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 $\omega_i$, while the entangling two-qubit CR gates are implemented by driving a control qubit $Q_c$ at the frequency $\omega_t$ 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 $\omega_{M_i}$. The reflected readout signals are amplified first by a JPC, pumped at a frequency $\omega_{P_1}$, 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 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 ($\epsilon_r$) for the six qubits discussed in the paper.

<table>
<thead>
<tr>
<th>Qubit</th>
<th>$\omega_{01}/2\pi$ (GHz)</th>
<th>$T_1$ ($\mu$s)</th>
<th>$T_2$ ($\mu$s)</th>
<th>$T_2^*$ ($\mu$s)</th>
<th>$\omega_r/2\pi$ (GHz)</th>
<th>$\delta/2\pi$ (GHz)</th>
<th>$\epsilon_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>5.3206</td>
<td>24.7 ± 3.2</td>
<td>31.1 ± 6.1</td>
<td>22.2 ± 4.8</td>
<td>6.6223</td>
<td>-0.311</td>
<td>0.0240</td>
</tr>
<tr>
<td>Q2</td>
<td>5.3567</td>
<td>42.0 ± 5.1</td>
<td>38.7 ± 12.5</td>
<td>28.6 ± 1.2</td>
<td>6.6892</td>
<td>-0.312</td>
<td>0.0544</td>
</tr>
<tr>
<td>Q3</td>
<td>5.2926</td>
<td>20.4 ± 4.4</td>
<td>35.3 ± 8.7</td>
<td>6.2 ± 0.9</td>
<td>6.5589</td>
<td>-0.315</td>
<td>0.0291</td>
</tr>
<tr>
<td>Q4</td>
<td>5.2455</td>
<td>42.3 ± 5.2</td>
<td>47.4 ± 14.0</td>
<td>36.7 ± 10.5</td>
<td>6.7154</td>
<td>-0.299</td>
<td>0.0469</td>
</tr>
<tr>
<td>Q5</td>
<td>5.2999</td>
<td>44.4 ± 4.9</td>
<td>60.5 ± 8.7</td>
<td>40.0 ± 3.2</td>
<td>6.6532</td>
<td>-0.311</td>
<td>0.0278</td>
</tr>
<tr>
<td>Q6</td>
<td>5.3882</td>
<td>20.6 ± 0.8</td>
<td>26.4 ± 4.3</td>
<td>27 ± 2.8</td>
<td>6.5885</td>
<td>-0.310</td>
<td>0.0507</td>
</tr>
</tbody>
</table>

**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 1.5 $\mu$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 $\sim 310$ MHz, while the asymmetric transmon has an anharmonicity of $\sim 300$ 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 $X_\pi$ 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 $X_{\pi}$ 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 $X_{\pi}$-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 $\theta_1$ 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 $\theta_k$ and evaluate $\langle H \rangle$ with $S$ samples
10:        Update and store the controls $\theta_k$
11:     end for
12:     Evaluate $E_f \equiv \langle H \rangle$ 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,

\[
H = H_1 + H_2 = \sum_{\alpha,\beta=1}^{M} t_{\alpha\beta} a^\dagger_{\alpha} a_{\beta} + \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta=1}^{M} u_{\alpha\beta\gamma\delta} a^\dagger_{\alpha} a^\dagger_{\gamma} a_{\delta} a_{\beta},
\]

where \(a^\dagger_{\alpha}(a_{\alpha})\) is the fermionic creation (annihilation) operator of the fermionic mode \(\alpha\), satisfying fermionic commutation rules \(\{a_{\alpha}, a_{\beta}\} = 0, \{a^\dagger_{\alpha}, a^\dagger_{\beta}\} = 0, \{a_{\alpha}, a^\dagger_{\beta}\} = \delta_{\alpha\beta}\). 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,

\[
l_{\alpha\beta} = \int d\mathbf{x}_1 \Psi_{\alpha}(\mathbf{x}_1) \left( -\frac{\nabla^2}{2} + \sum_i Z_i \right) \Psi_{\beta}(\mathbf{x}_1),
\]

\[
u_{\alpha\beta\gamma\delta} = \int \int d\mathbf{x}_1 d\mathbf{x}_2 \Psi^\dagger_{\alpha}(\mathbf{x}_1) \Psi_{\beta}(\mathbf{x}_1) \frac{1}{|\mathbf{r}_{12}|} \Psi^\dagger_{\gamma}(\mathbf{x}_2) \Psi_{\delta}(\mathbf{x}_2),
\]

