Direct Randomized Benchmarking for Multiqubit Devices

Timothy J. Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young

Phys. Rev. Lett. **123**, 030503 — Published 19 July 2019

DOI: 10.1103/PhysRevLett.123.030503

Direct randomized benchmarking for multi-qubit devices

Timothy J. Proctor,[1] Arnaud Carignan-Dugas,[2] Kenneth Rudinger,[3]
Erik Nielsen,[3] Robin Blume-Kohout,[3] and Kevin Young[1]

[1] *Quantum Performance Laboratory, Sandia National Laboratories, Livermore, CA 94550, USA*
[2] *Institute for Quantum Computing and the Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[3] *Quantum Performance Laboratory, Sandia National Laboratories, Albuquerque, NM 87185, USA*

(Dated: June 3, 2019)

Benchmarking methods that can be adapted to multi-qubit systems are essential for assessing the overall or "holistic" performance of nascent quantum processors. The current industry standard is Clifford randomized benchmarking (RB), which measures a single error rate that quantifies overall performance. But scaling Clifford RB to many qubits is surprisingly hard. It has only been performed on 1, 2, and 3 qubits as of this writing. This reflects a fundamental inefficiency in Clifford RB: the n-qubit Clifford gates at its core have to be compiled into large circuits over the 1- and 2-qubit gates native to a device. As n grows, the quality of these Clifford gates quickly degrades, making Clifford RB impractical at relatively low n. In this Letter, we propose a *direct* RB protocol that mostly avoids compiling. Instead, it uses random circuits over the native gates in a device, seeded by an initial layer of Clifford-like randomization. We demonstrate this protocol experimentally on 2 – 5 qubits, using the publicly available IBMQX5. We believe this to be the greatest number of qubits holistically benchmarked, and this was achieved on a freely available device without any special tuning up. Our protocol retains the simplicity and convenient properties of Clifford RB: it estimates an error rate from an exponential decay. But it can be extended to processors with more qubits – we present simulations on 10+ qubits – and it reports a more directly informative and flexible error rate than the one reported by Clifford RB. We show how to use this flexibility to measure separate error rates for distinct sets of gates, which includes tasks such as measuring an average CNOT error rate.

With quantum processors incorporating 5 – 20 qubits now commonplace [1–12], and 50+ qubits expected soon [13–15], efficient, holistic benchmarks are becoming increasingly important. Isolated qubits or coupled pairs can be studied in detail with tomographic methods [16–20], but the required resources scale exponentially with qubit number n, making these techniques infeasible for $n \gg 2$ qubits. And while an entire device could be characterized two qubits at a time, this often results in over-optimistic estimates of device performance that ignore crosstalk and collective dephasing effects. What is needed instead is a family of *holistic* benchmarks that quantify the performance of a device as a whole. Randomized benchmarking (RB) methods [21–29] avoid the specific scaling problems that afflict tomography – in RB, both the number of experiments [30] and the complexity of the data analysis [25] are independent of n – but introduce a new scaling problem in the form of *gate compilation*.

Although a quantum processor's native gates typically include only a few one- and two-qubit operations, the "gates" benchmarked by RB are elements of an exponentially large n-qubit group 2-design (e.g., the Clifford group). These gates must be *compiled* into the native gate set [31, 32]. As the number of qubits increases, the circuit depth and infidelity of these compiled group elements grow rapidly, rendering current RB protocols impractical for relatively small n, even with state-of-the-art gates. The industry-standard protocol laid out by Magesan *et al.* [24, 25] – which we will refer to as *Clifford randomized benchmarking* (CRB) – has been widely used to benchmark [33–44] and calibrate [45, 46] both individual qubits and pairs of qubits, but we are aware of just one reported application to three qubits [47], and none to four or more.

Another consequence of compilation is that, instead of quantifying native gate performance, CRB measures the error *per compiled group element*. Although this is sometimes translated into a native gate error rate, e.g., by dividing it by the average circuit size of a compiled Clifford [42–44], this is ad hoc and not always reliable [48]. Moreover, error rates obtained this way are hard

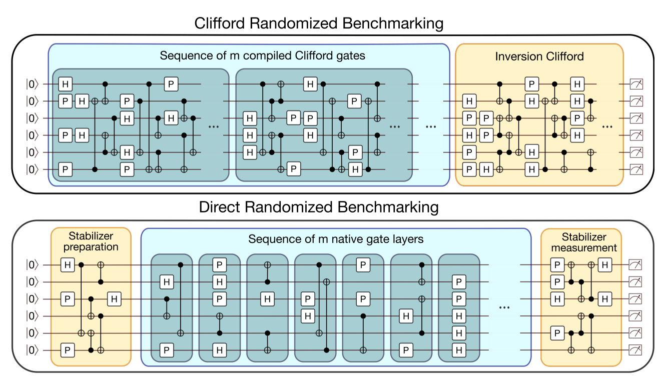


FIG. 1. A cartoon illustrating the circuits used in Clifford RB and the streamlined *direct* RB protocol that we propose.

to interpret for $n \gg 1$ CRB, where error rates can vary widely between native gates.

