Experimental accreditation of outputs of noisy quantum computers

Samuele Ferracin, Seth T. Merkel, David McKay, and Animesh Datta

Phys. Rev. A **104**, 042603 — Published 11 October 2021

DOI: 10.1103/PhysRevA.104.042603

Experimental accreditation of outputs of noisy quantum computers

Samuele Ferracin,[1,2,*] Seth T. Merkel,[3,†] David McKay,[3,‡] and Animesh Datta[2,§]

[1] *Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[2] *Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom*
[3] *IBM Quantum, T.J. Watson Research Center, Yorktown Heights, NY 10598, USA*

(Dated: September 22, 2021)

We provide and experimentally demonstrate an accreditation protocol that upper-bounds the variation distance between noisy and noiseless probability distributions of the outputs of arbitrary quantum computations. We accredit the outputs of twenty-four quantum circuits executed on programmable superconducting hardware, ranging from depth nine circuits on ten qubits to depth twenty-one circuits on four qubits. Our protocol requires implementing the "target" quantum circuit along with a number of random Clifford circuits and subsequently post-processing the outputs of these Clifford circuits. Importantly, the number of Clifford circuits is chosen to obtain the bound with the desired confidence and accuracy, and is independent of the size and nature of the target circuit. We thus demonstrate a practical and scalable method of ascertaining the correctness of the outputs of arbitrary-sized noisy quantum computers—the ultimate arbiter of the utility of the computer itself.

1. Introduction— The utility of noisy quantum computers in simulation and optimisation will be determined by our ability to ascertain if the solutions provided are correct or close to correct. This is a challenging task for problems that are outside the complexity class NP. The current methods are based on evaluating single-valued metrics such as the Cross Entropy [1, 2] or the Quantum Volume [3], which can be linked to the performance of the quantum hardware being used. These methods require simulating the relevant quantum circuits on classical computers. Though practical at present, they are not scalable and consequently useless for problems that cannot already be simulated classically. On the contrary, the proposals based on quantum cryptography and interactive proof systems are scalable in principle, but have an overhead in width (qubits) and depth (gates) that makes them impractical for the foreseeable future [4–11]. This calls for new methods that are both practical in the short term and scalable in the long term.

In this work we present and experimentally demonstrate an Accreditation Protocol (AP) that achieves this goal. This AP provides an upper bound on the variation distance (VD) between the probability distribution of the experimental outputs of a noisy quantum circuit $\{p_{\text{exp}}(\bar{s})\}$ and its ideal, noiseless counterpart $\{p_{\text{ideal}}(\bar{s})\}$, where $\bar{s}$ denotes the bit strings that may be obtained as output. In our AP, the "target" quantum circuit the correctness of whose outputs we wish to ascertain is executed along with a number v of random Clifford circuits (the "traps"). The trap circuits have the same width and depth as the target circuit and are designed such that in the *absence* of noise they always return a fixed known output. This enables us, in the *presence* of noise, to measure the probability p_{inc} that a trap's output is incorrect. Our AP guarantees that the VD between $\{p_{\text{exp}}(\bar{s})\}$ and $\{p_{\text{ideal}}(\bar{s})\}$ is bounded as (see section 1 of the Appendix for a proof)

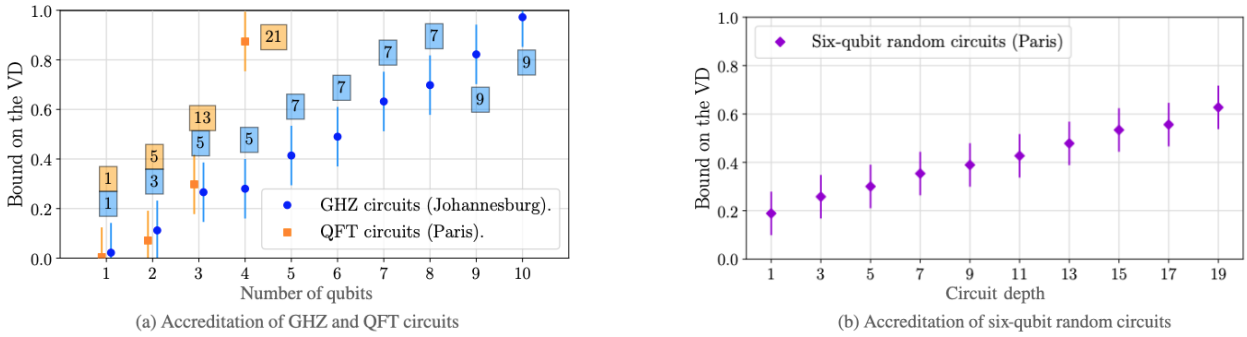
$$\text{VD} := \frac{1}{2}\sum_{\bar{s}} \left|p_{\text{ideal}}(\bar{s}) - p_{\text{exp}}(\bar{s})\right| \leq 2p_{\text{inc}}. \qquad (1)$$

The value $2p_{\text{inc}}$ is estimated experimentally with accuracy $\theta \in (0,1)$ and confidence $\alpha \in (0,1)$ chosen by the user. The number v of trap circuits is determined by the desired θ and α, but is independent of the size and nature of the target circuit (Eq. 2).

We implement our AP on ibmq_johannesburg and ibmq_paris, two two-dimensional arrays of superconducting transmon qubits. Fig. 1 shows the bounds provided by our AP for twenty-four different circuits. Of these, fourteen are structured circuits—ten circuits to generate Greenberger-Horne-Zeilinger (GHZ) states [12] and four to perform the quantum Fourier transform (QFT) [13], both important primitives in quantum computation—and ten are six-qubit random circuits of varying depth. The widest of these circuits has ten qubits and depth nine, the deepest has four qubits and depth twenty-one. The widths of our circuits compare favourably to that reached in the experimental demonstrations of some of the main protocols for noise characterization—three qubits in Process Tomography [14], five in Randomized Benchmarking [15], seven in Direct Fidelity Estimation [16] and ten in Cycle Benchmarking [17].

Our AP is designed to ascertain the correctness of a noisy quantum computation rather than the performance of individual gates or families thereof. Thus, it can detect noise (such as location-dependant noise acting on the whole register of qubits) that may arise when gates are put together to form a circuit and may be missed by the protocols for characterizing individual gates [14, 16–24]. An alternate accreditation protocol can detect even more complex noise such as temporally-correlated qubit-environment couplings, albeit at the cost of looser bounds

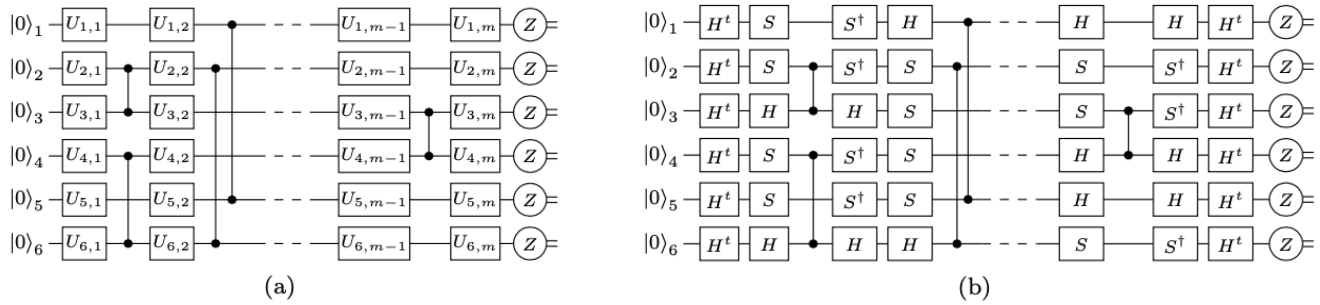
* samuele.ferracin@gmail.com
† seth.merkel@ibm.com
‡ dcmckay@us.ibm.com
§ animesh.datta@warwick.ac.uk



(a) Accreditation of GHZ and QFT circuits

(b) Accreditation of six-qubit random circuits

Figure 1. Experimental bounds on the VD (r.h.s. of Eq. 1) provided by our AP. In (a) the numbers inside the rectangles indicate the depths of the various target circuits. The bounds in (a) are calculated by implementing $v = 450$ trap circuits, those in (b) by implementing $v = 900$ trap circuits. The bars correspond to confidence levels above 95% on our estimates of $2p_{\text{inc}}$—specifically, in (a) we set $\theta = 13\%$ and $\alpha = 95\%$, in (b) $\theta = 9\%$ and $\alpha = 95\%$, see Eq. 2.



