

THE TROTTER STEP SIZE REQUIRED FOR ACCURATE QUANTUM SIMULATION OF QUANTUM CHEMISTRY

DAVID POULIN

Département de Physique, Université de Sherbrooke, Québec, Canada

MATTHEW B. HASTINGS

Station Q, Microsoft Research, Santa Barbara, CA 93106-6105, USA

Quantum Architectures and Computation Group

Microsoft Research, Redmond, WA 98052, USA

DAVE WECKER NATHAN WIEBE

Quantum Architectures and Computation Group

Microsoft Research, Redmond, WA 98052, USA

ANDREW C. DOHERTY

SCentre for Engineered Quantum Systems, School of Physics

the University of Sydney, Sydney, NSW 2006, Australia

MATTHIAS TROYER

Theoretische Physik, ETH Zurich, 8093 Zurich, Switzerland

Received June 25, 2014

Revised September 1, 2014

The simulation of molecules is a widely anticipated application of quantum computers. However, recent studies [1, 2] have cast a shadow on this hope by revealing that the complexity in gate count of such simulations increases with the number of spin orbitals N as N^8 , which becomes prohibitive even for molecules of modest size $N \sim 100$. This study was partly based on a scaling analysis of the Trotter step required for an ensemble of random artificial molecules. Here, we revisit this analysis and find instead that the scaling is closer to N^6 in worst case for real model molecules we have studied, indicating that the random ensemble fails to accurately capture the statistical properties of real-world molecules. Actual scaling may be significantly better than this due to averaging effects. We then present an alternative simulation scheme and show that it can sometimes outperform existing schemes, but that this possibility depends crucially on the details of the simulated molecule. We obtain further improvements using a version of the coalescing scheme of [1]; this scheme is based on using different Trotter steps for different terms. The method we use to bound the complexity of simulating a given molecule is efficient, in contrast to the approach of [1, 2] which relied on exponentially costly classical exact simulation.

Keywords:

Communicated by: R Jozsa & B Terhal

1 introduction

It has been 30 years since Feynman suggested that a quantum information processor could in principle simulate the dynamics of quantum systems efficiently [3], and this idea has since been formalized and studied in great detail [4, 5, 6, 7, 8, 9, 10]. Based on this knowledge, it has been advocated that one of the first practical applications of quantum information processors will be the simulation of molecules [11, 12, 13, & references therein]. This is motivated by the fact that state-of-the-art, high-precision numerical simulations are limited to molecules with at most 50-70 spin orbitals, where a spin orbital denotes a choice of both orbital and spin quantum numbers [14, 15, 16, 17, 18]. Thus, a quantum computer using as little as 100 logical qubits has enough storage capacity to efficiently perform a simulation which is otherwise intractable classically.

However, closer scrutiny of the problem has recently revealed that, while the memory requirements are indeed relatively modest, the duration of such quantum simulations using the proposed techniques are far too demanding [1]. Significant improvements were obtained by optimizing the quantum simulation circuitry [2], but the required time-resources remain prohibitive.

To understand the origin of this problem, recall that the time-evolution operator associated to a Hamiltonian H is $U_H(t) = e^{-iHt}$. For a Hamiltonian expressed as the sum of m terms $H = \sum_{\alpha=1}^m H_{\alpha}$, we can use the Trotter-Suzuki (TS) decomposition to approximate the “infinitesimal” time-evolution operator $U_H(\Delta_t)$ by a product of m infinitesimal time-evolution operators $U_{\alpha}(\Delta_t) = e^{-iH_{\alpha}\Delta_t}$, each generated by a single term H_{α} from the Hamiltonian. Repeating $1/\Delta_t$ times yields the time-evolution operator for a unit time. We can deduce two immediate consequences of this approach. On the one hand, the number of gates N_g required to implement a single infinitesimal time step will scale at least proportionally to the number of terms m in the Hamiltonian. On the other hand, the error in the TS approximation also increases as some power of m , forcing us to adopt a smaller time step Δ_t , and hence a slower simulation [8, 10, 1].

In the case of a molecule, the Coulomb force generates quartic terms $c_p^{\dagger}c_q^{\dagger}c_r c_s$ in fermion creation and annihilation operators. For small molecules, there can be as many as $\sim N^4$ such distinct terms, where N is the number of relevant spin orbitals of the molecule. The standard approach to the problem [11, 12, 13, 1, 2] applies the TS decomposition directly to these $m = \mathcal{O}(N^4)$ terms, each of which can be implemented using (at best) a constant number of gates [2] on average given a gate set containing one- and two-qubit Clifford operations as well as arbitrary single-qubit controlled rotations. Thus, this technique unavoidably entails at least a $\sim N^4$ cost per infinitesimal time-step.

Furthermore, to achieve a constant accuracy, the time evolution operator needs to be broken into a number of steps which increases as some power of N . In [1], a rigorous upper bound on the TS error was derived which indicates that $1/\Delta_t = \mathcal{O}(N^5)$ infinitesimal time-steps are sufficient to achieve a constant accuracy. This scaling was confronted with exact numerical simulations which revealed that $1/\Delta_t = \mathcal{O}(N^{4-5})$ infinitesimal time-steps were indeed sufficient, resulting in a total complexity of $\mathcal{O}(N^8)$ at best. This result is already exorbitant for a molecule with $N \sim 100$ spin orbitals.

These exact simulations were not carried on real-world molecules, but instead used artificial molecules drawn from a random ensemble meant to reproduce the statistical properties of real

molecules. This is motivated by the fact that exact numerical simulations are restricted to small molecules $N \sim 24$ and the limited availability of interesting real-world molecules of small size.

In this article, we assess the complexity of quantum simulations without resorting to costly exact simulations, but instead directly and efficiently evaluate an upper bound derived in [1]. Because quantum simulations will be used precisely for those molecules that are too large to be amenable to classical simulations, this efficient and rigorous error assessment is also of independent interest. In this way, we are able to predict the complexity for the quantum simulations of real-world molecules of size up to $N \sim 100$, and find that $1/\Delta_t = \mathcal{O}(N^{1.5-2.5})$ time-steps are sufficient to achieve a constant-accuracy simulation, and the true cost may be even lower. This result clearly indicates that the statistical properties of those molecules are not accurately reproduced by the random ensemble used in [1], and that the complexity of simulating real-world molecules is substantially lower than anticipated. The bounds we use are given in section 3, and numerical estimates are in section 5.

In section 4, we also propose an alternative simulation scheme based on a decomposition of the Hamiltonian into a sum of $m = \mathcal{O}(N^2)$ terms, each of which can be implemented with $\mathcal{O}(N^2)$ gates. While this leads to an identical gate count $N_g \sim N^4$ per infinitesimal time step, reducing the number of terms in the TS decomposition can significantly reduce the resulting error, thus enabling a larger time step Δ_t . We find that this alternative simulation scheme can sometimes outperform existing schemes, but that the performance of each scheme depends greatly on the details of the simulated molecule. Our efficient error assessment technique comes in handy at this point because it enables us to determine which simulation technique is best suited for a given molecule. This illustrates that other decompositions of the Hamiltonians could lead to substantial gains.

These estimates for the performance of both the standard and alternative decomposition schemes are given in section 5, using the bounds of section 3 to analyze a variety of real-world molecules. We further present numerical experiments on TS error from exact simulations using a re-ordering of terms.

Finally, in section 6 we present a scheme which uses different TS steps for different terms. This improves on the coalescing scheme of Ref. [1] by modifying how terms are chosen to be coalesced. Numerical simulations show a significant reduction in work without increasing the error.

In general, in this paper when estimating work we will count the number of *gates* required. The nesting scheme of Ref. HWBT14a means that in many cases we can parallelize such that the *depth* of the circuit will be proportional to the number of gates divided by N . In a few places we comment on this more explicitly.

2 Background

In this section, we present the problem more formally, and review the standard simulation approach [11, 12, 13, 1, 2].

2.1 Hamiltonian

Our starting point is a Hamiltonian of the form

$$H = \sum_{pq} h_{pq} c_p^\dagger c_q + \sum_{pqrs} h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s, \quad (1)$$

where c_p^\dagger and c_p are fermion creation and annihilation operators for the spin orbital p . There are N spin orbitals which have been chosen using, e.g. Hartree-Fock calculations. To get a constant-accuracy estimate of the ground-state energy of the corresponding molecule, we need to simulate the time-evolution operator $U_H(t)$ for some constant time t , which we will set to unity in what follows and drop the explicit t variable when unnecessary.

The first term of Eq. (1) describes free fermions. We will use the shorthand notation $H_{pq} = h_{pq} c_p^\dagger c_q$ for these terms and note that $H_{pq}^\dagger = H_{qp} \Leftrightarrow h_{pq}^* = h_{qp}$ is required for H to be Hermitian. A nice property of free-fermion operators is that they form a closed Lie algebra, i.e., for $H = \sum_{pq} h_{pq} c_p^\dagger c_q$ and $H' = \sum_{pq} h'_{pq} c_p^\dagger c_q$, we have

$$[H, H'] = \sum_{pq} [h, h']_{pq} c_p^\dagger c_q, \quad (2)$$

where $[h, h']_{pq}$ simply refers to the (p, q) matrix element of $[h, h']$. Throughout, we will use upper-case letters H, U, \dots to denote operators on the 2^N -dimensional Hilbert space, and lower-case letters h, u, \dots for matrices on the N -dimensional orbital space. The group $\mathcal{U}(N)$ acts on the orbital space, and Eq. (2) simply shows that free fermion Hamiltonians form a (reducible) representation of $\mathcal{U}(N)$. In other words, $e^{-iHt} c_p e^{iHt} = \sum_q u_{pq} c_q$ where $u = e^{-iht} \in \mathcal{U}(N)$.

