

Error mitigation extends the computational reach of a noisy quantum processor

Abhinav Kandala^{1*}, Kristan Temme¹, Antonio D. Córcoles¹, Antonio Mezzacapo¹, Jerry M. Chow¹ & Jay M. Gambetta¹

Quantum computation, a paradigm of computing that is completely different from classical methods, benefits from theoretically proved speed-ups for certain problems and can be used to study the properties of quantum systems¹. Yet, because of the inherently fragile nature of the physical computing elements (qubits), achieving quantum advantages over classical computation requires extremely low error rates for qubit operations, as well as substantial physical qubits, to realize fault tolerance via quantum error correction^{2,3}. However, recent theoretical work^{4,5} has shown that the accuracy of computation (based on expectation values of quantum observables) can be enhanced through an extrapolation of results from a collection of experiments of varying noise. Here we demonstrate this error mitigation protocol on a superconducting quantum processor, enhancing its computational capability, with no additional hardware modifications. We apply the protocol to mitigate errors in canonical single- and two-qubit experiments and then extend its application to the variational optimization^{6–8} of Hamiltonians for quantum chemistry and magnetism⁹. We effectively demonstrate that the suppression of incoherent errors helps to achieve an otherwise inaccessible level of accuracy in the variational solutions using our noisy processor. These results demonstrate that error mitigation techniques will enable substantial improvements in the capabilities of near-term quantum computing hardware.

Quantum computation can be extended indefinitely if decoherence and inaccuracies in the implementation of gates can be brought below an error correction threshold^{2,3}. However, the resource requirements for a fully fault-tolerant architecture lie beyond the scope of near-term quantum hardware¹⁰. In the absence of quantum error correction, the dominant sources of noise in current hardware are unitary gate errors and decoherence, both of which set a limit on the size of the computation that can be carried out. In this context, hybrid quantum algorithms^{7,8,11} with short-depth quantum circuits have been designed to perform computations within the available coherence window, while also demonstrating some robustness to coherent unitary errors^{9,12}. However, even when computations are restricted to short-depth circuits, the effect of decoherence already becomes evident in small experiments⁹. The recently proposed zero-noise extrapolation method^{4,5,13} presents a route to mitigating incoherent errors and greatly improving the accuracy of the computation. It is important to note that, unlike quantum error-correction, this technique does not allow for an indefinite extension of the computation time, and provides corrections only to expectation values, without correcting for the full statistical behaviour. However, since it does not require any additional quantum resources, the technique is extremely well suited for practical implementations with near-term hardware.

We shall first briefly describe the proposal of ref. ⁴ and discuss important considerations for our implementation with superconducting qubits. Any quantum circuit can be expressed in terms of evolution under a time-dependent drive Hamiltonian $K(t) = \sum_{\alpha} J_{\alpha}(t) P_{\alpha}$ for a time T , where each P_{α} represents a N -qubit Pauli operator, and J_{α} is the strength of the associated interaction. The expectation value of an observable of interest $E_K(\lambda)$ for a state prepared by the drive K in the

presence of noise can be expressed as a power series around its zero-noise value E^* as

$$E_K(\lambda) = E^* + \sum_{k=1}^n a_k \lambda^k + \mathcal{O}(\lambda^{n+1}) \quad (1)$$

Here $\lambda \ll 1$ is a small noise parameter, and the coefficients in the expansion a_k are dependent on specific details of the noise model. The primary objective of this Letter is to experimentally obtain improved estimates to the noise-less expectation value E^* despite using noisy quantum hardware. A powerful numerical technique for suppressing the higher-order noise terms in equation (1) is Richardson's deferred approach to the limit¹⁴. If n additional estimates to the expectation value $\hat{E}_K(c_i \lambda)$ can be obtained when precisely amplifying the noise rate by factors c_i for $i = 1, 2, \dots, n$, an improved approximation to E^* with a reduced error of order $\mathcal{O}(\lambda^{n+1})$ can be constructed as

$$\hat{E}_K^n(\lambda) = \sum_{i=0}^n \gamma_i \hat{E}_K(c_i \lambda) \quad (2)$$

For a chosen set of c_i , the coefficients γ_i are solutions to $\sum_{i=0}^n \gamma_i = 1$ and $\sum_{i=0}^n \gamma_i c_i^k = 0$ for $k = 1 \dots n$. Now, precisely amplifying the strength of the noise is a very challenging experimental task. It can be approximated by the insertion of noisy gates in the quantum circuit, but relies heavily on assumptions of the noise mechanisms at play¹⁵. However, if the noise is 'time invariant', it can be shown⁴ that measurements of the expectation value after evolution under a scaled drive $K^I(t) = \sum_{\alpha} J_{\alpha}^I(t) P_{\alpha}$ for a time $c_i T$ (within the coherence window) is equivalent to a measurement under an amplified noise strength $c_i \lambda$. The scaled strength of the interactions in the drive is given as $J_{\alpha}^I(t) = J_{\alpha}(t/c_i)/c_i$, requiring a good understanding and control of the gates used in the circuit. We use the term 'time invariant' for noise that can be described by a dynamical equation that is invariant under time rescaling—such as constant-amplitude damping and dephasing noise or other noise that arises from a constant coupling to environment. Beyond this requirement, we emphasize that the method is completely agnostic to the details of the noise model, making it extremely attractive for implementations on near-term noisy hardware. However, for superconducting qubits, which often show fluctuations¹⁶ in relaxation T_1 and coherence times T_2 , the requirement of time invariant noise implies that the measurements under the scaled dynamics need to be made within the typical timescales of these fluctuations (see Methods).

The experiments described in this Letter are performed on a 5-qubit superconducting processor depicted in Fig. 1a. The device comprises fixed-frequency Josephson-junction-based transmon¹⁷ qubits, with individual superconducting co-planar waveguide resonators for qubit control and readout, and another pair of co-planar waveguide resonators providing the qubit connectivity. This fixed-frequency architecture is favourable for obtaining long coherence times, and the qubit control and readout is solely by microwave pulses.

In our device architecture, arbitrary quantum circuits are implemented using combinations of single-qubit gates, and two-qubit gates between nearest-neighbour qubits. We first discuss the implementation

¹IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. *e-mail: akandala@us.ibm.com

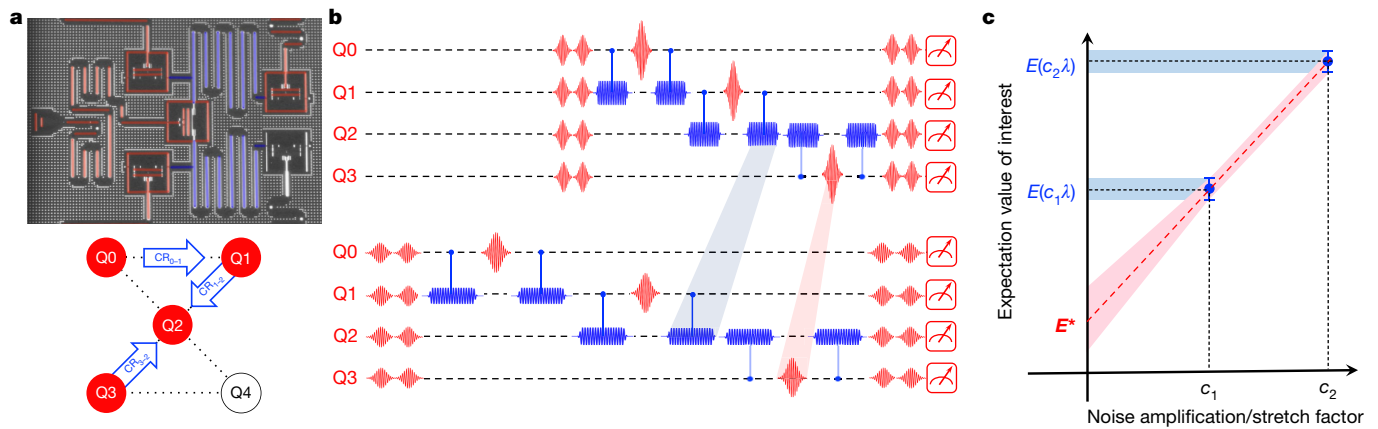


Fig. 1 | Device and experimental protocol. **a**, False-coloured optical micrograph (top) of the superconducting quantum processor and schematic (bottom) of the qubits (Q0, Q1, Q2, Q3) and gates (CR₀₋₁, CR₁₋₂, CR₂₋₃, CR₃₋₄) used in the experiment. The device is composed of five transmon qubits, with the coupling provided by two superconducting co-planar waveguide resonators, in blue. **b**, A measurement of the expectation value

