Variational ansatz-based quantum simulation of imaginary time evolution

Sam McArdle,[1] Tyson Jones,[1] Suguru Endo,[1] Ying Li,[2] Simon Benjamin,[1] and Xiao Yuan[1, *]

[1]*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*
[2]*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

Imaginary time evolution is a powerful tool for studying quantum systems. While it is possible to simulate with a classical computer, the time and memory requirements generally scale exponentially with the system size. Conversely, quantum computers can efficiently simulate quantum systems, but not non-unitary imaginary time evolution. We propose a variational algorithm for simulating imaginary time evolution on a hybrid quantum computer. We use this algorithm to find the ground state energy of many-particle systems; specifically molecular Hydrogen and Lithium Hydride, finding the ground state with high probability. Our method can also be applied to general optimisation problems and quantum machine learning. As our algorithm is hybrid, suitable for error mitigation, and can exploit shallow quantum circuits, it can be implemented with current quantum computers.

INTRODUCTION

Imaginary time is an unphysical, yet powerful, mathematical concept. It has been utilised in numerous physical domains including: quantum mechanics, statistical mechanics, and cosmology. Often referred to as performing a 'Wick rotation' [1], replacing real time with imaginary time connects Euclidean and Minkowski space [2], quantum and statistical mechanics [3], and static problems to problems of dynamics [4]. In quantum mechanics, propagating a wavefunction in imaginary time enables: the study of finite temperature properties [5–7], finding the ground state wavefunction and energy (such as in density matrix renormalisation group) [8–11], and simulating real time dynamics (such as time dependent Hartree) [12, 13]. For a system with Hamiltonian, H, evolving in real time, t, the propagator is given by e^{-iHt}. The corresponding propagator in imaginary time, $\tau = it$, is given by $e^{-H\tau}$; a non-unitary operator.

Using a classical computer, we can simulate imaginary time evolution by evaluating the propagator and applying it to the system wavefunction. There also exist various related classical methods, such as quantum Monte Carlo [14, 15] and density matrix renormalization group [16, 17] for solving different problems. However, because the dimension of the wavefunction grows exponentially with the number of particles, classical simulation of many-body quantum systems is generally hard [18]. While efficient variational trial states have been developed for a number of applications [19], powerful trial wavefunctions typically require classical computational resources which scale exponentially with the system size [11].

Quantum computing can naturally and efficiently store many-body quantum states, and hence is suitable for simulating quantum systems [20]. We can map the system Hamiltonian to a qubit Hamiltonian, and simulate real time evolution (as described by the Schrödinger equation) by realising the corresponding unitary evolution with a quantum circuit [21]. Using Trotterization [22], the real time propagator can be decomposed into a sequence of single and two qubit gates [23]. The ability to represent the real time propagator with a sequence of gates stems from its unitarity. In contrast, because the imaginary time operator is non-unitary, it is not straightforward to decompose it into a sequence of unitary gates using Trotterization, and thus directly realise it with a quantum circuit. As a result, alternative methods are required to implement imaginary time evolution using a quantum computer.

Classically, we can simulate real (imaginary) time evolution of parametrised trial states by repeatedly solving the (Wick-rotated) Schrödinger equation over a small timestep, and updating the parameters for the next timestep [8, 9, 11, 24–27]. This method has recently been extended to quantum computing, where it was used to simulate real time dynamics [28]. Closely related are the variational quantum eigensolver (VQE) [29–35] and the quantum approximate optimisation algorithm (QAOA) [36], which update the parameters using a classical optimisation routine, to find the minimum energy eigenvalue of a given Hamiltonian. As 'hybrid quantum-classical methods', these algorithms use a small quantum computer to carry out a classically intractable subroutine, and a classical computer to solve the higher level problem. The quantum subroutine may only require a small number of qubits and a low depth circuit, presenting a potential use for noisy intermediate-scale quantum hardware [37].

In this paper, we propose a method to simulate imaginary time evolution on a quantum computer, using a hybrid quantum-classical variational algorithm. The proposed method thus combines the power of quantum computers to efficiently represent many-body quantum states, with classical computers' ability to simulate arbitrary (including unphysical) processes. We discuss using this method to find the ground state energy of many-body quantum systems, and to solve optimisation problems. We then numerically test the performance of our algorithm at finding the ground state energy of both the Hydrogen molecule (H_2) and Lithium Hydride (LiH). We compare our results for LiH to those obtained

using the VQE with gradient descent. As our algorithm only requires a low depth circuit, it can be realised with current and near-term quantum processors.

RESULTS

Variational imaginary time evolution. We focus on many-body systems that are described by Hamiltonians $H = \sum_i \lambda_i h_i$, with real coefficients, λ_i, and observables, h_i, that are tensor products of Pauli matrices. We assume that the number of terms in this Hamiltonian scales polynomially with the system size, which is true for many physical systems, such as molecules or the Fermi-Hubbard model. Given an initial state $|\psi\rangle$, the normalised imaginary time evolution is defined by

$$|\psi(\tau)\rangle = A(\tau)e^{-H\tau}|\psi(0)\rangle, \qquad (1)$$

where $A(\tau) = 1/\sqrt{\langle\psi(0)|\,e^{-2H\tau}\,|\psi(0)\rangle}$ is a normalisation factor. In the instance that the initial state is a maximally mixed state, the state at time τ is a thermal or Gibbs state $\rho_{T=1/\tau} = e^{-H\tau}/\text{Tr}[e^{-H\tau}]$, with temperature $T = 1/\tau$. When the initial state has a non-zero overlap with the ground state, the state at $\tau \to \infty$ is the ground state of H. Equivalently, the Wick rotated Schrödinger equation is,

$$\frac{\partial|\psi(\tau)\rangle}{\partial\tau} = -(H - E_\tau)|\psi(\tau)\rangle, \qquad (2)$$

where the term $E_\tau = \langle\psi(\tau)|H|\psi(\tau)\rangle$ results from enforcing normalisation. Even if $|\psi(\tau)\rangle$ can be represented by a quantum computer, the non-unitary imaginary time evolution cannot be naively mapped to a quantum circuit.

