Towards Practical Quantum Variational Algorithms

Dave Wecker,[1] Matthew B. Hastings,[2,1] and Matthias Troyer[3]

[1] *Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA*
[2] *Station Q, Microsoft Research, Santa Barbara, CA 93106-6105, USA*
[3] *Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland*

The preparation of quantum states using short quantum circuits is one of the most promising near-term applications of small quantum computers, especially if the circuit is short enough and the fidelity of gates high enough that it can be executed without quantum error correction. Such quantum state preparation can be used in variational approaches, optimizing parameters in the circuit to minimize the energy of the constructed quantum state for a given problem Hamiltonian. For this purpose we propose a simple-to-implement class of quantum states motivated by adiabatic state preparation. We test its accuracy and determine the required circuit depth for a Hubbard model on ladders with up to 12 sites (24 spin-orbitals), and for small molecules. We find that this ansatz converges faster than previously proposed schemes based on unitary coupled clusters. While the required number of measurements is astronomically large for quantum chemistry applications to molecules, applying the variational approach to the Hubbard model (and related models) is found to be far less demanding and potentially practical on small quantum computers. We also discuss another application of quantum state preparation using short quantum circuits, to prepare trial ground states of models faster than using adiabatic state preparation.

PACS numbers:

I. INTRODUCTION

Variational classes of states such as matrix product states (MPS) [1], multiscale entanglement renormalization (MERA) [2], and projected entangled pair states (PEPS) [3] play a key role in studying many-body quantum systems. An ideal class of states should be large enough to approximate the ground state and it must be possible to evaluate the energy and observables. Unfortunately for many potential classes of states, including PEPS, evaluation may be difficult on a classical computer, and for other classes it can be very computationally expense. However, on a *quantum computer*, large classes of PEPS states can be prepared efficiently [4, 5], (although likely not all PEPS can be prepared this way [6]). Once the state is prepared, the energy and observables can be measured by sampling using short quantum circuits. Similar quantum circuits exist for MPS and MERA states [7]. Recently, Ref. 8 proposed variational methods for studying quantum chemistry problems on a quantum computer, and demonstrated one method on a model with a four-dimensional Hilbert space.

Preparation of variational states is an attractive application of small quantum computers, since many classically intractable states can be prepared with quite short quantum circuits and thus do not pose stringent requirements on coherence times and gate fidelities. While variation over all possible states that can be produced by quantum circuits of a given maximum depth (as done in the demonstration experiments of Ref. 8) might sound ideal, this approach does not scale efficiently. Ref. 8 also proposed using the unitary coupled cluster method (UCC) [9], where variational states are prepared using unitary evolution under a sum of fermionic terms including $c_p^\dagger c_q$, $c_p^\dagger c_q^\dagger c_r c_s$, and higher order terms That approach involves many parameters: even if truncated at the level of four fermion operators, the number of variational parameters may scale as the fourth power of the number of orbitals. We will consider this method below and find that this then requires a large number of evaluations of the energy to obtain accurate results and for many systems accuracy is still limited if truncated at this level.

In this paper, we present an analysis of a different class of variational states, using a modest gate depth and a very modest number of variational parameters compared to the system size. We present a detailed numerical analysis of this variational technique applied to a Hubbard model, ranging up to systems of 12 sites (equivalently, 24 spin-orbitals) at half-filling, to quantify its accuracy. The difficulty in general increases with system size, but also fluctuates in complicated ways, depending upon the spectrum of the non-interacting model. For certain cases, including our largest size, the non-interacting model is highly degenerate, leading to strong interaction effects and a very poor overlap of an initial Slater determinant state with the true ground state. We find that even then the variational approach works well. Finally, we also analyze the effects of sampling error in measuring energy and give concrete estimates for time scales to implement this procedure in practice. A direct comparison to UCC shows higher accuracy for a given number of evaluations of the energy, and significantly fewer parameters are required.

The class of variational states that we consider is inspired by both adiabatic state preparation and the quantum optimization algorithm of Refs. 10, 11. Consider a family of Hamiltonians $H = \sum_a J_a h_a$ with the J_a being scalars and the h_a being operators. Let J_a^0 and J_a^1 be two choices of these J_a, with corresponding Hamiltonians H_0, H_1. Suppose it is easy to prepare a ground

state Ψ_I of H_0 (for example, in our study of the Hubbard model, H_0 is non-interacting). Then, assuming that no gap closes, it is possible to adiabatically evolve from Ψ_I to the ground state of H_1. If we break the annealing into short time steps dt and evolve for a total time T, this annealing is a sequence of $(T/\mathrm{d}t)$ different unitary rotations by Hamiltonians interpolating between H_0 and H_1. This could be implemented on a quantum computer using a Trotter-Suzuki method which further decomposes this sequence into a sequence of unitary rotations by individual terms in the Hamiltonian.

The idea of the variational method that we consider is still restricted to a sequence of unitary rotations by terms in the Hamiltonian, but considers arbitrary angles for the rotations in the sequence, rather than choosing them from a Trotterization of an annealing process. By choosing these angles arbitrarily, this allows us to take a much shorter sequence. Thus, we consider a trial state, which we call the "Hamiltonian variational" state

$$\Psi_T = \exp(i\theta_n h_{a_n}) \dots \exp(i\theta_2 h_{a_2}) \exp(i\theta_1 h_{a_1}) \Psi_I, \quad (1)$$

where larger n increases the accuracy. The angles θ_k are variational parameters. The optimziation algorithm of Refs. 10, 11 considered only rotation by two different types of operators, while we consider rotation by a larger number of terms.

If all the terms in the Hamiltonian have some symmetry (such as spin-rotation symmetry), then Ψ_I and Ψ_T transform in the same way under this symmetry. This allows us to find ground states with different quantum numbers by picking initial states Ψ_I with the desired quantum number. In cases of symmetries such as translation invariance, it is helpful to try to construct the sequence of terms a_k to approximately preserve this symmetry (see a more detailed discussion below).

To be useful for variational state preparation, it must be possible to optimize over the given parameters, without getting stuck in false minima. In Sec. II D we discuss techniques which succeed in finding good optima for the Hubbard model using access to numerically exact values of the energy obtained from a classical simulation. Such a classical approach may be useful on a quantum computer as it finds short circuits that prepare specific highly entangled states, albeit limited to system sizes where classical simulation is possible.