of our error mitigation scheme for single-qubit gates in Fig. 2a, b. Single-qubit control is achieved using 4σ Gaussian pulses with a scaled derivative in quadrature to reduce leakage to higher transmon energy levels, and software-implemented Z-gates¹⁸. Each microwave pulse is followed by a buffer time to ensure separation from subsequent pulses. Arbitrary single-qubit rotations are constructed using an interleaved sequence of calibrated $X_{\pi/2}$ pulses and Z rotations $U(\theta) = Z_{\theta} X_{\pi/2} Z_{\theta} X_{\pi/2} Z_{\theta}$ where θ represents the Euler angles. For the error mitigation experiments, the lengths of the 4σ Gaussian pulses as well as the buffer times are stretched by the desired stretch factors c_i and calibrated. We first consider sequences of identity-equivalent random single-qubit Clifford operations that ideally return the qubit to the $|0\rangle$ state. We study the decay of the expectation value of the ground-state projector $\langle \Pi_0 \rangle$ with increasing length of the Clifford sequence for the different stretch factors c_i , in Fig. 2a. The data are obtained using 10^5 samples. Using measurements for each stretch factor c_i and Clifford sequence, we demonstrate suppression of higher-order errors in the estimates of $\langle \Pi_0 \rangle$, using up to a third-order Richardson extrapolation to the zero-noise limit. Obtaining tight error bounds on the extrapolated estimates that encompasses all the sources of error remains a challenge. One obvious approach to capture the effect of finite sampling would be to perform a large number of independent experimental runs—however, this is an extremely time-consuming task. Instead, as detailed in the Methods, we employ a bootstrapping technique to simulate the error associated with finite sampling, given an experimental dataset. The distributions of numerical outcomes in Fig. 2a show that higher-order extrapolations are increasingly sensitive to the variance of the unmitigated measurements, as can be seen from the conservative error bounds (see Methods) for the zero-noise extrapolation technique derived in ref. ⁴. This produces a run-time cost for the technique associated not only with the additional stretch experiments, but also with the need for increased sampling. In Fig. 2b, the error mitigation technique may also be visualized by a Bloch sphere picture of a trajectory that begins in the $|0\rangle$ state and, in the absence of noise, brings the qubit to its $|1\rangle$ state, along the surface of the sphere. Every j th point along the trajectory is implemented by the unitary operation U_j defined as

$$U_{j+1} = Z_{4\theta(j+1)} X_{\theta(j+1)} X_{-\theta(j)} Z_{-4\theta(j)} U_j \quad (3)$$

Here, $\theta(j) = j\pi/30$, the X rotations are implemented as $X_{\theta(j)} = Y_{\pi/2} Z_{\theta(j)} Y_{-\pi/2}$ for $j = 0, 1, \dots, 29$ and U_0 is the identity gate. We construct the trajectories using measurements of the projectors $\langle X \rangle$, $\langle Y \rangle$ and $\langle Z \rangle$ for each c_i at every point j . Although the effect of relaxation and

after rescaled state preparation is equivalent to a measurement under an amplified noise strength, if the noise is time invariant. **c**, An illustration of the error mitigation method, shown here for a first-order Richardson extrapolation to the zero-noise limit, highlights that the variance of the mitigated estimate E^* is crucially dependent on the variance of the unmitigated measurements and the stretch factors c_i .

dephasing is apparent in the experimental trajectories, the error-mitigated trajectory approaches the final $|1\rangle$ state.

The implementation of this method with two-qubit gates is more challenging for superconducting qubit architectures, since control is often more complex and the fidelities of two-qubit gates are typically an order of magnitude worse than those of single-qubit gates. Furthermore, the stretching of the gates requires an understanding of the drive Hamiltonian. For our fixed-frequency qubits, we use the all-microwave cross-resonance (CR) gate^{19,20}, which is implemented by driving a control qubit Q_c with a Gaussian-square-modulated microwave pulse that is resonant with the frequency of the target qubit Q_t . For the error mitigation experiments discussed here, we operate our CR gates in a low-power regime where the strengths of the interaction terms in the drive scale linearly with drive amplitude. This is crucial for our implementation, since operating the CR gates in a nonlinear regime can result in interactions strengths that do not scale appropriately, leading to unphysical values for the mitigated estimates (see Methods). As with single-qubit gates, the pulse lengths, rise-fall times and buffer times of the CR drives are all stretched by the chosen stretch factors and calibrated to a $ZX_{\pi/2}$ gate. As a model experiment that spans the two-qubit Hilbert space, we consider identity-equivalent sequences of random two-qubit Clifford operators that are applied on a maximally entangled Bell state. Figure 2c depicts the decay of ZZ parity for these sequences, and its mitigation using a first-order Richardson extrapolation to the zero-noise limit. The deviations of the mitigated estimates from the ideal values for the data in Fig. 2 can be accounted for by readout assignment errors. It is important to note that the eventual decay of the mitigated curves in Fig. 2 shows that the method cannot be applied indefinitely and is ultimately limited by the quantum coherence of the device, as predicted by the error analysis of ref. ⁴. Also, gate sequences such as those depicted in Fig. 2a, c are often used in standard randomized benchmarking protocols²¹, highlighting the potential applicability of the zero-noise extrapolation technique to improved gate characterization schemes.

The ability to apply the zero-noise extrapolation techniques to random single-qubit and two-qubit circuits enables us now to address multi-qubit variational eigensolvers^{7,9,12,22,23}. Here, variational approximations to the ground state of Hamiltonians of interest are parametrized by experimental controls and prepared on the quantum processor, with the parameters updated iteratively in conjunction with a classical optimization routine. We first address an interacting spin problem with highly entangled ground states considered in