The second term of Eq. (1) represents interactions. We will use the shorthand notation $H_{pqrs} = h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s$. We note that the substitution $h_{pqrs} \leftarrow \frac{h_{pqrs} + h_{qpsr}}{2}$ leaves the Hamiltonian invariant, so we will henceforth assume that $h_{pqrs} = h_{qpsr}$.

With the exception of section 4, where all four fermion terms are considered on an equal footing, we will reserve H_{pqrs} to refer to terms where p, q, r, s are all distinct and otherwise refer to H_{prrq} terms and H_{pqqp} terms to refer to the case that only 3 or 2 of the indices are distinct. Note that the terms H_{pqqp} are diagonal in the occupation number basis. Similarly, we use H_{pp} to refer to terms proportional to $c_p^\dagger c_p$ and H_{pq} to refer to terms proportional to $c_p^\dagger c_q$ for $p \neq q$.

2.2 Simulating time evolution

The general strategy to simulate the time-evolution generated by a Hamiltonian which is the sum of m simple terms $H = \sum_{\alpha=1}^m H_\alpha$ proceeds in two phases. First, we decompose the total time evolution into a sequence of infinitesimal steps $U_H(1) = [U_H(\Delta_t)]^{1/\Delta_t}$. Second, we use the second-order (or higher) TS decomposition to approximate each infinitesimal steps

$$U_H(\Delta_t) \approx U_H^{\text{TS}}(\Delta_t) \quad (3)$$

$$:= U_m(\frac{\Delta_t}{2}) \dots U_2(\frac{\Delta_t}{2}) U_1(\Delta_t) U_2(\frac{\Delta_t}{2}) \dots U_m(\frac{\Delta_t}{2}), \quad (4)$$

where $U_\alpha(t) = e^{-iH_\alpha t}$. For the Hamiltonian of Eq. (1), the index α would range over all the pairs $\alpha = (p, q)$ and quartets $\alpha = (p, q, r, s)$ of spin orbitals, so $m \sim N^4$.

We claim that the evolution generated by any free Hamiltonian can be implemented exactly with $\mathcal{O}(N^2)$ gates. This follows from the Householder transformation [19] which shows that we can decompose $u = e^{-ih} \in \mathcal{U}(N)$ into a sequence of N^2 unitary matrices $u = v_1 v_2 \dots v_{N^2}$, where each v_α is trivial everywhere except on a 2×2 block (p_k, q_k) . Writing $v_k = e^{-ig^k}$, we can express $U_H = V_1 V_2 \dots V_{N^2}$ where $V_k = \exp(-ig_{p_k q_k}^k c_{p_k}^\dagger c_{q_k} + h.c.)$ is a time-evolution operator generated by a free-fermion operator acting only on 2 modes. These operators V_k can be implemented with a constant number of gates on average using the technique of Ref. [2] to cancel Jordan-Wigner strings.^a We note that this simulation of free Hamiltonians is exact, in contrast to previous approaches [11, 12, 13, 1, 2] that rely on TS approximations. This is crucial for the alternative simulation scheme we will present in Sec. 4 because it makes frequent use of such free evolution operators to implement spin orbital basis changes. Further, it is possible to choose the ordering of the free-fermion evolution operators such that nesting as in Ref. HWBT14a can be used to reduce the depth to $\mathcal{O}(N)$.

Simulating the interaction terms is more demanding. In Ref. HWBT14a, it was shown how the time evolution generated by each term H_{pqrs} can be implemented using a constant number of gates on average. Since there are far more interaction terms than free terms, the overall circuit complexity is set by them, and the number of gates N_g required to implement a single infinitesimal time-evolution operator Eq. (4) is therefore $\sim N^4$.

In practice, although it is possible to simulate all the free fermion terms in Eq. (1) exactly without any TS error as described above, it would likely be preferred to use the scheme of Ref. HWBT14a in which these terms are interleaved with terms H_{prrq} so that after a term $c_p^\dagger c_q$ is executed, it is followed by terms $c_p^\dagger c_r^\dagger c_r c_q$. In this way, in a Hartree-Fock basis these terms tend to cancel each other, reducing the TS error.

2.3 Error per infinitesimal time-step

Following the analysis of [1, appendix B], the approximation in Eq. (4) results in an error bounded by

$$\delta^{\text{TS}} := \|U_H(\Delta_t) - U_H^{\text{TS}}(\Delta_t)\| \tag{5}$$

$$\leq \sum_{\alpha=1}^m \left\| \left[[H_\alpha, H_{>\alpha}], H_\alpha \right] + \left[[H_{>\alpha}, H_\alpha], H_{>\alpha} \right] \right\| \Delta_t^3 \tag{6}$$

where $H_{>\alpha} = \sum_{\beta>\alpha} H_\beta$. Note that $[H_{pqrs}, H_{p'q'r's'}] = 0$ unless one of the indices is repeated. Thus, of all the $m^3 = \mathcal{O}(N^{12})$ terms $[[H_\alpha, H_\beta], H_\gamma]$ appearing in Eq. (6), only mK^2 will be non-zero, where $K = \mathcal{O}(N^3)$ is the maximum number of terms H_β with which a given H_α does not commute. Defining $\Lambda = \max_\alpha \|H_\alpha\|$, which is a constant in the present case, this leads to the upper bound

$$\delta^{\text{TS}} \leq mK^2 \Lambda^3 \Delta_t^3 = \mathcal{O}(N^{10}) \Delta_t^3. \tag{7}$$

The error in TS evolution gives an upper bound to the error in the eigenvalues of the unitary operator to evolve for a time step Δ_t . This translates to an error in the ground state

^aThis constant average cost can only be achieved when an appropriate sequence of free evolution operators are executed, which will always be the case here. Otherwise the cost is $\sim N$ due to the gates that are required to implement the Jordan-Wigner mapping of fermion orbitals into qubits.

energy

$$\begin{aligned}\Delta E^{\text{TS}} &\leq \sum_{\alpha=1}^m \left\| [[H_{\alpha}, H_{>\alpha}], H_{\alpha}] + [[H_{>\alpha}, H_{\alpha}], H_{>\alpha}] \right\| \Delta_t^2 \\ &= \mathcal{O}(N^{10})\Delta_t^2.\end{aligned}\tag{8}$$

This implies that the time step Δ_t needs to be as little as $\Delta_t = \mathcal{O}(N^{-5})$ to achieve a constant precision, which is the rigorous upper bound derived in [1].

An alternate route to understanding error is to use the Baker-Campbell-Hausdorff (BCH) formula to compute the error in the TS approximation as a power series in Δ_t . Remarkably, the lowest order term in the power series gives an error that is within a constant factor of that resulting from the bound above, implying that the bound is close to optimum. The advantage of the BCH formula is that it gives a tighter estimate of error for small Δ_t . The advantage of the bound above is that it works for all Δ_t while using the BCH formula it would be necessary also to consider higher-order terms in the power series.

The accuracy of the upper bounds in (6) can be assessed by comparing the values it predicts to the asymptotic formula for the error. The BCH formula can provide this by giving the leading order behavior of the *effective Hamiltonian* that the TS simulation evolves under. Applying the formula iteratively to (4) to order Δt^3 yields

$$H_{\text{eff}} = H - \frac{1}{12} \sum_{\alpha \leq \beta} \sum_{\beta} \sum_{\alpha' < \beta} [H_{\alpha}(1 - \frac{\delta_{\alpha,\beta}}{2}), [H_{\beta}, H_{\alpha'}]] \Delta_t^2.\tag{9}$$

The error in the TS expansion for a single time step is therefore

$$\|e^{-iH\Delta_t} - e^{-iH_{\text{eff}}\Delta_t}\| \leq \|H - H_{\text{eff}}\| \Delta_t.\tag{10}$$

Similarly, perturbation theory gives that the error in the ground state energy invoked by using the TS formula is

$$\sum_{\alpha \leq \beta} \sum_{\beta} \sum_{\alpha' < \beta} \frac{1}{12} \langle \Psi_0 | [H_{\alpha}(1 - \frac{\delta_{\alpha,\beta}}{2}), [H_{\beta}, H_{\alpha'}]] \Delta_t^2 | \Psi_0 \rangle,\tag{11}$$

where $|\Psi_0\rangle$ is the ground state of the true Hamiltonian H and terms of order $\mathcal{O}(\Delta_t^3)$ have been neglected. These formulas are valuable because they exactly predict the errors in the simulation as Δ_t approaches zero.

If we apply the triangle inequality to (10) then we obtain a comparable result to (6) to within roughly a factor of 12 if we neglect the $\mathcal{O}(\Delta t^3)$ terms. This means that (6) can be expected to be reasonably tight in the regime where the error scales proportional to Δ_t^2 , up to errors incurred by using the triangle inequality.

2.4 *Random ensemble*

While the bound derived in the previous section is rigorous, it is possible that the actual accuracy achieved in a quantum simulation is much better. In Ref. WBCH13a, this question was addressed using full classical simulations of the quantum simulation algorithm itself. While such a numerical simulation are certainly not tractable (this is the entire point of using

a quantum information processor), they can be realized for molecules of modest sizes, and this can provide an idea of the general scaling. Due to the limited availability of interesting real-world molecules of small size, the simulations of [1] were carried on artificial molecules drawn from a random ensemble.

Specifically, for Hamiltonians of this ensemble, only a fraction $F \approx 0.8\%$ of the entries h_{pqrs} of the Hamiltonian are assigned non-zero values, those non-zero values have random signs and magnitudes following the distribution

$$\text{Prob}(|h_{pqqp}|) = \text{Uniform}(0, 0.5) \tag{12}$$

$$\text{Prob}(|h_{pqqr}|) = \text{Exponential}(0.2) \tag{13}$$

$$\text{Prob}(|h_{pqrs}|) = \text{Exponential}(0.1). \tag{14}$$

These simulations revealed that $\Delta_t \sim N^{-4.32} N^{-5.08}$ depending on the electronic filling factor. In our numerical studies, we have also considered the ensemble obtained with $F = 1$.

