

# Supplementary Information: Variational ansatz-based quantum simulation of imaginary time evolution

Sam McArdle,<sup>1</sup> Tyson Jones,<sup>1</sup> Suguru Endo,<sup>1</sup> Ying Li,<sup>2</sup> Simon C. Benjamin,<sup>1</sup> and Xiao Yuan<sup>1,\*</sup>

<sup>1</sup>*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*

<sup>2</sup>*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

(Dated: July 31, 2019)

## I. VARIATIONAL SIMULATION OF IMAGINARY TIME EVOLUTION

McLachlan's variational principle [1], applied to imaginary time evolution, is given by

$$\delta \|(\partial/\partial\tau + H - E_\tau) |\psi(\tau)\rangle\| = 0 \quad (1)$$

where

$$\|(\partial/\partial\tau + H - E_\tau) |\psi(\tau)\rangle\| = ((\partial/\partial\tau + H - E_\tau) |\psi(\tau)\rangle)^\dagger (\partial/\partial\tau + H - E_\tau) |\psi(\tau)\rangle, \quad (2)$$

and  $E_\tau = \langle\psi(\tau)|H|\psi(\tau)\rangle$ . For a general quantum state, McLachlan's variational principle recovers the imaginary time evolution

$$\frac{\partial |\psi(\tau)\rangle}{\partial\tau} = -(H - E_\tau) |\psi(\tau)\rangle. \quad (3)$$

If we consider a subspace of the whole Hilbert space, which can be reached using the ansatz  $|\phi(\tau)\rangle = |\phi(\theta_1, \theta_2, \dots, \theta_N)\rangle$ , we can project the imaginary time evolution onto the subspace using McLachlan's variational principle. Replacing  $|\psi(\tau)\rangle$  with  $|\phi(\tau)\rangle$ , yields

$$\begin{aligned} \|(\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle\| &= ((\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle)^\dagger (\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle, \\ &= \sum_{i,j} \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_j} \dot{\theta}_i \dot{\theta}_j + \sum_i \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} (H - E_\tau) |\phi(\tau)\rangle \dot{\theta}_i \\ &\quad + \sum_i \langle\phi(\tau)| (H - E_\tau) \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i} \dot{\theta}_i + \langle\phi(\tau)| (H - E_\tau)^2 |\phi(\tau)\rangle. \end{aligned} \quad (4)$$

Focusing on  $\dot{\theta}_i$ , we obtain

$$\begin{aligned} \frac{\partial \|(\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle\|}{\partial\dot{\theta}_i} &= \sum_j \left( \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_j} + \frac{\partial \langle\phi(\tau)|}{\partial\theta_j} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i} \right) \dot{\theta}_j \\ &\quad + \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} (H - E_\tau) |\phi(\tau)\rangle + \langle\phi(\tau)| (H - E_\tau) \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i}. \end{aligned} \quad (5)$$

Considering the normalisation condition for the trial state  $|\phi(\tau)\rangle$ ,

$$\langle\phi(\tau)|\phi(\tau)\rangle = 1, \quad (6)$$

we have

$$E_\tau \frac{\partial \langle\phi(\tau)|\phi(\tau)\rangle}{\partial\theta_i} = E_\tau \left( \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} |\phi(\tau)\rangle + \langle\phi(\tau)| \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i} \right) = 0, \quad (7)$$

and the derivative is simplified to

$$\frac{\partial \|(\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle\|}{\partial\dot{\theta}_i} = \sum_j A_{ij} \dot{\theta}_j - C_i. \quad (8)$$

---

\* xiao.yuan.ph@gmail.com

where

$$\begin{aligned} A_{ij} &= \Re \left( \frac{\partial \langle \phi(\tau) |}{\partial \theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial \theta_j} \right), \\ C_i &= -\Re \left( \frac{\partial \langle \phi(\tau) |}{\partial \theta_i} H |\phi(\tau)\rangle \right). \end{aligned} \quad (9)$$

