

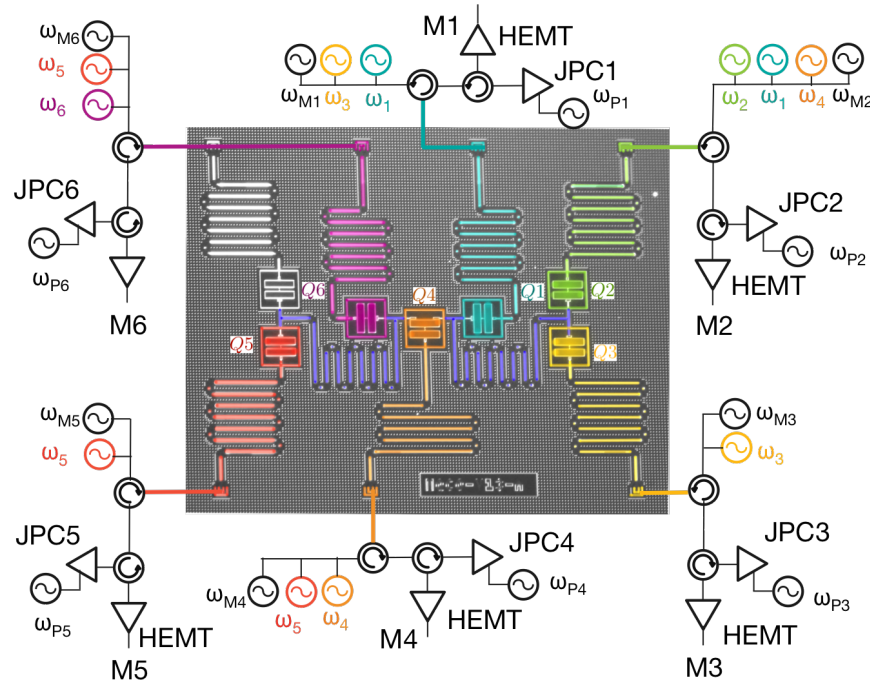


FIG. S1. Device and circuit schematic False colored optical micrograph depicts the components of our superconducting quantum processor: seven transmon qubits, two shared CPW resonators (in blue) for qubit-qubit coupling, and seven individual CPW resonators used for both, qubit control and readout. The qubits are controlled solely by microwave pulses that are delivered from the room temperature electronics via attenuated coaxial lines. The single qubit gates are implemented by microwave drives at the specific qubit Q_i 's frequency ω_i , while the entangling two-qubit CR gates are implemented by driving a control qubit Q_c at the frequency ω_c of the target qubit Q_t , where $i, c, t \in \{1, 2, 3, 4, 5, 6\}$. The state of each qubit is measured at its readout resonator frequency ω_{Mi} . The reflected readout signals are amplified first by a JPC, pumped at a frequency ω_{Pi} , followed by HEMT amplifiers at 4K.

I. DEVICE AND CHARACTERIZATION

The fundamental building blocks of our quantum hardware are superconducting Josephson junction (JJ) based qubits. The physical device includes 6 fixed frequency transmon qubits and a central flux-tunable asymmetric transmon qubit [1]. For the experiments discussed in this paper, we use 6 of these qubits, including the central flux-tunable qubit. The device connectivity is provided by two superconducting coplanar waveguide (CPW) resonators acting as quantum information buses, each of which couples four qubits, with the central asymmetric transmon coupled to both buses (see Fig. S1). Each qubit has its own individual CPW resonator for control and readout. The device is fabricated on a Si wafer using a single step of photolithography and sputtering for the superconducting Nb resonators and qubit capacitor pads, followed by e-beam lithography and double angle evaporation to define the Al-based JJ's. Refer to [2, 3] for further fabrication details.

Frequency crowding is an important issue for large networks of fixed frequency qubits employing cross resonance (CR) as an entangling gate, leading to crosstalk, leakage out of the computational sub-space or very slow gate times. Furthermore, current fabrication capabilities make it challenging to control the frequencies of transmons to within 200 MHz. In this context, we designed our central qubit Q4, which is directly coupled to all other qubits on the chip, to be weakly frequency tunable for reduced sensitivity to flux noise [1]. The qubit is referred to as an 'asymmetric transmon', and uses a superconducting quantum interference device (SQUID) as its inductive element. The two junctions in the SQUID however have different Josephson energies, engineered by varying the size of the junctions. An external superconducting coil is used to tune Q4 to its upper sweet spot, which, in the current experiment, is the optimal point for CR gates to its neighbors. The flux-tuning curve is shown in Fig. S2. At its upper sweet spot, Q4 is operated as a fixed frequency transmon, with coherence times that are comparable to other qubits on the chip Table S1.

The qubits are readout by dispersive measurements through independent readout resonators, with each readout line having a sequence of low temperature amplifiers — a Josephson parametric converter (JPC) [4, 5] followed by a high electron mobility transistor (model : LNF-LNC4 8A) — for achieving high assignment fidelity. For a measurement

Qubit	Q ₁	Q ₂	Q ₃	Q ₄	Q ₅	Q ₆
$\omega_{01}/2\pi$ (GHz)	5.3206	5.3567	5.2926	5.2455	5.2999	5.3882
T_1 (μ s)	24.7 ± 3.2	42.0 ± 5.1	20.4 ± 4.4	42.3 ± 5.2	44.4 ± 4.9	20.6 ± 0.8
T_2 (μ s)	31.1 ± 6.1	38.7 ± 12.5	35.3 ± 8.7	47.4 ± 14.0	60.5 ± 8.7	26.4 ± 4.3
T_2^* (μ s)	22.2 ± 4.8	28.6 ± 1.2	6.2 ± 0.9	36.7 ± 10.5	40.0 ± 3.2	27 ± 2.8
$\omega_r/2\pi$ (GHz)	6.6223	6.6892	6.5589	6.7154	6.6532	6.5885
$\delta/2\pi$ (GHz)	-0.311	-0.312	-0.315	-0.299	-0.311	-0.310
ϵ_r	0.0240	0.0544	0.0291	0.0469	0.0278	0.0507

TABLE S1. **Qubit and readout characterization.** Qubit transitions ($\omega_{01}/2\pi$), average relaxation times (T_1), average coherence times (T_2 , T_2^*), readout resonator frequencies ($\omega_r/2\pi$), qubit anharmonicity ($\delta/2\pi$), readout assignment errors (ϵ_r) for the six qubits discussed in the paper.

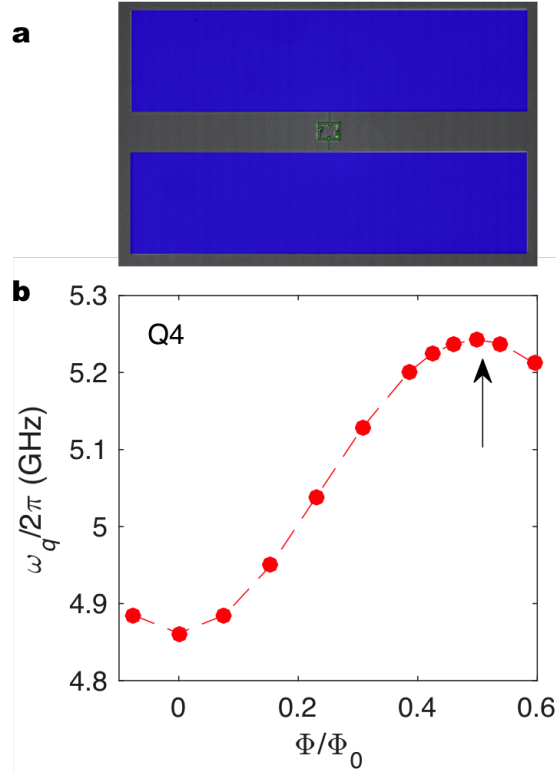


FIG. S2. **Asymmetric transmon and tuning curve** **a** False-colored optical micrograph of an asymmetric transmon, with an Al SQUID loop (in green), shunted by Nb capacitor pads (in blue). **b** Qubit frequency versus flux for the asymmetric transmon Q4. A constant flux offset is subtracted, and the flux is expressed in units of the flux quantum $\Phi_0 = h/2e$, where h is Planck's constant, and e is electric charge. The qubit is operated at its upper flux sweet spot, indicated by the arrow. The dashed line is a guide to the eye.

time of 1.5μ s, the joint readout assignment errors on Q2, Q4, Q6 are < 0.06 , and < 0.03 for Q1, Q3, and Q5. The anharmonicity of the fixed frequency qubits are ~ 310 MHz, while the asymmetric transmon has an anharmonicity of ~ 300 MHz. Further details of the device parameters are listed in Table S1.

The experimental implementation of variational quantum algorithms requires stability of the gates used for trial state preparation. Given the long times associated with optimization of large Hamiltonians, we periodically calibrate the amplitude and phase of our single-qubit and two-qubit gates during the course of the experiment. In order to estimate the time scale and magnitude of drifts in pulse amplitude and phase, we repeatedly calibrate our gates over several hours. For instance, Fig. S3 shows the drifts in the pulse amplitude for calibrated X_π pulses, expressed as angle deviations from the starting 180° X -rotation. Over the course of 18 hours, the deviations are less than 1.5° .

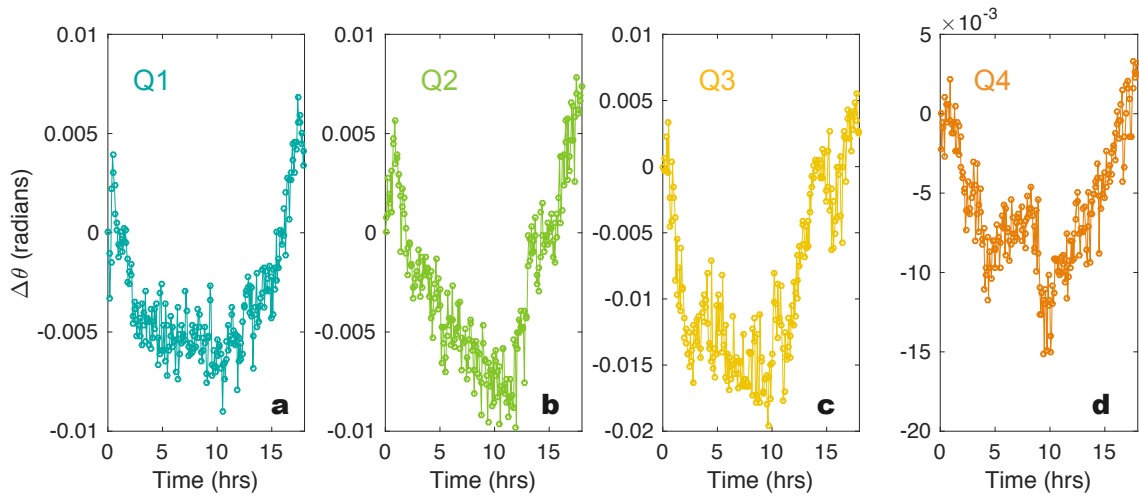


FIG. S3. **Single qubit gate drifts** Repeated calibrations of the amplitude for a X_π pulse over 18 hours for Q1-4 (a-d) reveal the magnitude and timescale for drifts in the amplitude of the single qubit gates. Here, the amplitude drifts are scaled as angle deviations $\Delta\theta$ from the starting X_π -rotation.

