Letter

Gaining confidence on the correct realization of arbitrary quantum computations

Jose Carrasco,[1,2] Marc Langer,[3,2,4] Antoine Neven,[2] and Barbara Kraus[3,2,4]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, 14195 Berlin, Germany*
[2]*University of Innsbruck, Institute for Theoretical Physics, A-6020 Innsbruck, Austria*
[3]*Technical University of Munich, TUM School of Natural Sciences, Physics Department, 85748 Garching, Germany*
[4]*Munich Center for Quantum Science and Technology (MCQST), Schellingstr. 4, 80799 Munich, Germany*

(Received 21 September 2023; accepted 15 August 2024; published 30 September 2024)

We present verification protocols to gain confidence in the correct performance of a device implementing an arbitrary quantum computation. The derivation of the protocols is based on the fact that matchgate computations, which are classically efficiently simulable, become universal if supplemented with additional resources. We combine tools from weak simulation, randomized compiling, and statistics to derive verification circuits that (i) strongly resemble the original circuit and (ii) can be classically efficiently simulated not only in the ideal, i.e., error free scenario, but also in the realistic situation where errors are present. In fact, in one of the protocols we apply exactly the same circuit as in the original computation, however, to a slightly modified input state.

DOI: 10.1103/PhysRevResearch.6.L032074

Introduction. With the advent of ever larger quantum processors, the question of how to evaluate their performance becomes increasingly relevant. A distinction is made between pro- tocols in which the quantum device is potentially untrusted by the user, such as a cloud computer, from those in which one does have direct access to the quantum processor. In the former scenario a solution based on interactive proofs and postquantum cryptography has been presented [1–3]. In the latter, several protocols to gain confidence in the performance of a quantum device have been put forward (for a recent review, see [4]). Randomized benchmarking [5,6] and several variants of it have been developed [7–9] where the average performance of, e.g., Clifford gates, is quantified with a single parameter, the average gate fidelity [10].

Also, due to their importance in fault tolerant quantum computation, the simulability of Clifford circuits has been utilized to study the performance of particular quantum computations [11,12] and bound the total variation distance of the erroneous to the ideal output state [12]. However, in order to accomplish the challenging task of verifying universal quantum computations and to also take into account the gates that naturally occur in, e.g., real-time Hamiltonian evolution, which might be difficult to benchmark with other methods [13], different gate sets need to be considered. Moreover, the problem of characterizing the reliability of implementing a particular (not on average) universal computation with more than a single error parameter is largely unexplored. It is precisely this problem that we will address in the present work: To check not only the computation itself but to also gain confidence in the correctness of an entire error model.

We consider an arbitrary quantum computation and analyze how the distribution of single or multiqubit measurement outcomes can be tested. One of the main obstacles here is, of course, that the correct outcome is unknown, as the computation can not be performed classically efficiently[1]. Moreover, as it is (in the near future) inevitable that errors will occur during the computation, a classical simulation of the ideal computation would only provide a limited amount of information on the quality of the quantum device. Only the simulation of the erroneous computation, which can be parametrized by a set of error parameters, allows not only to establish trust in the outcome, but also in the error model. A natural issue arising here is that even if the computation itself were efficiently simulable, inclusion of arbitrary—possibly coherent—errors likely renders the simulation hard.

We will show that combining methods from weak simulation of quantum computation [16] with randomized compiling [17] and classical statistics allows one to overcome these obstacles. For a given (universal) quantum computation we will introduce verification protocols, which test computations which differ only in some gates, or the input state, compared to the original computation (see Fig. 1). We will show that via the notion of randomized compiling, the output state of the erroneous quantum computation can be tailored into one which is parametrized by a few error parameters and is, crucially, still weakly efficiently simulable. This holds under mild conditions on the error model. The circuits we verify must be classically efficiently simulable and therefore must, in general, differ from the original circuit. However, we choose our verification circuits very similar to the original circuit. In fact, we either only add some additional gates,

[1]Note that there are however methods to compare the output of two computations (quantum and/or classical), such as crossplatform verifications [14].

