

## Doubling the Size of Quantum Simulators by Entanglement Forging

Andrew Eddins<sup>1,\*</sup>, Mario Motta,<sup>1</sup> Tanvi P. Gujarati,<sup>1</sup> Sergey Bravyi,<sup>2</sup> Antonio Mezzacapo,<sup>2</sup>  
Charles Hadfield,<sup>2</sup> and Sarah Sheldon<sup>1</sup>

<sup>1</sup>IBM Quantum, Almaden Research Center, San Jose, California 95120, USA

<sup>2</sup>IBM Quantum, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA



(Received 30 August 2021; accepted 17 December 2021; published 14 January 2022)

Quantum computers are promising for simulations of chemical and physical systems, but the limited capabilities of today's quantum processors permit only small, and often approximate, simulations. Here we present a method, classical entanglement forging, that harnesses classical resources to capture quantum correlations and double the size of the system that can be simulated on quantum hardware. Shifting some of the computation to classical postprocessing allows us to represent ten spin orbitals of the water molecule on five qubits of an IBM Quantum processor in the most accurate variational simulation of the H<sub>2</sub>O ground-state energy using quantum hardware to date. We discuss conditions for applicability of classical entanglement forging and present a roadmap for scaling to larger problems.

DOI: [10.1103/PRXQuantum.3.010309](https://doi.org/10.1103/PRXQuantum.3.010309)

Simulating quantum systems is an especially hard task for classical computers, making the realization of quantum computers potentially revolutionary for the study of chemistry, materials science, and fundamental physics. However, techniques like quantum phase estimation, which promises accurate chemical simulations, require hardware well beyond the present state of the art. While hardware capabilities continue to steadily advance, limitations on both quantity and quality of qubits are giving rise to new algorithms that leverage additional classical resources to enable quantum computations requiring more qubits than physically available [1–13]. Such algorithms commonly benefit from a strategy of partitioning a problem into weakly interacting clusters, then correlating the results of each on a classical computer. A range of important systems naturally possess suitable partitions, including low-energy eigenstates of chemical and lattice-model Hamiltonians [14–17], systems embedded in a quantum bath [18,19], and static correlations associated with chemical bond-breaking processes [20–22]. Here we introduce a scheme, *classically forged entanglement*, which represents a  $2N$ -qubit wave function as multiple  $N$ -qubit states embedded in a classical computation. Besides halving the requisite qubit number, offloading entanglement synthesis to classical

processing permits the constituent  $N$ -qubit quantum circuits to be shallower, relaxing requirements on error rates and connectivity, at the cost of an increased number of circuit executions.

As a general purpose quantum-classical representation of a state, entanglement forging may be utilized in a variety of algorithms including fault-tolerant methods like quantum phase estimation. We theoretically describe the scheme, then demonstrate it with a variational quantum eigensolver [23] (VQE) simulation in which five qubits capture the behavior of ten spin orbitals of the water molecule [24].

Theoretical description of classically forged entanglement begins with Schmidt decomposition [Fig. 1(a)], a standard application of singular value decomposition (SVD) that allows one to write any state  $|\psi\rangle$  of a bipartite  $N + N$  qubit system as

$$|\psi\rangle = (U \otimes V) \sum_{n=1}^{2^N} \lambda_n |b_n\rangle \otimes |b_n\rangle. \quad (1)$$

Here  $|b_n\rangle$  are the  $N$ -qubit bit-string states, also known as computational basis states,  $U, V$  are unitary operators respective to the two subsystems, and the Schmidt coefficients  $\lambda_n$  may be taken to be non-negative. The flatter the distribution of Schmidt coefficients, the stronger the entanglement; a uniform distribution  $\lambda_n = 1/\sqrt{2^N}$  indicates the two halves of the system are maximally entangled, while only one nonzero coefficient corresponds to no entanglement.

Mixing classical and quantum information, entanglement forging is more naturally expressed using density

\*aeddins@ibm.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

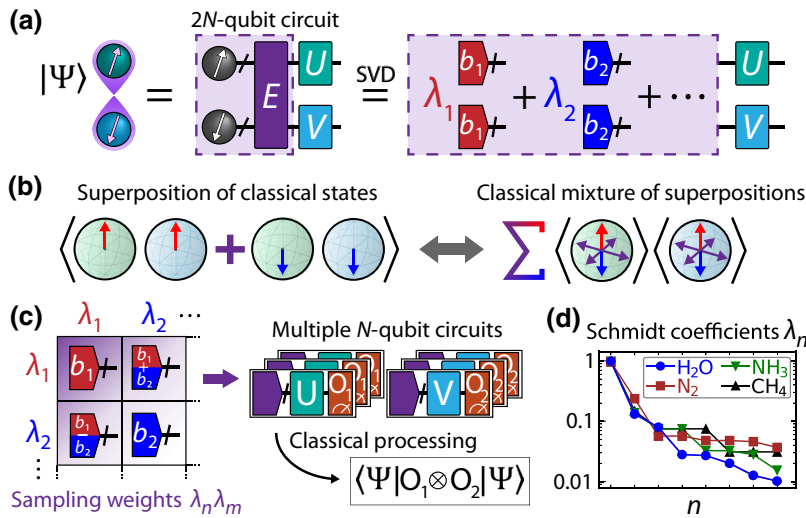


FIG. 1. Classically forged entanglement. (a) A state  $|\Psi\rangle$  of a bipartite quantum system, labeled with arrows alluding to spin polarization, can be defined by gates  $E$ ,  $U$ , and  $V$ , where  $E$  outputs a combination of bit-string states  $|b_n\rangle|b_n\rangle$ . (b) A two-qubit entangled state can be rewritten using one-qubit superposition states. Changing labels  $0, 1 \rightarrow b_n, b_m$  gives a transformation acting on components of the  $2N$ -qubit state. (c)  $|\Psi\rangle$  can be reconstructed from  $N$ -qubit circuits initialized as bit strings and pairwise superpositions thereof. Circuits associated with small  $\lambda_n \lambda_m$  can be estimated adequately from few samples. (d) Rapid (slow) decay of the leading Schmidt coefficients in the decomposition of a molecular ground state signals weak (strong) entanglement between spin-up and spin-down particles.

operators rather than wave functions, even for pure states such as  $|\psi\rangle$ . As shown in Appendix A, one can write the density operator as

$$|\psi\rangle\langle\psi| = (U \otimes V) \sum_{n=1}^{2^N} \left( \lambda_n^2 |b_n\rangle\langle b_n|^{\otimes 2} + \sum_{m=1}^{n-1} \lambda_n \lambda_m \right. \\ \left. \times \sum_{p \in \mathbb{Z}_4} (-1)^p |\phi_{b_n b_m}^p\rangle\langle\phi_{b_n b_m}^p|^{\otimes 2} \right) (U^\dagger \otimes V^\dagger), \quad (2)$$

where we use the definition  $|\phi_{xy}^p\rangle = (|x\rangle + i^p |y\rangle)/\sqrt{2}$  with  $p \in \{0, 1, 2, 3\} = \mathbb{Z}_4$ . For example, in the minimal case of two qubits [Fig. 1(b)],  $|\phi_{01}^p\rangle$  correspond to four equatorial points on the Bloch sphere, rewriting a quantum superposition of product states in terms of classical products of superposition states. Equation (2) generalizes methods proposed in Refs. [1,25,26], and is connected to tensor-network representations of quantum circuits [8,27], variational simulation of open quantum systems [28], and the encoding of open-shell singlet and triplet states [29]. The expectation of a  $2N$ -qubit operator  $O = O_1 \otimes O_2$  is now

$$\langle O \rangle = \sum_{n=1}^{2^N} \left( \lambda_n^2 \langle b_n | \tilde{O}_1 | b_n \rangle \langle b_n | \tilde{O}_2 | b_n \rangle + \sum_{m=1}^{n-1} \lambda_n \lambda_m \right. \\ \left. \times \sum_{p \in \mathbb{Z}_4} (-1)^p \langle \phi_{b_n b_m}^p | \tilde{O}_1 | \phi_{b_n b_m}^p \rangle \langle \phi_{b_n b_m}^p | \tilde{O}_2 | \phi_{b_n b_m}^p \rangle \right), \quad (3)$$

where  $\tilde{O}_1 = U^\dagger O_1 U$  and  $\tilde{O}_2 = V^\dagger O_2 V$ , and each constituent requires only  $N$  qubits to evaluate.

The resulting summation for  $\langle O \rangle$  is not obviously scalable, involving as many as  $2^{N+1}(2^{N+1} - 1)$  distinct  $N$ -qubit quantum circuits. Nonetheless, if one restricts to

simulations of sufficiently weak entanglement,  $\langle O \rangle$  can be efficiently estimated, as shown in Appendix E, by sampling each circuit in proportion to the associated coefficients  $\lambda_n \lambda_m$  in Eq. (3), with a total number of samples  $S$  for target precision  $\epsilon$  scaling as

$$S \sim \left( \frac{1}{\epsilon} \sum_{n,m} |\lambda_n \lambda_m| \right)^2 = \frac{\|\vec{\lambda}\|_1^4}{\epsilon^2}, \quad \|\vec{\lambda}\|_1 = \sum_n |\lambda_n|. \quad (4)$$

Executing a quantum circuit once provides one sample of the corresponding expectation value, such that total runtime scales linearly with  $S$ . Since the one-norm decreases toward 1 in the limit of weak entanglement, the overhead cost of entanglement forging is smaller for simulations of states divisible into weakly entangled halves, such as the spin-up and spin-down components of some molecular ground states [Fig. 1(d)], and scales efficiently when the one-norm is at most polynomial in the problem size. For example, in some statically correlated ground states,  $S$  can be independent of the number of basis orbitals. Outside of the domain of scalability, entanglement forging still enables useful heuristic simulations beyond the standard capacity of given quantum hardware, which may be realized with precision by truncating the list of bit-string states retained in the Schmidt decomposition.

