Entropy density benchmarking of near-term quantum circuits

Marine Demarty [1,*] James Mills [1,2] Kenza Hammam [1] and Raúl García-Patrón [1,3]
[1]*School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom*
[2]*Quandela, 7 Rue Léonard de Vinci, 91300 Massy, France*
[3]*Phasecraft Ltd., 77-79 Charlotte Street, London W1T 4PW, United Kingdom*

(Received 15 January 2026; accepted 5 August 2026; published 20 August 2026)

Understanding the limitations imposed by noise on current and next-generation quantum devices is a crucial step toward demonstrating practical quantum advantage. In this work, we investigate the accumulation of entropy density as a benchmark to monitor the performance of quantum processing units (QPUs). We provide a proof-of-principle demonstration of our methodology which entails developing simple heuristic models of how entropy accumulates, testing them against real QPU data, and finally using these models to determine a circuit volume threshold above which quantum advantage is unattainable. Monitoring entropy density not only offers an alternative approach that complements existing circuit-level benchmarking techniques, but, more importantly, it bridges the gap between circuit-level and application-level benchmarking protocols. In particular, our heuristic model of entropy accumulation allows us to outperform existing techniques that bound the circuit-size threshold for quantum advantage.

DOI: 10.1103/swvt-3pkh

I. INTRODUCTION

One of the main limitations of current and near-term quantum devices remains the imperfect control of the qubits and the challenge of isolating the system from its environment, leading to noise levels beyond those required to reach fault-tolerant quantum computation [1–3]. This regime is known as noisy intermediate-scale quantum (NISQ) computation. Even if noise greatly limits the computational power of such computers, their practical applications are worth investigating as their size is already above the limit of 50 qubits—which is typically when error-free quantum devices become highly nontrivial to simulate by a classical digital computer [4,5].

To harness the full power of NISQ technology, however, one must look for tasks that are suitable for intermediate-scale devices with high error rates; this means, for instance, designing quantum algorithms that are resilient to noise, and finding tests of quantum advantage that do not require more than a few hundred qubits. Examples of such tasks include random circuit sampling [6] and Gaussian boson sampling [7,8], which have been the target of recent quantum supremacy claims. However, these sampling tasks have very limited useful application (see, e.g., [9–11]).

A popular way to design applications for NISQ devices is to reformulate a given computational task as an optimization problem which can be solved using so-called variational quantum algorithms (VQAs) [12]. These hybrid quantum-classical algorithms are believed to be good candidates to implement on NISQ devices affected by errors and low coherence times. Solving combinatorial optimization problems [13] or finding the ground-state energy of a Hamiltonian [14] are examples among an extensive literature on the topic [12]. Whether quantum advantage can be achieved in the near term on *practical* tasks remains a question of scientific debate, as seen after the recent simulation of the behavior of a magnetic material on a 127-qubit quantum processor [15] and its subsequent rebuttals (see, e.g., Refs. [16,17]).

To understand and evaluate the potential of noisy quantum devices to solve specific tasks, it is crucial to be able to assess the effect of errors on performance. With this in mind, a wide range of benchmarking techniques have been developed. Those fall into three main categories: gate-, circuit-, and application-level protocols. Gate-level benchmarking protocols aim at characterizing the performance of low-level components such as gates and sequences of gates (e.g., randomized benchmarking [18], gateset tomography [19]). Circuit-level protocols are concerned with assessing the performance of a quantum processing unit (QPU) when running specific classes of circuits (e.g., cross-entropy benchmarking [6], quantum volume [20]). Finally, application-level benchmarking techniques characterize the performance of a QPU when solving practical problems (e.g., Q-score for Max-Cut with quantum approximate optimization algorithm (QAOA) [21], fermionic depth for many-body problems [22]). However, to the best of our knowledge, those benchmarking categories are relatively isolated and there are no benchmarks that exploit benchmarking tools of one level to provide information at a higher level.

We believe entropy density benchmarking provides an excellent bridge between circuit-level and application-level

*Contact author: marine.demarty@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

benchmarking protocols, enabling us to estimate the circuit size above which quantum advantage is unattainable for solving a given problem of interest. A significant part of the errors affecting quantum circuits, such as stochastic Pauli channels, leads to an increase of entropy. Intuitively, the accumulation of entropy is bad for the performance of the system, as the entropy quantifies closeness to a maximally mixed quantum state. The link between entropy accumulation and the performance of variational quantum algorithms was recently formalized in Ref. [23], providing a rigorous formalism that bridges the gap between circuit- and application-level benchmarking. We envision the use of entropy density as a central tool of a benchmarking methodology that consists of three main steps: (i) combination of analytical and numerical methods to build simple yet effective heuristic models of entropy accumulation, (ii) validation and adjustment after comparison with real hardware entropy data, and (iii) bounding the circuit volume accessible before losing any potential quantum advantage.

In this work, we aim to provide a proof-of-principle demonstration of this benchmarking methodology allowing to bridge the gap between circuit-level and application-level benchmarking protocols, enabling us to derive more accurate thresholds than in Ref. [23] on the circuit size above which quantum advantage is unattainable.

The structure of this paper is as follows. In Sec. II, we analyze numerically the Renyi-2 entropy scaling of the output register of a hardware-efficient VQA circuit with the circuit size (width and depth) and noise parameters, motivating the use of an analytical model for entropy accumulation in such circuits based on global depolarizing noise. In Sec. III, we test the model against entropy accumulation data collected on a real superconducting device and discuss further refinements of the model to bring it closer to actual hardware observations. We then argue that global depolarizing, despite some limitations, still provides a simple heuristic approach to decide whether quantum advantage is attainable with a typical superconducting NISQ device for a specific optimization problem in Sec. IV. Finally, we summarize the main results of this paper and discuss open questions and possible future directions to this work in Sec. V.

II. CLASSICAL SIMULATION

We are interested in how the entropy of the quantum register evolves in a variational quantum circuit as layers of gates are applied to the input state, causing errors to accumulate, and in the dependency of the entropy on the circuit width (number of qubits). As a worked example, we consider a typical hardware-efficient VQA ansatz: each layer of this circuit consists of the concatenation of random single-qubit rotations about the X axis, $R_X(\theta) := \exp(-i\frac{\theta}{2}X)$, with random single-qubit rotations about the Y axis, $R_Y(\theta) := \exp(-i\frac{\theta}{2}Y)$, followed by a set of nearest-neighbor CZ gates distributed into two layers of nonintersecting gates, as shown in Fig. 1.

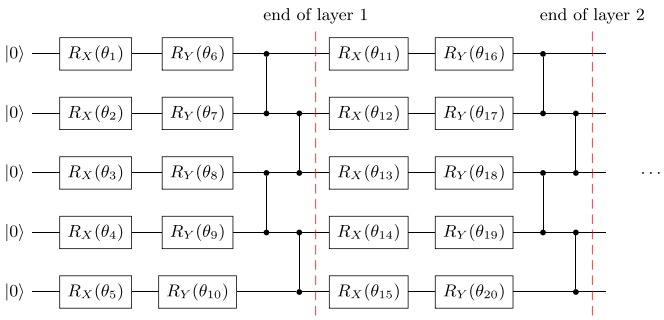


FIG. 1. First two circuit layers of a five-qubit hardware-efficient parametrized quantum circuit $U(\vec{\theta})$. Each circuit layer is composed of a single-qubit $R_X(\theta)$ rotation followed by an $R_Y(\theta)$ rotation and a final layer of CZ gates between nearest neighbors. The angles θ_i are drawn from the uniform distribution over $[0, 2\pi)$.

A. Entropy density accumulation under local depolarizing noise

We start by numerically investigating the entropy accumulation of our circuit of interest for different circuit widths and number of circuit layers (depth), under a commonly used local depolarizing noise model. Let p_1 (p_2) be the depolarizing probability for single-qubit (two-qubit) gates. The noise model we choose for the numerical simulation consists of applying a single-qubit (two-qubit) depolarizing channel with depolarizing probability p_1 (p_2) after each single-qubit (two-qubit) gate of the quantum circuit from Fig. 1. For a single-qubit state ρ, our single-qubit depolarizing channel is $\mathcal{N}_{p_1}(\rho) := (1-p_1)\rho + p_1 I/2$, while for a two-qubit state ρ, the two-qubit depolarizing channel considered is $\mathcal{N}_{p_2}(\rho) := (1-p_2)\rho + p_2 I/4$.

We use the software development kit qiskit [24] to calculate the entropy of the quantum register at depth D by evolving the density matrix of n qubits starting with the input state $\rho_0 := (|0\rangle\langle 0|)^{\otimes n}$. For each circuit layer, we compute the second-order Renyi entropy defined as

$$S^{(2)}(\rho_D) = -\log_2\left(\mathrm{Tr}\left[\rho_D^2\right]\right), \tag{1}$$

where ρ_D is the output density matrix (after applying D layers of gates). The obtained evolution of the Renyi entropy density $S^{(2)}(\rho_D)/n$ for different numbers of qubits, n (widths), and different depths can be found in Fig. 2, where for p_1 and p_2 we take deduced values from QPU calibration data (see Appendix D). We focus our interest on entropy density in order to allow a meaningful comparison between circuits of different width with the additional benefit of potentially inferring universal rules independently of the system size. Note that for every circuit size and depth, the dotted results correspond to the classical simulation using local depolarizing noise, while the solid lines correspond to an interpolation given by a global depolarizing model discussed in the section below.

As expected, local depolarizing noise causes the system to converge toward the maximally mixed state as the error occurrence increases, approaching a density of 1 at high depth. Interestingly, the larger the width n of the circuit, the faster the entropy density converges, where for large systems ($n \to \infty$) the speed of convergence of the Renyi entropy density seems to converge to an asymptotic bound, as one can clearly see from Fig. 2.

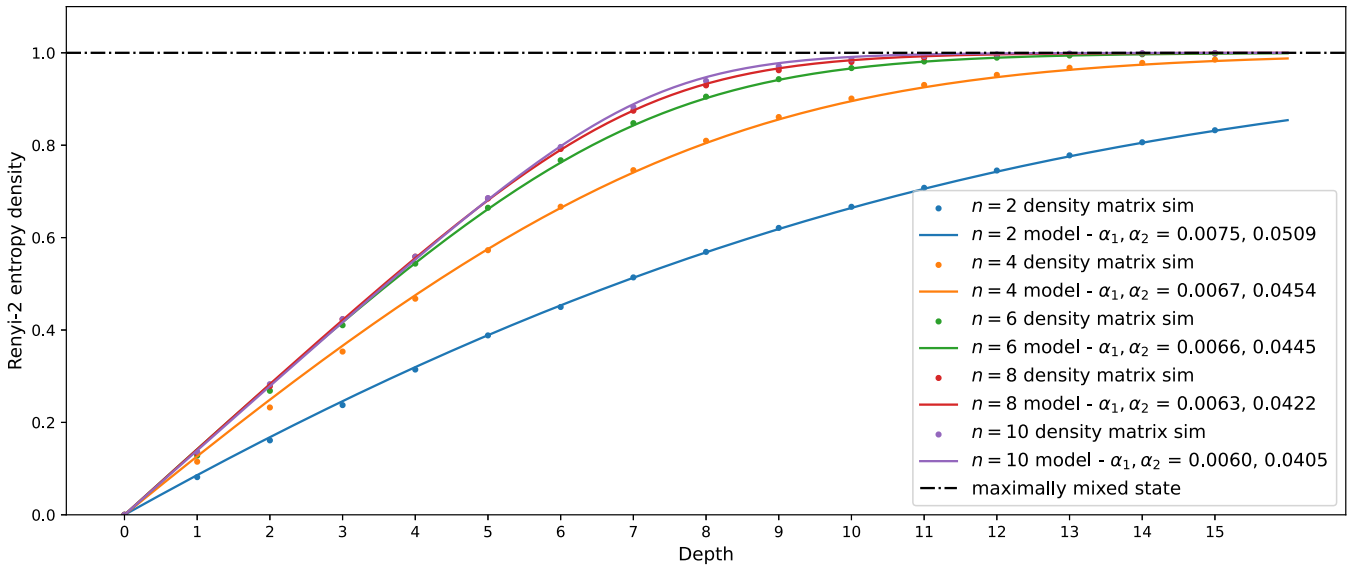


FIG. 2. Numerical simulation and heuristic model. We simulate a single VQA circuit, represented in Fig. 1, with randomly selected parameters [random seed `np.random.seed(837)`] affected by local depolarizing noise ($p_1 = 0.008$ and $p_2 = 0.054$ from calibration data of Rigetti's QPU, Aspen-M-3). The dots characterize the Renyi-2 entropy density evolution of the quantum register after each circuit layer for different number of qubits, n (see color label). The black horizontal dash-dotted line corresponds to the entropy density of the maximally mixed state $\sigma_0 := I/2^n$. The solid lines correspond to interpolations using a global depolarizing heuristic model from Eq. (6) with fitting parameters α_1 and α_2 ($\alpha_i \approx p_i$), as detailed in Sec. II B 1.