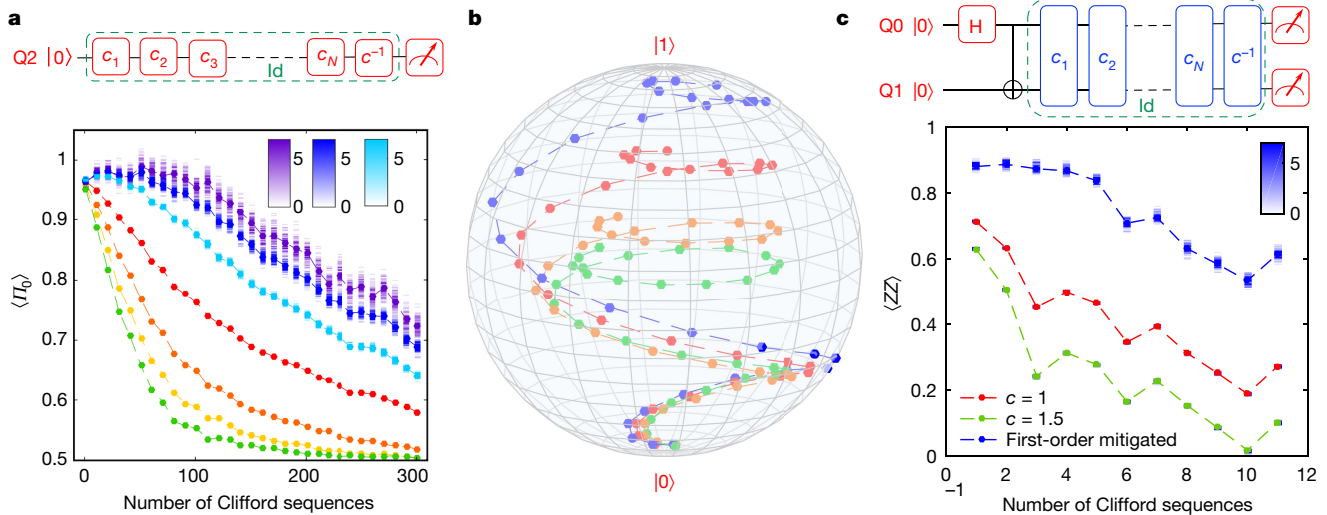


Fig. 2 | Error mitigation of random single-qubit and two-qubit circuits.

a, Expectation value of the ground-state projector for identity-equivalent ('Id') single-qubit Clifford sequences for stretch factors $c = 1$ (red), $c = 2$ (orange), $c = 3$ (yellow), $c = 4$ (green) and the corresponding Richardson extrapolations to first (light blue), second (dark blue) and third order (violet). **b**, Experimental implementation of trajectories described by equation (3), represented on a Bloch sphere for stretch factors $c = 1$ (red), $c = 2$ (orange), $c = 3$ (green) and the corresponding first-order Richardson extrapolation (blue). The ideal theoretical trajectory is one that takes the

qubit from its ground state to its excited state along the surface of the Bloch sphere. **c**, Expectation value of the ZZ parity for identity-equivalent two-qubit Clifford sequences applied on a Bell state for stretch factors $c = 1$ (red), $c = 1.5$ (green) and the corresponding first-order Richardson extrapolations (dark blue). The H gate in the sequence represents a Hadamard gate. The colour density plots of **a** and **c** represent histograms of outcomes of 100 numerical experiments obtained by bootstrapping of each experimental data point, with the histogram frequencies quantified by the colour bars.

ref.⁹, specifically, an anti-ferromagnetic four-qubit Heisenberg model on a square lattice, in an external magnetic field:

$$H = J \sum_{\langle ij \rangle} (X_i X_j + Y_i Y_j + Z_i Z_j) + B \sum_i Z_i \quad (4)$$

Here, J is the strength of the spin–spin interaction for nearest-neighbour pairs $\langle ij \rangle$, and B is the magnetic field in the Z direction. We employ a hardware-efficient variational ansatz⁹, constructed as an interleaved sequence of arbitrary single-qubit rotations and entanglers composed of gates that are natural to the hardware architecture. In the experiment, each entangler is composed of a sequence of echo CR pulses (see Methods) and the Euler rotation angles serve as the variational parameters. For the classical optimization routine, we use the simultaneous perturbation stochastic approximation (SPSA) algorithm²⁴ that approximates the gradient at each iteration using only two measurements of the cost function, irrespective of the dimensionality of the parameter space. At every iteration, the energies obtained from trial states prepared at stretch factors $c = 1$ and 1.5 are used to obtain a mitigated energy estimate that is supplied to the SPSA routine. This is shown for the trial circuit of depth $d = 5$, using 10^4 samples along the optimization in Fig. 3a. For the final set of controls, we increase the sampling to 10^5 , and obtain measurements at additional stretch factors $c = 1, 1.1, 1.25$ and 1.5 to reduce the variance on the error-mitigated energy estimate.

In ref.⁹, it was numerically shown that the accuracy of a hardware-efficient variational quantum eigensolver is affected by the competing effects of insufficient circuit depth and incoherent noise. Having discussed the application of zero-noise extrapolation techniques to a variational quantum eigensolver, we now demonstrate how the suppression of incoherent errors helps to discern improvements with increasing circuit depth. First, we consider the energy error $\epsilon_1 = |E_{\text{exp}} - E_{\text{exact}}|$ and compare it with the energy error of the lowest-energy separable state $\epsilon_{1,d=0}$. The depth dependence of ϵ_1 is depicted in Fig. 3b for the specific Heisenberg Hamiltonian with $J/B = 1$ after optimizations of the kind depicted in Fig. 3a are run for each d . For the $c = 1$ state preparations, improvements in ϵ_1 with respect to $\epsilon_{1,d=0}$ are limited to $d = 2$, owing to the increased decoherence associated with longer trial circuits.

However, the ability to suppress the effect of incoherent errors enables us to discern improvements in ϵ_1 for up to $d = 5$ trial states. Additionally, we also quantify the errors in the expectation values of the individual Pauli operators $\langle P_i \rangle$ in the Hamiltonian $H = \sum_i \alpha_i P_i$ by the quantity $\epsilon_2 = \sum_i |\alpha_i|^2 (\langle P_{i,\text{exp}} \rangle - \langle P_{i,\text{exact}} \rangle)^2$. The improvements in ϵ_2 with increasing circuit depth shown in Fig. 3c demonstrate that the error mitigation also improves convergence towards the ground-state wavefunction, which may not be captured merely by the energy. We map the energy for a range of J/B values in Fig. 3d, algorithmically increasing the circuit depth for every Hamiltonian until no further improvement in energy is obtained. In a similar spirit, one can anticipate the use of such mitigation techniques also to benefit Trotter-based quantum simulations on noisy hardware, enabling improvements in accuracy with decomposition of the time evolution into finer steps.

In contrast to alternative error mitigation schemes²⁵ that rely on specific features of the target Hamiltonian, this zero-noise extrapolation technique is independent of the simulation problem considered. We now demonstrate its general applicability by considering problems in quantum chemistry^{7,9,12,22,23,26} (see Fig. 4). We map the interacting fermion problem for H_2 and LiH on two and four qubits respectively, taking advantage of fermionic spin-parity symmetries²⁷ and the freezing of core-shell orbitals, as detailed in ref.⁹. The accuracy of the variational solutions to these problems, obtained on a similar device with comparable coherence times⁹, was severely limited by incoherent errors and insufficient circuit depth for trial state preparations. However, as in the Heisenberg model discussed above, Fig. 4 demonstrates far superior accuracies without considerable improvements in the coherence properties of the hardware, and the ability to benefit from longer circuit depths for trial state preparation.

A crucial aspect affecting the accuracy and variance of the mitigated estimates is that of sampling errors, which are amplified by the extrapolation. In this context, the integration of fast initialization schemes²⁸ would enable much faster sampling rates. The benefits of this are manyfold: a reduced variance on the mitigated estimates, enabling more accurate classical optimization, reduced experimental run times, the ability to apply higher-order Richardson extrapolations, and a reduced effect of coherence time fluctuations in the hardware.