To go beyond the limits of classical simulation on a quantum computer one must measure the energy using a quantum circuit. One way is to write the Hamiltonian as a sum of sets of terms, so that all terms in a given set commute with each other and then measure each set in one run, doing many runs for each set of terms, with the error in the energy going as the inverse square-root of the number of samples. The required number of samples to achieve high accuracy may be large and limiting the number of samples significantly impacts the ability to find the optimum as discussed later in Sec. II E.

Another way to measure energy is to to use phase estimation to compute the energies, rather than sampling. In this case, the error in energy goes inversely with the phase estimation time. This procedure still is probabilistic, returning on each run an energy chosen randomly from a distribution. Thus many runs are still required to estimate the average energy. For many distributions, this procedure provides much more acccurate access to the energy for a given run time than sampling would, at the cost of requiring longer coherence time. Additionally, it allows one to optimize on the probability of finding the ground state, rather than on the energy, which may improve the optimization.

One may wonder why this would be useful, since if phase estimation finds an energy equal to the ground state, one knows that one has successfully prepared the ground state and can now make a measurement, without any need to improve the variational state. However, once the optimum variational parameters are found, these parameters are classical information that can be used to quickly re-create the state. This is useful if one wishes to make many measurements on the state, if the measurements destroy the state. Also, techniques in Ref. 12 show how to quadratically speed up the sampling of properties of the state, assuming access by measuring a projector onto the state. However, if we have access to a projector onto Ψ_I, then conjugating this projector by the unitaries in Eq. (1) gives us a short-depth circuit which projects Ψ_T. Finally, in many applications such as chemical reactions where one studies the same Hamiltonian along a path of parameters, it may be possible to use the variational solution for a given Hamiltonian to find a solution of a nearby Hamiltonian rapidly.

II. THE HUBBARD MODEL

A. The Hubbard Hamiltonian

To study performance scaling across a range of sizes, we consider a sequence of Hubbard models on N-sites, with the sites arranged in a two-leg ladder (an $N/2$-by-2 square lattice) for $N = 4, 6, 8, 10, 12$. The Hamiltonian of the Hubbard model is

$$H = -t \sum_{\langle i,j \rangle} \sum_{\sigma} t_{ij} c^\dagger_{i,\sigma} c_{j,\sigma} + U \sum_i c^\dagger_{i,\uparrow} c_{i,\uparrow} c^\dagger_{i,\downarrow} c_{i,\downarrow}, \quad (2)$$

with $t = 1, U = 2$. Here $c^\dagger_{i,\sigma}$ and ci, σ create and annihilate an electron at site i with spin σ respectively. We use periodic boundary conditions along the long "horizontal" direction of the lattice which is $N/2$ sites long and open boundary conditions in the short "vertical" direction so that there are a total of N bonds in the horizontal direction and $N/2$ in the vertical direction, all with the same strength. We write this as $H = h_h + h_v + h_U$, where h_h is the sum of hopping terms in the horizontal direction, h_v is the sum of hopping terms in the vertical direction, and h_U is the repulsion term.

B. Ground state degeneracies and initial states

We studied the system at half-filling with a N electrons on N sites. For an equal number of $N/2$ up and down electrons the single particle spectrum for $N = 4, 6, 10$ has $N/2-1$ states below the Fermi energy and is doubly degenerate at the Fermi energy, giving a $2^2 = 4$-fold degenerate many-body ground state for the non-interacting ($U = 0$) Hamiltonian.. For $N = 8$ the $U = 0$ Hamiltonian has a unique ground state, with $N/2$ states below the Fermi energy and an excitation gap. For $N = 12$, there are $N/2-2$ single particle states below the Fermi energy, and four states at the Fermi energy, so that for $N/2$ up electrons and $N/2$ down electrons, the $U = 0$ ground state is $\binom{4}{2}^2 = 36$-fold degenerate. These different degeneracies of the non-interacting problem impact the difficulty of solving it. They also reduce the overlap of a Slater determinant with the true ground state. The degeneracies at $N = 4$ and 12 are due to the special choice of the vertical coupling; for generic vertical coupling, degeneracies are only seen for $N = 4n + 2 = 6, 10, 14, \ldots$, where n is a positive integer.

The spin of the ground state at $U = 2$ also depends upon N. It is a singlet for $N = 4n = 4, 8, 12, \ldots$ and is a triplet for $N = 4n + 2 = 6, 10, \ldots$. All the results that we report below are for Ψ_I in the correct spin sector. In the case of $N = 6, 10$ we did this by choosing $N/2 + 1$ up electrons and $N/2-1$ down electrons. In that sector, the free fermion ground state becomes non-degenerate. Otherwise, we chose $N/2$ up electrons and $N/2$ down electrons. To choose a unique ground state for $N = 4, 12$ we prepared the ground state of the Hamiltonian $th_h + (1-\epsilon)th_v$ for small $\epsilon > 0$; this state is independent of ϵ for ϵ small and this simply picks out one of the ground states of H.

C. The variational ansatz

We choose the terms a_k in Eq. (1) in a repeating pattern and perform S repetitions, which we call "steps", In each step, we have three variational parameters, $\theta_h^b, \theta_v^b, \theta_U^b$, where $b = 1, \ldots, S$. We thus obtain

$$\Psi_T = \prod_{b=1}^{S} \Big(U_U(\frac{\theta_U^b}{2}) U_h(\theta_h^b) U_v(\theta_v^b) U_U(\frac{\theta_U^b}{2}) \Big) \Psi_I, \qquad (3)$$

where $U_X(\theta) = \exp(i\theta h_X)$ for $X \in \{U, h, v\}$ and the product is ordered by decreasing b. This form approximates evolution under $H(b) \equiv \theta_U^b h_U + \theta_v^b h_v + \theta_h^b h_h$, (note that $[U_h, U_v] = 0$; this property holds in general for free fermion hopping terms on any square lattice because U_h and U_v are diagonal operators in a basis for the Fock space obtained from a momentum basis for single-particle states).

