High Ground State Overlap via Quantum Embedding Methods

Mihael Erakovic [1,*] Freek Witteveen [2,†] Dylan Harley [2] Jakob Günther [2] Moritz Bensberg [1]
Oinam Romesh Meitei [3] Minsik Cho [3] Troy Van Voorhis [3] Markus Reiher [1,‡] and Matthias Christandl [2,§]
[1]*ETH Zürich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland*
[2]*Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*
[3]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 4 September 2024; accepted 16 December 2024; published 14 January 2025)

Quantum computers can accurately compute ground state energies using phase estimation, but this requires a guiding state that has significant overlap with the true ground state. For large molecules and extended materials, it becomes difficult to find guiding states with good ground state overlap for growing molecule sizes. Additionally, the required number of qubits and quantum gates may become prohibitively large. One approach for dealing with these challenges is to use a quantum embedding method, which allows a reduction to one or multiple smaller quantum cores embedded in a larger quantum region. In such situations, it is unclear how the embedding method affects the hardness of constructing good guiding states. In this work, we therefore investigate the preparation of guiding states in the context of quantum embedding methods. We extend previous work on quantum impurity problems, a framework in which we can rigorously analyze the embedding of a subset of orbitals. While there exist results for optimal active orbital space selection in terms of energy minimization, we rigorously demonstrate how the same principles can be used to define selected orbital spaces for state preparation in terms of the overlap with the ground state. Moreover, we perform numerical studies of molecular systems relevant to biochemistry, one field in which quantum embedding methods are required due to the large size of biomacromolecules such as proteins and nucleic acids. We investigate two different embedding strategies which can exhibit qualitatively different orbital entanglement. In all cases, we demonstrate that the easy-to-obtain mean-field state will have a sufficiently high overlap with the target state to perform quantum phase estimation.

DOI: 10.1103/PRXLife.3.013003

I. INTRODUCTION

Some of the most promising applications for quantum computers arguably lie in their utility for molecular and materials science. These fields face important computational challenges, in both academic and industrial settings. Prominent examples can be found in catalysis, battery and drug design [1–4], and biochemistry [5,6]. True quantum advantage will likely only emerge in the fault-tolerant regime, where quantum phase estimation (QPE) algorithms will allow for energy calculations of a quantum system (i.e., a molecule or material) with controlled and guaranteed accuracy.

One of the central problems of quantum chemistry is the computation of ground-state electronic energies. We consider a system with a fixed number n of electrons and a discretization of the electronic structure problem in the Born-Oppenheimer approximation into N spin-orbitals, giving a Hamiltonian of the form

$$H = \sum_{pq=1}^{N} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs=1}^{N} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q,$$

where a_p for $p = 1, \ldots, N$ are the fermionic annihilation operators for the N spin-orbitals, and $g_{pqrs} = (pq|rs)$ are the two-body integrals; see, for example, Ref. [7] for details. A system with N spin-orbitals and n electrons has a Hilbert space of dimension $\binom{N}{n}$ which is exponential in N if n scales linearly with N. High accuracy classical methods struggle with the size of this Hilbert space. The key advantage of quantum computing is that the quantum computer can natively represent states in the Hilbert space using only N qubits. When considering the scaling with system size, one should keep in mind that at fixed electron number n, but an increasing number of orbitals N (the continuum limit), the Hilbert space dimension increases only as poly(N) (polynomially in N), albeit with an in practice prohibitive exponent n.

There are two major challenges in quantum computing for chemistry and many-body physics:

(i) The first challenge is the *orthogonality catastrophe*. Quantum phase estimation requires an initial *guiding state* which has sufficient overlap with the ground state. However, for systems with large N and scaling electron number, small local errors in the guiding state lead to an exponential decay in the overlap with the global ground state. As a result, good guiding states become hard to find; see Appendix A.

*Contact author: mihael.erakovic@phys.chem.ethz.ch
†Contact author: fw@math.ku.dk
‡Contact author: mreiher@ethz.ch
§Contact author: christandl@math.ku.dk

(ii) The second challenge is to perform Hamiltonian simulation for a time $O(\varepsilon^{-1})$ to achieve quantum phase estimation precision ε, and to reduce the dependence of the computational cost on N.

The first challenge is a fundamental and *theoretical* obstruction to computing ground state energies efficiently on a quantum computer. However, for many molecules, conventional computational methods are observed to find good guiding states in practice (though this may be different for strongly correlated materials with a lattice structure), so for practical applications to quantum chemistry there is evidence that this is not the main problem [8,9], and our work supports this claim.

For the second challenge, phase estimation based on Trotterization of the time evolution unitary operator scales as $O(N^4)$ for the full electronic Hamiltonian because of the two-electron interaction terms in its second-quantized form. However, truncation strategies can be employed to reduce the number of relevant parameters and significantly reduce the scaling of the phase estimation. In the case of Trotterization, the scaling can be reduced to $O(N^3)$ for a given molecule and increasing basis set size and to $O(N^2 \log N)$ for increasing molecule size [11]. Alternatively, qubitization allows for the factorization of the Hamiltonian, which can reduce the scaling down to $O(N)$ [2] . However, this comes at the cost of a large number of ancillary qubits, and the overall scaling of the phase estimation also depends on a normalization factor λ which scales with N. Hence, the scaling can lead to quantum circuits with a prohibitively large gate count in practice, even for different truncation and factorization strategies, and it represents a serious bottleneck for useful quantum computations. Specifically, from state-of-the-art resource estimates it appears that, with fault-tolerant devices, computations for 100–200 spin orbitals already require very extensive quantum resources [2,3,10,12], on the order of 10^{10} Toffoli gates [13] (with the caveat that it is difficult to predict the hardware specifications of future fault-tolerant quantum computers).

Given a choice of orbital basis, the electronic ground state can be written as

$$|\Psi\rangle = \sum_i C_i |\Phi_i\rangle, \quad \sum_i |C_i|^2 = 1, \tag{1}$$

where the sum runs over Slater determinants $|\Phi_i\rangle$. For example, one can choose Hartree-Fock (HF) orbitals. In that case, the expansion in Eq. (1) consists of the HF state $|\Phi_0\rangle$ and corrections to it, which represent electron correlations. One can also choose a different orbital basis, and this may lead to a significant increase in the largest weight on a single determinant. If the HF basis state qualitatively misrepresents the ground state, and there are more determinants with weights of the same order of magnitude, this is known as *static correlation*. The remaining determinants with small weights then account for the so-called *dynamical correlation*. While useful notions, there is no sharp distinction between dynamical and static correlation.

The question of finding and preparing good guiding states is closely related to the computational complexity of the electronic structure problem. The problem of finding ground state energies of local Hamiltonians in full generality is QMA-hard [14,15], so this task is believed to be intractable even for quantum computers. In particular, this implies that it is hard to prepare guiding states with nontrivial ground state overlap—however, numerical evidence suggests that finding good guiding states is feasible for many realistic not-too-large chemical systems [8,16]. This is the case if the correlation is of a dynamical nature, and also in systems with a static correlation; a small number of determinants [9], choosing a different orbital basis (see, for instance, Ref. [17]), or symmetry-respecting configuration state functions [18] may suffice to reach good ground state overlap and hence yield efficient quantum phase estimation algorithms. For quantum phase estimation, a constant (or even inverse polynomially small) ground state overlap suffices for efficient ground state energy computation. For the total cost, it does not make a significant difference whether one has a constant ground state overlap, or even ground state overlap close to 1. However, if the guiding state has ground state overlap close to 1, this allows the possibility to parallelize the phase estimation into multiple circuits of shorter depth [19,20], as we briefly review in Appendix A, and which may be particularly useful for early fault-tolerant quantum computers.

In many large molecules where static quantum correlations are important, so that a mean-field treatment does not suffice to represent the target state even qualitatively well, these correlations can be assigned to a relatively small number of orbitals. If these orbitals are localized in three-dimensional space, static correlation will be of a short-ranged nature. In such situations, both challenges (ground state overlap and Hamiltonian simulation scaling with N) can be addressed by using an appropriate *embedding method*, which singles out a spatial region (or lengthscale) that is treated fully quantum mechanically, and an environment (or longer range correlations) which can be treated with an approximate electronic structure model such as (mean-field) HF or some other electronic structure model of low computational complexity such as (variants of) Kohn-Sham density functional theory (DFT). By focusing on a smaller embedded quantum subsystem (denoted as a quantum core in the following), the system to which one applies phase estimation is sufficiently small so that the exponential scaling of the orthogonality catastrophe has not yet kicked in, and one can still find high-overlap guiding states using classical methods. Additionally, the smaller number of orbitals leads to a reduced gate count in the phase estimation quantum circuit. For very large molecules (such as proteins, protein complexes, or interacting biomolecules in biochemistry), embedding strategies can be of a multilevel nature, where the description of the interatomic interactions ranges from classical force fields to different quantum-mechanical approximations. We provide a brief introduction to relevant aspects of quantum embedding theory in Sec. II.

In this work, we focus on quantum-in-quantum embedding schemes, where a large system is described by an approximate quantum-mechanical model (such as HF or Kohn-Sham DFT), while one or more subsystems embedded into this large system define the quantum cores. The energy contribution of the quantum cores to the full potential energy surface of the large system can be rigorously obtained by a (future) fault-tolerant quantum computer using a representation of the full wave function in the restricted orbital space of the quantum cores.

It is for this reason that we consider state preparation in the context of quantum embedding approaches.

While wave-function-based approaches in traditional computing (such as coupled cluster theory or multiconfigurational approaches with multireference perturbation theory) can deliver accurate quantum core energies, an important drawback is that precise and controllable error bounds are not known for any of these methods [21]. By contrast, quantum computation based on phase estimation can deliver total electronic energies to a given precision (that is, typically chemical accuracy between 1 and 0.1 mHartree, which is sufficient to ensure that relative energies are sufficiently accurate for the evaluation of valence-shell properties such as relative energies of molecular structures or rate constants)—provided that an initial guiding state can be efficiently prepared which will have a large overlap with the target state that is to be determined. We note that the growth of the absolute electronic energy with molecular size is due to the low-lying core shell orbitals, which do not contribute to such valence shell properties (apart from the fact that an embedding that restricts the orbital space would not allow for arbitrary growth of the electronic energy). We emphasize that rigorous error estimates will be a key advantage over traditional approaches [1,2,22], apart from the fact that a quantum computer with a sufficiently large number of qubits for the representation of an electronic state will tame the curse of dimensionality posed to traditional approaches.

The use of embedding methods for quantum computing in quantum chemistry raises important questions:

(i) What is the interplay between an embedding method and the *guiding state*? For example, if the choice of the embedding method affects the type of correlation on the resulting orbital space in a quantum core, then this may have consequences for the character of the guiding state up to the point where it might be difficult to determine.

(ii) What is the *computational complexity* of problems with localized quantum correlations?

The second question concerns the general *computational complexity* of problems with localized quantum correlations. In Sec. III, we study this question theoretically in the context of *quantum impurity models*, specifically by extending the work of [23] on the nature of *quantum impurity problems* to shed light on these questions. A quantum impurity model is a system with a scaling number of N orbitals, of which only a constant number M participate in two-body interactions. Such models may be considered a useful description for systems where electron correlation is localized to a small subset of orbitals. Based on [23], we show that there exist good quantum embeddings for such systems. Based on this, we propose an approach to the guiding state problem where one uses an embedding method to find a guiding state that is an arbitrary state on the quantum core, and a Slater determinant on the environment. We show that if one chooses a good embedding or if the one-body Hamiltonian is gapped, this guiding state has at least an inverse polynomial overlap with the true ground state of the embedding Hamiltonian, giving a polynomial quantum algorithm for the ground state energy. This is in contrast to the best-known rigorous classical algorithms, which in these cases have a quasipolynomial scaling in the precision [23].

These results motivate a general strategy for preparing guiding states for problems with localized quantum correlations on large-scale fault-tolerant quantum computers. We propose that one searches for guiding states through quantum-in-quantum embeddings.

Here, one identifies an active orbital space A and an environment C, and takes as a guiding state a state of the form $|\Phi_A\rangle \wedge |\Psi_C\rangle$, where $|\Phi_A\rangle$ is the solution to the (small) active space problem, $|\Psi_C\rangle$ is a Slater determinant on the environment, and $\wedge$ denotes the antisymmetrized product. A mean-field calculation is used to obtain the wave function on the environment, with which the embedded Hamiltonian can be constructed. An approximation of the ground state energy and wave function can then be calculated using one of the classical approaches. Tensor network-based methods, such as the density matrix renormalization group (DMRG), stand out in particular as they provide an adequate description even in the case of strongly correlated embedded systems. These approaches, however, introduce errors in the energy due to approximations (e.g., limited bond dimension in the case of DMRG), but may provide a good approximation to the wave function. Quantum phase estimation then allows for accurate computation of the ground state energy using the classically calculated approximate wave function as a guiding state.

We demonstrate numerically in Sec. IV that for two conceptually different embedding methods, one can find low-complexity guiding states, even though the two different methods produce rather different orbital bases. For this demonstration, we focus on small biomolecules (tryptophan and its oligopeptide structures), which occur as monomers and oligomers in biomacromolecules (proteins), and on a transition-metal drug that can form a host-guest complex with a protein. Such complexes are key to molecular recognition processes. While we focus on biomolecules, our approach is similarly relevant in other application areas involving larger molecules or materials.

II. QUANTUM EMBEDDING METHODS

To make computations feasible for large systems, for which physical or chemical properties are governed by a relatively small spatial region within the larger system, an established approach (see, for instance, Refs. [24,25]) is to treat different subsystems or lengthscales at different levels of accuracy. For example, in the case of molecular recognition, it may be sufficient to focus on the interaction seam of host and guest molecules and, hence, to take into account electron correlation accurately only in the binding region. *Embedding theory* deals with approaches to embed approximations of different levels of accuracy into one another; see Refs. [26–31] for reviews of quantum embedding theory. In general, one decomposes the system into a *quantum core*, where one uses a *high-level* electronic structure model, which takes into account the more complex nature of the wave function, and a (quantum or classical) environment. If feasible, the environment can be treated using a *low-level* model such as HF or DFT. It is possible to have multiple quantum cores and multiple embedding layers, treating parts of the system at the level of classical force fields, semiempirical methods, or DFT at different levels of

accuracy. To take the effect of the environment into account, an embedding scheme requires the construction of an effective Hamiltonian on the quantum cores, given the solution of the low-level solver. Given the result of minimizing this Hamiltonian on the quantum cores, the embedding scheme should provide a prescription for a global energy estimate.

We will discuss two different embedding schemes, which we compare in detail with molecular examples in Sec. IV. The first approach is based on density matrices. This is natural for quantum computing, since quantum computation naturally interfaces with quantum states. For example, in the bootstrap embedding method (a particular density matrix-based embedding scheme), one can enforce the consistency conditions between overlapping fragments on the quantum computer [32]. As an alternative, we also consider projection-based embedding methods. These have the advantage of being naturally compatible with DFT, which may be useful for multiscale approaches.

