 1 - (1 - r_\Omega)^{1/n}]$ versus the average composite layer error rate per qubit $[\epsilon_{\Omega, \mathrm{perQ}} = 1 - (1 - \epsilon_\Omega)^{1/n}]$ for each randomly sampled error model. The MRB error rate r_Ω closely approximates ϵ_Ω, and the agreement is closest under purely stochastic errors. (d-f): The relative error $\delta_{\mathrm{rel}} = {}^{(r_{\Omega, \mathrm{perQ}} - \epsilon_{\Omega, \mathrm{perQ}})}/_{\epsilon_{\Omega, \mathrm{perQ}}}$ divided by its uncertainty $\sigma_{\delta_{\mathrm{rel}}}$ for each randomly sampled error model ($\sigma_{\delta_{\mathrm{rel}}}$ is calculated via a standard non-parametric bootstrap). The MRB error rate r_Ω is biased towards very slightly underestimating ϵ_Ω for $n > 2$ qubits, which is expected from our theory (see main text).

where

$$\bar{\Gamma}_d = \mathop{\mathbb{E}}_{\mathsf{C}_d} \gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{C}_d)) \qquad (26)$$

and

$$\Delta_\Omega = \mathop{\mathbb{E}}_{\mathsf{C}_{d/2}} \gamma\left(\mathcal{U}(\mathsf{C}_{d/2}) \bar{\mathcal{E}}_{\mathrm{eff}}(\mathsf{C}_{d/2}^{-1}) \mathcal{U}(\mathsf{C}_{d/2})^{-1} \mathcal{E}_{\mathrm{eff}}(\mathsf{C}_{d/2})\right)$$
$$- \left(\mathop{\mathbb{E}}_{\mathsf{C}_{d/2}} \gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{C}_{d/2}))\right)^2. \qquad (27)$$

Equation (25) shows that $\bar{S}_d \approx \gamma(\mathcal{E}_{\mathrm{SPAM}}) \bar{\Gamma}_{d/2}^2$ if $|\Delta_\Omega|$ is small. If $\bar{S}_d$ and $\bar{\Gamma}_{d/2}$ decay exponentially, Eq. (25) relates their decay rates—i.e., $r_\Omega = \epsilon_\Omega$ if $\Delta_\Omega = 0$. Δ_Ω quantifies the correlation between the overall error map of a depth-$d/2$ Ω-distributed random circuit and the overall error map of its randomly compiled inverse. We conjecture that $|\Delta_\Omega|$ is typically small for physically relevant errors, which is supported by our simulations (see Section IV C) and the theory below.

To show that $\bar{S}_d$ decays exponentially, we will assume that the error on each layer $\mathcal{E}(\mathsf{L})$ is a stochastic Pauli channel [Eq. (6)] (we consider the case of general Markovian errors in Section IV B). This assumption means that $\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_d)$ [Eq. (19)] is the composition of a stochastic Pauli channel for each circuit layer of $\tilde{\mathsf{M}}_d$, each rotated by a unitary. This allows us to relate the polarization of $\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_d)$ to the polarizations of the error channels of individual circuit layers. To first order in the

layer infidelities ε, the polarization of the effective error channel is equal to the product of the polarizations of the layers' error channels [42]:

$$\gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)) = \gamma(\mathcal{E}_{\mathrm{SPAM}}) \prod_{i=1}^{d/2} \Big[\gamma(\mathcal{E}(\mathsf{L}_i)) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}, \mathsf{P}_i))\big)$$
$$\gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}^{-1}, \mathsf{P}_{d-i+1}))\big) \Big] + O(d^2 \varepsilon^2). \quad (28)$$

By substituting Eq. (28) into Eq. (24) (see Appendix A), we find that $\bar{S}_d = A p^d + O(d^2 \varepsilon^2)$, with $A = \gamma(\mathcal{E}_{\mathrm{SPAM}})$ and p given by

$$p^2 = \mathop{\mathbb{E}}_{\mathsf{L}, \mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}^{-1})) \gamma(\mathcal{E}(\mathsf{L})) \gamma(\mathcal{E}(\mathsf{L}_\theta)) \bar{\gamma}(\mathsf{L}_\theta), \qquad (29)$$

where $\bar{\gamma}(\mathsf{L}_\theta) = \mathbb{E}_{\mathsf{P}'} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}'))\big)$ is the expected polarization of a two-qubit gate layer L_θ after inversion and randomized compiling. The $O(d^2 \varepsilon^2)$ term (and higher-order corrections) quantifies the rate that errors in different layers cancel, and it is negligible whenever the probability of two or more errors cancelling is negligible [42]. This is the case in highly scrambling circuits [5, 42] (when $n \gg 1$), and our circuits are highly scrambling by construction. Therefore $\bar{S}_d \approx A p^d$, i.e., the observed polarization of depth-d randomized mirror circuits decays exponentially with d.

We have shown that the observed polarization of randomized mirror circuits decays exponentially, and we now relate

its decay rate to the decay rate of the expected polarization of the overall error map of an Ω-distributed random circuit, thereby relating r_Ω and ϵ_Ω. An analogous argument to above implies that the expected polarization of the overall error map of an Ω-distributed random circuit [Eq. (26)] decays exponentially. By rewriting Eq. (26) as a product of layer polarizations, we find that the expected polarization of the overall error map is

$$\bar{\Gamma}_{d,\mathrm{rc}} = p_{\mathrm{rc}}^d + O(d^2\varepsilon^2), \tag{30}$$

where p_{rc}, which is given by

$$p_{\mathrm{rc}} = \mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}_i))\gamma(\mathcal{E}(\mathsf{L}_{\theta_i})), \tag{31}$$

sets the decay rate of the expected polarization of the overall error map on an Ω-distributed random circuit, and therefore determines ϵ_Ω. Substituting Eq. (31) into Eq. (29), we have $p^2 = p_{\mathrm{rc}}^2 + \Delta'_\Omega$, where

$$\Delta'_\Omega = \mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}))\gamma(\mathcal{E}(\mathsf{L}^{-1}))\gamma(\mathcal{E}(\mathsf{L}_\theta))\bar{\gamma}(\mathsf{L}_\theta) - p^2. \tag{32}$$

Therefore $p \approx p_{\mathrm{rc}}$, which implies that $r_\Omega \approx \epsilon_\Omega$, if $|\Delta'_\Omega|$ is small. Δ'_Ω quantifies the correlation of the polarization of the overall error map of a Ω-random composite layer with the average polarization of the overall error map of composite layers of the form $\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})\mathsf{L}^{-1}$ (averaged over all n-qubit Pauli layers P). When these quantities are perfectly uncorrelated ($\Delta'_\Omega = 0$), $p = p_{\mathrm{rc}}$, which implies $r_\Omega = \epsilon_\Omega$. This holds, for example, when errors are gate-independent. In practice, the polarizations will be somewhat correlated ($\Delta'_\Omega > 0$), implying that $p > p_{\mathrm{rc}}$, and so r_Ω will underestimate ϵ_Ω [44]. Prior work [6] and our simulations (see Section IV C) suggest that $|\Delta'_\Omega|$ is typically small for physically realistic errors, so r_Ω only slightly underestimates ϵ_Ω.

B. MRB with general errors

The theory presented above (Section IV A) shows that MRB is reliable whenever stochastic Pauli errors dominate over all other possible errors (e.g., coherent errors). In practice, stochastic error is not always dominant, which our protocol addresses with the randomized compilation step [see Fig. 2]. The purpose of this step is to, upon averaging, convert all types of errors into stochastic Pauli errors [43]—in which case the theory presented above can be used to infer that $r_\Omega \approx \epsilon_\Omega$. When MRB is implemented on a gate set in which all of the two-qubit gates are Clifford gates, this noise tailoring follows from standard randomized compilation theory [43]. In Appendix B, we show that with a Clifford two-qubit gate set, the error in MRB circuits is twirled into Pauli stochastic noise under the assumption that the error map on the single-qubit gates is independent of the Pauli gates with which they are compiled. In actual devices it is common for the single-qubit gate layers to have errors that are gate-dependent but much smaller than the two-qubit gate errors, in which case this result holds approximately [43].

Our MRB protocol can be applied to all controlled rotations around Pauli axes, i.e., all CP_θ gates. When the two-qubit gates are not all Clifford gates (i.e., when $\theta \neq 0, \pi$), the randomized compilation method used in our circuits is not equivalent to standard randomized compilation. In this case, we cannot use standard randomized compilation theory to guarantee that all coherent errors on the two-qubit gates are twirled into stochastic Pauli errors. Ineffective twirling of coherent errors on two-qubit gates could result in coherent cancellation of the errors in a layer of two-qubit gates and its inversion layer in the second half of the mirror circuit (as happens in a simple mirror circuit, or standard Loschmidt echo [7]). In Appendix C 1 we prove that our randomized compilation method largely—but not entirely—prevents this error cancellation. We consider the sensitivity of our method to general Hamiltonian errors on each gate $g \in \mathbb{G}_2$. We model these errors by an error map $\mathcal{E}(g) = e^{M_g}$, where

$$M_g = \sum_{P_a, P_b} \varepsilon^g_{P_a, P_b} H_{P_a, P_b}, \tag{33}$$

and H_{P_a, P_b} is the two-qubit Hamiltonian error generator indexed by the Pauli operators P_a and P_b, as defined in Ref. [45]. We show that r_Ω depends on all Hamiltonian errors in CP_θ gates *except* one particular linear combination of the Hamiltonian errors on CP_θ and $CP_{-\theta}$ gates, when $\theta \neq 0, \pi$ (i.e., when CP_θ is not a Clifford gate). In particular, r_Ω is insensitive (at first order) to $\varepsilon^{CP_\theta}_{P,P} + \varepsilon^{CP_{-\theta}}_{P,P}$ when $\theta \neq 0, \pi$. This is the sum of over- and under-rotation Hamiltonian errors in the CP_θ gate and its inverse. In Appendix C 2 we discuss how our technique could be adapted to remove this limitation. Note that if $\mathbb{G}_2 = \{\mathsf{cs}, \mathsf{cs}^\dagger\}$, as is the case in our simulations (below) and some of our experiments (Section V), then r_Ω is insensitive (at first order) to $\varepsilon^{\mathsf{cs}}_{ZZ} + \varepsilon^{\mathsf{cs}^\dagger}_{ZZ}$. However, it is sensitive to all other linear combinations of the Hamiltonian errors on the cs and $\mathsf{cs}^\dagger$ gates.

C. Simulations

We now use numerical simulations to investigate the robustness of MRB, studying whether the MRB error rate (r_Ω) closely approximates the error rate of Ω-distributed layers (ϵ_Ω). Our theory for MRB suggests that MRB is particularly robust when the two-qubit gates are Clifford gates and when all errors are stochastic Pauli errors. Therefore we simulated MRB with non-Clifford two-qubit gates and for both stochastic and coherent errors. We simulated MRB for n-qubit layer sets constructed from the gate set $\mathbb{G}_1 = \mathbb{SU}(2)$ and $\mathbb{G}_2 = \{\mathsf{cs}, \mathsf{cs}^\dagger\}$ and $n = 1, 2, 4$, with all-to-all connectivity. We used a sampling distribution Ω_2 for which the two-qubit gate density is $\xi = 1/2$ [46]. In these simulations (and our experiments) each single-qubit gate is decomposed into the following sequence of $\mathsf{x}_{\pi/2}$ and z_θ gates:

$$\mathsf{u}(\theta, \phi, \lambda) = \mathsf{z}_{-\phi-\pi/2}\,\mathsf{x}_{\pi/2}\,\mathsf{z}_{\pi-2\theta}\,\mathsf{x}_{\pi/2}\,\mathsf{z}_{-\lambda+\pi/2}. \tag{34}$$

Here $\mathsf{x}_{\pi/2}$ is a $\pi/2$ rotation around the X axis and z_θ is a rotation around the Z axis by $\theta \in [0, 2\pi)$. Note that even when a

shorter sequence of gates can implement the required unitary (e.g., $u(0, 0, 0)$ implements the identity so it could be implemented with no gates) we always use this sequence of five gates. Therefore, the only difference between any two single-qubit gates is the angles of the z_θ gates.

We simulated three different families of error model: stochastic Pauli errors, Hamiltonian errors, and stochastic and Hamiltonian errors. These error models are specified using the error generator framework of Ref. [45], and they consist of gate-dependent errors specified by randomly sampling error rates for each type of error and each gate. We simulated error models that are crosstalk free (note that our theory encompasses crosstalk errors) so each error model is specified by the rates of each type of local error on each gate. In particular, for an m-qubit gate we randomly sample $4^m - 1$ stochastic error generators, or $4^m - 1$ Hamiltonian error generators, or both, depending on the error model family. We sampled the error rates so that the infidelity of each two-qubit gate was approximately q, and the infidelity of each one-qubit gate was approximately $0.1q$, where q is a parameter swept over a range of values (See Appendix D).

Figure 3 shows the results of these simulations. It compares the true average layer error rate per qubit

$$\epsilon_{\Omega,\,\text{perQ}} = 1 - (1 - \epsilon_\Omega)^{1/n} \approx \epsilon_\Omega/n \qquad (35)$$

to the observed MRB error rate per qubit

$$r_{\Omega,\,\text{perQ}} = 1 - (1 - r_\Omega)^{1/n} \approx r_\Omega/n \qquad (36)$$

in each simulation, separated into the three families of error model (1σ error bars are shown, computed using a standard bootstrap). Figure 3(a)-(c) shows that $r_\Omega \approx \epsilon_\Omega$ in every simulation, which means that our method closely approximates the error rate of Ω-distributed layers for all of these error models.

For stochastic error models [Fig. 3 (a)], the relative error $\delta_{\text{rel}} = (r_{\Omega,\,\text{perQ}} - \epsilon_{\Omega,\,\text{perQ}})/\epsilon_{\Omega,\,\text{perQ}}$ in the MRB estimate of $\epsilon_{\Omega,\text{perQ}}$ is small: $|\delta_{\text{rel}}| < 0.04$ and the mean $|\delta_{\text{rel}}|$ is 0.007 for all sampled error models. This is consistent with, and supports, our theory for MRB with stochastic errors. The relative error is larger for Hamiltonian error models—the mean relative error is 0.04 and $|\delta_{\text{rel}}| < 0.21$ for all error models. We expect larger relative error for some Hamiltonian error models, because MRB is insensitive to some Hamiltonian errors (see Section IV B)—but note that the uncertainty due to finite sample fluctuations (σ) are larger in these simulations. For stochastic Pauli errors [Fig. 3 (a)], the uncertainty in $r_{\Omega,\text{perQ}}$ is small, because there is little variation in the performance of circuits of the same depth (the mean uncertainty in $r_{\Omega,\text{perQ}}$ is 0.5%). For Hamiltonian errors [Fig. 3 (c)], the uncertainty in $r_{\Omega,\text{perQ}}$ is larger (the mean uncertainty is 3%), as individual circuit performance varies widely due to coherent addition or cancellation of error being highly dependent on the circuit structure (as in all RB methods, we expect coherent errors to add or cancel in individual MRB circuits).

Arguably the most relevant simulations for real-world quantum computers are those with both stochastic and coherent errors [Fig. 3 (b)]. In these simulations we sampled random combinations of stochastic and Hamiltonian errors (so the dominant source of error varies across these models). We find that $r_\Omega \approx \epsilon_\Omega$ holds to a good approximation for typical error models sampled from this ensemble (the mean relative error is 0.017, and $|\delta_{\text{rel}}| < 0.11$ for all models, and the mean uncertainty in $r_{\Omega,\text{perQ}}$ is 1.4%).

To investigate whether there is evidence for r_Ω systematically under (or over) estimating ϵ_Ω we plot the relative error divided by its uncertainty $\sigma_{\delta_{\text{rel}}}$ [Fig. 3 (d-f)]. For $n = 1$ qubit, there is no evidence that MRB is significantly biased towards under or overestimating ϵ_Ω with these error models. In contrast, we find that MRB slightly but systematically underestimates ϵ_Ω for $n > 1$ qubits. This underestimate can be explained by the correlation between the error in an Ω-distributed circuit and its randomly-compiled inverse, which determines the difference between r_Ω and ϵ_Ω (see Section IV A). When the circuits contain two-qubit gates—which in our simulations (and in most real systems) have higher error rates than one-qubit gates—the error in a circuit is typically highly correlated with the number of two-qubit gates in the circuit. As a result, the correlation between a circuit and its randomly-compiled inverse is typically larger when the circuits contain a variable number of two-qubit gates, causing r_Ω to *slightly* underestimate ϵ_Ω.

V. EXPERIMENTS ON THE ADVANCED QUANTUM TESTBED

We used MRB to benchmark universal gate sets on the Advanced Quantum Testbed (AQT) [33], a quantum computing testbed platform based on superconducting qubits. We performed our experiments on four qubits (Q4-Q7) of an eight-qubit superconducting transmon processor (`AQT@LBNL Trailblazer8-v5.c2`). These four qubits are coupled to their nearest neighbors in a linear geometry (see Fig. 4). Below and throughout this paper, estimated quantities include error bars where possible [47]. All error bars are 1σ and are written using standard concise notation, i.e., $r = 1.2(3)\%$ means $r = 1.2\%$ with a standard error of 0.3%.

A. Experiment design

One of the advantages of MRB is that it can benchmark a wide variety of n-qubit layer sets, and we used this flexibility to explore the performance of three distinct layer sets on AQT. Each layer set is defined by a set of single-qubit gates $\mathbb{G}_1$, a set of two-qubit gates $\mathbb{G}_2$, a two-qubit gate density ξ, and the connectivity of the qubit subset (see Section II). In our experiments we investigated three different choices for $(\mathbb{G}_1, \mathbb{G}_2)$: $(\mathbb{SU}(2), \{cs, cs^\dagger\})$, $(\mathbb{SU}(2), \{cz\})$, and $(\mathbb{C}_1, \{cz\})$, where $\mathbb{C}_1$ is the set of all 24 single-qubit Clifford gates.

MRB enables benchmarking each layer set on any connected set of qubits, and the error rates on subsets of a device can be used to learn about the location and type of errors. We benchmarked n-qubit layer sets for every possible connected set $\mathbb{Q} \subseteq \{Q4, Q5, Q6, Q7\}$ of n qubits with $n = 1, 2, 3, 4$, resulting in 10 different qubit subsets. Independently benchmark-
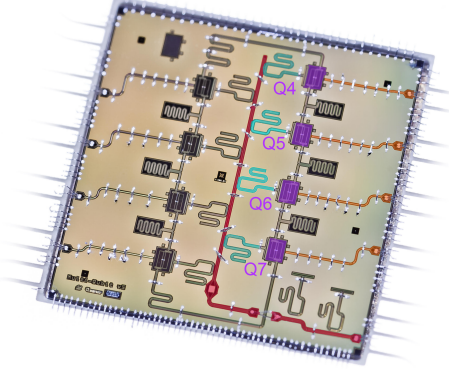
FIGURE 4. **The Advanced Quantum Testbed.** We performed MRB experiments on four qubits (Q4-Q7) of AQT's eight-qubit superconducting transmon processor (`AQT@LBNL Trailblazer8-v5.c2`). The processor includes 8 fixed frequency transmons coupled in a ring geometry. Each qubit (purple) has its own control line (orange) and readout resonator (cyan) coupled to a shared readout bus (red) for multiplexed readout.

ing every connected subset of qubits allows us to study the spatial variation in gate performance in detail and determine the size of crosstalk error on circuits with 3 and 4 qubits (see Section V C).

