

# Initial State Preparation for Quantum Chemistry on Quantum Computers

Stepan Fomichev<sup>1,\*</sup>, Kasra Hejazi<sup>1,†</sup>, Modjtaba Shokrian Zini<sup>1</sup>, Matthew Kiser<sup>2,3</sup>,  
Joana Fraxanet<sup>1,4</sup>, Pablo Antonio Moreno Casares<sup>1</sup>, Alain Delgado<sup>1</sup>, Joonsuk Huh<sup>1,5,6,7</sup>,  
Arne-Christian Voigt<sup>2</sup>, Jonathan E. Mueller<sup>2</sup> and Juan Miguel Arrazola<sup>1</sup>

<sup>1</sup>*Xanadu, Toronto, Ontario M5G2C8, Canada*

<sup>2</sup>*Volkswagen AG, Berliner Ring 2, Wolfsburg 38440, Germany*


<sup>3</sup>*TUM School of Natural Sciences, Technical University of Munich, Garching, Germany*

<sup>4</sup>*ICFO - Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona) 08860, Spain*

<sup>5</sup>*Department of Chemistry, Sungkyunkwan University, Suwon 16419, Republic of Korea*

<sup>6</sup>*SKKU Advanced Institute of Nanotechnology (SAINT), Sungkyunkwan University, Suwon 16419, Korea*

<sup>7</sup>*Institute of Quantum Biophysics, Sungkyunkwan University, Suwon 16419, Korea*

 (Received 9 February 2024; revised 21 June 2024; accepted 29 October 2024; published 9 December 2024)

Quantum algorithms for ground-state energy estimation of chemical systems require a high-quality initial state. However, initial state preparation is commonly either neglected entirely, or assumed to be solved by a simple product state like Hartree-Fock. Even if a nontrivial state is prepared, strong correlations render ground-state overlap inadequate for quality assessment. In this work, we address the initial state preparation problem with an end-to-end algorithm that *prepares* and *quantifies* the quality of initial states, accomplishing the latter with a new metric—the energy distribution. To be able to prepare more complicated initial states, we introduce an implementation technique for states in the form of a sum of Slater determinants that exhibits significantly better scaling than all prior approaches. We also propose low-precision quantum phase estimation (QPE) for further state quality refinement. The complete algorithm is capable of generating high-quality states for energy estimation, and is shown in select cases to lower the overall estimation cost by several orders of magnitude when compared with the best single product state ansatz. More broadly, the energy distribution picture suggests that the goal of QPE should be reinterpreted as generating improvements compared to the energy of the initial state and other classical estimates: such an improvement can still be achieved even if QPE does not project directly onto the ground state. Finally, we show how the energy distribution can help in identifying potential quantum advantage.

DOI: [10.1103/PRXQuantum.5.040339](https://doi.org/10.1103/PRXQuantum.5.040339)

## I. INTRODUCTION

One of the main contenders for useful applications of quantum computers is the simulation of many-body physics, in particular, for quantum chemistry and materials science. Of special interest is the determination of ground-state energies, which have broad application [1–9]. Many different quantum methods for ground-state energy determination have been proposed, ranging from quantum phase estimation (QPE) and its variants [10–14], to more

recent developments [15–17]. We refer to these methods as *quantum energy estimation* algorithms. Each of these methods requires a high-quality initial state, where quality is traditionally understood in terms of the overlap with the ground state. The quality of the initial state directly impacts the performance and runtime of any energy estimation algorithm, making it crucial to develop advanced methods for initial state preparation.

While decisive for the success of quantum algorithms, initial state preparation is often treated as separate from the energy estimation algorithm, and has not received as much attention as other aspects of quantum algorithms in the literature. A common approach to preparing an initial state is to take an approximate wavefunction from a traditional quantum chemistry method and encode it on a quantum computer. The Hartree-Fock state is the simplest and computationally cheapest choice. Even though it has been found to have high overlap with the ground state in small molecules [2,4,18,19], it is seriously lacking

\*Contact author: [stepan.fomichev@xanadu.ai](mailto:stepan.fomichev@xanadu.ai)

†Contact author: [kasra.hejazi@xanadu.ai](mailto:kasra.hejazi@xanadu.ai)

‡These two authors contributed equally.

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

for strongly correlated systems [20], such as molecules with stretched bonds [18,19,21–24] and more complex molecules with transition-metal centers [25]. Beyond the Hartree-Fock state, a variety of methods have been proposed to encode sums of Slater determinants (SOS) [18, 26–29] or matrix product states (MPS) [30,31]. Ground-state energy estimation was explored using SOS states obtained from configuration interaction singles and doubles [28], active space methods [19,21,22], and selective configuration interaction methods [18]. However, the performance of these approaches has mostly not been evaluated or compared beyond small, uncorrelated molecules.

Other approaches that can be categorized as quantum heuristic methods have also been considered widely. While adiabatic state preparation is likely the most well known [1,32–35], other heuristics include variational methods [36–42] and quantum imaginary time evolution [43–46]. While promising, to date most of these methods have been demonstrated only for small molecules [1,23,37–40,42], and suffer from various shortcomings such as long runtimes, expensive classical optimization, or costly state tomography. More broadly, the absence of any guarantee of their success in state preparation is problematic.

The variety of state-preparation approaches raises the question of which method is best suited to which situations. Furthermore, it is not even clear how one should compare different possible state-preparation schemes for actual problems of interest. For example, in general, the overlap with the ground state cannot serve as a practical metric for comparison or quality assessment since we typically do not know what the ground state is. These issues also make it difficult to quantify the total runtime of quantum algorithms and to understand their actual potential to outperform classical methods. Overall, there is a need for a framework that encompasses the most powerful methods for initial state preparation, provides tools to evaluate their quality, and allows us to make informed statements about the prospects for quantum advantage.

In this work, we present a complete algorithm for preparing high-quality states for quantum energy estimation. Our state-preparation algorithm begins with using quantum chemistry methods to obtain classical descriptions of approximate ground-state wavefunctions, either in SOS or MPS form. We then introduce a novel quantum algorithm for preparing SOS states on a quantum computer, with a better cost compared to all previous methods. This is complemented with resource estimation formulas quantifying the number of qubits and gates needed for implementation, both for our new SOS algorithm and for previously developed techniques for implementing MPS states [30]. To assess and compare the quality of the several candidate states in hand, we develop a methodology that works based on their associated *energy distributions*. These are projections of the candidate wavefunction on the

eigenspectrum of the system Hamiltonian: while obtaining them exactly is more difficult than computing the overlap with the ground state, they can be approximated—a task for which we propose new classical and quantum methods. Once the state quality is assessed, and the state with the highest quality as per the energy distribution metric is implemented, it can also be further refined with the use of a quantum filtering algorithm. On the basis of our analyses, we find that coarse QPE with postselection—that is, QPE performed with low precision—generally outperforms other filtering methods.

The concept of the energy distribution has utility beyond state quality assessment. First, it suggests an alternative interpretation of quantum energy estimation algorithms—not as a means of projecting onto the ground state, but as a way to improve classical estimates of the ground-state energy. Second, when QPE is performed, the energy distribution can help address the problem of the contribution of higher-energy states towards low outcome values—what we call the *leakage problem* in QPE. We show that this problem can be diagnosed when the energy distribution is at hand, and can furthermore be mitigated through quantum refining mentioned above. Finally, the energy distribution picture can be a guide towards potential quantum advantage: the amount of low-energy support of the initial state below a classical target energy estimate can be a proxy for the likelihood that quantum energy estimation algorithms can obtain lower-energy estimates than the classical reference. With this in mind, we discuss the concept of *Goldilocks problems*. By these, we mean energy estimation tasks where the initial state is neither too good, such that classical methods cheaply obtain the ground state with high accuracy (like in a small, uncorrelated molecule); nor too bad, as in the orthogonality catastrophe regime where even quantum algorithms fail (like with large clusters with multiple transition metals [25]). These extremes suggest the presence of an intermediate-complexity system where classical methods cannot obtain the ground state accurately, but the approximate ground states they can produce, if used as initial states, can allow the quantum computer to obtain the ground-state energy accurately with reasonable cost. Our energy distribution techniques can be used to search for such problems, which are candidates for quantum advantage, as we illustrate with numerical examples.

All of the subroutines discussed above combine to give an initial state preparation algorithm, which can be applied for quantum energy estimation in any quantum chemical system. The complete algorithm, illustrated in Fig. 1 consists of the following steps:

- (1) Classical computation of a candidate initial state.
- (2) Converting the candidate initial state to either SOS or MPS form.

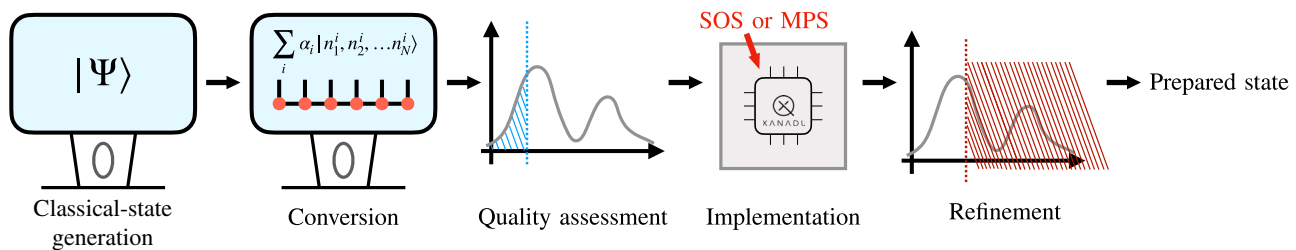


FIG. 1. The initial state preparation algorithm consists of the following steps: classical search for a low energy state, conversion of the state into a standardized form, i.e., SOS or MPS, quality assessment performed based on the energy distribution of the candidate state, implementation of the highest-quality state on the quantum computer, and quantum refining. The end-to-end procedure results in a high-quality state prepared on the quantum computer.

- (3) Assessing the quality of different candidates through the energy distribution.
- (4) Implementing the highest-quality state from (3) into the quantum computer.
- (5) Quantum refining of the state with an energy-filtering method.
- (6) With the implemented state, execute QPE or any other quantum energy estimation algorithm.

Using our algorithm, we find that for nontrivial problems of interest in quantum chemistry such as estimating ground-state energies of iron-sulfur complexes, by improving state quality we can reduce the total algorithm cost by several orders of magnitude, compared to using a single product state.

The rest of the paper is structured as follows. In Sec. II, we briefly review the traditional quantum chemistry methods we use for calculating candidate initial states. In the following Sec. III, we describe our state-of-the-art algorithm for implementing such initial states in their SOS form, as well as a separate way of implementing a state in the MPS form, providing resource estimations for both approaches. To assess and compare our candidate states, in Sec. IV we introduce the concept of the energy distribution, pioneering new methods to estimate the quality of candidate states. As a case study, we apply our new techniques to address the QPE leakage problem. Once a state is prepared, assessed, and implemented, we can use a quantum algorithm to refine it by filtering out high-energy contributions, as we discuss in Sec. V. Having described each step of our algorithm individually, in Sec. VI we look at the entire pipeline, and showcase numerical experiments that demonstrate the state-preparation algorithm for different molecules.

## II. OBTAINING CLASSICAL DESCRIPTIONS OF INITIAL STATES

The initial state preparation algorithm starts from executing a traditional quantum chemistry method to generate a candidate wavefunction. It is likely that the choice of

which method to run is highly situation-dependent, so we consider a wide variety of techniques, which we now briefly describe. We also review the concept of electronic correlation and its implications for choosing a quantum chemistry method for a given molecule. Experts are welcome to skip this section, but concepts from it will be employed throughout our paper.

Here and in the rest of the paper, we work in the second quantized language and represent states with Fock occupation number vectors. By a Slater determinant we mean a many-particle product state built by distributing the available  $n_e$  electrons over  $2N$  single-particle spin-orbitals. We will use either spatial orbital occupation numbers  $n_i = 0, \alpha, \beta, 2$  or spin-orbital occupation numbers  $n_i = 0, 1$ , typically ordered according to their Hartree-Fock energy (increasing from left to right): the choice of whether orbital or spin-orbital occupations are used will be clear from context. For example, a generic  $S_z = 0$  Hartree-Fock state is written as  $|222 \dots 200 \dots\rangle$  in the first scheme and  $|1111 \dots 100 \dots\rangle$  in the second scheme, respectively. More broadly, a generic Slater determinant reads

$$|\psi_{\text{Slater}}\rangle = |n_1, n_2, \dots, n_N\rangle \text{ or } |n_1, n_2, \dots, n_{2N}\rangle. \quad (1)$$

### A. Quantum chemistry methods for obtaining approximate ground states

The strategies we consider can be split into two groups. The first includes methods in the configuration interaction family, building the wavefunction as a superposition of Slater determinants (SOS). The second includes just one approach: representing the wavefunction using the matrix product state (MPS) ansatz, and variationally optimizing it over a series of sweeps using the density matrix renormalization group algorithm (DMRG) [47,48]. We focus first on methods based directly on the basis of Slater determinants.

The configuration interaction with single and double excitations (CISD) aims to prepare a wavefunction of the

form

$$|\psi_{\text{CISD}}\rangle = |\text{HF}\rangle + \sum_i c_i |S_i\rangle + \sum_i d_i |D_i\rangle, \quad (2)$$

with singly and doubly excited determinants  $S_i, D_i$  that are parametrized by CI coefficients  $c_i, d_i$ . The coupled cluster with the singles and doubles (CCSD) technique instead builds a wavefunction of the form

$$|\psi_{\text{CCSD}}\rangle = e^{\hat{T}_1 + \hat{T}_2} |\text{HF}\rangle, \quad (3)$$

where the excitation operators  $\hat{T}_1, \hat{T}_2$  are single and double excitation operators parametrized by amplitudes  $t^{(1,2)}$  that connect only occupied orbitals to virtual ones. Both methods are built on top of the Hartree-Fock reference state  $|\text{HF}\rangle$  and are hence termed single-reference methods.

On the other hand, instead of restricting the many-body Hilbert space in terms of excitations, we can recognize that in many cases the majority of correlation—the majority of the difference between different determinants entering the wavefunction—is restricted to a few orbitals, for example, the  $d$  orbitals of transition metals or the  $\pi$  orbitals in organic molecules. The remaining orbitals—say high-energy virtual orbitals, or deep core orbitals—are merely spectators, maintaining the same occupancy in all determinants entering the wavefunction. Given this observation, we could consider a wavefunction ansatz that freezes the occupancy of those spectator orbitals to zero (for virtual orbitals) or to doubly occupied (for core orbitals), and then perform full diagonalization in the remaining, much smaller Hilbert space. This is the nature of the complete active space configuration interaction (CASSCI) method. With  $N$  spatial orbitals and  $1 \leq l < L \leq N$ , we fix the first  $l - 1$  orbitals to be fully occupied and set the final  $N - L - 1$  orbitals to be unoccupied. Then the wavefunction takes the form

$$|\psi_{\text{CASSCI}}\rangle = \sum_{n_l, n_{l+1}, \dots, n_L} c_{n_l, n_{l+1}, \dots, n_L} |2, \dots, 2, n_l, n_{l+1}, \dots, n_L, 0, \dots, 0\rangle, \quad (4)$$

parametrized by coefficients  $c_{n_l, n_{l+1}, \dots, n_L}$ . CASSCI entails exact diagonalization within the active space of orbitals  $\{l, l + 1, \dots, L\}$ . While this lets CASSCI produce complex multireference wavefunctions that can have many determinants with similar weights, there are two limitations. First, the choice of active orbitals is widely acknowledged to be challenging [49] (although automated approaches [50,51] are gaining popularity). Second, exact diagonalization scales too prohibitively to be useful in non-trivial molecules. The impact of frozen orbitals on the CASSCI wavefunction could be partially taken into account

with multireference perturbation theory (MRPT), which is applied on top of a CASSCI calculation. The perturbative correction modifies the coefficients  $c_{n_l, \dots, n_L}$  of the CASSCI wavefunction, typically using standard second-order (Moller-Plesset) perturbation theory. In practice, MRPT is usually much better at improving the energy estimate than at improving the wavefunction: largely used for recovering dynamic correlation energy, it is not capable of adding multireference character from the nonactive space states. However, note that, in general, the idea of an active space may be freely combined with any of the other techniques described in this section: so while CASSCI-style exact diagonalization within the active space could be too prohibitive, techniques like selective configuration interaction and DMRG are performant enough to handle a large active space.

Selective configuration interaction (SCI) methods are inspired by the observation that for many wavefunctions of interest, written in the full basis as

$$|\psi_{\text{SCI}}\rangle = \sum_{n_1, n_2, \dots, n_N} c_{n_1, n_2, \dots, n_N} |n_1, n_2, \dots, n_N\rangle, \quad (5)$$

most coefficients  $c_{n_1, n_2, \dots, n_N}$  vanish. The goal is to identify an efficient way or criterion to search for these nonzero coefficients. There are a broad variety of methods in the SCI family [52–58]: among the most recent approaches that include code implementation are adaptive configuration interaction (ACI) [53], adaptive sampling configuration interaction (ASCI) [52,59], and heat-bath configuration interaction (HCI) [54,60]. In this work, we focus on HCI [54,60], owing to its open-source nature and Python interface.

At a high level, HCI starts with an initial collection of Slater determinants  $|v_i\rangle$  (or even just the Hartree-Fock determinant by itself), and finds the ground state in the subspace spanned by them, namely  $|\Psi_0^{(0)}\rangle = \sum_i c_i |v_i\rangle$ . Next, an iterative procedure begins: consider a pool of all determinants  $|v_k\rangle$  that are connected to one of  $|v_i\rangle$  by the Hamiltonian, meaning the determinants differ by only a single or a double excitation. Such a determinant  $|v_k\rangle$  will be added to the pool only if it satisfies the relatively simple criterion  $\max_i (\langle v_k | H | v_i \rangle c_i) > \epsilon_1$ , i.e., when the *largest* Hamiltonian matrix element between the candidate  $|v_k\rangle$  and some determinant  $|v_i\rangle$  from the pool, times the coefficient  $c_i$  of that determinant in the ground-state expansion, is above a user-specified cutoff  $\epsilon_1$ . (Note that this particular criterion is unique to HCI: other methods like ASCI and ACI have different criteria, which could be more useful for different classes of systems.) Once the new pool is determined this way, a new ground state  $|\Psi_0^{(1)}\rangle$  is calculated within this pool, and the process of pool augmentation begins again. The iterative loop continues until no more determinants are being added to the pool. The threshold  $\epsilon_1$



acts as a convergence parameter: setting it very high limits the pool to the original determinants, while continually decreasing it allows more and more determinants to enter the pool.

For completeness, we note that there are post-HCI perturbative approaches to further refine the energy estimate. In this work, however, we will employ only the variational version of the method as described above, as we are mainly interested in the wavefunction rather than the energy estimate, and perturbative corrections do not affect the wavefunction in the HCI implementation we are using.

Finally, we summarize the DMRG approach for obtaining wavefunctions in MPS form. DMRG has proven to be a reliable, robust, and efficient method for constructing approximate ground states for a wide variety of molecules [4,25,61–66]. An MPS can be seen as an efficient way of factorizing the general  $N$ -tensor coefficient  $c_{n_1, \dots, n_N}$  of a Slater determinant series into a product of matrices, whose size is limited by the bond dimension  $\chi$  [67]. The MPS wavefunction can be written as

$$|\psi_{\text{MPS}}\rangle = \sum_{\substack{\alpha_1, \dots, \alpha_{N-1} \\ n_1, \dots, n_N}} A_{1; \alpha_1}^{n_1} A_{2; \alpha_1 \alpha_2}^{n_2} \dots A_{N; \alpha_{N-1}}^{n_N} \\ \times |n_1, n_2, \dots, n_N\rangle, \\ n_i \in \{0, \alpha, \beta, 2\}, \quad \alpha_i \leq \chi. \quad (6)$$

This factorization scales polynomially with system size for a fixed bond dimension [68,69]. Combined with the DMRG algorithm, it gives a wavefunction-based variational approach that provably converges to the exact solution in the limit  $\chi \rightarrow \infty$ . To apply the DMRG method to molecules, which exhibit inherently nonlocal interactions between molecular orbitals and do not resemble the spin chains that DMRG was originally developed for, orbitals must be arranged along a one-dimensional chain. Ideally, the arrangement is such that it minimizes the amount of long-range nonzero molecular integrals. The standard choice is to arrange the molecular orbitals according to their Hartree-Fock energy; more sophisticated reordering schemes are also considered [65].

Throughout the paper we will find it convenient to convert between the SOS and MPS formats regardless of the original method (e.g., CISD or HCI or DMRG) used to obtain the wavefunction: we describe how the conversion is done in Appendices A and B.

The diversity of wavefunction forms resulting from different methods presented above means that in practice it is difficult to compare and evaluate them. To do the comparison, we will use the publicly available software package Overlapper [70]. On the other hand, this variety suggests that there are trade offs that could be exploited depending on the molecule being considered. The strengths and weaknesses of different approaches are largely determined by

the amount and type of correlation present in the system, as we discuss next.

## B. Electronic correlations

Correlation energy is the portion of total system energy that is not accounted for by the Hartree-Fock ansatz. It is due to Coulomb interactions between electrons.

Strong correlation necessitates the use of additional determinants in the many-body wavefunction for an accurate description. Depending on the amount of correlation and its type, different methods might be preferable. Correlation energy is usually partitioned into static and dynamic types. Each of these has a unique origin, and thus different quantum chemistry methods will provide different-quality initial states for each situation.

Dynamic correlation is a consequence of electrons avoiding each other due to Coulomb repulsion. When electrons approach each other in real space—as when sharing a spatial orbital—their wavefunctions acquire nonanalytic cusps approximately  $e^{-|x|}$  due to the Coulomb potential  $1/r$  divergence. Resolving these cusps with smooth, Gaussian-type orbitals requires a large basis set, and many determinants. Dynamic correlation leads to the many-body wavefunction having one dominant determinant with a large weight, often the Hartree-Fock state, together with many small-weight contributions necessary to account for the nonanalytic cusps. While the weights of these additional determinants are small, there are many of them: if they are neglected, such as when using a simple initial state like Hartree-Fock, large energy errors relative to the true ground-state energy will result. Fast single-reference methods such as CISD and CCSD are usually capable of recovering dynamic correlation energy well, as they are able to work with a larger basis set and their single-reference nature is conducive to the task. On the other hand, scale limitations of CASCI, DMRG, and HCI make them comparatively worse at recovering this type of correlation.

By contrast, static correlation arises in the presence of nearly degenerate eigenstates: in such a situation, multiple determinants with roughly similar weights are needed for an accurate description of the ground-state wavefunction. Typical situations where static correlation arises are nonequilibrium geometries, low-spin states of open-shell molecules (spin-state degeneracy), excited states and molecules containing transition-metal atoms (due to high degeneracy of  $d$ -type orbitals). In such situations a single-reference method, such as the Hartree-Fock state or CISD and CCSD, will not be a good starting point, because fundamentally they are both built around a single reference, which will remain dominant despite optimization. Instead, a multireference method—one that does not give preference to any particular reference—is needed, for example, CASCI, DMRG, or HCI. In practice, CASCI is strongly

limited by the active space size; and while MRPT improves the CASCI energy appreciably through adding dynamic correlation, it provides only minor improvements to the wavefunction itself, which is the object of interest in quantum algorithms. This leaves DMRG and HCI as the leading contenders. DMRG can boast polynomially efficient representation of even strongly multireference states and relatively straightforward convergence with bond dimension; by contrast, HCI's lower computational demands make it easier to run calculations for larger spaces, without sacrificing accuracy even for correlated systems such as the chromium dimer [66].

### III. EFFICIENT ANSATZ IMPLEMENTATION ON QUANTUM COMPUTERS

In Sec. II, we detailed how approximate ground states can be found using classical computational methods and expressed in two standardized forms: SOS (Appendix A) and MPS (Appendix B). In this section, we detail how SOS and MPS states can be implemented on a quantum computer and estimate the number of qubits and gates required for these tasks. Typically the cost of encoding classical states is lower than the cost of the energy estimation algorithm, even for sophisticated states with many Slater determinants or large bond dimension. This results in considerable runtime reductions of the full algorithm by lowering the number of repetitions needed to achieve a target accuracy, while incurring only small increases in the cost of each independent run.

#### A. Sum of Slater determinants (SOS)

The goal is to prepare the normalized state

$$|\psi\rangle = \sum_{i=1}^D \alpha_i |v_i\rangle, \quad (7)$$

where  $\alpha_i$  are the given amplitudes and  $|v_i\rangle = |n_{i,1}n_{i,2}\dots n_{i,2N}\rangle$  are states of  $2N$  qubits. The bits  $n_{i,j}$  denote the occupation number of spin-orbital  $j$  for the  $i$ th Slater determinant  $|v_i\rangle$ . We use  $N$  to denote the total number of spatial orbitals, each supporting one spin-up and one spin-down particle.

