nature communications

Article

https://doi.org/10.1038/s41467-022-34279-5

Cross-platform comparison of arbitrary quantum states

Received: 28 March 2022

Accepted: 18 October 2022

Published online: 04 November 2022

Check for updates

D. Zhu[1,2,3,4,9], Z. P. Cian[1,2,5,9] ✉, C. Noel[1,2,5,6,7], A. Risinger[1,2,3], D. Biswas[1,2,5], L. Egan[1,2,5], Y. Zhu[1,5], A. M. Green[1,5], C. Huerta Alderete[1,5], N. H. Nguyen[1,5], Q. Wang[1,2,8], A. Maksymov[4], Y. Nam[4,5], M. Cetina[1,2,5,6], N. M. Linke[1,5], M. Hafezi[1,2,3,5] & C. Monroe[1,2,4,5,6,7]

As we approach the era of quantum advantage, when quantum computers (QCs) can outperform any classical computer on particular tasks, there remains the difficult challenge of how to validate their performance. While algorithmic success can be easily verified in some instances such as number factoring or oracular algorithms, these approaches only provide pass/fail information of executing specific tasks for a single QC. On the other hand, a comparison between different QCs preparing nominally the same arbitrary circuit provides an insight for generic validation: a quantum computation is only as valid as the agreement between the results produced on different QCs. Such an approach is also at the heart of evaluating metrological standards such as disparate atomic clocks. In this paper, we report a cross-platform QC comparison using randomized and correlated measurements that results in a wealth of information on the QC systems. We execute several quantum circuits on widely different physical QC platforms and analyze the cross-platform state fidelities.

Cross-platform quantum state comparisons are critical in the early stages of developing QC systems, as they may expose particular types of hardware-specific errors and also inform the fabrication of next-generation devices. There are straightforward methods for comparing generic output from different quantum computers, such as coherently swapping information between them[1–5], and full quantum state tomography[6]. However, these schemes require either establishing a coherent quantum channel between the systems[7], which may be impossible with highly disparate hardware types; or transforming quantum states to classical measurements, requiring resources that scale exponentially with system size.

Recently, a new type of cross-platform comparison based on randomized measurements has been proposed[8,9]. While this approach still scales exponentially with the number of qubits, it has a significantly smaller exponent prefactor compared with full quantum state tomography[10], allowing scaling to larger quantum computer systems.

Here, we demonstrate a cross-platform comparison based on randomized-measurement[8,9,11], obtained independently over different times and locations on several disparate quantum computers built by different teams using different technologies, comparing the outcomes of four families of quantum circuits.

To quantify the comparison, we use the cross-platform fidelity defined as[8,12]

$$\mathcal{F}(\rho_1,\rho_2)=\frac{\mathrm{tr}[\rho_1\rho_2]}{\sqrt{\mathrm{tr}[\rho_1^2]\mathrm{tr}[\rho_2^2]}}, \quad (1)$$

[1]Joint Quantum Institute, University of Maryland, College Park, MD 20742, USA. [2]Center for Quantum Information and Computer Science, University of Maryland, College Park, MD 20742, USA. [3]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA. [4]IonQ, College Park, MD 20740, USA. [5]Department of Physics, University of Maryland, College Park, MD 20742, USA. [6]Duke Quantum Center and Department of Physics, Duke University, Durham, NC 27708, USA. [7]Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA. [8]Chemical Physics Program and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA. [9]These authors contributed equally: D. Zhu, Z. P. Cian. ✉e-mail: zpcian@umd.edu

where ρ_i is the density matrix of the desired N qubits quantum state produced by system i. To evaluate this fidelity, for each system, we first initialize N qubits in the state $|0,0,\ldots,0\rangle$ and apply the unitary V to nominally prepare the desired quantum states on each platform. In order to perform randomized-measurement, we measure the quantum states in M_U different bases. In particular, we sample M_U distinct combinations of random single-qubit rotations $U = u_1 \otimes u_2 \otimes \cdots \otimes u_N$ and append them to the circuit that implements V as shown in Fig. 1a. Finally, we perform projective measurements in the computational basis. For each rotation setting U, the measurements are repeated M_S times("shots") on each platform. We infer the cross-platform fidelity defined in Eq. (1) from the randomized measurement results via either the statistical correlations between the randomized measurements[8] (Protocol I in Method) or constructing an approximate classical representation of a quantum state using randomized measurements, the so-called the classical shadow[11,13] (Protocol II in Method).

We use four ion-trap platforms, the University of Maryland (UMD) EURIQA system[14] (referred to as UMD_1), the University of Maryland TIQC system[15] (UMD_2), and two IonQ quantum computers[16,17] (IonQ_1, IonQ_2), as well as five separate IBM superconducting quantum computing systems hosted in New York, *ibmq_belem* (IBM_1), *ibmq_casablanca* (IBM_2), *ibmq_melbourne* (IBM_3), *ibmq_quito* (IBM_4), and *ibmq_rome* (IBM_5)[18]. See Supplementary Information Sec. S4 for more details of these systems, which includes refs. 14, 18–22.

We first demonstrate the application of randomized measurements for comparing 5-qubit GHZ (Greenberger-Horne-Zeilinger) states[23] generated on different platforms and the ideal 5-qubit GHZ state obtained from classical simulation. Using the same protocol, we also compare states generated with three random circuits of different width and depth, each sharing a similar construction to circuits used in quantum volume (QV) measurements[24].

Results

We first measure the cross-platform fidelity to compare 5-qubit GHZ states. Specifically, the circuit that prepares the GHZ states is appended with a total of $243 = 3^5$ different sets of single-qubit Clifford gates. These appended circuits complete all the measurements needed for quantum state tomography. Each appended circuit is repeated for $M_S = 2000$ shots. We sample $M_U = 100$ out of the 243 different Us to calculate the cross-platform fidelity defined in Eq. (1) (Fig. 1d). We see that our method has good enough resolution to reveal the performance difference between platforms. In Supplementary Information Sec. S2, we benchmark our method against full quantum state tomography by computing the fidelity as a function of M_U. The comparison shows that the fidelity obtained via randomized measurements approaches that obtained via the full quantum state tomography rapidly.

We present cross-platform fidelity results for 7- and 13-qubit QV-like circuits[24]. QV circuits have been studied extensively, both theoretically and experimentally[24–26], making them an ideal choice for the cross-platform comparison. Also, quantum volume provides a single-number metric for the overall performance of a quantum computer. However, in our randomized measurement scheme, we can obtain more information for the state we prepare. In particular, by using the classical post-processing scheme presented in[11], we can estimate many observables from the randomized measurement data. An N-qubit QV circuit consists of $d = N$ layers : each layer contains a random permutation of the qubit labels, followed by random two-qubit gates among every other neighboring pair of qubits. In our study, we call circuits of such construction but different circuit depth d QV-like circuits. Specifically, a QV-like circuit can be written as a unitary operation $V = \prod_{i=1}^{d} V^{(i)}$, where $V^{(i)} = V^i_{\pi_i(N'-1),\pi_i(N')} \otimes \cdots \otimes V^i_{\pi_i(1),\pi_i(2)}$ and $N' = 2\lfloor N/2 \rfloor$. The operation $\pi(a)$ is a random permutation sampled from the permutation group S_N. The unitary operation $V^i_{a,b}$ is a random two-qubit gate acting on qubits a and b and sampled from $SU(4)$. The circuit diagram of an example QV-like circuit is shown in Fig. 2a. In this experiment, we infer the fidelity for 7-qubit QV-like states with $d = 2$ and $d = 3$ and a 13-qubit QV-like state with $d = 2$.