For each of the three gate sets $(\mathbb{G}_1, \mathbb{G}_2)$, and each qubit subset $\mathbb{Q}$, we ran experiments with a two-qubit gate density of $\xi = 1/2$. To investigate the effect of varying ξ, we also ran experiments with $\xi = 1/8$ for one of the gate sets—$(\mathbb{SU}(2), \{cs, cs^\dagger\})$—and every $\mathbb{Q}$. For each qubit subset we therefore ran 4 MRB experiments, defined by [48]:

1. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{cs, cs^\dagger\}$, and $\xi = 1/8$.

2. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{cs, cs^\dagger\}$, and $\xi = 1/2$.

3. $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{cz\}$, and $\xi = 1/2$.

4. $\mathbb{G}_1 = \mathbb{C}_1$, $\mathbb{G}_2 = \{cz\}$, and $\xi = 1/2$.

Further experiment details are provided in Appendix E.

B. Estimating average error rates of universal layer sets

Figure 5 summarizes the results of the 3×10 MRB experiments in which we vary the gate set $(\mathbb{G}_1, \mathbb{G}_2)$—corresponding to each row of Fig. 5—and the subset of qubits benchmarked $\mathbb{Q}$, but we keep the expected two-qubit gate density constant ($\xi = 1/2$). The main output of an MRB experiment is an average layer error rate (r_Ω), obtained by fitting the mean observed polarization [$\bar{S}_d$, defined in Eq. (14)] to an exponential decay. This error rate is a function of $(\mathbb{G}_1, \mathbb{G}_2, \mathbb{Q}, \xi)$, so we denote our estimated error rates by $r(\mathbb{G}_1, \mathbb{G}_2, \mathbb{Q}, \xi)$ whenever we need to refer to a particular error rate. These error rates quantify the performance of random circuits on this device and enable us to compare the average performance of the gate sets we tested.

Figure 5 (a-c) shows MRB data and fits to an exponential, for each of the three gate sets and $\xi = 1/2$. For each MRB experiment, we show the mean observed polarization ($\bar{S}_d$) versus benchmark depth, the distribution of the observed polarization versus benchmark depth, and the fit of $\bar{S}_d$ to $\bar{S}_d = Ap^d$. Data for a single representative subset of qubits of each size ($n = 1, 2, 3, 4$) are shown. In all cases, we observe that $\bar{S}_d$ is consistent with an exponential decay in d, providing experimental evidence for our claim that $\bar{S}_d$ will decay exponentially under a broad range of conditions.

Figure 5 (d-f) shows the estimated error rates (r_Ω) for each qubit subset that we benchmarked, for each of the three different gate sets. Each r_Ω is a rescaling of the decay rate of the fitted exponential [see Eq. (15)]. By comparing Fig. 5 (d), (e) and (f) we can compare the average error rates of n-qubit layers constructed from three different gate sets, two of which are universal and one of which contains only Clifford gates and therefore is not. By comparing (e) and (f), we find that the average error rate of a layer set is approximately independent of whether single-qubit gates are sampled from $\mathbb{SU}(2)$ or from $\mathbb{C}_1$ (the single-qubit Clifford group)—that is, $r(\mathbb{SU}(2), \{cz\}, \mathbb{Q}, 1/2) \approx r(\mathbb{C}_1, \{cz\}, \mathbb{Q}, 1/2)$ for all ten subsets of qubits $\mathbb{Q}$. All single-qubit gates in our experiments are implemented using a composite $u(\theta, \phi, \lambda)$ gate [see Eq. (34)] that contains two $x_{\pi/2}$ gates and three z_θ gates. This is the case even for unitaries that do not require two $x_{\pi/2}$ pulses, such as the identity. The difference between any two single-qubit gates is therefore only in the angles of the three z_θ gates within $u(\theta, \phi, \lambda)$. These gates are implemented by in-software phase updates on later pulses [49], so it is expected that these "virtual gates" cause negligible errors. The observed similarity between the average performance of these two gate sets is consistent with this expectation (numerical values for all estimated r_Ω are included in Table I). Note, however, that the observed similarity between the average success rates of circuits in which the single qubit-gate gates $u(\theta, \phi, \lambda)$ are sampled from two different distributions does *not* imply that the success rate of an individual circuit is independent of the values of θ, ϕ and λ in its $u(\theta, \phi, \lambda)$ gates — see Appendix E 2 for further discussions.

Our experiments included MRB on n-qubit layers containing two non-Clifford two-qubit gates—cs and $cs^\dagger$—and we now turn to these results. Comparing Figs. 5 (d) and (f), we observe that the error rates for layers containing cs and $cs^\dagger$ gates are all almost equal to, but slightly larger than, the error rates for layers containing cz gates. The largest relative difference is in the experiments on the 3-qubit set $\{Q4, Q5, Q6\}$: $r(\mathbb{SU}(2), \{cs, cs^\dagger\}, \{Q4, Q5, Q6\}, 1/2) = 1.64(5)\%$ and $r(\mathbb{SU}(2), \{cz\}, \{Q4, Q5, Q6\}, 1/2) = 1.48(4)\%$. The three different two-qubit gates (cs, $cs^\dagger$, and cz) on each qubit pair were *a priori* expected to have similar error rates, due to their similar calibration procedures. The slightly larger error rates for cs and $cs^\dagger$ were cross-validated using cycle benchmarking [9] (see Section V D for a quantitative comparison). Therefore, these results are experimental evidence for the robustness of MRB with non-Clifford two-qubit gates (see Sections IV B and IV C for discussion of and theory for MRB of non-Clifford two-qubit gates).
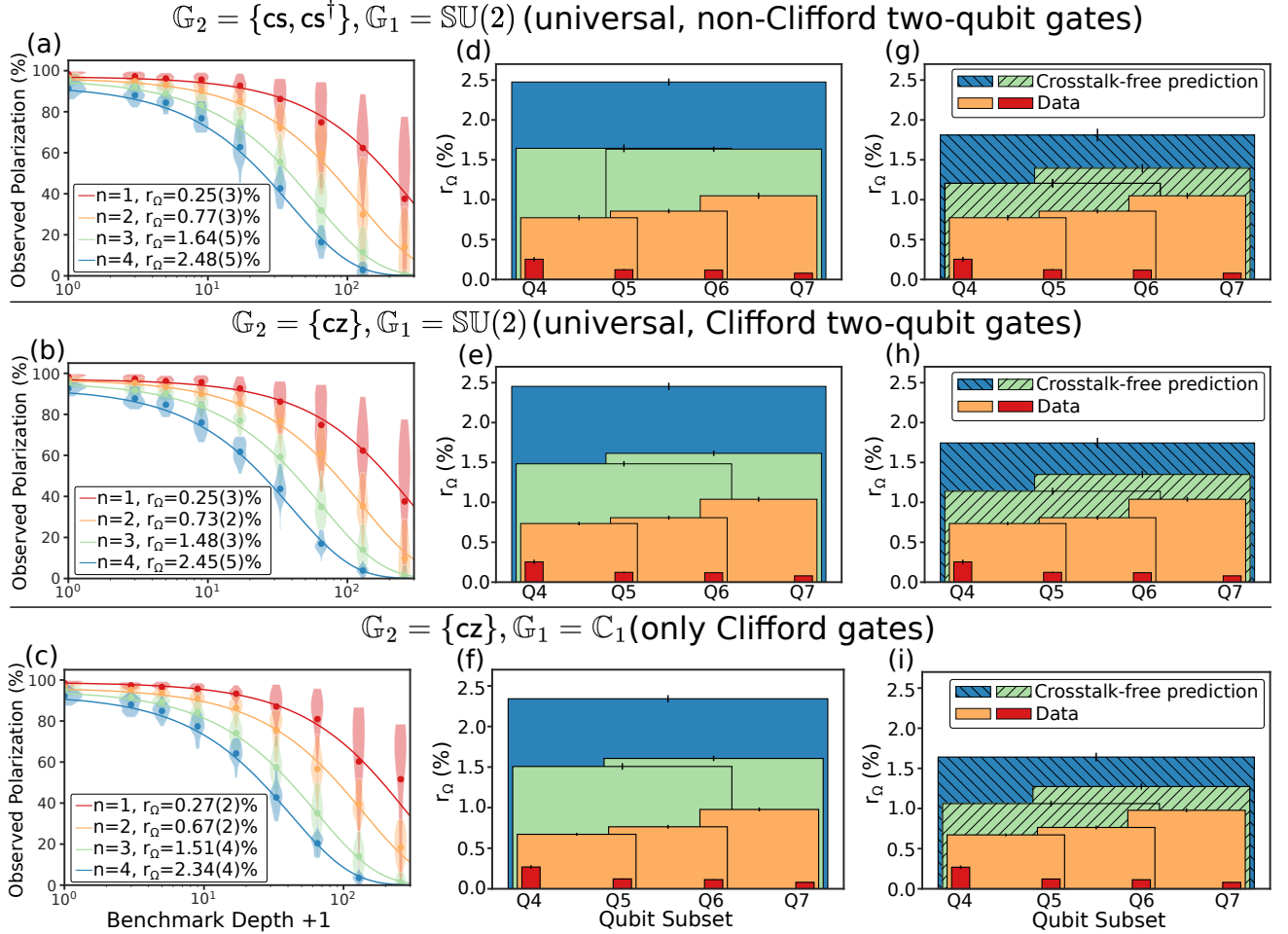
FIGURE 5. **Randomized benchmarking of universal gate sets on four qubits of the Advanced Quantum Testbed.** We used MRB to benchmark n-qubit layers constructed from three different gate sets, on each connected n-qubit subset of a linearly-connected set of four qubits $\{Q4, Q5, Q6, Q7\}$ in an eight-qubit superconducting transmon processor (`AQT@LBNL Trailblazer8-v5.c2`). The rows correspond to results from three different choices of gate set, each consisting of a two-qubit gate set $\mathbb{G}_2$ and a single-qubit gate set $\mathbb{G}_1$. From top to bottom, the rows correspond to: a universal gate set containing two non-Clifford entangling gates and the set of all single-qubit gates $[\mathbb{G}_2 = \{cs, cs^\dagger\}$, $\mathbb{G}_1 = \mathbb{SU}(2)]$; a universal gate set containing a Clifford entangling gate and the set of all single-qubit gates $[\mathbb{G}_2 = \{cz\}, \mathbb{G}_1 = \mathbb{SU}(2)]$; and a non-universal, Clifford gate set $[\mathbb{G}_2 = \{cz\}, \mathbb{G}_1 = \mathbb{C}_1$ where $\mathbb{C}_1$ is the one-qubit Clifford group]. (a-c): MRB decays for the qubit subsets $\{Q4\}$, $\{Q4, Q5\}$, $\{Q4, Q5, Q6\}$, and $\{Q4, Q5, Q6, Q7\}$. Violin plots and points show the distribution and mean, respectively, of the MRB circuit's observed polarization (S_d) versus benchmark depth (d). The curve is a fit of the mean of S_d ($\bar{S}_d$) to $\bar{S}_d = Ap^d$. The average error rate of an n-qubit layer (r_Ω) is given by $r_\Omega = (4^n - 1)(1 - p)/4^n$. The observed $\bar{S}_d$ decays exponentially, as predicted by our theory for MRB. (d-f): The estimated error rate r_Ω for each qubit subset that we benchmarked. (g-i): Predictions for the average layer error rate of 3- and 4-qubit subsets (hatched) based on the experimental 1- and 2-qubit error rates (un-hatched) and the assumption of no crosstalk errors. The difference between (d-f) and (g-i) quantifies the contribution of crosstalk errors to the average error rate of an n-qubit layer, for $n = 3, 4$. For all three gate sets and $n = 4$, we see that crosstalk errors are contributing approximately 0.7% error to r_Ω, which is approximately $^1/_3$ of r_Ω.

C.   Estimating crosstalk errors

Crosstalk is an important type of error in current quantum processors, but it is challenging to quantify [4]. Multi-qubit MRB captures crosstalk errors, and it enables us to quantify the contribution of crosstalk errors to the average error rate of n-qubit layers. To do so, we compare the observed increase in r_Ω with n [Fig. 5 (d-f)] to predictions for r_Ω that assume no crosstalk errors. The excess observed error above these predictions is then attributed to crosstalk.

We predict r_Ω for sets of three or more qubits from the observed r_Ω values for each one- and two-qubit subset (note, however, that this is not the only possible way for predicting r_Ω). This prediction is built on a simple theory for MRB. We model r_Ω by

$$r_\Omega = \sum_{L \in \mathbb{L}} \Omega(L)\epsilon_L, \qquad (37)$$

where ϵ_L is the infidelity of a $\mathbb{G}_1$-*dressed layer* L, which consists of a specific layer of two-qubit gates—i.e., L is labelled by the two-qubit gate layer—followed by a layer of random
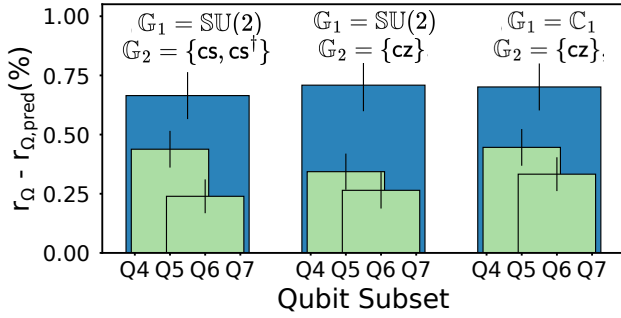
FIGURE 6. **Estimating crosstalk errors on AQT.** We estimate the contribution of crosstalk errors to the layer error rate r_Ω for $n = 3, 4$ qubits by taking the difference between each experimental error rate (r_Ω) and a corresponding prediction ($r_{\Omega,\text{pred}}$) obtained from the experimental one- and two-qubit error rates and the assumption of no crosstalk. We find that crosstalk contributes approximately $0.2\% - 0.4\%$ to r_Ω for $n = 3$ (which is $1/8 - 1/4$ of r_Ω), and approximately 0.7% to r_Ω for $n = 4$ (which is $1/3$ of r_Ω).

single-qubit gates (either from $\mathbb{SU}(2)$ or $\mathbb{C}_1$). Equation (37) is justified by our theory for MRB (see Section IV), but note that it only holds approximately, unless each layer's error channel is an n-qubit depolarizing channel. The fidelity $F = 1 - \epsilon$ of a tensor product of channels is the product of those channels' fidelities. So, under the assumption that there are no crosstalk errors, the infidelity of L is given by $\epsilon_L = \prod_{g \in L} F_g$, where g are the $\mathbb{G}_1$-dressed gates in the $\mathbb{G}_1$-dressed layer L, and F_g is the fidelity of g. Therefore,

$$\epsilon_L = 1 - \prod_{g \in L}(1 - \epsilon_g), \qquad (38)$$

where $\epsilon_g = 1 - F_g$.

To predict ϵ_L using Eq. (38) [and then r_Ω using Eq. (37)] we need estimates for ϵ_g for every possible $\mathbb{G}_1$-dressed gate g. That is, we need estimates for (1) $\epsilon_{\text{idle(Qi)}}$ for each qubit $\text{Qi} \in \{\text{Q4}, \text{Q5}, \text{Q6}, \text{Q7}\}$ where idle(Qi) is the $\mathbb{G}_1$-dressed idle gate on Qi, and (2) $\epsilon_{g(\text{Qi},\text{Qj})}$ for each connected pair of qubits $\{\text{Qi}, \text{Qj}\}$ where $g(\text{Qi},\text{Qj})$ is a two-qubit gate on $\{\text{Qi}, \text{Qj}\}$ uniformly sampled from $\mathbb{G}_2$. Each of these quantities can be estimated from the observed one- and two-qubit MRB error rates. Using Eq. (37) we have

$$r(\mathbb{G}_1, \{\text{Qi}\}) = \epsilon_{\text{idle(Qi)}}, \qquad (39)$$

because each single-qubit MRB circuit simply consists of repeating the $\mathbb{G}_1$-dressed idle gate. Similarly, using Eq. (37) we have

$$r(\mathbb{G}_1, \mathbb{G}_2, \{\text{Qi}, \text{Qj}\}, \xi) = \xi\epsilon_{g(\text{Qi},\text{Qj})} + (1 - \xi)(1 - \epsilon_{\text{idle(Qi)}})(1 - \epsilon_{\text{idle(Qj)}}), \quad (40)$$

because each $\mathbb{G}_1$-dressed layer in a two-qubit MRB circuit is either (with probability ξ) a $\mathbb{G}_1$-dressed two-qubit gate sampled uniformly at random from $\mathbb{G}_2$, or a $\mathbb{G}_1$-dressed idle on each qubit (with probability $1 - \xi$).

Using Eqs. (38)–(40) and explicit expressions for $\Omega(L)$, we obtain analytic expressions for our crosstalk-free predictions of r_Ω for the 3- and 4-qubit layers. These predictions
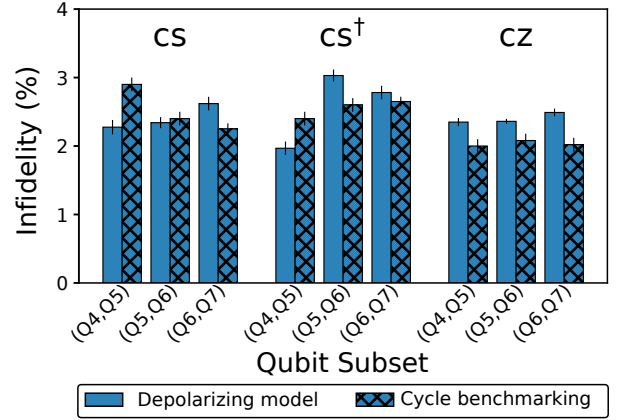

FIGURE 7. **Estimating the infidelity of dressed 4-qubit layers.** Estimates of the error rates of individual $\mathbb{G}_1$-dressed layers containing a single 2-qubit gate (cs, $\text{cs}^\dagger$, or cz), obtained by fitting an n-qubit depolarizing model to the 4-qubit MRB data. This scalable analysis technique enables extraction of additional information about each layer's error from MRB data. To validate our results against an established technique, we compare to infidelities independently estimated using cycle benchmarking [9]. We observe qualitative agreement. The cycle benchmarking experiments measure the infidelities of layers dressed with one-qubit gates sampled from a different gate set (the Pauli group) to those used in our MRB experiments, so exact agreement is not expected.

are shown in Fig. 5 (g-i). The crosstalk-free predictions are significantly smaller than the observed experimental values, shown in Fig. 5 (d-f). For each gate set, the predicted 4-qubit r_Ω is approximately 25% smaller than the observed value. The crosstalk-free predictions for $\{\text{Q4}, \text{Q5}, \text{Q6}\}$ are 13%–19% smaller than their observed values, and the crosstalk-free predictions for $\{\text{Q5}, \text{Q6}, \text{Q7}\}$ are 20%–27% smaller than their observed values. The difference between the experimental error rates and the crosstalk free predictions, shown in Fig. 6, is a quantification of the contribution of crosstalk errors to the average rate of errors in 3- and 4-qubit random circuits in this system [50].