In this Letter we propose and demonstrate *direct randomized benchmarking* (DRB), an RB protocol that directly benchmarks the native gates of a device. Like CRB, our DRB protocol utilizes random circuits of variable length, but these circuits consist of the native gates of the device, rather than compiled Clifford operations (see Fig. 1). Our protocol is not infinitely scalable, but the simplified structure enables DRB to be successfully implemented on significantly more qubits than CRB. Moreover, DRB preserves the core simplicity of CRB: it estimates an error rate from an exponential decay.

We anticipate that DRB will be an important tool for characterizing current multi-qubit devices. For this reason, this Letter focuses on the practical applications of DRB. We present experiments on $2-5$ qubits and simulations on $2-10+$ qubits. These examples show that DRB works, demonstrate how our protocol improves on current methods, and show that DRB can be implemented on significantly more than two qubits on current devices. We follow these demonstrations with arguments for why DRB is broadly reliable, but this Letter does not contain a *comprehensive* theory for DRB – that will be presented in a series of future papers.

Direct randomized benchmarking – DRB is a protocol to directly benchmark the native gates in a device. There is flexibility in defining a device's "native gates". For DRB we only require that they generate the n-qubit Clifford group $\mathbb{C}_n$ [49]. Normally, they will be all the n-qubit Clifford operations that can be implemented by depth-1 circuits, e.g., by parallel 1- and 2-qubit gates (see Fig. 1). We call these *circuit layers* or (n-qubit) *native gates*.

Just as CRB uses sequences of random Cliffords, DRB uses sequences of random circuit layers. But whereas the Cliffords in CRB are supposed to be uniformly random, DRB allows the circuit layers to be sampled according to a *user-specified* probability distribution Ω. Many distributions are permissible, but, to ensure reliability, Ω must have support on a subset of the gates that generates $\mathbb{C}_n$ and Ω-random circuits must quickly spread errors (see later).

The n-qubit DRB protocol is defined as follows (note that all operations are assumed to be imperfect):

1. For a range of lengths $m \geq 0$, repeat the following $k_m \gg 1$ times:
   1.1. Sample a uniformly random n-qubit stabilizer state $|\psi\rangle$.
   1.2. Sample an m-layer circuit $\mathcal{U}_m$, where each layer is drawn independently from some user-specified distribution Ω over all n-qubit native gates.
   1.3. Repeat the following $N \geq 1$ times:
      1.3.1 Initialize the qubits in $|0\rangle^{\otimes n}$.
      1.3.2 Implement a circuit to map $|0\rangle^{\otimes n} \to |\psi\rangle$.
      1.3.3 Implement the sampled $\mathcal{U}_m$ circuit.
      1.3.4 Implement a circuit that maps $\mathcal{U}_m|\psi\rangle$ to a known computational basis state $|s\rangle$.
      1.3.5 Measure all n qubits and record whether the outcome is s (success) or not (failure).
2. Calculate the average probability of success P_m at each length m, averaged over the k_m randomly sampled circuits and the N trials for each circuit.
3. Fit P_m to $P_m = A + Bp^m$, where A, B and p are fit parameters.
4. The Ω-averaged DRB error rate of the native gates is $r = (4^n - 1)(1-p)/4^n$.

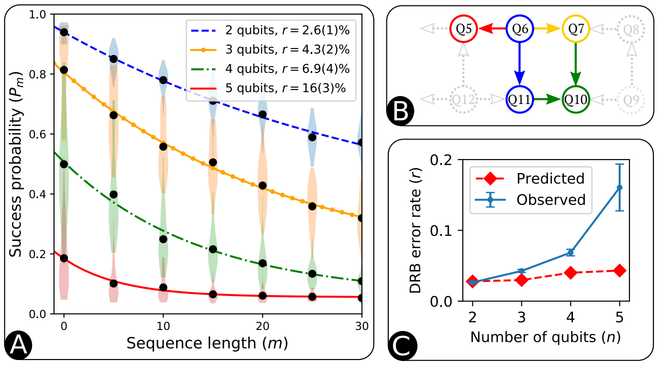
The n-dependent rescaling used above is different from that in common usage [23–25]. Using our convention, r corresponds to the probability of an error when the errors are stochastic (see later). This is particularly convenient when varying n.

DRB is similar to the earliest implementations of RB. Both the 1-qubit RB experiments of Knill *et al.* [23] and the 3-qubit experiments of Ryan *et al.* [50] utilize random sequences of group generators, and so are specific examples of DRB *without* the stabilizer state preparation step and flexible sampling. These additional features, however, are essential to DRB: they make DRB provably reliable under broad conditions, and allow us to separate the error rate into contributions from distinct sets of gates.

What DRB measures – To interpret DRB results it is important to understand what DRB measures. Assume that the gate errors are stochastic, which can be enforced to a good approximation by, e.g., Pauli-frame randomization [51–53] or by following each layer in DRB with a random n-qubit Pauli gate. Then, whenever Ω-random circuits quickly increase the weight of errors, r is a good estimate of the probability that an error happens on a Ω-average native gate. That is, $r \approx \epsilon_\Omega \equiv \sum_i \Omega(\mathcal{G}_i)\epsilon_i$, where ϵ_i is the probability of an error on the n-qubit native gate $\mathcal{G}_i$. Later, we derive this relationship.