B. Global depolarizing heuristic model

1. Analytical model

We wish to build a heuristic model for the evolution of the second-order Renyi entropy as a function of the quantum circuit size, i.e., depth and width. Recent random circuit sampling quantum supremacy demonstrations [25], as well as some error mitigation techniques run on superconducting hardware [26], have relied on a seemingly successful global depolarizing noise (or white noise) assumption. It is therefore natural to explore the approximation of the chosen local depolarizing noise model described previously by a global depolarizing noise channel applied after the last layer of the ideal quantum circuit $U_D(\vec{\theta})$; i.e., the output of our noisy VQA circuit can be written as

$$\rho_D = (1-P)U_D(\vec{\theta})\rho_0 U_D(\vec{\theta})^\dagger + P\frac{I}{2^n}, \tag{2}$$

where P is the global depolarizing probability, and $\rho_0 = (|0\rangle\langle 0|)^{\otimes n}$ is the input state. The analytical model for the purity then reads

$$\mathrm{Tr}\left[\rho_D^2\right] = (1-2^{-n})((1-P)^2 - 1) + 1. \tag{3}$$

To link the global depolarizing probability P of the model with the circuit size (n, D) and local noise parameters (p_1, p_2), we exploit the standard assumption that the probability of no error occurring in the circuit under local depolarizing noise should match the probability of no error occurring in the model with global depolarizing noise. This yields

$$1 - P = (1-p_1)^{|1\text{Q gates}|}(1-p_2)^{|2\text{Q gates}|}, \tag{4}$$

$$= e^{-\alpha_1 2nD} e^{-\alpha_2 (n-1)D}, \tag{5}$$

where $|1\text{Q gates}|$ ($|2\text{Q gates}|$) is the number of one-qubit (two-qubit) gates of the circuit, corresponding to $2n$ and $n-1$, respectively, for the circuit of Fig. 9 (see Appendix D).

We have introduced the parameters $\alpha_1 := \ln((1-p_1)^{-1})$ and $\alpha_2 := \ln((1-p_2)^{-1})$, which are expected to model the depolarizing probabilities (p_1, p_2) in the low-noise limit: $\alpha_1 \approx p_1$ and $\alpha_2 \approx p_2$. Substituting Eq. (3) with Eq. (5), we get the following model for the purity:

$$\mathrm{Tr}\left[\rho_D^2\right] = (1-2^{-n})(e^{-2(2\alpha_1 n + \alpha_2(n-1))D} - 1) + 1, \tag{6}$$

with fitting parameters α_1 and α_2.

We subsequently performed a nonlinear least-squares fitting of the purity data obtained from numerical simulation; the corresponding Renyi-2 entropy density data (dots) with the Renyi-2 entropy density of the fitted purity model are shown in Fig. 2. We explored three different strategies for the use of the free parameters α_i: (i) leaving both parameters free (α_i), (ii) fixing $\alpha_1 = 0$, and (iii) assuming $\frac{\alpha_1}{\alpha_2} = \frac{p_1}{p_2}$, where the last two assumptions are different approaches to model the gap between single-qubit and two-qubit gate errors. If all three assumptions lead to very similar qualities of interpolation, the first leads to parameters α_1 and α_2 that have little connection with the physical values p_1 and p_2. Therefore, we use the constraint $\frac{\alpha_1}{\alpha_2} = \frac{p_1}{p_2}$ in all the fittings presented in this paper, to guarantee the same order of magnitude between all pairs (α_i, p_i).

Interestingly, a global depolarizing model provides a good heuristic description of the entropy accumulation behavior at considered noise levels and for small systems without having to average over the whole family of circuits (variational gate parameters); it achieves a good fit already for a given specific circuit.

2. Asymptotic behaviors

It is easy to see that for the limit $D \to \infty$, the global depolarizing model leads to $\mathrm{Tr}[\rho_D^2] \sim 1/2^n$, which results in $S^{(2)}(\rho_D)/n \sim 1$, reaching the entropy density of the maximally mixed state as observed in Fig. 2.

For all system sizes, at small enough circuit depth, one can observe in Fig. 2, for local depolarizing noise, that the Renyi entropy density growth is well approximated by a linear function with the circuit depth. Moreover, in the limit of large system sizes, $n \to \infty$, the Renyi entropy density growth seems to converge to an upper bound.

A similar feature can be derived for the global depolarizing model as the Renyi-2 entropy density bound

$$\lim_{n\to+\infty} \frac{S^{(2)}(\rho_D)}{n} = \begin{cases} \frac{2(2\alpha_1+\alpha_2)}{\ln(2)} D & \text{if } D \leqslant \frac{\ln(2)}{2(2\alpha_1+\alpha_2)} \\ 1 & \text{if } D > \frac{\ln(2)}{2(2\alpha_1+\alpha_2)}, \end{cases} \tag{7}$$

which defines the depth threshold

$$D^* = \frac{\ln(2)}{2(2\alpha_1+\alpha_2)}, \tag{8}$$

above which the output is indistinguishable from the maximally mixed state. For instance, taking typical QPU values for noise parameters $\alpha_1 = p_1 = 10^{-4}$ and $\alpha_2 = p_2 = 10^{-3}$ [27], we get a depth threshold $D^* = 288$, beyond which the quantum device's output is no better than a random guess.

It is important to note that the obtained bound to the Renyi-2 entropy density for the global depolarizing noise model in the limit of large system sizes assumes fixed α_1 and α_2 values. When the heuristic model is fitted on the local depolarizing simulations, we can see from Fig. 2 that the fitting parameters α_1 and α_2 vary with the system size n. Nonetheless, assuming the fitting parameters saturate to constant values as $n \to \infty$ as Fig. 2 suggests, this asymptotic result for global depolarizing noise captures the local depolarizing behavior.

III. ANALYSIS ON A REAL QPU

We are now ready to study entropy accumulation as a function of circuit size on an actual quantum processing unit, and explore whether our model for the Renyi-2 entropy from Eq. (6) matches QPU results. In Sec. III B, we provide detailed analytical, numerical, and QPU data over a superconducting platform using classical shadows, a technique that has proven extremely useful for the estimation of expectation values on NISQ devices. Despite not being scalable, the classical shadows protocol provides a good compromise for small circuit sizes like those explored in this work and more generally for systems made of a handful of qubits. We observe that the QPU data diverge from a simple global depolarizing model and are able to explain most of this gap by incorporating a model of T_1 relaxation at the gate and measurement level in Sec. III B 3.

A. Classical shadows

1. Background

In their 2020 paper, Huang *et al.* introduce a new method for efficient estimation of many properties of a quantum state ρ known as the classical shadows protocol [28]. Instead of performing full tomography of the quantum state, which requires a number of measurements that scale exponentially with the system size n, they proposed to obtain several classical snapshots of the quantum state by measuring the state in M randomly chosen bases K times. Classical postprocessing of the measurement outcomes then allows to deduce the expectation value of N_{obs} observables of interest $\mathrm{Tr}[O_i\rho]$, in a sample efficient manner for specific families of observables.

Huang *et al.* propose two versions of this protocol: (i) random single-qubit Pauli measurements and (ii) random n-qubit Clifford circuit measurements. Pauli measurements are particularly efficient for predicting a large collection of local observables, and only require short-depth measurement circuits, which are very suitable for NISQ devices. On the other hand, n-qubit Clifford measurements should be preferred for efficiently predicting many global observables with a bounded Hilbert-Schmidt norm; the main drawback is that gates of order $n^2/\log(n)$ are needed for their implementation, resulting in circuits that are less practical for current quantum devices.

More precisely, to estimate N_{obs} observables O_i with additive error ϵ, the number M of randomly chosen bases, measured K times, needs to scale as

$$O\big(\log(N_{\mathrm{obs}}) \max_i ||O_i||^2_{\mathrm{shadow}}/\epsilon^2\big), \tag{9}$$

where $||O_i||_{\mathrm{shadow}}$ is the so-called shadow norm, which is upper bounded by $\max_i 4^{k_i}||O_i||^2_\infty$ in scenario (i) of single-qubit Clifford measurements, and by $\max_i \mathrm{Tr}[O_i^2]$ in scenario (ii) of global Clifford measurements (see Ref. [28]), where k_i is the locality of observable O_i and $||\cdot||_\infty$ is the operator norm.

2. Renyi-2 entropy estimation

The above framework also extends to nonlinear functions of ρ, such as quadratic terms $\mathrm{Tr}[O_i\rho\otimes\rho]$. By noting that the second-order Renyi entropy of a quantum state, $S^{(2)}(\rho)$, is a function of the purity $\mathrm{Tr}[\rho^2]$, which can be rewritten as

$$\mathrm{Tr}[\rho^2] = \mathrm{Tr}[S\rho\otimes\rho], \tag{10}$$

where S is the SWAP matrix, one can use the classical shadows protocol to build an estimate for the second-order Renyi entropy where the corresponding purity is estimated as the median among N_g measurement groups each containing M_g Pauli measurements, i.e., $\hat{P}_{\mathrm{shadows}} = \mathrm{median}\{\hat{P}^{(M_g,K)}_{\mathrm{shadows},1}, \ldots, \hat{P}^{(M_g,K)}_{\mathrm{shadows},N_g}\}$ with

$$\hat{P}^{(M_g,K)}_{\mathrm{shadows}} = \frac{2}{M_g(M_g-1)} \sum_m \sum_{m'|m'<m} \frac{1}{K^2} \times \sum_{k,k'=1}^{K} \prod_{n=1}^{N} (9\gamma_{n,m,m',k,k'} - 4), \tag{11}$$

where n labels the qubits, k, k' is the snapshot of a given measurement m, and $\gamma_{n,m,m',k,k'} \in \{0, 1/2, 1\}$ depends on the measurement and its outcome as given in Appendix B in the case of Pauli measurements, and the total number of shadows is $M = N_g M_g$. Because the operator S is not local, the use of classical shadows to estimate purities requires an exponential scaling of the number of samples with system size.

Although the Renyi-2 entropy or equivalently purity is a global property, here we choose to consider Pauli basis

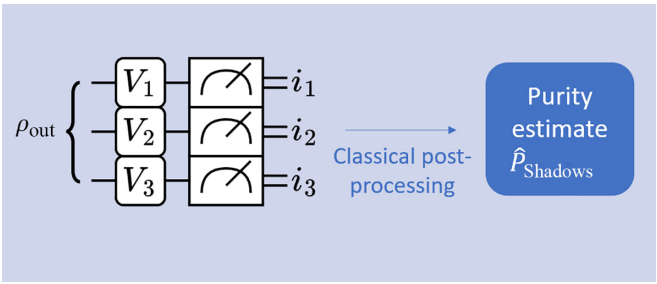


FIG. 3. Classical shadows protocol. Each V_i is a single-qubit Clifford gate such that its combination with a computational basis measurement corresponds to a random measurement in the X, Y, or Z basis.

measurements for entropy estimation on *near-term* devices. This is because a single layer of single-qubit gates is easy to implement and introduces fewer errors compared to an n-qubit Clifford circuit. Figure 3 shows the corresponding measurement circuit. For large quantum systems, the Pauli version of the method is very inefficient in terms of sampling cost, as the purity is a global property thus implying an $O(4^n)$ scaling with the system size n: fixing $\epsilon, \delta > 0$, obtaining an ϵ-accurate purity estimate with success probability at least $1-\delta$ requires $O(\ln(\frac{1}{\delta}) \times \frac{4^n}{\epsilon^2})$ (Pauli) measurements (see Appendix B). Note that the n-qubit Clifford version of the protocol for purity estimation would only provide a quadratic improvement of the scaling, as the number of measurements scales as $\sqrt{\text{Tr}[S^2]} = 2^n$. We expect that the added complexity of the circuit will make this approach unpractical for sizes beyond tens of qubits.