II. HARDWARE-EFFICIENT OPTIMIZATION OF QUANTUM HAMILTONIAN PROBLEMS

We present here a compact scheme describing the whole optimization algorithm. The individual subroutines of the method will be described in the following sections.

Algorithm Hardware-efficient optimization of quantum Hamiltonian problems

- 1: Map the quantum Hamiltonian problem to a qubit Hamiltonian H
 - 2: Choose a depth d for the quantum circuit that prepares the trial state
 - 3: Choose a set of variational controls θ_1 that parametrize the starting trial state
 - 4: Choose a number of samples S for the feedback loop and one S_f for the final estimation
 - 5: Choose a number of maximal control updates k_L
 - 6: **while** E_f has not converged **do**
 - 7: **procedure** QUANTUM FEEDBACK LOOP
 - 8: **for** $k = 1$ **to** k_L **do**
 - 9: Prepare trial states around θ_k and evaluate $\langle H \rangle$ with S samples
 - 10: Update and store the controls θ_k
 - 11: **end for**
 - 12: Evaluate $E_f = \langle H \rangle$ using the best controls with S_f samples
 - 13: **end procedure**
 - 14: Increase d, k_L, S, S_f
 - 15: **end while**
 - 16: **return** E_f
-

In the above algorithm, the first item describes the encoding of quantum Hamiltonians on a set of qubits. In the case of addressing a fermionic problem, we use an encoding and qubit reduction scheme from Ref. [6], explained in Section III, which is convenient for the molecular problems considered in this work. In general, different encodings could be considered, such as ones based on first-quantization methods. The outcome of the optimization depends on the parameters d, k_L, S, S_f , and in general will be better as these are increased, up to a point in which either one saturates the quantum resources available (e.g. decoherence limit, sampling time), or the optimization outcome E_f has converged: in this case increasing d, k_L, S, S_f will not improve the final answer E_f . In Section IV we describe the specific entangling gates we have used in the experiment to prepare trial states. In Section V we give details on the evaluation of the mean energy $\langle H \rangle$, and its dependence on the total number of samples S and experimental assignment errors. The energies measured in this way are then fed to a classical optimizer, described in Section VI. In Section VII we numerically estimate the resources required (circuit depth d , number of control updates k_L , number of samples S) to improve the accuracy of the optimization outcome.

III. MOLECULAR HAMILTONIANS

The molecular Hamiltonians considered in this work are computed in the STO-3G basis, using the software PyQuante [7] to obtain the one and two-electron integrals. The STO-3G minimal basis is obtained by fitting three gaussians to the Slater atomic orbitals, and commonly used in quantum chemistry because of the efficiency in obtaining electronic integrals [8]. For the H_2 molecule, each atom contributes a 1s orbital, for a total of 4 spin-orbitals. We set the X axis as the interatomic axis for the LiH and BeH_2 molecules, and consider the orbitals 1s for each H atom and 1s, 2s, $2p_x$ for the Li and Be atoms, assuming zero filling for the $2p_y$ and $2p_z$ orbitals, which do not interact strongly with the subset of orbitals considered. This choice of orbitals amounts to a total of 8 spin-orbitals for LiH and 10 for BeH_2 . The Hamiltonians are expressed using the second quantization language,

$$H = H_1 + H_2 = \sum_{\alpha,\beta=1}^M t_{\alpha\beta} a_{\alpha}^{\dagger} a_{\beta} + \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta=1}^M u_{\alpha\beta\gamma\delta} a_{\alpha}^{\dagger} a_{\gamma}^{\dagger} a_{\delta} a_{\beta}, \quad (1)$$

where $a_{\alpha}^{\dagger}(a_{\alpha})$ is the fermionic creation(annihilation) operator of the fermionic mode α , satisfying fermionic commutation rules $\{a_{\alpha}, a_{\beta}\} = 0$, $\{a_{\alpha}^{\dagger}, a_{\beta}^{\dagger}\} = 0$, $\{a_{\alpha}, a_{\beta}^{\dagger}\} = \delta_{\alpha\beta}$. Here $M = 4, 8, 10$ is the number of spin-orbitals for H_2 , LiH and BeH_2 respectively, and we have used the chemists' notation [8] for the two-body integrals,

$$t_{\alpha\beta} = \int d\mathbf{x}_1 \Psi_{\alpha}(\mathbf{x}_1) \left(-\frac{\nabla^2}{2} + \sum_i \frac{Z_i}{|\mathbf{r}_{1i}|} \right) \Psi_{\beta}(\mathbf{x}_1), \quad (2)$$

$$u_{\alpha\beta\gamma\delta} = \int \int d\mathbf{x}_1 d\mathbf{x}_2 \Psi_{\alpha}^*(\mathbf{x}_1) \Psi_{\beta}(\mathbf{x}_1) \frac{1}{|\mathbf{r}_{12}|} \Psi_{\gamma}^*(\mathbf{x}_2) \Psi_{\delta}(\mathbf{x}_2), \quad (3)$$

where we have defined the nuclei charges Z_i , the nuclei-electron and electron-electron separations $\mathbf{r}_{1i}$ and $\mathbf{r}_{12}$, the α -th orbital wavefunction $\Psi_{\alpha}(\mathbf{x}_1)$, and we have assumed that the spin is conserved in the spin-orbital indices α, β and $\alpha, \beta, \gamma, \delta$. In the case of LiH and BeH_2 , we then consider perfect filling for the inner 1s orbitals, dressed in the basis in which H_1 is diagonal. To this extent, we first implement a Bogoliubov transformation on the modes $a'_{\alpha} = \sum_{\beta} U_{\alpha\beta} a_{\beta}$, such that

$$H_1^d = U^{\dagger} H_1 U, \quad H_1^d = \sum_{\alpha=1}^M \omega'_{\alpha} a'_{\alpha}{}^{\dagger} a'_{\alpha}. \quad (4)$$

We then consider the “dressed” 1s modes of Li and Be to be filled, efficiently obtaining an effective Hamiltonian acting on generic states of the form $|\Psi\rangle = a'_{1s\uparrow}{}^{\dagger} a'_{1s\downarrow}{}^{\dagger} \left(\sum_{\beta \neq 1s\sigma} \psi_{\beta} a'_{\beta}{}^{\dagger} \right) |0\rangle$, where ψ_{β} are generic normalized coefficients, and $1s\sigma = \{1s\uparrow, 1s\downarrow\}$ refers to the inner 1s orbitals of Li and Be. Note that this approximation is valid whenever $-\omega'_{1s\sigma} \gg |u'_{\alpha\beta\gamma\delta}| \forall \sigma, \alpha, \beta, \gamma, \delta$, i.e. in the case of very low-energy orbitals that do not interact strongly with the higher-energy ones. The ansatz $|\Psi\rangle = a'_{1s\uparrow}{}^{\dagger} a'_{1s\downarrow}{}^{\dagger} \left(\sum_{\beta \neq 1s\sigma} \psi_{\beta} a'_{\beta}{}^{\dagger} \right) |0\rangle$ allows to define an effective screened Hamiltonian on the 1s orbitals for the hydrogen atoms, and 2s and $2p_x$ for Lithium and Berillium, for a total of 6 and 8 spin-orbitals for LiH and BeH_2 , respectively. According to this ansatz, the one-body fermionic terms containing the filled orbitals will now contribute as a shift to the total energy (I here is the identity operator)

$$\omega'_{1\uparrow} a'_{1\uparrow}{}^{\dagger} a'_{1\uparrow} \rightarrow \omega'_{1\uparrow} I, \quad \omega'_{1\downarrow} a'_{1\downarrow}{}^{\dagger} a'_{1\downarrow} \rightarrow \omega'_{1\downarrow} I, \quad (5)$$

while some of the two-body interactions, containing the set F of 1s filled modes of Li and Be, $F = \{1s\uparrow, 1s\downarrow\}$, become effective one-body or energy shift terms,

$$\frac{u'_{\alpha\beta\gamma\delta}}{2} a'_{\alpha}{}^{\dagger} a'_{\gamma}{}^{\dagger} a'_{\delta} a'_{\beta} \rightarrow \begin{cases} \frac{u'_{\alpha\beta\gamma\delta}}{2} a'_{\gamma}{}^{\dagger} a'_{\delta}, & \alpha = \beta, \alpha \in F, \{\gamma, \delta\} \notin F \\ \frac{u'_{\alpha\beta\gamma\delta}}{2} a'_{\alpha}{}^{\dagger} a'_{\beta}, & \gamma = \delta, \gamma \in F, \{\alpha, \beta\} \notin F \\ -\frac{u'_{\alpha\beta\gamma\delta}}{2} a'_{\gamma}{}^{\dagger} a'_{\beta}, & \alpha = \delta, \alpha \in F, \{\beta, \gamma\} \notin F \\ -\frac{u'_{\alpha\beta\gamma\delta}}{2} a'_{\alpha}{}^{\dagger} a'_{\delta}, & \gamma = \beta, \gamma \in F, \{\alpha, \delta\} \notin F \\ \frac{u'_{\alpha\beta\gamma\delta}}{2} I, & \alpha = \beta, \gamma = \delta, \alpha \neq \gamma, \{\alpha, \gamma\} \in F \\ -\frac{u'_{\alpha\beta\gamma\delta}}{2} I, & \alpha = \delta, \gamma = \beta, \alpha \neq \gamma, \{\alpha, \gamma\} \in F \end{cases} \quad (6)$$

while the two-body operators containing an odd number of modes in F will be neglected. We then map the fermionic Hamiltonians $H = \sum_{\alpha,\beta \neq 1s\sigma} t_{\alpha\beta} a'_{\alpha}{}^{\dagger} a'_{\beta} + 1/2 \sum_{\alpha,\beta,\gamma,\delta \neq 1s\sigma} u'_{\alpha\beta\gamma\delta} a'_{\alpha}{}^{\dagger} a'_{\gamma}{}^{\dagger} a'_{\delta} a'_{\beta}$ obtained in this way to our qubits. The H_2