Because r depends on the sampling distribution, they should be reported together. A similar, but hidden variability also exists in CRB – the CRB r depends on the Clifford compiler. This compiler-dependence in CRB is inconvenient, as the properties of multi-qubit Clifford compilers are difficult to control. In contrast, because we directly choose Ω, we can control how often each gate appears in the random circuits, to estimate error rates of particular interest.

Experiments on 2 – 5 qubits – To demonstrate that DRB is useful and behaves correctly on current multi-qubit devices, we used it to benchmark $2-5$ qubit subsets of the publicly accessible IBMQX5 [1, 2]. The IBMQX5



D

Number of qubits in circuits	2	3	4	5
Qubits in circuits	Q6+Q11	+Q7	+Q10	+Q5
Mean CNOT ϵ estimated via n-qubit DRB (%)	3.4(1)	5.4(4)	8.2(9)	20(6)
Mean CNOT ϵ estimated via 2-qubit CRB (%)	3.6	3.4	4.4	4.5

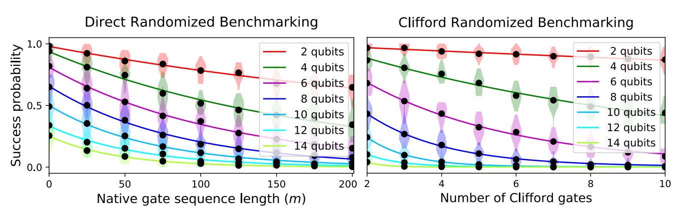
FIG. 2. Experimental 2 – 5 qubit DRB on IBMQX5. **A.** Success probability decays. The points are average success probabilities P_m, and the violin plots show the distributions of the success probabilities at each length over circuits (there are 28 circuits per length). The curves are obtained from fitting to $P_m = A + Bp^m$, and $r = (4^n - 1)(1-p)/4^n$. **B.** A schematic of IBMQX5. The colors match those in A and correspond to the additional qubits/CNOTs added from $n \to n+1$ qubit DRB (see also D). **C.** Observed r versus n, and predictions from 1- and 2-qubit CRB calibration data. **D.** Estimates of the average CNOT error rate in n-qubit circuits, obtained by comparing the data in A with additional DRB data that used circuits with fewer CNOTs per layer.

native gates comprise CNOTs and arbitrary 1-qubit gates [2, 54]; we benchmarked a set of n-qubit gates consisting of parallel applications of all directly available CNOTs and all 1-qubit Clifford gates.

Fig. 2 summarizes our results. Fig. 2 A demonstrates that DRB was successful on 2 – 5 qubits: an exponential decay is observed and r is estimated with reasonable precision (bootstrapped 2σ uncertainties are shown). To our knowledge, this is the largest number of qubits holistically benchmarked to date, which was made possible by the streamlined nature of DRB (see Fig. 1). To interpret these results it is necessary to specify the circuit sampling. Each layer was sampled as follows: with probability p_{CNOT} we uniformly choose one of the CNOTs and add it to the sampled layer; for all n or $n-2$ remaining qubits we independently and uniformly sample a 1-qubit gate and add it to the layer. For the data in Fig. 2 A, $p_{\text{CNOT}} = 0.75$. We also implemented experiments with $p_{\text{CNOT}} = 0.25$; see the Supplemental Material [55] for this data and further experimental details.

Using this sampling, the average number of CNOTs per layer is p_{CNOT}, independent of n. Therefore r would vary little with n if CNOT errors dominate, the error rates are reasonably uniform over the CNOTs, and n-qubit benchmarks are predictive of benchmarks on more than n qubits. Instead, the observed r increases quickly with n. This is quantified in Fig. 2 C, where we compare each observed r to a prediction r_{cal} obtained from the IBMQX5 CRB calibration data (1-qubit error rates from simultaneous 1-qubit CRB [2, 56, 57] and CNOT error rates from CRB on isolated pairs [56]). These predictions are calculated both using $r \approx \epsilon_\Omega$ and via a DRB simulation using a crosstalk-free error model that is consistent with the calibration data. Both methods agree, confirming that the increase in r with n is *not* due to a failure of DRB. For $n = 2$, r_{cal} and r are similar, demonstrating that n-qubit DRB and CRB are consistent. But, as n increases, r diverges from r_{cal}. This shows that the effective error rates of the 1-qubit and/or 2-qubit gates in the device change as we implement circuits over more qubits, demonstrating that $n > 2$-qubit DRB can detect errors that are not predicted by 1- and 2-qubit CRB (calibration data) or 2-qubit DRB (our data). This highlights the value of holistic benchmarking for multi-qubit devices.