B. Study of entropy accumulation on a QPU

Using the Pauli version of the classical shadows protocol whose estimate is given in Eq. (11), we estimate the purity and Renyi-2 entropy density of the quantum state along the chosen VQA circuit, i.e., as functions of the circuit depth, on Rigetti's Aspen-M-3 QPU (see Appendix D for the device's description). QPU results for a circuit width $n = 3$ can be found in Fig. 4(a), using two different estimation strategies (see discussion in Sec. III B 1 for details) corresponding to light and dark gold error bars, where, for each instance, a purity estimate was calculated by taking $K = 1000$ classical snapshots of the output state for each of the $M = 5\times4^3 = 320$ measurement settings considered, then taking the median of means over five groups. Before we discuss a comparison of the QPU data to a global depolarizing model and its later refinement with additional T_1 relaxation, we explain the reasoning behind having two sets of QPU data in Figs. 4(a) and 4(b).

1. Randomization of measurements

A standard classical shadows protocol should ideally randomly select a new measurement for every measurement sample. Changing circuits being relatively costly in execution time for some platforms like superconducting qubits, we often implement a variant where M measurement settings are randomly selected and each one is executed K times its number of occurrence. Because our QPU data were limited in number of qubits ($n = 3$) and therefore we were expecting to see the occurrence of most measurements and in similar number, we initially opted to derandomize the measurement choice and scan the $3^3 = 27$ possible measurement settings with $K = 1000$ classical snapshots of the output state for each of them [29]. We were surprised to observe that this strategy generated a systematic error in the QPU entropy estimation that was later corroborated by numerical simulations (see Appendix B 4).

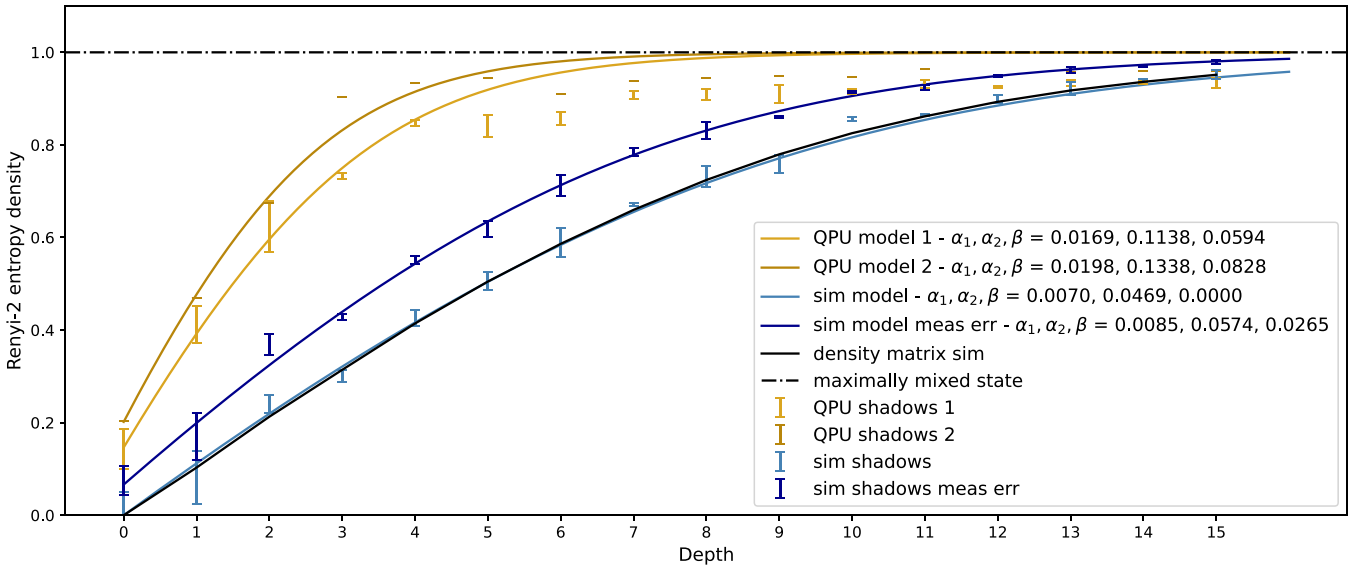
To mitigate this bias, we reintroduced a random choice of measurements in our QPU results (light and dark gold error bars in Fig. 4) through two different methods. We first select $M = 320$ random measurements and count how many times each measurement is chosen. In method 1 (light gold error bars), for a fixed run of the protocol and a given choice of measurement, we build the list of shadows by replicating the measurement outcomes associated with that measurement from the corresponding derandomized run the given number of times the measurement setting was obtained. Since we still have three sets of data, we can obtain errors bars, exactly as done before. In method 2 (dark gold crosses), we sacrifice the information on errors bars in order to cluster the data of the previous three runs into a single one that contains more realistic randomized QPU data; i.e., we build the purity estimate for a single run of the protocol by exploiting the three available sets of snapshots for each measurement setting. More precisely, for a given setting, the first three times a measurement occurs, we use one of the three sets of data, and if a measurement occurs more times, we cycle over them again.

2. Gap with respect to the global depolarizing model

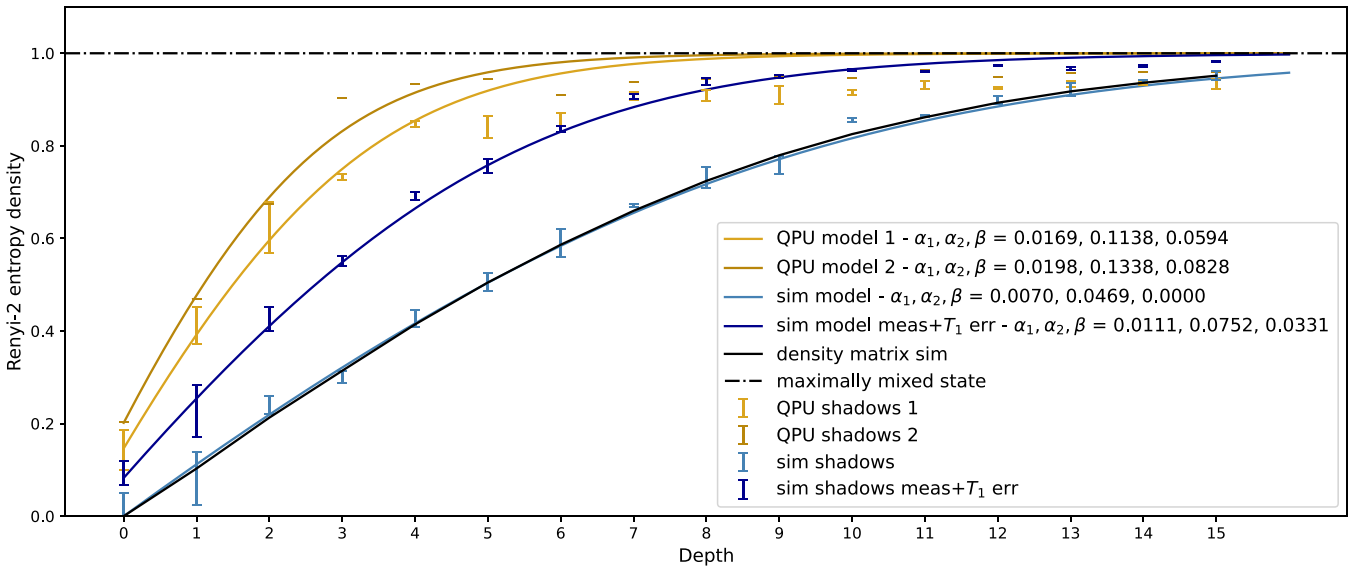
Figure 4(a) includes a comparison of the hardware entropy results with a density matrix simulation of the same VQA circuit under a local depolarizing noise model (see solid black line) with noise parameters $p_1 = 0.008$ and $p_2 = 0.054$ deduced from calibration data of Aspen-M-3 (see Appendix D). The QPU data show a faster-than-expected increase of the entropy density after only a few layers of gates when compared to a global depolarizing model with parameters resulting from calibration data. We also observe convergence to a value strictly below the threshold of 1 corresponding to the maximally mixed state (dash-dotted line) with some minor oscillations for intermediate time. The nonzero entropy density at depth zero most probably indicates the presence of errors in the entropy estimation protocol.

One observes a clear gap between the QPU data and the density matrix simulation under local depolarizing noise, which can be easily interpolated with a global depolarizing model (light blue solid line), as done in Sec. II B 1.

To discard the possibility of the observed gap being solely a systematic error resulting from the classical shadows protocol, we perform a simulation of the target circuit under local depolarizing noise from calibration data followed by its noiseless classical shadows postprocessing under the same sample constraints as for the QPU data [see light blue error bars in Fig. 4(a)]. We confirm that classical shadows do not seem to produce any significant systematic shift. Even if a very minor systematic error could be observed, it can be removed by further increasing the number of samples, as discussed in Appendix B 5.



(a) Without T_1 errors



(b) With T_1 errors

FIG. 4. QPU results and validation of our heuristic model. We consider a VQA circuit as in Fig. 1 applied to $n = 3$ qubits, with fixed circuit parameters [fixed random seed `np.random.seed(837)`]. Gold error bars give the purity and Renyi-2 entropy evolution as functions of circuit depth on Rigetti's QPU Aspen-M-3 using two different methods. Each error bar was obtained by running the classical shadows protocol three times, and computing the average estimate and the standard deviation over those three samples. Results obtained in the same setting but using a local depolarizing noise model ($p_1 = 0.008$ and $p_2 = 0.054$ from calibration data of Aspen-M-3) and without measurement errors correspond to the light blue error bars. The dark blue error bars in (a) correspond to the same simulation with added measurement errors from calibration data (measurement circuit gate errors $p_1 = 0.008$ and errors in the detectors $P(0|1) = 0.03$ and $P(1|0) = 0.02$). The dark blue error bars in (b) also include T_1 relaxation errors, both at the gate level ($\gamma_1, \gamma_2 = 0.003, 0.015$) right after depolarization ($p_1, p_2 = 0.006, 0.038$) and in the measurement stage ($\gamma_{\text{meas}} = 0.05$) just before readout errors ($P(0|1) = P(1|0) = 0.03$) based on calibration data. The black solid lines show the purity and Renyi-2 entropy density evolution obtained via density matrix simulation under our local depolarizing noise model without measurement errors (similar to the light blue error bars). The black horizontal dash-dotted lines correspond to the purity or entropy density of the maximally mixed state $\sigma_0 := I/2^n$. Fit of our global depolarizing heuristic model, where we have imposed that $\alpha_2/\alpha_1 = p_2/p_1$, and where β models readout errors and noise in the measurement circuit, corresponds to the solid gold and blue lines.

To explain part of the observed gap between the real QPU data and the model, we incorporate errors in the gates of the measurement circuit (cf. Fig. 3) and readout errors into our numerical simulations, both deduced from calibration data (see Appendix D). The numerical results correspond to the dark blue error bars in Fig. 4(a). To account for the additional measurement circuit gate error and additional readout errors, which may artificially increase the entropy estimation, we introduce an additional fitting parameter β to our global depolarizing model from Eq. (6) (see Appendix A). The resulting fitted model corresponds to the dark blue solid line in Fig. 4(a). We observe that its fitted parameters α_1, α_2, and β closely match the calibration data $p_1 = 0.008$, $p_2 = 0.054$, and $p_m \approx P(0|1) = 0.03$ used in the local depolarizing noise simulations with measurement circuit and readout errors (dark blue error bars). Although the added measurement circuit and readout noise account for part of the gap between the real hardware data and the depolarizing noise model, a significant part remains unexplained.

3. Refinement with T_1 relaxation

One can notice in Fig. 4(a) that the heuristic global depolarizing noise model (dark blue solid line), despite capturing part of the QPU entropy accumulation behavior, also significantly differs. We observe three different features in the hardware results (gold error bars): (1) faster initial entropy growth, (2) potential saturation below the maximum entropy density (value of 1) in the large-depth limit (confirmed by additional QPU data for circuits with up to depth 30), and (3) potential oscillations in the intermediate regime. As discussed previously, neither measurement errors nor a deviation of the classical shadows estimate is the reason for the observed differences.