In our variational method, instead of directly encoding the quantum state $|\psi(\tau)\rangle$ at time τ, we approximate it using a parametrised trial state $|\phi(\vec{\theta}(\tau))\rangle$, with $\vec{\theta}(\tau) = (\theta_1(\tau), \theta_2(\tau), \ldots, \theta_N(\tau))$. This stems from the intuition that the physically relevant states are contained in a small subspace of the full Hilbert space [38]. The trial state is referred to as the ansatz. In condensed matter physics and computational chemistry, a wide variety of ansätze have been proposed for both classical and quantum variational methods [11, 20, 39, 40].

Using a quantum circuit, we prepare the trial state, $|\phi(\vec{\theta})\rangle$, by applying a sequence of parametrised unitary gates, $V(\vec{\theta}) = U_N(\theta_N)\ldots U_k(\theta_k)\ldots U_1(\theta_1)$ to our initial state, $|\bar{0}\rangle$. We express this as $|\phi(\vec{\theta})\rangle = V(\vec{\theta})|\bar{0}\rangle$ and remark that $V(\vec{\theta})$ is also referred to as the ansatz. We refer to all possible states that could be created by the circuit V as the 'ansatz space'. Here, $U_k(\theta_k)$ is the kth unitary gate, controlled by parameter θ_k, and the gate can be regarded as a single or two qubit gate.

To simulate the imaginary time evolution of the trial state, we use McLachlan's variational principle [41, 42],

$$\delta\|(\partial/\partial\tau + H - E_\tau)|\psi(\tau)\rangle\| = 0, \qquad (3)$$

where $\|\rho\| = \text{Tr}[\sqrt{\rho\rho^\dagger}]$ denotes the trace norm of a state. By replacing $|\psi(\tau)\rangle$ with $|\phi(\tau)\rangle = |\phi(\vec{\theta}(\tau))\rangle$, we effectively project the desired imaginary time evolution onto the manifold of the ansatz space. The evolution of the parameters is obtained from the resulting differential equation

$$\sum_j A_{ij}\dot{\theta}_j = C_i, \qquad (4)$$

where

$$A_{ij} = \Re\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}\frac{\partial|\phi(\tau)\rangle}{\partial\theta_j}\right),$$
$$C_i = \Re\left(-\sum_\alpha \lambda_\alpha \frac{\partial\langle\phi(\tau)|}{\partial\theta_i}h_\alpha|\phi(\tau)\rangle\right), \qquad (5)$$

and h_α and λ_α are the Pauli terms and coefficients of the Hamiltonian, as described above. The derivation of Eq. (4) can be found in the Supplementary Materials. As both A_{ij} and C_i are real, the derivative $\dot{\theta}_j$ is also real, as required for parametrising a quantum circuit. Interestingly, although the average energy term E_τ appears in Eq. (2), it does not appear in Eq. (4). This is because the ansatz applied maintains normalisation, as it is composed of unitary operators.

Imaginary time evolution with quantum circuits. By following a similar method to that introduced in Ref. [28], we can efficiently measure A_{ij} and C_i using a quantum computer. We assume that the derivative of a unitary gate $U_i(\theta_i)$ can be expressed as $\partial U_i(\theta_i)/\partial\theta_i = \sum_k f_{k,i}U_i(\theta_i)\sigma_{k,i}$, with unitary operator $\sigma_{k,i}$. The derivative of the trial state is given by $\partial|\phi(\tau)\rangle/\partial\theta_i = \sum_k f_{k,i}\tilde{V}_{k,i}|\bar{0}\rangle$, with $\tilde{V}_{k,i} = U_N(\theta_N)\ldots U_{i+1}(\theta_{i+1})U_i(\theta_i)\sigma_{k,i}\ldots U_1(\theta_1)$. There are typically only one or two terms resulting from each derivative. As an example, when $U_i(\theta_i)$ is a single qubit rotation $R_z(\theta_i) = e^{-i\theta_i\sigma_z/2}$, the derivative $\partial U_i(\theta_i)/\partial\theta_i = -i/2 \times \sigma_z e^{-i\theta_i\sigma_z/2}$. The coefficients A_{ij} and C_i are given by

$$A_{ij} = \Re\left(\sum_{k,l} f_{k,i}^* f_{l,j}\langle\bar{0}|\tilde{V}_{k,i}^\dagger\tilde{V}_{l,j}|\bar{0}\rangle\right),$$
$$C_i = \Re\left(\sum_{k,\alpha} f_{k,i}^*\lambda_\alpha\langle\bar{0}|\tilde{V}_{k,i}^\dagger h_\alpha V|\bar{0}\rangle\right). \qquad (6)$$

All of these terms are of the form $a\Re(e^{i\theta}\langle\bar{0}|U|\bar{0}\rangle)$ and can be evaluated using the circuits shown in the Supplementary Materials.

With $A(\tau)$ and $\vec{C}(\tau)$ at time τ, the imaginary time evolution over a small interval $\delta\tau$ can be simulated by evaluating $\dot{\vec{\theta}}(\tau) = A^{-1}(\tau)\cdot\vec{C}(\tau)$, and using a suitable

update rule, such as the Euler method,

$$\vec{\theta}(\tau + \delta\tau) \simeq \vec{\theta}(\tau) + \dot{\vec{\theta}}(\tau)\delta\tau = \vec{\theta}(\tau) + A^{-1}(\tau) \cdot \vec{C}(\tau)\delta\tau. \tag{7}$$

By repeating this process $N_T = \tau_{total}/\delta\tau$ times, we can simulate imaginary time evolution over a duration τ_{total}. Often, the satisfying parameter evolution is not unique and Eq. (4) is underdetermined. In that case, we can employ truncated singular value decomposition to approximately invert A, or Tikhonov regularisation to additionally constrain the parameters to vary smoothly. We elaborate upon these strategies in the Supplementary Materials.