McLachlan's variational principle requires

$$\frac{\partial \|(\partial/\partial\tau + H - E_\tau) |\phi(\tau)\rangle\|}{\partial \dot{\theta}_j} = 0, \quad (10)$$

which is equivalent to the differential equation of the parameters

$$\sum_j A_{ij} \dot{\theta}_j = C_i. \quad (11)$$

Denoting  $E(\tau) = \langle \phi(\tau) | H | \phi(\tau) \rangle$ , we can show that the average energy always decreases by following our imaginary time evolution algorithm, for a sufficiently small stepsize;

$$\begin{aligned} \frac{dE(\tau)}{d\tau} &= \Re \left( \langle \phi(\tau) | H \frac{d|\phi(\tau)\rangle}{d\tau} \right), \\ &= \sum_i \Re \left( \langle \phi(\tau) | H \frac{\partial |\phi(\tau)\rangle}{\partial \theta_i} \dot{\theta}_i \right), \\ &= -\sum_i C_i \dot{\theta}_i, \\ &= -\sum_i C_i A_{ij}^{-1} C_j, \\ &\leq 0. \end{aligned} \quad (12)$$

The third line follows from the definition of  $C_i$ ; the fourth line follows from the differential equation of  $\dot{\theta}$ ; the last line is true when  $A^{-1}$  is positive. First, we show matrix  $A$  is positive. We consider an arbitrary vector  $x = (x_1, x_2, \dots, x_N)^T$ , and calculate  $x^\dagger \cdot A \cdot x$ ,

$$\begin{aligned} x^\dagger \cdot A \cdot x &= \sum_{i,j} x_i^* A_{ij} x_j, \\ &= \sum_{i,j} x_i^* \Re \left( \frac{\partial \langle \phi(\tau) |}{\partial \theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial \theta_j} \right) x_j, \\ &= \sum_{i,j} x_i^* \frac{\partial \langle \phi(\tau) |}{\partial \theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial \theta_j} x_j + \sum_{i,j} x_i^* \frac{\partial \langle \phi(\tau) |}{\partial \theta_j} \frac{\partial |\phi(\tau)\rangle}{\partial \theta_i} x_j, \end{aligned} \quad (13)$$

Denote  $|\Phi\rangle = \sum_i x_i \frac{\partial |\phi(\tau)\rangle}{\partial \theta_i}$ , then the first term equals

$$\sum_{i,j} x_i^* \frac{\partial \langle \phi(\tau) |}{\partial \theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial \theta_j} x_j = \langle \Phi | \Phi \rangle \geq 0. \quad (14)$$

Similarly, we can show that the second term is also nonnegative. Therefore,  $x^\dagger \cdot A \cdot x \geq 0, \forall x$  and  $A$  is nonnegative. In practice, when  $A$  has eigenvalues with value zero,  $A$  is not invertible. However, in our simulation, we define the inverse of  $A$  to be only the inverse of the nonnegative eigenvalues. Suppose  $U$  is the transformation that diagonalises  $A$ , i.e.,  $G_{i,j} = (U A U^\dagger)_{i,j} = 0, \forall i \neq j$ . Then, we define  $G^{-1}$  by

$$G_{i,j}^{-1} = \begin{cases} \frac{1}{G_{i,j}} & i = j, G_{i,j} \neq 0, \\ 0 & i = j, G_{i,j} = 0, \\ 0 & i \neq j. \end{cases} \quad (15)$$

The inverse of  $A$  is thus defined by

$$A^{-1} = U^\dagger G^{-1} U. \quad (16)$$

Because  $A$  has nonnegative eigenvalues,  $G$ ,  $G^{-1}$ , and hence  $A^{-1}$  all have nonnegative eigenvalues.

## II. EVALUATING $A$ AND $C$ WITH QUANTUM CIRCUITS

In this section, we review the quantum circuit that can efficiently evaluate the coefficients  $A$  and  $C$  introduced in Ref. [2–4].