where we have defined the nuclei charges \(Z_i\), the nuclei-electron and electron-electron separations \(\mathbf{r}_{11}\) and \(\mathbf{r}_{12}\), the \(\alpha\)-th orbital wavefunction \(\Psi_{\alpha}(x_1)\), and we have assumed that the spin is conserved in the spin-orbital indices \(\alpha, \beta\) 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

\[
H^d_1 = U^\dagger H_1 U, \quad H^d_1 = \sum_{\alpha=1}^{M} \omega'_{\alpha} a'^\dagger_{\alpha} a'_{\alpha},
\]

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 \(|\Psi\rangle = a_{1s}^\dagger a_{1s}^\dagger \left( \sum_{\beta \neq 1s} \psi_{\beta} a_{\beta}^\dagger \right) |0\rangle\), where \(\psi_{\beta}\) 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 \(-\omega'_{1s\sigma} \gg |u'_{\alpha\beta\gamma\delta}| \forall \alpha, \sigma, \alpha, \beta, \gamma, \delta, \) i.e. in the case of very low-energy orbitals that do not interact strongly with the higher-energy ones. The ansatz \(|\Psi\rangle = a_{1s}^\dagger a_{1s}^\dagger \left( \sum_{\beta \neq 1s} \psi_{\beta} a_{\beta}^\dagger \right) |0\rangle\) allows to define an effective screened Hamiltonian on the 1s orbitals for the hydrogen atoms, and 2s and 2p\(_x\), for Lithium and Berilium, 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),

\[
\omega'_{1s} a_{1s}^\dagger a_{1s} \rightarrow \omega'_{1s} I, \quad \omega'_{1s} a_{1s}^\dagger a_{1s} \rightarrow \omega'_{1s} I,
\]

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,

\[
\frac{u'_{\alpha\beta\gamma\delta}}{2} a_{\alpha}^\dagger a_{\beta}^\dagger a_{\gamma} a_{\delta} \rightarrow \begin{cases} 
\omega_{\alpha\beta} a_{\alpha}^\dagger a_{\beta}, & \alpha = \beta, \alpha \in F, \{\gamma, \delta\} \notin F \\
\frac{u_{\alpha\beta}}{2} a_{\alpha}^\dagger a_{\beta}, & \gamma = \delta, \gamma \in F, \{\alpha, \beta\} \notin F \\
\frac{u_{\alpha\beta}}{2} a_{\alpha}^\dagger a_{\beta}, & \alpha = \delta, \alpha \in F, \{\beta, \gamma\} \notin F \\
\frac{u_{\alpha\beta}}{2} a_{\alpha}^\dagger a_{\beta}, & \gamma = \beta, \gamma \in F, \{\alpha, \delta\} \notin F \\
\frac{u_{\alpha\beta}}{2} I, & \alpha = \delta, \gamma = \delta, \alpha \neq \gamma, \{\alpha, \gamma\} \in F \\
\frac{u_{\alpha\beta}}{2} I, & \alpha = \delta, \gamma = \beta, \alpha \neq \gamma, \{\alpha, \gamma\} \in F \\
\frac{u_{\alpha\beta}}{2} I, & \alpha = \delta, \gamma = \beta, \alpha \neq \gamma, \{\alpha, \gamma\} \in F \\
\end{cases}
\]

while the two-body operators containing an odd number of modes in \(F\) will be neglected. We then map the fermionic Hamiltonians \(H = \sum_{\alpha,\beta \neq 1s} t_{\alpha\beta} a^\dagger_{\alpha} a_{\beta} + 1/2 \sum_{\alpha,\beta,\gamma,\delta \neq 1s} u_{\alpha\beta\gamma\delta} a^\dagger_{\alpha} a^\dagger_{\gamma} a_{\delta} a_{\beta}\) 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 \( Z_2 \) symmetries encoded in those qubits \[6\]. For the LiH and BeH\(_2\) Hamiltonians we use the parity mapping, which has the two \( Z_2 \) 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 \( H_2 \). 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

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

The \(+1(-1)\) on \( Z_M \) for even(odd) \( m \) implies an even(odd) total electron parity. The values \(+1\), \(-1\) and \(\pm1\) 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 \( H_2 \), LiH and BeH\(_2\), 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 \( U_{\text{ENT}} \) 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 \( \sim 160 \) 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