A limitation of our variational method is that the ansatz may not be able to faithfully describe all states on the desired trajectory, much like its real time counterpart [28]. Even though such states lie in a small subspace of the full Hilbert space [38], it is difficult to prove that they can be generated by a given ansatz, despite promising numerical results [28]. However, our numerical results are similarly promising for imaginary time, and demonstrate it to be a robust routine for energy minimisation. Moreover, we believe that when tasked with finding the ground state using imaginary time evolution, a small deviation from the true evolution is less problematic than when trying to simulate real time evolution. This is because imaginary time evolution always drives a state towards the ground state (or one of the lowest eigenstates), whereas the real time evolution of two closely separated states may be very different. Consequently, as long as errors due to an imperfect ansatz do not cause the simulation to become trapped in local minima, we do not mind if the evolution deviates from the path of true imaginary time evolution, as ultimately, it will still be driven towards the ground state. Nevertheless, designing ansätze that are well suited to imaginary time evolution is an interesting open problem.

Ground state energy via imaginary time evolution. We apply our method to the problem of finding the ground state energy of a many-body Hamiltonian, H. As with the VQE, our goal is to find the values of the parameters, $\vec{\theta}$, which minimise the expectation value of the Hamiltonian

$$E_{\min} = \min_{\vec{\theta}} \langle\phi(\vec{\theta})| H |\phi(\vec{\theta})\rangle, \tag{8}$$

where $|\phi(\vec{\theta})\rangle = V(\vec{\theta}) |\bar{0}\rangle$ is our variational trial state. The VQE solves this problem by using a quantum computer to construct a good ansatz and measure the expectation value of the Hamiltonian, and a classical optimisation routine to obtain new values of the parameters. In order to preserve the exponential speedup of the VQE over classical methods, the trial state is constructed using a number of parameters that scales polynomially with the

system size. However, because we may need to consider many possible values for each parameter, the total size of the parameter space still scales exponentially with the system size. Moreover, many optimisation algorithms, such as gradient descent, are liable to becoming trapped in local minima. This combination can make the classical optimisation step of the VQE very difficult [43].

As described above, if the initial state has a non-zero overlap with the ground state, true propagation in imaginary time will evolve the system into the ground state, in the limit that $\tau \to \infty$. Classically, this has been leveraged as a powerful tool to find the ground state energy of quantum systems [8, 9, 11]. Using our method, we can efficiently simulate ansatz-based imaginary time evolution to find the ground state, using a quantum computer. In the numerical simulations described below, we use the Euler method to solve differential equations, which corresponds to the update rule for the parameters shown in Eq. (7). We prove in the Supplementary Materials that when $\delta\tau$ is sufficiently small, the average energy of the trial state, $E(\tau) = \langle\phi(\tau)| H |\phi(\tau)\rangle$, always decreases when following the Euler update rule: $E(\tau + \delta\tau) \leq E(\tau)$.

In this work, we consider gradient descent, a canonical classical optimisation method

$$\vec{\theta}(\tau + \delta\tau) = \vec{\theta}(\tau) + \vec{G}(\tau)\delta\tau = \vec{\theta}(\tau) + \vec{C}(\tau)\delta\tau, \tag{9}$$

where $\vec{G}(\tau) = -\nabla E(\tau)$ is the gradient of $E(\tau)$ and $\vec{C}(\tau) \equiv -\nabla E(\tau)$ is the same vector in Eq. (4). Classical optimisation methods only consider information about the average energy, and not about the ansatz itself, which is encoded in the matrix A, used only in variational imaginary time evolution.

Toy example. Here we present two simple toy examples which highlight the difference between variational imaginary time evolution and gradient descent for finding the ground state energy of Hamiltonians. Consider the following Hamiltonians

$$H_A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad H_B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{10}$$

with ansätze

$$|\psi_A(\theta_1, \theta_2, \theta_3)\rangle = e^{i\theta_3} CR_Y^{0,1}(\theta_2) R_X^0(\theta_1) |00\rangle, \tag{11}$$

$$|\psi_B(\theta_1, \theta_2, \theta_3)\rangle = e^{i\theta_3} CR_Y^{0,1}(\theta_2) R_X^0(\theta_1) R_X^1(\theta_1) |01\rangle, \tag{12}$$

prepared by circuits ((A) and (B) respectively)

$$(A) \qquad\qquad (B)$$

Here, $CR_Y^{0,1}$ is a controlled Y rotation with control qubit 0 and target qubit 1, $R_X^q(\theta_1)$ is a rotation of qubit q around the X axis, and the rotation about the j axis is $R_{\sigma_j}(\theta) = e^{-i\theta\sigma_j/2}$ with Pauli matrices σ_j. Note that θ_3 is a *fictitious* parameter which corresponds to the global phase. This is present only so that the evolution of the other parameters are not constrained to produce an oscillating global phase in time, as recently studied in Ref. [44].

We study the ability of variational imaginary time and gradient descent to navigate the energy landscapes of the toy systems A and B, and present the results in Figure 1. Figure 1(a) shows imaginary time robustly discovering the global minima of the energy landscape of system A, while gradient descent becomes trapped in local minima. Figure 1(b) shows imaginary time performing comparably to gradient descent for system B, despite it being only slightly more complicated than system A. This shows that it is still possible for imaginary time evolution to become trapped in local minima, when the ansatz is not sufficiently powerful for certain Hamiltonians.

Simulation of H_2 and LiH. We use our method to find the ground state energy of the H_2 and LiH molecules in their minimal spin-orbital basis sets. We map the molecular fermionic Hamiltonians to qubit Hamiltonians using the procedure described in the Supplementary Materials. The H_2 Hamiltonian acts on two qubits, and considers the space of two electrons in four spin-orbitals. The LiH Hamiltonian acts on eight qubits, and considers an active space of two electrons in eight spin-orbitals. There are numerous possible choices for the ansatz circuit; we use a universal ansatz for H_2 [29] and an ansatz inspired by the low-depth circuit ansatz [45] for LiH, as shown in the Supplementary Materials. The simulation results for H_2 are shown in Fig. 2. We have used a universal ansatz, which is capable of representing all states along the imaginary time trajectory to confirm that our method can recover true imaginary time evolution, when the ansatz is sufficiently powerful. We attribute deviation from the true evolution to the use of an Euler update rule, and finite step size. Our simulations were able to converge to the ground state in all trials.