Without loss of generality, we can assume that each unitary gate  $U_i(\theta_i)$  in our circuit depends only on parameter  $\theta_i$  (since multiple parameter gates can be decomposed into this form). Suppose each  $U_i$  is a rotation or a controlled rotation gate, its derivative can be expressed by

$$\frac{\partial U_i(\theta_i)}{\partial \theta_i} = \sum_k f_{k,i} U_i(\theta_i) \sigma_{k,i}, \quad (17)$$

with unitary operator  $\sigma_{k,i}$  and scalar parameters  $f_{k,i}$ . The derivative of the trial state is

$$\frac{\partial |\phi(\tau)\rangle}{\partial \theta_i} = \sum_k f_{k,i} \tilde{V}_{k,i} |\bar{0}\rangle, \quad (18)$$

with

$$\tilde{V}_{k,i} = U_N(\theta_N) \dots U_{i+1}(\theta_{i+1}) U_i(\theta_i) \sigma_{k,i} \dots U_2(\theta_2) U_1(\theta_1). \quad (19)$$

In practice, there are only one or two terms,  $f_{k,i} \sigma_{k,i}$ , for each derivative. For example, when  $U_i(\theta_i)$  is a single qubit rotation  $R_{\theta_i}^Z = e^{-i\theta_i \sigma_z/2}$ , the derivative  $\partial U_i(\theta_i)/\partial \theta_i = -i/2 \times Z e^{-i\theta_i Z/2}$ , and the derivative of the trial state  $\partial |\phi(\tau)\rangle/\partial \theta_i$  can be prepared by adding an extra  $Z$  gate with a constant factor  $-i/2$ . When  $U_i(\theta_i)$  is a control rotation such as  $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes R_{\theta_i}^Z$ , the derivative  $\partial U_i(\theta_i)/\partial \theta_i = |1\rangle\langle 1| \otimes \partial R_{\theta_i}^Z/\partial \theta_i = -i/2 \times |1\rangle\langle 1| \otimes Z e^{-i\theta_i Z/2}$ . By choosing  $\sigma_{1,i} = I \otimes Z$ ,  $\sigma_{2,i} = Z \otimes Z$ ,  $f_{1,i} = -i/4$ , and  $f_{2,i} = i/4$ , we can show Eq. (17).

Therefore, the coefficients  $A_{ij}$  and  $C_i$  are given by

$$\begin{aligned} A_{ij} &= \Re \left( \sum_{k,l} f_{k,i}^* f_{l,j} \langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle \right), \\ C_i &= \Re \left( \sum_{k,l} f_{k,i}^* \lambda_l \langle \bar{0} | \tilde{V}_{k,i}^\dagger h_l V | \bar{0} \rangle \right). \end{aligned} \quad (20)$$

All the terms of the summation follow the general form  $a \Re(e^{i\theta} \langle \bar{0} | U | \bar{0} \rangle)$  and can be evaluated with a quantum circuit.

In practice, we do not need to realize the whole controlled-U gate and can instead use a much easier circuit. For example, for the term  $\Re(f_{k,i}^* f_{l,j} \langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle)$ , we can let  $f_{k,i}^* f_{l,j} = a e^{i\theta}$  and

$$\langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle = \langle \bar{0} | U_1^\dagger \dots U_{i-1}^\dagger \sigma_{k,i}^\dagger U_i^\dagger \dots U_N^\dagger U_N \dots U_j \sigma_{l,j} U_{j-1} \dots U_1 | \bar{0} \rangle. \quad (21)$$

Suppose  $i < j$ , then

$$\langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle = \langle \bar{0} | U_1^\dagger \dots U_{i-1}^\dagger \sigma_{k,i}^\dagger U_i^\dagger \dots U_{j-1}^\dagger \sigma_{l,j} U_{j-1} \dots U_i \dots U_1 | \bar{0} \rangle, \quad (22)$$

and  $\Re(e^{i\theta} \langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle)$  can be measured by the circuit in Fig. S1. The terms for  $C$  can be measured similarly.

