arXiv:2108.03747v1 [quant-ph] 8 Aug 2021

A Quantum Hamiltonian Simulation Benchmark

Yulong Dong[1,3], K. Birgitta Whaley[1,2,5], and Lin Lin[3,4,5]*
[1] *Berkeley Center for Quantum Information and Computation, Berkeley, California 94720 USA*
[2] *Department of Chemistry, University of California, Berkeley, California 94720 USA*
[3] *Department of Mathematics, University of California, Berkeley, California 94720 USA*
[4] *Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA and*
[5] *Challenge Institute of Quantum Computation, University of California, Berkeley, California 94720 USA*
(Dated: August 10, 2021)

Hamiltonian simulation is one of the most important problems in quantum computation, and quantum singular value transformation (QSVT) is an efficient way to simulate a general class of Hamiltonians. However, the QSVT circuit typically involves multiple ancilla qubits and multi-qubit control gates. We propose a drastically simplified quantum circuit called the minimal QSVT circuit, which uses only one ancilla qubit to simulate a class of n-qubit random Hamiltonians. We formulate a simple metric called the quantum unitary evolution score (QUES), which is a scalable quantum benchmark and can be verified without any need for classical computation. We demonstrate that QUES is directly related to the circuit fidelity, and the classical hardness of an associated quantum circuit sampling problem. Theoretical analysis suggests under suitable assumptions, there exists an "optimal" simulation time $t^{\rm opt} \approx 4.81$, at which even a noisy quantum device may be sufficient to demonstrate the classical hardness.

Introduction

Recent years have witnessed tremendous progress in quantum hardware and quantum algorithms. As near-term quantum devices become increasingly accessible, the need for holistic benchmarking of such devices is also rapidly growing. Indeed, while most of the frequently used quantum benchmarks, such as randomized benchmarking [48] and gateset tomography [14], still focus on the performance of one or a few qubits, over the past three years a number of "whole machine" benchmarks have been proposed that aim at assessing the performance of quantum devices beyond a small number of qubits [7, 15, 28–30, 33, 55].

While results from such generic benchmarks certainly provide important characteristics of the quantum devices themselves, we are ultimately interested in applying the devices to carry out specific computational tasks. However, the circuit structure of quantum algorithms can be vastly different for different algorithms. Generic quantum benchmarks can miss structural information that is specific to a particular algorithm and which may amplify either quantum errors of certain types or errors amongst a certain group of qubits, and/or reduce errors elsewhere. In this work we address the benchmarking of quantum simulations for time-independent Hamiltonians. Such a simulation can be stated as follows: given an initial state $|\psi_0\rangle$ and a Hamiltonian $\mathfrak{H}$, evaluate the quantum state at time t according to $|\psi(t)\rangle = \exp(-it\mathfrak{H})\,|\psi_0\rangle$. Hamiltonian simulation is of immense importance in characterizing quantum dynamics for a diverse range of systems and situations in quantum physics, chemistry and materials science. Simulation of one quantum Hamiltonian by another quantum system was also one of the motivations of Feynman's 1982 proposal for design of quantum computers [34]. Hamiltonian simulation is also used as a subroutine in numerous other quantum algorithms, such as quantum phase estimation [39] and solving linear systems of equations [38].

Following the conceptualization of a universal quantum simulator using a Trotter decomposition of the time evolution operator $e^{-it\mathfrak{H}}$ [44], a number of new quantum algorithms for Hamiltonian simulation have been proposed [10, 12, 18, 46, 47]. Detailed assessment of these algorithms, with continued improvement of theoretical error bounds, has since emerged as a very active area of research [6, 9, 11, 13, 19, 21–24, 45, 56, 58]. In this context, one of the most significant developments in recent years is the quantum signal processing (QSP) method [46], and its generalization, the quantum singular value transformation (QSVT) method [36]. For sparse Hamiltonian simulation, the query complexity of QSVT matches the complexity lower bound with respect to all parameters [36, 46]. The QSVT method also enjoys another advantage, namely that the circuit quantum circuit is relatively simple, and requires very few ancilla qubits. QSVT allows one to use essentially the same parameterized quantum circuit to perform a wide range of useful computational tasks, including Hamiltonian simulation [31], solution of linear systems [36, 41, 59], and finding eigenstates of quantum Hamiltonians [42]. In this sense, it provides a "grand unification" of a large class of known quantum algorithms [49].

Despite these advantages, QSVT is generally not viewed as a suitable technique for near-term quantum devices today. This is largely because these techniques rely on an input model called "block encoding" which views the Hamiltonian $\mathfrak{H}$ as a submatrix of an enlarged unitary matrix $U_{\mathfrak{H}}$. For Hamiltonians arising from realistic applications (e.g., linear combination of products of Pauli or fermionic operators, and sparse matrices in general), the construction of $U_{\mathfrak{H}}$ often involves multiple ancilla qubits

* linlin@math.berkeley.edu

and multi-qubit control gates. Taken together, these requirements can make QSVT very difficult to implement with high fidelity and to date there has been no QSVT based Hamiltonian simulation on realistic devices.

Results

In this work we remedy this situation by identifying and demonstrating an application for QSVT on near term quantum devices that allows benchmarking of Hamiltonian simulation for a class of Hamiltonians that are relevant to recent efforts to demonstrate supremacy of quantum computation over classical computation [7]. This is the class of random Hamiltonians generated from block encoding of random unitary operators that correspond to random unitary circuits. We show that for this class of Hamiltonians it is possible to formulate a simple metric, called the quantum unitary evolution score (QUES), for the success of quantum unitary evolution. This metric is the primary output from the Hamiltonian simulation benchmark, and is directly related to the circuit fidelity. This allows verification of Hamiltonian simulation on near-term quantum devices without any need for classical computation, and the approach can be scaled to a large number of qubits.

The main result of this paper is a *very simple* quantum circuit (Figure 1), called the minimal QSVT (mQSVT) circuit. With proper parameterization, the mQSVT circuit is able to perform a Hamiltonian simulation to an arbitrary accuracy specified *a priori*. In fact, we argue that the mQSVT circuit is not only *the simplest* quantum circuit for carrying out a QSVT based Hamiltonian simulation, but that it is actually the simplest possible circuit for all tasks based on QSVT. The tradeoff is that $\mathfrak{H}$ is no longer a Hamiltonian corresponding to a given physical system, but a random Hamiltonian generated using a simple random unitary circuit, called a Hermitian random circuit block encoded matrix (H-RACBEM) [30]. However, for the purpose of benchmarking the capability of a quantum device to perform arbitrary Hamiltonian simulations, averaging over a distribution of the underlying arbitrary Hamiltonians is precisely what is required to generate a holistic benchmark protocol that samples from all possible instantiations.

The quantum circuit in Figure 1 consists of two components: an arbitrary random unitary matrix U_A that implicitly defines the Hamiltonian $\mathfrak{H}$, together with its Hermitian conjugate $U_A^\dagger$ and a series of R_z gates with carefully chosen phase factors $\{\varphi_i\}_{i=0}^{2d}$ (see Appendix C in the Supplementary Information). The mQSVT circuit makes d queries to U_A and $U_A^\dagger$, two of which are shown explicitly in Figure 1. For an n-qubit matrix $\mathfrak{H}$, the total number of qubits needed is always $n+1$, i.e., only 1 ancilla qubit, hereafter referred to as the signal qubit, is required. This is even smaller than the simplest QSVT circuit [46], which requires at least 2 ancilla qubits. However, more important than the reduction of the number of qubits is the fact that Figure 1 removes all two-qubit and multi-qubit gates outside of the unitary U_A. This means that one can choose any convenient entangling two-qubit gate (e.g.. CZ, CNOT, $\sqrt{\text{iSWAP}}$) and any coupling map that is native to a quantum device to construct the random U_A. Combining this with the sequence of single qubit R_z gates then makes the resulting benchmarking quantum circuit of Figure 1 instantly executable.

Quantum unitary evolution score (QUES)

Figure 1 implements $f_t(\mathfrak{H})\,|0^n\rangle$ on the system qubits, where $f_t(\mathfrak{H})$ is a matrix polynomial (see Supplementary Information for details), with approximation error in the operator norm upper bounded by $\left\|f_t(\mathfrak{H}) - e^{-it\mathfrak{H}}\right\|_2 \leq \epsilon$. Therefore in the absence of quantum errors, after applying the circuit to the input state $|0^{n+1}\rangle$, the probability $P_t(U_A) := \|f_t(\mathfrak{H})\,|0^n\rangle\|^2$ of measuring the top ancilla qubit with outcome 0 will be close to 1, indicating that the underlying Hamiltonian evolution is unitary.

From now on, we will primarily consider mQSVT circuits with a fixed set of phase factors $\{\varphi_i\}$ and hence fixed simulation time t. For notational simplicity, we will drop the t-dependence in quantities such as $P_t(U_A)$, unless specified otherwise.

On a real quantum device, the probability $P(U_A)$ should be replaced by $P_{\text{exp}}(U_A)$, which is the experimentally measured probability. We define the quantum unitary evolution score (QUES) by

$$\text{QUES}(n,d) := \mathbb{E}\left(P_{\text{exp}}(U_A)\right), \tag{1}$$

where the expectation is taken over the ensemble of random quantum circuit instances U_A. The deviation of QUES from 1 then measures the average performance of the quantum computer under a Hamiltonian simulation task.

There is no unique prescription for constructing random quantum circuits. To fix the choice of U_A, we employ here the model random quantum circuit construction used to analyze the concept of quantum volume in [29]. Here, given a number of qubits n, U_A is constructed to contain n layers, each consisting of a random permutation of the qubit labels followed by random two-qubit gates between the n qubits. Given this construction, the QUES in Eq. (1) will then depend only on n and d, and the overall depth of the circuit is approximately $2d$ times the circuit depth of U_A. Note that given the basic quantum gate set of a particular quantum device, alternative constructions of U_A using random choices of specific one- and two-qubit gates are possible.

Figure 2 shows the results of computing the QUES across 8 different IBM Q quantum devices [1], each having 5 qubits and one of three distinct coupling maps (panel b). When the number of qubits $n \leq 3$, the QUES on all devices is relatively high ($\gtrsim 0.7$) but it decreases sharply for $n \geq 4$. In contrast, the QUES decreases only relatively mildly as d increases. This is particularly noticeable for $n = 2$, which may indicate that the quantum

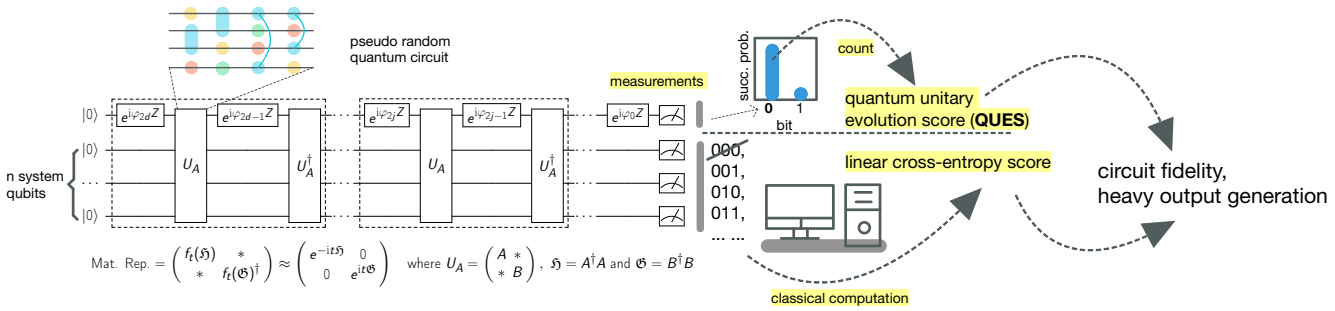


FIG. 1: Illustration of the minimal quantum singular value transformation (mQSVT) circuit for the Hamiltonian simulation benchmark. The overall circuit implements a complex matrix polynomial $f_t(\mathfrak{H})$ of degree d on the Hamiltonian $\mathfrak{H}$ that is defined in terms of a pseudo random quantum circuit U_A. The circuit acts on $n+1$ qubits, consisting of n system qubits and 1 ancilla qubit. After measuring the top ancilla qubit and post-selecting on the 0 outcome of this, the action on the bottom n system qubits accurately approximates $\exp(-it\mathfrak{H})\,|0^n\rangle$.

circuit transpiler provided by the IBM Q may be particularly effective for this device with very small qubit number. We emphasize that compared to generic benchmark measures such as the quantum volume, the QUES is specific to the computational task of the Hamiltonian simulation, and any information specific to this is not diluted by additional averaging over output distributions from other computational tasks. In particular, we find that even for quantum devices with relatively small quantum volume (8-QV), the performance in terms of QUES is only mildly worse than for those with a larger quantum volume (32-QV).

Circuit fidelity and system linear cross-entropy score (sXES)

The quality of a noisy implementation of a quantum circuit is often characterized by the circuit fidelity. Loosely speaking, the output quantum state of a noisy circuit can be characterized as a convex combination of the correct result and the result obtained under noise, i.e., "output" $= \alpha \times$ "correct result" $+ (1-\alpha) \times$ "noise", where $0 \leq \alpha \leq 1$ is the circuit fidelity. Let $p(U_A, x)$ be the noiseless bit-string probability of measuring the mQSVT circuit with outcome 0 in the ancilla qubit and an n-bit binary string x in the n system qubits. Let $p_{\text{exp}}(U_A, x)$ be the corresponding experimental bit-string probability, which can be estimated from the frequency of occurrence of the bit-string $0x$ in the measurement outcomes. Under the assumption that the noise model is a fully depolarized channel [7], i.e., the noise completely randomizes the computation, we have

$$p_{\text{exp}}(U_A, x) = \alpha p(U_A, x) + \frac{1-\alpha}{2^{n+1}}. \tag{2}$$

We now analyze the effect of noise on the circuit and show how measuring the QUES allows the circuit fidelity to be extracted. Prior work has made use of a combination of quantum and classical computation to obtain the circuit fidelity α. Such analysis relies on the possibility of evaluating the noiseless bit-string probability $p(U_A, x)$ classically, given U_A and x, e.g., via tensor network contraction [60]. This enabled the estimation of α from measurements of cross-entropy, referred to as XEB in this setting [7, 15]. We adapt this approach to the Hamiltonian simulation problem by defining a system linear cross-entropy score (sXES):

$$\text{sXES}(U_A) := \sum_{x \neq 0^n} p(U_A, x) p_{\text{exp}}(U_A, x). \tag{3}$$

The prefix "system" is added because the ancilla qubit is fixed to be the $|0\rangle$ state in the definition of $p(U_A, x), p_{\text{exp}}(U_A, x)$, and the $|x\rangle$ state belongs to the system register. In order to connect to the problem of generating heavy weight samples later, our definition of sXES excludes the bit-string 0^n. Taking the expectation with respect to the distribution of U_A, and rearranging Eq. (2) then gives an expression for the circuit fidelity:

$$\alpha = \frac{\mathbb{E}\left(\text{sXES}(U_A)\right) - \frac{1}{2^{n+1}}\mathbb{E}\left(\sum_{x\neq 0^n} p(U_A, x)\right)}{\mathbb{E}\left(\sum_{x\neq 0^n} p(U_A, x)^2\right) - \frac{1}{2^{n+1}}\mathbb{E}\left(\sum_{x\neq 0^n} p(U_A, x)\right)}. \tag{4}$$

This expression holds for any ensemble of random matrices, and relies only on the assumption that the noise model is depolarizing.

Once the probability distribution of U_A is specified (e.g., the Haar measure [51]), the only term in α that requires a quantum computation is $\text{sXES}(U_A)$, and all other terms in Eq. (4) can be evaluated classically. However, evaluation of the right-hand side of Eq. (4) often requires a significant amount of classical computation when n becomes large [15].

Inferring circuit fidelity from QUES

Based on the discussion so far, it might seem surprising that an alternative, very good approximation to the circuit fidelity can readily be obtained from the QUES metric in Eq. (1). This is arrived at by first defining $P_{\text{exp}}(U_A) = \sum_x p_{\text{exp}}(U_A, x)$, i.e., the average over all possible output bit strings x of the probability of measuring

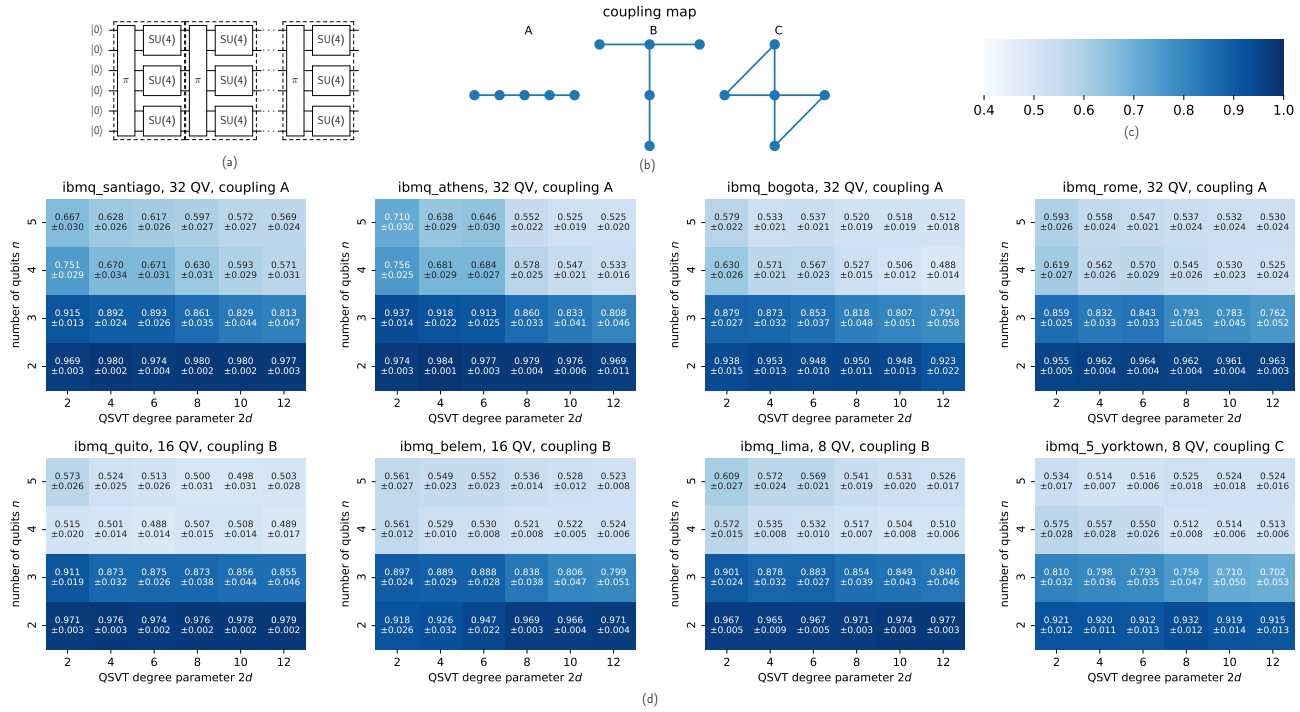


FIG. 2: Quantum unitary evolution score (QUES) of the 5-qubit quantum devices provided by the IBM Q platform [1]. (a) Visualization of the quantum circuit U_A used in computing QUES. When the number of qubits is n, there are n layers of the dashed boxes consists of the random permutation of the qubits labels followed by random two-qubit gates. After calling the transpiler, the circuit U_A is decomposed with respect to the basic gate set $\Gamma = \{\mathrm{Rz}, \sqrt{\mathrm{X}}, \mathrm{X}, \mathrm{CNOT}\}$ and the coupling map which indicates the available qubit pairs on which CNOT can act. (b) Layouts of coupling maps. (c) Color bar of the heatmap. (d) Each heatmap displays the benchmarking result of a specific quantum device, with the title showing the name of the device, its quantum volume, and its coupling map. Each QUES is estimated from 50 circuit instances. Each circuit instance is measured with $1,000$ measurement shots. The number displayed in each heatmap is the QUES value and its 95% confidence interval.

a given bit string as outcome of the action of U_A on the input state $|0^{n+1}\rangle$. Then summing both sides of Eq. (2) with respect to all bit strings x, further taking the expectation value of both sides over all possible U_A yields a fidelity estimate

$$\alpha_{\mathrm{QUES}} = 2 \times \mathrm{QUES} - 1. \tag{5}$$

The approximation error ϵ determines the extent of deviation of α_{QUES} from α. Specifically, we have the following bound (Appendix E)

$$|\alpha_{\mathrm{QUES}} - \alpha| \leq 16\epsilon + \mathcal{O}(\epsilon^2). \tag{6}$$

It is evident that, unlike Eq. (4), there is no classical overhead for evaluating α_{QUES} for any n. Since the circuit fidelity α should be non-negative, combining Eq. (5) and Eq. (6) also indicates that under the assumption of the depolarizing noise model, we have $\mathrm{QUES} \geq 0.5 - 8\epsilon + \mathcal{O}(\epsilon^2)$.