Since h_h is a sum of non-commuting terms, we used a second-order Trotter-Suzuki method to implement U_h, applying the terms in sequence from left to right in each row and then reversing the order. This Trotterization helps approximately preserve momentum, useful if Ψ_I has the right momentum quantum numbers. The sequence length can be reduced by combining $U_U(\frac{\theta_U^{b+1}}{2}) U_U(\frac{\theta_U^b}{2}) = U_U \frac{\theta_U^b + \theta_U^{b+1})}{2})$; other orderings, such as first applying odd terms and then applying even terms will slightly reduce the gate depth.

It is important to emphasize that we are *not* concerned with Trotter error other than for a desire to preserve quantum numbers such as momentum; while this choice of unitaries gives evolution similar to the evolution under $H(b)$, we are optimizing the parameters for evolution with these unitaries. If we instead did exact evolution under a time-varying Hamiltonians (which is possible for atomic quantum gases in optical lattices [13]), we would instead optimize to a different optimum in the parameters. We expect that the difference in these evolutions can be largely absorbed into a small shift in the variational parameters.

D. Optimization with Exact Energies

We first consider the case where we can exactly calculate energies on a classical computer and use a simple optimization procedure. Six points (a "point" is a set of parameters) are chosen at random near the origin. For each point, we first follow a greedy noisy search, where we slightly perturb the values of the points, accepting whenever this reduces the energy, for a total of 150 evaluations of the energy. We then use Powell's conjugate direction [14] method until it converges. After following this procedure for each of the six points, we keep the point whose energy is lowest at the end of the procedure, and we discard the other five points. For the point we keep, we alternate greedy noisy search and Powell search until neither can find an improvement. Our greedy noisy search uses a simple algorithm to determine step size: every thirty trials we count the number of acceptances. If that number is large, the step size is increased; if the number is small, the step size is reduced.

We call this optimization algorithm the "global variational" method as it involves optimizing all parameters simultaneously. Using the global variational approach and an ansatz with $S = 3$, we obtain energy errors of 2.0×10^{-8}, 0.019,0.029, 0.083, and 0.59 for $N = 4$, 6, 8, 10, and 12. Increasing S improves the error, but at the larger sizes convergence was slow and results varied greatly from run to run, suggesting that the minimization was getting stuck in local minima. One clue to the origin of this difficulty is that in some cases the energy reduced but the overlap with the ground state also reduced. This is possible if it also reduces the amplitude of some highly excited state but increases the amplitude of a low excited state.

To overcome this problem, we used an alternative procedure to find the optimum. This procedure is inspired

S	ΔE^s	ΔE^f	P^s	P^f	ΔE^s	ΔE^f	P^s	P^f
	4				6			
3	0.23	0.011	0.7311	0.9878	0.053	0.033	0.9853	0.9903
5	0.14	8×10^{-8}	0.8402	1.0000	0.041	0.0025	0.9899	0.9994
7	0.16	1×10^{-8}	0.8219	1.0000	0.034	0.0011	0.9920	0.9997
9	0.038	3×10^{-8}	0.9604	1.0000	0.023	0.00032	0.9945	0.9999
11	0.016	3×10^{-8}	0.9841	1.0000	0.016	0.00013	0.9968	1.0000
	8				10			
3	0.078	0.033	0.9869	0.9934	0.12	0.083	0.9288	0.9374
5	0.019	0.0046	0.9954	0.9984	0.11	0.040	0.9385	0.9585
7	0.016	0.0029	0.9956	0.9988	0.10	0.022	0.9444	0.9713
9	0.016	0.0022	0.9956	0.9991	0.099	0.016	0.9490	0.9792
11	0.016	0.0018	0.9956	0.9993	0.097	0.013	0.9538	0.9855
13	0.016	0.0018	0.9956	0.9993	0.090	0.011	0.9645	0.9869
15	0.016	0.0016	0.9956	0.9995	0.061	0.0043	0.9827	0.9980
17	0.016	0.0014	0.9956	0.9994	0.052	0.0039	0.9873	0.9984
19	0.016	0.0012	0.9957	0.9995	0.045	0.0027	0.9870	0.9991
	8^*				12			
3	0.61	0.53	0.4598	0.5232	0.67	0.59	0.3533	0.4102
5	0.48	0.17	0.5773	0.8727	0.53	0.39	0.4841	0.6154
7	0.25	0.065	0.7505	0.9346	0.44	0.18	0.5570	0.8522
9	0.14	0.034	0.8687	0.9601	0.34	0.086	0.6553	0.9367
11	0.087	0.025	0.9103	0.9679	0.26	0.050	0.7452	0.9615
13	0.069	0.021	0.9245	0.9712	0.23	0.030	0.7789	0.9733
15	0.064	0.018	0.9304	0.9761	0.21	0.023	0.7973	0.9769
17	0.063	0.015	0.9336	0.9806	0.20	0.016	0.8092	0.9862
19	0.060	0.010	0.9403	0.9908	0.19	0.013	0.8223	0.9900

TABLE I: Performance using the annealed variational method. Numbers $4, 6, 8, 10, 12$ indicate different values of N. 8^* is described in text. The left-hand column is number of steps, S. Quantities $\Delta E^s, \Delta E^f$ indicate error in ground state energy after sequential and full optimization, respectively. Quantities P^s, P^f indicate absolute squared ground state overlap in those two cases, respectively.

by annealing and we call it the "annealed variational" method. Let $H_s = th_H + th_V + sUh_U$ interpolate from H_0 to H_1 by changing the coupling U. We use the same ansatz (3), but we use a different procedure to select a starting point for the search. We first do a single step ansatz using as Ψ_I the ground state of H_0 and targetting $H_{1/S}$, using the same optimization as above, calling the resulting parameters θ_X^1. We then use the Ψ_T from that optimization as Ψ_I for another single step targetting $H_{2/S}$, giving θ_X^2. We continue, targetting $H_{3/S}, \ldots,$, using Ψ_T from one step as Ψ_I for the next. This then gives a trial state using S steps for H_1 using parameters θ_X^b for $b = 1, \ldots, S$. We call this state the result of sequential optimization. We use those parameters as the starting point for a further search as before (i.e., a global variational search using the sequential optimization as the starting point), calling the resulting parameters the result of full optimization.

