JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 3510−3521

pubs.acs.org/JCTC

Efficient Ab Initio Auxiliary-Field Quantum Monte Carlo Calculations in Gaussian Bases via Low-Rank Tensor Decomposition

Mario Motta,[*,†] James Shee,[‡] Shiwei Zhang,[§,¶] and Garnet Kin-Lic Chan[*,†]

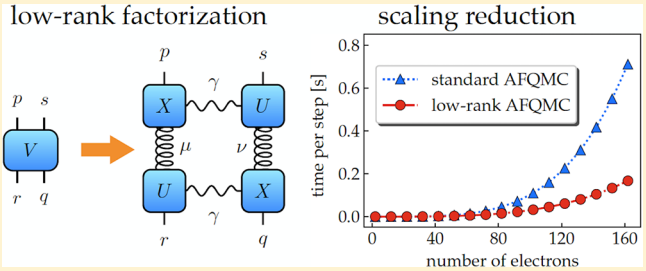
[†]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States

[‡]Department of Chemistry, Columbia University, New York, New York 10027, United States

[¶]Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, United States

[§]Department of Physics, College of William and Mary, Williamsburg, Virginia 23187-8795, United States

ABSTRACT: We describe an algorithm to reduce the cost of auxiliary-field quantum Monte Carlo (AFQMC) calculations for the electronic structure problem. The technique uses a nested low-rank factorization of the electron repulsion integral (ERI). While the cost of conventional AFQMC calculations in Gaussian bases scales as $O(N^4)$, where N is the size of the basis, we show that ground-state energies can be computed through tensor decomposition with reduced memory requirements and subquartic scaling. The algorithm is applied to hydrogen chains and square grids, water clusters, and hexagonal BN. In all cases, we observe significant memory savings and, for larger systems, reduced, subquartic simulation time.



1. INTRODUCTION

Correlated electronic structure calculations often require one to store and manipulate tensors that have high rank and act on vector spaces of high dimension. Frequently, the input−output and algebraic operations involving such high-rank tensors constitute a computational bottleneck of the calculations.

The cost of tensor manipulations and storage can be significantly reduced by low-rank decompositions,[1−5] in which a higher-rank tensor is represented by contractions of lower-rank tensors. The most common tensor appearing in Gaussian basis calculations is the rank-4 electron-repulsion integral (ERI)

$$V_{prqs} = \int d\mathbf{r}d\mathbf{r}' \chi_p(\mathbf{r})\chi_q(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}\chi_r(\mathbf{r})\chi_s(\mathbf{r}') \tag{1}$$

where the real-valued Gaussian atomic orbitals (AOs) $\{\chi_p(\mathbf{r})\}_{p=1}^{M}$ form a nonorthogonal basis for the one-electron Hilbert space. In eq 1, and in the remainder of the paper, we employ atomic units (a.u.) unless otherwise specified. Density-fitting (DF)[1,6−8] and modified Cholesky (CD)[2,9,10] are commonly applied to obtain a low-rank decomposition of the ERI in the AO basis in terms of a rank-3 tensor L^{γ}_{pr}

$$v_{prqs} \simeq \sum_{\gamma=1}^{N_\gamma} L^{\gamma}_{pr} L^{\gamma}_{qs} \tag{2}$$

(To obtain the form eq 2 in DF, one can apply a Cholesky decomposition or eigenvalue decomposition to the inverse density fitting metric, as is done in density fitted exchange algorithms.[11,12]) Importantly, it is known that the error in such approximations of the ERI decays exponentially with the number of vectors N_γ and requires only $M = O(N)$ vectors for a fixed error per atom as a function of increasing system size.[13] Using the DF or CD approximations reduces the cost of storing the ERI from $O(N^4)$ to $O(N^3)$,[13] although the computational scaling of most electronic structure methods using DF or CD integrals is not changed.

More recently, several strategies to represent the ERI by contractions of rank-2 tensors have been introduced. One well-known scheme is tensor hyper-contraction, which[4,14−18] unlike CD or DF, can be used to obtain lower-computational scaling in many different electronic structure methods, including coupled-cluster[17,19−22] and Møller−Plesset perturbation theory.[18] Another recently proposed scheme is nested matrix diagonalization, introduced in ref 23. This has been used to improve quantum computing algorithms for simulating the electronic structure Hamiltonian.[24]

In the present work, we explore nested matrix diagonalization in the context of the auxiliary-field quantum Monte Carlo (AFQMC) method in a Gaussian basis.[25−27] While the cost of conventional Gaussian basis AFQMC scales as $O(N^4)$ even after using CD or DF,[28] we find that low-rank nested matrix diagonalization reduces the computational complexity to subquartic (asymptotically cubic), while retaining the $O(N^3)$ storage cost of CD and DF. As we show, this is because nested matrix diagonalization effectively implements a form of integral screening, by exposing it as a low-rank tensor structure. While cubic scaling is only achievable for very large systems, in the applications presented we always observe a reduction of computation time and subquartic scaling.

Received: October 3, 2018
Published: May 15, 2019

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

It is well-known that cubic computational scaling is also achieved in AFQMC calculations with plane-waves as the one-electron basis, where the ERI is naturally represented in a factorized form, and the fast Fourier transform leads to the reduced scaling.[29] However, in most scenarios, Gaussian basis sets are more compact than plane-wave bases.[30] Thus, the current algorithm has the potential to exhibit reduced computational times because of a smaller prefactor than plane wave implementations.

The rest of the paper is organized as follows. In Section 2, we provide a brief description of the AFQMC method. In Section 3, we describe some of the properties of the nested matrix decomposition and show how a low-rank approximation can be used to accelerate the most expensive part of an AFQMC simulation, namely, the calculation of the local energy. In Section 4, we assess the performance and accuracy of AFQMC calculations using Gaussian bases and low-rank decompositions, and conclusions are drawn in Section 5.

2. THE AFQMC METHOD

In this section, we introduce the AFQMC method and illustrate that the origin of its quartic cost for general electronic structure problems lies in the local energy calculation. Throughout the rest of the paper, we use letters $pqrs$ to indicate a general basis function χ_p (part of an orthogonal or nonorthogonal set over the range $1\cdots N$), $ijkl$ for particles (indices range from $1\cdots O$), $\gamma\mu\nu$ for auxiliary indices associated with the low-rank decompositions (range $1\cdots M$ for γ, $1\cdots \rho_\gamma$ for $\mu\nu$). Spin labels are suppressed for compactness.

AFQMC[25,27] is a projective quantum Monte Carlo (QMC) method, which estimates the ground-state properties of a many-Fermion system by statistically sampling the ground-state wave function

$$|\Psi_\beta\rangle = \frac{e^{-\beta\hat{H}}|\Phi_T\rangle}{\langle\Phi_T|e^{-\beta\hat{H}}|\Phi_T\rangle} \xrightarrow{\beta\to\infty} \frac{|\Psi_0\rangle}{\langle\Phi_T|\Psi_0\rangle} \tag{3}$$

In eq 3, Ψ_0 is the ground-state wave function of the system, Φ_T is an initial wave function not orthogonal to Ψ_0, which for simplicity we assume to be a single Slater determinant, and $\hat{H}$ is the Hamiltonian of the system, which without loss of generality[27] can be written in the form

$$\hat{H} = E_0 + \sum_{pq} t_{pq}\hat{E}_{pq} + \frac{1}{2}\sum_{prqs} V_{prqs}\hat{E}_{pr}\hat{E}_{qs} \tag{4}$$

$\hat{H}$ is composed of a constant term, a one-body part written in terms of the excitation operator $\hat{E}_{pq} = \hat{a}_p^\dagger \hat{a}_q$, and a two-body part. The underlying single-particle basis in eq 4 must be an orthonormal basis. Thus, when employing a Gaussian AO basis, the AO ERI in eq 1 must first be transformed to an orthogonal basis, as must the DF or CD vectors in the decomposition (eq 2). Using the transformed CD vectors, the two-body part can be written as a sum of squares of one-body operators

$$\sum_{prqs} V_{prqs}\hat{E}_{pr}\hat{E}_{qs} = \sum_\gamma \hat{v}_\gamma^2, \quad \hat{v}_\gamma = \sum_{pr} L_{pr}^\gamma \hat{E}_{pr} \tag{5}$$

These are illustrated in Figure 1a,b. For sufficiently large β, expectation values computed over Ψ_β yield ground-state averages. AFQMC projects Ψ_T toward Ψ_0 iteratively, writing

$$e^{-\beta\hat{H}} = (e^{-\Delta\tau\hat{H}})^n \tag{6}$$

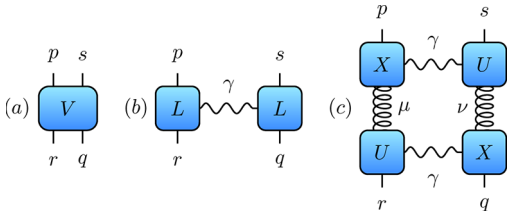


Figure 1. Pictorial illustrations (a) of the rank-4 electron repulsion integral (ERI) tensor V_{prqs}, (b) of its Cholesky (CD) or density-fitting (DF) decomposition $V_{prqs} = \sum_{\gamma=1}^{M} L_{pr}^\gamma L_{qs}^\gamma$, and (c) of the low-rank decomposition $V_{prqs} = \sum_{\gamma=1}^{M}\sum_{\mu\nu=1}^{\rho_\gamma} X_p^{\gamma\mu} U_r^{\gamma\mu} X_q^{\gamma\nu} U_s^{\gamma\nu}$ used in the present work. Lines emerging from colored blocks indicate free indices, and lines connecting blocks, indices summed over. Approximate decompositions in (b, c) break the original ERI into tensors of low rank, decreasing the memory requirements and cost to evaluate the local energy.