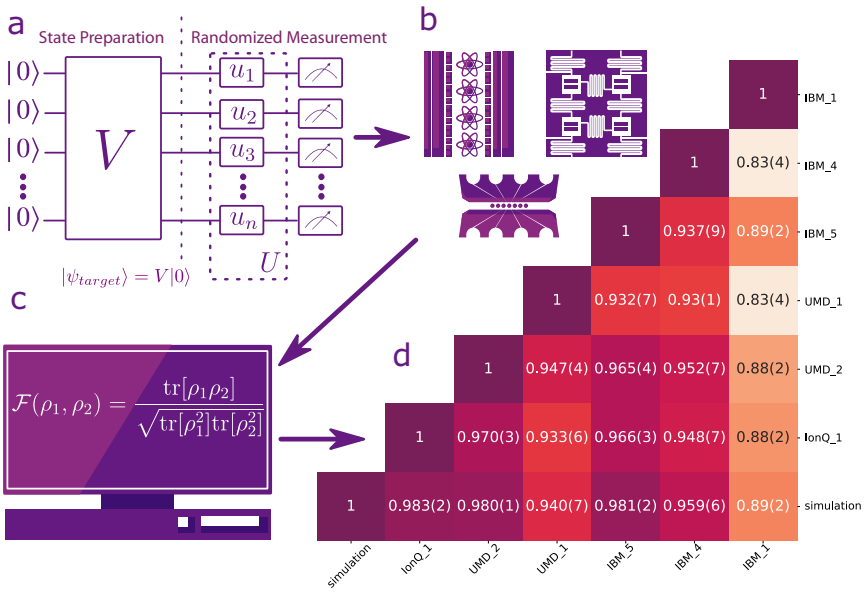

Fig. 1 | Schematic diagram of the cross-platform comparison. a Test quantum circuit, represented by unitary operator V for state preparation, with appended random rotations u_i to each qubit i for measurements in a random (particular) basis. **b** The circuits are transpiled for different quantum platforms into their corresponding native gates. Each of the M_U circuits is repeated M_S times for each platform. **c** The measurement results are sent to a central data repository for processing the fidelities defined in Eq. (1). As an example, **d** The cross-platform fidelity results for a 5-qubit GHZ state, including a row of comparisons between each of the six hardware systems and theory (labeled "simulation"). Entry i, j corresponds to the cross-platform fidelity between platform-i and platform-j. The cross-platform fidelity is inferred from $M_U = 100$ randomized measurements and $M_S = 2000$ repetitions for each U.

https://doi.org/10.1038/s41467-022-34279-5

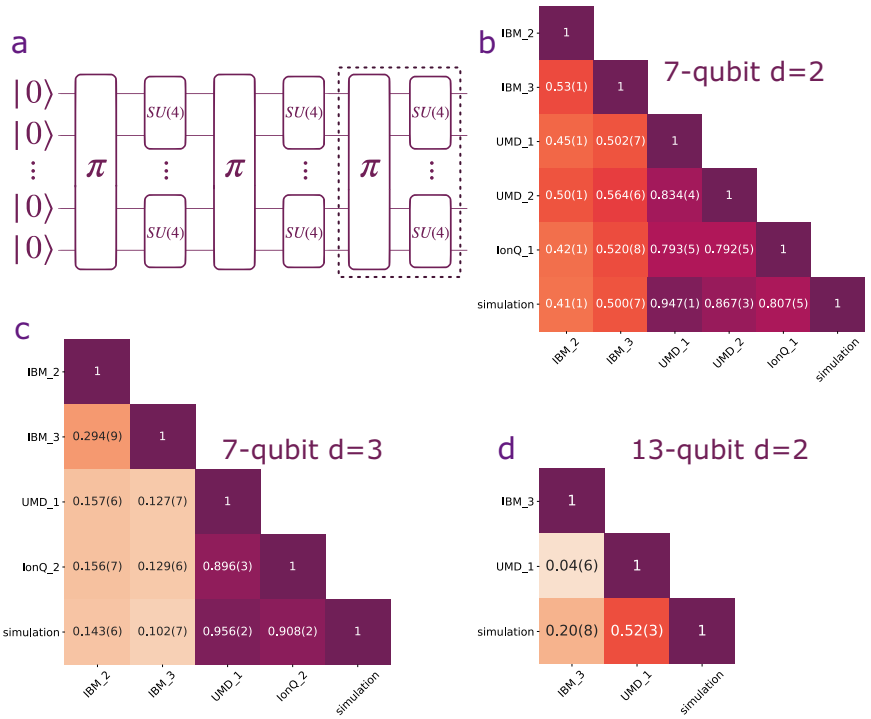

Fig. 2 | The cross-platform fidelity for 7-qubit and 13-qubit QV-like circuit. a The quantum volume circuit diagram for $d = 3$. The $d = 2$ case does not have the operations in the dashed rectangle. **b**–**d** Cross-platform fidelity between different quantum computers. Entry i, j corresponds to the cross-platform fidelity $\mathcal{F}(\rho_i, \rho_j)$ between platform-i and platform-j as defined in Eq. (1). **b** $N = 7$ and $d = 2$; **c** $N = 7$ and $d = 3$; **d** $N = 13$ and $d = 2$.

Similar to the GHZ case, we first distribute the circuits, synthesize them into device-specific native gates, and allow optimizations/error-mitigation that satisfies the aforementioned state-preparation rule.

On each platform, we append the circuit with $M_U = 500$ different Us sampled using the greedy method. Outcomes are measured in the computational basis for $M_S = 2000$ shots. The cross-platform fidelities for $d = 2$ and $d = 3$ are shown in Fig. 2b, c. Our results verify that with only a fraction of the number of measurements required to perform full quantum state tomography, we can estimate the fidelities to sufficiently high precision to be able to see clear differences among them.

We also infer the cross-platform fidelity with a 13-qubit QV-like circuit with $d = 2$. The results are shown in Fig. 2d. Here we use $M_U = 1000$ and $M_S = 2000$, in contrast with the much larger $M_U = 3^{13} = 1{,}594{,}323$ needed for full quantum state tomography.

We find several interesting features by analyzing the cross-platform fidelity of 7-qubit QV-like results. First, we observe that the cross-platform fidelity drops significantly when the number of layers d increases from $d = 2$ to $d = 3$ for the IBM quantum computers. The drop may be due to the restricted nearest-neighbor connectivity of superconducting quantum computers[27], requiring additional SWAP gates overhead for the execution of the permutation gates. In Supplementary Information Sec. S3, we numerically evaluate the number of entangling gates as function of the number of layers d with different connectivity graphs. We see that, according to IBM's native compiler QISKit (see Supplementary Information Sec. S3 and Sec. S6 for measurement error calibration) extra entangling gates are used to perform two-qubit gates for non-nearest-neighbor qubits on superconducting platforms, resulting in extra errors.

The cross-platform fidelity between IBM_2 and IBM_3 is higher than the cross-platform fidelity between either of them and the ion-trap systems (and classical simulation) as shown in Fig. 2c. This motivates us to study whether quantum states generated from different devices tend to be similar to each other if the underlying technology of the two devices is the same. Therefore, we perform a further analysis to investigate this phenomenon, which we refer to as intra-technology similarity.

We first study the fidelity between subsystems of the 7-qubit QV-like states prepared on different quantum computers for both $d = 2$ and $d = 3$. The subsystem fidelity provides a scalable way to estimate the upper bound for the full system fidelity, since the cost of measuring all possible subsystem fidelities of a fixed subsystem size scales polynomially with the full system size. For a given subsystem, we use the same data collected for the full system, but trace out qubits not within the subsystem of interest. The results are presented in Fig. 3a. We observe that the cross-platform fidelity between for all subsystem sizes from the same technology is higher for a given subsystem size.