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

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, \( \Delta \) is the frequency separation between the qubits, and \( \mu \) 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 CR\(_{2-4}\) 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 \( ||R|| \) discussed in \[13\], whose time evolution indicates points of maximal entanglement at \( ||R|| = 0 \); 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 (\( \approx 0.0016 \) 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 \( U_{\text{ENT}} \). For these simulations, we choose \( ZX \) gates for \( U_{\text{ENT}} \), 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 \( 5 \times 10^3 \). 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

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

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

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 $h_{\alpha}$ being real coefficients. We are interested in estimating the mean energy $\langle \Phi(\theta_{k}) \vert H \vert \Phi(\theta_{k}) \rangle = \langle H \rangle_{k}$ for the k-th control updates (more specifically for two sets of angles close to $\theta_{k}$, 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 $\theta_{k}$, 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 $P_{\alpha}$, estimating the mean values and

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

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

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 $h_{\alpha}$ being real coefficients. We are interested in estimating the mean energy $\langle \Phi(\theta_{k}) \vert H \vert \Phi(\theta_{k}) \rangle = \langle H \rangle_{k}$ for the k-th control updates (more specifically for two sets of angles close to $\theta_{k}$, 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 $\theta_{k}$, 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 $P_{\alpha}$, estimating the mean values and
variances \( \langle P_\alpha \rangle \), \( \langle \Delta P_\alpha^2 \rangle = \langle P_\alpha^2 \rangle - \langle P_\alpha \rangle^2 \) from the measurements outcomes of the \( \alpha \)-th Pauli operator. The energy and Hamiltonian variance can then be obtained as

\[
\langle H \rangle = \sum_{\alpha=1}^{T} h_\alpha \langle P_\alpha \rangle, \tag{10}
\]

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

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

\[
\epsilon = \sqrt{\frac{\text{Var}[H]}{S}} \leq \sqrt{\frac{T h_{\text{max}}^2}{S}} \tag{12}
\]

where \( h_{\text{max}} = \max_\alpha |h_\alpha| \) 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 fist 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 \( P_\alpha \) in \( A \) sets \( s_1, s_2, \ldots, s_A \), 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,

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

where \( s_{\text{max}} = \max_i |s_i| \) 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

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

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

\[
\langle \hat{P}_\alpha \rangle = \frac{1}{S} \sum_{i=1}^{S} X_{i,\alpha}, \tag{15}
\]

\[
\text{Var}^G[H] = \sum_{i=1}^{A} \sum_{\alpha, \beta \in s_i} h_\alpha h_\beta \text{cov}(\langle \hat{P}_\alpha \rangle, \langle \hat{P}_\beta \rangle), \tag{16}
\]

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

\[
\text{cov}(\langle \hat{P}_\alpha \rangle, \langle \hat{P}_\beta \rangle) = \frac{1}{S-1} \sum_{i=1}^{S} (X_{i,\alpha} - \langle \hat{P}_\alpha \rangle)(X_{i,\beta} - \langle \hat{P}_\beta \rangle). \tag{17}
\]
FIG. S5.  **Energy variance** Numerical computation of the variance of the mean energy $\epsilon^2$, as in Eq. (16), with $S = 10^3$ samples, for the molecular Hamiltonians of $\text{H}_2$ (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 $P_\alpha$ 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 $\Pi_0, \Pi_1$ on the $|0\rangle, |1\rangle$ states for the qubit,

$$\hat{\Pi}_0 = (1 - \eta_0 + \eta_1)\Pi_0 + (1 - \eta_0 - \eta_1)\Pi_1 = (1 - \eta_0)I + \eta_1 Z$$

$$\hat{\Pi}_1 = (\eta_0 - \eta_1)\Pi_0 + (\eta_0 + \eta_1)\Pi_1 = \eta_0 I - \eta_1 Z,$$

via the two parameters $\eta_0, \eta_1$ (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 \( g_k(\theta_k) \) are obtained if the energy difference \( |\langle \Phi(\theta^+_k) | H | \Phi(\theta^+_k) \rangle - \langle \Phi(\theta^-_k) | H | \Phi(\theta^-_k) \rangle| \) is larger than the stochastic fluctuations on the energy \( \epsilon_k \). 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 \( E(\theta^+_k) = \langle \Phi(\theta^+_k) | H | \Phi(\theta^+_k) \rangle \), measured here for the LiH molecule at the bond distances, from the starting angles \( \theta_1 \), for different random gradients approximations. 

c The energy difference \( \Delta E = |\langle \Phi(\theta^+_k) | H | \Phi(\theta^+_k) \rangle - \langle \Phi(\theta^-_k) | H | \Phi(\theta^-_k) \rangle| \) 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

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

(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

\[
Z^{\otimes w} \propto \frac{\hat{Z}^{\otimes w}}{(2\eta_1)^w},
\]  

(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 \( \theta \), and then correcting our sampling outcome with the calibrations.

