ARTICLE OPEN

Demonstration of quantum advantage in machine learning

Diego Ristè[1], Marcus P. da Silva[1], Colm A. Ryan[1], Andrew W. Cross[2], Antonio D. Córcoles[2], John A. Smolin[2], Jay M. Gambetta[2], Jerry M. Chow[2] and Blake R. Johnson[1]

The main promise of quantum computing is to efficiently solve certain problems that are prohibitively expensive for a classical computer. Most problems with a proven quantum advantage involve the repeated use of a black box, or oracle, whose structure encodes the solution. One measure of the algorithmic performance is the query complexity, i.e., the scaling of the number of oracle calls needed to find the solution with a given probability. Few-qubit demonstrations of quantum algorithms, such as Deutsch–Jozsa and Grover, have been implemented across diverse physical systems such as nuclear magnetic resonance, trapped ions, optical systems, and superconducting circuits. However, at the small scale, these problems can already be solved classically with a few oracle queries, limiting the obtained advantage. Here we solve an oracle-based problem, known as learning parity with noise, on a five-qubit superconducting processor. Executing classical and quantum algorithms using the same oracle, we observe a large gap in query count in favor of quantum processing. We find that this gap grows by orders of magnitude as a function of the error rates and the problem size. This result demonstrates that, while complex fault-tolerant architectures will be required for universal quantum computing, a significant quantum advantage already emerges in existing noisy systems.

npj Quantum Information (2017)3:16 ; doi:10.1038/s41534-017-0017-3

INTRODUCTION

The limited size of engineered quantum systems and their extreme susceptibility to noise sources have made it hard so far to establish a clear advantage of quantum over classical computing. Although the classical success probability has been exceeded in two-qubit demonstrations of the Deutsch–Jozsa[1] and Grover[2] algorithms, the required number of oracle queries has so far remained comparable. A promising avenue to highlight a quantum advantage is offered by a new family of algorithms designed for machine learning.[3–6] In this class of problems, artificial intelligence methods are employed to discern patterns in large amounts of data, with little or no knowledge of underlying models. A particular learning task, known as binary classification, is to identify an unknown mapping between a set of bits onto 0 or 1. An example of binary classification is identifying a hidden parity function,[7, 8] defined by the unknown bit-string $\boldsymbol{k}$, which computes $f(\boldsymbol{D},\boldsymbol{k}) = \boldsymbol{D}\cdot\boldsymbol{k}$ mod 2 on a register of n data bits $\boldsymbol{D} = \{D_1,D_2\ldots,D_n\}$ (Fig. 1a). The result, i.e., 0 (1) for even (odd) parity, is mapped onto the state of an additional bit A. The learner has access to the output register of an *example oracle* circuit that implements f on random input states, on which he/she has no control. Repeated queries of the oracle allow the learner to reconstruct $\boldsymbol{k}$. However, any physical implementation suffers from errors, both in the oracle execution itself and in readout of the register. In the presence of errors, the problem becomes hard. Assuming that every bit introduces an equal error probability, the best known algorithms have a number of queries growing as $\mathcal{O}(n)$ and runtime growing almost exponentially with n.[7–9] In view of the classical hardness of learning parity with noise (LPN), parity functions have been suggested as keys for secure and computationally easy authentication.[10, 11]

The picture is different when the oracle is implemented by a quantum circuit and the algorithm can process quantum superpositions of input states. In this case, applying a coherent operation on all qubits after an oracle query ideally creates the entangled state

$$\left(\left|0_A 0^n_{\mathbf{D}}\right\rangle + \left|1_A \boldsymbol{k}_{\mathbf{D}}\right\rangle\right)/\sqrt{2}. \tag{1}$$

In particular, when A is measured to be in $|1\rangle$, $|D\rangle$ will be projected onto $|\boldsymbol{k}\rangle$. With constant error per qubit, learning from a quantum oracle requires a number of queries that scales as $\mathcal{O}(\log n)$, and has a total runtime that scales as $\mathcal{O}(n)$.[12] This gives the quantum algorithm an exponential advantage in query complexity and a super-polynomial advantage in runtime.

In this work, we implement a LPN problem in a superconducting quantum circuit using up to five qubits, realizing the experiment proposed in Ref. 12. We construct a parity function with bit-string $\boldsymbol{k}$ using a series of CNOT gates between the ancilla and the data qubits (Fig. 1b). We then present two classes of learners for $\boldsymbol{k}$ and compare their performance. The first class simply measures the output qubits in the computational basis and analyzes the results. The measurement collapses the state into a random $\{\boldsymbol{D}, f(\boldsymbol{D},\boldsymbol{k})\}$ basis state, reproducing an example oracle of the classical LPN problem. The second class performs some quantum computation (coherent operations), followed by classical analysis, to infer the solution. We show that, beyond a minimum complexity of the problem, the quantum approach outperforms the classical one. Furthermore, as the classical problem becomes rapidly intractable as noise is added to the output register, the performance gap widens.