A. Embedding based on density matrices

A natural approach toward embedding emerges from the framework of open-system quantum theory and is based on density matrices; an example is density matrix embedding theory (DMET) [33–36], see [37] for a precise mathematical description. Here we consider a system with N orbitals, where we identify subsets $A_1, \ldots, A_n$ of these orbitals, which may be called "fragments" or "impurities." We let M_i be the number of orbitals in A_i. Depending on the method, these fragments may or may not overlap. Whereas the formal basis of DMET is the Schmidt decomposition, its key practical idea is to combine a high-level solver on the fragments with a mean-field solution on the global level and enforce consistency conditions. We start with a global state given by a single Slater determinant $|\Psi\rangle$, which can be obtained as the Hartree-Fock solution by the low-level solver. For each of the subsystems A_i we can perform a Schmidt decomposition, which for each i gives a decomposition of the full N-dimensional single-particle space into A_i (the fragment orbitals), a "bath" B_i for A_i of size at most M_i (the bath orbitals) and its complement C_i, such that the state

$$|\Psi\rangle = |\Psi_{A_iB_i}\rangle \wedge |\Psi_{C_i}\rangle$$

is decomposed into Slater determinants $|\Psi_{A_iB_i}\rangle$ and $|\Psi_{C_i}\rangle$ on A_iB_i and C_i, respectively. Now, one may look for wave functions of the form $|\Psi_i\rangle = |\Phi_{A_iB_i}\rangle \wedge |\Psi_{C_i}\rangle$, where $|\Phi_{A_iB_i}\rangle$ is now an arbitrary fermionic state on A_iB_i, determined by a high-precision solver (such as a quantum computer). The high-level solver minimizes the energy over states of this form. Since we have a fixed Slater determinant $|\Psi_{C_i}\rangle$ on C_i, this gives an effective Hamiltonian on A_iB_i,

$$H_{A_iB_i} = (I \otimes \langle\Psi_{C_i}|)H(I \otimes |\Psi_{C_i}\rangle).$$

This Hamiltonian is such that

$$\min_{|\Psi_i\rangle} \langle\Psi_i|H|\Psi_i\rangle = \min_{|\Phi_{A_iB_i}\rangle} \langle\Phi_{A_iB_i}|H_{A_iB_i}|\Phi_{A_iB_i}\rangle,$$

where the minimization on the left-hand side is over states of the form

$$|\Psi_i\rangle = |\Phi_{A_iB_i}\rangle \wedge |\Psi_{C_i}\rangle.$$

There are different schemes for estimating a global energy based on density matrix embedding. The standard approach is to try to achieve self-consistency between the global Slater determinant and the fragment wave functions. Here, one takes nonoverlapping fragments and searches for a solution where the Slater determinant $|\Psi\rangle$ has the same 1-RDMs on the fragments A_i as the fragment solutions $|\Phi_{A_iB_i}\rangle$. This can be obtained by iteratively solving for the fragments and updating them to a global mean-field solution with constrained 1-RDMs. A different approach is the *bootstrap embedding*, where one takes overlapping fragments and iteratively tries to enforce consistency between the fragments on their overlap instead of with a global state; see [38–40] for details. However, we note that in both cases, there is no guarantee of the existence of a *global* wave function compatible with these schemes, and the results are not variational. This means that the obtained value is not necessarily a strict upper bound on the true ground state energy.

B. Embedding based on projection

An alternative approach is given by *projection-based* embeddings [41,42]. Here, we use a method based on the Huzinaga equation [41,43]. We describe the case with a single quantum core. We start the theory discussion with HF as the low-level solver, highlighting the similarities to DMET. Later will we describe how DFT retrieves the missing electron correlation in Hartree-Fock, making the approach formally exact. The exactness of the approach means that when using the exact (unknown) exchange correlation functional for the DFT computation, and the exact wave-function solution on the embedded fragment, the result of following the embedding procedure equals the exact ground state energy. This embedding method starts from a HF (or from a Kohn-Sham density functional theory calculation), yielding a collection of molecular orbitals indexed by a set J and a single Slater determinant $|\Phi\rangle$ with the orbitals in $\mathcal{N} \subseteq J$ occupied. Given these orbitals ϕ_j, they are partitioned into two sets A and C of active and environment orbitals. We write $\mathcal{N}_A$ and $\mathcal{N}_C$ for the subsets of occupied orbitals in $|\Phi\rangle$. The division into A and C is often based on the localization of the orbitals on specific atoms in the molecule [41,42]. Alternatively, a unitary transformation in $\mathcal{N}$ may be transformed to enforce a localization on specific atoms [44], or an automated orbital analysis can be performed to identify orbitals contributing strongly to energy differences [45]. The idea of the projection-based embedding is now again to try to solve the minimization problem

$$\min_{|\Psi\rangle} \langle\Psi|H|\Psi\rangle, \quad |\Psi\rangle = |\Phi_A\rangle \wedge |\Psi_C\rangle,$$

where $|\Phi_A\rangle$ is an arbitrary state on A, and $|\Psi_C\rangle$ is the Slater determinant on the environment where the orbitals in $\mathcal{N}_C$ are occupied. Similarly to the embedding based on density matrices, the Slater determinant on the environment is kept fixed, which yields the effective Hamiltonian for the fragment,

$$\begin{aligned} H_A &= (I \otimes \langle\Psi_C|)H(I \otimes |\Psi_C\rangle) \\ &= \bar{E} + \sum_{p,q\in A} h_{pq}^{(A\text{in}C)} a_p^\dagger a_q + \frac{1}{2} \sum_{\substack{p,q \\ r,s\in A}} (pq|rs) a_p^\dagger a_r^\dagger a_s a_q, \end{aligned}$$

where

$$\bar{E} = 2\sum_{p\in\mathcal{N}_C} h_{pp} + \sum_{p,q\in\mathcal{N}_C} [2(pp|qq) - (pq|qp)]$$

is the mean-field energy of the environment, and

$$h_{pq}^{(\mathrm{AinC})} = h_{pq} + \sum_{c\in\mathcal{N}_C} [2(pq|cc) - (pc|qc)]$$

is the effective one-electron operator that includes the interaction between the fragment and the environment. In this case, both the environment energy and the fragment-environment interactions are captured on a mean-field level only, neglecting the correlation, while the overall energy is represented as $E = \bar{E} + E^{\mathrm{AinC}}$, where E^{AinC} is the energy of $h^{(\mathrm{AinC})}$. Note that the Hamiltonian H_A on the quantum core has as two-body terms precisely the two-body terms acting on A, and only the one-body term is modified.

The key idea of the projection-based embedding is to recover the missing correlation using DFT. For this purpose, densities γ from the initial DFT calculation are used to define the interaction terms between the fragment and the environment as

$$v_{\mathrm{emb}}[\gamma^A, \gamma^C] = g[\gamma^A + \gamma^C] - g[\gamma^A],$$

where g is defined by

$$(g[\gamma])_{pq} = \sum_{ij}\gamma_{ij}\left[(pq|ij) - \frac{1}{2}x(pi|qj)\right] + (v_{\mathrm{xc}}[\gamma])_{pq}.$$

Here, x is the fraction of the exact exchange in the exchange-correlation functional, and v_{xc} is the exchange-correlation potential matrix. This interaction term is used to redefine the effective one-electron operator as

$$\begin{aligned}\tilde{h}_{pq}^{(\mathrm{AinC})} &= h_{pq} + (v_{\mathrm{emb}}[\gamma^A, \gamma^C])_{pq} \\ &\quad - ((F^A - \mu)P^C - P^C(F^A - \mu))_{pq},\end{aligned}$$

where P^C is the matrix of the projection operator onto the subspace C, μ is a positive constant energy shift, and F^A is the Fock matrix for the embedded fragment. The last two terms are used to ensure that there is no mixing of the fragment and the environment orbitals. The overall energy of the system can be written as

$$\begin{aligned}E = &\langle\Psi_A|\tilde{H}_A|\Psi_A\rangle - \langle\Psi_A|\hat{v}_{\mathrm{xc}}[\gamma^A + \gamma^C] - \hat{v}_{\mathrm{xc}}[\gamma^C]|\Psi_A\rangle \\ &+ E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma^A + \gamma^C] - E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma^A] - E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma^C] \\ &+ E^{\mathrm{DFT}}[\gamma^C],\end{aligned}$$

where $E^{\mathrm{DFT}}[\gamma^C]$ is the Kohn-Sham DFT energy of C, $E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma]$ is the exchange-correlation functional evaluated for the density γ, and $\hat{v}_{\mathrm{xc}}[\gamma]$ is the operator belonging to v_{xc}. Note that $E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma]$ is typically nonlinear in γ, which means that it cannot be calculated as the expectation value of $\hat{v}_{\mathrm{xc}}[\gamma]$. Therefore, we formally subtracted the term $\langle\Psi_A|\hat{v}_{\mathrm{xc}}[\gamma^A + \gamma^C] - \hat{v}_{\mathrm{xc}}[\gamma^C]|\Psi_A\rangle$ and added the differences in $E_{\mathrm{xc}}^{\mathrm{DFT}}[\gamma]$.

C. Embedding methods and quantum computing

Using appropriate embedding methods will be crucial for the application of quantum computers to chemical systems, especially for biochemistry, nanochemistry, and materials science. Identifying the right quantum-mechanical subsystem of a molecule and using the quantum chemistry calculation on this fragment to derive conclusions about the full system is unavoidable for practical applications of quantum computation. The quantum computer then serves as the high-level solver in the embedding scheme.

Previous work using density matrix-based embedding schemes includes Refs. [32,46–49], approaches using projection-based embeddings include Refs. [50–52], and approaches based on Green's functions are presented in Refs. [53,54] (which we do not discuss in this work). These previous works have mostly focused on reducing the problem size, and running heuristic variational algorithms on (near-term) quantum computers. How the embedding method influences the problem of finding good guiding states, and thereby the prospects for phase estimation, has not been explored systematically in previous work to the best of our knowledge; one work in this direction is [5], which computes guiding state overlaps for a range of active space sizes for phase estimation for a protein system.

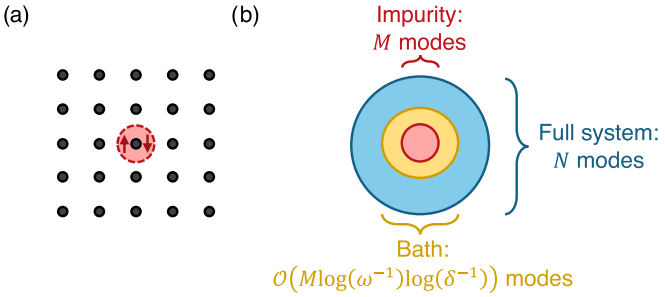

FIG. 1. (a) A lattice model for the electrons in a metal, with an impurity at one site. (b) Illustration of Theorem III 1: the ground state can be accurately approximated by a state which is a Slater determinant on the environment and some unconstrained state on the fragment and bath modes.

III. ORBITAL SELECTION FOR STATE PREPARATION IN QUANTUM IMPURITY PROBLEMS

In this section, we aim at deriving rigorous results for the state preparation problem in an embedding situation (see Fig. 1). To accomplish this, we exploit the framework of quantum impurity problems. A *quantum impurity model* is a fermionic Hamiltonian that has arbitrary one-body interactions, and two-body interactions on a constant-sized subset of orbitals. Quantum impurity models are relevant as models in material science where there is an impurity in the material. A famous example is the Anderson impurity model [55], which describes a magnetic impurity in a metal. This model gives an explanation for the Kondo effect and has been instrumental in the development of numerical renormalization group techniques [56]. In addition to the important application to impurities in metals, one can also consider them as a toy model for large molecules where mean-field approximations are accurate outside of a small interaction region. For example, the "impurity" may consist of the orbitals that are localized in the binding region. Note that in the quantum impurity model,

we assume that away from the impurity we have a one-body Hamiltonian (which is a different assumption from having an accurate mean-field approximation).

More formally, a quantum impurity model is defined by a fermionic Hamiltonian of the form $H_{\text{free}} + H_{\text{imp}}$, where H_{free} is a one-body Hamiltonian on N fermionic modes, whereas H_{imp} is an interacting fermionic Hamiltonian acting on a subset of M modes,

$$H_{\text{free}} = \sum_{pq=1}^{N} h_{pq} a_p^\dagger a_q,$$

$$H_{\text{imp}} = \sum_{pqrs=1}^{M} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s.$$

Typically, we consider constant M and a scaling total system size N. We assume that H_{free} has single-particle energies in a constant range (which we can always achieve by rescaling). We write ω for the ground state energy gap of H_{free}. If the Hamiltonian is number-conserving, we can study the lowest-energy state in the subspace where we fixed the number of particles to n.

Quantum impurity models are a simple but nontrivial class of models that capture the orthogonality catastrophe and the quality of embedding methods. The Anderson derivation of the orthogonality catastrophe [57], see Appendix A, suggests that the scaling of the orthogonality catastrophe may be polynomial rather than exponential for quantum impurity problems. For quantum phase estimation, an inverse polynomial overlap suffices to get a polynomial algorithm, so this suggests that quantum impurity problems may be in BQP.

Second, from the perspective of embedding methods, since the impurity is small compared to the system size, one may hope that one can identify a *bath* of a relatively small number of orbitals that are influenced by the impurity, and one can solve the full system by restricting to this bath.

Quantum impurity problems have been studied rigorously by Bravyi and Gosset in [23]. Their main technical result is a bound on the decay of the eigenvalues of the 1-RDM. This gives rise to approximations of the ground state by a relatively small number of Gaussian states. As an application, Ref. [23] gives a quasipolynomial classical algorithm for the ground state energy, and it shows that the quantum impurity problem is in QCMA, which is the class of problems that can be efficiently verified by a quantum computer given a classical witness, and which is between NP and QMA.

Our results build on this work in two ways. First, we argue that under certain conditions one can find an efficiently preparable state that has inverse polynomial overlap with the ground state, giving an efficient quantum algorithm for the ground state energy.

Second, the results in [23] are formulated in terms of Gaussian states rather than Slater determinants. To relate their results to usual approaches to embedding methods in quantum chemistry, we derive the analogous statements using Slater determinants, working at a fixed particle number. A detailed description of the setup, as well as formal statements and proofs, are provided in Appendix C.

A. Decay properties of the one-body reduced density matrix and good embeddings

One-electron basis states (orbitals) are key to understanding different embedding techniques [58]. Given an electronic Hamiltonian on a set of orbitals and a fixed number of electrons, orbitals are often partitioned into sets that are called *frozen* (doubly occupied), *active*, and *virtual* (unoccupied). The idea is that, in an expansion of the wave function as in Eq. (1), only the frozen and a few additional orbitals (that is, the active ones) give rise to determinants with large weights α_i. This means that the problem reduces to an effective problem defined only on the active space. The frozen and virtual spaces may contribute to the dynamical correlation, but this can be accounted for by a computationally less intensive method (for example, by perturbation theory). The selection of the active orbitals can be based on various measures such as natural orbital occupation numbers [59–62] or information-entropy-based [63–66]. Our formal analysis will focus on the former in the following, whereas for practical reasons, our numerical results will exploit the latter, which can be based on localized orbitals that can have advantages over natural orbitals. In any case, as we review in the Appendix A, orbital rotations are always possible, which would allow one to exploit natural orbitals for state preparation, which can then be rotated into a local basis.

Given a ground state $|\Psi\rangle$, the *one-body reduced density matrix* (1-RDM), or covariance matrix, is the matrix γ defined by $\gamma_{pq} = \langle\Psi|a_p^\dagger a_q|\Psi\rangle$. Its eigenvectors define a set of orbitals known as the *natural orbitals*. We order the eigenvalues of γ as $1 \geqslant \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_N \geqslant 0$. The eigenvalue σ_j equals the expected particle number for natural orbital j. This means that the natural orbitals for which σ_j equals 0 or 1 are always unoccupied or occupied, respectively, in the ground state. If the eigenvalue σ_j is close to 0 or 1, we may approximate the ground state by a state where the corresponding natural orbital is unoccupied or occupied, respectively. In other words, if we can bound the number of orbitals for which σ_j is *not* close to 0 or 1 by K, then we can approximate the ground state $|\Psi\rangle$ by a state $|\tilde{\Psi}\rangle = |\Phi\rangle \wedge |\Theta\rangle$, where $|\Theta\rangle$ is a Slater determinant and $|\Phi\rangle$ is an arbitrary state on K modes. This means that these K modes define a good active space for the problem at hand, and we can reduce the problem of (approximately) computing the ground state energy to this subspace.

The main result of [23] is a bound on the eigenvalues of the 1-RDM of quantum impurity problems, which is used to approximate the ground state in terms of Gaussian states. For chemistry problems, we typically have a fixed number of electrons, and it is more convenient to work with Slater determinants. We adapt the argument of [23] to show the following result, which we prove in Appendix C.

Theorem III 1. Let $\omega > 0$ be the ground state energy gap of H_{free}, and let $\varepsilon > 0$. Then for

$$K = O(\log(\omega^{-1})(\log(\varepsilon^{-1}) + \log\log(\omega^{-1})))$$

there exists a Slater determinant $|\Theta\rangle$ on $N - K$ modes and an arbitrary state $|\Phi\rangle$ on K modes such that the state

$$|\tilde{\Psi}\rangle = |\Phi\rangle \wedge |\Theta\rangle$$

has overlap $|\langle\tilde{\Psi}|\Psi\rangle| \geqslant 1-\varepsilon$ with a ground state $|\Psi\rangle$ of $H_{\text{free}} + H_{\text{imp}}$.