We compare the LiH results to those obtained using the VQE, with gradient descent as the classical optimisation routine. We use the low-depth circuit ansatz shown in the Supplementary Materials for our simulation, with 137 parameters. This is approximately a quarter the number needed in a universal ansatz. We consider starting from a good initial state (the Hartree-Fock state for LiH), and also random initial states. We believe that the latter simulations provide a more thorough test of both methods.

We use the maximum stable stepsize $\delta\tau$ for each method such that energy monotonically decreases in the first 200 iterations. The stable timestep for imaginary

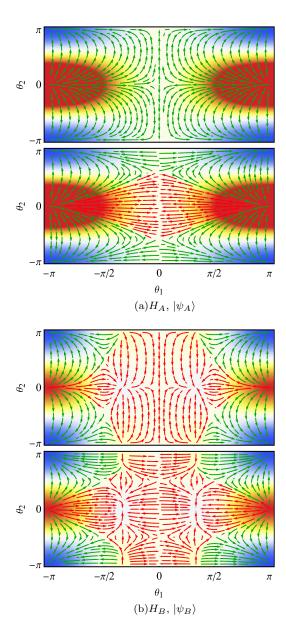(a)H_A, $|\psi_A\rangle$

(b)H_B, $|\psi_B\rangle$

FIG. 1. Comparison of variational imaginary time (top plot in each panel) and gradient descent (bottom plot in each panel) discovering the ground state in toy systems A (top panel) and B (bottom panel). The background colour indicates the energy $\langle\psi(\theta_1,\theta_2,\theta_3)|H|\psi(\theta_1,\theta_2,\theta_3)\rangle$ with red and blue corresponding to the global maximum and ground state energies, respectively. The arrows indicate the trajectories of the methods, and are coloured green if they converge to the true ground state, and red otherwise. While imaginary time avoids all local minima in system A, both methods can become trapped in local minima for the adversarial system B.

time was 0.225, and for gradient descent it was 0.886. Fig. 3 shows the imaginary time method outperforming gradient descent. It is able to locate the ground state more quickly, and accurately. This advantage is most noticeable for the case of random start states, where
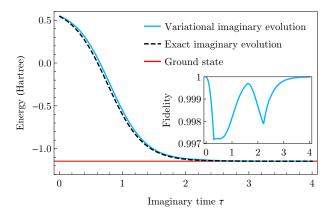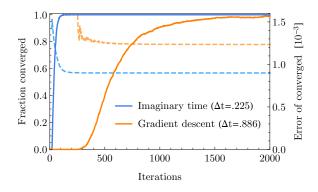
FIG. 2. Simulations of H_2 with random initial parameters and timestep $\delta\tau = 0.01$. The red line is the exact ground state energy. The dashed black line is the exact imaginary time evolution. The blue line is the variational imaginary time evolution. The inset plot shows the fidelity of variational imaginary time to true imaginary time evolution. Here we consider an internuclear distance of $R = 0.75$ Å. The inset plot and main plot share the same x axis label.

the obtained convergence rate is significantly higher than gradient descent. This may be most relevant for solving optimisation problems using the QAOA algorithm, where it is often harder to motivate a good initial starting state.

It is natural to question whether the resource requirements of variational imaginary time evolution are comparable to those of gradient descent. We assess this by performing a simple resource estimation, and by examining its sensitivity to shot noise and to gate errors within the quantum computer. At each iteration, populating the gradient vector requires $\mathcal{O}(N_C N_H N_p)$ measurements, where N_C is the number of measurements required to ascertain a Hamiltonian term to the required precision, N_H is the number of terms in the Hamiltonian, and N_p is the number of parameters used in the ansatz. For imaginary time, the total cost is $\mathcal{O}(N_C N_H N_p + N_p^2 N_A)$ where N_A is the number of measurements required to ascertain an element of the A matrix to the required precision. Sensible ansätze typically have fewer parameters N_p than there are Hamiltonian terms N_H (in our LiH simulations, $N_p = 137$ and $N_H = 181$). If the number of Hamiltonian terms is considerably larger than the number of parameters used, then the additional cost $\mathcal{O}(N_p^2 N_A)$ of imaginary time can be dominated by the cost of calculating the gradient vector. While this is not true for our LiH simulations, we find that it is possible to further reduce the cost of imaginary time by using fewer measurements for each term in the A matrix than for each element of the gradient vector ($N_A \ll N_C$), while still maintaining imaginary time's superior performance. We demonstrate this in Fig. 4, where we vary the number of measurements used to populate A, and simulate the methods under the effect
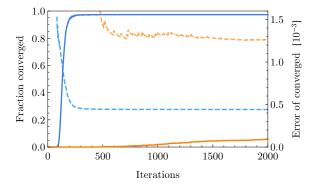

FIG. 3. Noise-free simulations of LiH at an internuclear distance of $R = 1.45$ Å. Simulations in the top plot begin in a small random perturbation (of at most, $\Delta\theta_j = \pi/50$) from the Hartree-Fock state. Simulations in the bottom plot begin with uniformly random parameters. The solid lines (against the left axis) indicate the fraction of 1280 simulations which, by the given iteration, have converged to within 1 mHartree of the true ground state. The dashed lines (against the right axis) indicate the average proximity to the true ground state of only the so-far converged simulations. Imaginary time and gradient descent use their maximum stable timesteps.

of decoherence. The results show that imaginary time can perform significantly better than gradient descent under the presence of noise, even when significantly fewer measurements are made. However, if the gradient is not known to sufficient accuracy ($N_C < 2 \times 10^4$), the reliability of imaginary time evolution cannot be improved by increasing N_A, and can even perform less effectively than gradient descent. Combined with imaginary time's faster convergence and the tendency of gradient based methods to become trapped in local minima, we expect finding the ground state to require substantially fewer measurements using imaginary time than gradient descent.