(a) (b)

Figure 2. (a) Example of target circuit. The target circuit must be compiled into m cycles of one-qubit gates, each one (apart from the last one) followed by a cycle of cZ gates (giving circuit depth $d = 2m - 1$). Input qubits are in the state $|0\rangle$ and measurements are in the Pauli-Z basis. (b) Example of trap circuit for the target circuit in (a). The trap circuit is obtained by replacing the one-qubit gates in the target circuit with one-qubit Clifford gates. Neighboring cycles of one-qubit gates can be recompiled into a single cycle. Thus, the trap circuit has the same circuit depth as the target.

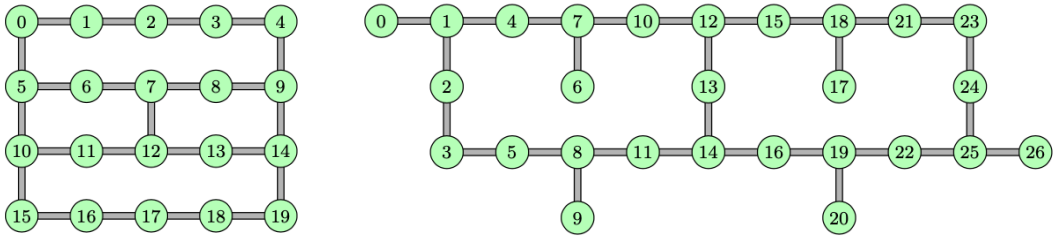
on the VD [25]—more details in section 3 in the Appendix. Due to its practicality, scalability and ability to capture a broad class of noise processes,we expect that in the future our AP will supplant the protocols based on classical simulations of quantum circuits.

We begin in section 2 by introducing the notation and the noise model, in section 3 we present our AP, in sections 4 and 5 we discuss the experimental results.

2. Notation and noise model—We indicate unitaries with capital letters and Completely Positive Trace Preserving (CPTP) maps with calligraphic letters. We use $I = \text{diag}(1,1)$ to denote the identity, X, Y and Z for the one-qubit Pauli matrices, $H = (X + Z)/\sqrt{2}$ for the Hadamard gate, $S = |0\rangle\langle 0| + i|1\rangle\langle 1|$ for the phase gate, $cZ = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes Z$ for the controlled-Z gate and $cX = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$ for the controlled-X gate. We indicate with "cycle" a set of gates acting on the entire system within a fixed period of time.

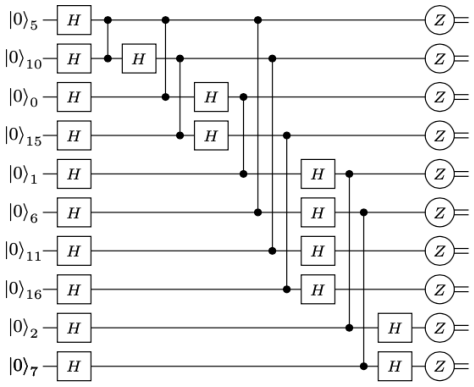
We model noise in state preparation, measurements and cycles as CPTP maps acting on all the qubits. Specifically, we assume that a noisy implementation of a cycle $\mathcal{G}$ on a state ρ at circuit depth j returns $\mathcal{E}_{\mathcal{G},j}\mathcal{G}(\rho)$, where $\mathcal{E}_{\mathcal{G},j}$ is a CPTP map that potentially acts on the whole system and depends on both $\mathcal{G}$ and on the depth j. This is a Markovian noise model that encompasses a broad class of noise processes afflicting current platforms, e.g., gate-dependent noise and cross-talk. It is more general than the noise models typically considered in the protocols for gate characterization, where the noise is represented by a static map $\mathcal{E}_{\mathcal{G}}$ independent of j [14, 16–24, 26].

We assume that the cycles of one-qubit gates suffer gate-independent noise, i.e. $\mathcal{E}_{\mathcal{U},j} = \mathcal{E}_j$ for all the cycles of one-qubit gates $\mathcal{U}$. In our analysis this assumption is required for two reasons: Firstly, to transform arbitrary noise processes into Pauli errors via a quantum one-time pad (QOTP, see section 3), and secondly, to ensure that the distributions of errors afflicting target and traps are identical. This is a common assumption in the literature on noise characterization protocols [14, 16–23, 26–32] and is motivated by the empirical observation that the one-qubit gates are the most accurate components in all the leading platforms [2, 33]. Nevertheless, we relax it by showing that the bound provided by our AP is robust to

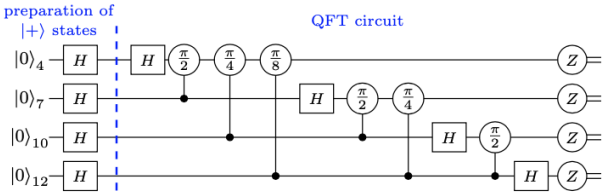


(a) ibmq_johannesburg. (b) ibmq_paris.

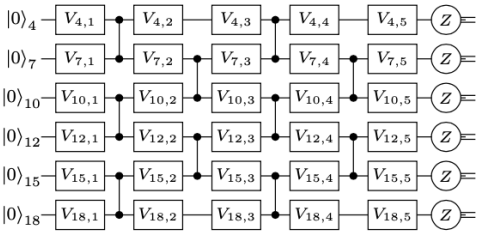
Figure 3. A graphical representation of the connectivity in ibmq_johannesburg and ibmq_paris. The circles represent qubits, the edges represent the available entangling gates.



(a) Our largest GHZ circuit.



(b) Our largest QFT circuit.



(c) Six-qubit depth nine pseudo-random circuit.

Figure 4. (a) Our largest GHZ circuit. After adding the QOTP, this circuit contains nine cycles of non-commuting gates. (b) Our largest QFT circuit. We apply the circuit to the state $|+\rangle^{\otimes 4}$. Each two-qubit gate is a controlled-phase gate, which we decompose into two cZ gates interleaved by one-qubit gates [34]. After adding the QOTP, this circuit contains twenty-one cycles of non-commuting gates. (c) Our six-qubit depth nine pseudo-random circuit. The various gates $V_{i,j}$ are random one-qubit gates.

noise that depends *weakly* on the cycles of one-qubit gates (section 2 of the Appendix).

