Efficient learning of quantum noise

Robin Harper 1, Steven T. Flammia 1,2✉ and Joel J. Wallman 3,4

Noise is the central obstacle to building large-scale quantum computers. Quantum systems with sufficiently uncorrelated and weak noise could be used to solve computational problems that are intractable with current digital computers. There has been substantial progress towards engineering such systems 9,10. However, continued progress depends on the ability to characterize quantum noise reliably and efficiently with high precision 10. Here, we describe such a protocol and report its experimental implementation on a 14-qubit superconducting quantum architecture. The method returns an estimate of the effective noise and can detect correlations within arbitrary sets of qubits. We show how to construct a quantum noise correlation matrix allowing the easy visualization of correlations between all pairs of qubits, enabling the discovery of long-range two-qubit correlations in the 14-qubit device that had not previously been detected. Our results are the first implementation of a provably rigorous and comprehensive diagnostic protocol capable of being run on state-of-the-art devices and beyond. These results pave the way for noise metrology in next-generation quantum devices, calibration in the presence of crosstalk, bespoke quantum error-correcting codes 10 and customized fault-tolerance protocols 11 that can greatly reduce the overhead in a quantum computation.

Useful large-scale quantum computers will require both careful calibration to reduce errors and some form of error correction before universal quantum computing can be realized. Due to crosstalk, optimal calibrations of gates depend on the other gates that are being implemented, which can reduce the overall system error rate by an order of magnitude 27. Furthermore, error-correction routines rely on knowing what the most likely error sources are. Error-correction routines that are optimized for the specific noise in a system can markedly outperform generic ones 11,12.

The calibration and error correction necessary for useful large-scale quantum computing therefore depends upon the ability to characterize the noise in large quantum systems. This characterization will become increasingly important as the field continues to progress 12. Unfortunately, current methods of characterizing noise are infeasible or report only a single summary statistic for how errors affect large-scale quantum devices. Some of the techniques related to RB provide more tomographic information 19–25. However, the protocol of ref. 26 is not robust to state preparation and measurement (SPAM) errors, while the protocols of refs. 26,27 lack the crucial digital processing steps needed to efficiently estimate correlations within arbitrary sets of qubits.

Here, we develop and experimentally implement a protocol based on the general method of ref. 26 that allows us to learn a complete description of the observed error rates (see Supplementary Information for a full definition) in a large-scale quantum device. Where the device is too large to allow a complete description of these error rates to be written or sufficient data to be practically gathered (say, ≳20 qubits), we show how to model the system in such a way that, with only mild and physically plausible assumptions, we can reconstruct the effective observed error rates to arbitrary model fidelity. The protocol is efficient in n, the number of qubits, and comes with mathematically rigorous guarantees on its convergence and performance, making only mild and physically plausible assumptions. Furthermore, the method is immune to systematic bias due to SPAM errors, and achieves both high precision and accuracy.

For any given noisy quantum system comprising n qubits, we can consider the average noise to have the special form of a Pauli channel 28. Although not every noise channel is a Pauli channel, practical methodologies have been developed to transform the noise to be exceptionally well approximated by a Pauli channel without introducing new errors 29,30. A Pauli channel $\mathcal{E}$ acting on a quantum state $\rho$ is of the form $\mathcal{E}(\rho) = \sum p(j) P_j \rho P_j^\dagger$, where $p(j)$ is the error rate associated with the Pauli operator $P_j$. The $p(j)$ form a probability distribution. They are closely related to, but distinct from, the eigenvalues of the Pauli channel, which are defined to be $\lambda(j) = 2^{-n} \text{Tr}(P_j \mathcal{E}(P_j))$. Thus, when a state $\rho$ is subjected to the noisy channel $\mathcal{E}$, $p(j)$ describes the probability of a multiqubit Pauli error $P_j$ affecting the system, while the respective eigenvalue describes how faithfully a given multispin Pauli operator is transmitted.

A protocol to measure a symmetrized characterization of the noise to additive precision was presented by Emerson et al. 26. Obtaining a complete description of the Pauli error rates requires learning more than just the single-Pauli eigenvalues or single-Pauli error rates 31, that is, those associated with the single-qubit Pauli operators such as $\sigma_x^{(1)}$ or $\sigma_z^{(1)}$. A complete description requires learning all of the noise correlations in the system, that is, how the probabilities of multiqubit Pauli operators, for example, $\sigma_x^{(1)} \otimes \sigma_z^{(2)} \otimes \sigma_x^{(3)}$, differ from the probabilities predicted under independent local noise. Knowing these correlations is essential for removing unwanted correlated errors 32, and for performing optimal quantum error correction 12. The number of all possible noise cor-

1Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney, New South Wales, Australia. 2Yale Quantum Institute, Yale University, New Haven, CT, USA. 3Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada. 4Quantum Benchmark Inc., Kitchener, Ontario, Canada. ✉e-mail: sflamnia@gmail.com
results 1/2 random, but is rather an inversion gate returning the system to its initial state in the ideal case. Step 2 gathers output statistics from the measurement free from SPAM error. With a final reverse transformation in step 5, we reconstruct the entire list of effective observed error rates.