Note that while this result shows that there *exists* a good active space, it only provides a good way to determine which orbitals make up this active space given the ground state 1-RDM (which one does not necessarily have access to without *a priori* knowledge of the ground state). In practice, finding good active spaces is based on using an efficient classical method to approximate the 1-RDM and using this estimate to select an appropriate accurate space as discussed in Sec. II.

B. Computational complexity of the quantum impurity problem

The general electronic structure problem is QMA-complete [15,67], meaning that it is likely a hard problem even for quantum computers. On the other hand, Hamiltonians with only one-body interactions, like H_{free}, can be efficiently simulated classically [68]. An important broad question in quantum computing is to determine the existence of (physically relevant) families of Hamiltonians for which the ground state problem is in BQP, while computing the ground state energy is hard classically. Proving such a separation is difficult (as this would imply a separation between P and BQP), but one can at least try to prove BQP containment for specific families of Hamiltonian ground state energy problems for which we do not know rigorous polynomial-time classical algorithms; see, for example, [69,70].

We investigate the quantum computational complexity of quantum impurity problems. Here, the goal is to approximate the ground state energy of an impurity problem on N modes, with a constant-sized impurity, to precision ε. As we saw in the previous section, for quantum impurity problems there exist relatively small active spaces, significantly reducing the size of the problem. It was shown in [23] that this can be leveraged to give a *classical* algorithm that scales polynomially in N and quasipolynomially in the precision ε, with overall scaling $O(N^3)\exp(O(\log(\varepsilon^{-1})^3))$. It is also shown that the ground state energy problem for quantum impurity models is contained in QCMA, the class of problems for which a quantum computer can efficiently verify solutions given a classical witness. QCMA-hard problems are unlikely to be solvable efficiently on a quantum computer. This leaves the interesting open problem of whether one can prove whether quantum computers can efficiently approximate the ground state energy of quantum impurity problems to polynomial precision. We argue that under two different restrictions, quantum impurity problems become easy for quantum computers (while not obviously so for classical algorithms).

The main observation is that when we apply Theorem III 1, it suffices to find an active space such that the ground state overlap is constant (but not close to 1). If this active space contains only a constant number of orbitals K, then we may consider the state ρ corresponding to choosing a random active space of size K and preparing the maximally mixed state for this active space. For constant K, this leads to a ground state overlap that is at least $\text{poly}(N)^{-1}$, and hence ρ can be used as a guiding state to yield a polynomial time algorithm. Indeed, Theorem III 1 gives a constant-sized active space for constant precision if the gap ω is constant. As a further comment, these assumptions change the scaling of the classical algorithm of [23] to $O(N^3)\exp(O(\log(\varepsilon^{-1})^2))$ in the case of constant gap. This provides evidence that quantum impurity models may provide an example class of systems with polynomial time quantum algorithm, but no polynomial time classical algorithm. However, the precise status of this remains unclear, in particular since there is already a classical algorithm with quasipolynomial scaling, and an improved analysis or better classical algorithm could lead to polynomial scaling. Moreover, the potential advantage disappears if the full quantum impurity Hamiltonian is gapped, as [23] provides a polynomial classical algorithm for that situation based on matrix product states (MPSs).

A second simplification arises in the case in which we have a sufficiently accurate approximation of the 1-RDM of a ground state. In this case, we can construct an active space on $\log(\omega^{-1})$ orbitals which suffices for a constant overlap with the ground state. Since this active space has polynomial dimension, the maximally mixed state on this subspace gives a sufficiently good guiding state for phase estimation. We summarize these findings with the following result, formally stated and proven as Theorem C 3 in Appendix C.

Theorem III 2. The quantum impurity problem for $H = H_{\text{free}} + H_{\text{imp}}$ is in BQP if either H_{free} has a constant gap, or if the ground state 1-RDM is given.

Of course, one generally needs the ground state in order to compute the 1-RDM, so if one requires it beforehand, then this does not directly lead to a useful algorithm. The above result should, however, be seen as evidence for the usefulness of iterative approaches, where one starts with some estimate of the 1-RDM using computationally inexpensive classical means, and then uses a quantum computer to perform a calculation that leads to a more accurate approximation of the 1-RDM. See [54] for a practical proposal for such a scheme.

The algorithms in Theorem III 2 and their analysis do not directly lead to *practically* useful approaches, as their polynomial scaling can be of high degree. However, these results should be seen as a proof of principle that in quantum impurity problems the "orthogonality catastrophe" is of a relatively mild nature and may not be an insurmountable problem, reducing a potentially exponential scaling to a polynomial one. In practice, one can likely find better guiding states than the maximally mixed states in the above theoretical algorithms. Of course, classical algorithms for impurity problems will also perform much better than currently proven by rigorous guarantees [71–73].

Given a large system of N orbitals, where we know that the problem has a strong electron correlation in some subsystem, one can use a quantum embedding method (such as described in Sec. II) to find a reasonable active space A of K modes, and an environment system C. One then uses a classical method to find a good guiding state $|\Psi_A\rangle$ on this small active space, and a Slater determinant $|\Psi_C\rangle$ on C, and uses $|\Psi_A\rangle \wedge |\Psi_C\rangle$ as a guiding state on the full system. Theorem III 1 shows that for a quantum impurity model there *exist* good guiding states of this form, but it does not guarantee that one finds the right subsystem A or state $|\Psi_A\rangle$ in this way.

Finally, we note that the concept of impurity models has also been used to compute properties of strongly correlated materials, as in the Hubbard model. In these cases, a local

patch of the system is treated as an impurity, and the interaction with the remainder is treated at a mean-field level [74]. It has been proposed to use this framework, with a quantum computer as the impurity solver, for strongly correlated systems [54]. It was demonstrated in [8] that this leads to a highly multiconfigurational problem on the quantum core, but that still, with only 10 determinants, an overlap of around 0.05 can be achieved.

IV. NUMERICAL EXAMPLES

We now proceed to a numerical investigation of two different embedding schemes following the concepts discussed above, namely of a scheme used in the bootstrap embedding (as an example of a density-matrix embedding) and of the Huzinaga (projection) embedding, to assess the influence of quantum embedding on the preparation of states with large ground state overlap. We investigate the relations between the overlap of the mean-field HF basis state, as well as more complex states, with the target state for different spatial sizes of a quantum core and for growing orbital spaces. As we will see, these different embedding strategies produce rather different orbital bases. We study two systems that are exemplary for biomolecules: tryptophan in Sec. IV A (one of the essential amino acids) and a compound containing ruthenium in Sec. IV B with prospects as an anticancer drug. Some of the computational methods are described in more detail in Appendix D.

In general, we observe that in all cases we encounter, the (easy-to-prepare) mean-field state has large overlap with the ground state of the embedded Hamiltonian, with overlaps larger than 0.9 in almost all cases. By using more complex states, the overlap can be brought closer to 1. This does not greatly influence the total cost of quantum phase estimation, but it does allow parallelization with reduced maximal depth (and in many cases this depth reduction will be larger than the additional circuit depth from the more complicated guiding state).

A. Tryptophan

Some of the 20 essential amino acids, the fundamental building blocks of biomacromolecules, contain unsaturated side chains that are therefore suitable candidates for probing the role of static correlation in biomacromolecules. Out of these aromatic amino acids, we chose tryptophan as an example. In addition to this monomeric protein building block, we also study a sequence of tryptophan molecules resembling an oligopeptide structure (Fig. 2) that can be considered a limiting case for proteins (noting that the primary sequence of amino acids will not necessarily show repetitions composed of only one amino acid). This situation is used to model a biomolecule of increasing size, so we can see how the overlap of the HF state or sum-of-Slater states decreases with increasing system size (peptide length).

An easy approximation is to consider the tryptophan residues in this series as noninteracting (so the ground state is a product state) so that we can easily estimate the exponential decline of the overlap of the HF determinant and the sum-of-Slater state with peptide length. Switching on the weak interaction only slightly changes the overlap, as we demonstrate for the di- and tripeptides of tryptophan.

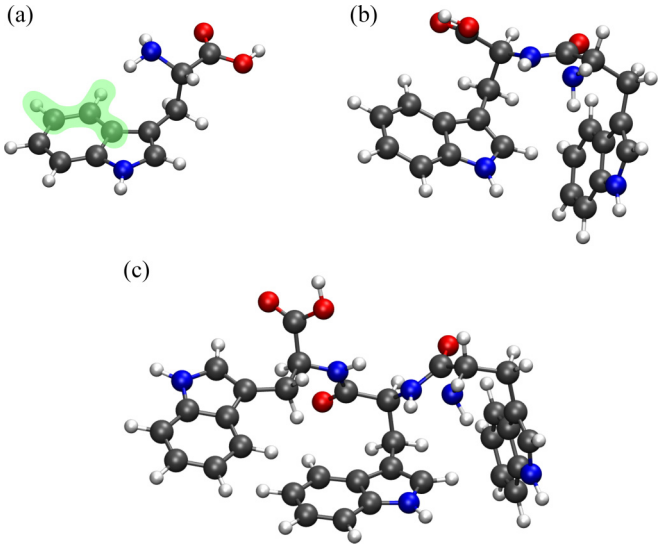

FIG. 2. Molecular structure of tryptophan and its derivatives: (a) monomer with the fragment used to compare different embedding approaches indicated in green, (b) dimer, and (c) trimer. Carbon atoms are indicated in black, hydrogen atoms in white, oxygen atoms in red, and nitrogen atoms in blue.

Although our results are specific to the examples chosen, we expect their electronic structures to be representative for a more general picture of electronic states. This is due to the regular structure of the amino-acid building blocks of proteins. The results obtained for tryptophan and its oligopeptides will be easily transferable to the homologous and also aromatic amino acid phenyl alanine and may be straightforwardly generalized to the other essential amino acids.

We first applied a density-matrix based embedding. Specifically, we used the bootstrap embedding algorithm for molecular systems [40] on a selected five-atom fragment that is part of the aromatic side chain, depicted in Fig. 2. This fragment was selected as it represents an intuitively challenging embedded system, as it is connected to the environment with both single and delocalized conjugate double bonds. For the bootstrap embedding, a set of orthogonal localized orbitals was constructed using the intrinsic atomic orbitals scheme [75]. This basis was then used to perform the Schmidt decomposition of the HF state and select the set of fragment and entangled bath orbitals. Fragment orbitals selected in this way were localized on atoms, making them a poor choice for the wave-function representation. Therefore, the selected fragment and bath orbitals were transformed to the eigenvectors of the Fock operator in the fragment-bath space, resulting in canonical orbitals depicted in Fig. 3(a). Due to the nature of Schmidt decomposition, these orbitals resemble the canonical HF orbitals of the entire system and, consequently, exhibit relatively small single-orbital entropies (defined as von Neumann entropies of the one-orbital reduced density matrix [63]). Hence, the multiconfigurational character of the fragment-bath system obtained with bootstrap embedding is comparable to the one of the entire system, even though fragmentation "cuts" through single and double bonds [Fig. 3(b)].

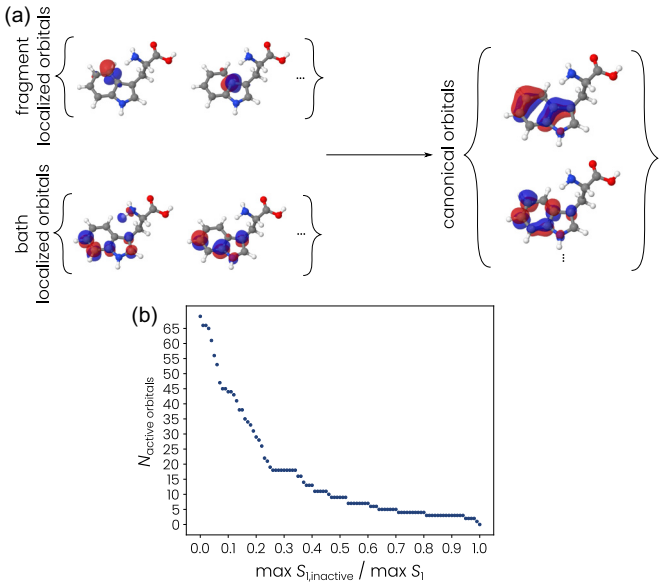

FIG. 3. (a) Left: fragment and bath orbitals obtained from the Schmidt decomposition of the HF state. Right: canonical orbitals obtained from the HF calculation in the fragment-bath space. (b) Threshold diagram corresponding to the maximal discarded single-orbital entropies relative to the largest value (max $S_1 = 0.191\,15$) for different active space sizes, as introduced in Ref. [66].

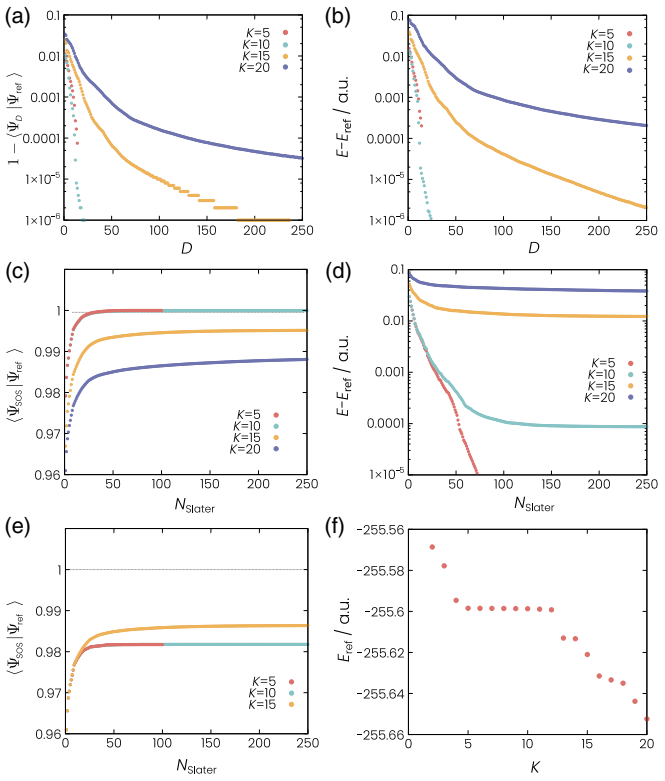

FIG. 4. (a) Overlap of truncated MPS wave functions with the reference MPS of bond dimension $D = 1024$ (taken as an approximate FCI result) for different active space sizes K. (b) Corresponding energy differences of the small-bond-dimension and the reference MPS. (c) Overlap of sum-of-Slater states with different numbers of Slater determinants with the reference MPS wave function. (d) Corresponding energy differences of the sum-of-Slater states and reference MPS state. (e) Overlap of sum-of-Slater states obtained from reference MPS wave functions in smaller active orbital spaces with the reference MPS wave function of the largest active space considered ($K = 20$). (f) Reference MPS energies for different active space sizes.

From the drop in the values of the single-orbital entropies, one can conclude that there are 12 orbitals that contribute the most to the multiconfigurational character and the static correlation [66,76] and include the reconstructed delocalized aromatic π system of the side chain (Fig. 3). Note that in the complete bootstrap embedding, one would apply this construction to multiple fragments, and then perform a matching procedure between the different fragments, a step we ignore for the purpose of this discussion.

Next, we performed DMRG calculations for different active space sizes in the fragment-bath space, with orbitals being selected based on the largest values of the single-orbital entropies, as described in Ref. [66]. We use a bond dimension of $D = 1024$ to yield a reference MPS that can be taken as a reliable approximation to the exact (full configuration interaction, or FCI) ground state. To assess the suitability of MPS wave functions with smaller bond dimensions for initial state preparation, we calculate both energy differences and overlaps of them with the reference MPS; the results are shown in Figs. 4(a) and 4(b).

From Fig. 4 one can see that the overlap of the truncated states with the reference MPS state approaches a value of 1 relatively quickly. This is expected, as the truncation of the bond dimension is performed in the optimal way using the singular value decomposition, selecting the state of the lower bond dimension that has maximal overlap with the nontruncated one. For similar reasons, the energies of the small-bond-dimension states quickly approach the corresponding reference energy. Furthermore, even states with bond dimension $D = 2$ already demonstrate excellent overlap with the reference MPS state ($\approx$0.97), making such an MPS a good candidate for state preparation.