3. Accreditation Protocol—Our AP takes as input the target circuit, and two numbers $\theta, \alpha \in (0,1)$ which quantify the desired accuracy and confidence on the final bound. The target circuit must (i) take as input n qubits in the state $|0\rangle$, (ii) contain $2m$ cycles alternating between a cycle of one-qubit gates and a cycle of two-qubit gates and (iii) end with measurements in the Pauli-Z basis (Fig. 2a). Our AP requires that all two-qubit gates in the target circuit be Clifford gates, so that arbitrary noise processes can be transformed into Pauli errors via QOTP. Without loss of generality, we assume that all the two-qubit gates in the circuit are cZ gates. Note that circuits containing different two-qubit Clifford gates (such as the cX gates implemented by IBM Quantum devices or the Mølmer-Sørensen gate implemented by trapped-ion quantum computers [35]) can be efficiently recompiled in this form without increasing the depth, while circuits containing two-qubit non-Clifford gates (such as those implemented by Google Sycamore [2]) require a linear increase in depth.

Our AP requires executing $v+1$ circuits sequentially, where $v = \lceil 2\ln(2/(1-\alpha))/\theta^2 \rceil$ and $\lceil \cdot \rceil$ is the ceiling function. One of these circuits (chosen at random) is the target circuit, the others are trap circuits. Each trap circuit is obtained by replacing the one-qubit gates in the target circuit with random one-qubit Clifford gates as per the following algorithm (Fig 2b):

1. For all $j \in \{1, \ldots, m-1\}$ and for all $i \in \{1, \ldots, n\}$:

   (i) If the j-th cycle of cZ gates connects qubit i to qubit i', randomly replace $U_{i,j}$ with S and $U_{i',j}$ with H, or $U_{i,j}$ with H and $U_{i',j}$ with S. Undo these gates after the cycle of cZ gates.

   (ii) If the j-th cycle of cZ gates does not connect qubit i to any other qubit, randomly replace $U_{i,j}$ with H or S. Undo this gate after the cycle of cZ gates.

2. Initialize a random bit $t \in \{0,1\}$. If $t=0$, do nothing. If $t=1$, append a cycle of Hadamard gates at the beginning and at the end of the circuit.

Since $(S^\dagger \otimes H)cZ(S \otimes H) = cX$, the trap circuits apply

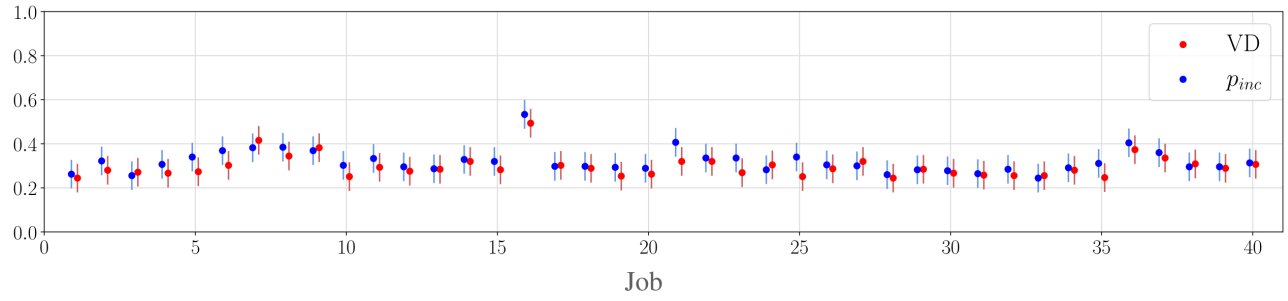


Figure 5. Values of VD (red points) and p_{inc} (blue points) measured in the experiment where the target circuit generates a six-qubit GHZ state. Each job contains 900 circuits (450 target circuits and 450 independently chosen trap circuits). The bars correspond to confidence levels above 95%—specifically, we set $\theta/2 = 6.5\%$ and $\alpha \gtrsim 95\%$, see Eq. 2.

a series of cX gates to n qubits in the state $|0\rangle$ (if $t = 0$) or $|+\rangle$ (if $t = 1$). Using $cX\,|00\rangle = |00\rangle$ and $cX\,|{+}{+}\rangle = |{+}{+}\rangle$ it can be seen that in the absence of noise all the trap circuits return the bit-string $(0, 0, \ldots, 0)$.

After initializing the $v+1$ circuits, we append a QOTP to each cycle of one-qubit gates in every circuit. This is done by appending a cycle of random Pauli gates after every cycle of one-qubit gates, and by appending a second cycle of Pauli gates before the following cycle of one-qubit gates that undoes the first. This randomizes the noise to stochastic Pauli errors [4, 10, 25, 36–38]. The trap circuits are designed to *detect* these Pauli errors, meaning that any Pauli error alters their outputs with a probability of at least 50% [25].

After appending the QOTP we recompile neighbouring cycles of one-qubit gates into a single cycle. This ensures that all the circuits (target and traps) contain the same number of cycles as the circuit given as input to the AP. Next, we implement all the circuits and subsequently estimate the probability p_{inc} as the fraction N_{inc}/v of traps that return an incorrect output. Since $v > 2\ln(2/(1-\alpha))/\theta^2$, the Hoeffding's inequality [39] guarantees that

$$\text{prob}\left(\left|p_{\text{inc}} - \frac{N_{\text{inc}}}{v}\right| \leq \frac{\theta}{2}\right) \geq \alpha\,. \tag{2}$$

Finally, we calculate the bound on the VD as $2N_{\text{inc}}/v$, and we have $\text{prob}(|2p_{\text{inc}} - 2N_{\text{inc}}/v| \leq \theta) \geq \alpha$ by Hoeffding's inequality.

The quantity $2p_{\text{inc}}$ grows linearly with the total probability p_{err} that the target circuit is afflicted by errors. More formally, we have

$$p_{\text{err}} \leq 2p_{\text{inc}} \leq 2p_{\text{err}}\,. \tag{3}$$

Here, the bound on the l.h.s. is proven in section 1 of the Appendix, while that on the r.h.s. is a consequence of the fact that in the absence of errors the traps always return the correct output. Since the VD is at most unity by construction, it follows that if $p_{\text{err}} \leq 50\%$ our AP always returns a *non-trivial* bound on the VD (i.e., below unity). Otherwise, it may return a trivial bound, indicating that the device is afflicted by such high levels of noise that its outputs are far enough from the ideal ones as to be unreliable.

As can be seen in Fig. 1, in our experiments we obtain non-trivial bounds for circuits with up to ten qubits. Larger circuits yield trivial bounds. However, Eq. 3 shows that improvements in the hardware will extend the reach of the AP beyond ten-qubit circuits. Being fully scalable, in the future the AP will be able to accredit the outputs of quantum circuits that will be intractable for the protocols relying on classical simulations.

4. Experimental Accreditation—We implement our AP on two superconducting quantum computers, ibmq_johannesburg and ibmq_paris. These quantum computers consist of superconducting transmon qubits dispersively coupled according to the topology given in Fig. 3, where each edge denotes a cX gate that can be implemented via the cross-resonance interaction. For a more comprehensive description of this architecture see Ref. [40] and for specific details about ibmq_johannesburg and ibmq_paris see Ref. [41].

We begin by conducting fourteen experiments to accredit the outputs of QFT and GHZ circuits of different widths (Figs. 4a and 4b). In every experiment we submit 40 jobs to the backend. Each job contains 450 trap circuits, each one chosen independently at random as described in section 3. At the end of each job we estimate p_{inc}, as illustrated in Fig. 5 for the preparation and measurement of the six-qubit GHZ state. (See [42] for more figures). To demonstrate the AP, in each job we also implement 450 instances of the target circuit and compute the VD between the ideal and experimentally obtained probability distributions. In our experiments this can be done within a reasonable amount of time given the size of the target circuits.