Hamiltonian is mapped first onto 4 qubits using a binary-tree mapping [9]. We order the M spin-orbitals by listing first the $M/2$ spin-up ones and then the $M/2$ spin-down ones. When using the binary-tree mapping, this produces a qubit Hamiltonian diagonal in the second and fourth qubit, which has the total particle and spin $\mathbb{Z}_2$ symmetries encoded in those qubits [6]. For the LiH and BeH₂ Hamiltonians we use the parity mapping, which has the two $\mathbb{Z}_2$ symmetries encoded in the $M/2$ -th and M -th mode, even if the total number of spin orbitals is not a power of 2, as in the case of H₂. We then assign to the Z Pauli operators of the $M/2$ - and M -th qubits a value based on the total number of electrons m in the system according to

$$\{Z_{M/2}, Z_M\} = \begin{cases} \{+1, +1\}, & \text{mod}(m, 4) = 0 \\ \{+1, -1\}, & \text{mod}(m, 4) = 1 \\ \{-1, +1\}, & \text{mod}(m, 4) = 2 \\ \{-1, -1\}, & \text{mod}(m, 4) = 3, \end{cases} \quad (7)$$

The $+1(-1)$ on Z_M for even(odd) m implies an even(odd) total electron parity. The values $+1$, -1 and ± 1 for $Z_{M/2}$ mean that the total number of electrons with spin-up in the ground state is even, odd, or there is an even/odd degeneracy, respectively. In the last case both $+1$ and -1 can be used equivalently for $Z_{M/2}$. The final qubit-tapered Hamiltonians consist of 4, 99 and 164 Pauli terms supported on 2, 4, 6 qubits, each having 2, 25 and 44 tensor product basis (TPB) sets (see Section V) for H₂, LiH and BeH₂, respectively. We explicitly list the Hamiltonians at the bond distance in Table S2.

IV. CHARACTERIZATION OF THE ENTANGLERS

The entanglers in our hardware-efficient approach are collective gates composed of individual two-qubit gates on a convenient connectivity. For our fixed frequency, multi-qubit architecture, a good choice of two-qubit entangling gate is the microwave-only cross resonance (CR) gate [10–12]. These gates constitute the entanglers U_{ENT} in the trial state preparation and are implemented by driving a control qubit Q_c with a microwave pulse that is resonant with a target qubit Q_t . With the addition of single qubit rotations, the CR gate can be used to construct a controlled NOT (CNOT), with fidelities exceeding 99% for gate time ~ 160 ns [13]. In the hardware-efficient approach, however, tuning up a high-fidelity CNOT gate is not required, as long as entanglement is delivered with the CR drive. A simplistic model of the CR drive Hamiltonian is given by

$$H_D \approx / \hbar \epsilon_{CR}(t) \left(mIX - (J/\Delta)ZX + (\mu)ZI \right) \quad (8)$$

Here, $\epsilon_{CR}(t)$ is the CR drive amplitude, m quantifies the strength of the classical cross-talk, J is the strength of the qubit-qubit coupling, Δ is the frequency separation between the qubits, and μ corresponds to the drive induced Stark-shift. However, a more detailed study [13] of the drive revealed additional terms, whose strengths are revealed by Hamiltonian tomography. For instance, in the CR₂₋₄ drive used in the experiment, these terms are ZX : 1.04 MHz, ZY : 0.07 MHz, ZZ : 0.05 MHz, IX : 0.68 MHz, IY : 0.12 MHz, IZ : 0.02 MHz. We measure the norm of the Bloch vector $\|\mathbf{R}\|$ discussed in [13], whose time evolution indicates points of maximal entanglement at $\|\mathbf{R}\| = 0$; see Fig. S4b.

As discussed in the main text, the entangling gate phase could be an additional variational parameter for the optimization. However we show by numerical simulations that chemical accuracy (≈ 0.0016 Hartree, the accuracy of the energy estimate required to predict the exponentially sensitive chemical reaction rates at room temperature to within an order of magnitude of the exact value) can be reached for a range of fixed gate phases around points of maximum concurrence. This is shown in Fig. S4a,d which shows the error in the energy estimates from numerical optimization of the LiH Hamiltonian at bond distance, as a function of the gate phase of the two-qubit gates that compose the entanglers U_{ENT} . For these simulations, we choose ZX gates for U_{ENT} , using the same connectivity as the experiment (2-1, 1-3, 2-4 for the case of 4-qubit experiments). In order to isolate the effect of the entangling phase in the optimization, we do not consider a decoherence model and stochastic fluctuations in these simulations (as opposed to Fig. 3 and 4 in the main text), and set a high total number of energy evaluations to 5×10^4 . The results show plateaus of minimum energy errors, correlated with regions around points of maximal concurrence (Fig. S4c) for the individual two-qubit gates. Instead of setting our gate times to points of maximal concurrence, we choose them such that the corresponding gate phases lie at the beginning of the minimal error plateaus, in order to minimize the effect of decoherence while delivering sufficient entanglement. For our chosen two-qubit gate time of 150 ns, we extrapolate the phases of all CR gates under the simple assumption of having a time independent ZX Hamiltonian with finite pulse ramping times, and indicate them in Fig. S4a. Also, CR drives for qubits on different buses are driven simultaneously, in order to reduce the time associated with state preparation.

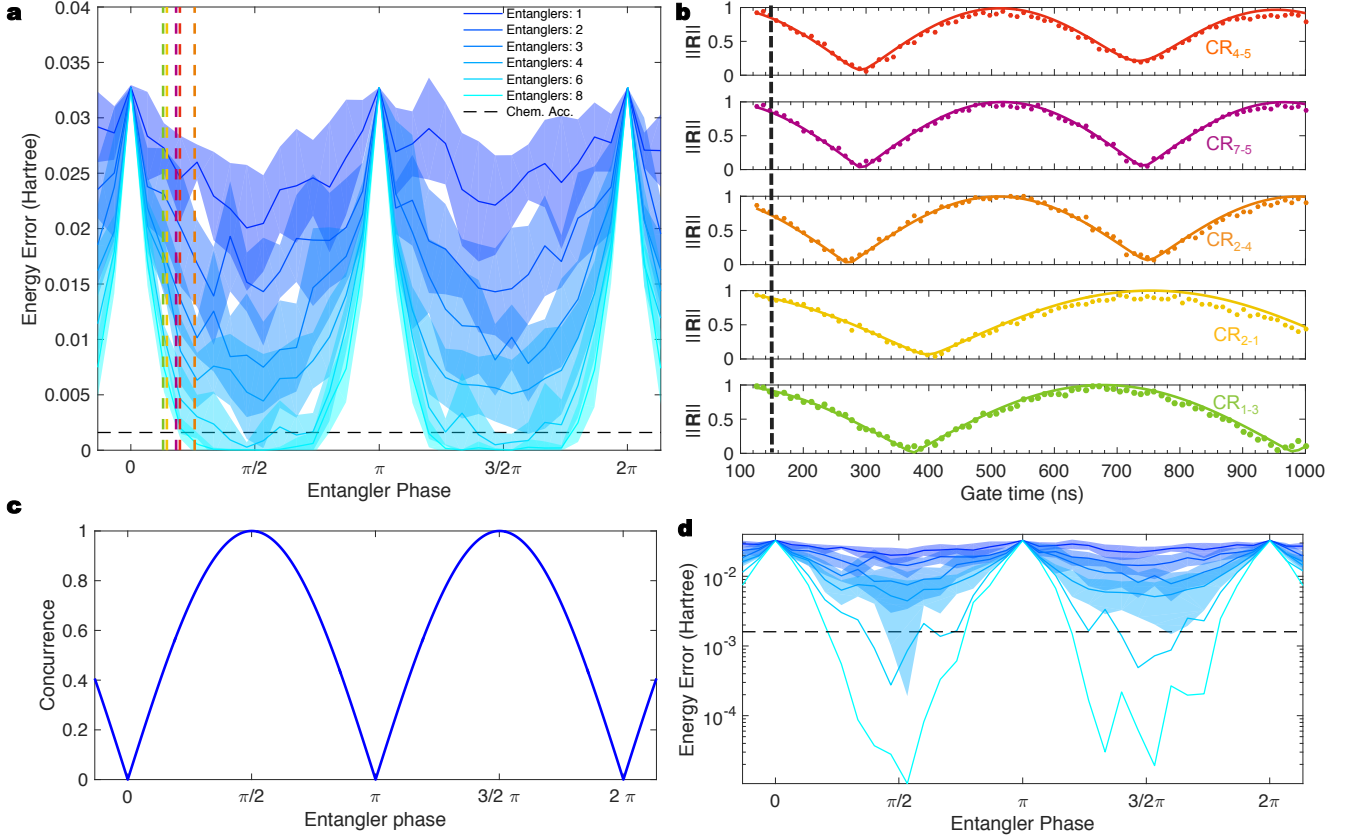


FIG. S4. **Dependence of energy error on entangler phase** **a** Energy error of numerical optimizations, as a function of the phase of the entanglers, for different depths $d = 1, 2, 3, 4, 6, 8$. The energy error is averaged over 10 optimization runs, for each depth, with bands represent the standard deviation of the distribution. The dashed vertical lines indicate approximate gate phases of the individual CR gates for the gate time of 150 ns, including finite pulse ramping times. **b** Norm of the Bloch vector $||\mathbf{R}||$ as a function of gate time for all the two-qubit entangling gates used in the experiment. The black dashed line corresponds to a gate time of 150 ns. The points where $||\mathbf{R}|| = 0$ indicate gate times of maximal entanglement. **c** Concurrence v/s gate phase of a ZX gate, starting from the state $(|10\rangle + |00\rangle)\sqrt{2}$. The energy error in **a** is least around points of maximal concurrence. **d** Energy error v/s entangler gate phase on a log linear scale. The dashed black line indicates chemical accuracy (0.0016 Hartree), showing that a critical depth $d = 6$ is required to achieve such accuracy. Color scheme follows from **a**.