Noticing the lengthy measurement time of the device and in an attempt to account for the above three features, Fig. 4(b) shows an additional numerical simulation (dark blue error bars) that includes the effect of T_1 relaxation, modeled as amplitude damping channels, both in the circuit implementation and in the measurement stage (from calibration data of the device; see Appendix D). We emphasize that the local depolarizing probabilities used in our simulations are fixed from the single- and two-qubit gate fidelities reported in the device calibration data from which we removed a T_1 contribution. These are thus effective error rates that group various error mechanisms degrading the gates, including the often significant dephasing (T_2) contribution. We single out only T_1 relaxation for separate treatment, modeling it with an amplitude damping channel, as a purely depolarizing description cannot reproduce two qualitative features of the measured entropy accumulation: its saturation below the maximally mixed value and the oscillations seen at large depth. We do not introduce a separate dephasing channel, for two reasons. First, we aim to keep the effective noise model as simple as possible. Second, the random single-qubit rotations of the hardware-efficient ansatz are expected to twirl dephasing errors toward an isotropic, depolarizinglike channel [30], so that their effect on entropy is well captured by the effective local depolarizing channel. This also supports the applicability of the global depolarizing heuristic underlying our analysis.

The addition of T_1 relaxation into the simulation explains a significant part of the fast entropy growth. Moreover, the saturation of the entropy density below the maximally mixed state threshold of 1, along with the observed oscillatory behavior, is partially accounted for by the added relaxation noise. However, a gap between the global depolarizing prediction (dark blue solid line) and the QPU data (gold error bars) remains. We believe that this could result from a more complex combination of errors present on the device, including coherent errors, crosstalk, or some additional damping not captured in our model, simulations, or in the original calibration data.

A natural direction for future work could be to incorporate a small number of additional effective noise parameters to model these other sources of error. To preserve the simplicity of the heuristic model, this could be done in a coarse-grained way, for example, by including nearest-neighbor correlated Pauli noise to capture crosstalk, or local coherent noise to model systematic overrotations. We note that while coherent unitary errors do not directly change the entropy they may interact with nonunital noise, such as amplitude damping, and thereby affect the entropy dynamics. Including such additional noise features would retain the analytical simplicity needed for circuit-volume estimates while more accurately capturing entropy saturation and oscillatory features in the QPU data.

4. Discussion on the update of the heuristic model

Our initial goal was to build simple heuristic models and prove their validity so they could later be exploited to make predictions at the application level. Therefore, a heuristic model that better captures the faster entropy increase would provide more stringent and accurate predictions on the performance of QPU hardware. Unfortunately, to the best of our knowledge, building a simple analytical model that captures this faster entropy increase resulting from the interplay of damping and depolarization is a nontrivial theoretical result.

While we wait for better models, we believe that a global depolarizing model built from calibration data provides a convenient compromise for the quantum device considered. Indeed, it is important to point out that such a model remains a lower bound to the QPU data for the most relevant part of the curve. This means that any prediction of effective available circuit volume done in the last step of the benchmarking methodology (presented in the next section) remains valid; it just becomes conservative and generous to the QPU performance.

IV. QUANTUM ADVANTAGE BENCHMARKING

After proposing a heuristic model, later tested and refined with numerical and hardware entropy accumulation analyses, we now present the last step of our benchmarking methodology. We will show how to combine the design of heuristic models of entropy accumulation in VQA circuits, such as the global depolarizing noise model of Eq. (6), with the application-level quantum advantage benchmarking technique introduced in Ref. [23]. This allows to make quick reliable predictions on whether the accumulation of errors precludes from reaching quantum advantage or the latter remains

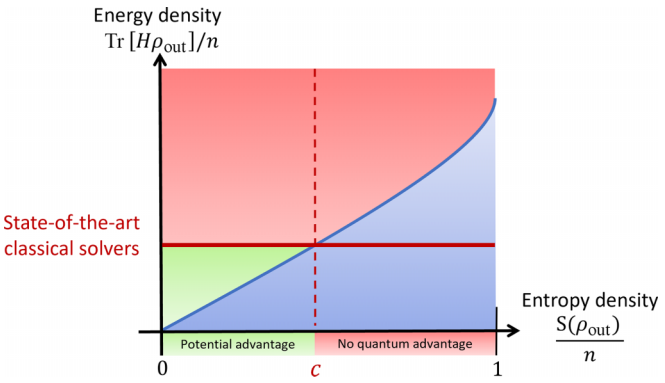


FIG. 5. Quantum advantage benchmarking framework from Ref. [23]. For an optimization task of the form $\min_\rho \text{Tr}[H\rho]$, the authors derive a lower bound (blue solid line) on the quantum device's solution to the energy density ($\text{Tr}[H\rho_{\text{out}}]/n$) as a function of the accumulated entropy density ($S(\rho_{\text{out}})/n$). The state-of-the-art classical solvers' solution to this optimization problem (red horizontal line) defines an entropy density threshold c such that if the quantum device's candidate solution ρ_{out} satisfies $\frac{S(\rho_{\text{out}})}{n} \geqslant c$ then quantum advantage is out of reach for this device and for this optimization problem.

possible. Most importantly, one does not need to run the quantum processing unit to solve the targeted application; only a model of how entropy accumulates on the problem-specific circuit ansatz is required. In what follows, we revisit the analysis of Ref. [23] for Max-Cut, but the technique is widely translatable to other optimization problems, whether defined by classical (e.g., combinatorial optimization problems [31]) or quantum Hamiltonians (e.g., quantum chemistry problems [32,33]).

A. Background

As detailed in Ref. [23], given a quantum device and known state-of-the-art classical solver solution to Max-Cut, we can certify whether there is some potential for a QPU to outperform the best classical solvers or conversely that quantum advantage is out of reach.

This is ascertained by an approach summarized in Fig. 5. Combining a lower bound on the quantum device energy density solution $\text{Tr}[H\rho_{\text{out}}]/n$ as a function of the von Neumann entropy $S(\rho_{\text{out}})/n$ (blue solid line) with the classical solver energy density prediction (red horizontal line) defines an entropy density threshold $S(\rho_{\text{out}})/n \geqslant c$ (red dotted vertical line) beyond which the QPU cannot provide a better solution than the classical solver.

Despite the benchmarking approach of Ref. [23] being constructed for von Neumann entropies and our gathered QPU data being for the Renyi-2 entropy, the well-known bound between the von Neumann and the Renyi entropies, $S(\rho) \geqslant S^{(2)}(\rho)$, $\forall\rho$, implies that any conclusion obtained by replacing $S(\rho)$ with $S^{(2)}(\rho)$ in the analysis summarized in Fig. 5 is a conservative and valid prediction.

B. Bounds on quantum advantage via circuit-size bounds

The next step of the approach is to combine the previous entropic threshold with the selected heuristic model of entropy accumulation in order to find a bound on the number of gates or maximum circuit depth before our QPU of interest can be certified to not provide better solutions than a standard classical solver. In what follows we revisit the analysis for MAX-CUT in Ref. [23] but now equipped with heuristic models of entropy accumulation.

1. Circuit bounds

Now using a global depolarizing heuristic model for the second-order Renyi entropy of the output of the VQA ansatz described in Eq. (6), we obtain the following bound on the circuit size (with width n and depth D):

$$2(2\alpha_1 n + \alpha_2(n-1))D \geqslant \ln\left(\frac{2^n-1}{2^{n(1-c)}-1}\right). \tag{12}$$

Assuming that two-qubit gate depolarizing noise is the dominant source of noise on the device ($\alpha_2 = p_2 \neq 0$), which is generally the case in practice, and thus neglecting single-qubit gate depolarizing noise ($\alpha_1 = 0$), this yields

$$(n-1)D \geqslant \frac{1}{2p_2}\ln\left(\frac{2^n-1}{2^{n(1-c)}-1}\right), \tag{13}$$

where we assume that p_2 is below 1%, a reasonable assumption for current QPUs. For large problem size $n \to \infty$, this leads to the bound on the circuit depth (number of circuit layers):

$$D \geqslant \frac{1}{2p_2} c \ln(2). \tag{14}$$

Note the similarity with the depth threshold obtained in Eq. (8) for large systems $n \to \infty$. In the special case where two-qubit gate depolarizing noise dominates ($\alpha_1 = 0$, $\alpha_2 = p_2$), both depth thresholds differ by the scalar factor c resulting from the entropy density threshold analysis above.

2. Comparison with previous work

Gset is a set of instances of Max-Cut popular for performance benchmarking for which classical solvers' solutions and times to solution are known. Following Ref. [23], we use the G12 instance of Gset consisting of 800 nodes and 1600 edges. Using Eq. (13) and the plot of the bound for this Max-Cut instance from Ref. [23] where the entropy density threshold against a heuristic classical solver reads $c = 0.3$, we conclude that, for $D \geqslant 0.104/p_2$, quantum advantage is out of reach for a quantum device with two-qubit gate depolarizing probability p_2; as we expect, this is an even smaller depth threshold than the $D \geqslant 0.18/p_2$ predicted in Ref. [23]. For a current state-of-the-art two-qubit gate error of $p_2 = 10^{-3}$ [27], our approach predicts that circuits with depth $D \geqslant 110$ are already in the regime of certified classical superiority.

In a more general setting, our approach also allows to predict an effective circuit volume, i.e., the total number of gates obtained from Eq. (13), beyond which quantum advantage for the specific problem under study is no longer possible. The effective circuit volume is visualized in Fig. 6 (dark blue line) for the G12 Max-Cut instance above for a QPU with two-qubit gate error given by $p_2 = 10^{-3}$.

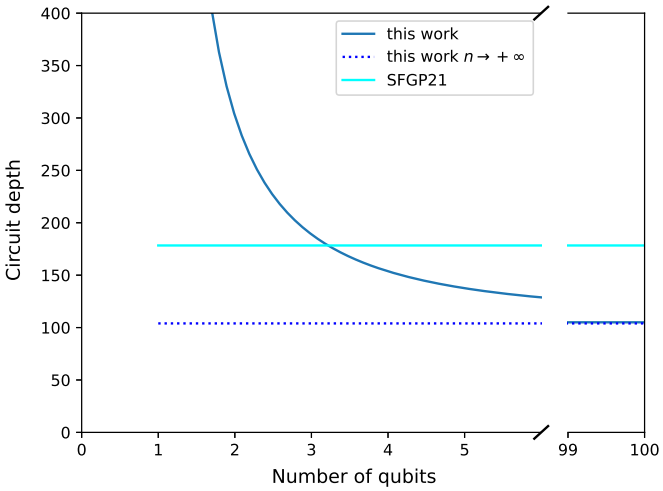


FIG. 6. Circuit size threshold above which quantum advantage is out of reach (for arbitrary c and p_2 values; here $c = 0.3$ and $p_2 = 10^{-3}$). "This work" corresponds to the newly derived circuit-size bound obtained from Eq. (13) (dark blue solid line) while SFGP21 (light blue solid line) is the depth condition previously derived in Ref. [23]. We cut the x axis to show the threshold behavior in the $n \to \infty$ limit.

For large system sizes, we see that our depth threshold saturates at a constant value $\frac{1}{2p_2} c \ln(2)$ for fixed c (dotted blue line), which is below the depth threshold derived in Ref. [23] based on relative entropy contraction results (cyan line). In Fig. 6, keeping $c = 0.3$ fixed as n increases is a simplifying assumption used to illustrate the large system behavior of our circuit-size bound and should not be interpreted as implying that c is generally independent of the problem size. This shows that quantum advantage is lost at a shorter depth than originally predicted for large problem sizes. In fact, it can be proven analytically that our depth threshold is tighter than that of Ref. [23] for any c and p_2 values (see Appendix C). One can also prove that in the event of the threshold depth still keeping some dependency on the number of qubits, n, our condition in Eq. (14) would still be tighter than the depth threshold from Ref. [23] for large systems (see Appendix C).

V. DISCUSSION

Inspired by the theoretical framework of Ref. [23], we have presented a QPU benchmarking methodology using the accumulation of entropy along the computation as a performance metric, which allows to bridge between circuit-level and application-level benchmarking techniques. The method works by first proposing simple but sufficiently explanatory models of entropy accumulation, and then by validating and refining those models via comparison to QPU data. These models are then used to estimate the available volume of computation before any potential quantum advantage is lost for our optimization problem of interest. We have carried out a proof-of-principle demonstration of the methodology centered on the global depolarizing model and, after comparison with QPU data, we have proposed some refinements accounting for the effect of T_1 relaxation time. Using this simple model for entropy accumulation, we have shown how to impose more stringent bounds than previously shown in Ref. [23] on the circuit size above which quantum advantage is lost for hardware-efficient VQA circuits run on a typical superconducting NISQ device.