In every job we find VD $\approx p_{\text{inc}}$, but we observe fluctuations across different jobs. These fluctuations indicate that different jobs suffer different noise due to e.g. automatic recalibration of the internal components of the device. VD $\approx p_{\text{inc}}$ also suggests that the factor 2 on the r.h.s. of Eq. 1 may be unnecessary. However, this factor 2 captures the effects of specific patterns of errors that are detected with probability 50% (such as single-cycle patterns afflicting a single qubit, see Fig. 6a). Thus, it

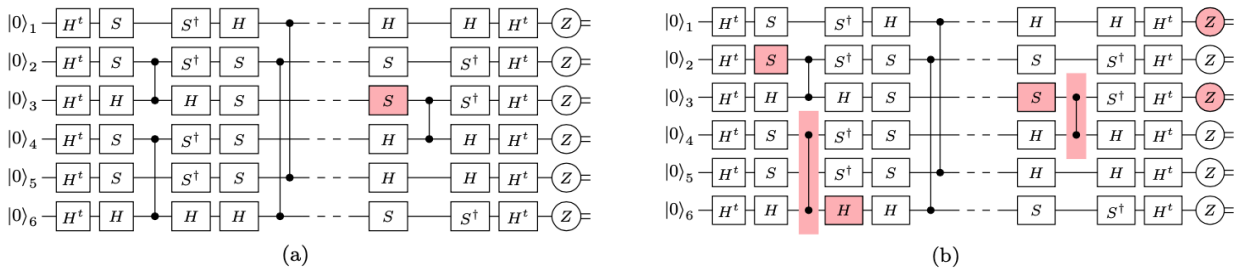


Figure 6. Examples of trap circuits affected by errors (the faulty gates and measurements are highlighted in red). (a) Single-cycle pattern affecting a single qubit. Patterns of this type are detected with probability 50% (see Ref. [25]). (b) Multi-cycle pattern. Patterns of this type are detected with probability greater than 50%.

is necessary to ensure the validity of Eq. 1 for arbitrary types of noise. In general, multi-cycle patterns (Fig. 6b) as well as patterns afflicting more than one qubit (such as those afflicting the today's devices [23, 38]) are detected with probability greater than 50% [25], and for these patterns we have $p_{\rm err} < 2p_{\rm inc}$.

Importantly, we find VD $< 2p_{\rm inc}$ for every job in each of our experiments [42]. This proves that in all the tests that we have conducted, our AP has correctly bounded the VD as expected from Eq. 1. Fig. 1a shows the smallest value of $2p_{\rm inc}$ obtained in the various experiments.

To study how $2p_{\rm inc}$ varies with the circuit depth we conduct ten more experiments on ibmq_paris. We target a set of six-qubit pseudo-random circuits of depths ranging from one to nineteen. These circuits alternate cycles of random one-qubit gates to cycles containing either two or three cZ gates (Fig. 4c). In every experiment we submit 20 jobs to the backend, each one containing 900 unique trap circuits. In Fig. 1b we show the smallest values of $2p_{\rm inc}$ obtained across the 20 jobs for each experiment.

5. Hardware diagnosis using AP—The trap circuits implement deterministic computations, designed to return the output $\overline{s} = (0, \ldots, 0)$ in the absence of errors and some other output in the presence of errors. Importantly, different errors alter the traps' outputs in different ways. Therefore, we expect that the probability distribution of the traps' outputs contains information regarding the nature of the noise afflicting the device in use. To corroborate this, in this section we focus on the traps' outputs collected in the experiments with six-qubit pseudo-random circuits. We show how these outputs can help identify the main sources of errors in circuits of different sizes implemented on ibmq_paris.

In ibmq_paris the error rates provided by the backend are around 0.05% for the one-qubit gates, 1.5% for the two-qubit gates and 2.3% for single-qubit measurements [42], while errors in state preparation are expected to be negligible. Therefore, we expect measurement noise to be the dominant source of error in shallow circuits and gate noise in deep circuits. To verify this, let us consider a noise model where the gates are noiseless, while measurement errors flip each bit $s_i \in \overline{s}$ with probability $p_{\rm flip}$. In this scenario, the probability that a trap returns an output $\overline{s}$ with Hamming weight $H_{\overline{s}} = h \in \{0, \ldots, n\}$ is

$$P_{\rm trap}\big(H_{\overline{s}} = h\big) = \binom{n}{h} p_{\rm flip}^{h} (1 - p_{\rm flip})^{n-h} \,, \qquad (4)$$

where the Hamming weight $H_{\overline{s}} = \sum_{s_i \in \overline{s}} s_i$ is the number of bits equal to 1 in the output string $\overline{s}$.

Setting $p_{\rm flip} = 2.3\%$, in Fig. 7a we compare the values of $P_{\rm trap}\big(H_{\overline{s}} = h\big)$ from our bit-flip noise model (striped bars) with the experimentally measured ones for pseudo-random circuits (solid bars). It can be seen that the bit-flip model accurately predicts the results obtained for shallow circuits (e.g. for circuits of depth one or three), indicating that measurement noise dominates short-depth circuits. It can also be seen that the bit-flip model becomes progressively disparate as the depth increases, indicating that in deep circuits measurements are no more the dominant contributor to noise.

The above inference may be challenged by positing that the measurement noise changes with the circuit's depth. To rule this possibility out, in Fig. 7b we set $p_{\rm flip} = 7.6\%$ such that the value of $P_{\rm trap}\big(H_{\overline{s}} = 0\big)$ calculated using the bit-flip model (left-most bar in the figure) equals the value measured in the experiment with depth nineteen random circuits. As can be seen in the figure, the bit-flip model still remains largely disparate. Overall, measurement errors alone cannot explain the distribution of outputs of our deepest trap circuits and gate noise can no longer be neglected.

This simple analysis builds upon the error rates provided by the backend and is thus device-specific. It shows that the probability distribution of the traps' outputs contains information regarding the noise afflicting ibmq_paris. Obvious questions as to how much of this information can be retrieved, and whether it can be retrieved in a device-agnostic manner remain open for future work.

5. Conclusions—We have presented an accreditation protocol that uses random Clifford circuits to ascertain the correctness of the outputs of quantum computations implemented on existing hardware. We have experimen-

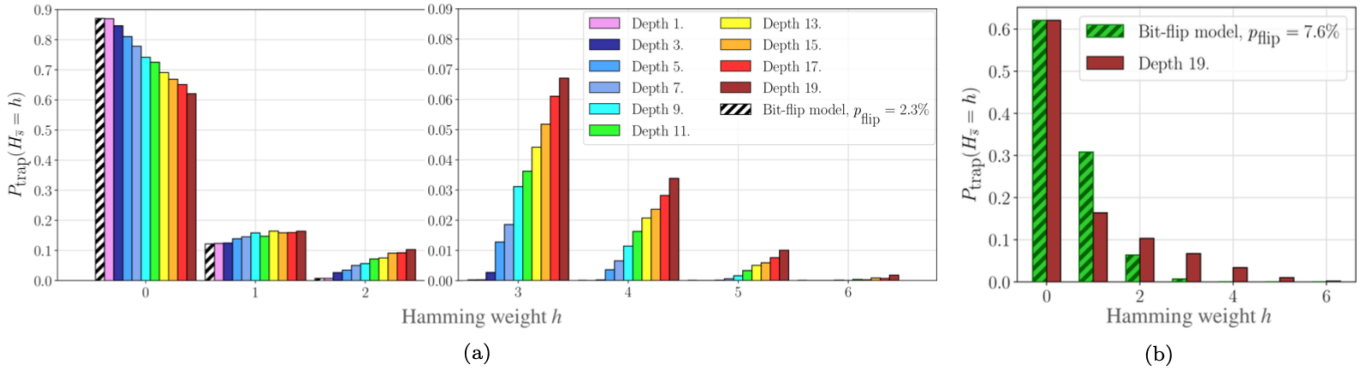


Figure 7. The values of $P_{\text{trap}}(H_{\bar{s}} = h)$ calculated using the bit-flip model defined in Eq. 4 (striped bars) and those measured in the experiments with pseudo-random circuits (solid bars)—for $h \geq 3$ the y axis is rescaled. At any given depth, the values of $P_{\text{trap}}(H_{\bar{s}} = h)$ are measured using the outputs of 17980 traps. In **(a)** we set $p_{\text{flip}} = 2.3\%$, which coincides with the error rate provided for the measurements by the backend. In **(b)** we set $p_{\text{flip}} = 7.6\%$.

tally demonstrated its present practicality and mathematically established its future scalability.