VI. OPTIMIZATION USING A SIMULTANEOUS PERTURBATION METHOD

The energy \( \langle \Phi(\theta_k) | H | \Phi(\theta_k) \rangle \equiv \langle H \rangle_k \) discussed in Section V, which needs to be evaluated before every update of the angles \( \theta \), 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 \( \epsilon_k \). 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 \( O(p) \) 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) $\Delta_k$, and use preassigned elements from two sequences converging to zero, $c_k$ and $a_k$. The gradient at $\theta_k$ is approximated using energy evaluations at $\theta_k^\pm = \theta_k \pm c_k \Delta_k$, and is constructed as

$$g_k(\theta_k) = \frac{\langle \Phi(\theta_k^\pm)| H|\Phi(\theta_k^\pm) \rangle - \langle \Phi(\theta_k^-)| H|\Phi(\theta_k^-) \rangle}{2c_k},$$

(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 $\theta$. The controls are then updated as

$$\theta_{k+1} = \theta_k - a_k g_k(\theta_k).$$

(22)

The convergence of $\theta_k$ to the optimal solution $\theta^*$ 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

$$c_k = \frac{c}{k^\gamma},$$

$$a_k = \frac{a}{k^\alpha}.$$

(23)

We pick the parameters $\alpha, \gamma$ 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

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

(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 $i = 1, 2, ..., p$. To achieve this, we use an inverse formula based on Eq. (22),

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

(25)

where the notation $\langle \rangle_{\Delta_1}$ indicates an average over different samples from the distribution $\Delta_1$ 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 $\theta_1$, 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. S6c.

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 $\theta_k$ are only measured at the end of the optimization, averaging over the last 25 $\theta_k^+$ and 25 $\theta_k^-$, 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 $\theta_k^+$ and $\theta_k^-$ 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.
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 $N(0, 1)$, and the $X$ angles set to $\pi/2$.

The effect of decoherence is taken into account by adding amplitude damping $(E^0_0(\tau), E^0_1(\tau))$ and dephasing $(E^d_0(\tau), E^d_1(\tau))$ channels acting on the system density matrix $\rho \rightarrow E^0_0(\tau)\rho E^0_0(\tau) + E^0_1(\tau)\rho E^0_1(\tau), \rho \rightarrow E^d_0(\tau)\rho E^d_0(\tau) + E^d_1(\tau)\rho E^d_1(\tau)$, 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,

$$E^0_0(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{e^{-\tau/T_1}} \end{bmatrix}, E^0_1(\tau) = \begin{bmatrix} 0 & \sqrt{1 - e^{-\tau/T_1}} \\ 0 & 0 \end{bmatrix}$$

$$E^d_0(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & e^{-\tau/T_\alpha} \end{bmatrix}, E^d_1(\tau) = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{1 - e^{-2\tau/T_\alpha}} \end{bmatrix}.$$  \hfill (26)

Here the time $\tau$ alternates between the duration of each single qubit gate sequence or entangler step, and the pure dephasing time is defined as $T_\alpha = 2T_2^ZT_1^I/(2T_1^I - T_2^Z)$, 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 parameterize the noise channels considering $T_1 = T_2^Z = 40 \mu s$ and set the length of $U_{\text{ENT}}$ to 150 ns, while for the 4 and 6-qubit simulations we use typical coherence values for the qubits of $T_1 = 30 \mu s$, $T_2^Z = 20 \mu s$ and a duration for $U_{\text{ENT}}$ 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 \( \text{H}_2 \), \( \text{LiH} \) and \( \text{BeH}_2 \) (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 \( U_{\text{ENT}} \) 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 \( 5 \times 10^4 \) (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 \( \text{H}_2, \text{LiH} \) and \( \text{BeH}_2 \), 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 \( 2 \times 10^3 \) function calls (\( 10^3 \) angle updates) are sufficient for reaching chemical accuracy on \( \text{H}_2 \), \( 2 \times 10^4 \) for \( \text{LiH} \) both for the all-connected and experiment connectivity, \( 2 \times 10^4 \) for \( \text{BeH}_2 \) in the all-connected case and approximately \( 3 \times 10^4 \) 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 \( \epsilon_A \) 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 \( \epsilon_A \to \epsilon_A \sqrt{10^3/S} \). 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 \( \rho \) as