## III. COMPUTATIONAL CHEMISTRY BACKGROUND

One of the central problems in computational chemistry is finding the ground state energy of molecules. This calculation is classically intractable, due to the exponential growth of Hilbert space with the number of electrons in the molecule. However, it has been shown that quantum computers are able to solve this problem efficiently [5]. The Hamiltonian of a molecule consisting of  $M$  nuclei (of mass  $M_I$ , position  $\mathbf{R}_I$ , and charge  $Z_I$ ) and  $N$  electrons (with position  $\mathbf{r}_i$ ) is

$$H = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_{i,I} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2} \sum_{I \neq J} \frac{e^2}{4\pi\epsilon_0} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \quad (23)$$

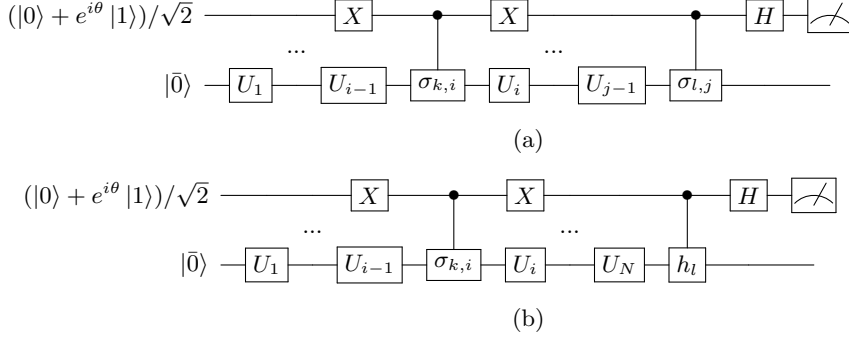


FIG. S1. Quantum circuits that evaluate (a)  $\Re(e^{i\theta} \langle \bar{0} | \tilde{V}_{k,i}^\dagger \tilde{V}_{l,j} | \bar{0} \rangle)$  and (b)  $\Re(e^{i\theta} \langle \bar{0} | \tilde{V}_{k,i}^\dagger h_l V | \bar{0} \rangle)$ . When  $\sigma_{k,i}$  is Hermitian, the  $X$  gates acting on the ancilla qubit can be also omitted.

Because the nuclei are orders of magnitude more massive than the electrons, we apply the Born-Oppenheimer approximation, and treat the nuclei as classical, fixed point charges. After this approximation, the eigenvalue equation we seek to solve (in atomic units) is given by

$$\left[ -\sum_i \frac{\nabla_i^2}{2} - \sum_{i,I} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] |\psi\rangle = E |\psi\rangle, \quad (24)$$

where  $|\psi\rangle$  is an energy eigenstate of the Hamiltonian, with energy eigenvalue  $E$ .

To solve this problem using a quantum computer, we first transform it into the second quantised form. We project the Hamiltonian onto a finite number of basis wave functions,  $\{\phi_p\}$ , which approximate spin-orbitals. Electrons are excited into, or de-excited out of, these spin-orbitals by fermionic creation ( $a_p^\dagger$ ) or annihilation ( $a_p$ ) operators, respectively. These operators obey fermionic anti-commutation relations, which enforces the antisymmetry of the wavefunction, a consequence of the Pauli exclusion principle. In the second quantised representation, the electronic Hamiltonian is written as

$$H = \sum_{p,q} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{p,q,r,s} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \quad (25)$$

with

$$h_{pq} = \int d\mathbf{x} \phi_p^*(\mathbf{x}) \left( -\frac{\nabla^2}{2} - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \right) \phi_q(\mathbf{x}), \quad (26)$$

$$h_{pqrs} = \int d\mathbf{x}_1 d\mathbf{x}_2 \frac{\phi_p^*(\mathbf{x}_1) \phi_q^*(\mathbf{x}_2) \phi_s(\mathbf{x}_1) \phi_r(\mathbf{x}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

where  $\mathbf{x}$  is a spatial and spin coordinate. This Hamiltonian in a molecular orbital basis contains  $O(N_{\text{SO}}^4)$  terms, where  $N_{\text{SO}}$  is the number of spin-orbitals considered. This fermionic Hamiltonian must then be transformed into a Hamiltonian acting on qubits. This can be achieved using the Jordan-Wigner (JW), or Bravyi-Kitaev (BK) transformations, which are described in Ref. [6].