Presently, the factor 2 in the r.h.s. of Eq. 1 represents the main obstacle towards increasing the number of qubits in our experiments beyond $n = 10$. Indeed, for target circuits with $n > 10$ qubits we find $p_{\text{inc}} > 50\%$, hence our AP returns a bound on the VD that exceeds unity. This is a trivial bound, since the VD is below unity by construction [13]. Nevertheless, better devices will extend the reach of our AP beyond 10-qubit circuits. Being fully scalable, we anticipate that in the future our AP will replace the protocols based on classical simulations of quantum circuits [1–3] and will become a standard routine to characterize the outputs of noisy quantum computers.

Acknowledgments—SF and AD were supported by the UK Networked Quantum Information Technologies (NQIT) Hub (EP/M013243/1) in the early stages of this work. SM and DM research was sponsored by the Army Research Office and was accomplished under Grant Numbers W911NF-14-1-0124 and W911NF-21-1-0002. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

[1] S. Boixo et al. Characterizing Quantum Supremacy in Near-Term Devices. *Nature Physics 14, 595-600*, 2018.
[2] Google AI and Collaborators. Quantum supremacy using a programmable superconducting processor. *Nature volume 574, pp 505–510*, 2019.
[3] A. Cross, L. Bishop, S. Sheldon, P. Nation, and J. Gambetta. Validating quantum computers using randomized model circuits. *Phys. Rev. A 100, 032328*, 2019.
[4] A. Childs. Secure Assisted Quantum Computation. *Quantum Information and Computation 5, 456*, 2005.
[5] J. Fitzsimons and E. Kashefi. Unconditionally Verifiable Blind Computation. *Phys. Rev. A 96, 012303*, 2017.
[6] B.W. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. *arXiv:1209.0448*, 2012.
[7] A. Broadbent. How to Verify a Quantum Computation. *Theory of Computing 14(11):1-37*, 2018.
[8] M. Hayashi and T. Morimae. Verifiable measurement-only blind quantum computing with stabilizer testing. *Phys. Rev. Lett. 115, 220502*, 2015.
[9] T. Morimae and J. Fitzsimons. Post hoc verification with a single prover. *Phys. Rev. Lett. 120, 040501*, 2018.
[10] S. Ferracin, T. Kapourniotis, and A. Datta. Reducing resources for verification of quantum computations. *Phys. Rev. A 98, 022323*, 2017.
[11] U. Mahadev. Classical Verification of Quantum Computations. *arXiv:1804.01082*, 2018.
[12] D. Greenberger, M. Horne, and A. Zeilinger. Going Beyond Bell's Theorem. *Bell's Theorem, Quantum Theory, and Conceptions of the Universe, pp 69-72*, 1989.
[13] M. Nielsen and I. Chuang. Quantum Computation and Quantum Information: 10th anniversary edition. *Cambridge University Press New York, NY, USA*, 2000.
[14] Y. Weinstein, T. Havel, J. Emerson, and N. Boulant. Quantum process tomography of the quantum Fourier transform . *J. Chem. Phys. 121, 6117*, 2004.
[15] T. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young. Direct randomized benchmarking for multi-qubit devices. *Phys. Rev. Lett. 123, 030503*, 2019.
[16] D. Lu et al. Experimental Estimation of Average Fidelity of a Clifford Gate on a 7-Qubit Quantum Processor. *Phys. Rev. Lett. 114, 140505*, 2015.
[17] A. Erhard et al. Characterizing Large-Scale Quantum Computers Via Cycle Benchmarking. *Nat. Commun. 10,*

5347, 2019.

[18] S. Merkel, J. Gambetta, J. Smolin, S. Poletto, and A. Corcoles. Self-consistent quantum process tomography. *Phys. Rev. A 87, 062119*, 2013.

[19] R. Blume-Kohout, J. King Gamble, E. Nielsen, J. Mizrahi, J. Sterk, and P. Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. *arXiv:1310.4492*, 2017.

[20] R. Blume-Kohout, J. 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. *Nature Communications 8, 14485*, 2017.

[21] S. Merkel et al. Self-consistent quantum process tomography. *Phys. Rev. A 87, 062119*, 2013.

[22] S. Flammia and K. Liu. Direct Fidelity Estimation from Few Pauli Measurements . *Phys. Rev. Lett. 106, 230501*, 2011.

[23] R. Harper, S. Flammia, and J. Wallman. Efficient learning of quantum noise. *Nat. Phys. 16, 1184-1188*, 2020.

[24] S. Flammia and J. Wallman. Efficient estimation of Pauli channels. *arXiv:1907.12976*, 2019.

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

[26] M. Lilly and T. Humble. Modeling noisy quantum circuits using experimental characterization. *arXiv:2001.08653*, 2020.

[27] I. Chuang and M. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics, Vol. 44, Issue 11-12*, 1997.

[28] D Greenbaum. Introduction to Quantum Gate Set Tomography. *arxiv:1509.0292*, 2015.

[29] M. Da Silva, O. Landon-Cardinal, and D. Poulin. Practical Characterization of Quantum Devices without Tomography. *Phys. Rev. Lett. 107, 210404*, 2011.

[30] O. Moussa, M. Da Silva, C. Ryan, and R. Laflamme. Practical Experimental Certification of Computational Quantum Gates Using a Twirling Procedure. *Phys. Rev. Lett. 109, 070504*, 2012.

[31] E. Knill et al. Randomized Benchmarking of Quantum Gates. *Phys. Rev. A 77, 012307*, 2007.

[32] E. Magesan, J. Gambetta, and J. Emerson. Scalable and Robust Randomized Benchmarking of Quantum Processes. *Phys. Rev. Lett. 106, 180504*, 2011.

[33] K. Wright et al. Benchmarking an 11-qubit quantum computer. *Nature Communications volume 10, Article number: 5464*, 2019.

[34] G. Barron et al. Microwave-based arbitrary CPHASE gates for transmon qubits. *Phys. Rev. B 101, 054508*, 2020.

[35] K. Mølmer and A. Sørensen. Multiparticle Entanglement of Hot Trapped Ions. *Phys. Rev. Lett. 82, 1835*, 1999.

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

[37] A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal Blind Quantum Computation. *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science, pp 517-526*, 2009.

[38] A. Hashim et al. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *arXiv:2010.00215*, 2020.

[39] W. Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association, 58 (301), pp 13–30*, 1963.

[40] P. Jurcevic et al. Demonstration of quantum volume 64 on a superconducting quantum computing system. *arXiv:2008.08571*, 2020.

[41] IBM Quantum Experience, https://quantum-computing.ibm.com/.

[42] S. Ferracin, S. Merkel, D. McKay, and A. Datta. Data and notebooks for "Experimental accreditation of outputs of noisy quantum computers", doi: 10.5281/zenodo.4592487, 2021.

APPENDIX

The Appendix is organized as follows: In section 1 we provide a derivation of the bound on the VD provided by our AP, in section 2 we show that our protocol is robust to noise processes that depend weakly on the choice of one-qubit gates, in section 3 we compare our AP with the AP in Ref. [25]. We refer the reader to section 2 of the main text for the notation.