3 Efficiently computable error bound

There is an obvious way of improving the upper-bound derived above using efficient numerical calculations. First, notice that the first term $[[H_\alpha, H_{>\alpha}], H_\alpha]$ of Eq. (6) contains only K terms in contrast to the second term $[[H_{>\alpha}, H_\alpha], H_{>\alpha}]$ which contains K^2 . Thus, we will henceforth drop the first term for simplicity. Then, using the Jacobi identity $[[A, B], C] = [A, [B, C]] - [B, [A, C]]$, we see that $[[H_\alpha, H_\beta], H_{\beta'}]$ is zero unless two of the three pairs of terms from H_α , H_β and $H_{\beta'}$ do not commute. Combined with the triangle inequality, we obtain the following upper bound to the second term of Eq. (6)

$$\delta^{\text{TS}} \leq 4 \sum_{\alpha} \|H_\alpha\| \left(\sum'_{\beta} \|H_\beta\| \right)^2 \Delta_t^3, \tag{15}$$

where Σ' is the sum restricted to the terms β for which $[H_\alpha, H_\beta] \neq 0$. This gives an error in ground state energy

$$\Delta E^{\text{TS}} \leq 4 \sum_{\alpha} \|H_\alpha\| \left(\sum'_{\beta} \|H_\beta\| \right)^2 \Delta_t^2, \tag{16}$$

In Eq. (15), we made double use of the inequality

$$\|[H_\alpha, H_\beta]\| \leq 2\|H_\alpha\| \cdot \|H_\beta\|. \tag{17}$$

We note however that in the case where each term H_α represents a H_{pqrs} term, this inequality is tight up to a factor of 2, provided that the terms do not commute. For instance, $\|[H_{pqrs}, H_{p'q'ps'}]\| = |h_{pqrs}| \cdot |h_{p'q'ps'}| \cdot \|c_p^\dagger c_q^\dagger c_s^\dagger c_r c_s\| = |h_{pqrs}| \cdot |h_{p'pr's'}|$. Thus, given a description of the molecule in terms of the m coefficients h_{pqrs} , the bound Eq. (15) can be evaluated with a complexity linear in m . Since $m \sim N^4$, the evaluation of this bound could become numerically demanding for large molecules $N \gg 100$. Moreover, the evaluation of a similar bound for the alternative simulation approach of Sec. 4 scales like N^9 , so it becomes necessary to develop more efficiently ways of evaluating Eq. (15). This can be achieved by

Monte Carlo sampling from the sum rather than evaluating every term. More precisely, we can generate M triples of indices α_k , β_k and β'_k such that both $[H_{\alpha_k}, H_{\beta_k}]$ and $[H_{\alpha_k}, H_{\beta'_k}]$ are non-zero, and estimate the bound in Eq. (15) by

$$\delta^{\text{MC}} = \frac{L}{M} \sum_{k=1}^M |h_{\alpha_k}| \cdot |h_{\beta_k}| \cdot |h_{\beta'_k}| \quad (18)$$

where L is the total number of triplets α_k , β_k and β'_k that obey the above conditions. The relative error on this estimate is $\sigma/\delta^{\text{MC}}\sqrt{M}$, where σ is the standard deviation of δ^{MC} and can also be estimated by sampling. In all the cases in which we were forced to use Monte Carlo sampling to estimate error upper bounds, we have used $M = 10^5$ samples and observed that the relative error on our estimate of δ^{MD} was about 1% or less. Moreover, Monte Carlo sampling is used only for the alternative simulation approach described in Sec. 4.

4 Alternative decomposition

In this section, we present an alternative way of simulating molecules described by Hamiltonians of the form Eq. (1) and show how to efficiently evaluate its accuracy.

4.1 Completing the square

The scheme we propose is based on the idea of expressing the interacting Hamiltonian as the sum of squares of free terms plus additional free terms, i.e., completing the squares. We begin by demonstrating how this is realized. By joining indices $(p, r) = \alpha$ and $(q, s) = \beta$, we can view the tensor h_{pqrs} as a matrix $h_{\alpha\beta}$. This matrix is symmetric given our convention $h_{pqrs} = h_{qpsr}$, so it can be diagonalized into $h_{\alpha\beta} = \sum_{\gamma} u_{\alpha\gamma} d_{\gamma} u_{\gamma\beta}^*$ with some unitary matrix u and real vector d .

Consider the operator $K_{\gamma} = \sum_{rs} u_{\gamma qs}^* c_q^{\dagger} c_s$, where we have partly converted back our notation $(q, s) = \beta$. The interaction Hamiltonian can now be expressed as $-\sum_{\gamma=1}^{N^2} d_{\gamma} K_{\gamma}^{\dagger} K_{\gamma} + H_1$, where $H_1 = \sum_{pqrs} h_{pqrs} c_p^{\dagger} c_s$ is a quadratic term. While K_{γ} are not Hermitian operators, they can be written as the sum of Hermitian and skew-Hermitian operators $K_{\gamma} = (\hat{K}_{\gamma} + i\tilde{K}_{\gamma})$, leading to $K_{\gamma}^{\dagger} K_{\gamma} = \hat{K}_{\gamma}^2 + \tilde{K}_{\gamma}^2 + i[\hat{K}_{\gamma}, \tilde{K}_{\gamma}]$. The commutator results in a free Hamiltonian. Defining $G_{\gamma} = \sqrt{|d_{\gamma}|} \hat{K}_{\gamma}$ and $G_{\gamma+N^2} = \sqrt{|d_{\gamma}|} \tilde{K}_{\gamma}$, we have expressed the Hamiltonian Eq. (1) as a sum of squares of free Hamiltonians

$$H = H_0 - \sum_{\gamma=1}^{2N^2} \eta_{\gamma} G_{\gamma}^2 \quad (19)$$

where $\eta_{\gamma} = \eta_{\gamma+N^2} = \text{sign}(d_{\gamma})$ and H_0 is a free Hamiltonian containing the initial free term of Eq. (1), the H_1 component above, plus the various commutators $i[\hat{K}_{\gamma}, \tilde{K}_{\gamma}]$.

4.2 Simulation

Now that we have expressed the Hamiltonian in the form Eq. (19), the simulation proceeds by using a second-order TS decomposition Eq. (4) as above, but this time using only $m = 2N^2 + 1$ terms; H_0 and the G_{γ}^2 . To implement an infinitesimal time evolution $U_{\gamma} = \exp(-i\eta_{\gamma} G_{\gamma}^2 \Delta_t)$ generated by a G_{γ}^2 term, we first change the basis of orbitals so as to diagonalize this term. Given $G_{\gamma} = \sum_{pq} g_{pq}^{\gamma} c_p^{\dagger} c_q$, we can diagonalize the matrix g^{γ} into $w^{\gamma} g^{\gamma} w^{\gamma\dagger} = \epsilon^{\gamma}$ where ϵ^{γ}

is a diagonal matrix. Thus, written in the orbital basis $f_p^\gamma = \sum_q w_{pq}^\gamma c_q$, the term $G_\gamma = \sum_p \epsilon_p^\gamma f_p^\gamma f_p^\gamma$ is diagonal. Using techniques of Ref. [2], it follows that in this orbital basis, the infinitesimal time evolution U_γ can be realized using $\mathcal{O}(N)$ gates.

The complexity of each U_γ therefore stems from the $\mathcal{U}(N)$ orbital basis change w^γ . But such a basis change is equivalent to time-evolution under a free Hamiltonian, so its complexity is $\mathcal{O}(N^2)$ as explained in Sec. 2.2. Thus, the overall complexity of simulating a single infinitesimal time evolution is the number of terms in the TS decomposition $\mathcal{O}(N^2)$, times the complexity of implementing a single term $\mathcal{O}(N^2)$, resulting in the same scaling $\mathcal{O}(N^4)$ as the method [2] outlined in Sec. 2.2.

4.3 Error per infinitesimal time-step

We now evaluate the general error bound Eq. (6) in the special case where each term $H_\alpha = \pm G_\alpha^2$ is the square of a free fermion Hamiltonian. While we could proceed the same way as what led to Eq. (15), we note that the bound Eq. (17) is not tight except in the special case explained in Sec. 3. So instead, we make use of the bound

$$[G_\alpha^2, G_\beta^2] \leq 4 \|G_\alpha\| \cdot \|G_\beta\| \cdot \|[G_\alpha, G_\beta]\|, \tag{20}$$

which, inserted into in Eq. (7) and combined with the triangle inequality, yields

$$\delta^{\text{TS}} \leq 8 \sum_{\beta, \beta' > \alpha} \|G_\alpha\| \cdot \|G_\beta\| \cdot \|G_{\beta'}\| \cdot \|[G_\alpha, G_\beta]G_{\beta'}\| \Delta_t^3 \tag{21}$$

Note that each of the terms $G_\alpha, G_\beta, G_{\beta'}$ and $[[G_\alpha, G_\beta], G_{\beta'}]$ are free fermion operators, so their norm can be computed efficiently numerically. Indeed, the operator norm of a free Hamiltonian $H = \sum_{pq} h_{pq} c_p^\dagger c_q$ can be computed from the spectrum ϵ_p of the corresponding h as $\|H\| = \max\{E_+, -E_-\}$, where E_+ (E_-) is the sum of the positive (negative) eigenvalues ϵ_p . It follows that computing the norm of such an operator has complexity $\mathcal{O}(N^3)$, and therefore evaluating the bound Eq. (21) has overall complexity $\mathcal{O}(N^9)$. For this reason, we have resorted to Monte Carlo sampling as explained in Sec. 3 to evaluate Eq. (21).