In step 1 we determine whether we wish to examine the correlations between qubits activated in single-qubit mode or when gates between the qubits are also used. Each red box in the quantum circuits of step 1 denotes a sequence of random Clifford gates, and the circuits themselves are run for varying lengths $m$ (with time moving from right to left). The final gate $C$ is not random, but is rather an inversion gate returning the system to its initial state in the ideal case. Step 2 gathers output statistics from the measurement results $x$, obtained in the experiments in step 1, and the empirically estimated probability distributions for these outcomes are transformed in step 3 via a Walsh–Hadamard transform. The transformed values are each fitted to an exponential decay in step 4, allowing us to reconstruct the averaged eigenvalues free from SPAM error. With a final reverse transformation in step 5, we reconstruct the entire list of effective observed error rates.

The protocol presented in Fig. 1 reconstructs the full probability distribution with a number of experiments that scales polynomially in $n$, but requires computational resources that scale with the (generally exponential) number of error rates to be estimated. The number of observed error rates scales as $2^n$, so to make our protocol truly scalable we need a method to estimate an efficient description of the noise that nonetheless captures the correlations in a transparent, systematic and physically motivated way.

To achieve an efficient protocol, scaling polynomially in $n$, we introduce the notion of a Gibbs random field (GRF) to describe the $p(j)$. A GRF is a strictly positive probability distribution that obeys certain conditional independence properties known as Markov conditions (Methods). These conditions restrict the range of possible correlations enough to make the problem of noise characterization tractable, but they allow sufficient expressive power that a GRF can accurately model noise in real devices. The underlying (but testable) assumption is that realistic devices will only have correlations between a bounded number of qubits.

When quantum noise is approximated by a GRF, the parameters of the GRF can be learned efficiently, precisely and accurately by only estimating the marginal distributions on the factors and their neighbours. When the factors have a small bounded size, then the protocol detailed in Fig. 1 applied to the subsets of qubits for each factor performs this estimation with aplomb. The specific methodology for estimating the global probability distribution as a GRF from the estimated marginals is discussed in Methods and Supplementary

---

### Table 1: Empirically Estimated Probabilities

| $n$ | $m$ | Probability Distribution Averages | $x_n$ | $x_1$ |
|-----|-----|-----------------------------------|-------|-------|
| 0   | 1   | 0.8956                           | 0.0025 | 0.18   |
| 1   | 1   | 0.0003                           | 0.0001 | 0.0000 |
Fig. 2 | Results from the protocol run in single-qubit mode. a. Spatial layout of the qubits in the 14-qubit Melbourne architecture. Edges in the schematic graph correspond to qubit pairs that can be coupled via a two-qubit gate. The factor graph below this is for a GRF that models quantum noise via spatially local correlations in the device. Each diamond-shaped node is a factor that can describe arbitrary correlations among the qubits connected to it. b. We can use the reconstructed probability distribution to look at the correlations between the probabilities of an error on each of the qubits, where the probability of an error on a particular qubit is treated as a random variable. These correlations can be plotted in the form of a correlation matrix calculated from the globally estimated distribution. Correlation matrices are always symmetric, so we plot separately the correlations from the global estimate (lower left) and the GrF reconstruction (upper right), assuming the factor graph in a. Grey background indicates a value of zero, and white (black) boxes indicate positive (negative) values in the range $[-1, 1]$ in proportion to their relative area. c. An example of how the spatial correlation of the qubits on the device translates to the layout in the correlation matrix. This example shows that, although qubit 1 and qubit 13 are spatially adjacent, they are not adjacent in the matrix. Right: the convention used to indicate error bars, using here $1\sigma$ bounds.

Information. Note that it is not necessary to know the topology of the GRF factor graph in advance; rather, it should be possible to learn the topology of a GRF that adequately describes the observed probabilities from the data themselves using techniques from classical machine learning. We discuss this briefly in Supplementary Information but it remains an open problem, although we aim to show how to learn the topology in future work.