DISCUSSION
In this work, we have proposed a method to efficiently simulate imaginary time evolution using hybrid quantum-classical computing. We have applied our method to finding the ground state energy of quantum
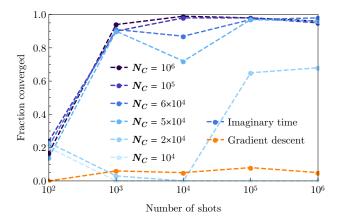
FIG. 4. Simulations of LiH in the presence of a 10^{-4} error rate per gate and varying amounts of shot noise. Each point indicates the fraction of 100 trials which, after 2000 iterations from uniformly random initial parameter states, finished within $1\,$mHartree of the true ground state of LiH. For imaginary time (the blue lines), the horizontal axis indicates the number of shots N_A used in sampling each element of the coefficient matrix A, every iteration. The number of shots N_C used in measuring each Hamiltonian *term* in each element of the gradient vector, as employed by imaginary time evolution, is varied between the blue lines. For gradient descent, the horizontal axis is N_C.

systems, and have tested its performance on H_2 and LiH. As imaginary time evolution outperformed gradient descent at this task, we believe our method provides a competitive alternative to conventional classical optimisation routines. We will examine this further in future work. We expect that our method would also be suitable for solving general optimisation problems, in conjunction with the QAOA, especially given its performance with randomly chosen initial states.

Our method can also be used to prepare a thermal (Gibbs) state, $\rho_T = e^{-H/T}/\text{Tr}[e^{-H/T}]$ of Hamiltonian H at temperature T. Sampling from a Gibbs distribution is an important aspect of many machine learning algorithms, and so we believe that our method is applicable to problems in quantum machine learning. Moreover, while previous methods to prepare the Gibbs state [46, 47] require long gate sequences (and hence, fault tolerance), our method can be implemented using a shallow circuit. Our algorithm can also be combined with recently proposed error mitigation techniques [28, 48–50], and so is suitable for current quantum hardware.

Although exact imaginary time evolution deterministically propagates a good initial state to the ground state in the limit that $\tau \to \infty$, our variational method may still converge to higher energy states, if the chosen ansatz is not sufficiently powerful. In future work, we will investigate how our method may be optimally applied to a variety of tasks in chemistry, optimisation and machine learning. This will include developing

suitable ansätze for a range of problems.

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

CODE AVAILABILITY
The code that supports the findings of this study are available from the corresponding author upon reasonable request.

AUTHOR CONTRIBUTIONS
S.M. and X.Y. conceived of the idea, developed the theory, and wrote the manuscript. T.J., S.M., and X.Y. carried out the numerical simulation. S.E. and Y.L. contributed to the theory. S.B. and X.Y. supervised the project. All authors discussed the results and contributed to the writing of the manuscript.
S.M. and T.J. contributed equally to this work.

ADDITIONAL INFORMATION
Competing Interests: The authors declare that there are no competing interests.

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

* xiao.yuan.ph@gmail.com
[1] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[2] M. H. Poincaré, Rendiconti del Circolo Matematico di Palermo (1884-1940) **21**, 129 (1906).
[3] J. J. Sakurai and J. Napolitano, Modern quantum mechanics (Cambridge University Press, 2017).
[4] J. C. Baez and B. S. Pollard, Entropy **17**, 772 (2015).
[5] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).
[6] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004).

[7] F. A. Wolf, A. Go, I. P. McCulloch, A. J. Millis, and U. Schollwöck, Phys. Rev. X **5**, 041032 (2015).

[8] L. Lehtovaara, J. Toivanen, and J. Eloranta, Journal of Computational Physics **221**, 148 (2007).

[9] C. V. Kraus and J. I. Cirac, New Journal of Physics **12**, 113004 (2010).

[10] J. R. McClean and A. Aspuru-Guzik, RSC Adv. **5**, 102277 (2015).

[11] T. Shi, E. Demler, and J. I. Cirac, Annals of Physics **390**, 245 (2018).

[12] J. R. McClean, J. A. Parkhill, and A. Aspuru-Guzik, **110**, E3901 (2013).

[13] J. R. McClean and A. Aspuru-Guzik, Phys. Rev. A **91**, 012311 (2015).

[14] W. Al-Saidi, S. Zhang, and H. Krakauer, The Journal of chemical physics **124**, 224101 (2006).

[15] M. Motta and S. Zhang, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1364 (2018).

[16] G. K.-L. Chan and S. Sharma, Annual review of physical chemistry **62**, 465 (2011).

[17] E. M. Stoudenmire and S. R. White, Physical review letters **119**, 046401 (2017).

[18] R. P. Feynman, International Journal of Theoretical Physics **21**, 467 (1982).

[19] F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).

[20] I. M. Georgescu, S. Ashhab, and F. Nori, Rev. Mod. Phys. **86**, 153 (2014).

[21] M. A. Nielsen and I. Chuang, "Quantum computation and quantum information," (2002).

[22] H. F. Trotter, Proceedings of the American Mathematical Society **10**, 545 (1959).

[23] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **79**, 2586 (1997).

[24] R. Jackiw and A. Kerman, Physics Letters A **71**, 158 (1979).

[25] P. Kramer, Journal of Physics: Conference Series **99**, 012009.

[26] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, Phys. Rev. Lett. **107**, 070601 (2011).

[27] Y. Ashida, T. Shi, M. C. Bañuls, J. I. Cirac, and E. Demler, arXiv preprint arXiv:1802.03861 (2018).

[28] Y. Li and S. C. Benjamin, Phys. Rev. X **7**, 021050 (2017).

[29] 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).

[30] Y. Wang, F. Dolde, J. Biamonte, R. Babbush, V. Bergholm, S. Yang, I. Jakobi, P. Neumann, A. Aspuru-Guzik, J. D. Whitfield, et al., ACS nano **9**, 7769 (2015).

[31] 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).

[32] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Phys. Rev. A **95**, 020501 (2017).

[33] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).

[34] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, Phys. Rev. Lett. **118**, 100503 (2017).

[35] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).

[36] E. Farhi, J. Goldstone, and S. Gutmann, arXiv preprint arXiv:1411.4028 (2014).

[37] J. Preskill, Quantum **2**, 79 (2018).

[38] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, Phys. Rev. Lett. **106**, 170501 (2011).

[39] F. Verstraete, V. Murg, and J. I. Cirac, Advances in Physics **57**, 143 (2008).