We are interested in cases where the number of Slater determinants  $D$  is much smaller than  $2^{2N}$ , which is the case in practice. Therefore, the problem is to prepare a summation of relatively few basis states picked from a very large Hilbert space. The gate cost of the algorithm will be measured in terms of the number of non-Clifford Toffoli gates, which is a standard complexity measure used in the literature. On a fault-tolerant architecture, Toffolis require orders of magnitude more qubit seconds and physical qubits due to the need for magic state distillation to implement them. Since our algorithm is explicitly meant

for state preparation on fault-tolerant quantum computers, we use Toffoli gate count (and not CNOT gate count) to measure the cost.

There has been previous work in this direction. Reference [18] proposes an iterative generation of the state using  $(2N - 1)(D - 1)$  Toffoli gates and  $2N - 1$  ancilla qubits (when  $D > 1$ ). Other algorithms include [26,28], but have Toffoli complexity even higher than  $O(ND)$ , or potentially exponential in  $N$  [27]. Instead, we present an algorithm with asymptotic runtime of  $O(D \log D)$ , where  $\log$  is in base two throughout this paper. This is a considerable improvement since the number of Slaters  $D$  is at most the full space  $2^{2N}$ , and so, it is often the case that  $\log D \ll 2N$ : Fig. 4 makes this comparison explicit for a few different model systems and molecules studied in this paper and elsewhere. Notice that the advantage is even more explicit for larger systems; for example, with  $N \sim 400$  spin-orbitals, our algorithm is an order of magnitude more efficient in Toffoli cost as long as  $D < 2^{40} \sim 10^{12}$ .

We aim to prepare the state in Eq. (7), where  $|v_i\rangle$  represents the occupation bitstrings of length  $2N$  mentioned above. Our main technical result is a mapping from the bitstrings  $v_i$  of  $2N$  bits, that identify each Slater determinant, to more compact and unique bitstrings  $b_i$  of only  $O(\log D)$  bits. This compression from size  $N$  to  $O(\log D)$  is the key reason behind the Toffoli complexity advantage of our method. The following lemma formalizes the compression scheme. We will assume that  $D$  is a power of two for convenience; otherwise,  $\log D$  should be replaced with its ceiling.

*Lemma 1.* Given as input a set  $\{v_i\}$  of bitstrings representing unique Slater determinants, there is a classical algorithm with complexity  $O(tD^2)$ , where  $t \leq \min(2N, D) - 2 \log D + 1$ , that outputs substrings  $\tilde{v}_i$  of  $v_i$  and  $2 \log D - 1$  bitstrings  $u_k$  of length  $O(\min(2N, D))$ , such that the bitstrings  $b_i := (u_1 \cdot \tilde{v}_i, \dots, u_{2 \log D - 1} \cdot \tilde{v}_i)$  presented as column vectors in

$$\begin{pmatrix} u_1^T \\ \vdots \\ u_{2 \log D - 1}^T \end{pmatrix} (\tilde{v}_1 \quad \dots \quad \tilde{v}_D) = (b_1 \quad \dots \quad b_D), \quad (8)$$

are mutually distinct, i.e.,  $b_i \neq b_j$  for all  $i$  and  $j$ .

The proof can be found in Appendix C. The algorithm to implement the SOS state in Eq. (7) is described below. We employ three registers: the system register, used to store  $v_i$  and where we wish to prepare the desired state, and two ancilla registers: an enumeration register with  $\log D$  qubits enumerating the Slater determinants, and an identification register with  $2 \log D - 1$  qubits storing  $b_i$ .

### Quantum algorithm for encoding SOS states

1. Prepare the state

$$\sum_{i=1}^D \alpha_i |0\rangle |i\rangle |0\rangle, \quad (9)$$

in the enumeration register using the quantum read-only memory (QROM) state-preparation method in Ref. [71].

2. Use a QROM oracle  $O$  of Toffoli cost  $D$  as in Ref. [71] that implements the transformation

$$O |0\rangle |i\rangle |0\rangle = |v_i\rangle |i\rangle |0\rangle. \quad (10)$$

This results in the state

$$\sum_{i=1}^D \alpha_i |v_i\rangle |i\rangle |0\rangle. \quad (11)$$

3. Using the output bitstrings  $u_k$  from Lemma 1, do the following. If the  $j$ th bit of  $u_1$  is equal to 1, apply a CNOT gate between the system register and the identification register. The CNOT is controlled on the  $j$ th qubit of  $|\tilde{v}_i\rangle$ . This results in the state

$$\sum_{i=1}^D \alpha_i |v_i\rangle |i\rangle (|u_1 \cdot \tilde{v}_i\rangle |0\rangle) \quad (12)$$

$$= \sum_{i=1}^D \alpha_i |v_i\rangle |i\rangle (|b_{i,1}\rangle |0\rangle). \quad (13)$$

Notice the bits in  $u_1$  are matched with the bits of the substring  $\tilde{v}_i$  obtained from Lemma 1.

4. Repeat the above step for all  $u_k$ . This results in the state

$$\sum_{i=1}^D \alpha_i |v_i\rangle |i\rangle |b_i\rangle, \quad (14)$$

which now contains the unique compact identifier  $b_i$  for the Slater determinant  $|v_i\rangle$ .

5. Using multicontrolled operations, apply the transformation  $|i\rangle \rightarrow |0\rangle$  to the first ancilla register conditioned on each unique  $b_i$ . This costs  $(2 \log D - 2)D$  Toffolis and results in the state

$$\sum_{i=1}^D \alpha_i |v_i\rangle |0\rangle |b_i\rangle. \quad (15)$$

6. The final step is to uncompute the sequence of CNOTs used to prepare the state  $|b_i\rangle$  in steps 3 and

(a)

$$\begin{matrix} \nu_1 & \nu_2 & \nu_3 & \nu_4 \\ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix} \longrightarrow \begin{matrix} \tilde{\nu}_1 & \tilde{\nu}_2 & \tilde{\nu}_3 & \tilde{\nu}_4 \\ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

(b)

$$\begin{matrix} u_1 \\ u_2 \\ u_3 \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} \tilde{\nu}_1 & \tilde{\nu}_2 & \tilde{\nu}_3 & \tilde{\nu}_4 \\ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix} = \begin{matrix} b_1 & b_2 & b_3 & b_4 \\ \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

(c)

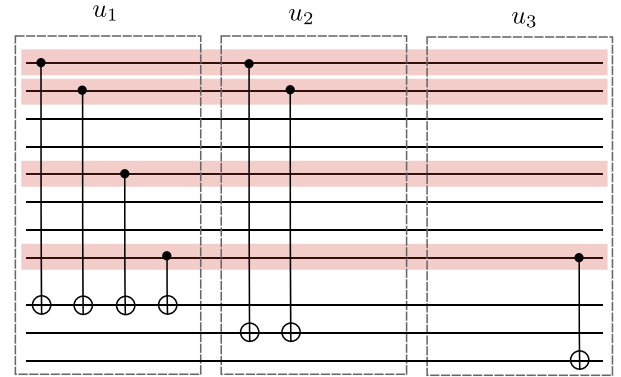


FIG. 2. Schematic representation of important steps in the SOS encoding algorithm: for illustrative purposes, we show a particular example with  $2N = 8$  orbitals and  $D = 4$  Slater determinants, for which the identification strings require  $2 \log D - 1 = 3$  bits. (a) Matrix of Slater determinant strings  $\nu_i$ . By selecting only rows 1, 2, 5, and 8, we can construct substrings  $\tilde{\nu}_i$  that form a matrix of full rank. This is the minimum number of rows that preserve the rank of the original matrix. (b) Using the result from Lemma 1, we construct bitstrings  $u_i$  that form a linear map transforming the substrings  $\tilde{\nu}_i$  to the identification bitstrings  $b_i$ . (c) The encoding quantum algorithm applies a series of CNOT operations for each  $u_i$ , acting only on qubits 1, 2, 5, and 8 in the system register, in accordance to the choice of bitstrings. These are responsible for setting every individual qubit in the identification register to the desired value.

4. This leads to the final output

$$\sum_{i=1}^D \alpha_i |v_i\rangle |0\rangle |0\rangle, \quad (16)$$

which contains the target state in the system register disentangled from all other registers, as desired.

Important steps of the algorithm are illustrated in Fig. 2.

Taking into account the first two usages of QROM for preparing  $\sum_{i=1}^D \alpha_i |i\rangle$  and in Eq. (10), with costs dominated by  $D$  and  $2^{\log D+1}$ , respectively, the overall Toffoli cost is dominated by

$$\begin{aligned} (2 \log D - 2)D + 2^{\log D+1} + D < \\ (2 \log D + 3)D = O(D \log D). \end{aligned} \quad (17)$$

The overall additional qubit cost due to use of ancillas is

$$4 \log D - 3 + \log D = 5 \log D - 3 = O(\log D). \quad (18)$$

One could trade off Toffolis with qubits, within the same volume cost of  $O((\log D)^2 D)$ . In most cases, this means using the SELSWAPDIRTY variant of QROM [71], also called QROAM in Ref. [72]. Importantly, this variant allows using uninitialized qubits for this trade off. The trade off leads to a Toffoli cost of  $\min(2\sqrt{32ND}, D) + (7 \log D + 2\sqrt{32 \log D})\sqrt{D}$ , a clean qubit cost of  $(2 \log D - 1)\sqrt{D}$ , and an additional uninitialized qubit cost of  $\sqrt{32ND}$ . We explain this in details in Appendix D, along with comments on how one could lower the expected Toffoli cost by combining our strategy with Ref. [18].

In the above, we employed the formalism of second quantization, but the implementation method is general and can also be used for algorithms employing a first quantization representation, or indeed in any situation where the states  $v_i$  are represented by bitstrings. More precisely, in first quantization quantum algorithms, the size of the Hilbert space scales as  $N^n$ , where  $N$  is the size of the basis set. This number is far larger than  $D$ , especially in industrially relevant applications of the quantum algorithms using plane waves [73], meaning that the technique should remain an advantageous method of state implementation in first quantization.

## B. Matrix product state (MPS)

Here we discuss how to implement initial states expressed in the MPS form. While it is always possible to transform an MPS into an SOS formulation, direct implementation of an MPS can be beneficial in certain cases. We consider mainly the method introduced in Ref. [30] and perform an estimation of the total Toffoli cost. We note that there are many variations of this technique considered in the literature [74–76] and also newer versions requiring lower-depth circuits [31] for short-range correlated MPSs.

First, we give a quick review of the method in Ref. [30]. We start with the MPS form shown in Eq. (6) and use standard graphical notation for representing the MPS.

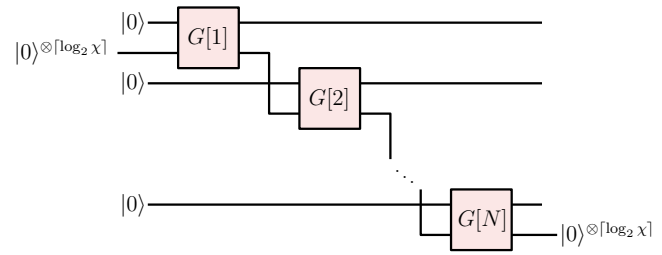


FIG. 3. Circuit for MPS implementation. Since the bond dimension can change as the circuit traverses the system, one can start with a number of ancillae equal to  $\lceil \log \chi_{\max} \rceil$  and use more or less of the available ancillae as the bond dimension dictates.

We denote its tensors as

$$A_{j; \alpha_{j-1} \alpha_j}^{n_j} = \alpha_{j-1} \boxed{A_j^{n_j}} \alpha_j. \quad (19)$$

The physical index  $n_j$  runs over  $d$  values, where  $d$  is the local Hilbert space dimension. The auxiliary indices  $\alpha_j$  run over  $\chi_j$  values, where  $\chi_j$  is called the bond dimension, which generally may be different for each index  $j$ . For the implementation, the MPS is turned into its left-canonical form:

$$\boxed{A_1^{n_1}} \boxed{A_2^{n_2}} \boxed{A_3^{n_3}} \cdots \boxed{A_N^{n_N}}. \quad (20)$$

It means that for all  $j > 1$ , we have

$$\sum_{\alpha_j, n_j} A_{j; \alpha_{j-1} \alpha_j}^{n_j} \left( A_{j; \alpha'_j \alpha'_{j-1}}^{n_j} \right)^* = \delta_{\alpha_{j-1} \alpha'_{j-1}}, \quad (21)$$

or diagrammatically

$$\boxed{A_j^{n_j}} \boxed{A_j^{n_j}} = \boxed{1}. \quad (22)$$

This is in general possible using singular value decomposition of tensors [69]. A note regarding notation: in the above equations and everywhere else, the summation over left and right auxiliary indices of the leftmost and the rightmost tensors, respectively, can simply be dropped. This is equivalent to setting  $\chi_0, \chi_{N+1} = 0$ .

The implementation works by first observing that the above tensors of the MPS, owing to the left-canonical form, can be directly used to define unitaries, which we



denote as  $G$ , that are used in a quantum circuit for preparing the MPS:

$$G[j]_{\alpha_j n_j, \alpha_{j-1} 0} = A_{j; \alpha_{j-1} \alpha_j}^{n_j},$$

$$\alpha_{j-1} \text{---} \boxed{A_j} \text{---} \alpha_j \rightarrow \alpha_{j-1} \text{---} \boxed{G[j]} \text{---} \alpha_j. \quad (23)$$

Each unitary acts on a  $d$ -level system composed of  $\lceil \log(d) \rceil$  qubits as well as  $\lceil \log \chi \rceil$  ancillae, where we remind the reader that  $\log$  is in base two throughout this paper. For example, for fermionic systems,  $d = 4$  and two qubits are required for each spatial orbital of the system. We note that  $G[j]$  has more degrees of freedom than  $A[j]$ . In particular, we note that in Eq. (23) the incoming physical index for the unitary is set to be equal to 0. Thus the above relation does not specify all the elements of  $G[j]$ , and there is some freedom in  $G[j]$  for when the incoming index is different from 0. However, such a situation does not occur in the MPS preparation circuit and thus as long as the rest of the elements are chosen so that  $G[j]$  remains unitary, the effect of  $G[j]$  on such states can be taken to be arbitrary. In fact in the unitary synthesis described below we have used this freedom and focused only on states with 0 on the incoming physical index.

A quantum circuit that implements the desired MPS and works with these unitaries is shown in Fig. 3. Note that an auxiliary register of a size, which we have denoted collectively as  $\lceil \log \chi \rceil$ , is required to reconstruct the MPS and that we are schematically moving it around to act with unitaries on this register and different qudits in the system. Also, note that for the first and the last unitaries the input and output auxiliary registers also have value 0.

For the above circuit, we need to synthesize the unitaries  $G[j]$ , for which we use the method in Ref. [71]. Details of how the synthesis can be performed is discussed in Appendix E. There, it is shown that each  $G[j]$  imposes a Toffoli cost of

$$\chi_{j-1} [8\chi_j d + (b+1) \log(\chi_j d)], \quad (24)$$

where  $b$  is the number of digits for storing the angle for implementing single-qubit rotations required in the synthesis process.

Assuming a number  $N$  of qudits in the physical system, the total cost will be the sum of the above over all  $j$ ; asymptotically and with using  $\chi$  selectively for all bond dimensions, the dominant Toffoli cost can be written as  $O(N\chi^2)$ . Using trade off schemes of Ref. [71], we can show that it is also possible to use  $O(N\chi^{3/2})$  Toffolis for implementation of the MPS.

### C. Discussion of the implementation methods

As with many questions concerning initial state preparation, the choice of SOS vs MPS is highly context dependent. The polynomial scaling (with system size) of the number of parameters in an MPS may provide it with a decisive advantage for strongly multireference systems; while in single-reference molecules (or ones not being highly multireference), the reduced cost of implementing the SOS form might be preferable. A concrete comparison between these two methods is done in Sec. VIB. Either way, these two forms are interconvertible, and we have outlined advanced quantum algorithms for implementing states of either form. This includes a novel method for SOS states with lower asymptotic cost than the previous state of the art.

Regardless of the method, the cost of implementing a classical state on a quantum computer is typically a lot lower than that of the full energy estimation algorithm. Our estimates earlier in this section (such as in Fig. 4) for preparing even quite complex initial states with hundreds or thousands of determinants for systems with over a hundred orbitals, suggest state-preparation costs of only around  $10^5$ – $10^6$  Toffoli gates. At the same time, estimates from the literature for the cost of QPE for industrially relevant contexts (for example, FeMoco [2,77], cytochrome P450 [78], electrolyte molecules in batteries [79], or battery cathodes [8,73] and general periodic crystals [80])

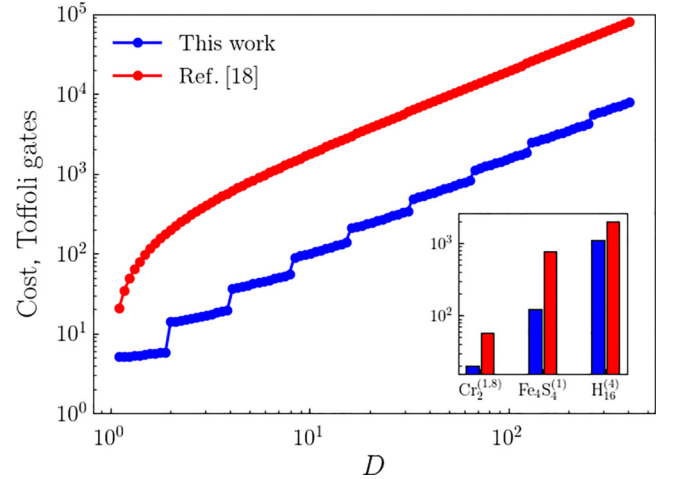


FIG. 4. Comparing the cost of implementing an SOS wavefunction with  $D$  determinants between the algorithm proposed here and that of Ref. [18] for a hypothetical system with  $N = 100$  spatial orbitals. Inset: comparing the cost of preparing an SOS wavefunction of quality 0.2, as measured through overlap with the highest-fidelity reference, for specific molecules. In the inset, the superscript on the molecule formula indicates how much the bond is stretched relative to the approximate equilibrium bond length as noted in more detail in Sec. VI. Note that system size  $N$  varies between the molecules: it comprises 10 spatial orbitals for  $\text{Cr}_2$ , 32 for  $\text{Fe}_4\text{S}_4$ , and 16 for  $\text{H}_{16}$ . The number of determinants required to achieve a quality of 0.2 also varies.

suggest a Toffoli gate cost of  $10^{10-14}$  gates for molecules and  $10^{12-19}$  for materials—for a single run of QPE. With such a high cost ratio, it is prudent to implement as good of an initial state as one can find, regardless of how complex it is, because the extra cost incurred in this way will be dwarfed by a corresponding cost reduction due to fewer iterations needed for the expensive QPE routine. This means effectively there is a large budget available for implementing sophisticated states with many Slater determinants or large enough bond dimension that better approximate the true ground state compared to simpler approaches like the Hartree-Fock state. This can lead to considerable runtime reductions for the entire quantum algorithm.

#### IV. ENERGY DISTRIBUTION OF THE INITIAL STATE

Using the overlap with the ground state as a way to assess the quality of an initial state is a challenging task, especially in strongly correlated systems. This is largely because the true ground state is generally not known: after all, this is the problem we are attempting to solve. Instead, we propose a new way to assess state quality through the use of the state's associated *energy distribution*. In this section, we first define the concept of an energy distribution in precise terms. We then discuss how the energy distribution picture can change our view of performing quantum phase estimation (QPE). Our main technical contribution is a description of methods for approximating energy distributions, as well as formalizing how the energy distribution picture can be used to predict statistics of QPE outcomes. Finally, we describe how the energy distribution of the initial state can be used to address the so-called QPE leakage problem.

##### A. Definition of the energy distribution

We define the energy distribution of the state  $|\psi\rangle$  with respect to the Hamiltonian  $H$  as

$$P(E) = \sum_n |\langle E_n | \psi \rangle|^2 f_\eta(E_n - E), \quad (25)$$

where  $E_n, |E_n\rangle$  are, respectively, eigenvalues and eigenstates of  $H$ , and  $f_\eta$  is a kernel, for example, Gaussian or Lorentzian, with width  $\eta$ , a copy of which is centered at each of the eigenvalues. The limit of  $\eta \rightarrow 0$  corresponds to a discrete distribution—the actual distribution of the state's overlaps with the Hamiltonian spectrum. With  $\eta \neq 0$ , each energy level is broadened and the result is a continuous distribution. While accessing the true  $\eta = 0$  distribution is as hard as determining the eigenvalues of the Hamiltonian, estimating it for a given  $\eta \neq 0$  turns out to be much cheaper—and much can be inferred about state quality from the broadened distribution.

We use the energy distribution in a number of applications. Most importantly, we formulate a simple criterion based on the energy distribution for assessing the quality of initial states. Suppose we are given a number of candidate states with similar energies, i.e., with similar expectation values of  $H$ . We can compare the quality of the states by focusing on the left-side tails of the states' energy distributions. Intuitively, whichever state has a distribution with more weight at lower energies is a better candidate, as it provides a higher probability for obtaining a low-energy estimate. We will make this statement more precise in Sec. IV D. The state thus identified becomes an input to the next stage of the initial state preparation pipeline (Fig. 1)—the implementation stage.

##### B. Quantum phase estimation through the lens of energy distributions

In many of the quantum routines for quantum energy estimation, the energy distribution has a very close relation with the distribution of outcomes. Most prominently, the outcomes of QPE are roughly sampled from the energy distribution of the initial state. More precisely, in a QPE measurement with  $k$  phase digits, the probability of an integer outcome  $x_m$  (that can be interpreted as an estimated energy of  $2^{-k}x_m$ ) in the phase register reads [11]

$$\sum_n |\langle E_n | \psi \rangle|^2 \frac{1}{2^{2k}} \left( \frac{\sin^2(\pi 2^k E_n)}{\sin^2(\pi [E_n - x_m/2^k])} \right). \quad (26)$$

Notice that in the above, we have assumed a normalization of the Hamiltonian such that  $0 \leq E_n \leq 1$  for all  $n$ . We similarly assume that the integer outcome  $x_m$  satisfies  $0 \leq x_m < 2^k$ . The discrete QPE kernel  $(1/2^{2k}) ((\sin^2(\pi 2^k E_n))/(\sin^2(\pi [E_n - x_m/2^k])))$  broadens each energy level, and is maximized when the integer  $x_m$  is closest to  $2^k E_n$  for each level. Thus, performing QPE can be thought of as sampling from a discrete distribution, which is obtained by spreading the weight of each energy level by the discrete QPE kernel. As the number of digits  $k$  increases, the sampling gets closer to sampling from the actual underlying distribution.

The traditional perspective on QPE is that the algorithm must be repeated enough times so that there is a high probability of sampling the ground-state energy (up to the allowed precision). This view tacitly implies that all samples other than the ground-state energy should be discarded as useless.

Instead, we recognize that the goal of any algorithm, whether classical or quantum, is to provide the best possible energy estimates. Ideally the estimate is exactly equal to the true ground-state energy, but that may be too ambitious in practice, especially for large systems with strong electronic correlations. The energy distribution picture presents an alternative, where QPE is viewed as a method

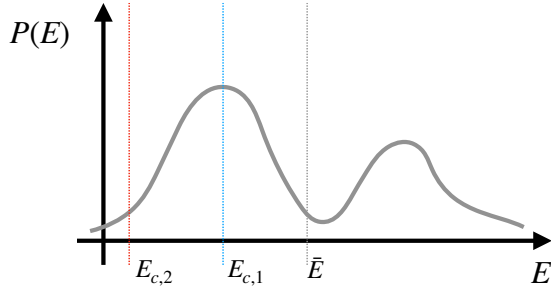


FIG. 5. Schematic depiction of an energy distribution for a particular initial state, illustrating the possibility of improvement over classically found energy estimates. If the best classical energy estimate is  $E_{c,1}$ , there is a high chance of obtaining better quantum estimates using quantum routines e.g., QPE. However, if the best classical estimate is  $E_{c,2}$ , the chance is quite slim.  $\bar{E}$ , the energy of the implemented state is also shown, but its value is irrelevant for predicting the likelihood of obtaining better energy estimates. Note that in practice, one must ensure that the weight to the left of each of these energy estimates is not due to the broadening of higher-energy weight; such broadening [ $\eta$  in Eq. (25)] is inevitable in any actual calculation, but one can examine the behavior of the tails as  $\eta$  is varied to determine whether the weight in the tail is due to such broadening or not.

to *improve* the energy estimate associated with the initial state, as we now describe.