Our first experiments were run using the single-qubit protocol on the 14-qubit superconducting quantum architecture Melbourne, made available by IBM through the IQX online quantum computing environment. After completion of stage 4 of the protocol, we
Our experiments give the first demonstration of a protocol that is practical, relevant and immediately applicable to characterizing error rates and correlated errors in present-day devices with a large (>10) number of qubits. This protocol opens myriad opportunities for novel diagnostic tools and practical applications. For example, structure learning heuristics can be applied to try to learn the most accurate and efficient GRF noise model that describes the error rates. In addition to the applications mentioned in the abstract, machine learning for fine-tuned error-correction decoders using the actual noise map of the device, quantum control for optimal gate synthesis, and noise-aware circuit compiling techniques are just some of the applications of this new method for characterizing quantum noise.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41567-020-0992-8.

Received: 25 October 2019; Accepted: 1 July 2020; Published online: 10 August 2020

References
1. Schindler, P. et al. Experimental repetitive quantum error correction. Science 332, 1059–1061 (2011).
2. Reed, M. D. et al. Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).
3. Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
4. Nigg, D. et al. Quantum computations on a topologically encoded qubit. *Science* **345**, 302–305 (2014).
5. Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).
6. Córcoles, A. D. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.* **6**, 6979 (2015).
7. Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441–445 (2016).
8. Linke, N. M. et al. Fault-tolerant quantum error detection. *Sci. Adv.* **3**, e1701074 (2017).
9. Martinis, J. M. Qubit metrology for building a fault-tolerant quantum computer. *npj Quantum Inf.* **1**, 15005 (2015).
10. Tuckett, D. K., Bartlett, S. D. & Flammia, S. T. Ultrahigh error threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
11. Aliferis, P. & Preskill, J. Fault-tolerant quantum computation against biased noise. *Phys. Rev. A* **78**, 052331 (2008).
12. Neill, C. et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science* **360**, 195–199 (2018).
13. Chuang, I. L. & Nielsen, M. A. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455–2467 (1997).
14. Blume-Kohout, R. et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.* **8**, 14485 (2016).
15. Kimmel, S., da Silva, M. P., Ryan, C. A., Johnson, B. R. & Ohki, T. Robust extraction of tomographic information via randomized benchmarking. *Phys. Rev. X* **4**, 011050 (2014).
16. Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105**, 150401 (2010).
17. Flammia, S. T., Gross, D., Liu, Y.-K. & Eisert, J. Quantum tomography via compressed sensing: error bounds, sample complexity, and efficient estimators. *New J. Phys.* **14**, 095022 (2012).
18. Ribeiro, C. A. et al. Experimental quantum compressed sensing for a seven-qubit system. *Nat. Commun.* **8**, 15305 (2017).
19. Emerson, J., Alicki, R. & Zyczkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B* **7**, S347 (2005).
Methods

Analysis of probability distributions. In this Letter we use various metrics to analyse the probability distribution returned by the protocol. Here, we formally define the terms we use. The relative entropy (also known as the Kullback–Leibler divergence) between two probability distributions is one measure, and is defined as

$$D(p||q) = \sum p(x) \ln \frac{p(x)}{q(x)}$$

(1)

The mutual information is a measure of the dependence between two random variables $X$ and $Y$, and it quantifies the amount of information obtained regarding one variable through observing the other. It is defined as

$$I(X;Y) = D(p(x,y)||p(x)p(y))$$

(2)

$$= \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

(3)

where $D(p(x,y)||p(x)p(y))$ is the relative entropy between the joint probability distribution $p(x,y)$ and the marginal distribution on each of the random variables.

For each run of the protocol, we have a partitioning of the set of 14 qubits into disjoint sets $s$ that are acted upon by independent twirls. For each $s$ we define a random variable $Q_s$ that takes on the value 1 if no error acts on $s$ and 0 otherwise. For the single-qubit protocol, we have 14 random variables $Q_0, \ldots, Q_{13}$. We then calculate the mutual information between each pair of random variables. For the two-qubit protocol, $s$ comprises one or two qubits, depending on whether two-qubit gates were used on that pair or not.

Conditional mutual information represents the expected value of the mutual information of two random variables conditioned on the value of the third. In the present case we have

$$I(X;Y|Z) = \sum_{x,y} p(x,y|z) \ln \frac{p(x,y|z)}{p(x|z)p(y|z)}$$

(4)

$$= \sum_{x,y,z} p(x,y,z) \ln \frac{p(x,y,z)}{p(x,z)p(y,z)}$$

(5)

One can use the probability distribution to compute the covariance matrix between the observed error random variables. In the experiment presented in this paper, where we have observed errors, we can treat the qubits as 0/1 random variables representing no error/error. Then, if $x$ is a column vector representing an error pattern, we compute the covariance matrix $\Sigma$ as

$$\Sigma = E[xx^T]$$

(6)

where $x = E[x]$ and $E$ denotes the expected value over the probability distribution.