To further characterize the intra-technology similarity, we perform principal component analysis[28] (PCA) on the randomized measurement data for the 7-qubit quantum volume states with $d = 2$ and $d = 3$ from all the platforms. PCA is commonly used to reduce the dimensionality of a dataset. It has been applied extensively in signal processing such as human face recognition and audio compression. When implementing PCA, we project the dataset onto the first few principal components to obtain lower-dimensional data while preserving as much of the variation as possible.

To prepare the data for PCA, we randomly sample 1000 shots from the randomized measurement data out of $M_U \times M_S = 1{,}000{,}000$ for each platform. We identify the set of Pauli strings whose expectation values can be evaluated using the sample. We then evaluate the expectation value of these identified Pauli strings by taking the average over the samples, and repeat the sampling $N_{\text{sample}} = 500$ times without replacement to make N_{sample} data points in the 4^N dimensional feature space. The feature vectors represent averaged classical shadow of the quantum state generated from the quantum computers[11,29]. We perform a rotation on the feature space and find the first two principal axes, which are the axes that show the two most significant variances on the dataset. Figure 3b shows the projection of the N_{sample} data points to the first two principal axes. We observe that the first principal

https://doi.org/10.1038/s41467-022-34279-5

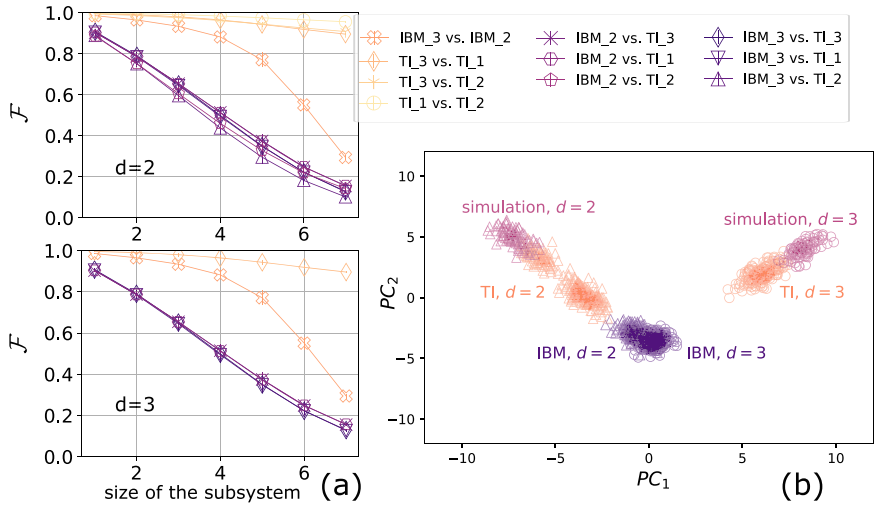

Fig. 3 | The cross-platform fidelity for subsystem and intra-technology similarity. a The cross-platform fidelity between subsystems prepared on different quantum computers. Left : 7-qubit quantum volume circuit of two layers. Right: 7-qubit quantum volume circuit of three layers. The mean for each subsystem size is calculated via bootstrap re-sampling. **b** The projection of randomized measurement dataset onto the first two principal axes, PC_1 and PC_2. Triangle marker is the 7-qubit quantum volume state with $d = 2$. Circle marker is the 7-qubit quantum volume state with $d = 3$. Magenta, orange, and violet correspond to simulation, trapped-ion, and IBM systems respectively.

component separates the two quantum volume states, and the second principal component can distinguish the technology that generates the states. The clustering of the data from the same technology indicates that each technology may share similar noise characteristics that can be distinguished through the cross-platform fidelity and machine-learning techniques.

Discussion

In this manuscript, we experimentally performed the cross-platform comparison of four quantum states allowing the characterization of the quantum states generated from different quantum computers with significantly fewer measurements than those required by full quantum state tomography. To expand our understanding of the intra-technology similarity, more quantum states, in particular those designed to probe the effect of different settings on the cross-platform comparison results, should be studied. Our method could be extended to additional technological platforms such as Rydberg atoms and photonic quantum computers[30]. With the large volume of quantum data generated from the randomized measurement protocol, we have only begun to explore the possibilities that machine learning techniques can offer. We envision extensions of our method will be indispensable in quantitatively comparing near-term quantum computers, especially across different qubit technologies.

Methods

Inference of cross-platform fidelity

Here we briefly introduce the two protocols used for inferring cross-platform fidelity (Eq. 1) from randomized measurements. In Protocol I, we calculate the second-order cross-correlations[8] between the outcomes of the two platforms i and j via the relation

$$\mathrm{Tr}[\rho_i\rho_j] = 2^N \sum_{s,s'} (-2)^{-D[s,s']} \overline{P_U^{(i)}(s) P_U^{(j)}(s')}, \tag{2}$$

where $i,j \in \{1,2\}$, $s = s_1, s_2, \ldots, s_N$ is the bit string of the binary measurement outcomes s_k of kth qubit, $D[s,s']$ is the Hamming distance between s and s', $P_U^{(i)}(s) = \mathrm{Tr}[U\rho_i U^\dagger |s\rangle\langle s|]$, and the overline denotes the average over random unitaries U.

For Protocol II, we reconstruct the classical shadow of the quantum state for each shot of measurement as $\hat{\rho} = \bigotimes_{k=1}^{N} (3u_k^\dagger |s_k\rangle\langle s_k| u_k - I)$, where I is the 2×2 identity matrix[11,13]. The overlap can be calculated as[11]

$$\mathrm{Tr}[\rho_i\rho_j] = \overline{\mathrm{Tr}[\hat{\rho}_i\hat{\rho}_j]}, \tag{3}$$

where $i,j \in \{1,2\}$ and the overline denotes the average over all the experimental realizations. We note that, for both protocols, unbiased estimators are necessary when calculating the purity $i = j$[8,11] using Eqs. (2) and (3).

While the fidelity inferred from the two protocols is identical in the asymptotic limit with $M = M_S \times M_U \to \infty$, the fidelity error inferred from Protocol II converges faster in the number of random unitaries[11]. Therefore, we implement Protocol II for 5- and 7-qubit experiments. However, this protocol is more costly for post-processing. Therefore, for the 13-qubit experiment, we post-process the result with Protocol I.

We explore two different schemes for sampling the single-qubit unitary rotations U, a random method and a greedy method. In the regime $M_S \gg 2^N$, we observe that the greedy method outperforms the random method (see Supplementary Information Sec. S1, which includes refs. 8, 11, 31). Therefore, for $N = 5, 7$, we sample the single-qubit unitary operation with the greedy method. For $N = 13$, we use the random method because to satisfy $M_S \gg 2^N$, the total number of measurements becomes too large. The specified target states and rotations are sent to each platform as shown in Fig. 1b, c. The circuit that implements the specified unitary UV are synthesized and optimized for each platform in terms of its native gates.

When preparing a quantum state on a quantum system, one can perform various error-mitigation and circuit optimization techniques. While these techniques can greatly simplify the circuit and reduce the noise of the measurement outcomes, they can make the definition of state preparation ambiguous. For example, when we prepare a GHZ state and perform the projective measurement in the computational basis, we can defer the CNOT gates right before the measurement to the post-processing, instead of physically applying them. Although one can still obtain the same expectation value for any observable using such a circuit optimization technique, the GHZ state is not actually prepared in the quantum computer. In order to standardize the comparison, in this study, we require that one can perform arbitrary error-mitigation and circuit optimization techniques provided that the target state $|\psi_{target}\rangle = V|0\rangle$ is prepared at the end of the state-preparation stage.

After performing the experiments, the results are sent to a data repository. Finally, we process the results and calculate the cross-platform fidelities. The statistical uncertainty of the measured fidelity is inferred directly from the measurement results via a bootstrap resampling technique[32]. The bootstrap resampling allows us to evaluate the statistical fluctuation of the measurements together with the system performance fluctuation within the duration of the data taking, which is typically two to three days.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Code availability

The code used for the analyses is available from the corresponding author upon request.