where $\Delta\tau = \frac{\beta}{n}$ is a small imaginary-time step. The propagator is represented, through a Hubbard−Stratonovich transformation,[31,32] as

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x} p(\mathbf{x})\hat{B}(\mathbf{x}) \tag{7}$$

where

$$\hat{B}(\mathbf{x}) = \exp\left(-\Delta\tau\hat{H}_1 + i\sqrt{\Delta\tau}\sum_{\gamma=1}^{N_\gamma} x_\gamma \hat{v}_\gamma\right) \tag{8}$$

is an independent-particle propagator that depends on the vector of fields $\mathbf{x}$, $p(\mathbf{x})$ is the standard normal M-dimensional probability distribution, and $\hat{H}_1 = \sum_{pq} t_{pq}\hat{E}_{pq}$ is the one-body part of $\hat{H}$. The representation (eq 7) maps the original interacting many-Fermion system onto an ensemble of non-interacting systems subject to a fluctuating potential. The imaginary-time projection can be realized as an open-ended random walk over paths of auxiliary-fields $\mathbf{x}$.[25] Importance sampling the trajectories of the random walk leads to a representation of Ψ_β as a stochastic weighted average of Slater determinants

$$|\Psi_\beta\rangle \simeq \frac{1}{\sum_w W_w}\sum_w W_w \frac{|\Phi_w\rangle}{\langle\Phi_T|\Phi_w\rangle} \tag{9}$$

Because the phase in $\hat{v}_\gamma$ can be complex for general two-body interactions, AFQMC suffers from a phase problem. This can be controlled using a trial state $|\Phi_T\rangle$ and imposing the phaseless approximation (Ph) and a real local energy estimator;[25,27] the error of these approximations vanishes if the trial state is exact.

The accuracy of Ph-AFQMC calculations of ground- and excited-state energies has been extensively benchmarked both in ab initio studies[33−36] and lattice models of correlated electrons.[37,38] The random walks take place in the overcomplete manifold of Slater determinants, in which Fermion antisymmetry is maintained by construction in each walker. Recently, the Ph-AFQMC has also been extended to the calculation of general ground-state properties, energy differences, and interatomic forces in realistic materials.[39−41]

In ab initio computations, the electron repulsion integrals entering into the AFQMC calculation can be obtained in different computational bases, such as plane-waves and pseudopotentials[25,42] or Gaussian type orbitals.[26] This choice of representation is important because it affects the cost of the AFQMC algorithm. When plane-waves are used, the standard

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

AFQMC methodology is known to scale as $\tilde{O}(N^3)$, as documented in Appendix A. (Note: Here we use the soft-O notation, well-established in complexity theory: $g(x) = \tilde{O}(f(x))$ if there exists an integer k such that $g(x) = O(f(x)\log^k(x))$.) When using a Gaussian basis, on the other hand, state-of-the-art calculations feature $O(N^4)$ cost. The computational bottleneck in both cases tends to be the local energy calculation, which we describe below.

2.1. Local Energy Calculation. AFQMC calculations require the computation of the following local energy functional for each sample

$$\mathcal{E}_{loc}(\Phi) = \frac{\langle \Phi_T|\hat{H}|\Phi\rangle}{\langle \Phi_T|\Phi\rangle} \tag{10}$$

from which the total energy is obtained as $E = \sum_w W_w \mathcal{E}_{loc}(\Phi_w)$. The local energy is also needed to determine the weights in Ph calculations.[25,27,34] The most demanding part of its calculation comes from the two-body term $\hat{H}_2$ which, from the generalized Wick's theorem,[43] can be written as

$$2\mathcal{E}_{loc,2}(\Phi) = 2\frac{\langle \Phi_T|\hat{H}_2|\Phi\rangle}{\langle \Phi_T|\Phi\rangle} = \\ = \sum_{prqs} V_{prqs}(G_{pr}G_{qs} - G_{ps}G_{qr}) \tag{11}$$

where the one-body reduced density matrix (RDM1)

$$G_{pr} = \frac{\langle \Phi_T|\hat{a}_p^\dagger \hat{a}_r|\Phi\rangle}{\langle \Phi_T|\Phi\rangle} = [\Phi(\Phi_T\Phi)^{-1}\Phi_T]_{rp} \\ = \sum_i \Theta_{ri}\Phi_{T_{ip}} \tag{12}$$

is defined in terms of the matrices Φ (of dimension $N \times O$) and Φ_T ($O \times N$) parametrizing the Slater determinant and trial wave function, respectively

$$|\Phi\rangle = \prod_i \hat{a}_{\phi_i}^\dagger|\varnothing\rangle \quad , \quad |\phi_i\rangle = \sum_p \Phi_{pi}|\chi_p\rangle \\ \langle\Phi_T| = \langle\varnothing|\prod_i \hat{a}_{\phi_i^T} \quad , \quad \langle\phi_i^T| = \sum_p \Phi_{T_{ip}}\langle\chi_p| \tag{13}$$

In eq 12, $\Theta = \Phi\,(\Phi_T\,\Phi)^{-1}$. Note that the expression eq 12 for the RDM1 sample resembles the expression for the RDM1 of the trial Slater determinant, $G_T = \Phi_T^\dagger\,\Phi_T$, with one Φ_T matrix (walker independent) replaced by Θ (dependent on the walker). Explicit evaluation of eq 12 costs $O(ON^2)$ per sample while the summation in the two-body local energy costs $O(N^4)$ per sample. For $N \gg O$, it is more efficient[26] to first contract the two-body matrix elements with Φ_T

$$\bar{V}_{irjs} = \sum_{pq} \Phi_{T_{ip}}\Phi_{T_{jq}}V_{prqs} \tag{14}$$

which may be carried out once and stored at the start of the AFQMC calculation at a cost of $O(ON^4 + O^2N^3)$. The local energy then follows as the sum

$$2\mathcal{E}_{loc,2}(\Phi) = \sum_{irjs} \bar{V}_{irjs}(\Theta_{ri}\Theta_{sj} - \Theta_{si}\Theta_{rj}) \tag{15}$$

at a cost of $O(O^2N^2)$ per sample. When memory is not a limitation, this is the most efficient conventional algorithm for local energy evaluation and is the one we compare against here. As mentioned in the introduction, the Cholesky decomposition (eq 2) allows one to significantly reduce the storage requirements by replacing the 4-index integrals by a truncated set of 3-index quantities. However, it does not reduce the computational cost of local energy evaluation. Inserting eq 12 into eq 11 and using the CD form in (eq 2) (after transformation to an orthogonal basis) gives

$$2\mathcal{E}_{loc,2}(\Phi) = \sum_{ij\gamma} f_{ii}^\gamma f_{jj}^\gamma - f_{ij}^\gamma f_{ij}^\gamma \tag{16}$$

with the intermediate f_{ij}^γ defined as

$$f_{ij}^\gamma = \sum_{pr} (\Phi_{T_{ip}}L_{pr}^\gamma)\Theta_{rj} \tag{17}$$

This is computed most efficiently by precomputing and storing the quantity in brackets, $\bar{L}_{ir}^\gamma = \sum_p \Phi_{Tip}\,L_{pr}^\gamma$, at the beginning of the AFQMC run, at cost $O(ON^2M)$, and subsequently carrying out the second contraction for each sample with $O(O^2NM) \sim O(N^4)$ cost. However, as $M > N$, the reduced memory cost afforded by CD is offset by an increased computational cost of the local energy evaluation, compared with the conventional algorithm in (11). The operations described so far are illustrated diagrammatically in Figure 2.

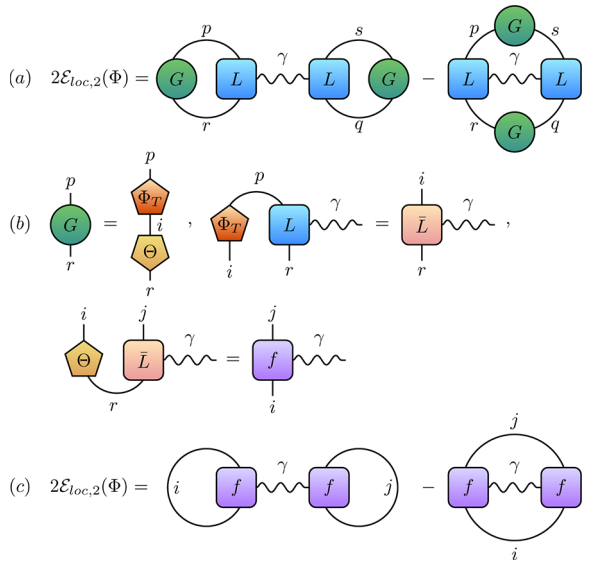


Figure 2. (a) Pictorial representation of the local energy calculation based on the CD decomposition of the ERI. (b) Separable structure of the RDM1, as used in precomputing the tensors $\bar{L}$ and f. (c) Expression of the local energy based on the f tensor.

To overcome this increased cost, we now describe how we can exploit additional structure in the Cholesky vector L_{pr}^γ.