The results of the annealed variational method are listed in Tab. I. For $S = 3, N > 4$, this algorithm is seen to be only marginally worse than the global variational method, but we found that it was faster at finding the optimum (the number of energy evaluations required depended on N, but the annealed variational method always required fewer than the original method, and in some cases required a factor of five times fewer evaluations). The most important advantage of the annealed variational method is that it becomes significantly more accurate than the global variational method for $S > 3$, where now the error drops consistently with increasing S, without the convergence issues seen using the global variational method. While generally larger N is more difficult, $N = 4$ shows worse performance than $N = 6$, 8, and 10 after sequential optimization and only shows better performance after full optimization, perhaps due to the degeneracy of the non-interacting Hamiltonian for $N = 4$.

An additional model shown in the table is called "8^*". This has eight sites with the same geometry as before, but we change the hopping strength in the horizontal direction to $1/\sqrt{2}$ and change the sign on the horizontal hoppings which are periodic, thus effectively inserting a π-flux for hopping around a loop. The resulting single particle spectrum has $N/2 - 2$ states below the Fermi energy and four states at the Fermi energy, giving the same many-body ground state degeneracy as for $N = 12$ sites. The flux model and $N = 12$ show very similar errors and are both distinctly more difficult than the others. Nonetheless, using only 33 parameters we obtain over 95% overlap and with 45 parameters over 98% overlap for $N = 12$, despite the Hilbert space having a much larger dimension: $\binom{12}{6}^2 = 853776$ at the given filling. For $N = 10$, a 93.7% overlap is obtained with only 9 parameters, in a 63504-dimensional space (symmetries such as total spin slightly reduce this dimension).

E. Inexact Optimization: Gate Count and Run Time

Next we consider the effects of sampling error assuming one measures individual terms in the Hamiltonian on a quantum computer, considering a tradeoff between runtime and accuracy. Clearly, given a large enough number of samples, one can reduce the statistical error to the point that one can obtain the same accuracy as above. However, for a smaller number of samples, it is useful to modify the optimization algorithm. The trouble is that a small change in a point leads to only a small change in energy, and it thus requires a large number of samples to determine whether or not the energy improves.

We thus used a different algorithm and tested it for $N = 8$. Starting with all parameters equal to zero we randomly order the parameters and try increasing or decreasing each parameter by a constant, looking for an improvement This is repeated (with parameters re-ordered again randomly) with slightly changed constants, until no further improvement is found. To determine if there is an improvement, we sample at the given points until the difference between the energies becomes twice the

standard deviation (or until a maximum number of samples is done). On an average over ten runs, we were able to obtain more than 98% overlap with the ground state using 506 different point evaluations and $8.5 \cdot 10^5$ average samples per point, for 4.3×10^7 total samples.

Now we wish to estimate the number of gates required to obtain a single sample, including preparing the state Ψ_I, implementing the unitaries, and finally measurement. Ψ_I is a Slater determinant and can be prepared using Givens rotations [12] (using fewer gates than the strategies in Ref. 15). The unitaries U_U, U_h, U_v can be implemented efficiently using Jordan-Wigner cancellation techniques [16]. Finally, for measurement, the terms in the Hamiltonian can be broken into at most four commuting sets. For $N = 4n = 8, 12, \ldots$ these are the Hubbard terms, the vertical hopping terms, and two sets for the horizontal hopping terms. Slightly more complicated divisions of the hopping terms are needed for $N = 4n + 2 = 6, 10, \ldots$, while a model with more horizontal rows will require five sets.

The cost of implementing the unitaries dominates the measurement cost. We can measure h_U with a cost (in gate count) that is almost identical to the cost to implement U_U; see, for example, the discussion around Fig. 12 of Ref. 12 and references therein. Similarly, we can measure h_v with cost almost identical to the cost to implement U_v, while we can measure both sets of commuting terms in h_h with cost almost identical to the cost to implement U_h. Hence, since in a single run we only measure *one* of these four sets of terms (i.e., h_U or h_v or one of the two terms in h_h, assuming $N = 4n$), the measurement cost in a single run is roughly one-quarter the cost of a single step $U_U(\frac{\theta_U^b}{2})U_h(\theta_h^b)U_v(\theta_v^b)U_U(\frac{\theta_U^b}{2})$. In general, the cost of implementing $\prod_{b=1}^{S}\Big(U_U(\frac{\theta_U^b}{2})U_h(\theta_h^b)U_v(\theta_v^b)U_U(\frac{\theta_U^b}{2})\Big)$ scales linearly with S.

The Slater determinant can be prepared using Givens rotations and the fast Fourier strategy [12], requiring on the order of $N \log_2(N)$ Givens rotations. These rotations will involve sites which are further separated than the nearest neighbor sites which appear in U_h, U_v, requiring longer Jordan-Wigner strings; ignoring the cost of the strings, the cost to execute these rotations will be comparable to the cost to execute $\log_2(N)$ rotations by U_h or U_v. For fixed S, this would eventually dominate at large N, but we expect that S will need to increase with N too and that the dominant cost will remain the cost of implementing the steps. Consider the case of $N = 8$, for example. The initial ground state has two particles in the zero momentum state in the horizontal direction, one in the symmetric state in the vertical direction and one in the anti-symmetric state. There are also two particles in momentum states $\pm\pi/2$ in the horizontal direction, in the symmetric state in the vertical direction. We can prepare the ground state as follows. Label sites in the top row $1, 2, 3, 4$ and label sites in the bottom row $5, 6, 7, 8$ with 1 directly above 5, and so on. Initialize with particles in sites $1, 3, 4, 5$. Then, applying Givens rotations between pairs of sites $3, 7$; $4, 8$ to place particles initially in sites $3, 4$ in the symmetric state in the vertical direction. Next apply Givens rotations between pairs $1, 2$; $5; 6$ to bring the particle in site 1 into a symmetric state between $1, 2$ and the particle in site 5 into a symmetric state between $5, 6$. Again apply Givens rotations between pairs $1, 3$; $2, 4$; $5, 7$; $6, 8$ so that now the particle initially in site 1 is in a symmetric state between sites $1, 2, 3, 4$ while the particle in site 5 is in a symmetric state between sites $5, 6, 7, 8$. Thus, we have successfully occupied both states with zero momentum in the horizontal direction. This same procedure in fact also produces the desired particles in momentum states $\pm\pi/2$ so it prepares the desired ground state. This uses a total of 8 Givens rotations. The gate count cost is comparable to that to implement U_h and U_v.