References

1. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2011).
2. Shor, P. W. Algorithms for quantum computation: discrete logarithms and factoring. In *Proc. 35th Annual Symposium on Foundations of Computer Science*, 124–134 (IEEE, 1994).
3. Jordan, S. Quantum algorithm zoo. https://quantumalgorithmzoo.org/.
4. Beloy, K. et al. Frequency ratio measurements at 18-digit accuracy using an optical clock network. *Nature* **591**, 564–569 (2021).
5. Buhrman, H., Cleve, R., Watrous, J. & De Wolf, R. Quantum fingerprinting. *Phys. Rev. Lett.* **87**, 167902 (2001).
6. Blume-Kohout, R. Optimal, reliable estimation of quantum states. *New J. Phys.* **12**, 043034 (2010).
7. Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
8. Elben, A. et al. Cross-platform verification of intermediate scale quantum devices. *Phys. Rev. Lett.* **124**, 010504 (2020).
9. Carrasco, J., Elben, A., Kokail, C., Kraus, B. & Zoller, P. Theoretical and experimental perspectives of quantum verification. *PRX Quantum* **2**, 010102 (2021).
10. Anshu, A., Landau, Z. & Liu, Y. Distributed quantum inner product estimation. In Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, 44–51 (2022).
11. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
12. Liang, Y.-C. et al. Quantum fidelity measures for mixed states. *Rep. Progr. Phys.* **82**, 076001 (2019).
13. Aaronson, S. Shadow tomography of quantum states. *SIAM J. Comput.* **49**, STOC18–368 (2019).
14. Egan, L. et al. Fault-tolerant operation of a quantum error-correction code. *Nature* **598**, 281–286 (2021).
15. Zhu, D. et al. Generation of thermofield double states and critical ground states with a quantum computer. *Proc. Natl Acad. Sci. USA* **117**, 25402–25406 (2020).
16. Wright, K. et al. Benchmarking an 11-qubit quantum computer. *Nat. Commun.* **10**, 5464 (2019).
17. Li, M. et al. Generalized hamiltonian to describe imperfections in ion-light interaction. *Phys. Rev. A* **102**, 062616 (2020).
18. Ibm quantum. https://quantum-computing.ibm.com/, 2021.
19. Cross, A. The ibm q experience and qiskit open-source quantum computing software. *APS March Meeting Abstracts* **2018**, L58–003 (2018).
20. Maunz, P. L. W. High optical access trap 2.0. (2016).
21. Debnath, S. et al. Demonstration of a small programmable quantum computer with atomic qubits. *Nature* **536**, 63–66 (2016).
22. Nam, Y., Ross, N. J., Su, Y., Childs, A. M. & Maslov, D. Automated optimization of large quantum circuits with continuous parameters. *npj Quantum Information* **4**, 1–12 (2018).
23. Greenberger, D. M., Horne, M. A. & Zeilinger, A. Going beyond bell's theorem. In *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, 69–72 (Springer, 1989).
24. Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* (2021).
25. Cross, A. W., Bishop, L. S., Sheldon, S., Nation, P. D. & Gambetta, J. M. Validating quantum computers using randomized model circuits. *Phys. Rev. A* **100**, 032328 (2019).
26. Pino, J. M. et al. Demonstration of the qccd trapped-ion quantum computer architecture. *Nature* **592**, 209–213 (2021).
27. Linke, N. M. et al. Experimental comparison of two quantum computing architectures. *Proc. Natl Acad. Sci. USA* **114**, 3305–3310 (2017).
28. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A: Math., Phys. Eng. Sci.* **374**, 20150202 (2016).
29. Huang, H.-Y., Kueng, R., Torlai, G., Albert, V. V. & Preskill, J. Provably efficient machine learning for quantum many-body problems. *Science* **377**, 6613 (2021).
30. Greganti, C. et al. Cross-verification of independent quantum devices. *Phys. Rev. X* **11**, 031049 (2021).
31. Rath, A., van Bijnen, R., Elben, A., Zoller, P., & Vermersch, B. Importance sampling of randomized measurements for probing entanglement. *Phys. Rev. Lett.* **127**, 200503 (2021).
32. Efron, B. & Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36–48 (1983).

Acknowledgements

We acknowledge Andreas Elben, Behtash Babadi, Benoi^t Vermersch and Peter Zoller for helpful discussions. We acknowledge the use of IBM Quantum services; the views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. This work was supported by the ARO through the IARPA LogiQ program (11IARPA1008), the NSF STAQ Program (PHY-1818914), the AFOSR MURIs on Dissipation Engineering in Open Quantum Systems (FA9550-19-1-0399) and Quantum Interactive Protocols for Quantum Computation (FA9550-18-1-0161), the ARO MURI on Modular Quantum Circuits (W911NF1610349), and the U.S. Department of Energy Quantum Systems Accelerator (QSA) Research Center (DE-FOA-0002253) and National Science Foundation QLCI grant OMA-2120757. N.M.L. acknowledges support from the Maryland-Army-Research-Lab Quantum Partnership (W911NF1920181), the Office of Naval Research (N00014-20-1-2695), and the NSF Physics Frontier Center at JQI (PHY-1430094). A.M.G. is supported by a JQI Postdoctoral Fellowship.

Author contributions

D.Z., Z.P.C., C.N., Y.N., M.C., N.M.L., M.H., and C.M. designed the research; D.Z., Z.P.C., C.N., A.R., D.B., L.E., Y.Z., A.M.G., C.H.A., N.H.N., Q.W., and A.M. performed experiments and collected the data; Z.P.C., A.M., and Y.N. compiled and optimized the circuit; D.Z. and Z.P.C. analyzed data; D.Z., Z.P.C., C.N., Q.W., Y.N. N.M.L., M.H., and C.M. contributed to the manuscript, with input from all authors.

Competing interests

C.M., A.M., and Y.N. are affiliated with IonQ. All other authors declare that they have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34279-5.

Correspondence and requests for materials should be addressed to Z. P. Cian.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Supplementary Information : Cross-Platform Comparison of Arbitrary Quantum States

S1 The greedy method in the regime $M_S \gg 2^N$

The parameters M_U and M_S can be optimized through minimizing the statistical error with grid search [1,2] or using the importance sampling with partial information on the quantum state [3]. Both approaches require prior knowledge or simulation of the target state. Here, we devise a greedy method for sampling the unitary operation U that reduces the statistical error without prior knowledge of the target state. The statistical error as a function of M_U converges faster than uniformly sampling the unitary operation when the number of shots $M_S \gg 2^N$, where N is the number of qubits. Therefore, the greedy method is particularly useful for the 5- and 7-qubit experiments. In this section, we demonstrate the comparison between the greedy method and random method for 5-qubit GHZ state.