3. LOW-RANK FACTORIZATION VIA NESTED MATRIX DIAGONALIZATION AND ACCELERATION OF LOCAL ENERGY EVALUATION

Ref 23 introduced a truncated nested matrix diagonalization, corresponding to an additional truncated factorization of L^γ. This starts from a truncated CD of the ERI, eq 2, such that all elements of the residual

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

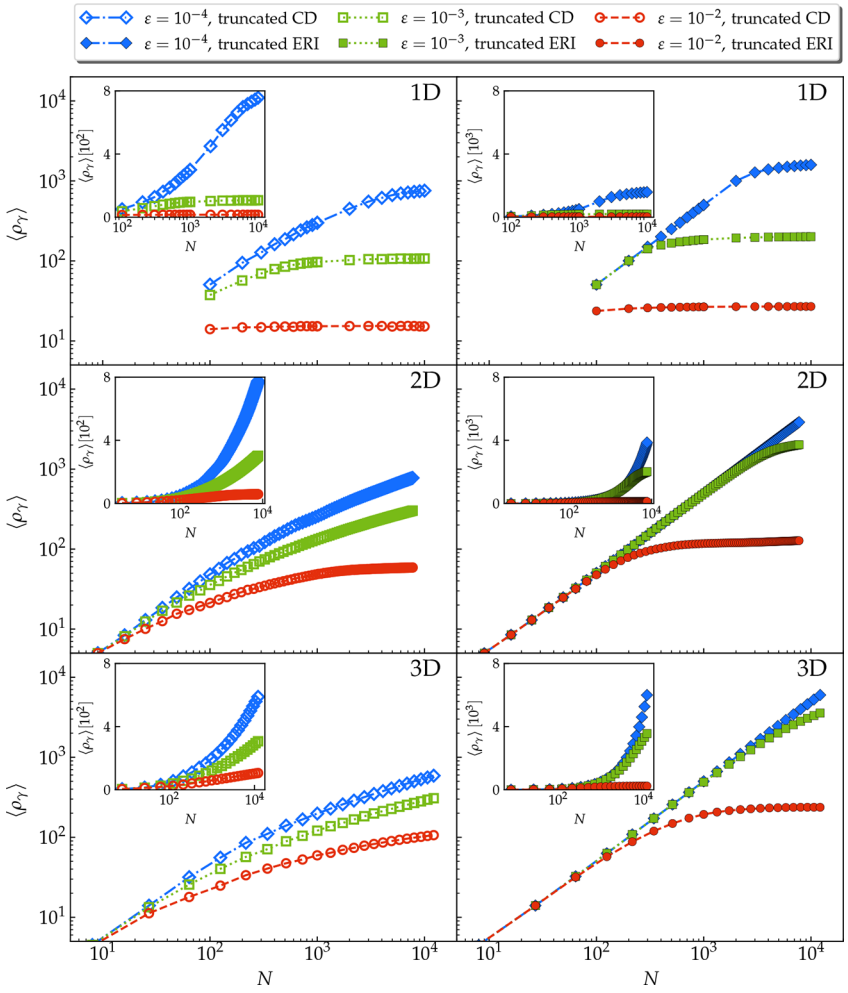


Figure 3. Main plots: Growth of the average number of eigenvalues $\langle \rho_\gamma \rangle$, on a log−log scale, for the model electron repulsion integral (eq 22) in 1D, 2D, 3D (top to bottom), using $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ (red, green, blue). The parameters in eq 22 are $c, V_0 = 1/2, 1$. We either perform a truncated CD over the exact ERI (left, empty symbols) or an untruncated CD over the truncated ERI (right, filled symbols). Note that on the log−log scale, the slope α gives $\langle \rho_\gamma \rangle \sim N^\alpha$; in all cases, $\alpha < 1$ and asymptotically approaches 0. Insets: the same quantities, on a log−linear scale.

$$R_{prqs} = V_{prqs} - \sum_\gamma L^\gamma_{pr} L^\gamma_{qs} \tag{18}$$

are kept smaller in absolute value than a predefined threshold ε_{CD}. Note that, since $|R_{prqs}| \le R_{prpr}$,[2] bounding R requires computing and bounding its diagonal only.

Then, after transformation to an orthogonal basis, we carry out an eigenvalue decomposition of the matrix L^γ_{pr} for each γ

$$L^\gamma_{pr} = \sum_\mu U^\gamma_{p\mu} \sigma^\gamma_\mu U^\gamma_{r\mu} \tag{19}$$

and only eigenvalues larger in absolute value than a predefined threshold ε_{ET} are kept, $|\sigma^\gamma_\mu| \ge \varepsilon_{ET}$. This additional eigenvalue truncation (ET) leads to the approximation

$$\begin{aligned} V_{prqs} &\simeq \sum_\gamma \sum_{\mu\nu} (U^\gamma_{p\mu}\sigma^\gamma_\mu) U^\gamma_{r\mu} (U^\gamma_{q\nu}\sigma^\gamma_\nu) U^\gamma_{s\nu} \\ &= \sum_\gamma \sum_{\mu\nu} X^\gamma_{p\mu} U^\gamma_{r\mu} X^\gamma_{q\nu} U^\gamma_{s\nu} \end{aligned} \tag{20}$$

where $\rho_\gamma \le N$ is the number of retained eigenvalues for the matrix L^γ and $X^\gamma_{p\mu} = U^\gamma_{p\mu}\ \sigma^\gamma_\mu$. The decomposition (eq 20) is diagrammatically illustrated in Figure 1c.

In ref 23, it was suggested, without detailed analysis, that the average number of eigenvalues of the Cholesky vectors

$$\langle \rho_\gamma \rangle = \frac{1}{M} \sum_\gamma \rho_\gamma \tag{21}$$

grows logarithmically with increasing system size.[23] In fact, we have found that the data presented in ref 23 can be fit just as well by a variety of functional forms, including by αN^β with $\beta \sim 1/2$. However, we now argue that in large systems, $\langle \rho_\gamma \rangle \to \tilde{O}(1)$, because the number of eigenvalues above a given threshold is related to $\frac{1}{N}$ times the number of Coulomb integrals above an integral threshold, which is asymptotically $O(1)$. To see this, assume the Gaussian basis has minimum exponent α. Then, on length scales longer than $\alpha^{-1/2}$, we can simplify V_{prqs} to a two-index quantity

$$V_{PQ} = \begin{cases} c|P - Q|^{-1} & P \ne Q \\ V_0 & P = Q \end{cases} \tag{22}$$

Here V_0 controls the strength of the on-site repulsion and c is related to the maximum exponent of the basis. In particular, c controls the rate at which V_{PQ} decays to zero so that, for smaller

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

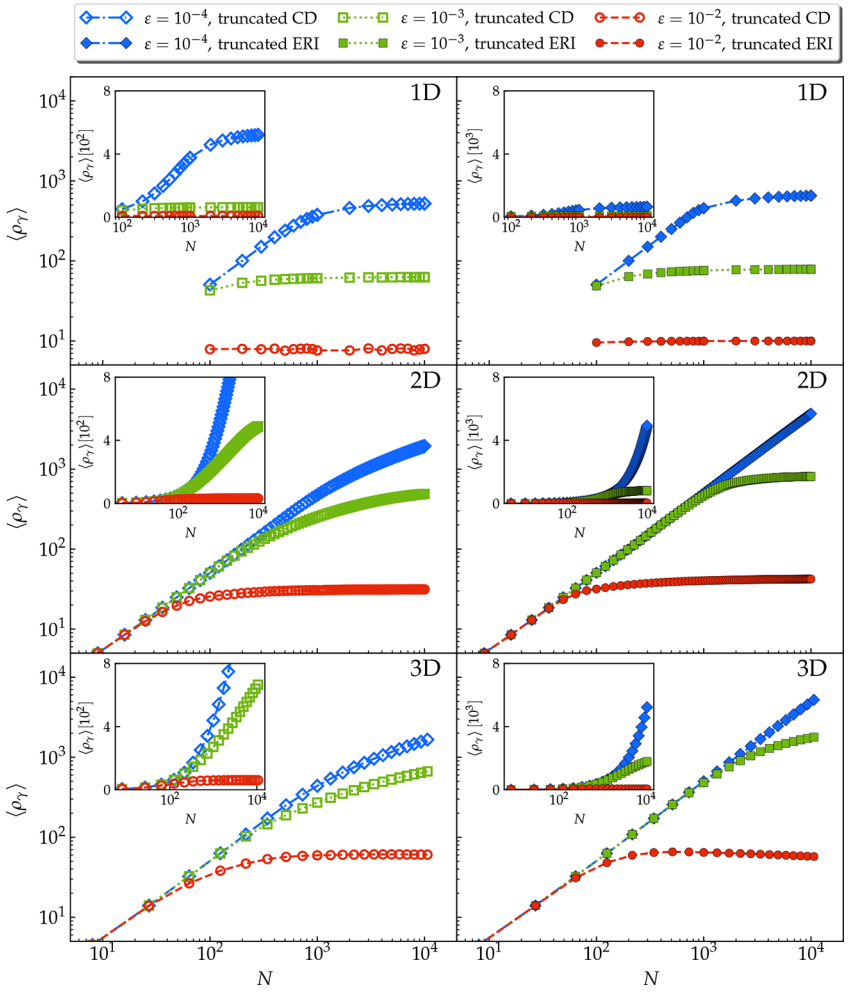


Figure 4. Same as Figure 3 but with parameters c, $V_0 = 1/10, 1$.

c, V_{PQ} is a more diagonally dominant matrix, and saturations of $\langle\rho_\gamma\rangle$ is reached faster. The Cholesky decomposition $V_{PQ} = \sum_\gamma L_P^\gamma L_Q^\gamma$ yields Cholesky vectors that are already "diagonal" for each γ (in other words, eq 22 expresses that V_{prqs} is approximately nonzero for $p = r$ and $q = s$). Thus, eigenvalue truncation in eq 20 truncates elements of the Cholesky vectors by absolute value, and the average number of significant eigenvalues ρ_γ is the average number of significant elements of the Cholesky vectors L_P^γ (controlled by the magnitude of c).

