

Spectral Functions from Auxiliary-Field Quantum Monte Carlo without Analytic Continuation: The Extended Koopmans' Theorem Approach

Joonho Lee,* Fionn D. Malone,* Miguel A. Morales,* and David R. Reichman*



Cite This: *J. Chem. Theory Comput.* 2021, 17, 3372–3387



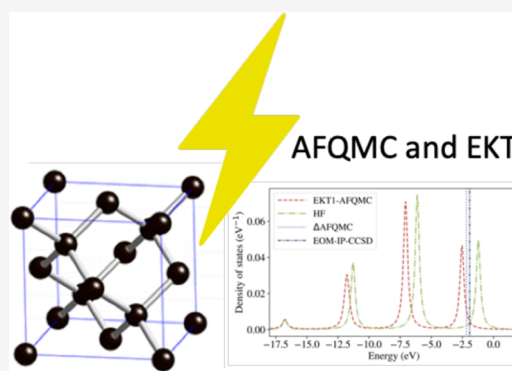
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: We explore the extended Koopmans' theorem (EKT) within the phaseless auxiliary-field quantum Monte Carlo (AFQMC) method. The EKT allows for the direct calculation of electron addition and removal spectral functions using reduced density matrices of the N -particle system and avoids the need for analytic continuation. The lowest level of EKT with AFQMC, called EKT1-AFQMC, is benchmarked using atoms, small molecules, 14-electron and 54-electron uniform electron gas supercells, and a minimal unit cell model of diamond at the Γ -point. Via comparison with numerically exact results (when possible) and coupled-cluster methods, we find that EKT1-AFQMC can reproduce the qualitative features of spectral functions for Koopmans-like charge excitations with errors in peak locations of less than 0.25 eV in a finite basis. We also note the numerical difficulties that arise in the EKT1-AFQMC eigenvalue problem, especially when back-propagated quantities are very noisy. We show how a systematic higher-order EKT approach can correct errors in EKT1-based theories with respect to the satellite region of the spectral function. Our work will be of use for the study of low-energy charge excitations and spectral functions in correlated molecules and solids where AFQMC can be reliably performed for both energy and back propagation.



1. INTRODUCTION

The dynamical response to external perturbation is one of the most powerful means of experimentally probing molecules and materials. Examples include angle-resolved photoemission spectroscopy,¹ electron energy loss spectroscopy,² and inelastic neutron scattering,³ each of which encodes the excitation spectrum of a many-body system. The theoretical description of such experiments can be modeled in the linear response regime by considering many-body Green's functions.^{4,5} For example, differential cross sections from direct and inverse photoemission experiments can be related to the retarded single-particle Green's function.⁶ In a general sense, these observables are connected to spectral functions describing electron removal and addition via the single-particle Green's function.^{6,7}

Given the above facts, the theoretical description of dynamical response properties have been dominated by Green's function-based approaches mainly due to the direct access to the spectral function that they afford.^{5,8} There has also been a sizable effort to construct spectral functions based on wave function methods. From the quantum chemistry community, there have been developments of coupled-cluster Green's function^{9–15} or equation-of-motion coupled-cluster methods.^{16,17} These and related approaches have distinct strengths and weaknesses in

terms of both cost and accuracy and continue to be actively pursued.

Another useful path to the description of spectral information is based on projector quantum Monte Carlo (PQMC) approaches.^{18,19} PQMC methods provide a highly accurate means to simulate the ground-state properties of correlated solids.²⁰ Unlike the aforementioned wave function-based approaches, PQMC methods do not provide direct access to real-time and real-frequency Green's functions. This is a direct consequence of the imaginary-time propagation at the heart of all PQMC approaches. A popular way around this hurdle is to first obtain the imaginary-time Green's function and then perform analytic continuation to obtain the real-frequency Green's function.^{21–27} Unfortunately, analytic continuation is numerically ill-conditioned, and the methods to perform analytic continuation such as the maximum entropy method^{21,22} can exhibit difficulties in resolving sharp features in the real-

Received: January 28, 2021

Published: May 13, 2021



frequency spectral function even if high-quality imaginary-time Green's functions are used as input.^{28–30} Therefore, it is highly desirable to develop an alternative means to obtain spectral functions which can work with PQMC methods without sacrificing its ground-state accuracy.

The approach that we will examine in this work is called the extended Koopmans' theorem (EKT), which was first proposed in refs 31–33. The EKT generalizes the KT in Hartree–Fock (HF) theory for arbitrary many-body wave functions. Its working ingredients are reduced density matrices (RDMs) for an N -particle system, and it produces approximate $(N - 1)$ -particle and $(N + 1)$ -particle wave functions even without the N -particle ground-state wave function. Due to this desirable feature, EKT methods have been widely used as a means to obtain spectral information for approaches for which one has access neither to real-time Green's functions nor wave functions. Examples include direct RDM-based methods,³⁴ density matrix functional theory,³⁵ natural orbital functional methods,^{36,37} and second-order Green's function methods.³⁸ EKT has also been explored with wave function methods such as configuration interaction methods,³⁹ Møller–Plesset perturbation methods,^{40,41} orbital-optimized methods,^{42,43} and coupled-cluster methods.⁴⁴ It is also a promising way for any QMC method to compute excited state and spectral information if the necessary RDMs can be constructed. The EKT has also been used to obtain the quasiparticle band structure of silicon,⁴⁵ ionization potentials (IPs), and electron affinities of atoms,⁴⁶ the Fermi velocity of graphene,⁴⁶ and quasiparticle bands of *trans*-polyacetylene⁴⁷ using VMC. Lastly, the EKT has been combined with DMC to study similar systems.⁴⁶

A PQMC approach that can be readily combined with the EKT is the phaseless auxiliary-field QMC (ph-AFQMC) method.^{48–50} ph-AFQMC has emerged as a flexible, accurate, and scalable many-body method. It imposes an approximate gauge boundary condition (i.e., the phaseless constraint) on the imaginary-time evolution of Slater determinant walkers, completely removing the Fermionic phase problem.⁵⁰ While the resulting energy is neither exact nor a variational upper bound to the exact ground-state energy,⁵¹ many benchmark studies have demonstrated the accuracy of ph-AFQMC and its related variants.^{52–63} Furthermore, with recent advances in local energy evaluation techniques in ph-AFQMC,^{64,65} the cost for obtaining each statistical sample scales cubically with the system size, which renders it less expensive than many other many-body methods. With the advent of the back-propagation (BP) method in ph-AFQMC,^{49,66,67} with some additional effort one can compute pure estimators for any operator, including those that do not commute with the Hamiltonian. Therefore, one can compute the relevant input for the EKT directly from ph-AFQMC using the BP algorithm. This is the direction we pursue in this work.