## A. Hydrogen

In our simulations, we consider the Hydrogen molecule in the minimal STO-3G basis. This means that only the minimum number of orbitals to describe the electrons are considered. ‘STO- $n$ G’ means that a linear combination of  $n$  Gaussian functions are used to approximate a Slater-type-orbital, which describes the electron wavefunction. Each Hydrogen atom contributes a single  $1S$  orbital. As a result of spin, there are four spin-orbitals in total. We are able

to construct the molecular orbitals for  $H_2$  by manually (anti)symmetrising the spin-orbitals. These are

$$\begin{aligned}
 |\phi_0\rangle &= |\sigma_{g\uparrow}\rangle = \frac{1}{\sqrt{\mathcal{N}_G}}(|1S_{1\uparrow}\rangle + |1S_{2\uparrow}\rangle), \\
 |\phi_1\rangle &= |\sigma_{g\downarrow}\rangle = \frac{1}{\sqrt{\mathcal{N}_G}}(|1S_{1\downarrow}\rangle + |1S_{2\downarrow}\rangle), \\
 |\phi_2\rangle &= |\sigma_{u\uparrow}\rangle = \frac{1}{\sqrt{\mathcal{N}_U}}(|1S_{1\uparrow}\rangle - |1S_{2\uparrow}\rangle), \\
 |\phi_3\rangle &= |\sigma_{u\downarrow}\rangle = \frac{1}{\sqrt{\mathcal{N}_U}}(|1S_{1\downarrow}\rangle - |1S_{2\downarrow}\rangle),
 \end{aligned} \tag{27}$$

where the subscripts on the  $1S$  orbitals denote the spin of the electron in that orbital, and which of the two hydrogen atoms the orbital is centred on, and  $\mathcal{N}_{G/U}$  are normalisation coefficients. By following the procedure in Ref. [6], the qubit Hamiltonian for  $H_2$  in the BK representation can be obtained. This 4 qubit Hamiltonian is given by

$$\begin{aligned}
 H = & h_0I + h_1Z_0 + h_2Z_1 + h_3Z_2 + h_4Z_0Z_1 + h_5Z_0Z_2 + h_6Z_1Z_3 + h_7X_0Z_1X_2 + h_8Y_0Z_1Y_2 \\
 & + h_9Z_0Z_1Z_2 + h_{10}Z_0Z_2Z_3 + h_{11}Z_1Z_2Z_3 + h_{12}X_0Z_1X_2Z_3 + h_{13}Y_0Z_1Y_2Z_3 + h_{14}Z_0Z_1Z_2Z_3.
 \end{aligned} \tag{28}$$

As this Hamiltonian only acts off diagonally on qubits 0 and 2 [7, 8], it can be reduced to

$$H = g_0I + g_1Z_0 + g_2Z_1 + g_3Z_0Z_1 + g_4Y_0Y_1 + g_5X_0X_1, \tag{29}$$

which only acts on two qubits.

In our work, we consider an internuclear distance of  $R = 0.75 \text{ \AA}$  and hence  $g_0 = 0.2252$ ,  $g_1 = 0.3435$ ,  $g_2 = -0.4347$ ,  $g_3 = 0.5716$ ,  $g_4 = 0.0910$ ,  $g_5 = 0.0910$ . We make use of the universal ansatz [9] shown in Fig. S2.