\[
\rho \to (1 - \xi)\rho + \frac{\xi}{3} \sum_{i=1,2,3} \sigma^i \rho \sigma^i,
\]

\[
\rho \to (1 - \xi)\rho + \frac{\xi}{15} \sum_{\{(i,j)\neq\{0,1,2,3\}, \{i,j\}\neq\{0,0\}}} \sigma^i_j \sigma^m_i \rho \sigma^m_j \sigma^i_j,
\]

(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 \( \text{LiH} \) Hamiltonian at the bond distance, for different number of entanglers and noise strengths, for a maximum of \( 5 \times 10^4 \) 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 \( \xi \), higher depths give better results, while as \( \epsilon \) 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 $5 \times 10^4$ 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 $U_{\text{ENT}}$, 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 $\sigma_x$, $\sigma_y$, $\sigma_z$ 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

<table>
<thead>
<tr>
<th></th>
<th>ZZ</th>
<th>ZI</th>
<th>IZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.011280</td>
<td>0.397936</td>
<td>0.397936</td>
<td></td>
</tr>
</tbody>
</table>

### LiH at bond distance

<table>
<thead>
<tr>
<th></th>
<th>ZIII</th>
<th>ZIZ</th>
<th>ZII</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.096022</td>
<td>0.364746</td>
<td>0.096022</td>
<td></td>
</tr>
<tr>
<td>-0.206128</td>
<td>0.145438</td>
<td>0.180931</td>
<td></td>
</tr>
</tbody>
</table>

### BeH₂ at bond distance

<table>
<thead>
<tr>
<th></th>
<th>XX</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.180931</td>
<td></td>
</tr>
</tbody>
</table>
BeH₂ at bond distance

<table>
<thead>
<tr>
<th>XXZZ</th>
<th>YYZX</th>
<th>ZZZX</th>
<th>ZXXZ</th>
<th>ZXXX</th>
<th>ZXYY</th>
<th>ZXZZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.007499</td>
<td>0.007499</td>
<td>0.009769</td>
<td>-0.001271</td>
<td>0.0011986</td>
<td>0.0008025</td>
<td>-0.007499</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ZIIIII</th>
<th>ZZZIZI</th>
<th>XZIII</th>
<th>XZIIIZX</th>
<th>ZIIIIX</th>
<th>ZXXIII</th>
<th>ZIIXIZX</th>
<th>YIYIXI</th>
<th>XZIXXZ</th>
<th>XXYXY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.143021</td>
<td>0.104962</td>
<td>0.059110</td>
<td>0.011986</td>
<td>0.161019</td>
<td>-0.002246</td>
<td>0.002246</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
</tr>
<tr>
<td>0.011094</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
<td>0.0011948</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ZZIIIX</th>
<th>ZIIIZI</th>
<th>IIXIX</th>
<th>IIXIXIX</th>
<th>ZIIIIX</th>
<th>ZXXIII</th>
<th>ZIIXIZX</th>
<th>YIYIXI</th>
<th>XZIXXZ</th>
<th>XXYXY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.123367</td>
<td>0.136055</td>
<td>0.136055</td>
<td>-0.003006</td>
<td>0.013745</td>
<td>0.002030</td>
<td>0.001124</td>
<td>0.000675</td>
<td>0.000675</td>
<td>0.000675</td>
</tr>
<tr>
<td>0.136055</td>
<td>0.136055</td>
<td>0.136055</td>
<td>-0.003006</td>
<td>0.013745</td>
<td>0.002030</td>
<td>0.001124</td>
<td>0.000675</td>
<td>0.000675</td>
<td>0.000675</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XIXX</th>
<th>XIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
<th>XIXIXIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
<td>0.011094</td>
</tr>
</tbody>
</table>

WWW.NATURE.COM/NATURE | 36