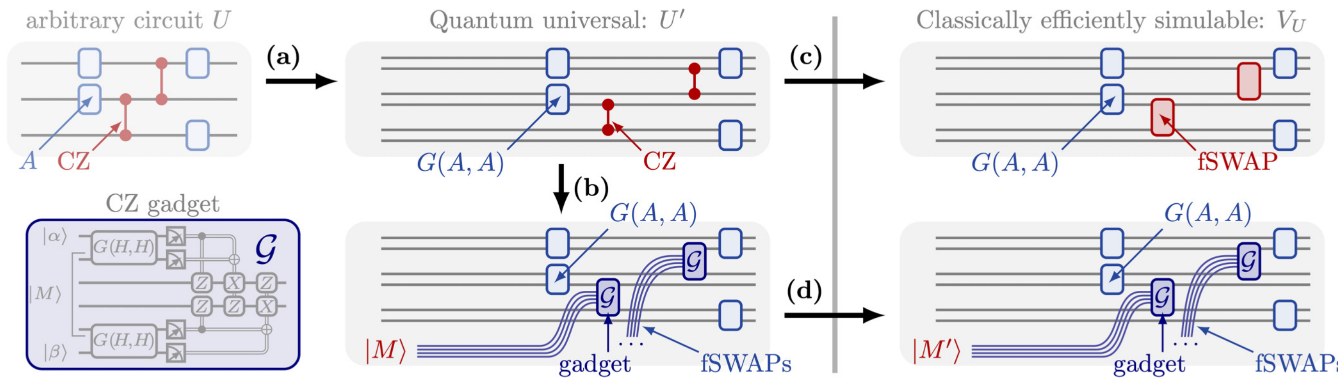

FIG. 1. Verification circuits: (a) U is mapped to its encoded version, U', which is acting on $2m$ qubits and is decomposed into n.n. MGs and CZ gates; (b) each resourceful CZ gates can be implemented deterministically using the magic states $|M\rangle$ and adaptive measurements via the gadget (bottom left) [15]. To verify the encoded realizations of U through efficient classical simulations, we consider verification circuits obtained either by (c) replacing all CZ gates with fSWAP gates but leaving the structure of the circuit intact, or (d) implementing exactly the same circuit but on a slightly modified input state.

or apply exactly the same computation on a slightly different input state. Then, we will show that tests from classical statistics allow us to estimate the error parameters of the (slightly modified) computation (see Fig. 1). Hence, as an output of these tests we obtain the error parameters which completely characterize the output state of the slightly modified (randomly compiled) circuit. This can then also be utilized to verify the error model, to detect, e.g., possible drifts during the computation, and might also be used in error mitigation. We illustrate the performance of the verification protocols with some examples, where we also present some additional tests.

Central to our approach are matchgate circuits. A matchgate (MG) is a two-qubit gate which can be written as $G(U_1, U_2) = U_1 \oplus U_2$. Here, U_1 (U_2) is a unitary acting on the even (odd) parity subspace, respectively, and their determinants coincide. Matchgate circuits (MGCs) are composed out of nearest neighbor (n.n.) MGs acting on computational basis states and the output corresponds to the measurement of a single (or several) qubit(s) in the computational basis. MGCs can be simulated classically efficiently [18–20], and even compressed into an exponentially smaller quantum computer [21]. Which resources can be added to MGCs such that the computation remains classically efficiently simulable has been studied in [22,23]. One distinguishes here between strong and weak simulation. Whereas strong simulation means that for any given output bitstring z on any subset of measured qubits, $p(z)$ can be determined classically efficiently, weak simulation implies that one can classically sample form the exact output probability distribution [16] (see also the Supplemental Material (SM) [24]).