This paper is organized as follows. We first present the general framework of the EKT, its most common form EKT-1, and its extension, EKT-3. We then discuss how to obtain the relevant input for EKT-1 using BP and ph-AFQMC. We assess the accuracy of EKT-1-AFQMC on atoms (He, Be, and Ne) and small molecules (FH, N₂, CH₄, and H₂O). We further show the qualitative failure of EKT-1 for the core spectra of a 14-electron 19-plane-wave model of uniform electron gas (UEG) and illustrate the drastic improvement upon this result from using EKT-3 on the same model. We also apply EKT1-AFQMC to a 54-electron 257-plane-wave model of UEG and a minimal unit

cell model of diamond at the Γ -point. We conclude and summarize our most important findings.

2. THEORY

2.1. Extended KT. In order to compute quasiparticle gaps and spectral functions, one must compute IP and electron attachment energies along with the associated wave functions (or at least squared amplitudes for spectral weights). While we focus in this work on electron removal processes, we keep our presentation of theory general so that it is also applicable to electron addition processes.

In the EKT approach, we consider wave functions

$$|\Psi_\nu^{N\pm 1}\rangle = \hat{O}_\nu^\pm |\Psi_0^N\rangle \quad (1)$$

where the electron addition operator $\hat{O}_\nu^+$ is

$$\hat{O}_\nu^+ = \sum_p (c_+)_p^\nu \hat{a}_p^\dagger \quad (2)$$

for the ν -th one-particle excitations, the electron removal operator $\hat{O}_\nu^-$ is

$$\hat{O}_\nu^- = \sum_p (c_-)_p^\nu \hat{a}_p \quad (3)$$

for the ν -th one-hole excitations, and $|\Psi_0^N\rangle$ is the N -particle ground state wavefunction. We obtain the linear coefficients $c_\pm^\nu$ by minimizing the following variational energy expression

$$\Delta E_\nu^\pm = E_\nu^{(N\pm 1)} - E_\nu^{(N)} = \frac{\langle \Psi_0^N | (\hat{O}_\nu^\pm)^\dagger [\hat{H}, \hat{O}_\nu^\pm] | \Psi_0^N \rangle}{\langle \Psi_0^N | (\hat{O}_\nu^\pm)^\dagger \hat{O}_\nu^\pm | \Psi_0^N \rangle} \quad (4)$$

where we have defined

$$E_\nu^{(N\pm 1)} = \frac{\langle \Psi_0^N | (\hat{O}_\nu^\pm)^\dagger \hat{H} \hat{O}_\nu^\pm | \Psi_0^N \rangle}{\langle \Psi_0^N | (\hat{O}_\nu^\pm)^\dagger \hat{O}_\nu^\pm | \Psi_0^N \rangle} \quad (5)$$

and assumed

$$\hat{H} |\Psi_0^N\rangle = E_0^{(N)} |\Psi_0^N\rangle \quad (6)$$

where $E_0^{(N)}$ is the N -particle ground state energy. We refer this approach to as EKT-1. The excitation levels in eqs 3 and 2 can be systematically increased to achieve a greater accuracy in principle at the expense of greater computational costs.^{34,68} The next level of theory would incorporate 2h1p and 2p1h excitations instead of eqs 3 and 2, respectively

$$\hat{O}_\nu^+ = \sum_{pqr} (c_+)^{\nu}_{pqr} \hat{a}_r \hat{a}_q^\dagger \hat{a}_p^\dagger \quad (7)$$

and

$$\hat{O}_\nu^- = \sum_{pqr} (c_-)^{\nu}_{pqr} \hat{a}_r^\dagger \hat{a}_q \hat{a}_p \quad (8)$$

These operators include EKT-1 excitations because when $r = q$, we recover the 1h and 1p excitations, as in eqs 3 and 2. We refer this higher level of theory to as EKT-3.

2.1.1. EKT-1. We consider the following Lagrangian for 1h and 1p excitations

$$\mathcal{L}[c^\nu] = \langle \Psi_0^N | (\hat{O}_\nu^\pm)^\dagger [\hat{H}, \hat{O}_\nu^\pm] | \Psi_0^N \rangle - \epsilon_\pm^\nu ((c^\nu)^\dagger \mathbf{S}_\pm c^\nu - 1) \quad (9)$$

where $\mathbf{S}_{\pm}$ is a pertinent metric matrix for normalization and $\epsilon_{\pm}^{\nu}$ is a Lagrange multiplier. We note that

$$\epsilon_{\pm}^{\nu} = \pm \Delta E_{\nu}^{\pm} \quad (10)$$

The normalization of $|\Psi_{\nu}^{N\pm 1}\rangle$ is ensured by the constraint in eq 9. The stationary condition of eq 9 with respect to $(\epsilon_{\pm}^{\nu})^{\dagger}$ then leads to a generalized eigenvalue equation

$$\mathbf{F}_{\pm} \mathbf{c}_{\pm}^{\nu} = \epsilon_{\pm}^{\nu} \mathbf{S}_{\pm} \mathbf{c}_{\pm}^{\nu} \quad (11)$$

where the generalized Fock matrix is defined as (assuming that $|\Psi_0^N\rangle$ is normalized)

$$(\mathbf{F}_{-})_{pq} = \langle \Psi_0^N | \hat{a}_p^{\dagger} [\hat{\mathcal{H}}, \hat{a}_q] | \Psi_0^N \rangle \quad (12)$$

and

$$(\mathbf{F}_{+})_{pq} = \langle \Psi_0^N | \hat{a}_p [\hat{\mathcal{H}}, \hat{a}_q^{\dagger}] | \Psi_0^N \rangle \quad (13)$$

and the corresponding metric matrix $\mathbf{S}_{\pm}$ is

$$(\mathbf{S}_{-})_{pq} = \Gamma_{pq} \quad (14)$$

and

$$(\mathbf{S}_{+})_{pq} = \delta_{pq} - \Gamma_{qp} \quad (15)$$

Here, Γ_{pq} is the one-body RDM (1-RDM)

$$\Gamma_{pq} = \langle \Psi_0^N | \hat{a}_p^{\dagger} \hat{a}_q | \Psi_0^N \rangle \quad (16)$$