For the case of quantum chemistry discussed below, the initial state is much simpler. It is a product state and so the cost to prepare is negligible compared to that to implement the unitaries.

A gate count estimate shows that about 1000 gates are required for $S = 2$ at $N = 8$ for a single run. This includes the cost to prepare Ψ_I, to implement the unitary rotations to create Ψ_T, and to perform measurement. To measure all terms, this must be multiplied by 4 as each run will give a measurement of the terms in *one* of the four commuting sets of terms. To give a rough estimate, if we ignore the possibility of parallelizing the circuits and assume a gate time of $1\mu s$ the total time would be $4.3 \times 10^7 \times 4 \times 1000 \times 10^{-6}$ seconds, or roughly 47 hours. This could be further improved using several devices in parallel to sample. Note that any given run requires only about 1000 gates, and thus poses only moderate demands on gate fidelities.

III. QUANTUM CHEMISTRY

A. The electronic structure Hamiltonian

We finally apply the method to three problems in quantum chemistry, where there are more terms than in the Hubbard model. Using the Born-Oppenheimer approximation by assuming that the nuclei behave classically and are localized, the Hamiltonian for the electronic degrees in second quantized form

$$H = \sum_{pq} h_{pq} c_p^\dagger c_q + \sum_{pqrs} h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s \tag{4}$$

has the most general form for two-body interactions, due to the long range nature of the Coulomb interaction. The indices p, q, r and s refer to spin-orbitals, combining the orbital and spin index.

As most basis sets used in quantum chemistry are non-orthogonal we follow the standard procedure [17] of first performing a (classical) Hartree-Fock calculation

S	ΔE [mHa]	P	ΔE [mHa]	P	ΔE [mHa]	P
	HeH^+		H_2O		BeH_2	
1	7.8	0.9970	23	0.9886	22	0.9825
2	1.7	0.9992	9.5	0.9950	6.6	0.9935
3	0.59	0.9998	7.6	0.9955	6.3	0.9937
4	0.26	0.9999	6.8	0.9959	5.8	0.9939
5	0.088	0.9999	3.2	0.9980	4.23	0.9954
6	0.14	0.9999	3.1	0.9982	1.85	0.9977

TABLE II: Variational Method applied to HeH^+ and H_2O. ΔE is the error in the ground state energy, P is absolute squared overlap with ground state. Note that the energy is slightly worse for HeH^+ at $S = 6$ than at $S = 5$; this indicates that the optimization algorithm did not find the true minimum in this case.

and then use the Hartree-Fock orbitals as an orthogonal basis set. On classical computers we can only simulate quantum algorithms for very small basis molecules.

Specifically we consider HeH^+ which has *two* electrons, using a P321 basis with $N_{\rm SO} = 8$ spin orbitals, H_2O which has 10 electrons, in an STO-6G basis with $N_{\rm SO} = 14$, and BeH_2 which has 6 electrons in a basis with $N_{\rm SO} = 14$. We use the Psi4 [18] quantum chemistry package to perform the Hartree Fock calculation and compute the matrix elements of the Hamiltonian (4). We also consider artificial hydrogen chains H_N, where we space hydrgen atoms along a line with distances of either 0.4614Å or 2Å. We used the global variational approach to optimize parameters for HeH^+, H_2O, and BeH_2 and we used the annealed variational approach to optimize H_N.

B. Variational ansatz for quantum chemistry

Instead of using $(O)(N_{\rm SO}^4)$ variational parameters in our ansatz, we group the matrix elements into three terms $H = H_{\rm diag} + H_{\rm hop} + H_{\rm ex}$, where

$$H_{\rm diag} = \sum_p \epsilon_p c_p^\dagger c_p + \sum_{p,q} h_{pqqp} c_p^\dagger c_p c_q^\dagger c_q \tag{5}$$

is a sum of diagonal terms,

$$H_{\rm hop} = \sum_{pq} h_{pq} c_p^\dagger c_q + \sum_{prq} h_{prrq} c_p^\dagger c_q c_r^\dagger c_r \tag{6}$$

contains normal and correlated hopping terms and

$$H_{\rm ex} = \sum_{pqrs} h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s \tag{7}$$

for p, q, r, s all distinct contains all other exchange terms.

The unitary $U_{\rm diag}(\theta) = \exp(i\theta H_{\rm diag})$ can be implemented exactly since all terms commute. We implemented the unitary $U_{\rm hop}(\theta)$ using an "interleaved" term ordering [16], approximating it as a product $U_{\rm hop}(\theta) = \prod_{p<q} U_{pq}(\theta)$ in arbitrary order, with

$$U_{pq}(\theta) = \exp\left[i\theta\left(h_{pq}c_p^\dagger c_q + \sum_r h_{prrq} c_p^\dagger c_r^\dagger c_r c_q + h.c.\right)\right], \tag{8}$$

where all terms in the exponential commute with each other. The unitary $U_{\rm ex}(\theta)$ was approximated as a product over factors

$$U_{pqrs}(\theta) = \exp\left[i\theta\left(h_{pqrs}c_p^\dagger c_r^\dagger q c_r c_s + h.c.\right)\right], \tag{9}$$

in an arbitrary order. As our final ansatz we use

$$\Psi_T = \prod_{b=1}^{S} U_{\rm diag}(\theta_{\rm diag}^b) U_{\rm hop}(\theta_{\rm hop}^b) U_{\rm ex}(\theta_{\rm ex}^b) \Psi_I. \tag{10}$$

The initial state Ψ_I is chosen to be the ground state of $H_{\rm diag}$ and the basis is a Hartree-Fock basis so that $H_{\rm hop}\Psi_I = 0$. Our results for the three benchmark molecules using various numbers of steps S are shown in Table II.

C. Unitary Coupled Cluster Ansatz

We have also performed simulations using UCC [9] for comparison. UCC method fixes $\Psi_T = \exp(T)\Psi_I$, for T an anti-Hermitian operator. The variational parameters are contained in the choice of T. Ref. 8 proposes using a Trotter-Suzuki method to implement the transformation $\exp(T)$. This increases the depth of the circuit, which we wish to avoid. We have found in our simulations that more accurate results are in fact obtained using large Trotter steps, with the most accurate results obtained using between two and four second-order Trotter-Suzuki steps; all results reported below are for two such steps. The method thus differs from the usual UCC, although we continue to refer to it as such. We used similar optimization procedures as before, exactly evaluating (to numerical error) the energy as a function of parameters.