Consider an initial state  $|\psi\rangle$ , for example, obtained by executing some classical computational method (Hartree-Fock, configuration interaction, or similar). This state has a representation  $|\psi\rangle = \sum_n c_n |E_n\rangle$  in terms of the eigenstates  $|E_n\rangle$  of the Hamiltonian. The overlaps  $c_n = \langle E_n | \psi \rangle$  define an energy distribution  $P(E)$  via Eq. (25)—suppose it has the shape as pictured in Fig. 5. The (variational) energy  $\bar{E}$  of the state  $|\psi\rangle$  is

$$\bar{E} = \langle \psi | H | \psi \rangle = \sum_n |c_n|^2 E_n = \sum_E P(E) E, \quad (27)$$

exactly equal to the mean of the energy distribution. Since the distribution has nonzero probability density on either side of the mean classical energy  $\bar{E}$ , QPE will, in general, sample energies that are higher or lower than  $\bar{E}$ : thus already after a handful repetitions, we will obtain an energy bound on the ground state below the classical estimate  $\bar{E}$ , even if we do not necessarily sample the ground-state energy  $E_0$ . In the particular example in Fig. 5, the large peak in the energy distribution to the left of  $\bar{E}$  means that it is quite likely for QPE sampling to yield a better energy estimate than  $\bar{E}$ , which would be the value add of the quantum algorithm. This identification of low-energy components within an initial state by projection is the unique aspect of QPE, unavailable to classical algorithms.

Now consider two possibilities. Suppose that within a given computational budget, the best possible classical calculation results in a classical energy estimate  $E_{c,1}$ , the blue

line in Fig. 5. Suppose also that the state  $|\psi_{c,1}\rangle$  associated with that calculation is too expensive to prepare in the quantum register, so we continue to use  $|\psi\rangle$  with energy distribution  $P(E)$  as the initial state. Looking at the energy distribution, we find that there is quite a large probability for QPE to sample energies lower than  $E_{c,1}$ : not too many iterations are needed, and thus the use of the quantum algorithm is advantageous. The knowledge of  $E_{c,1}$  helps decide whether or not the energy distribution of  $|\psi\rangle$  is good enough.

On the other hand, if the best obtainable classical energy is  $E_{c,2}$  (red line in Fig. 5), then the energy distribution analysis shows that it is quite unlikely to sample an energy with QPE from the initial state  $|\psi\rangle$  that will be better than the classical estimate. This situation can arise if either (a) the system is simple enough that classical methods give a very good solution, in which case quantum algorithms are not needed, or (b) the initial state chosen for the problem is insufficient to improve on the best classical estimate, meaning one should look for another state  $|\psi'\rangle$  with a better energy distribution. (There are two main ways to search for a better state: one is to try different ansatz, as discussed in Sec. II, and/or optimize those ansatz further, for example, increase the bond dimension of the MPS ansatz; the other is to apply quantum filtering techniques, as will be discussed in Sec. V.) In either case, the energy distribution allows this analysis to be performed, unlike the ground-state overlap metric.

Generating a larger number of samples increases the probability of observing more dramatic improvements, with the ultimate goal of obtaining precisely the ground-state energy. Employing quantum algorithms is advantageous whenever there is a sizeable probability of obtaining an energy estimate that is lower than any classical method, including more powerful ones than those used for the initial state. This concept is illustrated in Fig. 5. The energy distribution picture moves the framing away from questions like “what resources are required for obtaining the ground state?” to ones like “with the given resources, how much improvement in the ground state energy is possible?”

The energy distribution also shows why the energy is an unreliable metric of initial state quality. Since the energy is only the first moment of the energy distribution [Eq. (27)], it only weakly constrains the distribution, leaving much freedom for improving the distribution without changing the energy—or even while making the energy worse. For instance, in a distribution with multiple peaks, one can increase the weight near the ground state while simultaneously shifting other peaks farther out to increase the energy. An example of this in a realistic  $\text{Cr}_2$  system is shown in Sec. VID.

The energy distribution picture is useful beyond helping reinterpret QPE. The QPE kernel discussed above has long algebraic tails on the two sides of each energy level: as a result, it is, in general, possible that an outcome indicating

a false low energy is obtained, that was actually in the long tail of much higher-energy levels. Note that this is contrary to the occurrence of an outcome due to the actual weight in its vicinity; we call this phenomenon the QPE leakage problem. In fact, in general, it is possible for the observed outcome to lie below the ground-state energy of the Hamiltonian, rendering it unphysical. We will discuss the problem in detail along with ways it can be diagnosed and avoided—by employing the energy distribution—in Sec. VD.

### C. Approximating energy distributions

As mentioned above, computing energy distributions exactly is exceedingly difficult: however, approximating them to a degree of accuracy within one's computational budget is always possible, and the information gained is useful for choosing initial states and beyond. We provide three different methods for approximating the energy distribution with respect to a Hamiltonian  $H$  for an initial state  $|\psi\rangle$ . Two of them are classical methods and one is quantum.

#### 1. Series expansion

This method employs moments of energy (expectation values of powers of  $H$ ) to obtain a series expansion for the energy distribution. We consider the Edgeworth series and the Gram-Charlier series, which both approximate a distribution as a Gaussian multiplied by different orders of the Hermite polynomials. The coefficients in the series can be written in terms of the moments of the distribution, i.e., the expectation values of powers of the Hamiltonian  $\langle\psi|H^n|\psi\rangle := \langle E^n\rangle$ . The lowest-order approximation is a Gaussian distribution with a variance proportional to that of the initial state, namely  $\langle E^2\rangle - \langle E\rangle^2$ . The series expansion method works best for distributions that are nearly Gaussian.

The Edgeworth and Gram-Charlier series have identical terms: the only difference is that the terms in an Edgeworth series are arranged in a way that the series constitutes a true asymptotic series [81]. The Gram-Charlier series expansion for the energy distribution  $P(E)$  can be written as

$$\tilde{P}(E) = \frac{\exp(-E^2/2)}{\sqrt{2\pi}} \left[ 1 + \sum_{n=3}^{\infty} (-1)^n c_n He_n(E) \right], \quad (28)$$

where  $He_n(E)$  is the Hermite polynomial in the probabilist's notation defined as

$$He_n(E) = (-1)^n \exp(E^2/2) \frac{d^n}{dE^n} \exp(-E^2/2). \quad (29)$$

The expansion in the form in Eq. (28) is used for a distribution function with zero mean and unit variance; any

TABLE I. Coefficients defined in Eq. (30).

Order	Gram-Charlier coefficient
3	$-\frac{1}{3!}\mu_3$
4	$\frac{1}{4!}[\mu_4 - 3]$
5	$\frac{1}{5!}[-\mu_5 + 10\mu_3]$
6	$\frac{1}{6!}[\mu_6 - 15\mu_4 + 30]$
7	$\frac{1}{7!}[-\mu_7 + 21\mu_5 - 105\mu_3]$
8	$\frac{1}{8!}[\mu_8 - 28\mu_6 + 210\mu_4 - 315]$

distribution can be cast in this form upon translating and rescaling. The coefficients of the Gram-Charlier expansion are given by

$$c_n = \frac{(-1)^n}{n!} \int dE P(E) He_n(E). \quad (30)$$

The coefficients can be written in terms of the moments of the distribution

$$\mu_n = \int dE P(E) E^n = \langle E^n \rangle, \quad (31)$$

once the Hermite polynomials are expanded. A list of the coefficients is given in Table I.

The Edgeworth series is obtained by regrouping the same terms from a Gram-Charlier:

$$\tilde{P}(E) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \left[ 1 + \frac{\kappa_3}{6} He_3(E) + \left( \frac{\kappa_4}{24} He_4(E) + \frac{\kappa_3^2}{72} He_6(E) \right) + \dots \right] \quad (32)$$

where  $\kappa_n$  is the  $n$ th cumulant of the distribution. For the general prescription for obtaining the terms and also explicit forms for more terms, see Appendix F.

In practice, the above series expansions may exhibit negative distribution values or artificial rapid oscillations. This happens mostly in cases where the energy distribution is far from Gaussian—for example, when large gaps are present in the (true) energy distribution of the initial state. Generally, the series will converge if the approximated distribution function falls faster than  $\exp(-E^2/4)$  [81].

The convergence condition is satisfied for a bounded energy spectrum, but the discreteness of the true energy distribution, from which the moments are calculated, can cause the series approximation to show rapid oscillations at higher orders. Moreover, we have seen in our numerics that



Gram-Charlier series is generally more well behaved than Edgeworth series. In general, it is challenging to quantify or bound the errors of the method precisely—the difficulty being that the series-expansion method works with an asymptotic series, and at present there are no reliable techniques for estimating the error in such cases. Thus the main way to judge the method’s accuracy is on empirical grounds. However, this method can still provide some valuable information at an affordable computational cost.

Given the above discussion, to obtain a series approximation for the energy distribution, it suffices to calculate the moments  $\langle E^n \rangle$ . This can be done directly from knowledge of the classical wavefunction  $|\psi\rangle$  and the system Hamiltonian  $H$ , although it can become computationally intensive. In this work, we mainly employ an MPS representation of the states and a matrix-product operator (MPO) expression for the Hamiltonian to compute moments. We find that this does not impose a large computational overhead, as acting with the Hamiltonian MPO on the solution MPS even multiple times is not a prohibitive computational task. In principle, the series expansion can work with any representation, the only requirement is being able to calculate low-order moments  $\langle H^n \rangle$  of the Hamiltonian. We opted for the MPS-MPO representation because we found it computationally more performant in practice. In general, quantifying the computational complexity of various ways of calculating  $\langle H^n \rangle$  is difficult due to the heuristic nature of the DMRG method, and also the fact that it is not predetermined how many moments are required and how costly they will be. However, we can expect that for MPS-style methods, only a polynomial cost is required.

## 2. The resolvent method

The energy distribution as defined in Eq. (25) can be thought of as the imaginary part of a particular Green’s function,  $P(E, \eta) = -\text{Im } G(E, \eta)$ , with the Green’s function defined as

$$G(E, \eta) = \frac{1}{\pi} \langle \psi | \frac{1}{H - E + i\eta} | \psi \rangle. \quad (33)$$

Transforming to the Lehmann representation by inserting the resolution of the identity, we find

$$P(E, \eta) = -\frac{1}{\pi} \sum_n |\langle E_n | \psi \rangle|^2 \text{Im} \left( \frac{1}{E_n - E + i\eta} \right), \quad (34)$$

and using the fact that

$$\text{Im} \frac{1}{E_n - E + i\eta} = \frac{-\eta}{(E_n - E)^2 + \eta^2}, \quad (35)$$

we can see that computing  $-\text{Im } G(E, \eta)$  is equivalent to Eq. (25) for a Lorentzian kernel with broadening  $\eta$ . Thus

any method that can calculate  $G(E, \eta)$  can be used to find the approximate energy distribution [52,53,59,82–84]. We refer to calculating the energy distribution through its associated Green’s function as the *resolvent method*. Note that the resolvent method works only with the Lorentzian kernel.

One approach to calculate the Green’s function above is through a DMRG-like variational method that was introduced in Ref. [83] and then improved in Ref. [84]. The MPS-based method uses the MPS wavefunction form, and performs DMRG-like sweeps to evaluate the Green’s function. This means that to use this approach as a Green’s function subroutine in the resolvent method to assess the quality of a candidate wavefunction, the state must first be transformed into MPS form. (Other methods for computing the Green’s function [52,53,59,82] could work directly with the SOS form.)

The MPS-based method of Ref. [84] works as follows: we define the state  $|\varphi\rangle$  that satisfies

$$\begin{aligned} |\varphi\rangle &= \frac{1}{\pi} \frac{1}{(H - E + i\eta)} |\psi\rangle, \\ \Rightarrow \pi(H - E + i\eta) |\varphi\rangle &= |\psi\rangle. \end{aligned} \quad (36)$$

The Green’s function can be written as the overlap  $G(E, \eta) = \langle \psi | \varphi \rangle$ . Now, defining  $|Y\rangle$  as [83]

$$[(H - E)^2 + \eta^2] |Y\rangle = -\frac{\eta}{\pi} |\psi\rangle. \quad (37)$$

From this equation, we see that  $\text{Im } G = \langle \psi | Y \rangle$ . Finding the overlap of the above equation with  $|Y\rangle$ , a DMRG-like algorithm is used to minimize the resulting functional [83,84]:

$$\langle Y | [(H - E)^2 + \eta^2] | Y \rangle + \frac{\eta}{\pi} \langle Y | \psi \rangle. \quad (38)$$

We rely on the package Block2 [85–87] to carry out the calculation of the Green’s function in Eq. (33). Even though the method is implemented in the MPS language, it is general because we have methods for transforming other forms of states into MPS (Appendix B). As discussed in the previous section, we expect the cost of this MPS-based method to scale at most polynomially.

## 3. Coarse QPE

Finally, we describe how using low-precision QPE, which we refer to as *coarse* QPE, can be used to build approximate energy distributions. QPE samples energies from the energy distribution of a given state, and this means that the sampled values can be used to reconstruct the underlying distribution if a sufficient number of them are obtained. We note that each of the energy levels is

broadened according to the QPE kernel:

$$f(E) = \frac{1}{2^{2k}} \frac{\sin^2(\pi 2^k E)}{\sin^2(\pi E)}, \quad (39)$$

in Eq. (25), for some integer  $k$  that is the number of digits in the coarse QPE measurement.

One possible way to obtain an energy distribution is through performing QPE multiple times and obtaining the statistics of the outcomes and then trying to reconstruct the energy distribution from that. As a refinement of this process and to diminish bias, for each QPE round, we add a random constant to the Hamiltonian. The constant  $c$  is chosen to lie in the interval  $[0, 2^{-k})$ . After the measurement is performed and an integer  $x_m$  is observed, we take the measured energy to have the value  $2^{-k}x_m - c$ . This ensures that all real values can be sampled, making it possible to approximate the energy distribution using smoothing methods such as kernel density estimation [88]. As is well known, in kernel density estimation with a suitable choice of the broadening factor, the error scales as  $M^{-0.8}$ , where  $M$  is the number of samples (see Appendix G for more details). There exist more advanced modern methods beyond kernel density estimation [89], and in practice they would likely offer even better performance. However, here we have focused on establishing a baseline with kernel density estimation, as it is a mature and well-understood technique. Finally, we note that the coarse QPE method can work with any kernel, not only the Lorentzian one, unlike

the resolvent method. The use of other kernels is discussed in more detail in Ref. [90].

#### 4. Numerical example

Here, we consider a hydrogen chain with six hydrogen atoms at a bond length of 5 a.u. (see Sec. VI B for more details on this class of systems). We calculate the exact spectrum and also find the energy distribution of an initial state of our choice using the above method. The state that is used has a form as a sum of Slater determinant as the following:

$$|\psi_0\rangle = 0.86 |2, 2, 2, 0, 0, 0\rangle - 0.36 (|\beta, 2, \alpha, \alpha, 0, \beta\rangle + |\alpha, 2, \beta, \beta, 0, \alpha\rangle), \quad (40)$$

in the basis of Hartree-Fock orbitals. The coefficients are directly taken from the corresponding terms in the exact ground state but the state is normalized.

The above three methods are used and energy distributions obtained are shown in Fig. 6 (see the caption for the details of implementation of the methods). All three methods for this particular example show that some useful information can be obtained. But among the two classical methods we see that the resolvent method is more reliable, especially because the series expansions can show uncontrolled oscillations for larger orders. However, we note that the series expansion method could be a measure of last resort for certain especially large problems due to its

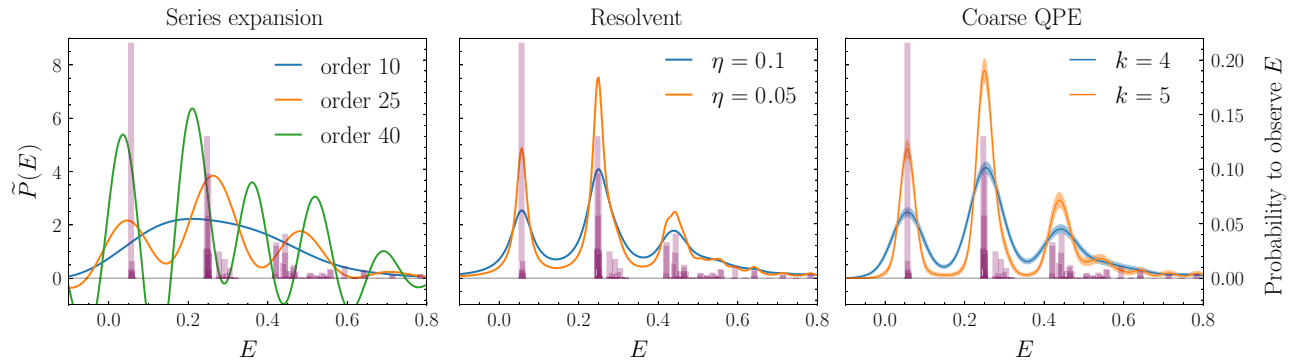


FIG. 6. Energy distribution of an initial state, which is the sum of three Slater determinants for the Hamiltonian of a hydrogen chain with six atoms within the STO-6G basis set. The bonds are uniformly stretched relative to equilibrium, to have the value of 5 a.u. The exact overlap of the state with all of the eigenstates of the Hamiltonian is also calculated for reference. For each panel, two vertical axes are used: the left vertical axis corresponds to the energy distribution and the right one corresponds to the probability ( $|\text{overlap}|^2$ ) of each Hamiltonian eigenstate in the expansion of the initial state. The horizontal axis energy is obtained from the actual energy in Hartree, by rescaling by a factor of  $1/3$  and then translating by the value of  $+1$ , so that it lies between 0 and 1. Left: energy distributions obtained through the Gram-Charlier series. As the order is increased, more features are captured but also for higher orders unphysical oscillations also start to occur. Middle: energy distributions obtained through the resolvent method for two  $\eta$  values. A bond dimension of 100 is used for the calculation of all the points of the two curves. For this particular system, the exact results with these  $\eta$  values closely match these curves. Right: energy distributions obtained through the coarse QPE approach for four and five QPE qubits. A total of 100 different realizations of the 50-measurement runs for both of the cases were considered and the mean and standard deviation were calculated. Kernel density estimation with a Gaussian kernel was also used for smoothing the curve with the broadening factor of  $2^{-k}$ . The resulting mean and error are shown as the solid curve and the shaded area around it, respectively.

relative lower cost compared to the other methods. Even though the results from it are approximate and might show oscillations and nonconverging asymptotic behavior, they would still yield useful partial data to inform the choice of the initial state.

#### D. Using the energy distribution for estimating lowest-energy outcomes

Assuming we have access to the approximate energy distribution of the initial state, we can use it to approximate the distribution of QPE outcomes. For a number of  $k$  digits and an integer outcome  $x_m$ , it can be calculated as follows:

$$\tilde{P}(x_m) = \int dE P(E) \frac{1}{2^{2k}} \left( \frac{\sin^2(\pi 2^k E)}{\sin^2(\pi [E - x_m/2^k])} \right). \quad (41)$$

Note that this is a discrete distribution function. We define the variable  $E_o = 2^{-k}x_m$  with the distribution function  $P(E_o) = \tilde{P}(x_m)/2^{-k}$ . In the limit of  $k \rightarrow \infty$ , this distribution approaches the underlying energy distribution and henceforth we denote this by  $P(E)$  for notational simplicity.

In the following, we formalize how the left tail of the energy distribution can be related to the best outcome obtained from QPE within a given number of runs. To this end, we study the statistics of the best energy achievable through repeated QPE measurements: we will see how the statistics depends on the number of measurements and thus the total budget for repeating the measurements.

Referring to  $K$  QPE outcomes as  $E^{(1)}, \dots, E^{(K)}$ , we focus on the distribution of the smallest observed energy,  $E_{\min,K} = \min(E^{(1)}, \dots, E^{(K)})$ . We may calculate the cumulative distribution function (CDF) of this variable at energy  $E$  as follows. First, note that the probability for any one measurement  $E^{(l)}$  from the set to be below energy  $E$ , denoted  $p_{<}(E)$ , is the integral of the distribution up to that energy

$$p_{<}(E) = \int_{-\infty}^E dE' P(E'). \quad (42)$$

Then the probability for a measurement  $E^{(l)}$  to be *above* energy  $E$  is  $1 - p_{<}(E)$ . For all  $K$  measurements to be above  $E$ , the probability is multiplicative, namely  $(1 - p_{<}(E))^K$ . Finally, the probability that *at least one* of the measurements  $E^{(l)}$  is below  $E$  is 1 minus the probability of all the measurements being above  $E$ , namely

$$C_{\min,K}(E) = 1 - (1 - p_{<}(E))^K. \quad (43)$$

This is the probability that at least one outcome from  $K$  rounds of QPE lies below  $E$ . Upon differentiating the CDF

with respect to  $E$ , we obtain the probability distribution function of  $E_{\min,K}$ , which reads

$$P_K(E) = KP(E) (1 - p_{<}(E))^{K-1}. \quad (44)$$

One simple measure of state quality is the mean value of this distribution:  $\int dE P_K(E)E$ .

There are two useful ways of applying this metric. First, we can use it to directly estimate how many times we need to run QPE to get an improvement over a given classical estimate, similarly to how we use the ground-state overlap to estimate the number of runs for sampling the ground state. In this sense, the energy distribution is complementary to the ground-state overlap metric. And secondly and more importantly, this weight can be compared between different initial states, to determine which one has a better chance of yielding improvements when used in QPE. This forms an essential part of our state-preparation pipeline (step 3), where we use the energy distribution to adjudicate between different possible initial states on the basis of the low-energy cumulative weight  $\int dE P_K(E)E$ .

### V. QUANTUM REFINING

After the classically optimized ansatz state has been analyzed with the energy distribution methods of Sec. IV and is implemented on the quantum computer with the methods of Sec. III, there is a possibility for further quality improvement by using a quantum algorithm to filter out some of the remaining high-energy weights. This is beneficial only if a cheap quantum refining procedure is possible. We show that this is indeed the case in this section. Also, we show that another subtlety with QPE, namely the leakage problem, can be addressed through the quantum refining process.

We focus on two main methods of filtering: coarse QPE [91] and quantum eigenvalue transformation of unitary matrices with real polynomials (QETU) [92]. The latter is chosen as a representative of the polynomial-based algorithms' family [15,17,92].

#### A. Coarse QPE with postselection

The idea behind coarse QPE filtering is to use QPE's ability to project an initial state with broad support across the eigenspectrum onto an individual eigenstate. Using QPE can already be viewed as a filtering procedure: one feeds in a state  $|\psi\rangle = \sum_n c_n |E_n\rangle$  with an energy distribution  $P_\psi(E) = \sum_n |c_n|^2 f_\eta(E_n - E)$ , and after QPE, with probability  $|c_n|^2$  one is left, roughly speaking, with an individual eigenstate  $|E_l\rangle$  in the system register and a state representing  $E_l$  (in binary form) in the ancilla register. Combining this with postselection, i.e., only continuing with the rest of the algorithm when the state we get from QPE filtering has our desired energy  $E_l$  and discarding and restarting otherwise, means we have a pipeline to reliably

generate  $|E_l\rangle$ . Thus we essentially filtered the initial state, which had broad support across many eigenstates, into a state with support only on  $|E_l\rangle$ .

The only complication is if the probability  $|c_n|^2$  to prepare a state with support in the desired energy range is very small, it will require running QPE many times before we encounter the desired state. This is where the “coarse” part of coarse QPE filtering comes in. If we consider an implementation of QPE with less digits of precision than are ultimately needed (for example, to accuracy of 50 mHa rather than the 1 mHa of chemical accuracy), this has the effect of binning neighboring eigenstates, thus boosting the probability of landing within a desired energy range  $[E_l - \Delta_E, E_l + \Delta_E]$  while simultaneously being cheap to execute as compared to the high-precision (1 mHa) QPE. In each coarse QPE measurement, if the outcome lies outside a set of predetermined low values, the state is discarded and the algorithm is restarted. The assessment of what QPE outcomes are considered small can be based on the energy distribution of the implemented state. Note that this method is based on the fact that after postselecting a particular QPE outcome, weights of different energy levels are suppressed based on the energy difference between a level and the postselected outcome; the larger the energy difference is, the higher the suppression (see below for more details). This means that upon postselecting for low-energy outcomes, the resulting state will have suppressed weight along higher energies of the spectrum.