To numerically verify the relation between QUES and circuit fidelity, we make use of the digital error model of [15] in which each quantum gate in the circuit is subject to a depolarizing error channel with a certain error rate, leading to the desired depolarizing error channel on the overall system. We test the resulting noisy quantum circuit with different two-qubit gate error rates r_2 and set the one-qubit gate error rate to $r_1 = r_2/10$. We also discard the rotation gate with phase factor φ_{2d}, since this just adds a global phase to the exact Hamiltonian simulation. Then, given U_A with a total of g_1 one-qubit gates and g_2 two-qubit gates, the reference value of the circuit fidelity is $\alpha_{\mathrm{ref}} := (1-r_1)^{2d(g_1+1)}(1-r_2)^{2dg_2}$. We assume U_A is Haar-distributed (numerically verified in Appendix F) to simplify the computation of classical expectations.

Figure 3 summarizes the estimated circuit fidelity for random quantum circuits with different depth parameter d, variable coupling maps, and a range of error parameters. In all cases, we find that the derived circuit fidelity from QUES (α_{QUES}) and the circuit fidelity α obtained from sXES are consistent with each other, and both fidelities agree well with the theoretical reference value α_{ref} computed by direct evaluation of the fidelity of the mQSVT circuit (see Table S5 for numerical values of the fidelities). We also see that for a given set of error rates r_1, r_2, the circuits with highest connectivity show the best performance. This is because random circuits on these architectures converge faster to the Haar measure, which reduces the circuit depth (see Appendix F).

In the next two subsections we show how to assess and

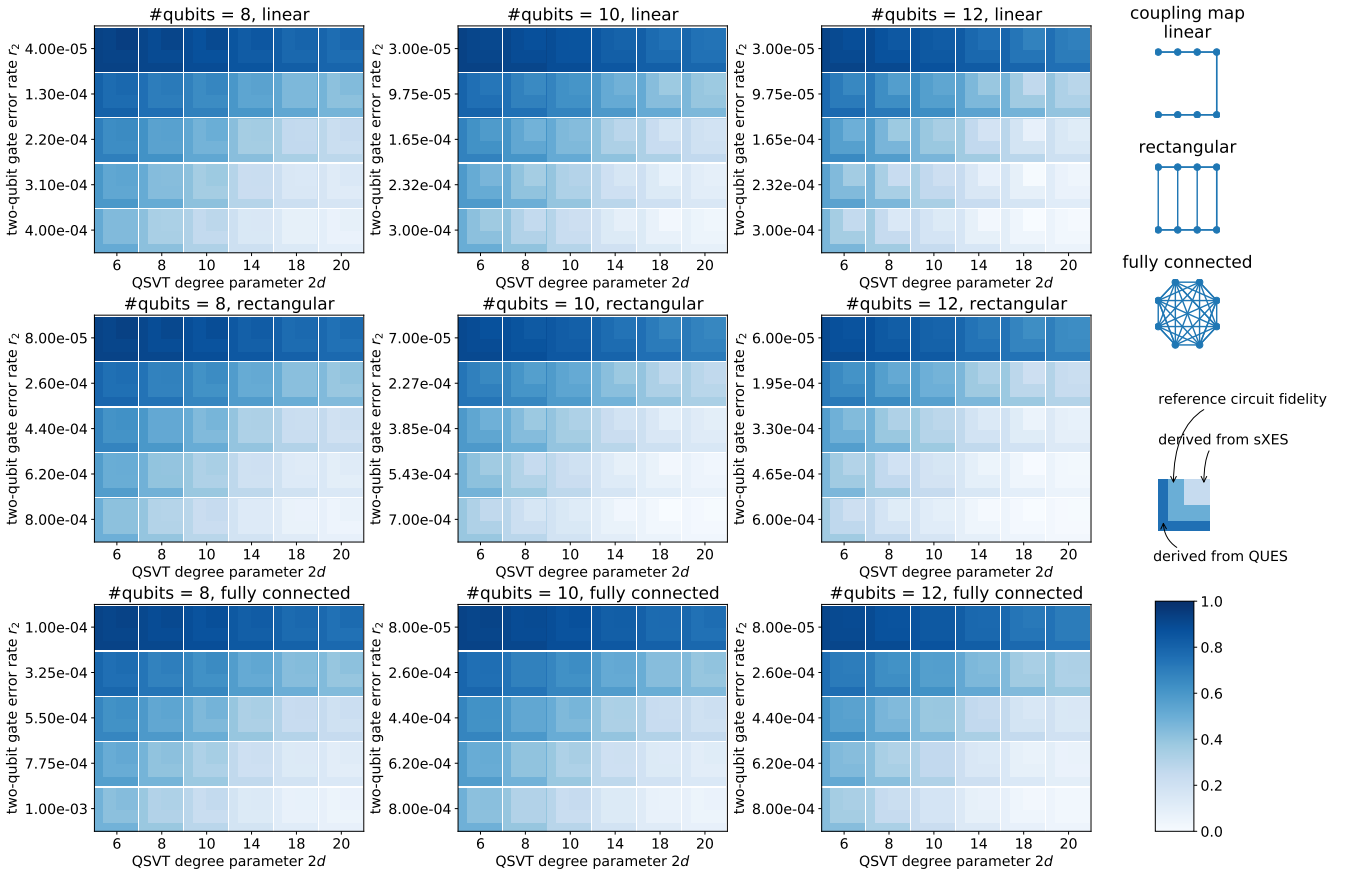


FIG. 3: Circuit fidelity estimated from the quantum Hamiltonian simulation benchmark. Colored grids represent the circuit fidelity estimated from ~ 100 circuit repetitions. The benchmarking is performed for circuits with a range of number of system qubits, having also variable types of couplings and a range of error parameters. The depth of the random circuit instances is set to the convergent depth deduced from the convergence to Haar measure (see Appendix F). The right column contains graphical depictions of the coupling maps, the layout of each grid, and the color bar.

evaluate the classical hardness of Hamiltonian simulation with the mQSVT circuit. We first define the analog of XHOG for Hamiltonian simulation, which we refer to as sXHOG, and give conditions for the hardness of this. We then show that classical hardness can be inferred directly from the value of the circuit fidelity obtained from the QUES, i.e. α_{QUES}.

Classical hardness and system linear cross-entropy heavy output generation (sXHOG)

The complexity-theoretic foundation of the Google claim of "quantum supremacy" in [7] is based on a computational task called linear cross-entropy heavy output generation (XHOG) with Haar-distributed unitaries [3, 4, 7, 15]. Specifically, given a number $b > 1$ and a random n-qubit unitary U, the task is to generate k nonzero bit-strings $x_1, x_2, \cdots, x_k \in \{0,1\}^n$ such that $\frac{1}{k}\sum_{j=1}^{k} p(U, x_j) \geq b \times 2^{-n}$, where $p(U,x) = |\langle x|U|0^n\rangle|^2$. For k randomly generated bit-strings, we expect that $\frac{1}{k}\sum_{j=1}^{k} p(U, x_j) \approx 2^{-n}$. Therefore any value $b > 1$ will correspond to a "heavy weight" output. When k is large enough, successful solution of the XHOG problem is considered to be classically hard for every value $b > 1$ [3, 4]. This holds for every circuit fidelity estimate $\alpha > 0$ obtained from the XEB metric, leading to the claim of supremacy in [7] based on extraction of a value $\alpha \approx 0.002$ from the experiments.

For the Hamiltonian simulation benchmark, we can define an analogous linear cross-entropy heavy output generation problem for the n system qubits. Note that the heavy weight samples are now defined only for the system qubits. We shall refer to this heavy output generation problem for Hamiltonian simulation the sXHOG problem, to emphasize this important feature and the difference from the standard XHOG problem. Specifically, given a number $b > 1$, a Hamiltonian simulation benchmark circuit with sufficiently small approximation error ϵ, and a random $(n+1)$-qubit unitary U_A defining a random Hamiltonian on the n qubits, the task is to generate k nonzero bit-strings $x_1, x_2, \cdots, x_k \in \{0,1\}^n \backslash \{0^n\}$ such that $\frac{1}{k}\sum_{j=1}^{k} p(U_A, x_j) \geq b \times 2^{-n}$. Now for the case of Hamiltonian simulation, $p(U_A, x) = \mathcal{O}(2^{-n})$ for any $x \neq 0^n$ at all t, but $p(U_A, 0^n)$ can be much larger

(for more details see Figure S5(b) in Appendix J). The state 0^n is then by definition "heavy" and we must therefore exclude this from the measure in order to avoid a trivial outcome. This is what distinguishes the sXHOG problem from the original XHOG problem. The classical hardness of the sXHOG problem is justified in Theorem 6 of Appendix H. In particular, if sXHOG can be solved efficiently on classical computers, then it would violate the system linear cross-entropy quantum threshold assumption (sXQUATH), which assumes that $p(U_A, x)$ for $x \neq 0^n$ cannot be efficiently estimated on classical computers to sufficient precision. The concept of sXQUATH directly parallelizes the linear cross-entropy quantum threshold assumption (XQUATH) in [4], with a similar restriction as above to exclude the output bit string 0^n (for more details see Appendix H).

Inferring classical hardness from QUES

In order to decide whether a noisy implementation of the Hamiltonian simulation benchmark is in the classically hard regime, we need to establish whether or not the sXHOG problem can be solved for $b > 1$.

Under the assumption that U_A is drawn from the Haar measure, and that the approximation error ϵ of the mQSVT circuit is sufficiently small, we derive the following relation between b and the circuit fidelity α:

$$b = 1 + \frac{\gamma(\alpha - \alpha^*)}{\alpha + 1}. \tag{7}$$

Here α^* is a fidelity threshold (not the complex conjugation of α) and γ a constant. Explicit expressions for the threshold value α^* and the constant γ are given in Appendix I. Both quantities are independent of the circuit fidelity α and depend only on the number of system qubits n and the simulation time t. Eq. (7) thus shows that when $\gamma > 0$ and $\alpha > \alpha^*$, we will have $b > 1$ so that the sXHOG problem solved by the mQSVT circuit is classically hard. This is qualitatively different from the situation for XEB experiments, for which every $\alpha > 0$ implies $b > 1$ [7].

Using the relation between QUES and α in Eqs. (5) and (6), and assuming that ϵ is sufficiently small, we immediately arrive at the conclusion that when

$$\text{QUES} \geq (1 + \alpha^*)/2, \quad \gamma > 0, \tag{8}$$

the corresponding sXHOG problem is classically hard for a sufficiently large value of n. This is a surprising result, since as noted above, the estimation of QUES does not require intensive classical computation. In fact it is not even necessary to actually generate any heavy weight samples - instead we just need to measure the value of QUES, Eq. (1), which is readily done by repeatedly running the circuit in Figure 1 with random circuit parameters as described above. Of course, should one wish to actually solve the sXHOG problem itself, the heavy weight samples would need to be generated using a quantum computer and intensive classical computation for computation of $\frac{1}{k}\sum_{j=1}^{k} p(U_A, x_j)$ would then also be required. But in order to demonstrate the regime of "quantum supremacy" for Hamiltonian simulation, i.e., the minimal values of n and d to reach this regime, this is not required.

To further investigate the implications of Eq. (7), we now explicitly indicate the time dependence of all quantities (i.e., we employ the notation $\gamma \to \gamma_t, \alpha^* \to \alpha_t^*$). In Figure 4 we plot the values of γ_t, α_t^* according to the expressions given in Appendix I as a function of the simulation time t, for $n = 4, 8, 12$ qubits. Figure 4 shows that $\gamma_t > 0$ for all t, so then we only need to determine whether it is possible to have fidelity $\alpha \geq \alpha_t^*$. It is evident from Figure 4 that both α_t^* (panel (a)) and γ_t (panel (b)) show oscillatory behavior. We now analyze this behavior to identify an optimal time at which the quantum supremacy of Hamiltonian simulation in this random Hamiltonian setting can be demonstrated for a sufficiently large number of qubits n.

For very short times, i.e., when $t \approx 0$, we have $\alpha_t^* > 1$. This means that we cannot have $b > 1$ for any value of the circuit fidelity $0 \leq \alpha \leq 1$. To see why this is the case, consider the limit $t = 0$. Here $p_t(U_A, 0^n) = 1$, and $p_t(U_A, x) = 0$ for any $x \neq 0^n$. By continuity, when t is very small, the magnitude of $p_t(U_A, x)$ for most bitstrings $x \neq 0^n$ is still very small and cannot reach the heavy output regime. Figure 4(a) also shows that there is a critical simulation time $t^{\text{thr}} \approx 2.26$, for which $\alpha_t^* < 1$ for any $t > t^{\text{thr}}$.

When $t > t^{\text{thr}}$, ideally we would like to have $\alpha_t^* \approx 0$, so that a very low experimental circuit fidelity α is sufficient to reach the heavy output regime. To this end we investigate what happens at the vanishing circuit fidelity, i.e., $\alpha = 0$. Detailed analysis shows that in the large n limit, we have $\gamma_t \alpha_t^* = \mathbb{E}\left(p_t(U_A, 0^n)\right)$, and Eq. (7) can be simplified as (see Appendix I)

$$b|_{\alpha=0} = 1 - \mathbb{E}\left(p_t(U_A, 0^n)\right), \tag{9}$$

where the expectation value is taken with respect to the random unitaries U_A as before. Thus when the expectation value is positive, i.e., $\mathbb{E}\left(p_t(U_A, 0^n)\right) > 0$, in the large n limit we again have $b < 1$ and there is no classical hardness. Moreover, since b is a continuous function of α, even if we now have finite circuit fidelity α, when this is small enough we can still find $b < 1$. This provides an alternative explanation of Eq. (7), namely, that the circuit fidelity α needs to be larger than the finite positive threshold value $\alpha_t^* > 0$ for *most* values of $t > t^{\text{thr}}$.

As a result of these considerations, in the supremacy-relevant setting when n is large enough, it is important to focus on the regimes where the expectation value $\mathbb{E}\left(p_t(U_A, 0^n)\right) \approx 0$, which from Eq. (7) implies that the threshold fidelity $\alpha_t^* \approx 0$. The numerical results shown in Figure 4 indicate that this can happen in two different scenarios. The first is when the simulation time $t \to \infty$ (see the analytic justification of this statement in Appendix K 3). Of course this requires a very large circuit

depth and is a physically "trivial" limit that is impractical on near-term quantum devices. The second scenario, which is much more relevant in practice, is when α_t^* reaches its first minimum, which defines an optimal time $t = t^{\rm opt}$. In the large n limit, the value of $t^{\rm opt}$ can be rationalized as the first node of the Bessel function $J_0(t/2)$ (see Appendix J). Figure 4 (a) shows that for $t^{\rm opt} \approx 4.81$, we already have $\mathbb{E}\left(p_t(U_A, 0^n)\right) \approx 0$ and $\alpha_t^* \approx 0$. Therefore simulating to the time $t = t^{\rm opt}$ is highly desirable, since this is a relatively short time at which the Hamiltonian simulation benchmark is nevertheless now guaranteed to solve the sXHOG problem even for a very small circuit fidelity. Our numerical results indicate that the values of t^* and $t^{\rm opt}$ depend only weakly on n, and their values are nearly converged for n as small as 12. Therefore this value of $t^{\rm opt}$ can be used in a future quantum simulation in the supremacy regime.

Discussion

We have presented a new quantum benchmark for Hamiltonian simulation on quantum computers. The Hamiltonian simulation problem is solved using a minimal quantum singular value transformation (mQSVT) circuit. The primary output of the Hamiltonian simulation benchmark is a single number called QUES, which can be verified without any classical computation, even in the quantum supremacy regime. Therefore the Hamiltonian simulation benchmark is scalable benchmark, and can be executed and verified on future quantum devices with a large number of qubits.

As the current quantum computing technologies advance, the possibility of implementing some error correction is improving [20]. Here the highly structured mQSVT circuit provides useful indications of where best to implement error correction under limited resources for this. Recall that the mQSVT circuit consists of a series of repetitions of a random circuit U_A and its conjugate $U_A^\dagger$, interleaved with single-qubit Z rotation operators characterized by carefully selected phase factors. Thus given a specific random Hamiltonian block encoded in U_A, the time dependent evolution operator for this Hamiltonian is defined entirely in terms of the phase angles for the single-qubit Z rotation operators. Since these phases should moreover be precisely determined, this suggests that on near-term quantum devices that may allow for some error correction but have overall limited resources, quantum error correction for these single-qubit rotations should be prioritized.

It is also useful to consider here the applicability of this Hamiltonian simulation approach to general Hamiltonians, i.e., not restricted to random Hamiltonians, on near-term quantum computers. Unfortunately it appears that for current quantum technologies there is potentially a large gap between the feasible simulation of a H-RACBEM given in this work and that of a general Hamiltonian relevant to e.g., molecular or solid-state physics. The main reason is that the block encoding of most Hamiltonians of practical interest will involve significant numbers of ancilla qubits, as well as multi-qubit control gates, all of which are extremely expensive on near-term quantum devices. In contrast to this general situation, the construction of H-RACBEM uses only whatever one-qubit and two-qubit gates are available for a given quantum device and is thus considerably easier. Nevertheless, it is possible that undertaking Hamiltonian simulation with H-RACBEM may also yield interesting physical applications to the various settings in which quantum chaotic dynamics are relevant. One immediate possibility in this direction is to use H-RACBEM to simulate the dynamics of quantum scrambling or quantum chaos in strongly interacting quantum systems. Scrambling dynamics can be studied by simulating put-of-time-order correlators (OTOCs) for effective Hamiltonians that can be defined implicitly in terms of a random circuit for time t (see e.g., [53]). We note that one can easily perform a Hamiltonian simulation backward in time, merely by reversing the sign of t, so the mQSVT circuit for an OTOC at any time t of a random Hamiltonian encoded in H-RACBEM can be readily constructed by adding local operators between forward and backward implementations of the mQSVT. Evaluation of the circuit at different times t can be implemented either by reevaluating the phase factors (which may required building a longer circuit depending on the accuracy required). The circuits can also be adapted to Hamiltonian simulation at finite temperatures and hence also to scrambling at finite temperatures. From a theoretical perspective it would also be useful to explore to what extent the structure of the H-RACBEM influences the speed of scrambling [17].

Our theoretical analysis of the sensitivity of the Hamiltonian benchmarking scheme in this work was based on a fully depolarized noise model, which is expected to be a good model for superconducting qubits [7]. However, in general the Pauli stochastic noise model on which is based may be biased or non-uniform across qubits. In addition, thermal noise and coherent errors are important for some qubit architectures. It will be useful to extend the current analysis to other noise models [43].