RESULTS

The quantum device used in our experiment consists of five superconducting transmon qubits, A, D_1, …, D_4, and seven microwave resonators (Fig. 1c). Five of the resonators are used

[1]Raytheon BBN Technologies, Cambridge, MA 02138, USA and [2]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
Correspondence: Diego Ristè (diego.riste@raytheon.com)

Received: 1 September 2016 Revised: 6 February 2017 Accepted: 10 March 2017
Published online: 13 April 2017

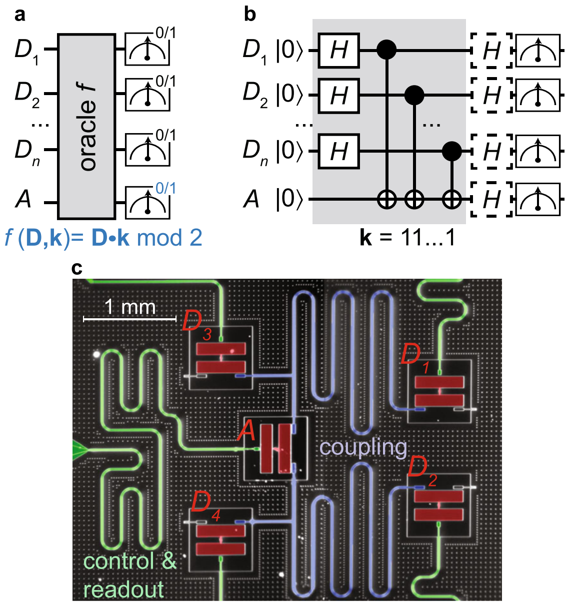

Fig. 1 Implementation of a parity function in a superconducting circuit. **a** Conceptual diagram of parity learning. The (classical or quantum) oracle f ideally maps the parity of a subset of n data bits (or qubits), defined by the bit string $\boldsymbol{k}$, into bit A. Repeated queries of the oracle allow the reconstruction of $\boldsymbol{k}$ by reading the output register. **b** Gate sequence implementing a quantum parity oracle with $\boldsymbol{k} = 11\ldots1$. Random examples are generated by preparing the data qubits $\{D_1,\ldots,D_n\}$ in a uniform superposition. *Vertical lines* indicate CNOT gates between each D_i (control) and the ancilla qubit A (target). Quantum learning differs from classical learning only by the addition of single-qubit gates (*dashed boxes*) applied before measurement (see also Supplementary Information). **c** Optical image of the superconducting quantum processor (*qubits in red*). A is coupled to each D_i by means of two bus resonators (*blue*). Each qubit is also coupled to a dedicated resonator for control and readout (*green*)[27]

for individual control and readout of the qubits, to which they are dispersively coupled.[13] The center qubit A plays the role of the result and is coupled to the data register $\{D_i\}$ via the remaining two resonators. This coupling allows the implementation of cross-resonance (CR) gates[14] between A (used as control qubit) and each D_i (target), constituting the primitive two-qubit operation for the circuit in Fig. 1b (full gate decomposition in the Supplementary Information). Each qubit is measured by probing its respective readout resonator with a near-resonant microwave pulse. The output signals are then demodulated and integrated at room temperature to produce the homodyne voltages $\{V_{D_1}, \ldots V_{D_n}, V_A\}$ (see Supplementary Information for the detailed experimental setup).

To implement a uniform random example oracle for a particular $\boldsymbol{k}$, we first prepare the data qubits in a uniform superposition (Fig. 1b). Preparing such a state ensures that all parity examples are produced with equal probability and is also key in generating a quantum advantage. We then implement the oracle as a series of CNOT gates, each having the same target qubit A and a different control qubit D_i for each $k_i = 1$. Finally, the state of all qubits is read out (with the optional insertion of Hadamard gates, see discussion below). The oracle mapping to the device is limited by imperfections in the two-qubit gates, with average fidelities 88–94%, characterized by randomised benchmarking[15] (see Supplementary Table S1). Readout errors in the register η_{D_i}, defined as the average probability of assigning a qubit to the wrong state, are limited to 20–40% by the onset of inter-qubit crosstalk at higher measurement power (see data in the Supplementary Information). A Josephson parametric amplifier[16] in front of the amplification chain of A suppresses its low-power readout error to $\eta_A = 5\%$.