To see to what extent large energies are suppressed after postselecting on low-energy QPE outcomes, consider a setting in which we postselect an outcome  $x_m$  when QPE with  $k$  digits is performed on a state  $|\psi\rangle = \sum_n c_n |E_n\rangle$ . The probability for each component  $E_n$  in the resulting state will be

$$P(n) = |c_n|^2 \frac{1}{2^{2k}} \left( \frac{\sin^2(\pi 2^k E_n)}{\sin^2(\pi [E_n - x_m/2^k])} \right). \quad (45)$$

If the standard deviation of the initial-state energy distribution is small compared to the span of the spectrum of the Hamiltonian, we can approximate the denominator in the above factor by a Taylor expansion. The weight after measurement then is suppressed as the inverse square distance from the measurement outcome: approximately  $|2^k E - x_m|^{-2}$ .

This shows that if the precision of the coarse QPE and the postselection values are chosen appropriately, the high-energy weight of the distribution can be well suppressed.

## B. QETU filtering

QETU [92] or other polynomial-based methods [15,17] aim to implement a function of the Hamiltonian that retains low energies and filters high energies. In short, the method consists of a quantum signal processing circuit [93,94] that

implements a unitary matrix that block encodes a function  $f(H) = P(\cos(H/2))$ , where  $H$  is the Hamiltonian of interest and  $P$  is an even polynomial of degree  $d_P$ . We need  $f(H)$  to be designed in a way so that low energies are retained and high energies are filtered, for instance, using an approximate step function. Details of how this can be done are discussed in Appendix H.

The cost of implementation is directly given by the degree of  $P$ : more precisely, the number of times that one queries the unitary  $U = e^{-iH}$  is exactly  $d_P$ . In order for the filtering to be successful, this degree should have a scaling  $O(\Gamma^{-1} \log \epsilon^{-1})$ , where the error  $\epsilon$  in the polynomial approximation is small enough, and  $\Gamma$  is the energy scale over which the transition in the function  $f(H)$  needs to occur. Apart from the above asymptotic scaling, in practice, we choose the degree by examining how good of a filtering function is achieved.

## C. Cost of implementing quantum refining methods

A simple analysis of the asymptotic cost of the above methods can be done by the following consideration: to suppress a high degree of weight at unwanted high energies, and to keep low-energy weights mostly intact, we need to differentiate energies separated by values of the order of the standard deviation  $\sigma_E$  of the energy distribution of the initial state.

This means that in a coarse QPE setting, we need the resolution  $2^{-k}$  to be of the order of  $\sigma_E$ ; on the other hand, in a QETU setting, again we need a function  $f(H)$  that discerns energies of the order of  $\sigma_E$ , i.e.,  $\Gamma \sim \sigma_E$ . Thus, both of the methods require a number of queries to  $e^{-iH}$  (or related unitary) that scales as  $O(1/\sigma_E)$ . We expect the target precision  $\epsilon$  to be considerably smaller than  $\sigma_E$ , so the cost of filtering should, in principle, be much lower than for the final energy estimation algorithm, which would have a scaling  $O(1/\epsilon)$ . This shows that quantum refining for further improvement of the state quality is viable through these methods.

While the asymptotic costs are the same, by examining simple concrete examples, we find that coarse QPE works appreciably more efficiently. In particular, we consider the case of a Gaussian energy distribution for the initial state, and study the effect of quantum refining via coarse QPE and QETU on it; even though this can be an artificial construction, we can capture the essential factors of comparing polynomial-based algorithms with coarse QPE in it. Plots of energy distribution after quantum refining are presented in Fig. 7, with coarse QPE on the left panel and QETU on the right. We can see that the ultimate state obtained through coarse QPE has a much higher quality when compared with the one obtained through QETU. More details on the processes are presented in Appendix I. There, we show that this performance of coarse QPE is achieved even with a lower cost compared with QETU.



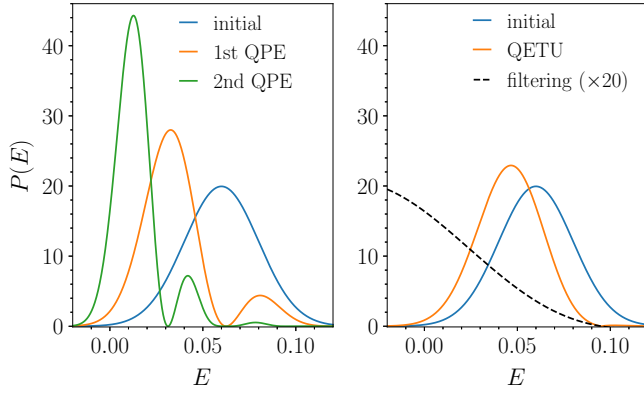


FIG. 7. An initial state with a Gaussian energy distribution is subject to quantum refining. The energy is expressed in dimensionless units. Left: two rounds of coarse QPE are performed and the resulting energy distributions are plotted. Right: QETU with a steplike filtering function is performed and the resulting distribution along with the filtering function are depicted. More details of this procedure is presented in Appendix I.

Given this example and similar other constructions, we come to the conclusion that QPE has generally a lower cost in practice. This is mostly due to the fact that constructing polynomials with sharp jumps for the QETU algorithm requires high degrees, resulting in high costs. Thus we pick coarse QPE as our method of choice for performing the quantum refining stage of the algorithm.

We note here that there are other methods that discussed a similar quantum refining of the state [95,96]. Even though we have not considered them here, it would be interesting to compare their performance with the methods discussed above in a future work.

#### D. Case study: the QPE leakage problem

In this section, we show how quantum refining can be used to prevent a problem that often arises when applying QPE in practice—a problem we call the QPE *leakage* problem. The consequence of QPE leakage is a significant increase in the cost scaling of QPE with the ground-state overlap  $p_0 = |\langle E_0 | \psi \rangle|^2$ , going from  $1/p_0$  to a  $1/p_0^2$  (see e.g., Sec. I.A of Ref. [16] and Appendix A of Ref. [15]).

At a high level, the QPE leakage problem arises because the QPE kernel—the distribution of energies sampled by QPE for a given eigenstate—is not a Dirac  $\delta$  but rather the  $\text{sinc}(x) = \sin(x)/x$  function (see Fig. 8). If an initial state has overlap with several eigenstates in the spectrum, the long tails of the sinc kernel can overlap, falsely enhancing the probability of observing a given eigenvalue—the probability *leaks* from higher energies, potentially even returning false results that lie *below* the true ground-state value.

It is argued in Ref. [15] that in order to prevent this from happening, longer evolution times that scale as  $(1/p_0)$  in

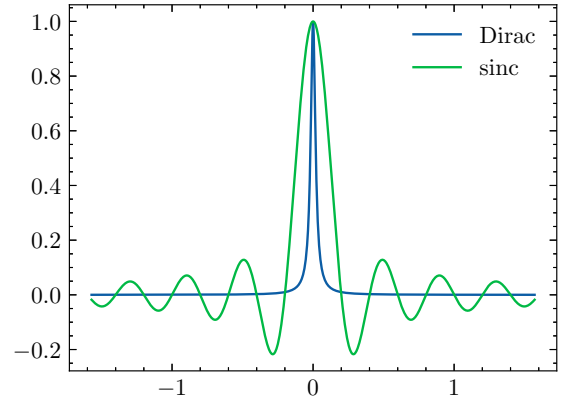


FIG. 8. The sinc kernel of QPE, as compared to (a finite approximation to) the Dirac  $\delta$  kernel.

each QPE round should be used. Since a total of  $(1/p_0)$  rounds are required to obtain precisely the ground state, the overall cost in this argument scales as  $(1/p_0^2)$ . Here we analyze this problem based on our energy distribution approach and discuss how it can be diagnosed. We also show that the problem can be circumvented using quantum refining of the initial state, without a need to resort to large evolution time: the idea is that if refining removes high-energy weights in the energy distribution, the long tails associated with those weights are also removed and the QPE leakage is significantly reduced.

We first consider the problem in the conventional setting: starting with an initial state  $|\psi\rangle = \sum_n c_n |E_n\rangle$ , we would like to perform QPE and estimate the ground-state energy  $E_0$  with a tolerated error  $\epsilon$ . The question is how many phase digits are required for this task. One requirement is to ensure that leakage is absent, meaning we need to estimate the probability of contributing an outcome below  $E_0 - \epsilon$  from all of the energy levels except the ground state:

$$p_{\text{leak}} = \sum_{n \neq 0} \sum_{x_j < x_{\text{upper}}} \frac{1}{2^{2k}} \left( \frac{\sin^2(\pi \delta_n)}{\sin^2\left(\frac{\pi}{2^k} [x_n + \delta_n - x_j]\right)} \right), \quad (46)$$

where  $x_{\text{upper}} = \lceil 2^k(E_0 - \epsilon) \rceil$ ,  $2^k E_n = x_n + \delta_n$  with  $x_n$  an integer satisfying  $0 \leq x_n < 2^k$ , and  $0 \leq \delta_n < 1$ . For the leakage to be improbable,  $p_{\text{leak}}$  should be small enough when compared with probability of the ground state in the initial-state composition  $p_0 = |c_0|^2$ . Note that for the ground state itself to not leak beyond the threshold one can add an  $O(1)$  number of more qubits to the phase register and discard their outcome [11].

Up to an additive error of  $O(\max[2^{-2k}, (x_n - x_{\text{upper}})^{-2}])$ , the single-level leakage probability in Eq. (46) can be

written as

$$p_{\text{leak}}(E_n) = \frac{\sin^2(\pi \delta_n)}{\pi^2} \frac{1}{x_n - x_{\text{upper}} + \delta_n}. \quad (47)$$

See Appendix J for more details. Having access to the energy distribution of the initial state, the total leakage probability in Eq. (46) can be approximated as

$$p_{\text{leak}} = \frac{1}{\pi^2 2^k} \int_{E_0 + \epsilon} dE P(E) \frac{\sin^2(\pi 2^k E)}{E - \frac{x_{\text{upper}}}{2^k}}. \quad (48)$$

A simple criterion for leakage, based on the above integral, can be derived for multimodal distributions; we focus on a unimodal distribution and the multimodal case is similar. Assuming an  $O(1)$  probability is concentrated close to the main peak of the distribution, and that this peak is located at  $E_p$ , from the above integral the probability of leakage beyond the ground state can be approximated as  $(1/(2\pi^2 2^k))(1/(E_p - E_0))$ . We need this probability to be smaller than the probability of the ground state: this means that the number of phase digits should be chosen large enough so that  $2^k = O\left([p_0(E_p - E_0)]^{-1}\right)$ . In general, we need to take the tolerated error  $\epsilon$  in QPE also into account for  $k$ , and as a result we have

$$2^k = O\left(\max\left([p_0(E_p - E_0)]^{-1}, \epsilon^{-1}\right)\right). \quad (49)$$

Apart from the above setting, where higher-energy states can contribute to QPE outcomes below the ground-state energy value, there is also a possibility for leakage when we are not aiming to necessarily obtain the ground-state energy but striving to obtain better energy estimates using QPE. In such a setting, if a small QPE outcome is obtained in an energy region where there is actually not an appreciable weight, it is more probable for the outcome to be invalid as it likely happened due to leakage from higher energies. Such an outcome should not be accepted as an estimate of the energy since it can, in general, be smaller than all the eigenvalues of  $H$  with which the initial state has non-negligible overlap: it can actually be below the ground-state energy of the system resulting in incorrect estimates.

To quantify such a possibility, we first note that the distribution of the lowest outcomes of QPE was considered in Eq. (41) through the use of the energy CDF: here, we use the CDF of QPE outcomes to study the possibility of leakage too. In particular, at an energy of interest, we can compare the CDFs of energy and QPE outcomes (with the desired number of digits  $k$ ). If the QPE outcomes of CDF is considerably larger than the energy CDF, this signifies a high probability of leakage contamination of results around or below that energy value. This is especially important for the region in which the energy CDF is of the order  $1/N$

for a QPE measurement with  $N$  repetitions as this is where the smallest outcome is expected to appear. An ultraprecise QPE will result in an energy close to this region but if the QPE outcomes of CDF is large, *lower precision* QPE can contribute smaller outcomes, and those can only be due to unwanted leakage from higher energies and thus should be avoided.

Given these two treatments of the possibility of leakage, we see that knowledge of the energy distribution function enables us to identify situations in which QPE leakage is not insignificant. If leakage is present, in principle, in both of the cases, with choosing QPE precision high enough, the leakage probability is managed; however, this induces extra cost for each QPE round, as with increasing the number of QPE digits  $k$ , the cost rises exponentially  $2^k$ . In the following, we show that quantum refining of the initial state can be used to manage the leakage possibility without a need to use higher precision.

### 1. Mitigating the leakage probability

The above analysis shows that if the spectral weight in some region, that is responsible for the leakage, is suppressed through some means, the leakage probability could also be suppressed with it. We have seen in Sec. VC that for a small cost compared with that of the most precise QPE measurement, high-energy weights can be filtered with quantum refining. Thus, QPE leakage can, in principle, be mitigated through quantum refining, but it is a matter to be studied on a case-by-case basis based on the details of the energy distribution at hand. This will be very helpful as it removes the need to perform time evolution for times of the order  $p_0^{-1}$  for the ultimate QPE measurement.

As an example, consider an energy distribution having a multimodal structure, one can identify the peaks—accumulating  $O(1)$  spectral weight in their vicinities—which are responsible for leakage through analyses discussed above; one can then perform a quantum refining in the form of coarse QPE, so that those peaks are close to discarded outcomes, and thus will lose a substantial weight after the process. This can substantially decrease the leakage probability. Such a procedure is illustrated in a concrete example in Appendix I (last paragraph).

Note that this means the quantum refining step of our state-preparation algorithm is capable of reducing the cost of the whole algorithm, not only by lowering the number of required repetitions of the most precise QPE measurement, but also by decreasing the cost of each single round by mitigating the leakage problem when it is present.

## VI. NUMERICAL DEMONSTRATIONS

In this section, we showcase our complete initial state preparation algorithm for a variety of molecules. Through the numerical examples, we explore how viable it is to prepare good-enough initial states for complex molecules to

be studied with a quantum algorithm. Of particular interest are situations where, on the one hand, classical methods struggle to give a good energy estimate, but at the same time one can still prepare a good-enough initial state for quantum energy estimation. We refer to such problems as Goldilocks problems, a concept we formalize through the energy distribution. All the numerical results shown in figures in subsequent sections are available in Tables III–V in Appendix K.

### A. Goldilocks molecules

As we argued in Sec. IV, the energy distribution picture can be used to estimate whether a quantum algorithm can improve a given initial state’s classical energy estimate. Using this concept, we can categorize energy estimation *problems* based on the hardness of preparing a good enough initial state for performing quantum energy estimation routines (e.g., QPE). Such a classification is of course tied to the available budget for initial state preparation: here we take the budget to be unspecified for the sake of generality. With a given budget for state preparation, one can have easy, intermediate, and hard problems: an easy problem is one in which a very high-quality initial state with large accumulated weight over low-energy parts of the spectrum of the Hamiltonian is possible to prepare on a quantum computer; on the other hand, a hard problem will be one in which with a given budget, it is only possible to prepare a poor-quality state with negligible weight over low energies of the Hamiltonian. Note that these depend on the given budget for state preparation. In between, there ought to be problems of intermediate hardness, where there is some non-negligible, but also not too large weight over low-energy parts of the spectrum.

We argue that in all likelihood, it is only *possible* to obtain quantum advantage in quantum ground energy estimation over classical computational methods in *intermediate* problems. This is because for hard problems, by definition one cannot perform quantum routines effectively; and for easy problems, it is very likely to find a good

classical energy estimate, which is highly challenging to beat using quantum algorithms. The question of whether there is quantum advantage in an intermediate problem remains open: the energy distribution allows this question to be explored computationally.

To make this more concrete, on an energy axis, let us mark the best classically achieved energy by  $E_T$ . Assuming we have access to the energy distribution of our initial state, we can calculate its accumulated weight for energies below  $E_T$  and call it  $p_{<E_T}$ . If  $p_{<E_T}$  is large enough—that is, large enough so that performing an accurate QPE measurement a number  $O(p_{<E_T}^{-1})$  times is within the available QPE budget—there is the possibility for quantum advantage. We call such a situation a *Goldilocks problem*—one in which there is a possibility for improving the classical ground energy estimate using QPE. Given our numerical results for the  $\text{Cr}_2$  and  $\text{Fe}_4\text{S}_4$  molecules in the following sections, we expect that complexes including several transition-metal centers, if studied within an appropriate active space, could indeed present such Goldilocks problems.

In the remainder of this section, we will study several different quantum chemical systems numerically and exhibit the ideas and discussions in the previous sections concretely in those systems. For all calculations in this section, we used the publicly available software library Overlapper [70], which implements the HF, CISD, CCSD, and CASCI methods by interfacing to PySCF [97], the HCI method through an interface to Dice [98], and the DMRG method through Block2 [86].

### B. Hydrogen chains

We begin by studying the hydrogen chain model system in the minimal STO-6G basis (Fig. 9), varying the bond length to increase correlations and evaluating the overlaps and energies of different methods relative to the exact solution from full configuration interaction (FCI). Hydrogen chains are a well-established and thoroughly studied benchmark system for electronic structure methods

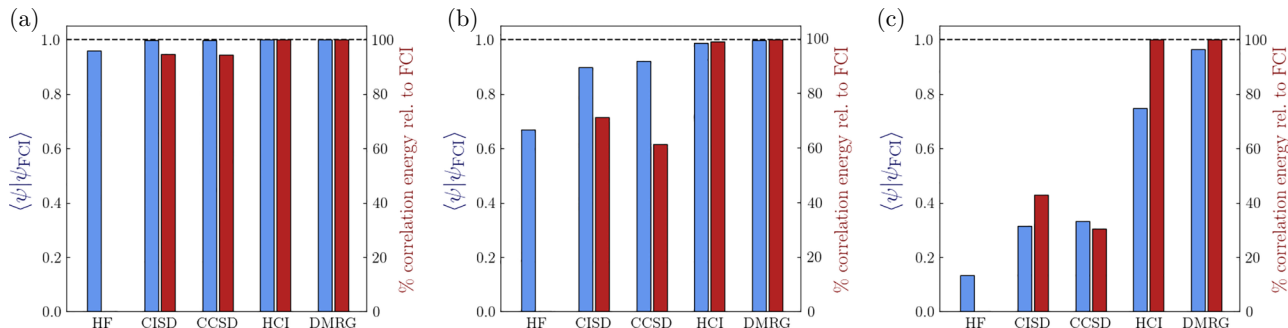


FIG. 9. Percent recovered correlation energy relative to FCI (red) and overlaps  $\langle\psi|\psi_{\text{FCI}}\rangle$  (blue) for the  $\text{H}_{12}$  chain in the STO-6G basis for (uniform) bond length (a) unstretched (0.71 Å); (b) stretched by a factor of two (1.42 Å) (c) stretched by a factor of 4 (2.84 Å). CASCI / MRPT are excluded because they give the exact solution. The final bond dimension used to obtain the DMRG state was  $\chi = 1000$ , and the final HCI cut-off criterion used was  $\epsilon_1 = 10^{-4}$ .

[99–108], with the advantage of allowing one to uniformly scale system size and also vary correlation by adjusting bond length: this makes them good systems for the current analysis. Throughout this section, we will use a uniform H to H bond length of 0.71 Å: we will stretch that bond length to increase the amount of correlation in the system by a scaling factor, mentioned explicitly in the relevant section.

All methods in Fig. 9 including DMRG are executed in the  $S_z$  symmetry mode, i.e., conserving the spin projection on the  $z$  axis. Here and in later figures, the overlap  $\langle \psi | \psi_{\text{FCI}} \rangle$  is computed by first bringing the output of all methods to the SOS form: in all cases (especially DMRG), we make sure that the SOS form of the wavefunction includes enough determinants to capture above 99% of the weight of the original state. Near equilibrium, as in panel (a), all methods perform equally well: but as the bonds get stretched and static correlation increases [panels (b) and (c)], HCI and DMRG clearly emerge as leaders in terms of overlap, while the Hartree-Fock state performs poorly.

We discuss the case of the CCSD ansatz separately. While in practice CCSD is able to recover a substantial portion of the dynamic correlation energy of the system, its energy estimates are not directly linked to the quality of the wavefunction because the energy CCSD computes is nonvariational. To place the CCSD energy estimates on equal footing with the other methods, we instead plot the variational energy of the associated CCSD ansatz that we aim to implement on a quantum computer. The way Overlapper currently prepares the CCSD ansatz and puts it into the SOS form is by expanding the CCSD exponential  $\exp(\hat{T}_1 + \hat{T}_2)$  in a Taylor series and truncating it to second order. We find that such a truncated CCSD wavefunction is only marginally improved compared to CISD, which is to be expected, as the higher-order components are what makes a CCSD ansatz obtain a better energy than CISD in many cases. To go beyond the second order, we have attempted to instead implement the CCSD ansatz via the MPS-MPO route and then convert the MPS to the SOS form, but this was ultimately unsuccessful within the code-implementation constraints, and we leave it to future work to properly benchmark the CCSD ansatz against other wavefunction forms. We note that ultimately the CCSD ansatz is single reference, and so there is a limit, even with the full expansion of the exponential, to the quality of the wavefunction that can be achieved—especially since all higher-order, unlinked contributions like  $\hat{T}_2^2$ ,  $\hat{T}_2\hat{T}_1$  rapidly decrease in magnitude, as their amplitudes approximately  $t_2^2, t_2 \cdot t_1$  are merely powers of  $t_1, t_2$  and are not allowed to vary independently.

Still, despite this truncation, our results for the CCSD wavefunction in intermediate bond-length hydrogen chains [Fig. 9(b)] appear in rough agreement with earlier work [41], where the authors also find that the CCSD wavefunction gives a fairly good initial state (when the

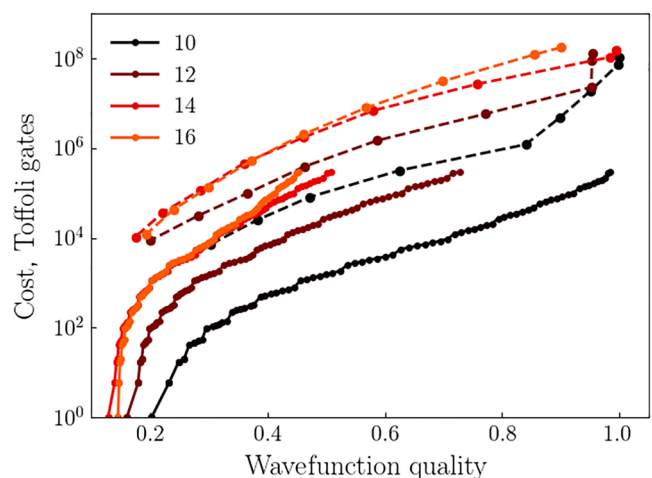


FIG. 10. Comparing the cost of implementation, in terms of Toffoli gates, of the DMRG solution in SOS form (solid lines) and MPS form (dashed lines), for the hydrogen chain with different numbers of atoms (as shown in the legend), as a function of the prepared wavefunction overlap with the ground state. The bond length is taken to be uniform along the chain and has been stretched to 4 times the original length (2.84 Å). As the system gets larger, the MPS form continues to allow the preparation of wavefunctions with high quality. The finite steps in the SOS cost curve are due to rounding of  $\log(D)$  factors to the nearest integer.

CCSD excitation amplitudes are used to instantiate a UCC ansatz). It would be interesting to compare our results to the UCC ansatz for the more extreme bond lengths in Fig. 9(c), as well as for the  $N_2$  molecule (see the following section), a task we leave for future work.

Hydrogen chains are also convenient for comparing the SOS and MPS forms of the initial state in terms of their quality versus implementation Toffoli cost curves. For this comparison, we first obtain the exact system ground state as an MPS with DMRG, then process it in two ways. First, to compute the cost of implementing the SOS form, we do reconstruction on the ground-state MPS we obtained—meaning we unpack the MPS wavefunction into its constituent Slater determinants (see Appendix A and Ref. [109])—and then truncate the number of Slater determinants to generate wavefunctions of varying quality (the  $x$  axis of Fig. 10). Second, to compute the cost of implementing the MPS form of varying quality directly, we compress the MPS to smaller and smaller bond dimensions. This gives a sequence of states with decreasing wavefunction quality: we evaluate the quality by computing overlaps with the original ground-state MPS. For implementation cost, we use the expressions derived in Sec. III: the cost is mainly set by the number of determinants  $D$  for SOS and by the bond dimension  $\chi$  for MPS. The results for hydrogen chains of varying sizes are shown in Fig. 10.