V. ENERGY ESTIMATION

The update of the angles in our optimization routine is based on measurements of the expectation value of the Hamiltonian operator. These measurements are then used to build an approximation of the gradient of the energy landscape, which is in turn used to get a better update of the angles (see Section VI). The energy estimation at every k -th trial state of the optimization is a central part of the optimization algorithm, since its accuracy affects the final outcome of the optimization. Once mapped to qubits (see Section III), every molecular Hamiltonian is expressed as a weighted sum of T Pauli terms supported on N qubits

$$H = \sum_{\alpha=1}^T h_{\alpha} P_{\alpha}, \quad (9)$$

where each $P_{\alpha} \in \{X, Y, Z, I\}^{\otimes N}$ is a tensor product of single-qubit Pauli operators X, Y, Z and the identity I , on N qubits, with h_{α} being real coefficients. We are interested in estimating the mean energy $\langle \Phi(\boldsymbol{\theta}_k) | H | \Phi(\boldsymbol{\theta}_k) \rangle \equiv \langle H \rangle_k$ for the k -th control updates (more specifically for two sets of angles close to $\boldsymbol{\theta}_k$, see Section VI). This can be done by averaging measurements outcomes from individual experiments, where one prepares the same initial state, applies the quantum gates parametrized by $\boldsymbol{\theta}_k$, and finally performs projective measurements on the individual qubits. In the experiment we do not have access to direct measurements of the Hamiltonian operator $\langle H \rangle$ and its variance $\langle \Delta H^2 \rangle = \langle H^2 \rangle - \langle H \rangle^2$. Instead, we sample the individual Pauli operators P_{α} , estimating the mean values and

variances $\langle P_\alpha \rangle$, $\langle \Delta P_\alpha^2 \rangle = \langle P_\alpha^2 - \langle P_\alpha \rangle^2 \rangle$ from the measurements outcomes of the α -th Pauli operator. The energy and Hamiltonian variance can then be obtained as

$$\langle H \rangle = \sum_{\alpha=1}^T h_\alpha \langle P_\alpha \rangle, \quad (10)$$

$$\text{Var}[H] = \sum_{\alpha=1}^T h_\alpha^2 \langle \Delta P_\alpha^2 \rangle \quad (11)$$

Note that the variance on the mean energy $\text{Var}[H]$ is different from $\langle \Delta H^2 \rangle$, since we are sampling the individual Pauli terms separately: for example, eigenstates of H will have $\langle \Delta H^2 \rangle = 0$, but a finite $\text{Var}[H] \neq 0$. The error on the mean energy $\langle H \rangle$ after taking S samples for each Pauli operator is

$$\epsilon = \sqrt{\frac{\text{Var}[H]}{S}} \leq \sqrt{\frac{T |h_{\max}^2|}{S}} \quad (12)$$

where $h_{\max} = \max_\alpha |h_\alpha|$ is the absolute value of the largest Pauli coefficient. Since sampling S times for a large number of trial states and Pauli operators comes with significant time overhead, one can instead use the same state preparations to measure different Pauli operators. This approach was considered in [14] for commuting operators. Here we use a stronger condition on grouping different Pauli terms, based on improving time efficiency. We first briefly describe how we sample an individual Pauli operator. The individual Pauli operators are measured by correlating measurement outcomes of single-qubit dispersive readouts in the Z basis, which can be done simultaneously since each qubit is provided with an individual readout resonator. In case a target multi-qubit Pauli operator contains non diagonal single-qubit Pauli operator, single-qubit rotations (post-rotations) are performed before the measurement in the Z basis. Specifically, a $-\pi/2(\pi/2)$ rotation along the $X(Y)$ axis to measure a $Y(X)$ single-qubit Pauli operator.

A. Grouping Pauli Operators

To minimize sampling overheads, we group the T Pauli operators P_α in A sets $s_1, s_2, \dots, s_A$, which have terms that are diagonal in the same tensor product basis. The post-rotations required to measure all the Pauli terms in a given TPB set are the same, and a unique state preparation can be used to sample all the Pauli operators in the same set.

By doing so, however, covariance effects in the same TPB set contribute to the variance of the total Hamiltonian,

$$\text{Var}^G[H] = \sum_{i=1}^A \sum_{\alpha, \beta \in s_i} h_\alpha h_\beta \langle (P_\alpha - \langle P_\alpha \rangle)(P_\beta - \langle P_\beta \rangle) \rangle \leq h_{\max}^2 (T + A s_{\max}^2), \quad (13)$$

where $s_{\max} = \max_i |s_i|$ is the number of elements in the largest TPB set. Keeping the same total number of measurements TS as in Eq. (12), the error on the mean in this case is given by

$$\epsilon = \sqrt{\frac{\text{Var}^G[H]}{S}} \leq \sqrt{\frac{A h_{\max}^2 (T + A s_{\max}^2)}{TS}}, \quad (14)$$

which can be compared to the case in which one samples the single Pauli terms individually, Eq. (12). The error contribution from the covariance (which can be positive or negative) has to be traded off against the use of less samples from grouping. The quantities in Eqs. (10) and (13) can be estimated in the experiment and in the numerical simulations as

$$\widehat{\langle P_\alpha \rangle} = \frac{1}{S} \sum_{i=1}^S X_{i,\alpha}, \quad (15)$$

$$\widehat{\text{Var}^G[H]} = \sum_{i=1}^A \sum_{\alpha, \beta \in s_i} h_\alpha h_\beta \text{cov}(\widehat{\langle P_\alpha \rangle}, \widehat{\langle P_\beta \rangle}), \quad (16)$$

where we have defined the outcome of the i -th measurement on the α -th Pauli term as $X_{i,\alpha}$. The covariance matrix element is defined after S measurements as

$$\text{cov}(\widehat{\langle P_\alpha \rangle}, \widehat{\langle P_\beta \rangle}) = \frac{1}{S-1} \sum_{i=1}^S (X_{i,\alpha} - \widehat{\langle P_\alpha \rangle})(X_{i,\beta} - \widehat{\langle P_\beta \rangle}). \quad (17)$$

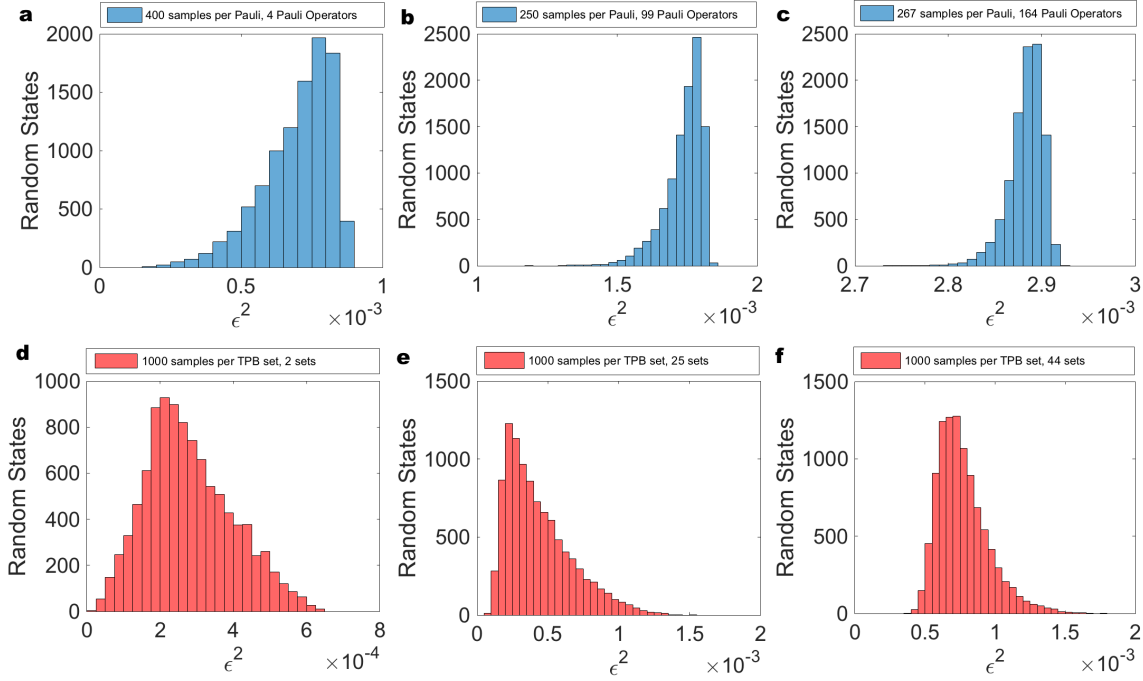


FIG. S5. **Energy variance** Numerical computation of the variance of the mean energy ϵ^2 , as in Eq. (16), with $S = 10^3$ samples, for the molecular Hamiltonians of H_2 (a, d), LiH (b, e) and BeH_2 (c, f) at their bond interatomic distances (see Table S2). The variances are computed sampling each Pauli operator P_α in H of Eq. (9) individually (a, b, c) and grouping them in TPB sets (d, e, f), keeping the total number of samples the same.

To evaluate whether grouping into TPB sets is convenient for the molecular Hamiltonians considered in this work, we perform numerical sampling experiments, shown in Fig. S5, using the Hamiltonians in Table S2. The variance of the mean energy is numerically sampled on 10^4 random states. In the “TPB sets” simulations (red histograms), the set of post-rotations associated to each TPB set is found by union of the set of post-rotations necessary to sample each Pauli in a given TPB set: for example, for the third TPB set of BeH_2 in Table S2 we have the post-rotations associated to ZZXXZX. Then, for each random state, a sample of $S = 10^3$ measurement outcomes are drawn for every TPB set. The total number of measurement is therefore AS . These measurements are then used to obtain the mean value and covariance for each Pauli operator in the TPB set. The variance of the mean total energy is then obtained as in Eq. (13). In the “No-TPB sets” simulations (blue histograms), the same measurements are drawn independently for each Pauli operator, with a number of samples per Pauli term SA/T , in order to keep the total number of samples in the TPB and No-TBP simulations the same. The results show the advantage of grouping into TPB sets for all the molecular Hamiltonians considered.