We set

$$T = \sum_{p<q}(T_{p,q}c_p^\dagger c_q - h.c.) + \sum_{p<q,r<s}(T_{p,q,r,s}c_p^\dagger c_q^\dagger c_r c_s - h.c.), \tag{11}$$

using all possible quadratic and quartic terms which are compatbile with the symmetries of the system. We considered two cases, taking the parameters $T_{p,q}$, $T_{p,q,r,s}$ either all real or all imaginary. More accurate results were found using real choices; this is likely due to the fact that the wavefunction is real. More flexibility could be obtained by using general complex values, but at the cost of a further increase in parameters. In the real case, we can drop diagonal terms such as $c_p^\dagger c_p$.

For molecules, to find terms compatible with symmetries, we included all terms which had nonzero coefficients in the Hamiltonian (this should work for small molecules, while for large molecules some terms compatible with

symmetry will have a small enough coefficient that they get truncated). For HeH^+ and H_20, UCC is able to find the ground state up to numerical precision. This is no surprise because in both cases the number of variational parameters exceeds the Hilbert space dimension. For HeH^+, there are 192 parameters in the imaginary case (almost as many in the real case) compared to a Hilbert space dimension of 64 with the given number of up and down electrons (possibly further reduced by symmetry) while for H_20 there are 595 parameters compared to a Hilbert space dimension of 441.

This is an obstacle that we encounter when trying to simulate small molecules on a classical computer. The number of variational parameters in this method grows as $N_{\rm SO}^4$ while the Hilbert space dimension grows exponentially with $N_{\rm SO}$ at fixed filling fraction; eventually the exponential growth beats the polynomial (indeed, this is the whole reason for interest in a quantum computer), but since the number of parameters grows at such a high power of $N_{\rm SO}$, one needs fairly large systems to see this.

For BeH_2, UCC is able to achieve an energy error of 3×10^{-4}, at the cost of over 3.6×10^5 energy evaluations; using $60,000$ evaluations the error was 1.28×10^{-3}. In contrast, the Hamiltonian variational method used between $5000 - 10000$ evaluations for the data above. For BeH_2, the Hilbert space dimension is 1225, compared to 595 terms and so the Hilbert state dimension is smaller than the number of terms, but still comparable.

D. Hydrogen chains

Real applications will be concerned with molecules with larger $N_{\rm SO}$, for which the Hilbert space dimension is much larger than the number of variational parameters in UCC. In order to access this regime we turn to a different system, H_N chains, varying the number of atoms from $N = 2$ to $N = 10$. For $N = 2$, 4 and 6 the number of terms exceeds the Hilbert space dimension, while for $N = 8$ there are 2964 terms with a Hilbert space dimension of 4900 and for $N = 10$ there are 7230 terms with a Hilbert space dimension of $N = 63504$. We performed simulations at separations of either 0.4614Å or 2ÅAngstroms between the atoms. For the smaller spacings, there is a much higher overlap between the true ground state and the Hartree-Fock ground state, while for the larger spacing, the overlap becomes much smaller; this is similar to the Hubbard chain at small U/t compared to large U/t.

For the larger spacing, UCC found the ground state energy to very high accuracy for $N = 2$, but had errors of 0.0357mHa and 0.349mHa for $N = 4$ and 6. Since we expect that it should be possible to find the ground state exactly using this number of parameters, this may be due to difficulty in optimizing the function. For $N = 8$, UCC had an error of 44mHa while the annealed variational approach obtained errors of 67mHa , 15mHa and 3.84mHa for $S = 3$, 6, and 15 steps, respectively. While this requires a slightly larger circuit depth (the UCC depth with our given Trotter number is equivalent to $S = 2$ for the Hamiltonian variational), the results are more accurate and again required significantly fewer circuit evaluations. For $N = 10$, UCC had an error of 168mHa compared to errors of 98mHa and 24mHa for the annealed variational approach with $S = 2$ and 6 respectively.

We believe that the scaling trend is that as system sizes increase, the annealed variational approach will get increasingly better than the UCC; the UCC shows good performance at small sizes, perhaps due to the large number of parameters compared to the Hilbert space dimension, but gets increasingly less accurate at large sizes and is more difficult to optimize. For the smaller spacing, UCC performed relatively better but a similar trend was found. For $N = 8$, it took $S = 9$ for annealed variational to outperform UCC while for $N = 10$ it took only $S = 5$. The number of evaluations for UCC was again larger, requiring about 50000 and 23000 evaluations for $N = 8$ and 10 to obtain that accuracy, while annealed variational used typically 5000-10000 evaluations for $N = 10$; reducing the number of evaluations reduced the accuracy.

A more flexible method for quantum chemistry is to use multiple steps (as in the Hamiltonian variational approach) but allow every term to have a separate parameters (as in UCC). Optimizing all these parameters may become difficult. Note that for Hubbard, the unitary coupled cluster method requires a large increase in circuit depth compared to the Hamiltonian variational approach due to all the additional terms in T and hence it is likely to not be suitable for that context unless particular useful choices of a small number of terms in T can be found. The Hamiltonian variational approach can be regarded as a useful choice of which terms to take and how to vary them.

IV. GUIDANCE FROM PERTURBATION THEORY

If $H_0 = h_0$ and $H_1 = h_0 + \epsilon h_1$ and Ψ_I is the ground state of h_0, then Eq. (1) allows one to approximate the first order response with respect to ϵ. Let $\Psi_T = \exp(i\sigma_S h_0)\exp(i\tau_S h_1)\dots\exp(i\sigma_1 h_0)\exp(i\tau_1 h_1)\Psi_I$, and take the τ_a to be small. Then, to linear order in τ_a we have

$$\Psi_T = \Psi_I + i\sum_{b=1}^{S}\exp(i\phi_b h_0)\tau_b h_0 \Psi_I, \qquad (12)$$

where $\phi_b = \sum_{c=b}^{S}\tau_b$. Thus, the amplitude on an excited state Ψ' with energy ω above the ground state is equal to $i\sum_{b=1}^{S}\exp(i\phi_b\omega)\tau_b\langle\Psi'|h_0|\Psi_I\rangle$, while the exact first order result should be $-\omega^{-1}\langle\Psi'|h_0|\Psi_I\rangle$. Thus, we are approximating $-\omega^{-1}$ by $i\sum_{b=1}^{S}\exp(i\phi_b\omega)\tau_b$, which in Fourier space corresponds to an approximation by a finite number of δ-functions at positions ϕ_b. This approximation

only needs to be accurate for frequencies ω that correspond to an energy difference between an excited state and the ground state, and so in some cases it may be possible to achieve high accuracy with only a limited number of δ-functions.