The MPS form appears to be more expensive relative to the SOS form for the cases shown even though the system is one dimensional, which means that MPSs should perform very well in representing the ground state. This signifies the efficiency of the SOS method developed in this work. Notice that DMRG can still be used as the method of choice for the classical ground-state search in such cases, however, it might be beneficial to transform the result into the SOS form and then implement on a quantum computer. Furthermore, we should also note that the SOS cost is seen to increase exponentially with wavefunction quality, especially in larger chains, as more and more determinants are needed to accurately represent the ground state. In fact, the reason why the SOS curve does not reach perfect wavefunction overlap for larger chains is because of the extreme memory requirements for storing all the determinants arising in reconstruction beyond a certain cutoff. Thus there could be advantages to implementing the state in the MPS form in large, strongly correlated systems. This needs to be explored more in future works.

We next study energy distributions for all of the states analyzed in Fig. 9. We compute the distributions using the resolvent method: for all non-MPS-based states, we first convert them to MPS form using subroutines from Overlapper [70]. Near equilibrium, all energy distributions are sharply peaked around the ground state (not shown). However, when bonds are stretched, the DMRG and HCI states have significantly more weight in the lower-energy portion of the spectrum, as seen in Fig. 11. The classical energies of the states are shown with vertical dashed lines of the same color. Given that we are implementing the SOS forms of the states, we report the associated SOS state energy, not the energy reported by the classical computational method.

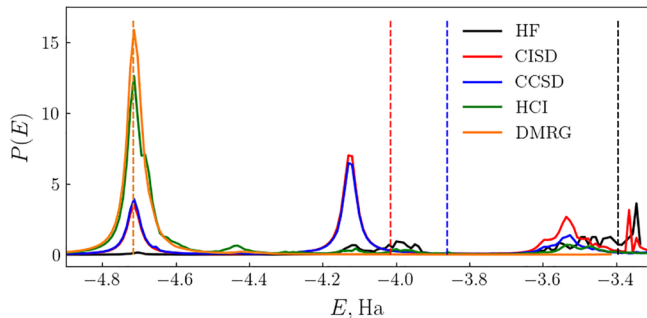


FIG. 11. Energy distributions for the initial states from the methods studied in Fig. 9, for the hydrogen chain with ten atoms with bonds stretched by a factor of 4 (2.84 Å). For CCSD, the energy reported is that of the truncated SOS state rather than the nonvariational energy evaluated by the CCSD method. The calculations are done with the resolvent method with  $\chi_{\text{calc}} = 200$  and  $\eta = 0.02$ . The DMRG and HCI parameters are as in Fig. 9. Here and in all later energy distribution plots, vertical dashed lines give the classical energy of the state of the same color. The DMRG and HCI energies coincide.

This applies most importantly to CCSD: because of this consideration, the CCSD energy shown is actually higher than the CISD energy. While the CCSD state redistributes some of the weight towards lower-energy states relative to CISD, this is marginal. At the same time, the Hartree-Fock state has very little weight in the low-energy parts of the spectrum, including the ground-state peak. Overall, we see that even though these energy distributions are approximate, by direct visual inspection the high-quality states can be identified—without any need for a reference state, as with the overlap metric. Beyond this, we also get a much richer picture of how the weight is distributed across the energy range.

### C. The N<sub>2</sub> molecule

Next we turn to molecules, starting directly with a system with an intermediate degree of correlations—the N<sub>2</sub> molecule with its bond stretched beyond equilibrium ( $r = 2.517$  Å) in the cc-pVDZ basis with the effective core potential ccECP [110], which is used to reduce the number of electrons in the problem. Since this system of 26 orbitals (28 orbitals without ccECP) and 10 electrons (14 electrons with ccECP) is now beyond our capability for FCI, we use a highly converged DMRG wavefunction ( $\chi = 2000$ ) as the reference state, although the HCI wavefunction is also exact with the cutoff  $\epsilon_1 = 10^{-5}$  (the HCI and DMRG solutions are less than a mHa apart). The CASCI and MRPT methods are carried out in an active space of CAS(10e, 12o).

The energy-overlap bar chart in Fig. 12 shows that most methods, while they struggle to generate an excellent wavefunction, still provide good-enough overlap to the reference state. Notice that while MRPT significantly improves the energy estimate, it does not improve the

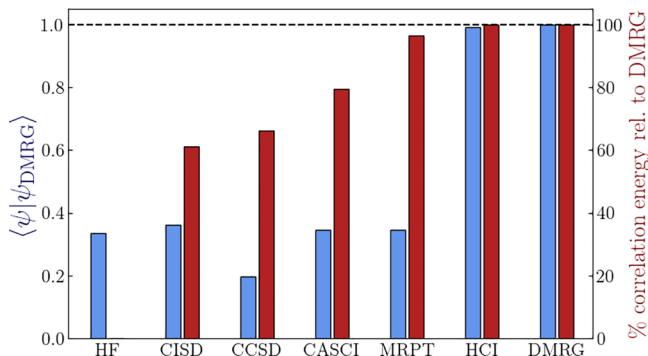


FIG. 12. Percent recovered correlation energy and overlaps for the N<sub>2</sub> molecule in the cc-pVDZ basis using the ccECP effective core potential from Ref. [110] as downloaded from Ref. [111], with its bond stretched beyond equilibrium ( $r = 2.517$  Å), relative to the DMRG solution, which obtained the lowest energy. The CASCI and MRPT methods are performed with the active space CAS(12o,10e), chosen directly from molecular orbitals around the Fermi level.

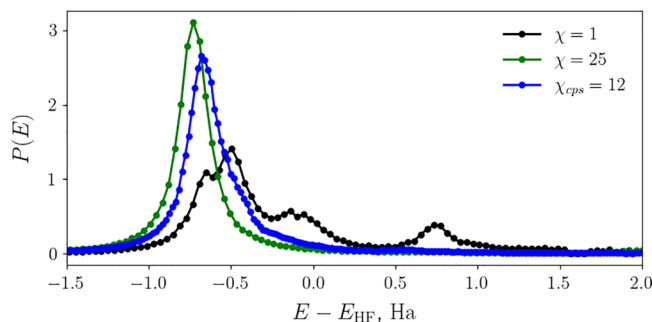


FIG. 13. Energy distributions for DMRG-generated states for the  $N_2$  molecule, with the bond length as in Fig. 12. The calculations are done with the resolvent method with the reduced parameters  $\chi_{\text{calc}} = 50$  and  $\eta = 0.1$  to speed up the calculations.

CASCI wavefunction quality. All of these observations reinforce the idea that energy is not a very reliable proxy for wavefunction quality.

For the energy distributions for this molecule, we choose to compare the states obtained from DMRG (see Fig. 13). Namely, we compare the Hartree-Fock state ( $\chi = 1$ ); an intermediate-quality state with bond dimensions  $\chi = 25$  obtained through an increasing bond-dimension schedule without compression; and finally  $\chi_{\text{cps}} = 12$ , a state, which was obtained initially at  $\chi = 250$  and then subsequently compressed to  $\chi_{\text{cps}} = 12$  before the energy distribution was computed. Once again, at a glance we notice that the Hartree-Fock state is of significantly worse quality than the other two states. Another interesting observation is that going to large bond dimensions and then compressing, as in the  $\chi_{\text{cps}} = 12$  state, gives results nearly as good as the  $\chi = 25$  state but with lower implementation cost. Beyond the details, this example shows that the energy distribution method allows us to characterize and compare the quality of initial states for physical molecules.

#### D. The $\text{Cr}_2$ dimer

We next turn to the  $\text{Cr}_2$  molecule, which is an example of a strongly multireference system. To increase correlations even further, we stretch the dimer bond length from its approximate equilibrium length [66] to 3.024 Å. To study such a many-electron molecule with a limited computational budget, we focus on active spaces built around a limited set of orbitals, namely the  $3d$  orbitals in  $\text{Cr}_2$ . We employ the atomic valence active space (AVAS) approach [50], where molecular orbitals with the largest  $d$ -orbital overlap above a given threshold are selected for the active space.

The active space for  $\text{Cr}_2$  focused on  $3d$  orbitals can be as small as ten active electrons in ten active orbitals, written CAS(10e,10o). In an active space this small, a reference FCI solution can be obtained. For concreteness we specifically targeted the total spin zero sector, searching

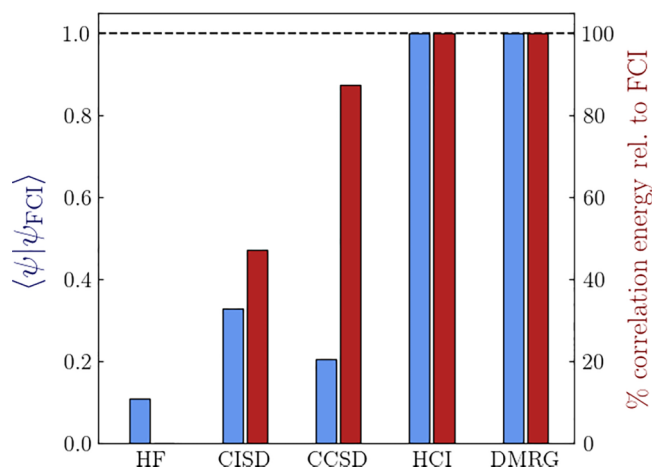


FIG. 14. Percent recovered correlation energy and overlaps for the  $\text{Cr}_2$  molecule with its approximate equilibrium bond length of 1.68 Å [66] stretched to 3.024 Å, in the cc-pVDZ basis in an active space of only  $d$ -type orbitals, relative to the FCI solution in the  $S^2 = S_z = 0$  spin sector. The final DMRG bond dimension used was  $\chi = 1000$ , and the final HCI cutoff used was  $\epsilon_1 = 10^{-7}$ .

for the lowest-energy solution within that subspace. As seen in Fig. 14, both the HCI and DMRG methods recover the entirety of the correlation energy and, more importantly, produce wavefunctions with perfect ground-state overlap, whereas both CISD and CCSD clearly struggle in this multireference situation. Note that the DMRG calculations are now being carried out in SU(2) mode, conserving total spin  $S$ , instead of only the projection  $S_z$ . Similar to before, the CCSD ansatz, as implemented in Overlapper, appears to recover most of the correlation energy whilst having a worse overlap than the CISD solution. Note that the implementation of HCI we had access to is nonspin adapted, meaning it finds wavefunctions that conserve the spin projection  $S_z$  but not the total spin  $S^2$ . It was thus necessary to tighten convergence thresholds and also compute several of the low-energy eigenstates before a solution was found in the total spin zero sector that could be meaningfully compared with the  $S^2 = 0$  solutions identified with the DMRG and CASCI approaches. Spin-adapted selective CI techniques, some of which have been implemented in the literature [52,59,112], could provide additional flexibility in initial state preparation: however, at the moment, there is no open-source implementation publicly available to the community.

For the energy distributions (Fig. 15), we again focus on the DMRG exact solution and its compressed versions with lower bond dimensions. The energy distribution gives an excellent account of the quality of the state during compression: not only are the different quality states easily distinguishable, it is also clear that even the poor-quality, highly compressed  $\chi_{\text{cps}} = 4$  state still has a wide range of energies covered in its distribution, showing weight

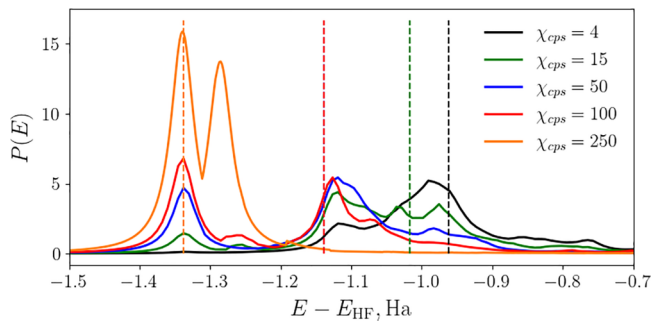


FIG. 15. Energy distributions for the  $\text{Cr}_2$  molecule within an active space, for different states of compression of the DMRG-generated ground state: starting with a well-converged state with  $\chi = 500$ , the state was slowly compressed to  $\chi_{\text{cps}} = 250$  (orange) and then further to  $\chi_{\text{cps}} = 100, 50, 15, 4$  (red, blue, green, and black respectively). As before, the calculator bond dimension used is  $\chi_{\text{calc}} = 200$ , with broadening  $\eta = 0.02$ . While this is a relatively large calculator bond dimension, the energy distribution seen here is still approximate and not fully converged with  $\chi_{\text{calc}}$ . However, even an approximate distribution already allows the evaluation of state quality, whilst having the advantage of being computable with reasonable resources. A key observation here is that compression, while reducing resources necessary for preparation, trades that off against the quality of the initial state—but only partially.

well below its mean (classical) energy. These energy distributions are an illustration of the Goldilocks state concept described in more detail in Sec. VI A: while the  $\chi = 4$  state is clearly too poor to do QPE with, and the  $\chi = 250$  state already has a good classical energy, the intermediate quality states with  $\chi = 50 - 100$  would allow QPE to improve on their associated best classical estimate with only a small number of iterations. Any molecule where states with quality like the  $\chi = 50 - 100$  above are the best we could get would likely represent a Goldilocks case. Notice also that while the quality of the  $\chi = 100$  state is greater than that of the  $\chi = 50$  state, their energies are nearly equal: this shows that it is possible to find initial states with improved quality without this corresponding to an improvement in the classical energy estimate, once again emphasizing that energy would be a poor metric to judge initial state quality.

### E. $[\text{Fe}_4\text{S}_4]$ core

Finally, we consider the  $[\text{Fe}_4\text{S}_4]$  molecule core—an 8-atom center extracted from the associated molecular iron-sulfur complex. The active space for this system is built around the Fe  $3d$  and S  $3p$  orbitals: we follow the procedure outlined in Refs. [25,64]. Instead of Pipek-Mezey, we use the Cholesky method to split-localize the  $\alpha$  molecular orbitals from a high-spin ( $S_z = 5/2$  per Fe atom) restricted open-shell Hartree-Fock calculation in the cc-pVDZ basis, then select orbitals with Fe  $3d$  and S  $3p$  character by visual inspection.

For this system, we consider four different states obtained with DMRG: three states obtained by converging DMRG at bond dimensions  $\chi = 20, 50, 100$ , respectively, and another obtained by converging a calculation at a high bond dimension of  $\chi = 1000$  and then compressing that wavefunction down to  $\chi_{\text{cps}} = 7$ . While the highly compressed state is the most high-quality of the four, it has the worst classical energy estimate—higher-energy admixtures balance out the weight at the lower energies. On the other hand, while the  $\chi = 50$  state has a significantly improved energy relative to  $\chi = 20$ , its energy distribution is mostly unchanged, and its quality not significantly improved. Finally, the states  $\chi_{\text{cps}} = 7$  and  $\chi = 20$  have nearly the same energies, but the former has much more weight in the low-energy part of the spectrum, making it much higher quality. At the same time, all these states for the  $\text{Fe}_4\text{S}_4$  core are built up from tens or hundreds of Slater determinants, with the coefficients of the largest contributing determinants being on the order of  $10^{-2}$ . This strongly suggests a single product state would be an exceedingly poor initial state in this situation. More specifically, the  $\chi_{\text{cps}} = 7$  state has an overlap of 0.69 with the ground state (obtained at  $\chi = 1000$ ), while the largest contributing determinant in the  $\chi = 1000$  state has weight 0.04. Then in terms of probabilities to project on the low-energy  $\chi = 1000$  state, the  $\chi_{\text{cps}} = 7$  state gives 0.47 while the best single product state gives 0.002. Thus at least 300 times fewer iterations of QPE will be needed to project on this state with the only marginally more complicated initial state—which translates into direct cost and runtime savings of more than 2 orders of magnitude. At the same time, the compression to  $\chi_{\text{cps}} = 7$  ensures that the cost of implementing the improved initial state continues to be negligible compared to the main costs of energy estimation [8].

The main conclusion to be drawn from Fig. 16 is that even though simple product states have low overlap with the low-energy subspace, it is generally possible to prepare a relatively cheap, relatively high-quality initial state (e.g., the  $\chi_{\text{cps}} = 7$  state) for this system with only moderate

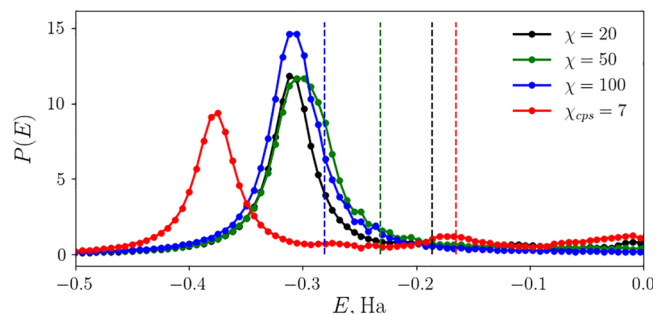


FIG. 16. Energy distributions for MPS states at different bond dimensions for the  $\text{Fe}_4\text{S}_4$  core in the active space. Energy appears to be a misleading guide for state quality.

additional effort. While this particular state is generated from an expensive classical calculation, it retains a considerable weight in the low-energy sectors postcompression, further suggesting that an expensive classical calculation followed by compression could be a good method to obtain a *cheap*-to-implement high-quality state.

These last two example molecules together suggest that preparing an inexpensive, good-quality state is possible for molecules with transition-metal centers. At the same time, such systems are known to be challenging cases for classical computational methods [25,66]. The combination of these facts implies that these systems are good candidates for Goldilocks systems, and motivates their further study for quantum computing applications. This conclusion is uniquely enabled by the energy distribution picture and data: the overlap metric would not give much insight into the relative quality of the states we considered here.

## VII. CONCLUSIONS

We have introduced a complete workflow for preparing initial states for quantum chemistry. Our results target a critical component of quantum algorithms for simulating chemical systems, which is essential to elucidate the potential for quantum advantage. Key technical contributions of this work include a state-of-the-art quantum algorithm for preparing states expressed as sums of Slater determinants, methods to construct approximate energy distributions for assessing state quality, and identification of coarse quantum phase estimation (QPE) as a leading technique for refining initial states and addressing the leakage problem. All calculations were carried out with the publicly available software library Overlapper [70]. We demonstrate the applicability and usefulness of our initial state preparation procedure with several numerical experiments on challenging molecules.

Our work indicates that it is worthwhile to employ advanced techniques for state preparation beyond simplistic approximations like the Hartree-Fock state. Quantum energy estimation algorithms such as QPE already incur a considerable cost, thus in effect leaving a large budget available for spending computational resources to prepare better initial states. This budget should be utilized as much as possible: improved initial states lead to higher probabilities of observing low-energy estimates, resulting in fewer repetitions of the energy estimation algorithm and an overall reduced cost. Our optimized technique for implementing sums of Slater determinants was designed precisely to enable the use of sophisticated approximate ground states such as those obtained from heat-bath configuration interaction (HCI) and the density matrix renormalization group (DMRG) methods, which we identify as leading strategies for initial state preparation.

The energy distribution approach that we propose suggests a rethinking of initial state preparation for quantum

chemistry. It provides a computationally tractable method for assessing and comparing the quality of initial states in a reference-free way: by contrast, this is usually out of reach when computing overlaps with the true ground state, because the latter is typically unknown. Energy distributions also help to shape our understanding of the prospects for quantum advantage: since the goal of QPE is to improve the energy estimates associated to the initial state, we can use approximate energy distributions to reason about the extent to which this is possible. We employ this perspective to formalize the concept of Goldilocks systems: molecules where the quality of the initial state is neither too high nor too low. This means that two conditions are met: (i) the difference between the best classical estimate and the true ground-state energy is large enough to leave room for improvements, and (ii) the quality of the initial state is sufficiently high to support a considerable probability of observing such improvements.

Numerical experiments support these findings. We observe that it is possible in practice to use even approximate, broadened energy distributions to infer quality of different initial states, meaningfully compare them and pick the highest-quality candidate. This is especially valuable in cases where the expectation values of the energy are very similar. The same evidence then also suggests that the average energy of an initial state can be a problematic proxy for state quality. Our studies hint that molecules with transition metals in nonequilibrium geometries are potentially Goldilocks systems, and therefore a quantum advantage in ground-state energy estimation could be possible.

Future work may focus on further optimizing quantum algorithms for implementing classical wavefunctions, and more generally, on further improving the proposed workflow. Of particular interest are quantum algorithms for refining initial states obtained from classical methods, which have not received much direct attention. It is possible that better methods than coarse QPE, equipped also with performance guarantees, could be discovered. Another direction that can be pursued is to extend our methodology to periodic systems. This is needed for simulating materials, which have many industrial use cases. It would also be of interest to explicitly compare the performance of state-preparation methods analyzed here with some of the leading variational contenders such as UCC and ADAPT-VQE. Finally, it is important to understand how to prepare initial states in circumstances where quantum hardware places restrictions in terms of available qubits and circuit depth, in preparation for the emergence of early fault-tolerant quantum computers.

## ACKNOWLEDGMENTS

We thank Huanchen Zhai, Yu Tong, and Soran Jahangiri for fruitful discussions. We gratefully acknowledge the



support and computational resources provided by the BC DRI Group through the Cedar supercomputing cluster and the Digital Research Alliance of Canada. We are also grateful for the use of computational resources of the LISA computational cluster at the Stewart Blusson Quantum Matter Institute. S.F. acknowledges support by Mitacs through the Mitacs Accelerate Program. J.F. acknowledges support from ERC AdG NOQIA; Ministro de Ciencia y Innovacion AEI (Plan Nacional FIDEUA PID2019-106901GB-I00, Plan Nacional STAMEENA) and Fundació Cellex. This work was also partly supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (Grants No. NRF-2021M3H3A1038085, No. NRF-2022M3H3A106307411, No. NRF-2023M3K5A1094805, and No. NRF-2023M3K5A109481311) and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (Grant No. 2019-0-00003, Research and Development of Core technologies for Programming, Running, Implementing and Validating of Fault-Tolerant Quantum Computing System). J.H. acknowledges Xanadu for hosting his sabbatical year visit.

#### APPENDIX A: CONVERSION TO SUM OF SLATER FORMATS FOR ALL WAVEFUNCTION METHODS

Converting all wavefunction-based methods explored in this paper to a sum of Slater format requires a number of specialized steps particular to each method.

The CISD wavefunction already comes in the sum of Slater format, so no conversion is required.

The CCSD ansatz is more challenging to convert to the unified sum of Slater format due to the fact that, in principle, excitations to all orders are being generated. However, since these decay quickly, in practice, going up to second or fourth order in excitations already captures most of the CCSD wavefunction. These can be obtained by Taylor expanding the exponential to the appropriate order and collecting like terms for excitations: in this way, coupled cluster amplitudes combine in various ways to become CI coefficients.

CASCI wavefunctions merely need to be padded to the full space with the occupied orbitals, which makes the conversion of these wavefunctions to the sum of Slater format almost immediate. The same applies to MRPT wavefunctions.

Being one of our standard formats, the MPS does not require form conversion. However, for the purpose of comparison with the other methods, it is also possible to start from an MPS and compute the equivalent Slater determinant representation of the wavefunction up to a specified tolerance—a process called reconstruction. A

deterministic approach to this involves partial resummation of the matrix products: the details can be found in Ref. [109]. On top of that, to switch to chemist convention of keeping all spin-up operators on the left, we evaluate the required parity conversion factor for each determinant.

The HCI method naturally returns the wavefunction as a sum of Slaters, so little conversion is required beyond postprocessing the results of the particular package we are employing.

#### APPENDIX B: SOS $\leftrightarrow$ MPS TRANSFORMATION

In this Appendix, we discuss how the two standardized formats, i.e., SOS and MPS are transformed to each other.

##### MPS to SOS:

The goal is to calculate the largest coefficients  $c(n_1, \dots, n_N) = \sum_{\alpha_1 \dots \alpha_{N-1}} A_{1; \alpha_1}^{n_1} \dots A_{N; \alpha_{N-1}}^{n_N}$  in a SOS expansion. Based on Appendix A of Ref. [109], we start from a left canonical form and set a threshold for keeping terms in the SOS. Partial coefficients such as  $c_{\alpha_p}^{(p)}(n_1, \dots, n_p) = \sum_{\alpha_1 \dots \alpha_{p-1}} A_{1; \alpha_1}^{n_1} \dots A_{p; \alpha_{p-1} \alpha_p}^{n_p}$  are formed and whenever a norm of the partial coefficient  $\sum_{\alpha_p} |c_{\alpha_p}^{(p)}(n_1, \dots, n_p)|^2$  goes below a threshold, all Slater determinants with the prefix  $(n_1, \dots, n_p)$ , i.e., of the form  $|n_1, \dots, n_{a_p}, \dots\rangle$ , are dropped from the SOS. This way owing to the left canonical form of the MPS, it is ensured that all the terms with coefficients above the threshold are recovered in the SOS.