When estimating circuit purities and their corresponding entropies, any imperfection in the measurement process—whether it is a noisy gate in the measurement protocol, T_1 relaxation, or bit flips occuring during the measurement—can cause the final entropy estimate to deviate from the target true circuit output entropy. Therefore, further work is needed to develop error mitigation techniques specific for purity and entropy estimation, for example, by adapting existing SPAM error mitigation techniques to the peculiarities of each estimation protocol. Another interesting direction would be to develop theoretical techniques that would provide guarantees for the outcome of a given, potentially error-mitigated, entropy estimation protocol.

For medium and large systems, we expect classical shadows to suffer from larger error bars and potentially stronger systematic errors, making it unpractical, leaving the swap test as the only known viable option. If sample-efficient single-copy techniques should not exist, following complexity arguments [34,35], an interesting open question is whether some relaxation of the problem, such as restricting to upper and/or lower bounds, leads to sample-efficient protocols.

Although depolarizing noise seems to be a decent first approximation model to lower-bound entropy accumulation in superconducting circuits run on a typical NISQ device (here a two-qubit gate fidelity of the order of 10^{-1}), our results hint that a more complete model for entropy accumulation is required to make tighter predictions for such quantum devices, for example, incorporating amplitude damping, coherent errors, and crosstalk. A challenge here is to have simple theoretical models, such as the global depolarization, that capture most of these additional features. From our experience, the interplay of amplitude damping and depolarizing seems to lead to nontrivial dynamics and regimes that require a more detailed investigation to build more powerful models.

As superconducting quantum computers scale up and noise levels are reduced, we expect, inspired by Ref. [25], that the global depolarizing heuristic model should provide a better approximation to the entropy accumulation behavior under local depolarizing noise, which in turn would become a better assumption to the real QPU entropy signature as the effect on entropy of other sources of errors such as amplitude damping become negligible. We leave the confirmation of this trend as a very relevant open question.

We expect our methodology to carry over to platforms beyond superconducting systems, and to topologies other than nearest-neighbor planar superconducting circuit architectures studied in this work, with the possible exception of photonic platforms in the pre-fault-tolerant regime. This is because photon loss, a dominant source of imperfection in photonic platforms, has the vacuum state as its fixed point, which has zero entropy. Therefore, our entropy metric will need to be complemented by further performance metrics for pre-fault-tolerant photonic devices. It is important to point out that our methodology can also be applied to quantum error-corrected and fault-tolerant architectures by focusing on the entropy of the logical qubits. In a fault-tolerant regime, we expect the

physical platform and architecture to become less relevant. Interestingly, the validity of our methodology could be restored for photonic platforms in the fault-tolerant regime, as such architectures are now meant to encode logical qubits being affected by standard qubit noise, which we expect to mostly increase the entropy. There are two open questions related to our methodology and the nature of the platform that are worth further investigation. The first is whether the accumulation of entropy density as the circuit depth increases has any dependency on the type of platform and its topology. On the one hand, our simulations seem to indicate that a global depolarizing model becomes more accurate as the gate quality improves (at least for the one-dimensional architecture studied here); on the other hand, an all-to-all connectivity architecture allows entanglement to spread faster and therefore a potentially stronger sensitivity to local imperfections, leading to a potentially faster increase in entropy. Settling this question is an interesting direction for future research. The second is whether the existence of simple models of entropy density accumulation (here, a global depolarizing model was a first approximation) depends on the choice of platform and its architecture topology in the pre-fault-tolerant era. While this is certainly true for photonic platforms, comparing ionsand neutral atoms to superconducting circuits is an interesting question for further investigation.

ACKNOWLEDGMENTS

The authors thank Pauline Besserve for useful feedback on this paper, and Marco Paini, Sean Thrasher, Chirag Wadhwa, and Qisheng Wang for valuable discussions. Results on the QPU were obtained using Rigetti's Aspen-M-3 system as part of the Innovate UK Project No. 44167: Quantum Computing Platform for NISQ Era Commercial Applications. R.G.-P. and K.H. were supported by the EPSRC-funded project Benchmarking Quantum Advantage, Reference No. EP/Y004418/1.

DATA AVAILABILITY

The data that support the findings of this article are openly available [36].

APPENDIX A: ANALYTICAL MODEL FOR THE RENYI ENTROPY DENSITY EVOLUTION IN A VQA CIRCUIT WITH CLASSICAL SHADOWS AND READOUT NOISE

In this section, we adapt the heuristic model from Eq. (6) to take into account the effect of noise in the measurement circuit of the classical shadows protocol and readout errors on the Renyi-2 entropy and purity evolution.

We assume that we can represent local depolarizing noise, and readout and measurement circuit noise, by two successive global depolarizing noise channels $\mathcal{N}_P : \rho \mapsto (1-P)\rho + P\frac{I}{2^n}$ and $\mathcal{N}_{P_M} : \rho \mapsto (1-P_M)\rho + P_M\frac{I}{2^n}$ (for an n-qubit quantum state ρ) with depolarizing probabilities P and P_M, respectively, placed at the very end of the ideal circuit. As we did in Eq. (5), we can link those two probabilities to fitting parameters α_1, α_2 (depolarizing) and β (readout) by equating the probabilities of no error occurring in both our local noise model and global noise model:

$$1 - P := (1-p_1)^{|1\text{Q gates}|}(1-p_2)^{|2\text{Q gates}|} \tag{A1}$$

$$= (1-p_1)^{2nD}(1-p_2)^{(n-1)D} \tag{A2}$$

$$:= e^{-2\alpha_1 nD}e^{-\alpha_2(n-1)D}, \tag{A3}$$

and

$$1 - P_M := (1-p_m)^n := e^{-\beta n}, \tag{A4}$$

where $\alpha_i = \ln((1-p_i)^{-1})$ and $\beta = \ln((1-p_m)^{-1})$ are the local depolarizing error probabilities p_i and p_m in the low-noise limit, $\alpha_i \approx p_i$ and $\beta \approx p_m$. Our model for the purity is given by

$$\mathrm{Tr}\left[\rho_D^2\right] = (1-2^{-n})(e^{-2(2\alpha_1 nD+\alpha_2(n-1)D+\beta n)} - 1) + 1. \tag{A5}$$

APPENDIX B: PURITY ESTIMATE FROM THE CLASSICAL SHADOWS PROTOCOL WITH PAULI BASIS MEASUREMENTS

In this section dedicated to purity estimation using the Pauli-basis classical shadows protocol from Ref. [28], we provide a simple expression for the corresponding purity estimate in Appendix B 1, give an upper bound on the sampling cost in Appendix B 2, analyze the performance of the protocol based on how mixed the target state is in Appendix B 3, and discuss the existence of a systematic error in the purity estimate of a derandomized version of the classical shadows technique in Appendix B 4.

1. Purity estimate

Consider the purity estimate from the classical shadows protocol given in Ref. [37]:

$$\hat{P}^{(M,K)}_{\text{shadows}} := \frac{1}{M(M-1)}\sum_{m\neq m'}\mathrm{Tr}[\hat{\rho}^{(m)}\hat{\rho}^{(m')}], \tag{B1}$$

where

$$\hat{\rho}^{(m)} := \frac{1}{K}\sum_{k=1}^{K}\bigotimes_{n=1}^{N}\left(3\left(U_n^{(m)}\right)^\dagger|s_n^{(m,k)}\rangle\langle s_n^{(m,k)}|U_n^{(m)} - I\right). \tag{B2}$$

For the special case of Pauli measurements, it is possible to further simplify the formula for the purity estimate. First, consider the product $\hat{\rho}^{(m)}\hat{\rho}^{(m')}$:

$$\hat{\rho}^{(m)}\hat{\rho}^{(m')} = \frac{1}{K^2}\sum_{k,k'=1}^{K}\bigotimes_{n=1}^{N}\left(3\left(U_n^{(m)}\right)^\dagger|s_n^{(m,k)}\rangle\langle s_n^{(m,k)}|U_n^{(m)} - I\right)\left(3\left(U_n^{(m')}\right)^\dagger|s_n^{(m',k')}\rangle\langle s_n^{(m',k')}|U_n^{(m')} - I\right), \tag{B3}$$

$$= \frac{1}{K^2}\sum_{k,k'=1}^{K}\bigotimes_{n=1}^{N} \tag{B4}$$

$$\big(9\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')} \tag{B5}$$

$$-3\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)} - 3\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')} + I\big). \tag{B6}$$

Taking the trace of the above product of shadows and exploiting the linearity of the trace and the fact that the trace of a tensor product is the product of the traces, we have

$$\mathrm{Tr}[\hat\rho^{(m)}\hat\rho^{(m')}] = \frac{1}{K^2}\sum_{k,k'=1}^{K}\prod_{n=1}^{N} \tag{B7}$$

$$\big(9\mathrm{Tr}\big[\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')}\big] \tag{B8}$$

$$-3\mathrm{Tr}\big[\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big] \tag{B9}$$

$$-3\mathrm{Tr}\big[\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')}\big] + \mathrm{Tr}[I]\big), \tag{B10}$$

where each of the traces above are traces of 2×2 matrices, so that

$$\mathrm{Tr}[I] = 2. \tag{B11}$$

As the trace is cyclic and the $U_n^{(m)}$ are unitary matrices and since the $|s_n^{(m',k')}\rangle$ form an orthonormal family of quantum states, we deduce that

$$\mathrm{Tr}\big[\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big] = 1, \tag{B12}$$

$$\mathrm{Tr}\big[\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')}\big] = 1. \tag{B13}$$

Simplifying Eq. (B8) needs further work; however, for the special case of Pauli measurements, possible values for this trace can be restricted to the three values $\{0, 1, 1/2\}$. To see this, first notice that since the trace is cyclic,

$$\mathrm{Tr}\big[\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\!\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')}\big] \tag{B14}$$

$$= \mathrm{Tr}\big[\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\!\big\langle s_n^{(m',k')}\big|U_n^{(m')}\big(U_n^{(m)}\big)^\dagger\big|s_n^{(m,k)}\big\rangle\big] \tag{B15}$$

$$= \mathrm{Tr}\big[\big|\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\big|^2\big] \tag{B16}$$

$$= \big|\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\big|^2 = \gamma_{n,m,m',k,k'}, \tag{B17}$$

where $s_n^{(m,k)}, s_n^{(m',k')} \in \{0, 1\}$ and $U_n^{(m)}, U_n^{(m')} \in \{H, HS^\dagger, I\}$. Exploring all possible combinations for $s_n^{(m,k)}, s_n^{(m',k')}, U_n^{(m)}, U_n^{(m')}$, it can then be shown that $\gamma_{n,m,m',k,k'}$ can only take three different kinds of values: this is summarized in Table I.

Finally, combining all previous results for the trace, we conclude that, for Pauli measurements, the purity estimate can be simplified as

$$\hat P_{\mathrm{shadows}}^{(M,K)} = \frac{1}{M(M-1)}\sum_{m\neq m'}\frac{1}{K^2}\sum_{k,k'=1}^{K}\prod_{n=1}^{N}(9\times\gamma_{n,m,m',k,k'} - 4). \tag{B18}$$

Noting that $\gamma_{n,m,m',k,k'} = \gamma_{n,m',m,k,k'}$, this further simplifies to

$$\boxed{\hat P_{\mathrm{shadows}}^{(M,K)} = \frac{2}{M(M-1)}\sum_{m}\sum_{m'|m'<m}\frac{1}{K^2}\sum_{k,k'=1}^{K}\prod_{n=1}^{N}(9\times\gamma_{n,m,m',k,k'} - 4)}, \tag{B19}$$

where

$$\gamma := \gamma_{n,m,m',k,k'} = \big|\big\langle s_n^{(m,k)}\big|U_n^{(m)}\big(U_n^{(m')}\big)^\dagger\big|s_n^{(m',k')}\big\rangle\big|^2 \tag{B20}$$

is given by Table I.

Considering N_{g} groups containing M_{g} measurement settings each, we may build an independent purity estimate for each group, then obtain the median of those purity estimates:

$$\boxed{\hat P_{\mathrm{shadows}} = \mathrm{median}\big\{\hat P_{\mathrm{shadows},1}^{(M_{\mathrm{g}},K)}, \ldots, \hat P_{\mathrm{shadows},N_{\mathrm{g}}}^{(M_{\mathrm{g}},K)}\big\}}, \tag{B21}$$

where we set $M = N_{\mathrm{g}}M_{\mathrm{g}}$.