Alternatively, this overhead may be reduced to a constant factor independent of qubit number via a complementary scheme (Appendix B) simulating quantum correlations between subsystems using those within a subsystem, rather than using classical correlations as above. This method can be seen as an application of forging in the Heisenberg picture, reinterpreting an observable acting on  $N + N$  qubits as a classical mixture of operators describing the forward and backward time evolution of  $N$  qubits, at a cost of deeper circuits. Provided certain sampling assumptions, this method is not limited to weakly

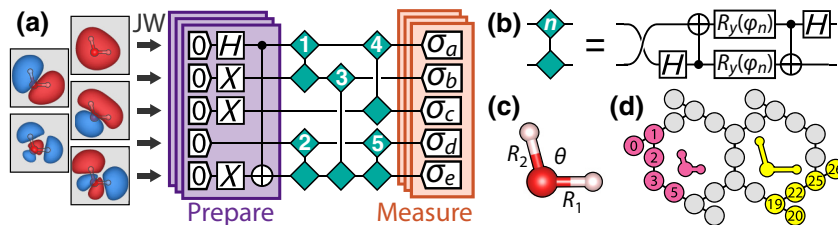


FIG. 2. Experimental ansatz structure. (a) The five active molecular orbitals are encoded in a line of qubits via the Jordan-Wigner mapping. The qubits are prepared in either a bit-string state, or a superposition of two bit strings (shown), and then acted upon by parameterized hop gates (b). In order to evaluate the Pauli strings of the Hamiltonian, each state is prepared many times, and rotated appropriately prior to measurement in the computational basis. (c) Parameterized molecular geometry from which molecular orbitals are determined. (d) Device map of *ibmq\_dublin*, with a highlighted example of how two geometries may be solved in parallel.

entangled states, so it may be applicable to a wider range of systems.

We use entanglement forging for a VQE simulation of the water molecule, as schematized in Figure 2. Starting from the minimal STO-6G basis, we freeze the core oxygen  $1s$  and the out-of-plane oxygen  $2p$  orbitals, leaving an active space of ten spin orbitals. The Jordan-Wigner mapping encodes either the spin-up or spin-down orbitals onto each five-qubit quantum circuit, and we simplify the problem structure by asserting  $V = U$  per the known symmetry between spin polarizations in the closed-shell singlet ground state.

How best to construct an ansatz circuit for VQE remains an open and active research topic [30]. To facilitate our demonstration of entanglement forging, we first ran classical simulations of VQE without entanglement forging to obtain a ten-qubit circuit that performs well for this problem. Based on this circuit, we select the quantum gates in  $U$ , and truncate the Schmidt decomposition to  $k = 3$  of the ten possible bit strings, namely  $|b_1\rangle = |11100\rangle$ ,  $|b_2\rangle = |01110\rangle$ , and  $|b_3\rangle = |01101\rangle$ . Such truncation is convenient due to a technical limitation on the number of unique circuits, distinct from the more fundamental limitation of overhead time scaling with the number of samples  $S$  discussed above; future implementations may variationally explore larger subsets of possible bit strings (Appendix E). The resulting ansatz is thus tailored to our particular problem and parameterized both by two-qubit gate-rotation angles and by the Schmidt coefficients, discussed further below. To realize each circuit required for Eq. (3), we prepare either one bit string, or a superposition of two. For example, Fig. 2(a) shows how to prepare  $|\phi_{b_1 b_3}^0\rangle$ , and a general construction is given in Appendix C.

Each initialized state is then acted upon by the unitary  $U$ , here comprised of five two-qubit “hop gates” [Figs. 2(a) and 2(b)], where each acts according to the matrix

$$h(\varphi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) & 0 \\ 0 & \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \quad (5)$$

In the simulated molecule, the rotation by  $\varphi$  moves particles between orbitals, while the entry  $-1$  provides a CPHASE-like two-particle interaction. A product of hop gates is universal with respect to real-valued wave functions of fixed particle number, making the gate appealing for simulating eigenstates of time-reversal invariant Hamiltonians. Details of compilation on quantum hardware appear in Appendix D.

The runtime cost of entanglement forging makes fast circuit execution critical. We implement the VQE routine using the Qiskit quantum-computing package, and execute the experiment via the IBM Quantum cloud computing service on the *ibmq\_dublin* processor [31,32]. Active qubit reset enabled execution at a 10-kHz repetition rate. Noting that distinct molecular geometries present embarrassingly parallel problems, we map two independent problems onto separate five-qubit chains [Fig. 2(d)], doubling throughput to 20 000 five-qubit circuits per second (Appendix H). While freedom to independently specify the number of samples for each unique circuit would best economize runtime, here we approximate this behavior by submitting multiple copies of circuits in proportion to the desired weighting (Appendix F).

Properties of the particular problem and ansatz permitted many circuits to be omitted, further accelerating execution. Generally, superposition states  $|\phi_{b_n b_m}^p\rangle$  may be skipped when evaluating Hamiltonian terms acting on only one of the two subsystems. Likewise, since here the operators  $O_1$  and  $O_2$  in the Hamiltonian are real valued in the bit-string basis, contributions from  $|\phi_{b_n b_m}^p\rangle$  with odd  $p$  must vanish. Finally, observing that our chosen hop gates nominally do not modify the Hartree-Fock (HF) state, we omit the corresponding circuits and use instead the classical mean-field result, fixing  $\langle b_1 | \tilde{O} | b_1 \rangle = E_{\text{HF}}$ . This simplification cannot capture excitations of unpaired electrons across the HF occupied-virtual threshold, and risks unphysical results as hardware errors impact intermediate calculations unevenly. Nonetheless, unburdening the quantum processor of the dominant, classically accessible part of the problem reduced  $S$  by roughly an order of magnitude for our near-equilibrium experiments, while

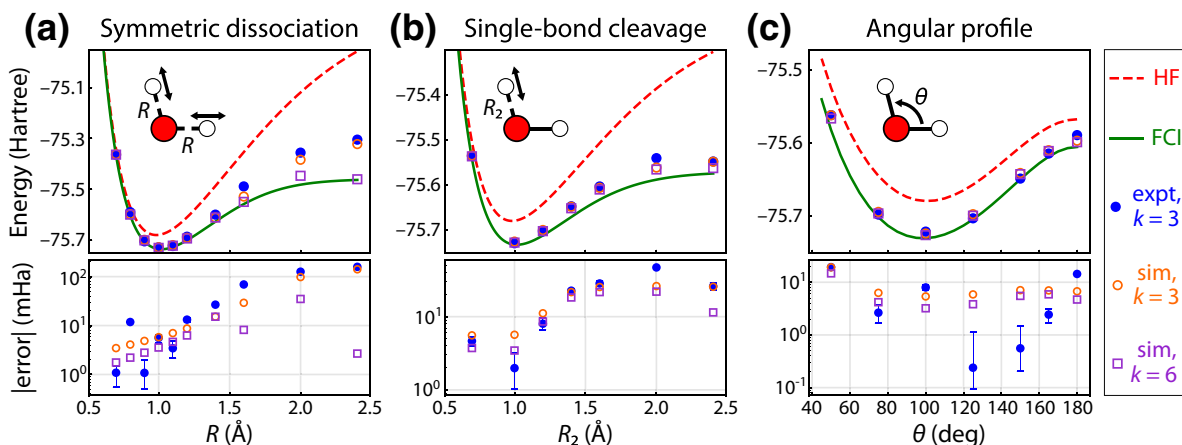


FIG. 3. VQE energies. Ground-state energies computed while varying (a) both O—H bond lengths, (b) a single O—H bond length, and (c) the H—O—H angle  $\theta$ . In the upper plots, VQE results using the entanglement-forging ansatz on *ibmq\_dublin* appear as filled blue circles, alongside curves indicating the classically computed Hartree-Fock (HF) and full configuration interaction (FCI) values for the active space. Unfilled shapes indicate results from running VQE on a noiseless classical simulator using the same three-bit-string ansatz used on *ibmq\_dublin* ( $k = 3$ ), and using a larger six-bit-string ansatz ( $k = 6$ ). Lower plots indicate absolute errors,  $|E - E_{\text{FCI}}|$ . Error bars are one standard deviation produced by bootstrapping measurement histograms, representing precision but not systematic error or drift.

retaining acceptable overall accuracy. Notably, this acceleration is enabled by the mixture of classical and quantum elements in the entanglement-forging representation.

VQE is repeated to study the behavior of entanglement forging across a variety of molecular geometries. For each geometry, optimization is performed in approximately 100 iterations of a simultaneous perturbation stochastic approximation (SPSA) algorithm [33,34]. Figure 3 displays the energies resulting from varying the length of both O—H bonds, the length of only one bond, and the H—O—H angle at fixed bond length. Where not otherwise specified, nominal equilibrium values of  $R_{\text{eq}} = 0.958 \text{ \AA}$  and  $\theta_{\text{eq}} = 104.478^\circ$  are used [35]. Deviations between the VQE results and exact full configuration interaction (FCI) energies trend from approximately 1–10 mHa to approximately 10–100 mHa as bond lengths are increased, in line with the expectation that entanglement forging should work best for problems with weakly entangled ground states. We repeat all experiments on a noiseless classical simulator, first verifying that this trend persists in the absence of gate errors (orange circles), and second observing that accuracy in the stretched regime improves substantially upon increasing the number of represented bit-string states  $k$  from three to six (purple squares), enabling the forging to describe stronger entanglement at the expense of sampling more distinct circuits. Finally, to emphasize that entanglement forging can extend to arbitrary accuracy with no increase in the number of qubits, we ran a noiseless VQE simulation at the nominal equilibrium geometry including all ten three-occupation bit strings along with more hop gates, which converged within 1.6 mHa of the FCI energy (Appendix K).

The Schmidt coefficients  $\lambda_n$  [Fig. 1(b)] are among the parameters optimized by VQE with entanglement forging (Appendix J). These coefficients are uniquely defined by the ground-state wave function and choice of partition, and thus, assuming adequate convergence of VQE to the ground state, do not depend on the selection of gates in  $U$  and  $V$  up to reordering of the index  $n$ . Figure 4 shows the Schmidt coefficients obtained on *ibmq\_dublin*, along with those obtained on a noiseless classical simulator with the same three-bit-string ansatz ( $k = 3$ ). The plots verify that weaker entanglement correlates with better performance in Fig. 3, and clarify how bond stretching strengthens spin-polarization entanglement. For instance, removing a single hydrogen [Fig. 4(b)] leads to a wave function dominated by two bit strings, consistent with a picture of dissociation into spin-1/2 fragments with opposite polarizations. Further discussion of Schmidt coefficient distributions supported by FCI calculations is provided in Appendix I.