D. Estimating the error rates of individual gates

An MRB experiment is primarily designed to estimate a single error rate (r_Ω) that quantifies the average error rate of an n-qubit layer. However, it is also often useful to quantify the error in specific layers, e.g., to identify high-error gates. Information about the error rates of individual layers is contained within the MRB data (e.g., RB data can even be used for full tomography [51, 52]), and we extract it using a scalable model fitting method. Specifically, we fit a 4-qubit depolarizing error model to the 4-qubit MRB data to estimate the error rates of individual $\mathbb{G}_1$-dressed layers [Fig. 7]. To validate our results, we compare the infidelities we estimate to independent estimates obtained from an established technique: cycle benchmarking [9], which is a method for estimating the infidelity of individual many-qubit gate layers. Fig. 7 shows that our
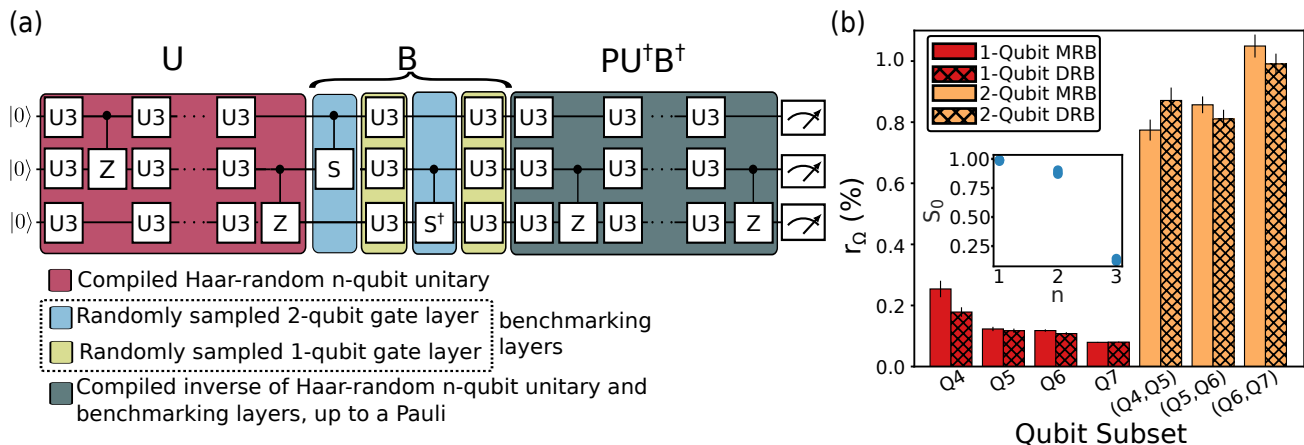
FIGURE 8. **Validating MRB by comparison to DRB.** (a) The structure of DRB circuits, which is a method for benchmarking an n-qubit layer set, when applied to a universal gate set. DRB is known to be reliable, but it is exponentially expensive in n for universal gate sets—because, for a universal gate set, its circuits start by implementing a Haar random unitary from $\mathbb{SU}(2^n)$. (b) The error rates obtained when running equivalent DRB and MRB experiments, on every one- and two-qubit subset of the 4 qubits we benchmarked on AQT. The close agreement between the DRB and MRB error rates is experimental evidence that MRB is reliable. The inset shows the polarization at benchmark depth $d = 0$ (S_0) for n-qubit DRB with $n = 1, 2, 3$. The rapid decay in S_0 is due to the overhead in implementing a Haar-random unitary, and it makes DRB of universal gate sets infeasible on more than around 2–3 qubits.

estimates are broadly similar to the those obtained from cycle benchmarking, differing by at most 23%, and note that we would not expect exact agreement [53]. This demonstrates the potential of MRB to go beyond average error rate estimation, and provides an alternative to, e.g., interleaved RB. In Appendix E 3 we discuss the depolarizing model fit as well as two additional methods for estimating the error rate of individual layers from MRB data, and compare their predictions.

E. Comparison to direct RB

One of the purposes of our experiments is to test the reliability of MRB. To investigate whether $r_\Omega \approx \epsilon_\Omega$ in experiment (as claimed by our theory), we compare the results of MRB to an alternative, established RB technique: direct RB (DRB) [5]. DRB is a streamlined variant of standard RB. Both DRB and standard RB are inefficient when applied to universal gate sets—as they have costs that scale exponentially with the number of qubits—but they are feasible in the very few qubit regime. We chose to compare MRB to DRB because these two methods have the same flexible circuit sampling and they are designed to measure the same error rate: ϵ_Ω. In contrast, standard RB benchmarks a gate set that forms a group, e.g., $\mathbb{SU}(2^n)$, and it measures an error rate for a uniformly random element of that group—so this error rate cannot be directly compared to r_Ω.

An n-qubit, benchmark depth d DRB circuit for a universal layer set is constructed by first sampling a depth-d circuit C with layers sampled from some distribution Ω—exactly as with MRB. As shown in Fig. 8 (a), this circuit C is then embedded between (1) a circuit that implements an n-qubit Haar random unitary, and (2) a circuit that returns the qubits to the computational basis. Note that both (1) and (2) require circuits

of one- and two-qubit gates whose size grows exponentially in n (we compile a $\mathbb{SU}(2^n)$ unitary into a circuit of $\mathsf{x}_{\pi/2}$, z_θ and cz gates using the Qsearch package [54, 55]). We therefore ran DRB on all n-qubit subsets only up to $n = 3$.

In our DRB experiments we used the same layer sampling distribution as in our $\mathbb{G}_1 = \mathbb{SU}(2)$, $\mathbb{G}_2 = \{\mathsf{cs}, \mathsf{cs}^\dagger\}$, and $\xi = 1/2$ MRB experiments. So the DRB error rates we are measuring—which we denote by $r_{\mathrm{DRB}}(\mathbb{SU}(2), \{\mathsf{cs}, \mathsf{cs}^\dagger\}, \mathbb{Q}, 1/2)$ for qubit subset $\mathbb{Q}$—will be equal to the equivalent MRB error rates $r(\mathbb{SU}(2), \{\mathsf{cs}, \mathsf{cs}^\dagger\}, \mathbb{Q}, 1/2)$ if both DRB and MRB are working correctly. Figure 8 compares these DRB and MRB error rates for each one- and two-qubit subset. For each of these qubit subsets, the two error rates differ by no more than 2σ. Due to the overhead in implementing a Haar-random unitary from $\mathbb{SU}(2^n)$, the 3-qubit DRB circuits were so large that the polarization of all $n = 3$ DRB circuits was $S_d \approx 0$, even for the $d = 0$ circuits, so we were not able to obtain reliable estimates of r_{DRB} for either 3-qubit subset. The rapid decrease in the $d = 0$ polarization (S_0) with increasing n is shown in the inset of Fig. 8 (b). This demonstrates that DRB cannot be used to benchmark universal gate sets on more than around 2-3 qubits (and note that standard RB requires running even larger circuits than those used in DRB).

VI. 27-QUBIT IBM Q EXPERIMENTS

To investigate what many-qubit MRB can reveal about errors in current many-qubit hardware, we ran MRB on a 27-qubit IBM Q device (ibmq_montreal, a Falcon r4 processor). We used the universal gate set $\mathbb{G}_1 = \mathbb{SU}(2)$ and $\mathbb{G}_2 = \{\mathsf{cnot}\}$, and we sampled layers with a two-qubit gate density of $\xi = 1/4$. We choose a single qubit subset $\mathbb{Q}$ containing n qubits for 15 exponentially spaced n up to $n = 27$. This
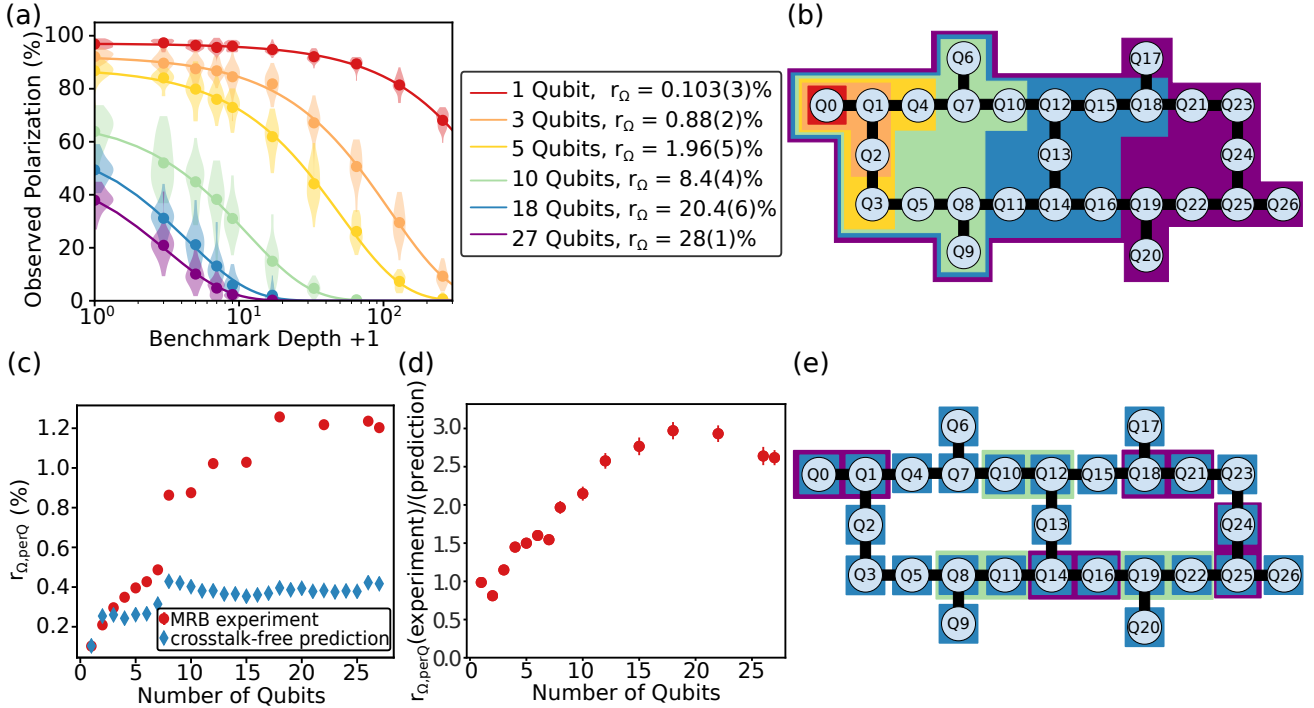
FIGURE 9. **Randomized benchmarking of a universal gate set on a 27-qubit IBM Q processor.** We ran MRB on n-qubit subsets of the `ibmq_montreal` processor, for 15 exponentially spaced n from $n = 1$ to $n = 27$. (a) The MRB decays and fits to an exponential for the six subsets of qubits illustrated in (b). The observed polarization decays exponentially in all cases. Due to the minimal overhead in MRB circuits, we obtain an exponential decay even for 27 qubits and can extract a low-uncertainty estimate of the average error rate of 27-qubit layers [$r_\Omega = 28(1)\%$]. (c) The observed error rate per qubit $r_{\Omega,\text{perQ}} = 1 - (1 - r_\Omega)^{1/n} \approx r_\Omega/n$ (red circles) versus n increases rapidly with n, even though the circuits have a constant expected two-qubit gate density $\xi = 1/4$. This increase in $r_{\Omega,\text{perQ}}$ is due to two-qubit gate crosstalk, not spatial variations in gate error rates. This is confirmed by comparison to predictions for $r_{\Omega,\text{perQ}}$ (blue diamonds) obtained from one- and two-qubit error rates, for each one-qubit and connected two-qubit subset, and the assumption of no crosstalk. (d) The ratio of the observed ($r_{\Omega,\text{perQ}}$) to predicted ($r_{\Omega,\text{perQ,pred}}$) per-qubit error rate shows that crosstalk errors cause the per-qubit error rate $r_{\Omega,\text{perQ}}$ to increase by approximately 250%–300% when $n \geq 15$. (e) The one- and two-qubit error rates were obtained using simultaneous one-qubit MRB on all 27 qubits (blue boxes), and two-qubit MRB, on all pairs of connected qubits, run simultaneously on the qubit pairs from eight distinct groupings (the purple and green boxes show two such groups).

is illustrated in Fig. 9 (b), for 6 of the 15 qubit subsets. For each qubit subset, we sampled and ran 30 circuits at each of a set of exponentially spaced depths.

Figure 9 (a) shows the observed polarization versus benchmark depth for six representative values of n. Even for $n = 27$, where we observe an average layer error rate of $r_\Omega = 28(1)\%$, we obtain a $d = 0$ average observed polarization of $S_0 \approx 40\%$. This demonstrates that MRB is practical on many qubits, even when the error rate per layer is $O(10\%)$. For all n, we observe that the mean observed polarization is consistent with an exponential decay, as expected. Fig. 9 (b) shows the error rate per qubit ($r_{\Omega,\text{perQ}} = 1 - (1 - r_\Omega)^{1/n} \approx r_\Omega/n$) versus n. Our circuits have a fixed expected two-qubit gate density (of $\xi = 1/4$). Therefore, $r_{\Omega,\text{perQ}}$ will be independent of n for $n \geq 2$ if (1) the error rate of one-qubit gates and the error rate of two-qubit gates is invariant across the device, and (2) there are no crosstalk errors. Instead, we observe that $r_{\Omega,\text{perQ}}$ rapidly increases from $r_{\Omega,\text{perQ}} \approx 0.2\%$ for $n = 2$ up to $r_{\Omega,\text{perQ}} \approx 1.2\%$—an increase of approximately 600%.

To quantify the contribution of crosstalk errors to the observed increase in the per-qubit error rate with n, we first need

to quantify the spatial variations in the one- and two-qubit gate error rates (meaning the error rates of those gates when all other qubits are idle). We used one- and two-qubit MRB to measure the error rates of each one-qubit subset and each connected two-qubit subset of the 27-qubits. Because of the large number of qubits, it would require running more circuits than was feasible to implement independent one-qubit MRB experiments on each qubit (27 MRB experiments) and independent two-qubit MRB experiments on each connected pair of qubits (30 MRB experiments). Instead, we implemented all 27 one-qubit MRB experiments simultaneously [3]. The resultant one-qubit MRB error rates therefore include contributions from single-qubit gate crosstalk errors. We ran the 30 two-qubit MRB experiments in eight groups, selected to minimize the closeness in frequency space of the qubits in each group. These two-qubit MRB error rates will therefore include some contributions from two-qubit gate crosstalk, but the experiments have been designed with the aim of minimizing this contribution. We also ran five isolated two-qubit MRB experiments and observed that the simultaneous two-qubit MRB error rates were a factor of between 1.5 and 2.5

times larger than the corresponding isolated MRB error rates (see Table III).

We use the set of measured one- and two-qubit MRB error rates to predict the n-qubit r_Ω that would be observed if there are no two-qubit gate crosstalk errors, using Eqs. (37)–(40) [56]. Figure 9 (c) shows the predictions for the per-qubit error rate $r_{\Omega,\text{perQ}}$. For $n \gg 1$ these predictions (blue diamonds) are much smaller than the observations (red circles). This prediction accounts for spatial variations in the one- and two-qubit error rates, and includes contributions from one-qubit gate crosstalk errors (and some contributions from two-qubit gate crosstalk). Therefore, we can conclude that the additional observed error is due to crosstalk caused by the two-qubit gates, and it lower bounds the total contribution of crosstalk errors to r_Ω. Figure 9 (d) shows the ratio R of the observed to the predicted error rate per qubit $r_{\Omega,\text{perQ}}$, versus n. R grows approximately linearly from $R \approx 0.2$ at $n = 2$ up to $R \approx 2.5$ at $n \approx 13$ and then saturates at between $R \approx 2.5$ and $R \approx 3.0$. One possible explanation for this is two-qubit gate crosstalk errors with finite spatial radius, i.e., two-qubit gates cause increased errors on other qubits within some distance of the target qubits.

VII. DISCUSSION

Scalable benchmarking methods are needed to quantify the integrated performance of medium- and large-scale quantum processors. In this paper, we introduced a scalable method for RB of universal gate sets that uses a novel and customizable family of randomized mirror circuit. We presented a theory for our method, showing that it reliably measures the error rate of a random n-qubit circuit layer sampled from a user-specified distribution Ω. We demonstrated MRB on multiple gate sets in both simulations and experiments, demonstrating that it is reliable and that it is a powerful tool for understanding errors in many-qubit circuits. Our method can be viewed as both an adaptation of standard RB and its variants, to enable efficient and scalable benchmarking of universal gate sets, and as an adaptation of XEB that removes XEB's inefficient circuit simulation step. It therefore provides a link between two widely used benchmarking methodologies, and so we anticipate that the ideas introduced here will lead to further advances in randomized benchmarking.

In our experiments, we demonstrated MRB of a gate set consisting of cnot and arbitrary single-qubit gates on up to 27 qubits and MRB of a gate set with non-Clifford two-qubit gates (cs and cs†) on up to 4 qubits. Our results provide evidence that MRB with non-Clifford gates is a robust method for determining a processor's error rate per gate layer, and that these error rates can be used to understand the magnitude of various types of errors. Our experiments show that MRB on many qubits reveals and quantifies errors not present in one- and two-qubit circuits, highlighting the importance of scalable benchmarks. Comparisons of RB error rates predicted from crosstalk-free models and our experimental results show evidence of large crosstalk errors in both of the devices we benchmarked and, importantly, our methods make it possible to quantify the size of these crosstalk errors.

We anticipate that a variety of interesting benchmarking methods can be constructed using MRB and extensions or adaptations of this method. For example, we anticipate that MRB can form the foundation of methods for estimating the error rates of individual gates and layers, within the context of many-qubit circuits. In this work we demonstrated a simple example of such a technique—fitting MRB data to a depolarizing model—and we expect that a variety of robust methods could be developed, that would complement or advance on existing methods for this task [9, 12, 35] such as interleaved RB. Alternatively, we anticipate that MRB can be adapted to construct scalable "full-stack" benchmarks based on random circuits, such as a scalable variant of the widely-used quantum volume benchmark [30].

REFERENCES

[1] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout, "Gate Set Tomography," Quantum **5**, 557 (2021).

[2] Kenneth Rudinger, Craig W. Hogle, Ravi K. Naik, Akel Hashim, Daniel Lobser, David I. Santiago, Matthew D. Grace, Erik Nielsen, Timothy Proctor, Stefan Seritan, Susan M. Clark, Robin Blume-Kohout, Irfan Siddiqi, and Kevin C. Young, "Experimental characterization of crosstalk errors with simultaneous gate set tomography," PRX Quantum **2**, 040338 (2021).

[3] Jay M Gambetta, AD Córcoles, Seth T Merkel, Blake R Johnson, John A Smolin, Jerry M Chow, Colm A Ryan, Chad Rigetti, S Poletto, Thomas A Ohki, *et al.*, "Characterization of addressability by simultaneous randomized benchmarking," Phys. Rev. Lett. **109**, 240504 (2012).

[4] Mohan Sarovar, Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout, "Detecting crosstalk errors in quantum information processors," Quantum **4**, 321 (2020).

[5] Timothy J Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young, "Direct randomized benchmarking for multiqubit devices," Phys. Rev. Lett. 123 (2019).

[6] Timothy Proctor, Stefan Seritan, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young, "Scalable randomized benchmarking of quantum computers using mirror circuits," (2020), arXiv:2112.09853 [quant-ph].