There are several ways to elevate the computational power of MGC to the one of a universal quantum computer. To review them, we consider here and throughout the paper an arbitrary, universal circuit, U, of width m (number of qubits it is acting on) and size $poly(m)$. U can be decomposed into single qubit and $r = \Omega(poly(m))$ controlled two-qubit phase gates, CZ. Slightly modifying the encoding used in [20,30] it is easy to show that such a circuit can be encoded into a circuit U' of width $n = 2m$, which is composed out of $poly(m)$ n.n. MGs and r resourceful n.n. CZ gates. This can be easily seen using the freely swappable logical states $|00\rangle$ and $|11\rangle$, the encoding of the single qubit gates A_i, $G(A, A)_{2i-1,2i}$, and $\mathrm{CZ}_{2i,2i+1}$ gates acting between logical states (see the SM [24]). Here and in the following, the subscript denotes the qubit(s) the gate is acting on. Thus, supplementing MGCs with n.n. CZ gates leads to universal quantum computation.

The resourceful gate can also be deterministically implemented via gate teleportation using magic states and adaptive measurements. As shown in [15], any non-Gaussian fermionic state is a magic state for MG computations, i.e., is resourceful. For instance, the magic state, $|M\rangle = \mathrm{CZ}_{2,3}|\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4}$ can be utilized together with adaptive measurements to implement the CZ gate deterministically[2] [15]. Crucial for our approach will be that supplementing MGCs with any of these ingredients separately, i.e., magic states with at most $\mathcal{O}(\log(n))$ adaptive measurements or adaptive measurements with at most $\mathcal{O}(\log(n))$ magic states, remains classically efficiently simulable [23].

Using the results summarized above we will now derive verification circuits, for which the erroneous output can be simulated classically efficiently[3]. Let us first explain two methods to map the fixed but arbitrary encoded circuit, U', to a slightly modified classically simulable circuit V_U (see Fig. 1). Starting with U' and replacing each of the CZ gates by a fSWAP gate, i.e., by CZ · SWAP, leads to a classically efficiently simulable circuit. Albeit this is a very simple mapping, it is clear that exchanging the CZ gates with fSWAP gates is a drastic change in the computation. A more sophisticated mapping in which the circuit is exactly the same, but

[2]Let us stress here that using this magic state, the CZ gate can also be implemented on non-n.n. qubits (see the SM [24]).

[3]Note that, in [11] a similar approach has been presented for Clifford gates and the measurement of a single qubit. However, there, the single qubit reduced states are completely mixed (or factorize). The here proposed method might, however, be applicable to Clifford circuits.

the gates are applied to a slightly different input state is the following. As explained above, U' can be realized via an adaptive circuit composed out of MGs applied to the input state $|0\rangle^{\otimes n}|M\rangle^{\otimes r}$. Considering exactly the same circuit, including adaptive measurements (using exactly the same correction operators, or modified ones to implement, e.g., deterministically the fSWAP gate, see the SM [24]), but applied to $|\Psi_{\rm in}\rangle = |0\rangle^{\otimes n}|M'\rangle^{\otimes r}$ with $|M'\rangle = \mathsf{CZ}_{2,3}|\Phi^+\rangle_{1,3}|\Phi^+\rangle_{2,4}$ leads to a circuit which is classically efficiently simulable. The reason for that is that the state $|M'\rangle$ can be generated with MGs (in contrast to the state $|M\rangle$) and that adaptive measurements on those states remain classically efficiently simulable [19,23].

We show next how the circuit can be transformed into one which allows for the efficient simulation of the erroneous realization of V_U. To this end, we employ the notion of randomized compiling (RC) [17]. RC does not only lead to a more robust implementation of the circuit, but, as we will show, allows us to tailor the output of the quantum computation to a state whose output probability distribution can be sampled from classically efficient. That is, we show now that we can weakly simulate the output of the randomly compiled, erroneous realization of V_U. Using statistical tests, such as the Kolmogorov-Smirnov (KS) [31] or the Epps-Singleton (ES) [32] test (see below), enables us to compare the samples and to gain confidence that they stem from the same probability distribution. Altogether, this allows us to estimate the error parameter(s) of the randomly compiled computation.