When performing the fidelity estimation using randomized measurement, there are two major sources of errors, the shot noise error and the error from the incomplete tomography. The shot noise error can be suppressed when the number of shots is $M_S \gg 2^N$. Instead of uniformly sampling the random unitary from a set of unitary operators U, we generate a sequence of unitary operators while maximizing the distance between each random unitary. Specifically, we define the distance between two unitary operators as $d(u_a, u_b) = \max_\rho ||u_a \rho u_a^\dagger - u_b \rho u_b^\dagger||_1$. We sequentially generate the M_U unitary operators $\{u_i\}$, where $1 \leq i \leq M_U$ sequentially. For $i = 1$, we sample a unitary

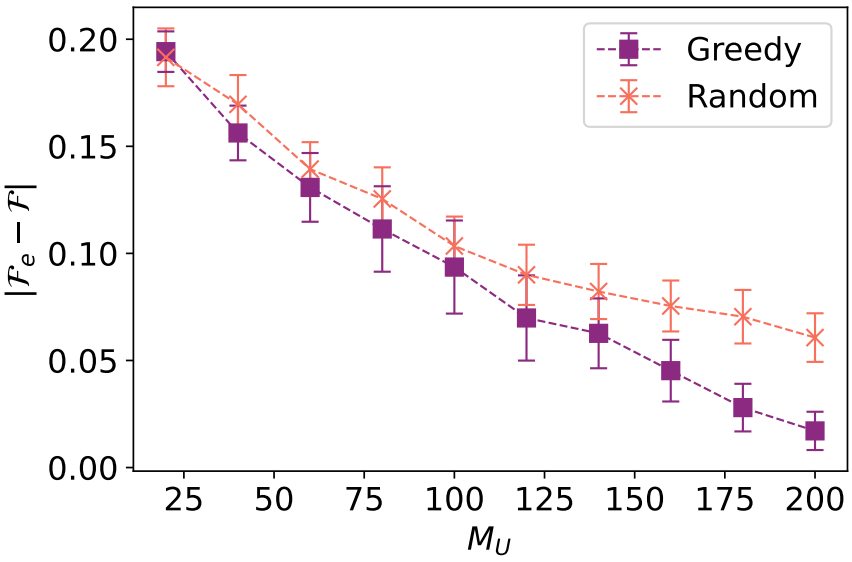

Figure S1: Comparison of error scaling for the fidelity of the GHZ states generated from UMD_1 vs IBM_1 with greedy or random sampling method for M_U.

operator randomly from V. For $i > 1$, we search for a unitary operator u_i that minimizes the cost function $C(u_i; u_1, \ldots, u_{i-1}) = -\sum_{j=1}^{i-1} d(u_i, u_j)$. In order to minimize the cost function efficiently, we randomly generate N_{sample} distinct unitary operators $u_{i,x}$, where $1 \leq x \leq N_{\text{sample}}$ and we define $u_i = \min_{u_{i,x}} C(u_{i,x}; u_1, \ldots, u_{i-1})$. In practice, we find that $N_{\text{sample}} = 200$ is enough to find the minimum for $N = 7$ and $V = Cl(2)^{\otimes N}$, where $Cl(2)$ is the single qubit Clifford group. The greedy method is summarized in Algorithm 1.

Algorithm 1 Greedy method for sampling random unitary

Input : Number of random unitaries M_U, a set of unitary operators S

Output : M_U random unitary operations for randomized measurement $\{u_i\}$, where $1 \leq i \leq M_U$.

1 : Sample u_1 randomly from S.

2 : **for** $i = 2$ **to** M_U **do**

3 : Find a unitary $u_i \in S$ to minimize the cost function $C(u_i; u_1, \ldots, u_{i-1})$.

4 : **end for**

5 : **return** $\{u_i\}$

We compare the two different methods of sampling the random unitary U: the randomized sampling and the greedy method. Using these two methods, we evaluate the fidelity between the states prepared on the UMD_1 system and the IBM_1 system, by sampling subsets of various sizes M_U from the full state tomography measurements. Figure S1 shows the error of the fidelity estimation between UMD_1 and IBM_1 as a function of M_U for $M_S = 2000$. We see that the greedy method outperforms the random method in this regime.

To further characterize the performance of the greedy method, we perform numerical simulation through Pauli basis measurements. Specifically, to generate the random Pauli measurement using the greedy method, we define a set of unitary operators $S = \{H, HSH, I\}$ to perform measurement in the x, y, and z basis. Using Algorithm. 1, we can sample the greedy random unitary operators.

We compare the measurement results for the greedy method and the random sampling method

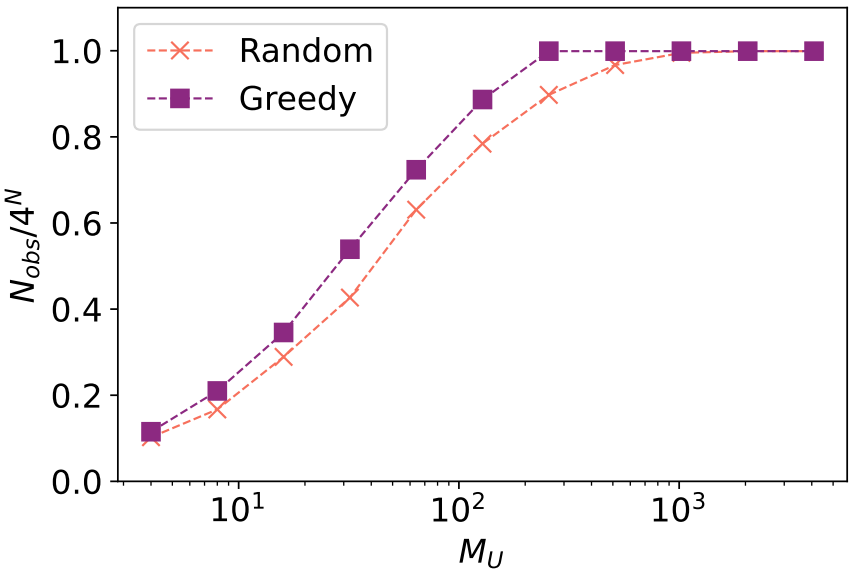

Figure S2: The number of Pauli-string operators that can be predicted through the set of measurements (N_{obs}) as function of M_U. In the large $M_U \gg 3^N$ limit, the measurement is equivalent to full state tomography and therefore, all the Pauli-string operator can be predicted ($N_{obs}/4^N = 1$). However, in the regime $M_U < 3^N$, we show that Greedy method can predict more observables than the random method.

in the regime $M_S \gg 2^N$. First, we compute the number of Pauli-string operators that can be predicted [2] through the set of measurements (N_{obs}). In Fig. S2, we present the N_{obs} as function of M_U, normalized by the total number of Pauli string operators 4^N. We see that the greedy method can predict more observables than the random method in the regime, $M_U < 3^N$. However, when $M_U \gg 3^N$, both greedy and random method can predict all 4^N Pauli string observables. Therefore, the shot noise is the dominant error source.

Second, we perform the simulation for the prediction of the linear observable $\mathrm{tr}(O\rho)$ and