5 Numerical results

We have evaluated the bounds Eq. (15) and Eq. (21) associated to the two simulation schemes. Since δ^{TS} is the error of a single infinitesimal TS time-step and that there are in total $1/\Delta_t$ such time-steps, a bound of the form $\delta^{\text{TS}} \leq \Gamma \Delta_t^3$ implies that $\sqrt{\Gamma}$ time-steps are required to achieve a constant accuracy. Fig. 1 therefore presents the numerical values obtained for Eq. (15) and Eq. (21) in terms of the total number of TS steps $1/\Delta_t$. The bounds were evaluated for real-world molecules, for artificial molecules chosen from the ensemble described at Eqs. (12-14), and for artificial molecules chosen from a different random ensemble where non-zero values were assigned to *all* h_{pqrs} coefficients following the distribution Eqs. (12-14). We refer to these two random ensembles as sparse and full respectively.

5.1 Artificial molecules

For the artificial molecules, we find that the fraction F of non-zero coefficients H_{pqrs} has little impact on the scaling of the complexity with N , and both simulation schemes display a complexity near $N^{4.5}$ - N^5 , in good agreement with the findings of [1] obtained from full

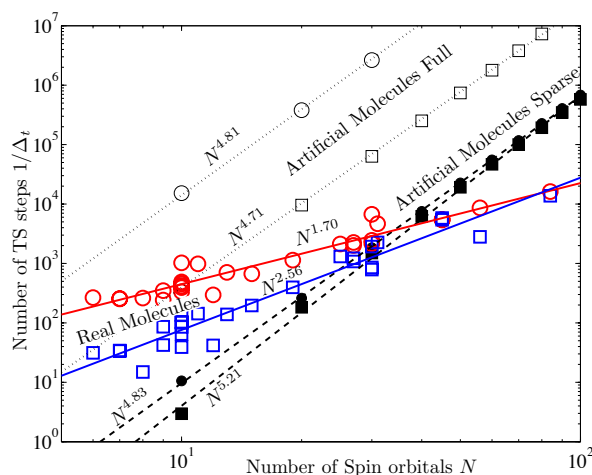


Fig. 1. Number of TS steps required to achieve constant accuracy on the energy measurement. Circles correspond to the standard approach described in Sec. 2 and is derived from Eq. (15). Squares correspond to the scheme described in Sec. 4 and is derived from Eq. (21). Coloured marks (full lines) are for a suite of real-world molecules. Filled black marks (dash lines) are for artificial molecules drawn from the sparse Hamiltonian ensemble $F = 0.008$ and hollow marks (dotted lines) are for artificial molecules drawn from the full Hamiltonian ensemble $F = 1$. The molecules (number of spin orbitals) studied are HF (6), NH₃ (8), CH₄ (9), N₂ (9), F₂ (10), HCl (10), LiF (10), NF (10), O₂ (10), H₂S (11), H₂CO (12), H₂O (7,11,13,19,25,27,31), O₃ (30), CO₂ (15,27,45), CO (10), F₂S₂ (56,84).

numerical simulations. However, as we discuss below, this is likely an artifact of small sizes and the true scaling even for the artificial molecules is likely much better.

Although the scaling with N is largely insensitive to the chosen random ensemble, we find that this choice greatly affects the constant pre factor. This is anticipated from the fact that a denser Hamiltonian will have a correspondingly higher norm, which will directly translate into a higher TS error bound. However, we observe that the two simulation schemes are not affected equally by this Hamiltonian density: while both schemes are have nearly identical complexity on the ensemble with $F = 0.008$, the estimates for the scheme of Sec. 4 are 100 times better on the ensemble with $F = 1$.

5.2 Real molecules

Despite the rather dispersed data, there appears to be a clear discrepancy with the results obtained from real and artificial molecules. This strongly suggests that the scaling with real molecules is much more favorable than the one anticipated from simulations of artificial molecules [1], with a scaling in the range $N^{1.5}$ - $N^{2.5}$ instead of N^4 - N^5 . The standard simulation scheme appears to offer a better scaling than the scheme of Sec. 4, but the data is too scattered to draw any firm conclusion. A case-by-case approach seems the most appropriate at this stage.

5.3 Analysis

The artificial molecule ensemble with $F = 1$ illustrate that, by choosing to decompose the Hamiltonian Eq. (1) into a sum of fewer but more complex terms Eq. (19), we can obtained

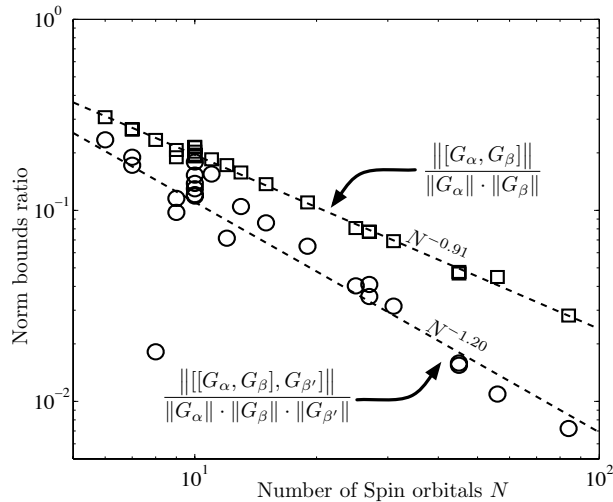


Fig. 2. Loss suffered from using bounds in Eq. (17) instead of Eq. (20) as a function of the number of spin orbitals.

significant improvements of the TS error in our quantum simulation algorithm. It is unclear at this stage how much of this gain is real, and how much is coming from a tighter upper bound on the error. On the one hand, all these bounds make use of the triangle inequality to bound the sum of M terms as follows

$$\left\| \sum_{i=1}^M X_i \right\| \leq \sum_{i=1}^M \|X_i\| \leq M \max_i \|X_i\|. \quad (22)$$

This bound may not be tight, since we might expect the true norm to scale like \sqrt{M} instead of M due to some averaging effect. However, surprisingly the agreement between the scaling of the bound agrees with the one found in [1] using exact simulations for the ensemble of artificial molecules. This initially suggests that not much is lost in triangle inequalities. We now analyze this in more detail using analytic estimates and additional simulations using LIQU*i*), a quantum simulator developed at Microsoft Research[20], and argue that at the sizes of molecules amenable to simulation (including those in [1]) the terms arising from the commutator of distinct H_{pqrs} terms are not yet important, but will become important at larger sizes. However, we further argue that the correct error analysis will scale only as \sqrt{M} due to average so that the true error estimate is significantly better than predicted by the triangle inequalities.

For this subsection, we primarily focus on the error estimates for the standard decomposition. For the alternative decomposition of section 4, another source of improvements comes from our ability to efficiently evaluate the operator norm of free fermion operators. For general operators G_α, G_β , the norm of their commutator is bounded by Eq. (17). However, when G_α, G_β are free fermion operators, we can directly evaluate their commutator and its norm. This idea is used to derive Eq. (21), which should be much tighter than the corresponding naive upper bound, and could also partly explain the observed gain. Fig. 2 illustrates the average advantage of evaluating the norm of the commutator instead of using a naive bound

Eq. (17) for the real molecules studied in Fig. 1.

Consider the commutator $[[H_{>\alpha}, H_\alpha], H_{>\alpha}]$ in Eq. (6). Above, we used a triangle inequality to upper bound this commutator by summing norms of double commutators $\|[[H_\alpha, H_\beta], H_\beta]\|$. There are at most $\mathcal{O}(N^{10})$ non-vanishing commutators, so that this estimate is at most $\mathcal{O}(N^{10})$. However, we also have available the bound that $\|[[H_{>\alpha}, H_\alpha], H_{>\alpha}]\| \leq 4\|H_\alpha\| \cdot \|H_{>\alpha}\|^2$. Let us now consider this bound in the specific context of the artificial molecule ensembles with h_{pqrs} assigned random signs, and show that in this case the triangle inequality bound will not give a tight bound on the scaling. In this ensemble, note that $H_{>\alpha}$ is a sum of $\mathcal{O}(N^4)$ terms with uncorrelated random signs and magnitude of order unity. In an appendix, we show that we can then bound the norm of $H_{>\alpha}$ by $\mathcal{O}(N^{5/2})$, which gives us the bound $\sum_\alpha \|[[H_{>\alpha}, H_\alpha], H_{>\alpha}]\| \leq \mathcal{O}(N^9)$, showing indeed that the triangle inequality estimate is not tight. However, likely even the bound of the appendix is not tight. Using a naive estimate that the norm will scale proportionally to the square-root of the number of terms so that $\|H_{>\alpha}\| \leq \mathcal{O}(N^2)$, we would have $\sum_\alpha \|[[H_{>\alpha}, H_\alpha], H_{>\alpha}]\| \leq \mathcal{O}(N^8)$ which further improves on the $\mathcal{O}(N^{10})$ estimate (we emphasize that we have not proven a $\mathcal{O}(N^2)$ bound on $\|H_{>\alpha}\|$ but only the weaker bound in the appendix; however, we conjecture that the $\mathcal{O}(N^2)$ bound is true in this ensemble). This estimate of $\mathcal{O}(N^8)$ completely ignores any considerations of which terms in $H_{>\alpha}$ commute with H_α and hence is still likely to be an overestimate; perhaps improved estimates could be obtained based on applying the trace method directly to the double commutator $[[H_{>\alpha}, H_\alpha], H_{>\alpha}]$ as this double commutator is a sum of $\mathcal{O}(N^6)$ terms with random signs but with some correlation between the signs. However, the estimate $\mathcal{O}(N^9)$ already shows that the upper bound from the triangle inequality is not tight. In this subsection, we further explore the possibility that averaging improves these estimates (the estimate for $\|H_\alpha\|$ is an example of a kind of averaging as the norm of a sum of terms may be much less than the sum of the norms).