We demonstrate entanglement forging by computing the ground-state energy of a water molecule for varying geometries using five qubits to represent ten spin orbitals, the most accurate VQE simulation of this molecule using quantum hardware to date. Entanglement forging opens a vista of experimental possibilities, such as dividing a lattice system using a spatial rather than spin partition; introducing an adaptive bisection based on orbital optimization; using entanglement forging in simulation algorithms beyond VQE, including error-corrected techniques such as quantum phase estimation; and more generally for any experiment otherwise inaccessible due to limited qubit number or connectivity where the overhead of forging may be tolerated. Where overhead remains prohibitive,

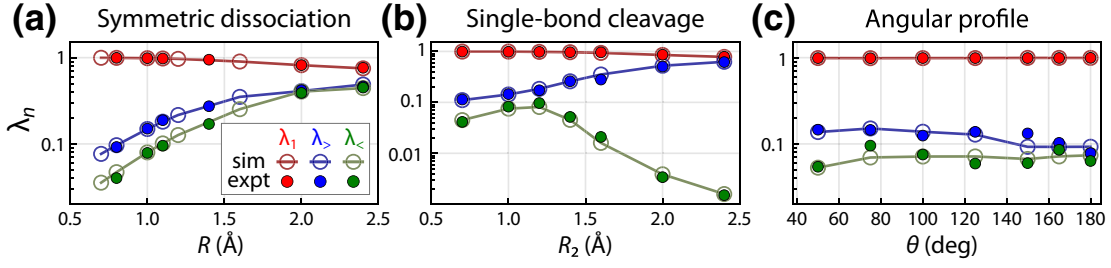


FIG. 4. Entanglement structure. Schmidt coefficients  $\lambda_n$  obtained from VQE run on *ibmq\_dublin* (solid circles) and on a noiseless classical simulator (empty circles) while varying molecular geometry as in Fig. 3. The largest component is the Hartree-Fock amplitude ( $\lambda_1$ , red); the two remaining values are sorted and plotted in blue and green. (a) Bit strings beyond Hartree Fock become increasingly significant as the molecular bonds are stretched. (b) Removing a single hydrogen produces a ground-state wave function dominated by two bit strings. (c) All choices of H—O—H bond angle yielded ground states characterized by weak entanglement.

Heisenberg-picture forging (Appendix B) may provide an alternative path to simulation of large systems. Additional exploration of techniques incorporating classical resources with quantum processing may further enhance the power of near-term quantum simulations and help realize quantum advantage for practical applications.

#### ACKNOWLEDGMENTS

We acknowledge Doug McClure, Thomas Alexander, Stephen Wood, Youngseok Kim, Daniel Egger, and the IBM Quantum backend team for technical assistance, and Julia Rice, John Lapeyre, Agata Branczyk, Lev Bishop, and Jay Gambetta for valuable discussions.

S.B. conceptualized the project. A.E., M.M., T.G., S.S. developed the main experiment and analysis software, and ran simulations and measurements. S.B., A.M., C.H. provided theoretical modeling and analyses. All authors contributed to the manuscript. Correspondence and requests for materials should be addressed to aeddins@ibm.com. Experiment data are available upon reasonable request to the authors.

S.B. is supported in part by the Army Research Office under Grant No. W911NF-20-1-0014 and by the IBM Research Frontiers Institute.

Elements of this work are included in a patent filed by the International Business Machines Corporation with the US Patent and Trademark Office.

#### APPENDIX A: DECOMPOSITION TO SUPERPOSITIONS OF $N$ -QUBIT BIT STRINGS

We would like to forge a  $2N$ -qubit state

$$|\Psi\rangle = \sum_n \lambda_n |b_n\rangle \otimes |b_n\rangle,$$

where  $\lambda_n$  are real coefficients. For each distinct pair of bit strings  $x, y$  and each  $p \in \{0, 1, 2, 3\} \equiv \mathbb{Z}_4$  define a state

$$|\phi_{xy}^p\rangle = \frac{|x\rangle + i^p |y\rangle}{\sqrt{2}}.$$

One gets

$$4 |\phi_{xy}^p\rangle \langle \phi_{xy}^p|^{\otimes 2} = i^{2p} |y\rangle \langle x|^{\otimes 2} + i^{-2p} |x\rangle \langle y|^{\otimes 2} + O + i^p O' + i^{-p} O'',$$

where  $O, O', O''$  are some operators. Note that  $\sum_{p \in \mathbb{Z}_4} i^{pq} = 4\delta_{q,0}$  for all  $q \in \mathbb{Z}_4$ . Thus

$$\sum_{p \in \mathbb{Z}_4} i^{2p} |\phi_{xy}^p\rangle \langle \phi_{xy}^p|^{\otimes 2} = |y\rangle \langle x|^{\otimes 2} + |x\rangle \langle y|^{\otimes 2}.$$

Applying the above identity to each off-diagonal term in  $|\Psi\rangle \langle \Psi|$  we arrive at

$$|\Psi\rangle \langle \Psi| = \sum_n \lambda_n^2 |b_n\rangle \langle b_n|^{\otimes 2} + \sum_n \sum_{m < n} \lambda_n \lambda_m \times \sum_{p \in \mathbb{Z}_4} (-1)^p |\phi_{b_n b_m}^p\rangle \langle \phi_{b_n b_m}^p|^{\otimes 2}.$$

#### APPENDIX B: ENTANGLEMENT FORGING IN THE HEISENBERG PICTURE

First let us define a class of quantum states whose entanglement can be efficiently forged. Consider a system of  $2N$  qubits and states of the form

$$|\psi\rangle = (U \otimes U) \sum_{x \in \{0,1\}^N} \lambda_x |x\rangle \otimes |x\rangle, \quad (\text{B1})$$

where the tensor product separates two  $N$ -qubit registers,  $\lambda_x$  are real coefficients, and  $U$  is a unitary operator with real matrix elements in the standard basis. The coefficients  $\lambda_x$  have to be normalized such that  $\sum_x \lambda_x^2 = 1$ . We assume

that  $U$  admits an efficient implementation by a quantum circuit. It is natural to consider quantum circuits composed of real gates such as the Hadamard,  $Y$ -rotation, CNOT, or a hop gate. This ensures that  $U$  has real matrix elements. The vector of  $2^N$  coefficients  $\lambda_x$  may be specified either explicitly (for small  $N$ ) or implicitly by a classical algorithm that can efficiently perform the following tasks:

- (i) sample a bit string  $x$  from the probability distribution  $\lambda_x^2$ ;
- (ii) compute the ratio  $\lambda_y/\lambda_x$  for a given pair of bit strings  $x, y$ .

Such an implicit description of exponentially large vectors is commonly used in quantum Monte Carlo simulations based on tensor networks [36] or neural-network states [37]. It follows the framework of computationally tractable quantum states introduced in Ref. [38].

Given Pauli observables  $O_1, O_2 \in \{I, X, Y, Z\}^{\otimes N}$ , our goal is to estimate the expected value

$$\mu = \langle \psi | O_1 \otimes O_2 | \psi \rangle. \quad (\text{B2})$$

We show how to accomplish this task by a series of  $N$ -qubit experiments. A typical experiment prepares a state  $U^\dagger C U |x\rangle$  for a suitable  $N$ -qubit Clifford circuit  $C$  and a bit string  $x$ . Then every qubit of the state  $U^\dagger C U |x\rangle$  is measured in the standard basis. Note that the circuit  $U^\dagger$  can be obtained from  $U$  by inverting the order of gates and replacing each gate by its inverse. The size of the Clifford circuit  $C$  depends on the form of the observables  $O_1, O_2$ . Let  $|O_j|$  be the Hamming weight of  $O_j$ , that is, the number of single-qubit terms  $X, Y, Z$  that appear in  $O_j$ . Then  $C$  contains at most  $2(|O_1| + |O_2|)$  CNOT gates and some single-qubit gates. The number of experiments required to approximate the expected value Eq. (B2) with a precision  $\epsilon$  is proportional to  $1/\epsilon^2$  with a constant prefactor. Crucially, the number of experiments does not depend on the number of qubits  $N$  or the amount of entanglement in the forged state  $|\psi\rangle$ .

The key idea behind this method is to convert a Pauli observable  $O_1 \otimes O_2$  describing a bipartite system of  $N + N$  qubits into a classical mixture of observables  $C^* \otimes C$ , where  $C$  is a self-adjoint  $N$ -qubit Clifford operator and  $C^*$  is the complex conjugate of  $C$ . By interpreting  $C$  and  $C^*$  as operators describing the forward and the backward time evolution of the same  $N$ -qubit register we express the expected value  $\mu$  as a mixture of quantum probabilities that can be measured on an  $N$ -qubit device. Since the analysis of the method is performed in the Heisenberg picture at the level of observables, we shall refer to it as the Heisenberg forging. Meanwhile the forging method described in the main text works primarily in the Schrödinger picture by decomposing an entangled state of  $N + N$  qubits into a classical mixture of product states. We refer to the latter as

the *Schrödinger forging*, to avoid a confusion between the two methods.

The Heisenberg forging roughly doubles the size of quantum circuits that need to be executed compared with the Schrödinger forging (the former needs to implement both  $U$  and  $U^\dagger$  while the latter only needs to implement  $U$ ). However, the Heisenberg forging requires shorter circuits compared with the preparation of the full state  $|\psi\rangle$  on a hardware with  $2N$  qubits. Indeed, a quantum circuit preparing the state  $|\psi\rangle$  includes two copies of  $U$  as well as some extra gates generating the initial entanglement between the  $N$ -qubit registers. Thus we expect that the Heisenberg forging reduces both the number of qubits and the number of gates per experiment while increasing the number of experiments only by a constant factor.

Using states Eq. (B1) as a variational ansatz in quantum simulations is justified if the exact ground state has real amplitudes in the standard basis and is invariant under exchanging the two  $N$ -qubit registers. As discussed in the main text, this is the case for molecular electronic structure Hamiltonians assuming that the  $N$ -qubit registers represent spin-up and spin-down orbitals. Specializing Eq. (B1) to the case  $U = I$  gives variational states in which every spatial orbital is either empty or occupied by a pair of electrons in the singlet state. Such states play the central role in the restricted Hartree-Fock space formalism [39]. As shown in Ref. [21], minimizing the energy of a molecular Hamiltonian over variational states of the form Eq. (B1) with  $U = I$  is equivalent to minimizing the energy of a quantum spin Hamiltonian with two-spin Heisenberg-type interactions and an external magnetic field over states  $|\lambda\rangle = \sum_x \lambda_x |x\rangle$ , where  $\lambda_x$  are the Schmidt coefficients that appear in Eq. (B1). We anticipate that the ansatz Eq. (B1) with  $U \neq I$  may improve upon the restricted Hartree-Fock formulation since it is based on more general variational states. We leave applications of the Heisenberg forging for a future work.

In the rest of this section we assume that  $O_1$  and  $O_2$  are nonidentity Pauli operators, which is the most interesting case. Otherwise we have  $O_1 = I$  or  $O_2 = I$ . For concreteness, assume  $O_2 = I$ . Then the expected value Eq. (B2) becomes  $\mu = \sum_x \lambda_x^2 \langle x | U^\dagger O_1 U | x \rangle$ . One can estimate  $\mu$  with a precision  $\epsilon$  by generating  $M \sim \epsilon^{-2}$  samples  $x^1, \dots, x^M \in \{0, 1\}^N$  from the distribution  $\lambda_x^2$ , measuring the eigenvalue of  $O_1$  on each state  $U|x^j\rangle$ , and computing the sample mean of the measured eigenvalues.