##### SOS to MPS:

For this task, we start with a bond dimension 1 MPS that corresponds to the largest coefficient Slater determinant in the SOS (could be the Hartree-Fock state or not); we make an auxiliary copy of it also. Using MPOs consisting of a number of  $c^\dagger$  and  $c$  operators, the auxiliary bond dimension 1 MPS is transformed to the Slater determinant with the second largest coefficient. The new auxiliary MPS is added to the main MPS and the procedure goes on until all coefficients are added. Note that the auxiliary MPS remains bond dimension 1, while the bond dimension of the main MPS grows, one can compress the main MPS as more and more terms are added to it.

#### APPENDIX C: PROOF OF LEMMA 1

We prove Lemma 1 by induction. To avoid cluttering, we shall replace  $\tilde{v}_i$  with  $v_i$ . From here onwards, the vectors  $v_i$  have length  $r$ . We recall the statement and notations of the lemma: in Eq. (8), one needs to prove the existence of  $2 \log D - 1$  vectors  $u_j$  of length  $r$ , forming a matrix called  $U$ , that helps to distinguish the  $D$  distinct vectors  $v_i$  of length  $r$  by mapping them to vectors  $b_i$  of length

$2 \log D - 1$ . Notice  $U$  is supposed to be found offline on the classical computer.

We can interpret  $U$  as a linear map acting on each  $v_i$ . We need to find a  $U: \mathbb{F}_2^r \rightarrow \mathbb{F}_2^{2 \log D - 1}$  such that  $U(v_i) \neq U(v_j) \Leftrightarrow v_i - v_j \notin \ker U, \forall i \neq j$ . We will efficiently construct such a linear map with the additional property that  $v_i \notin \ker U, \forall i$  unless  $v_i = \mathbf{0}$ ; this additional property will help us in proving the next inductive step from the induction hypothesis. In summary,  $U$  has to satisfy the following properties:

$$v_i - v_j \notin \ker U, \forall i \neq j, \quad v_i \notin \ker U, \forall i, \text{ unless } v_i = \mathbf{0}. \quad (\text{C1})$$

*Proof.* Since  $\log D - 1 < \log(D) \Rightarrow \dim \mathbb{F}_2^{\log D - 1} = 2^{\log D - 1} < D$ , there are at least  $\log D$  many linearly independent vectors among the  $v_i$ 's. Therefore,  $r \geq \log D$ . First assume  $\log D \leq r \leq 2 \log D - 1$ . Without loss of generality, assume  $v_1, \dots, v_r$  are linearly independent and generate all the  $v_i$ 's. Note that finding these linearly independent generators is an efficient classical algorithm in linear algebra. In this case, we can distinguish the  $v_i$ 's using  $r \leq 2 \log D - 1$  many  $u_j$ 's; we simply choose  $U$  to be the  $r \times r$  identity matrix. This choice yields  $b_i = v_i$ , with length  $r \leq 2 \log D - 1$ , fulfilling the same purpose.

When  $r > 2 \log D - 1$ , we perform induction on  $t \in \mathbb{N}$  where  $r = 2 \log D - 1 + t$ . According to the rank theorem,  $\dim \text{Im } U + \dim \ker U = \dim \mathbb{F}_2^r = r = 2 \log D - 1 + t$ . We find  $U$  by first constructing a subspace  $\mathcal{W}$  of dimension  $t$ , satisfying the same properties in Eq. (C1), followed by building a  $U$  with kernel equal to  $\mathcal{W}$ . More precisely, we will find  $t$  many linearly independent vectors  $w_1, \dots, w_t \in \mathcal{V}_r := \text{span}\langle v_1, \dots, v_r \rangle$  that would define such a  $\mathcal{W}$ . Then, by basic linear algebra, there is an efficient classical algorithm that finds linearly independent vectors  $u_1, \dots, u_{2 \log D - 1} \in \mathcal{V}_r$  that satisfy  $u_j \cdot w_i = 0, \forall i, j$ . Because of their linear independence, a matrix  $U$  defined by such  $u_j$ 's would have rank  $2 \log D - 1$ , so  $\dim \text{Im } U = 2 \log D - 1 \Rightarrow \dim \ker U = t$ . Finally, since  $\forall i: w_i \in \ker U$  and  $\dim \ker U = \dim \mathcal{W}$ , it follows  $\ker U = \mathcal{W}$ . Therefore,  $\ker U$  satisfies Eq. (C1), as desired.

*Note.* Going forward, as operations are over the field  $\mathbb{F}_2$ , we may play loose with subtraction and addition, as  $v_i - v_j = v_i + v_j$ .

Let us start by proving the base of induction  $t = 1 \Rightarrow r = 2 \log D$ . We need to find a single vector  $w_1$  such that  $v_i \neq w$  and  $v_i - v_j \neq w$ . The number of distinct vectors in the set  $\{v_i, v_i - v_j\}_{i,j}$  is at most  $D + \binom{D}{2} = (D^2 + D)/2 \leq 2^{2 \log D - 1} + 2^{\log D - 1} < |\mathbb{F}_2^r| = 2^r = 2^{2 \log D}$ . Therefore, there exists  $w \in \mathbb{F}_2^r - \{v_i, v_i - v_j\}_{i,j}$ , and this vector can be found after a search over  $(D^2 + D)/2 + 1$  vectors picked from  $\mathbb{F}_2^r$ . Thus,  $w_1$  for the base of induction can be found efficiently.

For the induction step, without loss of generality assume  $v_1, \dots, v_{2 \log D - 1 + t}$  are all linearly independent and generate

the rest of the  $v_i$ 's (we note again that finding these generators can be done efficiently). By induction hypothesis  $w_1, \dots, w_{r-1} \in \mathcal{V}_{r-1} = \text{span}\langle v_1, \dots, v_{r-1} \rangle$  form a desired subspace  $\mathcal{W}_{r-1}$  for the previous induction step. Note that clearly  $\mathcal{V}_{r-1} \subset \mathcal{V}_r = \text{span}\langle v_1, \dots, v_r \rangle$ .

We can partition the set of all vectors  $\{v_i\}_{i=1}^D$  into three subsets: (1)  $\mathcal{M} := \{v_i \mid v_i \in \mathcal{V}_{r-1}\}$ , which elements will be referred to as  $m_i$ , (2) the single-element subset  $\{v_r\}$ , and (3)  $\mathcal{N} := \{v_i \mid v_i \in \mathcal{V}_r - \mathcal{V}_{r-1}\}$ . The latter will have vectors that look like  $v_i = m'_i + v_r$  where  $0 \neq m'_i \in \mathcal{V}_{r-1}$ . Because of the partitioning,  $|\mathcal{M}| + 1 + |\mathcal{N}| = D$ . Note any future use of  $m_i, m'_i$  will refer to a  $v_i$  inside  $\mathcal{M}, \mathcal{N}$ , respectively. We emphasize that  $\mathcal{M}, \mathcal{N}$  are sets and not necessarily a linear subspace.

We would like to invent a new set of  $D$  vectors with rank  $r - 1$ , so that we can apply the induction hypothesis. To do so, let us replace  $v_r$  with some  $l \in \mathcal{V}_{r-1}$ , and similarly substitute every  $v_r$  in the linear expansion of any  $v_i = m'_i + v_r \in \mathcal{N}$ , meaning  $v_i$  becomes  $l + m'_i$ . This vector  $l$  needs to satisfy some properties:

$$l \neq \mathbf{0}, m'_i + l \neq \mathbf{0}, \quad (\text{C2})$$

$$l \neq m_j, m'_i + l \neq m_j. \quad (\text{C3})$$

The first two conditions ensure that after replacement, we do not obtain any zero vector. The second line ensures that we do not obtain any repeated vector. All these conditions amount to  $l \notin \{\mathbf{0}, m'_i, m_j, m_j + m'_i\}$ , the size of this set being (at most)  $1 + |\mathcal{N}| + |\mathcal{M}| + |\mathcal{M}| \cdot |\mathcal{N}|$ . We recall  $1 + |\mathcal{N}| + |\mathcal{M}| = D$ , so the size is  $\leq D + (D - 1 - |\mathcal{N}|)|\mathcal{N}| \leq D + D^2/4 \leq 2^{\log D} + 2^{2 \log D - 2} < 2^{r-1} = 2^{2 \log D - t - 2}$  as  $t > 1$ . Hence, there exists  $l \in \mathcal{V}_{r-1}$  that satisfies Eqs. (C2) and (C3).

Now the induction hypothesis for the new set of  $D$  vectors apply, since the rank has clearly been decreased by one to  $r - 1 = 2 \log D - 1 + t - 1$ . Therefore, there exists a subspace  $\mathcal{W} = \text{span}\langle w_1, \dots, w_{t-1} \rangle \subset \mathcal{V}_{r-1}$  satisfying Eq. (C1) for this new set of vectors. Now let us bring back  $v_r$  by undoing the replacement by  $l$ . After this change, we need to verify that  $\mathcal{W}$  still satisfies Eq. (C1), and then, in order to finish the proof, extend  $\mathcal{W}$  by a vector  $w_t$  while satisfying said properties. We verify the properties as follows:

- We first check that  $v_i, v_i + v_j \notin \mathcal{W}$ . Note that  $v_i \notin \mathcal{W}$  needs to be checked only for  $v_r$  and  $v_i \in \mathcal{N}$  (as they are the only ones impacted by bringing back  $v_r$ ). For both, this is in fact obvious as  $(\{v_r\} \cup \mathcal{N}) \subset \mathcal{V}_r - \mathcal{V}_{r-1} \Rightarrow (\{v_r\} \cup \mathcal{N}) \cap \mathcal{V}_{r-1} = \emptyset$  while  $\mathcal{W} \subset \mathcal{V}_{r-1}$ .
- For the property  $v_i + v_j \notin \mathcal{W}$ , this needs to be checked only when at least one of  $v_i, v_j$  is inside  $(\{v_r\} \cup \mathcal{N})$ .

- Assume that  $v_i \in (\{v_r\} \cup \mathcal{N})$  and  $v_j = m_j \in \mathcal{M}$ . If  $v_i = v_r$ , we need to show  $v_r + m_j \notin \mathcal{W}$  and for  $v_i \in \mathcal{N}$ , we need to prove  $m'_i + v_r + m_j \notin \mathcal{W}$ . Both these cases, due to  $v_r$ , are outside of  $\mathcal{V}_{r-1}$  and  $\mathcal{W} \subset \mathcal{V}_{r-1}$ .
- Assume that  $v_i, v_j \in (\{v_r\} \cup \mathcal{N})$ . Then we need to check  $v_r + m'_j + v_r = m'_j \notin \mathcal{W}$ , and also  $m'_i + v_r + m'_j + v_r = m'_i + m'_j \notin \mathcal{W}$ . However, by induction hypothesis, we already know that  $l + m'_j + l = m'_j \notin \mathcal{W}$  and  $l + m'_i + l + m'_j = m'_i + m'_j \notin \mathcal{W}$ , so this is also guaranteed.

Finally, we need to find a new  $w_t$  to add to  $\mathcal{W}$  while preserving its properties. Let us define  $w_t = v_r + l$  and let  $\mathcal{W}' = \mathcal{W} \oplus w_t \subset \mathcal{V}_r = \mathcal{V}_{r-1} \oplus v_r$ . To prove that  $\mathcal{W}'$  satisfies Eq. (C1), one needs to verify that  $v_i \notin \mathcal{W}' - \mathcal{W}$  or  $v_i - v_j \notin \mathcal{W}' - \mathcal{W}$ , as  $\mathcal{W}$  has already been shown to satisfy said properties. If  $v_i \in \mathcal{W}' - \mathcal{W}$  or  $v_i - v_j \in \mathcal{W}' - \mathcal{W}$  then  $w_t$  must be “involved”:

- For  $v_i \in \mathcal{W}' - \mathcal{W}$ , we must have  $v_i = w_t + w = v_r + l + w$  for some  $w \in \mathcal{W}$ , in which case  $v_i - l = v_r + w$ . However, the latter is inside  $\mathcal{V}_r - \mathcal{V}_{r-1}$ . Therefore, since  $l \in \mathcal{V}_{r-1}$ , we have  $v_i \in \mathcal{V}_r - \mathcal{V}_{r-1}$ . So  $v_i \in (\{v_r\} \cup \mathcal{N})$ . If  $v_i = v_r$  then  $l = w \in \mathcal{W}$ , which violates our induction hypothesis. If  $v_i = m'_i + v_r \in \mathcal{N}$  then  $m'_i + l = w \in \mathcal{W}$ , which again violates the construction of  $\mathcal{W}$ .
- For  $v_i - v_j = v_r + l + w$  for some  $w \in \mathcal{W}$ , exactly one of  $v_i$  or  $v_j$  must be inside  $(\{v_r\} \cup \mathcal{N})$ . Without loss of generality assume  $v_i = v_r$  or  $v_i = m'_i + v_r$ . Then this simplifies to  $l - v_j = w \in \mathcal{W}$  or  $m'_i + l - v_j = w \in \mathcal{W}$ , both violating the induction hypothesis for  $\mathcal{W}$ .

This shows  $\mathcal{W}'$  satisfies Eq. (C1) and finishes the induction. The significant cost in each inductive step is the search to find  $l$ , taking  $O(D^2/2 + D)$  steps.

**Resource estimation.** The total complexity is found by applying this for each induction step, thus  $O(tD^2)$ . Note that  $t \leq \min(2N, D) - 2 \log D + 1$ , with equality when all the vectors  $v_i$  are linearly independent; so the total cost of the classical algorithm used to find  $U$  is at most  $O(D^2(\min(2N, D) - 2 \log D + 1))$ . It should be noted that the search process can be fully parallelized, using all cores on an available machine, and given the nature of this search, the expected runtime could be much less. ■

#### APPENDIX D: TRADING OFF TOFFOLIS WITH QUBITS IN THE SOS ALGORITHM

As explained in the main text, trading off Toffolis with qubits can be done by using an alternative version of QROM. This variant, which we shall call SELSWAPDIRTY [71, Fig. 1(d)], has a parameter  $\lambda$  that allows for trading

off qubits with Toffolis. For a QROM loading  $L$  many data points  $|x_i\rangle$ , indexed by  $i = 1, \dots, L$  and of size  $c$ , the trade off  $\lambda \in [1, L]$  can be applied to change the Toffoli complexity from  $O(L)$  to  $O(L/\lambda + \lambda c)$  while increasing the uninitialized (so-called dirty) qubit cost to  $O(\lambda c)$ . Notice that the volume cost stays as  $O(Lc)$ , although technically, as we traded gates with dirty qubits, this volume is not a clean volume, so it is an overall improvement. To keep our discussion focused on the novelties and following the convention in previous resource estimations [72], we will select  $\lambda = \sqrt{L/c}$  in our applications. This strikes a balance in the trade off, using “equally” many ( $O(\sqrt{Lc})$ ) Toffolis as dirty qubits.

The first expensive QROM is employed when outputting the system register in Eq. (10). Using the SELSWAPDIRTY variant with Toffoli cost  $2D/\lambda + 8(2N)\lambda$  with  $\lambda = \sqrt{2D/16N}$ , this QROM Toffoli cost can be lowered to  $2\sqrt{32ND}$  while also using  $\sqrt{32ND}$  dirty qubits.

We also could have chosen to use SELSWAPDIRTY QROM to flip the register  $|i\rangle$  using  $|b_i\rangle$  in

$$\sum_{i=1}^D \alpha_i |i\rangle |v_i\rangle_v |b_i\rangle_b, \quad (\text{D1})$$

where we have denoted the registers by subscripts. We need to employ the inverse  $O^\dagger$  of the SELSWAPDIRTY operator  $O |b_i\rangle_b |0\rangle^{\otimes \log D} = |b_i\rangle_b |i\rangle$ . The naive implementation of  $O$  would read  $2 \log D - 1$  qubits of  $b_i$ , therefore its optimized Toffoli cost would scale as  $O(\sqrt{D^2 \cdot \log D}) = O(D\sqrt{\log D})$ . However, know that we only have  $D \sim 2^{\log D}$  many  $b_i$ 's and we would like to exploit this fact, achieving a Toffoli cost that is sublinear in  $D$ .

First we note that we have knowledge of the value of  $b_i$  as we computed them classically. Let  $\lambda = \lceil \sqrt{D} \rceil$ . Let us now order  $b_{l_1} < \dots < b_{l_D}$ . We compute

$$|b_j\rangle |0\rangle^{\otimes \lceil \log D/2 \rceil} \rightarrow |b_j\rangle |f(j)\rangle, \quad (\text{D2})$$

where  $f(j)$  is the unique index such that  $b_{l_{\lambda f(j)-1}} < b_j < b_{l_{\lambda f(j)}}$ , where for  $\lambda f(j) > D$ , we set  $l_{\lambda f(j)} = D$ . This computation requires comparing  $b_j$  to  $\lceil D/\lambda \rceil$  many other  $b_i$ 's. This means a Toffoli cost of  $(2 \log D - 1 + 2 \log D - 1)$  (the comparator cost of two  $2 \log D - 1$  bits integers) multiplied by the  $\lceil D/\lambda \rceil$  comparisons that we have to make. More precisely, starting from  $b_{q\lambda}$  for  $q = 1$ , we compare  $b_j$  with  $b_{q\lambda}$  stored in an auxiliary register, and if  $b_j > b_{q\lambda}$ , we store  $|1\rangle$  in another auxiliary register and otherwise  $|0\rangle$ . Then we add the value of that register to another register holding  $|q\rangle$ , which will eventually become  $|f(j)\rangle$ . This summation itself costs  $\lceil \log D/2 \rceil + 1$  Toffolis as the size of the register holding  $q$  is  $\lceil \log D/2 \rceil$  (the qubit cost is effectively zero, as any ancilla used is immediately liberated without Toffolis). As this summation must be performed for all  $q \in \{1, \dots, \lceil D/\lambda \rceil\}$ ,

we have an additional  $\lceil \log D/2 \rceil \cdot \lceil (D/\lambda) \rceil$  to account for. In total, the cost for this computation is  $(4 \log D - 2 + \lceil \log D/2 \rceil \lceil (D/\lambda) \rceil) \sim ((9 \log D \sqrt{D})/2)$ , and we note that its (later) uncomputation can be done with Clifford gates. After this, our state becomes

$$\sum_{i=1}^D \alpha_i |i\rangle |v_i\rangle_v |b_i\rangle_b |f(i)\rangle. \quad (\text{D3})$$

Next, we read the register  $f(i)$  using a SELECT QROM with cost  $2^{\log D/2}$  and output the  $\lambda$  block of  $b_j$ 's in which we know  $b_i$  lives in, i.e.,  $|b_{\lambda(f(i)-1)+1}, \dots, b_{\lambda f(i)}\rangle$ .

$$\sum_{i=1}^D \alpha_i |i\rangle |v_i\rangle_v |b_i\rangle_b |f(i)\rangle |b_{\lambda(f(i)-1)+1}, \dots, b_{\lambda f(i)}\rangle. \quad (\text{D4})$$

Notice the significant clean qubit cost  $(2 \log D - 1)\lambda \sim (2 \log D - 1)\sqrt{D}$ . The uncomputation of this step is later done with Clifford gates. CNOT-ing the register  $b_i$  into the block yields

$$\sum_{i=1}^D \alpha_i |i\rangle |v_i\rangle_v |b_i\rangle_b |f(i)\rangle \otimes |b_{\lambda(f(i)-1)+1} \oplus b_i, \dots, b_{\lambda f(i)} \oplus b_i\rangle. \quad (\text{D5})$$

By using  $\lambda$  many  $(2 \log D - 1)$ -MCCNOTs, with total Toffoli cost  $(2 \log D - 1)\lambda$ , we compute the index  $1 \leq g(i) \leq \lambda$  for which  $b_{\lambda(f(i)-1)+g(i)} = b_i$ . More precisely, we hold in an auxiliary register the index  $|k\rangle$  at which we are implementing the MCCNOT, and using Toffolis, bit by bit for every  $\lceil (\log D/2) \rceil$  bits, add that to the register designed for holding  $g(i)$  controlled on the result of the  $(2 \log D - 1)$ -MCCNOT. This requires an additional  $\lceil \log D/2 \rceil$  Toffolis and ensures that only the true  $k = g(i)$  is added to the register designed for holding  $g(i)$ . Therefore, the total Toffoli cost of this step is  $(2 \log D + \lceil \log D/2 \rceil - 1)\lambda \sim ((5 \log D \sqrt{D})/2)$ . We now have the following state:

$$\sum_{i=1}^D \alpha_i |i\rangle |v_i\rangle_v |b_i\rangle_b |f(i)\rangle \otimes |b_{\lambda(f(i)-1)+1} \oplus b_i, \dots, b_{\lambda f(i)} \oplus b_i\rangle |g(i)\rangle. \quad (\text{D6})$$

Notice the register  $|f(i), g(i)\rangle$  has  $2\lceil (\log D/2) \rceil \sim \log D$  qubits, and determines  $i$  uniquely. Tracing this back to our original goal of distinguishing the  $D$  many  $v_i$ 's, this is the ideal we can hope for, as we have computed  $D$  many  $\log(D)$ -bits integers  $|f(i), g(i)\rangle$  that are distinct. The rest can be done using the inverse  $O^\dagger$  of a SELSWAPDIRTY

QROM which computes

$$O |f(i), g(i)\rangle |0\rangle^{\otimes \log D} = |f(i), g(i)\rangle |i\rangle \quad (\text{D7})$$

Notice that we have access to where  $b_i$  is in the list  $b_{l_1} < \dots < b_{l_D}$ , and therefore can classically compute the values  $f(i), g(i)$ . Hence the QROM has the required classical lookup data. This SELSWAPDIRTY QROM has Toffoli cost  $2\sqrt{16 \log(D)2^{\log D}} < 2\sqrt{32 \log(D)D}$  with dirty qubit cost  $< \sqrt{32 \log(D)D}$ . Note that these dirty qubits are already available using the  $\lambda$  block, which has  $(2 \log D - 1)\sqrt{D}$  qubits.

**Resource estimation.** Overall, the clean qubit cost increases to  $(2 \log D - 1)\sqrt{D}$ , and an additional dirty qubit cost  $\sqrt{32ND}$ . The clean qubit cost may be improved if one can find a way to compute  $g(i)$  without outputting the entire  $\lambda$  block. While the dirty qubit cost can be avoided if we do not apply SELSWAPDIRTY QROM in Eq. (10). In practice, the dirty qubit cost, though significantly larger than the clean  $(5 \log D)$  qubit cost in our previous method, may be actually available if a future computation (such as qubitization) to simulate the evolution of the Hamiltonian requires that much qubit. The Toffoli cost changes to  $\min(2\sqrt{32ND}, D) + ((9 \log D \sqrt{D})/2) + ((5 \log D \sqrt{D})/2) + 2^{\log D/2} + 2\sqrt{32 \log(D)D} \sim \min(2\sqrt{32ND}, D) + (7 \log D + 2\sqrt{32 \log D})\sqrt{D}$ . In case  $D$  is chosen from the minimum, the dirty qubit cost  $\sqrt{32ND}$  is lifted.

**Is the algorithm optimal?** Our algorithm is general in the sense that it simply assumes a given set of amplitudes  $\{\alpha_i\}$  and bitstrings  $\{v_i\}$ , with no assumption on the nature of either of the two sets. Furthermore, a dimensionality argument can show that the compression in Lemma 1 of  $v_i$  to  $2 \log D - 1$  bits is very likely to be tight, and we conjecture that the volume cost cannot be asymptotically lower than  $\tilde{O}(D)$ , and more strongly,  $O((\log D)^2 D)$ . We remark that this conjecture is for creating the superposition  $\sum_{i=1}^D \alpha_i |v_i\rangle$ , without any junk register, i.e.,  $\sum_{i=1}^D \alpha_i |v_i\rangle | \text{junk}_i \rangle$  is not acceptable. The only “approximation” allowed is in the amplitudes, and the distance to the state should be on the order of chemical accuracy. Both these restrictions are necessary as this preparation problem concerns the system register, and not, say, a PREP state in a qubitization protocol derived from an LCU, which *can* include a junk register, and hence why preparation methods such as coherent alias sampling [72] can be employed in that instance.