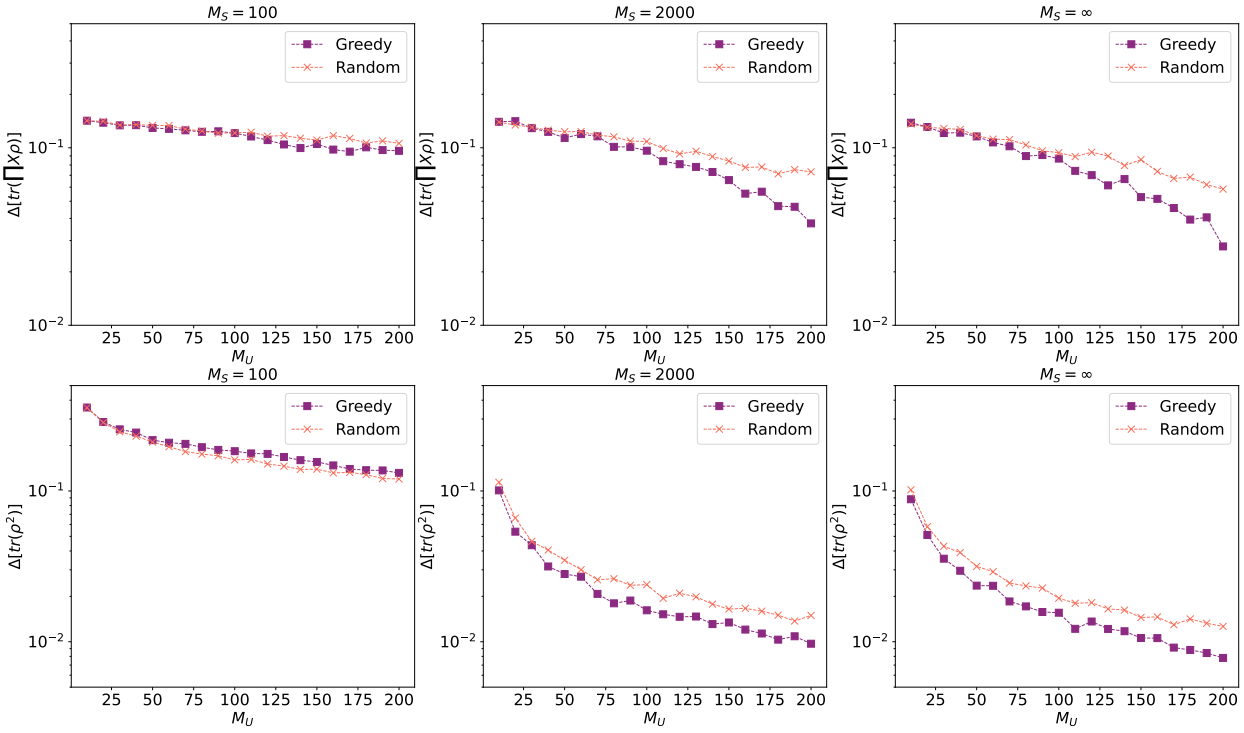

Figure S3: The measurement error Δ for linear observable $tr(O\rho)$, where $O = \bigotimes_{i=1}^{N} X_i$ and non-linear observable $tr(\rho^2)$. In the regime $M_S \sim O(2^N)$, the greedy and the random methods have similar performance. However, when the number of shots $M_S \gg O(2^N)$, the greedy method shows better performance against the randomized measurement method.

non-linear observable $\mathrm{tr}(\rho^2)$ [2], using both the greedy method and the random method. We first generate a random wave function $|\psi\rangle = U|0\rangle$, where U is a unitary operator sampled from a Haar random distribution. We then generate M_U unitary operators via both greedy and random methods and simulate the measurements implied by the unitary operators with M_S shots. In Fig. S3, we present the average error $\Delta[O] = avg(|O_{measure} - O_{exact}|)$, where $O_{measure}$ is the expectation value of $O = \bigotimes_{i=1}^{N} X_i$ using randomized measurements and O_{exact} is the exact expectation value. To obtain the average performance, we perform 100 independent numerical experiments and average

over the error from each numerical experiment.

S2 Full state tomography vs. randomized measurement for 5-qubit GHZ state

Here, we compare the cross-platform fidelity obtained from full-state tomography and that from the randomized measurement on the 5-qubit GHZ state prepared on different platforms. We perform the full-state-tomography on a platform by measuring all of the 243 independent 5-qubit Pauli operators. To do so, we first independently generate the 5-qubit GHZ state circuits on each platform. Then we append different single-qubit rotations to the circuit to create the 243 different circuits. Each of the circuits gives the projective measurement result of one of the 243 independent 5-qubit Pauli operators. We set $M_S = 2000$ for all the platforms. For the randomized measurement, because a random Pauli basis measurement is equivalent to a randomized measurement with single qubit Clifford gates [2], we directly sample from the 243 Pauli basis measurements used for the full state tomography.

We calculate the cross-platform fidelity as a function of the number of randomized measurements M_U. The fidelity error $|\mathcal{F}_e - \mathcal{F}|$ is defined as the difference between the fidelity estimated by the randomized measurement $\mathcal{F}_e$ and the fidelity calculated through full state tomography $\mathcal{F}$. The average of $|\mathcal{F}_e - \mathcal{F}|$ and the standard deviation are calculated through a bootstrap resampling method [4]. The result (Fig. S4) shows that with only a fraction of the full state tomography measurements, one can estimate the cross-platform fidelity accurately.

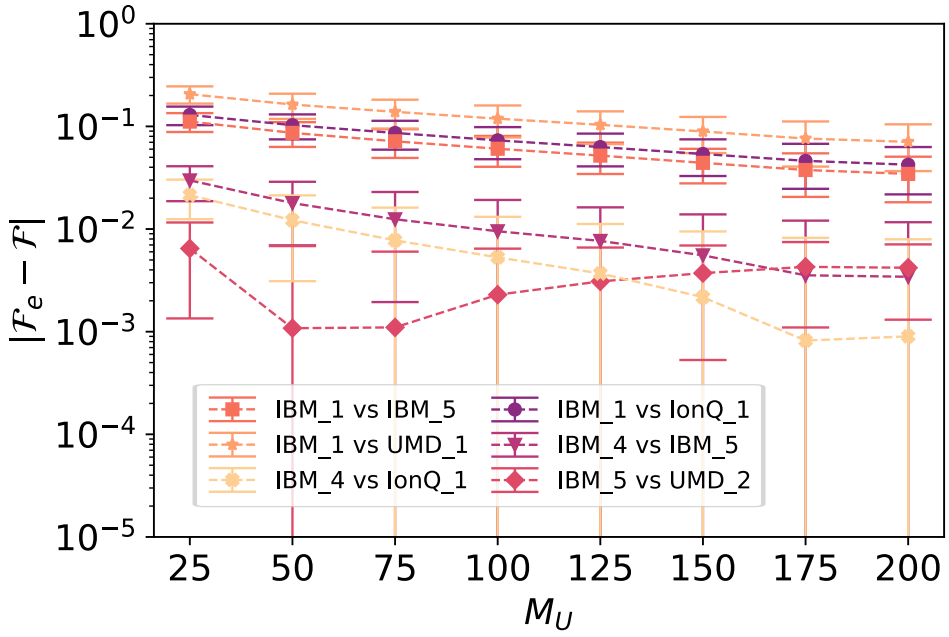

Figure S4: Fidelity error, $|\mathcal{F}_e - \mathcal{F}|$, for six randomly selected 5-qubit GHZ state cross-platform fidelities implemented on different platforms vs. the number of randomized measurements M_U. The number of measurement is $M_S = 2000$ for all cases.

S3 SWAP overhead for quantum volume circuits

Two-qubit gates on non-nearest-neighbor pairs are not directly available on superconducting quantum computers. To realize such non-nearest-neighbor two-qubit gates effectively, SWAP gates are necessary. Note each SWAP gate consists of three CNOT gates. Thus, when used, non-trivial degradation to the overall fidelity of the computational output state is incurred.

Optimizing the so-called qubit routing can effectively decrease the number of involved non-nearest-neighbor two-qubit gates in evaluating the quantum volume circuits. As the number of layers d increases though, the non-nearest-neighbor two-qubit gate becomes unavoidable. In Fig.

S5 we show the mean value of two-qubit gates used to implement quantum volume circuits of d layers on different platforms. The value grows linearly with d.

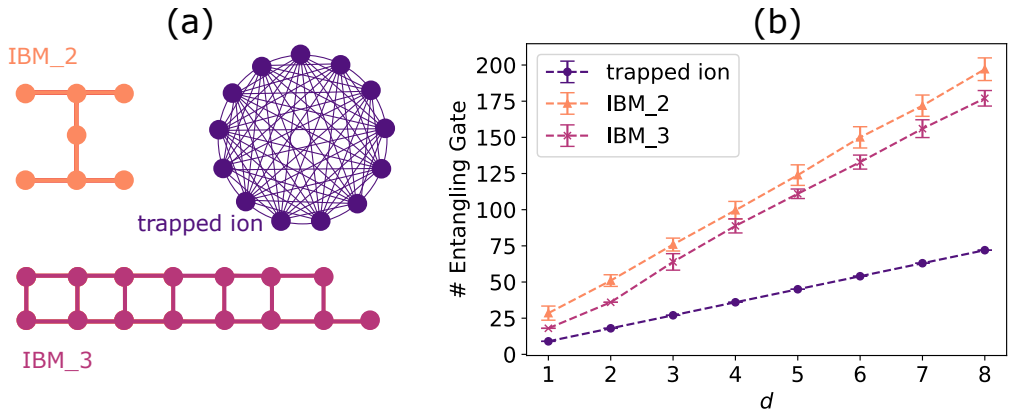

Figure S5: (a) Connectivity graphs of IBM_2, IBM_3, and trapped ion (UMD_1 as an example) (b) Average number of two-qubit (entangling) gates used to implement quantum volume circuits with d layers, on different quantum computers. The trapped ion quantum computers have an all-to-all connectivity.