B. Assignment Errors

An important aspect to take into account when sampling is the presence of assignment errors at the qubit readout. A qubit-independent assignment error can be modeled by a deformation $\hat{\Pi}_0, \hat{\Pi}_1$, of the ideal projectors Π_0, Π_1 on the $|0\rangle, |1\rangle$ states for the qubit,

$$\begin{aligned}\hat{\Pi}_0 &= (1 - \eta_0 + \eta_1)\Pi_0 + (1 - \eta_0 - \eta_1)\Pi_1 = (1 - \eta_0)I + \eta_1 Z \\ \hat{\Pi}_1 &= (\eta_0 - \eta_1)\Pi_0 + (\eta_0 + \eta_1)\Pi_1 = \eta_0 I - \eta_1 Z,\end{aligned}\tag{18}$$

via the two parameters η_0, η_1 (note that in the absence of errors $\eta_0 = \eta_1 = 1/2$), such that $\hat{\Pi}_0 + \hat{\Pi}_1 = I$. With these definitions, the assignment error of reading a qubit in $|1\rangle(|0\rangle)$ when it is in $|0\rangle(|1\rangle)$ is given by $1 - \eta_0 - \eta_1$, or $(\eta_0 - \eta_1)$. The measured readout assignment error, averaged on preparations of $|0\rangle$ and $|1\rangle$ in Table S1, can be expressed with the parametrization considered as $\epsilon_r = 1/2 - \eta_1$. The projectors in Eqs. (18) define an effective deformed $\hat{Z}$ operator,

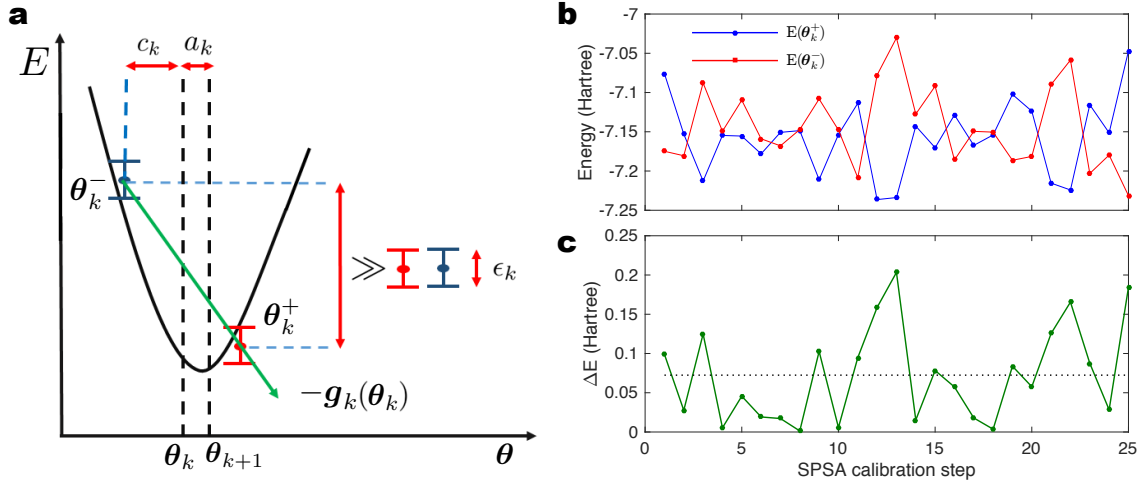


FIG. S6. **Calibration of the classical optimizer** **a** Good gradient approximations $\mathbf{g}_k(\boldsymbol{\theta}_k)$ are obtained if the energy difference $|\langle \Phi(\boldsymbol{\theta}_k^+) | H | \Phi(\boldsymbol{\theta}_k^+) \rangle - \langle \Phi(\boldsymbol{\theta}_k^-) | H | \Phi(\boldsymbol{\theta}_k^-) \rangle|$ is larger than the stochastic fluctuations on the energy ϵ_k . The parameter c in Eq. (23) is heuristically chosen to meet this condition. **b** The parameter a in Eq. (23) is calibrated by measuring 25 times the energies $E(\boldsymbol{\theta}_1^\pm) = \langle \Phi(\boldsymbol{\theta}_1^\pm) | H | \Phi(\boldsymbol{\theta}_1^\pm) \rangle$, measured here for the LiH molecule at the bond distances, from the starting angles $\boldsymbol{\theta}_1$, for different random gradients approximations. **c** The energy difference $\Delta E = |\langle \Phi(\boldsymbol{\theta}_1^+) | H | \Phi(\boldsymbol{\theta}_1^+) \rangle - \langle \Phi(\boldsymbol{\theta}_1^-) | H | \Phi(\boldsymbol{\theta}_1^-) \rangle|$ is measured for each random instance of the gradient (solid green line), averaged (black dotted line), and then used to calibrate the parameter a , according to Eq. (25).

related to the ideal one Z via

$$\hat{Z} = \hat{\Pi}_0 - \hat{\Pi}_1, \quad Z = \frac{\hat{Z} - (1 - 2\eta_0)\mathbf{I}}{2\eta_1}. \quad (19)$$

Note that the measured value $\langle \hat{Z} \rangle$ is affected by the contrast factor $2\eta_1$, and shifted by the amount $1 - 2\eta_0$. Generalizing this to a Pauli operator with weight w , one has that

$$Z^{\otimes w} \propto \frac{\hat{Z}^{\otimes w}}{(2\eta_1)^w}, \quad (20)$$

revealing an exponential loss in contrast in the weight w . When addressing larger systems, it will then be important to use the binary tree encoding [9], for its logarithmic scaling in locality with the system size, to combat the exponential scaling in (20). Note that the error model in Eq. (18) only takes into account independent readout errors, while in general correlated readout errors may happen. In our experiments we take into account assignment errors by running readout calibrations before sampling for every update of the angles $\boldsymbol{\theta}$, and then correcting our sampling outcome with the calibrations.

VI. OPTIMIZATION USING A SIMULTANEOUS PERTURBATION METHOD

The energy $\langle \Phi(\boldsymbol{\theta}_k) | H | \Phi(\boldsymbol{\theta}_k) \rangle \equiv \langle H \rangle_k$ discussed in Section V, which needs to be evaluated before every update of the angles $\boldsymbol{\theta}$, has a number of parameters $p = N(3d - 1)$ that grows linearly with the depth of the circuit d and the number of qubits N . As the number of parameters increases the classical optimization component of the algorithm comes with increasing overheads. The accuracy of the optimization may also be significantly lowered by the presence of energy fluctuations at the k -th step ϵ_k . Furthermore, on real quantum hardware, there are time overheads associated with loading of pulse waveforms on the electronics, resonator and qubit reset, and repeated sampling of the qubit readout. Ideally, one would like to use an optimizer robust to statistical fluctuations, that uses the least number of energy measurements per iteration. The simultaneous perturbation stochastic approximation (SPSA) algorithm, introduced in [15], is a gradient-descent method that gives a level of accuracy in the optimization of the cost function that is comparable with finite-difference gradient approximations, while saving an order $\mathcal{O}(p)$ of cost function evaluations. It has been recently used in the context of quantum control and quantum tomography [16–18].

In the SPSA approach, for every step k of the optimization, we sample from p symmetrical Bernoulli distributions (coin flips) Δ_k , and use preassigned elements from two sequences converging to zero, c_k and a_k . The gradient at θ_k is approximated using energy evaluations at $\theta_k^\pm = \theta_k \pm c_k \Delta_k$, and is constructed as

$$\mathbf{g}_k(\theta_k) = \frac{\langle \Phi(\theta_k^+) | H | \Phi(\theta_k^+) \rangle - \langle \Phi(\theta_k^-) | H | \Phi(\theta_k^-) \rangle}{2c_k} \Delta_k, \quad (21)$$

as illustrated in Fig. S6a. Note that this gradient approximation only requires two estimations of the energy, regardless of the number p of variables in θ . The controls are then updated as

$$\theta_{k+1} = \theta_k - a_k \mathbf{g}_k(\theta_k). \quad (22)$$

The convergence of θ_k to the optimal solution θ^* can be proven even in the presence of stochastic fluctuations, if the starting point is in the domain of the attraction of the problem [15]. Convergence remains an open issue if the starting point for the controls is not in a domain of attraction. In this case strategies like multiple competing starting points can be adopted [19]. The sequences c_k, a_k can be chosen as

$$\begin{aligned} c_k &= \frac{c}{k^\gamma}, \\ a_k &= \frac{a}{k^\alpha}. \end{aligned} \quad (23)$$

We pick the parameters α, γ optimally at $\{\alpha, \gamma\} = \{0.602, 0.101\}$ [20], ensuring the smoothest descent along the approximate gradients defined in Eq. (22). We then tune the value of c to adjust the robustness of the gradient evaluation with respect to the magnitude of the energy fluctuations. In fact, large fluctuations of the energy require gradient evaluations with large c_k (21), so that the fluctuations do not substantially affect the gradient approximation. This condition is valid in the regime

$$|\langle \Phi(\theta_k^+) | H | \Phi(\theta_k^+) \rangle - \langle \Phi(\theta_k^-) | H | \Phi(\theta_k^-) \rangle| \gg \epsilon_k, \quad (24)$$

depicted visually in Fig. S6a. Keeping these considerations in mind, we have used $c = 10^{-1}$ to ensure robustness in all the experiments and in the realistic simulations that include decoherence noise and energy fluctuations, while the smaller $c = 10^{-2}$ factor is used in the numerical optimizations where the energy is evaluated without fluctuations. The parameter a is then calibrated experimentally in order to achieve a reasonable angle update on the first step of the optimization, which we chose to be $|\theta_2^{(i)} - \theta_1^{(i)}| = 2\pi/10$, for all the angles $i = 1, 2, \dots, p$. To achieve this, we use an inverse formula based on Eq. (22),

$$a = \frac{2\pi}{5} \frac{c}{\left\langle |\langle \Phi(\theta_1^+) | H | \Phi(\theta_1^+) \rangle - \langle \Phi(\theta_1^-) | H | \Phi(\theta_1^-) \rangle| \right\rangle_{\Delta_1}}, \quad (25)$$

where the notation $\langle \rangle_{\Delta_1}$ indicates an average over different samples from the distribution Δ_1 that generates the first gradient approximation. In fact, by averaging along different directions, we can measure the average slope of the functional landscape of $\langle \Phi(\theta) | H | \Phi(\theta) \rangle$ in the vicinity of the starting point θ_1 , and calibrate the experiment accordingly. In the experiment and in the numerics the average $\langle \rangle_{\Delta_1}$ is realized over 25 random gradient directions. The gradient averaging is shown for the optimization of the LiH Hamiltonian at bond distance with a $d = 1$ circuit, in Fig. S6b,c.