[7] Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout, "Measuring the capabilities of quantum computers," Nature Physics 18, 75–79 (2022).

[8] David C McKay, Andrew W Cross, Christopher J Wood, and Jay M Gambetta, "Correlated randomized benchmarking," (2020), arXiv:2003.02354 [quant-ph].

[9] Alexander Erhard, Joel James Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban Adrian Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt, "Characterizing large-scale quantum computers via cycle benchmarking," Nat. Commun. 10, 5347 (2019).

[10] Robin Harper, Steven T. Flammia, and Joel J. Wallman, "Efficient learning of quantum noise," Nat. Phys. 16, 1–5 (2020).

[11] Steven T. Flammia and Joel J. Wallman, "Efficient estimation of Pauli channels," ACM Trans. Quant. Comp. 1, 3 (2020).

[12] Steven T Flammia, "Averaged circuit eigenvalue sampling," (2021), arXiv:2108.05803 [quant-ph].

[13] Joseph Emerson, Robert Alicki, and Karol Życzkowski, "Scalable noise estimation with random unitary operators," J. Opt. B Quantum Semiclass. Opt. 7, S347 (2005).

[14] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G Cory, and Raymond Laflamme, "Symmetrized characterization of noisy quantum processes," Science 317, 1893–1896 (2007).

[15] Easwar Magesan, Jay M Gambetta, and Joseph Emerson, "Scalable and robust randomized benchmarking of quantum processes," Phys. Rev. Lett. 106, 180504 (2011).

[16] Easwar Magesan, Jay M Gambetta, and Joseph Emerson, "Characterizing quantum gates via randomized benchmarking," Phys. Rev. A 85, 042311 (2012).

[17] Emanuel Knill, D Leibfried, R Reichle, J Britton, RB Blakestad, JD Jost, C Langer, R Ozeri, S Seidelin, and DJ Wineland, "Randomized benchmarking of quantum gates," Phys. Rev. A 77, 012307 (2008).

[18] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson, "Characterizing universal gate sets via dihedral benchmarking," Phys. Rev. A 92, 060302 (2015).

[19] Andrew W Cross, Easwar Magesan, Lev S Bishop, John A Smolin, and Jay M Gambetta, "Scalable randomised benchmarking of non-clifford gates," NPJ Quantum Inf. 2, 16012 (2016).

[20] Winton G. Brown and Bryan Eastin, "Randomized benchmarking with restricted gate sets," Phys. Rev. A 97, 062323 (2018).

[21] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, "Real randomized benchmarking," Quantum 2, 85 (2018).

[22] Jonas Helsen, Xiao Xue, Lieven MK Vandersypen, and Stephanie Wehner, "A new class of efficient randomized benchmarking protocols," arXiv preprint arXiv:1806.02048 (2018).

[23] Jonas Helsen, Sepehr Nezami, Matthew Reagor, and Michael Walter, "Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates," (2020), arXiv:2011.13048 [quant-ph].

[24] Jahan Claes, Eleanor Rieffel, and Zhihui Wang, "Character randomized benchmarking for non-multiplicity-free groups with applications to subspace, leakage, and matchgate randomized benchmarking," PRX Quantum 2, 010351 (2021).

[25] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H Werner, and Jens Eisert, "A general framework for randomized benchmarking," (2020), arXiv:2010.07974 [quant-ph].

[26] A. Morvan, V. V. Ramasesh, M. S. Blok, J. M. Kreikebaum, K. O'Brien, L. Chen, B. K. Mitchell, R. K. Naik, D. I. Santiago, and I. Siddiqi, "Qutrit randomized benchmarking," Phys. Rev. Lett. 126, 210504 (2021).

[27] Sergio Boixo, Sergei V Isakov, Vadim N Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J Bremner, John M Martinis, and Hartmut Neven, "Characterizing quantum supremacy in near-term devices," Nat. Phys. 14, 595 (2018).

[28] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al., "Quantum supremacy using a programmable superconducting processor," Nature 574, 505–510 (2019).

[29] Yunchao Liu, Matthew Otten, Roozbeh Bassirianjahromi, Liang Jiang, and Bill Fefferman, "Benchmarking near-term quantum computers via random circuit sampling," (2021), arXiv:2105.05232 [quant-ph].

[30] Andrew W Cross, Lev S Bishop, Sarah Sheldon, Paul D Nation, and Jay M Gambetta, "Validating quantum computers using randomized model circuits," Phys. Rev. A 100, 032328 (2019).

[31] Karl Mayer, Alex Hall, Thomas Gatterman, Si Khadir Halit, Kenny Lee, Justin Bohnet, Dan Gresh, Aaron Hankin, Kevin Gilmore, and John Gaebler, "Theory of mirror benchmarking and demonstration on a quantum computer," arXiv [quant-ph] (2021), arXiv:2108.10431 [quant-ph].

[32] Vivek V. Shende, Igor L. Markov, and Stephen S. Bullock, "Minimal universal two-qubit controlled-not-based circuits," Physical Review A 69 (2004), 10.1103/physreva.69.062321.

[33] "AQT@LBL - SC Qubit Testbed," https://aqt.lbl.gov/, accessed: 2020-03-01.

[34] IBM Quantum https://quantum-computing.ibm.com, (2021).

[35] Easwar Magesan, Jay M Gambetta, Blake R Johnson, Colm A Ryan, Jerry M Chow, Seth T Merkel, Marcus P da Silva, George A Keefe, Mary B Rothwell, Thomas A Ohki, et al., "Efficient measurement of quantum gate error by interleaved randomized benchmarking," Phys. Rev. Lett. 109, 080505 (2012).

[36] Shelly Garion, Naoki Kanazawa, Haggai Landa, David C. McKay, Sarah Sheldon, Andrew W. Cross, and Christopher J. Wood, "Experimental implementation of non-clifford interleaved randomized benchmarking with a controlled-s gate," Phys. Rev. Research 3, 013204 (2021).

[37] Robin Harper and Steven T Flammia, "Estimating the fidelity of T gates using standard interleaved randomized benchmarking," Quantum Sci. Technol. 2, 015008 (2017).

[38] Michael A Nielsen, "A simple formula for the average gate fidelity of a quantum dynamical operation," Phys. Lett. A 303, 249–252 (2002).

[39] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout, "What randomized benchmarking actually measures," Phys. Rev. Lett. 119, 130502 (2017).

[40] Arnaud Carignan-Dugas, Kristine Boone, Joel J Wallman, and Joseph Emerson, "From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters," New Journal of Physics 20, 092001 (2018).

[41] Xiao Mi, Pedram Roushan, Chris Quintana, Salvatore Mandrà, Jeffrey Marshall, Charles Neill, Frank Arute, Kunal Arya, Juan Atalaya, Ryan Babbush, et al., "Information scrambling in quantum circuits," Science 374, 1479–1483 (2021),

https://www.science.org/doi/pdf/10.1126/science.abg5029.

[42] Timothy Proctor, Stefan Seritan, Erik Nielsen, Kenneth Rudinger, Kevin Young, Robin Blume-Kohout, and Mohan Sarovar, "Establishing trust in quantum computations," (2022), arXiv:2204.07568 [quant-ph].

[43] Joel J Wallman and Joseph Emerson, "Noise tailoring for scalable quantum computation via randomized compiling," Phys. Rev. A 94, 052325 (2016).

[44] It is also possible for these polarizations to be anti-correlated, although this is much less likely in practice, as a gate and its inverse do not typically have anti-correlated error rates.

[45] Robin Blume-Kohout, Marcus P. da Silva, Erik Nielsen, Timothy Proctor, Kenneth Rudinger, Mohan Sarovar, and Kevin Young, "A taxonomy of small markovian errors," PRX Quantum 3, 020335 (2022).

[46] Here and throughout this paper we use the "edge grab" sampler, which is parameterized by the two-qubit gate density ξ, defined in the supplementary material of Ref. [7].

[47] All uncertainties are calculated using a standard bootstrap.

[48] For the four one-qubit subsets, three of the cases coincide—as they differ only by the two-qubit gate set or the two-qubit gate density, which are unused parameters in one-qubit circuits. In that case we only sample and run only of the three identical MRB designs.

[49] David C McKay, Christopher J Wood, Sarah Sheldon, Jerry M Chow, and Jay M Gambetta, "Efficient Z gates for quantum computing," Phys. Rev. A 96, 022330 (2017).

[50] We note that one contribution to the difference between the observed r_Ω and the crosstalk-free prediction is the difference between idle gates that occur in parallel with a two-qubit gate and idle gates that occur in single-qubit circuits. The idle that occurs in parallel with a two-qubit gate is a 200 ns idle, whereas the idle gate that occurs in a one-qubit circuit is a 60 ns idle. Our prediction methodology implicitly assumes that these two idle gates have the same error rate. However, we conjecture that the contribution from this difference is small, because idle gates in this system are relatively low error.

[51] Shelby Kimmel, Marcus P da Silva, Colm A Ryan, Blake R Johnson, and Thomas Ohki, "Robust extraction of tomographic information via randomized benchmarking," Phys. Rev. X 4, 011050 (2014).

[52] Erik Nielsen, Kenneth Rudinger, Timothy Proctor, Kevin Young, and Robin Blume-Kohout, "Efficient flexible characterization of quantum processors with nested error models," New J. Phys. 23, 093020 (2021).

[53] The cycle benchmarking experiments measure the infidelities layers dressed with a different single-qubit gate set (the Pauli group) to those used in our MRB experiments, and these experiments were implemented on a different day than the MRB experiments, so exact agreement is not expected.

[54] Costin Lancu, Marc Davis, Ethan Smith, and USDOE, "Quantum search compiler (qsearch) v2.0, version v2.0," (2020).

[55] Marc G. Davis, Ethan Smith, Ana Tudor, Koushik Sen, Irfan Siddiqi, and Costin Iancu, "Towards optimal topology aware quantum circuit synthesis," in 2020 IEEE International Conference on Quantum Computing and Engineering (QCE) (2020) pp. 223–234.

[56] In this case we do not have a closed form expression for $\Omega(L)$. Instead, we sample 10000 layers from Ω, compute ϵ_L for each layer from the one- and two-qubit MRB error rates, and then average all 10000 ϵ_L to estimate r_Ω.

[57] Ryan Shaffer, Hang Ren, Emiliia Dyrenkova, Christopher G Yale, Daniel S Lobser, Ashlyn D Burch, Matthew N H Chow, Melissa C Revelle, Susan M Clark, and Hartmut Häffner, "Efficient verification of continuously-parameterized quantum gates," (2022), arXiv:2205.13074 [quant-ph].

[58] Kenneth Rudinger, Timothy Proctor, Dylan Langharst, Mohan Sarovar, Kevin Young, and Robin Blume-Kohout, "Probing context-dependent errors in quantum processors," Phys. Rev. X 9, 021045 (2019).

[59] S Mavadia, C L Edmunds, C Hempel, H Ball, F Roy, T M Stace, and M J Biercuk, "Experimental quantum verification in the presence of temporally correlated noise," npj Quantum Information 4, 7 (2018).

[60] Harrison Ball, Thomas M Stace, Steven T Flammia, and Michael J Biercuk, "Effect of noise correlations on randomized benchmarking," Phys. Rev. A 93, 022303 (2016).

[61] Arnaud Carignan-Dugas, Joel J Wallman, and Joseph Emerson, "Efficiently characterizing the total error in quantum circuits," arXiv preprint arXiv:1610.05296 (2016).

[62] The two densities heuristic is not the only method for estimating ϵ_2 from Eq. (40). An alternative is to first estimate ϵ_1 from single-qubit MRB error rates, and to then put this ϵ_1 into Eq. (40) to find ϵ_2 (this is what we did in Section V C). However, this will underestimate ϵ_1 if there is crosstalk errors in the one-qubit gates (causing an overestimate of ϵ_2), because the single-qubit MRB error rates are not impacted by these errors (as only one qubit is driven). This would not be the case if the single-qubit MRB experiments were run simultaneously [3]. This is what we did in the experiments reported in Section VI.

Appendix A: MRB with stochastic Pauli errors

In this appendix, we provide further details on the theory presented in Section IV A, showing that the MRB error rate (r_Ω) approximately equals ϵ_Ω [Eq. (8)] under the assumption of stochastic Pauli errors.

1. Determining the observed polarization of MRB circuits

In this appendix, we prove Eq. (23), which says that the mean observed polarization [Eq. (14)] of randomized mirror circuits equals the mean polarization of the overall error map of a randomized mirror circuit. We start by defining an overall error map for our mirror circuit, which captures all error in the circuit. We can define an overall error map for a circuit C by rewriting $\phi(\mathsf{C})$

with all error moved to the beginning of the circuit. For a general depth-l circuit with gate-dependent errors,

$$\begin{aligned}
\phi(\mathsf{C}) &= \mathcal{E}(\mathsf{L}_l)\mathcal{U}(\mathsf{L}_l)\cdots\mathcal{E}(\mathsf{L}_2)\mathcal{U}(\mathsf{L}_2)\mathcal{E}(\mathsf{L}_1)\mathcal{U}(\mathsf{L}_1) \\
&= \mathcal{U}(\mathsf{L}_l)\cdots\mathcal{U}(\mathsf{L}_2)\mathcal{U}(\mathsf{L}_1)\mathcal{E}'_{\mathsf{L}_l}\cdots\mathcal{E}'_{\mathsf{L}_2}\mathcal{E}'_{\mathsf{L}_1},
\end{aligned} \tag{A1}$$

where

$$\mathcal{E}'_{\mathsf{L}_i} = \mathcal{U}(\mathsf{L}_1)^{-1}\cdots\mathcal{U}(\mathsf{L}_i)^{-1}\mathcal{E}(\mathsf{L}_i)\mathcal{U}(\mathsf{L}_i)\cdots\mathcal{U}(\mathsf{L}_1). \tag{A2}$$

Applying Eq. (A1) to our randomized mirror circuit allows us to express the error in $\tilde{\mathsf{M}}_d$ (which is the mirror circuit without the initial and final layers), as a single error channel following the initial randomized state preparation layer L_0. We find that $\phi(\mathsf{M}_d)$ can be expressed as

$$\begin{aligned}
\phi(\mathsf{M}_d) &= \phi(\mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c))\phi(\tilde{\mathsf{M}}_d)\phi(\mathcal{R}(\mathsf{P}_0\mathsf{L}_0)) \\
&= \mathcal{U}(\mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c)\tilde{\mathsf{M}}_d)\mathcal{U}(\mathsf{P}_0)\mathcal{E}_{\mathrm{SPAM}}\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_d)\mathcal{U}(\mathsf{L}_0) \\
&= \mathcal{U}(\mathsf{P}_{d+1})\mathcal{U}(\mathsf{L}_0^{-1})\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)\mathcal{U}(\mathsf{L}_0),
\end{aligned} \tag{A3}$$

where

$$\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d) = \mathcal{E}_{\mathrm{SPAM}}\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_d) \tag{A4}$$

$$= \mathcal{E}_{\mathrm{SPAM}}\mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1}^{-1},\mathsf{P}_d)}\cdots\mathcal{E}'_{\mathsf{L}_{d/2}^{-1}}\mathcal{E}'_{\mathsf{L}_{d/2}}\cdots\mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1},\mathsf{P}_0)}. \tag{A5}$$

To obtain Eq. (A3), we use the reflection structure of randomized mirror circuits—in particular, $\mathcal{U}(\mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c)\tilde{\mathsf{M}}_d) = \mathcal{U}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_0)$, where P_0 and P_{d+1} are the Pauli gates that are recompiled into L_0 and L_0^{-1}, respectively, in the randomized compilation step. The Pauli gate P_{d+1} determines the target bit string of M_d—i.e., $\mathcal{U}(\mathsf{M})|0\rangle\!\rangle = \mathcal{U}(\mathsf{P}_{d+1})|0\rangle\!\rangle = |b\rangle\!\rangle$. The overall error map $\mathcal{E}_{\mathrm{eff}}(\tilde{\mathsf{M}}_\mathsf{d})$ [Eq. (A5)] contains the error from the $d/2$ Ω-random circuit layers and their inverses (after randomized compilation), and it is composed of unitary rotations of the error channels associated with each circuit layer.

In the MRB protocol, we compute each circuit's observed polarization S [Eq. (14)]. We now show that the observed polarization $S(\mathsf{M}_d)$ is related to the polarization [Eq. (5)] of M_d's overall error map (introduced above). Using the expression for $\phi(\mathsf{M}_d)$ in Eq. (A3), the probability of measuring bit string x on circuit M_d is given by

$$\mathcal{P}_x = \langle\!\langle x|\,\mathcal{U}(\mathsf{P}_{d+1})\mathcal{U}(\mathsf{L}_0^{-1})\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)\mathcal{U}(\mathsf{L}_0)|0\rangle\!\rangle \tag{A6}$$

$$= \langle\!\langle x+b|\,\mathcal{U}(\mathsf{L}_0^{-1})\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)\mathcal{U}(\mathsf{L}_0)|0\rangle\!\rangle. \tag{A7}$$

The layer L_0 consists of single-qubit gates independently sampled from single-qubit unitary 2-designs. We now average over the initial circuit layer L_0, making use of a fidelity estimation technique based on single-qubit gates: the fidelity of any error channel $\mathcal{E}$ can be found by averaging over a tensor product of single-qubit 2-designs [42]. In particular, for any bit string $y \in \{0,1\}^n$,

$$\gamma(\mathcal{E}) = \frac{4^n}{4^n-1}\sum_{x\in\{0,1\}^n}(-1/2)^{h(x,y)}\langle\!\langle x+y|\,\bar{\mathcal{E}}\,|0\rangle\!\rangle - \frac{1}{4^n-1}, \tag{A8}$$

where $\bar{\mathcal{E}} = \mathbb{E}_\mathsf{L}[\mathcal{U}(\mathsf{L})^\dagger\mathcal{E}\mathcal{U}(\mathsf{L})]$ [42] and $\mathsf{L} = \otimes_{i=1}^n\mathsf{L}_i$, where each L_i is a independent, single-qubit 2-design. This implies that the expected observed polarization of M_d over L_0 is

$$\begin{aligned}
\mathbb{E}_{\mathsf{L}_0} S(\mathsf{M}_d) &= \frac{4^n}{4^n-1}\left[\sum_{k=0}^n\sum_{\langle x,b\rangle=k}\left(-\frac{1}{2}\right)^k\mathbb{E}_{\mathsf{L}_0}\mathcal{P}_x\right] - \frac{1}{4^n-1} \\
&= \frac{4^n}{4^n-1}\left[\sum_{x\in\{0,1\}^n}\left(-\frac{1}{2}\right)^{h(x,b)}\mathbb{E}_{\mathsf{L}_0}\mathcal{P}_x\right] - \frac{1}{4^n-1} \\
&= \gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d))
\end{aligned} \tag{A9}$$

where $\gamma(\mathcal{E})$ denotes the *polarization* of $\mathcal{E}$ [Eq. (5)]. Eq. (A9) follows from Eq. (A8). Averaging over all depth-d randomized mirror circuits, the mean observed polarization is

$$\bar{S}_d = \mathbb{E}_{\mathsf{M}_d}\gamma(\mathcal{E}_{\mathrm{eff}}(\mathsf{M}_d)). \tag{A10}$$

Equation (A10) says that the average observed polarization $\bar{S}_d$, which is estimated by the MRB protocol, is equal to the expected polarization of the error channel of a depth-d mirror circuit.