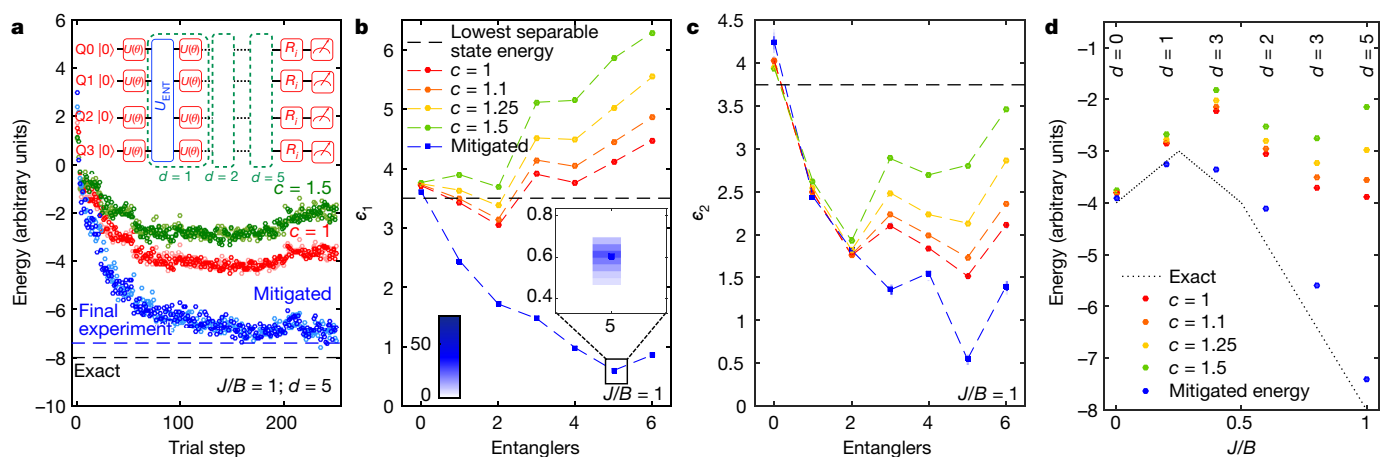


Fig. 3 | Error-mitigated variational optimization of interacting spin Hamiltonians. A hardware-efficient variational eigensolver applied to an anti-ferromagnetic four qubit Heisenberg model in a magnetic field. All colour density plots represent histograms of outcomes of 100 numerical experiments obtained by bootstrapping of each experimental dataset, with the frequencies quantified by the colour bar in **b**. **a**, Energy minimization for $J/B = 1$ using a depth $d = 5$ hardware-efficient trial circuit (depicted in the inset) involves the simultaneous optimization of 68 variational parameters. The circuit is constructed as an interleaved sequence of arbitrary single qubit rotations $U(\theta)$ and entanglers U_{ENT} that entangle all the qubits in the circuit. A set of post-rotations R_i are used to sample the expectation values of the Pauli operators in the target Hamiltonian. At each iteration, the energies of trial states prepared using stretch factors

$c = 1$ and $c = 1.5$ are measured using 10^4 samples, and the mitigated energies obtained by first-order Richardson extrapolation are fed to the classical optimization routine. The control parameters from the final 25 iterations are averaged to obtain the controls for the final state measurement. The final energy is then obtained using a linear extrapolation to the energies associated with the final state preparations with stretch factors $c = 1$, $c = 1.1$, $c = 1.25$ and $c = 1.5$, obtained using 10^5 samples. **b**, **c**, Final energy error ϵ_1 (**b**) and a measure of the error on the individual Pauli terms in the Hamiltonian ϵ_2 (**c**) as a function of the number of entanglers in the state preparation circuit for the final stretch factors and the mitigated values. **d**, Experimental results from the extrapolation of energies obtained from the final stretch factors, compared to the exact energy, for a range of J/B values.

The work presented here highlights the important considerations for hardware and algorithmic implementations of the zero-noise extrapolation technique, and demonstrates tremendous improvements in the accuracy of variational eigensolvers implemented by a noisy superconducting quantum processor. Further improvements in coherence times will make these methods even more effective, enabling the applicability to even longer quantum circuits, as well as the ability to reduce the variance of the mitigated estimates with longer stretch factors. Gate calibrations particularly tailored for error mitigation could enable a more accurate rescaling of the dynamics. Alternatively, the insertion

of twirling gates can be used to randomize systematic errors²⁹ in the rescaling, albeit at the cost of additional measurements. We note that other general purpose schemes such as probabilistic error cancellation^{4,13} exist that do not use extrapolation to mitigate the effect of noise. However, the experimental challenges for their practical implementation differ from those discussed here and are left for subsequent research. Finally, while the experiments discussed here address problems in quantum simulation, the techniques presented have great applicability to improved gate characterization, quantum optimization and quantum machine learning³⁰ in near-term hardware.

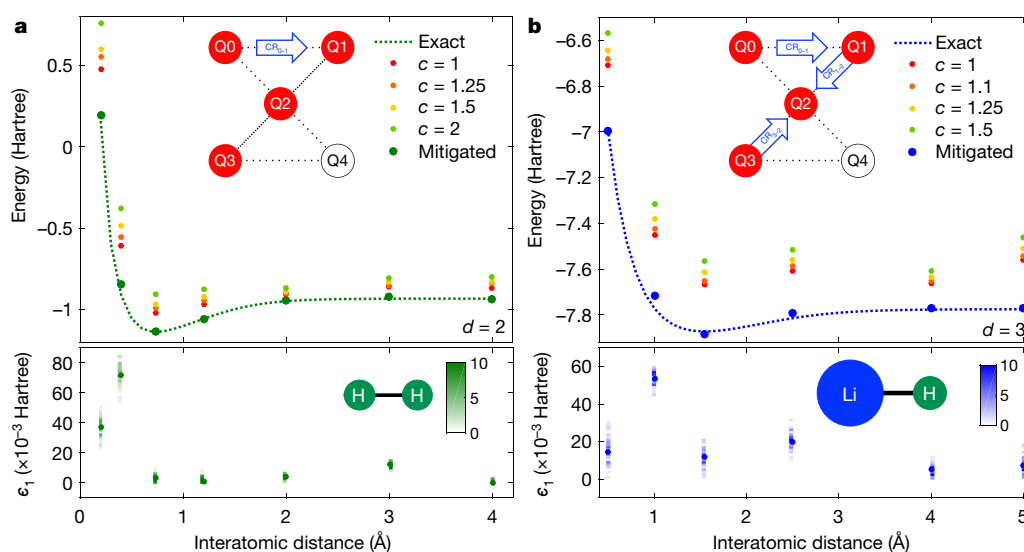


Fig. 4 | Error-mitigated variational optimization of molecular Hamiltonians. **a**, **b**, A hardware-efficient variational eigensolver applied to the electronic structure problem of H_2 (**a**) and LiH (**b**), using trial state preparation circuit depths $d = 2$ and $d = 3$, respectively. The experimental results from the extrapolation of energies obtained from the final stretch factors, compared to the exact energy, for a range of interatomic distances. The bottom panels represent the energy errors ϵ_1 for the Hamiltonians

considered with histograms of outcomes of 50 numerical experiments obtained by bootstrapping of each experimental dataset and frequencies quantified by the respective colour bars. The insets in the top panel depict the qubits used and the gates that compose the entangler U_{ENT} . The insets in the bottom panel represent schematics of the molecular geometry, not drawn to scale. 1 Hartree = 4.359×10^{-18} J.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-1040-7>.

Received: 9 May 2018; Accepted: 4 January 2019;

Published online 27 March 2019.

1. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
2. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, 2493–2496 (1995).
3. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
4. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
5. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
6. Yung, M.-H. et al. From transistor to trapped-ion computers for quantum chemistry. *Sci. Rep.* **4**, 3589 (2014).
7. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
8. Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. Preprint at <https://arxiv.org/abs/1411.4028> (2014).
9. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
10. Preskill, J. Quantum computing in the NISQ era and beyond. Preprint at <https://arxiv.org/abs/1801.00862> (2018).
11. McClean, J., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 023023 (2016).
12. O'Malley, P. J. J. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
13. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
14. Richardson, L. F. IX. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Phil. Trans. R. Soc. Lond. A* **210**, 307–357 (1911).
15. Dumitrescu, E. et al. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.* **120**, 210501 (2018).
16. Müller, C., Lisenfeld, J., Shnirman, A. & Poletto, S. Interacting two-level defects as sources of fluctuating high-frequency noise in superconducting circuits. *Phys. Rev. B* **92**, 035442 (2015).
17. Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
18. McKay, D. C., Wood, C. J., Sheldon, S., Chow, J. M. & Gambetta, J. M. Efficient Z gates for quantum computing. *Phys. Rev. A* **96**, 022330 (2017).
19. Paoanau, G. S. Microwave-induced coupling of superconducting qubits. *Phys. Rev. B* **74**, 140504 (2006).
20. Sheldon, S., Magesan, E., Chow, J. M. & Gambetta, J. M. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A* **93**, 060302 (2016).
21. Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).
22. Colless, J. I. et al. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Phys. Rev. X* **8**, 011021 (2018).
23. Hempel, C. et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Phys. Rev. X* **8**, 031022 (2018).
24. Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Contr.* **37**, 332–341 (1992).
25. McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. & de Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A* **95**, 042308 (2017).
26. Lanyon, B. P. et al. Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2**, 106–111 (2010).
27. Bravyi, S., Gambetta, J. M., Mezzacapo, A. & Temme, K. Tapering off qubits to simulate fermionic Hamiltonians. Preprint at <https://arxiv.org/abs/1701.08213> (2017).
28. Egger, D. J. et al. Pulsed reset protocol for fixed-frequency superconducting qubits. *Phys. Rev. X* **8**, 031022 (2018).
29. Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).
30. Havlicek, V. et al. Supervised learning with quantum enhanced feature spaces. Preprint at <https://arxiv.org/abs/1804.11326> (2018).

Acknowledgements We thank J. Rozen, M. Takita, and O. Jinka for experimental contributions, B. Abdo for design and characterization of the Josephson parametric converters, and M. Brink for device fabrication. We acknowledge discussions with S. Bravyi, E. Magesan, S. Sheldon, and M. Takita. We acknowledge support from the IBM Research Frontiers Institute. The research is based on work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via the Army Research Office contract W911NF-10-1-0324.

Author contributions A.K., K.T. and J.M.G. developed the experimental protocol, planned the experiments and analysed the experimental data. A.K. performed the experiments. A.D.C. contributed to the experimental set-up. A.M. contributed to the data analysis. A.K., K.T., J.M.C., and J.M.G. wrote the manuscript, with contributions from all authors.

Competing interests The authors declare that elements of this work are included in a patent filed by the International Business Machines Corporation with the US Patent and Trademark office.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-019-1040-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1040-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to A.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Device and gates. The quantum processor is composed of five fixed-frequency transmon¹⁷ qubits, and superconducting coplanar waveguide resonators that are employed for qubit–qubit coupling, as well as qubit control and readout, all fabricated on a Si wafer. Each transmon is a single Al–Al₂O_x–Al Josephson junction, capacitively shunted by Nb capacitor pads. Additional details of the device fabrication may be found in refs.^{31,32}. The qubit frequencies lie in the $\omega_0/2\pi \approx 5$ –5.3 GHz range, and are read out by dispersive measurements through their individual readout resonators at a frequencies close to $\omega_r/2\pi \approx 6.5$ GHz. The qubit anharmonicities are about 0.330 GHz. The readout signal is amplified by a Josephson parametric converter^{33,34} at the mixing chamber stage of a dilution refrigerator followed by a high-electron-mobility transistor at the 4-K stage, to achieve typical readout errors $\epsilon_r < 0.05$ for integration times of 2 μ s. Typical relaxation (T_1) and dephasing (T_2 echo) times of the qubits in the device are in the range 40–70 μ s.

All single-qubit operations are implemented using software Z gates and/or 4 σ Gaussian-modulated microwave pulses employing a derivative removal via adiabatic gate (DRAG) protocol³⁵. The shortest pulse time (for $c = 1$) used in the experiment is 83.3 ns and a buffer time of 6.7 ns is used for ample separation from subsequent pulses. The pulse times and the buffer times are rescaled by the stretch factor c , with calibrated pulse amplitudes and DRAG parameters. Extended Data Fig. 1a depicts the gate fidelities of the single qubit gates obtained by simultaneous randomized benchmarking, for a range of different stretch factors. The two-qubit gates are implemented using Gaussian-square CR pulses^{36,37}. $ZX_{\pi/2}$ operations are constructed using an echo CR sequence, where a X_{π} gate on the control qubit is sandwiched between two CR pulses of identical pulse time τ and calibrated amplitude, but opposite sign (see Methods section on stretching of two-qubit gates). The shortest $ZX_{\pi/2}$ gate employed in the experiment employs 2 CR pulses each of width $\tau = 500$ ns, a 3 σ Gaussian rise fall profile (included in τ) with $\sigma = 10$ ns, and a buffer time of 6.7 ns. In combination with the X_{π} gate, this brings the total $ZX_{\pi/2}$ gate time for $c = 1$ to 1103.4 ns. We note that this is much slower than typical operations of the CR gate (https://github.com/Qiskit/qiskit-backend-information/blob/master/backends/yorktown/V1/version_log.md#110); this is detailed in the Methods section ‘Stretching of two-qubit gates’. As with the single-qubit gates, for every stretch factor, the pulse times, σ , and the buffer times are rescaled, and the CR pulse amplitudes calibrated to a $ZX_{\pi/2}$ gate. Extended Data Fig. 1b depicts the gate fidelities of the two-qubit gates for a range of stretch factors.

For a hardware-efficient ansatz, it was shown in ref.⁹ that convergence to the ground state of small molecular and spin systems can be obtained for a range of entangling gate phases around points of maximum concurrence. For the variational trial states implemented in this work, the entangler is constructed as a series of pairwise $ZX_{\pi/4}$ gates, each employing an echo CR pulse sequence. These entanglers are so chosen as to deliver sufficient entanglement, while reducing the effect of decoherence during state preparation.

Time invariant noise and coherence fluctuations in superconducting qubits. The dynamical time evolution of a physical system with initial state $\rho(t=0)$ in the presence of noise is described by

$$\frac{\partial}{\partial t}\rho = -i[K(t), \rho] + \lambda\mathcal{L}(\rho) \quad (5)$$

where $K(t) = \sum_{\alpha} J_{\alpha}(t)P_{\alpha}$ generates the coherent circuit that we want to implement with interactions P_{α} and their associated time-dependent drive strengths $J_{\alpha}(t)$, \mathcal{L} represents the noise that we want to mitigate, and λ is a small noise parameter. The expectation value of a quantum observable of interest A is then obtained from the evolved state $\rho_{\lambda}(t=T)$ with noise factor λ as $E_{\lambda} = \text{tr}[A\rho_{\lambda}(T)]$. Now, if \mathcal{L} is not explicitly time dependent, we can obtain the expectation values with an amplified noise factor $c\lambda$ by stretching the evolution to a time cT with rescaled drive amplitudes $J_{\alpha}^I(t) = J_{\alpha}(t/c)/c$. This can be seen by the realization that a simple rescaling of $t \rightarrow t/c$ in the stretched evolution recovers the original differential equation (5) with $\lambda \rightarrow c\lambda$. The experimental advantage of performing noise amplification in this way is that the exact form of \mathcal{L} and the magnitude of λ may be completely unknown to the experimentalist, and the noise itself can be quite general.