In computing the error in numerical simulation, in all cases we chose the time step Δ_t sufficiently small to enter the regime that error scaled proportional to Δ_t^2 . The first piece of numerical evidence that the bounds are not yet relevant at the available N is that an attempt to correlate the errors resulting from the triangle inequality above with actual errors observed in simulation showed no correlation at all for a wide range of available molecules. Further, replacing the bound from the triangle inequality with an alternate estimate using the square-root of the sum of terms continued to show no correlation. Indeed, we found in the numerical simulations that the error in fact tended to *decrease* with larger N for a range of molecules studied.

More precise numerical evidence was obtained from a numerical experiment in which the term order was randomized for the molecule H_2O using a basis with 14 spin orbitals. We used the interleaved term order of Ref. HWBT14a, keeping the ordering of $H_{pp}, H_{pqqp}, H_{pq}, H_{prrq}$ terms fixed, while randomizing the order of the H_{pqrs} terms (randomizing the order of *all* terms, not just H_{pqrs} led to a significant increase in numerical error). We used a second-order TS formula to compare to Eq. (11). From this equation, we can see the effect at order Δ_t^2 on the ground state energy of a term $[H_\alpha, [H_\beta, H_{\alpha'}]]$ has a random sign depending upon term order. Thus, this term ordering randomizes the sign of any term involving three distinct H_{pqrs} terms (or involving the commutator of a non- H_{pqrs} term with the commutator of two distinct H_{pqrs} terms). Further, the signs of distinct terms in Eq. (11) are decorrelated for each other

for most choices: the average over term orderings of

$$= \langle \Psi_0 | [H_\alpha, [H_\beta, H_{\alpha'}]] | \Psi_0 \rangle \times \langle \Psi_0 | [H_\mu, [H_\nu, H_{\mu'}]] | \Psi_0 \rangle \quad (23)$$

vanishes if $\alpha, \beta, \alpha', \mu, \nu, \nu'$ are all distinct from each other and are all H_{pqrs} terms.

Thus, for a typical term order, we expect the errors to add proportional to \sqrt{M} . Fig. 3 shows a histogram of the actual TS error for 1000 instances. The Trotter number was set equal to 8, meaning a time step $\Delta_t = 1/8$. The curve is reasonably close to a Gaussian distribution with a non-zero mean. To quantify the Gaussianity of the curve, the ratio of the fourth moment to the square of the second moment is equal to $3.15\dots$, rather than the expected 3 and the ratio of the third moment to the three-halves power of the second moment is equal $0.32\dots$ rather than 0. The ratio of the root-mean-square width of the curve to the mean is $0.033\dots$, indicating that the terms with non-zero average still give the dominant contribution to the error at this size. However, for sufficiently larger sizes, the dominant contribution to the error should indeed arise from double commutators involving three distinct H_{pqrs} terms (as the number of these terms increases rapidly with N) and these terms will add with random signs.

It would be interesting to extend this analysis to higher order. We expect that there will still continue to be an averaging effect. The order Δ_t^4 correction to the ground state energy is the sum of two terms. First, there is the ground state expectation value of the order Δ_t^4 correction to Eq. (9); this term will still vanish on average over term order. Second, at order Δ_t^4 there is a correction to the ground state energy which is second order in the order Δ_t^2 Hamiltonian given in Eq. (9). This correction requires summing over intermediate states. Note, however, that if Ψ_i is an excited state, then

$$= \langle \Psi_0 | [H_\alpha, [H_\beta, H_{\alpha'}]] | \Psi_i \rangle \times \langle \Psi_i | [H_\mu, [H_\nu, H_{\mu'}]] | \Psi_0 \rangle \quad (24)$$

vanishes on averaging over term orders if $\alpha, \beta, \alpha', \mu, \nu, \nu'$ are all distinct from each other and are all H_{pqrs} terms. This holds because, by the Jacobi identity, the term $[H_\alpha, [H_\beta, H_{\alpha'}]]$ vanishes identically on averaging over term orders. However, beyond this treatment of each term order-by-order, it would be very interesting if an averaging estimate could be given to all orders, similar to the way that Ref. WBCH13a gave an upper bound in terms of double commutators that was valid to all orders.

As a further test, to see if the bounds were in any way sensitive to molecule geometry or closeness to Hartree-Fock, we studied the molecule ozone, O_3 . The kinetics of the recombination of O and O_2 to form the O_3 ozone molecule depend sensitively on the potential energy surface. In particular the height of a barrier, separating a shallow van der Waals minimum from the ground state, has a big influence on the reaction rate. Accurately calculating the barrier height of this transition state is a challenge for classical calculations, since the full basis is too large to be treated in a full-configuration interaction calculations and truncated basis sets introduce large approximation errors [21]. Calculating the energy of various configuration of the ozone molecule is thus a useful early benchmark problem for a quantum computer.

For our estimates we considered three distinct configurations of ozone: a) the ground state, b) the transition state and c) a metastable state at a van der Waals minimum between an O_2 molecule and a free oxygen atom. The distances and angles between the atoms at these configurations was obtained from Ref.[21]. Direct evaluation of the double commutator bound

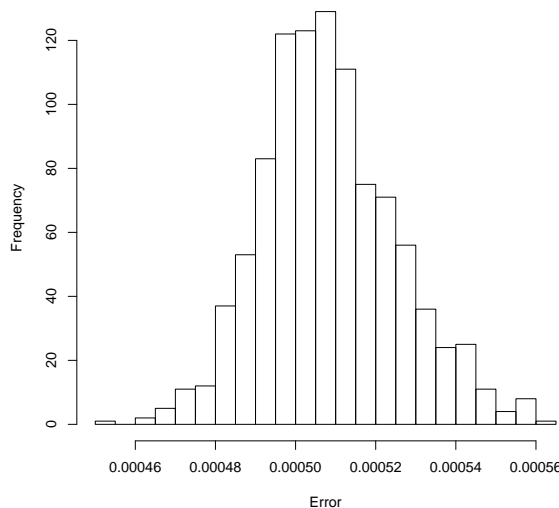


Fig. 3. Histogram of TS error for H_2O with H_{pqrs} term order randomized, 1000 samples.

showed that it was much larger in the ground state than anywhere else. In particular, the value at the transition state was 7.5 times smaller than in the ground state and at the van der Waals minimum even 9.6 times smaller. Interestingly, this means that the bound is not in any significant way worse for the transition point, which is hard to obtain classically. We used a basis with 60 spin orbitals for ozone; compared to H_2O in a large basis with 62 spin orbitals, the bound for the ground state configuration of ozone was roughly twice as large, and it was roughly 0.6 times as large as that for Fe_2S_2 in a basis with 112 spin orbitals.

5.4 *Cauchy–Schwarz bounds and decay of $|h_{pqrs}|^2$*

The Cauchy–Schwarz inequality gives an alternative method for bounding the error in quantum simulations that can be easily computed for large values of n . Let $W(\vec{x})$ be an indicator function that is 1 if and only if the vector $\vec{x} = (\alpha, \beta, \beta')$ corresponds to a triple of Hamiltonians $H_\alpha, H_\beta, H_{\beta'}$ that contributes to the simulation error in Eq. (8). That is, $W(\vec{x})$ is zero if $[[H_\beta, H_\alpha], H_{\beta'}] = 0$; alternately, if we are content to evaluate the error to order Δ_t^2 , we can set $W(\vec{x})$ to zero if the ground state expectation value of the triple product is known to be zero from symmetry arguments since (11) shows that the ground state energy is unaffected by such terms to order Δ_t^3 . Given these assumptions, the use of the Cauchy–Schwarz inequality applied to triplets which give a non-vanishing contribution to Eq. (8) shows us that the error can be bounded by the root-mean-square values of $\|H_\alpha\|$

$$\left\| \sum_{\vec{x}} [[H_\beta, H_\alpha], H_{\beta'}] W(\vec{x}) \right\| \leq 4 \left(\sum_{\alpha} \|H_\alpha\|^2 \right)^{3/2} \sqrt{N_W}, \quad (25)$$

where $N_W = \sum_{\vec{x}} W(\vec{x})$. Similar Cauchy–Schwarz bounds can also be found for Eq. (15). This bound has the advantages that it depends on the RMS value of $\|H_\alpha\|$ which is easy

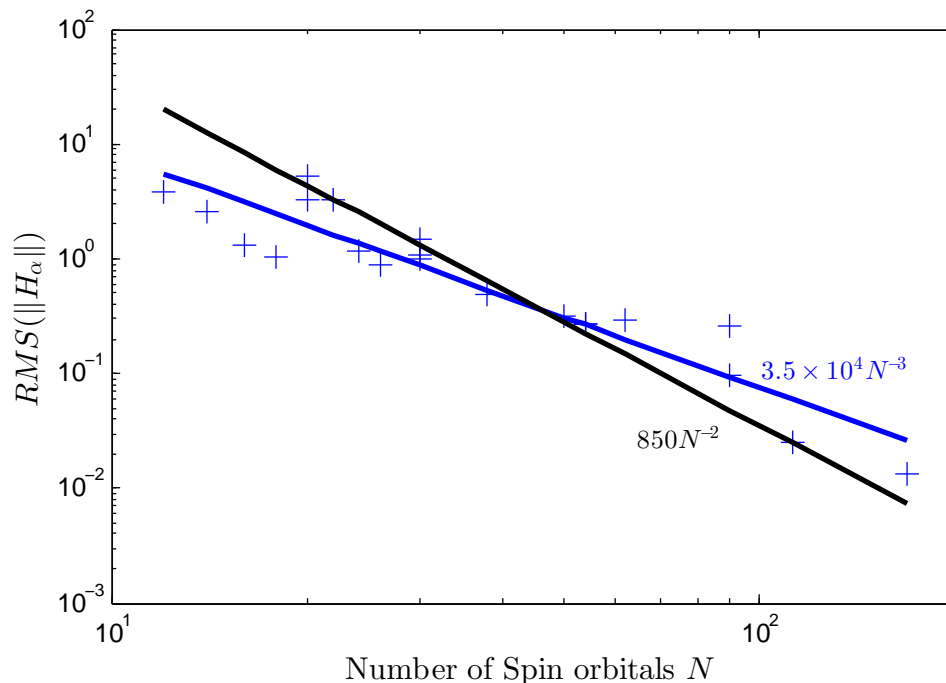


Fig. 4. RMS value of $\|H_\alpha\|$ for a set of small molecules as a function of the number of spin orbitals N .

to compute directly or from Monte–Carlo sampling and that it also handles constraints in a natural way.