S4 Quantum systems

In this section we detail the quantum systems used in this study.

IBM Quantum Experience

We use IBM Quantum Experience service to access several of their superconducting quantum computers [5]. The ones we used are *ibmq_belem* (IBM_1), *ibmq_casablanca* (IBM_2), *ibmq_melbourne* (IBM_3), *ibmq_quito* (IBM_4), and *ibmq_rome* (IBM_5). All the IBM systems we used use superconducting transmon qubits. Their gate set is made of arbitrary single-qubit rotations on every

qubit and two-qubit CNOT gates over qubit pairs that are connected according to their respective connectivity graphs. The size of of single-qubit gate errors in the IBM systems ranges from 3.32×10^{-4} to 5.03×10^{-2}, and the size of two-qubit gate errors ranges from 7.47×10^{-3} to 1.07×10^{-1}. Detailed specifications of each quantum device including qubit-connectivity diagrams can be found on (https://quantum-computing.ibm.com/). We used QISKit open-source software [6] for the circuit synthesis and optimization for all of our experiments conducted on the IBM systems.

TI_EURIQA (UMD_1)

Error-corrected Universal Reconfigurable Ion-trap Quantum Archetype (EURIQA) is a trapped-ion quantum computer currently located at the University of Maryland. This quantum computer supports up to 13 qubits in a single chain of 15 trapped $^{171}\mathrm{Yb}^{+}$ ions in a microfabricated chip trap [7]. The system achieves native single-qubit gate fidelities of 99.96% and two-qubit XX gate fidelities of 98.5-99.3%[8]. On this platform, we compile the circuits to its native gate set through KAK decomposition. We optimize the qubit assignment through exhaustive search to minimize the anticipated noise of entangling gates. No SPAM correction was applied in post-processing.

TI_UMD (UMD_2)

The second trapped-ion quantum computer system at Maryland is part of the TIQC (Trapped Ion Quantum Computation) team. This quantum computer supports up to nine qubits made of a single chain of $^{171}\mathrm{Yb}^{+}$ ions trapped in a linear Paul trap with blade electrodes [9]. Typical single- and two-qubit gate fidelities are $99.5(2)\%$ and $98-99\%$, respectively. On this platform, we compile the

quantum volume to its native gate set through KAK decomposition. We apply SPAM correction to mitigate the detection noise assuming that the preparation noise is negligible.

IonQ (IonQ_1 and IonQ_2)

The commercial trapped-ion quantum systems used by IonQ contain eleven fully connected qubits in a single chain of $^{171}Yb^{+}$ ions trapped in a linear Paul trap with surface electrodes [9]. The single-qubit fidelities are 99.7% for both systems at the time of measurement, while two-qubit fidelities are $95-96\%$ and $96-97\%$ for IonQ_1 and IonQ_2 respectively. On this platform, we apply the technique described in Ref. [10] to optimize the circuit. No SPAM correction was applied in post-processing.

S5 Intra-Technology Similarity through Principal component analysis

We consider quantum states reconstructed from the shadow tomography measurement

$$\rho = \frac{1}{T}\sum_{U,b}\hat{\rho}_{U,b}, \tag{S1}$$

where $\hat{\rho} = \bigotimes_{i=1}^{n}(3U_i|b\rangle\langle b|U_i^{\dagger} - I)$ and T is the number of classical shadows. After averaging over T classical shadows, we decompose the density matrix defined in Eq. (S1) as a linear combination of Pauli string operators, $\rho = \sum_{k=0}^{4^n-1} v_k P_k$, where P_k is the Pauli string operator and $v_k = \frac{1}{2^n}tr(\rho P_k)$. Therefore, a density matrix ρ can be represented by a 4^n-dimensional vector $\vec{v} = \sum_{k=0}^{4^n-1} v_k \hat{e}_k$. We define the vector $\vec{v}$ as the feature vector for the principal component analysis.

In the noiseless limit, the vector that represents the target state $|\psi\rangle$ is $\vec{v}_t = \sum_{k=0}^{4^n-1}\langle\psi|P_k|\psi\rangle\hat{e}_k$.

However, in the presence of noise, the vector can deviate from $\vec{v}_t$. Specifically, there are two main sources of noise: the noise from the finite sampling of the classical shadow and the noise from the imperfections in the quantum devices such as coherent and incoherent errors. The first type of noise gives a random fluctuation of the feature vector, which scales as $O(\frac{1}{\sqrt{T}})$. The second type of noise is highly platform specific. We conjecture the intra-technology similarity, i.e. quantum states prepared on similar quantum devices suffer from similar noise: The feature vectors prepared on similar devices cluster together in the 4^n- dimensional space.

In order to observe the intra-technology similarity, we prepare the feature vectors $\vec{v}$ of all the platforms, for the 7-qubit QV states with $d = 2$ and $d = 3$. For each platform and each QV state, we independently sample N_{sample} feature vectors. We define a $N_{\text{sample}} \times 4^n$ data matrix M. Each row of M contains a feature vector $\vec{v}$. We perform the principal component analysis to project the data matrix M onto a lower dimensional space in order to reduce the dimensionality of the feature vector space while preserving as much of the variation of the data as possible. The implementation of principal component analysis is based on scikit-learn. Each low-dimensional vector corresponds to a state reconstructed from the classical shadow. In the main text Figure 3(c), we can see that the low-dimensional feature vectors cluster by platform. In addition, we also perform PCA analysis for 5-qubit GHZ states. The projection result to first two principal components is shown in Fig. S6. We observe the intra-technology similarity since the data generated from similar platforms are clustered together.

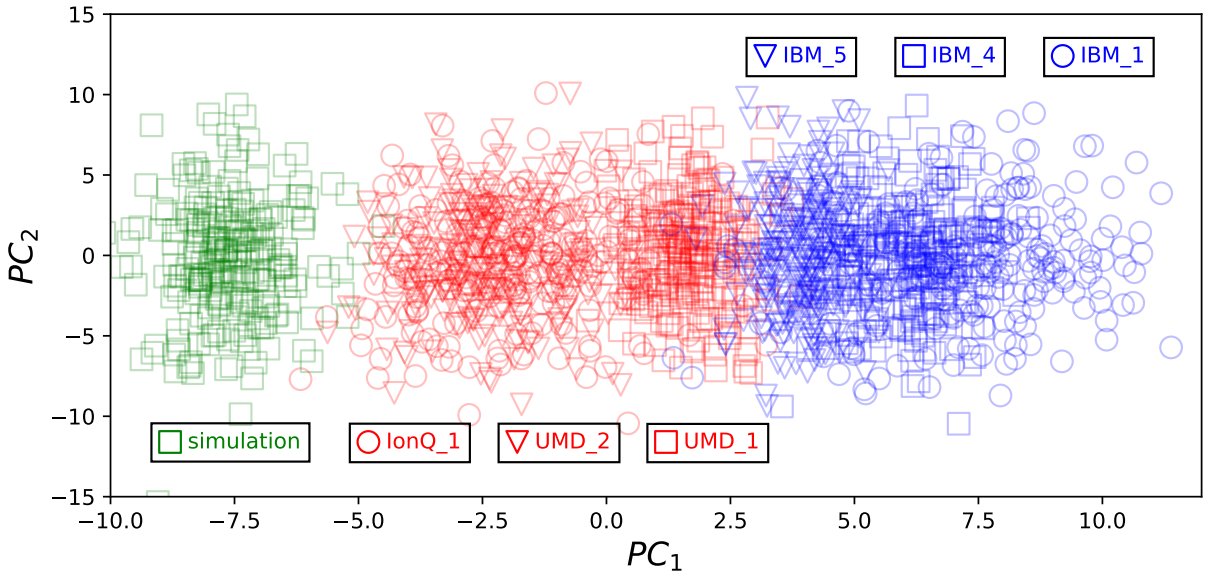

Figure S6: The projection of randomized measurement dataset of the GHZ state onto the first two principal axes, PC_1 and PC_2.