We also consider a different *Ansatz*, namely sum-of-Slater states (see Appendix A and [9] for a discussion of such states on quantum computers). We obtain these by selecting the most important determinants from the FCI-type expansion represented by the reference MPS wave function, keeping the $N_{\text{Slater}} = L$ terms with the largest amplitudes:

$$|\Psi_{\text{SOS}}\rangle = \frac{1}{\sqrt{\sum_{i=1}^{L} |C_i|^2}} \sum_{i=1}^{L} C_i |\Phi_i\rangle, \tag{2}$$

where $|\Phi_i\rangle$ are Slater determinants in a fixed basis of orbitals, and we have ordered so $|C_1| \geqslant |C_2| \geqslant \cdots$. These states are both useful as a guiding state *Ansatz* (for small L), and to understand the correlation structure of the ground state. From Fig. 4(c), it can clearly be seen that the overlap obtained for these sum-of-Slater expansions with the reference MPS wave function are again relatively large. However, convergence of these overlaps with the number of included determinants is much slower than the convergence of small-bond-dimension MPS wave functions with increasing bond dimension.

Moreover, convergence is slowed down even further for increasing active space sizes K, which may be taken as an indication of the orthogonality catastrophe in the large active space limit. It is thus necessary to weigh both the gate count for the preparation of MPS and sum-of-Slater states and the quality of the overlap that they provide for the specific system of interest.

With respect to the correlation structure of the ground state, it is noteworthy that the change of overlap exhibits two different regimes, with a large jump at a small number of Slater determinants, followed by a slow convergence to unity. The initial jump corresponds to the inclusion of the Slater determinants describing excitations into the virtual orbitals with the largest values of the single-orbital entropies, which carry the largest coefficients in the FCI expansion and account for static correlation. Although overlaps with the reference MPS wave function are comparatively large, energies corresponding to the sum-of-Slater states converge slowly to the reference MPS energy [Fig. 4(d)]. Such behavior is expected, as truncation of a Slater-determinant expansion results in a neglect of a significant portion of the dynamical correlation energy. By contrast, an MPS, even if truncated to very low bond dimension, can represent a large number of Slater determinants, resulting in much better convergence behavior, which makes them a good candidate for guiding states with high ground state overlap [9].

We now investigate the quality of the sum-of-Slater states obtained for smaller active spaces as initial states for the larger active space ($K = 20$). As can be seen in Fig. 4(e), the sum-of-Slater *Ansatz* prepared in an active space with $K = 15$ orbitals provides overlaps that are virtually indistinguishable from the reference. In the case of $K = 5$ and 10, the overlap is somewhat lower when compared to the sum-of-Slater state prepared in the $K = 15$ case, but it still represents an improvement over the Hartree-Fock state. This is due to the fact that not all highly entangled orbitals are included in these cases [as can be seen from Fig. 3(b)], resulting in a lack of several determinants with non-negligible coefficients. However, it can be seen that as long as the most entangled orbitals are included, these active spaces provide sum-of-Slater states that are of the same quality in terms of overlap as the ones obtained from the calculation on the larger active space.

Finally, from Fig. 4(f), it is evident that many orbitals must be considered to reach chemical accuracy in absolute energies, as they carry a significant portion of the correlation energy. The fact that the initial state can be prepared in smaller active spaces therefore significantly reduces the computational overhead for the classical part of the initial state preparation.

We also investigated a projection-based Huzinaga embedding [41,77], using the same five-atom fragment, in order to assess how different embedding strategies affect the state preparation problem. There are several key differences with respect to bootstrap embedding.

First, the orbitals used for the Huzinaga embedding are obtained from a DFT calculation on the entire molecule, while the basis for the density-matrix approach in the bootstrap embedding is a HF calculation. This results in a different Hilbert space for the fragment problem. Second, in the Huzinaga embedding, interaction of the fragment orbitals with the occupied environment is described with a DFT potential, which approximates the correlation contributions from the occupied environment orbitals that are neglected in the case of bootstrap embedding. This interaction modifies the one-electron part of the fragment Hamiltonian. Third, a split orbital localization scheme is applied within the Huzinaga embedding, based on intrinsic bond orbitals [75,78], while canonical orbitals are used in the bootstrap embedding, which results in a qualitatively different multiconfigurational character of the fragment wave function in these cases. Unlike bootstrap embedding, the entire virtual space is considered in Huzinaga embedding, and the canonical orbitals are not recomputed after the embedding. Due to the local nature of the orbitals, the values of the single-orbital entropies are significantly different from the bootstrap embedding case (Figs. 5 and 3). In Huzinaga embedding, orbital entropies are larger when compared to the bootstrap case, which can be attributed to the usage of localized orbitals [79]. This results in a larger number of highly entangled orbitals, while the contribution of the virtuals that

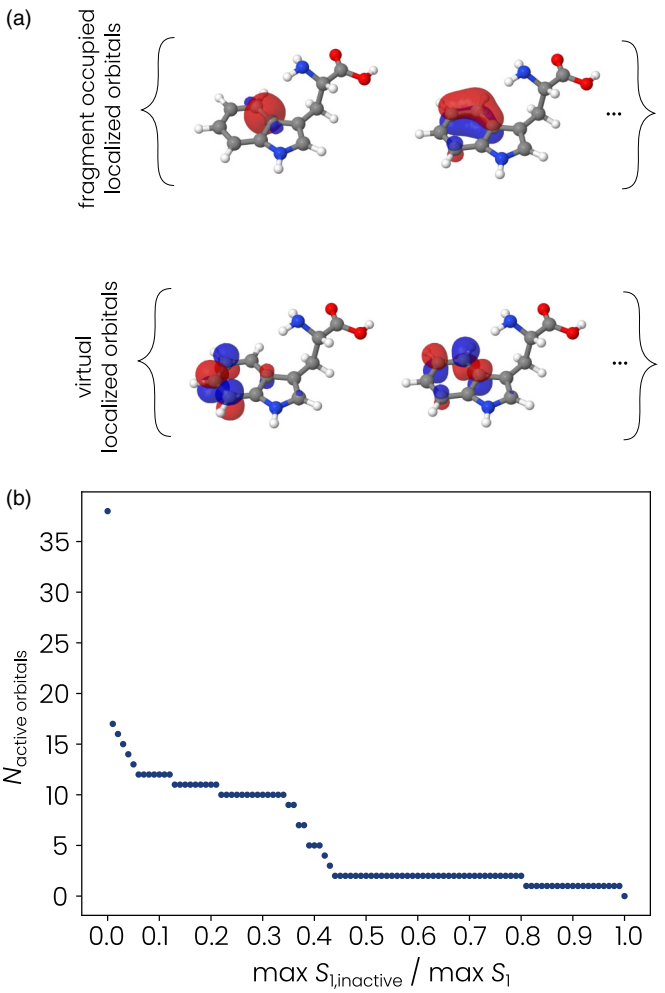

FIG. 5. (a) Localized occupied fragment orbitals and localized virtual orbitals obtained in Huzinaga embedding. (b) Threshold diagram corresponding to the maximal discarded single-orbital entropies relative to the largest value (max $S_1 = 0.218\,66$) for different active space sizes, as introduced in Ref. [66]. This may be contrasted with Fig. 3.

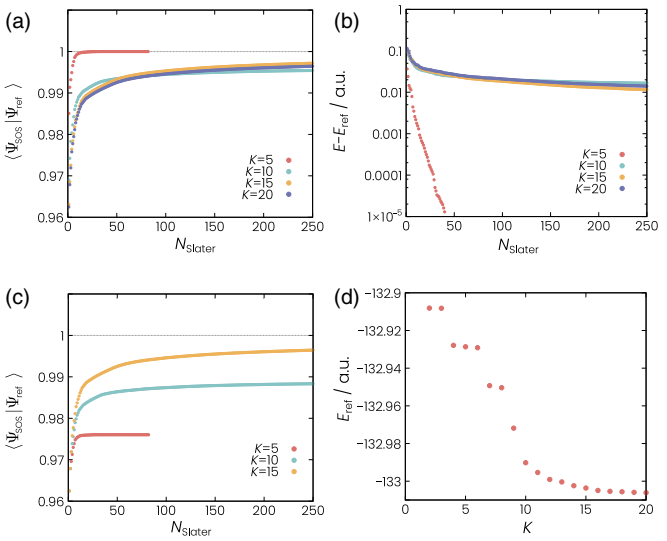

FIG. 6. (a) Overlap of sum-of-Slater states (with N_{Slater} Slater determinants corresponding to the largest coefficients) with the reference MPS state of bond dimension $D = 1024$ for different active space sizes K. (b) Overlap of sum-of-Slater states for different active space sizes K with the reference MPS state ($D = 1024$) for the largest active space considered $K = 20$. (c) Energies of sum-of-Slater states, with lines corresponding to the energies of the full MPS wave function. (d) Energy of the MPS ground state energy approximation, as a function of active space size K.

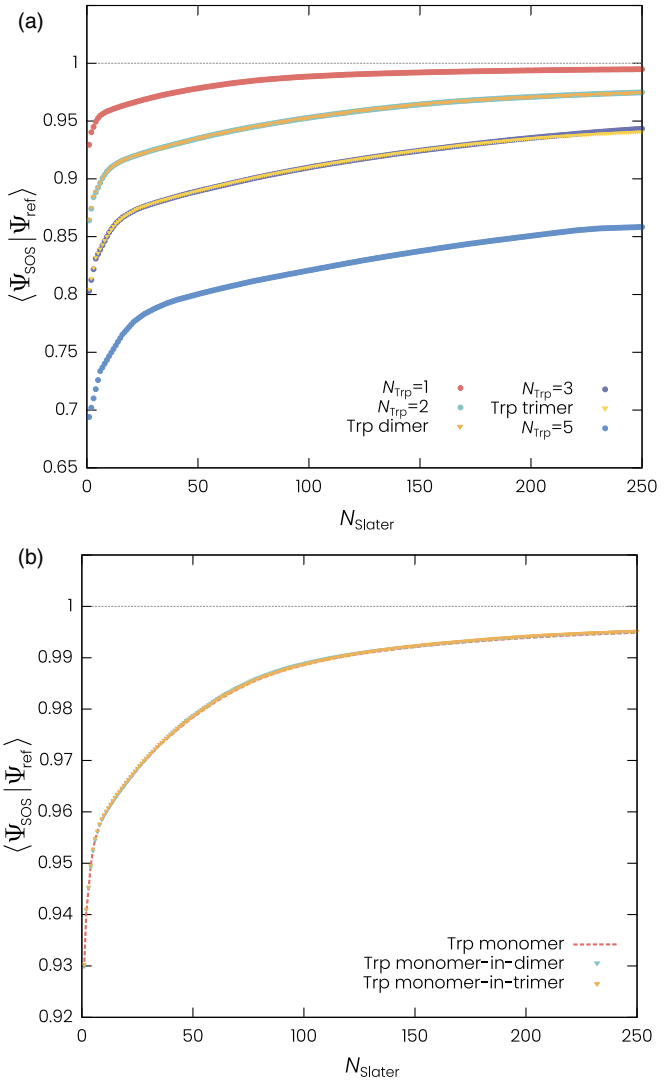

FIG. 7. (a) Overlap of sum-of-Slater states containing N_{Slater} Slater determinants with the reference MPS state of bond dimension $D = 1024$ for different numbers of noninteracting tryptophan residues (circles) and interacting residues in ditryptophan and tritryptophan (triangles; note: the data for the interacting dimer and trimer overlap with the data for the noninteracting cases). (b) Same as (a) for a single monomer embedded into the ditryptophan and tritryptophan (triangles). Values for the isolated monomer are given as a dashed line.

are spatially separated from the embedded fragment decreases significantly.

Due to the small entanglement of some virtual orbitals, convergence of the sum-of-Slater states both in terms of overlap and energy is faster than in the bootstrap embedding case [Figs. 6(a) and 6(b)]. Hence, localized orbitals might present a more suitable basis for initial state preparation. In classical quantum chemical methods one aims for a pronounced single-configurational character of the wave function and accounts for dynamic correlation *a posteriori*. In contrast, for quantum computers, it may be more desirable to have a multiconfigurational wave function with little residual dynamic correlation [Fig. 6(d)], since given a guiding state with only constant overlap, phase estimation gives accurate energies on the fragment orbitals. Similar to our bootstrap embedding results, initial states with large overlap can be prepared in a smaller active space, reducing the cost of a classical calculation for the state preparation step also in the case of Huzinaga embedding [Fig. 6(c)].

Finally, we turn to the sequence of tryptophan molecules mimicking an oligopeptide structure. We chose the active spaces to be residing entirely on the side chain (see Appendix D for details). First, we consider the limit of noninteracting side chains corresponding to separated tryptophan molecules in the sequence. In this case, the overlaps of the HF state with the reference MPS states decay exponentially with the number of tryptophan molecules [Fig. 7(a)]. In the case of sum-of-Slater states, the decay is approximately exponential as well. The exact exponential dependence occurs when the Slater determinants included correspond to the wave function that is a product of sum-of-Slater states on each monomer. Since the monomer ground state has relatively high overlap with the HF state, we see that the overlap of the HF state is still substantial (at around 0.7) for a sequence of five tryptophane molecules.

We also compute the overlap with the ground state of the *interacting* tryptophan dimer and trimer. It can be seen that weak interactions (between the side chains) only marginally perturb the overlap behavior observed for the separately treated side chains, so the exponential decay from the noninteracting case is a good approximation. We performed embedding calculations for a single side chain in ditryptophan (at the C-terminus side chain) and tritryptophan (side chain at the amino acid in the middle) and investigated the

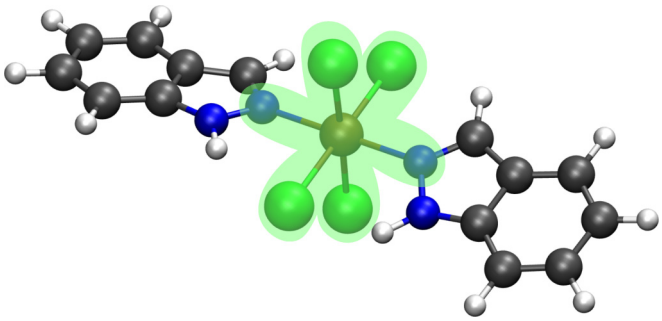

FIG. 8. Molecular structure of the ruthenium complex with the fragment used for the Huzinaga embedding indicated in transparent green. Carbon atoms are indicated in black, hydrogen atoms in white, nitrogen atoms in blue, chlorine atoms in green, and the ruthenium atom in brown.

resulting overlaps of the sum-of-Slater states. As can be seen in Fig. 7(b), the behavior of the overlap in both cases of the embedded monomers is essentially the same as in the case of a single tryptophan side chain. From these observations, we may conclude for proteins that embedding of relevant amino acid side chains represents a good strategy for mitigation of the orthogonality catastrophe caused by large system size.

B. Ruthenium anticancer drug

As a second example, we consider a system containing elements beyond the second period of the Periodic Table. This compound (see Fig. 8 for a ball-and-stick representation of its structure) is an anticancer drug [80,81], which can bind as an inhibitor to a glucose-regulating protein [82]. Such a binding is a typical example of small-molecule drug recognition by biomacromolecules and therefore an example for a molecular recognition application. Here, we focus on the isolated Ru-based drug molecule and consider its central region, the Ru ion containing moiety, as a quantum core (as highlighted in green in Fig. 8). We leave an investigation of how the ground state overlap problem changes when including the target protein in the embedding to future work.