Thus, one might wonder whether it is worth directly adding additional terms which better approximate the first order result with fewer steps. We have tried this in quantum chemistry above using both a rotation by $\sum_{pqrs}(1 - f(\omega_{pqrs}))h_{pqrs}c_p^\dagger c_q^\dagger c_r c_s$ and $\sum_{pqrs} f(\omega_{pqrs})h_{pqrs}c_p^\dagger c_q^\dagger c_r c_s$ where $f(x)$ is a function that is $1/x$ for sufficiently large x and 0 otherwise, and ω_{pqrs} is the change in energy obtained by applying h_{pqrs} to the ground state. This led to only a minor improvement over the above, at the cost of additional parameters. Thus, it seems like we are already well able to approximate the desired first order results.

We have also tried allowing multiple diagonal parameters in quantum chemistry, rotating by $c_p^\dagger c_p$ with different angles for different p. This also led to only a minor improvement.

V. RESOURCE ESTIMATES FOR PRACTICAL APPLICATIONS

We want to end with a discussion about resource requirements for practical applications that might go beyond what can currently be done classically.

A. Hubbard model

For the Hubbard model interesting results can be obtained for lattice sizes beyond 10×10, and thus with $N \geq 100$ sites, which requires about 200 qubits (or slightly more for ancillas if we want to parallelize the circuits). As the circuits to implement the various terms in the Hamiltonian can be efficiently parallelized [12], the parallel circuit depth will not substantially increase except for the potential need to go to a larger number of steps S. In order to be competitive to the best classical variational wave functions and distinguish competing ground states (such as stripe oder versus uniform superconducting ground states) we need to aim for an error of $\epsilon \approx 10^{-3}t$ in the energy *per site* [19].

Assuming a variance of order 1 for the measurement of the hopping terms $c_{i,\sigma}^\dagger c_{j,\sigma} + c_{j,\sigma}^\dagger c_{i,\sigma}$ and a reduced variance of order $1/U$ for the double occupancy term $c_{i,\uparrow}^\dagger c_{i,\uparrow} c_{i,\downarrow}^\dagger c_{i,\downarrow}$ (due to a suppression of double occupancy to about t/U for large U) we get an error estimate of

$$\epsilon^2 \approx U^2 \frac{t/U}{MN} + 4t^2 \frac{1}{MN}, \tag{13}$$

assuming M measurements for each of five terms (double occupancy, horizontal and vertical hopping for each of the spins) and using that we can perform N measurements in parallel in each of the runs. This is consistent with the number of samples that was found to be necessary comparing this to the values measures for small systems in Sec. II E. Using relevant values of $t = 1$, $U = 8$, $N = 100$, we obtain $M \approx 12/\epsilon^2 N \approx 120,000$ samples for each of the five terms, or about a total of 600,000 samples per energy evaluation. Again assuming gate times of order $1\mu s$, as in Sec. II E, the total estimated times needed to achieve convergence remains of the order of days when parallelizing the circuits, even considering that larger S and more energy evaluations might be required. While not trivial on a small quantum computer, this is potentially practical in the not too distant future.

B. Quantum chemistry

For quantum chemistry applications the required resources will be more demanding. Writing the Hamiltonian as $H = \sum_{i=1}^{N_{\text{terms}}} h_i \mathcal{O}_i$, where N_{terms} is the number of terms, and the numbers h_i the coefficients of the terms $\mathcal{O}_i$ we obtain for the error

$$\epsilon^2 = \sum_i \frac{|h_i|^2 \text{Var}(\mathcal{O}_i)}{M_i}, \tag{14}$$

assuming M_i measurements of each of the operators $\mathcal{O}_i$, where $\text{Var}(\ldots)$ denotes the variance in the measurement of the operator for the given trial state. Minimizing the error for a total number of measurements M, we choose $M_i \propto |h_i|$, and bounding the variances by $\text{Var}(\mathcal{O}_i) \leq 1$ we get

$$M \approx \frac{\left(\sum_i |h_i|\right)^2}{\epsilon^2} \tag{15}$$

As the variances of the (large) diagonal terms ϵ_p and h_{pqqp} will be small for orbitals where the occupation number is close to 1 or to 0 (this holds for all orbitals in the molecules studied above), we will drop these terms from the error estimates and consider just the off-diagonal terms in the following order of magnitude estimate. Some of the off-diagonal terms have small variance. For example, any term h_{pq} in which p, q both have occupation number close to 1 or close to 0 will have small variance. However many off-diagonal terms have large variance; for example, a term h_{pq} where p has occupation number close to 1 and q has occupation number close to 0 has large variance. Hence, as a rough estimate, we treat the variance of the off-diagonal as a constant of order unity.

We find for the sums of matrix elements $\sum_i |h_i|$ the values 11.3Ha for HeH^+, 12.3Ha for BeH_2 and 36Ha for H_2O, which results in 10^8 to 10^9 required samples for each energy evaluation to achieve an error of 1mHa. This is about a factor 1000 larger than for the Hubbard model, which already needed on the order of a few days to be optimized at an assumed $1\mu s$ gate time and measurement time. Considering additionally that the circuits are more complex, this will require a massively parallel cluster of

quantum computers to perform the simulations for small molecules in a reasonable time.