The electron attachment and IP simply follow $\epsilon_{+} = -\text{EA}$ and $\epsilon_{-} = -\text{IP}$ (assuming that ν corresponds to the lowest energy state for each particle sector). Then, the quasiparticle gap is given as $\Delta E_{qp} = \epsilon_{+} - \epsilon_{-}$. We note that these Fock matrices are not Hermitian unless $|\Psi_0^N\rangle$ is an exact eigenstate of $\hat{\mathcal{H}}$.

To provide more detailed expressions, let us define a generic *ab initio* Hamiltonian

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_1 + \hat{\mathcal{H}}_2 \quad (17)$$

with

$$\hat{\mathcal{H}}_1 = \sum_{pq} h_{pq} \hat{a}_p^{\dagger} \hat{a}_q \quad (18)$$

$$\hat{\mathcal{H}}_2 = \frac{1}{2} \sum_{pqrs} \langle pq|rs \rangle \hat{a}_p^{\dagger} \hat{a}_q^{\dagger} \hat{a}_s \hat{a}_r \quad (19)$$

where h_{pq} is the one-body Hamiltonian matrix element and $\langle pq|rs \rangle$ is the two-electron integral tensor in the Dirac notation. Substituting eq 17 into eqs 12 and 13, it can be shown that $\mathbf{F}_{\pm}$ can be evaluated with the 1-RDM and the two-body RDM (2-RDM)

$$(\mathbf{F}_{-})_{pq} = - \sum_q h_{qr} (\mathbf{S}_{-})_{pr} + \frac{1}{2} \sum_{trs} \langle tq|rs \rangle \Gamma_{pt}^{rs} \quad (20)$$

and

$$(\mathbf{F}_{+})_{pq} = \sum_q h_{qr} (\mathbf{S}_{+})_{pr} + \frac{1}{2} \sum_{trs} \langle rt|qs \rangle \Gamma_{rt}^{sp} + \sum_{rs} (\mathbf{S}_{-})_{rs} \langle pr||qs \rangle \quad (21)$$

where the 2-RDM Γ is

$$\Gamma_{pt}^{rs} = \langle \Psi_0^N | \hat{a}_p^{\dagger} \hat{a}_t^{\dagger} \hat{a}_s \hat{a}_r | \Psi_0^N \rangle \quad (22)$$

and the antisymmetrized two-electron integral tensor is defined as

$$\langle pq||rs \rangle = \langle pq|rs \rangle - \langle pq|sr \rangle \quad (23)$$

2.1.2. EKT-3. For many solid-state systems, including 1h or 1p excitations only is not sufficient as the restriction to such excitations would not be capable of describing satellite peaks.⁶⁹ A straightforward way to obtain satellite peaks in addition to dominant quasiparticle peaks is to include higher-order excitations in the EKT ansatz. Thus, EKT-3 is the next level in this hierarchy that can be attempted. Although EKT-3 has been mentioned in the literature^{34,68,70} and approximately implemented (neglecting the opposite spin term) before,³⁴ to the best of our knowledge, this work presents the first complete implementation of EKT-3 along with numerical results.

The corresponding generalized Fock operator for the IP problem reads

$$(\mathbf{F}_{-})_{pqr,stu} = \langle \Psi_0^N | \hat{a}_p^{\dagger} \hat{a}_q^{\dagger} \hat{a}_r [\hat{\mathcal{H}}, \hat{a}_u^{\dagger} \hat{a}_t \hat{a}_s] | \Psi_0^N \rangle \quad (24)$$

Using the SecondQuantizationAlgebra package⁷¹ developed by Neuscamman and others,^{72,73} we derived a complete spin-orbital equation of the generalized Fock operator

$$\begin{aligned} (\mathbf{F}_{-})_{ijk,lmn} = & -h_{kn} \Gamma_{ij}^{ml} - \sum_a (h_{la} \Gamma_{ij}^{am} \delta_{kn} + h_{ma} \delta_{kn} \Gamma_{ij}^{al} \\ & + h_{la} \Gamma_{ijn}^{amk} - h_{ma} \Gamma_{ijn}^{alk} + h_{na} \Gamma_{ija}^{mlk}) \\ & + \frac{1}{2} \sum_{ab} (\langle lm||ab \rangle \delta_{kn} \Gamma_{ij}^{ba} - \langle kl||ab \rangle \Gamma_{ijn}^{bam} + \langle km||ab \rangle \Gamma_{ijn}^{bal} \\ & - 2 \langle ka||nb \rangle \Gamma_{ija}^{bml} - \langle lm||ab \rangle \Gamma_{ijn}^{bak}) \\ & + \frac{1}{2} \sum_{abc} (\langle ma||bc \rangle \delta_{kn} \Gamma_{ija}^{cbl} - \langle la||bc \rangle \delta_{kn} \Gamma_{ija}^{cbm} \\ & - \langle la||bc \rangle \Gamma_{ijna}^{cbmk} + \langle ma||bc \rangle \Gamma_{ijna}^{cbkl} \\ & - \langle na||bc \rangle \Gamma_{ijbc}^{amlk}) \end{aligned} \quad (25)$$

where the three-body RDM (3-RDM) is

$$\Gamma_{ijk}^{npq} = \langle \Psi_0^N | \hat{a}_i^{\dagger} \hat{a}_j^{\dagger} \hat{a}_k^{\dagger} \hat{a}_p \hat{a}_q \hat{a}_n | \Psi_0^N \rangle \quad (26)$$

and the four-body RDM (4-RDM) is

$$\Gamma_{ijkl}^{mnop} = \langle \Psi_0^N | \hat{a}_i^{\dagger} \hat{a}_j^{\dagger} \hat{a}_k^{\dagger} \hat{a}_l^{\dagger} \hat{a}_p \hat{a}_q \hat{a}_r \hat{a}_s | \Psi_0^N \rangle \quad (27)$$