Using the data from Fig. 2 A ($p_{\text{CNOT}} = 0.75$) alongside additional data with $p_{\text{CNOT}} = 0.25$ sampling [58], we can estimate the average CNOT error rate in n-qubit circuits. For each n and using $r \approx \sum_i \Omega(\mathcal{G}_i)\epsilon_i$, we have $\vec{r} \approx M\vec{\epsilon}$ where: $\vec{r} = (r_{0.75}, r_{0.25})$ with $r_{0.75}$ (resp., $r_{0.25}$) the r obtained with $p_{\text{CNOT}} = 0.75$ (resp., $p_{\text{CNOT}} = 0.25$) sampling; $\vec{\epsilon} = (\epsilon_A, \epsilon_B)$ with ϵ_A (resp., ϵ_B) the average error rate of those n-qubit gates containing one CNOT in parallel with 1-qubit gates on the other qubits (resp., n parallel 1-qubit gates); $M = \frac{1}{4}\left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix}\right)$. Therefore, ϵ_A and ϵ_B can be estimated using $\vec{\epsilon} = M^{-1}\vec{r}$, and so – by estimating the average 1-qubit gate error rate from ϵ_B and removing this contribution from ϵ_A – we can estimate the mean CNOT error rate versus n. Estimates are given in Fig. 2 D. For two qubits, our estimate of the CNOT error rate is similar to the prediction from the calibration data, so our methodology seems consistent with CRB techniques. In contrast, our results show that CNOTs perform substantially worse in $n > 2$ qubit circuits than in 2-qubit circuits. This is likely due to CNOT crosstalk, i.e., CNOTs affect "spectator" qubits.



Number of qubits	2	4	8	14
Ω-averaged error rate ϵ_Ω (%)	0.300	0.599	1.19	2.08
DRB r (%)	0.29(1)	0.59(2)	1.17(4)	2.2(2)
CRB r (%)	1.8(2)	9.9(4)	38(2)	Failed
CNOT-rescaled CRB r (%)	0.83(9)	1.02(4)	0.97(5)	Failed
Depth-rescaled CRB r (%)	0.16(2)	0.39(2)	0.65(3)	Failed

FIG. 3. Simulation of DRB and CRB for 2 – 14 qubits with a simple error model. The n-qubit DRB error rate is $r \approx n \times 0.15\%$, consistent with the simulated sampling-averaged native gate error rate ϵ_Ω.

DRB simulations – We have shown that DRB works on current multi-qubit devices, and so we now demonstrate with simulations that $r \approx \epsilon_\Omega \equiv \sum_i \Omega(\mathcal{G}_i)\epsilon_i$. Assume n qubits with native gates consisting of parallel CNOT, idle I, Hadamard H and phase P gates ($P|x\rangle = i^x|x\rangle$), and all-to-all connectivity. We model gate errors by assuming that after each CNOT (resp., 1-qubit gate) the qubits involved in the gate are independently, with probability 0.25% (resp., 0.05%), subject to a random σ_x, σ_y or σ_z error. So the CNOT error rate is ≈ 0.5%. We simulated DRB with a sampling distribution defined by randomly pairing up the qubits, applying a CNOT to a pair with probability $\tilde{p}_{\text{CNOT}} = 0.5$, and applying uniformly random 1-qubit gates (H, P or I) to all qubits that do not have a CNOT acting on them. Fig. 3 shows simulated 2 – 14 qubit DRB and CRB data. DRB has succeeded: the decay is exponential and $r \approx \epsilon_\Omega = 1-(0.5\times0.9975^2+0.5\times0.9995^2)^{\frac{n}{2}} \approx n\times0.15\%$. In contrast, the CRB r grows rapidly with n – for only 4-qubit CRB $r \approx 10\%$ – and CRB fails for $n > 12$, demonstrating that DRB can be implemented on more qubits than CRB. Moreover, the CRB error rates r rescaled to $r_{\text{RCRB}} = 1-(1-r)^{1/\alpha}$ [42–44, 48], with α the average compiled Clifford circuit-depth or CNOT-count, are not simple functions of the native gate error rates (see Fig. 3).

This example is illustrative, but simplistic. So, in the Supplemental Material we present additional simulations with large, non-uniform CNOT error rates, and limited qubit connectivity. We also simulate the CNOT error rate estimation method used on the IBMQX5 data, validating the technique.

DRB theory – We now provide a theory for DRB of gates with Pauli-stochastic errors. DRB circuits consist of preparing a uniformly random n-qubit stabilizer state ψ, a circuit $\mathcal{U}_m = \mathcal{G}_{s_m}\cdots\mathcal{G}_{s_1}$ with m layers $\mathcal{G}_i$ sampled according to Ω, and a stabilizer measurement projecting onto $\mathcal{U}_m|\psi\rangle$. For now, assume that the stabilizer state preparation and measurement (SSPAM) are perfect. In the stochastic error model, each time $\mathcal{U}_m$ is applied there is some faulty implementation, $\tilde{\mathcal{U}}_m = \mathcal{P}_{s_m}\mathcal{G}_{s_m}\cdots\mathcal{P}_{s_1}\mathcal{G}_{s_1}$, with $\mathcal{P}_{s_i}$ some Pauli error or the identity. DRB aims to capture the rate that these $\mathcal{P}_{s_i}$ deviate from the identity. Because ψ is a stabilizer state, the measurement will register success iff one of the following holds:

(S1) No errors occur in $\tilde{\mathcal{U}}_m$, i.e., all $\mathcal{P}_i = \mathbb{1}$.

(S2) 2+ errors occur in $\tilde{\mathcal{U}}_m$, but when they propagate through the circuit they cancel, i.e., multiple $\mathcal{P}_i \neq \mathbb{1}$ but $\tilde{\mathcal{U}}_m = i^k\mathcal{U}_m$ (for some $k = 0, 1, 2, 3$).