We will make the following assumptions on the error model: (i) Instead of a MG, M, the map Λ_M is implemented, where $\Lambda_M(\rho) = \mathcal{E}_M(M\rho M^\dagger)$ and the error $\mathcal{E}_M$ can depend on M, but is assumed to be Markovian; (ii) Pauli operators can be implemented with negligible error; (iii) a measurement with projectors $\Pi_t, t \in \{0, 1\}$ is modeled by $\Pi_t\mathcal{E}(\cdot)\Pi_t$; (iv) for any MG, M and any k-fold Pauli operator P, it holds that $\Lambda_{M(P)} = \mathcal{E}_M(M(P)\cdot M(P)^\dagger)$, where $M(P) = PMP$. Here, $\mathcal{E}_M$ depends on M, but is independent of P. Additionally, we assume that any error channel acts on at most $\mathcal{O}(\log(n))$ qubits and that the initial state $|0^n\rangle$ can be prepared perfectly. The first three assumptions are not very stringent and are commonly used [17,33]. To see that assumption (iv) is justified, note that any MG is up to local phase gates ($e^{i\alpha_i Z}$) of the form $e^{i\beta XX + i\gamma YY}$ [34]. Thus, for any Pauli operator P, acting on arbitrary many qubits, we have $PM(P)P = M$, where the local and nonlocal parts of M and $M(P)$ coincide up to changing the signs of the phases (α_i, β, and γ) randomly, which justifies assumption (iv). It follows that error models $\mathcal{E}_M$ depending on the absolute value of the mentioned angles do satisfy assumption (iv). This encompasses coherent errors represented by over-rotations in the form of $e^{\pm i|\alpha_i|\epsilon Z}$, $e^{\pm i|\beta|\epsilon XX}$, or $e^{\pm i|\gamma|\epsilon YY}$. Additionally, it includes stochastic errors where each over-rotation occurs with a certain probability. Let us conclude this discussion on the error model by noting that our assumptions can be relaxed, such that the noisy state remains simulable. This can be achieved in two ways: First, after randomized compilation, any Pauli channel would be admissible, provided the total number of parameters is at most $poly(n)$ (including some errors that are correlated in time). Second, one can allow for errors that are convex combinations of MGs, e.g., certain over-rotation errors, as this remains simulable.

Next, we show that under these assumptions on the error model, one can depolarize the error of any MGC to a Pauli channel. For each MG, M_i we choose a random Pauli operator, $P_i \in \mathcal{P}_n$, and apply the gate sequence $P_iM_i(P_i)P_i$. In the error-free case we obtain the final pure state $\prod_{i=1}^s P_iM_i(P_i)P_i|0\cdots0\rangle$ which coincides with the ideal state $\prod_{i=1}^s M_i|0\cdots0\rangle$ as $P_iM_i(P_i)P_i = M_i$ for each i. To analyze the erroneous case we consider first a single gate: M_i with corresponding error channel $\mathcal{E}_i$. As shown in the SM [24], $\mathcal{E}_i(\cdot)$ is transformed to a Pauli channel $\mathcal{S}_i$, i.e., $|\mathcal{P}_n|^{-1}\sum_{P_i} P_i\mathcal{E}_M(P_i\cdot P_i)P_i = \mathcal{S}_i(\cdot)$. Concatenating the channels for the whole circuit leads to the output state

$$\rho_{\rm exp} = \sum_{\{P_i\}} c_1(P_1)\cdots c_s(P_s)W(P_1,\ldots,P_s), \tag{1}$$

where $W(P_1,\ldots,P_s)$ denotes the projector onto the state

$$\begin{aligned}|\psi(P_1,\ldots,P_s)\rangle &= P_sM_s\cdots P_2M_2P_1M_1|0^n\rangle \\ &= P'_sM_s(P'_{s-1})\cdots M_2(P'_1)M_1|0^n\rangle \end{aligned}\tag{2}$$