Finally, we note that while this Hamiltonian simulation benchmark is restricted to the specific class of random Hamiltonians, it might also provide information relevant to more general Hamiltonian simulations. Efforts to analyze the complexity of analog Hamiltonian simulations have often focused on the relation of such simulations to classical sampling tasks [2, 16, 37], and are closely related to the cross-entropy analysis for sampling of random quantum circuits of [7, 15]. As noted recently [37], this enables proof of the classical hardness of certain classes of analog quantum Hamiltonian simulation [8, 35]. It could be useful to explore generalizations of other classical sampling tasks to the QSVT setting, as was done here for the cross-entropy heavy output generation, to help guide the search for Hamiltonians whose simulation by QSVT would show an exponential gap to the corresponding classical simulation. Finally, the current

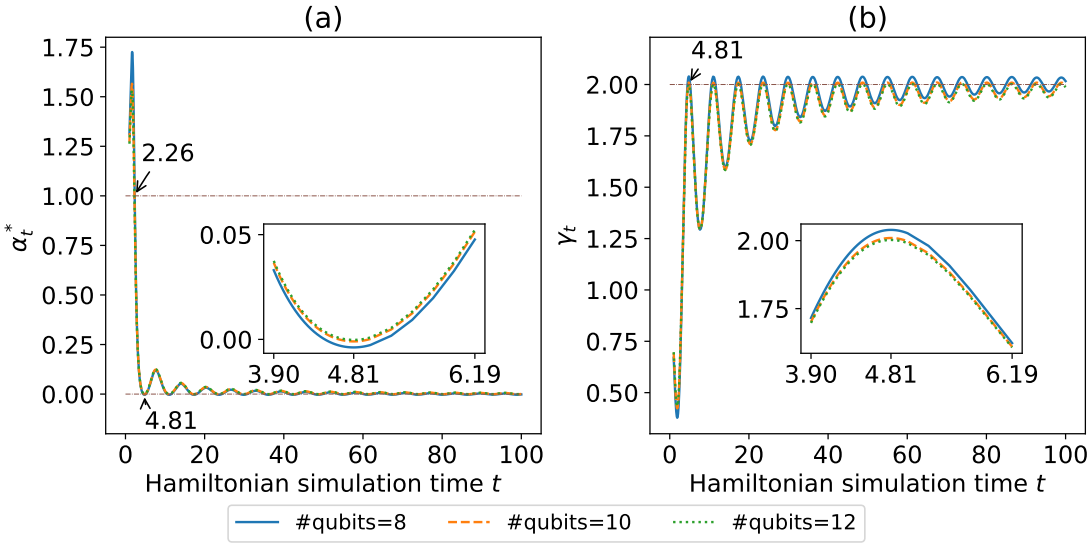


FIG. 4: Quantities relevant to the system linear cross-entropy heavy output generation (sXHOG) problem, evaluated using the explicit expressions given in Appendix I. (a) The threshold fidelity α_t^* as a function of Hamiltonian simulation time t. The upper value noted on the plot indicates the time value $t^{\text{thr}} \approx 2.26$ where $\alpha^*(t^{\text{thr}}) = 1$. The lower value noted on the plot indicates the regime at finite time $t^{\text{opt}} \approx 4.81$ with the first minimal value of threshold fidelity. (b) The parameter γ_t as a function of Hamiltonian simulation time t. The value noted on the plot indicates the value $\gamma_t \approx 2$ at the optimal time $t^{\text{opt}} \approx 4.81$. The averages in (a) and (b) are estimated numerically from ~ 100 instances of the mQSVT circuit encoding random Hamiltonians drawn from the Haar measure. Insets in each panel show the behavior of α_t^* and γ_t near the optimal time $t^{\text{opt}} \approx 4.81$.

approach of analysis of alternative fidelity measures under Hamiltonian simulation using mQSVT may provide useful for analysis of recent fidelity based experimental studies of analog Hamiltonian simulations that followed the emergent random nature of a projected ensemble of states [25].

Methods

All numerical tests are implemented in python3.7 and Qiskit [5]. Quantum circuits in Figure 2 are optimized by the transpiler provided by Qiskit before being executed on a real quantum device. The number of measurements (shots) is fixed to be $1,000$ for the experiments on real quantum devices in Figure 2, and it is set to $1,000,000$ for those on classical simulators in Figure 3. The classical generation of Haar random unitaries in Figure 4 is performed by QR factorization to random complex matrices with i.i.d. Gaussian entries according to the recipe in [52].

Data availability

The experimental data that support the finding are available from the authors upon request.

Acknowledgments

This work was partially supported by the Department of Energy under the Quantum Systems Accelerator Program and a Google Quantum Research Award (YD,BW,LL), by the Department of Energy under grant DE-SC0017867, and by the Department of Energy under the Center for Advanced Mathematics for Energy Research Applications (CAMERA) program (LL). We thank András Gilyén, Yunchao Liu, Murphy Niu and Jiahao Yao for helpful discussions.

Author contributions

YD and LL designed the Hamiltonian simulation benchmark and proved its theoretical properties. YD, BW and LL designed the experiments. YD carried out classical simulations and IBM-Q experiments. All authors contributed to the discussion of results and writing of the manuscript.

Competing interests

The authors declare no competing interests.

[1] URL https://quantum-computing.ibm.com.

[2] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 333–342, 2011.

[3] S. Aaronson and L. Chen. Complexity theoretic foundations of quantum supremacy experiments. *arXiv:1612.05903*, 2016.

[4] S. Aaronson and S. Gunn. On the Classical Hardness of Spoofing Linear Cross-Entropy Benchmarking. pages

1–7, 2019.

[5] H. Abraham et al. Qiskit: An open-source framework for quantum computing, 2019.

[6] D. An, D. Fang, and L. Lin. Time-dependent unbounded hamiltonian simulation with vector norm scaling. *Quantum*, 5:459, 2021.

[7] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[8] J. Bermejo-Vega, D. Hangleiter, M. Schwarz, R. Raussendorf, and J. Eisert. Architectures for quantum simulation showing a quantum speedup. *Physical Review X*, 8(2):021010, 2018.

[9] D. W. Berry and A. M. Childs. Black-box Hamiltonian simulation and unitary implementation. *Quantum Information & Computation*, 12(1-2):29–62, 2012.

[10] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Commun. Math. Phys.*, 270(2):359–371, 2007.

[11] D. W. Berry, R. Cleve, and S. Gharibian. Gate-efficient discrete simulations of continuous-time quantum query algorithms. *Quantum Information and Computation*, 14 (1-2):1–30, 2014.

[12] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. Somma. Simulating Hamiltonian dynamics with a truncated Taylor series. *Phys. Rev. Lett.*, 114:090502, 2015.

[13] D. W. Berry, A. M. Childs, and R. Kothari. Hamiltonian simulation with nearly optimal dependence on all parameters. *Proceedings of the 56th IEEE Symposium on Foundations of Computer Science*, pages 792–809, 2015.

[14] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.*, 8(1):1–13, 2017.

[15] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven. Characterizing quantum supremacy in near-term devices. *Nat. Phys.*, 14(6):595–600, 2018.

[16] M. J. Bremner, A. Montanaro, and D. J. Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.*, 117(8):080501, 2016.

[17] W. Brown and O. Fawzi. Scrambling speed of random quantum circuits. *arXiv preprint arXiv:1210.6644*, 2012.

[18] E. Campbell. Random compiler for fast Hamiltonian simulation. *Phys. Rev. Lett.*, 123(7):070503, 2019.

[19] C.-F. Chen, H.-Y. Huang, R. Kueng, and J. A. Tropp. Quantum simulation via randomized product formulas: Low gate complexity with accuracy guarantees. 2020.

[20] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, et al. Exponential suppression of bit or phase flip errors with repetitive error correction. *arXiv preprint arXiv:2102.06132*, 2021.

[21] A. M. Childs and Y. Su. Nearly optimal lattice simulation by product formulas. *Phys. Rev. Lett.*, 123(5):050503, 2019.

[22] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su. Toward the first quantum simulation with quantum speedup. *Proc. Nat. Acad. Sci.*, 115:9456–9461, 2018.

[23] A. M. Childs, A. Ostrander, and Y. Su. Faster quantum simulation by randomization. *Quantum*, 3:182, 2019.

[24] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu. A theory of trotter error. *To appear in Physical Review X*, 2020.

[25] J. Choi, A. L. Shaw, I. S. Madjarov, X. Xie, J. P. Covey, J. S. Cotler, D. K. Mark, H.-Y. Huang, A. Kale, H. Pichler, et al. Emergent randomness and benchmarking from many-body quantum chaos. *arXiv preprint arXiv:2103.03535*, 2021.

[26] B. Collins. *Intégrales matricielles et probabilités non-commutatives*. PhD thesis, 2003.

[27] B. Collins and P. Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264 (3):773–795, 2006.

[28] A. Cornelissen, J. Bausch, and A. Gilyén. Scalable benchmarks for gate-based quantum computers. *arXiv preprint arXiv:2104.10698*, 2021.

[29] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta. Validating quantum computers using randomized model circuits. *Physical Review A*, 100(3): 032328, 2019.

[30] Y. Dong and L. Lin. Random circuit block-encoded matrix and a proposal of quantum linpack benchmark. *Phys. Rev. A*, 103(6):062412, 2021.

[31] Y. Dong, X. Meng, K. B. Whaley, and L. Lin. Efficient phase factor evaluation in quantum signal processing. *Phys. Rev. A*, 103:042419, 2021.

[32] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43:5830–5847, 2002.

[33] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *Nat. Commun.*, 10(1):5347, 2019.

[34] R. P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys*, 21(6/7), 1982.

[35] X. Gao, S.-T. Wang, and L.-M. Duan. Quantum supremacy for simulating a translation-invariant ising spin model. *Physical review letters*, 118(4):040502, 2017.

[36] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 193–204, 2019.

[37] J. Haferkamp, D. Hangleiter, A. Bouland, B. Fefferman, J. Eisert, and J. Bermejo-Vega. Closing gaps of a quantum advantage with short-time hamiltonian dynamics. *Physical Review Letters*, 125(25):250501, 2020.

[38] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, 2009.

[39] A. Y. Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.

[40] H. S. Leff. Class of ensembles in the statistical theory of energy-level spectra. *Journal of Mathematical Physics*, 5:763–768, 1964.

[41] L. Lin and Y. Tong. Optimal quantum eigenstate filtering with application to solving quantum linear systems. *Quantum*, 4:361, 2020.

[42] L. Lin and Y. Tong. Near-optimal ground state prepara-

tion. *Quantum*, 4:372, 2020.

[43] Y. Liu, M. Otten, R. Bassirianjahromi, L. Jiang, and B. Fefferman. Benchmarking near-term quantum computers via random circuit sampling. pages 1–29, 2021.

[44] S. Lloyd. Universal quantum simulators. *Science*, pages 1073–1078, 1996.

[45] G. H. Low. Hamiltonian simulation with nearly optimal dependence on spectral norm. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 491–502, 2019.

[46] G. H. Low and I. L. Chuang. Optimal hamiltonian simulation by quantum signal processing. *Phys. Rev. Lett.*, 118:010501, 2017.

[47] G. H. Low and N. Wiebe. Hamiltonian simulation in the interaction picture. *arXiv:1805.00675*, 2019.

[48] E. Magesan, J. M. Gambetta, and J. Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106(18):180504, 2011.

[49] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang. A grand unification of quantum algorithms. *arXiv:2105.02859*, 2021.

[50] C. Mastrodonato and R. Tumulka. Elementary proof for asymptotics of large Haar-distributed unitary matrices. *Letters in Mathematical Physics*, 82(1):51–59, 2007.

[51] M. L. Mehta. *Random matrices*. Elsevier, 2004.

[52] F. Mezzadri. How to generate random matrices from the classical compact groups. *arXiv preprint math-ph/0609050*, 2006.

[53] X. Mi, P. Roushan, C. Quintana, S. Mandra, J. Marshall, C. Neill, F. Arute, K. Arya, J. Atalaya, R. Babbush, et al. Information scrambling in computationally complex quantum circuits. *arXiv:2101.08870*, 2021.

[54] D. Petz and J. Réffy. On asymptotics of large Haar distributed unitary matrices. *Periodica Mathematica Hungarica*, 49(1):103–117, 2004.

[55] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout. Measuring the capabilities of quantum computers. 2020.

[56] B. Şahinoğlu and R. D. Somma. Hamiltonian simulation in the low energy subspace. *arXiv:2006.02660*, 2020.

[57] A. Selberg. Berkninger om et multilet integral. *Norsk, Mat. Tidsskr.*, 26:71–78, 1944.

[58] Y. Su, D. W. Berry, N. Wiebe, N. Rubin, and R. Babbush. Fault-tolerant quantum simulations of chemistry in first quantization. *arXiv:2105.12767*, 2021.

[59] Y. Tong, D. An, N. Wiebe, and L. Lin. Fast inversion, preconditioned quantum linear system solvers, and fast evaluation of matrix functions. *arXiv:2008.13295*, 2020.

[60] B. Villalonga, D. Lyakh, S. Boixo, H. Neven, T. S. Humble, R. Biswas, E. G. Rieffel, A. Ho, and S. Mandrà. Establishing the quantum supremacy frontier with a 281 pflop/s simulation. *Quantum Sci. Tech.*, 5:034003, 2020.

[61] K. Zyczkowski and H.-J. Sommers. Truncations of random unitary matrices. *Journal of Physics A: Mathematical and General*, 33(10):2045, 2000.

Supplementary Information

Appendix A: Notations

We first introduce the definition of block encoding. Let $A \in \mathbb{C}^{N\times N}$ be an n-qubit Hermitian matrix ($N = 2^n$). If we can find an $(n+1)$-qubit unitary matrix U_A such that ($*$ stands for a matrix block whose entries are not of interest)

$$U_A = \begin{pmatrix} A & * \\ * & * \end{pmatrix} \tag{A1}$$

holds, i.e. A is the upper-left matrix block of U_A, then we may get access to the action of A on an n-qubit state $|\psi\rangle$ via the unitary matrix U_A by

$$U_A |0\rangle |\psi\rangle = |0\rangle (A|\psi\rangle) + |\perp\rangle ,$$

where $|\perp\rangle$ is an unnormalized $(n+1)$-qubit state not of interest and satisfies $(|0\rangle\langle 0| \otimes I_n)|\perp\rangle = 0$. Here we follow the row-major order convention. For instance,

$$|0\rangle |\psi\rangle \equiv \begin{pmatrix} \psi \\ 0^n \end{pmatrix}, \quad |1\rangle |\psi\rangle \equiv \begin{pmatrix} 0^n \\ \psi \end{pmatrix},$$

and Eq. (A1) can also be written as $A = (\langle 0| \otimes I_n) U_A (|0\rangle \otimes I_n)$.

Clearly when the operator norm $\|A\|_2$ is larger than 1, A cannot be encoded by any unitary U_A as in Eq. (A1). Generally if we can find $\alpha, \epsilon' \in \mathbb{R}_+$, and an $(m+n)$-qubit matrix U_A such that

$$\|A - \alpha (\langle 0^m| \otimes I_n) U_A (|0^m\rangle \otimes I_n)\|_2 \leq \epsilon', \tag{A2}$$

then U_A is called an (α, m, ϵ')-block-encoding of A. Here m is called the number of ancilla qubits for block encoding. We refer to [36] for more details on block encoding. When the block encoding is exact with $\epsilon' = 0$, U_A is called an (α, m)-block-encoding of A. The special case of the $(1,1)$-block-encoding may also be called a 1-block-encoding.

In the Supplementary Information, for notational simplicity, we may use U without a subscript to represent a $(n+1)$-qubit quantum circuit drawn at random from a certain probability distribution. Unless otherwise noted, A denotes the upper-left n-qubit submatrix of U, i.e. U is the 1-block-encoding of A. This matrix A is also called a random circuit block encoded matrix (RACBEM), and $\mathfrak{H} = A^\dagger A$ is corresponding Hermitian random circuit block encoded matrix (H-RACBEM) [30].

We use $N = 2^n$ to represent the dimension of the Hilbert space of the system qubits, and I_n to denote the n-qubit identity matrix. For a complex square matrix A with singular value decomposition (SVD) $A = W\Sigma V^\dagger$, its singular value transformation through an even function g is defined as $g^{\triangleright}(A) = V g(\Sigma) V^\dagger$. Here, the right triangle in the notation means only the right singular vectors V are kept in the transformation. If we consider $|A| := \sqrt{A^\dagger A} = V\Sigma V^\dagger$, then the singular value transformation of A is equal to the eigenvalue transformation of the Hermitian matrix $|A|$, namely, $g^{\triangleright}(A) = g(|A|)$. Furthermore, due to the even parity of g, there is a function f so that $g(x) = f(x^2)$ and $g^{\triangleright}(A) = f(|A|^2) = f(\mathfrak{H})$. In particular, when $g_t(x)$ is an even polynomial approximation to $s_t(x) = e^{-itx^2}$, we can define $g_t(x) = f_t(x^2)$. Hence $f_t(x)$ approximates e^{-itx}, and $g_t^{\triangleright}(A) = f_t(\mathfrak{H})$ approximates the Hamiltonian evolution $e^{-it\mathfrak{H}}$.

We use $\mathcal{U}_{f,U}$ to represent the minimal quantum singular value transformation (mQSVT) circuit in Figure 1, which has only a single ancilla qubit, $m = 1$. For any n-qubit input state $|\psi\rangle$, the mQSVT circuit performs the following transformation of the input quantum state,

$$\mathcal{U}_{f,U} |0\rangle \otimes |\psi\rangle = |0\rangle \otimes \left(g^{\triangleright}(A) |\psi\rangle\right) + |1\rangle \otimes |\perp\rangle ,$$

where $|\perp\rangle \in \mathbb{C}^N$ is an unnormalized quantum state. In other words, $\mathcal{U}_{f,U}$ is the 1-block-encoding of $f(\mathfrak{H}) \equiv g^{\triangleright}(A)$. $\|A\|_2 := \sigma_{\max}(A)$ is the operator norm of a matrix which is equal to its maximal singular value. $\|f\|_\infty := \max_{x\in[-1,1]} |f(x)|$ is the infinity norm of continuous functions on $[-1,1]$. $\mathbb{E}(\cdot)$ stands for the average over the random matrix ensemble (most commonly, the ensemble of U). Both $\overline{z}$ and z^* stands for the complex conjugate of a complex number z. For a complex polynomial $P(x) = \sum_i c_i x^i \in \mathbb{C}[x]$, its complex conjugate as $P^*(x) = \sum_i c_i^* x^i$. For a matrix A, the transpose, Hermitian conjugate and complex conjugate are denoted by $A^\top$, $A^\dagger$, A^*, respectively. Without otherwise noted, an n-bit binary string $x \in \{0,1\}^n$ is identified to its decimal representation. Specifically, when an

n-bit binary string appears in the subscript of a matrix or a vector, it is identified to be its decimal representation (we use a zero-based indexing). For example, $A_{0^n,1^n} := A_{0,2^n-1}$.

Table S1 summarizes the main notations used in the Supplementary Information. In the context of Hamiltonian simulation, many quantities depend on the value of the simulation time t. Such a t-dependence is usually added as a subscript such as $p_t(U,x)$. Most of the discussion focuses on the simulation at a fixed time t. Therefore when the context is clear, for simplicity we may drop the t dependence.

Symbol	Definition
$\mathcal{U}_{f,U}$	mQSVT circuit in Figure 1 implementing a 1-block-encoding of $f(\mathfrak{H})$
A	Upper-left n-qubit submatrix of a $(n+1)$-qubit random unitary matrix U
$\mathfrak{H}$	$A^\dagger A$, also called a H-RACBEM
$s_t(x)$	e^{-itx^2}
$g_t(x)$	an even polynomial approximation to $s_t(x)$, also denoted by $P(x,\Phi)$ with phase factor Φ
$f_t(x)$	$g_t(x^2)$, which is a polynomial approximation to e^{-itx}
$\mathbb{P}(\cdot)$	probability density function of random quantum circuits
$\mathbb{P}_{\exp}(\cdot)$	probability density function associated with the noisy implementation of random quantum circuits
p_j	probability associated with the matrix element at the 0-th column and the j-th row of a unitary matrix V, i.e., $p_j := \lvert V_{j0}\rvert^2$
p_{ij}	probability associated with the matrix element at the i-th column and the j-th row of a unitary matrix V, i.e., $p_j := \lvert V_{ij}\rvert^2$
$p(U,x)$	noiseless bit-string probability of measuring $\mathcal{U}_{f,U}$ with outcome 0 in the ancilla qubit and an n-bit binary string x in the n system qubits (dependence on f is omitted)
$P(U)$	noiseless probability of measuring $\mathcal{U}_{f,U}$ with outcome 0 in the ancilla qubit, satisfying $P(U)=\sum_x p(U,x)$ (dependence on f is omitted)
$p_{\exp}(U,x)$	bit-string probability of measuring the noisy implementation of $\mathcal{U}_{f,U}$ with outcome 0 in the ancilla qubit and an n-bit binary string x in the n system qubits (dependence on f is omitted)
$P_{\exp}(U)$	probability of measuring the noisy implementation of $\mathcal{U}_{f,U}$ with outcome 0 with 0 in the ancilla qubit, satisfying $P_{\exp}(U)=\sum_x p_{\exp}(U,x)$ (dependence on f is omitted)

TABLE S1: Summary of notations used in the Supplementary Information.

Appendix B: Equivalence between minimal and standard QSVT circuits

The standard implementation of the QSVT circuit [36] uses one extra ancilla qubit, called the signal ancilla qubit. In this section, we show that when the system matrix A is block encoded with only one ancilla qubit, the signal ancilla qubit is no longer needed. Therefore the only ancilla qubit is due to the block encoding of A, and the circuit is called the minimal QSVT (mQSVT) circuit in Figure 1. Furthermore, Figure 1 removes all two-qubit and multi-qubit gates outside of the unitary U_A, which greatly simplifies the implementation for a given quantum device. We prove the equivalence between the mQSVT and the standard QSVT circuits in this section for completeness.

For any $(n+1)$-qubit unitary U, let the singular value decomposition of its upper-left n-qubit submatrix A be $A = W_1 \Sigma V_1^\dagger$. Following the cosine-sine decomposition (CSD), there exists n-qubit unitaries W_2, V_2 so that U can be decomposed as follows,

$$U = \begin{pmatrix} A & * \\ * & B \end{pmatrix} = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \begin{pmatrix} \Sigma & S \\ -S & \Sigma \end{pmatrix} \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}^\dagger,$$

where $S = \sqrt{I - \Sigma^2}$. This decomposition also implies any n-qubit non-unitary matrix A, up to a scaling factor, can in principle be block encoded using only one ancilla qubit.