For this complex, we applied the Huzinaga embedding of the ruthenium ion and its first (nearest-neighbor) coordination sphere (Fig. 8). We consider two different charges of this complex: the anion, $q = -1$, corresponds to a doublet state, and the neutral complex, $q = 0$, we considered as a triplet ground state. For the doublet, the overlap of the HF state and the sum-of-Slater states is very large, indicating a single-configurational character of the wave function. By contrast, the triplet state exhibits smaller overlap of the HF determinant with the reference MPS state of bond dimension $D = 1024$, which is quickly cured by including a second Slater determinant to yield a large overlap; see Fig. 9. This behavior is a consequence of the triplet nature of the wave function, which cannot be described by a single spin-restricted determinant. However, it can be described with a configuration state function, which is a symmetry-adjusted basis state. In the case of the triplet, the configuration state function is a linear combination of two restricted Slater determinants (and therefor multiconfigurational). If the system of interest

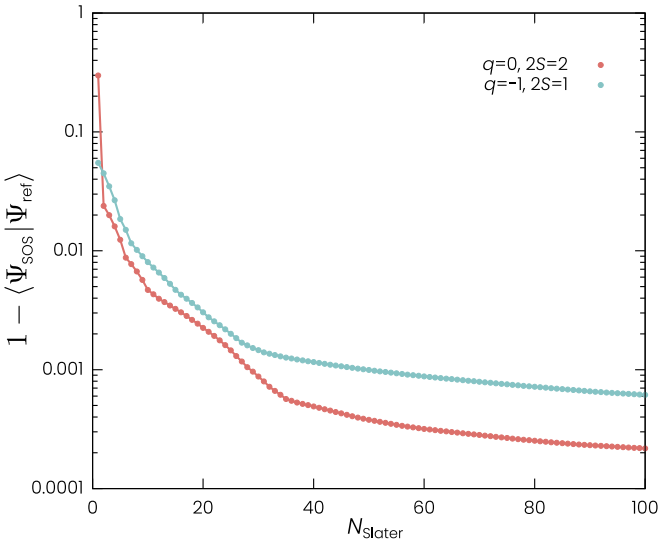

FIG. 9. Overlap of sum-of-Slater states (with N_{Slater} Slater determinants corresponding to the largest coefficients) with a reference MPS state of bond dimension $D = 1024$ for the doublet (net charge $q = -1$) and triplet (net charge $q = 0$) electronic states of the embedded fragment in the ruthenium complex.

is comprised of several high-spin regions (e.g., in the case of metal clusters), the overlap of the HF determinant will further decrease, depending on the number and spins of these regions [16]. It has already been argued that spin coupling to produce configuration state functions effectively solves this problem [18]. However, we note that such situations are not at all common in biomolecular recognition situations.

V. CONCLUSIONS

In this work, we have demonstrated that embedding methods effectively avoid the orthogonality catastrophe in practice for problems with local spatial structures, specifically in biochemistry. We show that for two conceptually different quantum embedding methods, low complexity states serve as good guiding states for quantum phase estimation. We demonstrate that this is the case for the fundamental building blocks of biomolecules, but also for other compounds such as small metal-containing drug molecules. As expected, the ground-state overlap of the mean-field HF basis state with the target state decreases with increasing active space size, but based on our estimates, phase estimation will likely face bottlenecks due to the polynomial scaling of simulation methods with the number of orbitals N before small or vanishing ground state overlaps become problematic.

If we consider embedded fragments with a fixed number of electrons and an increasing number of orbitals, in principle FCI on a classical computer scales polynomially as $O(N^n)$. This means that the speed-up provided by quantum phase estimation is polynomial [as the Hamiltonian simulation subroutine has polynomial complexity in N, with $O(N^4)$ scaling without any truncation and factorization strategies]. However, for a modest number of electrons, on the order of $\sim$20 for exact solvers and on the order of up to $\sim$100 for approximate solvers such as DMRG, FCI becomes unfeasible as a classical

method. Therefore, even though the quantum advantage of phase estimation algorithms in this case is in principle only polynomial, it is nevertheless significant. Additionally, with increasing quantum resources we can also increase the number of electrons in the fragment in order to reduce the edge effects of the embedding. This will restore the exponential scaling size of the Hilbert space.

For general applicability, we have also studied embedding and guiding state preparation for impurity models. Here we argued for the existence of accurate small active orbital spaces and for mild scaling of ground state overlap. This is clearly a positive prospect for quantum algorithms for ground state energy estimation, especially for biochemical applications. Additionally, the quantum description of a small reactive region in a macromolecular structure will always be sufficient for the investigation of relevant chemical processes as these can be considered to occur locally in a molecular structure (bond-breaking/forming involves a limited number of atoms). Therefore, focusing on the embedded region only is expected to be sufficient for such cases. If the quantum description of a larger region is required, this will also be possible. For instance, bootstrap embedding can be utilized to reconstruct the total energy from calculations on many smaller, overlapping embedded fragments [83] or stitch local energies for (small) embedded fragments together with a regression model or a machine learning potential. We emphasize that these procedures have already been established within traditional computing, and the insight is that they alleviate the orthogonality catastrophe and allow for efficient state preparation required for quantum phase estimation to obtain energies with controlled accuracy eventually.

To provide a more general framework for formal analysis, here we considered also embedding and guiding state preparation for impurity models. In this context, we could argue in favor of the existence of accurate small active orbital spaces and for mild scaling of ground state overlap. We further demonstrated that orbital selection based on the quantum information theory and single-orbital entropies provides such a small active space in which state preparation can be performed. An interesting open question is whether the quantum impurity problem without any further restrictions is in BQP. The mathematical analysis of embedding methods and of quantum impurity problems is based on the natural occupation numbers of the 1-RDM. Future work could explore whether the structure of 1-RDMs, which satisfy additional linear constraints [84,85], can be used to improve quantum embedding schemes [86–88] and help to find good guiding states for phase estimation.

Various interesting challenges remain to assess the utility of fault-tolerant quantum computers for ground state energy estimation of electronic structure in chemistry. As we have emphasized, one of the main advantages of quantum computing based on phase estimation over conventional methods is its accuracy guarantees. However, embedding methods introduce an (uncontrolled) error, even if small in practice. A better understanding of this error and the effect of the approximate treatment of the environment on the quantum core will be important for applications of quantum computers to macromolecules, and it requires further work.

ACKNOWLEDGMENTS

This work is part of the Research Project "Molecular Recognition from Quantum Computing." Work on this project is supported by Wellcome Leap as part of the Quantum for Bio (Q4Bio) Program. We also acknowledge financial support from the Novo Nordisk Foundation (Grant No. NNF20OC0059939 "Quantum for Life").

APPENDIX A: GUIDING STATES IN QUANTUM CHEMISTRY

This Appendix contains a review of the guiding state problem for quantum chemistry. We start by briefly reviewing the dependence of quantum phase estimation on the ground state overlap of the guiding state. Next we give an overview of different *Ansätzes* for guiding states. We discuss different types of correlation in quantum chemistry, and why one expects that the ground state overlap vanishes exponentially for increasing system size (the orthogonality catastrophe).

1. Quantum circuits and guiding state preparation

To estimate the ground state energy, one would like to apply quantum phase estimation with the time-evolution operator $U(t) = \exp(itH)$. Then, using a guiding state $|\psi\rangle$ with ground state overlap $|\langle\psi|\psi_0\rangle| \geqslant \eta$, one finds an approximation to the ground state energy with probability η^2. Therefore, the ground state overlap is related to the number of repetitions needed to find the ground state energy through phase estimation. An approximation of accuracy ε requires a circuit with time evolution for time $O(\varepsilon^{-1})$, leading to a total cost of $\tilde{O}(\varepsilon^{-1}\eta^{-2})$. The scaling with ε^{-1} is known as Heisenberg scaling and is optimal [89], but different approaches to phase estimation have different scalings with the ground state overlap. Given access to a quantum circuit preparing $|\psi\rangle$, one can find the ground state energy using $O(\eta^{-1})$ uses of this state preparation unitary, and time evolution for time $\tilde{O}(\varepsilon^{-1}\eta^{-1})$. This scaling is optimal [90] but does, however, lead to deeper circuits. On the other hand, there are also approaches that only require a single ancilla qubit and time evolution for time $\tilde{O}(\varepsilon^{-1})$ [91]; these require $\tilde{O}(\eta^{-4})$ circuit repetitions. Finally, in the regime where $\eta \to 1$, it is possible to reduce the maximal circuit depth, at the cost of an increased number of repetitions. If $\delta = 1 - \eta^2$, the maximal required time evolution can be reduced to $O(\delta\varepsilon^{-1})$, with a total evolution time of $O(\delta\varepsilon^{-2})$ [19,20]. This means that if one can find guiding states with ground state overlap surpassing $\eta \geqslant 0.9$ using cheap conventional methods, this may be used to significantly reduce the required circuit depth for accurate energy estimates using quantum phase estimation, which may be helpful for realizing such simulations on devices with a limited number of qubits and maximal circuit depth.

There are three standard classes of guiding states for chemistry problems, which we will now briefly review, together with the cost of preparing them on a quantum computer. We refer to [9] for an extensive overview. For convenience, we assume we are using a Jordan-Wigner mapping for the fermion-to-qubit mapping of the electronic structure problem. We quote the best known gate counts for two measures: the

number of two-qubit gates required (and ignoring the number of single-qubit gates), or alternatively the number of Toffoli gates (and ignoring the number of Clifford gates). The first measure is relevant for current and near-term devices where two-qubit gates are typically much slower and more noisy than single-qubit gates. The second measure applies to a fault-tolerant model where the main cost comes from the nontransversal (non-Clifford) operations.

(i) *Single Slater determinant:* This is the most basic case. Hartree-Fock yields a single Slater determinant, and under a fermion-to-qubit mapping this state is represented by a product state (which requires no two-qubit gates or T gates to prepare). If we use the standard Jordan-Wigner mapping with canonical orbitals (compatible with the HF state), then for N spin-orbitals and n electrons this is simply the state $|1\rangle^n|0\rangle^{N-n}$. It is also possible to choose a single Slater determinant in a different orbital basis than the one used for the fermion-to-qubit mapping, which can significantly improve the ground state overlap [8,17]. A single-particle basis change can be implemented as a quantum circuit using so-called Givens rotations. Using the Jordan-Wigner representation, this requires $n(N-n)$ two-qubit gates [92].

(ii) *Sum-of-Slater determinants:* A direct extension is to take a state that is a superposition of a (small) number L of Slater determinants, as in Eq. (2). Such states can, for example, be found through configuration interaction with single and double excitations (CISD) methods or selective configuration interaction (SCI) methods [9]. Alternatively, they can be extracted as the dominant coefficients of a matrix product state (see below). Again, if we assume that the fermion-to-qubit mapping uses the same orbital basis, such states are mapped to a superposition of L standard basis states. These can be prepared using at most $O(NL)$ two-qubit gates, see [93] for explicit gate counts, or $O(L\log(L))$ Toffoli gates and $[\log(L)]$ ancilla qubits [9]. Additionally, if the number of excitations from the HF state is bounded by k, $O(Lk)$ two-qubit gates suffice [94].

(iii) *Matrix product states:* A final power class of states are matrix product states (MPSs). These can be found using the DMRG algorithm, and they are characterized by their bond dimension D. Accurate quantum chemistry calculations typically require a large bond dimension, but alternatively one can either try to minimize over low bond dimension MPSs, or truncate a high bond dimension MPS to a low bond dimension state, which has less accurate energy but still significant ground state energy overlap. In many cases, this leads to high-quality guiding states [9], but at a relatively high classical processing cost. MPSs can be prepared in sequential fashion, using $O(ND^2)$ gates (either two-qubit or Toffoli) [95] and depth scaling with N. Heuristic methods for short depth circuits preparing states with high overlap with a given MPS can be found in [96,97]. Other depth reduction techniques are based on the correlation length of the MPSs [98], or by using midcircuit measurements and adaptive circuits [99].

Other methods for finding and preparing guiding states, not discussed in this work, include adiabatic state preparation [100,101], thermal state preparation algorithms [102], and variational approaches based on unitary coupled cluster methods [103].

2. The orthogonality catastrophe

In most cases in many-body physics, requiring approximations at the level of the wave function that have large overlap with the true ground state is a very strong requirement. As we will explain below in more detail, small (local) errors in the approximation lead to essentially orthogonal states on the many-body level, a phenomenon known as the orthogonality catastrophe. Nevertheless, for quantum algorithms it is in fact important to prepare guiding states that have a significant overlap with the true many-body ground state in order to guarantee that the quantum phase estimation algorithm will give a good estimate of the ground state energy.

We collect four basic arguments for the orthogonality catastrophe, based on *accumulation of error* and *Anderson's impurity argument* in the thermodynamic limit, the *electron-electron cusp* in the continuum limit, and finally an argument from the *computational complexity of the ground state problem*.

The most basic intuition for the orthogonality catastrophe is that if we have an extended system consisting of N subsystems without any correlation between the systems, the global ground state overlap decays exponentially if there is a local approximation error. If the true state is $|\phi\rangle^{\otimes n}$, and we have on-site estimates $|\psi\rangle$ of $|\phi\rangle$, then the global squared overlap decays exponentially as $|\langle\phi|\psi\rangle|^{2N}$ (and the same is true for uncorrelated fermionic systems) [104]. While this gives a reasonable intuition, it is also a very artificial scenario.

In [57] Anderson showed that a similar phenomenon occurs for a realistic physical system of an electron gas with an impurity. The impurity is a local-sized perturbation, which is constant-sized compared to the number of electrons. The ground state of the perturbed system has an overlap with the ground state of the unperturbed system, which, in this case, decays *polynomially* with the system size. Note that this differs from the uncorrelated example, where the decay arises from the fact that *every* local fragment incurs an error. This phenomenon is known as the Anderson orthogonality catastrophe. This can be derived in perturbation theory for a one-body perturbation [57]. This means that for impurity models, the orthogonality "catastrophe" may only be of a polynomial nature and therefore need not be an obstruction to an efficient quantum algorithm as one may hope that the ground state has a polynomially decaying overlap with the mean-field state. We partially make this idea rigorous in Sec. III and Appendix C, based on the work of [23].

Another fundamental reason for the orthogonality catastrophe lies in the continuum limit rather than the thermodynamic limit. Here we keep the number of electrons n fixed, but we increase the spatial resolution of the second-quantized Hamiltonian by raising the number of orbitals N. This leads to convergence to the true wave function in $L^2(\mathbb{R}^{3n})\otimes\mathbb{C}^2$. Note that the Hilbert space of a fixed number of electrons n but increasing N scales polynomially (but with an exponent scaling with n). It is well known that resolving the electron-electron cusp of the wave function of any system with Coulomb interactions requires a diverging number of Slater determinants (see, for example, the discussion in Chap. 7 of [7]). This is well understood analytically [105–107], and is true for arbitrary eigenfunctions of the electronic Hamiltonian. In

particular, it is known that the 1-RDM in the continuum limit has eigenvalues (i.e., natural occupation numbers) decaying as $\lambda_k \sim k^{-\frac{8}{3}}$ as k goes to infinity [105], lower bounding the required number of (natural) orbitals needed to approximate the exact wave function.

A final argument for a version of the orthogonality catastrophe comes from complexity theory. The problem of computing ground state energies is known to be QMA-complete, which means that it is strongly believed to be hard for quantum computers. This remains true in the case of electronic Hamiltonians [15,67]. As a result, there should be no efficient (quantum) algorithm for finding states with at most polynomially decaying overlap with the true ground state: if such an algorithm existed, it could be used to prepare an initial state for the quantum phase estimation algorithm and thus efficiently estimate the ground state energy. On the other hand, it is known that the ground state estimation problem remains BQP-hard even if one is provided a guiding state with very high ground state overlap [108]. This means that in general, having classical methods which achieve reasonable overlap does not imply the ability to improve this precision arbitrarily in an efficient manner (although this could still be the case in practice for problems in chemistry, as suggested by [16]).

Finally, for the general class of Hamiltonians that have an (unknown) efficiently describable guiding state whose expectation values can be computed efficiently using a classical computer (such as MPS), the ground state energy problem is QCMA complete [109]. This provides evidence that in general guiding states that can efficiently be described classically may still be hard to find.

3. Guiding states in chemistry

Electronic structure theory is the main quantum-mechanical problem in chemical and materials science. We distinguish three regimes:

(i) *The weak correlation limit:* In the weak correlation limit the mean-field Hartree-Fock state is already an excellent qualitative approximation. As discussed, these are easy to prepare on a quantum computer, and this suffices as the choice of guiding state. Most chemical processes belong to this class. However, for these systems, coupled cluster models based on a Hartree-Fock reference state deliver reliable results because they can efficiently account for the lacking dynamic correlation (yet with unknown system-focused error for a specific application [21]). Small molecules [8] belong to this class, as well as electronic host-guest binding energy calculations such as those required in drug design problems. Quantum computers potentially exhibit an advantage in this regime for larger numbers of orbitals N. Then, quantum phase estimation can, in principle, obtain an energy of guaranteed accuracy that can be taken as a reference for standard coupled cluster models.