with $P'_k = P_kP_{k-1}\cdots P_2P_1$ for $k = 1, 2, \ldots, s$. Note that the output probability distribution of each pure state $W(P_1,\ldots,P_s)$ can be weakly simulated and the coefficients $c_i(P)$ can be measured experimentally via gate tomography (for single MGs). Moreover, errors which occur during intermediate measurements can be similarly taken into account by using the fact that during the computation the qubits are only measured in the computation basis. Hence, only bit-flip errors, which can be applied to the classical output, have to be taken into account (phase-flip errors commute with the measurement). Taking also the measurement errors into account (see the SM [24]), the output state has a similar form as $\rho_{\rm exp}$ and can therefore be weakly simulated.

Running the verification circuits on the quantum computers gives us a sample $\{y_1,\ldots y_k\}$ that we want to compare with the output $\{z_1,\ldots z_l\}$ of the weak simulation, where $k, l \in poly(n)$. Then, using the KS [31] or ES [32] test (see the SM [24]) allows one to gain confidence that the two samples stem from the same (unknown) distribution by computing a distance between their empirical distribution functions. Up to our knowledge, these tests are among the most widely used tests for the two-sample problem.

The protocols. Our protocols aim to verify a realization U' of an arbitrary universal quantum computation U. To obtain U', one decomposes U into single qubit and CZ gates and maps those to n.n. MGs and CZ gates in U'. Furthermore, one could implement each CZ gate in U' by consuming one copy of the magic state $|M\rangle$ and adaptive measurements. Our protocols can then be summarized as follows: 1. Use one of the options to construct a classically efficiently simulable verification circuit: 1a. Replace each gate CZ by CZ $\cdot$ SWAP = fSWAP (or any other MG) to obtain a MG circuit. 1b. Consider the realization of the CZ gates via the magic state $|M\rangle$ and adaptive measurements. Apply the exact same computation to the input state where each magic state $|M\rangle$ is replaced by the state $|M'\rangle = \mathsf{fSWAP}_{2,3}|\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4} = \mathsf{CZ}_{2,3}|\Phi^+\rangle_{1,3}|\Phi^+\rangle_{2,4}$ (or any other resourceless state). Note