1. Derivation of the bound on the VD—In this section we derive the bound on the VD provided by our AP (Eq. 1). Before presenting the mathematical proof, we calculate the state of the system at the end of a noisy implementation of the k-th circuit executed in our AP, with $k \in \{1,\dots,v+1\}$.

Under the assumptions that noise is Markovian and that the cycles of one-qubit gates suffer gate-independent noise, the state of the system at the end of a noisy implementation of circuit k is

$$\widetilde{\rho}_{\text{out}}^{(k)} = \mathcal{M}\,\mathcal{E}_m \mathcal{U}_m^{(k)}\,\mathcal{E}_{c\mathcal{Z}_{m-1},m-1}\,c\mathcal{Z}_{m-1}\mathcal{U}_{m-1}^{(k)}\cdots \\ \cdots \mathcal{E}_{c\mathcal{Z}_1,1}\,c\mathcal{Z}_1\mathcal{U}_1^{(k)}\,\mathcal{R}\big(|0\rangle\langle 0|^{\otimes n}\big)\,, \tag{5}$$

where $\mathcal{R}$ is the noise in state preparation, $\mathcal{U}_j^{(k)}$ ($c\mathcal{Z}_j$) is the j-th cycle of one-qubit gates (two-qubit gates), $\mathcal{E}_{c\mathcal{Z}_j,j}$ is the noise due to $c\mathcal{Z}_j\mathcal{U}_j^{(k)}$ (which depends only on $c\mathcal{Z}_j$ and not on $\mathcal{U}_j^{(k)}$) and finally, $\mathcal{M}$ is the round of measurements. To simplify the structure of the noise, a QOTP is appended to each cycle of one-qubit gates in all the circuits. This randomizes the noise into stochastic Pauli errors [4, 10, 25, 36, 37] and allows rewriting $\widetilde{\rho}_{\text{out}}^{(k)}$ as

$$\widetilde{\rho}_{\text{out}}^{(k)} = \sum_{\mathcal{P}_0,\dots,\mathcal{P}_m} q_0(\mathcal{P}_0)\cdots q_m(\mathcal{P}_m)\mathcal{M}\mathcal{P}_m c\mathcal{Z}_m\mathcal{U}_m^{(k)}\cdots \\ \cdots \mathcal{P}_1 c\mathcal{Z}_1\mathcal{U}_1^{(k)}\mathcal{P}_0\big(|0\rangle\langle 0|^{\otimes n}\big)\,, \tag{6}$$

where $q_0(\mathcal{P}_0)\cdots q_m(\mathcal{P}_m)$ is the probability that the "pattern of Pauli errors" $\mathcal{P}_0,\dots,\mathcal{P}_m \in \{\mathcal{I},\mathcal{X},\mathcal{Y},\mathcal{Z}\}^{\otimes n}$ occurs.

Importantly, note that the cycles of two-qubit gates are identical in all the circuits (target and traps), as well as the input state and measurements. Therefore, under the assumption that the one-qubit gates suffer gate-independent noise, the probabilities $q(\mathcal{P}_0)\cdots q(\mathcal{P}_m)$ are the same in all the circuits, and so is the total probability of error per circuit

$$p_{\text{err}} = \sum_{\mathcal{P}_0,\dots,\mathcal{P}_m\neq\mathcal{I},\dots,\mathcal{I}} q_0(\mathcal{P}_0)\cdots q_m(\mathcal{P}_m)\,. \tag{7}$$

We can now establish the bound on the VD.

Proof. (Eq. 1). To prove the inequality we make use of the following two statements:

Statement 1. *(Proof in Appendix B of Ref. [25]). Suppose that a trap circuit is afflicted by a "single-cycle" pattern of errors, i.e. a pattern such that $\mathcal{P}_{j_0}\neq\mathcal{I}$ for some $j_0\in\{0,\dots,m\}$ and $\mathcal{P}_j=\mathcal{I}$ for all $j\neq j_0$. Then, summed over the random one-qubit gates in the trap, the trap returns an incorrect output with probability 50% or above.*

Statement 2. *(Proof at the end of this section). Let $q_{\text{tot}}(j)=\sum_{\mathcal{P}_j\neq\mathcal{I}} q_j(\mathcal{P}_j)$ be the error rate of cycle $j\in\{0,\dots,m\}$. Denoting by p_{canc} the probability that errors in different cycles of a trap circuit cancel with each other, we have $p_{\text{canc}}\leq C$, where*

$$C = O\Big(\sum_{j,j'\neq j} q_{\text{tot}}(j)q_{\text{tot}}(j')\Big)\,. \tag{8}$$

Statements 1 and 2 ensure that the trap circuits can detect errors with probability greater that 50%. To see this, consider the state of the system at the end of a trap circuit (Eq. 6). Since all the gates in the trap circuits are Clifford, we can map arbitrary patterns of errors into single-cycle patterns. That is, we can commute the errors with the various cycles and *merge* them into a single error $\mathcal{Q}_{(\mathcal{P}_0,\dots,\mathcal{P}_m)}$ (which depends on the initial errors $\mathcal{P}_0,\dots,\mathcal{P}_m$), obtaining

$$\widetilde{\rho}_{\text{out}}^{(k)} = \sum_{\mathcal{P}_0,\dots,\mathcal{P}_m} q_0(\mathcal{P}_0)\cdots q_m(\mathcal{P}_m)\,\mathcal{M}c\mathcal{Z}_m\mathcal{U}_m^{(k)}\cdots \\ \dots\,\mathcal{Q}_{(\mathcal{P}_0,\cdots,\mathcal{P}_m)}c\mathcal{Z}_{j_0}\mathcal{U}_{j_0}^{(k)}\dots c\mathcal{Z}_1\mathcal{U}_1^{(k)}\big(|0\rangle\langle 0|^{\otimes n}\big) \tag{9}$$

for some $j_0\in\{0,\dots,m\}$. In principle, the errors in the trap may cancel with each other, yielding $\mathcal{Q}_{(\mathcal{P}_0,\dots,\mathcal{P}_m)}=\mathcal{I}$. In particular, denoting by p_{canc} the probability of error cancellation, we obtain $\mathcal{Q}_{(\mathcal{P}_0,\dots,\mathcal{P}_m)}\neq\mathcal{I}$ with probability $p_{\text{err}}(1-p_{\text{canc}})$.

Having mapped the original pattern into a single-cycle pattern, Statement 1 ensures that if errors do not cancel (i.e., if $\mathcal{Q}_{(\mathcal{P}_0,\dots,\mathcal{P}_m)}\neq\mathcal{I}$), the trap circuit returns the incorrect output with probability greater than 50%. This proves that

$$p_{\text{inc}} \geq \frac{p_{\text{err}}(1-p_{\text{canc}})}{2}\,, \tag{10}$$

where p_{inc} is the probability that a trap returns an incorrect output.