Given the covariance, it is simple to calculate the Pearson product–moment correlation coefficient matrix $Q$, obtained by dividing the covariance of the two variables by their standard deviation. This is the correlation matrix plotted in this Letter. In this case let $V = \text{diag}(\Sigma)$. Then we have

$$Q = V^{-1/2} \Sigma V^{-1/2}$$

(7)

In the case where we average over the Pauli group (instead of the Clifford group) to characterize the Pauli noise of the device (rather than the basis-averaged Pauli noise of the device presented here in our experiments), then the relevant qubit random variable can be characterized by the 2-qubit noise of the device presented here in our experiments), then the relevant group) to characterize the Pauli noise of the device (rather than the basis-averaged variables representing no error/error. Then, if this paper, where we have observed errors, we can treat the qubits as 0/1 random variables.

For the current experiment we can use the topology of the device to define a Markov network, where the connections in the graph are identical to the resonators between the qubits. The graph appears in Fig. 2. If we wish to enforce short-range correlations then we have

$$p(0,1,13,2,12,3,11,4,10,5,9,6,8,7) = p(0,1,13)$$

(12)

$$p(1,13,0,12,3,11,4,10,5,9,6,8,7) = p(1,13,0,12,3,11,4,10,5,9,6,8,7)$$

(12)

$$p(2,12,0,13,3,11,4,10,5,9,6,8,7) = p(2,12,0,13,3,11,4,10,5,9,6,8,7)$$

(12)

$$p(3,11,0,1,13,2,12,4,10,5,9,6,8,7) = p(3,11,0,13,2,12,4,10,5,9,6,8,7)$$

(12)

$$\ldots$$

(12)

where we have used numbers to represent random variables, for example 0 for $x_0$, to declutter the notation.

We can then use the chain rule to write the joint probability distribution as follows:

$$p(0,1,13) = p(0,13)p(1,3)$$

(12)

$$p(0,1,13) = p(0,13)p(1,3)$$

(12)

$$\ldots$$

(12)

then using equation (12), this simplifies to

$$p(0,1,13)p(1,3) = p(0,13)p(1,3)$$

(12)

$$\ldots$$

(12)

which we can calculate as

$$p(0,1,13)p(1,3) = p(0,13)p(1,3)$$

(12)

$$\ldots$$

(12)

Finally, we can use the probability distribution defined in equation (13) to reconstruct the GRF, that is, the probability distribution induced by the condition imposed by our chosen Markov conditions. A comparison between the observed probability distribution and that induced by, say, the GRF gives us a metric to observe how closely our device corresponds to a device that only has short-range correlations.

GRFs. Associated with any GRF is a factorization of the error rates, where each factor is a positive function depending only on a subset of the qubits. The factor graph, depicted in Fig. 2, has two types of node: one for the random variables and a subset

$$H(p, q) = \left(1 - \sum \sqrt{p(q)|q|}\right)^{1/2}$$

(9)

This distance is efficient to compute and is related to the more commonly used but difficult to compute notion of statistical distance (or total variational distance

$$\delta(p, q)$$

as

$$H^2(p, q) \leq \delta(p, q) \leq 2H(p, q)$$

(10)

where

$$\delta(p, q) = \frac{1}{2} \sum |p(j) - q(j)|$$

(11)

IBM Quantum Experience. The experiments reported here were conducted on the IBM Quantum Experience Melbourne device. Jobs were submitted via Qiskit” in two separate runs. The single-qubit experiment consisted of 1,000 different submissions. Each submission contained 11 single-qubit Clifford twirled on each of the 14 qubits for gate lengths $1, 5, 10, 15, 20, 30, 45, 60, 75, 90, 105$, with each gate length sequence requesting 1,024 shots. This meant in total 1,024,000 x 2 jobs were performed. For step 1 of the protocol (Fig. 1) simple least-squares fitting was used, although for any sequence length where the value was less than $(p + 1/16)$ those data and the data from longer sequences were discarded for the purposes of the fit.