(S3) 1+ errors occur in $\tilde{\mathcal{U}}_m$ that do not cancel, but they are nonetheless unobserved by the stabilizer measurement, i.e., $\tilde{\mathcal{U}}_m \neq i^k\mathcal{U}_m$ but $\tilde{\mathcal{U}}_m|\psi\rangle = i^l\mathcal{U}_m|\psi\rangle$.

The DRB average success probability P_m is obtained by averaging $P(\mathcal{U}_m, \psi) = |\langle\psi|\mathcal{U}_m^\dagger\tilde{\mathcal{U}}_m|\psi\rangle|^2$ over the possible Pauli errors, ψ and $\mathcal{U}_m$. We may then write $P_m = s_1 + (1-s_1)(s_2 + (1-s_2)s_3)$, where s_1 is the probability of S1, i.e., no errors, s_2 is the probability of S2 conditioned on 1+ errors occurring, and s_3 is the probability of S3 conditioned on 1+ errors occurring and the errors not canceling. Because ϵ_Ω is the Ω-averaged error rate per layer, $s_1 = (1-\epsilon_\Omega)^m$. A uniformly random stabilizer state ψ is an eigenstate of *any* Pauli error with probability $(2^n-1)/(4^n-1)$, so $s_3 = (2^n-1)/(4^n-1) \approx 2^{-n}$. This is one of the motivations for the state preparation step in DRB.

In order to understand the effect of s_2 on P_m, we consider two regimes: small n ($\lesssim 3$ qubits) and not-so-small n ($\gtrsim 3$). In both regimes, we expect Pauli errors to occurring at most once every several layers and to be *low-weight*, with support on only a few qubits. In the not-so-small n regime, errors propagating through a sequence of one- and two-qubit gates are likely to quickly increase in weight [59–61] (due to the demands we made of Ω earlier). Subsequent errors are therefore very unlikely to cause error cancellation. If each layer is a uniformly random Clifford (as in uncompiled CRB), any Pauli error is randomized to one of the 4^n-1 possible n-qubit Pauli errors at each step. So the probability that another error cancels with an earlier error is $\simeq 1/4^n$, implying that $s_2 \lesssim 1/4^n$. In DRB, we expect error cancellation at a rate only slightly above this. Therefore, s_2 contributes negligibly to P_m, so $P_m \approx 2^n + (1-2^n)(1-\epsilon_\Omega)^m$. This is an exponential with decay rate ϵ_Ω. Verifying this error scrambling process for a given sampling distribution, Ω, is computationally efficient in qubit number. Distributions that do *not* scramble the errors quickly (e.g., if 2-qubit gates are rare) can yield decays that are not simple exponentials. These should be avoided.

For small n, the probability of cancellation (s_2) is *not* negligible for *any* distribution. But because n is small, we only need a few random circuit layers of Clifford-group generators to implement approximate Clifford twirling, so P_m may be computed using the resulting effective depolarizing channel. Such channels are well-known to lead to exponential decays [25]. However, s_2 (a function of m) now contributes significantly to the DRB decay constant p, so $p \not\approx 1-\epsilon_\Omega$. This motivates $r = (4^n-1)(1-p)/4^n$, which removes the unwanted s_2 contribution in $1-p$. Let each layer be followed by a depolarizing map $\mathcal{D}_\lambda$ where $\mathcal{D}_\lambda[\rho] = \lambda\rho + (1-\lambda)\mathbb{1}/2^n$. Then $P_m = (1-2^{-n})\lambda^m + 2^{-n}$, but the error rate of $\mathcal{D}_\lambda$ is $\epsilon = (4^n-1)(1-\lambda)/4^n$. Of course, in the large-n limit, $\epsilon \to 1-\lambda$.

Above, we assumed perfect SSPAM which is unrealistic. The SSPAM operations are *almost* m-independent, and so errors in SSPAM are almost entirely absorbed into A and B in $P_m = A + Bp^m$, as normal in RB [24, 25]. The only m-dependent impact is from correlations in the stabilizer state that is prepared and measured – they are perfectly correlated (resp., uncorrelated) at $m = 0$ (resp., $m \to \infty$). This causes an inconsequentially small tendency to over-estimate the gate error rate – because SS-

PAM contributes an error of $1 - \mathrm{avg}_i[(1-\epsilon_{i,\mathrm{SSPAM}})^2]$ at $m = 0$ but a smaller error of $1 - (\mathrm{avg}_i[1-\epsilon_{i,\mathrm{SSPAM}}])^2$ at $m \to \infty$, where $\epsilon_{i,\mathrm{SSPAM}}$ is the error in creating or measuring the i^{th} stabilizer state.

DRB remains effective with coherent errors – with any 1-qubit gates that generate the 1-qubit Clifford group, independently random 1-qubit gates on each qubit are sufficient to quickly twirl coherent errors to Pauli-stochastic errors, implying that errors can only coherently combine between a few layers in a DRB circuit (in contrast to the uncontrolled coherent addition *within* a compiled Clifford gate in CRB). But linking r to a formal notion of gate error rate is subtler with coherent errors, in direct analogy with CRB [62–64], as will be discussed in future work.