2. Relating the observed polarization of MRB circuits and Ω-distributed random circuits

Above, we related the mean observed polarization ($\bar{S}_d$), which determines the MRB error rate, to the expected polarization of the overall error map of a depth-d randomized mirror circuit. We now use this result to derive Eq. (25), which relates the mean observed polarization of depth-d randomized mirror circuits to the expected polarization of the error map of a depth-$d/2$ Ω-distributed circuit. In combination with the theory in Section A 3—which shows that $\bar{S}_d$ and the mean polarization of the overall error map of Ω-distributed random circuits decay exponentially—the relationship we derive here implies that $r_\Omega \approx \epsilon_\Omega$.

Our goal is to relate the rate of decay of $\bar{S}_d$ to the rate of decay of the fidelity of Ω-distributed circuits ($\bar{F}_d$) [Eq. (7)]. We start by expressing $\bar{F}_d$ in terms of the expected polarization of the overall error map of a depth-d Ω-distributed circuit. Applying Eq. (A1) to a depth-d, Ω-distributed random circuit $\mathsf{C}_d = \mathsf{L}_d \mathsf{L}_{\theta_d} \cdots \mathsf{L}_0 \mathsf{L}_{\theta_1}$, we obtain an overall error map for C_d, $\mathcal{E}_{\text{eff}}(\mathsf{C}_d)$, which is defined by $\phi(\mathsf{C}_d) = \mathcal{U}(\mathsf{C}_d)\mathcal{E}_{\text{eff}}(\mathsf{C}_d)$. We define Γ_d to be the average polarization of the error map of a depth-d mirror circuit:

$$\bar{\Gamma}_d = \mathop{\mathbb{E}}_{\mathsf{C}_d} \gamma(\mathcal{E}_{\text{eff}}(\mathsf{C}_d)). \tag{A11}$$

To relate $\bar{S}_d$ to $\bar{\Gamma}_d$, we use the fact that a depth-d randomized mirror circuit consists of randomized compilation of a depth-$d/2$ Ω-distributed random circuit followed by its inverse. These two depth-$d/2$ circuits are both Ω-distributed (even after randomized compilation), but they are correlated. Below, we show that the polarization of the mirror circuit's overall error map depends on the covariance between the error in a depth-$d/2$ Ω-distributed circuit and its randomly compiled inverse. We can write the overall error map in Eq. (A5) as a composition of two error maps—an overall error map for a random circuit and an overall error map for its randomly compiled inverse:

$$\mathcal{E}_{\text{eff}}(\mathsf{M}_d) = \mathcal{E}_{\text{SPAM}}\mathcal{E}_{\text{eff},2}(\mathsf{M}_d)\mathcal{E}_{\text{eff},1}(\mathsf{M}_d), \tag{A12}$$

where

$$\mathcal{E}_{\text{eff},1}(\mathsf{M}_d) = \mathcal{E}'_{\mathcal{R}(\mathsf{P}_{d/2}\mathsf{L}_{d/2}\mathsf{P}^c_{d/2-1})} \cdots \mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1},\mathsf{P}_0)}$$

$$\mathcal{E}_{\text{eff},2}(\mathsf{M}_d) = \mathcal{E}'_{\mathcal{T}(\mathsf{L}_{\theta_1}^{-1},\mathsf{P}_d)} \cdots \mathcal{E}'_{\mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2})}$$

$$= \mathcal{U}(\mathsf{C})\mathcal{E}_{\text{eff}}(\mathcal{T}(\mathsf{L}_{\theta_1}^{-1},\mathsf{P}_\mathsf{d}) \cdots \mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2}))\mathcal{U}(\mathsf{C})^{-1},$$

and

$$\mathsf{C} = \mathcal{R}(P_{d/2+1}L_{d/2}^{-1}P_{d/2})\mathcal{R}(P_{d/2}L_{d/2}P^c_{d/2-1}) \cdots \mathcal{R}(P_2 L_1 P^c_1)\mathcal{T}(L_{\theta_1},P_0). $$

C is the first half of $\tilde{\mathsf{M}}_d$, and it is a depth-$d/2$ Ω-distributed random circuit that has had randomized compilation applied to it. By substituting Eq. (A12) into Eq. (A10), we obtain

$$\bar{S}_d = \mathop{\mathbb{E}}_{\mathsf{M}_d} \gamma(\mathcal{E}_{\text{SPAM}}\mathcal{E}_{\text{eff},2}(\mathsf{M}_d)\mathcal{E}_{\text{eff},1}(\mathsf{M}_d)) \tag{A13}$$

$$= \gamma(\mathcal{E}_{\text{SPAM}}) \mathop{\mathbb{E}}_{\mathsf{M}_d} \gamma(\mathcal{E}_{\text{eff},2}(\mathsf{M}_d)\mathcal{E}_{\text{eff},1}(\mathsf{M}_d)), \tag{A14}$$

where, to go from Eq. (A13) to Eq. (A14), we have used the assumption that $\mathcal{E}_{\text{SPAM}}$ is a global depolarizing channel.

Applying randomized compilation to an Ω-distributed random circuit creates a new random circuit that is also Ω-distributed. This is due to the conditions we require of Ω_1 and Ω_2 (Ω_1 is the uniform distribution, and Ω_2 is invariant under replacing a subset of a layer's gates with their inverses). Therefore, we can replace the average over all depth-d randomized mirror circuits in Eq. (A14) with an average over all depth-$d/2$ Ω-distributed random circuits:

$$\bar{S}_d = \gamma(\mathcal{E}_{\text{SPAM}}) \mathop{\mathbb{E}}_{\mathsf{C}_{d/2}} \gamma(\mathcal{U}(\mathsf{C}_{d/2})\bar{\mathcal{E}}_{\text{eff}}(\mathsf{C}_{d/2}^{-1})\mathcal{U}(\mathsf{C}_{d/2})^{-1}\mathcal{E}_{\text{eff}}(\mathsf{C}_{d/2})), \tag{A15}$$

where $\bar{\mathcal{E}}_{\text{eff}}(\mathsf{C}_{d/2}^{-1})$ denotes the average over all possible circuits C' resulting from applying randomized compilation to $\mathsf{C}_{d/2}^{-1}$. Expressing Eq. (A15) in terms of $\bar{\Gamma}_{d/2}$ [Eq. (A11)], we have

$$\bar{S}_d = \gamma(\mathcal{E}_{\text{SPAM}})\left(\bar{\Gamma}_{d/2}^2 - \Delta_\Omega\right), \tag{A16}$$

where

$$\Delta_\Omega = \mathop{\mathbb{E}}_{\mathsf{C}_{d/2}} \gamma(\mathcal{U}(\mathsf{C}_{d/2})\bar{\mathcal{E}}_{\text{eff}}(\mathsf{C}_{d/2}^{-1})\mathcal{U}(\mathsf{C}_{d/2})^{-1}\mathcal{E}_{\text{eff}}(\mathsf{C}_{d/2})) - \left(\mathop{\mathbb{E}}_{\mathsf{C}_{d/2}} \gamma(\mathcal{E}_{\text{eff}}(\mathsf{C}_{d/2}))\right)^2. \tag{A17}$$

3. Deriving the exponential decay model

In this appendix, we show that the mean observed polarization of randomized mirror circuits [$\bar{S}_d$, Eq. (A15)] and the mean polarization of the overall error map of a Ω-distributed random circuit [$\bar{\Gamma}_{d/2}$, Eq. (A11)] decay exponentially. To show that $\bar{S}_d$ and $\bar{\Gamma}_{d/2}$ decay exponentially, we will assume that the error on each layer $\mathcal{E}(\mathsf{L})$ is a Pauli stochastic channel [Eq. (6)]. Because $S_d \approx \gamma(\mathcal{E}_{\text{SPAM}})\bar{\Gamma}_{d/2}^2$, this theory implies that $r_\Omega \approx \epsilon_\Omega$. Assuming stochastic Pauli error, we further show that $r_\Omega \approx \delta_\Omega$, where δ_Ω is the average infidelity of a layer sampled from Ω.

We start by considering the rate of decay of the observed polarization of randomized mirror circuits. Because each layer's error map is a stochastic Pauli channel, we can relate the polarization of $\mathcal{E}_{\text{eff}}(\tilde{\mathsf{M}}_d)$ [Eq. (A5)] to the polarizations of the error channels of individual circuit layers. $\mathcal{E}_{\text{eff}}(\tilde{\mathsf{M}}_d)$ is the composition of a stochastic Pauli channel for each circuit layer of $\tilde{\mathsf{M}}_d$, each rotated by a unitary. To first order in the layer infidelities ε, the polarization of the effective error channel is equal to the product of the polarizations of the layers' error channels [42]:

$$\gamma(\mathcal{E}_{\text{eff}}(\mathsf{M}_d)) = \gamma(\mathcal{E}_{\text{SPAM}}) \prod_{i=1}^{d/2} \left[\gamma(\mathcal{E}(\mathsf{L}_i)) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}, \mathsf{P}_i))\big) \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}^{-1}, \mathsf{P}_{d-i+1}))\big) \right] + O(d^2\varepsilon^2). \tag{A18}$$

Equation (A18) relates $\bar{S}_d$ to the polarizations of the $2d$ individual layer error maps. By substituting Eq. (A18) into Eq. (A15) and using the uniformity conditions on Ω specified in Section II C (see Appendix A 4 for details), we find that $\bar{S}_d = Ap^d + O(d^2\varepsilon^2)$, with $A = \gamma(\mathcal{E}_{\text{SPAM}})$ and p given by

$$p^2 = \mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}^{-1})) \gamma(\mathcal{E}(\mathsf{L})) \gamma(\mathcal{E}(\mathsf{L}_\theta)) \bar{\gamma}(\mathsf{L}_\theta). \tag{A19}$$

where $\bar{\gamma}(\mathsf{L}_\theta) = \mathbb{E}_{\mathsf{P}'} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}'))\big)$ is the expected polarization of a two-qubit gate layer L_θ after inversion and randomized compiling. The $O(d^2\varepsilon^2)$ term (and higher-order corrections) are negligible whenever the probability of two or more errors cancelling is negligible [42]. This is the case in highly scrambling circuits [5, 42] (when $n \gg 1$), and our circuits are highly scrambling by construction. Therefore $S \approx Ap^d$, i.e., the observed polarization of depth-d randomized mirror circuits decays exponentially with d.

We now show that the expected polarization of the overall error map on a Ω-distributed random circuit decays exponentially in depth, which will allow us to relate its decay rate to the decay rate of the observed polarization of randomized mirror circuits. By approximating $\gamma(\mathcal{E}_{\text{eff}}(\mathsf{C}))$ as a product of polarizations (as Ω-distributed random circuits are highly scrambling), we obtain

$$\gamma(\mathcal{E}_{\text{eff}}(\mathsf{C})) = \prod_{i=1}^{d} \gamma(\mathcal{E}(\mathsf{L}_i)) \gamma(\mathcal{E}(\mathsf{L}_{\theta_i})) + O(d^2\varepsilon^2). \tag{A20}$$

Combined with Eq. (A11), Eq. (A20) implies that the expected polarization of the overall error map of an Ω-distributed random circuit decays exponentially. Averaging over all depth-d Ω-distributed random circuits, the expected polarization of the overall error map is

$$\bar{\Gamma}_{d,\text{rc}} = \left(\mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L})) \gamma(\mathcal{E}(\mathsf{L}_\theta)) \right)^d + O(d^2\varepsilon^2). \tag{A21}$$

$$= p_{\text{rc}}^d + O(d^2\varepsilon^2), \tag{A22}$$

where p_{rc} is given by

$$p_{\text{rc}} = \mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}_i)) \gamma(\mathcal{E}(\mathsf{L}_{\theta_i})) \tag{A23}$$

and sets the decay rate of the expected polarization of the overall error map on an Ω-distributed random circuit. Substituting Eq. (A23) into Eq. (A19), we have $p^2 = p_{\text{rc}}^2 + \Delta'_\Omega$, where

$$\Delta'_\Omega = \mathop{\mathbb{E}}_{\mathsf{L},\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L})) \gamma(\mathcal{E}(\mathsf{L}^{-1})) \gamma(\mathcal{E}(\mathsf{L}_\theta)) \bar{\gamma}(\mathsf{L}_\theta) - p^2.$$

Other RB protocols, including Clifford MRB [6], have been formulated as protocols measuring the average layer infidelity of random circuits, which we denote as δ_Ω. The average layer infidelity can, in general, be very different from ϵ_Ω, and it is not a gauge-invariant quantity [39, 40]. Under the assumption of stochastic Pauli errors, Eqs. (A22) and (A23) imply that $\epsilon_\Omega \approx \delta_\Omega$ (due to the $O(\varepsilon^2)$ probability of errors on two layers canceling, this equality is not exact). However, for more general errors δ_Ω is ill-defined [39, 40]. Furthermore, even for stochastic Pauli errors, the approximations required to relate r_Ω to ϵ_Ω are smaller than those needed to relate r_Ω to δ_Ω—relating r_Ω to ϵ_Ω requires the assumption that the correlation between the error on a Ω-distributed random circuit and its randomly compiled inverse [Δ_Ω, Eq. (A17)] is small, whereas relating r_Ω to δ_Ω requires this assumption *and* the assumption that the rate of error cancellation between individual circuit layers is low.

4. Computing the average observed polarization

In this appendix, we derive the exponential decay of the average observed polarization of randomized mirror circuits [Eq. (A19)] from Eq. (A18). Our proof makes use of the following properties of the transformation applied to a two-qubit gate layer L_θ in our randomized compilation procedure (see Section III A):

$$\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}), \mathsf{P}) = \mathsf{L}_\theta \tag{A24}$$

$$\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}) = \mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1} \tag{A25}$$

$$\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}_1), \mathsf{P}_2) = \mathcal{T}(\mathsf{L}_\theta, \mathsf{P}_2 \mathsf{P}_1) \tag{A26}$$

$$\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}) = \mathsf{L}_\theta^{-1}. \tag{A27}$$

Eq. (A27) follows from Eqs. (A24) and (A25).

We start with Eq. (A18), which gives the polarization of the overall error map of a depth-d mirror circuit M_d,

$$\gamma(\mathcal{E}_{\text{eff}}(\mathsf{M}_d)) = A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i)) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}, \mathsf{P}_i))\big) \gamma\big(\mathcal{E}(\mathsf{L}_i^{-1})\big) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}^{-1}, \mathsf{P}_{d-i+1}))\big) + O(d^2 \varepsilon^2), \tag{A28}$$

where $A = \gamma(\mathcal{E}_{\text{SPAM}})$. We will average over all depth-d randomized mirror circuits to show that the expected polarization of the overall error map decays exponentially. Our randomized mirror circuits consist of $d/2$ randomly-sampled two-qubit gate layers from Ω_2 and d randomly-sampled Paulis that get recompiled with these (for randomized compilation). We now average the polarization over all of these layers and Paulis:

$$\mathop{\mathbb{E}}_{\mathsf{P}_1,\dots\mathsf{P}_d} \mathop{\mathbb{E}}_{\mathsf{L}_{\theta_i}} \gamma(\mathcal{E}_{\text{eff}}(\mathsf{M}_d)) = A \mathop{\mathbb{E}}_{\mathsf{P}_1,\dots\mathsf{P}_d} \mathop{\mathbb{E}}_{\mathsf{L}_{\theta_i}} \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i)) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}, \mathsf{P}_i))\big) \gamma\big(\mathcal{E}(\mathsf{L}_i^{-1})\big) \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_{\theta_i}^{-1}, \mathsf{P}_{d-i+1}))\big) + O(d^2 \varepsilon^2) \tag{A29}$$

$$= A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \mathop{\mathbb{E}}_{\mathsf{P}_1,\dots\mathsf{P}_{d/2}} \mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}_i))\big) \mathop{\mathbb{E}}_{\mathsf{P}_{d/2+1}\dots\mathsf{P}_d} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}_{d-i+1}))\big) + O(d^2 \varepsilon^2) \tag{A30}$$

$$= A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \left(\mathop{\mathbb{E}}_{\mathsf{P}} \mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}))\big) \mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}'))\big) \right)^{d/2} + O(d^2 \varepsilon^2) \tag{A31}$$

$$= A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \left(\mathop{\mathbb{E}}_{\mathsf{P}} \mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}))\big) \mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}\big(\mathcal{T}(\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}), \mathsf{P}')\big)\big) \right)^{d/2} + O(d^2 \varepsilon^2) \tag{A32}$$

$$= A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \left(\mathop{\mathbb{E}}_{\mathsf{P}} \mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}_i))\big) \mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}\big(\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}'\mathsf{P})\big)\big) \right)^{d/2} + O(d^2 \varepsilon^2) \tag{A33}$$

$$= A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \left(\mathop{\mathbb{E}}_{\mathsf{P}} \mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P}))\big) \mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}\big(\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}')\big)\big) \right)^{d/2} + O(d^2 \varepsilon^2). \tag{A34}$$

Eq. (A32) follows from Eq. (A31) and Eq. (A27), and Eq. (A33) follows from Eq. (A32) and Eq. (A26) . To get from Eq. (A33) to Eq. (A34), we use the fact that multiplying the uniformly random Pauli P' by P gives another uniformly random Pauli, so

$$\mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}\big(\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}'\mathsf{P})\big)\big) = \mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}\big(\mathcal{T}(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})^{-1}, \mathsf{P}')\big)\big). \tag{A35}$$

We can simplify Eq. (A34) using the uniformity conditions on the distribution Ω_2 (see Section II C) that the two-qubit gate layers are sampled from. The effect of randomized compiling on a 2-qubit gate layer is to invert some subset of the gates in the layer, leaving the rest unchanged. Because Ω_2 is invariant under inverting any subset of the gates in a layer, $\Omega_2(\mathsf{L}_{\theta_i}) = \Omega_2(\mathcal{T}(\mathsf{L}_{\theta_i}, \mathsf{P}_i))$. Additionally, the randomized compilation transformation just permutes the elements of $\mathbb{L}_2$. Therefore, $\mathbb{E}_{\mathsf{L}_\theta} f(\mathcal{T}(\mathsf{L}_\theta, \mathsf{P})) = \mathbb{E}_{\mathsf{L}_\theta} f(\mathsf{L}_\theta)$ for any function f. Applying this fact to Eq. (A34),

$$\mathop{\mathbb{E}}_{\mathsf{P}_1,\dots\mathsf{P}_d} \mathop{\mathbb{E}}_{\mathsf{L}_{\theta_i}} \gamma(\mathcal{E}_{\text{eff}}(\mathsf{M}_d)) = A \prod_{i=1}^{d/2} \gamma(\mathcal{E}(\mathsf{L}_i^{-1})) \gamma(\mathcal{E}(\mathsf{L}_i)) \left(\mathop{\mathbb{E}}_{\mathsf{L}_\theta} \gamma(\mathcal{E}(\mathsf{L}_\theta)) \left(\mathop{\mathbb{E}}_{\mathsf{P}'} \gamma\big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1}, \mathsf{P}'))\big) \right) \right)^{d/2} + O(d^2 \varepsilon^2). \tag{A36}$$

We now finish averaging over all depth-d randomized mirror circuits by averaging over the single-qubit gate layers. Averaging over the $d/2$ Ω_1-random single-qubit gate layers, the expected polarization of the error map, over all depth-d randomized mirror

circuits, is

$$\underset{\mathsf{M}_d}{\mathbb{E}}\,\gamma(\mathcal{E}_{\text{eff}}(\mathsf{M}_d)) = A\,\underset{\mathsf{L}_i}{\mathbb{E}}\left[\prod_{i=1}^{d/2}\gamma(\mathcal{E}(\mathsf{L}_i^{-1}))\gamma(\mathcal{E}(\mathsf{L}_i))\left(\underset{\mathsf{L}_\theta}{\mathbb{E}}\,\gamma(\mathcal{E}(\mathsf{L}_\theta))\left(\underset{\mathsf{P}'}{\mathbb{E}}\,\gamma\Big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1},\mathsf{P}'))\Big)\right)\right)^{d/2}\right] + O(d^2\varepsilon^2) \tag{A37}$$

$$= A\prod_{i=1}^{d/2}\left(\underset{\mathsf{L}}{\mathbb{E}}\,\underset{\mathsf{L}_\theta}{\mathbb{E}}\,\gamma(\mathcal{E}(\mathsf{L}^{-1}))\gamma(\mathcal{E}(\mathsf{L}))\gamma(\mathcal{E}(\mathsf{L}_\theta))\left(\underset{\mathsf{P}'}{\mathbb{E}}\,\gamma\Big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1},\mathsf{P}'))\Big)\right)\right)^{d/2} + O(d^2\varepsilon^2). \tag{A38}$$

Therefore, combining Eq. (A38) with Eq. (A15), we have that the average observed polarization decays exponentially as $\bar{S}_d = Ap^d + O(d^2\varepsilon^2)$, with

$$p^2 \approx \underset{\mathsf{L}}{\mathbb{E}}\,\underset{\mathsf{L}_\theta}{\mathbb{E}}\,\gamma(\mathcal{E}(\mathsf{L}))\gamma(\mathcal{E}(\mathsf{L}_\theta))\gamma(\mathcal{E}(\mathsf{L}^{-1}))\left(\underset{\mathsf{P}'}{\mathbb{E}}\,\gamma\Big(\mathcal{E}(\mathcal{T}(\mathsf{L}_\theta^{-1},\mathsf{P}'))\Big)\right). \tag{A39}$$

Appendix B: MRB with Clifford Two-Qubit Gates

The theory presented in Section IV A (which is presented in detail in Appendix A) assumes stochastic Pauli noise on each circuit layer to derive the exponential decay of the observed polarization of mirror circuits. However, stochastic error is not always the dominant error in a processor. Our method uses a randomized compilation procedure to convert error into stochastic Pauli error. In this appendix, we prove that when all two-qubit gates in an MRB experiment are Clifford, the error in randomized mirror circuits is twirled into Pauli stochastic error, under the assumption that the error map on the one-qubit gates is independent of the Paulis with which they are compiled.