We now introduce a simple model to understand the behavior of ρ_γ. First consider a one-dimensional lattice of Gaussian functions (e.g., 1s functions) evenly spaced for simplicity. If integral screening is used with threshold ε, we truncate V_{PQ} in such a way as to obtain a banded matrix of width $w(\varepsilon)$. Then, the Cholesky vectors are also strictly banded, that is, $L_P^\gamma = 0$, $|\gamma - P| > w(\varepsilon)$, and we rigorously obtain $\rho_\gamma = O(1)$. This is not precisely a statement about truncating Cholesky elements of the full (untruncated) Coulomb matrix, but as we see in Figure 3 and Figure 4, the behavior of $\langle\rho_\gamma\rangle$ in these two settings is exactly the same. For V_{PQ} corresponding to a general graph, the number of nonzeros of the Cholesky vectors above a threshold is well-studied as the problem of fill-in generated by threshold-based incomplete Cholesky factorization.[44] While rigorous bounds are difficult to prove, we numerically compute $\langle\rho_\gamma\rangle$ for 2D and 3D cubic lattices for V_{PQ} truncated by a threshold ε and for the untruncated V_{PQ}. The behavior is very similar in both cases. While we cannot rule out nested logarithmic factors such as $\log(N)$ or $\log(\log(N))$, this numerical evidence strongly suggests that $\langle\rho_\gamma\rangle$ saturates at $\tilde{O}(1)$, just as it does in 1D. Finally, the same numerical behavior can be seen when decomposing the 4-index integral tensor, which we show in Figure 5 for hydrogen chains, where we can reach sufficiently large sizes to see saturation unambiguously for sufficiently large thresholds. For a fixed truncation accuracy and up to possible logarithmic factors, we consider the evidence to be strong that $\langle\rho_\gamma\rangle$ saturates to become independent of system size.

3.1. Accelerated Local Energy Evaluation. The low-rank structure revealed in the Cholesky vectors directly reduces the computational and memory costs of the AFQMC algorithm. In the present work, we choose $\varepsilon_{CD} = \varepsilon_{ET}$, although the two thresholds can in principle be chosen separately.[23]

In the case of the local energy, the intermediate f_{ij}^γ can be built as (see also Figure 6)

$$f_{ij}^\gamma = \sum_{pr\mu} (\Phi_{T_{ip}} U_{p\mu}^\gamma \sigma_\mu^\gamma)(U_{r\mu}^\gamma \Theta_{rj}) = \sum_\mu A_i^{\gamma\mu} B_j^{\gamma\mu} \tag{23}$$

where A can be evaluated at the beginning of the AFQMC run with cost $O(NM\langle\rho_\gamma\rangle)$, B is evaluated for each sample with cost $O(ONM\langle\rho_\gamma\rangle)$, and the assembly into f_{ij}^γ is $O(O^2M\langle\rho_\gamma\rangle)$ per

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

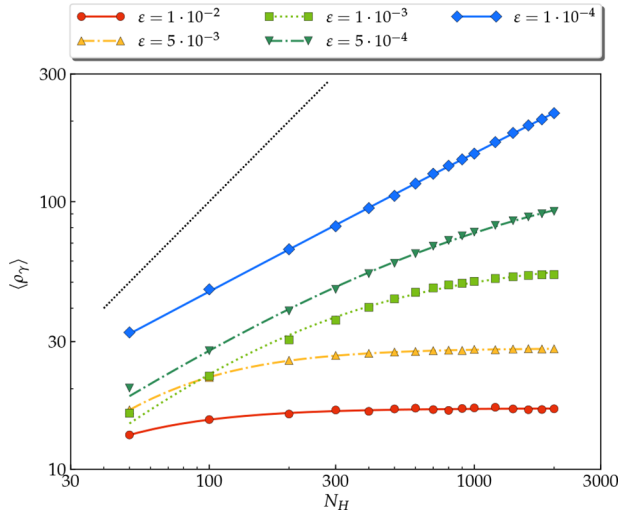


Figure 5. Average number $\langle \rho_\gamma \rangle$ of eigenvalues for H chains on a log−log scale, at the representative bond length R = 1.8 a_B, at the STO-6G level, using thresholds ε between 10^{-4} and 10^{-2} a.u. The black dotted line represents the number N of basis functions, providing an upper bound for $\langle \rho_\gamma \rangle$. Colored lines are the result of a fit to $\frac{x^\alpha}{\beta + \gamma x^\alpha}$. Sublinear growth is visible in all cases, and saturation is reached for the looser thresholds.

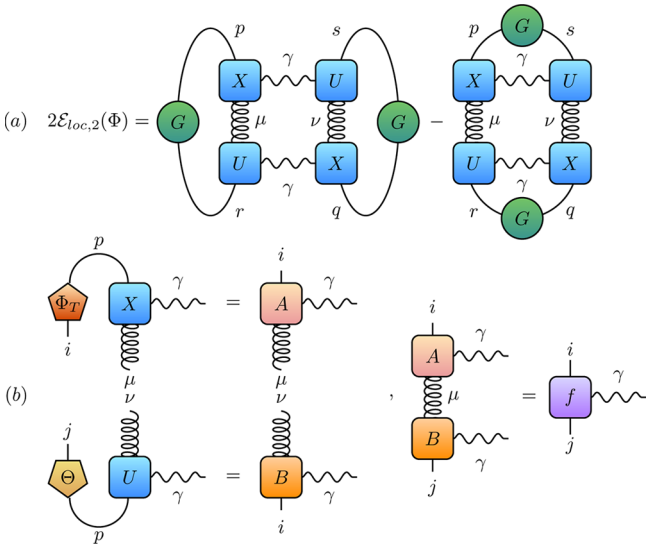


Figure 6. Pictorial illustrations (a) of the local energy calculation based on the CD+ET decomposition of the ERI, (b) of the precomputed and intermediate tensors involved in the calculation. The final expression for the local energy coincides with the one in Figure 2c.

sample. For a sublinear (constant) $\langle \rho_\gamma \rangle$, this gives subquartic (cubic) cost for the energy evaluation.

The memory reduction from the low-rank factorization is shown in Figure 7, where the ratio between the size of the tensors $\bar{V}$ and A, B is shown for hydrogen chains and grids. As seen, for a large system, the size of A, B is only ≃5% of that of $\bar{V}$.

Note that eigenvalue truncation should not be performed for the mean-field energy, because it effectively truncates the Coulomb interaction, leading to an incorrect treatment of the classical electrostatics of the electron distribution (overall charged without the nuclei) because of the truncation of charge−charge terms. Consequently, we compute the mean-field energy with and without CD+ET, and add the

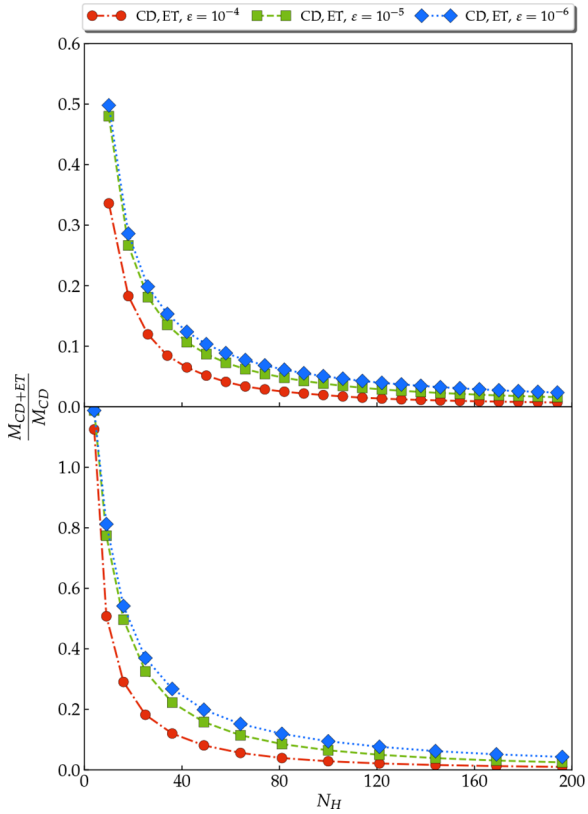


Figure 7. Ratio between the memory required for local energy precomputing for AFQMC with CD (M_{CD}) and AFQMC with CD+ET (M_{CD+ET}), as a function of the number N_H of hydrogen atoms for H chains (top) and square grids (bottom) at the representative bond length R = 1.8 au at STO-6G level of theory. Three different truncation thresholds, $\varepsilon = 10^{-4}$, 10^{-5}, 10^{-6} a.u. are explored (red circles, green squares, blue diamonds).

corresponding correction to the local energy computed from eq 23

$$E_{AFQMC} = (E^{CD+ET}_{AFQMC} - E^{CD+ET}_{RHF}) + E_{RHF} \tag{24}$$

Since we achieve a reduction in cost because the truncated low-rank factorization implements a form of integral screening, we briefly compare the low-rank factorization to directly screening the local energy evaluation in eq 11. A direct screening of the sum in eq 11 will give $O(N^2)$ cost for moderate sized and systems, and asymptotically $O(N)$ cost (although formation of the Green's function and other operations still require $O(N^3)$ cost). However, unless one evaluates the ERI on the fly, this requires $O(N^4)$ storage, and, for larger basis sets, there will be a very large number of significant ERIs, as one does not achieve the basis compression afforded by CD or DF. Thus, except for large systems in small basis sets, we expect the CD+ET approach to be superior to simple integral screening.

A hybrid strategy would be to use the sparsity of the Cholesky vectors directly (e.g., implement the construction of all intermediates in the local energy using sparse matrix multiplication). Compared with low-rank factorization, sparse matrix multiplication often incurs overhead for smaller problem sizes. However, a detailed comparison between the direct use of sparsity in the Cholesky vectors versus the nested matrix factorization is an interesting question to explore in the future.

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

4. RESULTS

We now apply the formalism outlined in Section 3 to several test systems, including both molecules and crystalline solids. In each case, we compare the local energy evaluation time T_{Eloc} from conventional AFQMC and AFQMC with CD+ET, assess the accuracy of the ET procedure, and investigate the scaling with system size. Timing calculations were performed on a cluster with nodes having 2 CPUs with 14 cores each (Intel E5−2680, 2.4 GHz).