Then, the unitary matrix representation of the quantum circuit in Figure 1 is

$$\text{Mat. Rep.} = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} \begin{pmatrix} e^{i\varphi_0} I_n & 0 \\ 0 & e^{-i\varphi_0} I_n \end{pmatrix} \prod_{j=1}^{d} \left[\begin{pmatrix} \Sigma & -S \\ S & \Sigma \end{pmatrix} \begin{pmatrix} e^{i\varphi_{2j-1}} I_n & 0 \\ 0 & e^{-i\varphi_{2j-1}} I_n \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} \Sigma & S \\ -S & \Sigma \end{pmatrix} \begin{pmatrix} e^{i\varphi_{2j}} I_n & 0 \\ 0 & e^{-i\varphi_{2j}} I_n \end{pmatrix} \right] \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}^\dagger.$$

Let K be the permutation matrix permuting the j-th and the $(N+j)$-th columns, and $V = \mathrm{diag}\,\{V_1,\ V_2\}$. The multiplicand is simplified as a direct sum of N 2-by-2 blocks upon conjugating $\widetilde{V} := VK$, i.e.

$$\widetilde{V}^\dagger\,(\text{Mat. Rep.})\,\widetilde{V} = \bigoplus_{q=0}^{N-1} e^{i\varphi_0 Z} \prod_{j=1}^{d} R_q e^{i\varphi_{2j-1}Z} R_q^\top e^{i\varphi_{2j}Z},$$

where

$$R_q = \begin{pmatrix} \sigma_q & -\sqrt{1-\sigma_q^2} \\ \sqrt{1-\sigma_q^2} & \sigma_q \end{pmatrix} = e^{i\frac{\pi}{4}Z} e^{i\arccos(\sigma_q)X} e^{-i\frac{\pi}{4}Z}.$$

Let $W(x) := e^{i\arccos(x)X}$, and

$$\widetilde{\varphi}_i = \begin{cases} \varphi_i + \frac{\pi}{4}, & i = 0 \text{ or } 2d, \\ \varphi_i + \frac{\pi}{2}, & i = 2, 4, \cdots, 2d-2, \\ \varphi_i - \frac{\pi}{2}, & i = 1, 3, \cdots, 2d-1. \end{cases}$$

The matrix representation is then

$$\widetilde{V}^\dagger\,(\text{Mat. Rep.})\,\widetilde{V} = \bigoplus_{q=0}^{N-1} e^{i\widetilde{\varphi}_0 Z} \prod_{j=1}^{2d} W(\sigma_q) e^{i\widetilde{\varphi}_j Z}.$$

It is straightforward to show that the following mapping from $[-1,1]$ to SU(2)

$$x \mapsto e^{i\widetilde{\varphi}_0 Z} \prod_{j=1}^{2d} W(x) e^{i\widetilde{\varphi}_j Z} = \begin{pmatrix} P(x) & i\sqrt{1-x^2}Q(x) \\ i\sqrt{1-x^2}Q^*(x) & P^*(x) \end{pmatrix}$$

defines an even polynomial $P(x)$ of degree at most $2d$, and an odd polynomial $Q(x)$ of degree at most $2d-1$, so that $|P(x)|^2 + (1-x^2)\,|Q(x)|^2 = 1$ holds for any $x \in [-1,1]$.

Then, the matrix representation of the quantum circuit is

$$\text{Mat. Rep.} = V \begin{pmatrix} P(\Sigma) & i\sqrt{I_n - \Sigma^2}Q(\Sigma) \\ i\sqrt{I_n - \Sigma^2}Q^*(\Sigma) & P^*(\Sigma) \end{pmatrix} V^\dagger = \begin{pmatrix} P^{\triangleright}(A) & * \\ * & (P^{\triangleright}(A))^\dagger \end{pmatrix}.$$

For example, when $g_t(x)$ is an even polynomial approximation to $s_t(x) = e^{-itx^2}$, we can define $g_t(x) = f_t(x^2)$, and the diagonal n-qubit submatrices are $g_t^{\triangleright}(A) = f_t(A^\dagger A)$ and $(g_t^{\triangleright}(A))^\dagger = (f_t(A^\dagger A))^\dagger$ respectively.

This remarkably simple structure of the QSVT circuit is due to the use of 1-block-encoding. In general, if an n-qubit matrix A is block encoded in an $(n+m)$-qubit unitary U_A, the standard QSVT circuit has the structure in Figure S1. In particular, even when $m = 1$, two CNOT gates are needed to implement each phase rotation. This introduces additional errors and can be practically cumbersome on near term devices the default two-qubit gate is not CNOT (e.g. $\sqrt{\text{iSWAP}}$).

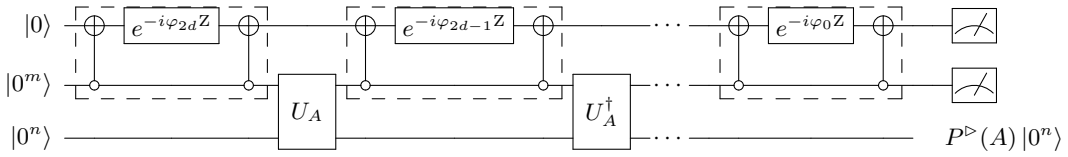


FIG. S1: Quantum circuit for quantum singular value transformation (QSVT) of an even complex matrix polynomial P of degree $2d$. The dashed boxes represent the controlled rotation with phase factor φ_j where two m-qubit Toffoli gates controlled at $|0^m\rangle$ are used. The QSVT circuit queries the $(n+m)$-qubit quantum circuit U_A and its inverse recursively for d times. For the given QSVT circuit, by measuring ancilla qubits with outcome 00^m, the action on the system qubits is the matrix polynomial.

In the mQSVT circuit in Figure 1, the phase factors $(\varphi_0, \cdots, \varphi_{2d})$ are determined by an optimization procedure that provides an even polynomial $g_t(x)$ satisfying $\|g_t(x) - s_t(x)\|_\infty \leq \epsilon$ for some given precision parameter ϵ (see Appendix C), and the evolution time t is encoded in the choice of phase factors. We then measure the top ancilla qubit and post-select on the 0 outcome of this measurement. This then ensures that the action on the lower n system qubits approximates the Hamiltonian evolution $e^{-it\mathfrak{H}}|0^n\rangle \approx f_t(\mathfrak{H})|0^n\rangle$, where $\mathfrak{H} = A^\dagger A$. Here $f_t(\mathfrak{H})$ is a matrix polynomial, and the approximation error in the operator norm is upper bounded by $\left\|f_t(\mathfrak{H}) - e^{-it\mathfrak{H}}\right\|_2 \leq \epsilon$.

In the absence of quantum errors the probability of measuring the top ancilla qubit with outcome 0, i.e. the $P_t(U) := \|f_t(\mathfrak{H})|\psi\rangle\|^2$, will be close to 1. Specifically, by the triangle inequality, the probability of measuring the top ancilla qubit with outcome 0 is lower bounded:

$$\begin{aligned} P_t(U) = \left\|g_t(\Sigma)V^\dagger|0^n\rangle\right\|_2^2 = \sum_{j=0}^{N-1} |g_t(\sigma_j)|^2 p_j = \left|1 + \sum_{j=0}^{N-1} \left((g_t(\sigma_j) - s_t(\sigma_j))\,\overline{g_t(\sigma_j)} + s_t(\sigma_j)\overline{(g_t(\sigma_j) - s_t(\sigma_j))}\right) p_j\right| \\ \geq 1 - 2\epsilon \sum_{j=0}^{N-1} p_j = 1 - 2\epsilon, \end{aligned} \tag{B1}$$

where $p_j = |V_{0,j}|^2$.

We also find that the probability $p_t(U,x) = |\langle 0x|\mathcal{U}_{f_t,U}|00^n\rangle|^2 \approx |\langle x|\exp(-it\mathfrak{H})|0^n\rangle|^2$ will characterize the dynamics of the propagation from 0^n to x for any n-bit string $x \in \{0,1\}^n$. If the simulation time t is short, then $\exp(-it\mathfrak{H}) \approx I$, and $p_t(U,0^n)$ can be much larger than $p_t(U,x)$ for any bitstring $x \neq 0^n$. This issue will be particularly important when defining the "heavy weight samples" in later discussions. Therefore we shall primarily focus on the case when $x \neq 0^n$.

As an illustrative example, Figure S2 shows a quantum circuit implementing a 2-qubit matrix A encoded by a 3-qubit unitary matrix U. The construction uses only the basic gate set $\{\mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3, \mathrm{CNOT}\}$. In Appendix C we describe an optimization based method to obtain the phase factors for a relatively short time t to a small approximation error ϵ. In this example we set $t = 1$. To obtain a theoretical error bound at large t, we can use the phase factor concatenation technique in Appendix K 2 to obtain the simulation at t from 2 to 10, and the error bound $\epsilon_t = \epsilon t^2$ is given by Theorem 14. Using these circuits, we may measure the outcome of the system qubits in the computational basis, and follow the dynamics of the probability $\sum_{x\neq 0^n} p_t(U,x)$, i.e. that of the quantum state moving away from the initial state $|0^n\rangle$. Figure S2(b) shows that this agrees very well with the result using the exact dynamics $\sum_{x\neq 0^n} |\langle x|e^{-it\mathfrak{H}}|0^n\rangle|^2$. Furthermore, according to Eq. (B1), the probability $P_t(U_A)$ in Figure S2(c) satisfies the theoretical bounds $1 - 2\epsilon_t \leq P_t(U) \leq 1$ and is very close to 1.

Appendix C: Optimization based method for finding phase factors in the Hamiltonian simulation benchmark

In order to implement the Hamiltonian simulation benchmark at time t, we need to find the phase factors Φ that generates an even polynomial $P(x,\Phi) = g_t(x)$ so that $\|g_t(x) - s_t(x)\|_\infty \leq \epsilon$ for a sufficiently small ϵ. For a large class of polynomials, the existence of such phase factors is established in [36, Theorem 4], and summarized in Theorem 1.

Theorem 1 (**Quantum signal processing in SU(2)**). *For any $P, Q \in \mathbb{C}[x]$ and a positive integer d such that (1) $\deg(P) \leq d, \deg(Q) \leq d-1$, (2) P has parity $(d \mod 2)$ and Q has parity $(d-1 \mod 2)$, (3) $|P(x)|^2+(1-x^2)|Q(x)|^2 = 1, \forall x \in [-1,1]$. Then, there exists a set of phase factors $\Phi := (\phi_0, \cdots, \phi_d) \in \mathbb{R}^{d+1}$ such that*

$$U(x,\Phi) = e^{i\phi_0 Z}\prod_{j=1}^{d}\left[W(x)e^{i\phi_j Z}\right] = \begin{pmatrix} P(x) & iQ(x)\sqrt{1-x^2} \\ iQ^*(x)\sqrt{1-x^2} & P^*(x) \end{pmatrix} \tag{C1}$$

where

$$W(x) = e^{i\arccos(x)X} = \begin{pmatrix} x & i\sqrt{1-x^2} \\ i\sqrt{1-x^2} & x \end{pmatrix}.$$

The phase factors in the theorem and those used in the quantum circuit in the main text is related by the following relation

$$\varphi_i = \begin{cases} \phi_0 + \frac{\pi}{4}, & i = 0, \\ \phi_i + \frac{\pi}{2}, & 1 \leq i \leq d-1, \\ \phi_d + \frac{\pi}{4}, & i = d. \end{cases} \tag{C2}$$

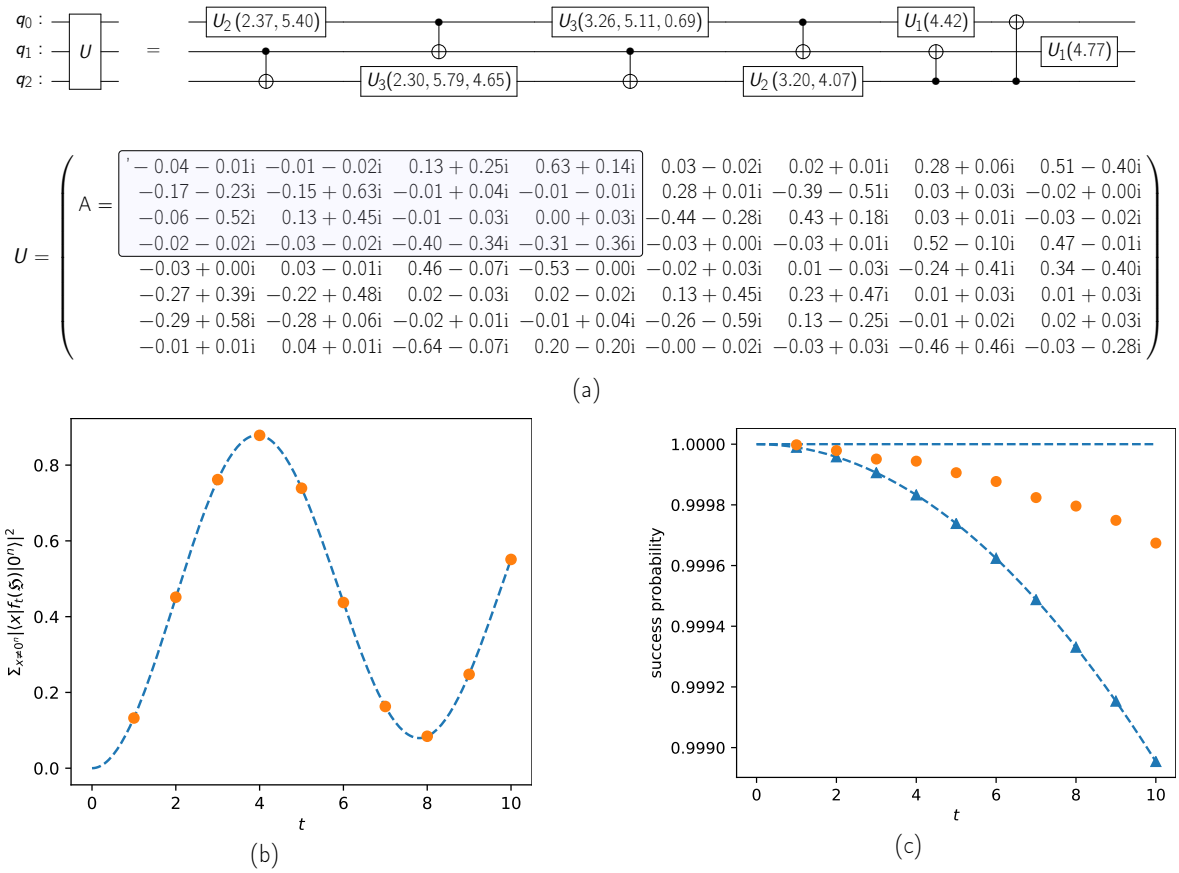


FIG. S2: An illustrative implementation from the quantum Hamiltonian simulation benchmark. (a) Top: A 3-qubit random quantum circuit constructed from the basic gate set $\{\mathrm{U}_1, \mathrm{U}_2, \mathrm{U}_3, \mathrm{CNOT}\}$. Bottom: The 3-qubit unitary matrix representation U of the quantum circuit and its upper-left 2-qubit submatrix A (in the shaded area). The top qubit q_0 is the ancilla qubit for block encoding. (b) Dynamics of the evolution away from the initial condition $\sum_{x\neq 0^n} p_t(U, x)$ implemented using the mQSVT circuit, compared to the reference solution $\sum_{x\neq 0^n} |\langle x|e^{-it\tilde{S}}|0^n\rangle|^2$. (c) The probability $P_t(U)$ obtained from 10^6 noiseless measurements (orange dots) and theoretical bounds (blue dashed).

In order to find the phase factors, the standard practice follows a two-step procedure. We first identify the approximate polynomial $P(x)$. Then the phase factors for $P(x)$ are computed following a recursive relation described in [36, Theorem 4]. In the case of the Hamiltonian simulation benchmark, it is highly nontrivial to find an approximate polynomial $P(x)$ satisfying the conditions in Theorem 1 while approximating the function e^{-itx^2} sufficiently well. Therefore we cannot follow the standard procedure to evaluate the phase factors.

On the other hand, the recently developed optimization based method [31] provides an alternative route to streamline this process. Instead of following a two-step procedure, the optimization based method allows one to find both the approximate polynomial and the phase factor sequence in a single step. Note that the optimization procedure in [31] only addresses the case when the target function is real. Here the target function $s_t(x)$ is complex. Below we present a modified optimization procedure to find the phase factor sequence for complex polynomials.

Specifically, given an arbitrary set of phase factors $\Phi \in \mathbb{R}^{d+1}$, Theorem 1 defines a mapping $\mathbb{R}^{d+1} \to \mathbb{C}[x]$ giving a complex polynomial of degree at most d via $P(x, \Phi) := \langle 0|U(x, \Phi)|0\rangle$. Note that given the parity constraint, the number of (complex) degrees of freedom is $\widetilde{d}$ where $\widetilde{d} := \lceil \frac{d+1}{2} \rceil$. Therefore, to fix the polynomial, one needs to sample $\widetilde{d}$ points given the polynomial is complex valued. In practice, to ensure numerical stability, we sample the function on $x_k = \cos\left(\frac{2k-1}{4\widetilde{d}}\pi\right), k = 1, \cdots, \widetilde{d}$, which are the positive Chebyshev nodes of $T_{2\widetilde{d}}(x)$. The optimization based method view P as a nonlinear approximation ansatz. Define the objective function as

$$F(\Phi) := \frac{1}{\widetilde{d}} \sum_{k=1}^{\widetilde{d}} |P(x_k, \Phi) - s_t(x_k)|^2 . \tag{C3}$$

Taking the (2π)-periodicity into account, the optimization problem is

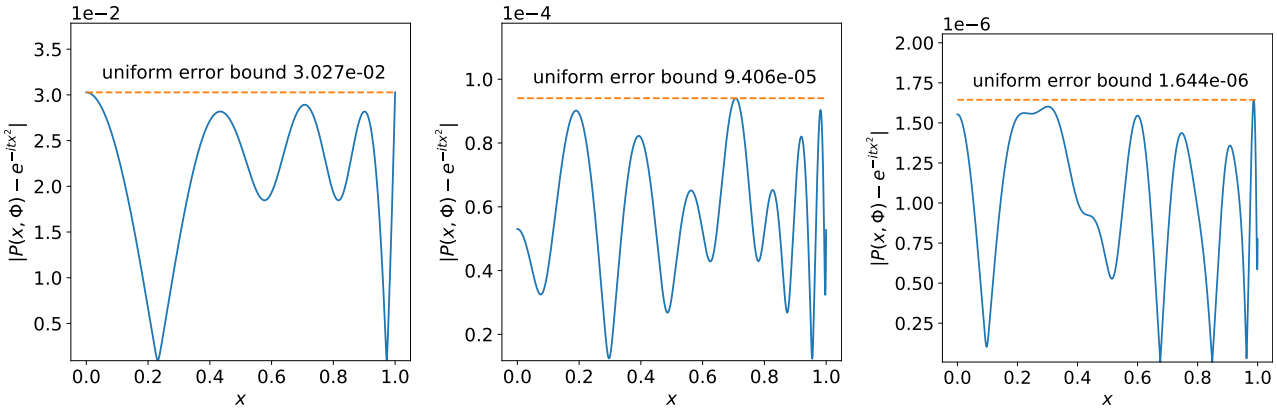
$$\Phi^* = \underset{\Phi\in[-\pi,\pi)^{d+1}}{\operatorname{argmin}} \; F(\Phi). \tag{C4}$$

The optimization problem is numerically solved by a quasi-Newton method. Table S2 describes the approximation error for polynomials measured by $\|P(x,\Phi^*) - s_t(x)\|_\infty$ at simulation time $t = 1$, for polynomial degrees between 6 and 20. When the polynomial degree is 20, the approximation error is as small as 10^{-8}, which demonstrates the effectiveness of the optimization based method.

degree d	approximation error
6	5.543×10^{-03}
8	5.805×10^{-04}
10	5.230×10^{-06}
14	3.332×10^{-06}
18	9.535×10^{-08}
20	1.107×10^{-08}

TABLE S2: Approximation error $\left\|P(x,\Phi^*) - e^{-itx^2}\right\|_\infty$ at time $t = 1$ with different polynomial degrees d.