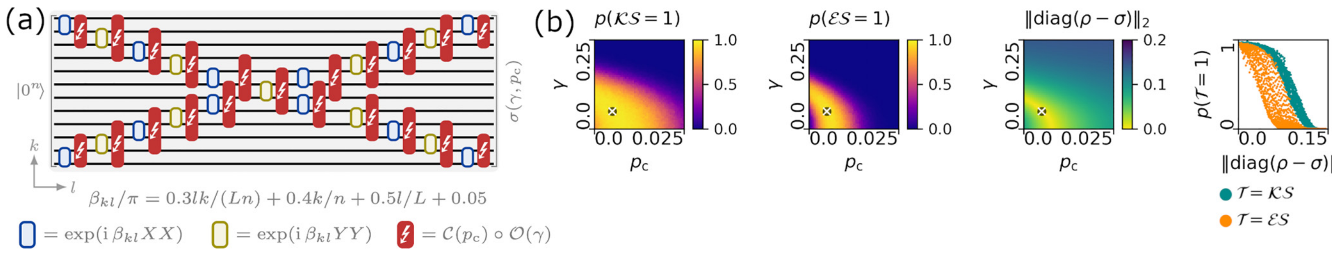

FIG. 2. (a) The MG circuit ($n = 12$ qubits, depth $L = 11$) we use to illustrate our method. To each gate $\exp(\mathrm{i}\beta H)$, we associate (i) a coherent overrotation $\mathcal{O}$ corresponding to $\exp(\mathrm{i}\gamma|\beta|H)$ and (ii) a stochastic crosstalk channel $\mathcal{C}$, which is constructed by concatenating in total four channels $\tilde{\mathcal{C}}(\rho) = (1 - p_c)\rho + p_c/4\sum_{P\in\mathcal{X}} P\rho P$, for each pair of targeted qubit and its neighbors, where $\mathcal{X} = \{XX, XY, YX, YY\}$ [35] (for details and additional examples, see the Supplemental Material [24]). In (b), we compare a reference erroneous output state $\rho \equiv \sigma(\gamma = 0.05, p_c = 0.005)$ (marked by the crosses) with various states $\sigma(\gamma, p_c)$. We find that the relation between the probability of passing the tests, $p(\mathcal{KS} = 1)$ and $p(\mathcal{ES} = 1)$, and $\|\operatorname{diag}(\rho - \sigma)\|_2$ behaves similarly across various examples.

that if there were no errors, both circuits can be simulated classically efficiently and can be directly compared to the output of the quantum computation. 2. Estimate the error parameters $c_i(\cdot)$ in Eq. (1) for each gate and measurement through process tomography. 3. Run the randomly compiled version of the verification circuits. Due to RC, the output state is given by Eq. (1). Confidence on the output and error of the whole circuit is gained by comparing the samples obtained from the quantum computer with those obtained via weak classical simulation thereof using, e.g., the KS or the ES test.

Before illustrating the performance of these verification protocols, the following comments are in order. First, we assess here the performance of a specific realization U' of the computation U (see Fig. 1). Hence, this provides a lower bound on the quality of the quantum device in realizing this particular computation. Second, one could use other gate replacements to make a circuit (including RC) classically efficiently simulable. For example, one could decompose a circuit into MGs and Hadamard gates, H [30], and replace the latter with, e.g., the MGs $G(H, H)$. Third, to improve the tests and to make the verification circuits even more similar to the original circuit, one can, e.g., use that $\mathcal{O}(\log(n))$ magic states can still be classically efficiently simulated [23]. Moreover, the probability with which the final state is in the even parity subspace gives additional information on the error. Fourth, the KS test can be applied to various mappings of the sampled bitstrings to numbers. Finally, using the obtained samples, various additional tests can be performed. To check whether our protocols could distinguish two error models at all, one can compare the outputs of two entirely classical simulations. This may also be used to determine the necessary sample sizes.

Illustration of the method. To illustrate the capability of our method to distinguish errors, we compare samples drawn from two classical simulations, one of which takes the place of the quantum computation. We specify the error of the latter and determine how much another set of errors must differ to be distinguished successfully. We consider the MG circuit depicted in Fig. 2(a), which includes physically relevant crosstalk and overrotation errors. Aiming to analyze the quality of the protocol proposed here, we compute the full density matrices of the output states ρ and σ, and calculate a distance between their probability distributions $\operatorname{diag}(\rho)$ and $\operatorname{diag}(\sigma)$. With $\mathcal{KS}(\rho, \sigma, M, \alpha)$, we denote the random variable taking the value 0 if the KS test rejects the null hypothesis $\operatorname{diag}(\rho) \approx \operatorname{diag}(\sigma)$ using M measurement shots with significance level α[4], and 1 otherwise. Likewise, we define $\mathcal{ES}$ for the ES test. To estimate $p(\mathcal{KS} = 1)$, we repeatedly (here, 1000 times, using $M = 400$ and $\alpha = 0.05$) apply the test to newly sampled data and estimate the expectation value $\mathbf{E}(\mathcal{KS} = 1)$. In our examples (see Fig. 2(b) and the SM [24], where we demonstrate, e.g., detection of error drift), the states ρ and σ can be distinguished by the statistical tests with high probability as soon as $\|\operatorname{diag}(\rho - \sigma)\|_2 \geqslant d_{\mathrm{crit}} \approx 0.1$. In such cases, the error model is rejected. Note that d_{crit} can be improved by, e.g., increasing M or applying the circuit twice. To demonstrate the applicability of our methods, we include in [24] numerical results for different types of circuits (brickwall layout) acting on up to 40 qubits. We also sometimes find that postprocessing the samples increases the power of the tests (see the SM [24]).