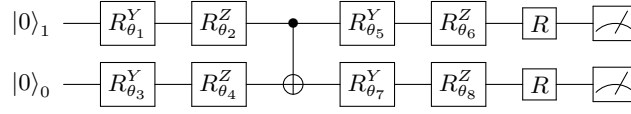


FIG. S2. The quantum circuit for preparing the hardware efficient ansatz with eight parameters. Here each  $R^Y_{\theta} = e^{-i\theta\sigma_Y/2}$ ,  $R^Z_{\theta} = e^{-i\theta\sigma_Z/2}$ , and the two  $R$  gates correspond to the rotation of the measurement.

We have eight parameters and an example circuit for measuring  $A_{2,7}$  is shown in Fig. S3

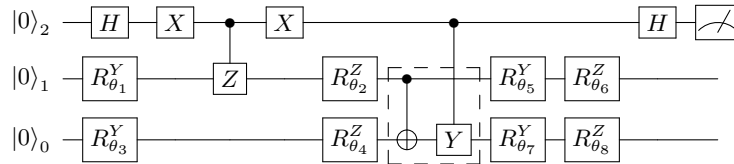


FIG. S3. The circuit to measure  $A_{2,7} = \Re\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_2}\frac{\partial\langle\phi(\tau)|}{\partial\theta_7}\right)$ . In practice, the gates in the dashed box may be omitted. The other terms of  $A$  and  $C$  can be measured using similar circuits.

## B. Lithium Hydride

In our simulations, we consider Lithium Hydride in the minimal STO-3G basis. The Lithium atom has 3 electrons, and so contributes a  $1S, 2S, 2P_x, 2P_y$  and  $2P_z$  orbital to the basis, while the Hydrogen atom contributes a single  $1S$  orbital. With spin, this makes 12 spin-orbitals in total. However, we are able to reduce the number of spin-orbitals required by considering their expected occupation. This reduces the qubit resources required for our calculation. In computational chemistry, the subset of spin-orbitals included in a calculation is called the active space.

We first obtain the one electron reduced density matrix (1-RDM) for LiH, using a classically tractable CISD (configuration interaction, single and double excitations) calculation. The 1-RDM for a distance of 1.45 Å is shown below,

$$\begin{pmatrix} 1.99991 & -0.00047 & 0.00047 & 0 & 0 & -0.00120 \\ -0.00047 & 1.95969 & 0.06691 & 0 & 0 & 0.00842 \\ 0.00047 & 0.06691 & 0.00968 & 0 & 0 & -0.01385 \\ 0 & 0 & 0 & 0.00172 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00172 & 0 \\ -0.00120 & 0.00842 & -0.01385 & 0 & 0 & 0.02728 \end{pmatrix}. \quad (30)$$

There are only six rows and columns in the 1-RDM because the spin-up and spin-down orbitals have been combined. The diagonal elements of the 1-RDM are the occupation numbers of the corresponding canonical orbitals (the Hartree-Fock orbitals). In order to reduce our active space, we first perform a unitary rotation of the 1-RDM, such that it becomes a diagonal matrix,

$$\begin{pmatrix} 1.99992 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.96201 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.03459 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.00005 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00172 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.00172 \end{pmatrix}. \quad (31)$$

This gives the 1-RDM in terms of natural molecular orbitals (NMOs). The diagonal entries are called the natural orbital occupation numbers (NOONs). The Hamiltonian of LiH must also be rotated, using the same unitary matrix used to diagonalise the 1-RDM. This is equivalent to performing a change of basis, from the canonical orbital basis to the natural molecular orbital basis.