Conclusions – Benchmarking methods for multi-qubit systems are essential for assessing the performance of current and near-term quantum processors. But currently there are no reliable methods that can be easily and routinely applied to more than two qubits with current device performance. In this Letter we have introduced and demonstrated *direct randomized benchmarking* (DRB), a method that streamlines the industry-standard Clifford randomized benchmarking (CRB) technique [24, 25] so that it can be applied to more qubits. DRB retains the core simplicity of CRB, our protocol directly measures the quantities of most interest – the error rates of the native gates in a device – and it is user-configurable, allowing a variety of important error rates to be estimated. Our experimental demonstrations were on 2 – 5 qubits, and, using a publicly accessible device [1, 2], set a record for the number of qubits holistically benchmarked. The tools we used are available as open-source code [65], and support any device connectivity. So, we anticipate that 5 – 10+ qubits will soon be benchmarked with our protocol, providing important insights into state-of-the-art device performance. Finally, the techniques of DRB can also be applied to extend and improve the full suite of RB methods [26, 27, 57, 66–74], and varied-sampling DRB provides an alternative to both interleaved CRB [74] and "interleaved DRB" for estimating individual error rates, demonstrating the broad applicability and impact of DRB.

Acknowledgements –

TJP thanks Scott Aaronson for helpful correspondence on how to efficiently generate random stabilizer states. KMR thanks Jay Gambetta and Diego Moreda for QISKit support. We acknowledge use of the IBM Q Experience for this work. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525. This research was funded, in part, by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

[1] IBM, "IBM Quantum Experience," URL: https://quantumexperience.ng.bluemix.net/qx/devices.
[2] Baleegh Abdo, Vivekananda Adiga, Lev Bishop, Markus Brink, Nicholas Bronn, Jerry Chow, Antonio Córcoles, Andrew Cross, Jay M. Gambetta, Jose Chavez-Garcia, Jared Hertzberg, Oblesh Jinka, George Keefe, David McKay, Salvatore Olivadese, Jason Orcutt, Hanhee Paik, Jack Rohrs, Sami Rosenblatt, Jim Rozen, Martin Sandberg, Dongbing Shao, Sarah Sheldon, Firat Solgun, and Maika Takita, "IBM Q Experience backend information," URL: https://github.com/QISKit/ibmqx-backend-information.
[3] Norbert M Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A Landsman, Kenneth Wright, and Christopher Monroe, "Experimental comparison of two quantum computing architectures," Proc. Natl. Acad. Sci. U.S.A. **114**, 3305–3310 (2017).
[4] C Figgatt, Dmitri Maslov, KA Landsman, NM Linke, S Debnath, and C Monroe, "Complete 3-qubit grover search on a programmable quantum computer," Nature communications **8**, 1918 (2017).
[5] JS Otterbach, R Manenti, N Alidoust, A Bestwick, M Block, B Bloom, S Caldwell, N Didier, E Schuyler Fried, S Hong, *et al.*, "Unsupervised machine learning on a hybrid quantum computer," arXiv preprint arXiv:1712.05771 (2017).
[6] Nicolai Friis, Oliver Marty, Christine Maier, Cornelius Hempel, Milan Holzäpfel, Petar Jurcevic, Martin B Plenio, Marcus Huber, Christian Roos, Rainer Blatt, *et al.*, "Observation of entangled states of a fully controlled 20-qubit system," Phys. Rev. X **8**, 021012 (2018).
[7] CA Riofrío, D Gross, ST Flammia, T Monz, D Nigg, R Blatt, and J Eisert, "Experimental quantum compressed sensing for a seven-qubit system," Nat. Commun. **8**, 15305 (2017).
[8] Julian Kelly, R Barends, AG Fowler, A Megrant, E Jeffrey, TC White, D Sank, JY Mutus, B Campbell, Yu Chen, *et al.*, "State preservation by repetitive error detection in a superconducting quantum circuit," Nature **519**, 66 (2015).
[9] C Neill, P Roushan, K Kechedzhi, S Boixo, SV Isakov, V Smelyanskiy, A Megrant, B Chiaro, A Dunsworth, K Arya, *et al.*, "A blueprint for demonstrating quantum supremacy with superconducting qubits," Science **360**, 195–199 (2018).
[10] Chao Song, Kai Xu, Wuxin Liu, Chui-ping Yang, Shi-Biao Zheng, Hui Deng, Qiwei Xie, Keqiang Huang, Qi-

ujiang Guo, Libo Zhang, *et al.*, "10-qubit entanglement and parallel logic operations with a superconducting circuit," Phys. Rev. Lett. **119**, 180511 (2017).

[11] X Fu, MA Rol, CC Bultink, J van Someren, N Khammassi, I Ashraf, RFL Vermeulen, JC De Sterke, WJ Vlothuizen, RN Schouten, *et al.*, "An experimental microarchitecture for a superconducting quantum processor," in *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture* (ACM, 2017) pp. 813–825.

[12] Intel, "Intel delivers 17-qubit superconducting chip with advanced packaging to qutech," (2018), URL: https://newsroom.intel.com/news/intel-advances-quantum-neuromorphic-computing-research/.

[13] Julian Kelly, "Engineering superconducting qubit arrays for quantum supremacy," Bulletin of the American Physical Society (2018).