In the future, it would be intriguing to see how our protocols can be combined with existing verification approaches, also in the context of fermionic quantum computation [36] and quantum simulation. Furthermore, it would be interesting to analyze whether our methods could be useful in the case of multiqubit measurements in Clifford circuits (see [11]), where the output can be weakly simulated [37]. However, symmetries of stabilizer states potentially hiding errors and possibly uniformly random measurement outcomes in the error-free case are obstacles one needs to overcome here.

Simulation codes and input parameters are available on Zenodo [38].

Acknowledgments. We thank Richard Jozsa and Richard Kueng for helpful discussions. J.C., A.N., M.L., and B.K. acknowledge financial support from the Austrian Science Fund (FWF) through the grants SFB BeyondC (Grant No. F7107-N38) and P 32273-N27 (Stand-Alone Project). JC, ML and BK acknowledge the support of the Munich Quantum Valley, which is supported by the Bavarian State

[4]The significance level α bounds the probability that the test rejects the null hypothesis, given that it is true.

Government with funds from the Hightech Agenda Bayern Plus, Teilprojekts K4 "Theoretical Quantum Computing" and K8 "Hardware Adapted Theory". JC acknowledges financial support from the research project BMBF (MuniQC-Atoms, DAQC). ML and BK acknowledge funding from the BMW endowment fund and the Horizon Europe programmes HORIZON-CL4-2022-QUANTUM-02-SGA via the project 101113690 (PASQuanS2.1) and HORIZON-CL4-2021-DIGITAL-EMERGING-02-10 under grant agreement No. 101080085 (QCFD).

[1] U. Mahadev, Classical verification of quantum computations, in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (Paris, 2018) pp. 259–267.

[2] Z. Brakerski, V. Koppula, U. Vazirani, and T. Vidick, Simpler proofs of quantumness, in *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, Leibniz International Proceedings in Informatics (LIPIcs), edited by S. T. Flammia (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020), Vol. 158, pp. 8:1–8:14.

[3] Z. Brakerski, P. Christiano, U. Mahadev, U. Vazirani, and T. Vidick, A cryptographic test of quantumness and certifiable randomness from a single quantum device, Journal of the ACM **68**, 1 (2021).

[4] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).

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

[6] D. Lu, H. Li, D.-A. Trottier, J. Li, A. Brodutch, A. P. Krismanich, A. Ghavami, G. I. Dmitrienko, G. Long, J. Baugh, and R. Laflamme, Experimental estimation of average fidelity of a clifford gate on a 7-qubit quantum processor, Phys. Rev. Lett. **114**, 140505 (2015).

[7] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[8] E. Onorati, A. H. Werner, and J. Eisert, Randomized benchmarking for individual quantum gates, Phys. Rev. Lett. **123**, 060501 (2019).

[9] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Scalable randomized benchmarking of quantum computers using mirror circuits, Phys. Rev. Lett. **129**, 150502 (2022).

[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, 2010).

[11] R. Jozsa and S. Strelchuk, Efficient classical verification of quantum computations, arXiv:1705.02817 [quant-ph].

[12] S. Ferracin, T. Kapourniotis, and A. Datta, Accrediting outputs of noisy intermediate-scale quantum computing devices, New J. Phys. **21**, 113038 (2019).

[13] N. Schuch, M. M. Wolf, K. G. H. Vollbrecht, and J. I. Cirac, On entropy growth and the hardness of simulating time evolution, New J. Phys. **10**, 033032 (2008).

[14] A. Elben, B. Vermersch, R. van Bijnen, C. Kokail, T. Brydges, C. Maier, M. K. Joshi, R. Blatt, C. F. Roos, and P. Zoller, Cross-platform verification of intermediate scale quantum devices, Phys. Rev. Lett. **124**, 010504 (2020).