2. Upper bound on the sampling cost

Fix accuracy parameters $\delta, \epsilon > 0$ and $K = 1$. Set $M = N_{\mathrm{g}}M_{\mathrm{g}}$, where

$$N_{\mathrm{g}} = 2\ln(2/\delta), \tag{B22}$$

TABLE I. Possible values for $\gamma_{n,m,m',k,k'}$ depending on the value of $s_n^{(m,k)}, s_n^{(m',k')}, U_n^{(m)}, U_n^{(m')}$, where $\gamma_{n,m,m',k,k'} := |\langle s_n^{(m,k)}|U_n^{(m)}(U_n^{(m')})^\dagger|s_n^{(m',k')}\rangle|^2$.

$U_n^{(m)} \times (U_n^{(m')})^\dagger$	$\lvert\langle 0\rvert\cdot\lvert 0\rangle\rvert^2$	$\lvert\langle 0\rvert\cdot\lvert 1\rangle\rvert^2$	$\lvert\langle 1\rvert\cdot\lvert 0\rangle\rvert^2$	$\lvert\langle 1\rvert\cdot\lvert 1\rangle\rvert^2$
$H \times H = I$	1	0	0	1
$H \times SH = HSH$	1/2	1/2	1/2	1/2
$H \times I = H$	1/2	1/2	1/2	1/2
$HS^\dagger \times H = HS^\dagger H$	1/2	1/2	1/2	1/2
$HS^\dagger \times SH = I$	1	0	0	1
$HS^\dagger \times I = HS^\dagger$	1/2	1/2	1/2	1/2
$I \times H = H$	1/2	1/2	1/2	1/2
$I \times SH = SH$	1/2	1/2	1/2	1/2
$I \times I = I$	1	0	0	1

and

$$M_g = \frac{272}{\epsilon^2} \times 4^n ||S||_\infty^2 = \frac{272}{\epsilon^2} \times 4^n = O(4^n/\epsilon^2). \quad \text{(B23)}$$

From Theorem S4 and Proposition S4 of the Supplemental Material of Ref. [28], we know that a collection of $M = N_g M_g$ independent classical shadows allows for accurately predicting the purity $\mathrm{Tr}[\rho^2]$ of state ρ:

$$|\hat{P}_{\text{shadows}} - \mathrm{Tr}[\rho^2]| \leqslant \epsilon, \quad \text{(B24)}$$

with probability at least $1 - \delta$. Note that setting $K > 1$ can only improve the estimate.

3. Theoretical bound on the variance of the purity estimate

To derive an upper bound on the variance of the purity estimate given by the classical shadows protocol, and show its dependency on the purity of the target state, we exploit results from the Supplemental Material of the original paper on the classical shadows protocol [28].

First, recall that the purity estimate is given by

$$\hat{P}_{\text{shadows}}^{(M,K)} := \frac{2}{M(M-1)} \sum_{m<m'} \mathrm{Tr}[\hat{\rho}^{(m)}\hat{\rho}^{(m')}] = \binom{M}{2}^{-1} \sum_{m<m'} \mathrm{Tr}[\hat{\rho}^{(m)}\hat{\rho}^{(m')}], \quad \text{(B25)}$$

where

$$\hat{\rho}^{(m)} = \frac{1}{K}\sum_{k=1}^{K} \hat{\rho}^{(m,k)} = \frac{1}{K}\sum_{k=1}^{K} \bigotimes_{n=1}^{N} \left(3\left(U_n^{(m)}\right)^\dagger |s_n^{(m,k)}\rangle\langle s_n^{(m,k)}| U_n^{(m)} - I\right). \quad \text{(B26)}$$

Performance bounds on the variance derived in Ref. [28] are for the single-shot limit $K = 1$. The variance of our purity estimate in the $K > 1$ case may be upper bounded by the single-shot variance as follows:

$$V\left[\hat{P}_{\text{Shadows}}^{(M,K)}\right] = V\left[\frac{1}{K^2}\sum_{k,k'} \hat{P}(k,k')\right] \quad \text{where} \quad \hat{P}(k,k') \equiv \frac{2}{M(M-1)} \sum_{m<m'} \mathrm{Tr}[\hat{\rho}^{(m,k)}\hat{\rho}^{(m',k')}] \quad \text{(B27)}$$

$$= \frac{1}{K^4} V\left[\sum_{k,k'} \hat{P}(k,k')\right] \quad \text{(B28)}$$

$$\leqslant \frac{1}{K^4} K \sum_{k} V\left[\sum_{k'} \hat{P}(k,k')\right] \quad \text{(B29)}$$

$$\leqslant \frac{1}{K^4} K^2 \sum_{k,k'} V[\hat{P}(k,k')] \quad \text{(B30)}$$

$$= \frac{1}{K^2} \sum_{k,k'} V[\hat{P}(1,1)] \quad \text{(B31)}$$

$$= V\left[\hat{P}_{\text{shadows}}^{(M,1)}\right], \quad \text{(B32)}$$

where we have used twice that the variance of a sum of random variables is upper bounded by the number of elements in the sum times the sum of the variances of those random variables. Indeed, for n discrete random variables X_k, we have

$$V\left[\sum_{k=1}^{n} X_k\right] = \sum_{i,j} \mathrm{Cov}[X_i, X_j] \quad \text{(B33)}$$

$$\leqslant \sum_{i,j} \sqrt{V[X_i]V[X_j]} \quad \text{(B34)}$$

$$\leqslant \sum_{i,j} \frac{V[X_i] + V[X_j]}{2} \quad \text{(B35)}$$

$$= \frac{1}{2}\left(n \sum_i V[X_i] + n \sum_j V[X_j] \right) \tag{B36}$$

$$= n \sum_k V[X_k], \tag{B37}$$

where we have used Cauchy-Schwarz inequality to upper bound the covariance with variances, as well as the fact that $\sqrt{xy} \leqslant (x+y)/2$ for positive x, y.

Now using a similar derivation as in Eq. (S65) of the Supplemental Material of Ref. [28], the single-shot variance satisfies

$$V\left[\hat{P}^{(M,1)}_{\text{shadows}}\right] = \binom{M}{2}^{-2} \sum_{m<m'} (\mathbf{E}[\mathrm{Tr}[\hat{\rho}^{(m)}\hat{\rho}^{(m')}]^2] - \mathrm{Tr}[\rho^2]^2) \tag{B38}$$

$$+ 2\binom{M}{2}^{-2} \sum_{m<m'} \sum_{n\neq m,m'} (\mathbf{E}[\mathrm{Tr}[\hat{\rho}^{(m)}\hat{\rho}^{(m')}]Tr[\hat{\rho}^{(m)}\hat{\rho}^{(n)}]] - \mathrm{Tr}[\rho^2]^2) \tag{B39}$$

$$= \binom{M}{2}^{-2} \sum_{m<m'} V[\mathrm{Tr}[\hat{\rho}^{(1)}\hat{\rho}^{(2)}]] + 2\binom{M}{2}^{-2} \sum_{m<m'} \sum_{n\neq m,m'} V[\mathrm{Tr}[\hat{\rho}^{(1)}\rho]] \tag{B40}$$

$$= \binom{M}{2}^{-1} V[\mathrm{Tr}[\hat{\rho}^{(1)}\hat{\rho}^{(2)}]] + 2\binom{M}{2}^{-1} (M-2) V[\mathrm{Tr}[\hat{\rho}^{(1)}\rho]] \tag{B41}$$

$$\leqslant \frac{4}{M^2} V[\mathrm{Tr}[\hat{\rho}^{(1)}\hat{\rho}^{(2)}]] + \frac{4}{M} V[\mathrm{Tr}[\hat{\rho}^{(1)}\rho]] \tag{B42}$$

$$= \frac{4}{M^2} V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\hat{\rho}^{(2,1)}]] + \frac{4}{M} V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\rho]], \tag{B43}$$

where in the line of Eq. (B42) we have used the fact that $\binom{M}{2}^{-1} \leqslant \frac{4}{M^2}$ and $2\binom{M}{2}^{-1}(M-2) \leqslant \frac{4}{M}$ for $M \geqslant 2$.

Noticing that

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\hat{\rho}^{(2,1)}]] = V[\mathrm{Tr}[O_2\hat{\rho}_2]], \quad \text{where } O_2 = S \quad \text{and} \quad \hat{\rho}_2 = \hat{\rho}^{(1,1)} \otimes \hat{\rho}^{(2,1)}, \tag{B44}$$

and

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\rho]] = V[\mathrm{Tr}[O_1\hat{\rho}_1]], \quad \text{where } O_1 = \rho \quad \text{and} \quad \hat{\rho}_1 = \hat{\rho}^{(1,1)}, \tag{B45}$$

we can use Lemma (S1) of the Supplemental Material of Ref. [28] to upper bound those variance terms with the shadow norm squared:

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\hat{\rho}^{(2,1)}]] \leqslant ||O_{2,0}||^2_{\text{shadow}}, \quad \text{where } O_{2,0} = S - \mathrm{Tr}[S]\frac{I_{2n}}{2^{2n}} = S - \frac{I_{2n}}{2^n}, \tag{B46}$$

and

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\rho]] \leqslant ||O_{1,0}||^2_{\text{shadow}}, \quad \text{where } O_{1,0} = \rho - \mathrm{Tr}[\rho]\frac{I_n}{2^n} = \rho - \frac{I_n}{2^n}. \tag{B47}$$

Finally, noticing that the swap operator has locality $2n$, and for a state ρ of unknown locality upper bounded by n, we use the last line of the proof of Proposition (S3) from the Supplemental Material of Ref. [28] to upper bound each shadow norm as a function of the two-norm (Hilbert-Schmidt norm) squared ($||O||_2^2 = \mathrm{Tr}[O^2]$):

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\hat{\rho}^{(2,1)}]] \leqslant 2^{2n}||O_{2,0}||_2^2 = 4^n \mathrm{Tr}\left[O_{2,0}^2\right] = 4^n \mathrm{Tr}\left[S^2 - 2S\frac{I_{2n}}{2^n} + \frac{I_{2n}}{2^{2n}}\right] = 4^n(4^n - 1) \leqslant 16^n, \tag{B48}$$

and

$$V[\mathrm{Tr}[\hat{\rho}^{(1,1)}\rho]] \leqslant 2^n||O_{1,0}||_2^2 = 2^n \mathrm{Tr}\left[O_{1,0}^2\right] = 2^n \mathrm{Tr}\left[\rho^2 - 2\frac{\rho}{2^n} + \frac{I_n}{2^{2n}}\right] = 2^n\left(\mathrm{Tr}[\rho^2] - \frac{1}{2^n}\right) \leqslant 2^n \mathrm{Tr}[\rho^2], \tag{B49}$$

where it may be interesting to note that the last term of the above equation would have become $2^{2k-n}\mathrm{Tr}[\rho^2]$ if we had decided to take into account the locality k of the state ρ in the derivation.

Combining everything together, we conclude that

$$V\left[\hat{P}^{(M,K)}_{\text{shadows}}\right] \leqslant V\left[\hat{P}^{(M,1)}_{\text{shadows}}\right] \leqslant \frac{4}{M^2}16^n + \frac{4}{M}2^n \mathrm{Tr}[\rho^2]. \tag{B50}$$

Therefore, the smaller the purity of the target state (i.e., the more mixed the state), the smaller the upper bound on the variance of the purity estimate output by the classical shadows protocol. Choosing M of order 4^n, we see that the effect of the purity of the state in the upper bound on the variance should become less noticeable as we consider larger system sizes and the 16^n term becomes dominant.

4. Systematic error in the derandomized version of the classical shadows protocol

Changing circuits is relatively costly in execution time for some platforms like superconducting qubits. Since our entropy accumulation QPU data were limited in number of qubits ($n = 3$) and therefore we were expecting to see the occurrence of most measurements and in similar number, we initially opted for a derandomization of the measurement choice, where we consider one by one all of the $3^3 = 27$ possible measurement settings with $K = 1000$ classical snapshots of the output state for each of them. The equivalent of Fig. 4 for QPU entropy accumulation but in the case of derandomized measurements of the classical shadows protocol corresponds to Fig. 7, where we can observe the presence of a systematic bias in the purity or Renyi-2 entropy density estimate provided by the classical shadows protocol.