Note that along the optimization we do not measure the value of the energy for the k -th optimized angles $\langle \Phi(\theta_k) | H | \Phi(\theta_k) \rangle$, instead we only measure and report the values $\langle \Phi(\theta_k^+) | H | \Phi(\theta_k^+) \rangle$ and $\langle \Phi(\theta_k^-) | H | \Phi(\theta_k^-) \rangle$, which serve to generate a new gradient approximation. The underlying optimized angles θ_k are only measured at the end of the optimization, averaging over the last 25 θ_k^+ and 25 θ_k^- , to further minimize stochastic fluctuations effect. Furthermore, this last average is done with 10^5 samples, as opposed to the 10^3 samples used to generate θ_k^+ and θ_k^- during the optimization, in order to reduce the error on the measurement.

VII. NUMERICAL SIMULATIONS AND SCALING OF RESOURCES

In this Section we first describe the numerical simulations used in Fig. 3 and Fig. 4, which include decoherence effects and stochastic fluctuations on the energy evaluation. We then show numerical results that indicate the scaling of the optimization outcome with the depth of the trial state preparation circuit, the number of angle updates considered in the optimization, and the sampling statistics. We estimate the resources necessary to achieve chemical accuracy for the three molecules considered. Last, we show the interplay between circuit depth and decoherence affecting the quantum circuit, using a depolarizing noise model.

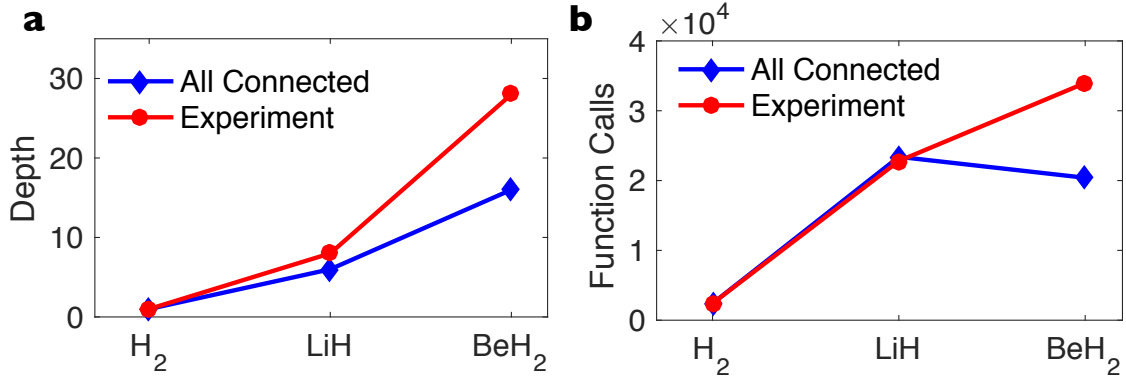


FIG. S7. **Scaling of resources to reach chemical accuracy.** **a** The critical depth required for reaching chemical accuracy for the 3 molecules discussed in the paper, using an all-to-all qubit connectivity (blue) and the experimental qubit connectivity (red). **b** The number of function calls for reaching chemical accuracy for the 3 molecules at their respective critical depths from **a**. Each data point in both plots is obtained by averaging over 10 optimization runs.

A. Numerical model of the experiment

In the numerical simulations in Fig. 3, Fig. 4 and Fig. S9, we have used entanglers made up of ZX two-qubit entangling gates, with a phase of $\pi/4$, and with additional terms ZY , ZZ , IX , IY , and IZ , whose relative phases are chosen according to the measurement reported in Section IV for CR_{2-4} . We use the same connectivity as in the experiment, with entangling gates between qubits 1 – 2, 2 – 4 and 1 – 3 in the 4-qubit simulations (LiH and quantum magnetism model) and gates between qubits 1 – 2, 2 – 4, 1 – 3, 4 – 5 and 5 – 6 in the 6-qubit simulations (BeH_2). The initial Z angles are distributed normally around zero according to $\mathcal{N}(0, 1)$, and the X angles set to $\pi/2$.

The effect of decoherence is taken into account by adding amplitude damping ($E_0^a(\tau), E_1^a(\tau)$) and dephasing ($E_0^d(\tau), E_1^d(\tau)$) channels acting on the system density matrix $\rho \rightarrow E_0^a(\tau)\rho E_0^{a\dagger}(\tau) + E_1^a(\tau)\rho E_1^{a\dagger}(\tau)$, $\rho \rightarrow E_0^d(\tau)\rho E_0^{d\dagger}(\tau) + E_1^d(\tau)\rho E_1^{d\dagger}(\tau)$, for all the qubits, after each round of Euler gates and entanglers, respectively. The strength of the channels is set by the experimental coherence times and the length of the gates,

$$E_0^a(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{e^{-\tau/T_1}} \end{bmatrix}, E_1^a(\tau) = \begin{bmatrix} 0 & \sqrt{1 - e^{-\tau/T_1}} \\ 0 & 0 \end{bmatrix} \quad (26)$$

$$E_0^d(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\tau/T_\phi} \end{bmatrix}, E_1^d(\tau) = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{1 - e^{-2\tau/T_\phi}} \end{bmatrix}. \quad (27)$$

Here the time τ alternates between the duration of each single qubit gate sequence or entangler step, and the pure dephasing time is defined as $T_\phi = 2T_2^*T_1/(2T_1 - T_2^*)$, see Table S1 for measured values on each qubit. In the H_2 simulations, since we use the most coherent qubits on the chip, we parametrize the noise channels considering $T_1 = T_2^* = 40 \mu s$ and set the length of U_{ENT} to 150 ns, while for the 4 and 6-qubit simulations we use typical coherence values for the qubits of $T_1 = 30 \mu s$, $T_2^* = 20 \mu s$ and a duration for U_{ENT} of 450 ns. Note that the duration for both 4 and 6-qubit entanglers is set to be the same because the two-qubit gates CR_{2-1} , CR_{4-5} and CR_{1-3} , CR_{6-5} are done in parallel, see Fig. 1c in the main text. To simulate the effect of finite sampling in the experiment, we first compute an average value of the standard deviation of the energy by sampling 10^3 times on 100 random states, as described in Section V. Then we add a normal-distributed error to each energy evaluation along the optimization, with the standard mean deviation computed previously on random states. On average, this will account for the energy fluctuations at the k -th step of the optimization. We fix the total number of angle updates to 250. For the final energy estimate, we average over the last 25 control updates, to mitigate the effect of stochastic fluctuation in the optimization. For every interatomic distance (for every J/B ratio in the case of Fig. 4), we show the outcome of 100 numerical simulations, in the form of a density plot, in Fig. 3 (Fig. 4) in the main text.

B. Scaling of resources: depth, function calls, sampling

In order to estimate resources required to reach chemical accuracy (i.e. an energy error of approximately 0.0016 Hartree), we consider molecular Hamiltonians at the bond distance for H_2 , LiH and BeH_2 (see Table S2), and declare convergence when the best energy estimate is close to the exact solution up to chemical accuracy. We assume that the resources required to reach chemical accuracy at the bond distance are comparable with the ones for any other interatomic distance, ensuring chemical accuracy also for the dissociation energy (defined as the molecular energy difference at the bond length and in the limit of infinite interatomic energy). In these simulations for determining the scaling of the resources, we consider ideal ZZ entangling gates with a phase of $\pi/2$. Note that any two qubit interaction can be mapped to a ZZ one via local rotations (i.e. our Euler angles). We use only the last two single-qubit rotation for each step, since Z rotations commute with the ZZ entangling gates, and consider two different topologies for the qubit connectivity: in addition to the experimental connectivity, we consider an “all connected” connectivity, where the entanglers U_{ENT} are composed of ZZ gates among all the qubit pairs in the system.

For the simulations outcomes plotted in Fig. S7a, we set a maximal number of function calls to 5×10^4 (i.e. evaluations of the energy as described in Section VI), ensuring convergence of the optimization beyond chemical accuracy for all the simulations considered. We start by not taking into account decoherence and stochastic fluctuations, run 10 optimizations for increasing circuit depths, average the final optimized energies, and report the shortest depth that has an average energy converged within chemical accuracy. Chemical accuracy is reached for depths $d = 1, 8, 28$ for the experimental connectivity, and $d = 1, 6, 16$ for the all connected case, for H_2 , LiH and BeH_2 , respectively. Having computed the shortest circuit depth for each molecule and connectivity, we now keep the circuit depth fixed and run optimizations, keeping track of the number of trial states sufficient to achieve chemical accuracy. We average the number of trial states obtained for 10 separate optimizations. The results are plotted in Fig. S7b. Approximately 2×10^3 function calls (10^3 angle updates) are sufficient for reaching chemical accuracy on H_2 , 2×10^4 for LiH_2 both for the all-connected and experiment connectivity, 2×10^4 for BeH_2 in the all-connected case and approximately 3×10^4 for the experiment connectivity.

We finally estimate the number of samples S required to reach chemical accuracy. We start by computing an average standard deviation ϵ_A for the energy on 10^2 random states, considering $S = 10^3$ samples, see Section V. Then we add the averaged deviation to the energies evaluated at the k -th step of the optimization. Then, we extrapolate standard deviations at higher samplings S , via $\epsilon_A \rightarrow \epsilon_A \sqrt{10^3/S}$. Using the depths indicated in Fig. S7a, we find that chemical accuracy is reached for all the three molecules when the number of samples is $S \approx 10^6$, i.e. approximately when all the energies in the optimization are evaluated at chemical accuracy. This can be understood by using values for the standard deviations of the mean energies as in Fig. S5, computed at 10^3 samples, and extrapolating to 10^6 samples. These results indicate a scaling of the resources with the problem size which is not very dramatic. If we set aside decoherence effects, both number of function calls and sampling could be increased in the near future by rapid reset protocols of the qubits [21–23].