Let  $\{O_1, O_2\} = O_1 O_2 + O_2 O_1$ . Below we construct a decomposition

$$O_1 \otimes O_2 + O_2 \otimes O_1 = \frac{a_0}{2} (\{O_1, O_2\} \otimes I + I \otimes \{O_1, O_2\}) + \sum_{j=1}^4 a_j C_j^* \otimes C_j, \quad (\text{B3})$$

where  $a_j$  are real coefficients such that  $|a_j| \leq 1$ , and  $C_j$  are  $N$ -qubit Clifford operators. Recall that the state  $|\psi\rangle$  is invariant under the swap of the two  $N$ -qubit registers. Thus the expected value  $\mu$  is invariant under swapping the observables  $O_1$  and  $O_2$ . Symmetrizing  $\mu$  over the swap of  $O_1$  and  $O_2$  one gets  $\mu = (\langle\psi|O_1 \otimes O_2|\psi\rangle + \langle\psi|O_2 \otimes O_1|\psi\rangle)/2$ . Combining this and Eq. (B3) gives

$$\begin{aligned} \mu &= a_0 \sum_x \lambda_x^2 \operatorname{Re}(\langle x|U^\dagger O_1 O_2 U|x\rangle) \\ &+ \frac{1}{2} \sum_{j=1}^4 a_j \langle\psi|C_j^* \otimes C_j|\psi\rangle. \end{aligned} \quad (\text{B4})$$

We now convert the expected value  $\langle\psi|C_j^* \otimes C_j|\psi\rangle$  into a mixture of quantum probabilities that can be measured on an  $N$ -qubit device. From Eq. (B1) one gets

$$\begin{aligned} \langle\psi|C_j^* \otimes C_j|\psi\rangle &= \sum_{x,y} \lambda_x \lambda_y \langle y|U^\dagger C_j^* U|x\rangle \langle y|U^\dagger C_j U|x\rangle \\ &= \sum_{x,y} \lambda_x \lambda_y |\langle y|U^\dagger C_j U|x\rangle|^2. \end{aligned} \quad (\text{B5})$$

Here we use the identity

$$\begin{aligned} \langle y|U^\dagger C_j^* U|x\rangle &= \langle x|(U^\dagger C_j^* U)^T|y\rangle = \langle x|U^T C_j^\dagger U^*|y\rangle \\ &= \langle x|U^\dagger C_j^\dagger U|y\rangle, \end{aligned}$$

which follows from the assumption that  $U$  has real matrix elements. Combining Eqs. (B4) and (B5) one gets

$$\mu = a_0 \sum_x \lambda_x^2 \operatorname{Re}(\langle x|U^\dagger O_1 O_2 U|x\rangle) + \frac{1}{2} \sum_{j=1}^4 a_j \mu_j, \quad (\text{B6})$$

where

$$\mu_j = \sum_{x,y} \lambda_x \lambda_y |\langle y|U^\dagger C_j U|x\rangle|^2. \quad (\text{B7})$$

The first term in Eq. (B6) can be estimated on an  $N$ -qubit device by sampling a bit string  $x$  from the distribution  $\lambda_x^2$ , preparing a state  $U|x\rangle$ , and measuring the eigenvalue of  $O_1 O_2$  [note that the first term in Eq. (B6) is nonzero only if  $O_1 O_2$  is a self-adjoint operator, that is, if  $O_1$  commutes with  $O_2$ ]. Below we focus on the second term in Eq. (B6) and show how to estimate the quantities  $\mu_j$ . Define a function

$$R(x,y) = \frac{\lambda_y}{\lambda_x}. \quad (\text{B8})$$

By assumption, one can efficiently compute  $R(x,y)$  for a given pair  $x,y$ . For each bit string  $x$  define a conditional

probability distribution

$$P_j(y|x) = |\langle y|U^\dagger C_j U|x\rangle|^2. \quad (\text{B9})$$

One can sample a bit string  $y$  from  $P_j(y|x)$  on a quantum device with  $N$  qubits by preparing the state  $U^\dagger C_j U|x\rangle$  and measuring every qubit in the standard basis. Furthermore, Eq. (B7) implies that  $\mu_j$  is the expected value of  $R(x,y)$  over the probability distribution  $\lambda_x^2 P_j(y|x)$ , that is,

$$\mu_j = \sum_{x,y} \lambda_x^2 P_j(y|x) R(x,y) \equiv \mathbb{E}(R). \quad (\text{B10})$$

The random variable  $R(x,y)$  has the variance at most one since

$$\begin{aligned} \mathbb{E}(R^2) &= \sum_{x,y} \lambda_x^2 P_j(y|x) R^2(x,y) = \sum_{x,y} \lambda_y^2 |\langle y|U^\dagger C_j U|x\rangle|^2 \\ &= \sum_y \lambda_y^2 = 1. \end{aligned} \quad (\text{B11})$$

Thus one can estimate  $\mu_j$  with an additive error  $\epsilon$  by generating roughly  $1/\epsilon^2$  samples  $(x,y)$  from the distribution  $\lambda_x^2 P_j(y|x)$  and computing the sample mean of  $R(x,y)$ . By estimating each term  $\mu_j$  in Eq. (B6) with a precision  $\epsilon/4$  and estimating the expected value  $\operatorname{Re}(\langle x|U^\dagger O_1 O_2 U|x\rangle)$  with a precision  $\epsilon/2$  gives the desired  $\epsilon$  estimate of  $\mu$  (use the triangle inequality and the assumption  $|a_j| \leq 1$ ). The total number of  $N$ -qubit experiments required to obtain these estimates is proportional to  $1/\epsilon^2$  with a constant prefactor.

In the rest of this section we explicitly construct the decomposition Eq. (B3) for nonidentity Pauli observables  $O_1, O_2 \in \{I, X, Y, Z\}^{\otimes N}$ . Let  $w_j$  be the number of single-qubit  $Y$  terms that appear in  $O_j$ . Note that  $O_j^* = (-1)^{w_j} O_j$ , that is, the matrix of  $O_j$  in the standard basis is real if  $w_j$  is even and imaginary if  $w_j$  is odd. Since the state  $|\psi\rangle$  is real, one has  $\mu = \langle\psi|O_1 \otimes O_2|\psi\rangle = 0$  unless  $w_1$  and  $w_2$  have the same parity. Below we assume that this is the case, that is,  $O_j^* = \sigma O_j$  for  $\sigma = (-1)^{w_1} = (-1)^{w_2}$ . Recall that Pauli observables either commute or anticommute.

*Case 1:*  $O_1$  anticommutes with  $O_2$ . Below we prove the following simple fact.

**Lemma 1.** *Let  $O_1, O_2 \in \{I, X, Y, Z\}^{\otimes N}$  be anticommuting Pauli observables. There exist  $N$ -qubit Clifford circuit  $V$  and a qubit  $q \in \{1, 2, \dots, N\}$  such that*

$$O_1 = VX_q V^\dagger \quad \text{and} \quad O_2 = VZ_q V^\dagger. \quad (\text{B12})$$

*The circuit  $V$  contains at most  $|O_1| + |O_2| - 2$  CNOT gates and some single-qubit Clifford gates. Here  $|O_j|$  is the Hamming weight  $O_j$ .*

Let  $V$  be the Clifford circuit from Lemma 1. Define Clifford operators

$$C_1 = V \frac{1}{\sqrt{2}} (X_q + Z_q) V^\dagger \quad \text{and} \quad C_2 = V \frac{1}{\sqrt{2}} (X_q - Z_q) V^\dagger. \quad (\text{B13})$$

Note that  $C_1 = (O_1 + O_2)/\sqrt{2}$  and  $C_2 = (O_1 - O_2)/\sqrt{2}$ . A simple algebra gives

$$O_1 \otimes O_2 + O_2 \otimes O_1 = C_1 \otimes C_1 - C_2 \otimes C_2. \quad (\text{B14})$$

Taking the partial transpose of Eq. (B14) over the first  $N$ -qubit register, using the fact that  $O_j$  and  $C_j$  are self-adjoint, and the identity  $O_j^* = \sigma O_j$  one gets

$$O_1 \otimes O_2 + O_2 \otimes O_1 = \sigma (C_1^* \otimes C_1 - C_2^* \otimes C_2). \quad (\text{B15})$$

This is the desired decomposition Eq. (B3) with  $a_1 = \sigma$ ,  $a_2 = -\sigma$ , and  $a_0 = a_3 = a_4 = 0$ . Lemma 1 and Eq. (B13) imply that  $C_1$  and  $C_2$  can be implemented by circuits with at most  $2(|O_1| + |O_2|)$  CNOTs.

*Case 2:*  $O_1$  commutes with  $O_2$ . If  $O_1 = O_2$  then choose  $C_1 = C_2 = O_1$ ,  $a_1 = a_2 = \sigma$ , and  $a_0 = a_3 = a_4 = 0$ . One can easily check that Eq. (B3) is satisfied. From now on we assume  $O_1 \neq O_2$ . Below we prove the following lemma.

**Lemma 2.** *Let  $O_1, O_2 \in \{I, X, Y, Z\}^{\otimes N}$  be commuting Pauli observables such that  $O_1 \neq O_2 \neq I$ . There exist  $N$ -qubit Clifford circuit  $V$  and a pair of qubits  $p, q \in \{1, 2, \dots, N\}$  such that*

$$O_1 = V Z_p V^\dagger \quad \text{and} \quad O_2 = V Z_q V^\dagger. \quad (\text{B16})$$

The circuit  $V$  contains at most  $|O_1| + |O_2| - 2$  CNOT gates and some single-qubit Clifford gates.