We consider a depth-d randomized mirror circuit (treated as a random variable), which we write as

$$\mathsf{M}_d = \mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c)\mathcal{T}(\mathsf{L}_{\theta_1}^{-1},\mathsf{P}_d)\mathcal{R}(\mathsf{P}_d\mathsf{L}_1^{-1}\mathsf{P}_{d-1}^c)\cdots\mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2})\mathcal{R}(\mathsf{P}_{d/2}\mathsf{L}_{d/2}\mathsf{P}_{d/2-1}^c)\cdots\mathcal{R}(\mathsf{P}_2\mathsf{L}_1\mathsf{P}_1^c)\mathcal{T}(\mathsf{L}_{\theta_1},\mathsf{P}_0)\mathcal{R}(\mathsf{P}_0\mathsf{L}_0). \tag{B1}$$

When the two-qubit gate layers consist of two-qubit Cliffords of the form CP_θ, they are not changed by the randomized compilation step of our circuit construction. Therefore,

$$\mathsf{M}_d = \mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c)\mathsf{L}_{\theta_1}^{-1}\mathcal{R}(\mathsf{P}_d\mathsf{L}_1^{-1}\mathsf{P}_{d-1}^c)\cdots\mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2})\mathcal{R}(\mathsf{P}_{d/2}\mathsf{L}_{d/2}\mathsf{P}_{d/2-1}^c)\cdots\mathcal{R}(\mathsf{P}_2\mathsf{L}_1\mathsf{P}_1^c)\mathsf{L}_{\theta_1}\mathcal{R}(\mathsf{P}_0\mathsf{L}_0), \tag{B2}$$

We will assume the error on the single-qubit gates is independent of the Paulis they are recompiled with—i.e., $\phi(\mathcal{R}(\mathsf{L}_iP'P)) = \mathcal{E}(\mathsf{L}_i)\mathcal{U}(\mathcal{R}(P'\mathsf{L}_iP))$. Using this assumption, an implementation of the circuit M_d can be written as

$$\phi(\mathsf{M}_d) = \mathcal{E}(\mathsf{L}_0^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c))\mathcal{E}(\mathsf{L}_{\theta_1}^{-1})\mathcal{U}(\mathsf{L}_{\theta_1}^{-1})\mathcal{E}(\mathsf{L}_1^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_d\mathsf{L}_1^{-1}\mathsf{P}_{d-1}^c))\cdots\mathcal{E}(\mathsf{L}_{\theta_{d/2}}^{-1})\mathcal{U}(\mathsf{L}_{\theta_{d/2}}^{-1})\mathcal{E}(\mathsf{L}_{d/2}^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2}))\mathcal{E}(\mathsf{L}_{d/2})$$

$$\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2}\mathsf{L}_{d/2}\mathsf{P}_{d/2-1}^c))\mathcal{E}(\mathsf{L}_{\theta_{d/2}})\mathcal{U}(\mathsf{L}_{\theta_{d/2}})\mathcal{E}(\mathsf{L}_{d/2-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2-1}\mathsf{L}_{d/2-1}\mathsf{P}_{d/2-2}^c))\cdots\mathcal{E}(\mathsf{L}_{\theta_1})\mathcal{U}(\mathsf{L}_{\theta_1})\mathcal{E}(\mathsf{L}_0)\mathcal{U}(\mathcal{R}(\mathsf{P}_0\mathsf{L}_0)). \tag{B3}$$

We now push the error on the single-qubit gate layers through the two-qubit gate layers, defining new error channels that represent the error on a composite layer. Eq. (B3) becomes

$$\phi(\mathsf{M}_d) = \mathcal{E}(\mathsf{L}_0^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c))\mathcal{E}_d'\mathcal{U}(\mathsf{L}_{\theta_1}^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_d\mathsf{L}_1^{-1}\mathsf{P}_{d-1}^c))\cdots\mathcal{E}_{d/2+1}'\mathcal{U}(\mathsf{L}_{\theta_{d/2}}^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2+1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2}))\mathcal{E}(\mathsf{L}_{d/2})$$

$$\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2}\mathsf{L}_{d/2}\mathsf{P}_{d/2-1}^c))\mathcal{E}_{d/2}'\mathcal{U}(\mathsf{L}_{\theta_{d/2}})\mathcal{U}(\mathcal{R}(\mathsf{P}_{d/2-1}\mathsf{L}_{d/2-1}\mathsf{P}_{d/2-2}^c))\cdots\mathcal{E}_1'\mathcal{U}(\mathsf{L}_{\theta_1})\mathcal{U}(\mathcal{R}(\mathsf{P}_0\mathsf{L}_0)), \tag{B4}$$

where

$$\mathcal{E}_i' = \mathcal{E}(\mathsf{L}_{\theta_i})\mathcal{U}(\mathsf{L}_{\theta_i})\mathcal{E}(\mathsf{L}_{i-1})\mathcal{U}(\mathsf{L}_{\theta_i})^{-1} \qquad\qquad 1 \le i \le \frac{d}{2} \tag{B5}$$

$$\mathcal{E}_i' = \mathcal{E}(\mathsf{L}_{\theta_{d-i+1}})\mathcal{U}(\mathsf{L}_{\theta_{d-i+1}})^{-1}\mathcal{E}(\mathsf{L}_{i+1})\mathcal{U}(\mathsf{L}_{\theta_{d-i+1}}) \qquad\qquad \frac{d}{2} < i \le d. \tag{B6}$$

We have grouped the error channels into error channels $\mathcal{E}_i'$ that represent the error in a composite layer. Now, we use the structure of the randomized compilation procedure to twirl the error. The dressed layers can be expanded in terms of the original sampled layer and the Paulis inserted in randomized compilation as

$$\mathcal{U}(\mathsf{L}_{\theta_i})\mathcal{U}(\mathcal{R}(\mathsf{P}_{i-1}\mathsf{L}_{i-1}\mathsf{P}_{i-2}^c)) = \mathcal{U}((\mathsf{P}_i^c)^{-1}\mathsf{L}_{\theta_i}\mathsf{L}_{i-1}\mathsf{P}_{i-1}^c) \qquad\qquad 1 \le i \le \frac{d}{2} \tag{B7}$$

$$\mathcal{U}(\mathsf{L}_{\theta_{d-i+1}}^{-1})\mathcal{U}(\mathcal{R}(\mathsf{P}_{i+1}\mathsf{L}_{d-i+1}^{-1}\mathsf{P}_i^c)) = \mathcal{U}((\mathsf{P}_{i+1}^c)^{-1}\mathsf{L}_{\theta_{d-i+1}}^{-1}\mathsf{L}_{d-i+1}^{-1}\mathsf{P}_i^c) \qquad\qquad \frac{d}{2} < i \le d. \tag{B8}$$

Rewriting Eq. (B4) using these expansions, we have

$$\phi(\mathsf{M}_d) = \mathcal{E}(\mathsf{L}_0^{-1})\mathcal{U}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1}\mathsf{P}_d^c)\mathcal{E}_d'\mathcal{U}((\mathsf{P}_d^c)^{-1}\mathsf{L}_{\theta_1}^{-1}\mathsf{L}_1^{-1}\mathsf{P}_{d-1}^c)\cdots\mathcal{E}_{d/2+1}'\mathcal{U}((\mathsf{P}_{d/2+1}^c)^{-1}\mathsf{L}_{\theta_{d/2}}^{-1}\mathsf{L}_{d/2}^{-1}\mathsf{P}_{d/2}^c)\mathcal{E}(\mathsf{L}_{d/2})$$

$$\mathcal{U}((\mathsf{P}_{d/2}^c)^{-1}\mathsf{L}_{\theta_{d/2}}\mathsf{L}_{d/2-1}\mathsf{P}_{d/2-1}^c)\mathcal{E}_{d/2}'\mathcal{U}((\mathsf{P}_{d/2-1}^c)^{-1}\mathsf{L}_{\theta_{d/2-1}}\mathsf{L}_{d/2-2}\mathsf{P}_{d/2-1}^c)\cdots\mathcal{E}_1'\mathcal{U}((\mathsf{P}_1^c)^{-1}\mathsf{L}_{\theta_1}\mathsf{L}_0), \quad \text{(B9)}$$

where each correction layer P_i^c is a uniform random Pauli, because the two-qubit gates are Clifford. Averaging over the uniform random n-qubit Paulis $\mathsf{P}_0, \mathsf{P}_1, \ldots, \mathsf{P}_d$, which equivalently averages over the correction Paulis $\mathsf{P}_1^c, \ldots, \mathsf{P}_{d+1}^c$, performs a Pauli twirl, converting the error channels into Pauli stochastic error channels. Performing this average, Eq. (B9) becomes

$$\phi(\mathsf{M}_d) = \mathcal{E}(\mathsf{L}_0^{-1})\mathcal{U}(\mathsf{P}_{d+1}\mathsf{L}_0^{-1})\mathcal{S}_d\mathcal{U}(\mathsf{L}_{\theta_1}^{-1}\mathsf{L}_1^{-1})\cdots\mathcal{S}_{d/2+1}\mathcal{U}(\mathsf{L}_{\theta_{d/2}}^{-1}\mathsf{L}_{d/2}^{-1})\mathcal{S}(\mathsf{L}_{d/2})\mathcal{U}(\mathsf{L}_{\theta_{d/2}}\mathsf{L}_{d/2-1})\mathcal{S}_{d/2}\mathcal{U}(\mathsf{L}_{\theta_{d/2-1}}\mathsf{L}_{d/2-2})\cdots\mathcal{S}_1\mathcal{U}(\mathsf{L}_{\theta_1}\mathsf{L}_0), \quad \text{(B10)}$$

where $\mathsf{S}_i = \mathbb{E}_\mathsf{P} \mathsf{P}\mathcal{E}_i'\mathsf{P}^{-1}$ and $\mathcal{S}(\mathsf{L}_{d/2}) = \mathbb{E}_\mathsf{P} \mathsf{P}\mathcal{E}(\mathsf{L}_{d/2})\mathsf{P}^{-1}$ are stochastic Pauli channels, each of which captures the error from one composite layer. All error, except the error on the final circuit layer, is twirled into stochastic Pauli noise by the random Paulis inserted in randomized compilation. Therefore, we expect our method to be sensitive to all errors when the two-qubit gates are chosen to be Clifford gates.

Appendix C: MRB with non-Clifford two-qubit gates

In this appendix we show that, when applied to a gate set containing non-Clifford two-qubit gates, MRB is sensitive to all Hamiltonian errors on those two-qubit gates except one linear combination of errors on a non-Clifford two-qubit gate and its inverse. We then discuss possible adaptations to our protocol that would guarantee sensitivity to all Hamiltonian errors on non-Clifford two-qubit gates.

1. Sensitivity of errors in non-Clifford two-qubit gates

In Appendix B, we showed that when the two-qubit gate set used in MRB contains only Clifford gates, the error in the two-qubit gates is twirled, upon averaging, into stochastic Pauli noise. This guarantees sensitivity to general errors on the two-qubit gates. We now consider circuits with non-Clifford two-qubit gates and show that randomized mirror circuits are sensitive to most Hamiltonian errors on the two-qubit gates, to first order. We will assume there is no crosstalk error, and all two-qubit layers are sampled independently, so that we expect the only systematic coherent cancellation of errors to come from a layer and its inverse. We will also assume there is no error on the single-qubit gates. To see the effect of error in a two-qubit gate on a randomized mirror circuit to first order, it is sufficient to consider mirror circuits resulting from applying our circuit construction procedure to a single two-qubit composite layer $L = L_1 L_\theta$, where $L_\theta = CP_\theta$ is a two-qubit gate and L_1 is a one-qubit gate layer. After mirroring and randomized compilation on L, we have the circuit $M = \mathcal{T}(L_\theta^{-1}, P_2)\mathcal{R}(P_2 L_1^{-1} P_1^c)\mathcal{R}(P_1 L_1 P_0^c)\mathcal{T}(L_\theta, P_0)$, where P_0, P_1, and P_2 are random two-qubit Pauli layers. The ideal operation M implements is $\mathcal{U}(M) = (P_2^c)^{-1}P_0$. An imperfect implementation of M can be expressed as

$$\phi(M) = \mathcal{E}(\mathcal{T}(L_\theta^{-1}, P_2))\mathcal{U}(\mathcal{T}(L_\theta^{-1}, P_2)\mathcal{R}(P_2 L_1^{-1} P_1^c)\mathcal{R}(P_1 L_1 P_0^c))\mathcal{E}(\mathcal{T}(L_\theta, P_0))\mathcal{U}(\mathcal{T}(L_\theta, P_0)) \quad \text{(C1)}$$

$$= \mathcal{E}(\mathcal{T}(L_\theta^{-1}, P_2))\mathcal{U}(\mathcal{T}(L_\theta^{-1}, P_2)P_2 P_0^c)\mathcal{E}(\mathcal{T}(L_\theta, P_0))\mathcal{U}(\mathcal{T}(L_\theta, P_0)) \quad \text{(C2)}$$

$$= \mathcal{E}(\mathcal{T}(L_\theta^{-1}, P_2))\mathcal{U}((P_2^c)^{-1}L_\theta^{-1}P_0^c)\mathcal{E}(\mathcal{T}(L_\theta, P_0))\mathcal{U}(\mathcal{T}(L_\theta, P_0)) \quad \text{(C3)}$$

$$= \mathcal{E}(\mathcal{T}(L_\theta^{-1}, P_2))\mathcal{U}((P_2^c)^{-1}P_0\mathcal{T}(L_{\theta_i}, P_0)^{-1}(P_0^c)^{-1}P_0^c)\mathcal{E}(\mathcal{T}(L_\theta, P_0))\mathcal{U}(\mathcal{T}(L_\theta, P_0)) \quad \text{(C4)}$$

$$= \mathcal{E}(\mathcal{T}(L_\theta^{-1}, P_2))\mathcal{U}((P_2^c)^{-1}P_0\mathcal{T}(L_\theta, P_0)^{-1})\mathcal{E}(\mathcal{T}(L_\theta, P_0))\mathcal{U}(\mathcal{T}(L_\theta, P_0)), \quad \text{(C5)}$$

where we have used the definitions of the two-qubit gate layer $\mathcal{T}(L_\theta, P)$ and the correction layers P_i^c to rewrite the unitary evolution.

We now consider the effect of general gate-dependent Hamiltonian errors on the two-qubit gates on $\phi(M)$. We will write the error in terms of elementary error generators, as defined in the error generator formalism of Ref. [45]. We model the error on each two-qubit gate g as $\mathcal{E}(g) = e^{M_g}$, where

$$M_g = \sum_{P_a, P_b} \varepsilon_{P_a, P_b}^g H_{P_a, P_b}, \quad \text{(C6)}$$

and where H_{P_a, P_b} is the two-qubit Hamiltonian error generator indexed by the Pauli operators P_a and P_b. Using this expression for the error and expanding Eq. (C5) to first order in the error rates ε_{P_a, P_b}^g, we have

$$\phi(M) \approx \mathcal{U}((P_2^c)^{-1}P_0)\left(\mathcal{U}(I) + \left(\sum_{P_a, P_b} \varepsilon_{P_a, P_b}^{\mathcal{T}(L_{-\theta}, P_2)} H_{P_a, P_b}\right) + \mathcal{U}(\mathcal{T}(L_\theta, P_0)^{-1})\left(\sum_{P_a, P_b} \varepsilon_{P_a, P_b}^{\mathcal{T}(L_\theta, P_0)} H_{P_a, P_b}\right)\mathcal{U}(\mathcal{T}(L_\theta, P_0))\right). \quad \text{(C7)}$$

Eq. (C7) expresses the implementation of M in terms of its target evolution and a first order correction. The circuit is insensitive to an error to first order when the correction term vanishes, which occurs when

$$\left(\sum_{P_a,P_b} \varepsilon_{P_a,P_b}^{\mathcal{T}(L_{-\theta},P_2)} H_{P_a,P_b}\right) + \mathcal{U}(\mathcal{T}(L_\theta,P_0)^{-1})\left(\sum_k \varepsilon_{P_a,P_b}^{\mathcal{T}(L_{\theta_i},P_{i-1})} H_{P_a,P_b}\right)\mathcal{U}(\mathcal{T}(L_\theta,P_0)) = 0. \quad \text{(C8)}$$

Satisfying Eq. (C8) requires that the coefficient of each elementary error generator H_{P_a,P_b} is 0, which results in a system of 15 linear equations for each of 16^2 choices of two-qubit random Paulis P_2, P_0 used in randomized compilation. The randomized mirror circuits are sensitive to an error if for some choice of P_0 and P_2, the system cannot be satisfied when the error is nonzero. The two-qubit gate set $\mathbb{G}_2$ is closed under inverses, so in addition to mirroring $L = L_1 L_\theta$ as we have done above, we can mirror $L_1 L_{-\theta}$ to get an analogous set of linear equations. Considering all of the equations from mirroring $L_1 L_\theta$ and $L_1 L_{-\theta}$, we have a system of $2 \times 16^2 \times 15$ linear equations. The solutions to this system are $\varepsilon_{P_a,P_b}^\theta = \varepsilon_{P_a,P_b}^{-\theta} = 0 \ \forall (P_a, P_b) \neq (P, P)$ and $\varepsilon_{P,P}^G = \varepsilon_{P,P}^{G^\dagger}$. This means that to first order, the mirror circuits are not sensitive to the sum of $H_{P,P}$ errors on CP_θ and $CP_{-\theta}$, as we can change $\varepsilon_{P,P}^G + \varepsilon_{P,P}^{G^\dagger}$ without changing the error in any of the mirror circuits. This is a result of the structure we use for our mirror circuits. Below, we discuss how our method can be adapted to address this insensitivity.