[14] IBM, "IBM announces advances to IBM quantum systems & ecosystem," (2017), URL: https://www-03.ibm.com/press/us/en/pressrelease/53374.wss.

[15] Intel, "2018 CES: Intel advances quantum and neuromorphic computing research," (2018), URL: https://newsroom.intel.com/news/intel-advances-quantum-neuromorphic-computing-research/.

[16] Seth T Merkel, Jay M Gambetta, John A Smolin, Stefano Poletto, Antonio D Córcoles, Blake R Johnson, Colm A Ryan, and Matthias Steffen, "Self-consistent quantum process tomography," Phys. Rev. A **87**, 062119 (2013).

[17] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz, "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography," Nat. Commun. **8**, 14485 (2017).

[18] Daniel Greenbaum, "Introduction to quantum gate set tomography," arXiv preprint arXiv:1509.02921 (2015).

[19] Shelby Kimmel, Guang Hao Low, and Theodore J Yoder, "Robust calibration of a universal single-qubit gate set via robust phase estimation," Phys. Rev. A **92**, 062315 (2015).

[20] Kenneth Rudinger, Shelby Kimmel, Daniel Lobser, and Peter Maunz, "Experimental demonstration of cheap and accurate phase estimation," Phys. Rev. Lett. **118**, 190502 (2017).

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

[22] 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).

[23] 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).

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

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

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

[27] 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).

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

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

[30] Jonas Helsen, Joel J. Wallman, Steven T. Flammia, and Stephanie Wehner, "Multi-qubit randomized benchmarking using few samples," arXiv preprint arXiv:1701.04299 (2017).

[31] Scott Aaronson and Daniel Gottesman, "Improved simulation of stabilizer circuits," Phys. Rev. A **70**, 052328 (2004).

[32] Ketan N Patel, Igor L Markov, and John P Hayes, "Efficient synthesis of linear reversible circuits," Quantum Inf. Comput. **8**, 282–294 (2008).

[33] Jun Yoneda, Kenta Takeda, Tomohiro Otsuka, Takashi Nakajima, Matthieu R Delbecq, Giles Allison, Takumu Honda, Tetsuo Kodera, Shunri Oda, Yusuke Hoshi, *et al.*, "A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%," Nat. Nanotechnol. **13**, 102 (2018).

[34] David M Zajac, Anthony J Sigillito, Maximilian Russ, Felix Borjans, Jacob M Taylor, Guido Burkard, and Jason R Petta, "Resonantly driven CNOT gate for electron spins," Science **359**, 439–442 (2018).

[35] TF Watson, SGJ Philips, Erika Kawakami, DR Ward, Pasquale Scarlino, Menno Veldhorst, DE Savage, MG Lagally, Mark Friesen, SN Coppersmith, *et al.*, "A programmable two-qubit quantum processor in silicon," Nature **555**, 633 (2018).

[36] John M Nichol, Lucas A Orona, Shannon P Harvey, Saeed Fallahi, Geoffrey C Gardner, Michael J Manfra, and Amir Yacoby, "High-fidelity entangling gate for double-quantum-dot spin qubits," NPJ Quantum Inf. **3**, 3 (2017).

[37] M Veldhorst, JCC Hwang, CH Yang, AW Leenstra, B De Ronde, JP Dehollain, JT Muhonen, FE Hudson, Kohei M Itoh, A Morello, *et al.*, "An addressable quantum dot qubit with fault-tolerant control-fidelity," Nat. Nanotechnol. **9**, 981–985 (2014).

[38] Antonio D Córcoles, Jay M Gambetta, Jerry M Chow, John A Smolin, Matthew Ware, Joel Strand, Britton LT Plourde, and Matthias Steffen, "Process verification of two-qubit quantum gates by randomized benchmarking," Phys. Rev. A **87**, 030301 (2013).

[39] T Xia, M Lichtman, K Maller, AW Carr, MJ Piotrowicz, L Isenhower, and M Saffman, "Randomized benchmarking of single-qubit gates in a 2D array of neutral-atom qubits," Phys. Rev. Lett. **114**, 100503 (2015).

[40] AD Córcoles, Easwar Magesan, Srikanth J Srinivasan, Andrew W Cross, M Steffen, Jay M Gambetta, and Jerry M Chow, "Demonstration of a quantum error detection code using a square lattice of four superconducting qubits," Nat. Commun. **6**, 7979 (2015).

[41] Zijun Chen, Julian Kelly, Chris Quintana, R Barends, B Campbell, Yu Chen, B Chiaro, A Dunsworth, AG Fowler, E Lucero, *et al.*, "Measuring and suppress-

ing quantum state leakage in a superconducting qubit," Phys. Rev. Lett. **116**, 020501 (2016).

[42] JT Muhonen, A Laucht, S Simmons, JP Dehollain, R Kalra, FE Hudson, S Freer, Kohei M Itoh, DN Jamieson, JC McCallum, *et al.*, "Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking," J. Phys. Condens. Matter **27**, 154205 (2015).

[43] R Barends, J Kelly, A Megrant, A Veitia, D Sank, E Jeffrey, TC White, J Mutus, AG Fowler, B Campbell, *et al.*, "Superconducting quantum circuits at the surface code threshold for fault tolerance," Nature **508**, 500–503 (2014).