	IBM_2 7q_2l (no calibrated)	IBM_2 7q_2l (calibrated)	IBM_2 7q_3l (no calibrated)	IBM_2 7q_3l (calibrated)
IBM_3	0.53(1)	0.542(7)	0.294(9)	0.30(4)
UMD_1	0.45(1)	0.461(8)	0.157(6)	0.159(5)
UMD_2	0.50(1)	0.511(5)	N/A	N/A
IonQ_1	0.42(1)	0.425(2)	N/A	N/A
IonQ_2	N/A	N/A	0.156(7)	0.161(4)
simulation	0.41(1)	0.418(6)	0.143(6)	0.150(5)

Table S1: The cross-platform fidelity between the platform IBM_2 and other platforms for the result with and without measurement error calibration.

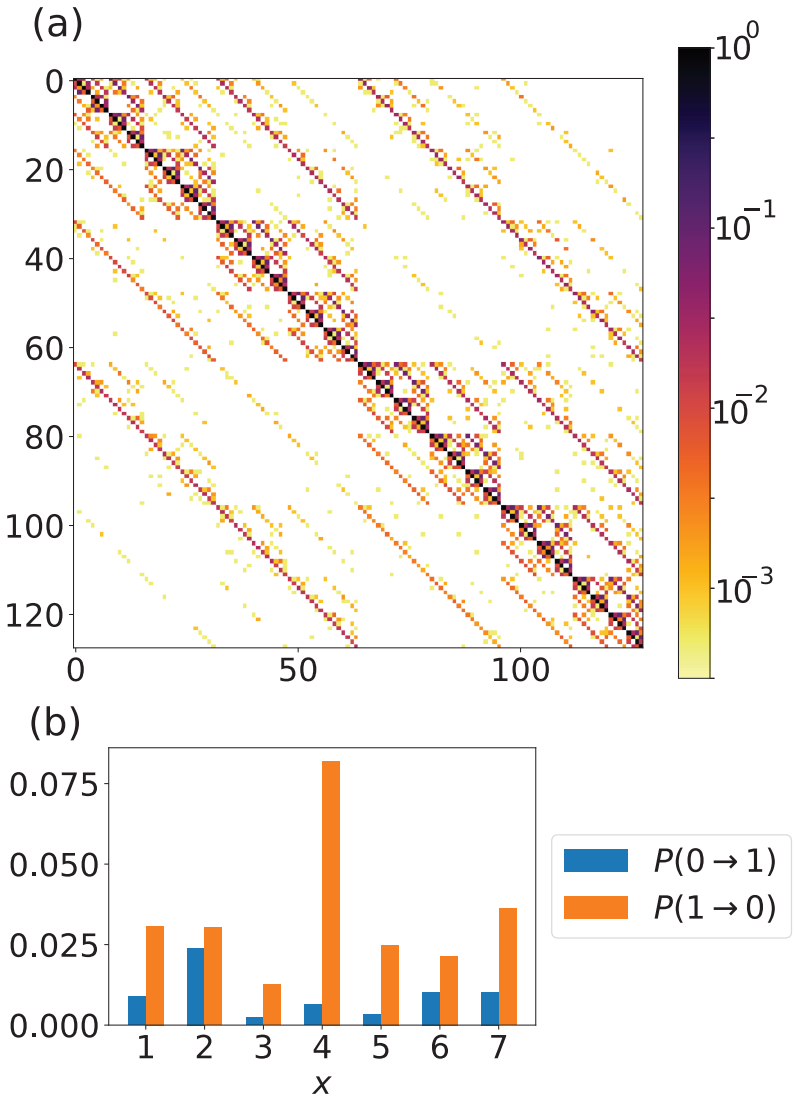

Figure S7: (a) The measurement error matrix for platform IBM_2. (b) The single-qubit measurement error for xth qubit.

S6 Measurement error mitigation

In this section, we introduce the measurement error of and the error mitigation technique used for the IBM superconducting qubits. Measurement error is one of the dominant errors in the superconducting qubit devices. A measurement error manifests itself as either a $|0\rangle$ state being read as a $|1\rangle$ state or vice versa. For a quantum computer with n qubits, the measurement error can be fully described by a $2^n \times 2^n$ measurement error matrix M. The matrix element $M[s, s']$ is the probability of measuring outcome s' when the quantum computer is in state $|s\rangle$. Therefore, if we take the probability vector P_{ideal} describing the ideal measurement results for a given circuit, applying the measurement matrix M gives a good approximation of the results when measurement noise is present. In particular,

$$P_{\text{noisy}} = MP_{\text{ideal}}. \tag{S2}$$

In order to approximate the P_{ideal}, we perform an optimization to minimize the cost function

$$||P_{\text{noisy}} - MP_{\text{ideal}}||_2^2, \tag{S3}$$

subject to constraints $0 \leq P_{\text{ideal}}(s) \leq 1$ and $\sum_s P_{\text{ideal}}(s) = 1$.

For the devices IBM_2 and IBM_3 with seven qubits, we measure the measurement error matrix M by initializing the qubits in all the 2^7 possible bit strings and measure each state with 2048 shots. The measurement error matrix for IBM_2 is shown in Fig. S7(a). The dominant measurement error is the single qubit flip. In particular, the error rate for measuring $|1\rangle$ when the state is $|0\rangle$ ranges from 1% to 8%. The error rate for measuring $|0\rangle$ if the state is $|1\rangle$ ranges from

0.1% to 2%, as shown in Fig. S7(b). We present the calibrated cross-platform fidelity between the platform IBM_2 and other platforms in Table S1. The calibrated result are close to the result without calibration up to the error bar.

Supplementary References

1. Elben, A. *et al.* Cross-platform verification of intermediate scale quantum devices. *Phys. Rev. Lett.* **124**, 010504 (2020).

2. Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics* **16**, 1050–1057 (2020).

3. Rath, A., van Bijnen, R., Elben, A., Zoller, P. & Vermersch, B. Importance sampling of randomized measurements for probing entanglement. *arXiv preprint arXiv:2102.13524* (2021).

4. Efron, B. & Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician* **37**, 36–48 (1983).

5. Ibm quantum. https://quantum-computing.ibm.com/, 2021 .

6. Cross, A. The ibm q experience and qiskit open-source quantum computing software. In *APS March Meeting Abstracts*, vol. 2018, L58–003 (2018).

7. Maunz, P. L. W. High optical access trap 2.0. (2016).

8. Egan, L. *et al.* Fault-tolerant operation of a quantum error-correction code. *arXiv preprint arXiv:2009.11482* (2020).

9. Debnath, S. *et al.* Demonstration of a small programmable quantum computer with atomic qubits. *Nature* **536**, 63–66 (2016).

10. Nam, Y., Ross, N. J., Su, Y., Childs, A. M. & Maslov, D. Automated optimization of large quantum circuits with continuous parameters. *npj Quantum Information* **4**, 1–12 (2018).