Having implemented parity functions with quantum hardware, we now proceed to interrogate an oracle N times and assess our capability to learn the corresponding $\boldsymbol{k}$. We start with oracles with register size $n = 2$, involving D_1, D_2, and A. We consider two classes of learning strategies, classical (C) and quantum (Q). In C, we perform a projective measurement of all qubits right after execution of the oracle. This operation destroys any coherence in the oracle output state, thus making any analysis of the result classical. The measured homodyne voltages $\{V_{D_1}, \ldots V_{D_n}, V_A\}$ are converted into binary outcomes, using a calibrated set of thresholds (see Methods). Thus, for every query, we obtain a binary string $\{a,d_1,d_2\}$, where each bit is 0 (1) for the corresponding qubit detected in $|0\rangle$ ($|1\rangle$). Ideally, a is the linear combination of d_1, d_2 expressed by the string $\boldsymbol{k}$ (Fig. 1a). However, both the gates comprising the oracle and qubit readout are prone to errors (see values in the Supplementary Information). To find the $\boldsymbol{k}$ that is most likely to have produced our observations, at each query m we compute the expected $\tilde{a}_{\mathbf{k},m} = \boldsymbol{d}_m{\cdot}\boldsymbol{k}$ mod 2 for the measured $\boldsymbol{D} = \{d_1,d_2\}_m$ and the 4 possible values of $\boldsymbol{k}$. We then select the $\boldsymbol{k}$ which minimizes the Hamming distance to the measured results $a_1,\ldots,a_N$ of N queries, i.e., $\sum_{m=1}^{N} |a_m - \tilde{a}_{\mathbf{k},m}|$.[7] In the case of a tie, $\boldsymbol{k}$ is randomly chosen among those producing the minimum distance. As expected, the error probability p of obtaining the correct answer decreases with N (Fig. 2a). Interestingly, the difficulty of the problem depends on $\boldsymbol{k}$ and increases with the number of $k_i = 1$. This can be intuitively understood as needing to establish a higher correlation between data qubits when the weight of $\boldsymbol{k}$ increases.

Our second approach (Q) takes advantage of the quantum correlations between ancilla and data qubits at the output of the oracle. Instead of directly measuring the qubits as above, we first apply a Hadamard gate on each. These local operations generate quantum interference between terms in the superposition state, ideally producing the desired result (Eq. (1)). This technique is widely used in quantum algorithms to increase the probability of obtaining the desired outcomes.[17] In this case, whenever A is measured to be in $|1\rangle$ (with 50% probability), the data register will ideally be projected onto the solution, $|D_1{,}D_2\rangle = |k_1{,}k_2\rangle$. We therefore digitize and postselect our results on the outcomes where $a = 1$ and perform a bit-wise majority vote on $\{d_1, d_2\}_{1\ldots\tilde{N}}$. Despite every individual query being subject to errors, the majority vote is effective in determining $\boldsymbol{k}$ (Fig. 2b). We assess

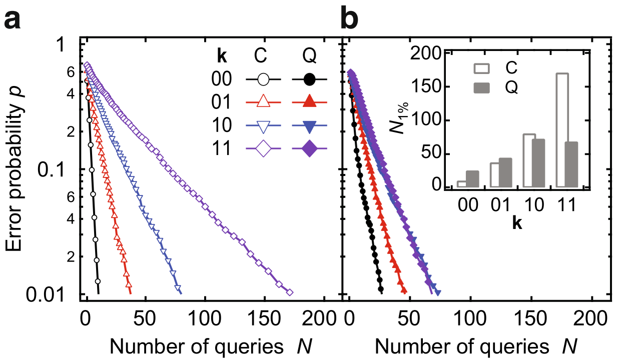

Fig. 2 Error probability p to identify a 2-bit oracle $\boldsymbol{k}$ as a function of the number of queries N. For both classical **a** and quantum **b** learners, one of the four oracles $\boldsymbol{k}$ is applied, followed by the simultaneous measurement of all qubits. Hadamard gates are applied prior to measurement in the quantum case (Fig. 1b). See text for a description of the solvers in the two scenarios. Inset: number of queries $N_{1\%}(\boldsymbol{k})$ required to reach 1% error for the classical (*empty bars*) and quantum (*solid*) solver

the performance of the two solvers by comparing the number of queries $N_{1\%}$ required to reach $p = 0.01$ (Fig. 2c). Whereas Q performs comparably or worse than C for $\boldsymbol{k} = 00$, 01 or 10, Q requires less than half as many queries as C for the hardest oracle, $\boldsymbol{k} = 11$. We note that, while these results are specific to the lowest oracle and readout errors we can achieve, a systematic advantage of quantum over classical learning will become clear in the following.