The scaling of the RMS values of $\|H_\alpha\|$ is given in Fig. 4. Since the RMS value decreases, the scaling predicted is better than the $\mathcal{O}(N^{10})$ bound trivially expected. If we only exclude commuting terms from the sum then $N_W = \mathcal{O}(N^{10})$. Using the scalings observed in Fig. 4 and the fact that H consists of $\mathcal{O}(N^4)$ terms, (25) shows that $\delta^{\text{TS}} = \mathcal{O}(N^\gamma)$, where we find empirically that $2 \leq \gamma \leq 5$, which means the number of Trotter steps needed to achieve a fixed error tolerance is $\mathcal{O}(N)$ - $\mathcal{O}(N^{2.5})$. Since $\mathcal{O}(N^4)$ gates are needed per TS step, the simulations require a number of gates that is $\mathcal{O}(N^5)$ - $\mathcal{O}(N^{6.5})$.

These scalings also depend strongly on the form of the ground state. If, for example, we were to assume that only $\mathcal{O}(N^6)$ terms lead to errors in the ground state energy (which holds when the error in the Hartree–Fock approximation is small) then scaling of the number of gates needed would further drop to $\mathcal{O}(N^{5.5})$ - $\mathcal{O}(N^4)$. Hence properties of the ground state can and should be used to reduce these bounds when possible.

6 Coalescing

In Appendix C of Ref. WBCH13a, the idea of “coalescing” was introduced. This idea can be regarded as a “multi-resolution Trotterization”: Rather than trying to determine the minimum TS time step which will work for all terms, we allow different terms to have a different TS step.

One simple realization of the approach within a first-order TS scheme is to pick a fixed

time step δ_t which represents the shortest time that we resolve. Consider a Hamiltonian $H = \sum_{\alpha} H_{\alpha}$. For each term, H_{α} , we choose some number n_{α} which reflects how infrequently the term is applied: it will be applied every n_{α} -th step with a strength proportional to n_{α} . Let K be the least common multiple of the n_{α} and let $\Delta_t = K\delta_t$. Then, this scheme gives an approximation to evolution over time Δ_t . For example, if $H = H_1 + H_2 + H_3$, with $n_1 = 1, n_2 = 2, n_3 = 4$ so that $\Delta_t = 4\delta_t$ then we approximate

$$\begin{aligned} \exp(i4\delta_t) &\approx \left(\exp(i\delta_t H_1) \exp(2i\delta_t H_2) \exp(4i\delta_t H_3) \right) \\ &\times \left(\exp(i\delta_t H_1) \right) \left(\exp(i\delta_t H_1) \exp(2i\delta_t H_2) \right) \\ &\times \left(\exp(i\delta_t H_1) \right), \end{aligned} \quad (26)$$

where the parenthesis (\dots) are used to separate the four different steps.

Clearly, such a coalescing scheme allows enormous flexibility. Even within the simple example above, there is room to choose the ordering of terms within each of the four steps (one might choose different orderings in each step). Further, with a term such as H_2 , we can choose to execute it on the first and third step as above or on the second and fourth step, and similarly we can choose to execute H_3 on any of the four steps.

An alternate more complicated coalescing scheme was also presented in Ref. WBCH13a. One (theoretical) advantage of this more complicated scheme is that it was defined in a way that allowed an inductive proof of tighter upper bounds on the TS error. For simplicity in this paper, we stick to the simpler first-order approach outlined above. Also, for simplicity, in all cases we choose the n_{α} to be powers of two, and we execute a term with given n_{α} on steps $1, n_{\alpha} + 1, 2n_{\alpha} + 1, \dots$. In practice we found no notable advantage to considering other options. Thus, the important question is how to choose the n_{α} for a given term.

Using this coalescing scheme, we otherwise continue to followed the interleaved term order of Ref. HWBT14a, so that in a given TS step we first execute the H_{pp} and H_{pqqp} terms, followed by the H_{pq} terms interleaved with the H_{prrq} terms. Then, we execute the H_{pqrs} terms. Coalescing is only applied to the H_{pqrs} terms; that is, all other terms will have $n_{\alpha} = 1$, while for the H_{pqrs} terms we have $n_{\alpha} = 1$ for some terms and $n_{\alpha} > 1$ for others. We used a first order TS scheme (as noted in Ref. HWBT14a the first order TS offers performance with the same error scaling as second order in this case).

Before going into details, it is worth noting two properties of this scheme. First, the scheme is exact if all terms H_{α} commute. Second, the scheme gives the correct response in the ground state energy to first order for any term H_{α} , regardless of the value of n_{α} . To make this more precise, suppose that $H_{\alpha} = \epsilon T$ for some α , for operator T and for some $\epsilon \ll 1$. Let ψ_0 be the approximation to the ground state resulting from TS evolution at $\epsilon = 0$. Then, regardless of n_{α} , the scheme gives a shift in ground state energy equal to $\epsilon \langle \psi_0 | T | \psi_0 \rangle + \mathcal{O}(\epsilon^2)$.

Finally, one may consider the question of circuits for coalescing. In Ref. HWBT14a, it was shown that a reduction in circuit depth could be obtained using modified circuits that enable the cancellation of much of the CNOT strings used to perform the Jordan-Wigner transformation and that also enable improved parallelism. Fortunately, these techniques are compatible with coalescing. We will find that most of the terms can be aggressively coalesced (choosing a large n_{α}); for these terms, since many terms will all be executed with the same

large n_α , most of the CNOT cancellation and parallelization benefits can still be obtained for those terms. Thus, in what follows, as a proxy for the total depth of the circuit required, we simply use the number of terms in the TS formula, while a more accurate estimate along the lines of previous work such as Ref. HWBT14a would require a detailed analysis of gate depth.

6.1 Prioritizing Terms

We now discuss how to choose the n_α . The main improvement here on Ref. WBCH13a is a different way to perform this choice which leads to significant numerical improvements. Unfortunately, we do not have mathematically rigorous upper bounds to justify our choice; our justification is instead based on extensive numerical simulation for small molecules within reach of a classical simulation. As explained below, we found a general, fairly simple rule which worked well for all such small molecules, so that we were able to decrease the simulation effort while reducing (or at worst, not increasing) the numerical error in the estimate of the ground state energy.

Previously, in Ref. WBCH13a, it was suggested that the choice of which terms to coalesce should be based solely on the *magnitude* of the term. Terms with a larger coefficient would be executed more frequently than those with smaller coefficient. We chose several different cutoffs $E_1, E_2, E_4, \dots, E_K$, and assigned all terms H_α with coefficient greater than or equal to E_1 to have $n_\alpha = 1$, while all terms H_α with coefficient smaller than E_1 but greater than or equal to E_2 had $n_2 = 2$, and so on. Unfortunately, despite extensive numerical exploration, we were unable to get this scheme to yield any significant improvement.

In this paper we propose an alternate scheme to choose the n_α . Heuristically, since the scheme gives the correct energy shift to first order, it is important to get the *second-order* response to a term correct. We estimate this effect as follows. Consider a term $H_\alpha = h_{pqrs}c_p^\dagger c_q^\dagger c_r c_s$. We define the importance of the term by a quantity I_α defined as

$$= I_\alpha = \frac{|h_{pqrs}|^2}{\Delta_{p,q,r,s}} \quad (27)$$

where Δ is an estimate of the energy denominator. We define Δ as follows. This is similar to ideas in Section 5 of Ref. HWBT14a. Assume we are working in a Hartree-Fock basis with diagonal terms

$$= \sum_p t_{pp} c_p^\dagger c_p + \frac{1}{2} \sum_{p,q} V_{pqqp} c_p^\dagger c_p c_q^\dagger c_q. \quad (28)$$

We let $\omega_p = t_{pp} + \sum_{q \in \text{occ.}} V_{pqqp}$, where the sum is over occupied orbitals q . We then let

$$= \Delta_{p,q,r,s} = \left| \omega_p + \omega_q - \omega_r - \omega_s \right|. \quad (29)$$

This expression describes the second-order response in ground state energy with respect to this perturbation about a Hartree-Fock state.

Our general strategy is to use I_α instead of h_{pqrs} . Thus, even if h_{pqrs} is small, we still regard terms as important if they have a small $\Delta_{p,q,r,s}$, so that terms with a larger I_α are assigned a smaller n_α .