4.1. Networks of H Atoms. We first consider the test case of hydrogen (H) chains,[36,45] at a representative bond length R = 1.8 a_B, using the minimal STO-6G basis and RHF trial wave function. We use identical thresholds for CD and eigenvalue truncation, $\epsilon_{CD} = \epsilon_{ET} = 10^{-4}, 10^{-5}, 10^{-6}$ a.u. The local energy evaluation time (per walker, averaged over many walkers) using the conventional AFQMC formula with Cholesky decomposition, eq 15, and CD+ET-based AFQMC (CD+ET-AFQMC) of Section 3, is shown in Figure 8. The reported times reflect the cumulative impact of floating-point and fundamental memory operations (e.g., allocations of arrays). The overhead due to memory operations, which we estimate of the order of ∼1 ms, becomes increasingly less important as the size of the studied system increases. Local energy calculations times are reproduced well by the formulas

$$T_{CD} \simeq t_0 N^{\alpha} \quad , \quad T_{CD+ET} \simeq t_0 N^{\beta} \tag{25}$$

We observe exponents α = 3.91(2), 3.99(1), 3.99(1) and β = 3.46(3), 3.54(2), 3.74(3) for $\epsilon = 10^{-4}, 10^{-5}, 10^{-6}$ a.u. respectively. The untruncated local energy calculation displays the anticipated quartic scaling, while the looser CD+ET truncation thresholds reach subquartic scaling for these system sizes. Given the relationship between ET and integral screening, it is unsurprising that saturation of $\langle \rho_\gamma \rangle$ (responsible for cubic scaling) is not reached for the tightest threshold. Nonetheless, the local energy evaluation time is still reduced relative to using only CD. The prefactors in the two functions determine the number N_H^* of H atoms required for the two curves to cross. We find that $N_H^* \simeq$ 25, 35,40 for the three thresholds we have considered.

In the insets, we compute the difference ΔE_c between the correlation energies per atom from AFQMC and CD+ET, as a function of the number of H atoms, using the estimator

$$\Delta E_c = \frac{1}{N_w} \sum_w [E_{loc,c}(\Phi_w) - E'_{loc,c}(\Phi_w)] \tag{26}$$

where $E_{loc,c}(\Phi) = E_{loc}(\Phi) - E_{HF}$ is defined in terms of the standard local energy functional (eq 15), but using integrals reconstructed from the CD vectors, while $E'_{loc,c}(\Phi) = E'_{loc}(\Phi) - E'_{HF}$ is formulated in terms of the CD+ET expression, Section 3, for the local energy. In Figure 8, ΔE_c is evaluated on six independently generated populations of walkers equilibrated for $\beta = 2\ E_{Ha}^{-1}$. Using all thresholds, the energies per atom agree to within 0.02% of the total correlation energy extrapolated to the thermodynamic limit (TDL), confirming the good accuracy of the CD+ET decomposition for conservative choices of the threshold ϵ.

As a simple application, we next study the asymmetric dissociation of the infinite H chain using the STO-6G basis in Figure 9. Note that, as seen in Figure 8, the cost of CD+ET calculations is reduced by a factor of roughly 10 for N_H = 100, so the full study of the asymmetric dissociation can be carried out at

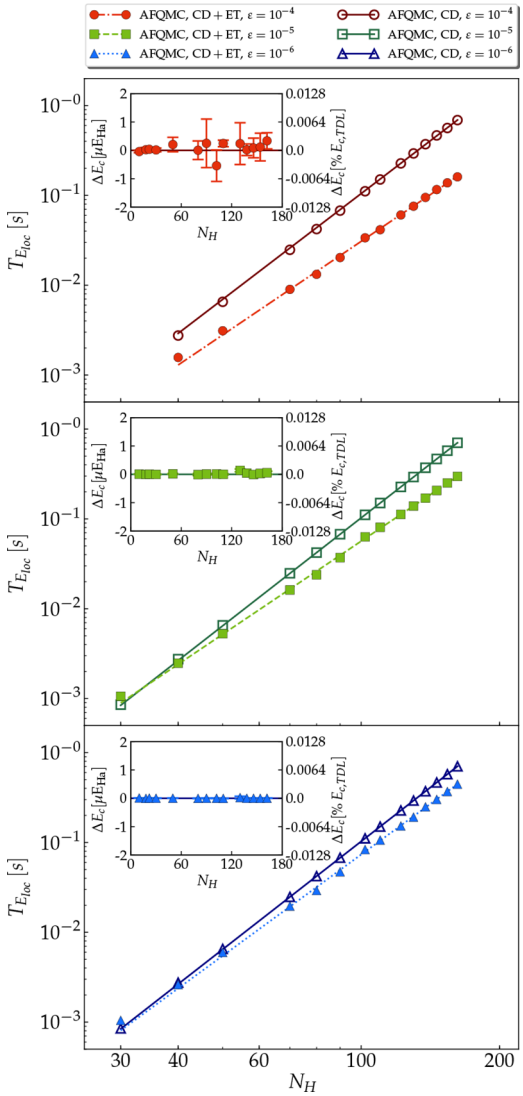


Figure 8. Main figures: Log−log plot of the local energy evaluation time $T_{E_{loc}}$ as a function of the number N_H of hydrogen atoms for H chains at the representative bond length R = 1.8 au at the STO-6G level of theory, from AFQMC with CD (empty markers) and AFQMC with CD+ET (filled markers). Truncation thresholds, $\epsilon = 10^{-4}, 10^{-5}, 10^{-6}$ a.u. (top to bottom) are explored. Solid, dashed lines are the result of fit of AFQMC with CD, CD+ET to eq 25. Insets: average difference in the correlation part of the local energy, per atom, between AFQMC with CD and CD +ET.

appreciably reduced computational cost, comparable with the cost of studying the symmetric dissociation only.

More specifically, we compute the potential energy surface of a network of H atoms at positions $\mathbf{R}_{k,\pm} = (0,0,z_{k,\pm})$ with $z_{k,\pm} = \pm\frac{R}{2} + k(R' + R), k = 0 \ldots \frac{N}{2} - 1$, for a total number of atoms between N_H = 10 and N_H = 100, as a function of the intrabond and interbond lengths R, R'. We use the UHF Slater determinant as a trial wave function. For all R, R' in a mesh of points between 1.2 and 3.6 a_B, we extrapolate the energy per atom $E(R, R', N)$ to the TDL using standard procedures,[36] and we compute correlation energies using AFQMC with CD+ET and the truncation threshold $\epsilon = 10^{-5}$ a.u. The extrapolated potential energy surface $E(R, R') = \lim_{N\to\infty} E(R, R', N)$ is shown in Figure 9, and values for R' =1.6, 2.4, 3.2 a_B are given in Table 1.

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

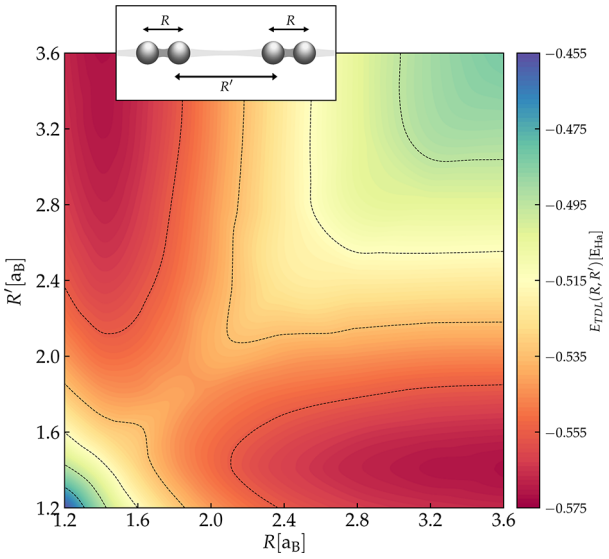


Figure 9. Energy per atom of the H chain at the STO-6G level of theory, as a function of the intrabond and interbond lengths R, R'. Results are obtained for R, R'= 1.2, 1.4, 1.6, 1.8, 2.0, 2.4, 2.8, 3.2, 3.6 au, and the potential energy surface is produced via cubic spline interpolation.

Table 1. Energy per Atom of the H Chain at the STO-6G Level of Theory, Extrapolated to the Thermodynamic Limit, as a Function of the Interbond Length R' for R = 1.6, 2.4, 3.2 a.u. (Left to Right)

R'	$E(R = 1.6, R')$	$E(R = 2.4, R')$	$E(R = 3.2, R')$
1.2	−0.51517(6)	−0.55578(9)	−0.5652(1)
1.4	−0.52857(9)	−0.5619(2)	−0.5704(1)
1.6	−0.53362(6)	−0.5582(2)	−0.5660(3)
1.8	−0.54288(8)	−0.5507(2)	−0.5569(4)
2.0	−0.5498(1)	−0.5411(1)	−0.5454(3)
2.4	−0.5582(2)	−0.5233(1)	−0.5223(1)
2.8	−0.5634(2)	−0.5219(1)	−0.5037(1)
3.2	−0.5660(2)	−0.5223(1)	−0.4915(1)
3.6	−0.5672(2)	−0.5228(1)	−0.4902(1)

The diagonal of Figure 9 corresponds to the symmetric dissociation of the chain, $R = R'$,[36] the minimum energy being reached at the saddle point $R = R' \simeq 1.83$ a_B. For large R, R' the potential energy surface increases toward the energy $E_H = 0.471$ E_{Ha} of a single H atom in the STO-6G basis, and the global minimum of the energy is reached for $R' \to \infty$, $R \simeq 1.4$ a_B, corresponding to a collection of uncoupled H_2 molecules, with energy $E_{H_2} = -0.573$ E_{Ha}. This illustrates the well-known Peierls instability of equally spaced atomic chains under lattice distorsions.