C. Scaling of resources: decoherence

In order to address the behavior of the optimization versus decoherence effects, we run numerical simulations that include a depolarizing noise model following each gate. We consider one-qubit and two-qubit depolarizing channels acting on the system density matrix ρ as

$$\begin{aligned} \rho &\rightarrow (1 - \xi)\rho + \frac{\xi}{3} \sum_{i=1,2,3} \sigma^i \rho \sigma^i, \\ \rho &\rightarrow (1 - \xi)\rho + \frac{\xi}{15} \sum_{\substack{\{i,j\}=\{0,1,2,3\} \\ \{i,j\} \neq \{0,0\}}} \sigma_l^j \sigma_m^i \rho \sigma_m^i \sigma_l^j, \end{aligned} \quad (28)$$

where $\sigma^1 = X, \sigma^2 = Y, \sigma^3 = Z, \sigma^0 = I$. The single-qubit depolarizing channels act on every qubit after the Euler rotations, while the two-qubit channels act on every qubit pair $\{l, m\}$ considered in a given connectivity. We run noisy optimizations for the LiH Hamiltonian at the bond distance, for different number of entanglers and noise strengths, for a maximum of 5×10^4 function calls. The results are shown in Fig. S8, averaged on 10 different optimizations. There is a clear interplay between the number of entanglers and the noise strength. For low noise rates ξ , higher depths give better results, while as ϵ increases lower depths perform better. Chemical accuracy is reached for noise rates of $\approx 10^{-5}$, for 6 and 8 entanglers. Such low noise rates emphasize that it will be important in the near future to explore error mitigation methods for short depth quantum circuits [24–26].

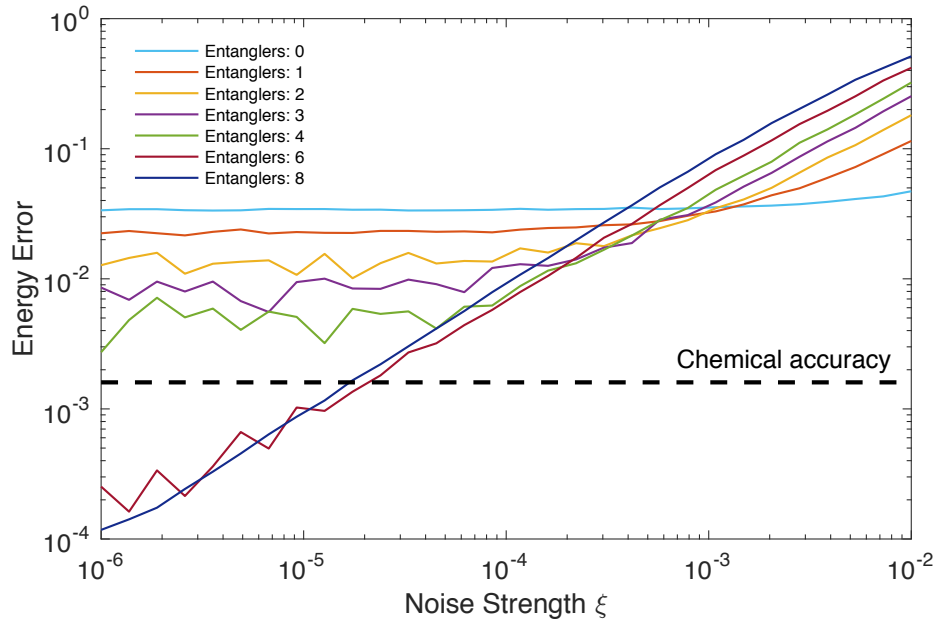


FIG. S8. **Scaling of energy error with noise strength** Error in the energy estimate for the 4-qubit LiH Hamiltonian at its bond length, for different depolarizing noise strengths of the model in Eq. (28), for different circuit depths used for trial state preparation, after 5×10^4 function calls. Each data point is obtained by averaging over 10 optimization runs. The black dashed line indicates the energy error for chemical accuracy.

When considering the combined effects of decoherence, stochastic fluctuations due to finite sampling and limited number of trial states, the advantages of using more entanglers may not be apparent anymore. This is the case for many of the molecular Hamiltonians discussed in this paper, whose energies are well approximated by separable states prepared using low-depth circuits. In Fig. S9 we show the experimental optimization for different depths, $d = 0, 1, 2$, for the Hamiltonian of LiH at the bond distance, compared with 100 outcomes of numerical simulations. The numerical histograms in Fig. S9b show large overlap between final energy distributions for $d = 0, 1, 2$, confirmed by the experiments presented in Fig. S9a. This overlap between outcomes of optimizations with different entanglers appear for most of the molecular Hamiltonians. In contrast, for the interacting spin Hamiltonians discussed in Fig. 4 of the main text, significantly better estimates are obtained with $d = 1, 2, 3$ circuits than $d = 0$ circuits.

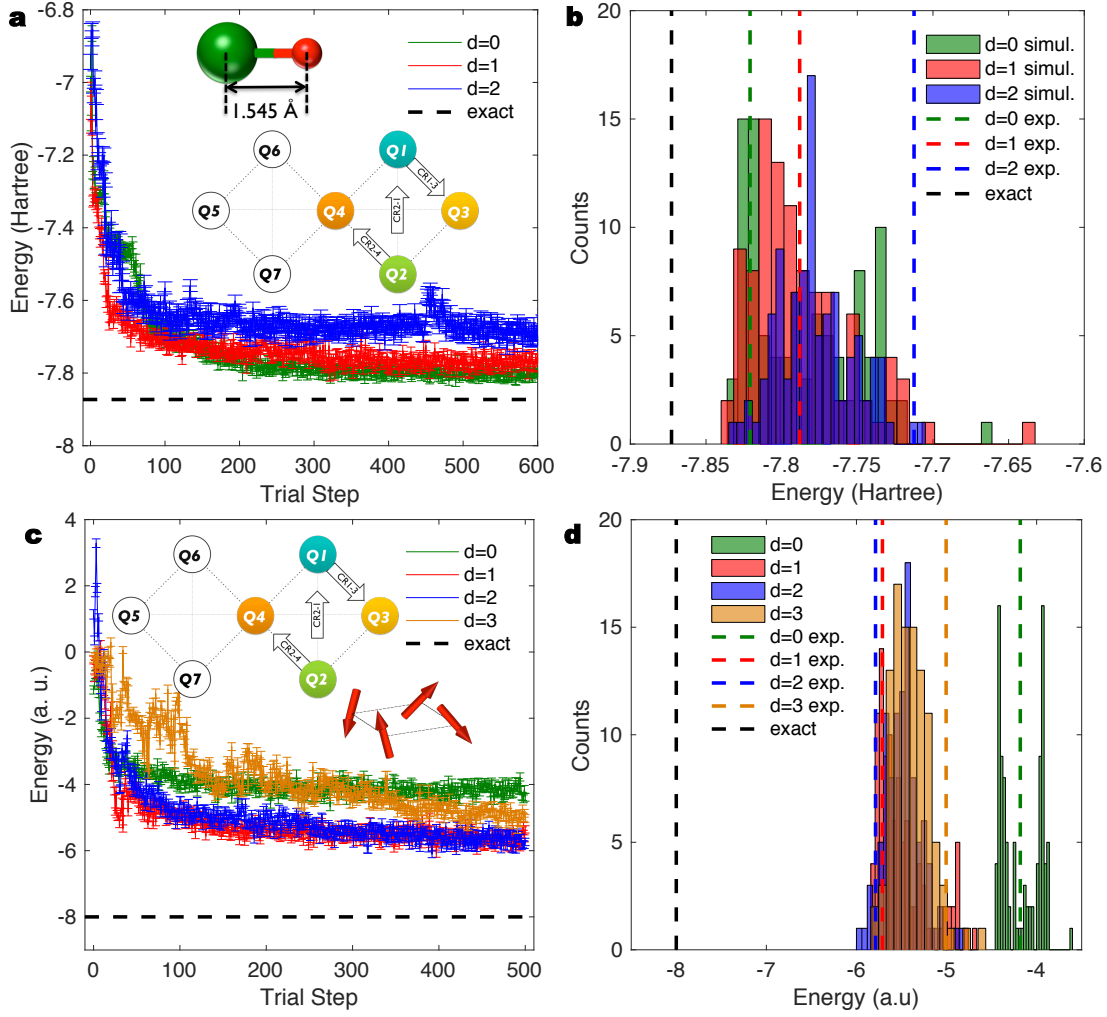


FIG. S9. **Experimental optimization for different depths: LiH Hamiltonian at bond distance and 4-qubit Heisenberg model** **a** Experimental optimization of the 4-qubit LiH Hamiltonian at bond distance, using depth $d = 0$ (green), 1 (red), 2 (blue) circuits for trial state preparation. The exact energy is indicated by the black dashed line. Bottom inset describes the qubits and the cross resonance gates that constitute U_{ENT} for this experiment. **b** Histograms of outcomes from 100 numerical simulations that account for decoherence and finite sampling effects show significant overlap for depth $d = 0$ (green), 1 (red), 2 (blue) circuits. The black dashed line indicates the exact energy and the green, red and blue dashed lines are the results from the single experimental runs of **a**, for $d = 0, 1$ and 2 circuits respectively. **c** Experimental optimization of the 4-qubit Heisenberg Hamiltonian for $J/B = 1$, using depth $d = 0$ (green), 1 (red), 2 (blue), 3 (orange) circuits for trial state preparation. The exact energy is indicated by the black dashed line. **d** Histograms of outcomes from 100 numerical simulations that account for decoherence and finite sampling effects show significant improvement over depth $d = 0$ circuits with $d = 1$ (red), 2 (blue), 3 (orange) circuits. The black dashed line indicates the exact energy and the green, red, blue and orange dashed lines are the results from the single experimental runs of **c**, for $d = 0, 1, 2$ and 3 circuits respectively.

TABLE S2: The H₂, LiH and BeH₂ Hamiltonians at the bond distance. Listed are all the Pauli operators, grouped in the different TPB sets, with the corresponding coefficients, not taking into account for the energy shifts due to the filling of inner orbitals and the Coulomb repulsion between nuclei. X,Y,Z,I here stand for the Pauli matrices σ^x , σ^y , σ^z and the identity operator on a single qubit subspace, respectively. There are 2,25,44 TPB sets for H₂, LiH and BeH₂, respectively with 4, 99 and 164 Pauli terms in total.