[40] K. B. Whaley, A. R. Dinner, and S. A. Rice, Quantum information and computation for chemistry (John Wiley & Sons, 2014).

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

[42] J. Broeckhove, L. Lathouwers, E. Kesteloot, and P. Van Leuven, Chemical physics letters **149**, 547 (1988).

[43] D. Wecker, M. B. Hastings, and M. Troyer, Phys. Rev. A **92**, 042303 (2015).

[44] X. Yuan, S. Endo, Q. Zhao, S. Benjamin, and Y. Li, arXiv e-prints , arXiv:1812.08767 (2018), arXiv:1812.08767 [quant-ph].

[45] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, arXiv preprint arXiv:1801.01053 (2018).

[46] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Nature **471**, 87 (2011).

[47] A. Riera, C. Gogolin, and J. Eisert, Phys. Rev. Lett. **108**, 080402 (2012).

[48] K. Temme, S. Bravyi, and J. M. Gambetta, Phys. Rev. Lett. **119**, 180509 (2017).

[49] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X **8**, 031027 (2018).

[50] S. McArdle, X. Yuan, and S. Benjamin, arXiv:1807.02467 (2018).

[51] J. Romero, R. Babbush, J. McClean, C. Hempel, P. Love, and A. Aspuru-Guzik, Quantum Science and Technology (2018).

[52] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Science **309**, 1704 (2005).

[53] 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.

[54] 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).

[55] T. Jones, A. Brown, I. Bush, and S. Benjamin, "Quest and high performance simulation of quantum computers," (2018), arXiv:1802.08032.

[56] Contributors and GSL Project, "GSL - GNU scientific library - GNU project - free software foundation (FSF)," http://www.gnu.org/software/gsl/ (2010).

[57] G. Golub and C. V. Loan, Matrix computations, Vol. 83 (Cambridge University Press, 1999) p. 556.

Supplementary Information

Variational simulation of imaginary time evolution

McLachlan's variational principle [41], applied to imaginary time evolution, is given by

$$\delta\|(\partial/\partial\tau + H - E_\tau)|\psi(\tau)\rangle\| = 0 \tag{13}$$

where

$$\|(\partial/\partial\tau + H - E_\tau)|\psi(\tau)\rangle\| = ((\partial/\partial\tau + H - E_\tau)|\psi(\tau)\rangle)^\dagger (\partial/\partial\tau + H - E_\tau)|\psi(\tau)\rangle, \tag{14}$$

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

$$\frac{\partial|\psi(\tau)\rangle}{\partial\tau} = -(H - E_\tau)|\psi(\tau)\rangle. \tag{15}$$

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

$$\begin{aligned}
\|(\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle\| &= ((\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle)^\dagger (\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle, \\
&= \sum_{i,j}\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}\frac{\partial|\phi(\tau)\rangle}{\partial\theta_j}\dot\theta_i\dot\theta_j + \sum_i\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}(H - E_\tau)|\phi(\tau)\rangle\dot\theta_i \\
&+ \sum_i\langle\phi(\tau)|(H - E_\tau)\frac{\partial|\phi(\tau)\rangle}{\partial\theta_i}\dot\theta_i + \langle\phi(\tau)|(H - E_\tau)^2|\phi(\tau)\rangle.
\end{aligned} \tag{16}$$

Focusing on $\dot\theta_i$, we obtain

$$\begin{aligned}
\frac{\partial\|(\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle\|}{\partial\dot\theta_i} &= \sum_j\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}\frac{\partial|\phi(\tau)\rangle}{\partial\theta_j} + \frac{\partial\langle\phi(\tau)|}{\partial\theta_j}\frac{\partial|\phi(\tau)\rangle}{\partial\theta_i}\right)\dot\theta_j \\
&+ \frac{\partial\langle\phi(\tau)|}{\partial\theta_i}(H - E_\tau)|\phi(\tau)\rangle + \langle\phi(\tau)|(H - E_\tau)\frac{\partial|\phi(\tau)\rangle}{\partial\theta_i}.
\end{aligned} \tag{17}$$

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

$$\langle\phi(\tau)|\phi(\tau)\rangle = 1, \tag{18}$$

we have

$$E_\tau\frac{\partial\langle\phi(\tau)|\phi(\tau)\rangle}{\partial\theta_i} = E_\tau\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}|\phi(\tau)\rangle + \langle\phi(\tau)|\frac{\partial|\phi(\tau)\rangle}{\partial\theta_i}\right) = 0, \tag{19}$$

and the derivative is simplified to

$$\frac{\partial\|(\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle\|}{\partial\dot\theta_i} = \sum_j A_{ij}\dot\theta_j - C_i. \tag{20}$$

where

$$\begin{aligned}
A_{ij} &= \Re\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}\frac{\partial|\phi(\tau)\rangle}{\partial\theta_j}\right), \\
C_i &= -\Re\left(\frac{\partial\langle\phi(\tau)|}{\partial\theta_i}H|\phi(\tau)\rangle\right).
\end{aligned} \tag{21}$$

McLachlan's variational principle requires

$$\frac{\partial\|(\partial/\partial\tau + H - E_\tau)|\phi(\tau)\rangle\|}{\partial\dot\theta_j} = 0, \tag{22}$$

which is equivalent to the differential equation of the parameters

$$\sum_j A_{ij}\dot{\theta}_j = C_i. \tag{23}$$

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;

$$\begin{aligned}
\frac{dE(\tau)}{d\tau} &= \Re\left(\langle\phi(\tau)| H \frac{d |\phi(\tau)\rangle}{d\tau}\right), \\
&= \sum_i \Re\left(\langle\phi(\tau)| H \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i}\dot{\theta}_i\right), \\
&= -\sum_i C_i\dot{\theta}_i, \\
&= -\sum_i C_i A_{ij}^{-1} C_j, \\
&\leq 0.
\end{aligned} \tag{24}$$