We continue our assessment of accuracy and performance by studying, in Figure 10, two-dimensional square grids of H atoms, where the H atoms occupy positions $\mathbf{R}_{ij} = (0, i\,R, j\,R)$, $i, j = 0 \cdots n-1$. Here n is related to the number N_H of atoms in the grid as $N_H = n^2$, and we work at the representative bond length $R = 1.8$ a_B.

The trends seen for H chains are confirmed: the standard and CD+ET-based local energy calculation times are well described by eq 25 with exponents α = 4.09(2), 4.14(3), 4.11(3) and β = 3.14(2), 3.25(1), 3.47(2) for $\varepsilon = 10^{-4}$, 10^{-5}, 10^{-6} a.u. respectively. (Note that the β exponents here are slightly lower than in 1D. This may seem surprising, but the model analysis shows that $\langle \rho_\gamma \rangle$ as a function of system size in 1D and 2D can actually crossover before saturation, with the 1D curve

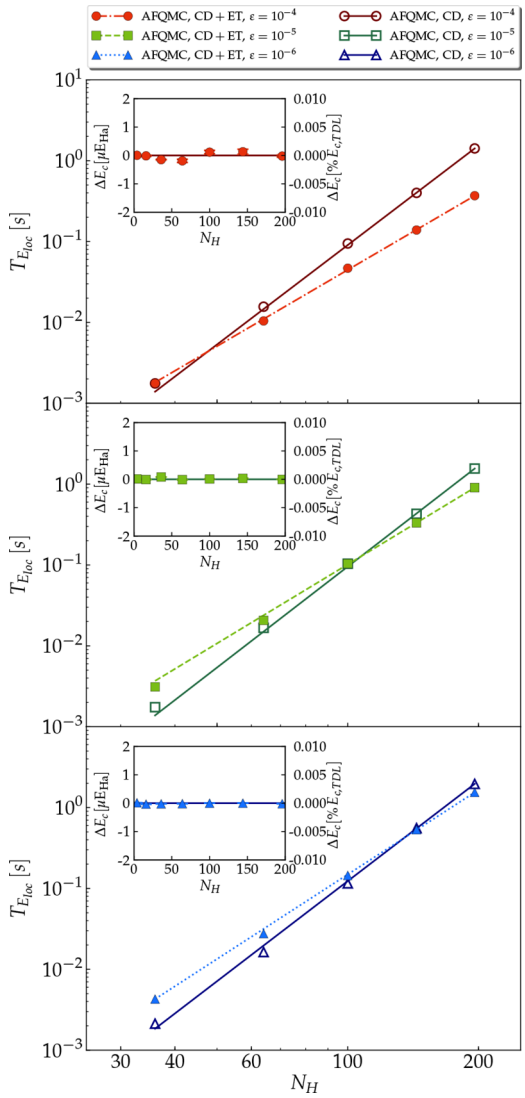


Figure 10. Main figures: local energy evaluation time as a function of the number N_H of hydrogen atoms, for H square grids at the representative bond length R = 1.8 au at the STO-6G level of theory, from AFQMC with CD (empty markers) and AFQMC with CD and ET (filled markers). Crossover between the two strategies is seen for N_H ≃50, 100, 150 for increasingly small threshold. Insets: average difference in the correlation part of the local energy, per atom, between state-of-the-art AFQMC and AFQMC with eigenvalue truncation.

growing more steeply, but saturating at smaller system size than the 2D curve.) Crossover between the two approaches is seen for $N_H^* \simeq 50, 120, 170$ for increasingly small threshold. The discrepancy ΔE_c between correlation energies based on AFQMC with CD and AFQMC with CD+ET is consistently below 0.01% of the correlation energy per atom extrapolated to the TDL, further confirming the accuracy of the truncation scheme.

4.2. Water Clusters. To test larger basis sets and heavier elements, in Figure 11, we investigate 38 water clusters (motivated by studies of water clusters in the terrestrial atmosphere) containing 2−10 water molecules,[46] using the heavy-augmented cc-pVDZ basis (aug-cc-pVDZ for O, cc-pVDZ for H), a truncation threshold $\varepsilon = 10^{-4}$ a.u. and a RHF trial wave function. Also in this case, the average number of retained eigenvalues $\langle \rho_\gamma \rangle$ grows sublinearly with the size of the system, as measured by the number of H_2O molecules, leading

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

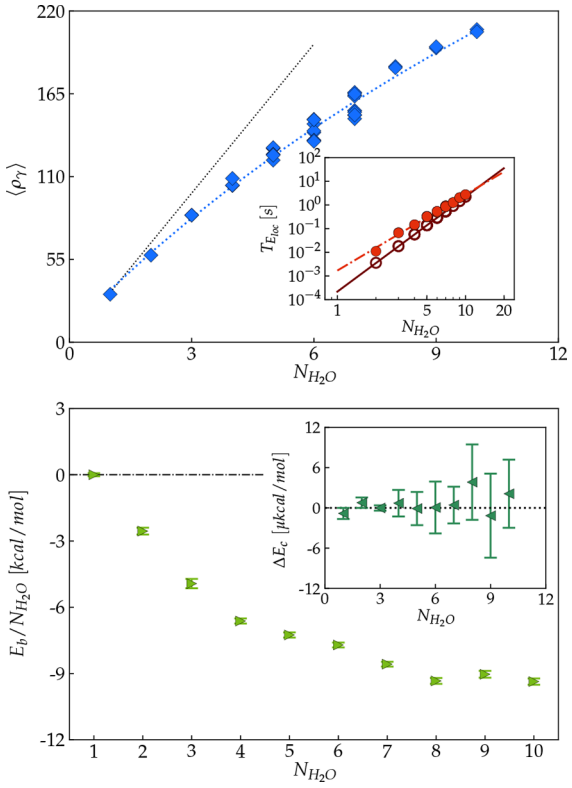


Figure 11. Top: average number $\langle \rho_\gamma \rangle$ of retained eigenvalues as a function of the number of H_2O molecules in small water clusters,[46] using $\varepsilon = 10^{-4}$ a.u. Inset: local energy calculation time from AFQMC with CD and CD+ET (empty, filled symbols). Solid, dot-dashed lines indicate fit to (25). Bottom: AFQMC binding energy per monomer, for the most stable water clusters with given number N_{H_2O} of monomers. Inset: difference ΔE_c between correlation energy per monomer from AFQMC with CD and with CD+ET, for all clusters.

to a subquartic scaling local energy evaluation (upper panel). The dependence of the local energy calculation time on the number of water molecules, shown in the inset of the upper panel, is again well represented by eq 25 with $\alpha = 4.01(1)$ and $\beta = 3.21(3)$, so that crossover between conventional AFQMC with CD and CD+ET local energy calculation times is seen at $N^*_{H_2O} \simeq 13$.

In the inset of the lower panel, we show the difference ΔE_c between the correlation energies per atom from AFQMC with CD integrals and CD+ET, as a function of the number of monomers. ΔE_c is evaluated on six independently generated populations of walkers, equilibrated for $\beta = 2\ E^{-1}_{Ha}$, and, for a given cluster size, ΔE_c is averaged over all cluster structures with the same number of monomers. For example, for $N = 5$, ΔE_c is averaged over the six water pentamers labeled CYC, CAA, CAB, CAC, FRA, FRB, FRC in ref 46. The error is only a few μ kcal/mol.

The binding energy per water molecule for the most stable clusters, labeled 2Cs, 3UUD, 4S4, 5CYC, 6PR, 7PR1, 8D2d, 9D2dDD, 10PP1 in ref 46, is shown in the lower panel of Figure 11. As seen, the binding energy per molecule decreases almost monotonically with the number of monomers in the cluster, reaching $E_b/N_{H_2O} \simeq -9$ kcal/mol for $N_{H_2O} \geq 8$.

Numerical data supplied in Table 2 provide a comparison with RHF, MP2, CCSD, and CCSD(T). Energies from these methodologies are computed without performing any truncation on the Hamiltonian, while AFQMC energies are estimated using eq 24. As seen, correlated methods are in relatively good agreement with each other. AFQMC is in good agreement with CCSD(T), with an average deviation of $\Delta = -0.59(29)$ kcal/mol. Data for the different water pentamers are shown in Figure 12. Binding energies from correlated methods are in good agreement with each other and display the same trends. The average deviation between CCSD(T) and AFQMC is $\Delta = 0.06(61)$ kcal/mol.

Table 2. Binding Energy for the Most Stable Water Clusters Reported in Ref 46, by RHF, MP2, CCSD, CCSD(T), and AFQMC(CD+ET), in kcal/mol, Using the Heavy-Augmented cc-pVDZ Basis

cluster	$E_{b,RHF}$	$E_{b,MP2}$	$E_{b,CCSD}$	$E_{b,CCSD(T)}$	$E_{b,AFQMC}$
2Cs	−3.815	−5.217	−4.912	−5.179	−5.11(31)
3UUD	−10.521	−15.833	−14.670	−15.619	−14.78(64)
4S4	−19.001	−28.358	−26.210	−27.865	−26.49(46)
5CYC	−25.297	−37.482	−34.627	−36.776	−36.27(59)
6PR	−29.917	−47.246	−43.727	−46.823	−46.25(66)
7PR1	−37.486	−59.149	−54.619	−58.470	−60.04(77)
8D2s	−46.650	−74.924	−69.023	−74.044	−74.7(1.1)
9D2dDD	−53.395	−84.816	−78.164	−83.739	−81.3(1.4)
10PP1	−60.449	−96.615	−89.029	−95.453	−93.7(1.4)

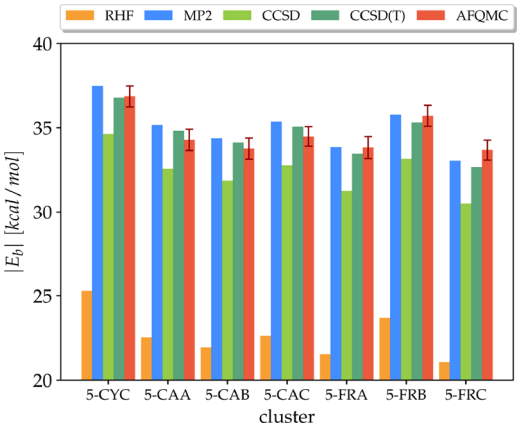


Figure 12. Binding energy for water pentamers by RHF, MP2, CCSD, CCSD(T), and AFQMC(CD+ET), in kcal/mol, using the heavy-augmented cc-pVDZ basis.