One might worry that this approach will not work well if the molecule is far from a Hartree-Fock solution. However, in strongly interacting Fermi systems, the most important deviations

from free fermion behavior arise for states near the Fermi energy. States sufficiently far above the Fermi energy have occupancy close to zero and those far below the Fermi energy have occupancy close to one, while those near the Fermi energy may have an occupancy far from zero or one in an interacting system. However, since this scheme ascribes a large importance to terms involving transitions with all orbitals close to the Fermi energy, we hope that it will continue to work well, although no hard evidence is presented here.

One important feature of this scheme, as explained below in the section on numerical results, is that we obtain a splitting of the histograms of I_α values. There are few terms with large I_α (which get $n_\alpha = 1$) and many terms with small I_α (which get larger values of n_α), with only few terms of intermediate importance. For these terms of intermediate importance, we have found it useful to define an additional heuristic rule. While this rule helps, it is not necessary as we have found that fewer terms fall into the intermediate region as molecules grow larger. This heuristic rule is explained in the next subsection.

6.2 Numerical Results

We now consider several small molecules, and explain specific choices of the cutoffs that lead to improvements using this scheme.

For reasons of the quantum circuits chosen, we make one slight modification to the scheme above. For us a term in the Hamiltonian is not simply the term $c_p^\dagger c_q^\dagger c_r c_s + \text{h.c.}$ but also includes all other terms involving four fermion operators on spin orbitals p, q, r, s such as $c_q^\dagger c_r^\dagger c_p c_s$. The reason for this is that some of the same circuits are used to execute both terms and so there is little to be gained by coalescing one such term but not the other. Hence, we instead choose $\Delta_{p,q,r,s}$ to be the minimum value of Δ over all such possible assignments of two creation and two annihilation operators to p, q, r, s .

Ref. whitfield2011 shows that all of the terms for a specific set of p, q, r, s may be gathered together to form a single circuit (greatly reducing the overall simulation depth). This leads to a re-write of the original h_{pqrs} values as a set of four strengths covering the eight basis directions ($xxxx + yyyy, xxyy + yyxx, yxyx + xyxy, yxxy + xyyx$). We take the maximum magnitude of these four as representative of the effect of the combined terms.

For those terms of intermediate importance, we use the following heuristic rule: a term is chosen to be more important and have $n_\alpha = 1$ if it does not annihilate the Hartree-Fock ground state, while it gives a larger value of n_α if it does annihilate the Hartree-Fock ground state. Whether or not a term annihilates the Hartree-Fock ground state can be determined fairly simply. For example, if a term $c_p^\dagger c_q^\dagger c_r c_s$ has p, s corresponding to spin up and q, r corresponding to spin down, then it annihilates the Hartree-Fock ground state unless r, s are occupied and p, q are unoccupied. In fact, since we always ensure that terms are Hermitian and since several different four fermion operators involving spin orbitals p, q, r, s are combined into a single term in the Hamiltonian, such a term will also be retained if one out of p, q is occupied in the Hartree-Fock state and the other is unoccupied and also one out of r, s is occupied and the other is unoccupied. Similarly, all four spins are up or all four are down, the term is retained if exactly two of the p, q, r, s are occupied in the Hartree-Fock state and the other two are unoccupied.

The final set of rules chosen after some numerical experimentation were: we imposed an upper cutoff C_U and a lower cutoff C_L . For any term with importance I_α larger than C_U , we

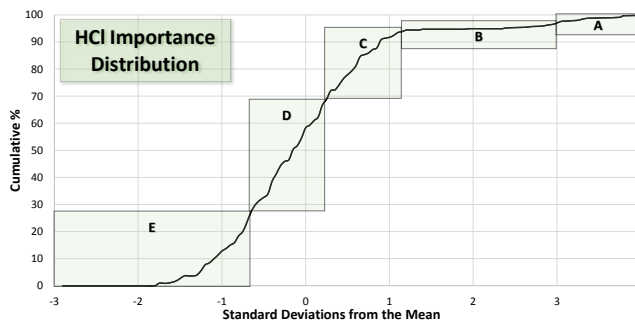


Fig. 5. Importance Distribution (h_{pqrs}^2/ω) for Hydrogen Chloride. Regions are: (A) terms that must be done every step (δ_t), (B) the “front porch” where terms are executed on every step or every 16 steps (based on occupancy), (C) less significant terms (25% of the remaining) that can be done every 16 steps, (D) 1/2 of the remaining terms that can be done every 32 steps and (E) all the remaining terms that can be done once every 64 steps.

set $n_\alpha = 1$. For terms with importance in the interval $[C_L, C_U]$, we set $n_\alpha = 1$ or $n_\alpha = 16$ depending on the heuristic in the above paragraph. Terms with importance smaller than C_L were always chosen to have $n_\alpha \geq 1$. Of these terms, we took the 25% with the largest importance and set those to $n_\alpha = 16$; of the remaining 75% of the terms, half (or 37.5%) were given $n_\alpha = 32$ and the remainder were given $n_\alpha = 64$.

This gives a choice of only 4 possible values of n_α : 1, 16, 32, 64. A more sophisticated rule with more possible choice might lead to even more speedup. Our initial studies considered first only two possible values, 1, 16, then later studies considered 3 values, 1, 16, 32; in both those cases, the speedup was not as large but still some speedup could be obtained. The effect of these rules is shown in Fig. 5 for the molecule *HCl*.

Our goal is to find a choice of C_U and C_L that will reduce the time effort without costing additional accuracy. Inevitably, the choice of $n_\alpha > 1$ for some will reduce the accuracy compared to a choice for $n_\alpha = 1$ for all terms, assuming both simulations are run with the same δ_t . Thus, what we did was to first run a simulation without any coalescing at a fixed value of Δ_t . Then, we ran a simulation using coalescing, with the above choice of n_α , with $\delta_t = \Delta_t/2$. This permits a more accurate treatment of the terms with highest importance since they use a shorter timestep. We ran these simulations with Δ_t sufficiently small that the simulations were in the regime that TS error was proportional to Δ_t^2 ; once we are in this regime, the relative accuracy of the two simulations (the one with and the one without coalescing) remains unchanged as Δ_t decreases to this order in Δ_t .

We found a set of rules for C_U and C_L that meant that in every molecule we tried, the error did not increase using coalescing, and in many cases it decreased. We chose $\log(C_U)$ equal to the average value of the log of I_α plus three times the standard deviation of the log of I_α , and we chose $\log(C_L)$ equal to the average value of the log of I_α plus 1.2 times the standard deviation. Specific molecules simulated using LIQU*i* were *HF*, *H₂O*, *NH₃*, *NCl*, *F₂*, and *H₂S*.

The result for the work required is shown in Fig. 6. By hand-tuning the choice of C_U , C_L for specific molecules it is possible to further reduce the work without increasing the error. However, this is clearly an unrealistic test: since our goal is to develop rules that will be useful

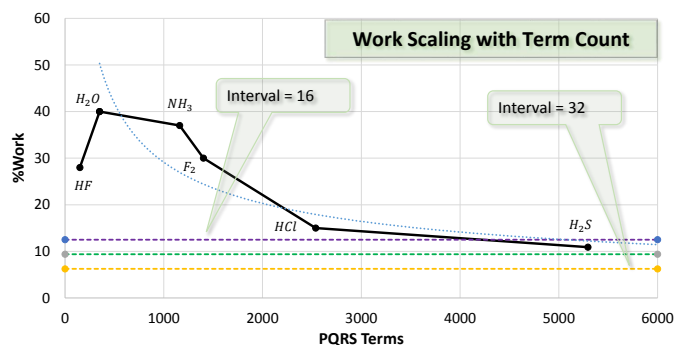


Fig. 6. Amount of work for various molecules using the current scheme at a Trotter number of 64. The three horizontal dotted lines are asymptotes for doing all terms at an interval of 16 steps (top) and 32 steps (bottom). The current approach (shown in the previous figure) is close to the middle asymptote for a 50/50 mixture of the two limits.

to reduce work on a real quantum computer, there is no way to know in advance what the most optimal values are. However, this general rule works well for all these molecules and is close to optimal.

We cannot simulate larger molecules, but assuming that the scheme does continue to work, we can investigate what speedups would be achieved. The first important point is that while the mean of $\log(I_\alpha)$ is observed to decrease with increasing N , no clear trend was observed for the standard deviation. Instead, the standard deviation was observed to vary between roughly 2 – 3 on a log-base 10 scale with no clear trend, considering larger molecules up to Fe_2S_2 in a basis with 168 spin-orbitals. After normalizing $\log(I_\alpha)$ by subtracting the mean value for the given molecule and dividing by the standard deviation (so that we use $(\log(I_\alpha) - \log(\bar{I}_\alpha))/sd(\log(I_\alpha))$ as the horizontal axis), the results are as shown in Fig. 7. One observes a range of importance (appearing as a flat spot in the curve, which we term the “front porch” in the figures) into which few terms fall. This range is observed to move to larger importance relative to the mean as the molecule size increases. The rules for C_U, C_L above were chosen such that for the smaller molecules, the interval $[C_L, C_U]$ is roughly the region of this front porch. Since the width of the front porch decreases, relatively fewer terms fall into this region. Further, as molecule size increases, one observes that the importance (relative to the mean) of the most important terms grows larger (see Fe_2S_2 for example, which extends furthest to the right on the curve). Since then these large molecules have a few terms with very high importance, this suggests that for these molecules it will be possible to coalesce almost all terms except these few high importance terms leading to potentially even larger gains in runtime for larger molecules.