We can now use Eq. 10 to upper-bound the VD between ideal and experimental outputs of the target circuit. Labeling the target circuit with $v_0\in\{1,\dots,v+1\}$, we rewrite the state of the system at the end of the target circuit (Eq. 6 with $k=v_0$) as

$$\widetilde{\rho}_{\text{out}}^{(v_0)} = (1-p_{\text{err}})\rho_{\text{out}}^{(v_0)} + p_{\text{err}}\sigma^{(v_0)}\,, \tag{11}$$

where $\rho_{\text{out}}^{(v_0)}$ is the state of the system at the end of an ideal implementation of the target circuit and $\sigma^{(v_0)}$ is a state encompassing the effects of noise. This leads to

$$\text{VD} = \frac{1}{2}\sum_{\bar{s}}\big|p_{\text{ideal}}(\bar{s}) - p_{\text{exp}}(\bar{s})\big| \tag{12}$$

$$= D\big(\rho_{\text{out}}^{(v_0)},\widetilde{\rho}_{\text{out}}^{(v_0)}\big) \leq p_{\text{err}} \leq 2\frac{p_{\text{inc}}}{1-p_{\text{canc}}}\,, \tag{13}$$

where $D(\tau,\tau') = \mathrm{Tr}|\tau-\tau'|/2$ is the trace distance between the states τ and τ'. Finally, since $p_{\rm canc} \leq C$ and C is quadratic in the cycles' error rates by Statement 2, we have $p_{\rm canc} \ll 1$ and

$$\mathrm{VD} \leq 2p_{\rm inc}(1+p_{\rm canc}) \approx 2p_{\rm inc}\,. \tag{14}$$

□

Relying on Statement 2, in the proof of Eq. 1 we used $p_{\rm canc} \leq C$, as well as $C \ll 1$. The latter can be corroborated empirically using calibration data. For example, our largest circuit (the ten-qubit GHZ circuit, Fig. 4a) contains five cycles of one-qubit gates with an error rate $\approx 0.1\%$ [42] and four cycles of two-qubit gates. Since each two-qubit gate has an error rate $\approx 1.5\%$ [42], we estimate an error rate $\approx 1.5\%$ for the first cycle, $\approx 3\%$ for the second and the fourth and $\approx 6\%$ for the third. One-qubit measurements have error rates $\approx 2\%$ [42], from which we estimate an error rate $\approx 20\%$ for the final cycle of measurements. Overall, using Eq. 8 we estimate $C \approx 3\%$. With the same strategy we estimate values of C below 3% for all the other circuits. As we point out at the end of this section, $p_{\rm canc} \leq C$ is a loose bound and we expect that $p_{\rm canc}$ be well below C in practice.

We now provide a proof of Statement 2.

Proof. (Statement 2.) For simplicity, let us first consider the case where errors afflict two neighbouring cycles j and $j+1$ and no other cycle. In this case, error cancellation happens when $\mathcal{P}_{j+1} = c\mathcal{Z}_{j+1}\mathcal{U}_{j+1}(\mathcal{P}_j)$. Therefore, indicating by $Q(j,j+1)$ the probability of error cancellation we have

$$\begin{aligned} Q(j,j+1) &= \sum_{\mathcal{P}_j \neq \mathcal{I}} q_j(\mathcal{P}_j) q_{j+1}\big(c\mathcal{Z}_{j+1}\mathcal{U}_{j+1}(\mathcal{P}_j)\big) && (15)\\ &\leq \sum_{\mathcal{P}_j, \mathcal{P}_{j+1} \neq \mathcal{I}} q_j(\mathcal{P}_j) q_{j+1}(\mathcal{P}_{j+1}) && (16)\\ &= q_{\rm tot}(j) q_{\rm tot}(j+1)\,, && (17)\end{aligned}$$

where to obtain Eq. 16 we use the fact that the probability of error cancellation is no more than the product of the probabilities of errors happening. With the same arguments we can upper-bound the probability $Q(j_1,j_2)$ of error cancellation for patterns afflicting any two cycles j_1 and j_2 as

$$Q(j_1,j_2) \leq q_{\rm tot}(j_1) q_{\rm tot}(j_2)\,. \tag{18}$$

This proves that the probability of error cancellation for patterns afflicting two cycles is at most quadratic in the cycles' error rates.

With the same strategy it can be shown that the probability of error cancellation for patterns afflicting $K > 2$ cycles $j_1, j_2, \ldots, j_K$ is higher order in the cycles' error rates. Specifically, indicating this probability by $Q(j_1, j_2, \ldots, j_K)$ we find

$$Q(j_1, j_2, \ldots, j_K) \leq q_{\rm tot}(j_1) q_{\rm tot}(j_2) \cdots q_{\rm tot}(j_K). \tag{19}$$

This leads to

$$\begin{aligned} p_{\rm canc} &= \sum_{K \in \{2,\ldots,m+1\}} \Big(\sum_{j_1, j_2 \ldots, j_K} Q(j_1, j_2 \ldots, j_K) \Big) && (20)\\ &= O\Big(\sum_{j_1, j_2 \neq j_1} Q(j_1, j_2) \Big) && (21)\\ &= O\Big(\sum_{j_1, j_2 \neq j_1} q_{\rm tot}(j_1) q_{\rm tot}(j_2) \Big)\,. && (22)\end{aligned}$$

□

We conclude the section by pointing out that $Q(j_1, j_2, \ldots, j_K) \leq q_{\rm tot}(j_1) q_{\rm tot}(j_2) \cdots q_{\rm tot}(j_K)$, and consequently $p_{\rm canc} \leq C$, is a loose bound. To see this, note that to upper-bound the r.h.s. of Eq. 15 we use $q_{j+1}\big(c\mathcal{Z}_{j+1}\mathcal{U}_{j+1}(\mathcal{P}_j)\big) \leq q_{\rm tot}(j+1)$. That is, we replace the probabilities of individual errors (including negligible probabilities) with the total probability of error in cycle $j+1$. Based on this observation, we expect that $p_{\rm canc}$ be well below C.

2. Robustness to weak gate-dependent noise. In the proof of Eq. 1 we have assumed that the cycles of one-qubit gates suffer gate-independent noise. In practice this assumption may be too stringent. To relax this assumption, in this section we analyse how gate-dependent noise may affect the effectiveness of our AP. Formally:

Theorem 1. *Let us consider a circuit implementing the operation*

$$\mathcal{C} = \sum_{\mathcal{U}_1,\ldots,\mathcal{U}_m} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) c\mathcal{Z}_m\mathcal{U}_m \ldots c\mathcal{Z}_1\mathcal{U}_1\,, \tag{23}$$

where the cycles of one-qubit gates $\mathcal{U}_1,\ldots,\mathcal{U}_m$ are chosen with probability $p(\mathcal{U}_1,\ldots,\mathcal{U}_m)$. Let

$$\mathcal{C}_{gi} = \sum_{\mathcal{U}_1,\ldots,\mathcal{U}_m} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) \mathcal{E}_{c\mathcal{Z}_m,m} c\mathcal{Z}_m\mathcal{U}_m \ldots \mathcal{E}_{c\mathcal{Z}_1,1} c\mathcal{Z}_1\mathcal{U}_1 \tag{24}$$

be a noisy implementation of $\mathcal{C}$ with noise $\mathcal{E}_{c\mathcal{Z}_j,j}$ that depends only on the cycle of two-qubit gates $c\mathcal{Z}_j$ and on the index j. Let

$$\mathcal{C}_{gd} = \sum_{\mathcal{U}_1,\ldots,\mathcal{U}_m} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) \mathcal{E}_{c\mathcal{Z}_m\mathcal{U}_m,m} c\mathcal{Z}_m\mathcal{U}_m \ldots \mathcal{E}_{c\mathcal{Z}_1\mathcal{U}_1,1} c\mathcal{Z}_1\mathcal{U}_1 \tag{25}$$

be a noisy implementation of $\mathcal{C}$ with noise $\mathcal{E}_{c\mathcal{Z}_j\mathcal{U}_j,j}$ that depends also on the cycle of one-qubit gates $\mathcal{U}_j$. Averaged over all possible choices of one-qubit gates we have

$$||\mathcal{C}_{gi} - \mathcal{C}_{gd}||_\diamond \leq \sum_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_m \\ j=1,\ldots,m}} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) ||\mathcal{E}_{c\mathcal{Z}_j,j} - \mathcal{E}_{c\mathcal{Z}_j\mathcal{U}_j,j}||_\diamond, \tag{26}$$

where $||\cdot||_\diamond$ is the diamond distance.