One wonders if the approach in Ref. [18] in iteratively generating the superposition can be combined with ours. This would involve ordering the states  $v_i$  and compressing  $v_1, \dots, v_l$  for each  $1 \leq l \leq D$ . Assuming that for each  $l$ , a compression of the bitstrings  $v_1, \dots, v_l$  to a length  $k_l$  is possible (and we know  $k_l \leq 2\lceil \log l \rceil - 1$ ), then the Toffoli cost of generating the superposition is  $\sum_{l=1}^D (k_l -$



1). This is upper bounded by  $\sum_{i=1}^{\lceil \log D \rceil - 2} (2i - 1)(2^i - 1) + (2\lceil \log D \rceil - 1)(D - 2^{\lceil \log D \rceil - 1})$  and can be seen to be smaller than our own cost  $(2\lceil \log D \rceil + 3)D$ . However, this method also involves a rotation *at each step* for  $1 \leq l \leq D$ , and even assuming access to a gradient state, the overall cost of these rotations is  $(c - 3)D$  Toffolis, where  $c$  is the required bit precision. Note that we also have to take into account the accuracy  $c$  for our rotations in generating the superposition in Eq. (9), however the associated cost is simply  $(c - 3) \log D$ , and according to Lemma E.1 in Ref. [73],  $c$  must satisfy  $c > \log(\log(D)\pi/\epsilon)$ , where  $\epsilon$  is chemical accuracy. Crucially,  $c$  is double logarithmic in terms of  $D$ , because there are only  $\log D$  rotations performed in generating Eq. (9). This is not the case when using the approach in Ref. [18], and for the overall error to be under chemical accuracy, we need  $c > \log(D\pi/\epsilon)$ , therefore making  $c$  logarithmic in  $D$ . Still, one could see the total cost  $\sum_{i=1}^{\lceil \log D \rceil - 2} (2i - 1)(2^i - 1) + (2\lceil \log D \rceil - 1)(D - 2^{\lceil \log D \rceil - 1}) + (\log(D\pi/\epsilon) - 3)D$  to be competitive within a constant factor of two with  $(2\lceil \log D \rceil + 3)D$ , but clearly, it does not lead to a significant constant-factor cost reduction. Nevertheless, this is a worst-case complexity analysis, and if  $k_i$ 's are significantly smaller, one might see benefits of this combination of the two methods. This could potentially be the case for a particular range of  $D$  and bitstrings  $v_i$ 's which have some common structure, such as being an excitation or two away from a reference state, and we leave that for future works.

## APPENDIX E: DETAILS OF RESOURCE ESTIMATION FOR MPS IMPLEMENTATION

In this Appendix we discuss the details of the MPS cost estimation presented in Sec. III B. In particular, we focus on the ancilla qubit and Toffoli costs of implementing  $G[j]$  defined in Eq. (23).

We represent the  $G$  operations as

$$G[j] = \sum_{\alpha_{j-1}} \left( |u_{\alpha_{j-1}}\rangle \langle \alpha_{j-1}, 0| \right) + \dots, \quad (\text{E1})$$

with

$$|u_{\alpha_{j-1}}\rangle = \sum_{\alpha_j, n_j} A_{j; \alpha_{j-1}\alpha_j}^{n_j} |\alpha_j, n_j\rangle. \quad (\text{E2})$$

In the states  $|\alpha_j, n_j\rangle$  the first and second arguments show the ancillae and system qudit indices, respectively. Note that the form of Eq. (E1) follows from the definition of  $G$  in Eq. (23). Each unitary is synthesized using a series of Householder reflections with the addition of a single

ancilla qubit as follows:

$$|0\rangle \langle 1| \otimes G[j] + |1\rangle \langle 0| \otimes G[j]^\dagger = \prod_{\alpha_{j-1}=1}^{\chi_{j-1}} (1 - 2P_{\alpha_{j-1}}), \quad (\text{E3})$$

where a total of  $\chi_{j-1}$  reflections are used, whose projectors are defined through an auxiliary state  $|w_{\alpha_{j-1}}\rangle$ :

$$P_{\alpha_{j-1}} = |w_{\alpha_{j-1}}\rangle \langle w_{\alpha_{j-1}}|, \quad (\text{E4})$$

with

$$\begin{aligned} |w_{\alpha_{j-1}}\rangle &= |1\rangle \otimes |\alpha_{j-1}, 0\rangle - |0\rangle \otimes |u_{\alpha_{j-1}}\rangle \\ &= W_{\alpha_{j-1}} |0\rangle \otimes |0, 0\rangle. \end{aligned} \quad (\text{E5})$$

On the second row, we have defined the operator  $W_{\alpha_{j-1}}$  that prepares  $|w_{\alpha_{j-1}}\rangle$ , also in the state  $|0\rangle \otimes |0, 0\rangle$  the 0's correspond to the reflection ancilla, MPS ancillae and the system qudit, respectively.

With the operator  $W_{\alpha_{j-1}}$ , each of the reflections in Eq. (E3) can be written as

$$1 - 2P_{\alpha_{j-1}} = W_{\alpha_{j-1}} [1 - 2|0, 0, 0\rangle \langle 0, 0, 0|] W_{\alpha_{j-1}}^\dagger, \quad (\text{E6})$$

and thus the cost of implementing each reflection is twice the cost of  $W_{\alpha_{j-1}}$  plus the cost of the simple reflection  $[1 - 2|0, 0, 0\rangle \langle 0, 0, 0|]$ . As a result, we need to also evaluate the cost of  $W_{\alpha_{j-1}}$ , which we synthesize as follows: we first define an operator that prepares the state  $|u_{\alpha_{j-1}}\rangle$ :

$$|u_{\alpha_{j-1}}\rangle = V_{\alpha_{j-1}} |0, 0\rangle. \quad (\text{E7})$$

It can be seen that  $W_{\alpha_{j-1}}$  as below can serve to satisfy the definition of  $|w_{\alpha_{j-1}}\rangle$  in Eq. (E5):

$$W_{\alpha_{j-1}} = (\bar{C}V_{\alpha_{j-1}}) (C_{\alpha_{j-1}}) ((ZH) \otimes \mathbb{I}), \quad (\text{E8})$$

where  $ZH$  (a Pauli  $Z$  and a Hadamard) acts on the reflection ancilla,  $C_{\alpha_{j-1}}$  is a product of CNOTs controlled on the reflection ancilla to prepare the state  $(1/\sqrt{2})[|0, 0, 0\rangle - |1, \alpha_j, 0\rangle]$ . With  $\bar{C}V_{\alpha_{j-1}}$  controlled negatively on the reflection ancilla, one can check that  $|w_{\alpha_{j-1}}\rangle$  is prepared up to a phase.

We are interested in the Toffoli cost of implementation and list all sources of Toffoli cost below:

- Simple reflections  $[1 - 2|0, 0, 0\rangle \langle 0, 0, 0|]$ : a number of  $\chi_{j-1}$  of them is required for  $G[j]$ .

- Operations  $\bar{C}V_{\alpha_{j-1}}$ : these are required for creating  $W_{\alpha_{j-1}}$  and are the only source of non-Clifford gates in their synthesis. A total of  $2\chi_{j-1}$  of such operators is required for  $G[j]$ .

First we note that the reflection  $[1 - 2|0\rangle^{\otimes \nu}\langle 0|^{\otimes \nu}]$  can essentially be thought of as a multicontrolled  $Z$  operation, and thus can be implemented using a circuit such as the one shown in Fig. 4.10 of Ref. [11]. However, the second half of the circuit, i.e., the uncomputing part can be done without Toffolis and in fact using measurements and Clifford gates as shown in Fig. 3 of Ref. [113]. This makes the total number of Toffoli gates and ancillae equal to  $\nu - 1$ .

Next, we discuss how  $\bar{C}V_{\alpha_{j-1}}$  can be implemented and estimate the required resources. For this we first discuss the implementation of  $V_{\alpha_{j-1}}$ .  $V_{\alpha_{j-1}}$  prepares the state

$$|u_{\alpha_{j-1}}\rangle = V_{\alpha_{j-1}} |0, 0\rangle, \quad (\text{E9})$$

which is defined as

$$|u_{\alpha_{j-1}}\rangle = \sum_{\alpha_j, n_j} A_{j; \alpha_{j-1}\alpha_j}^{n_j} |\alpha_j, n_j\rangle. \quad (\text{E10})$$

First, we take the above subspace of interest to consist of  $\nu$  qubits. The preparation of a generic state  $|u_{\alpha_{j-1}}\rangle$  can be done using the methods discussed in Ref. [71]; the state is carved qubit by qubit in  $\nu$  steps; in each step a single-qubit rotations on one qubit being controlled on all the previous entries is performed and with consecutive application of this procedure all the correct probabilities for bitstrings are reproduced at the end; one last multicontrolled single-qubit rotation is required to recover the complex phases corresponding to the components of the state in question (see p. 3 of Ref. [71] for details). Thus a total of  $\nu + 1$  of such single-qubit rotations are required for reproducing the state.

The rotation will be performed with the method given in Ref. [113], where access to a phase gradient state  $2^{-b/2} \sum_{k=0}^{2^b-1} e^{-2\pi i k/2^b} |k\rangle$  is assumed.  $b = \log(1/\delta_r)$  is the number of digits in the binary representation of the rotation angle and thus  $\delta_r$  is the error in rotation. The Toffoli cost of each single-qubit rotation is given by  $b + O(1)$  [113]; note that this also means we need an additional  $\log(1/\delta_r)$  additional qubits to store the phase of each single-qubit rotation.

Considering first the SELECT variant of implementation in Ref. [71], we now discuss the cost of control operations for the above single-qubit rotations. Since multicontrolling over a sequence of  $0, 1, 2, \dots, \nu$  qubits is required to store the respective rotation angles, we will, respectively, have a sequence of Toffoli costs of  $2^0 - 1, 2^1 - 1, 2^2 - 1, \dots, 2^\nu - 1$  according to Ref. [114] (see e.g., Fig. 7 therein). Note that we are interested in implementing

$\bar{C}V_{\alpha_{j-1}}$ , and Ref. [114] also considers a controlled operation for the above cost. After each single-qubit rotation, the qubits storing the rotation angle should be uncomputed and this adds a multiplicative factor of 2. As a result, for the SELECT variant we have a total Toffoli cost equal to

$$2^{\nu+2} + \nu b, \quad (\text{E11})$$

where we have dropped an additive  $-\nu$  term in the sum as it is subdominant.

In our particular case of interest, i.e., synthesis of  $G[j]$ , we have a number of reflections shown in Eq. (E3), equal to  $\chi_{j-1}$ . On the other hand, the Hilbert space over which each of the reflections acts is  $\chi_j d$  and as a result,  $N$  in the above treatments should be  $2^\nu = \chi_j d$ . This means that in the SELECT variant, the total Toffoli cost reads:

$$\chi_{j-1} [8\chi_j d + b \log(\chi_j d) + \log(\chi_j d)], \quad (\text{E12})$$

Next turning to the other variant SELSWAPDIRTY of Ref. [71], which is capable of reducing the Toffoli cost if dirty qubits are available; we saw above that a number  $b$  qubits is required for storing the rotation angles, however with the addition of a number  $\lambda b$  dirty qubits, we can use this variant of the algorithm. A total of  $\nu + b$  extra clean qubits are also required (excluding the ancillae required for performing single-qubit rotations like the phase gradient state); note that this is the same number as the SELECT variant. Moreover, for SELSWAPDIRTY, one needs to perform swaps also which will add to the total Toffoli cost. It is straightforward to see that the Toffoli gate cost in this case reads

$$2 \frac{2^{\nu+2}}{\lambda} + 4 \cdot 2\lambda \nu b + \nu b. \quad (\text{E13})$$

The first term corresponds to multiqubit controls, the second term swaps and the third-term single-qubit rotations. The factors 2 and 4 in the first and the second terms appear because SELECT and SWAP need to be done twice and four times in SELSWAPDIRTY (see Fig. 1(d) of [71]). The factor of 2 in the second term comes from uncomputing the rotation angles.

As is discussed in Ref. [71], it is best for Toffoli gate count to be  $\lambda = O(\sqrt{2^\nu})$ , but we will keep it unspecified for the rest of the discussion.

Gathering all the above costs together for synthesizing  $G[j]$  in Eq. (E1), we see that the Toffoli cost reads using the SELSWAPDIRTY variant:

$$\chi_{j-1} \left[ 8 \frac{\chi_j d}{\lambda} + 8\lambda b \log(\chi_j d) + b \log(\chi_j d) + \log(\chi_j d) \right]. \quad (\text{E14})$$

In total, assuming a number  $N$  of qudits in the physical system, the total cost will be the sum of the above;

TABLE II. Edgeworth series terms.

$s$	Edgeworth term
1	$(\kappa_3/6) He_3(x)$
2	$(\kappa_4/24) He_4(x) + (\kappa_3^2/72) He_6(x)$
3	$(\kappa_5/120) He_5(x) + (\kappa_4\kappa_3/144) He_7(x) + (\kappa_3^3/1296) He_9(x)$
4	$(\kappa_6/720) He_6(x) + (\kappa_5\kappa_3/720 + \kappa_4^2/1152) He_8(x) + (\kappa_4\kappa_3^2/1728) He_{10}(x) + (\kappa_3^4/31104) He_{12}(x)$
5	$(\kappa_7/5040) He_7(x) + (\kappa_6\kappa_3/4320 + \kappa_5\kappa_4/2880) He_9(x) + (\kappa_5\kappa_3^2/8640 + \kappa_4^2\kappa_3/6912) He_{11}(x) + (\kappa_4\kappa_3^3/31104) He_{13}(x) + (\kappa_3^5/933120) He_{15}(x)$

asymptotically and with using  $\chi$  selectively for all bond dimensions, the dominant Toffoli cost can be written as  $O(N\chi^{3/2})$ .

## APPENDIX F: EDGEWORTH SERIES TERMS

In general, the Edgeworth series terms can be written as [81]

$$p_E(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \left[ 1 + \sum_{s=1}^{\infty} \sum_{\{k_m\}} He_{s+2r}(x) \prod_{m=1}^s \frac{1}{k_m!} \left( \frac{\kappa_{m+2}}{(m+2)!} \right)^{k_m} \right], \quad (\text{F1})$$

where the summation over  $\{k_m\}$  in the above denotes summation over all non-negative integer solutions of the Diophantine equation

$$k_1 + 2k_2 + \dots + sk_s = s, \quad (\text{F2})$$

and  $r$  is the sum of these integers for each solution:  $r = \sum k_m$ . The explicit forms for a few of the orders of the Edgeworth expansion can be found in Table II.

## APPENDIX G: KERNEL DENSITY APPROXIMATION

Here, we give a quick overview of the kernel density approximation method. Supposing we have access to a finite number of samples drawn from a distribution function, the goal is to approximate the distribution function. To this end, a broadening kernel is placed at the position of each of the outcomes and a normalized sum approximates the underlying distribution:

$$\hat{p}(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - X_i}{h}\right), \quad (\text{G1})$$

where  $K$  is a kernel (e.g., Gaussian, Lorentzian, etc.) with mean of 0 and variance of 1,  $X_i, i = 1, \dots, M$  are the outcomes of sampling and  $h$  is the broadening factor.

The analysis of error in reconstructing the above QPE-kernel energy distribution with kernel density estimation follows a standard approach [88]. First, the error is quantified by the quantity mean integrated square error ( $R_{\text{MISE}}$ ):

$$R_{\text{MISE}} = \mathbb{E} \left( \int dx (\hat{p}(x) - p(x))^2 \right), \quad (\text{G2})$$

where  $\hat{p}(x)$  is the approximated distribution for the underlying distribution  $p(x)$ . When a sample of size  $M$  is used, it is well known that with an appropriate choice of  $h$ , i.e.,  $h_{\text{opt}} \sim 1/M^{1/5}$ , the error also shows the behavior  $R_{\text{MISE}} \sim 1/M^{4/5}$ .

## APPENDIX H: DETAILS OF THE QUANTUM EIGENVALUE TRANSFORMATION OF UNITARY MATRICES METHOD

The method consists of a quantum signal-processing circuit [93,94] that implements a unitary matrix that block encodes a polynomial function  $f(H) = P(\cos(H/2))$ , where  $H$  is the Hamiltonian of interest and  $P$  is an even polynomial of degree  $d$ . A schematic of the quantum circuit is shown in Fig. 17. The circuit works by implementing  $U = e^{-iH}$  and its Hermitian conjugate controlled on a single ancilla, a total number of  $d$  times. The parameters  $\varphi_0, \varphi_1, \dots, \varphi_{d/2}$  are determined based on the polynomial of interest. Upon measuring the ancilla qubit at the end and obtaining the outcome 0, the implementation has been successful, the probability of success is given by  $\|P(\cos(H/2))|\psi\rangle\|$ .

A scheme of the quantum circuit is shown in Fig. 17. The circuit works by implementing  $U = e^{-iH}$  and its Hermitian conjugate controlled on a single ancilla, a total number of  $d$  times. The parameters  $\varphi_0, \varphi_1, \dots, \varphi_{d/2}$  are determined based on the polynomial of interest. Upon measuring the ancilla qubit at the end and obtaining the

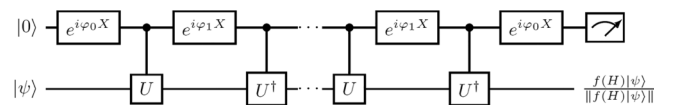


FIG. 17. THE QETU circuit, see the main text for explanation. The figure is taken from Ref. [92].

outcome 0, the implementation has been successful, the probability of success is given by  $\|P(\cos(H/2))|\psi\rangle\|$ .

The polynomial that needs to be implemented for our energy-filtering task should be a symmetric function that retains low energies and filters high energies. We take the spectrum of the Hamiltonian to lie within the interval  $[-\pi + \eta, 0 - \eta]$ , if necessary this can be done by adding a constant to and/or rescaling the Hamiltonian before performing QETU. Note that this is contrary to the original setting of Ref. [92] (the spectrum is contained in  $[\eta, \pi - \eta]$ ) and coarse QPE mentioned above; the reason for this change is better performance. We need a polynomial  $P$ , which when expressed as  $P(\cos(H/2))$  can filter high energies; it is straightforward to see that using the following combination of error functions, which we will try to imitate using the polynomial  $P$ , it is possible to filter high energies:

$$\xi_{k,\mu}(x) = \frac{1}{2} [\text{erf}(-k(x - \mu)) + \text{erf}(k(x + \mu))], \quad (\text{H1})$$

with  $0 < k$  and  $0 < \mu < 1$  determining the steepness and position of the transitions, i.e., position of energy filtering in the function. We use the prescription in Appendix A of Ref. [115] to reconstruct the error function  $\text{erf}(kx)$  in terms of Chebyshev polynomials as follows:

$$p_{\text{erf},k,n} = \frac{2k e^{-k^2/2}}{\sqrt{\pi}} \left( I_0(k^2/2) + \sum_{j=1}^{(n-1)/2} (-1)^j I_j(k^2/2) \left[ \frac{T_{2j+1}(x)}{2j+1} - \frac{T_{2j-1}(x)}{2j-1} \right] \right), \quad (\text{H2})$$

where  $T_j$  is the degree  $j$  Chebyshev polynomial,  $I_j$  is the modified Bessel function of the first kind. Note that  $p_{\text{erf},k,n}$  is an odd polynomial of degree  $n$ ; it is the degree  $n$  that controls the error in approximating  $\text{erf}(kx)$  and thus ensuring that low energies are retained and high energies filtered, and therefore it should be chosen large enough (see below). An example constructing polynomials like this is shown in Fig. 18.

Applying a successful round of QETU filtering to a state  $|\psi\rangle = \sum_E \varphi_E |E\rangle$ , we end up with the following unnormalized state:

$$\sum_E \varphi_E P(\cos(E/2)) |E\rangle |0\rangle. \quad (\text{H3})$$

This shows that supposing we want to keep energies below  $E_l$  and filter energies above  $E_u$ , we can choose a filtering function in Eq. (H1) (to be approximated by  $P$ ) with  $\mu = ((\cos(E_u/2) + \cos(E_l/2))/2)$  and  $(1/k) = \zeta ((\cos(E_u/2) - \cos(E_l/2))/2)$ . The factor  $\zeta$  is added so

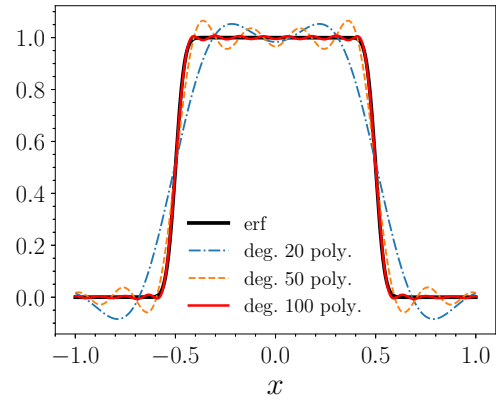


FIG. 18. The even filtering functions required for QETU. Polynomial approximations of the same function are also shown.

that it is possible to control the intensity of filtering while keeping the degree of the polynomial and the cost down.

The degree of a polynomial that needs to be used for this task will have a scaling  $O(\Gamma^{-1} \log \epsilon^{-1})$ , where  $\epsilon$  is the error in the polynomial approximation and  $\Gamma$  is the energy scale over which the transition in the error functions in Eq. (H1) happens, and thus should scale as  $1/k$ . Apart from the above asymptotic scaling, in practice, we choose  $n$  by examining how good of an approximation is achieved for degree  $n$ .

## APPENDIX I: SIMPLIFIED NUMERICAL EXAMPLE FOR THE QUANTUM REFINING STEP

Here, as a simple concrete model, we consider a Gaussian energy distribution for our initial state. This Gaussian distribution can be characterized by a mean value  $\bar{E}$  and a width  $\sigma_E$ :

$$A(E) = \frac{1}{\sqrt{2\pi}\sigma_E} e^{-\frac{(E-\bar{E})^2}{2\sigma_E^2}}, \quad (\text{I1})$$

Even though the energy distribution of an initial state might not actually be close to Gaussian in general, but we expect at least some variational states to show qualitatively similar behavior.

We work with following concrete example of a Gaussian distribution:  $\bar{E} = 0.06$ ,  $\sigma = 0.02$ . We would like to estimate the resources required to obtain a result close to 0 using QPE. With the above choice of parameters for the energy distribution, the accumulated weight below 0 is  $p_{<}(0) = 0.0013$ . This means that we need roughly  $1/p_{<}(0)$  measurements to obtain a value around 0. We can perform quantum refining to decrease the number of times the most expensive quantum energy estimation routine is performed. We take this most precise routine to be a QPE with  $k = 10$  digits for this example, however, we tolerate an error of  $2^{-8}$ , discarding the last two digits in any QPE outcome. The number of digits  $k$  furthermore determines



the total evolution time required as  $T \sim 2^k$ . We characterize the cost of different operations by the number of queries they require to the unitary  $e^{-iH}$ , which for a  $k$ -digit QPE measurement becomes  $2^k$ . This means that each round of the ultimate QPE measurement brings in a cost of  $2^{10}$ .

We first consider coarse QPE for energy filtering in this setting. Concretely, we do a coarse QPE measurement with 4 digits and only keep the results that show outcome 0 in the measured phase register, we can filter out some part of the weight as shown in Fig. 7 (left). This outcome happens with a probability of  $W_{k=4} = 0.10$ . In this new energy distribution the total weight below zero now reads  $p'_{\text{QPE},<}(0) = 0.012$ . This means that, after such measurement, roughly ten times less rounds of the precision QPE will be required compared to the initial state.

One can do this procedure one more time with a coarse QPE with 5 digits now and postselecting on the outcome 0 again; the resulting weight distribution can be seen in Fig. 7 (left). The probability of such outcome (given the previous outcome of 0 with 4 digits) is now  $W_{k=4;5} = 0.13$  (this means that the probability of obtaining 0 in the 4-digit measurement and then also 0 in the 5-digit measurement is  $W_{k=4}W_{k=4;5} = 0.013$ ). Remarkably with this measurement, the total weight below zero becomes  $p''_{\text{QPE},<}(0) = 0.083$ ; this results in close to 2 orders of magnitude decrease in the number of precision QPE measurement required for obtaining outcomes close to 0. This is achieved for a cost of  $2^4 + 2^5$ , which is an insignificant overhead compared with the cost of the most precise QPE measurement.

For QETU, we shift the energies so that low energies are located close to  $-\pi$  as discussed in Appendix H. We use a degree 200 polynomial to approximate a step function as shown in Fig. 7. The normalized distribution after QETU has been performed moves to the left and thus some of the higher energies are filtered. Probability of success in this case is  $W_{\text{QETU}} = 0.21$  and the total weight below 0 after the procedure can also be calculated as  $p'_{\text{QETU},<} = 0.0056$ . This means decreasing the number of repetitions roughly by a factor of 4. As the polynomial that is used is order 200, the number of required queries to  $e^{-iH}$  is also 200.

We see that both coarse QPE and QETU refining methods can be helpful for a cost that is an insignificant fraction of the ultimate QPE cost, but coarse QPE acts considerably better for a lower cost. As creating steep polynomials like the one used here is generally a hard task, we believe this result should hold generically even though we tested it here for a simple model.