[44] J Raftery, A Vrajitoarea, G Zhang, Z Leng, SJ Srinivasan, and AA Houck, "Direct digital synthesis of microwave waveforms for quantum computing," arXiv preprint arXiv:1703.00942 (2017).

[45] MA Rol, CC Bultink, TE O'Brien, SR de Jong, LS Theis, X Fu, F Luthi, RFL Vermeulen, JC de Sterke, A Bruno, *et al.*, "Restless tuneup of high-fidelity qubit gates," Phys. Rev. Appl. **7**, 041001 (2017).

[46] J Kelly, R Barends, B Campbell, Y Chen, Z Chen, B Chiaro, A Dunsworth, AG Fowler, I-C Hoi, E Jeffrey, *et al.*, "Optimal quantum control using randomized benchmarking," Phys. Rev. Lett. **112**, 240504 (2014).

[47] David C McKay, Sarah Sheldon, John A Smolin, Jerry M Chow, and Jay M Gambetta, "Three qubit randomized benchmarking," Phys. Rev. Lett. **122**, 200502 (2019).

[48] Jeffrey M Epstein, Andrew W Cross, Easwar Magesan, and Jay M Gambetta, "Investigating the limits of randomized benchmarking protocols," Phys. Rev. A **89**, 062321 (2014).

[49] Extensions to other group 2-designs are possible.

[50] CA Ryan, Martin Laforest, and Raymond Laflamme, "Randomized benchmarking of single-and multi-qubit control in liquid-state nmr quantum information processing," New J. Phys. **11**, 013034 (2009).

[51] Emanuel Knill, "Quantum computing with realistically noisy devices," Nature **434**, 39–44 (2005).

[52] Matthew Ware, Guilhem Ribeill, Diego Riste, Colm A Ryan, Blake Johnson, and Marcus P da Silva, "Experimental demonstration of pauli-frame randomization on a superconducting qubit," arXiv preprint arXiv:1803.01818 (2018).

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

[54] "Qiskit," URL: https://qisqkit.org.

[55] See Supplemental Material for additional data and details, which includes Refs. [56, 75–77].

[56] Private communication with the IBMQX team.

[57] 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).

[58] This data was obtained after the device had been recalibrated. This is not ideal, but the error rates did not change substantially; details in the Supplemental Material.

[59] Winton Brown and Omar Fawzi, "Decoupling with random quantum circuits," Commun. Math. Phys. **340**, 867–900 (2015).

[60] Yasuhiro Sekino and Leonard Susskind, "Fast scramblers," J. High Energy Phys. **2008**, 065 (2008).

[61] Fernando GSL Brandão, Aram W Harrow, and Michał Horodecki, "Local random quantum circuits are approximate polynomial-designs," Commun. Math. Phys. **346**, 397–434 (2016).

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

[63] Joel J Wallman, "Randomized benchmarking with gate-dependent noise," Quantum **2**, 47 (2018).

[64] 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 J. Phys. **20**, 092001 (2018).

[65] Erik Nielsen, Robin Blume-Kohout, Lucas Saldyt, Jonathan Gross, Travis Scholten, Kenneth Rudinger, Timothy Proctor, and John King Gamble, "PyGSTi version 0.9.7.2," (2019), 10.5281/zenodo.2598776.

[66] Christopher J Wood and Jay M Gambetta, "Quantification and characterization of leakage errors," Phys. Rev. A **97**, 032306 (2018).

[67] T Chasseur and FK Wilhelm, "Complete randomized benchmarking protocol accounting for leakage errors," Phys. Rev. A **92**, 042333 (2015).

[68] Joel J Wallman, Marie Barnhill, and Joseph Emerson, "Robust characterization of loss rates," Phys. Rev. Lett. **115**, 060501 (2015).

[69] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia, "Estimating the coherence of noise," New J. Phys. **17**, 113020 (2015).

[70] Guanru Feng, Joel J Wallman, Brandon Buonacorsi, Franklin H Cho, Daniel K Park, Tao Xin, Dawei Lu, Jonathan Baugh, and Raymond Laflamme, "Estimating the coherence of noise in quantum control of a solid-state qubit," Phys. Rev. Lett. **117**, 260501 (2016).

[71] Sarah Sheldon, Lev S Bishop, Easwar Magesan, Stefan Filipp, Jerry M Chow, and Jay M Gambetta, "Characterizing errors on qubit operations via iterative randomized benchmarking," Phys. Rev. A **93**, 012301 (2016).

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

[73] Tobias Chasseur, Daniel M Reich, Christiane P Koch, and Frank K Wilhelm, "Hybrid benchmarking of arbitrary quantum gates," Phys. Rev. A **95**, 062335 (2017).

[74] 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).

[75] Jeroen Dehaene and Bart De Moor, "Clifford group, stabilizer states, and linear and quadratic operations over GF(2)," Phys. Rev. A **68**, 042318 (2003).

[76] Erik Hostens, Jeroen Dehaene, and Bart De Moor, "Stabilizer states and clifford operations for systems of arbitrary dimensions and modular arithmetic," Phys. Rev. A **71**, 042315 (2005).

[77] Robert Koenig and John A Smolin, "How to efficiently select an arbitrary Clifford group element," J. Math. Phys. **55**, 122202 (2014).