According to Figure 4, there exists an "optimal" simulation time $t^{\text{opt}} = 4.8096$, for which the threshold fidelity $\alpha^*(t^{\text{opt}}) \approx 0$ (the derivation is in Appendix J). Table S3 describes the phase factor sequences that can be directly used in Figure 1 to perform Hamiltonian simulation at time t^{opt}. In order to reach low (3.0×10^{-2}), medium (9.4×10^{-5}), and high (1.6×10^{-6}) accuracy, the degrees of the polynomial found by the optimization procedure are $10, 18, 26$, respectively. Figure S3 further shows the pointwise approximate error on the interval $[0, 1]$ (the error on $[-1, 0]$ is the same due to the even parity). Compared to Table S2, in order to reach precision $\epsilon = 3.3 \times 10^{-6}$ at simulation time $t = 1$, the polynomial degree still needs to be 14. So even though t^{opt} is nearly 5 times larger, the polynomial degree only increases by less than twofold to reach similar accuracy. Since t^{opt} is still relatively small, this does not violate the "no-fast-forwarding" theorem of Hamiltonian simulation [10].



(a) (b) (c)

FIG. S3: Point-wise approximation error of phase factors for Hamiltonian simulation at time $t^{\text{opt}} = 4.8096$. The corresponding sets of phase factors are listed in Table S3.

Appendix D: Structure of the probability space of measuring noisy random quantum circuits and sXES

There are two sources of randomness when measuring noisy random quantum circuits. The first is due to the random choice of U with probability density $\mathbb{P}(U)$. The second is due to the Monte Carlo nature of the quantum measurement. Specifically, given the choice of U and a noisy implementation of the quantum circuit, the probability to obtain $0x$ as the measurement outcome is $p_{\text{exp}}(U, x) := \langle 0x|\varrho_{\text{exp}}|0x\rangle$. The joint probability of U and the measurement outcome $0x$ is

$$\mathbb{P}_{\text{exp}}(U, x) = p_{\text{exp}}(U, x)\mathbb{P}(U).$$

approximation error $\left\|P(x,\Phi)-e^{-itx^2}\right\|_\infty = 3.027\times10^{-2}$						
φ_0	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
-2.7731963	2.7942520	-1.5707963	2.5930970	-1.5707963	-0.6434012	-1.5707963
φ_7	φ_8	φ_9	φ_{10}			
2.5930970	-1.5707963	2.7942520	-2.7731963			

approximation error $\left\|P(x,\Phi)-e^{-itx^2}\right\|_\infty = 9.406\times10^{-5}$						
φ_0	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
-2.7731963	2.8229351	-1.5707963	-2.5716144	-1.5707963	-3.1056796	-1.5707963
φ_7	φ_8	φ_9	φ_{10}	φ_{11}	φ_{12}	φ_{13}
-1.1677625	1.5707963	-0.6437954	1.5707963	-1.1677625	-1.5707963	-3.1056796
φ_{14}	φ_{15}	φ_{16}	φ_{17}	φ_{18}		
-1.5707963	-2.5716144	-1.5707963	2.8229351	-2.7731963		

approximation error $\left\|P(x,\Phi)-e^{-itx^2}\right\|_\infty = 1.644\times10^{-6}$						
φ_0	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
-1.5893341	-0.3207550	2.8668325	-2.9662972	-1.1921175	-0.4528806	1.5270366
φ_7	φ_8	φ_9	φ_{10}	φ_{11}	φ_{12}	φ_{13}
1.6658052	-0.2379487	-2.9130657	0.3245889	0.7863552	-1.3306612	-0.2863103
φ_{14}	φ_{15}	φ_{16}	φ_{17}	φ_{18}	φ_{19}	φ_{20}
-1.3306612	0.7863552	0.3245889	-2.9130657	-0.2379487	1.6658052	1.5270366
φ_{21}	φ_{22}	φ_{23}	φ_{24}	φ_{25}	φ_{26}	
-0.4528806	-1.1921175	-2.9662972	2.8668325	-0.3207550	-1.5893341	

TABLE S3: Phase factors for Hamiltonian simulation at time $t^{\rm opt} = 4.8096$. The table lists three sets of phase factors with different approximation errors.

Here for simplicity, we only focus on the measurement result whose ancilla qubit is measured with outcome 0. When $\mathbb{P}(U)$ is given by the Haar measure, the noise channel is depolarized, and the noisy (experimental) bit-string probability is

$$p_{\rm exp}(U,x) = \alpha p(U,x) + \frac{1-\alpha}{2N}, \tag{D1}$$

which is the convex combination of the exact bit-string probability and the uniform distribution [15].

The information of the circuit fidelity can then be encoded in the experimental average of various quantities. For example, the bit-string probability for nonzero bit-strings is given by

$$\mathbb{E}_{\rm exp}\left(p(U,x); x\neq 0^n\right) = \mathbb{E}\left(\sum_{x\neq 0^n} p(U,x)p_{\rm exp}(U,x)\right) = \sum_{x\neq 0^n}\alpha\mathbb{E}\left(p(U,x)^2\right) + \frac{1-\alpha}{2N}\mathbb{E}\left(p(U,x)\right). \tag{D2}$$

Eq. (D2) connects quantities evaluated from quantum experiments and classical computation on the left hand side and those from classical computations on the right hand side. The left-hand side is given by the system linear cross-entropy score (sXES) in Eq. (3), which can be evaluated from multiplying the bit-string frequency $p_{\rm exp}(U,x)$ obtained from the quantum experiment and the bit-string probability $p(U,x)$ computed classically. The quantities on the right-hand side can be evaluated fully classically. The circuit fidelity α is then the only unknown and can be solved for by substituting the quantum experimental and classically computed quantities into Eq. (4) of the main text.

Appendix E: Estimating circuit fidelity from quantum unitary evolution score

The experimental average of the probability of measuring the ancilla qubit with outcome 0 is

$$\mathbb{E}_{\rm exp}\left(P(U)\right)=\mathbb{E}\left(\sum_x P(U)p_{\rm exp}(U,x)\right)=\mathbb{E}\left(P(U)P_{\rm exp}(U)\right)=\alpha\mathbb{E}\left(P(U)^2\right)+\frac{1-\alpha}{2}\mathbb{E}\left(P(U)\right). \tag{E1}$$

Here $P(U):=\sum_x p(U,x)$, and $P_{\rm exp}(U)$ is the probability which can be approximately determined by the bit frequency of the measurement outcome in the experiment. Rearranging the terms in Eq. (E1), the circuit fidelity can be alternatively estimated via

$$\alpha=\frac{\mathbb{E}\left(P(U)P_{\rm exp}(U)\right)-\frac{1}{2}\mathbb{E}\left(P(U)\right)}{\mathbb{E}\left(P(U)^2\right)-\frac{1}{2}\mathbb{E}\left(P(U)\right)}. \tag{E2}$$

From Eq. (B1), we have $P(U)\in[1-2\epsilon,1]$. Hence a lower and upper bound on the fidelity follows:

$$\underline{\alpha}:=\frac{2\left(1-2\epsilon\right)\mathbb{E}\left(P_{\rm exp}(U)\right)-1}{1+2\epsilon}\leq\alpha\leq\frac{2\mathbb{E}\left(P_{\rm exp}(U)\right)-(1-2\epsilon)}{1-8\epsilon}=:\overline{\alpha}. \tag{E3}$$

The difference between the upper and lower bound is

$$\overline{\alpha}-\underline{\alpha}\leq 16\epsilon+\mathcal{O}(\epsilon^2). \tag{E4}$$

Therefore, $\lim_{\epsilon\to 0}(\overline{\alpha}-\underline{\alpha})=0$ and the derived bounds are tight. Let us choose the form of the estimate as ϵ-independent

$$\alpha_{\rm QUES}:=2\mathbb{E}\left(P_{\rm exp}(U)\right)-1\in[\underline{\alpha},\overline{\alpha}]. \tag{E5}$$

Then, $|\alpha_{\rm QUES}-\alpha|\leq\overline{\alpha}-\underline{\alpha}\leq 16\epsilon+\mathcal{O}(\epsilon^2)$. Furthermore, the estimate can be determined using only the experimentally measurable quantity $P_{\rm exp}(U)$ and is independent of the classical computation of $P(U)$, which may be hard to evaluate for large n. This remarkable fact, namely the evaluation of circuit fidelity without any classical computation, arises from the approximate implementation of Hamiltonian simulation of the overall circuit. Since only one ancilla qubit is measured, the QUES defined in Eq. (E5) cannot entirely capture whether the circuit is implemented correctly. However, when the assumption that the noise channel is depolarized and when the polynomial approximation to $s_t(x)$ is sufficiently accurate, $\alpha_{\rm QUES}$ provides a very good estimate to the circuit fidelity.

Appendix F: Algorithm for constructing random quantum circuits and numerical convergence to Haar measure

In order to theoretically analyze the circuit fidelity, we need the additional assumption that $\mathbb{P}(U)$ is the Haar measure. This has the advantage that several terms in Eq. (4) can be evaluate analytically. Using the Haar measure, the statistics of an ensemble of random Hamiltonians is much simplified and can be computed by the statistics of the truncation matrix of Haar unitaries [26, 27, 50, 54, 61]. Details of the statistics are given in Appendix K 1. Furthermore, if U is Haar-distributed, then the noise effect of directly sampling U is well captured by a fully depolarized error channel, due to the nearly maximal entanglement in the output state [7, 15]. The need to choose an appropriate circuit depth ℓ such that the circuit statistics approximate those of Haar unitaries motivates an investigation of the statistics of Haar random quantum circuits of finite number of qubits.

We first construct random quantum circuits by using the algorithm given in Appendix F. It follows a similar recipe in Ref. [30]. We set the basic one-qubit gates to U1, U2 and U3 gates. Up to a global phase factor, the U3 gate is

$$\mathrm{U}_3(\theta,\phi,\lambda)=R_z(\phi+3\pi)R_x(\pi/2)R_z(\theta+\pi)R_x(\pi/2)R_z(\lambda),$$

which is a generic single-qubit operation parameterized by three Euler angles. The U1 and U2 gates are defined by restricting to one or two Z-rotation angles respectively, i.e.

$$\mathrm{U}_1(\lambda)=R_z(\lambda),\quad \mathrm{U}_2(\phi,\lambda)=R_z(\phi+\pi/2)R_x(\pi/2)R_z(\lambda-\pi/2).$$

Taken together with the CNOT gate, these form a continuously parameterized gate set that is universal.

Although we specify the choice of one-qubit gates and the use of the CNOT gate here, Appendix F can be directly generalized to an arbitrary basic gate set. The random quantum circuit generated by the algorithm respects the

ALGORITHM 1: Constructing random quantum circuits

Input: Coupling map $G = \langle V, E\rangle$ where V is the set of n qubits, E is the set of qubit pairs on which CNOT is available, basic gates $\Gamma = \{\mathrm{U1}, \mathrm{U2}, \mathrm{U3}, \mathrm{CNOT}\}$, the number of total one-qubit gates g_1, and the density of one-qubit gates $p_1 \in (0, 1)$.

Set the number of two-qubit gates to $g_2 = \lceil \frac{1-p_1}{2p_1} g_1 \rceil$.
Set the maximal number of two-qubit gates in each layer to $y_2 = \lceil \frac{1-p_1}{2} n \rceil$.
Set $m_1 = m_2 = 0$, initialize an empty quantum circuit $\mathcal{C}$.
while $m_1 \le g_1$ **and** $m_2 \le g_2$ **do**
  Draw $x_2 \le y_2$ pairs of qubits from E so that each pair (q_1, q_2) and its permutation (q_2, q_1) are not selected in the previous layer. The choice of x_2 also satisfies $m_2 + x_2 \le g_2$.
  Draw $x_1 = \min\{n - 2x_2, g_1 - m_1\}$ one-qubit gates uniformly at random from $\Gamma\backslash\{\mathrm{CNOT}\}$ and act them on the rest of qubits in this layer.
  Update the numbers of one- and two-qubit gates, $m_1 \leftarrow m_1 + x_1$ and $m_2 \leftarrow m_2 + x_2$.
end while
if $m_1 < g_1$ **then**
  Append layers of random $g_1 - m_1$ one-qubit gates sampled uniformly at random from $\Gamma\backslash\{\mathrm{CNOT}\}$.
else if $m_2 < g_2$ **then**
  Append layers of $g_2 - m_2$ CNOT gates acting on random operands.
end if
Return: A random quantum circuit $\mathcal{C}$ with g_1 one-qubit gates and g_2 two-qubit gates.

architecture of a quantum computer. In practice, we set the density of one-qubit gates to $p_1 = 0.5$. Then, for an n-qubit random quantum circuit with ℓ layers, the number of one-qubit gates is $g_1 = \frac{\ell n}{2}$ and that of two-qubit gates is $g_2 = \frac{\ell n}{4}$.

To measure the numerical convergence of random circuits to the Haar measure, we first summarize some of the statistical properties of the Haar measure. Given an n-qubit Haar-distributed unitary U, we denote $p_{ij} := |U_{ij}|^2$. As a special case of the more general Theorem 7 (to be presented in Appendix K 1), the p_{ij}'s are identically Beta-distributed.

Theorem 2. *The probability density of p_{ij} is* $\mathrm{Beta}(1, N-1)$*,*

$$\mathbb{P}(p_{ij}) = (N-1)(1-p_{ij})^{N-2} \mathbb{1}_{0 \le p_{ij} \le 1}.$$

Proof. Let the submatrix of interest be the upper left 1-by-1 block, namely, a single matrix element $a := U_{00}$. Note that $p_{00} := |a|^2$. Then, Eq. (K2) indicates that the probability density of a is

$$\mathbb{P}(a) \propto (1-p_{00})^{N-2} \mathbb{1}_{0 \le p_{00} \le 1}.$$

The polar decomposition of the complex number $a = re^{i\theta}$ yields the Jacobian $\mathrm{d}a = r\mathrm{d}r\mathrm{d}\theta \propto \mathrm{d}p_{00}\mathrm{d}\theta$. Then, integrating with respect to $\mathrm{d}\theta$, the marginal distribution of p_{00} is

$$\mathbb{P}(p_{00}) = (N-1)(1-p_{00})^{N-2} \mathbb{1}_{0 \le p_{00} \le 1}.$$

This is the $\mathrm{Beta}(1, N-1)$ distribution. When $i \neq 0$ or $j \neq 0$, let K_1 be the matrix permuting the i-th row and the 0-th row by left multiplication, and let K_2 be the matrix permuting the j-th column and the 0-th column by right multiplication. Then, $\widetilde{U} := K_1 U K_2$ is Haar distributed by the bi-invariance of the Haar measure. Furthermore, $\widetilde{U}_{00} = U_{ij}$. Therefore, the previous proof shows that p_{ij} is also $\mathrm{Beta}(1, N-1)$ distributed. □

Note that in the limit $N \gg 1$, the distribution of p_{ij} is well approximated by the exponential distribution $\mathrm{Exp}(N)$, a.k.a. the Porter-Thomas distribution derived in [15]. The statistics follows straightforward computation by integrating with respect to the probability density.

Theorem 3. *Let $M_k := \sum_{i=0}^{N-1} p_{ij}^k$ be the k-th moment, $S := \sum_{i=0}^{N-1} -p_{ij} \ln(p_{ij})$ be the entropy. Their averages with respect to the Haar distribution take the form*

$$M_k^{\mathrm{Haar}} := \mathbb{E}(M_k) = \prod_{i=1}^{k-1} \frac{1+i}{N+i}, \quad S^{\mathrm{Haar}} := \mathbb{E}(S) = \sum_{i=2}^{N} i^{-1}.$$

The variance of the k-th moment $V_k^{\mathrm{Haar}} := \sum_{i=0}^{N-1} \mathrm{Var}(p_{ij}^k)$ is

$$V_k^{\mathrm{Haar}} = \left(\frac{1}{N}\binom{2k}{k} + \frac{N-1}{N}\right) \prod_{i=k}^{2k-1} \frac{N-k+i}{N+i} - 1.$$

We remark that in the limit $N \gg 1$, $M_k^{\text{Haar}} \approx \frac{k!}{N^{k-1}}$ and $S^{\text{Haar}} \approx \ln(N) + \gamma - 1$ where γ is Euler's constant. The asymptotic results are the same as those derived in [15]. The variance is asymptotically $V_k^{\text{Haar}} \approx \frac{1}{N}\left(\binom{2k}{k} - 1\right)$. From the variance, we conclude two important features about the statistics. Given N, the variance (i.e. fluctuation) increases with respect to the order of the moment. For each moment, the statistics becomes concentrated as N increases, namely the variance V_k^{Haar} vanishes as $N \to \infty$. By Taylor expansion, the entropy has the same concentrated behavior which can be numerically observed in Figure S4.

Figure S4 presents the statistics of random quantum circuits for several different structures of the circuit coupling map and shows that for all three coupling maps studied in the main text, the distribution of circuits of sufficient depth converges to the Haar measure. In Figure S4(a), we plot the normalized entropy S/S^{Haar}. The figure shows that the entropy converges to that of Haar measure, i.e., $S/S^{\text{Haar}} \to 1$ after the circuit depth of U increases beyond specific values that depend only weakly on the number of qubits n. Since the random quantum circuit is constructed by combining layers of random one- and two-qubit gates, we test the convergence of random quantum circuits for different coupling maps, thereby varying the qubit pairs on which the two-qubit gates can act. The numerical results in Figure S4(a) show that coupling maps with greater connectivity converge significantly faster to the Haar measure. We attribute this to the larger number of possible allocations of two-qubit gates enhancing state entanglement within the system and thereby leading to faster mixing of information.

In addition to showing the convergence in terms of circuit entropy, we also quantify the convergence to the Haar measure for the first five moments in Figure S4(b). The minimal depth to achieve approximate Haar random circuits deduced from the convergence in moments is highly consistent with that derived from the convergence in entropy. We list the depth used in the computation of the sXES in Table S4.

coupling map	n (number of system qubits)		
	7	9	11
linear	140	160	160
rectangular	76	94	100
fully connected	60	60	60

TABLE S4: Depth for random quantum circuits used in the computation of the system linear cross-entropy score. Each depth is chosen so that both entropy and moments are close to these derived from the Haar measure.

Appendix G: Circuit fidelity from the system linear cross-entropy score

The system linear cross-entropy score is based on Hamiltonian simulation. Note that in the circuit fidelity of Eq. (4), only the terms involved the system linear cross entropy sXES in the numerator contain quantities that must be experimentally evaluated. All other terms can be simplified by using the statistical property of an ensemble of random matrices inherited from the Haar measure of U in Appendix K 1 (in particular Theorem 13).

The procedure of computing the system linear cross-entropy score can be summarized as follows.

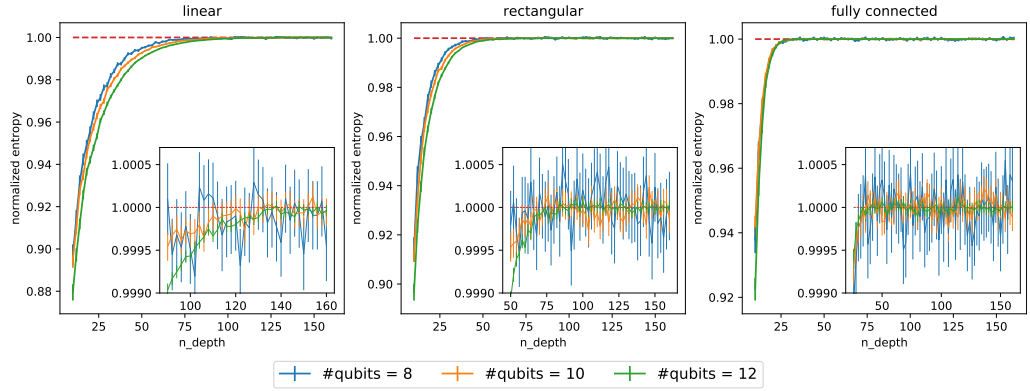
1. Draw quantum circuits U_i's approximately from the Haar measure at random.

2. For each U_i, build the mQSVT circuit for the Hamiltonian simulation benchmark, and measure all qubits to count the bit-string frequencies of $0x$ where $x \in \{0,1\}^n$. The bit-string frequency is an estimate to $p_{\text{exp}}(U_i, x)$. Furthermore, the sum of bit-string frequencies for all x's is an estimate to $P_{\text{exp}}(U_i) = \sum_{x\in\{0,1\}^n} p_{\text{exp}}(U_i, x)$.