The pertinent metric, $(\mathbf{S}_{-})_{pqr,stu}$ for this generalized eigenvalue problem is

$$\begin{aligned} (\mathbf{S}_{-})_{pqr,stu} = & \langle \Psi_0^N | \hat{a}_p^{\dagger} \hat{a}_q^{\dagger} \hat{a}_r \hat{a}_u^{\dagger} \hat{a}_t \hat{a}_s | \Psi_0^N \rangle \\ = & \delta_{ur} \Gamma_{pq}^{st} - \Gamma_{pqu}^{str} \end{aligned} \quad (28)$$

The storage requirement of the 4-RDM scales as $O(N^8)$, and it becomes prohibitively expensive for more than 16 orbitals. To circumvent this problem, we approximate the 4-RDM via a cumulant expansion. The cumulant approximation to the 4-RDM has been used in multireference perturbation theory and configuration interaction methods previously.^{73,74} In essence, the 4-RDM is approximately constructed from four classes of terms: (1) 1-RDM $\times$ 1-RDM $\times$ 1-RDM $\times$ 1-RDM, (2) 2-RDM $\times$ 1-RDM $\times$ 1-RDM, (3) 2-RDM $\times$ 2-RDM, and (4) 1-RDM $\times$ 3-RDM. Interested readers are referred to ref 74 for more details. To construct the cumulant terms, we wrote a

Python code based on the Fortran code presented in ref 73. For the systems we have investigated, we have found the error of the cumulant approximation is insignificant, and we present results with the reconstructed cumulant 4-RDM later in this work. We further note that Farnum and Mazziotti have used a cumulant expansion for both 3- and 4-RDMs in their EKT-3 calculations.³⁴

Practical implementations may be achieved using spin-orbital expressions where we consider two spin blocks in $c_-[(\alpha\alpha\alpha)$ and $(\alpha\beta\beta)]$ for removing an α electron

$$\hat{O}_\nu^- = \sum_{p_\alpha q_\alpha r_\alpha} (c_-)_{p_\alpha q_\alpha r_\alpha}^\nu \hat{a}_{r_\alpha}^\dagger \hat{a}_{q_\alpha} \hat{a}_{p_\alpha} + \sum_{p_\alpha q_\beta r_\beta} (c_-)_{p_\alpha q_\beta r_\beta}^\nu \hat{a}_{r_\beta}^\dagger \hat{a}_{q_\beta} \hat{a}_{p_\alpha} \quad (29)$$

Consequently, this leads to four distinct spin blocks for $\mathbf{F}$: $(\alpha\alpha\alpha\alpha\alpha)$, $(\alpha\beta\beta\alpha\alpha)$, $(\alpha\alpha\alpha\alpha\beta\beta)$, and $(\alpha\beta\beta\alpha\beta\beta)$. These explicit spin equations can be derived from the spin-orbital expression in eq 25 by carrying out spin integrations for each term.⁷⁵

2.2. Spectral Functions from EKT. We write the retarded single-particle Green's function in a finite basis as⁴

$$iG_{pq}^R(t, t') = \theta(t - t') \langle \Psi_0^N | \{ \hat{a}_p(t), \hat{a}_q^\dagger(t') \} | \Psi_0^N \rangle \quad (30)$$

where $\theta(t)$ is the Heaviside step function. Assuming the $\hat{\mathcal{H}}$ is time-independent, we can write the Green's function in the frequency domain as

$$G_{pq}^R(\omega + i\eta) = \langle \Psi_0^N | \hat{a}_p \frac{1}{\omega - (\hat{\mathcal{H}} - E_0^{(N)}) + i\eta} \hat{a}_q^\dagger | \Psi_0^N \rangle + \langle \Psi_0^N | \hat{a}_q^\dagger \frac{1}{\omega - (\hat{\mathcal{H}} - E_0^{(N)}) + i\eta} \hat{a}_p | \Psi_0^N \rangle \quad (31)$$

where η is a small positive constant and $E_0^{(N)}$ is the ground-state energy of the N -particle system.

The EKT approach offers a systematically improvable way to approximate the evaluation of eq 31. This is because one can form projection operators on the subspace of EKT excitations

$$\hat{P}_\pm = \sum_{\mu\nu} |\Psi_\mu^{N\pm 1}\rangle \langle \Psi_\nu^{N\pm 1}| \quad (32)$$

where $|\Psi_\nu^{N\pm 1}\rangle$ are approximate wave functions obtained via the EKT as defined in eq 1 and $\mathbf{S}^\pm$ is a metric in the pertinent space. Using eq 32, we obtain an approximate $\mathbf{G}^R$

$$G_{pq}^R(\omega + i\eta) \simeq \langle \Psi_0^N | \hat{a}_p \hat{P}_+ \frac{1}{\omega - (\hat{\mathcal{H}} - \tilde{E}_0^{(N)}) + i\eta} \hat{P}_+ \hat{a}_q^\dagger | \Psi_0^N \rangle + \langle \Psi_0^N | \hat{a}_q^\dagger \hat{P}_- \frac{1}{\omega - (\hat{\mathcal{H}} - \tilde{E}_0^{(N)}) + i\eta} \hat{P}_- \hat{a}_p | \Psi_0^N \rangle \quad (33)$$

where $\tilde{E}_0^{(N)}$ is the approximate ground-state energy. This approximation can be systematically improved as higher-order excitations are included in eqs 3 and 2. It is exact when $|\Psi_\nu^{N\pm 1}\rangle$ spans the entire $(N \pm 1)$ -particle Hilbert space. Substituting eq 32 into eq 33 and using $\hat{P}_\pm \hat{\mathcal{H}} \hat{P}_\pm |\Psi_\nu^{N\pm 1}\rangle = \tilde{E}_\nu^{N\pm 1} |\Psi_\nu^{N\pm 1}\rangle$ (from