2. Adaptations of MRB

While our simulations and experiments suggest non-Clifford MRB is a robust method, when our randomized mirror circuits contain non-Clifford two-qubit gates they are not sensitive to some coherent errors on these gates. In Appendix C 1 we showed that MRB circuits containing non-Clifford two-qubit gates are not sensitive to one linear combination of the Hamiltonian errors in these gates because of the correlations between the randomized compilation and the two-qubit gate that is applied, which prevent error from being perfectly twirled into stochastic noise. This shortcoming in our method is due to our choice of structure for our randomized mirror circuits. However, circuit mirroring is a flexible technique that can be applied to a variety of circuit structures, and here we discuss several adaptations of our method utilizing this flexibility that would address the error insensitivity in MRB.

Our method involves sampling random circuits with layers sampled from a user-specified distribution Ω over circuit layers. Different choices of circuit structure can address the shortcomings of our method and make other scalable benchmarks. We could guarantee sensitivity to all errors with more complex sampling of the Ω-distributed random circuit. For example, to benchmark a two-qubit gate set $\mathbb{G}_2 = \{\text{cs}, \text{cs}^\dagger\}$ we could generate circuits containing cs, $\text{cs}^\dagger$ and cz gates and implement the cz gate by two consecutive cs or $\text{cs}^\dagger$ gates. This MRB experiment would be sensitive to the $H_{Z,Z}$ errors on the cs and $\text{cs}^\dagger$ gates that our MRB experiment is insensitive to (see above).

Our MRB protocol performs inversion layer-by-layer, and an alternative method to guarantee sensitivity to all errors is to use more complex inversion strategies that reduce the correlation in the gate layers in the two halves of a mirror circuit. One option is to invert multiple circuit layers at a time, through computing the inverse of the layers and compiling an inverse circuit—and similar ideas to this have recently been used to implement RB of continuously parameterized gates [57]. However, compilation can be computationally-intensive with many qubits. Alternatively, we can modify the inversion layers by adding in additional gates, while maintaining a circuit that is logically equivalent to the inverse.

Appendix D: Error Models for MRB Simulations

We simulated MRB with three classes of error models—stochastic, Hamiltonian, and stochastic+Hamiltonian. Our models are defined based on the error generator formalism in [45]. Error rates are specified as elementary error generators of a post-gate error map. We include qubit-dependent Hamiltonian errors and Pauli stochastic errors on the $x_{\pi/2}$ and single-qubit idle gates with Hamiltonian error rates sampled in the range $[0, h/10]$, and stochastic Pauli error rates sampled in the range $[0, s/10]$. The stochastic and Hamiltonian errors are each split randomly across the three Paulis. We also include qubit-dependent Hamiltonian errors and Pauli stochastic errors on the cs and $\text{cs}^\dagger$ gates with Hamiltonian error rates sampled in the range $[0, h]$, and Pauli stochastic error rates sampled in the range $[0, s]$, spread at random across the 15 two-qubit Pauli errors.

To generate error models, we start with an overall error parameter p and select s, h such that $h^2 + s = p$. We generate models with $p \in [0.001, 0.2475]$ for 150 evenly-spaced values for the 1-qubit models and $p \in [0.0001, 0.075]$ for 150 evenly-spaced values for the 2- and 4-qubit models. In the stochastic error models, we set $h = 0$. In the Hamiltonian error models, we set $s = 0$. In the stochastic+Hamiltonian error models, we generate $s \in [0, p]$ at random, and set $h = \sqrt{p - s}$.

For each error model, we run a randomly-generated set of MRB circuits consisting of $K = 300$ circuits at each benchmark depth $d \in \{2^j \mid 0 \leq j \leq 8\}$. We approximate the error rate in Ω-distributed random circuits (ϵ_Ω) via sampling. For each depth $d \in \{2^j \mid 0 \leq j \leq 8\}$, we ran K randomly-generated depth-$d/2$ Ω-distributed random circuits, each followed by a perfect projective measurement onto the target state.

Appendix E: AQT experiments

In this appendix we provide further details about our experiments on AQT, which are discussed in Section V.

1. Experiment design

A specific set of MRB circuits for a given $(\mathbb{G}_2, \mathbb{G}_1, \mathbb{Q}, \xi)$ is obtained by sampling K circuits at a set of benchmark depths. We used $K = 30$ and a set of exponentially spaced benchmarking depths ($d = 0, 2, 4, 8 \ldots$). For the MRB designs in which $\xi = 1/2$ we did not independently sample the circuits for the three different gate sets. Instead, to sample a depth-d circuit on qubits $\mathbb{Q}$ for each of our three gate sets:

1. We sampled a depth $d/2$ circuit C for the $(\{\mathsf{cz}\}, \mathbb{SU}(2))$ gate set.

2. We created a correlated sample for the $(\{\mathsf{cz}\}, \mathbb{C}_1)$ gate set by replacing the $\mathbb{SU}(2)$ gates in C with gates sampled from $\mathbb{C}_1$.

3. We created a correlated sample for the $(\{\mathsf{cs}, \mathsf{cs}^\dagger\}, \mathbb{SU}(2))$ gate set by replacing each cz gates in C with either cs or $\mathsf{cs}^\dagger$ at random.

4. We independently converted each of the three circuits in (1)-(3) into a randomized mirror circuit.

Because the marginal distribution for the sampling of each circuit set is unaffected by this procedure, it does not impact the RB error rates we estimate (except by correlating their uncertainties), but it allows us to perform an interesting per-circuit comparison (see Appendix E 2).

To enable comparison to an established technique, we also ran *direct* RB circuits, which are described in Section V E. This resulted in a total of 16,194 circuits. In the experiment, we randomized the order of this circuit list, and ran each circuit 1000 times in turn. We repeated this three times in succession, to enable us to look for substantial changes in a circuit's success probability that signify drift. We used standard statistical testing methods [58] to identify circuits in which the success probability changed between the three runs, and discarded that data. We performed qutrit classification in the readout, i.e., the readout was calibrated to resolve the '2' leakage state from the two computational basis states. Whenever a circuit output '2', the result was discarded.

2. Comparing MRB circuits with Clifford and Haar-random single-qubit gates

The observed similarity between the average success rates of circuits in which the single qubit-gate gates $\mathsf{u}(\theta, \phi, \lambda)$ are sampled from two different distributions (see Section V B) does *not* imply that the success rate of an individual circuit is independent of the values of θ, ϕ and λ in its $\mathsf{u}(\theta, \phi, \lambda)$ gates. We designed our experiments so that each circuit in the $(\mathbb{SU}(2), \{\mathsf{cz}\}, \mathbb{Q}, 1/2)$ experiment is identical to a circuit in the $(\mathbb{C}_1, \{\mathsf{cz}\}, \mathbb{Q}, 1/2)$ experiment except for the values of each of the single-qubit gate's parameters. We can therefore use our data for each such circuit pair (C_1, C_2) to investigate whether circuit success rates depend on the values of the phase shifts. Figure 10 shows the observed polarization S for C_1 versus S for C_2 for each pair of circuits (C_1, C_2) that differ only by the values of the phases in z_θ gates. There are many circuit pairs that have very different S, e.g., there is a circuit pair for which $S \approx 0.9$ for one circuit and $S \approx -0.3$ for the other (note that $-1/2 \leq S \leq 1$). Figure 10(b) shows that the spread in the differences between the observed polarization of circuit pairs is largest for single-qubit circuits ($\sigma = 0.147$) and decreases as the number of qubits increases ($\sigma = 0.052$ for $n = 4$). The substantial variance in observed polarization differences implies strongly structured errors, e.g., coherent errors. Even for perfect z_θ gates, the value of each phase shift impacts how errors in other gates propagate through a circuit [59, 60].

3. Estimating the error rates of individual gates

A single MRB experiment is designed to estimate a single error rate r_Ω that quantifies the average rate at which an n-qubit layer causes an error in Ω-random circuits. But we can also use MRB to extract information about the error rates of particular layers. In Section V D we present one method for doing so—fitting to a depolarizing model. In this appendix we explain this method and present two alternative methods. These methods are complementary, as they trade off rigor for complexity. Note that one possible method for estimating the error rates of individual gates using MRB is to run an *interleaved* [35] version of MRB (and interleaved standard RB has been previously used to measure the error rate of a cs gate [36]). We do not explore this here, although we note that interleaved MRB would inherit all of the known problems with interleaved standard RB [51, 61].
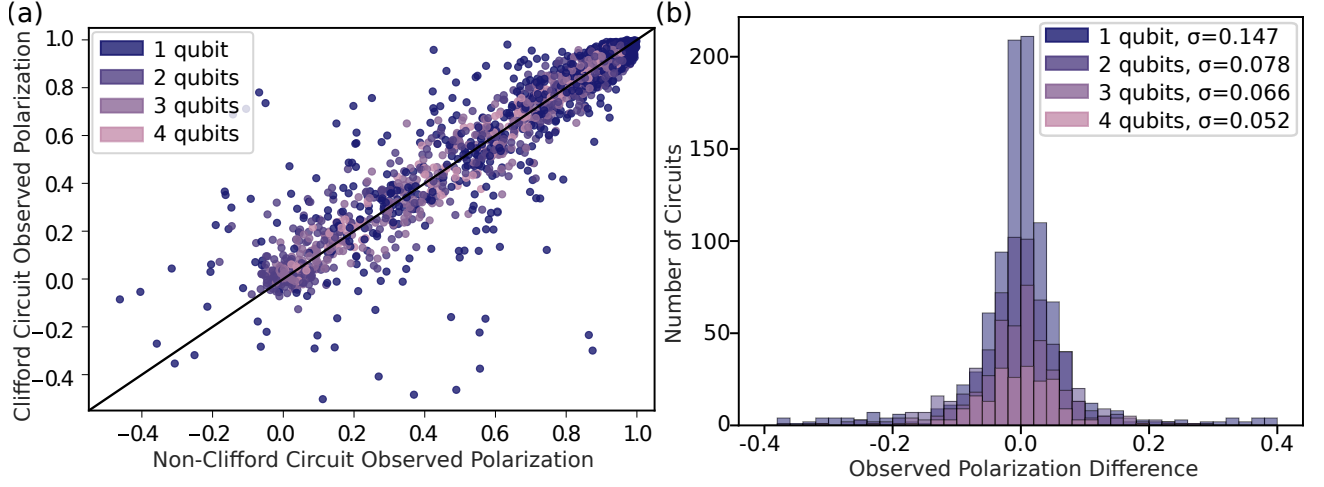
FIGURE 10. **Comparing the success rates of circuits that differ only by virtual phase gates**. Our experiments were designed so that each randomly sampled circuit containing cz and single-qubit Clifford gates [the ($\{$cz$\}$, $\mathbb{C}_1$) gate set] differs from a randomly sampled circuit containing cz and Haar-random single-qubit unitaries [the ($\{$cz$\}$, $\mathbb{SU}(2)$) gate set] only by the angles in its z_θ gates. Here we compare the observed polarization of each pair of circuits that differ only by the angles in these virtual z_θ gates. For many of these circuit pairs (C_1, C_2), C_1 and C_2 have very different observed polarizations, meaning that they have very different success rates. (a): The observed polarization for the ($\{$cz$\}$, $\mathbb{SU}(2)$) circuits and their corresponding Clifford circuits. (b): The difference in observed polarization between ($\{$cz$\}$, $\mathbb{SU}(2)$) circuits and their corresponding Clifford circuits.

a.   Estimating gate error rates using a varied-densities heuristic

MRB uses flexible sampling of the circuit layers, as each composite layer is sampled from some distribution Ω. By running MRB experiments with the same layer set $\mathbb{L}$ but different sampling distributions $\{\Omega_1, \Omega_2, \dots\}$ over $\mathbb{L}$, we can (approximately) ascertain the average error rates of different subsets of gates by applying basic linear algebra to $\{r_{\Omega_1}, r_{\Omega_2}, \dots\}$ [5]. In our experiments, we ran MRB for the gate set ($\{$cs, cs$^\dagger\}$, $\mathbb{SU}(2)$) with two different Ω defined by two different two-qubit gate densities: $\xi = 1/2$ and $\xi = 1/8$. We focus on the three two-qubit sets of connected qubits. Using Eq. (40), for each two-qubit subset $\mathbb{Q}$, we have that

$$\begin{pmatrix} r_{1/2} \\ r_{1/8} \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 7/8 & 1/8 \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \tag{E1}$$

where $r_\xi = r(\mathbb{G}_1, \{cs, cs^\dagger\}, \mathbb{Q}, \xi)$, ϵ_1 is the infidelity of the dressed layer consisting of dressed idles on each qubit in $\mathbb{Q}$, and ϵ_2 is the mean of the infidelities of $\mathbb{G}_1$-dressed cs and cs† gates applied to the qubits $\mathbb{Q}$. We solve these linear equations to estimate ϵ_2, for all three connected qubit pairs. These estimates are shown in Fig. 11(a), and we call this method the *two densities heuristic*, as it is based on the approximate relation of Eq. (37) [62].

Qubit subset ($\mathbb{Q}$)	$r(\{$cs, cs$^\dagger\}, \mathbb{SU}(2), \mathbb{Q}, 1/8)$	$r(\{$cs, cs$^\dagger\}, \mathbb{SU}(2), \mathbb{Q}, 1/2)$	$r_{DRB}(\{$cs, cs$^\dagger\}, \mathbb{SU}(2), \mathbb{Q}, 1/2)$	$r(\{$cz$\}, \mathbb{SU}(2), \mathbb{Q}, 1/2)$	$r(\{$cz$\}, \mathbb{C}, \mathbb{Q}, 1/2)$
Q4		0.25(3)%	0.18(2)%		0.27(2)%
Q5		0.12(1)%	0.12(1)%		0.12(1)%
Q6		0.118(4)%	0.108(5)%		0.113(3)%
Q7		0.079(1)%	0.080(2)%		0.080(1)%
(Q4, Q5)	0.50(3)/	0.77(3)%	0.87(4)%	0.73(2)%	0.67(2)%
(Q5, Q6)	0.54(3)%	0.86(3)%	0.81(3)%	0.81(2)%	0.76(2)%
(Q6, Q7)	1.05(4)%	1.05(4)%	0.99(4)%	1.04(3)%	0.98(2)%
(Q4, Q5, Q6)	1.04(4)%	1.64(5)%		1.48(4)%	1.51(4)%
(Q5, Q6, Q7)	0.97(3)%	1.63(3)%		1.61(4)%	1.61(3)%
(Q4, Q5, Q6, Q7)	1.36(3)%	2.48(5)%		2.45(5)%	2.34(5)%

TABLE I. **RB error rates on AQT.** The RB error rates for every RB experiment we ran on AQT. We benchmarked each connected subset of four linearly-connected qubits and used three different gate sets.
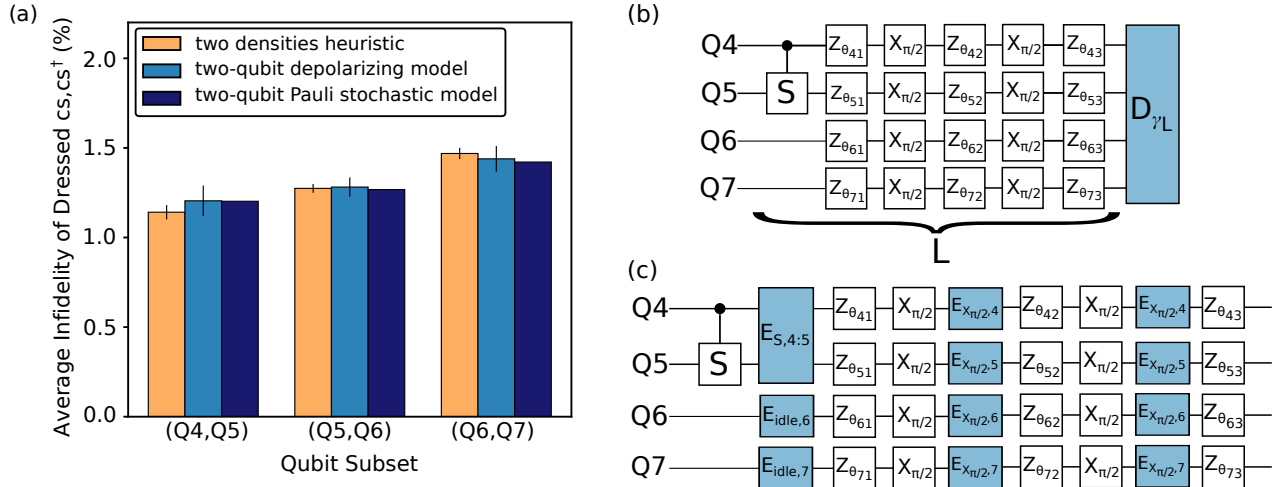
FIGURE 11. **Fitting error models to MRB data and estimating gate error rates.** We fit two types of error models to MRB data to estimate the infidelity of individual circuit layers. (a) By running two MRB experiments with two different two-qubit gate densities ξ, we can estimate the mean infidelity of a set of one or more two-qubit gates—here cs and cs†—using basic linear algebra (see Appendix E 3 a). We call this procedure the *two densities heuristic*. The estimates of the average gate error obtained from the two densities heuristic (orange) are compared to independent estimates obtained from two more rigorous but more complex and computationally intensive procedures: fitting each set of two-qubit MRB data to (1) a depolarizing model (light blue), and (2) a stochastic Pauli errors model (dark blue). (b) To fit a depolarizing model, we assign an error rate to each dressed layer and an error rate to each qubit's readout. (c) To fit a Pauli stochastic model, we assign a Pauli stochastic channel to each possible gate except the virtual z_θ gates.
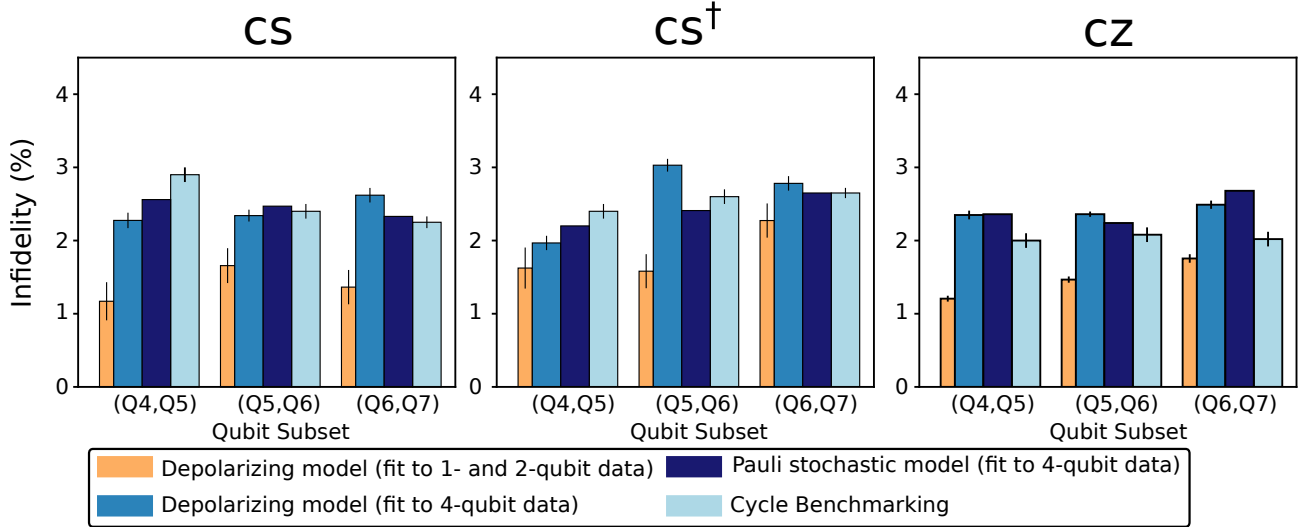