(ii) *The intermediate correlation limit:* Here, static electron correlation can become important and multiconfigurational approaches are more suitable. In this regime, traditional computation begins to face severe problems. Only coupled cluster models that can deal with a multiconfigurational reference or that are of high excitation degree (including at least quadruple excitations) are applicable. However, the former are not unambiguously defined, whereas the latter are too costly for all but the smallest molecules. At the same time, generic multiconfigurational models such as the complete active space self-consistent field (CASSCF) wave function and the MPS wave function optimized by DMRG are natural choices, but suffer from a lack of similarly accurate dynamical correlation methods to account for the fact that CASSCF and DMRG approaches are restricted to a few dozen orbitals only. Nevertheless, only a small number of determinants (say, on the order of a dozen) will represent the state qualitatively well. Their superposition can be initialized efficiently as a guiding state on a quantum computer (provided that knowledge about these determinants can be obtained at comparatively little cost before a quantum computation).

(iii) *The strong correlation limit:* A large number of Slater determinants will be required in order to achieve a sufficiently high overlap with the target state. There are only a few examples known in ground state chemistry of this kind. A prominent class of examples are iron-sulfur clusters, where the overlap of the optimal Slater determinant with the ground state becomes very small [16], while DMRG optimized MPSs may still have large overlap. Such cases can be more routinely found in materials science, where materials are built from many units with half-filled single-particle states, or in electronically excited states.

Generally, to assess the *quantum advantage* of quantum phase estimation, a better understanding of which systems have polynomially scaling conventional methods in practice on the one hand, and an understanding of which systems allow for good guiding states on the other, is required [9,16].

APPENDIX B: FERMIONIC FORMALISM

We start by defining notation and recalling the formalism of fermionic quantum systems. We consider a single-particle space $\mathcal{H}$ of dimension N, which we may identify with $\mathbb{C}^N$. We refer to elements of the single-particle space $\mathcal{H}$ as *modes*, or equivalently in chemistry terminology as *orbitals*. The full Hilbert space is the Fock space

$$\bigoplus_{n=0}^{N} \mathcal{H}^{\wedge n} \cong (\mathbb{C}^2)^{\otimes N},$$

which can be mapped to N qubits. Numbering an orthonormal basis of modes $j = 1, 2, \ldots, N$, we associate creation and annihilation operators a_j and $a_j^\dagger$, respectively, which satisfy the fermionic anticommutation relations

$$\{a_j, a_k\} = 0 = \{a_j^\dagger, a_k^\dagger\}, \quad \{a_j, a_k^\dagger\} = \delta_{jk}.$$

We define the vacuum state $|\Omega\rangle$ to be the unique mutual kernel of all the operators a_j, from which the creation operators $a_j^\dagger$ generate the full Hilbert space as a Fock space. Given any normalized vector $x = (x_1, \ldots, x_N) \in \mathbb{C}^N$, we let

$$a^\dagger(x) = \sum_{j=1}^{N} x_j a_j^\dagger, \quad a(x) = \sum_{j=1}^{N} \overline{x_j} a_j$$

be the operators that create and annihilate the mode x.

The operators

$$\hat{n}(x) = a^\dagger(x)a(x), \quad \hat{n} = \sum_{j=1}^{N} a_j^\dagger a_j$$

are the number operator for mode x and the total number operator. If $|\Psi\rangle$ is an eigenvector of $\hat{n}(x)$ with eigenvalue 0 or 1, we say that mode x is, respectively, unoccupied or occupied in the state $|\Psi\rangle$. Given a state $|\Psi\rangle$, its *covariance matrix* or *one-body reduced density matrix* (1-RDM) is an operator on $\mathcal{H}$ defined by

$$\langle x, \gamma y\rangle = \langle\Psi|a^\dagger(x)a(y)|\Psi\rangle.$$

After a choice of basis, this gives an $N \times N$ matrix with entries $\gamma_{jk} = \langle\Psi|a_j^\dagger a_k|\Psi\rangle$.

A *Slater determinant* is a state of the form

$$a^\dagger(x_n)\cdots a^\dagger(x_1)|\Omega\rangle,$$

where $x_1, \ldots, x_n$ are a collection of orthonormal modes in $\mathcal{H}$. Up to a phase, the Slater determinant only depends on the subspace $\mathcal{X}$ spanned by the $x_1, \ldots, x_n$, and it is the state where the modes in $\mathcal{X}$ are occupied and the modes in $\mathcal{X}^\perp$ are unoccupied.

We can also define the $2N$ Majorana operators

$$c_{2j-1} = a_j + a_j^\dagger, \quad c_{2j} = -i(a_j - a_j^\dagger).$$

A *Gaussian unitary* is a unitary U acting on the Fock space such that as an orthogonal transformation $O \in O(2N)$,

$$U c_p U^\dagger = \sum_q O_{pq} c_q$$

for an orthogonal transformation $O \in O(2N)$. A *Gaussian state* is a state of the form $U|\Omega\rangle$, where U is a Gaussian unitary operator. In particular, any Slater determinant is a Gaussian state.

A *noninteracting* (or *free*, or *one-body*) Hamiltonian is a Hamiltonian of the form

$$H_{\text{free}} = \sum_{j,k} h_{jk} a_j^\dagger a_k,$$

where h is a Hermitian operator on $\mathcal{H}$. By choosing a basis for $\mathcal{H}$ in which h is diagonal, we can always transform this to the form

$$H_{\text{free}} = \sum_{j=1}^{N} \epsilon_j \tilde{a}_j^\dagger \tilde{a}_j$$

with $\epsilon_1 \leqslant \epsilon_2 \leqslant \cdots$. The ground state space of a noninteracting Hamiltonian is spanned by a set of Slater determinants. If $\epsilon_j < 0$ for $j \leqslant n$, and $\epsilon_j > 0$ for $j > n + k$, then the ground state space is spanned by the set of Slater determinants for which modes $1, \ldots, n$ are occupied and $n + k, \ldots, N$ are unoccupied. In particular, if $\epsilon_j \neq 0$ for all j, then the ground state is unique.

In this work, we are concerned with quantum impurity Hamiltonians, where a free Hamiltonian is perturbed by a non-negligible but spatially localized impurity term H_{imp}. The term H_{imp} is localized in the sense that, written in terms of the fermionic operators a_j, $a_j^\dagger$, it only contains the modes $j \leqslant M$ for some constant M.

Definition B 1. A *quantum impurity Hamiltonian* is a Hamiltonian of the form

$$H = H_{\text{free}} + H_{\text{imp}},$$

where H_{free} is a noninteracting Hamiltonian on N modes, and H_{imp} is an interacting Hamiltonian on a subset of M of the modes. The *quantum impurity problem* is the problem of computing the ground state energy of $H = H_{\text{free}} + H_{\text{imp}}$ to accuracy ε, where H_{free} has bounded single-particle energies ϵ_j.

In this computational problem, note that H_{imp} and M are taken to be constant parameters: we seek an efficient solution in terms of the total system size N and the precision ε.

There are a few ways in which this problem can be simplified. First, note that we can assume without loss of generality that the single-particle energies are non-negative. This follows by defining a new set of fermionic operators b_j by

$$b_j = \tilde{a}_j^\dagger \quad \text{for } j = 1, \ldots, n,$$
$$b_j = \tilde{a}_j \quad \text{for } j = n+1, \ldots, N,$$

where n is maximal such that $\epsilon_n < 0$, as above. More generally, this corresponds to the transformation

$$b(x) = \sum_{j=1}^{n} x_j a_j^\dagger + \sum_{j=n+1}^{N} \overline{x_j} a_j.$$

Under this transformation, the free Hamiltonian takes the form (after a constant energy shift)

$$H_{\text{free}} = \sum_{j=1}^{N} |\epsilon_j| b_j^\dagger b_j,$$

and the state vacuum state with respect to the b_j, $|\Theta\rangle$, is a ground state. Note that $|\Theta\rangle$ is a Slater determinant with respect to $\tilde{a}_j$, via

$$|\Theta\rangle = \prod_{j=1}^{n} \tilde{a}_j^\dagger |\Omega\rangle.$$

For convenience, we can normalize the energy scale to assume without loss of generality that $\epsilon_j \in [0, 1]$ for all j.

We denote by ω the energy gap of H_{free}, which is equal to the smallest nonzero $|\epsilon_j|$,

$$\omega = \min_{j:\epsilon_j \neq 0} |\epsilon_j|.$$

In fact, for the purposes of estimating the ground state energy of H to accuracy ε, it turns out that we may assume $\omega > \varepsilon/m = \Omega(\varepsilon)$. This fact follows by truncating the low-energy modes of H_{free}, and it is proved as Lemma 5 of [23].

In our analysis of the quantum impurity problem, we will make particular use of the covariance matrix with respect to the ground state $|\Psi\rangle$ of H, whose entries are given by

$$\gamma_{jk} = \langle\Psi|b_j^\dagger b_k|\Psi\rangle.$$

APPENDIX C: COVARIANCE MATRIX ANALYSIS FOR QUANTUM IMPURITY MODELS

In this Appendix, $H = H_{\text{free}} + H_{\text{imp}}$ is a quantum impurity Hamiltonian on N modes, with an impurity on M modes, and

we use the notation of Sec. III. We assume that H_{free} has the form

$$H_{\text{free}} = \sum_{j=1}^{N} \epsilon_j a_j^\dagger a_j,$$

where $\epsilon_j \in [-1, 1]$ for all j. As in Appendix B, we may write

$$H_{\text{free}} = \sum_{j=1}^{N} |\epsilon_j| b_j^\dagger b_j$$

for some new fermionic operators b_j. In this Appendix, we denote by γ and γ' the 1-RDMs of the ground state $|\Psi\rangle$ with respect to the b_j and a_j modes, respectively, that is,

$$\gamma_{jk} = \langle\Psi|b_j^\dagger b_k|\Psi\rangle \qquad \gamma'_{jk} = \langle\Psi|a_j^\dagger a_k|\Psi\rangle.$$

In [23] it is shown that for the ground state $|\Psi\rangle$ of H, the covariance matrix γ has exponentially decaying eigenvalues. To be precise, Theorem 2 of [23] shows that there exists a constant c such that there is a ground state $|\Psi\rangle$ such that the eigenvalues $\sigma_1 \geqslant \sigma_2 \geqslant \cdots$ of γ are bounded as

$$\sigma_j \leqslant c \exp\left(-\frac{j}{14M \log(2\omega^{-1})}\right). \tag{C1}$$

This bound can be used to show that $|\Psi\rangle$ can be approximated by a state that is a superposition over a limited number of Gaussian states. We revisit the argument of [23] and show that (with respect to the original choice of fermionic operators a_j) we can also obtain a ground state approximation using Slater determinants. Note that the class of Gaussian quantum states is strictly larger than the class of Slater determinants; for certain models, ground state approximations using Gaussian states can be much better than those by Slater determinants [110].

The key computational step is summarized in the following lemma. Informally, this guarantees that by assuming that modes with eigenvalues close to 1 in the 1-RDM are filled, and assuming that those with eigenvalues close to 0 are unfilled, one can obtain a reasonable approximation to the true state.

Lemma C 1. Let γ be the 1-RDM of a state $|\Psi\rangle$, and let $x_1, x_2, \ldots, x_N$ be an orthonormal basis for $\mathbb{C}^N$. Given disjoint subsets $I^+, I^- \subseteq [n]$, define the projectors

$$\Pi^- = \prod_{j\in I^-} a(x_j)^\dagger a(x_j) \quad \Pi^+ = \prod_{j\in I^+} a(x_j) a(x_j)^\dagger.$$

Then there exists a state $|\tilde{\Psi}\rangle$ in the image of both Π^- and Π^+ such that

$$|\langle\Psi|\tilde{\Psi}\rangle| \geqslant 1 - \delta,$$

where

$$\delta \leqslant \sum_{j\in I^-} \sqrt{1-\gamma_{jj}} + \sum_{j\in I^+} \sqrt{\gamma_{jj}},$$

where γ_{jk} are the elements of γ with respect to the basis $\{x_j\}$.

Proof. We can directly compute

$$\begin{aligned}\langle\Psi|\Pi^+\Pi^-|\Psi\rangle &= 1 - \langle\Psi|I - \Pi^+\Pi^-|\Psi\rangle \\ &\geqslant 1 - \|I - \Pi|\Psi\rangle\| \\ &\geqslant 1 - \|I - \Pi^-|\Psi\rangle\| - \|I - \Pi^+|\Psi\rangle\|.\end{aligned}$$

The first norm may be bounded by

$$\begin{aligned}\|I - \Pi^-|\Psi\rangle\| &\leqslant \sum_{j\in I^-} \|I - a(x_j)^\dagger a(x_j)|\Psi\rangle\| \\ &= \sum_{j\in I^-} \sqrt{1-\gamma_{jj}},\end{aligned}$$

and the second may be bounded by

$$\begin{aligned}\|I - \Pi^+|\Psi\rangle\| &\leqslant \sum_{j\in I^+} \|I - a(x_j) a(x_j)^\dagger|\Psi\rangle\| \\ &= \sum_{j\in I^+} \sqrt{\gamma_{jj}}.\end{aligned}$$

We now let

$$|\tilde{\Psi}\rangle = \frac{\Pi^+\Pi^-|\Psi\rangle}{\|\Pi^+\Pi^-|\Psi\rangle\|}, \tag{C2}$$

which satisfies the theorem by the above calculation. ■

We now restate Theorem III 1, with some additional details.

Theorem C 1. Let $\omega > 0$ be the ground state energy gap of H_{free}, and let $\varepsilon > 0$. Then for

$$K = O(\log(\omega^{-1})(\log(\varepsilon^{-1}) + \log\log(\omega^{-1})))$$

there exists a Slater determinant $|\Theta\rangle$ on $N - K$ modes and an arbitrary state $|\Phi\rangle$ on K modes such that the state

$$|\tilde{\Psi}\rangle = |\Phi\rangle \wedge |\Theta\rangle$$

has ground state overlap $|\langle\tilde{\Psi}|\Psi\rangle| \geqslant 1 - \varepsilon$. Moreover, we may choose the Slater determinant $|\Theta\rangle$ to be defined on a set of modes that commute with H_{imp}.

Proof. The full single-particle space is $\mathbb{C}^N$. Let $\mathcal{J}_+$ and $\mathcal{J}_-$ be the subspaces of non-negative and negative energy modes (of dimensions $N - n$ and n, respectively) and let $\mathcal{M}$ be the subspace of modes on which H_{imp} acts (of dimension M), and let $\mathcal{L}$ be the orthogonal complement of $\mathcal{M}$. By definition, H_{imp} commutes with $a(x)$ for any $x \in \mathcal{L}$. We then define

$$\mathcal{L}_+ = \mathcal{J}_+ \cap \mathcal{L}, \quad \mathcal{L}_- = \mathcal{J}_- \cap \mathcal{L}.$$

By dimension counting,

$$\dim(\mathcal{L}_+) \geqslant N - M - n, \quad \dim(\mathcal{L}_-) \geqslant n - M.$$

We define fermionic operators b_j by $b_j = a_j^\dagger$ for $j \leqslant n$ and $b_j = a_j$ for $j > n$, so the free Hamiltonian is given by

$$H_{\text{free}} = \sum_{j=1}^{N} |\epsilon_j| b_j^\dagger b_j$$

after subtracting a constant term. Since H_{free} is assumed to be normalized, the single-particle energies $|\epsilon_j|$ now lie in [0,1]. We let γ be the 1-RDM for the ground state $|\Psi\rangle$ with respect to the fermionic operators b_j, with eigenvalues $\sigma_1 \geqslant \sigma_2 \geqslant \cdots$. Let $\Lambda^\pm$ be the projection onto $\mathcal{L}_\pm$ and let $\lambda_1^\pm \geqslant \lambda_2^\pm \geqslant \cdots$

be the eigenvalues of $\Lambda^{\pm}\gamma\Lambda^{\pm}$. By the Cauchy interlacing theorem and Eq. (C1), we have

$$0 \leqslant \lambda_j^{\pm} \leqslant \sigma_j \leqslant c\exp\left(-\frac{j}{14M\log(2\omega^{-1})}\right).$$

Let γ' be the 1-RDM with respect to the original fermionic operators a_j, so that

$$\gamma'_{jk} = \delta_{jk} - \gamma_{kj} \quad \text{for } j,k \leqslant n,$$
$$\gamma'_{jk} = \gamma_{jk} \quad \text{for } j,k > n.$$

In particular, this means that $\Lambda^{+}\gamma'\Lambda^{+}$ has eigenvalues λ_j^{+} and $\Lambda^{-}\gamma'\Lambda^{-}$ has eigenvalues $1-\lambda_j^{-}$.