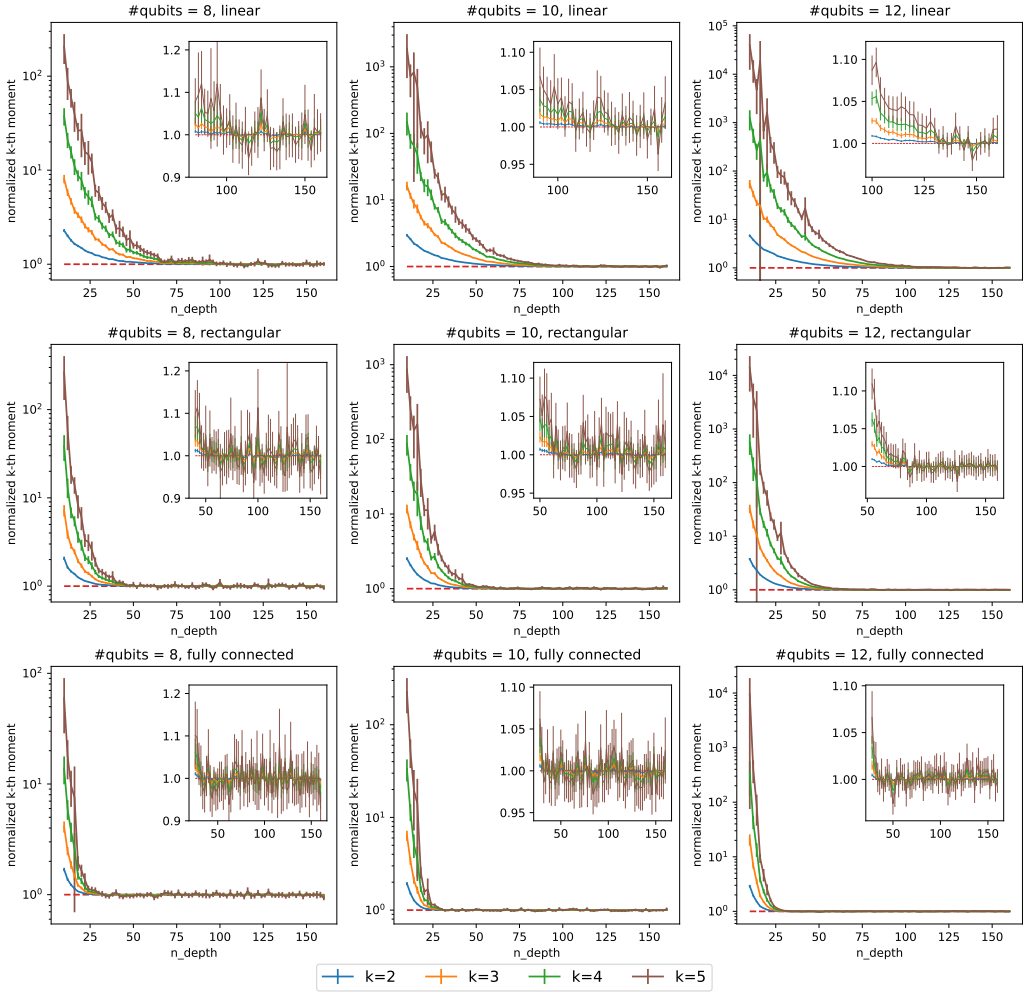
3. For each U_i and bit-string $0x$, compute the noiseless bit-string probability $p(U_i, x)$ on classical computers.

4. Compute estimates of the fidelity according to Eqs. (4) and (5).

We list the circuit fidelity estimated by different methods in Table S5. The agreement shows the consistency of the quantum Hamiltonian simulation benchmark. Here, the theoretical reference value is estimated from the depolarization noise model. Given U with a total of g_1 one-qubit gates and g_2 two-qubit gates, the value $\alpha_{\text{ref}} := (1-r_1)^{2d(g_1+1)}(1-r_2)^{2dg_2}$ follows approximately assuming each quantum error fully mixes the quantum state.



(a)



(b)

FIG. S4: Convergence to the Haar measure. (a) Convergence in terms of entropy. Normalized entropy S/S^{Haar} as a function of the depth for random quantum circuits with different number of system qubits and coupling map. (b) Convergence in terms of moments. Each panel is the first five normalized moments M_k/M_k^{Haar} as a function of the depth for random quantum circuits with different number of system qubits and coupling map. The convergence of curves to the dashed line at 1 shows that the random quantum circuit with a modest circuit depth can well approximate the Haar measure. Error bars correspond to the 95% confidence interval estimated from ~ 1000 circuit instances.

two-qubit gate error rate r_2	QSVT degree parameter $2d$					
	6	8	10	14	18	20
4.00×10^{-5}	0.95	0.92	0.90	0.85	0.80	0.79
	0.92	0.90	0.87	0.83	0.79	0.76
	0.93	0.92	0.89	0.86	0.82	0.80
1.30×10^{-4}	0.78	0.71	0.67	0.55	0.45	0.44
	0.77	0.70	0.65	0.54	0.46	0.42
	0.81	0.76	0.69	0.61	0.53	0.48
2.20×10^{-4}	0.65	0.56	0.50	0.36	0.26	0.25
	0.64	0.55	0.48	0.36	0.26	0.23
	0.70	0.62	0.54	0.43	0.34	0.28
3.10×10^{-4}	0.54	0.43	0.38	0.24	0.15	0.15
	0.54	0.43	0.35	0.23	0.15	0.12
	0.60	0.51	0.41	0.30	0.22	0.17
4.00×10^{-4}	0.45	0.34	0.28	0.16	0.087	0.090
	0.45	0.34	0.26	0.15	0.089	0.068
	0.52	0.42	0.32	0.21	0.14	0.10

TABLE S5: Circuit fidelity estimated from quantum Hamiltonian simulation benchmark. The total number of qubits is 8, namely, there are 7 system qubits and 1 ancilla qubit. The coupling map is linear. In each cell of the table, the top data is estimated from sXES, the middle data is the theoretical reference value, and the bottom data is estimated from QUES.

Appendix H: Classical hardness of sXHOG

Definition 4 (sXHOG, or System Linear Cross-entropy Heavy Output Generation)**.** *Given as input a number $b > 1$, a random $(n+1)$-qubit unitary U, and the mQSVT circuit for the Hamiltonian simulation benchmark with sufficiently small approximation error ϵ, output nonzero bit-strings $x_1, x_2, \cdots, x_k \in \{0,1\}^n \backslash \{0^n\}$ so that*

$$\frac{1}{k}\sum_{j=1}^{k} p(U, x) \geq b \times 2^{-n}. \tag{H1}$$

The classical hardness of the XEB experiment is justified by reducing the XHOG problem to a complexity assumption referred to as Linear Cross-entropy Quantum Threshold Assumption (XQUATH)[4]. Similarly, the hardness of the sXHOG problem can be reduced to an assumption that we refer to by analogy as sXQUATH.

Definition 5 (sXQUATH, or System Linear Cross-entropy Quantum Threshold Assumption)**.** *Given a random $(n+1)$-qubit unitary U, and the mQSVT circuit for the Hamiltonian simulation benchmark with sufficiently small approximation error ϵ, for a uniformly random $x \in \{0,1\}^n \backslash \{0^n\}$, there is no polynomial-time classical algorithm that produces an estimate p of $p_x := p(U, x)$ so that*

$$\mathbb{E}\left((p_x - p)^2\right) = \mathbb{E}\left((p_x - 2^{-n})^2\right) - \Omega(2^{-3n}).$$

Here, the expectation is taken over random circuits U, the internal randomness of the algorithm, and the random bit-string x.

The reduction of the XHOG problem is given in the following theorem, which is directly parallel to that in [4, Theorem 1].

Theorem 6 (Classical hardness of sXHOG)**.** *Assuming sXQUATH, no polynomial-time classical algorithm can solve the XHOG problem in Definition 4 with probability $s > \frac{1}{2} + \frac{1}{2b}$, and*

$$k \geq \frac{1}{((2s-1)b-1)(b-1)}.$$

Proof. Suppose that A is a classical algorithm solving sXHOG in Definition 4 with a success probability s as stated in the theorem. Given U and the mQSVT circuit as the input of A, it outputs $\mathsf{S} := \{x_i \neq 0^n : i = 1, \cdots, k\}$. When A successfully solves the sXHOG problem, the set S satisfies Eq. (H1). Specifically, let $x \in \{0,1\}^n \backslash \{0^n\}$ be a bit-string

sampled uniformly at random. We now construct an algorithm to produce an estimate p of $p(U,x)$. Given such a bit-string x, the algorithm outputs an estimate $p = b2^{-n}$ if $x \in \mathsf{S}$ and $p = 2^{-n}$ if $x \notin \mathsf{S}$.

Consider a random variable

$$X(U,x) := \left(p(U,x) - 2^{-n}\right)^2 - \left(p(U,x) - p\right)^2 = \left(2p(U,x) - (p + 2^{-n})\right)\left(p - 2^{-n}\right).$$

Here, the randomness comes from the uniformly random bit-string x, the random unitary U and its corresponding mQSVT circuit, and whether the classical algorithm A succeeds. We write them explicitly as the subscript of the expectation. Furthermore, we denote by $\mathsf{S}_U^{(s)} := \mathsf{S}$ when A succeeds, and $\mathsf{S}_U^{(f)} := \mathsf{S}$ when A fails. Let $\mathbb{1}_E$ be the indicator function which gives 1 if the condition E is satisfied and gives 0 otherwise. According to the algorithm,

$$\begin{aligned}
\mathbb{E}_{x,U}\left(X(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}|\mathsf{A}\text{ succeeded}\right) =& 2\cdot 2^{-n}(b-1)\cdot\mathbb{E}_{x,U}\left(p(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\middle|\mathsf{A}\text{ succeeded}\right)\\
&+2^{-2n}(1-b^2)\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\right),\\
\mathbb{E}_{x,U}\left(X(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\middle|\mathsf{A}\text{ failed}\right) =& 2\cdot 2^{-n}(b-1)\cdot\mathbb{E}_{x,U}\left(p(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\middle|\mathsf{A}\text{ failed}\right)\\
&+2^{-2n}(1-b^2)\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\right)\\
\geq& 2^{-2n}(1-b^2)\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\right).
\end{aligned} \tag{H2}$$

Furthermore, $X(U,x)\equiv 0$ if $x\notin\mathsf{S}$, regardless of whether A succeeds or not. By the law of total expectation,

$$\begin{aligned}
&\mathbb{E}_{x,U,\mathsf{A}}\left(X(U,x)\right) = s\cdot\mathbb{E}_{x,U}\left(X(U,x)|\mathsf{A}\text{ succeeded}\right) + (1-s)\cdot\mathbb{E}_{x,U}\left(X(U,x)|\mathsf{A}\text{ failed}\right)\\
=& s\cdot\mathbb{E}_{x,U}\left(X(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\middle|\mathsf{A}\text{ succeeded}\right) + (1-s)\cdot\mathbb{E}_{x,U}\left(X(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\middle|\mathsf{A}\text{ failed}\right)\\
\geq& s\cdot\left(2\cdot 2^{-n}(b-1)\mathbb{E}_{x,U}\left(p(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}|\mathsf{A}\text{ succeeded}\right) + 2^{-2n}(1-b^2)\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\right)\right)\\
&+(1-s)\cdot 2^{-2n}(1-b^2)\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\right).
\end{aligned} \tag{H3}$$

Here

$$\mathbb{E}_{x,U}\left(p(U,x)\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}|\mathsf{A}\text{ succeeded}\right) = \frac{k}{2^n-1}\mathbb{E}_U\left(\frac{1}{k}\sum_{x\in\mathsf{S}_U^{(s)}}p(U,x)\right) \geq \frac{bk2^{-n}}{2^n-1}.$$

Note that $\mathsf{S}_U^{(s)}$ and $\mathsf{S}_U^{(f)}$ are sets of k distinct bit-strings. Following that x is uniformly distributed, we have

$$\mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\right) = \mathbb{E}_U\left(\mathbb{E}_x\left(\mathbb{1}_{x\in\mathsf{S}_U^{(s)}}\right)\right) = \frac{k}{2^n-1} \text{ and } \mathbb{E}_{x,U}\left(\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\right) = \mathbb{E}_U\left(\mathbb{E}_x\left(\mathbb{1}_{x\in\mathsf{S}_U^{(f)}}\right)\right) = \frac{k}{2^n-1}.$$

Then

$$\begin{aligned}
\mathbb{E}_{x,U,\mathsf{A}}\left(X(U,x)\right) &\geq \frac{2^{-2n}}{2^n-1}\left(ks\cdot(b-1)^2 + k(1-s)\cdot(1-b^2)\right)\\
&\geq 2^{-3n}k\left((2s-1)b-1\right)(b-1) = \Omega(2^{-3n})
\end{aligned} \tag{H4}$$

when $k \geq \frac{1}{((2s-1)b-1)(b-1)}$. This violates sXQUATH and thereby proves the classical hardness of sHOG. □

Appendix I: Circuit fidelity and sXHOG

In this section we demonstrate that the success of sXHOG can be verified by experimental evaluation of the circuit fidelity. Due to the relation between the circuit fidelity and QUES in Appendix G, it means that the success of sXHOG can be verified by QUES, which does not involve any classical computation.

First, since we are only interested in the measurement outcome whose ancilla qubit returns 0, we normalize the bit-string probability as a conditional probability

$$p_{\text{exp}}(U,x|\text{ancilla}=0) := p_{\text{exp}}(U,x)/P_{\text{exp}}(U). \tag{I1}$$

The physical interpretation of the normalization of the bit-string probability is to discard the measurement result whose ancilla is measured with 1. We also remark that when the Hamiltonian simulation benchmark circuit is sufficiently accurate, we have $P(U) \approx 1$, and it is not necessary to normalize the noiseless bit-string probability in Eq. (H1).

The probability that the experimental measurement on the ancilla qubit outputs 0 is

$$\mathbb{P}_0 := \sum_{x\in\{0,1\}^n} \int p_{\text{exp}}(U,x)\mathrm{d}U = \sum_{x\in\{0,1\}^n} \int \alpha p(U,x) + \frac{1-\alpha}{2N}\mathrm{d}U = \frac{1+\alpha}{2} \approx P_{\text{exp}}(U).$$

Given the circuit fidelity α, we denote the conditional probability density as

$$\mathbb{Q}_\alpha(U,x) := \frac{1}{\mathbb{P}_0} p_{\text{exp}}(U,x),$$

and the corresponding expectation is denoted as $\mathbb{E}_{\mathbb{Q}_\alpha}(\cdot)$. Then, the conditional average bit-string probability, which is directly related to the parameter b in determining the sXHOG problem as

$$b(\alpha) := N\mathbb{E}_{\mathbb{Q}_\alpha}\left(\sum_{x\neq 0^n} p(U,x)\right) = \frac{\mathbb{E}\left(\text{sXES}(U)\right)}{\mathbb{P}_0} = \left(1 + \frac{\alpha(2-5\mathcal{H}_1+4\mathcal{H}_2)-\mathcal{H}_1}{\alpha+1}\right) + \mathcal{O}\left(\frac{1}{N}\right). \tag{I2}$$

The last equality is derived using results in Appendix K 1. Here

$$\mathcal{H}_1 = \int \mathbb{P}^{(2)}_{\text{eig}}(\lambda_1,\lambda_2)\cos\left(t\left(\lambda_1-\lambda_2\right)\right)\mathrm{d}\lambda_1\mathrm{d}\lambda_2, \tag{I3}$$

and

$$\mathcal{H}_2 = \int \mathbb{P}^{(4)}_{\text{eig}}(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\cos\left(t\left(\lambda_1-\lambda_2+\lambda_3-\lambda_4\right)\right)\mathrm{d}\lambda_1\mathrm{d}\lambda_2\mathrm{d}\lambda_3\mathrm{d}\lambda_4 \tag{I4}$$

are cosine transformations of $\mathbb{P}^{(2)}_{\text{eig}}$ and $\mathbb{P}^{(4)}_{\text{eig}}$, which are the 2-marginal and the 4-marginal distribution of eigenvalues corresponding to the ensemble of random Hermitian matrices, respectively. The values of $\mathcal{H}_1$ and $\mathcal{H}_2$ can be evaluated on classical computers according to Theorem 11 and Algorithm 2.

Thus for large n, we have

$$b(\alpha) \approx 1 + \frac{\alpha(2-5\mathcal{H}_1+4\mathcal{H}_2)-\mathcal{H}_1}{\alpha+1} =: 1 + \frac{\gamma(\alpha-\alpha^*)}{\alpha+1}. \tag{I5}$$

Here $\gamma = 2-5\mathcal{H}_1+4\mathcal{H}_2$ and $\alpha^* = \mathcal{H}_1/\gamma$. When $\mathcal{H}_1, \mathcal{H}_2$ are sufficiently small, $b(\alpha)$ is monotonically increasing.

The hardness of classical spoofing also requires $b(\alpha) \geq 1$ (see Theorem 6). Thus, we define the threshold $\alpha^* := \frac{\mathcal{H}_1}{2-5\mathcal{H}_1+4\mathcal{H}_2}$ be the fidelity so that $b(\alpha^*) = 1$. To achieve supremacy, the fidelity is required to satisfy $\alpha \geq \alpha^*$. To see the existence of the threshold fidelity, let us consider a fully contaminated noise where $\alpha = 0$. Then, the average bit-string probability is

$$\mathbb{E}_{\mathbb{Q}_0}\left(\sum_{x\neq 0^n} p(U,x)\right) = \frac{1}{N}\left(1-\mathbb{E}\left(p(U,0^n)\right)\right) \Rightarrow b(\alpha)|_{\alpha=0} = 1-\mathbb{E}\left(p(U,0^n)\right) \leq 1. \tag{I6}$$

By continuity, a threshold fidelity α^* exists to ensure $b(\alpha) > 1$. It also indicates that the threshold is very close to zero when the diagonal elements of the time evolution vanish simultaneously in the ensemble, namely $\mathbb{E}\left(p(U,0^n)\right) \approx 0$. The threshold can be suppressed by choosing a larger simulation time t since $\alpha^* \to 0$ as $t \to \infty$. Furthermore, when $\alpha^* \ll 1$, the conditional average bit-string probability is

$$\mathbb{E}_{\mathbb{Q}_\alpha}\left(\sum_{x\neq 0^n} p(U,x)\right) = \frac{1}{N}\left(1+\frac{2(\alpha-\alpha^*)}{1+\alpha}\right) + \mathcal{O}\left(\frac{1}{N^2}\right).$$

Note that at t^{opt}, the threshold $\alpha^*|_{t^{\text{opt}}} \approx \mathcal{H}_1/2$. Eq. (K15) implies that $\mathcal{H}_1 \geq -\frac{2}{N-1}$. Therefore, when the number of qubits n is not sufficiently large, $\alpha^*|_{t^{\text{opt}}}$ can possibly be negative. However, as n increases, $\alpha^*|_{t^{\text{opt}}}$ converges to 0 exponentially fast because the lower bound $-\frac{2}{N-1} \to 0$ in the large n-limit. This agrees with the numerical behavior of the threshold α^* shown in Figure 4.