eq 5), we obtain the (approximate) Lehmann representation of Green's function

$$G_{pq}^R(\omega) = \sum_\nu \frac{\langle \Psi_0^N | \hat{a}_p | \tilde{\Psi}_\nu^{N+1} \rangle \langle \tilde{\Psi}_\nu^{N+1} | \hat{a}_q^\dagger | \Psi_0^N \rangle}{\omega - (\tilde{E}_\nu^{(N+1)} - \tilde{E}_0^{(N)}) + i\eta} + \sum_\nu \frac{\langle \Psi_0^N | \hat{a}_q^\dagger | \tilde{\Psi}_\nu^{N-1} \rangle \langle \tilde{\Psi}_\nu^{N-1} | \hat{a}_p | \Psi_0^N \rangle}{\omega + (\tilde{E}_\nu^{(N-1)} - \tilde{E}_0^{(N)}) + i\eta} \quad (34)$$

where we orthogonalized the eigenvectors by

$$|\Psi_\nu^{N\pm 1}\rangle = \sum_\mu |\Psi_\mu^{N\pm 1}\rangle (\mathbf{S}^\pm)_{\mu\nu}^{-1/2} \quad (35)$$

Details concerning the numerical implementation of this orthogonalization procedure are given in Appendix B.

A spectral function in a finite basis set can be computed from

$$\mathbf{A}_{pq}(\omega) = -\frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \text{Im}[G_{pq}^R(\omega + i\eta)] = \mathbf{A}_{pq}^>(\omega) + \mathbf{A}_{pq}^<(\omega) \quad (36)$$

where

$$\mathbf{A}_{pq}^>(\omega) = \sum_\nu \langle \Psi_0^N | \hat{a}_p | \Psi_\nu^{N+1} \rangle \langle \Psi_\nu^{N+1} | \hat{a}_q^\dagger | \Psi_0^N \rangle \times \delta(\omega - (E_\nu^{(N+1)} - E_0^{(N)})) \quad (37)$$

and

$$\mathbf{A}_{pq}^<(\omega) = \sum_\nu \langle \Psi_0^N | \hat{a}_q^\dagger | \Psi_\nu^{N-1} \rangle \langle \Psi_\nu^{N-1} | \hat{a}_p | \Psi_0^N \rangle \times \delta(\omega + (E_\nu^{(N-1)} - E_0^{(N)})) \quad (38)$$

where $\mathbf{A}^>$ and $\mathbf{A}^<$ are the addition and removal single-particle spectral functions, which describe inverse and direct photoemission experiments, respectively, in the sudden approximation.

Using the definition of eq 1, eqs 37 and 38 can be expressed in terms of directly computable quantities, $\tilde{\mathbf{c}}_\pm^\nu$ (orthogonalized eigenvectors) and Γ_{pq} in the case of EKT-1

$$\mathbf{A}_{pq}^>(\omega) = \sum_{rs} \sum_\nu \tilde{\mathbf{c}}_+^\nu (\tilde{\mathbf{c}}_+^\nu)^\dagger \delta(\omega - \epsilon_+^\nu) \times (\delta_{pr} - \Gamma_{rp}) (\delta_{sq} - \Gamma_{qs}) \quad (39)$$

and

$$\mathbf{A}_{pq}^<(\omega) = \sum_{rs} \sum_\nu \tilde{\mathbf{c}}_-^\nu (\tilde{\mathbf{c}}_-^\nu)^\dagger \delta(\omega + \epsilon_-^\nu) \Gamma_{qr} \Gamma_{sp} \quad (40)$$

Similarly, for EKT-3, 2-RDM naturally arises in the evaluation of the spectral functions. The working equation for IP states is as follows

$$\mathbf{A}_{pq}^<(\omega) = \sum_\nu \sum_{ijklmn} \Gamma_{qk}^{ij} \Gamma_{lm}^{pn} \tilde{\mathbf{c}}_{ijk}^\nu (\tilde{\mathbf{c}}_{lmn}^\nu)^* \delta(\omega + \epsilon_-^\nu) \quad (41)$$

It is straightforward to find similar equations for the EA states. Furthermore, we note that the density of state (DOS) is simply defined as

$$g(\omega) = \frac{\text{tr}(\mathbf{A}(\omega))}{M} \quad (42)$$

where M is the number of single-particle basis functions. We note that for solid-state applications, single-particle basis carries an additional index for the crystalline momentum $\mathbf{k}$, which can be straightforwardly incorporated into the above formalism. In such applications, it is useful to compute the momentum-dependent DOS, $g(\mathbf{k}, \omega)$.

2.3. Phaseless Auxiliary-Field Quantum Monte Carlo.

We refer readers to [Appendix A](#) for a short summary of the ground-state computation with ph-AFQMC. In ph-AFQMC, the simplest way to evaluate the expectation value of an operator $\hat{O}$ is using the mixed estimator

$$\langle \hat{O} \rangle_{\text{mixed}} := \frac{\langle \Psi_T | \hat{O} | \Psi(\tau) \rangle}{\langle \Psi_T | \Psi(\tau) \rangle} \quad (43)$$

$$= \frac{\sum_i w_i(\tau) \frac{\langle \Psi_T | \hat{O} | \psi_i(\tau) \rangle}{\langle \Psi_T | \psi_i(\tau) \rangle}}{\sum_i w_i(\tau)} \quad (44)$$

where all relevant variables are defined in [Appendix A](#). The mixed estimator is an unbiased estimator only for operators that commute with the Hamiltonian. For operators that do not commute with the Hamiltonian, the mixed estimator can introduce significant biases due to the approximate trial wave functions that can be practically used. To overcome this, we use the BP algorithm^{48,66,67,76} and write

$$\langle \hat{O} \rangle \approx \lim_{\kappa \rightarrow \infty} \frac{\langle \Psi_T | e^{-\kappa \hat{H}} \hat{O} | \Psi(\tau) \rangle}{\langle \Psi_T | e^{-\kappa \hat{H}} | \Psi(\tau) \rangle} \quad (45)$$

$$= \lim_{\kappa \rightarrow \infty} \frac{\sum_i w_i(\tau + \kappa) \frac{\langle \psi_i(\kappa) | \hat{O} | \psi_i(\tau) \rangle}{\langle \psi_i(\kappa) | \psi_i(\tau) \rangle}}{\sum_i w_i(\tau + \kappa)} \quad (46)$$

To summarize, we propagate $|\Psi\rangle$ until $\kappa + \tau$, storing the walker wave function at time τ . We can then split the propagation into κ BP and τ forward propagation, as in [eq 46](#). The back-propagated wave function is constructed by applying walker's propagators to the trial wave function from the κ portion of the path. Practically, the convergence of the expectation value has to be monitored with respect to the BP time κ . It should be emphasized that in ph-AFQMC, the walker wave function is a single determinant wave function.