5. Systematic error for low number of samples

When using the classical shadows protocol for Renyi-2 entropy or purity estimation, it is crucial to choose a sufficiently large number of samples, MK, and more precisely a number of measurement settings, M, to avoid any systematic error in the estimate. We illustrate the importance of the choice of value for M in Fig. 8. In that figure, as an example, we consider an ($n = 3$)-qubit VQA circuit as in Fig. 1, with fixed circuit parameters, under local depolarizing noise from calibration data of Aspen-M-3. Then, we estimate the output Renyi-2 entropy density as a function of circuit depth using the classical shadows protocol with all parameters fixed except the number of measurement settings, M. We plot the resulting entropy density estimate for $M \in \{50; 350; 800; 1000\}$, where each value corresponds to a different color of error bars in the figure. As a reference, we show the exact entropy density evolution as a solid black line. We observe that, for very small M value (e.g., $M = 50$), the estimate is above the exact value. As M increases, the systematic error decreases, and the estimate converges to the exact value (e.g., $M = 350$). A minor systematic error can sometimes be observed for larger M values (e.g., $M = 800$) where the estimated entropy evolution is slightly below the exact curve; however, we have observed that increasing M even further (e.g., $M = 1000$) removes this small deviation.

APPENDIX C: CIRCUIT-SIZE BOUNDARY FOR REACHABLE QUANTUM ADVANTAGE

In this section, we prove that our bound on the circuit size boundary for reachable quantum advantage [Eq. (C2)] is tighter than that of Ref. [23] for a large number of qubits.

The circuit-size condition for classical superiority from Ref. [23] is

$$D \geqslant \frac{1}{2\alpha} \ln (1-c)^{-1}, \tag{C1}$$

while at $n \to \infty$, ours is

$$D \geqslant \frac{1}{2\alpha} c \ln (2). \tag{C2}$$

Let $f(c) := c \ln(2) - \ln((1-c)^{-1}) = c \ln(2) + \ln(1-c)$, $\forall c \in [0,1]$. We have $\frac{df}{dc}(c) = \ln(2) - \frac{1}{1-c}$, which can be shown to satisfy $\frac{df}{dc}(c) < 0$, $\forall c \in [0,1]$. Therefore, f is a strictly decreasing function on its domain $c \in [0,1]$. Since $f(0) = 0$, we conclude that $f(c) \leqslant 0$ for all values of $c \in [0,1]$. This means that for large problem sizes, i.e., $n \to \infty$, quantum advantage will be lost at a shorter depth [Eq. (C2)] than that predicted in Ref. [23]. Note that the above argument remains valid even for the more realistic case where the entropy density threshold c keeps a dependency with the system size n, i.e., $c = c(n)$. This is because the value $c(n)$ remains the same input for both bounds and $f(c) \leqslant 0$ for all values of $c \in [0,1]$ remains true irrespective of the variables that may define c (n in our case or others).

APPENDIX D: RIGETTI'S ASPEN-M-3 AND LINK BETWEEN CALIBRATION DATA AND NOISE PARAMETERS

We link noise parameters of the noise model used for numerical simulations with typical calibration data given for a quantum device.

All QPU results shown in this article were obtained on Rigetti's eight-qubit Aspen-M-3 device (see Fig. 9). For one-qubit circuits, qubit 100 was used; for two-qubit circuits, qubits 100 and 101 were used; for three-qubit circuits, qubits 100, 101, and 102 were used; and so on. Calibration data from the day of our circuit runs on the QPU is provided in Table II. Typical gate times for the device are $T_{\text{gate},1} = 40$ ns and $T_{\text{gate},2} = 240$ ns for single-qubit and two-qubit gates. The measurement process can take more than $T_{\text{meas}} = 1$ μs [38].

1. Gate noise

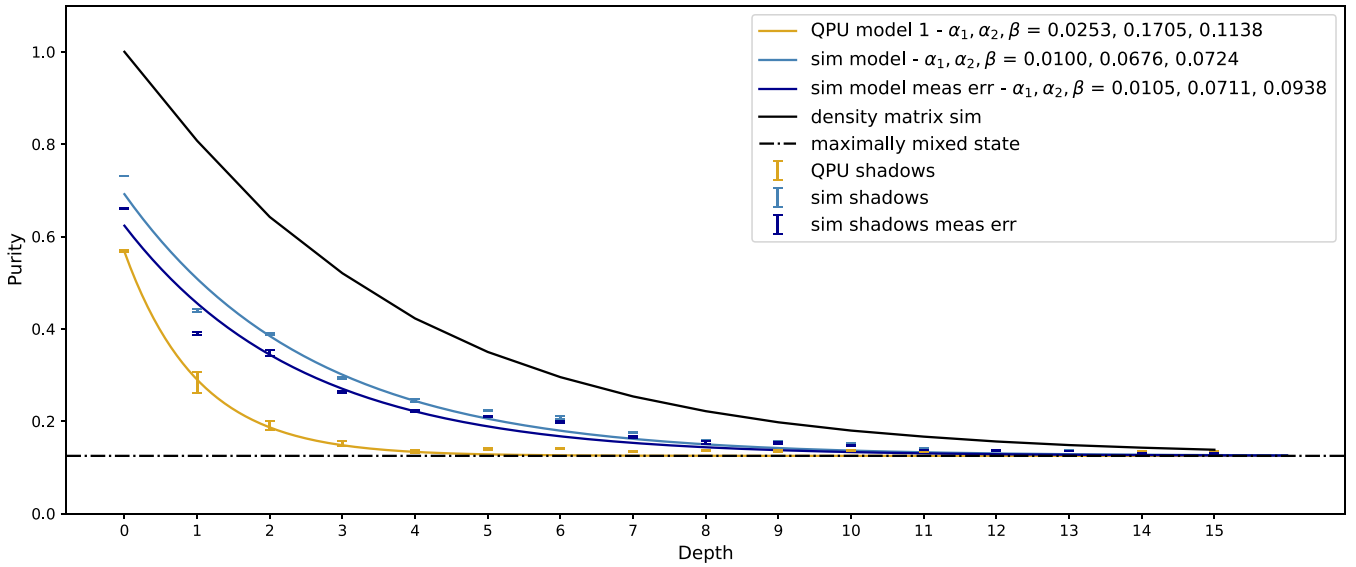
a. Local depolarizing noise

Assuming local depolarizing noise as the noise model on the device for classical simulations, it is possible to link one-qubit and two-qubit gate fidelities (F_1 and F_2) from calibration data to one-qubit and two-qubit depolarizing probabilities (p_1 and p_2) as follows:

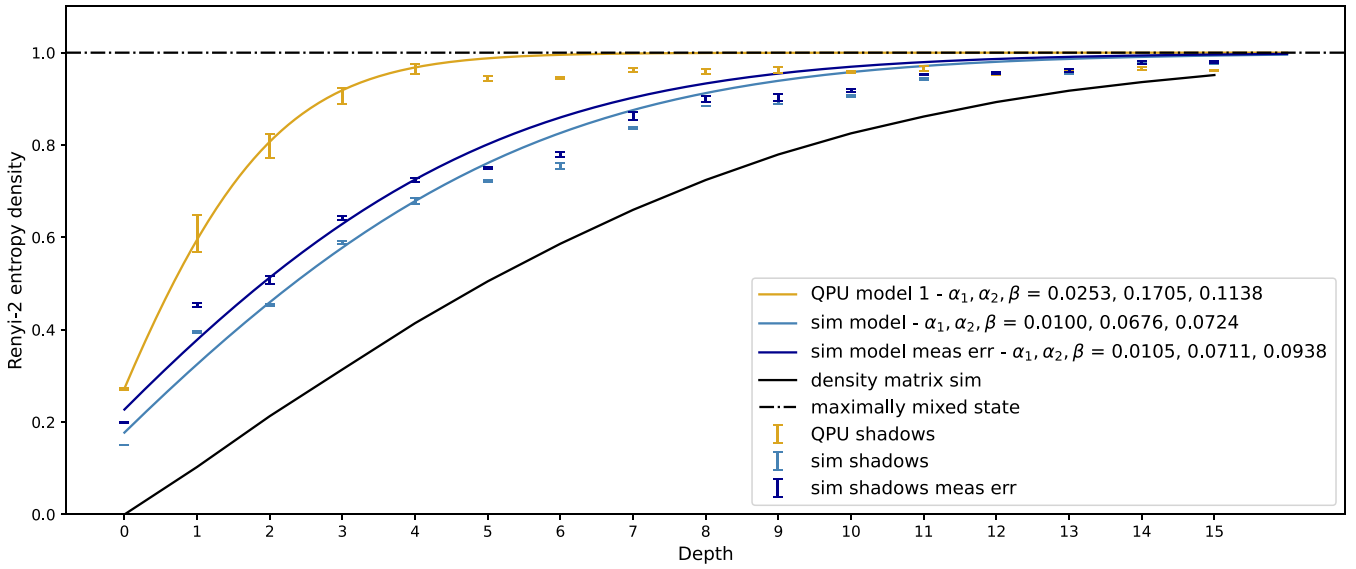
$$p_1 = 2(1 - F_1), \tag{D1}$$

$$p_2 = \tfrac{4}{3}(1 - F_2), \tag{D2}$$

where we have used the fact that the fidelity between a pure state $|\Psi\rangle$ and a mixed state ρ is $F = \langle\Psi|\rho|\Psi\rangle$ and taking for $|\Psi\rangle$ a pure state evolved with some ideal gate U, and for ρ a pure state evolved with U followed by a depolarizing channel (e.g., see Ref. [39]). For our classical simulations, we considered the median sim RB fidelity (one-qubit gate fidelity) and the median cz fidelity (two-qubit gate fidelity) in agreement with the two-qubit gates of our VQA ansatz.



(a) Purity



(b) Renyi entropy density

FIG. 7. QPU results and validation of our heuristic model. We consider a VQA circuit as in Fig. 1 applied to $n = 3$ qubits, with fixed circuit parameters [fixed random seed `np.random.seed(837)`]. Gold error bars give the purity and Renyi-2 entropy evolution as functions of circuit depth on Rigetti's QPU Aspen-M-3 using the derandomized version of the protocol (with $M = 3^3 = 27$ measurement settings considered in order). Each error bar was obtained by running the classical shadows protocol three times, and computing the average estimate and the standard deviation over those three samples. Results obtained in the same setting but using a local depolarizing noise model ($p_1 = 0.008$ and $p_2 = 0.054$ from calibration data of Aspen-M-3) correspond to the blue error bars; dark (light) blue error bars were obtained with (without) readout errors from calibration data. The black solid lines show the purity and Renyi-2 entropy density evolution obtained via density matrix simulation under our local depolarizing noise model. The black horizontal dash-dotted lines correspond to the purity or entropy density of the maximally mixed state $\sigma_0 := I/2^n$. Fit of our global depolarizing heuristic model, where we have imposed that $\alpha_2/\alpha_1 = p_2/p_1$, corresponds to the solid gold and blue lines.

b. Amplitude damping (T_1 relaxation) and local depolarizing noise

In the case where gate noise in the device can be modeled via a *combination* of local depolarizing noise followed by amplitude damping, one may link one-qubit and two-qubit gate fidelities (F_1 and F_2), the relaxation time T_1, and single-qubit and two-qubit gate times $T_{\text{gate},1}$ and $T_{\text{gate},2}$—as obtained from calibration data—to the corresponding one-qubit and two-qubit depolarizing probabilities (p_1 and p_2) as well as the

TABLE II. Aspen-M-3 calibration data on 9 December 2023 at 9:59 pm (UTC+00:00).