[15] M. Hebenstreit, R. Jozsa, B. Kraus, S. Strelchuk, and M. Yoganathan, All pure fermionic non-Gaussian states are magic states for matchgate computations, Phys. Rev. Lett. **123**, 080503 (2019).

[16] M. Van Den Nest, Simulating quantum computers with probabilistic methods, Quantum Inf. Comput. **11**, 784 (2011).

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

[18] L. G. Valiant, Quantum circuits that can be simulated classically in polynomial time, SIAM J. Comput. **31**, 1229 (2002).

[19] B. M. Terhal and D. P. DiVincenzo, Classical simulation of noninteracting-fermion quantum circuits, Phys. Rev. A **65**, 032325 (2002).

[20] R. Jozsa and A. Miyake, Matchgates and classical simulation of quantum circuits, Proc. R. Soc. A **464**, 3089 (2008).

[21] R. Jozsa, B. Kraus, A. Miyake, and J. Watrous, Matchgate and space-bounded quantum computations are equivalent, Proc. R. Soc. A **466**, 809 (2010).

[22] D. J. Brod, Efficient classical simulation of matchgate circuits with generalized inputs and measurements, Phys. Rev. A **93**, 062332 (2016).

[23] M. Hebenstreit, R. Jozsa, B. Kraus, and S. Strelchuk, Computational power of matchgates with supplementary resources, Phys. Rev. A **102**, 052604 (2020).

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.L032074 for more details on MGCs, randomized compilation for MGCs including measurements and classical simulation thereof, an overview on the hypothesis tests we use, and a selection of additional numerical demonstrations, which also includes Refs. [25–29].

[25] J. E. Walsh, Bounded probability properties of Kolmogorov-Smirnov and similar statistics for discrete data, Ann. Inst. Stat. Math. **15**, 153 (1963).

[26] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[27] N. Smirnov, Table for estimating the goodness of fit of empirical distributions, Ann. Math. Stat. **19**, 279 (1948).

[28] D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms* (Addison-Wesley, 1997).

[29] C. Spee, K. Schwaiger, G. Giedke, and B. Kraus, Mode entanglement of Gaussian fermionic states, Phys. Rev. A **97**, 042325 (2018).

[30] D. J. Brod and E. F. Galvão, Extending matchgates into universal quantum computation, Phys. Rev. A **84**, 022310 (2011).

[31] A. Kolmogorov, Sulla determinazione empirica di una legge di distributione, Giornale dell' Istituto Italiano degli Attuari **4**, 83 (1933).

[32] T. Epps and K. J. Singleton, An omnibus test for the two-sample problem using the empirical characteristic function, J. Stat. Comput. Simul. **26**, 177 (1986).

[33] T. Proctor, S. Seritan, E. Nielsen, K. Rudinger, K. Young, R. Blume-Kohout, and M. Sarovar, Establishing trust in quantum computations, arXiv:2204.07568.

[34] A. Soeda, S. Akibue, and M. Murao, Two-party locc convertibility of quadpartite states and Kraus–Cirac number of two-qubit unitaries, J. Phys. A: Math. Theor. **47**, 424036 (2014).

[35] S. Heußen, L. Postler, M. Rispler, I. Pogorelov, C. D. Marciniak, T. Monz, P. Schindler, and M. Müller, Strategies for a practical advantage of fault-tolerant circuit design in noisy trapped-ion quantum computers, Phys. Rev. A **107**, 042422 (2023).

[36] A. Montanaro and S. Stanisic, Error mitigation by training with fermionic linear optics, arXiv:2102.02120.

[37] R. Jozsa and M. Van Den Nest, Classical simulation complexity of extended Clifford circuits, Quantum Inf. Comput. **14**, 633 (2014).

[38] Code and setup for the simulations is available at https://zenodo.org/record/8364056.