Let  $V$  be the Clifford circuit from Lemma 2. Define  $N$ -qubit operators

$$C_{\alpha,\beta} = V X_p^\alpha X_q^\beta \mathbf{CZ}_{p,q} X_p^\alpha X_q^\beta V^\dagger,$$

where  $\alpha, \beta \in \{0, 1\}$  and  $\mathbf{CZ}_{p,q}$  denotes the controlled- $Z$  gate acting on qubits  $p, q$ . Using the identity

$$\begin{aligned} X_p^\alpha X_q^\beta \mathbf{CZ}_{p,q} X_p^\alpha X_q^\beta \\ = (1/2)[I + (-1)^\alpha Z_p + (-1)^\beta Z_q - (-1)^{\alpha+\beta} Z_p Z_q] \end{aligned}$$

one gets

$$C_{\alpha,\beta} = (1/2)[I + (-1)^\alpha O_1 + (-1)^\beta O_2 - (-1)^{\alpha+\beta} O_1 O_2].$$

A simple algebra shows that

$$\begin{aligned} O_1 \otimes O_2 + O_2 \otimes O_1 &= O_1 O_2 \otimes I + I \otimes O_1 O_2 \\ &+ \sum_{\alpha,\beta=0,1} (-1)^{\alpha+\beta} C_{\alpha,\beta} \otimes C_{\alpha,\beta}. \end{aligned}$$

Taking the partial transpose over the first  $N$ -qubit register, using the identity  $O_j^* = \sigma O_j$ , and the fact that  $C_j$  and  $O_j$  are self-adjoint one gets

$$\begin{aligned} O_1 \otimes O_2 + O_2 \otimes O_1 &= \sigma (O_1 O_2 \otimes I + I \otimes O_1 O_2) \\ &+ \sigma \sum_{\alpha,\beta=0,1} (-1)^{\alpha+\beta} C_{\alpha,\beta}^* \otimes C_{\alpha,\beta}. \end{aligned}$$

This is the desired decomposition Eq. (B3) with  $C_1 = C_{0,0}$ ,  $C_2 = C_{0,1}$ ,  $C_3 = C_{1,0}$ ,  $C_4 = C_{1,1}$ ,  $a_0 = a_1 = a_4 = \sigma$ , and  $a_2 = a_3 = -\sigma$ .

It remains to prove Lemmas 1,2.

**Proof of Lemma 1.** We convert  $O_1$  and  $O_2$  to single-qubit Pauli operators  $X_q$  and  $Z_q$ , respectively, by a sequence of steps  $O_1 \leftarrow W^\dagger O_1 W$  and  $O_2 \leftarrow W^\dagger O_2 W$ , where  $W$  is a Clifford circuit composed of CNOT, CZ, and single-qubit gates. We choose  $W$  such that at each step the combined weight  $|O_1| + |O_2|$  is reduced at least by a number of two-qubit gates in  $W$ . The desired circuit  $V$  is then obtained as the composition of the circuits  $W$  applied at each step.

The first step converts  $O_1$  and  $O_2$  into the standard form such that their action on any qubit falls into one of five cases shown below.

Case	$A$	$B$	$C$	$D$	$E$
$O_1$	$X$	$Z$	$I$	$Z$	$I$
$O_2$	$Z$	$I$	$Z$	$Z$	$I$

The corresponding circuit  $W$  can be easily constructed as a product of single-qubit Clifford gates. This gives rise to a partition of  $N$  qubits into five disjoint subsets,  $[N] = ABCDE$ . For example,  $A$  contains all qubits  $j$  such that  $O_1$  and  $O_2$  act on  $j$  by Pauli  $X$  and  $Z$ , respectively. Note that  $A$  has an odd size since otherwise  $O_1$  and  $O_2$  would commute. In particular,  $A \neq \emptyset$ .

Suppose  $B \neq \emptyset$ . Apply  $W = \mathbf{CZ}_{a,b}$  for some  $a \in A$  and  $b \in B$ . This reduces  $|O_1|$  by one without changing  $O_2$ .

Suppose  $CD \neq \emptyset$ . Apply  $W = \text{CNOT}_{b,a}$  for some  $a \in A$  and  $b \in CD$ . This reduces  $|O_2|$  by one without changing  $O_1$ .

In the remaining case  $B = C = D = \emptyset$  and  $A \neq \emptyset$ . Assume without loss of generality that  $A = \{1, 2, \dots, 2k+1\}$  for some integer  $k$ . Then  $O_1 = X_1 X_2 \cdots X_{2k+1}$  and  $O_2 = Z_1 Z_2 \cdots Z_{2k+1}$ . Set  $q = 1$ . Choose

$$W = \prod_{a=1}^k \text{CNOT}_{1,2a+1} \text{CNOT}_{2a,1} \text{CNOT}_{2a+1,2a}.$$

One can easily check that  $W^\dagger O_1 W = X_1$  and  $W^\dagger O_2 W = Z_1$ . Thus  $W$  reduces the combined weight  $|O_1| + |O_2|$  by  $4k$ . Furthermore,  $W$  contains  $3k \leq 4k$  two-qubit gates. ■

**Proof of Lemma 2.** We use the notations introduced in the proof of Lemma 1. Consider the standard form of  $O_1$



and  $O_2$ . Suppose  $A \neq \emptyset$ . Note that  $|A|$  is even since  $O_1$  and  $O_2$  commute. Assume without loss of generality that  $A = \{1, 2, \dots, 2k\}$  for some integer  $k$ . Choose

$$W = \prod_{a=1}^k \text{H}_{2a-1} \text{CNOT}_{2a,2a-1} \text{CNOT}_{2a-1,2a}.$$

One can easily check that  $W$  maps  $X_1 X_2 \dots X_{2k}$  and  $Z_1 Z_2 \dots Z_{2k}$  to  $Z_1 Z_3 \dots Z_{2k-1}$  and  $Z_2 Z_4 \dots Z_{2k}$ , respectively. The combined weight  $|O_1| + |O_2|$  is reduced by  $2k$  and  $W$  contains  $2k$  two-qubit gates. From now on we can assume  $A = \emptyset$ .

If  $B \neq \emptyset$  and  $D \neq \emptyset$  then apply  $W = \text{CNOT}_{d,b}$  for some  $b \in B$  and  $d \in D$ . This reduces  $|O_1|$  by one without changing  $O_2$ .

If  $C \neq \emptyset$  and  $D \neq \emptyset$  then apply  $W = \text{CNOT}_{d,c}$  for some  $c \in C$  and  $d \in D$ . This reduces  $|O_2|$  by one without changing  $O_1$ .

After a sequence of steps as above we have  $A, D = \emptyset$ . Note that  $B \neq \emptyset$  and  $C \neq \emptyset$  since we assume that  $O_1 \neq I$  and  $O_2 \neq I$ .

If  $|B| \geq 2$  then apply  $W = \text{CNOT}_{b,b'}$  for some qubits  $b, b' \in B$ . This reduces  $|O_1|$  by one without changing  $O_2$ . If  $|C| \geq 2$  then apply  $W = \text{CNOT}_{c,c'}$  for some qubits  $c, c' \in C$ . This reduces  $|O_2|$  by one without changing  $O_1$ .

After a sequence of steps as above we have  $A, D = \emptyset$ ,  $|B| = 1$ , and  $|C| = 1$ , that is,  $O_1 = Z_p$  and  $O_2 = Z_q$  for some pair of qubits  $p \neq q$ . ■

### APPENDIX C: STATE INITIALIZATION ROUTINES

For the entanglement-forging scheme used in the experiment, we need to initialize  $N$  qubits in a superposition state  $|\phi_{xy}^p\rangle = (|x\rangle + i^p |y\rangle)/\sqrt{2}$ , where  $x$  and  $y$  are  $N$  qubit bit strings, and  $x \neq y$ . A general construction of  $|\phi_{xy}^p\rangle$  proceeds as follows:

1. Find an index  $k$  where  $x_k \neq y_k$ .
2. If  $x_k = 1$  ( $y_k = 0$ ), swap the definitions of  $x$  and  $y$ , and substitute  $p \rightarrow (-p \pmod{4})$ , using the fact that  $|\phi_{xy}^p\rangle$  is the same as  $|\phi_{yx}^{-p}\rangle$  up to an inconsequential global phase.
3. Find the sets of indices  $S = \{l \neq k : x_l \neq y_l\}$  and  $T = \{l : x_l = 1\}$ .
4. Given a set of  $N$  qubits prepared in  $|0\rangle^{\otimes N}$ , apply the single-qubit gates  $\otimes_{i \in T} X_i$ .
5. Apply single-qubit gate  $G_p$  to qubit  $k$ , defined as  $(G_0, G_1, G_2, G_3) = (H, SH, ZH, SZH)$ . Note the Hadamard acts first in each case.
6. For all qubits  $l \in S$ , apply  $\text{CNOT}_{kl}$ .

The above procedure may become difficult to execute with high fidelity in simulations of large systems as limited processor connectivity may lead to large numbers of

costly swap operations, or simply as the Hamming distance between pairs of included bit strings grows and necessitates more CNOTs in step 6. Moreover, circuits with different initialization gates will experience different errors, and these uneven error rates can lead to inference of nonphysical average behaviors, e.g., energies below the correct ground-state energy. To circumvent this issue, one may replace the circuits initialized as  $N$ -qubit superposition states by a larger number of circuits initialized as products of single-qubit states, which are readily prepared with high fidelity. Suppose  $x \neq y$  are  $N$ -bit strings that differ on  $d$  bits. For each qubit  $j$  and an integer  $p$  define a single-qubit state

$$|\psi_{xy}^{pj}\rangle = \begin{cases} |x_j\rangle & \text{if } x_j = y_j \\ \frac{1}{\sqrt{2}} (|x_j\rangle + e^{i\pi p/2d} |y_j\rangle) & \text{if } x_j \neq y_j. \end{cases} \quad (\text{C1})$$

Define an  $N$ -qubit tensor-product state

$$|\psi_{xy}^p\rangle = |\psi_{xy}^{p1}\rangle \otimes |\psi_{xy}^{p2}\rangle \otimes \dots \otimes |\psi_{xy}^{pN}\rangle. \quad (\text{C2})$$

Such a state can be easily prepared starting from the basis vector  $|0^N\rangle$  by applying Hadamard gates, Pauli- $X$  gates, and single-qubit  $Z$  rotations by the angle  $\pm\pi p/2d$ . After simple algebra one gets a decomposition

$$|x\rangle\langle y|^{\otimes 2} + |y\rangle\langle x|^{\otimes 2} = \frac{4^d}{4d} \sum_{p=0}^{4d-1} (-1)^p |\psi_{xy}^p\rangle\langle\psi_{xy}^p|^{\otimes 2}. \quad (\text{C3})$$

Note that the right-hand side is a linear combination of  $2N$ -qubit tensor-product states. We can use this decomposition to classically forge entanglement for an arbitrary state

$$|\psi\rangle = (U \otimes V) \sum_n \lambda_n |b_n\rangle \otimes |b_n\rangle.$$