A concern with superconducting qubits is the observed fluctuations in coherence and relaxation times, which is strictly speaking, in violation of the assumption of ‘time invariant’ noise. Extended Data Fig. 2a depicts fluctuations in the T_1 and T_2 times over a period of approximately 2 h that are clearly larger than the error bars on the individual data points. This apparent violation can however be remedied by an appropriate ordering sequence, such that circuits implemented with different stretch factors, for a desired expectation value, are performed shortly after one another. Over the entire duration of averaging the measurements, the average decay time for the different stretched experiments is then the same, as seen in Supplementary Fig. 2b, c for typical T_1 and echo T_2 sequences for $c_i = 1$, 2 obtained with 10^5 samples. However, when the different stretched experiments

are performed separately, with larger time intervals between them, the decays may differ for different c_i , potentially leading to Richardson extrapolated expectation values that are out of bounds. The experiments discussed here employed a sampling rate of 2 kHz, set by the time required for the qubits to naturally relax to the ground state ($5T_1$ – $10T_1$) for initialization. Integration with fast initialization schemes will improve the accuracy of the mitigated estimates even further by allowing a larger number of samples in a shorter time, while reducing the susceptibility to coherence fluctuations.

Stretching of two-qubit gates. We employ the CR interaction for our two-qubit entangling gate^{19,36,38}. This is particularly well suited for our fixed-frequency, all microwave control hardware architecture, and is implemented by driving a control qubit with a microwave tone that is resonant with a nearest-neighbour target qubit. Advances in the understanding of the CR drive Hamiltonian have led to controlled-NOT (CNOT) gate fidelities exceeding 99%, with gate times²⁰ less than 200 ns. An effective Hamiltonian model of the CR drive³⁹ that accounts for higher energy levels of the transmon reveals the following interactions: IX , IY , IZ , ZX , ZY , ZZ and a Stark shift ZI arising due to the non-resonant tone. The ZX and IX terms are predicted to be the dominant interactions, and ZZ and IZ are small and independent of the strength of the CR drive. This is also seen experimentally, in Extended Data Fig. 3, which depicts the strengths of the different interactions for a range of drive amplitudes, obtained by tomography of the CR drive Hamiltonian for the Q3 (control)–Q2 (target) pair. Isolating the ZX interaction enables the construction of a CNOT gate, merely requiring additional single-qubit gates. This can be achieved by using a standard echo CR sequence that applies a X_{π} on the control qubit between CR pulses of opposite sign, to refocus the IX , ZZ and ZI terms. In the absence of classical cross-talk, the IY is negligible, or else a cancellation may be employed as in ref.²⁰. Finally, the phase of the CR drive may be set such that the sole conditional interaction is ZX .

As seen in Extended Data Fig. 3, the amplitude dependence of the interactions in the CR drive can depict small nonlinearities that are important to consider for the implementation of our error mitigation scheme. These nonlinearities have been detailed in the recent work of ref.³⁹. For instance, the amplitude dependence of the ZX interaction, obtained by a perturbative model to third order, takes the following form

$$J_{ZX} = -\Omega \left(\frac{J\delta_1}{\Delta(\delta_1 + \Delta)} \right) + \Omega^3 \left(\frac{J\delta_1^2(3\delta_1^3 + 11\delta_1^2\Delta + 15\delta_1\Delta^2 + 9\Delta^3)}{4\Delta^3(\delta_1 + \Delta)^3(\delta_1 + 2\Delta)(3\delta_1 + 2\Delta)} \right) \quad (6)$$

Here, Δ is the frequency difference between the qubits, δ_1 represents their anharmonicity, J is the qubit–qubit coupling and Ω is the amplitude of the CR drive. If the gates are not tuned-up appropriately this nonlinearity can lead the stretched evolution to be out of phase. To model the effect of this nonlinearity on our error mitigation scheme, we use a simplistic form of the CR drive $J_{ZX} = -0.0159J\Omega + 1.0541 \times 10^{-6}J\Omega^3$, obtained from equation (1) using $\delta_1 = 320$ MHz and $\Delta = 50$ MHz and add fixed strength $\lambda = 2 \times 10^{-3}$ amplitude damping and de-phasing noise for each qubit. The simplistic CR model system evolves as

$$\begin{aligned} \partial_t \rho = & -iJ_{ZX}[ZX, \rho] \\ & + \lambda \sum_{i=1}^2 \left(\sigma_i^- \rho \sigma_i^+ - \frac{1}{2} \{ \sigma_i^+ \sigma_i^-, \rho \} + Z_i \rho Z_i - \rho \right) \end{aligned} \quad (7)$$

for a total time $T = 100J^{-1}$ with $\sigma^{\pm} = 2^{-1/2}(X \pm iY)$. In Extended Data Fig. 4 we compare the decay of oscillations in the $\langle IZ \rangle$ expectation value for three different drive amplitudes of this model. These were chosen to implement a $ZX_{\pi/2}$ gate with three different gate times $T_{\text{gate}}^a = 2J^{-1}$, $T_{\text{gate}}^b = 3J^{-1}$ and $T_{\text{gate}}^c = 6J^{-1}$ in Extended Data Fig. 4a, b, c respectively. We used the relationship $\Omega = \pi\Delta(\delta_1 + \Delta)(2T_{\text{gate}}\delta_1)^{-1}$ to fix the drive amplitudes. The drive time and amplitudes are scaled for the stretch factors $c_i = 1, 2$ and the mitigated estimates are then obtained using a first-order Richardson extrapolation. Extended Data Fig. 4a shows that the large nonlinearity for fast gates drives the mitigated estimates severely out of bounds. This may be visualized as a consequence of the ‘out-of-phase’ oscillations in $\langle IZ \rangle$ for the two stretch factors. However, the effect of the nonlinearity may be suppressed by using a weaker drive amplitude, as seen in Extended Data Fig. 4c, where the extrapolation is well within bounds, although at the expense of slower gate times. In the absence of the nonlinear term in the model, it can be seen that the mitigated estimates remain in bounds even at high drive powers.

Typical operations of high-fidelity CR gates are often in a fast, nonlinear regime. This highlights the need for gate calibrations specifically tailored for error mitigation, since a simple rescaling of the amplitude would not suffice, as shown above. In the experiment, the CR pulse amplitudes are calibrated to a $ZX_{\pi/2}$, for each stretch factor c_i . To reduce the effect of nonlinearities in the other terms in the

drive Hamiltonian, we employ the echo sequence described above to isolate the ZX interaction, as well as to conservatively operate the gate in a more linear, low-power regime. The two-qubit gate times are therefore considerably slower than normal operation. We emphasize that the accuracies of the computations achieved in this work with these slow gates are otherwise inaccessible with faster gates in the absence of error mitigation. It is also important to note that the error mitigation protocol cancels error terms that are invariant under rescaling⁴. This also includes constant terms in the Hamiltonian evolution such as the static ZZ interaction.

Choice of stretch factors. There are several considerations regarding the choice of stretch factors c_i . While small stretch factors have the advantage that the rescaled gates probably have systematic errors that scale appropriately with respect to the $c = 1$ gates, they can also lead to an amplification of the error bounds on the mitigated estimate. While this may be naively visualized in Fig. 1, this is also seen in the ‘worst case’ error bounds derived in ref. ⁴ for the zero-noise estimator after $n + 1$ measurements

$$|\hat{E}_K^n(\lambda) - E^*| \leq \tilde{I}_n(\delta^* + \mathcal{O}(\lambda^{n+1})) \quad (8)$$

Here δ^* is the largest experimental error (sampling, systematic, etc.), and the stretch factors explicitly affect the error bounds via $\tilde{I}_n = \sum_{i=0}^n |\gamma_i| c_i^{n+1}$. It is important to note that the coefficients γ_i are also dependent on the choice of stretch factors, as noted in the main text—small stretch factors result in large γ_i . The limit on choosing particularly large stretch factors is set by decoherence affecting the ability to accurately calibrate our gates using error amplification sequences similar to those described in ref. ⁴⁰. Furthermore, large stretch factors could also potentially result in systematic errors in the gates that do not scale appropriately.