It was found in [ref 67](#) that the standard BP algorithm described in [eq 46](#) can yield poor results in ph-AFQMC when applied to *ab initio* systems. The authors devised a number of additional steps to reduce the phaseless error, the most accurate of which was to partially restore the phase and cosine factors along the BP portion of the path. In this work, we restore phases along the back propagated path and along the forward direction. Practically, this amounts to storing the phases and cosine factors between $[\tau - \kappa, \tau + \kappa]$ and multiplying these by the weights appearing in [eq 46](#). This additional restoration of paths along the forward direction was not described in [ref 67](#) but was used in practice,^{77,78} and we found it necessary to obtain more accurate results for the systems studied here.

In EKT1-AFQMC, we directly sample $F_{\pm}$ using the back-propagated estimator form. This boils down to the evaluation of the 2-RDM appearing in [eq 22](#) using the back propagated 1-RDM via Wick's theorem:

$$\Gamma_{pt}^{rs} = \Gamma_{pr} \Gamma_{ts} - \Gamma_{ps} \Gamma_{tr} \quad (47)$$

With these ingredients, we can evaluate $F_{\pm}$ by contracting the one- and two-body matrix elements with the back-propagated 1- and 2-RDMs.

While efficient implementations are not the focus of our efforts in this work, we mention the computational cost of producing one back-propagated sample of $F^{\pm}$ and $S^{\pm}$. A sample of $S^{\pm}$ has the same overhead as computing a back-propagated 1-RDM sample which scales as $O(NM^2)$, where M is the number of orbitals and N is the number of occupied orbitals. The cost for producing a sample of $F^{\pm}$ is more involved and depends on the integral factorization that one chooses to use. Using the most common integral factorization, that is, the Cholesky factorization, it can be shown that the cost scales as $O(M^3X)$, where X is the number of Cholesky vectors (also note that the 2-RDM is never explicitly formed). With tensor hypercontraction,^{64,79–82} the cost can be brought down to overall cubic. If one were to just implement a matrix-vector product for iterative eigensolvers, the Cholesky factorization can achieve cubic scaling per matrix-vector product as well. The cost is increased in EKT3-AFQMC where each matrix-vector product sample costs $O(M^5)$. It is potentially possible to reduce this cost further by also factorizing EKT amplitudes in a THC format, as is done in [ref 81](#).

We leave the exploration of EKT3-AFQMC for the future study and focus on EKT1-AFQMC in this work.

2.4. Uniform Electron Gas. Aside from small atomic and molecular benchmarks, we also study the spectral properties of finite-basis supercell models of the UEG. We assume atomic units and use the spin-orbital representation. For readers unfamiliar with the spin-orbital representation, we present the spin-explicit Hamiltonian in [Appendix C](#). The UEG Hamiltonian is usually defined in the plane-wave basis set, which gives the one-body operator

$$\hat{H}_1 = \sum_{\mathbf{K}} \frac{|\mathbf{K}|^2}{2} a_{\mathbf{K}}^{\dagger} a_{\mathbf{K}} \quad (48)$$

and the electron-electron interaction operator is (in a spin-orbital basis)

$$\hat{H}_2 = \frac{1}{2\Omega} \sum_{\mathbf{K} \neq 0, \mathbf{K}_1, \mathbf{K}_2} \frac{4\pi}{|\mathbf{K}|^2} a_{\mathbf{K}_1 + \mathbf{K}}^{\dagger} a_{\mathbf{K}_2 - \mathbf{K}}^{\dagger} a_{\mathbf{K}_2} a_{\mathbf{K}_1} \quad (49)$$

where $\mathbf{K}$ here is a plane wave vector and Ω is the volume of the unit cell. In addition to $\hat{H}_1$ and $\hat{H}_2$, there is a constant term that arises due to a finite-size effect. Specifically, the Madelung energy E_M should be included to account for self-interactions associated with the Ewald sum under periodic boundary conditions⁸³ via

$$E_M = \frac{N}{2} \xi \quad (50)$$

with

$$\xi = -2 \times 2.837297 \times \left(\frac{3}{4\pi} \right)^{1/3} N^{-1/3} r_s^{-1} \quad (51)$$

where N is the number of electrons in the unit cell and r_s is the Wigner-Seitz radius. We define the UEG Hamiltonian as a sum of these three terms

$$\hat{H}_{\text{UEG}} = \hat{H}_1 + \hat{H}_2 + E_M \quad (52)$$

The Madelung constant can be either included in the Hamiltonian as written in [eq 52](#) or it can be included as an a

a posteriori correction to the simulation done without it. When the latter choice is made, the spectral functions have to be shifted accordingly in order to compare the results obtained from the former approach. The corresponding shift can be derived from a shift, Δ , in the poles

$$\Delta_{\text{IP}} = E_{\text{M}}(N-1) - E_{\text{M}}(N) = -\frac{\xi}{2} \quad (53)$$

$$\Delta_{\text{EA}} = E_{\text{M}}(N+1) - E_{\text{M}}(N) = \frac{\xi}{2} \quad (54)$$

Regardless of whether we included the Madelung constant in the Hamiltonian, there is an additional correction of $-\xi/2$ for both IP and EA to remove spurious image interactions coming from the excess charge created.⁸⁴ Therefore, overall electron removal poles are shifted by $-\xi$ and electron addition poles remain the same.