So far we have adhered to a literal implementation of the classical LPN problem, where each output can only be either 0 or 1. However, the actual measurement results are the continuous homodyne voltages $\{V_{D_1}, \dots V_{D_n}, V_A\}$, each having mean and variance determined by the probed qubit state and by the measurement efficiency.[13] This additional resource can be exploited to improve the learner's capabilities. A more effective strategy for C uses Bayesian estimation to calculate the probability of any possible $\boldsymbol{k}$ for the measured output voltages, and select the most probable (see Methods). This approach is expensive in classical processing time (scaling exponentially with n), but drastically reduces the error probability $\bar{p}$, averaged over all $\boldsymbol{k}$, at any N (Fig. 3). To improve on Q, we still postselect the oracle queries on the digitized outcome $a = 1$. Then, instead of digitizing the corresponding $\{V_{D_i}\}$ as above, we digitize their averages $\{\langle V_{D_i}\rangle\}$, obtaining our best guess for $\boldsymbol{k}$ (see Methods). This procedure simply replaces the majority vote between multiple noisy observations with a single observation, with variance reduced by the number of postselected queries. Using the analog results, not only does Q retain an advantage over C (smaller p for given N), but it does so without introducing an overhead in classical processing.

The superiority of Q over C becomes even more evident when the oracle size n grows from 2 to 3 data qubits (Fig. 3b). Whereas Q solutions are marginally affected, the best C solver demands almost an order of magnitude higher N to achieve a target error. Maximizing the resources available in our quantum hardware, we observe an even larger gap for oracles with $n = 4$ (data in the Supplementary Information), suggesting a continued increase of quantum advantage with the problem size.

As predicted, quantum parity learning surpasses classical learning in the presence of noise. To investigate the impact of noise on learning, we introduce additional readout error on either A or on all D_i. This can be easily done by tuning the amplitude of the readout pulses, effectively decreasing the signal-to-noise ratio.[18] When the ancilla assignment error probability η_A grows (Fig. 4a), the number of queries $\overline{N}_{1\%}$ (the average of $N_{1\%}$ over all $\boldsymbol{k}$) required by the C solver increases by up to 2 orders of magnitude in the measured range (see also data in the Supplementary Information). Conversely, using Q, $\overline{N}_{1\%}$ only changes by a factor of ~3. Key to this performance gap is the optimization of the digitization threshold for $\{\langle V_{D_i}\rangle\}$ at each value of η_A (see Methods). When η_A is increased, an interesting question is whether postselection on V_A remains always beneficial. In fact, for $\eta_\mathrm{A} > 0.25$, it becomes more convenient to ignore V_A and use the totality of the queries (Q' in Fig. 4a).

Similarly, we step the readout error of the data qubits, with average η_D, while setting η_A to the minimum. Not only does Q outperform C at every step, but the gap widens with increasing η_D.

The computational advantage of quantum learning, which appears in the reduction of the number of oracle calls, is even more significant when accounting for the post-processing time. For example, finding $\boldsymbol{k}$ at $\eta_\mathrm{A} = 0.44$ with 1% error (Fig. 4a) takes C about 10 times longer in post-processing relative to Q. Moreover, the processing time for C grows exponentially with n, as the Bayesian solver must track probabilities for each possible $\boldsymbol{k}$. Conversely, Q consists only of binary comparisons (for postselection on A), averages, and a final digitization (for $\boldsymbol{D}$), thus scaling linearly with n.