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 $x = (x_1, x_2, \ldots, x_N)^T$, and calculate $x^\dagger \cdot A \cdot x$,

$$\begin{aligned}
x^\dagger \cdot A \cdot x &= \sum_{i,j} x_i^* A_{ij} x_j, \\
&= \sum_{i,j} x_i^* \Re\left(\frac{\partial \langle\phi(\tau)|}{\partial\theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_j}\right) x_j, \\
&= \sum_{i,j} x_i^* \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_j} x_j + \sum_{i,j} x_i^* \frac{\partial \langle\phi(\tau)|}{\partial\theta_j} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_i} x_j,
\end{aligned} \tag{25}$$

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

$$\sum_{i,j} x_i^* \frac{\partial \langle\phi(\tau)|}{\partial\theta_i} \frac{\partial |\phi(\tau)\rangle}{\partial\theta_j} x_j = \langle\Phi|\Phi\rangle \geq 0. \tag{26}$$

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., $G_{i,j} = (UAU^\dagger)_{i,j} = 0, \forall i \neq j$. Then, we define G^{-1} by

$$G_{i,j}^{-1} = \begin{cases} \frac{1}{G_{i,j}} & i = j, G_{i,j} \neq 0, \\ 0 & i = j, G_{i,j} = 0, \\ 0 & i \neq j. \end{cases} \tag{27}$$

The inverse of A is thus defined by

$$A^{-1} = U^\dagger G^{-1} U. \tag{28}$$

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

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. [28, 45, 51].

Without loss of generality, we can assume that each unitary gate $U_i(\theta_i)$ in our circuit depends only on parameter θ_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

$$\frac{\partial U_i(\theta_i)}{\partial \theta_i} = \sum_k f_{k,i} U_i(\theta_i)\sigma_{k,i}, \tag{29}$$

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

$$\frac{\partial |\phi(\tau)\rangle}{\partial \theta_i} = \sum_k f_{k,i} \tilde{V}_{k,i} |\bar{0}\rangle, \tag{30}$$

with

$$\tilde{V}_{k,i} = U_N(\theta_N)\ldots U_{i+1}(\theta_{i+1})U_i(\theta_i)\sigma_{k,i}\ldots U_2(\theta_2)U_1(\theta_1). \tag{31}$$

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 $R^Z_{\theta_i} = e^{-i\theta_i \sigma_Z/2}$, 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 $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes R^Z_{\theta_i}$, the derivative $\partial U_i(\theta_i)/\partial \theta_i = |1\rangle\langle 1| \otimes \partial R^Z_{\theta_i}/\partial \theta_i = -i/2 \times |1\rangle\langle 1| \otimes Z e^{-i\theta_i Z/2}$. 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. (29).

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

$$A_{ij} = \Re\left(\sum_{k,l} f^*_{k,i} f_{l,j} \langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle\right),$$

$$C_i = \Re\left(\sum_{k,l} f^*_{k,i}\lambda_l \langle \bar{0}| \tilde{V}^\dagger_{k,i}h_l V |\bar{0}\rangle\right). \tag{32}$$

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 $\Re(f^*_{k,i} f_{l,j} \langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle)$, we can let $f^*_{k,i} f_{l,j} = ae^{i\theta}$ and

$$\langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle = \langle \bar{0}| U^\dagger_1 \ldots U^\dagger_{i-1}\sigma^\dagger_{k,i}U^\dagger_i \ldots U^\dagger_N U_N \ldots U_j\sigma_{l,j}U_{j-1}\ldots U_1 |\bar{0}\rangle. \tag{33}$$

Suppose $i < j$, then

$$\langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle = \langle \bar{0}| U^\dagger_1 \ldots U^\dagger_{i-1}\sigma^\dagger_{k,i}U^\dagger_i \ldots U^\dagger_{j-1}\sigma_{l,j}U_{j-1}\ldots U_i \ldots U_1 |\bar{0}\rangle, \tag{34}$$

and $\Re(e^{i\theta} \langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle)$ can be measured by the circuit in Fig. 5. The terms for C can be measured similarly.
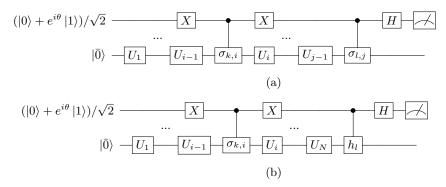

FIG. 5. Quantum circuits that evaluate (a) $\Re(e^{i\theta} \langle \bar{0}| \tilde{V}^\dagger_{k,i}\tilde{V}_{l,j} |\bar{0}\rangle)$ and (b) $\Re(e^{i\theta} \langle \bar{0}| \tilde{V}^\dagger_{k,i}h_l V |\bar{0}\rangle)$. When $\sigma_{k,i}$ is Hermitian, the X gates acting on the ancilla qubit can be also omitted.

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 [52]. 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

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_{i,I} \frac{e^2}{4\pi\epsilon_0}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{4\pi\epsilon_0}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{I\neq J}\frac{e^2}{4\pi\epsilon_0}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \tag{35}$$

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

$$\left[-\sum_i \frac{\nabla_i^2}{2} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] |\psi\rangle = E|\psi\rangle, \tag{36}$$

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

$$H = \sum_{p,q} h_{pq} a_p^\dagger a_q + \frac{1}{2}\sum_{p,q,r,s} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \tag{37}$$

with

$$\begin{aligned} h_{pq} &= \int d\mathbf{x}\,\phi_p^*(\mathbf{x})\left(-\frac{\nabla_i^2}{2} - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \right)\phi_q(\mathbf{x}), \\ h_{pqrs} &= \int d\mathbf{x}_1 d\mathbf{x}_2 \frac{\phi_p^*(\mathbf{x}_1)\phi_q^*(\mathbf{x}_2)\phi_s(\mathbf{x}_1)\phi_r(\mathbf{x}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \end{aligned} \tag{38}$$

where $\mathbf{x}$ is a spatial and spin coordinate. This Hamiltonian in a molecular orbital basis contains $O(N_{\mathrm{SO}}^4)$ terms, where N_{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. [53].