Indeed, write the density matrix  $|\psi\rangle\langle\psi|$  as a linear combination of diagonal terms  $|b_n\rangle\langle b_n|^{\otimes 2}$  and off-diagonal terms  $|b_n\rangle\langle b_m|^{\otimes 2} + |b_m\rangle\langle b_n|^{\otimes 2}$  with  $n \neq m$ . Applying Eq. (C3) with  $x = b_n$  and  $y = b_m$  to each off-diagonal term one finally arrives at

$$\begin{aligned} \langle\psi|O_1 \otimes O_2|\psi\rangle &= \sum_n \lambda_n^2 \langle b_n|U^\dagger O_1 U|b_n\rangle \times \langle b_n|V^\dagger O_2 V|b_n\rangle \\ &+ \sum_{n < m} \frac{\lambda_n \lambda_m 4^{d_{n,m}}}{4d_{n,m}} \sum_{p=0}^{4d_{n,m}-1} (-1)^p \langle\psi_{b_n b_m}^p|U^\dagger O_1 U|\psi_{b_n b_m}^p\rangle \\ &\times \langle\psi_{b_n b_m}^p|V^\dagger O_2 V|\psi_{b_n b_m}^p\rangle, \end{aligned} \quad (\text{C4})$$

where  $O_i$  are arbitrary  $N$ -qubit observables and  $d_{n,m}$  is the Hamming distance between the bit strings  $b_n$  and  $b_m$  (the

number of bit flips separating  $b_n$  and  $b_m$ ). Note that each term in Eq. (C4) can be estimated on a device with only  $N$  qubits by initializing each individual qubit in the state  $|0\rangle$ ,  $|1\rangle$ , or  $(|0\rangle + e^{\pm i\pi p/2d}|1\rangle)/\sqrt{2}$  according to Eqs. (C1) and (C2) with  $x = b_n$ ,  $y = b_m$ , applying the circuit  $U$  or  $V$ , and measuring the eigenvalue of  $O_1$  or  $O_2$ . However, the exponential factor  $4^{d_{n,m}}$  in the decomposition Eq. (C4) may lead to a loss of accuracy. For concreteness, suppose the observables  $O_i$  are normalized such that the operator norm of  $O_i$  is at most one (for example,  $O_i$  are  $N$ -qubit Pauli operators). Suppose each  $N$ -qubit expected value in Eq. (C4) can be estimated within an additive error  $\epsilon_0$ . Using the triangle inequality one can easily check that the right-hand side of Eq. (C4) approximates the expected value  $\langle \psi | O_1 \otimes O_2 | \psi \rangle$  within an additive error

$$\epsilon = O(\epsilon_0) \left( 1 + \sum_{n < m} |\lambda_n \lambda_m| 4^{d_{n,m}} \right). \quad (\text{C5})$$

Thus the method is practical only if  $d_{n,m}$  is sufficiently small for all bit strings  $b_n$  and  $b_m$  that contribute to  $|\psi\rangle$ .

**APPENDIX D: CONNECTIVITY AND SWAPS**

Figure 5 outlines how we utilize the structure of the particular forged ansatz in this demonstration to compile the required circuits onto a five-qubit line without introducing any additional gate-based swap operations. Panel (a) shows the definition of the hop gate, denoted by green diamonds, which we may also abbreviate as a swap followed by the gate denoted by green circles. Panel (b) likewise

defines a “modified hop gate,” which does not involve a swap, and acts like the hop gate except that it leaves the  $|11\rangle$  state unchanged. Following panel (c), we start from the set of hop gates in Fig. 2, use the fact that in the experiment we fix  $\varphi_3 = 0$  to reduce that hop gate to a CPHASE gate, and unpack gates 1, 4, and 5 according to the hop gate definition. (The CPHASE gate is compiled using Hadamards and a CNOT.)

We now compile hop-gate 2 differently depending on the initialization subcircuits (purple rectangles) of the given circuit. Unpacking gate 2 per its definition and untwisting all crossed wires produces the circuit in the top right of panel (d). The wires have been permuted to reveal the linear connectivity; in terms of the letters labeling the qubits in the figure, the required connectivity is  $C-B-A-D-E$ . To this set of gates we need to prepend initialization subcircuits preparing superpositions of bit strings. Each such subcircuit in this experiment contains one additional CNOT (red); to avoid any gate-based swaps, the two involved qubits must also be connected. The two purple subcircuits in **d**,  $|\phi_{b_1 b_2}^0\rangle$  and  $|\phi_{b_2 b_3}^0\rangle$ , respectively, require connections  $A-D$  and  $D-E$ . Since this is compatible with the connectivity  $C-B-A-D-E$ , we compile hop-gate 2 as per panel (d) whenever these states are prepared.

In contrast, superposing the third pair of bit strings by the  $|\phi_{b_1 b_3}^0\rangle$  subcircuit in **e** requires  $A-E$  connectivity. To realize this, we note that this state preparation ideally never provides  $|11\rangle$  as an input to hop-gate 2, and thus hop-gate 2 may be replaced by a modified hop gate with no nominal change in behavior. With the swap removed from hop-gate 2, the desired linear connectivity of the hop-gate

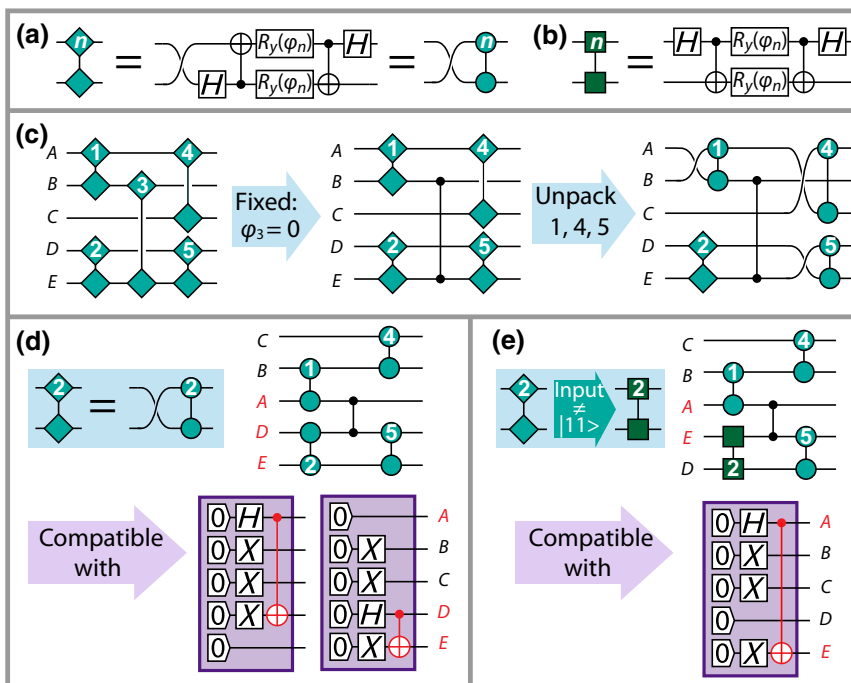


FIG. 5. Compilation onto a line of qubits. (a) Definition of the hop gate (green diamonds) in terms of a swap followed by standard gate operations, the latter abbreviated by the gate with green circles. (b) Definition of the modified hop gate, equivalent to a hop gate except that it leaves the  $|11\rangle$  state unchanged. (c) Initial compilation steps for the hop gates used in the experiment. (d) Likewise unpacking hop-gate 2 per its definition leads to the circuit in the upper right, the connectivity of which is compatible with the two state-initialization subcircuits shown (purple rectangles). (e) Alternatively, replacing hop-gate 2 with a modified hop gate provides a connectivity solution for the state-initialization subcircuit shown in this panel.

circuit becomes  $C$ - $B$ - $A$ - $E$ - $D$ , providing the  $A$ - $E$  connection facilitating  $|\phi_{b_1 b_3}^0\rangle$ .

Nonphysical results may be inferred in cases where hardware noise acts unevenly across the set of executed circuits, as will occur to some extent when varying gate compilations depending on the state preparation. However, the total error in the circuit is reduced by the avoidance of swaps, and we accept the trade-off for this experiment. An intriguing question is how generally entanglement forging and other circuit decomposition techniques may be used to circumvent limitations of device connectivity [9]. Future entanglement-forging experiments may reduce connectivity requirements using the decomposition to single-qubit stabilizer states presented in Appendix C.

### APPENDIX E: WEIGHTED SAMPLING OF CIRCUITS: THEORY

Suppose  $O_1, O_2 \in \{I, X, Y, Z\}^{\otimes N}$  are  $N$ -qubit Pauli observables and  $|\psi\rangle$  is a  $2N$ -qubit state. Entanglement forging provides a decomposition

$$\langle \psi | O_1 \otimes O_2 | \psi \rangle = \sum_{a=1}^{\ell} \mu_a \text{Tr}(O_1 \rho_a) \text{Tr}(O_2 \rho_a), \quad (\text{E1})$$

where  $\rho_a$  are  $N$ -qubit pure states and  $\mu_a$  are real coefficients simply related to the Schmidt coefficients of  $|\psi\rangle$ . Let  $V_i$  be a Clifford circuit that maps each Pauli term  $X$  or  $Y$  that appears in  $O_i$  to the Pauli  $Z$ . Note that  $V_i$  includes only single-qubit Clifford gates. Then

$$O_i = V_i^\dagger \left( \sum_{x \in \{0,1\}^N} O_i(x) |x\rangle \langle x| \right) V_i,$$

where  $O_i(x)$  takes value  $+1$  or  $-1$  if the parity of  $x$  restricted to the support of  $O_i$  is even or odd, respectively. We would like to estimate the quantity  $\langle \psi | O_1 \otimes O_2 | \psi \rangle$  on a quantum device with only  $N$  qubits by a series of experiments such that each experiment prepares a state  $V_i \rho_a V_i^\dagger$  for some pair  $(i, a)$  and measures each qubit in the computational basis. We claim that estimating  $\langle \psi | O_1 \otimes O_2 | \psi \rangle$  with an additive error  $\epsilon$  and a confidence level 99% requires at most  $S$  experiments, where

$$S = \frac{200 (\|\mu\|_1)^2}{\epsilon^2}, \quad \|\mu\|_1 \equiv \sum_{a=1}^{\ell} |\mu_a|. \quad (\text{E2})$$

For large problems, Monte Carlo methods may be used to generate the necessary  $M = S/2$  state preparations  $\rho_a$ . Indeed, define a normalized probability distribution  $\pi_a = |\mu_a| / \|\mu\|_1$ , where  $a = 1, 2, \dots, \ell$ , and let  $a_1, a_2, \dots, a_M \in \{1, 2, \dots, \ell\}$  be  $M$  independent samples from the distribution. For each  $j = 1, 2, \dots, M$  perform two experiments:

- (1) prepare the state  $V_1 \rho_{a_j} V_1^\dagger$ , measure every qubit in the  $Z$  basis, and record the measured bit string  $x^j \in \{0, 1\}^N$ ;
  - (2) prepare the state  $V_2 \rho_{a_j} V_2^\dagger$ , measure every qubit in the  $Z$  basis, and record the measured bit string  $y^j \in \{0, 1\}^N$ .
- Define a random variable

$$f = \frac{\|\mu\|_1}{M} \sum_{j=1}^M \text{sgn}(\mu_{a_j}) O_1(x^j) O_2(y^j). \quad (\text{E3})$$

One can easily check that the mean and the variance of  $f$  are

$$\begin{aligned} \mathbb{E}(f) &= \langle \psi | O_1 \otimes O_2 | \psi \rangle \quad \text{and} \\ \mathbb{E}(f^2) - \mathbb{E}(f)^2 &= \frac{\|\mu\|_1^2 - \langle \psi | O_1 \otimes O_2 | \psi \rangle^2}{M}. \end{aligned} \quad (\text{E4})$$

Here the expectation values are taken over the random choice of  $a_1, \dots, a_M$  and the random measurement outcomes. We conclude that  $f$  is an unbiased estimator of  $\langle \psi | O_1 \otimes O_2 | \psi \rangle$  with the variance at most  $\|\mu\|_1^2 / M = \epsilon^2 / 100$ . By the Chebyshev inequality,  $|f - \langle \psi | O_1 \otimes O_2 | \psi \rangle| \leq \epsilon$  with probability at least 0.99.