H₂ at bond distance

ZZ 0.011280 ZI 0.397936 IZ 0.397936	XX 0.180931
--	----------------

LiH at bond distance

ZIII -0.096022 ZZII -0.206128 IZII 0.364746 IIZI 0.096022 IIZZ -0.206128 IIIZ -0.364746 ZIZI -0.145438 ZIZZ 0.056040 ZIIZ 0.110811 ZZZI -0.056040 ZZZZ 0.080334 ZZIZ 0.063673 IZZI 0.110811 IZZZ -0.063673 IZIZ -0.095216	XZII -0.012585 XIII 0.012585 IIXZ 0.012585 IIXI 0.012585 XZXZ -0.002667 XZXI -0.002667 XIXZ 0.002667 XIXI 0.002667 XZIZ 0.007265 XIIZ -0.007265 IZXZ 0.007265 IZXI 0.007265	XXII -0.029640 IXII 0.002792 IIXX -0.029640 IIIX 0.002792 XIXX -0.008195 XIIH -0.001271 XXXI -0.008195 XXXX 0.028926 XXIX 0.007499 IXXI -0.001271 IXXX 0.007499 IXIX 0.009327	YYII 0.029640 IIYY 0.029640 YYYY 0.028926	ZXII 0.002792 IIZX -0.002792 ZIZX -0.016781 ZIIX 0.016781 ZXZI -0.016781 IXZI -0.016781 ZXZX -0.009327 ZXIX 0.009327 IXZX -0.009327	ZIXZ -0.011962 ZIXI -0.011962 ZZXZ 0.000247 ZZXI 0.000247	ZIXX 0.039155 ZZXX -0.002895 ZZIX -0.009769 IZXX -0.024280 IZIX -0.008025	ZIYY -0.039155 ZZYY 0.002895 IZYY 0.024280	XZZI -0.011962 XIZI 0.011962 XZZZ -0.000247 XIZZ 0.000247
XZXX 0.008195 XZIX 0.001271	XZYY -0.008195 XIYY 0.008195	XZZX -0.001271 XIZX 0.001271 IZZX 0.008025	XXZI -0.039155 XXZZ -0.002895 XXIZ 0.024280 IXZZ -0.009769 IXIZ 0.008025	YYZI 0.039155 YYZZ 0.002895 YYIZ -0.024280	XXXZ -0.008195 IXXZ -0.001271	YYXZ 0.008195 YYXI 0.008195	XXYY -0.028926 IXYY -0.007499	YYXX -0.028926 YYIX -0.007499

XXZX -0.007499	YYZX 0.007499	ZZZX 0.009769	ZXXZ -0.001271 ZXXI -0.001271 ZXIZ 0.008025	ZXXX 0.007499	ZXYY -0.007499	ZXZZ -0.009769		
-------------------	------------------	------------------	--	------------------	-------------------	-------------------	--	--

BeH₂ at bond distance

ZIIII -0.143021 ZZIII 0.104962 IZZII 0.038195 IIZII 0.038195 -0.325651 IIZII -0.143021 IIZZI 0.104962 IIZZZ 0.038195 IIIZZ -0.325651 IZIII 0.172191 ZZZII 0.174763 ZIZII 0.136055 ZIIIZ 0.116134 ZIIZI 0.094064 ZIIZZ 0.099152 ZIIHZ 0.123367	ZZIZII 0.094064 ZZIZZI 0.098003 ZZIIZ 0.102525 ZZIIZ 0.097795 IZZZI 0.099152 IZZZZI 0.102525 IZZZZ 0.112045 IZZIZ 0.105708 IIZZI 0.123367 IIZZI 0.097795 IIZIZ 0.105708 IIZIZ 0.133557 IIZZI 0.172191 IIZZZ 0.174763 IIZIZ 0.136055	XZIII 0.059110 XIIII -0.059110 IZXII 0.161019 IIXII -0.161019 IIXZI 0.059110 IIXII -0.059110 IIXZX 0.161019 IIXIX -0.161019 XIXII -0.038098 XZXII -0.003300 XZIXZI 0.013745 XZIXI -0.013745 XIXZI -0.013745 XIXII 0.013745	XZIIZX 0.011986 XZIIIX -0.011986 XIIHZ -0.011986 XIIIX 0.011986 IZXXZI 0.011986 IZXXII -0.011986 IIXXZI -0.011986 IIXXII 0.011986 IZXIZX 0.013836 IZXII -0.013836 IIXIZX -0.013836 IIXIX 0.013836 IIXIX -0.038098 IIXZX -0.003300	ZZXIII -0.002246 ZIXII 0.002246 ZIIIZ 0.014815 ZIIIZ -0.014815 ZIIIZ 0.009922 ZIIIX -0.009922 ZZIXZI -0.002038 ZZIXII 0.002038 ZZIIZX -0.007016 ZZIIX 0.007016	XIZIII -0.006154 XZZII 0.006154 XZIZI 0.014815 XIIIZ -0.014815 XZIZZI -0.002038 XIIIZ 0.002038 XZIIIZ 0.001124 XIIIZ -0.001124 XZIIIZ 0.017678 XIIIZ -0.017678	YIYIII -0.041398 YYIXXZ 0.011583 YIYIXI -0.011094 IYIXXZ -0.011094 IYIXXZ 0.010336 IYIYI -0.005725 IIXIZ -0.017678	XXZXXZ 0.011583 XXZIXI -0.011094 IXIXXZ -0.011094 IXIXI 0.026631 IIZXII -0.017678	XXZYIYI 0.011583 XXZIYY 0.010336 IXIYYI -0.011094 IXIYY -0.005725 IHYIY -0.041398
YIYIYI 0.011583 YIYIYY 0.010336 IYIYIY 0.010336 IYIYIY 0.010600	XXZXXX 0.024909 IXIXXX -0.031035 IIZIIX -0.010064	XXZYXY 0.024909 IXIYXY -0.031035	YYIXXX 0.024909 IYYXXX 0.021494	YIYIXY 0.024909 IYIYXY 0.021494	XXZZXZ 0.011094 IXIZXZ -0.026631	YYIZXZ 0.011094 IYIYZX 0.005725	XXZZXX 0.010336 IXIZXX -0.005725 IIZIX 0.002246	YYIZXX 0.010336 IYIYZX 0.010600
XXXXXZ 0.024909 XXXIXI -0.031035 IIXIIZ -0.010064	XXXYIYI 0.024909 XXXIYY 0.021494	YXYXXXZ 0.024909 YXYIXI -0.031035	YXYIYI 0.024909 YXYIYY 0.021494	XXXXXX 0.063207	XXXYXY 0.063207	YXYXXX 0.063207	YXYIYI 0.063207	XXXXXZ 0.031035 IIXIIZ -0.009922
YXYZXX 0.031035	XXXZXX 0.021494	YXYZXX 0.021494	ZXZXXZ 0.011094 ZXZIXI -0.026631	ZXZYIY 0.011094 ZXZIYY 0.005725	ZXZXXX 0.031035	ZXZYXY 0.031035	ZXZZXZ 0.026631	ZXZZXX 0.005725

ZXXXXXZ 0.010336 ZXXIXI -0.005725	ZXXYYI 0.010336 ZXXIYY 0.010600	ZXXXXXX 0.021494	ZXXYXY 0.021494	ZXXZXZ 0.005725	ZXXZXX 0.010600	IZZXXI 0.001124 IZZXXI -0.001124 IZZIZX -0.007952 IZZIX 0.007952 IIXZI 0.017678 IIZIX 0.010064	IZXZII 0.009922 IZXZZI -0.007016 IIXZZI 0.007016 IZXIZZ -0.007952 IIXIZZ 0.007952 IZXIIZ 0.010064	IIIZZ -0.002246
IIIXZZ 0.006154								

-
- [1] Hutchings, M. *et al.* Tunable superconducting qubits with flux-independent coherence. *arXiv preprint arXiv:1702.02253* (2017).
- [2] Chow, J. M. *et al.* Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nat. Commun.* **5**:4015 doi: 10.1038/ncomms5015 (2014).
- [3] Córcoles, A. D. *et al.* Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.* **6**:6979 doi: 10.1038/ncomms7979 (2015).
- [4] Bergeal, N. *et al.* Analog information processing at the quantum limit with a josephson ring modulator. *Nat. Phys.* **6**, 296–302 (2010).
- [5] Abdo, B., Schackert, F., Hatridge, M., Rigetti, C. & Devoret, M. Josephson amplifier for qubit readout. *Appl. Phys. Lett.* **99**, 162506 (2011).
- [6] Bravyi, S., Gambetta, J. M., Mezzacapo, A. & Temme, K. Tapering off qubits to simulate fermionic hamiltonians. *arXiv preprint arXiv:1701.08213* (2017).
- [7] Muller, R. P. Python quantum chemistry, version 1.6.0. <http://pyquante.sourceforge.net/>.
- [8] Szabo, A. & Ostlund, N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Corporation, 1989).
- [9] Bravyi, S. & Kitaev, A. Fermionic quantum computation. *Ann. Phys.* **298**, 210–226 (2002).
- [10] Paraoanu, G. S. Microwave-induced coupling of superconducting qubits. *Phys. Rev. B* **74**, 140504 (2006).
- [11] 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).
- [12] Chow, J. M. *et al.* Simple all-microwave entangling gate for fixed-frequency superconducting qubits. *Phys. Rev. Lett.* **107**, 080502 (2011).
- [13] 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).
- [14] McClean, J., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 023023 (2016).
- [15] Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control* **37**, 332 (1992).
- [16] Ferrie, C. Self-guided quantum tomography. *Phys. Rev. Lett.* **113**, 190404 (2014).
- [17] Ferrie, C. & Combes, J. Robust and efficient in situ quantum control. *Phys. Rev. A* **91**, 052306 (2015).
- [18] Chapman, R., Ferrie, C. & Peruzzo, A. Experimental demonstration of self-guided quantum tomography. *Phys. Rev. Lett.* **117**, 040402 (2016).
- [19] Wecker, D., Hastings, M. B. & Troyer, M. Progress towards practical quantum variational algorithms. *Phys. Rev. A* **92**, 042303 (2015).
- [20] Spall, J. C. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Trans. Aerosp. and Electron. Syst.* **34**, 817–823 (1998).
- [21] Risté, D., Bultink, C. C., Lenhart, K. W. & DiCarlo, L. Feedback control of a solid-state qubit using high-fidelity projective measurement. *Phys. Rev. Lett.* **109**, 240502 (2012).
- [22] McClure, D. T., Paik, H., Bishop, L. S., Chow, J. M. & Gambetta, J. M. Rapid driven reset of a qubit readout resonator. *Phys. Rev. Applied* **5**, 011001 (2016).
- [23] Bultink, C. C. *et al.* Active resonator reset in the nonlinear dispersive regime of circuit qed. *Phys. Rev. Applied* **6**, 034008 (2016).

- [24] McClean, J. R., 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).
- [25] Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimisation. *arXiv 1611.09301* (2016).
- [26] Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short depth quantum circuits. *arXiv preprint arXiv:1612.02058* (2016).