4.3. Two-Dimensional Hexagonal Boron Nitride. We now consider a crystalline solid, 2D hexagonal boron nitride (BN). To perform these calculations, we used an underlying single-particle basis of crystalline Gaussian-based atomic orbitals, which are translational-symmetry-adapted linear combinations of Gaussian atomic orbitals.[47] Core electrons were replaced with separable norm-conserving GTH-LDA pseudopotentials,[48,49] removing sharp nuclear densities. Matrix elements for the Hamiltonian of the system were computed with the PySCF[50] package, using the GTH-DZV Gaussian basis set.[51] The RHF state was used as trial wave function.

Size effects were removed studying increasingly large supercells at the Γ point. Supercells were obtained repeating the primitive, two-atom cell $N_x = N_y = 1, \ldots, 5$ times along directions a_x, a_y sketched in Figure 13, and we operated at the representative bond length $R_{BN} = 2.5$ Å to illustrate the effects of the eigenvalue truncation on top of the DF approximation. The ERI was obtained using the Gaussian DF approximation[52]

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

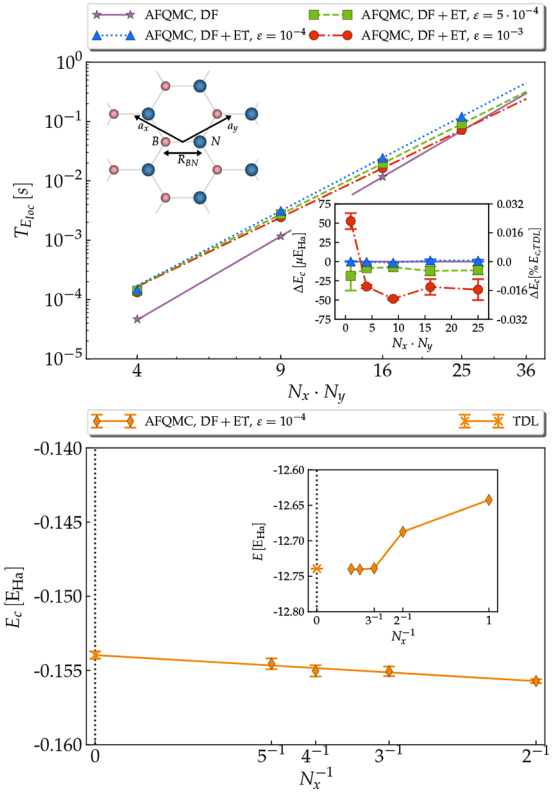


Figure 13. Top: Local energy evaluation time $T_{E_{loc}}$ from AFQMC with DF approximation for the ERI (purple stars), and AFQMC with DF +ET (red circles, green squares, blue triangles for $\varepsilon = 10^{-3}, 5\cdot10^{-4}, 10^{-4}$ a.u. respectively), for 2D hexagonal BN at $R_{BN} = 2.5$ Å, as a function of supercell size. Inset: difference in the correlation part of the local energy per unit cell. Bottom: extrapolation to the thermodynamic of AFQMC correlation (main plot) and total (inset) energy (orange diamonds) per unit cell. Extrapolations to the thermodynamic limit are indicated by orange crosses.

$$(pr|qs) = \sum_{L} (pr|L)\Omega^{-1}_{LM}(qs|M) = \sum_{\gamma} D^{\gamma}_{pr} D^{\gamma}_{qs} \tag{27}$$

with $D^{\gamma}_{pr} = \sum_L (pr|L)\ \Omega^{-1/2}_{L\gamma}$, and eigenvalue truncation was performed on the DF operators D^{γ}_{pr} with truncation thresholds $\varepsilon = 10^{-4}, 5\cdot10^{-4}, 10^{-3}$ a.u.

In the upper panel of Figure 13, we illustrate the local energy evaluation time from AFQMC with DF and DF+ET as a function of supercell size $N_x \cdot N_y$. Crossover is seen for $N_x \simeq 5,6$ for increasingly small thresholds. For a wide-bandgap semiconductor like BN, as discussed below, supercells of this size are sufficient to converge mean-field and correlation energies to the thermodynamic limit. We thus expect the DF+ET approach to be even more beneficial for materials with smaller or vanishing gap (e.g., metals), that require even larger supercells or Brillouin zone meshes to reliably converge energies to the TDL. In the lower panel of Figure 13, we extrapolate the AFQMC correlation energy to the TDL using the power-law Ansatz $E_c(N_x) = \alpha + \beta N_x^{-1}$.[47] We add to the extrapolated AFQMC correlation energy α the extrapolated RHF energy, obtained following the procedure in ref 47. The extrapolated total energy is shown in the inset of the lower panel of Figure 13.

In the inset of the upper panel, we illustrate the difference ΔE_c between the correlation energy per cell from AFQMC with DF and DF+ET, estimated on three populations of walkers equilibrated for $\beta = 4\ E_{Ha}^{-1}$. As naturally expected, ΔE_c increases monotonically with the truncation threshold, though remaining consistently below 0.03% of the AFQMC correlation energy extrapolated to the TDL.

5. CONCLUSIONS

In the present work we have shown that, through a simple and efficient low-rank factorization of the ERI, it is possible to reduce the asymptotic complexity of AFQMC calculations for electronic structure problems in Gaussian bases from its conventional quartic scaling with system size. While the asymptotic cubic scaling is attained only for large systems, at the intermediate sizes studied, we nevertheless observed subquartic scaling accompanied by significant memory savings and high accuracy. This reduction arises from exposing the sparsity of ERI in Gaussian bases through a nested matrix diagonalization. This approach will be useful in studies of larger molecules, and of crystalline solids requiring extrapolations to the thermodynamic limit. We also find that the memory requirements using this approach are significantly reduced from the conventional AFQMC algorithm using Cholesky decomposition alone. The algorithmic advances may also be used in conjunction with parallel efforts to accelerate AFQMC through improved hardware implementations.[53] While more work is necessary to establish the relative benefits in AFQMC of exposing sparsity through low rank, as in the current work, versus the direct utilization of sparse operations, we expect such combinations to greatly advance the practical possibilities for AFQMC calculations on large systems.

RELATIONSHIP WITH PLANE-WAVE FORMULATIONS

Many calculations in solid-state systems are performed using a plane wave basis. AFQMC simulations using this computational basis[25,29] have an $O(N^3)$ scaling. We here briefly outline the relationship between the cubic scaling achieved in the plane-wave basis and that achieved using the factorization techniques in this paper.

In the plane-wave basis, the Hamiltonian with pseudopotentials takes the form

$$\hat{H} = H_0 + \sum_{\mathbf{G}\mathbf{G}'} t_{\mathbf{G}\mathbf{G}'}\hat{E}_{\mathbf{G}\mathbf{G}'} + \frac{1}{2}\sum_{\mathbf{G}\mathbf{G}'\mathbf{q}} V_{\mathbf{q}}\hat{E}_{\mathbf{G}+\mathbf{q}\mathbf{G}}\hat{E}^{\dagger}_{\mathbf{G}'\mathbf{G}'+\mathbf{q}} \tag{28}$$

where $\mathbf{G}$ is a wave-vector in the reciprocal lattice, corresponding to the plane-wave state $\langle \mathbf{r}|\mathbf{G}\rangle = \frac{e^{i\mathbf{G}\cdot\mathbf{r}}}{\sqrt{\Omega}}$ where Ω is the computational cell volume. The vectors $\mathbf{q}$ are the transfer momenta, and, due to momentum conservation, their number is proportional to number of plane-waves N. Thus, eq 28 is a low-rank factorization of the integrals (with $\mathbf{q}$ playing the role of γ) but it is not a Cholesky factorization, because the analogous quantity

$$L^{\mathbf{q}}_{\mathbf{G}\mathbf{G}'} = \sqrt{V_{\mathbf{q}}}\,\delta_{\mathbf{G},\mathbf{G}'+\mathbf{q}} \tag{29}$$

is not a lower triangular matrix for each $\mathbf{q}$.

The local energy formula can be written analogously to eq 16

$$\mathcal{E}_{\mathrm{loc},2}(\Phi) = \sum_{\mathbf{ij}\mathbf{q}} f^{\mathbf{q}}_{ii} f^{-\mathbf{q}}_{jj} - f^{\mathbf{q}}_{ij} f^{-\mathbf{q}}_{ji} \tag{30}$$

with $f^{\mathbf{q}}_{ij}$ defined formally as

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

$$f_{ij}^{\mathbf{q}} = \sum_{\mathbf{G}\mathbf{G}'} L_{\mathbf{G}\mathbf{G}'}^{\mathbf{q}} \Phi_{T_i\mathbf{G}} \Theta_{\mathbf{G}'j} \tag{31}$$

Unlike in the case of the atomic orbital basis, the $L^{\mathbf{q}}$ matrices contain N, rather than $O(\log N)$, elements, and do not display the same low-rank structure. However, $L^{\mathbf{q}}$ encodes a periodic delta function, which means that eq 29 is a convolution

$$f_{ij}^{\mathbf{q}} = \sum_{\mathbf{G}} \Phi_{i\mathbf{G}}^{\sigma} \Theta_{\mathbf{G}+\mathbf{q}j}^{\sigma} \tag{32}$$

Consequently, using the fast Fourier transform, computing $f_{ij}^{\mathbf{q}}$ requires only $O(O^2 N \log N) \sim \tilde{O}(N^3)$ time.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: mmotta@caltech.edu.
*E-mail: gkc1000@gmail.com.