With the two-qubit protocol the topology of the Melbourne device does not allow seven qubit pairs to be operated simultaneously. The qubits were grouped in a GRF as a distinct qubit pair, the remaining two qubits being operated in single-qubit-protooc mode. To determine which qubit pairs to activate, attention was paid to the reported fidelities of the CNOT gates. For instance, on 18 February
2019 the two-qubit gate between qubits 13 and 1 had a reported fidelity of 84.1% (other gates could be as high as 97%). The groupings shown in Fig. 3 were chosen to attempt to ensure wide coverage of the various two-qubit gates available while avoiding any gates with a reported fidelity close to or below 90%. Only one of the three configuration settings is reported in this Letter; the others appear in Supplementary Information. Because of the possibility that correlations within the device might change with recalibration, all the different runs shown in Fig. 3 were interleaved. In total there were 1,924 different submissions (of 11 different sequence lengths) for each of the three different configurations. One recalibration cycle did occur during the gathering of the data, although the fidelities of the two-qubit gates did not appear to change substantially as a result. The data do, however, represent an average of the noise that occurred in the machine during that time period. Given the reduced fidelity of the runs (since two-qubit gates have an infidelity an order of magnitude greater than that of the single-qubit gates), sequence lengths were reduced to 0...10. The 0 sequence, representing one single-qubit Clifford gate (that is, no two-qubit gate), was added to allow a more accurate determination of the A constant in the fit. As with the single-qubit runs, data with a value less than \( p_0 + \frac{1}{16}/4 \) were discarded for the purposes of the fit, although a minimum of three data points were retained.

Finally, in all cases an X gate was randomly compiled into the qubits of each sequence submitted, with the probability distribution interpreted accordingly to eliminate any bias in the SPAM\(^6\).

**Error bars.** All error bars shown here were calculated using non-parametric bootstrap methods. For each sequence length in each run the observed probability distributions of the measurement counts (step 2 in the protocol shown in Fig. 1) were resampled (with replacement) for the same number of measurements as used to originally gather the data. This was repeated 1,000 times. Each of these 1,000 sets of resampled distributions was then analysed in a manner identical to the original (steps 3, 4 and 5 of the protocol), to provide 1,000 bootstrapped samples of the SPAM-free probability vector (the bootstrapped probabilities). From this the appropriate confidence intervals to provide error bars can be constructed. With the mutual information estimates the mutual information between the qubits in question can be calculated 1,000 times and ordered, and by extracting the values at the appropriate location of the ordered values the confidence intervals are obtained (so for 1σ confidence intervals the 159th and 841st values are used). With the error bars on the correlation matrices the following conservative approach was adopted. Using the bootstrapped probabilities, 1,000 correlation matrices were constructed. Since there is no clear way to order such matrices, each individual cell on the matrices was treated separately, with the possible values for that cell location being ordered and the appropriate high/low values being extracted as before. While a matrix constructed from all the low (or high) values would not in itself be a valid correlation matrix, it is believed that the error bars still, conservatively, convey the confidence intervals for each of the individual values in the matrix. Finally, with the calculation of the JSD two different resampling techniques were utilized. In the first the GRF-reconstructed probability distribution for the originally observed distribution was compared with each of the bootstrapped probability distributions, and in the second a GRF distribution was constructed from each bootstrapped probability distribution and compared with the full bootstrapped distribution from which it was constructed. In all cases the error ranges were broadly similar and in this Letter the uncertainties quoted were taken from largest error ranges from either of the methodologies.

**Data availability**
Source data are available for this paper at https://github.com/rharper2/EfficientLearningDataSet. All other data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

**Code availability**
Details of code that was used to analyse the data is available from the corresponding author upon reasonable request.

**References**
37. Harper, R., Hincks, I., Ferrie, C., Flammini, S. T. & Wallman, J. J. Statistical analysis of randomized benchmarking. *Phys. Rev. A* **99**, 052330 (2019).

**Acknowledgements**
We thank S. Bartlett, A. Doherty, J. Emerson and T. Monz for comments on an earlier draft. This work was supported in part by US Army Research Office grant nos. W911NF-14-1-0098 and W911NF-14-1-0103, the Australian Research Council Centre of Excellence for Engineered Quantum Systems (EQUUS) CE170100009, the Government of Ontario, and the Government of Canada through the Canada First Research Excellence Fund (CFREF) and Transformative Quantum Technologies (TQT), Natural Sciences and Engineering Research Council (NSERC), Industry Canada.

**Author contributions**
R.H., S.T.F. and J.J.W. conceived the experiments, and S.T.F. and J.J.W. conceived the original methodology. The implementation was carried out by R.H. R.H. wrote the initial draft and all authors contributed to the revisions and editing of the manuscript.

**Competing interests**
J.J.W. is the chief technology officer of the company Quantum Benchmark, Inc., and S.T.F. and R.H. were both consultants for it for part of the duration of this project.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41567-020-0992-8.

Correspondence and requests for materials should be addressed to S.T.F.

Peer review information *Nature Physics* thanks Jonas Bylander, Diego Risté, Marcus da Silva and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.