For many molecular quantum chemistry methods, often the two-electron integral tensor is assumed to be eightfold symmetric. Practical implementations utilize this symmetry to simplify equations as well. As such, without the eightfold symmetry, molecular quantum chemistry methods would not produce correct answers even though the UEG Hamiltonian only contains real-valued matrix elements. This complication of the UEG Hamiltonian becomes more obvious once we write it in the form of eq 19 with

$$\langle pq|rs \rangle = \frac{1}{\Omega} \frac{4\pi}{|\mathbf{K}_p - \mathbf{K}_r|^2} \delta_{\mathbf{K}_p - \mathbf{K}_r, \mathbf{K}_q - \mathbf{K}_s} \times (1 - \delta_{\mathbf{K}_p, \mathbf{K}_r}) \quad (55)$$

The permutation between p and r or between q and s alters the value of the integral tensor because of the Kronecker delta term. This is a direct consequence of using a plane wave basis which is complex, unlike the usual Gaussian orbitals. To circumvent any complications due to this, we perform a unitary transformation that rotates the plane wave basis into a real-valued basis. Namely, for given $\mathbf{K}$ and $-\mathbf{K}$ (assuming $\mathbf{K} \neq \mathbf{0}$), we use

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \quad (56)$$

We apply this transformation to every pair of $\mathbf{K}$ and $-\mathbf{K}$ in the two-electron integral tensor in eq 55. The resulting transformed integral tensor now recovers the full eightfold symmetry. One can also transform observables such as spectral functions back to the original basis using $\mathbf{U}^\dagger$ when necessary.

3. COMPUTATIONAL DETAILS

All quantum chemistry calculations are performed with PySCF,⁸⁵ which include mean-field (HF) calculations, coupled-cluster with singles and doubles (CCSD), and CCSD with perturbative triples (CCSD(T)). All one- and two-electron integrals needed for ph-AFQMC were also generated with PySCF. ph-AFQMC calculations were mostly performed with QMCPACK,⁸⁶ and PAUXY⁸⁷ was used to cross-check some results.

We use a selected configuration interaction (CI) method called heat-bath CI (HCI)^{88–90} to produce numerically exact IPs within a basis whenever possible. Furthermore, we compute the HCI IPs within EKT1 (i.e., EKT1-HCI) using the 1- and 2-RDM from variational HCI wave functions along with eq 20. Since there can be an inherent bias of EKT1 itself, we provide

EKT1-HCI as an “exact” result for IPs within the limits of the EKT1 approach. This can be used to quantify the phaseless and BP errors in EKT1-AFQMC. In HCI, there is a single tunable parameter, ϵ_1 that controls the variational energy which is used to select determinants to be included in the variational expansion. We also use the 3-RDM of the variational wave function to compute the 4-RDM via the cumulant construction.^{73,74} These 3- and 4-RDMs are further used to construct the EKT3 Fock matrix in eq 25. This approach is referred to as EKT3-HCI. The eigenvalues and eigenvectors of the EKT3 Fock matrix (with its pertinent metric) can then be used to produce IPs and spectral functions of EKT3. We tuned ϵ_1 to be such that the resulting second-order Epstein–Nesbet perturbation energy is no greater than 1 mE_h for every system, except the 14-electron 19-plane-wave model of the UEG, where we observed a PT2 correction of 3 mE_h at $r_s = 4$. This was found to be sufficient to produce accurate EKT IPs for systems studied here. All calculations are performed with a locally modified version of Dice.^{88–90}

We used a time step of 0.01 au, and the pair branch population control method⁹¹ used the hybrid propagation scheme⁶⁶ for all ph-AFQMC simulations. For the small atoms and molecules and 14-electron UEG examples, we used 2880 walkers, while for the 54-electron UEG, we used 1152 walkers. All calculations used restricted HF (RHF) trial wave functions except for the charged species where we instead used unrestricted HF (UHF) wave functions. The exception to this was CH₄, where we used the same RHF orbitals for the charged species as we found that the UHF solution broke spatial symmetry and led to a large phaseless constraint bias. All AFQMC results are performed with the phaseless approximation, so we simply refer ph-AFQMC to as AFQMC in the following sections.

We adapted the standard dynamical Lanczos algorithm^{92,93} to obtain spectral functions of eq 52. Even by exploiting symmetry and using a distributed sparse Hamiltonian, dynamical Lanczos results could only be obtained for the smallest UEG system with 14 electrons in 19 plane waves, corresponding to an N -electron Hilbert space size of 2.5×10^9 determinants. In the dynamical Lanczos algorithm, one first obtains the N -particle ground state, $|\Psi_0^N\rangle$, iterating within the Lanczos Krylov subspace. Ultimately, our goal is to compute eq 36, which then requires another run of the Lanczos algorithm. For an electron removal problem, we pick an orbital index i and generate an initial vector in the $(N-1)$ -electron sector, $|f_0\rangle = \hat{a}_i |\Psi_0^N\rangle / \langle \Psi_0^N | \hat{a}_i^\dagger \hat{a}_i | \Psi_0^N \rangle$. Each Lanczos iteration then generates coefficients, $\{a_k\}$ and $\{b_k\}$, for the following continued fraction expression^{92,93}

$$A_{ii}(\omega, \eta) = -\frac{1}{\pi} \text{Im} \frac{\langle \Psi_0^N | \hat{a}_i^\dagger \hat{a}_i | \Psi_0^N \rangle}{z - a_0 - \frac{b_1^2}{z - a_1 - \frac{b_2^2}{z - a_2 - \dots}}} \quad (57)$$

where $z = E_0^{(N)} - \omega + i\eta$ with some spectral broadening constant η . We take a total of 50 Lanczos iterations to generate the continued fraction coefficients and this was enough to converge the low-energy spectrum within the energy scale that is relevant in this work.

4. RESULTS AND DISCUSSION

While the EKT approach is valid for both electron removal and electron addition, for numerical results, we focus on electron removal processes (IP energies and electron removal spectral functions) for simplicity. We benchmark the IP energies from the proposed EKT1-AFQMC approach over several small

chemical systems and the finite-basis supercell model of UEG. Furthermore, we also show promising improvements over EKT1 using EKT3-HCI in cases where satellite peaks are prominent in the 14-electron 19-plane-wave UEG model. We use a “ Δ method” to denote a scheme, where we run the pertinent method for both N - and $(N - 1)$ -electron systems and obtain the IP as an energy difference.