### a. Mitigating the leakage

Another thing that can be studied in this simple model is the probability of leakage before and after the refining is performed. We consider only coarse QPE for this. Before

any of the measurements are performed the total probability of leakage is equal to  $p_{\text{leak}} = 0.00097$  which close to  $p_{<}(0) = 0.0013$  and this can be problematic by contributing outcomes below the actual energy levels of the system. Upon performing 4-digit and 5-digit QPE measurements discussed above, the probability of leakage becomes  $p'_{\text{leak}} = 0.0019$  and, respectively,  $p''_{\text{leak}} = 0.0036$ . These two values when compared with  $p'_{\text{QPE},<}(0) = 0.012$  and  $p''_{\text{QPE},<}(0) = 0.083$  show that the probability of leakage has decreased substantially enough compared to the probability of obtaining results of interest, so that its occurrence has become improbable, and thus quantum refining has suppressed the possibility of leakage also.

## APPENDIX J: ERROR ANALYSIS OF THE QPE LEAKAGE APPROXIMATE FORM

In this Appendix, we analyze Eq. (46) and in particular how the approximation in Eq. (47) can be performed.

$$p_{\text{leak}}(E_n) = \sum_{x_j < x_{\text{upper}}} \frac{1}{2^{2k}} \left( \frac{\sin^2(\pi \delta_n)}{\sin^2\left(\frac{\pi}{2^k}[x_n + \delta_n - x_j]\right)} \right), \quad (\text{J1})$$

we take the lower bound in the summation over  $x_j$  to be  $-2^{k-1}$ , and we are using the periodicity of the QPE results. A lower bound and an upper bound for the above sum can be found by using the following integral form:

$$\begin{aligned} I(x_0) &= \frac{\sin^2(\pi \delta_n)}{2^{2k}} \int_{-2^{k-1}}^{x_0} \frac{dx}{\sin^2\left(\frac{\pi}{2^k}(x_n + \delta_n - x)\right)} \\ &= \frac{\sin^2(\pi \delta_n)}{2^{2k}} \left[ \frac{2^k}{\pi} \cot\left(\frac{\pi}{2^k}(x_n + \delta_n - x)\right) \right]_{-2^{k-1}}^{x_0}. \end{aligned} \quad (\text{J2})$$

It is easy to check that

$$I(x_{\text{upper}} - 1) \leq p_{\text{leak}}(E_n) \leq I(x_{\text{upper}}). \quad (\text{J3})$$

This readily results in Eq. (47), and the error can also be shown to have the form  $O\left(\max\left[2^{-2k}, (x_n - x_{\text{upper}})^{-2}\right]\right)$  by evaluating  $I(x_{\text{upper}}) - I(x_{\text{upper}} - 1)$ .

## APPENDIX K: ENERGY AND OVERLAP TABLES

In this Appendix, we provide the energy and overlap numbers from the numerical studies section (specifically Figs. 9, 12, and 14) for general reference.

TABLE III. Energies (Ha) and overlaps for the H<sub>12</sub> hydrogen chains shown originally in Fig. 9. The CCSD energy reported is that returned by the CCSD method in PySCF.

Quantity	Bond length (Å)	HF	CISD	CCSD	HCI	DMRG	FCI
$E$ , Ha	0.71	−5.9836	−6.0994	−6.0991	−6.1057	−6.1059	−6.1059
	1.42	−5.6987	−5.9992	−5.9573	−6.1158	−6.1188	−6.1203
	2.84	−4.0757	−4.7544	−4.5582	−5.6572	−5.6577	−5.6578
$\langle\psi \psi_{\text{FCI}}\rangle$	0.71	0.960	0.999	0.999	1.000	1.000	
	1.42	0.669	0.899	0.921	0.987	0.999	
	2.84	0.133	0.314	0.333	0.747	0.966	

TABLE IV. Energies (Ha) and overlaps for the N<sub>2</sub> nitrogen molecules with stretched bond lengths shown originally in Fig. 12. The CCSD energy reported is that returned by the CCSD method in PySCF.

Quantity	Bond length (Å)	HF	CISD	CCSD	CASCI	MRPT	HCI	DMRG
$E$ , Ha	2.517	−18.6047	−19.1193	−19.1628	−19.2749	−19.4169	−19.4473	−19.4467
$\langle\psi \psi_{\text{DMRG}}\rangle$	2.517	0.336	0.363	0.197	0.345	0.345	0.992	1.000

TABLE V. Energies (Ha) and overlaps for the Cr<sub>2</sub> dimer within the active space of the 10 3*d* orbitals, with stretched bond lengths, shown originally in Fig. 14. The CCSD energy reported is that returned by the CCSD method in PySCF.

Quantity	Bond length (Å)	HF	CISD	CCSD	HCI	DMRG
$E$ , Ha	3.024	−2085.2675	−2085.9015	−2086.4395	−2086.6067	−2086.6067
$\langle\psi \psi_{\text{DMRG}}\rangle$	3.024	0.110	0.328	0.205	1.000	1.000

- [1] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* **309**, 1704 (2005).
- [2] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, *Proc. Natl. Acad. Sci.* **114**, 7555 (2017).
- [3] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya *et al.*, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
- [4] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, *Phys. Rev. Res.* **3**, 033055 (2021).
- [5] A. Ho, J. McClean, and S. P. Ong, The promise and challenges of quantum computing for energy storage, *Joule* **2**, 810 (2018).
- [6] J. E. Rice, T. P. Gujarati, M. Motta, T. Y. Takeshita, E. Lee, J. A. Latone, and J. M. Garcia, Quantum computation of dominant products in lithium–sulfur batteries, *J. Chem. Phys.* **154**, 134115 (2021).
- [7] L. Clinton, T. Cubitt, B. Flynn, F. M. Gambetta, J. Klassen, A. Montanaro, S. Piddock, R. A. Santos, and E. Sheridan, Towards near-term quantum simulation of materials, *Nat. Commun.* **15**, 211 (2024).
- [8] A. Delgado, P. A. M. Casares, R. dos Reis, M. S. Zini, R. Campos, N. Cruz-Hernández, A.-C. Voigt, A. Lowe, S. Jahangiri, M. A. Martin-Delgado, J. E. Mueller, and J. M. Arrazola, Simulating key properties of lithium-ion batteries with a fault-tolerant quantum computer, *Phys. Rev. A* **106**, 032428 (2022).
- [9] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, *Chem. Rev.* **120**, 12685 (2020).
- [10] A. Y. Kitaev, arXiv preprint [arXiv:quant-ph/9511026](https://arxiv.org/abs/quant-ph/9511026).
- [11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge university press, Cambridge, 2010).
- [12] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, *Proc. R. Soc. London. Ser. A: Math., Phys. Eng. Sci.* **454**, 339 (1998).
- [13] E. Knill, G. Ortiz, and R. D. Somma, Optimal quantum measurements of expectation values of observables, *Phys. Rev. A* **75**, 012328 (2007).
- [14] D. Poulin and P. Wocjan, Sampling from the thermal quantum Gibbs state and evaluating partition functions with a quantum computer, *Phys. Rev. Lett.* **103**, 220502 (2009).
- [15] Y. Ge, J. Tura, and J. I. Cirac, Faster ground state preparation and high-precision ground energy estimation with fewer qubits, *J. Math. Phys.* **60**, 022202 (2019).
- [16] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, *PRX Quantum* **3**, 010318 (2022).
- [17] L. Lin and Y. Tong, Near-optimal ground state preparation, *Quantum* **4**, 372 (2020).
- [18] N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, D. Hait, D. S. Levine, W. Huggins, Z. Jiang, J. R. McClean, R. Babbush, M. Head-Gordon *et al.*, arXiv preprint [arXiv:1809.05523](https://arxiv.org/abs/1809.05523).

- [19] L. Veis, J. Višňák, T. Fleig, S. Knecht, T. Saue, L. Visscher, and J. Pittner, Relativistic quantum chemistry on quantum computers, *Phys. Rev. A* **85**, 030304(R) (2012).
- [20] W. Kohn, Nobel Lecture: Electronic structure of matter—wave functions and density functionals, *Rev. Mod. Phys.* **71**, 1253 (1999).
- [21] H. Wang, S. Kais, A. Aspuru-Guzik, and M. R. Hoffmann, Quantum algorithm for obtaining the energy spectrum of molecular systems, *Phys. Chem. Chem. Phys.* **10**, 5388 (2008).
- [22] L. Veis and J. Pittner, Quantum computing applied to calculations of molecular energies: CH<sub>2</sub> benchmark, *J. Chem. Phys.* **133**, 194106 (2010).
- [23] L. Veis and J. Pittner, Adiabatic state preparation study of methylene, *J. Chem. Phys.* **140**, 214111 (2014).
- [24] K. Sugisaki, S. Yamamoto, S. Nakazawa, K. Toyota, K. Sato, D. Shiomi, and T. Takui, Quantum chemistry on quantum computers: A polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules, *J. Phys. Chem. A* **120**, 6459 (2016).
- [25] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu *et al.*, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, *Nat. Commun.* **14**, 1952 (2023).
- [26] G. Ortiz, J. E. Gubernatis, E. Knill, and R. Laflamme, Quantum algorithms for fermionic simulations, *Phys. Rev. A* **64**, 022319 (2001).
- [27] H. Wang, S. Ashhab, and F. Nori, Efficient quantum algorithm for preparing molecular-system-like states on a quantum computer, *Phys. Rev. A* **79**, 042335 (2009).
- [28] R. Babbush, J. McClean, D. Wecker, A. Aspuru-Guzik, and N. Wiebe, Chemical basis of Trotter-Suzuki errors in quantum chemistry simulation, *Phys. Rev. A* **91**, 022311 (2015).
- [29] S. Bravyi, D. Gosset, R. König, and K. Temme, Approximation algorithms for quantum many-body problems, *J. Math. Phys.* **60**, 032203 (2019).
- [30] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, Sequential generation of entangled multiqubit states, *Phys. Rev. Lett.* **95**, 110503 (2005).
- [31] D. Malz, G. Styliaris, Z.-Y. Wei, and J. I. Cirac, Preparation of matrix product states with log-depth quantum circuits, *Phys. Rev. Lett.* **132**, 040404 (2024).
- [32] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem, *Science* **292**, 472 (2001).
- [33] E. Farhi, J. Goldstone, and S. Gutmann, arXiv preprint [arXiv:quant-ph/0007071](https://arxiv.org/abs/quant-ph/0007071).
- [34] L. A. Wu, M. S. Byrd, and D. A. Lidar, Polynomial-time simulation of pairing models on a quantum computer, *Phys. Rev. Lett.* **89**, 057904 (2002).
- [35] T. Albash and D. A. Lidar, Adiabatic quantum computation, *Rev. Mod. Phys.* **90**, 015002 (2018).
- [36] M.-H. Yung, J. Casanova, A. Mezzacapo, J. McClean, L. Lamata, A. Aspuru-Guzik, and E. Solano, From transistor to trapped-ion computers for quantum chemistry, *Sci. Rep.* **4**, 3589 (2014).
- [37] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [38] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, *Quantum Sci. Technol.* **4**, 014008 (2018).
- [39] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer, *Quantum Sci. Technol.* **4**, 045005 (2019).
- [40] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nat. Commun.* **10**, 3007 (2019).
- [41] M. R. Hirsbrunner, D. Chamaki, J. W. Mullinax, and N. M. Tubman, arXiv preprint [arXiv:2301.05666](https://arxiv.org/abs/2301.05666).
- [42] U. Baek, D. Hait, J. Shee, O. Leimkuhler, W. J. Huggins, T. F. Stetina, M. Head-Gordon, and K. B. Whaley, Say NO to optimization: A nonorthogonal quantum eigensolver, *PRX Quantum* **4**, 030307 (2023).
- [43] M. Motta, C. Sun, A. T. Tan, M. J. O’Rourke, E. Ye, A. J. Minnich, F. G. Brandao, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, *Nat. Phys.* **16**, 205 (2020).
- [44] N. Gomes, F. Zhang, N. F. Berthussen, C.-Z. Wang, K.-M. Ho, P. P. Orth, and Y. Yao, Efficient step-merged quantum imaginary time evolution algorithm for quantum chemistry, *J. Chem. Theory. Comput.* **16**, 6256 (2020).
- [45] H. Kamakari, S.-N. Sun, M. Motta, and A. J. Minnich, Digital quantum simulation of open quantum systems using quantum imaginary-time evolution, *PRX Quantum* **3**, 010320 (2022).
- [46] K. Hejazi, M. Motta, and G. K.-L. Chan, Adiabatic quantum imaginary time evolution, *Phys. Rev. Res.* **6**, 033084 (2024).
- [47] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [48] S. R. White, Density-matrix algorithms for quantum renormalization groups, *Phys. Rev. B* **48**, 10345 (1993).
- [49] V. Veryazov, P. Å. Malmqvist, and B. O. Roos, How to select active space for multiconfigurational quantum chemistry?, *Int. J. Quantum Chem.* **111**, 3329 (2011).
- [50] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, Automated construction of molecular active spaces from atomic valence orbitals, *J. Chem. Theory Comput.* **13**, 4063 (2017).
- [51] C. J. Stein and M. Reiher, autoCAS: A program for fully automated multiconfigurational calculations, *J. Comput. Chem.* **40**, 2216 (2019).
- [52] N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, A deterministic alternative to the full configuration interaction quantum Monte Carlo method, *J. Chem. Phys.* **145**, 044112 (2016).
- [53] J. Schriber and F. Evangelista, Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy, *J. Chem. Phys.* **144**, 161106 (2016).
- [54] A. A. Holmes, N. M. Tubman, and C. Umrigar, Heat-bath configuration interaction: An efficient selected

- configuration interaction algorithm inspired by heat-bath sampling, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- [55] R. J. Harrison, Approximating full configuration interaction with selected configuration interaction and perturbation theory, *J. Chem. Phys.* **94**, 5021 (1991).
- [56] N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling CI method, *J. Chem. Theory Comput.* **16**, 2139 (2020).
- [57] C. F. Bender and E. R. Davidson, Studies in configuration interaction: The first-row diatomic hydrides, *Phys. Rev.* **183**, 23 (1969).
- [58] B. Huron, J. Malrieu, and P. Rancurel, Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions, *J. Chem. Phys.* **58**, 5745 (1973).
- [59] E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff *et al.*, Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package, *J. Chem. Phys.* **155**, 084801 (2021).
- [60] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory, *J. Chem. Theory Comput.* **13**, 1595 (2017).
- [61] S. R. White and R. L. Martin, Ab initio quantum chemistry using the density matrix renormalization group, *J. Chem. Phys.* **110**, 4127 (1999).
- [62] K. H. Marti, I. M. Ondík, G. Moritz, and M. Reiher, Density matrix renormalization group calculations on relative energies of transition metal complexes and clusters, *J. Chem. Phys.* **128**, 014104 (2008).
- [63] Y. Kurashige, G. K.-L. Chan, and T. Yanai, Entangled quantum electronic wavefunctions of the  $\text{Mn}_4\text{CaO}_5$  cluster in photosystem II, *Nat. Chem.* **5**, 660 (2013).
- [64] S. Sharma, K. Sivalingam, F. Neese, and G. K.-L. Chan, Low-energy spectrum of iron–sulfur clusters directly from many-particle quantum mechanics, *Nat. Chem.* **6**, 927 (2014).
- [65] R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. Chan, The ab-initio density matrix renormalization group in practice, *J. Chem. Phys.* **142**, 034102 (2015).
- [66] H. R. Larsson, H. Zhai, C. J. Umrigar, and G. K.-L. Chan, The chromium dimer: Closing a chapter of quantum chemistry, *J. Am. Chem. Soc.* **144**, 15932 (2022).
- [67] G. K.-L. Chan, J. J. Dorando, D. Ghosh, J. Hachmann, E. Neuscamman, H. Wang, and T. Yanai, *Frontiers in Quantum Systems in Chemistry and Physics* (Springer, 2008), Chap. An introduction to the density matrix renormalization group ansatz in quantum chemistry, p. 49.
- [68] G. K.-L. Chan and S. Sharma, The density matrix renormalization group in quantum chemistry, *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
- [69] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys. (N. Y.)* **326**, 96 (2011).
- [70] S. Fomichev, K. Hejazi, J. Fraxanet, and J. M. Arrazola, “Overlapper,” <https://github.com/XanaduAI/Overlapper/> (2024).
- [71] G. H. Low, V. Kliuchnikov, and L. Schaeffer, Trading T gates for dirty qubits in state preparation and unitary synthesis, *Quantum* **8**, 1375 (2024).
- [72] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, *Quantum* **3**, 208 (2019).
- [73] M. S. Zini, A. Delgado, R. dos Reis, P. A. M. Casares, J. E. Mueller, A.-C. Voigt, and J. M. Arrazola, Quantum simulation of battery materials using ionic pseudopotentials, *Quantum* **7**, 1049 (2023).
- [74] S.-J. Ran, Encoding of matrix product states into quantum circuits of one- and two-qubit gates, *Phys. Rev. A* **101**, 032310 (2020).
- [75] M. S. Rudolph, J. Chen, J. Miller, A. Acharya, and A. Perdomo-Ortiz, Decomposition of matrix product states into shallow quantum circuits, *Quantum Sci. Technol.* **9**, 015012 (2023).
- [76] M. B. Dov, D. Shnaiderov, A. Makmal, and E. G. D. Torre, Approximate encoding of quantum states using shallow circuits, *npj Quantum Inf.* **10**, 65 (2024).
- [77] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, *PRX Quantum* **2**, 030305 (2021).
- [78] J. J. Goings, A. White, J. Lee, C. S. Tautermann, M. Degroote, C. Gidney, T. Shiozaki, R. Babbush, and N. C. Rubin, Reliably assessing the electronic structure of cytochrome P450 on today’s classical computers and tomorrow’s quantum computers, *Proc. Natl. Acad. Sci.* **119**, e2203533119 (2022).
- [79] I. H. Kim, Y.-H. Liu, S. Pallister, W. Pol, S. Roberts, and E. Lee, Fault-tolerant resource estimate for quantum chemical simulations: Case study on Li-ion battery electrolyte molecules, *Phys. Rev. Res.* **4**, 023019 (2022).
- [80] N. C. Rubin, D. W. Berry, F. D. Malone, A. F. White, T. Khattar, A. E. DePrince III, S. Siculo, M. Kühn, M. Kaicher, J. Lee *et al.*, Fault-tolerant quantum simulation of materials using Bloch orbitals, *PRX Quantum* **4**, 040303 (2023).
- [81] S. Blinnikov and R. Moessner, Expansions for nearly Gaussian distributions, *Astron. Astrophys. Suppl. Ser.* **130**, 193 (1998).
- [82] C. Mejuto-Zaera, G. Weng, M. Romanova, S. J. Cotton, K. B. Whaley, N. M. Tubman, and V. Vlček, Are multi-quasiparticle interactions important in molecular ionization?, *J. Chem. Phys.* **154**, 193205 (2021).
- [83] E. Jeckelmann, Dynamical density-matrix renormalization-group method, *Phys. Rev. B* **66**, 045114 (2002).
- [84] E. Ronca, Z. Li, C. A. Jimenez-Hoyos, and G. K.-L. Chan, Time-step targeting time-dependent and dynamical density matrix renormalization group algorithms with ab initio hamiltonians, *J. Chem. Theory Comput.* **13**, 5560 (2017).
- [85] H. Zhai and G. K.-L. Chan, Low communication high performance ab initio density matrix renormalization group algorithms, *J. Chem. Phys.* **154**, 224116 (2021).



- [86] H. Zhai *et al.*, “Block2,” <https://github.com/block-hczhai/block2-preview> (2023).
- [87] H. Zhai, H. R. Larsson, S. Lee, Z.-H. Cui, T. Zhu, C. Sun, L. Peng, R. Peng, K. Liao, J. Tölle *et al.*, Block2: A comprehensive open source framework to develop and apply state-of-the-art DMRG algorithms in electronic structure and beyond, *J. Chem. Phys.* **159**, 234801 (2023).
- [88] M. Wand and M. Jones, in *Chapman & Hall/CRC Monographs on Statistics & Applied Probability* (60) (Chapman & Hall/CRC, New York, 1994).
- [89] A. Gramacki, *Nonparametric Kernel Density Estimation and its Computational Aspects* (Springer, Cham (Switzerland), 2018), Vol. 37.
- [90] S. Fomichev, K. Hejazi, I. Loaiza, M. S. Zini, A. Delgado, A.-C. Voigt, J. E. Mueller, and J. M. Arrazola, “Simulating x-ray absorption spectroscopy of battery materials on a quantum computer,” [arXiv:2405.11015](https://arxiv.org/abs/2405.11015).
- [91] D. W. Berry, M. Kieferová, A. Scherer, Y. R. Sanders, G. H. Low, N. Wiebe, C. Gidney, and R. Babbush, Improved techniques for preparing eigenstates of fermionic Hamiltonians, *npj Quantum Inf.* **4**, 22 (2018).
- [92] Y. Dong, L. Lin, and Y. Tong, Ground-state preparation and energy estimation on early fault-tolerant quantum computers via quantum eigenvalue transformation of unitary matrices, *PRX Quantum* **3**, 040305 (2022).
- [93] G. H. Low and I. L. Chuang, Optimal hamiltonian simulation by quantum signal processing, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [94] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand unification of quantum algorithms, *PRX Quantum* **2**, 040203 (2021).
- [95] G. Wang, S. Sim, and P. D. Johnson, State preparation boosters for early fault-tolerant quantum computation, *Quantum* **6**, 829 (2022).
- [96] G. Wang, D. S. França, R. Zhang, S. Zhu, and P. D. Johnson, Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision, *Quantum* **7**, 1167 (2023).
- [97] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma *et al.*, PySCF: The Python-based simulations of chemistry framework, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1340 (2018).
- [98] S. Sharma *et al.*, “Dice,” <https://github.com/sanshar/Dice> (2023).
- [99] J. Hachmann, W. Cardoen, and G. K. Chan, Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group, *J. Chem. Phys.* **125**, 144101 (2006).
- [100] W. A. Al-Saidi, S. Zhang, and H. Krakauer, Bond breaking with auxiliary-field quantum Monte Carlo, *J. Chem. Phys.* **127**, 144101 (2007).
- [101] T. Tsuchimochi and G. E. Scuseria, Strong correlations via constrained-pairing mean-field theory, *J. Chem. Phys.* **131**, 121102 (2009).
- [102] A. V. Sinitskiy, L. Greenman, and D. A. Mazziotti, Strong correlation in hydrogen chains and lattices using the variational two-electron reduced density matrix method, *J. Chem. Phys.* **133**, 014104 (2010).
- [103] D. A. Mazziotti, Large-scale semidefinite programming for many-electron quantum mechanics, *Phys. Rev. Lett.* **106**, 083001 (2011).
- [104] N. Lin, C. Marianetti, A. J. Millis, and D. R. Reichman, Dynamical mean-field theory for quantum chemistry, *Phys. Rev. Lett.* **106**, 096402 (2011).
- [105] L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, Strong electronic correlation in the hydrogen chain: A variational Monte Carlo study, *Phys. Rev. B* **84**, 245117 (2011).
- [106] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma *et al.*, Towards the solution of the many-electron problem in real materials: Equation of state of the hydrogen chain with state-of-the-art many-body methods, *Phys. Rev. X* **7**, 031059 (2017).
- [107] N. H. Stair and F. A. Evangelista, Exploring Hilbert space on a budget: Novel benchmark set and performance metric for testing electronic structure methods in the regime of strong correlation, *J. Chem. Phys.* **153**, 104108 (2020).
- [108] M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis *et al.*, Ground-state properties of the hydrogen chain: Dimerization, insulator-to-metal transition, and magnetic phases, *Phys. Rev. X* **10**, 031058 (2020).
- [109] S. Lee, H. Zhai, S. Sharma, C. J. Umrigar, and G. K.-L. Chan, Externally corrected CCSD with renormalized perturbative triples (R-ecCCSD(T)) and the density matrix renormalization group and selected configuration interaction external sources, *J. Chem. Theory Comput.* **17**, 3414 (2021).
- [110] M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitos, A new generation of effective core potentials for correlated calculations, *J. Chem. Phys.* **147**, 224106 (2017).
- [111] “Pseudopotential library,” <https://pseudopotentiallibrary.org/>, accessed: 2022-10-01.
- [112] A. W. Prentice, J. P. Coe, and M. J. Paterson, Modular approach to selected configuration interaction in an arbitrary spin basis: Implementation and comparison of approaches, *J. Chem. Theory Comput.* **19**, 9161 (2023).
- [113] C. Gidney, Halving the cost of quantum addition, *Quantum* **2**, 74 (2018).
- [114] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, *Phys. Rev. X* **8**, 041015 (2018).
- [115] G. H. Low and I. L. Chuang, [arXiv preprint arXiv:1707.05391](https://arxiv.org/abs/1707.05391).