Let $\{v_j^{+}\} \subseteq \mathcal{L}^{+}$ and $\{v_j^{-}\} \subseteq \mathcal{L}^{-}$ be eigenbases of $\Lambda^{+}\gamma\Lambda^{+}$ and $\Lambda^{-}\gamma\Lambda^{-}$, respectively, corresponding to eigenvalues of descending size. The union of these bases can be extended to an orthonormal basis $\mathcal{B}$ for $\mathbb{C}^N$. Let I^{-} index all but the first $K/2-M$ elements of $\{x_j^{-}\}$, and let I^{+} index all but the first $K/2-M$ elements of $\{x_j^{+}\}$. Note that either of these sets may be empty, but $|I^{-}\cup I^{+}| \geqslant N-K$.

Applying Theorem C 1 with I^{+} and I^{-} as above, we obtain a state $|\tilde{\Psi}\rangle$ as in Eq. (C2) which has ground state overlap bounded by

$$|\langle\Psi|\tilde{\Psi}\rangle| \geqslant 1 - \sum_{j\geqslant K/2-M}(\sqrt{\lambda_j^{+}} + \sqrt{\lambda_j^{-}}).$$

By construction, the state $|\tilde{\Psi}\rangle$ is of the form

$$|\tilde{\Psi}\rangle = |\Phi\rangle \wedge |\Theta\rangle,$$

where $|\Theta\rangle$ is the state with modes v_j^{+} not filled and modes v_j^{-} filled for $j \geqslant K/2-M$, and where $|\Phi\rangle$ is a state on at most K modes. By choosing K as in the theorem statement, the ground state overlap can be lower bounded by $1-\delta$, and hence $|\tilde{\Psi}\rangle$ fulfills all the desired requirements. ■

In the case where H_{free} has a constant spectral gap ω, the exponential decay of the 1-RDM spectrum places a constant upper bound on the number of excitations for a good approximation to the ground state $|\Psi\rangle$. This leads to a space of polynomial dimension which may be readily searched via quantum phase estimation. If the matrix γ itself is known—but with no assumptions on the spectral gap ω—a similar argument can be applied. Although one cannot place a constant bound on the total number of excitations as before, knowing the eigenvectors of γ allows one to predict *which* modes are likely to be excited. Following the same approach as Corollary 2 in [23], this again leads to a search space of polynomial dimension. Based on the above discussion, we now restate and prove Theorem III 2.

Theorem C 2. Consider a quantum impurity problem with $M = O(1)$. Suppose that either $\omega = O(1)$, or we are given the covariance matrix γ for a ground state satisfying Eq. (C1). Then the quantum impurity problem can be solved by a quantum computer using $\text{poly}(N, \varepsilon^{-1})$ gates.

It is not necessary that γ is given precisely; for the proof below it is sufficient merely to have knowledge of an upper bound $\tilde{\gamma} \geqslant \gamma$ such that the spectrum of $\tilde{\gamma}$ decays exponentially as in Eq. (C1).

Proof. We may choose fermionic operators such that

$$H_{\text{free}} = \sum_j \epsilon_j b_j^{\dagger} b_j$$

with $0 \leqslant \epsilon_j \leqslant 1$. We represent the Hamiltonian on N qubits using the Jordan-Wigner transformation. If $\omega = O(1)$, by Theorem III 1, for constant $\varepsilon = 1/2$, for *constant* K there exists a ground state $|\Psi\rangle$ that has overlap

$$\langle\Psi|\tilde{\Psi}\rangle \geqslant \tfrac{1}{2}$$

for a state $|\tilde{\Psi}\rangle$ that is a superposition of Slater determinants with at most K excitations. We now consider the subspace V of the full Hilbert space, which is spanned by all states with at most K excitations. The dimension of this space is

$$\dim(V) = \sum_{k=0}^{K}\binom{N}{k}. \tag{C3}$$

For $N/2 \geqslant K$,

$$\sum_{k=0}^{K}\binom{N}{k} \leqslant K\binom{N}{K} \leqslant KN^{K} = \text{poly}(N).$$

For small $N/2 < K$, Eq. (C3) is bounded by $2^N \leqslant 2^{2K}$, which is constant. If we let τ denote the maximally mixed state on V, then

$$\langle\Psi|\tau|\Psi\rangle \geqslant \frac{|\langle\Psi|\tilde{\Psi}\rangle|^2}{\dim(V)},$$

which has at least inverse polynomial magnitude. The mixed state τ can be efficiently prepared on a quantum computer. In the qubit representation, it corresponds to the uniform mixture of standard basis states $|x_1,\ldots,x_N\rangle$, $x_1,\ldots,x_N \in \{0,1\}^N$ with Hamming weight $x_1+\cdots+x_N \leqslant K$. One can prepare this state by uniformly sampling from bit strings $x_1,\ldots,x_n$ with Hamming weight at most K, and then prepare $|x_1,\ldots,x_n\rangle$. Applying quantum phase estimation, using (approximate) time evolution along H and initial state τ now gives an efficient algorithm for computing the ground state energy to precision ε using $\text{poly}(N, \varepsilon^{-1})$ gates.

The case in which we are given γ (but no assumption on ω) proceeds similarly; it suffices to find an efficiently preparable state with at least inverse polynomial overlap.

First we note that, as discussed in Appendix B, we may without loss of generality restrict ourselves to the case in which $\omega = \Omega(\varepsilon)$. Let $x_1, x_2, \cdots \in \mathbb{C}^n$ be the orthonormal eigenvectors of the 1-RDM γ corresponding to eigenvalues $\sigma_1 \geqslant \sigma_2 \geqslant \cdots$, respectively. Now take $Q = \lceil 14M\log(2\omega^{-1})\rceil$, and define sets

$$I_1 = \{1,2,\ldots,Q\},$$
$$I_2 = \{Q+1,\ldots,2Q\},$$
$$I_3 = \{2Q+1,\ldots,3Q\},$$
$$\vdots$$

For each $s \in \mathbb{N}$, define the partial number operator

$$N_s = \sum_{j\in I_s} b(x_j)^{\dagger} b(x_j),$$

which counts how many of the modes corresponding to the subset I_s of eigenvectors of γ are excited. Note that the N_s mutually commute. We can upper-bound the expected value of these observables by

$$\langle\Psi|N_s|\Psi\rangle = \sum_{j\in I_s}\sigma_j \leqslant c_0 M\log(2\omega^{-1})e^{-s},$$

where c_0 is some universal constant, using Eq. (C1). For each s, let n_s be a random variable corresponding to the measurement distribution of the observable N_s induced by the state $|\Psi\rangle$, and let $R_s = c_0 M\log(2\omega^{-1})e^{-s/2}$. Then Markov's inequality implies

$$\mathrm{P}[n_s \geqslant R_s] \leqslant e^{-s/2},$$

and applying a union bound we see that, for any $S\in\mathbb{N}$,

$$\mathrm{P}[n_s \leqslant R_s \text{ for all } s\geqslant S] \geqslant 1-\sum_{s\geqslant S} e^{-s/2}.$$

Now choose $S_0\in\mathbb{N}$ such that

$$\sum_{s\geqslant S_0} e^{-s/2} \leqslant \frac{1}{2}$$

and

$$\frac{R_s}{Q} < \frac{1}{2} \text{ for all } s\geqslant S_0.$$

Note that such an S_0 can be chosen as a universal constant, independent of M, ω, and N. Then

$$\mathrm{P}[n_s \leqslant R_s \text{ for all } s\geqslant S_0] \geqslant \tfrac{1}{2}.$$

Equivalently, letting P denote the projection onto the subspace spanned by Fock basis states $|\Phi\rangle$ [with respect to the modes $b(x_j)$] satisfying $N_s|\Phi\rangle \leqslant R_s$ for all $s > S_0$,

$$\langle\Psi|P|\Psi\rangle \geqslant \tfrac{1}{2}.$$

In particular, this implies that the maximally mixed state on P, τ', has squared ground state overlap

$$\langle\Psi|\tau'|\Psi\rangle \geqslant \frac{1}{2\mathrm{Tr}[P]}.$$

By an identical combinatorial argument to the one presented in Corollary 2 of [23], $\mathrm{Tr}[P]$ can be bounded by $e^{O(m\log(\omega^{-1}))} = \mathrm{poly}(\varepsilon^{-1})$, completing the proof. ■

In fact, even in the absence of a spectral gap or knowledge of the 1-RDM as in Theorem C 3, one may still obtain a quasipolynomial quantum speedup for the impurity problem. In particular, whereas the classical algorithm of [23] requires time $\mathrm{poly}(N)\exp(O(\log(\varepsilon^{-1})^3))$, a naive quantum adaptation of this approach can reduce the time complexity to $\mathrm{poly}(N)\exp(O(\log(\varepsilon^{-1})^2))$. This speedup arises from the step analogous to Theorem C 1, in which the search space for candidate ground states is reduced to a limited set of active excitations. In the classical case, it is necessary to choose $\delta = O(\varepsilon)$ to ensure that $|\tilde{\Psi}\rangle$ approximates the ground state energy sufficiently well, however in the quantum case δ can be taken as constant for the phase estimation step. In the algorithm of [23], this leads to reduction of the search space dimension by a factor of $\log(\varepsilon)$. Although this is not sufficient to show BQP containment of the quantum impurity problem, it provides intuition that a polynomial time algorithm for the fully general case may be attainable through more detailed analysis.

APPENDIX D: NUMERICAL METHODS

Geometries of all systems were optimized with RI-DFT [111,112] and a Perdew, Burke, and Ernzerhof exchange-correlation functional (PBE) [113], D3 dispersion correction [114], and Becke-Johnson damping [115] using a def2-SVP basis set [116]. Structures were optimized with TURBOMOLE [117,118].

For the (oligo)tryptophan systems, HF molecular orbitals were obtained for the bootstrap embedding and for the sequence of tryptophan side chains. As we were interested in the valence shell for the exploration of multiconfigurational character, the molecular orbitals were first obtained using the def2-SVP basis set and then used to construct intrinsic bond orbitals [75,78] (IBOs) separately for occupied and virtual space. For the Huzinaga embedding [41,77] in the tryptophan and ruthenium system, Kohn-Sham orbitals calculated with PBE were used instead. IBOs were again constructed separately for the occupied and virtual space. These calculations were run with Serenity [119,120]. For the bootstrap embedding, Hartree-Fock molecular orbitals were obtained using the def2-SVP basis set. Following the Hartree-Fock calculation, intrinsic atomic orbitals [75] were constructed and used to represent the 1-RDM. From there, the procedure described in Ref. [121] was followed to obtain the fragment and the entangled bath orbitals. These calculations were run with the PySCF program [122–124].

Orbitals selected for the active space calculations in the case of the tryptophan sequence comprised the entire π system of the tryptophan side chain, as well as the σ and σ^* orbitals of the C–H bond in the five-membered ring that forms a weak CH-π hydrogen bond. For the embedding examples, single-orbital entropies, mutual information, and orbital ordering were obtained with autoCAS [66,125] and the active spaces were constructed by choosing K orbitals corresponding to the largest values.

Next, DMRG calculations with bond dimension $D = 1024$ were performed to obtain an approximation of the target ground state of the system as an MPS. Slater determinants with the leading contributions were constructed from the MPS using sampling-reconstruction of the complete active space (SR-CAS) [126] and used to construct the sum-of-Slater states. MPS wave functions of lower bond dimensions were obtained by truncating the bond dimension using the singular value decomposition. DMRG calculations were run with the QCMaquis software package [127].

[1] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proc. Natl. Acad. Sci. USA **114**, 7555 (2017).

[2] 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).

[3] 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).

[4] R. Santagati, A. Aspuru-Guzik, R. Babbush, M. Degroote, L. Gonzalez, E. Kyoseva, N. Moll, M. Oppel, R. M. Parrish, N. C. Rubin, M. Streif, C. S. Tautermann, H. Weiss, N. Wiebe, and C. Utschig-Utschig, Drug design on quantum computers, Nat. Phys. **20**, 549 (2024).

[5] 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. USA **119**, e2203533119 (2022).

[6] A. Baiardi, M. Christandl, and M. Reiher, Quantum computing for molecular biology, ChemBioChem **24**, e202300120 (2023).

[7] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*, 1st ed. (John Wiley & Sons, Ltd, Chichester, 2014).

[8] 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, and K. B. Whaley, Postponing the orthogonality catastrophe: Efficient state preparation for electronic structure simulations on quantum devices, arXiv:1809.05523.

[9] S. Fomichev, K. Hejazi, M. S. Zini, M. Kiser, J. F. Morales, P. A. M. Casares, A. Delgado, J. Huh, A.-C. Voigt, J. E. Mueller, and J. M. Arrazola, Initial state preparation for quantum chemistry on quantum computers, PRX Quantum **5**, 040339 (2024).

[11] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. Kin-Lic Chan, Low rank representations for quantum simulation of electronic structure, npj Quantum Inf. **7**, 83 (2021).

[10] 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).

[12] N. S. Blunt, G. P. Gehér, and A. E. Moylett, Compilation of a simple chemistry application to quantum error correction primitives, Phys. Rev. Res. **6**, 013325 (2024).

[13] 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).

[14] J. Kempe, A. Kitaev, and O. Regev, The complexity of the local Hamiltonian problem, SIAM J. Comput. **35**, 1070 (2006).

[15] B. O'Gorman, S. Irani, J. Whitfield, and B. Fefferman, Intractability of electronic structure in a fixed basis, PRX Quantum **3**, 020322 (2022).

[16] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, M. Kastoryano, R. Babbush, J. Preskill, D. R. Reichman, E. T. Campbell, E. F. Valeev, L. Lin, and G. Kin-Lic Chan, Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry, Nat. Commun. **14**, 1952 (2023).

[17] P. J. Ollitrault, C. L. Cortes, J. F. Gonthier, R. M. Parrish, D. Rocca, G.-L. Anselmetti, M. Degroote, N. Moll, R. Santagati, and M. Streif, Enhancing initial state overlap through orbital optimization for faster molecular electronic ground-state energy estimation, Phys. Rev. Lett. **133**, 250601 (2024).

[18] D. Marti-Dafcik, H. G. A. Burton, and D. P. Tew, Spin coupling is all you need: Encoding strong electron correlation on quantum computers, arXiv:2404.18878.

[19] Z. Ding and L. Lin, Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation, PRX Quantum **4**, 020331 (2023).

[20] H. Ni, H. Li, and L. Ying, On low-depth algorithms for quantum phase estimation, Quantum **7**, 1165 (2023).

[21] M. Reiher, Molecule-specific uncertainty quantification in quantum chemical studies, Isr. J. Chem. **62**, e202100101 (2022).

[22] H. Liu, G. H. Low, D. S. Steiger, T. Häner, M. Reiher, and M. Troyer, Prospects of quantum computing for molecular sciences, Mater. Theor. **6**, 11 (2022).

[23] S. Bravyi and D. Gosset, Complexity of quantum impurity problems, Commun. Math. Phys. **356**, 451 (2017).

[24] A. Warshel and M. Levitt, Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, J. Mol. Biol. **103**, 227 (1976).

[25] T. A. Wesolowski and A. Warshel, Frozen density functional approach for ab initio calculations of solvated molecules, J. Phys. Chem. **97**, 8050 (1993).

[26] H. M. Senn and W. Thiel, QM/MM methods for biomolecular systems, Angew. Chem. Int. Ed. **48**, 1198 (2009).

[27] B. Mennucci, Polarizable continuum model, WIREs Comput. Mol. Sci. **2**, 386 (2012).

[28] F. Libisch, C. Huang, and E. A. Carter, Embedded correlated wavefunction schemes: Theory and applications, Acc. Chem. Res. **47**, 2768 (2014).

[29] Q. Sun and G. Kin-Lic Chan, Quantum embedding theories, Acc. Chem. Res. **49**, 2705 (2016).

[30] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, Embedding methods for quantum chemistry: Applications from materials to life sciences, J. Am. Chem. Soc. **142**, 3281 (2020).

[31] C. R. Jacob and J. Neugebauer, Subsystem density-functional theory (update), WIREs Comput. Mol. Sci. **14**, e1700 (2024).