Moving from these toy problems that can easily be simulated classically to a more challenging problem, such as Fe_2S_2 in a small single particle basis (STO-3G with $N = 112$ spin-orbitals) we have $\sum_i |h_i| \approx 3$kHa, thus requiring about 10^{13} samples per energy evaluation. Even assuming that still only about 10^6 energy evaluations could be sufficient to optimize the ansatz, the required number of 10^{19} total samples is too large for the variational algorithm to be practical in this form. For this model, a single instance of the Hamiltonian circuit is on the order of 2×10^8 gate executions, which is an estimate of the cost to prepare a sample using either a UCC method truncated at all four fermi terms or the Hamiltonian variational method. This amounts to a total of 10^{26} gate operations assuming $20 - 30$ commuting terms are measured on each sample. This estimate assumed that all of the energy evaluations in the optimization were conducted at the same accuracy as the final estimate; potentially some of the earlier energy evaluations could be conducted at lower energy accuracy, but following a noisy search to the minimum may require that many steps of the evaluation be conducted at *higher* accuracy than the final evaluation, since we have found that often only a very small improvement in energy is obtained on each step of the search.

VI. DISCUSSION

We have presented a method of constructing variational states for arbitrary Hamiltonians, and tested it numerically. Preparation of quantum states using short quantum circuits is likely the lowest hanging fruit for small quantum computers to achieve quantum supremacy, and outperforming classical computers. No expensive quantum error correction may be needed if the gate fidelity is high enough to allow a short quantum circuit to succeed in preparing a state.

We find that a modest number of parameters enables us to obtain a very large overlap with the ground state, helping reduce the number of evaluations required to obtain good values of the parameters. Of course, the minimum possible circuit depth is attained simply by declaring the class of variational states to be "all states that can be constructed with unitary quantum circuits of given depth". However, this would require an impractical optimization. The method here allows us a flexible class of states with a small number of parameters, with the circuits chosen from the terms in the Hamiltonian. We find that for larger systems this approach becomes more efficient than ansatz wave functions based on the unitary coupled cluster formalism.

Although the state preparation is fast, the number of measurements required to estimate the energy with sufficient accuracy is a huge challenge. While the estimates for interesting quantum chemistry applications are astronomical and the variational algorithm for these problems thus impractical in its current form, the demands for the Hubbard model are less challenging. There, the smaller number of terms, limited energy range and translation invariance help reduce the required number of measurements to make a quantum variational approach potentially practical, although still demanding.

A less demanding application may be to use these ansatz wave functions for state preparation, which uses shorter circuits than would be needed if one instead prepared the state by adiabatic preparation. We have attempted to compare this by optimizing the parameters in an anneal for the Hubbard model: we consider an anneal using a linear path from the model at $U = 0$ to the final model at $U = 2$. This linear path was implemented by discrete time steps $\mathrm{d}t$ on the quantum computer, evolving using a second order Trotter-Suzuki for each time step (hence, the individual rotations in this Trotter-Suzuki evolve by time step $\mathrm{d}t/2$. Thus, there are two parameters that quantify the anneal: the total annealing time T, and the time step $\mathrm{d}t$. The ratio $T/\mathrm{d}t$ is the number of second order Trotter-Suzuki steps that must be implemented. We optimized these two parameters separately to obtain the minimum ratio that achieved 0.99 or higher overlap with the ground state. For the $N = 8$ model, this minimum was 8, to be compared to $S = 3$ to attain this accuracy using the variational method. For $N = 12$, the minimum was 640, to be compared to $S = 19$ using the variational method. We expect that larger system sizes may lead to even more significant gains in depth.

This state preparation may be useful for measuring properties of the states, in particular when combined with the quadratic speedup of Ref. 12. Additionally, this state preparation may be useful for annealing to even larger states. For example, if we construct a short circuit to create a Hubbard model ground state on 12 sites (and indeed, one outcome of this work is that we have been able to find such a circuit using a classical computer), then one could create two or more copies of this ground state and then anneal from the Hubbard Hamiltonian describing 24 sites partitioned into 2 decoupled sets of 12 sites to the Hubbard Hamiltonian for 24 coupled sites. This is an example of the "joining" procedure of Ref. 12, using the variational method to simplify the creation of the decoupled systems.

We optimized the parameters by a search procedure which treated the energy as a function of parameters of a black-box function to be optimized. It is in fact possible to also determine the derivative of this function using a quantum circuit, but this requires roughly doubling the depth of the circuit. It is not clear whether or not access to the derivative would improve the optimization.

One might further speculate whether optimizing over parameters by searching would be useful for the optimization problems of Refs. 10, 11, improving over the performance of an adiabatic algorithm. Further numerical work may help answer this.

Acknowledgments

This work was supported by Microsoft Research. MH and MT acknowledge hospitality of the Aspen Center for Physics, supported by NSF grant PHY-1066293.

[1] U. Schollwoeck, Ann. Phys. **326**, 96 (2011).
[2] G. Vidal,, Phys. Rev. Lett. **101**, 110501 (2008).
[3] F. Verstraete and J. I. Criac, cond-mat/0407066 (2004).
[4] M. Schwarz, K. Temme, and F. Verstraete, Phys. Rev. Lett. **108**, 110502 (2012).
[5] M. Schwarz, T. S. Cubitt, K. Temme, F. Verstraete, and D. Perez-Garcia, Physical Review A **88**, 032321 (2013).
[6] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. 98, 140506 (2007).
[7] C. M. Dawson, J. Eisert, and T. J. Osborne, Phys. Rev. Lett. **100**, 130501 (2008).
[8] A. Peruzzo et. al., Nature Comm. **5**, 4213 (2014).
[9] A. Taube and R. J. Bartlett, Int. J. Quant. Chem. **106**, 33933401 (2006).
[10] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1411.4028.
[11] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1412.6062.
[12] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak, and M. Troyer, arXiv:1506.05135.
[13] D. Jaksch and P. Zoller, Annals of Physics, **315** 52 (2005).
[14] M. J. D. Powell, Computer Journal **7**, 155 (1964).
[15] R. Somma, G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Phys. Rev. A **65**, 042323 (2002).
[16] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, QIC **15**, 1 (2015).
[17] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Molecular Physics **109**, 735 (2011).
[18] "Psi4: An open-source ab initio electronic structure program," J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, WIREs Comput. Mol. Sci. **2**, 556 (2012). (doi: 10.1002/wcms.93).
[19] P. Corboz, T. M. Rice, and M. Troyer, Phys. Rev. Lett. **113**, 046402 (2014).