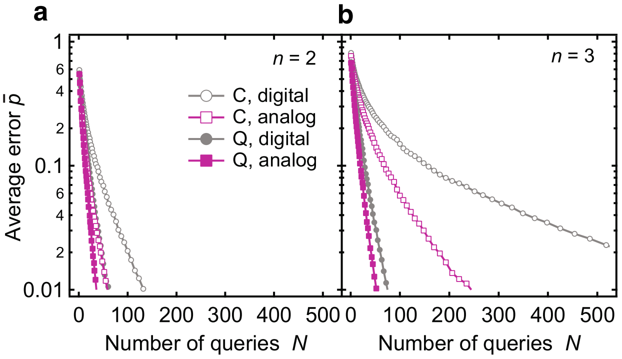

Fig. 3 Learning error probability $\bar{p}$ averaged over all the n-bit oracles $\boldsymbol{k}$, for different n and solvers. **a** $n = 2$, **b** $n = 3$. Making use of the analog measurements $\{V_{D_1}, \dots V_{D_n}, V_A\}$ (*squares*) improves over the digital solvers in Fig. 2 (*circles*) for both classical (*empty symbols*) and quantum (*solid symbols*) learning. The analog solver in Q proves to be the most efficient solution. Moreover, the gap between Q and C grows with n. The same dataset is used in Figs 2 and 3, with D_3 ignored in the analysis for $n = 2$. See Supplementary Information for the $p(N)$ corresponding to each 3-bit $\boldsymbol{k}$

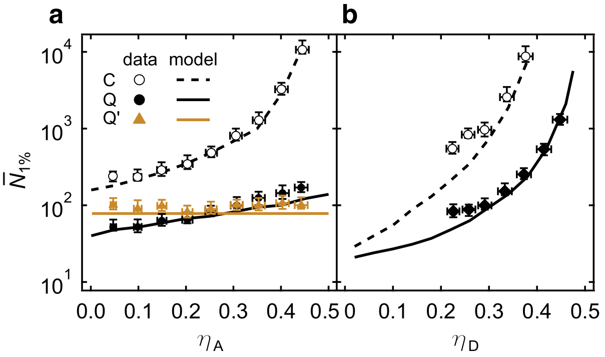

Fig. 4 Robustness of quantum parity learning to noise. Number of queries $\overline{N}_{1\%}$ for $\bar{p} = 0.01$ for variable readout error η of ancilla **a** or data **b** qubits, with $n = 3$. η is tuned by setting the readout power of the corresponding qubit(s). *Empty* (*solid*) *circles* correspond to the analog C (Q) solver. **a**, $\overline{N}_{1\%}$ diverges for $\eta_\mathrm{A} \to 0.5$ for C, while it stays limited for Q. When $\eta_\mathrm{A} \gtrsim 0.25$, it is preferable to ignore V_A altogether (Q', triangles). **b** Whereas both C and Q are severely affected by a noisy data register, Q remains superior and the performance gap increases with η_D. Results are out of scale for C and $\eta_\mathrm{D} \gtrsim 0.4$. The corresponding $N_{1\%}$ are not computed, due to the processing time of several hours that would be required. See Methods for an explanation of the error bars

Finally, to verify that our results are not limited to highly noisy systems, we have implemented all 4-bit $\boldsymbol{k}$ on a second device with lower gate and readout errors, particularly for η_{D_i}. Whereas the required number of queries is greatly reduced for both learners, the performance gap remains in favor of Q (see Supplementary Information).

DISCUSSION

A numerical model including the measured η_A,η_D, qubit decoherence, and gate errors (see Supplementary Information) modeled as depolarization noise is in very good agreement with the measured $N_{1\%}$ at all η_A,η_D. This model allows us to extrapolate $N_{1\%}$ to the extreme cases of zero and maximum noise. Obviously, when $\eta_\mathrm{D} = 0.5$, readout of the data register contains no information, and $N_{1\%}$ consequently diverges. On the other hand a random ancilla result ($\eta_A = 0.5$) does not prevent a quantum learner from obtaining $\boldsymbol{k}$. In this limit, the predicted factor of ~2 in $\overline{N}_{1\%}$ between Q and Q' can be intuitively understood as Q indiscriminately discards half of the queries, while Q' uses all of them. (See Supplementary Information for theoretical bounds on the scaling of $\overline{N}_{1\%}$ for different solvers.)

It is worth noting that the quantum advantage here demonstrated is not limited to a noisy realization of the oracle. Lower gate errors, as achieved by a future fault-tolerant processor, will reduce $N_{1\%}$ for both Q and C solvers. Nevertheless, for a given oracle, classical learning will remain more susceptible to measurement errors (η_A,η_D), preserving the performance gap with Q.[12]

In conclusion, we have implemented a LPN algorithm in a quantum setting. We have demonstrated a superior performance of quantum learning compared to its classical counterpart, where the performance gap increases with added noise in the query outcomes. A quantum learner, with the ability of physically manipulating the output of a quantum oracle, is expected to find the hidden $\boldsymbol{k}$ with a logarithmic number of queries and linear runtime as function of the problem size, whereas a passive classical observer would require a linear number of queries and nearly exponential runtime. We have shown that the difference in classical and quantum queries required for a target error rate grows with the oracle size in the experimentally accessible range, and that quantum learning is much more robust to noise. We expect that future experiments with increased oracle size will further demarcate a quantum advantage, in support of the predicted asymptotic behavior. Furthermore, our experiment provides a novel method to benchmark the performance of a quantum algorithm using the same hardware to construct the equivalent classical problem. As prototype quantum computers continue to grow, we expect this approach to become increasingly useful in determining the quantum advantage attainable in complex problems.