Appendix J: Analytic estimation of t^{opt} for large n

According to the result in Figure 4, at $t = t^{\mathrm{opt}}$, we have

$$\mathbb{E}\left(p_t(U, 0^n)\right) = \mathbb{E}|\left\langle 0^n|e^{-i\mathfrak{H}t}|0^n\right\rangle|^2 \approx 0.$$

Jensen's inequality gives

$$\left|\mathbb{E}\left\langle 0^n|e^{-i\mathfrak{H}t}|0^n\right\rangle\right|^2 \leq \mathbb{E}|\left\langle 0^n|e^{-i\mathfrak{H}t}|0^n\right\rangle|^2 \approx 0.$$

If $\mathfrak{H}$ is a H-RACBEM and the corresponding U is drawn from the Haar measure, then

$$\mathbb{E}\left\langle 0^n|e^{-i\mathfrak{H}t}|0^n\right\rangle = \int_0^1 e^{-i\lambda t}\mathbb{P}_{\mathrm{eig}}^{(1)}(\lambda)\,\mathrm{d}\lambda.$$

Here $\mathbb{P}_{\mathrm{eig}}^{(1)}(\lambda)$ is defined in defined in Eq. (K7) with $\ell = 1$. It is also the 1-marginal (a.k.a. the level density) of the joint probability distribution of all eigenvalues in Eq. (K4), which is called a β-Jacobi ensemble with $\beta = 2$ [32]. With the block encoding of one ancilla qubit (i.e. $M = 2$), the level density follows the Beta$(0.5, 0.5)$ distribution in the large n-limit [40], i.e. in the sense of weak convergence, we have

$$\lim_{n\to\infty}\mathbb{P}_{\mathrm{eig}}^{(1)}(\lambda) = \frac{1}{\pi}\lambda^{-\frac{1}{2}}(1-\lambda)^{-\frac{1}{2}}.$$

Therefore for any t,

$$\begin{aligned}\lim_{n\to\infty}\int_0^1 e^{-i\lambda t}\mathbb{P}_{\mathrm{eig}}^{(1)}(\lambda)\,\mathrm{d}\lambda &= \int_0^1 e^{-i\lambda t}\frac{1}{\pi}\lambda^{-\frac{1}{2}}(1-\lambda)^{-\frac{1}{2}}\,\mathrm{d}\lambda \\ &\overset{\lambda=\sin^2\left(\frac{\theta}{2}\right)}{=\!=\!=} \frac{1}{\pi}\int_0^\pi \exp\left(-it\sin^2\frac{\theta}{2}\right)\mathrm{d}\theta \\ &= e^{-\frac{it}{2}}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\frac{t}{2}\cos\theta}\,\mathrm{d}\theta \\ &= e^{-\frac{it}{2}}J_0(t/2).\end{aligned} \tag{J1}$$

Here we have used the integral representation of the Bessel function of the first kind

$$J_0(t/2) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\frac{t}{2}\sin\theta}\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\frac{t}{2}\cos\theta}\,\mathrm{d}\theta.$$

Therefore in the large n limit, $\mathbb{E}\left\langle 0^n|e^{-i\mathfrak{H}t}|0^n\right\rangle$ approximately vanishes at the first node of $J_0(t/2)$, which gives

$$t^{\mathrm{opt}} \approx 4.81.$$

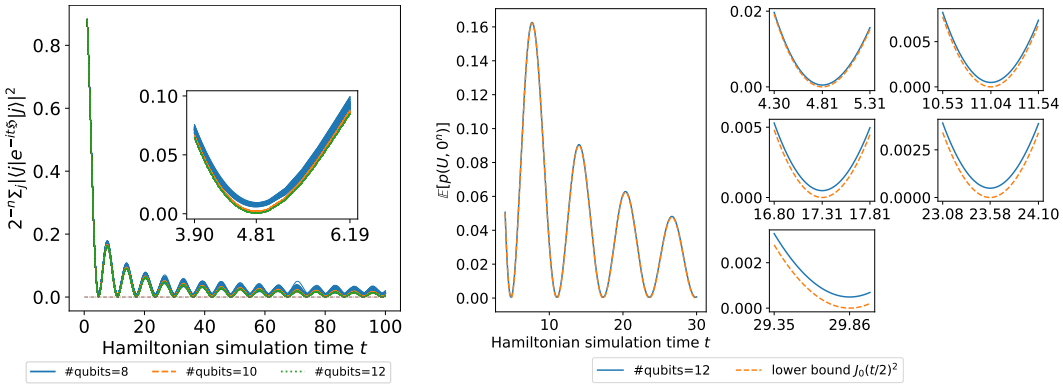
This agrees very well with the numerical results in Figures 4 and S5.

Appendix K: Additional analytical computations and proofs

1. Statistical property of the random-matrix ensemble inherited from Haar measure

The solution of the system heavy output generation problem and analytic evaluation of the system linear cross-entropy score require the use of statistical properties of the ensemble of random matrices obtained from the Haar measure. In this section, we derive the statistical properties of the ensemble. We consider a generic block encoding with m extra ancilla qubits, namely, an n-qubit matrix A is a submatrix of an $(n+m)$-qubit unitary U. We use $M = 2^m$ to represent the dimension of the Hilbert space generated by m ancilla qubits. We assume that U is drawn from an $(n+m)$-qubit Haar measure.

Given the identification $\mathbb{C}^{N\times N} \simeq \mathbb{C}^{N^2}$, the uniform measure on the space of complex matrices is identified as the pushforward of the Lebesgue measure on $\mathbb{C}^{N^2}$, for example, by taking the coordinate system as matrix elements. We denote this uniform measure as $\mathrm{d}A$. Assuming that A is an n-qubit submatrix of a Haar-distributed $(n+m)$-qubit unitary U, the first theorem gives a characterization of the induced probability distribution of A.



(a) (b)

FIG. S5: Numerical justification of t^{opt}. (a) The trajectory of the average diagonal probability $2^{-n}\sum_j \left|\langle j|e^{-it\mathfrak{H}}|j\rangle\right|^2$ as a function of Hamiltonian simulation time t. The broadening is due to plotting ~ 100 instances individually. The subfigure in the box shows the behavior near $t^{\text{opt}} \approx 4.81$. (b) The average probability by measuring all qubits with 0 and the analytical lower bound on it. Zooming into the first five zeros of the Bessel function, the minima of the average probability well agree these zeros.

Theorem 7 ([26, Theorem 1.3.1]). *Let $A \in \mathbb{C}^{N\times N}$ be a submatrix block encoded in a Haar unitary. Then the probability density is*

$$\mathbb{P}(A) = \mathcal{Z}^{-1} \det\left(I - A^\dagger A\right)^{N(M-2)} \mathbb{1}_{\|A\|_2\le 1}, \tag{K1}$$

where $\mathcal{Z} := \int_{\|A\|_2\le 1} \det\left(I - A^\dagger A\right)^{N(M-2)} \mathrm{d}A$ is a normalization constant. Here, $\mathbb{1}$ is an indicator function. It gives 1 when the condition in the subscript is satisfied, and gives 0 otherwise. Generically, let A be an n_1-by-n_2 submatrix of an n-by-n Haar-distributed unitary U, and $n \ge n_1 + n_2$. Then, the probability density is

$$\mathbb{P}(A) \propto \det\left(I - A^\dagger A\right)^{n-n_1-n_2} \mathbb{1}_{\|A\|_2\le 1}. \tag{K2}$$

In particular, for 1-block-encoded matrix with $m = 1$, the exponent of the determinant is 0, and A is uniformly distributed in the unit ball $\{A \in \mathbb{C}^{N\times N} : \|A\|_2 \le 1\}$. Let us consider $A = W\Sigma V^\dagger$ where $W \in \mathsf{U}(N)/\mathsf{U}(1)^N, V \in \mathsf{U}(N)$ and $\mathrm{diag}\,\Sigma = (\sigma_1, \cdots, \sigma_N)$. The Jacobian of this decomposition is $\mathrm{d}A \propto \mathrm{d}V\mathrm{d}W\left(\Delta(\sigma_1^2, \cdots, \sigma_N^2)^2 \prod_{j=1}^N \sigma_j \mathrm{d}\sigma_j\right)$ where $\mathrm{d}W, \mathrm{d}V$ are the Haar measure on their compact manifolds respectively and $\Delta(x_1, \cdots, x_n) := \prod_{i<j}(x_i - x_j)$ is the Vandermonde determinant. Then, the distribution of V, W, Σ follow immediately the theorem.

Corollary 8. *Let W, V be Haar-distributed. The joint distribution of all singular values has the density*

$$\mathbb{P}\left(\sigma_1, \cdots, \sigma_N\right) \propto \Delta(\sigma_1^2, \cdots, \sigma_N^2)^2 \prod_{i=1}^N \sigma_i \left(1 - \sigma_i^2\right)^{N(M-2)} \mathbb{1}_{\sigma_i\in[0,1]}. \tag{K3}$$

Since $\mathfrak{H} = A^\dagger A$, the eigenvalue λ_j of the random Hermitian matrix $\mathfrak{H}$ and the singular value σ_j of the complex matrix A is related by $\lambda_j = \sigma_j^2$. By a direct change-of-variable, the joint distribution of all eigenvalues has the density

$$\mathbb{P}\left(\lambda_1, \cdots, \lambda_N\right) = \mathcal{Z}_{\text{eig}}^{-1} \Delta(\lambda_1, \cdots, \lambda_N)^2 \prod_{i=1}^N (1 - \lambda_i)^{N(M-2)} \mathbb{1}_{\lambda_i\in[0,1]}. \tag{K4}$$

The normalization constant is precisely given by the Selberg's integral [57]

$$\mathcal{Z}_{\text{eig}} = \prod_{j=0}^{N-1} \frac{\Gamma(j+1)\Gamma(j+2)\Gamma(j+N(M-2)+1)}{\Gamma(j+N(M-1)+1)}. \tag{K5}$$

The distribution is invariant under the relabeling of eigenvalues $(\lambda_1, \cdots, \lambda_N) \mapsto (\lambda_{\pi(1)}, \cdots, \lambda_{\pi(N)})$ for any permutation π. This feature is inherited from the bi-invariance of the Haar measure on the compact Lie group.

In practice, only the marginal distribution involving a few eigenvalues will be used. However, the Vandermonde determinant in the joint distribution couples all eigenvalues together, which makes it hard to compute the marginal

distribution analytically. Nonetheless, a semi-analytical representation by orthogonal polynomial expansion can be derived as follows.

Let $w(x) := (1-x)^{N(M-2)}$ be the weight function, $(f,g)_w := \int_0^1 f(x)g(x)w(x)\mathrm{d}x$ be the weighted inner product on $[0,1]$, and $\|f\|_w := \sqrt{(f,f)_w}$ be the weighted norm.

Theorem 9 ([51, Theorem 5.7.1])**.** *Let $\{C_i(x) : \deg C_i = i,\ i = 0,\cdots,N-1\}$ be a set of linearly independent monic polynomials such that they are orthogonal with respect to $(\cdot,\cdot)_w$. Let $c_i := \|C_i\|_w^2$. Define a bivariate function*

$$K(x,y) := w(x)\sum_{i=0}^{N-1}\frac{1}{c_i}C_i(x)C_i(y). \tag{K6}$$

Then, the joint distribution for ℓ eigenvalues follows a determinantal process

$$\mathbb{P}_{eig}^{(\ell)}(\lambda_1,\cdots,\lambda_\ell) = \frac{(N-\ell)!}{N!}\det\left[K(\lambda_{j_1},\lambda_{j_2})\right]_{j_1,j_2=1,\cdots,\ell}. \tag{K7}$$

The orthogonal polynomial can be generated by 3-point recursion formula. Specifically, for 1-block-encoding (i.e. $m=1$ and $M=2$), the weight function is $w\equiv 1$, and the orthogonal polynomial is the shifted Legendre polynomial

$$C_i(x) \propto P_i(2x-1). \tag{K8}$$

Corollary 10. *For 1-block-encoding, the joint eigenvalue distribution can be expressed in terms of the bivariate function*

$$K(x,y) = \sum_{i=0}^{N-1}(2i+1)P_i(2x-1)P_i(2y-1), \tag{K9}$$

where P_i is the i-th Legendre polynomial.

With Corollary 10, the averages of bit-string probability can be evaluated semi-analytically. For a generic complex even polynomial, we define

$$\mathcal{R}^{q_1,\cdots,q_{\ell_2}}_{k_1,\cdots,k_{\ell_1}|r_1,\cdots,r_{\ell_3}} := \mathbb{E}\left(\prod_{j=1}^{\ell_1} g^{k_j}(\sigma_j)\prod_{j=1}^{\ell_2}\overline{g^{q_j}(\sigma_{\ell_1+j})}\prod_{j=1}^{\ell_3}|g^{r_j}(\sigma_{\ell_1+\ell_2+j})|\right). \tag{K10}$$

A complex polynomial $f\in\mathbb{C}[x]$ can be determined by setting $f(x^2)=g(x)$. Note that $g(\sigma_j)=f(\lambda_j)$ relates the singular value transformation and the eigenvalue transformation. The expectation can be expresses exactly by the integration with joint distribution,

$$\mathcal{R}^{q_1,\cdots,q_{\ell_2}}_{k_1,\cdots,k_{\ell_1}|r_1,\cdots,r_{\ell_3}} = \int_{[0,1]^{\ell_1+\ell_2+\ell_3}}\mathbb{P}_{\mathrm{eig}}^{(\ell_1+\ell_2+\ell_3)}\prod_{j=1}^{\ell_1}f^{k_j}(\lambda_j)\prod_{j=1}^{\ell_2}\overline{f^{q_j}(\lambda_{\ell_1+j})}\prod_{j=1}^{\ell_3}|f^{r_j}(\lambda_{\ell_1+\ell_2+j})|\,\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_{\ell_1+\ell_2+\ell_3}. \tag{K11}$$

For Hamiltonian simulation, e^{-itx^2} has unit absolute value for all $x\in\mathbb{R}$. For simplicity we assume the approximation error is sufficiently small, and $g(x)=s_t(x)=e^{-itx^2}$, or equivalently $f(x)=e^{-itx}$. This allows us to omit the terms due to $|f^{r_j}(\lambda_{\ell_1+\ell_2+j})|$. Furthermore, when the upper and lower indices of $\mathcal{R}$ in Eq. (K11) are the same $k_j=q_j=1$, the relabeling invariance of eigenvalues implies that the defined quantity is reduced to

$$\mathcal{H}_\ell(t) = \mathbb{E}\left(\prod_{j=1}^{\ell}g(\sigma_j)\overline{g(\sigma_{\ell+j})}\right) = \int_{[0,1]^{2\ell}}\mathbb{P}_{\mathrm{eig}}^{(2\ell)}(\lambda_1,\cdots,\lambda_{2\ell})\cos\left(t\sum_{j=1}^{\ell}\lambda_j-\lambda_{\ell+j}\right)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_{2\ell}, \tag{K12}$$

which is directly related to the Hamiltonian simulation. When the time dependence is irrelevant to the analysis, we drop the t dependence in $\mathcal{H}_\ell(t)$ by writing it as $\mathcal{H}_\ell$ for simplicity. We can compute $\mathcal{H}_\ell$ as follows.

Theorem 11. *Let the degree-d complex polynomial $f(x)=\sum_{q=0}^d c_q P_q(2x-1)$ be decomposed in terms of Legendre polynomials. Then,*

$$\mathcal{H}_\ell = \frac{(N-2\ell)!}{N!}\sum_{k_1=0}^{N-1}\cdots\sum_{k_{2\ell}=0}^{N-1}\sum_{\varsigma\in\mathsf{S}_{2\ell}}\mathrm{sgn}(\varsigma)\prod_{j=1}^{2\ell}\left((2k_j+1)\sum_{q=0}^{d}C_q^{(j)}F_{q,k_j,k_{\varsigma^{-1}(j)}}\right). \tag{K13}$$

Here, $\mathsf{S}_{2\ell}$ is the symmetric group,

$$C_q^{(j)} = \begin{cases} c_q \ , & \text{if } j \le \ell, \\ \overline{c_q} \ , & \text{otherwise,} \end{cases}$$

and

$$F_{i,j,k} = \frac{1}{2}\int_{-1}^{1} P_i(x)P_j(x)P_k(x)\mathrm{d}x = \begin{cases} \frac{(2s-2i)!(2s-2j)!(2s-2k)!}{(2s+1)!}\left(\frac{s!}{(s-i)!(s-j)!(s-k)!}\right)^2, \\ \quad \text{if } 2s = i+j+k \text{ is even and } |i-j| \le k \le i+j, \\ 0, \ \text{otherwise.} \end{cases} \tag{K14}$$

Proof. Let $f^{(j)}(x) = \sum_{q=0}^{d} C_q^{(j)} P_q(2x-1)$ so that $f^{(j)}(x) = f(x)$ when $j \le \ell$ and $f^{(j)}(x) = \overline{f(x)}$ when $j > \ell$. Then, directly applying Theorem 9 and Corollary 10, the quantity can be evaluated immediately,

$$\begin{aligned} \frac{N!}{(N-2\ell)!}\mathcal{H}_\ell &= \sum_{\varsigma\in\mathsf{S}_{2\ell}} \mathrm{sgn}(\varsigma) \int_{[0,1]^{2\ell}} \prod_{j=1}^{2\ell} f^{(j)}(x_j) K(x_j, x_{\varsigma(j)})\mathrm{d}x_j \\ &= \sum_{\varsigma\in\mathsf{S}_{2\ell}} \mathrm{sgn}(\varsigma) \sum_{k_1=0}^{N-1} \cdots \sum_{k_{2\ell}=0}^{N-1} \prod_{j=1}^{2\ell} (2k_j+1) \int_0^1 f^{(j)}(x) P_{k_j}(2x-1) P_{k_{\varsigma^{-1}(j)}}(2x-1)\mathrm{d}x. \end{aligned}$$

The conclusion follows. □

The constraint in Eq. (K14) will be referred to as the triangle rule. Due to the triangle rule, $F_{i,j,k}$ is a sparse tensor. Many terms in the (2ℓ)-fold summation vanishes, which can be used to accelerate the evaluation. Note that Legendre polynomials are bounded by 1 on $[-1,1]$, which implies that $|F_{i,j,k}| \le 1$. To circumvent the numerical instability arising from factorials, $F_{i,j,k}$ can be evaluated recursively,

$$\begin{aligned} F_{0,0,0} = 1, \ F_{i,j+1,k+1} &= \frac{2s+1-2i}{s+1-i}\frac{s+1}{2s+3}F_{i,j,k}, \\ F_{i,j,k+2} &= \frac{2s+1-2i}{s+1-i}\frac{2s+1-2j}{s+1-j}\frac{s-k}{2s-2k-1}\frac{s+1}{2s+3}F_{i,j,k}, \end{aligned}$$

where $2s = i+j+k$. Using Algorithm 2, $F_{i,j,k}$ can be evaluated stably with s recursions.

ALGORITHM 2: A stable recursive algorithm for computing $F_{i,j,k}$

Input: A triplet (i,j,k) satisfying the triangle rule.

Order and relabel the triplet so that $i \le j \le k$.
Set $2s = i+j+k$.
if $k \ge j-i+2$ **then**
  Recursively call the algorithm to compute $F_{i,j,k-2}$. Note $(i,j,k-2)$ preserves the triangle rule.
  Return: $F_{i,j,k} = \frac{2s-1-2i}{s-i}\frac{2s-1-2j}{s-j}\frac{s-k+1}{2s-2k+1}\frac{s}{2s+1}F_{i,j,k-2}$.
else if $k < j-i+2$ and $i < j$ **then**
  Recursively call the algorithm to compute $F_{i,j-1,k-1}$. Note $(i,j-1,k-1)$ preserves the triangle rule.
  Return: $F_{i,j,k} = \frac{2s-1-2i}{s-i}\frac{s}{2s+1}F_{i,j-1,k-1}$.
else $i = j = k = 0$
  Return: $F_{0,0,0} = 1$.
end if

The measurement on all qubits gives an $(n+1)$-bit string. We are interested in the bit-string $0x$ which means the outcome of the ancilla qubit is 0 and that of n system qubits is $x \in \{0,1\}^n$. The probability of measuring the bit-string $0x$ is

$$p(U,x) = |\langle 0x|\mathcal{U}_{f,U}|00^n\rangle|^2 = \left|\langle x|g^{\triangleright}(A)|0^n\rangle\right|^2 = \sum_{j,k=0}^{N-1} g(\sigma_j)\overline{g(\sigma_k)}V_{j0}\overline{V_{jx}}\,\overline{V_{k0}}V_{kx}.$$

When $x = 0^n \equiv 0$, $p(U,x) = \sum_{j,k} g(\sigma_j)\overline{g(\sigma_k)}\,|V_{j0}|^2\,|V_{k0}|^2$ involves only one column of a Haar unitary V. Yet when $x \ne 0^n$, the probability involves two columns. For $x \ne 0^n$ or 1^n, let us consider another unitary $\widetilde{V}$ by permuting the column 1 and column x of V. By the bi-invariance of Haar measure on unitary group, $\widetilde{V}$ and V are identically distributed. Therefore, we conclude the following lemma.

Lemma 12. *For any nonzero bit-string $0^n \neq x \in \{0,1\}^n$, $p(U,x)$ is identically distributed.*

According to Corollary 8, the distributions of Σ and V are decoupled. The average over the singular values can be evaluated semi-analytically, and the average over the Haar unitary can be analytically computed by using representation theory. We conclude the relevant average values as follows.