The above theorem shows that if the noise depends *weakly* on the choice of one-qubit gates (i.e., $||\mathcal{E}_{c\mathcal{Z}_j,j} - \mathcal{E}_{c\mathcal{Z}_j\mathcal{U}_j,j}||_\diamond$ is small for all j), the outputs of a circuit affected by gate-dependent noise remain close to those of the same circuit affected by gate-independent noise. This theorem is valid for any circuit where the one-qubit gates are selected at random. Applied to the target and trap circuits discussed in this paper, it guarantees our AP is robust to noise that depends weakly on the choice of one-qubit gates.

Proof. (Theorem 1). Our proof follows the same arguments as those in Ref. [36]. Let

$$\mathcal{F}_j = \mathcal{E}_{c\mathcal{Z}_j,j} c\mathcal{Z}_j\mathcal{U}_j \tag{27}$$
$$\mathcal{G}_j = \mathcal{E}_{c\mathcal{Z}_j\mathcal{U}_j,j} c\mathcal{Z}_j\mathcal{U}_j \tag{28}$$

and

$$\mathcal{F}_{j:1} = \mathcal{F}_j \cdots \mathcal{F}_1 \tag{29}$$
$$\mathcal{G}_{j:1} = \mathcal{G}_j \cdots \mathcal{G}_1 \, . \tag{30}$$

By induction it can be proven that

$$\mathcal{F}_{m:1} - \mathcal{G}_{m:1} = \sum_{j=1}^{m} \mathcal{F}_{m:j+1}(\mathcal{F}_j - \mathcal{G}_j)\mathcal{G}_{j-1:1} \tag{31}$$

Noting that

$$\mathcal{C}_{\rm gi} = \sum_{\mathcal{U}_1,\ldots,\mathcal{U}_m} p(\mathcal{U}_1,\ldots,\mathcal{U}_m)\mathcal{F}_{m:1} \tag{32}$$
$$\mathcal{C}_{\rm gd} = \sum_{\mathcal{U}_1,\ldots,\mathcal{U}_m} p(\mathcal{U}_1,\ldots,\mathcal{U}_m)\mathcal{G}_{m:1} \tag{33}$$

we have

$$||\mathcal{C}_{\rm gi} - \mathcal{C}_{\rm gd}||_\diamond \tag{34}$$
$$= \Big|\Big| \sum_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_m \\ j=1,\ldots,m}} p(\mathcal{U}_1,\ldots,\mathcal{U}_m)\mathcal{F}_{m:j+1}(\mathcal{F}_j - \mathcal{G}_j)\mathcal{G}_{j-1:1} \Big|\Big|_\diamond \tag{35}$$
$$\leq \sum_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_m \\ j=1,\ldots,m}} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) \big|\big|\mathcal{F}_{m:j+1}(\mathcal{F}_j - \mathcal{G}_j)\mathcal{G}_{j-1:1}\big|\big|_\diamond \tag{36}$$
$$\leq \sum_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_m \\ j=1,\ldots,m}} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) \big|\big|\mathcal{F}_j - \mathcal{G}_j\big|\big|_\diamond \tag{37}$$
$$= \sum_{\substack{\mathcal{U}_1,\ldots,\mathcal{U}_m \\ j=1,\ldots,m}} p(\mathcal{U}_1,\ldots,\mathcal{U}_m) \big|\big|\mathcal{E}_{c\mathcal{Z}_j,j} - \mathcal{E}_{c\mathcal{Z}_j\mathcal{U}_j,j}\big|\big|_\diamond \, , \tag{38}$$

where we used the fact that $||\mathcal{F}_j||_\diamond, ||\mathcal{G}_j||_\diamond \leq 1$ for all $\mathcal{F}_j, \mathcal{G}_j$. □

3. Comparing the present AP with the AP in Ref. [25]—In this section we compare the AP demonstrated in this paper (which we name "present AP") with the AP in Ref. [25] (which we name "original AP"), demonstrating that the present AP leads to a significantly tighter bound on the VD.

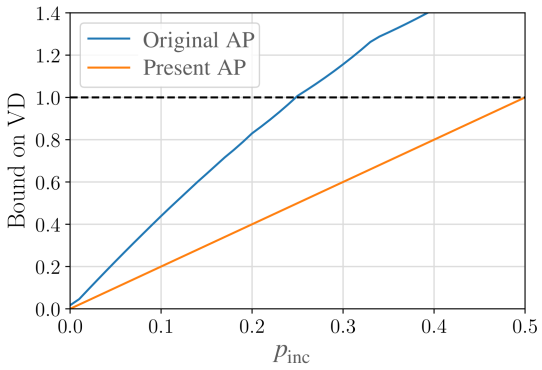


Figure 8. The best value of provided by the original AP (equation 41) and the bound provided by the refined AP (r.h.s. of Eq. 1) as functions of $p_{\rm inc}$.

In the original AP the user implements the target circuit together with a v trap circuits, initialized in the same way as the trap circuits in the present AP. After implementing all the circuits, the output of the target circuit is accepted only if all the trap circuits return the correct output, otherwise it is discarded. The main result proven in Ref. [25] is that the VD between the probability distribution of the accepted outputs $\{p^{\rm acc}_{\rm exp}(\bar{s})\}$ and the ideal probability distribution $\{p_{\rm ideal}(\bar{s})\}$ can be bounded as

$$\frac{1}{2}\sum_{\bar{s}} \left|p_{\rm ideal}(\bar{s}) - p^{\rm acc}_{\rm exp}(\bar{s})\right| \leq \frac{\kappa}{(v+1){\rm prob(acc)}} \, , \tag{39}$$

where $\kappa \approx 1.7$ is a constant and prob(acc) is the probability that the output of the target circuit is accepted (which can be measured by running the AP multiple times with the same target and the same number of traps).

To prove that the present AP leads to a better bound than the original AP, we now rewrite the r.h.s. of Eq. 39 as a function of the total probability $p_{\rm inc}$ that a trap circuit returns an incorrect output. The probability that all the traps return the correct output is prob(acc) $= (1-p_{\rm inc})^v$, which gives

$$\frac{1}{2}\sum_{\bar{s}} \left|p_{\rm ideal}(\bar{s}) - p^{\rm acc}_{\rm exp}(\bar{s})\right| \leq \frac{\kappa}{(v+1)(1-p_{\rm inc})^v} \tag{40}$$

The r.h.s. of the above inequality depends on the number v of traps. All the values obtained at different v are valid upper bounds on the VD. The smallest value

$$\eta_{\rm best} = \min_v \frac{\kappa}{(v+1)(1-p_{\rm inc})^v} \tag{41}$$

corresponds to the best upper bound and can be calculated by implementing the AP many times for different values of v.

In Fig. 8 we plot the bounds provided by present AP and original AP as functions of $p_{\rm inc}$. As it can be seen,

for all the values of p_{inc} the latter bound is larger than the former one approximately by a factor 2. Moreover, the bound provided by the original AP exceeds unity for all $p_{\text{inc}} \gtrsim 0.25$, while that provided by the present AP only exceeds unity for $p_{\text{inc}} \geq 0.5$.

While the present AP yields tighter bounds on the VD, the original AP has been proven to be robust to a more general noise model. Indeed, the noise model assumed in Ref. [25] encompasses arbitrary coupling between system and environment, allowing for time-correlated noise. There is thus a trade-off between the generality of noise models captured and the tightness of the bounds obtained.