METHODS

Pulse calibration

Single- and two-qubit pulses are calibrated by an automated routine, executed periodically during the experiments. For each qubit, first the transition frequency is calibrated with Ramsey experiments. Second, π and $\pi/2$ pulse amplitudes are calibrated using a phase estimation protocol.[19] The pulse amplitudes, modulating a carrier through an I/Q mixer (diagram in the Supplementary Information) are adjusted at every iteration of the protocol until the desired accuracy or signal-to-noise limit is reached. Pulses have a Gaussian envelope in the main quadrature and derivative-of-Gaussian in the other, with DRAG parameter[20] calibrated beforehand using a sequence amplifying phase errors.[21] A CR$_i$ gate[14, 22] on qubits $\{A,D_i\}$ consists of two pulses applied on A at the D_i frequency, separated by a refocusing π pulse on A. For some frequency conditions (mainly that the qubit-qubit detuning is smaller than their anharmonicity), this sequence implements a D_i rotation, controlled by A. The gate is calibrated in a two-step procedure, determining first the optimum duration and then the optimum phase corresponding to the unitary $CR_i = Z_A X_{D_i}(\pi/2)$.

Experimental setup

A detailed schematic of the experimental setup is illustrated in the Supplementary Information. For each qubit, signals for readout and control are delivered to the corresponding resonator through an individual line through the dilution refrigerator. For an efficient use of resources, we apply frequency division multiplexing[23] to generate the five measurement tones by sideband modulation of three microwave sources. Moreover, the same pair of BBN APS (arbitrary waveform generators) channels produce the readout pulses for $\{D_1,D_2\}$, and another one for $\{D_3,D_4\}$. Similarly, the output signals are pairwise combined at base temperature, limiting the number of HEMTs and digitizer channels to three. The attenuation on the input lines, distributed at different temperature stages, is a compromise between suppression of thermal noise impinging on the resonators (affecting qubit coherence) and the input power required for CR gates.

Gate sequence

CNOT gates can be decomposed in terms of CR gates using the relation $\mathrm{CNOT}_{12} = (Z^{-}_{90} \otimes X^{-}_{90})\mathrm{CR}_{12}$.[24] Moreover, the role of control and target qubits are swapped, using $\mathrm{CNOT}_{12} = (H_1 \otimes H_2)\mathrm{CNOT}_{21}(H_1 \otimes H_2)$. The first of these H gates is absorbed into state preparation for the LPN sequence (Fig. 1a and Supplementary Information). Similarly, when two CNOTs are executed back to back, two consecutive H gates on A are canceled out. In order to maintain the oracle identical in C and Q, we do not compile the H gates in the CNOTs with those applied before measurement in Q.

Sample size

For each set of oracle $\boldsymbol{k}$, readout errors η_A,η_D, solver type, and register size n, we measure the result of 100,000 oracle queries. Each set is accompanied by n+2 calibration points (averaged 10,000 times), providing the distributions of $V_A, V_{D_1}, \ldots, V_{D_n}$ for the collective ground state and for single-qubit excitations (n data and 1 ancilla qubit). These distributions are then used to determine the optimum digitization threshold (for digital solvers) or as input to the Bayesian estimate in C. To obtain $p(N)$, we resample the full data set with 2000–4000 random subsets of each size N.

Statistical analysis

Error bars are obtained by first computing the credible intervals for p at each set $\{N,\boldsymbol{k},\eta_A,\eta_D\}$. These intervals are computed with Jeffreys beta distribution prior Beta$(\frac{1}{2},\frac{1}{2})$ for Bernoulli trials, with a credible level of 100% −(100−95%)/8≈99.36%. This ensures that, under a union bound, the average of estimates for 8 different $\boldsymbol{k}$ is inside the credible interval with a probability of at least 95%. We then perform antitonic regression on the upper and lower bounds of the credible intervals to ensure monotonicity as function of N, and find the intercept to $p = 0.01$ for each $\boldsymbol{k}$. The bounds on the value $\overline{N}_{1\%}$ averaged over $\boldsymbol{k}$ is computed by interval arithmetic on the credible intervals of $N_{1\%}$ for each $\boldsymbol{k}$.