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-nG' 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

$$
\begin{aligned}
|\phi_0\rangle = |\sigma_{g\uparrow}\rangle &= \frac{1}{\sqrt{\mathcal{N}_G}}(|1S_{1\uparrow}\rangle + |1S_{2\uparrow}\rangle), \\
|\phi_1\rangle = |\sigma_{g\downarrow}\rangle &= \frac{1}{\sqrt{\mathcal{N}_G}}(|1S_{1\downarrow}\rangle + |1S_{2\downarrow}\rangle), \\
|\phi_2\rangle = |\sigma_{u\uparrow}\rangle &= \frac{1}{\sqrt{\mathcal{N}_U}}(|1S_{1\uparrow}\rangle - |1S_{2\uparrow}\rangle), \\
|\phi_3\rangle = |\sigma_{u\downarrow}\rangle &= \frac{1}{\sqrt{\mathcal{N}_U}}(|1S_{1\downarrow}\rangle - |1S_{2\downarrow}\rangle),
\end{aligned}
\tag{39}
$$

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. [53], the qubit Hamiltonian for H_2 in the BK representation can be obtained. This 4 qubit Hamiltonian is given by

$$
\begin{aligned}
H = {}& h_0 I + h_1 Z_0 + h_2 Z_1 + h_3 Z_2 + h_4 Z_0 Z_1 + h_5 Z_0 Z_2 + h_6 Z_1 Z_3 + h_7 X_0 Z_1 X_2 + h_8 Y_0 Z_1 Y_2 \\
& + h_9 Z_0 Z_1 Z_2 + h_{10} Z_0 Z_2 Z_3 + h_{11} Z_1 Z_2 Z_3 + h_{12} X_0 Z_1 X_2 Z_3 + h_{13} Y_0 Z_1 Y_2 Z_3 + h_{14} Z_0 Z_1 Z_2 Z_3.
\end{aligned}
\tag{40}
$$

As this Hamiltonian only acts off diagonally on qubits 0 and 2 [31, 35], it can be reduced to

$$
H = g_0 I + g_1 Z_0 + g_2 Z_1 + g_3 Z_0 Z_1 + g_4 Y_0 Y_1 + g_5 X_0 X_1,
\tag{41}
$$

which only acts on two qubits.

In our work, we consider an internuclear distance of $R = 0.75$ Å 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 [29] shown in Fig. 6.
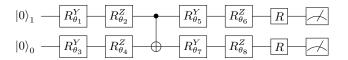

FIG. 6. The quantum circuit for preparing the hardware efficient ansatz with eight parameters. Here each $R_\theta^Y = e^{-i\theta\sigma_Y/2}$, $R_\theta^Z = e^{-i\theta\sigma_Z/2}$, and the two R gates correspond to the rotation of the measurement.

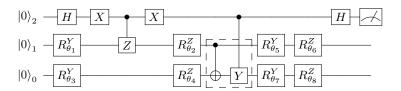We have eight parameters and an example circuit for measureing $A_{2,7}$ is shown in Fig. 7

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

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,

$$
\begin{pmatrix}
1.99991 & -0.00047 & 0.00047 & 0 & 0 & -0.00120 \\
-0.00047 & 1.95969 & 0.06691 & 0 & 0 & 0.00842 \\
0.00047 & 0.06691 & 0.00968 & 0 & 0 & -0.01385 \\
0 & 0 & 0 & 0.00172 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.00172 & 0 \\
-0.00120 & 0.00842 & -0.01385 & 0 & 0 & 0.02728
\end{pmatrix}.
\tag{42}
$$

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,

$$
\begin{pmatrix}
1.99992 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.96201 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.03459 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.00005 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.00172 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.00172
\end{pmatrix}.
\tag{43}
$$

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 [54], an electronic structure package to transform computational chemistry problems into a form that is suitable for investigation using a quantum computer.

LiH ansatz

We used an ansatz inspired by the low depth circuit ansatz (LDCA) [45]. 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. 8.

Numerical simulation

We simulate the aforementioned quantum circuits using the Quantum Exact Simulation Toolkit (QuEST) [55]. True ground states are found by diagonalising the considered Hamiltonians with the GNU Scientific Library (GSL) [56], which employs a complex form of the symmetric bidiagonalisation and QR reduction method [57].

While our H_2 tests simulate the full experimental routine, our LiH tests are optimised by individually computing each ansatz derivative $\partial \left| \psi(\vec{\theta}) \right\rangle / \partial \theta_j$. 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 $\hat{H} \left| \psi(\theta) \right\rangle$ can be computed
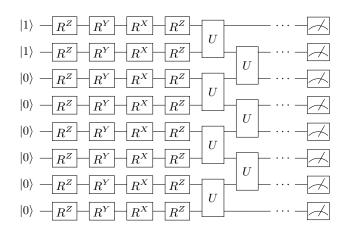
FIG. 8. The general structure of the ansatz used in our LiH simulations. We repeat the circuit structure of the two-qubit rotations block to depth $M = 3$. The form of the U gate is $U = e^{i\alpha YX} e^{i\beta XY} e^{i\gamma ZZ} e^{i\delta YY} e^{i\epsilon XX}$. In total there are $(3 \times 5 \times 7) + (4 \times 8) = 137$ parameters.

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.

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

$$\|C - A\dot{\vec{\theta}}\|^2 + \lambda\|\dot{\vec{\theta}}\|^2. \tag{44}$$

Here, the Tikhonov parameter λ can be varied to trade accuracy for keeping $\dot{\vec{\theta}}$ small and the parameter evolution smooth. We estimate an ideal λ at each time-step by selecting the corner of a 3-point L-curve [56, 57], though force $\lambda \in [10^{-4}, 10^{-2}]$. This is because too large a λ 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

$$a_{ij} = \Re \left(\frac{\partial \langle \psi(\vec{\theta})|}{\partial \theta_i} \frac{\partial |\psi(\vec{\theta})\rangle}{\partial \theta_j} \right) \in \left[-\frac{1}{4}, \frac{1}{4} \right]. \tag{45}$$

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

$$A_{ij} \sim N \left(a_{ij}, \frac{1}{N_A} \left(1/16 - a_{ij}^2 \right) \right). \tag{46}$$

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 ρ as

$$\rho_{\text{final}} = \left(1 - 10^{-4} \right)^D \rho + \left(1 - \left(1 - 10^{-4} \right)^D \right) \mathbb{1}. \tag{47}$$

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.