7 Discussion and conclusion

The Hamiltonian of a small molecule contains a number of terms scaling as N^4 with the number of orbitals N in the simulation, which represents a major bottleneck to quantum simulations. Previous analysis [1, 2] predict a general complexity in gate count (which can be parallelized saving a factor of N) scaling as N^8 - N^9 . By numerically evaluating an upper bound on the simulation error entailed by the TS approximation, we have demonstrated that

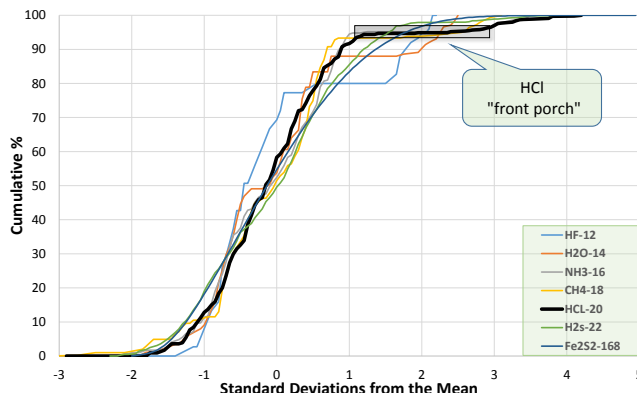


Fig. 7. Results for several molecules of different sizes. Note that the “front porch” disappears as N increases and may thus be ignored as we scale (terms are either coalesced or not, there is no need for occupancy calculations). Further, the most important terms become more important relative to the mean (and hence the typical terms are less important relative to the most important ones) as N increases.

the scaling is in fact much more favorable for real-world molecules, in the $N^{5.5}$ - $N^{6.5}$ range in worst case. Evaluation of this bound on a greater number of molecules would be necessary to reach firm conclusions; further, a better understanding would be needed of whether the error terms should be added in absolute value or whether they add with random signs.

Because the scaling analysis of [1] uses molecules drawn from a random ensemble, our observations strongly indicate that this ensemble fails to accurately reproduce the statistical properties of real molecules. As a simple indication of this discrepancy, we find that the sum of the magnitude of all the Hamiltonian coefficients $\sum_{pqrs} |h_{pqrs}|$ scales like N^2 for real molecules, while the random ensemble yields N^4 by design.

We have explored the consequences of breaking up the Hamiltonian into a different sum of terms to implement the TS decomposition. With artificial molecules and a few real molecules, we have seen that this alternative decomposition can offer significant savings. Thus, we believe that exploring the different ways in which this decomposition can be realized is a good approach to obtain further improvements. This decomposition is usually guided by our ability to simulate sparse Hamiltonians [7]. It is noteworthy that in the decomposition we use, each term G_α^2 is no sparser than the full Hamiltonian H . This illustrates that other criteria should be envisioned to guide this decomposition.

We have also developed the coalescing technique, showing large gains for small molecules. It seems likely based on our numerical data that even more aggressive coalescing will be possible on larger molecules, leading to further gains. We have separately explored the possibility of combining coalescing with the sum of squares decomposition of the Hamiltonian, but have not found any improvement this way. Combining the coalescing technique with the improved scaling here suggests that simulation of large molecules with the order of a hundreds of spin orbitals will be much more practical than indicated in [1].

Lastly, our study, as well as previous ones [11, 12, 13, 1, 2], have focused on low-order TS decompositions. While schemes based on random walks [9] or techniques from simulating continuous query algorithms [10] promise lower query complexity than TS based simulations

for general purpose quantum simulations, their reliance on a quantum oracle makes a comparison of their time complexity to that of TS algorithms for quantum chemistry challenging. Their concrete realization would therefore require a circuit which, given inputs (i, j) , returns the matrix element $\langle i|H|j\rangle$; see also Ref. [22]. Since an arbitrary Hamiltonian may contain $\sim N^4$ non-zero terms, such a circuit would at best require $\sim N^4$ gates, yielding an overall complexity greater than what we have observed here. This can be parallelized using, for example, a QRAM[23], but at the cost of an enormous space overhead. Despite these difficulties in implementing the oracles, it remains possible that these simulation techniques could be competitive because they may perform better than what their upper bounds suggest.

Acknowledgements

DP acknowledges the hospitality of the University of Sydney where this research project was realized. DP and ACD were supported by the ARC via the Centre of Excellence in Engineered Quantum Systems (EQuS), project number CE110001013.

References

- [1] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Can quantum chemistry be performed on a small quantum computer?, December 2013.
- [2] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer. Improving quantum algorithms for quantum chemistry, 2014.
- [3] R. P. Feynman. Simulating physics with computers. *Int. J. of Theor. Phys.*, 21:467, 1982.
- [4] S. Lloyd. Universal quantum simulators. *Science*, 273:1073–1078, 1996.
- [5] C. Zalka. Efficient simulation of quantum systems by quantum computers. *Fortsch. Phys.*, 46:877–879, 1998.
- [6] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme. Quantum algorithms for fermionic simulations. *Phys. Rev. A*, 64:22319, 2001.
- [7] D. Aharonov and A. Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. *Proc. 35th Annual ACM Symp. on Theo. Comp.*, page 20, 2003.
- [8] Nathan Wiebe, Dominic W. Berry, Peter Høyer, and Barry C. Sanders. Simulating quantum dynamics on a quantum computer. *J. Phys. A*, 44:445308, 2011.
- [9] D.W. Berry and A.M. Childs. Black-box hamiltonian simulation and unitary implementation. *Quant. Info. and Comp.*, 12:29, 2012.
- [10] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Exponential improvement in precision for simulating sparse hamiltonians, 2013.
- [11] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.

- [12] James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics*, 109(5):735–750, 2011.
- [13] Ivan Kassal, James D. Whitfield, Alejandro Perdomo-Ortiz, Man-Hong Yung, and Alán Aspuru-Guzik. Simulating chemistry using quantum computers. *Annual Review of Physical Chemistry*, 62(1):185–207, 2011.
- [14] Z. Gan and R. J. Harrison. Supercomputing. In *Proc. ACM/IEEE SC 2005 Conference*, page 22, 2005.
- [15] Hiroki Nakano and Toru Sakai. Numerical-diagonalization study of spin gap issue of the kagome lattice heisenberg antiferromagnet. *J. Phys. Soc. Japan*, 80:053704, 2011.
- [16] Andreas M. Läuchli, Julien Sudan, and Erik S. Sørensen. Ground-state energy and spin gap of spin-12 kagomé-heisenberg antiferromagnetic clusters: Large-scale exact diagonalization results. *Phys. Rev. B*, 83:212401, 2011.
- [17] Sylvain Capponi, Oleg Derzhko, Andreas Honecker, Andreas M. Läuchli, and Johannes Richter. Numerical study of magnetization plateaus in the spin-12 kagome heisenberg antiferromagnet. *Phys. Rev. B*, 88:144416, 2013.
- [18] Yuki Kurashige, Garnet Kin-Lic Chan, and Takeshi Yanai. Entangled quantum electronic wavefunctions of the Mn4CaO5 cluster in photosystem II. *Nature Chemistry*, 5:660, 2013.
- [19] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.
- [20] D. Wecker and K. M. Svore. Liqui—j: A software design architecture and domain-specific language for quantum computing.
- [21] Reinhard Schinke and Paul Fleurat-Lessard. The transition-state region of the $\text{o}(^3p)+\text{o}_2(^3\sigma_g^-)$ potential energy surface. *The Journal of Chemical Physics*, 121(12):5789–5793, 2004.
- [22] B. Toloui and P. J. Love. Quantum algorithms for quantum chemistry based on the sparsity of the ci-matrix.
- [23] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical review letters*, 100(16):160501, 2008.

Appendix A Estimate for Norm of Random Hamiltonian

We now briefly discuss the extent to which we can prove tighter bounds on $\|H_{>\alpha}\|$ in a random ensemble.

We conjecture that $\|H_\alpha\| \leq \mathcal{O}(N^2)$ with high probability. However, proving this bound $\mathcal{O}(N^2)$ might be difficult but it is possible using a version of the trace method to prove the weaker bound $\|H_{>\alpha}\| \leq \mathcal{O}(N^{5/2})$ using a version of the trace method. Consider the average (over random choices of the Hamiltonian) of $\text{tr}(H_{>\alpha}^n)$ for a constant n chosen later.

Expanding the trace as a sum $\sum_{\alpha_1 > \alpha} \dots \sum_{\alpha_n > \alpha} \text{tr}(H_{\alpha_1} \dots H_{\alpha_n})$, the only terms that do not vanish on average are where the sequence $\alpha_1, \dots, \alpha_n$ repeats each α an *even* number of times. Hence, of the at most m^n terms in the sum, only at most $(mn)^{n/2}$ are non-vanishing. Let us use an overline to denote the average over Hamiltonians. Thus $\overline{\text{tr}(H_{>\alpha}^n)}$ is bounded by $2^N \cdot (\text{const.} \times mn)^{n/2}$ and $\left(\overline{\text{tr}(H_{>\alpha}^n)}\right)^{1/n} \leq \text{const.} \times 2^{N/n} (mn)^{1/2}$. We will choose $n = 2N \ln(2)$ so that $\left(\overline{\text{tr}(H_{>\alpha}^n)}\right)^{1/n} \leq \mathcal{O}(N^{5/2})$ and so $\overline{\|H_{>\alpha}\|} \leq \mathcal{O}(N^{5/2})$. Further, using similar estimates, one can show that with high probability $\|H_{>\alpha}\| = \mathcal{O}(N^{5/2})$; to see this, by Markov's inequality, the probability that $\|H_{>\alpha}\|$ is, for example, twice as big as $\left(\overline{\text{tr}(H_{>\alpha}^n)}\right)^{1/n}$ is at most 2^{-n} .

Even using this weaker bound $\|H_{>\alpha}\| \leq \mathcal{O}(N^{5/2})$ we still find the bound $\sum_{\alpha} \|[[H_{>\alpha}, H_{\alpha}], H_{>\alpha}]\| \leq \mathcal{O}(N^9)$ which still improves on the triangle inequality estimate $\mathcal{O}(N^{10})$.