We note that estimating  $\langle \psi | O_1 \otimes O_2 | \psi \rangle$  on a  $2N$ -qubit device with a precision  $\epsilon$  and confidence level 99% would require at most  $100\epsilon^{-2}$  experiments (each experiment prepares the state  $V_1 \otimes V_2 |\psi\rangle$  and measures every qubit in the  $Z$  basis). From Eq. (E2) one infers that entanglement forging increases the required number of experiments roughly by the factor  $2(\|\mu\|_1)^2$ . However, each experiment requires  $N$  instead of  $2N$  qubits. In addition, as argued in the main text, quantum circuits preparing the states  $\rho_a$  are much simpler compared with the circuit preparing the full state  $\psi$ .

Let  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{2^N})$  be the Schmidt coefficients of  $|\psi\rangle$ . Equation (2) from the main text gives

$$\|\mu\|_1 = 1 + 4 \left( \sum_{i=1}^{2^N} |\lambda_i| \right)^2. \quad (\text{E5})$$

Thus entanglement forging is mostly useful for weakly entangled states such that  $\|\lambda\|_1 = \sum_{i=1}^{2^N} |\lambda_i|$  is a constant or a slowly growing function of  $N$  (note that  $\|\lambda\|_1 = \sqrt{2^N}$  in the worst case when  $|\psi\rangle$  is a maximally entangled state of  $N + N$  qubits). In practice, one may wish to limit the maximum number of experiments  $S$  by some specified cutoff  $S_{\max}$ . Combining Eqs. (E2), (E5) that determine the number of experiments  $S = S(\epsilon, \lambda)$  one obtains a nonlinear constraint  $S(\epsilon, \lambda) \leq S_{\max}$  on the vector of Schmidt coefficients  $\lambda$ . This constraint should be incorporated into the classical optimizer that minimizes the variational energy.

## APPENDIX F: WEIGHTED SAMPLING OF CIRCUITS: IMPLEMENTATION

As the number of bit strings included in the experimental demonstration (three bit strings) is computationally tractable, Monte Carlo methods are not required. To minimize the variance of our estimate of the sum in Eq. (3), we wish to allocate the total experimental samples  $S$  to the various circuits  $\langle b_n | \tilde{O}_k | b_m \rangle$  in proportion to the respective coefficients  $\lambda_n \lambda_m$ . However, at the time of the experiment Qiskit did not yet support nonuniform sampling of a list of distinct quantum circuits, instead sampling all  $J$  circuits in the submitted list, or job, an equal number of times  $s = S/J$ . We thus wrote a small “copysampling” function to approximate nonuniform sampling by populating the job with proportionally more copies of more important circuits. The function takes as inputs the desired normalized statistical weight  $w_c$  for each circuit  $c$ , and the final size  $J$  of the job to be submitted. Thus a target expression for the number of copies desired for circuit  $c$  is  $w_c J$ , though this is not necessarily integer valued. For simplicity, the copysampling function first ensures the job includes at least one copy of every possible circuit. Next, it deterministically appends  $\text{floor}(w_c J - 1)$  additional copies of each circuit  $c$ . There are then  $J - \sum_c [1 + \text{floor}(w_c J - 1)]$  remaining spots in the job. These are allocated to circuits randomly according to the remaining residuals, without replacement, to approximate the desired weighting distribution. A similar partially deterministic shot allocation scheme was studied in Ref. [40]. The job is then executed, and the results (“counts”) for all copies of  $c$  are merged before analysis.

The above is limited in that the weighted sampling is coarse grained by the number of samples  $s$  per copy of each circuit. The desired weighted-sampling distribution could be better approximated by setting  $s = 1$  and  $J = S$ . However, at present there is appreciable overhead in execution time scaling with the job size  $J$ . In our experiment, we find  $J \approx 800$  and  $s \approx 2000$  gives an acceptable compromise between job-size overhead time and efficiency of weighted sampling.

Our sampling weights accounted for both the coefficients  $\lambda_n \lambda_m$  and the zero-noise extrapolation stretch factors discussed below. We note that possible extensions could further optimize sampling weights for each quantum circuit by accounting for weights of the Pauli strings in the Hamiltonian, optionally further informed by the expectations of Pauli strings obtained in the previous VQE iteration. Optimally assigning samples accounting for Pauli strings is nontrivial as Hamiltonian terms contain pairwise products of  $N$ -qubit Pauli-string expectations, and moreover groups of compatible Paulis are in our experiment measured jointly via the tensor-product basis grouping method [41] standard in Qiskit. Nonetheless, preliminary simulations suggest such weighting schemes can further

reduce the required number of samples  $S$  by more than approximately 25%.

## APPENDIX G: REPEATED-GATE ZERO-NOISE EXTRAPOLATION

The accuracy of energies computed on the quantum hardware is significantly improved by mitigating gate errors via zero-noise extrapolation, or ZNE [42,43]. To avoid the overhead of calibrating stretched gates, we implement a repeated-gate ZNE routine similar to that in Ref. [44]. For each quantum circuit, a copy is made by applying the transformation  $G \rightarrow GG^{-1}G$  to each primitive gate, approximating a noise-amplification factor of 3. We use first-order extrapolation of the results from the original and copied circuits to infer the approximate zero-noise result for each distinct circuit [i.e., for each distinct (state-preparation, Pauli-string) combination] used to compute the desired observable. Additionally, because the extrapolated result depends more sensitively on the original circuit than on the noise-amplified circuit, to improve overall precision we assign a proportionally higher weight to the original circuit in the weighted-sampling routine described above.

## APPENDIX H: PARALLEL CIRCUIT EXECUTION

To reduce the total data-acquisition time, we execute pairs of VQE problems simultaneously on different subsets of the Dublin processor. Physical qubits are selected based on automated device benchmarking performed shortly before to the start of each VQE run. To reduce the risk of crosstalk between problems, we require that no direct connectivity exists between the two sets of physical qubits, with a buffer of at least one idle physical qubit. A similar parallelization scheme was recently characterized in Ref. [45]. Future parallelized experiments might suppress readout crosstalk as needed by choosing sets of qubits disjoint with respect to readout-line multiplexing, or by applying efficient mitigation methods such as in Ref. [46].

This parallelization routine is implemented in software to be a modular wrapper of the experiment code. Using the standard PYTHON multiprocessing package, the wrapper first launches a PYTHON process executing one copy of the experiment code for each independent VQE problem. In each iteration of the VQE optimization, each of these processes generates a list of circuits, or job, it needs executed. Rather than passing the job directly to the backend as usual, each process passes its job to a common “multiplexer” PYTHON process. The multiplexer, respectively, merges the circuits in the separate jobs into a single new job, sends that to the backend for execution, and then separates the component results and returns them to the corresponding VQE processes. This framework is convenient as the multiplexer can be swapped in as a virtual

backend with little modification of the original experiment code.

### APPENDIX I: SCHMIDT COEFFICIENTS AND TRUNCATION

For technical convenience in the forging demonstration, rather than sampling all possible bit strings, we truncate the space to include only the leading  $k = 3$  bit strings, that is, set  $\lambda_{n>k} = 0$ . Noiseless simulations with  $k = 3$  and  $k = 6$  are performed for comparison (Fig. 3). Extended data, Fig. 6, illustrates how the choice of  $k$  limits accuracy of describing the various ground state wave functions, here obtained via FCI calculations. For the case of symmetric stretching, a qualitative change occurs for  $k \geq 6$  such that accuracy improves as the molecule dissociates further. Single-bond cleavage exhibits a similar, though less pronounced, change for  $k \geq 2$ , as per the discussion in the main text. In contrast, sweeping the bond angle does not reveal any critical value of  $k$ . Note that unlike the experiments in the main text, the FCI calculations here do not freeze the oxygen  $2p$  orbital; including this orbital increases the number of allowed bit strings from 10 to 15, such that a nonzero residual remains for  $k = 10$ .

### APPENDIX J: UPDATING SCHMIDT COEFFICIENTS

The Schmidt coefficients  $\lambda_n$  must be updated each iteration of the VQE optimization. One way to do this is by treating them on the same footing as the gate parameters ( $\varphi$  in Fig. 2), such that the optimizer explores a space with dimension equal to the number of Schmidt coefficients plus the number of gate parameters. Here however, since the problems are small, we instead can and do remove the Schmidt coefficients from the optimizer search space by having the classical processor exactly minimize the energy as a function of the Schmidt coefficients based on the latest measurement results. This process is detailed below.

In each iteration of the VQE, quantum circuits are sampled to estimate the expectation values  $\langle b_n | \tilde{O}_i | b_n \rangle$  and

$\langle \phi_{b_n b_m}^p | \tilde{O}_i | \phi_{b_n b_m}^p \rangle$  in Eq. (3), where  $O_i$  here represents an  $N$ -qubit Pauli string. Computing the energy involves plugging these values into a product over  $i$  (indicating spin-up or -down), and summations over  $p, n, m$ , and the weighted list of  $2N$ -qubit Pauli strings defining the Hamiltonian. Deferring the operations over Schmidt-coefficient indices  $n, m$  and performing all others leaves

$$\langle H \rangle = \sum_{n,m} \lambda_n \lambda_m h_{nm}, \quad (\text{J1})$$

where  $h$  is a symmetric matrix of known values. By the variational principle, the eigenvector of this matrix with the lowest eigenvalue minimizes the energy. Thus we update  $\lambda$  by computing this eigenvector, such that the VQE optimizer need only minimize  $E(\varphi)$  rather than  $E(\varphi, \lambda)$ .