Classical solver with Bayesian estimate

An improved classical solver for the LPN problem can be constructed when the oracle provides an analog output. Approximating the distributions of each bit value as Gaussian[25] (neglecting qubit transitions during readout), this solver corresponds to a Bayesian estimate of $\boldsymbol{k}$ after a series of observations of the data and ancilla bits. More formally, taking a uniform prior distribution for all binary strings produced by the oracle, one computes the (unnormalized) posterior $p(D_i)$ distribution for each data bit D_i the output of the oracle,

$$p(D_i = b|V_{D_i}) = \frac{1}{2}\exp\left[-\frac{(V_{D_i} - b)^2}{2\sigma_i^2}\right]$$

The (unnormalized) posterior distribution $p_m(\boldsymbol{k}|\boldsymbol{V}_\mathrm{D}, V_A)$ for $\boldsymbol{k}$ after the mth query, on the other hand, is given by

$$p_m(\boldsymbol{k}|\boldsymbol{V}_\mathrm{D}, V_A) = \exp\left[-\frac{(V_A - \boldsymbol{D}\cdot\boldsymbol{k})^2}{2\sigma_A^2}\right]p(\boldsymbol{D}|\boldsymbol{V}_\mathrm{D})p_{m-1}(\boldsymbol{k}),$$

where $p_0(\boldsymbol{k})$ is the prior distribution. Here and above, $\{V_{D_1}, \ldots V_{D_n}, V_A\}$ are rescaled to have mean 0 and 1 for the corresponding qubit in $|0\rangle$ and $|1\rangle$, respectively. Iterating this procedure (while updating $p(\boldsymbol{k})$ at each iteration), and then choosing the most probable $\boldsymbol{k}_\mathrm{Bayes} = \mathrm{argmax}_\mathbf{k} p(\boldsymbol{k})$, one obtains an estimate for $\boldsymbol{k}$.

Analog quantum solver with postselection on A

While postselection on A is performed equally on both digital (Fig. 2) and analog (Figs. 3 and 4) Q solvers, in the analog case all postselected $\{V_{D_i}\}$ are averaged together. Finally, the results $\{\langle V_{D_i}\rangle\}$ are digitized to determine the most likely $\boldsymbol{k}$. The choice of digitization threshold for each D_i depends on: a) the readout voltage distributions ρ_0 and ρ_1 for the two basis states, each characterized by a mean μ and a variance σ^2; b) η_A. Ideally ($\eta_A = 0$ and perfect oracle), the distribution of each query output V_{D_i} matches ρ_0 (ρ_1) for $k_i = 0(1)$. When $\eta_A > 0$, the distribution for $k_i = 1$ becomes the mixture $\rho_{k_i=1} = \eta_A\rho_0 + (1-\eta_A)\rho_1$. This mixture has mean $(1-\eta_A)\mu_1 + \eta_A\mu_0$ and variance $(1-\eta_A)\sigma_1^2 + \eta_A\sigma_0^2 - 2\eta_A(1-\eta_A)\mu_0\mu_1$. Instead, $\rho_{k_i=0} = \rho_0$ independently of η_A. We approximate the expected distribution of the mean $\langle V_{D_i} \rangle$ with a Gaussian having average and variance obtained from $\rho_{k_i=0}(\rho_{k_i=1})$ for $k_i = 0(1)$. Finally, we choose the digitization threshold for V_{D_i} which maximally discriminates these two Gaussian distributions. We note that the number of queries scales the variance of both distributions equally and therefore does not affect the optimum threshold. Furthermore, this calibration protocol is independent of the oracle (see Supplementary Information).

Analog quantum solver without postselection

The analysis without ancilla (Q') closely follows the steps outlined in the last paragraph. For the purpose of extracting the optimum digitization thresholds, we consider $\eta_A = 0.5$ in the expressions above. This corresponds to an equal mixture of ρ_0 and ρ_1 when $k_i = 1$.

Data deposition and code availability

The full dataset and the Julia[26] code used for this analysis are available at https://doi.org/10.5281/zenodo.268731.