Pair	fXY	fXY std err	fCZ	fCZ std err	Average T1 (μs)	Average T2 (μs)	Average fActiveReset	Average fRO
100-101	0.9614	0.007829	0.9598	0.0058	19.3	33.22	0.9943	0.941
100-107	0.8553	0.007008	0.9769	0.005897	27.2	45.47	0.996	0.944
101-102	0.9626	0.009064			21.36	32.92	0.996	0.9165
102-103	0.8409	0.01479	0.9711	0.007058	26.6	29.48	0.9858	0.9135
103-104	0.8534	0.006161	0.909	0.003757	31.31	43.49	0.8827	0.9455
104-105	0.812	0.007665	0.8917	0.004528	29.04	35.61	0.8945	0.945
105-106	0.9756	0.003552			27.46	15.55	0.9885	0.9735
106-107	0.9828	0.002881	0.9948	0.002183	32.39	39.13	0.9888	0.98
Qubit	T1 (μs)	T2 (μs)	f1QRB	f1QRB std err	f1Q sim. RB	f1Q sim. RB std err	fActiveReset	fRO
100	11.82	20.1	0.997	0.000175	0.9957	0.000109	0.9955	0.908
101	26.78	46.34	0.9997	0.000382	0.9989	5.61×10^{-5}	0.993	0.974
102	15.95	19.5	0.9977	0.00014	0.9959	0.000148	0.999	0.859
103	37.25	39.46	0.9953	0.000827	0.9869	0.000571	0.9725	0.968
104	25.36	47.52	0.8209	0.04649	0.8665	0.0576	0.793	0.923
105	32.72	23.71	0.9984	6.81×10^{-5}	0.9949	0.000161	0.996	0.967
106	22.2	7.401	0.999	4.9×10^{-5}	0.9973	0.000185	0.981	0.98
107	42.58	70.85	0.9995	6.44×10^{-5}	0.9989	2.96×10^{-5}	0.9965	0.98

amplitude damping probabilities (γ_1 and γ_2), respectively, as follows:

$$\gamma_1 = 1 - e^{-\frac{T_{\text{gate},1}}{T_1}}, \tag{D3}$$

$$\gamma_2 = 1 - e^{-\frac{T_{\text{gate},2}}{T_1}}, \tag{D4}$$

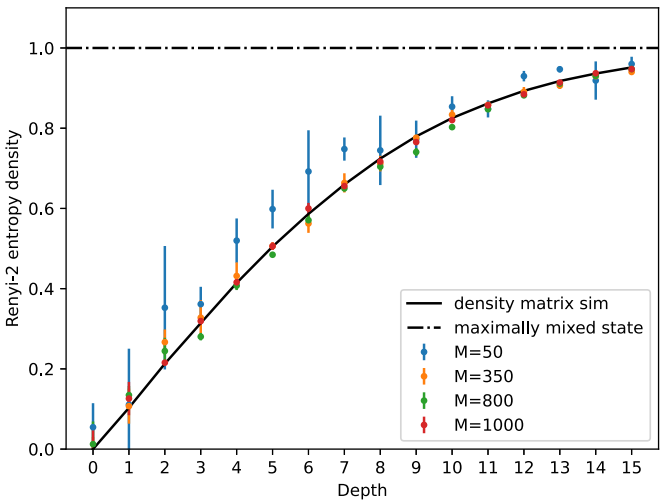


FIG. 8. Systematic error in the classical shadows estimate for low sample complexity. We consider an ($n = 3$)-qubit VQA circuit as in Fig. 1, with fixed circuit parameters, under local depolarizing noise (with parameters $p_1 = 0.008$ and $p_2 = 0.054$ from calibration data of Aspen-M-3). The solid black line corresponds to the Renyi-2 entropy density evolution of the circuit output for different circuit depths. We also estimate the entropy density using the Pauli version of the classical shadows protocol for a varying number of measurement settings $M \in \{50; 350; 800; 1000\}$ but with fixed parameters $K = 1000$ and $N_g = 5$ and each time running the protocol three times to obtain error bars. The black horizontal dash-dotted line corresponds to the entropy density of the maximally mixed state $\sigma_0 := I/2^n$.

$$p_1 = 1 - \frac{2(3F_1 - 1) - x^2}{1 + 2x}, \tag{D5}$$

$$p_2 = 1 - \frac{4(5F_2 - 1) - 1}{y(4 + 6y + 4y^2 + y^3)}, \tag{D6}$$

where $x = \sqrt{1 - \gamma_1}$ and $y = \sqrt{1 - \gamma_2}$. Here, we have used Eq. (3) from Ref. [40]—which links average gate fidelity and entanglement fidelity—to easily express average gate fidelities F_1 and F_2 as functions of noise parameters p_1, γ_1 and p_2, γ_2, respectively. In particular, we defined single-qubit and two-qubit noise channels:

$$\mathcal{E}_1 = A_{\gamma_1} \circ D_{p_1}, \tag{D7}$$

$$\mathcal{E}_2 = A_{\gamma_2} \circ D_{p_2}, \tag{D8}$$

where the single-qubit depolarizing and amplitude damping channels are

$$D_{p_1}\colon \rho \mapsto (1 - p_1)\rho + p_1 \frac{I}{2}, \tag{D9}$$

$$A_{\gamma_1}\colon \rho \mapsto E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger, \tag{D10}$$

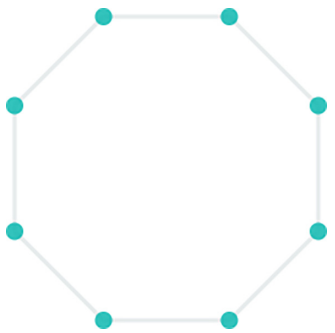


FIG. 9. Eight-qubit Aspen-M-3 QPU.

and the two-qubit depolarizing and amplitude damping channels are

$$D_{p_2}\colon \rho \mapsto (1-p_2)\rho + p_2\frac{I}{4}, \tag{D11}$$

$$A_{\gamma_2}\colon \rho \mapsto F_0\otimes F_0\rho(F_0\otimes F_0)^\dagger + F_0\otimes F_1\rho(F_0\otimes F_1)^\dagger + F_1\otimes F_0\rho(F_1\otimes F_0)^\dagger + F_1\otimes F_1\rho(F_1\otimes F_1)^\dagger, \tag{D12}$$

with Kraus operators

$$E_0=\begin{pmatrix}1 & 0\\ 0 & \sqrt{1-\gamma_1}\end{pmatrix},\quad E_1=\begin{pmatrix}0 & \sqrt{\gamma_1}\\ 0 & 0\end{pmatrix},\quad F_0=\begin{pmatrix}1 & 0\\ 0 & \sqrt{1-\gamma_2}\end{pmatrix},\quad F_1=\begin{pmatrix}0 & \sqrt{\gamma_2}\\ 0 & 0\end{pmatrix}. \tag{D13}$$

2. Readout noise

a. Amplitude damping (T_1 relaxation)

In classical simulations with assumed T_1 relaxation in the measurement stage, we add a layer of single-qubit amplitude damping channels with error probability $\gamma_{\text{meas}} = 1 - e^{-\frac{T_{\text{meas}}}{T_1}} = 0.05$ (based on calibration data of Aspen-M-3) just before the computational basis measurement. More precisely, the single-qubit amplitude damping channel in the measurement stage is defined as

$$A_{\gamma_{\text{meas}}}\colon \rho \mapsto G_0\rho G_0^\dagger + G_1\rho G_1^\dagger, \tag{D14}$$

where

$$G_0=\begin{pmatrix}1 & 0\\ 0 & \sqrt{1-\gamma_{\text{meas}}}\end{pmatrix},\quad G_1=\begin{pmatrix}0 & \sqrt{\gamma_{\text{meas}}}\\ 0 & 0\end{pmatrix}. \tag{D15}$$

b. Classical bit flip

For classical simulations where readout errors were included, we estimated the probabilities $P(0|1)$ and $P(1|0)$ on Aspen-M-3, and obtained

$$P(0|1) = 0.03484722222,\quad P(1|0) = 0.022177777, \tag{D16}$$

where $P(m|n)$ is the probability of obtaining outcome m instead of n due to readout error. In our heuristic model from Eq. (A5), adapted to take into account readout errors, p_m is related to $P(0|1)$ and $P(1|0)$ in the case where those have similar values (symmetric readout error). For the more accurate classical simulations with added T_1 relaxation effects, we take $P(0|1) = P(1|0) = 0.03$, which is a reasonable assumption based on the observed asymmetry and to make the classical simulation more practical since standard publicly available calibration data usually provide readout fidelities $F_{\text{RO}} = 1 - P(0|1) = 1 - P(1|0)$, which do not give information on the error asymmetry.

[1] D. Gottesman, Class of quantum error-correcting codes saturating the quantum Hamming bound, Phys. Rev. A **54**, 1862 (1996).

[2] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, in *Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics* (American Mathematical Society, Providence, 2010), Vol. 68, pp. 13–58.

[3] J. Roffe, Quantum error correction: An introductory guide, Contemp. Phys. **60**, 226 (2019).

[4] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[5] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[6] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[7] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[8] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen, L.-C. Peng, Y.-H. Luo, D. Wu, S.-Q. Gong, H. Su, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, *et al.*, Phase-programmable Gaussian boson sampling using stimulated squeezed light, Phys. Rev. Lett. **127**, 180502 (2021).

[9] S. Yu, Z.-P. Zhong, Y. Fang, R. B. Patel, Q.-P. Li, W. Liu, Z. Li, L. Xu, S. Sagona-Stophel, E. Mer, S. E. Thomas, Y. Meng, Z.-P. Li, Y.-Z. Yang, Z.-A. Wang, N.-J. Guo, W.-H. Zhang, G. K. Tranmer, Y. Dong, Y.-T. Wang, *et al.*, A universal programmable Gaussian boson sampler for drug discovery, Nat. Comput. Sci. **3**, 839 (2023).

[10] J. M. Arrazola and T. R. Bromley, Using Gaussian boson sampling to find dense subgraphs, Phys. Rev. Lett. **121**, 030503 (2018).

[11] S. Aaronson and S.-H. Hung, Certified randomness from quantum supremacy, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2023), pp. 933–944.

[12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[13] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[14] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[15] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[16] J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, Efficient tensor network simulation of IBM's Eagle kicked Ising experiment, PRX Quantum **5**, 010308 (2024).

[17] T. Begušić, J. Gray, and G. K.-L. Chan, Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, Sci. Adv. **10**, eadk4321 (2024).

[18] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[19] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[20] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A **100**, 032328 (2019).

[21] S. Martiel, T. Ayral, and C. Allouche, Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way, IEEE Trans. Quantum Eng. **2**, 1 (2021).

[22] P.-L. Dallaire-Demers, M. Stęchły, J. F. Gonthier, N. T. Bashige, J. Romero, and Y. Cao, An application benchmark for fermionic quantum simulations, arXiv:2003.01862.

[23] D. Stilck França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17**, 1221 (2021).

[24] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, arXiv:2405.08810.

[25] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão, Random quantum circuits transform local noise into global white noise, Commun. Math. Phys. **405**, 78 (2024).

[26] M. Urbanek, B. Nachman, V. R. Pascuzzi, A. He, C. W. Bauer, and W. A. de Jong, Mitigating depolarizing noise on quantum computers with noise-estimation circuits, Phys. Rev. Lett. **127**, 270502 (2021).

[27] R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya, *et al.*, Quantum error correction below the surface code threshold, Nature (London) **638**, 920 (2025).

[28] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[29] The original protocol [28] takes a single shot per random setting ($K = 1$), but on platforms where reconfiguring the measurement circuit dominates the runtime, such as the superconducting QPU used here, a multishot variant ($K > 1$) is more time efficient [37,41]. Reported values of K in the literature range from 1 [42] to 10^4–10^5[43]. In this work, we use $K = 1000$ for the hardware experiments, and take the same value for simulations to make a fair comparison. Although analyzed for specific observables [44], we note that the optimal split between the number of settings, M, and shots per setting, K, for purity estimation via Pauli-basis shadows is left as an open question.

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

[31] D. F. Perez-Ramirez, Variational quantum algorithms for combinatorial optimization, arXiv:2407.06421.

[32] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, *et al.*, Scalable quantum simulation of molecular energies, Phys. Rev. X **6**, 031007 (2016).

[33] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer, Nat. Commun. **13**, 5743 (2022).

[34] J. Acharya, I. Issa, N. V. Shende, and A. B. Wagner, Estimating quantum entropy, IEEE J. Sel. Areas Inf. Theory **1**, 454 (2020).

[35] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, Exponential separations between learning with and without quantum memory, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2022), pp. 574–585.

[36] M. Demarty, Entropy density benchmarking: Analysing the accumulation of Rényi-2 entropy density due to noise in quantum circuits, GitHub (2026), https://github.com/MDemarty/Entropy_Benchmarking.

[37] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, Nat. Rev. Phys. **5**, 9 (2023).

[38] Rigetti Computing, Noise and quantum computation—pyQuil documentation (v4.17.0), https://pyquil-docs.rigetti.com/en/stable/noise.html (2021).

[39] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, Randomized benchmarking of multiqubit gates, Phys. Rev. Lett. **108**, 260503 (2012).

[40] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).

[41] P. Zeng, Tailoring randomized-measurement schemes for practical devices, Quantum Views **7**, 74 (2023).

[42] T. Zhang, J. Sun, X.-X. Fang, X.-M. Zhang, X. Yuan, and H. Lu, Experimental quantum state measurement with classical shadows, Phys. Rev. Lett. **127**, 200501 (2021).

[43] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, Experimental estimation of quantum state properties from classical shadows, PRX Quantum **2**, 010307 (2021).

[44] Y. Zhou and Q. Liu, Performance analysis of multi-shot shadow estimation, Quantum **7**, 1044 (2023).