As can be seen, the first orbital has a NOON close to two, and so is very likely to be doubly occupied. As a result, we ‘freeze’ this core orbital, and consider it to always be doubly filled. We can then remove any terms containing  $a_0^\dagger, a_0, a_1^\dagger, a_1$  from the LiH fermionic Hamiltonian, where 0 and 1 denote the spin-orbitals that correspond to the core spatial orbital. We also notice that the fourth spatial orbital has a NOON close to zero. As a result, we assume that this orbital is never occupied by either a spin-up or spin-down electron, and so remove another two fermionic spin-orbital operators from the Hamiltonian. This leaves a fermionic Hamiltonian acting on 8 spin-orbitals. We then map this fermionic Hamiltonian to a qubit Hamiltonian, using the JW transformation. All of these steps were carried out using OpenFermion [10], an electronic structure package to transform computational chemistry problems into a form that is suitable for investigation using a quantum computer.

### 1. LiH ansatz

We used an ansatz inspired by the low depth circuit ansatz (LDCA) [4]. This ansatz is both hardware efficient (in the sense that it only uses nearest neighbour gates), and chemically motivated. The specific ansatz used for simulating LiH is shown in Fig. S4.

### 2. Numerical simulation

We simulate the aforementioned quantum circuits using the Quantum Exact Simulation Toolkit (QuEST) [11]. True ground states are found by diagonalising the considered Hamiltonians with the GNU Scientific Library (GSL) [12], which employs a complex form of the symmetric bidiagonalisation and QR reduction method [13].

While our H<sub>2</sub> tests simulate the full experimental routine, our LiH tests are optimised by individually computing each ansatz derivative  $\partial|\psi(\vec{\theta})\rangle/\partial\theta_j$ . While obtaining this wavefunction from an experiment would require a number of measurements that grows exponentially with the number of qubits, we can access it directly in our numerical simulations. The observables of the experimental routine are then directly calculated via inner products of these derivative wavefunctions. This allows us to populate the  $A$  matrix with only  $N_p$  evaluations of the ansatz circuit, in contrast to the  $N_p^2$  evaluations involved in a full experimental routine. Similarly, the state  $\hat{H}|\psi(\theta)\rangle$  can be computed once each iteration, and the  $C$  vector populated via the inner product of  $\hat{H}|\psi(\theta)\rangle$  with the derivative wavefunctions. This reduces the number of simulated gates by  $N_p N_H$  from the full experimental routine described in the main text.

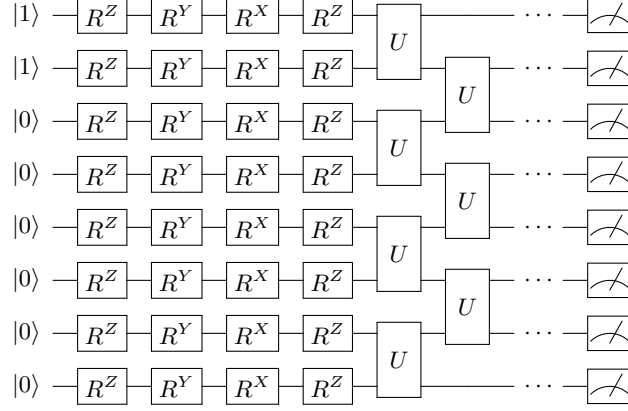


FIG. S4. The general structure of the ansatz used in our LiH simulations. We repeat the circuit structure of the two-qubit rotations block to depth  $M = 3$ . The form of the  $U$  gate is  $U = e^{i\alpha YX} e^{i\beta XY} e^{i\gamma ZZ} e^{i\delta YY} e^{i\epsilon XX}$ . In total there are  $(3 \times 5 \times 7) + (4 \times 8) = 137$  parameters.