FIGURE 12. **Estimating the infidelity of dressed 4-qubit layers.** By fitting error models to MRB data, we can estimate the infidelity of each $\mathbb{G}_1$-dressed layer used in the MRB circuits. Here we show four different estimates of the infidelities of 4-qubit layers containing a single cs, cs† or cz gate on one of the three connected pairs of qubits. We fit a simple n-qubit depolarizing model to (1) the 4-qubit data, and (2) the 1- and 2-qubit data, and use both models to estimate the infidelity of 4-qubit $\mathbb{G}_1$-dressed layers. The estimates from fitting to the 1- and 2-qubit data do not account for any additional crosstalk errors that occur in 4-qubit layers, so the additional error estimated when fitting to the 4-qubit data is a quantification of crosstalk. We also fit a more sophisticated stochastic Pauli error model to the 4-qubit circuit data, resulting in comparable estimates to those obtained from the simple depolarizing model (which uses a scalable, less computationally intensive analysis). To validate our results against an established technique, we compare to infidelities independently estimated using cycle benchmarking [9]. We observe qualitative agreement. The cycle benchmarking experiments measure the infidelities of layers dressed with one-qubit gates sampled from a different gate set (the Pauli group) to that used in our MRB experiments [$\mathbb{SU}(2)$ or $\mathbb{C}_1$, the single-qubit Clifford group], and these experiments were implemented on a different day than the MRB circuits, so exact agreement is not expected.

b. Estimating gate error rates by fitting depolarizing error models

The two densities heuristic is built on the standard MRB data analysis, which extracts a single error rate (r_Ω) from each MRB experiment design. But data from even a single MRB experiment contains a lot more information about each gate's errors than is contained in r_Ω, e.g., RB data can contain sufficient information for complete tomography [51, 52]. In principle, this information can be extracted by fitting error models to MRB data—as has been demonstrated in simulations with 2-qubit standard RB [52]. However, fitting an error model to data typically requires simulating the circuits under that error model and this simulation is, in general, exponentially expensive in the number of qubits. Simplified, scalable approximations are therefore useful. One model that satisfies these criteria is a model in which each gate's error is modelled by a single error rate, and the error map for a layer of gates is a depolarizing channel [7].

Our depolarizing model summarizes the errors in each dressed one- and two-qubit gate (g) with a single, independent error rate ϵ_g. The error channel for each dressed n-qubit layer L is modelled by an n-qubit depolarizing channel with an infidelity ϵ_L given by $\epsilon_L = 1 - \prod_{g \in L}(1 - \epsilon_g)$. This means modelling the error channel for each dressed n-qubit layer L by

$$\mathcal{D}_{\gamma_L}[\rho] = \gamma_L \rho + (1 - \gamma_L)\frac{\mathbb{I}}{2^n}, \tag{E2}$$

with

$$\gamma_L = \frac{1}{4^n - 1}\left[4^n \prod_{g \in L}(1 - \epsilon_g) - 1\right]. \tag{E3}$$

This error model is illustrated in Fig 11(b). We also model the readout on each qubit with an independent error rate ϵ_{Qi}, where the readout error on an n-qubit circuit is an n-qubit depolarizing channel with infidelity $\epsilon_R = 1 - \prod_{Qi \in \mathbb{Q}}(1 - \epsilon_{Qi})$. Under this error model, the observed polarization of a circuit $C = L_1 L_2 \ldots L_d$ is

$$S(C) = \gamma(L_1)\gamma(L_2)\ldots\gamma(L_d)\gamma(R). \tag{E4}$$

The parameters of this depolarizing model are a set of error rates—an ϵ_g for each $\mathbb{G}_1$-dressed one- and two-qubit gate g and an ϵ_{Qi} for the readout on each qubit Qi. To estimate these parameters we use a least-squares fit of Eq. (E4) to the observed polarizations of the MRB circuits. We separately fit the parameters of the depolarizing model to the data from MRB circuits on different numbers of qubits ($n = 1, 2, 3, 4$), so that we can study how the error rates of the gates change with n, due to crosstalk errors.

Fitting Eq. (E4) to the data from two-qubit MRB circuits results in estimates of the infidelity of each two-qubit dressed layer containing a two-qubit gate from $\mathbb{G}_2$ (and an estimate for ϵ_1, the dressed two-qubit idle). We can therefore use these fits to compare to the two densities heuristic (above). Figure 11(a) compares the mean of the entanglement infidelities of the dressed cs and cs† gates, obtained from this fit, with the estimate from the two densities heuristic. The estimates of both methods are between 1.1% and 1.5%, with the estimates differing by 0.5%–2.1%, which cross-validates the two methods.

Fitting Eq. (E4) to the data from 4-qubit MRB circuits provides estimates for the entanglement infidelities of all 15 dressed 4-qubit layers used in our experiments. Figure 12 shows the estimated infidelity for each of the nine $\mathbb{G}_1$-dressed layers that consist of a single dressed two-qubit gate applied to one of the three connected qubit pairs (in parallel with dressed idles on the other two qubits). These infidelities are between 2% and 3.1%, and they vary between qubit pairs and between gates (cs, cs† and cz). We quantify the contribution of crosstalk errors to these infidelities by also predicting the infidelities of these 4-qubit layers from the dressed gate error rates obtained from fitting the depolarizing model to the one- and two-qubit data. Shown in Fig. 12, these predicted infidelities are smaller than those estimated from the 4-qubit data by up to 60%. Crosstalk errors are a large proportion of the total infidelity in these 4-qubit layers.

To validate our results, we compare the infidelities we estimate to independent estimates obtained from an established technique: *cycle benchmarking* [9], a technique for estimating the infidelity of individual many-qubit gate layers. Fig. 12 shows that our estimates are broadly similar to those obtained from cycle benchmarking, differing by at most 23%. Moreover, we only expect rough agreement with the cycle benchmarking estimates, for two reasons: (1) the cycle benchmarking experiments were implemented on a different day (they were run immediately after the gates were calibrated), and (2) cycle benchmarking estimates the error rate of layers that are dressed by random Pauli gates (whereas our layers are dressed by Haar-random gates or random single-qubit Clifford gates).

c. Estimating gate error rates by fitting Pauli error models

Fitting data to an n-qubit depolarizing model is scalable, but the actual error map for each layer is unlikely to be global depolarization. For example, a global depolarizing channel causes highly correlated errors, whereas physically we expect many
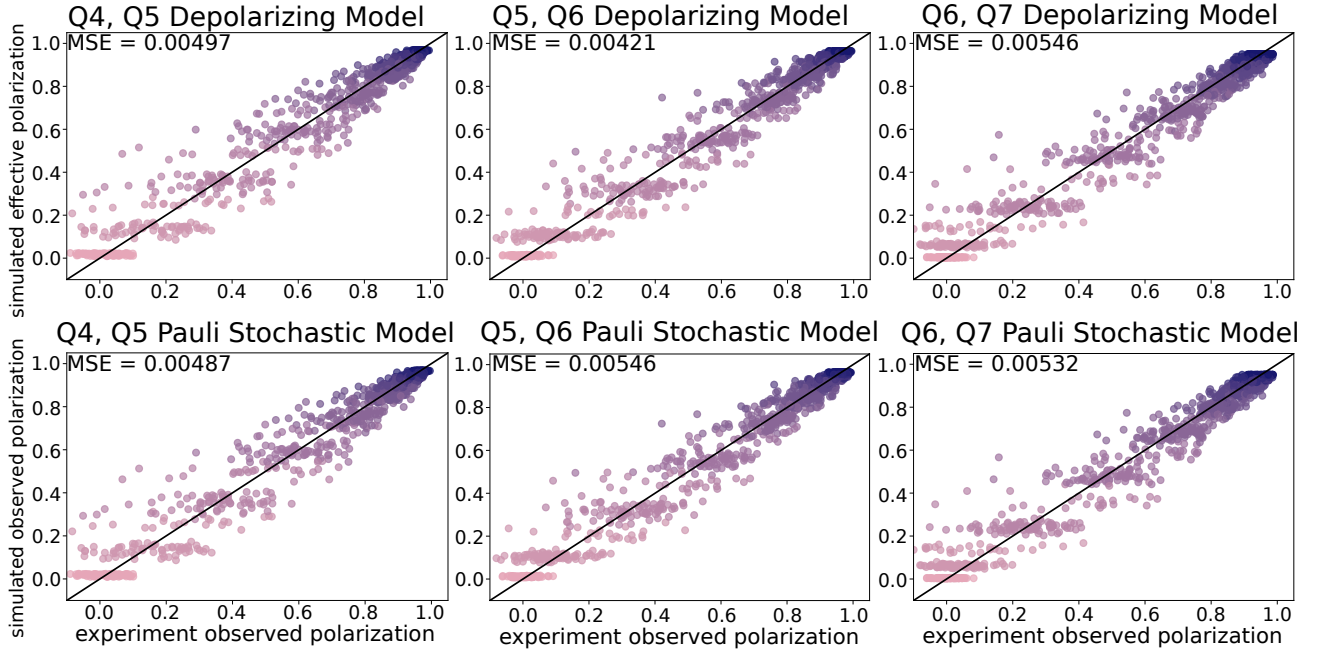
FIGURE 13. Comparing Error Models for 2-Qubit MRB Data. For each pair of qubits we benchmarked, we fit two error models, a depolarizing error model and a Pauli stochastic error model, to the data from the 2-qubit RB experiments. Here, we compare the simulated observed polarization based on the fit models to the observed observed polarization for each circuit.

FIGURE 14. Comparing Error Models for 4-Qubit MRB Data. We fit two error models, a depolarizing error model and a Pauli stochastic error model, to the data from our 4-qubit MRB experiments. Here, we compare the simulated observed polarization based on the fit models to the obs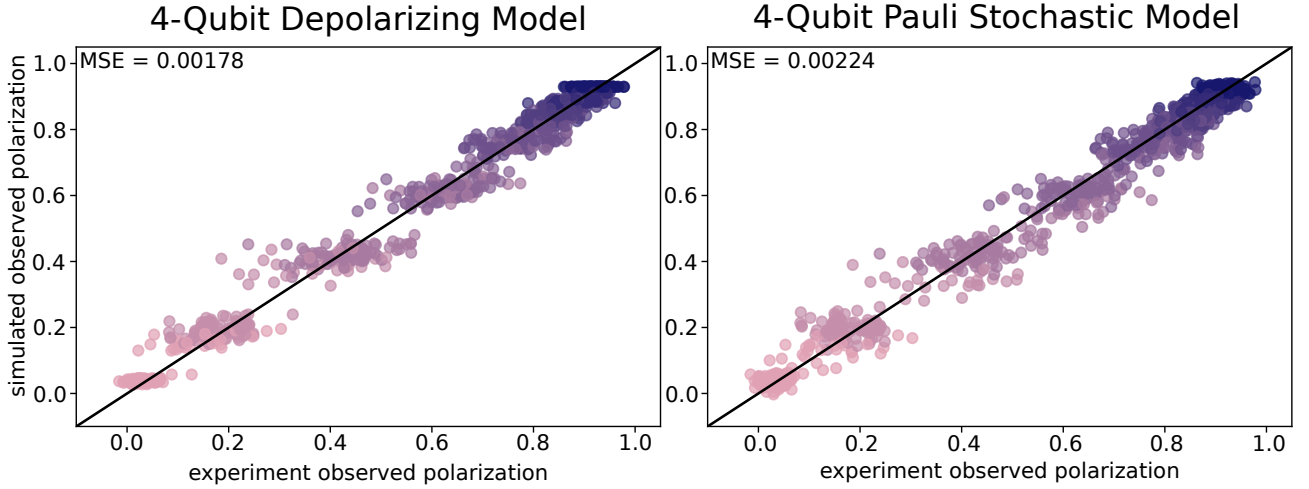erved observed polarization for each circuit. The mean-squared error in the observed polarization is approximately 40% smaller for the Pauli stochastic model than the depolarizing model.

errors to be local errors. We therefore fit a more physically well-motivated model against which to compare our estimates for each dressed layer's infidelity. Arbitrary Markovian errors on a set of n-qubit layers $\mathbb{L}$ can be modelled by an n-qubit process matrix for each $L \in \mathbb{L}$ [1]. But each of these process matrices has $O(16^n)$ parameters, resulting in an infeasible number of parameters to estimate when $n = 4$. Instead, we fit to a process matrix model of reduced complexity. This error model is illustrated in Fig. 11(c). We model the error in each one- or two-qubit native gate (i.e., each $x_{\pi/2}$ etc, not each dressed gate, or each element of $\mathbb{G}_1$) by a one- or two-qubit stochastic Pauli channel [Eq. (6)], respectively. We allow the Pauli error rates $\{\gamma_P\}$ to be gate- and qubit-dependent. We fix the error rates of the z_θ gate on each qubit to zero (because it is a virtual gate). We model state preparation error as a tensor product of single-qubit stochastic Pauli channels before the circuit, and we model measurement error as a tensor product of single-qubit stochastic Pauli channels immediately before readout. We estimate the

error rates of all gates besides z_θ gates (cz, cs, cs† on each of three qubit pairs, and $x_{\pi/2}$ and idle on each of the four qubits) and on state preparation and readout. We fit this model (which has 159 gate error parameters and 24 SPAM error parameters) to the data from the 4-qubit MRB circuits, using maximum likelihood estimation.

The best-fit model contains a process matrix for each gate present in our circuits. These process matrices imply infidelities for each dressed layer. Figure 12 shows the estimated infidelity for each of the 9 four-qubit $\mathbb{G}_1$-dressed layers containing a single two-qubit gate. The estimates we obtained from the stochastic Pauli errors model are comparable to those obtained from the depolarizing model. Maximum likelihood estimation of a Pauli stochastic model using data from general circuits is exponentially expensive in n (due to the circuit simulation cost), whereas fitting a global depolarizing model is not. Note, however, that there are a variety of powerful techniques for efficient estimation of Pauli errors on Clifford gates, using data from Clifford circuits [9–12] (including a technique that uses data from Clifford randomized mirror circuits [12]).

Appendix F: IBM Q experiments

The error rates from all of our MRB experiments on `ibmq_montreal` are shown in Tables II and III.

qubit subset	r_Ω
(Q0, Q1, Q2)	0.88(2)
(Q0, Q1, Q2, Q3)	1.39(3)
(Q0, Q1, Q2, Q3, Q4)	1.96(5)
(Q0, Q1, Q2, Q3, Q4, Q5)	2.53(6)
(Q0, Q1, Q2, Q3, Q4, Q5, Q7)	3.36(8)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6 ,Q7)	6.7(3)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9)	8.4(4)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11)	11.6(4)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14)	14.4(7)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q18)	20.4(6)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q18, Q19, Q20, Q21)	23.6(9)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q18, Q19, Q20, Q21, Q22, Q23, Q24, Q25)	28(1)
(Q0, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q18, Q19, Q20, Q21, Q22, Q23, Q24, Q25, Q26)	27.9(9)

TABLE II. **Many-Qubit MRB on IBM Q.** The MRB error rates for every MRB experiment with $n > 2$ qubits we ran on `ibmq_montreal`. We benchmarked a single qubit subset $\mathbb{Q}$ containing n qubits for 13 exponentially spaced n.

qubit subset	r_Ω (isolated MRB)	r_Ω (simultaneous MRB)
Q0	0.103(3)	0.104(2)
Q1		0.113(3)
Q2		0.106(2)
Q3		0.105(1)
Q4		0.087(2)
Q5		0.113(3)
Q6		0.33(1)
Q7		0.38(1)
Q8		0.19(1)
Q9		0.149(2)
Q10		0.30(1)
Q11		0.135(3)
Q12		0.165(6)
Q13		0.118(4)
Q14		0.106(3)
Q15		0.208(7)
Q16		0.30(2)
Q17		0.127(3)
Q18		0.31(1)
Q19		0.148(4)
Q20		0.097(2)
Q21		0.205(7)
Q22		0.118(3)
Q23		0.140(5)
Q24		0.121(2)
Q25		0.391(9)
Q26		0.168(2)

qubit subset	r_Ω (isolated MRB)	r_Ω (simultaneous MRB)
(Q0, Q1)	0.418(5)	0.82(2)
(Q24, Q25)		2.7(1)
(Q14, Q16)		1.23(5)
(Q18, Q21)		1.61(7)
(Q3, Q5)	0.389(5)	0.75(2)
(Q4, Q7)		1.26(4)
(Q12, Q15)		0.99(3)
(Q19, Q20)		0.72(1)
(Q1, Q2)		0.81(2)
(Q12, Q13)		0.92(3)
(Q22, Q25)		3.15(1)
(Q2, Q3)		0.64(1)
(Q8, Q9)	0.499(8)	0.85(2)
(Q25, Q26)		1.95(6)
(Q1, Q4)		1.34(3)
(Q6, Q7)		3.31(3)
(Q15, Q18)		1.98(7)
(Q23, Q24)		1.19(2)
(Q7, Q10)		1.30(4)
(Q11, Q14)		0.98(2)
(Q16, Q19)	0.77(2)	1.74(6)
(Q21, Q23)		1.45(6)
(Q8, Q11)		1.10(4)
(Q10, Q12)		0.66(1)
(Q19, Q22)		0.78(2)
(Q5, Q8)		1.00(4)
(Q17, Q18)	0.530(8)	1.29(5)
(Q13, Q14)		0.79(2)

TABLE III. **1- and 2-qubit isolated and simultaneous MRB on IBMQ.** We performed simultaneous one-qubit MRB on all 27 individual qubits of ibmq montreal. We also performed simultaneous two-qubit MRB on each connected qubit pair of IBMQ Montreal, in eight groups. We ran isolated MRB experiments on five qubit pairs to compare the error rates from simultaneous and isolated two-qubit mirror RB. Isolated MRB experiments had an error rate approximately 50% smaller than simultaneous MRB experiments.