For the shorter, high-fidelity single qubit gates, we have employed stretch factors as large as $c = 4$, as discussed in Fig. 2a. However, for our longer, two-qubit gates the stretch factors were limited to $c \leq 2$ in this work. Although not done here, once precise estimates of systematic errors are known, it will be worthwhile to optimize the stretch factors to the derived error bound.

We also point out that the choice of stretch factors, through γ_i , also affect the number of samples S required to achieve a desired error bound on the mitigated estimates. This can be seen from equation (8), where in the limit of dominant error arising from finite sampling, $\delta^* = \mathcal{O}(1/\sqrt{S})$.

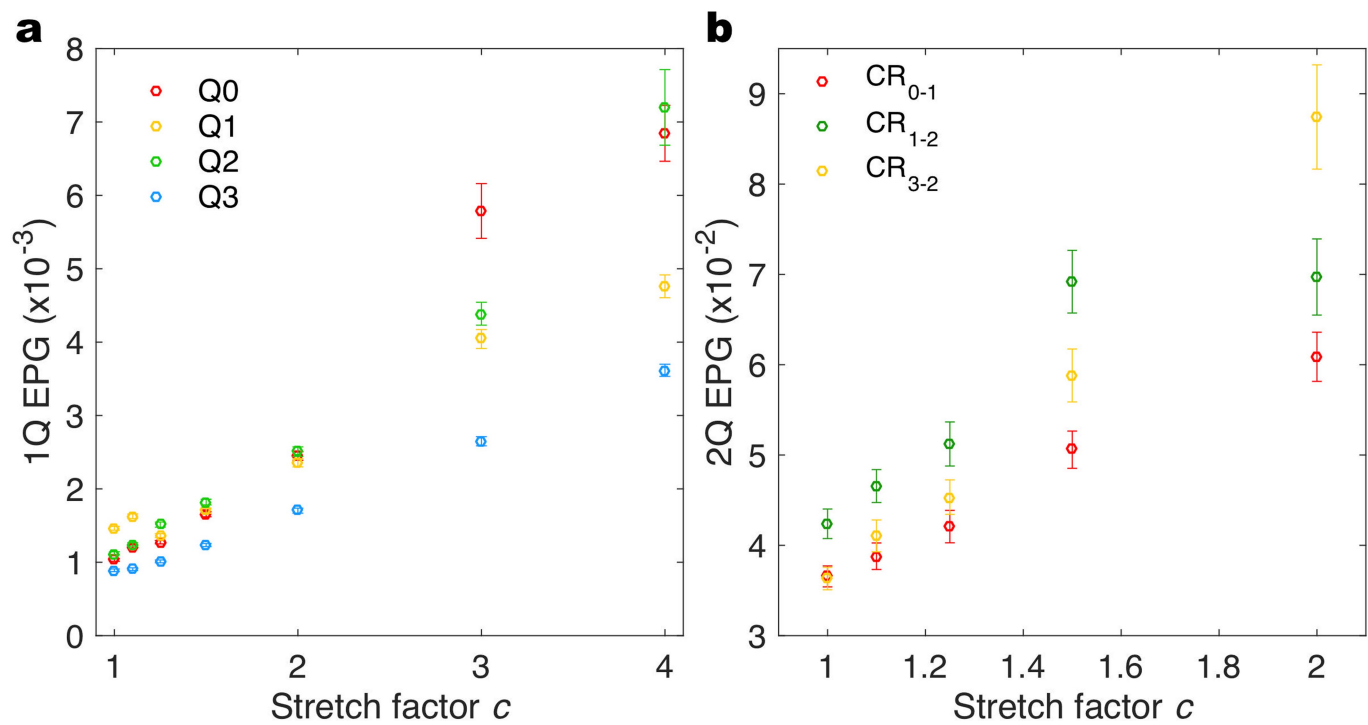
Sampling and bootstrapping. In addition to the effect of finite sampling, the accuracy of measured expectation values is sensitive to the readout assignment infidelity, in particular for large-weight Pauli operators. For the variational eigensolver described in Figs. 3, 4, the expectation values are therefore corrected using

a readout calibration of the possible outcomes, at every iteration, as in ref. ⁹. The variance of the measured expectation values for the different stretch factors then translates to error bounds on the mitigated estimates obtained from a zero-noise extrapolation. One obvious way to estimate the effect of finite sampling on the error bounds of the mitigated estimates is simply to repeat the experiment a large number of times. However, since this is an extremely time-consuming task, we simulate the spread arising from finite sampling using a well known statistical technique, bootstrapping⁴¹. Using the experimentally measured probability distributions, we resample both the readout calibrations and the measurements of the quantum state of interest. The resampled probability distributions are then used to evaluate the assignment-error-corrected expectation values, for each stretch factor, and consequently, the mitigated estimates. Running this bootstrapping protocol 50–100 times is then used to obtain a distribution of numerical outcomes for the mitigated expectation value.

Data availability

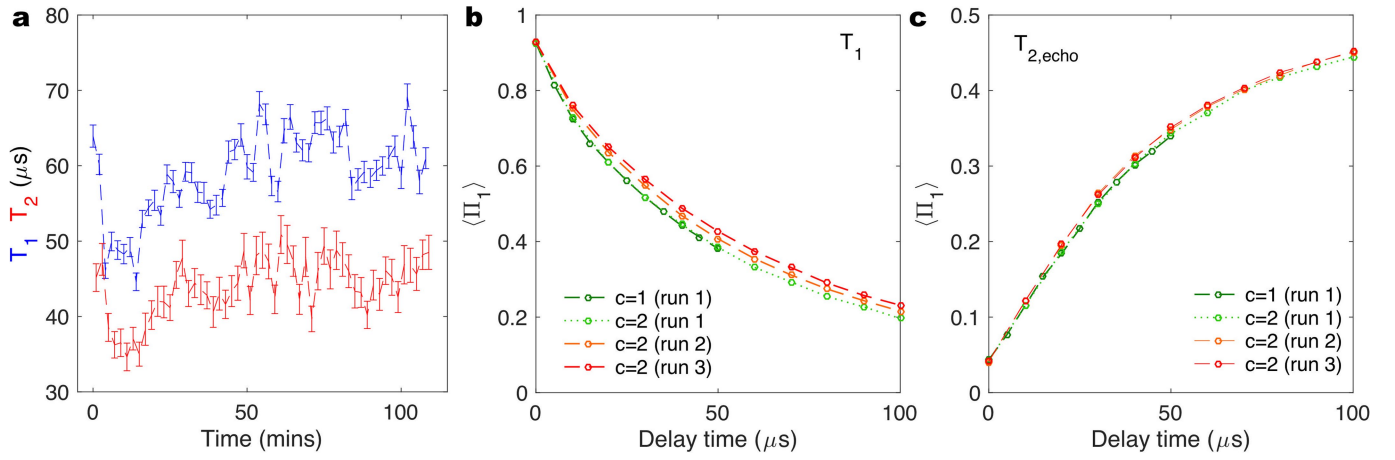
The data that support the findings of this study are available from the corresponding author on reasonable request.