After populating  $A$  and  $C$ , we then update the parameters under the variational imaginary time evolution scheme described in the main text. Since  $A\dot{\vec{\theta}} = C\delta t$  is generally underdetermined and leaves us unable to invert  $A$ , we instead update the parameters under Tikhonov regularisation, which minimises

$$\|C - A\dot{\vec{\theta}}\|^2 + \lambda\|\dot{\vec{\theta}}\|^2. \quad (32)$$

Here, the Tikhonov parameter  $\lambda$  can be varied to trade accuracy for keeping  $\dot{\vec{\theta}}$  small and the parameter evolution smooth. We estimate an ideal  $\lambda$  at each time-step by selecting the corner of a 3-point L-curve [12, 13], though force  $\lambda \in [10^{-4}, 10^{-2}]$ . This is because too large a  $\lambda$  over-restricts the change in the parameters and was seen to lead to eventual convergence to non-ground states. Meanwhile, no regularisation ( $\lambda = 0$ ) saw residuals in  $A^{-1}$  disrupt the monotonic decrease in energy. Using Tikhonov regularisation affords us a larger time-step than other tested methods, which included LU decomposition, least squares minimisation, singular value decomposition (SVD) and truncated SVD.

We use a basic model of how shot noise and gate errors affect the variational algorithm, by modifying the elements in  $A$ , and  $C$ . We here illustrate the procedure for an element of  $A$ , which is found by  $N_A$  samples of a binary-outcome measurement with expected value

$$a_{ij} = \Re \left( \frac{\partial \langle \psi(\vec{\theta}) |}{\partial \theta_i} \frac{\partial | \psi(\vec{\theta}) \rangle}{\partial \theta_j} \right) \in \left[ -\frac{1}{4}, \frac{1}{4} \right]. \quad (33)$$

For sufficiently many samples  $N_A$ , the central limit theorem informs  $A_{ij}$  is normally distributed.

$$A_{ij} \sim N \left( a_{ij}, \frac{1}{N_A} (1/16 - a_{ij}^2) \right). \quad (34)$$

We model the effect of decoherence, viewed as a mixing of each underlying measurement distribution with the fully mixed state, as a skewing of observable expectation values toward their centre (zero). That is,  $a_{ij} \rightarrow \epsilon a_{ij}$  for some  $\epsilon \in [0, 1]$ . We assume an experimental error rate of  $10^{-4}$  per gate, and approximate the state after  $D$  gates to differ from the noise-free state  $\rho$  as

$$\rho_{\text{final}} = (1 - 10^{-4})^D \rho + \left( 1 - (1 - 10^{-4})^D \right) \mathbb{1}. \quad (35)$$

For our LiH simulations  $D \approx 100$ , so the effect of gate error is to skew our expected values by  $\epsilon \approx 0.99$ . We follow a similar procedure to introduce noise into the other components of the simulation.

---

[1] A. McLachlan, Molecular Physics **8**, 39 (1964).

- [2] J. Romero, R. Babbush, J. McClean, C. Hempel, P. Love, and A. Aspuru-Guzik, *Quantum Science and Technology* (2018).
- [3] Y. Li and S. C. Benjamin, *Phys. Rev. X* **7**, 021050 (2017).
- [4] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, arXiv preprint arXiv:1801.01053 (2018).
- [5] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, *Science* **309**, 1704 (2005).
- [6] J. T. Seeley, M. J. Richard, and P. J. Love, *The Journal of Chemical Physics* **137**, 224109 (2012), <https://doi.org/10.1063/1.4768229>.
- [7] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, *Phys. Rev. X* **6**, 031007 (2016).
- [8] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Nature* **549**, 242 (2017).
- [9] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *Nature communications* **5** (2014).
- [10] J. R. McClean, I. D. Kivlichan, D. S. Steiger, Y. Cao, E. S. Fried, C. Gidney, T. Häner, V. Havlíček, Z. Jiang, M. Neeley, *et al.*, arXiv preprint arXiv:1710.07629 (2017).
- [11] T. Jones, A. Brown, I. Bush, and S. Benjamin, "Quest and high performance simulation of quantum computers," (2018), arXiv:1802.08032.
- [12] Contributors and GSL Project, "GSL - GNU scientific library - GNU project - free software foundation (FSF)," <http://www.gnu.org/software/gsl/> (2010).
- [13] G. Golub and C. V. Loan, *Matrix computations*, Vol. 83 (Cambridge University Press, 1999) p. 556.