[32] Y. Liu, O. R. Meitei, Z. E. Chin, A. Dutt, M. Tao, I. L. Chuang, and T. V. Voorhis, Bootstrap embedding on a quantum computer, J. Chem. Theor. Comput. **19**, 2230 (2023).

[33] G. Knizia and G. Kin-Lic Chan, Density matrix embedding: A simple alternative to dynamical mean-field theory, Phys. Rev. Lett. **109**, 186404 (2012).

[34] S. Wouters, C. A. Jiménez-Hoyos, Q. Sun, and G. K.-L. Chan, A practical guide to density matrix embedding theory in quantum chemistry, J. Chem. Theor. Comput. **12**, 2706 (2016).

[35] H. Q. Pham, V. Bernales, and L. Gagliardi, Can density matrix embedding theory with the complete activate space self-consistent field solver describe single and double bond

breaking in molecular systems? J. Chem. Theor. Comput. **14**, 1960 (2018).
[36] S. Sekaran, M. Tsuchiizu, M. Saubanère, and E. Fromager, Householder-transformed density matrix functional embedding theory, Phys. Rev. B **104**, 035121 (2021).
[37] E. Cancès, F. Faulstich, A. Kirsch, E. Letournel, and A. Levitt, Some mathematical insights on density matrix embedding theory, arXiv:2305.16472.
[38] M. Welborn, T. Tsuchimochi, and T. Van Voorhis, Bootstrap embedding: An internally consistent fragment-based method, J. Chem. Phys. **145**, 074102 (2016).
[39] H.-Z. Ye, N. D. Ricke, H. K. Tran, and T. V. Voorhis, Bootstrap embedding for molecules, J. Chem. Theor. Comput. **15**, 4497 (2019).
[40] H.-Z. Ye, H. K. Tran, and T. V. Voorhis, Bootstrap embedding for large molecular systems, J. Chem. Theor. Comput. **16**, 5035 (2020).
[41] B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, Exact density functional and wave function embedding schemes based on orbital localization, J. Chem. Phys. **145**, 064107 (2016).
[42] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, A simple, exact density-functional-theory embedding scheme, J. Chem. Theor. Comput. **8**, 2564 (2012).
[43] S. Huzinaga and A. A. Cantu, Theory of separability of many-electron systems, J. Chem. Phys. **55**, 5543 (1971).
[44] D. Claudino and N. J. Mayhall, Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories, J. Chem. Theor. Comput. **15**, 1053 (2019).
[45] M. Bensberg and J. Neugebauer, Direct orbital selection for projection-based embedding, J. Chem. Phys. **150**, 214106 (2019).
[46] L. Mineh and A. Montanaro, Solving the Hubbard model using density matrix embedding theory and the variational quantum eigensolver, Phys. Rev. B **105**, 125117 (2022).
[47] J. Tilly, P. V. Sriluckshmy, A. Patel, E. Fontana, I. Rungger, E. Grant, R. Anderson, J. Tennyson, and G. H. Booth, Reduced density matrix sampling: Self-consistent embedding and multiscale electronic structure on current generation quantum computers, Phys. Rev. Res. **3**, 033230 (2021).
[48] Y. Kawashima, E. Lloyd, M. P. Coons, Y. Nam, S. Matsuura, A. J. Garza, S. Johri, L. Huntington, V. Senicourt, A. O. Maksymov *et al.*, Optimizing electronic structure simulations on a trapped-ion quantum computer using problem decomposition, Commun. Phys. **4**, 245 (2021).
[49] C. Cao, J. Sun, X. Yuan, H.-S. Hu, H. Q. Pham, and D. Lv, *Ab initio* quantum simulation of strongly correlated materials with quantum embedding, npj Comput. Mater. **9**, 78 (2023).
[50] A. Ralli, M. W. de la Bastida, and P. V. Coveney, Scalable approach to quantum simulation via projection-based embedding, Phys. Rev. A **109**, 022418 (2024).
[51] M. Rossmannek, F. Pavosevic, A. Rubio, and I. Tavernelli, Quantum embedding method for the simulation of strongly correlated systems on quantum computers, J. Phys. Chem. Lett. **14**, 3491 (2023).
[52] S. Battaglia, M. Rossmannek, V. V. Rybkin, I. Tavernelli, and J. Hutter, A general framework for active space embedding methods with applications in quantum computing, npj Comput. Mater. **10**, 297 (2024).
[53] C. Vorwerk, N. Sheng, M. Govoni, B. Huang, and G. Galli, Quantum embedding theories to simulate condensed systems on quantum computers, Nat. Comput. Sci. **2**, 424 (2022).
[54] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid quantum-classical approach to correlated materials, Phys. Rev. X **6**, 031045 (2016).
[55] P. W. Anderson and Anderson, Localized magnetic states in metals, Phys. Rev. **124**, 41 (1961).
[56] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, Rev. Mod. Phys. **47**, 773 (1975).
[57] P. W. Anderson, Infrared catastrophe in Fermi gases with local scattering potentials, Phys. Rev. Lett. **18**, 1049 (1967).
[58] A. H. Mühlbach and M. Reiher, Quantum system partitioning at the single-particle level, J. Chem. Phys. **149**, 184104 (2018).
[59] H. J. A. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, Second-order møller-plesset perturbation theory as a configuration and orbital generator in multiconfiguration self-consistent field calculations, J. Chem. Phys. **88**, 3834 (1988).
[60] J. M. Bofill and P. Pulay, The unrestricted natural orbital–complete active space (uno-cas) method: An inexpensive alternative to the complete active space-self-consistent-field (cas–scf) method, J. Chem. Phys. **90**, 3637 (1989).
[61] F. Krausbeck, D. Mendive-Tapia, A. J. W. Thom, and M. J. Bearpark, Choosing rasscf orbital active spaces for multiple electronic states, Comput. Theor. Chem. **1040-1041**, 14 (2014).
[62] S. Keller, K. Boguslawski, T. Janowski, M. Reiher, and P. Pulay, Selection of active spaces for multiconfigurational wavefunctions, J. Chem. Phys. **142**, 244104 (2015).
[63] Ö. Legeza and J. Sólyom, Optimizing the density-matrix renormalization group method using quantum information entropy, Phys. Rev. B **68**, 195116 (2003).
[64] J. Rissler, R. M. Noack, and S. R. White, Measuring orbital interaction using quantum information theory, Chem. Phys. **323**, 519 (2006).
[65] K. Boguslawski, P. Tecmer, Ö. Legeza, and M. Reiher, Entanglement measures for single- and multireference correlation effects, J. Phys. Chem. Lett. **3**, 3129 (2012).
[66] C. J. Stein and M. Reiher, Automated selection of active orbital spaces, J. Chem. Theor. Comput. **12**, 1760 (2016).
[67] N. Schuch and F. Verstraete, Computational complexity of interacting electrons and fundamental limitations of density functional theory, Nat. Phys. **5**, 732 (2009).
[68] S. Bravyi, Lagrangian representation for fermionic linear optics, Quantum Inf. Comp. **5**, 216 (2005).
[69] C.-F. Chen, A. M. Dalzell, M. Berta, F. G. S. L. Brandão, and J. A. Tropp, Sparse random Hamiltonians are quantumly easy, Phys. Rev. X **14**, 011014 (2024).
[70] M. B. Hastings and R. O'Donnell, Optimizing strongly interacting fermionic Hamiltonians, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (ACM, New York, 2022), pp. 776–789.
[71] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, Rev. Mod. Phys. **78**, 865 (2006).
[72] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, Rev. Mod. Phys. **83**, 349 (2011).

[73] R. Bulla, T. A. Costi, and T. Pruschke, Numerical renormalization group method for quantum impurity systems, Rev. Mod. Phys. **80**, 395 (2008).

[74] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Rev. Mod. Phys. **68**, 13 (1996).

[75] G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts, J. Chem. Theor. Comput. **9**, 4834 (2013).

[76] C. J. Stein and M. Reiher, Measuring multi-configurational character by orbital entanglement, Mol. Phys. **115**, 2110 (2017).

[77] D. V. Chulhai and J. D. Goodpaster, Projection-based correlated wave function in density functional theory embedding for periodic systems, J. Chem. Theor. Comput. **14**, 1928 (2018).

[78] B. Senjean, S. Sen, M. Repisky, G. Knizia, and L. Visscher, Generalization of intrinsic orbitals to kramers-paired quaternion spinors, molecular fragments, and valence virtual spinors, J. Chem. Theor. Comput. **17**, 1337 (2021).

[79] O. Weser, L. Freitag, K. Guther, A. Alavi, and G. L. Manni, Chemical insights into the electronic structure of Fe(ii) porphyrin using fciqmc, dmrg, and generalized active spaces, Int. J. Quantum Chem. **121**, e26454 (2021).

[80] R. Trondl, P. Heffeter, C. R. Kowol, M. A. Jakupec, W. Berger, and B. K. Keppler, NKP-1339, the first ruthenium-based anticancer drug on the edge to clinical application, Chem. Sci. **5**, 2925 (2014).

[81] W. Peti, T. Pieper, M. Sommer, B. K. Keppler, and G. Giester, Synthesis of tumor-inhibiting complex salts containing the aniontrans-tetrachlorobis(indazole)ruthenate(III) and crystal structure of the tetraphenylphosphonium salt, Eur. J. Inorg. Chem. **1999**, 1551 (1999).

[82] A. T. Macias, D. S. Williamson, N. Allen, J. Borgognoni, A. Clay, Z. Daniels, P. Dokurno, M. J. Drysdale, G. L. Francis, C. J. Graham *et al.*, Adenosine-derived inhibitors of 78 kDa glucose regulated protein (Grp78) ATPase: Insights into isoform selectivity, J. Med. Chem. **54**, 4034 (2011).

[83] L. P. Weisburn, M. Cho, M. Bensberg, O. R. Meitei, M. Reiher, and T. V. Voorhis, Multiscale embedding for quantum computing, arXiv:2409.06813.

[84] M. Altunbulak and A. Klyachko, The Pauli principle revisited, Commun. Math. Phys. **282**, 287 (2008).

[85] C. Schilling, M. Altunbulak, S. Knecht, A. Lopes, J. D. Whitfield, M. Christandl, D. Gross, and M. Reiher, Generalized Pauli constraints in small atoms, Phys. Rev. A **97**, 052503 (2018).

[86] I. Theophilou, N. N. Lathiotakis, M. A. L. Marques, and N. Helbig, Generalized pauli constraints in reduced density matrix functional theory, J. Chem. Phys. **142**, 154108 (2015).

[87] C. Schilling, C. L. Benavides-Riveros, A. Lopes, T. Maciążek, and A. Sawicki, Implications of pinned occupation numbers for natural orbital expansions: I. Generalizing the concept of active spaces, New J. Phys. **22**, 023001 (2020).

[88] F. M. Faulstich, R. Kim, Z.-H. Cui, Z. Wen, G. Kin-Lic Chan, and L. Lin, Pure state ν-representability of density matrix embedding theory, J. Chem. Theor. Comput. **18**, 851 (2022).

[89] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum metrology, Phys. Rev. Lett. **96**, 010401 (2006).

[90] N. S. Mande and R. de Wolf, Tight bounds for quantum phase estimation and related problems, arXiv:2305.04908.

[91] L. Lin and Y. Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, PRX Quantum **3**, 010318 (2022).

[92] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. Kin-Lic Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, Phys. Rev. Lett. **120**, 110501 (2018).

[93] E. Malvetti, R. Iten, and R. Colbeck, Quantum circuits for sparse isometries, Quantum **5**, 412 (2021).

[94] T. M. L. de Veras, L. D. da Silva, and A. J. da Silva, Double sparse quantum state preparation, Quant. Inf. Proc. **21**, 204 (2022).

[95] 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).

[96] S.-J. Ran, Encoding of matrix product states into quantum circuits of one- and two-qubit gates, Phys. Rev. A **101**, 032310 (2020).

[97] 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 (2024).

[98] 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).

[99] K. C. Smith, A. Khan, B. K. Clark, S. M. Girvin, and T.-C. Wei, Constant-depth preparation of matrix product states with adaptive quantum circuits, arXiv:2404.16083.

[100] L. Veis and J. Pittner, Adiabatic state preparation study of methylene, J. Chem. Phys. **140**, 214111 (2014).

[101] T. Albash and D. A. Lidar, Adiabatic quantum computation, Rev. Mod. Phys. **90**, 015002 (2018).

[102] C.-F. Chen, M. J. Kastoryano, F. G. S. L. Brandão, and A. Gilyén, Quantum thermal state preparation, arXiv:2303.18224.

[103] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, J. Chem. Theor. Comput. **15**, 311 (2019).

[104] W. Kohn, Nobel lecture: Electronic structure of matter—wave functions and density functionals, Rev. Mod. Phys. **71**, 1253 (1999).

[105] A. V. Sobolev, Eigenvalue asymptotics for the one-particle density matrix, Duke Math. J. **171**, 3481 (2022).

[106] P. Hearnshaw and A. V. Sobolev, Analyticity of the one-particle density matrix, Ann. Henri Poincaré **23**, 707 (2022).

[107] P. Hearnshaw and A. V. Sobolev, The diagonal behaviour of the one-particle Coulombic density matrix, Probab. Math. Phys. **4**, 935 (2023).

[108] C. Cade, M. Folkertsma, S. Gharibian, R. Hayakawa, F. L. Gall, T. Morimae, and J. Weggemans, Improved hardness results for the guided local Hamiltonian problem, In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)* (Schloss Dagstuhl-Leibniz-Zentrum für Informatik, Wadern, 2023).

[109] J. Weggemans, M. Folkertsma, and C. Cade, Guidable local Hamiltonian problems with implications to heuristic ansatze state preparation and the quantum PCP conjecture, arXiv:2302.11578.

[110] S. Bravyi, D. Gosset, R. König, and K. Temme, Approximation algorithms for quantum many-body problems, J. Math. Phys. **60**, 032203 (2019).

[111] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Auxiliary basis sets to approximate Coulomb potentials (Chem. Phys. Letters 240 (1995) 283-290), Chem. Phys. Lett. **242**, 652 (1995).

[112] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials, Theor. Chem. Acc. **97**, 119 (1997).

[113] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[114] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. **132**, 154104 (2010).

[115] S. Grimme, S. Ehrlich, and L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, J. Comput. Chem. **32**, 1456 (2011).

[116] F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[117] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Electronic structure calculations on workstation computers: The program system turbomole, Chem. Phys. Lett. **162**, 165 (1989).

[118] TURBOMOLE V7.0.2 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

[119] J. P. Unsleber, T. Dresselhaus, K. Klahr, D. Schnieders, M. Böckers, D. Barton, and J. Neugebauer, Serenity: A subsystem quantum chemistry program, J. Comput. Chem. **39**, 788 (2018).

[120] N. Niemeyer, P. Eschenbach, M. Bensberg, J. Tölle, L. Hellmann, L. Lampe, A. Massolle, A. Rikus, D. Schnieders, J. P. Unsleber, and J. Neugebauer, The subsystem quantum chemistry program SERENITY, WIREs Comput. Mol. Sci. **13**, e1647 (2023).

[121] H.-Z. Ye, H. K. Tran, and T. V. Voorhis, Accurate electronic excitation energies in full-valence active space via bootstrap embedding, J. Chem. Theor. Comput. **17**, 3335 (2021).

[122] 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, S. Wouters, and G. Kin-Lic Chan, Pyscf: The python-based simulations of chemistry framework, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[123] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen *et al.*, Recent developments in the pyscf program package, J. Chem. Phys. **153**, 024109 (2020).

[124] Q. Sun, Libcint: An efficient general integral library for gaussian basis functions, J. Comput. Chem. **36**, 1664 (2015).

[125] M. Bensberg, M. Mörchen, C. J. Stein, J. P. Unsleber, T. Weymuth, and M. Reiher, "qcscine/autocas: Release 2.3.0 (version 2.3.0)," Zenodo (2024), doi:10.5281/zenodo.11440088.

[126] K. Boguslawski, K. H. Marti, and M. Reiher, Construction of casci-type wave functions for very large active spaces, J. Chem. Phys. **134**, 224101 (2011).

[127] S. Keller, M. Dolfi, M. Troyer, and M. Reiher, An efficient matrix product operator representation of the quantum chemical hamiltonian, J. Chem. Phys. **143**, 244118 (2015).