Theorem 13. *For Hamiltonian simulation benchmark, the averages of bit-string probability are*

$$\mathbb{E}\left(p(U,0^n)\right) = \frac{N-1}{N+1}\mathcal{H}_1 + \frac{2}{N+1}, \quad \mathbb{E}\left(\sum_{x\neq 0^n} p(U,x)\right) = \frac{N-1}{N+1}\left(1-\mathcal{H}_1\right), \tag{K15}$$

and

$$\begin{aligned}
\mathbb{E}\left(p(U,0^n)^2\right) &= \frac{12}{(N+2)(N+3)} + \frac{12N(N-1)\mathcal{H}_1}{(N+1)(N+2)(N+3)} + \frac{(N-1)(N-2)(N-3)}{(N+1)(N+2)(N+3)}\mathcal{H}_2, \\
\mathbb{E}\left(\sum_{x\neq 0^n} p(U,x)^2\right) &= \frac{2(N-1)(N^2+3N+6)}{N(N+1)(N+2)(N+3)} - \frac{4(N-1)(N^2-N+6)\mathcal{H}_1}{N(N+1)(N+2)(N+3)} + \frac{2(N-1)(N-2)(N-3)}{N(N+1)(N+2)(N+3)}\mathcal{H}_2.
\end{aligned} \tag{K16}$$

Proof. We first evaluate the first order moments in Eq. (K15). Let $p_j =: |V_{j0}|^2$. Directly applying Theorem 7 to the 0-th column of V, the joint probability density of k distinct success probabilities is

$$\mathbb{P}(p_1,\cdots,p_k) = \prod_{j=0}^{k-1}(N-k+j)\left(1-\sum_{j=1}^{k} p_j\right)^{N-k-1} \mathbb{1}_{\sum_{j=1}^{k} p_j<1}.$$

Given an index set $\boldsymbol{\alpha} := (\alpha_1,\cdots,\alpha_k)$ and $|\boldsymbol{\alpha}| := \sum_{j=1}^{k}\alpha_j$, the average $\mathcal{I}_{\boldsymbol{\alpha}} := \mathbb{E}\left(\prod_{j=1}^{k} p_j^{\alpha_j}\right) = \left(\prod_{j=1}^{k}\alpha_j!\right)\prod_{j=0}^{|\boldsymbol{\alpha}|-1}\frac{1}{N+j}$ follows direct computation,

$$\begin{aligned}
\left(\prod_{j=0}^{k-1}\frac{1}{N-k+j}\right)\mathcal{I}_{\boldsymbol{\alpha}} &= \int_{\sum_{j=1}^{k-1} p_j<1}\prod_{j=1}^{k-1} p_j^{\alpha_j}\mathrm{d}p_j \int_0^{1-\sum_{j=1}^{k-1} p_j} p_k^{\alpha_k}\left(1-\sum_{j=1}^{k-1} p_j - p_k\right)^{N-k-1}\mathrm{d}p_k \\
&= \alpha_k!\prod_{j=0}^{\alpha_k}\frac{1}{N-k+j}\int_{\sum_{j=1}^{k-1} p_j<1}\left(1-\sum_{j=1}^{k-1} p_j\right)^{N-k+\alpha_k}\prod_{j=1}^{k-1} p_j^{\alpha_j}\mathrm{d}p_j = \cdots = \left(\prod_{j=1}^{k}\alpha_j!\right)\prod_{j=0}^{|\boldsymbol{\alpha}|+k-1}\frac{1}{N-k+j}.
\end{aligned}$$

For the second equal sign, we use the identity

$$\int_0^y x^\alpha (y-x)^\beta \mathrm{d}x = y^{\alpha+\beta+1}\mathrm{B}(\alpha+1,\beta+1) = y^{\alpha+\beta+1}\frac{\alpha!\beta!}{(\alpha+\beta+1)!}, \tag{K17}$$

where the Beta function is

$$\mathrm{B}(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}\mathrm{d}t = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

By definition, $p(U,0^n) = \sum_{j,k} g(\sigma_j)\overline{g(\sigma_k)}p_j p_k$. Applying these results for average and using the fact that singular values and singular vectors are independent, the average bit-string probability is

$$\mathbb{E}\left(p(U,0^n)\right) = N\mathbb{E}\left(|g(\sigma_1)|^2\right)\mathcal{I}_{(2)} + N(N-1)\mathbb{E}\left(\frac{1}{2}\left(g(\sigma_1)\overline{g(\sigma_2)} + \overline{g(\sigma_1)}g(\sigma_2)\right)\right)\mathcal{I}_{(1,1)} = \frac{N-1}{N+1}\mathcal{H}_1 + \frac{2}{N+1}.$$

Then,

$$\mathbb{E}\left(\sum_{x\neq 0^n} p(U,x)\right) = 1 - \mathbb{E}\left(p(U,0^n)\right) = \frac{N-1}{N+1}\left(1-\mathcal{H}_1\right).$$

Now we evaluate the second order moments in Eq. (K16). Let $\pi \in \mathsf{S}_4$ be any permutation, and R be any set of constraints on four n-bit binary strings i, j, k, l. We denote the action of the symmetric group as $\pi \cdot \mathsf{R} := \{(\pi(i), \pi(j), \pi(k), \pi(l)) : (i, j, k, l) \in \mathsf{R}\}$. Due to the relabeling invariance of the joint distribution of singular values, we have

$$\sum_{(i,j,k,l)\in\pi\cdot\mathsf{R}} \mathbb{E}\left(g(\sigma_i)\overline{g(\sigma_j)}g(\sigma_k)\overline{g(\sigma_l)}\right) = \sum_{(i,j,k,l)\in\mathsf{R}} g(\sigma_{\pi(i)})\overline{g(\sigma_{\pi(j)})}g(\sigma_{\pi(k)})\overline{g(\sigma_{\pi(l)})} = \sum_{(i,j,k,l)\in\mathsf{R}} g(\sigma_i)\overline{g(\sigma_j)}g(\sigma_k)\overline{g(\sigma_l)}. \tag{K18}$$

For example, let $\mathsf{R}_1 := \{(i = j) \neq (k = l)\}, \mathsf{R}_2 := \{(i = k) \neq (j = l)\}, \mathsf{R}_3 := \{(i = l) \neq (j = k)\}$, and $\pi_1 := \begin{pmatrix} i & j & k & l \\ i & k & j & l \end{pmatrix}, \pi_2 := \begin{pmatrix} i & j & k & l \\ i & l & j & k \end{pmatrix}$. Then, $\pi_1 \cdot \mathsf{R}_1 = \mathsf{R}_2$ and $\pi_2 \cdot \mathsf{R}_1 = \mathsf{R}_3$ holds. Therefore

$$\sum_{(i,j,k,l)\in\mathsf{R}_\ell} \mathbb{E}\left(g(\sigma_i)\overline{g(\sigma_j)}g(\sigma_k)\overline{g(\sigma_l)}\right) = N(N-1)\mathbb{E}\left(|g(\sigma_1)|^2 |g(\sigma_2)|^2\right) = N(N-1), \quad \ell = 1, 2, 3.$$

For bit-string $x = 0^n$,

$$\mathbb{E}\left(p(U, 0^n)^2\right) = \mathbb{E}\left(\sum_{i,j,k,l} g(\sigma_i)\overline{g(\sigma_j)}g(\sigma_k)\overline{g(\sigma_l)}p_i p_j p_k p_l\right)$$

Using Eq. (K18), the four-fold summation can be categorized into five equivalent classes. We define the partition $\boldsymbol{\alpha}$ be the array with at most 4 entries whose sum is exactly 4.

1. All indices are distinct with partition $\boldsymbol{\alpha} = (1, 1, 1, 1)$: $i \neq j \neq k \neq l$. The contribution is $N(N-1)(N-2)(N-3)\mathcal{H}_2\mathcal{I}_{(1,1,1,1)}$ where $\mathcal{I}_{(1,1,1,1)} = \frac{1}{N(N+1)(N+2)(N+3)}$.
2. Only three indices are distinct with partition $\boldsymbol{\alpha} = (2, 1, 1)$: $i = j \neq k \neq l$ and its other five permutations. The contribution is $6N(N-1)(N-2)\mathcal{H}_1\mathcal{I}_{(2,1,1)}$ where $\mathcal{I}_{(2,1,1)} = \frac{2}{N(N+1)(N+2)(N+3)}$.
3. Only two indices are distinct with partition $\boldsymbol{\alpha} = (3, 1)$: $i = j = k \neq l$ and its other three permutations. The contribution is $4N(N-1)\mathcal{H}_1\mathcal{I}_{(3,1)}$ where $\mathcal{I}_{(3,1)} = \frac{6}{N(N+1)(N+2)(N+3)}$.
4. Only two indices are distinct with partition $\boldsymbol{\alpha} = (2, 2)$: $(i = j) \neq (k = l)$ and its other two permutations. The contribution is $3N(N-1)\mathcal{I}_{(2,2)}$ where $\mathcal{I}_{(2,2)} = \frac{4}{N(N+1)(N+2)(N+3)}$.
5. All indices are the same with partition $\boldsymbol{\alpha} = (4)$: $i = j = k = l$. The contribution is $N\mathcal{I}_{(4)}$ where $\mathcal{I}_{(4)} = \frac{24}{N(N+1)(N+2)(N+3)}$.

Collecting all the terms together, we have

$$\mathbb{E}\left(p(U, 0^n)^2\right) = \frac{12}{(N+2)(N+3)} + \frac{12N(N-1)\mathcal{H}_1}{(N+1)(N+2)(N+3)} + \frac{(N-1)(N-2)(N-3)}{(N+1)(N+2)(N+3)}\mathcal{H}_2.$$

The evaluation of the second moment of the probability with nonzero bit-strings x follows a similar procedure but is more involved. Using Lemma 12, $\mathbb{E}\left(\sum_{x\neq 0^n} p(U, x)^2\right) = (N-1)\mathbb{E}\left(p(U, 1^n)^2\right)$. Let

$$H(i, j, k, l) := \mathbb{E}\left(\overline{V_{1^n,i}}V_{0^n,i}V_{1^n,j}\overline{V_{0^n,j}}\,\overline{V_{1^n,k}}V_{0^n,k}V_{1^n,l}\overline{V_{0^n,l}}\right).$$

Then,

$$\mathbb{E}\left(p(U, 1^n)^2\right) = \sum_{ijkl} \mathbb{E}\left(g(\sigma_i)\overline{g(\sigma_j)}g(\sigma_k)\overline{g(\sigma_l)}\right) H(i, j, k, l).$$

Let us consider a linear map $\phi : \left(\mathbb{C}^N\right)^{\otimes 4} \to \left(\mathbb{C}^N\right)^{\otimes 4}$,

$$\phi_{ijkl} := \mathbb{E}\left(\bigotimes_{q\in\{i,j,k,l\}} V |q\rangle \langle q| V^\dagger\right),$$

Then

$$H(i,j,k,l) = \langle 0^n, 1^n, 0^n, 1^n | \phi_{ijkl} | 1^n, 0^n, 1^n, 0^n \rangle .$$

Note that the linear map ϕ_{ijkl} commutes with the action of the symmetric group S_4 and that of the unitary group $\mathsf{U}(N)$ on the tensor product space $\left(\mathbb{C}^N\right)^{\otimes 4}$. Then, by Schur's lemma over $\mathbb{C}$, ϕ_{ijkl} is a scalar on each subspace decomposed with respect to Schur-Weyl duality. By taking trace on each subspace, the linear map is determined as the linear combination of projectors. The generic formula was derived in [27] by which $H(i,j,k,l)$ is evaluated. Let the four-fold sum in terms of $ijkl$ be broken into five equivalent classes as follows.

1. All indices are distinct with partition $(1,1,1,1)$: $i \neq j \neq k \neq l$. The contribution is $N(N-1)(N-2)(N-3)\mathcal{H}_2 \frac{2}{N^2\,(N-1)(N+1)(N+2)(N+3)}$.
2. Only three indices are distinct with partition $(2,1,1)$: $i = j \neq k \neq l$ and its other five permutations. The contribution is $6N(N-1)(N-2)\mathcal{H}_1 \left(-\frac{2\,(N-3)}{3\,N^2\,(N-1)(N+1)(N+2)(N+3)}\right)$.
3. Only two indices are distinct with partition $(3,1)$: $i = j = k \neq l$ and its other three permutations. The contribution is $4N(N-1)\mathcal{H}_1 \left(-\frac{4}{N(N-1)(N+1)(N+2)(N+3)}\right)$.
4. Only two indices are distinct with partition $(2,2)$: $(i = j) \neq (k = l)$ and its other two permutations. The contribution is $3N(N-1)\frac{2\left(N^2+N+6\right)}{3\,N^2\,(N-1)(N+1)(N+2)(N+3)}$.
5. All indices are the same with partition (4): $i = j = k = l$. The contribution is $N\frac{4}{N\,(N+1)\,(N+2)\,(N+3)}$.

To conclude, the second moment of the probability with nonzero bit-strings is

$$\mathbb{E}\left(\sum_{x\neq 0^n} p(U,x)^2\right) = \frac{2(N-1)(N^2+3N+6)}{N(N+1)(N+2)(N+3)} - \frac{4(N-1)(N^2-N+6)\mathcal{H}_1}{N(N+1)(N+2)(N+3)} + \frac{2(N-1)(N-2)(N-3)}{N(N+1)(N+2)(N+3)}\mathcal{H}_2.$$

□

The asymptotic behavior of the ensemble is discussed in Appendix K 3.

2. Concatenating phase factors for long time Hamiltonian simulation

Although simulation at very long time is certainly beyond the regime of near term applications, the unitarity of Hamiltonian simulation provides an alternative method to obtain phase factors. Specifically, given the phase factor sequence at some short time t, the phase factor sequence at a long simulation time rt can be easily constructed for some integer $r > 1$. The procedure, called phase factor concatenation, is given in Eq. (K19), and the quality of the approximation is describe in Theorem 14.

Theorem 14. *Let $\Phi = (\phi_0, \cdots, \phi_d)$ be a set of phase factors so that $\|P(x,\Phi) - s_t(x)\|_\infty \leq \epsilon$. Given an integer $r \geq 1$, define*

$$\Phi^{(r)} := \left(\phi_0, \phi_1, \cdots, \phi_{d-1}, \underbrace{\phi_d + \phi_0, \phi_1, \cdots, \phi_{d-1}}_{\text{repeat } r-1 \text{ times}}, \phi_d\right). \tag{K19}$$

Then,

$$\left\| P(x,\Phi^{(r)}) - s_{tr}(x) \right\|_\infty \leq r^2\epsilon. \tag{K20}$$

Proof. For simplicity, let $P(x) := P(x,\Phi)$ and $P^{(r)}(x) := P(x,\Phi^{(r)})$. According to the normalization condition,

$$(1-x^2)\,|Q(x)|^2 = 1 - P(x)\overline{P(x)} = (s_t(x) - P(x))\,\overline{P(x)} + s_t(x)\overline{(s_t(x) - P(x))}.$$

By the triangle inequality, $\left\|(1-x^2)\,|Q(x)|^2\right\|_\infty \leq 2\epsilon$.

Let ϵ_r be the approximation error of $P^{(r)}$, and let $Q^{(r)}$ be the complementing polynomial in as Theorem 1. Similarly, we have $\left\|(1-x^2)\left|Q^{(r)}(x)\right|^2\right\|_\infty \leq 2\epsilon_r$. By the construction of the phase factor sequence in Eq. (K19), we have

$$P^{(r)}(x) = P^{(r-1)}(x)P(x) - (1-x^2)Q^{(r-1)}(x)\overline{Q(x)}.$$

This gives

$$P^{(r)}(x) - s_t^r(x) = \left(P^{(r-1)}(x) - s_t^{r-1}(x)\right)P(x) + s_t^{r-1}(x)\left(P(x) - s_t(x)\right) - (1-x^2)Q^{(r-1)}(x)\overline{Q(x)}.$$

Note that $s_t^r(x) = s_{tr}(x)$. By the triangle inequality, $\epsilon_r \leq \epsilon_{r-1} + \epsilon + 2\sqrt{\epsilon_{r-1}\epsilon} = (\sqrt{\epsilon_{r-1}} + \sqrt{\epsilon})^2$. Then, $\sqrt{\epsilon_r} \leq \sqrt{\epsilon_{r-1}} + \sqrt{\epsilon} \leq \cdots \leq r\sqrt{\epsilon}$. Therefore

$$\epsilon_r \leq r^2\epsilon, \tag{K21}$$

which proves the theorem. □

Following the construction of Eq. (K19), in order to perform Hamiltonian simulation at time rt, the length of the phase factor sequence increases by a factor of r, and the error grows at most quadratically with respect to r according to Theorem 14. However, from Tables S2 and S3 we observe that this estimate can be significantly improved if we can construct the phase factor sequence at time rt directly using the optimization based method. Even when Eq. (K19) is used, numerical results in Figure S2 also indicate that $\epsilon_r \leq r^2\epsilon$ is only an upper bound of the numerical error, which only grows linearly at least until $t = 10$.

3. Asymptotic behavior of the long time Hamiltonian simulation benchmark

The first and second moments of the bit-string probability are directly relevant to the construction of the system linear cross-entropy benchmarking based on Hamiltonian simulation. For simplicity, we define

$$K_1(x,t) := \mathbb{E}\left(p_t(U,x)\right), \text{ and } K_2(x,t) := \mathbb{E}\left(p_t(U,x)^2\right), \tag{K22}$$

where the dependence on t is encoded in the implementation of the quantum circuit. In this section, we will investigate the behavior of these moments in different regimes in terms of the Hamiltonian simulation time t when N is sufficiently large.

According to Lemma 12, $K_1(x,t)$ and $K_2(x,t)$ are constant for any nonzero bit-string $x \neq 0^n$. Therefore, by using Theorem 13, we have

$$K_1(0^n,t) = \mathcal{H}_1(t) + \mathcal{O}\left(\frac{1}{N}\right), \; K_2(0^n,t) = \mathcal{H}_2(t) + \frac{12}{N}\left(\mathcal{H}_1(t) - \mathcal{H}_2(t)\right) + \mathcal{O}\left(\frac{1}{N^2}\right),$$

and for any $x \neq 0^n$,

$$K_1(x,t) = \frac{1}{N}\left(1 - \mathcal{H}_1(t)\right) + \mathcal{O}\left(\frac{1}{N^2}\right), \; K_2(x,t) = \frac{1}{N^2}\left(2 - 4\mathcal{H}_1(t) + 2\mathcal{H}_2(t)\right) + \mathcal{O}\left(\frac{1}{N^3}\right).$$

Note that by definition, $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ are the cosine transformation of the joint eigenvalue distribution $\mathbb{P}_{\text{eig}}^{(2)}$ and $\mathbb{P}_{\text{eig}}^{(4)}$ respectively. According to Theorem 9, these joint distributions are polynomials. Then both $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$ converge to zero as $t \to \infty$. In particular, there exists t^* such that $\mathcal{H}_1(t), \mathcal{H}_2(t) = \mathcal{O}\left(\frac{1}{N}\right)$ for any $t > t^*$. In this regime, for any nonzero bit-string $x \neq 0^n$,

$$K_1(x,t) = \frac{1}{N} + \mathcal{O}\left(\frac{1}{N^2}\right), \text{ and } K_2(x,t) = \frac{2}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right).$$

Note that the bit-string probability of a Haar-distributed unitary U, $p_{ij} := |U_{ij}|^2$, has the first and second moment $\mathbb{E}\left(p_{ij}\right) = \frac{1}{N}$ and $\mathbb{E}\left(p_{ij}^2\right) = \frac{2}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)$. Furthermore, by using Lemma 12 and Theorem 13, the cross moment of two different nonzero bit-strings $x \neq y$ is

$$\mathbb{E}\left(p_t(U,x)p_t(U,y)\right) = \frac{1 - 2\mathcal{H}_1(t) + \mathcal{H}_2(t)}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right).$$

Thus, in the regime where $\mathcal{H}_1(t), \mathcal{H}_2(t) = \mathcal{O}\left(\frac{1}{N}\right)$, the cross moment is $\frac{1}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)$. Following Theorem 2, the cross moment of success probabilities of a Haar-distributed unitary is exactly the same up to higher order $\mathbb{E}\left(p_{ij}p_{kj}\right) = \frac{1}{N(N+1)} = \frac{1}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right)$. Remarkably, the correlation between different nonzero bit-strings is small, which is quantified by the covariance $\mathrm{Cov}\left(p(U,x)p(U,y)\right) = \mathbb{E}\left(p(U,x)p(U,y)\right) - \mathbb{E}\left(p(U,x)\right)\mathbb{E}\left(p(U,y)\right) = \mathcal{O}\left(\frac{1}{N^3}\right)$. We conclude that in the defined regime, the ensemble of the time evolution matrix induced by our construction has approximately the same statistics as that of Haar-distributed unitaries up to at least the second moment.

Note that the threshold $\alpha_t^* := \frac{\mathcal{H}_1(t)}{2-5\mathcal{H}_1(t)+4\mathcal{H}_2(t)}$ is directly related to $\mathcal{H}_1(t)$ and $\mathcal{H}_2(t)$. When $t = t^{\mathrm{opt}}$ for small time or $t > t^*$ in the long time regime, we have $\mathcal{H}_1(t), \mathcal{H}_2(t) = \mathcal{O}\left(\frac{1}{N}\right)$ which leads to an exponentially small threshold $\alpha_t^* = \mathcal{O}\left(\frac{1}{N}\right)$. It allows the quantum supremacy to be achieved for a small circuit fidelity $\alpha > \alpha_t^*$. Note t^* can be impractically large for near-term applications. Hence it is crucial that the simulation at $t = t^{\mathrm{opt}} \approx 4.81$ is equally effective and much more tractable.