31. Chow, J. M. et al. Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nat. Commun.* **5**, 4015 (2014).
32. 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).
33. Bergeal, N. et al. Analog information processing at the quantum limit with a josephson ring modulator. *Nat. Phys.* **6**, 296–302 (2010).
34. Abdo, B., Schackert, F., Hatridge, M., Rigetti, C. & Devoret, M. Josephson amplifier for qubit readout. *Appl. Phys. Lett.* **99**, 162506 (2011).
35. 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).
36. Chow, J. M. et al. Simple all-microwave entangling gate for fixed-frequency superconducting qubits. *Phys. Rev. Lett.* **107**, 080502 (2011).
37. Córcoles, A. D. et al. Process verification of two-qubit quantum gates by randomized benchmarking. *Phys. Rev. A* **87**, 030301 (2013).
38. 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).
39. Magesan, E. & Gambetta, J. M. Effective Hamiltonian models of the cross-resonance gate. Preprint at <https://arxiv.org/abs/1804.04073> (2018).
40. Sheldon, S. et al. Characterizing errors on qubit operations via iterative randomized benchmarking. *Phys. Rev. A* **93**, 012301 (2016).
41. Efron, B. in *Breakthroughs in Statistics* 569–593 (Springer, 1992).



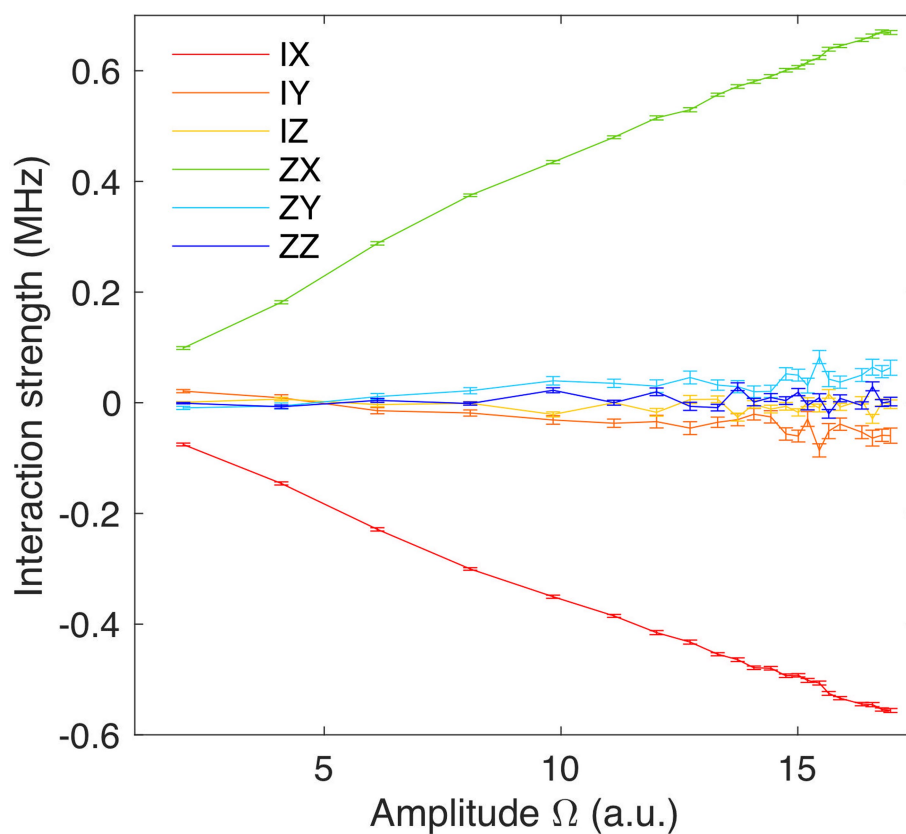
Extended Data Fig. 1 | Gate characterization. a, b, Single-qubit (1Q, **a**) and two-qubit (2Q, **b**) error per gate (EPG), for a range of stretch factors c , obtained by randomized benchmarking (RB). The error bars represent the standard deviation of the exponential fits to the RB decay, for each stretch

factor. The considered 1Q gates in **a** are for qubits Q0 (red), Q1 (yellow), Q2 (green), Q3 (blue) and the 2Q gates in **b** are CR₀₋₁ (red), CR₁₋₂ (green), CR₃₋₂ (yellow).



Extended Data Fig. 2 | Coherence and relaxation time fluctuations in superconducting qubits. **a**, Repeated measurements of T_1 and echo T_2 show fluctuations that are larger than the error bars to the fits. The error bars represent the standard deviation of the exponential fits to each instance of the T_1 and echo T_2 decay curves. **b**, **c**, Decay of excited-state

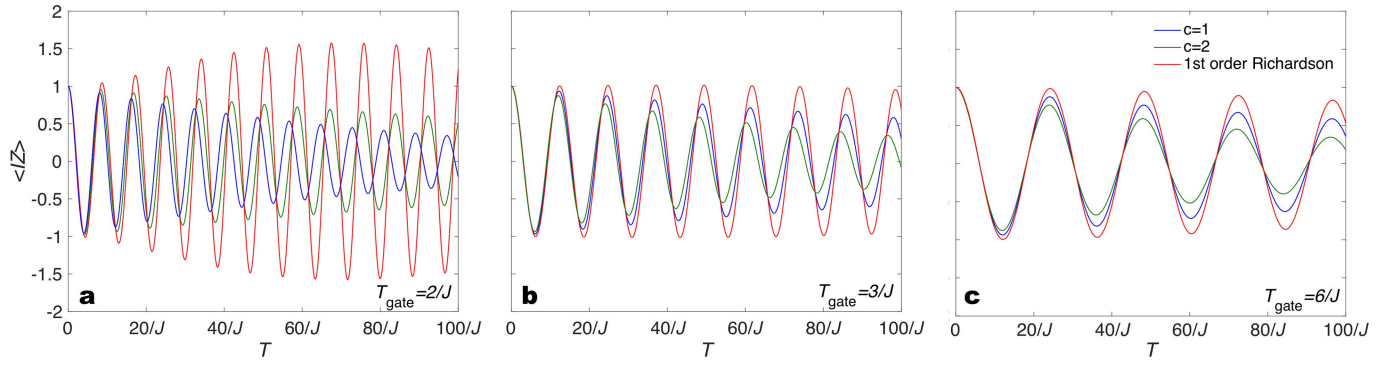
projector $\langle \Pi_1 \rangle$ for a standard T_1 sequence (**b**) and T_2 echo sequence (**c**), for stretch factors $c = 1, 2$ obtained using 10^5 samples. In both cases, the average decay over the total sampling duration for the different stretch factors is the same only when their measurements are grouped together (run 1), in contrast to when sampled separately (run 2 and run 3).



Extended Data Fig. 3 | Hamiltonian tomography of the CR drive.

Amplitude dependence of the interaction terms in the cross-resonance drive CR_{3-2} . The error bars represent the standard deviation of the interaction strengths, estimated from fits to Hamiltonian tomography²⁰ for

each CR drive amplitude. The dominant interactions in the drive are ZX and IX, with discernible nonlinearities in the strong drive limit. (a.u., arbitrary units.).



Extended Data Fig. 4 | Error mitigation with a nonlinear CR drive.

Time evolution of $\langle IZ \rangle$ under a simplistic CR drive $J_{ZX} = -0.0159J\Omega + 1.0541 \times 10^{-6}J\Omega^3$, and fixed amplitude damping and dephasing channels, for drive amplitudes calibrated to $ZX_{\pi/2}$ gate times T_{gate} values $2J^{-1}$ (a),

$3J^{-1}$ (b) and $6J^{-1}$ (c) and with stretch factors $c = 1$ (blue) and $c = 2$ (green). The mitigated time evolution (red) goes out of bounds for fast gates (large drive amplitudes), owing to the nonlinearity in the amplitude dependence.