ORCID

Mario Motta: 0000-0003-1647-9864
James Shee: 0000-0001-8333-8151

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.M. and G.K.C. were supported by the U.S. NSF (Grant No. 1665333). S.Z. acknowledges support from DOE (Grant No. DE-SC0001303). Additional software developments for the AFQMC and PySCF periodic calculations were supported by U.S. NSF (Grant No. 1657286). Computations were carried out on facilities supported by the National Energy Research Scientific Computing Center (NERSC), on facilities supported by the Scientific Computing Core at the Flatiron Institute, a division of the Simons Foundation, on the Pauling cluster at the California Institute of Technology, and on the Storm and SciClone Clusters at the College of William and Mary. M.M. acknowledges Narbe Mardirossian, Yuliya Gordiyenko and Qiming Sun for useful discussion about electronic structure calculations for H_2O clusters and BN.

REFERENCES

(1) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496−4501.
(2) Beebe, N. H. F.; Linderberg, J. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683−705.
(3) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. Tensor factorizations of local second-order Møller-Plesset theory. *J. Chem. Phys.* **2011**, *134*, 044123.
(4) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 044103.
(5) Malone, F. D.; Zhang, S.; Morales, M. A. Overcoming the memory bottleneck in auxiliary field quantum Monte Carlo simulations with interpolative separable density fitting. *J. Chem. Theory Comput.* **2019**, *15*, 256−264.
(6) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On the applicability of LCAO-Xa methods to molecules containing transition metal atoms: The nickel atom and nickel hydride. *Int. J. Quantum Chem.* **1977**, *12*, 81−87.
(7) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993−4999.
(8) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149−8160.
(9) Koch, H.; Sanchez de Meras, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481−9484.
(10) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-A.; Neogrády, P.; Pedersen, T. B.; Pitonák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224−247.
(11) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285−4291.
(12) Aquilante, F.; Pedersen, T. B.; Lindh, R. Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J. Chem. Phys.* **2007**, *126*, 194106.
(13) Sierka, M.; Hogekamp, A.; Ahlrichs, R. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118*, 9136−9148.
(14) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
(15) Parrish, R. M.; Hohenstein, E. G.; Schunck, N. F.; Sherrill, C. D.; Martínez, T. J. Exact tensor hypercontraction: a universal technique for the resolution of matrix elements of local finite-range N-body potentials in many-body quantum problems. *Phys. Rev. Lett.* **2013**, *111*, 132505.
(16) Hohenstein, E. G.; Kokkila, S. I. L.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction equation-of-motion second-order approximate coupled cluster: electronic excitation energies in $O(N^4)$ time. *J. Phys. Chem. B* **2013**, *117*, 12972−12978.
(17) Parrish, R. M.; Sherrill, C. D.; Hohenstein, E. G.; Kokkila, S. I. L.; Martínez, T. J. Communication: Acceleration of coupled cluster singles and doubles via orbital-weighted least-squares tensor hypercontraction. *J. Chem. Phys.* **2014**, *140*, 181102.
(18) Kokkila Schumacher, S. I. L.; Hohenstein, E. G.; Parrish, R. M.; Wang, L.-P.; Martínez, T. J. Tensor hypercontraction second-order Møller-Plesset perturbation theory: grid optimization and reaction energies. *J. Chem. Theory Comput.* **2015**, *11*, 3042−3052.
(19) Benedikt, U.; Auer, A. A.; Espig, M.; Hackbusch, W. Tensor decomposition in post-Hartree-Fock methods. I. Two-electron integrals and MP2. *J. Chem. Phys.* **2011**, *134*, 054118.
(20) Shenvi, N.; van Aggelen, H.; Yang, Y.; Yang, W.; Schwerdtfeger, C.; Mazziotti, D. The tensor hypercontracted parametric reduced density matrix algorithm: coupled-cluster accuracy with $O(r^4)$ scaling. *J. Chem. Phys.* **2013**, *139*, 054110.
(21) Hummel, F.; Tsatsoulis, T.; Grüneis, A. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **2017**, *146*, 124105.
(22) Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*, 184113.
(23) Peng, B.; Kowalski, K. Highly efficient and scalable compound decomposition of two-electron integral tensor and its application in coupled cluster calculations. *J. Chem. Theory Comput.* **2017**, *13*, 4179−4192.
(24) Motta, M.; Ye, E.; McClean, J. R.; Li, Z.; Minnich, A.; Babbush, R.; Chan, G. K.-L. Low-rank representations for quantum simulation of electronic structure. 2018, *arXiv:1808.02625*. arXiv.org e-Print archive. https://arxiv.org/abs/1808.02625.
(25) Zhang, S.; Krakauer, H. Quantum Monte Carlo Method using phase-free random walks with Slater determinants. *Phys. Rev. Lett.* **2003**, *90*, 136401.
(26) Al-Saidi, W. A.; Zhang, S.; Krakauer, H. Auxiliary-field quantum Monte Carlo calculations of molecular systems with a Gaussian basis. *J. Chem. Phys.* **2006**, *124*, 224101.

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521

(27) Motta, M.; Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *WIREs: Comput. Mol. Sci.* **2018**, *8*, e1364.
(28) Purwanto, W.; Krakauer, H.; Virgus, Y.; Zhang, S. Assessing weak hydrogen binding on Ca^+ centers: An accurate many-body study with large basis sets. *J. Chem. Phys.* **2011**, *135*, 164105.
(29) Suewattana, M.; Purwanto, W.; Zhang, S.; Krakauer, H.; Walter, E. J. Phaseless auxiliary-field quantum Monte Carlo calculations with plane waves and pseudopotentials: Applications to atoms and molecules. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75*, 245123.
(30) Booth, G. H.; Tsatsoulis, T.; Chan, G. K.-L.; Grüneis, A. From plane waves to local Gaussians for the simulation of correlated periodic systems. *J. Chem. Phys.* **2016**, *145*, 084111.
(31) Stratonovich, R. L. On a method of calculating quantum distribution functions. *Sov. Phys. Doklady* **1958**, *2*, 416.
(32) Hubbard, J. Calculation of partition functions. *Phys. Rev. Lett.* **1959**, *3*, 77−78.
(33) Al-Saidi, W. A.; Zhang, S.; Krakauer, H. Bond breaking with auxiliary-field quantum Monte Carlo. *J. Chem. Phys.* **2007**, *127*, 144101.
(34) Purwanto, W.; Krakauer, H.; Zhang, S. Pressure-induced diamond to β-tin transition in bulk silicon: A quantum Monte Carlo study. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 214116.
(35) Virgus, Y.; Purwanto, W.; Krakauer, H.; Zhang, S. Stability, energetics, and magnetic states of cobalt adatoms on graphene. *Phys. Rev. Lett.* **2014**, *113*, 175502.
(36) Motta, M.; et al. Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods. *Phys. Rev. X* **2017**, *7*, 031059.
(37) LeBlanc, J. P. F.; et al. Solutions of the two-dimensional Hubbard model: benchmarks and results from a wide range of numerical algorithms. *Phys. Rev. X* **2015**, *5*, 041041.
(38) Qin, M.; Shi, H.; Zhang, S. Coupling quantum Monte Carlo and independent-particle calculations: self-consistent constraint for the sign problem based on the density or the density matrix. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 235119.
(39) Motta, M.; Zhang, S. Computation of ground-state properties in molecular systems: back-propagation with auxiliary-field quantum Monte Carlo. *J. Chem. Theory Comput.* **2017**, *13*, 5367−5378.
(40) Motta, M.; Zhang, S. Communication: Calculation of interatomic forces and optimization of molecular geometry with auxiliary-field quantum Monte Carlo. *J. Chem. Phys.* **2018**, *148*, 181101.
(41) Shee, J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. Chemical Transformations Approaching Chemical Accuracy via Correlated Sampling in Auxiliary-Field Quantum Monte Carlo. *J. Chem. Theory Comput.* **2017**, *13*, 2667−2680.
(42) Ma, F.; Zhang, S.; Krakauer, H. Auxiliary-field quantum Monte Carlo calculations with multiple-projector pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 165103.
(43) Balian, R.; Brezin, E. Nonunitary Bogoliubov transformations and extension of Wick's theorem. *Nuovo Cimento B* **1969**, *64*, 37.
(44) Lin, C.-J.; Moré, J. J. Incomplete Cholesky factorizations with limited memory. *SIAM J. Sci. Comput.* **1999**, *21*, 24−45.
(45) Hachmann, J.; Cardoen, W.; Chan, G. K.-L. Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group. *J. Chem. Phys.* **2006**, *125*, 144101.
(46) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034−12046.
(47) McClain, J.; Sun, Q.; Chan, G. K.-L.; Berkelbach, T. C. Gaussian-based coupled-cluster theory for the ground-state and band structure of solids. *J. Chem. Theory Comput.* **2017**, *13*, 1209−1218.
(48) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(49) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 3641−3662.
(50) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *WIREs: Comput. Mol. Sci.* **2018**, *8*, e1340.
(51) Lundqvist, B. Single-particle spectrum of the degenerate electron gas. *Eur. Phys. J. B* **1967**, *6*, 193−205.
(52) Sun, Q.; Berkelbach, T. C.; McClain, J. D.; Chan, G. K.-L. Gaussian and plane-wave mixed density fitting for periodic systems. *J. Chem. Phys.* **2017**, *147*, 164119.
(53) Shee, J.; Arthur, E. J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. Phaseless auxiliary-field quantum Monte Carlo on graphical processing units. *J. Chem. Theory Comput.* **2018**, *14*, 4109−4121.

DOI: 10.1021/acs.jctc.8b00996
J. Chem. Theory Comput. 2019, 15, 3510−3521