Although Schmidt coefficients are typically taken to be non-negative for definiteness, in this computation we allow them to take any real values, facilitating the exact-minimization update routine above and increasing ansatz flexibility for fixed  $U, V$ . In the main text, empirical  $\lambda$  values (Fig. 4) are represented by their absolute values for simplicity of discussion.

### APPENDIX K: NOISELESS CLASSICAL SIMULATION WITH LARGER ANSATZ

We execute an additional, noiseless classical simulation with a larger ansatz at the nominal equilibrium geometry (defined in the main text) to emphasize that the reduction in qubit number does not necessarily entail a loss of accuracy. Moreover, the  $U$  (and  $V$ ) part of the ansatz can be constructed exclusively from hop gates. We targeted a discrepancy of 1.6 mHa or less from the active-space FCI, as this value is sometimes quoted as a standard for quantum chemistry calculations. The ansatz included all ten bit strings consistent with assigning three electron pairs to the five molecular orbitals in the active space ( $k = 10$ ), and the gates are found via a heuristic empirical search using a noiseless classical simulator. The optimization converged to a ground-state energy of  $-75.726303$

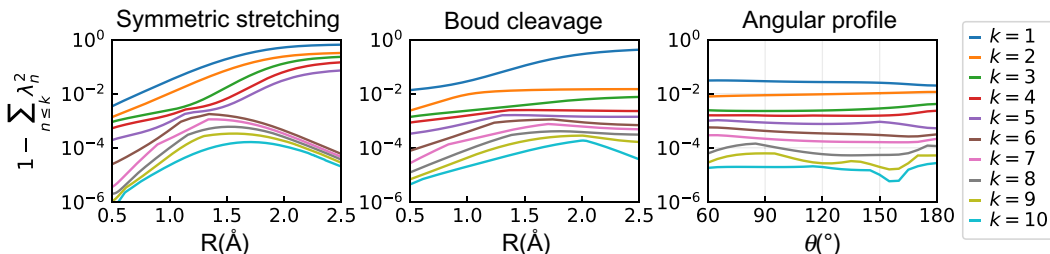


FIG. 6. Accuracy of truncated Schmidt decompositions. The three panels correspond to the three sweeps of molecular geometries as in Figs. 3, 4. For each choice of bit-string truncation  $k$ , we plot one minus the projection of the FCI wave function into the space defined by the leading  $k$  bit strings. In other words, we plot the sum of squares of the FCI Schmidt coefficients excluded by the truncation. Values near 1 (0) indicate that the FCI wave function cannot (can) be accurately represented in the truncated space. Unlike in the rest of the text, these FCI calculations do not freeze the oxygen  $2p$  orbital.

Ha, leaving a discrepancy of 1.47 mHa relative to the FCI value in this active space,  $-75.727775$  Ha. Explicitly, the  $N$ -qubit bit strings, including frozen orbitals, are 1111100, 1011101, 1011110, 1101110, 1101101, 1110110, 1110101, 1001111, 1010111, and 1100111. The

hop gates, listed in terms of the ordered orbital pair they acted upon (indexing from 0 to 6, such that here 0 and 4 are frozen), are given in the table. This ad hoc sequence of gates is presumably not optimal in terms of circuit depth.

	Orbitals	$\varphi$ (rad)		Orbitals	$\varphi$ (rad)
1	1,2	1.571 070 08e+00	18	3,5	1.799 008 33e-03
2	5,6	7.856 313 57e-01	19	3,5	0
3	2,6	0	20	1,3	-6.898 391 95e-02
4	1,3	-1.641 240 47e-01	21	2,6	2.452 420 43e-03
5	5,6	6.949 461 36e-01	22	2,6	0
6	2,3	0	23	1,2	1.205 658 41e-03
7	3,5	-1.326 983 09e-03	24	5,6	-1.233 294 17e-02
8	3,5	0	25	2,6	0
9	1,3	7.475 390 70e-02	26	1,3	1.684 458 71e-02
10	2,6	2.737 337 21e-04	27	5,6	-9.592 901 16e-03
11	2,6	0	28	2,3	0
12	1,2	8.049 942 86e-01	29	3,5	-1.113 789 80e-02
13	5,6	8.852 498 94e-01	30	3,5	0
14	2,6	0	31	1,3	7.224 900 09e-03
15	1,3	-1.010 798 74e+00	32	2,6	-8.256 569 96e-04
16	5,6	7.986 107 96e-01	33	2,6	0
17	2,3	0			

- [1] S. Bravyi, G. Smith, and J. A. Smolin, Trading Classical and Quantum Computational Resources, *Phys. Rev. X* **6**, 021043 (2016).
- [2] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating Large Quantum Circuits on a Small Quantum Computer, *Phys. Rev. Lett* **125**, 150504 (2020).
- [3] T. Yamazaki, S. Matsuura, A. Narimani, A. Saidmuradov, and A. Zaribafiyani, *ArXiv:1806.01305* (2018).
- [4] J. M. Kreula, L. García-Álvarez, L. Lamata, S. R. Clark, E. Solano, and D. Jaksch, Few-qubit quantum-classical simulation of strongly correlated lattice fermions, *EPJ Quantum Technol* **3**, 1 (2016).
- [5] N. C. Rubin, *ArXiv:1610.06910* (2016).
- [6] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid Quantum-Classical Approach to Correlated Materials, *Phys. Rev. X* **6**, 031045 (2016).
- [7] S. Bravyi and D. Gosset, Complexity of quantum impurity Problems, *Comm. Math. Phys* **356**, 451 (2017).
- [8] X. Yuan, J. Sun, J. Liu, Q. Zhao, and Y. Zhou, *ArXiv:2007.00958* (2020).
- [9] K. Mitarai and K. Fujii, Constructing a virtual two-qubit gate by sampling single-qubit operations, *New J. Phys* **23**, 023021 (2020).
- [10] Y. Kawashima, M. P. Coons, Y. Nam, E. Lloyd, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov, J. H. V. Nguyen, J. Kim, N. Alidoust, A. Zaribafiyani, and T. Yamazaki, *ArXiv:2102.07045* (2021).
- [11] W. Tang, T. Tomesh, M. Suchara, J. Larson, and M. Martonosi, Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (2021). doi:10.1145/3445814.3446758.
- [12] T. Ayril, F.-M. Le Regent, Z. Saleem, Y. Alexeev, and M. Suchara, 2020 IEEE Computer Society Annual Symposium on VLSI (ISVLSI) (2020). doi:10.1109/isvlsi49217.2020.00034.
- [13] S. E. Smart and D. A. Mazziotti, Quantum Solver of Contracted Eigenvalue Equations for Scalable Molecular Simulations on Quantum Computing Devices, *Phys. Rev. Lett.* **126**, 070504 (2021).
- [14] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys* **92**, 015003 (2020).
- [15] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: area laws for the entanglement entropy, *Rev. Mod. Phys* **82**, 277 (2010).
- [16] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys* **326**, 96 (2011).
- [17] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Variational quantum eigensolver with fewer qubits, *Phys. Rev. Res.* **1**, 023025 (2019).

- [18] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. Marianetti, Electronic structure calculations with dynamical mean-field theory, *Rev. Mod. Phys.* **78**, 865 (2006).
- [19] Q. Sun and G. K.-L. Chan, Quantum embedding Theories, *Acc. Chem. Res.* **49**, 2705 (2016).
- [20] L. Bytautas, T. M. Henderson, C. A. Jiménez-Hoyos, J. K. Ellis, and G. E. Scuseria, Seniority and orbital symmetry as tools for establishing a full configuration interaction hierarchy, *J. Chem. Phys.* **135**, 044119 (2011).
- [21] V. E. Elfving, M. Millaruelo, J. A. Gámez, and C. Gogolin, *ArXiv:2002.00035* (2020).
- [22] K. Gunst, D. V. Neck, P. Limacher, and S. D. Baerdemacker, The seniority quantum number in tensor network States, *SciPost Chem* **1**, 1 (2021).
- [23] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun* **5**, 1 (2014).
- [24] Y. Nam *et al.*, IBM Q experience as a versatile experimental testbed for simulating open quantum systems, *Npj Quantum Inf.* **6**, 1 (2020).
- [25] H. J. García, I. L. Markov, and A. W. Cross, On the geometry of stabilizer states, *Quantum Info. Comput* **14**, 683 (2014).
- [26] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver, *Phys. Rev. Lett.* **122**, 230401 (2019).
- [27] I. L. Markov and Y. Shi, Simulating quantum computation by contracting tensor Networks, *SIAM J. Comput* **38**, 963 (2008).
- [28] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Variational quantum simulation of general Processes, *Phys. Rev. Lett* **125**, 010501 (2020).
- [29] G. Greene-Diniz and D. Muñoz Ramo, Generalized unitary coupled cluster excitations for multireference molecular states optimized by the variational quantum eigensolver, *Int. J. Quantum Chem* **121**, e26352 (2021).
- [30] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *arXiv:2012.09265* [quant-ph] (2020).
- [31] G. Aleksandrowicz *et al.*, Qiskit: An Open-source Framework for Quantum Computing, (2019).
- [32] *ibmq\_dublin* v1.0.0-1.3.5, IBM Quantum Team (2020). Retrieved from <https://quantum-computing.ibm.com>.
- [33] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, *Johns Hopkins APL Tech. Dig.* **19**, 482 (1998).
- [34] K. Ito and T. Dhaene, Adaptive initial step size selection for simultaneous perturbation stochastic Approximation, *SpringerPlus* **5**, 200 (2016).
- [35] NIST Computational Chemistry Comparison and Benchmark Database (2020), doi:10.18434/T47C7Z.
- [36] A. W. Sandvik and G. Vidal, Variational Quantum Monte-Carlo Simulations with Tensor-Network States, *Phys. Rev. Lett* **99**, 220602 (2007).
- [37] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [38] J. Bermejo-Vega and M. Van Den Nest, Classical simulations of Abelian-group normalizer circuits with intermediate measurements, *Quantum Info. Comput* **14**, 181 (2014).
- [39] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley & Sons, New York, 2014).
- [40] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, Operator sampling for shot-frugal optimization in variational algorithms, *arXiv:2004.06252* [quant-ph] (2020).
- [41] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [42] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, *Nature* **567**, 491 (2019).
- [43] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum Circuits, *Phys. Rev. Lett* **119**, 180509 (2017).
- [44] R. LaRose, A. Mari, P. J. Karalekas, N. Shammah, and W. J. Zeng, *ArXiv:2009.04417*.
- [45] S. Niu and A. Todri-Sanial, *ArXiv:2102.05321*.
- [46] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay, and J. M. Gambetta, *ArXiv:2006.14044*.