ACKNOWLEDGEMENTS

We thank George A. Keefe and Mary B. Rothwell for device fabrication, T. Ohki for technical assistance, H. Krovi for discussions, and I. Siddiqi for providing the Josephson parametric amplifier. This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office contract no. W911NF-10-1-0324. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

AUTHOR CONTRIBUTIONS

C.A.R. and B.R.J. developed the BBN APS and the data acquisition software, D.R. and A.D.C. carried out the experiment, D.R., M.P.S., and B.R.J. performed the data analysis, M.P.S. implemented the solvers and developed the theoretical models, D.R. and M.P.S. wrote the manuscript with comments from the other authors, A.W.C. and J.A.S. contributed to the initial design of the experiment, B.R.J., J.M.C., and J.M.G. supervised the project.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Yamamoto, T. et al. Quantum process tomography of two-qubit controlled-Z and controlled-NOT gates using superconducting phase qubits. *Phys. Rev. B* **82**, 184515 (2010).
2. Dewes, A. et al. Quantum speeding-up of computation demonstrated in a superconducting two-qubit processor. *Phys. Rev. B* **85**, 140503 (2012).
3. Schuld, M., Sinayskiy, I. & Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys*. **56**, 172–185 (2015).
4. Manzano, D., Pawowski, M. & Brukner, Č. The speed of quantum and classical learning for performing the k-th root of NOT. *New J. Phys*. **11**, 113018 (2009).
5. Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum algorithms for supervised and unsupervised machine learning. *arXiv:quant-ph/1307.0411* (2013).
6. Wiebe, N., Granade, C., Ferrie, C. & Cory, D. G. Hamiltonian learning and certification using quantum resources. *Phys. Rev. Lett*. **112**, 190501 (2014).
7. Angluin, D. & Laird, P. Learning from noisy examples. *Mach. Learn*. **2**, 343–370 (1988).
8. Blum, A., Kalai, A. & Wasserman, H. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM* **50**, 506–519 (2003).
9. Lyubashevsky V. The Parity Problem in the Presence of Noise, Decoding Random Linear Codes, and the Subset Sum Problem. In Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques. Lecture Notes in Computer Science, vol. 3624 (eds Chekuri, C., Jansen, K., Rolim, J. D. P. & Trevisan, L.) (Springer, Berlin, Heidelberg, 2005).
10. Hopper, N. J. & Blum, M. Secure human identification protocols. In *Advances in Cryptology — ASIACRYPT 2001*, vol. 2248. Lecture Notes in Computer Science, (eds Boyd, C.) 52–66 (Springer, Berlin, Heidelberg, 2001).
11. Pietrzak, K. Cryptography from Learning Parity with Noise. In *SOFSEM 2012: Theory and Practice of Computer Science. SOFSEM 2012*. Lecture Notes in Computer Science, vol. 7147. (eds Bieliková, M., Friedrich, G., Gottlob, G., Katzenbeisser, S. & Turán, G.) (Springer, Berlin, Heidelberg, 2012).
12. Cross, A. W., Smith, G. & Smolin, J. A. Quantum learning robust against noise. *Phys. Rev. A* **92**, 012327 (2015).
13. Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
14. Rigetti, C. & Devoret, M. Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies. *Phys. Rev. B* **81**, 134507 (2010).
15. Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).
16. Hatridge, M., Vijay, R., Slichter, D. H., Clarke, J. & Siddiqi, I. Dispersive magnetometry with a quantum limited SQUID parametric amplifier. *Phys. Rev. B* **83**, 134501 (2011).
17. Cleve, R., Ekert, A., Macchiavello, C. & Mosca, M. Quantum algorithms revisited. *Proc. R. Soc. Lond. A* **454**, 339–354 (1998).
18. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett*. **106**, 110502 (2011).
19. Kimmel, S., Low, G. H. & Yoder, T. J. Robust calibration of a universal single-qubit gate set via robust phase estimation. *Phys. Rev. A* **92**, 062315 (2015).
20. Motzoi, F., Gambetta, J. M., Rebentrost, P. & Wilhelm, F. K. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett*. **103**, 110501 (2009).
21. Lucero, E. et al. Reduced phase error through optimized control of a superconducting qubit. *Phys. Rev. A* **82**, 042339 (2010).
22. Chow, J. M. et al. Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits. *Phys. Rev. Lett*. **109**, 060501 (2012).
23. Jerger, M. et al. Frequency division multiplexing readout and simultaneous manipulation of an array of flux qubits. *Appl. Phys. Lett*. **101**, 042604 (2012).
24. Chow, J. M. et al. Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nature Comm* **5**, 4015 (2014).
25. Gambetta, J., Braff, W. A., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement. *Phys. Rev. A* **76**, 012325 (2007).
26. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *arXiv:cs/1411.1607* (2014).
27. Córcoles, A. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Comm* **6**, 6979 (2015).

Supplementary Information accompanies the paper on the *npj Quantum Information* website (doi:10.1038/s41534-017-0017-3).