4.1. Small Chemical Systems in the aug-cc-pVDZ Basis.

In this section, we study seven small chemical systems (He, Be, Ne, FH, N₂, CH₄, and H₂O) that have well-documented experimental IPs.⁹⁴ We use the nomenclature FH for the hydrogen fluoride molecule to distinguish it from the abbreviation for HF. All geometries were taken from ref 94. We used a relatively small basis set, namely, aug-cc-pVDZ,⁹⁵ to obtain good statistics in back-propagated estimators. We have used more than 3000 back-propagated estimator samples in all cases considered here, each of which requires a BP time of greater than 4 a.u. This results in a total propagation time longer than 1,2000 a.u., which is unusually long for standard AFQMC calculations. The use of this basis set also allows for a direct comparison between AFQMC and numerically exact HCI within this basis set.

The goal of this numerical section is to quantify the three sources of error in addition to the basis set incompleteness error in EKT1-AFQMC based on simple examples where exact simulations are possible. These three sources of error are as follows:

1. Phaseless constraint errors. As mentioned, the phaseless constraint is necessary to remove the phase problem that arises in the imaginary-time propagation. However, due to this constraint, the resulting ground-state energies and properties (e.g., RDMs) are biased.
2. BP errors. The BP algorithm incurs additional errors. This was noted and studied in detail in ref 67. For instance, in ref 67, it was shown that for neon, the phaseless error with a simple trial wave function is negligible (below 1 mE_h), but the error in the one-body energy from the back-propagated 1-RDM was around 5 mE_h.
3. EKT1 errors. While systematically improvable with higher-order excitations, EKT1 is not an exact approach to quasiparticle spectra unless all orders of excitations are included. Nonetheless, for the first IP, it has been numerically and analytically suggested that EKT1 approaches the exact IP in the basis set limit if the exact 1- and 2-RDMs are used.^{39,96–98} Beyond the first IP, we will show that EKT1 qualitatively fails to capture satellite peaks that arise in the case of the core spectrum of this model.

In Table 1, we present numerically exact first IPs of molecules within this basis set using Δ HCI. The basis set incompleteness error can be as large as 0.3 eV in these molecules, and therefore, we will only compare AFQMC results to these numerically exact results in the same basis set as opposed to comparing to the experimental data. We do not expect the qualitative conclusions of our study to change with larger basis sets.

4.1.1. Phaseless constraint Errors. Next, we assess the phaseless bias in the N -electron system ground-state energy and the error in the Δ AFQMC IP energies compared to the Δ HCI IPs. In Table 2, we present numerical data that detail the phaseless bias in these quantities. The ground-state energy error is less than 0.04 eV, which is in the neighborhood of the usual standard of accuracy, 1 mE_h. Unfortunately, in many cases, ($N -$

Table 1. Experimental First IPs (eV) and Deviation (eV) of the Numerically Exact Δ HCI from These Results for Chemical Species Considered^a

	experiment	Δ HCI
He	24.59	−0.23
Be	9.32	−0.03
Ne	21.56	−0.13
FH	16.19	−0.12
N ₂	15.6	−0.34
CH ₄	14.35	−0.06
H ₂ O	12.62	−0.07

^a Δ HCI employed the aug-cc-pVDZ basis.

Table 2. Deviation (eV) in the AFQMC N -Electron System Ground-State Energy and in the First IP with Respect to the Corresponding HCI Results^a

	deviation in the ground state	deviation in IP
He	0.00	0.00
Be	0.01	−0.01
Ne	−0.03	0.07
FH	−0.03	0.07
N ₂	−0.04	0.06
CH ₄	−0.03	0.06
H ₂ O	−0.03	0.04

^aStatistical error bar of AFQMC is less than 0.01 eV, and therefore, we do not present them here.

1)-electron systems incur as large a phaseless bias as do the N -electron systems, albeit with an opposite sign of the error. Thus, AFQMC does not benefit from a cancellation of errors for the IP energy, which results in IP errors that are larger than those for the ground-state energy. The largest IP error we find is around 0.07 eV.

4.1.2. BP Errors. We present the EKT1 results in Table 3. We refer readers to Appendix B, where a detailed description of the

Table 3. Error (eV) in the First IP Obtained from EKT1-AFQMC and the KT Relative to EKT1-HCI^a

	EKT1-HCI	KT	EKT1-AFQMC
He	24.36	0.60	0.02
Be	9.29	−0.87	0.06
Ne	21.48	1.73	−0.04
FH	16.13	1.58	0.01
N ₂	15.34	1.92	0.15
CH ₄	14.13	0.68	0.19
H ₂ O	12.60	1.27	−0.04

^aStatistical error bar of EKT1-AFQMC cannot be estimated without bias (see the main text for discussion).^{99,100}

EKT1 calculation is given. Theoretically “exact” EKT1 results can be obtained by using exact RDMs from HCI. To quantify the BP error of AFQMC (given that the phaseless error is very small for these systems), we shall compare EKT1-AFQMC to EKT1-HCI. We also computed simple KT IPs using HF and report these results in Table 3. The error of EKT1-AFQMC is small for most chemical species, but it becomes as large as 0.19 eV for CH₄. Even though the phaseless bias in the ground-state energy was found to be very small (0.04 eV or less), EKT1-AFQMC errors from using the back-propagated 1-RDM and the EKT Fock matrix can be five times larger. Therefore, we attribute this

error mainly to the BP error. Nonetheless, the comparison against the simpler KT suggests that EKT1-AFQMC can readily recover the correlation contribution to quasiparticle energies with errors less than 0.2 eV in these examples.

We note that EKT1-AFQMC IPs do not have any statistical error bars. This is not because these numbers are deterministic but arises because EKT eigenvalues are not unbiased estimators. This was also observed in a similar approach called Krylov-projected full CI QMC. Interested readers are referred to ref 100 for details. In essence, eigenvalues of a noisy matrix where each element is normally distributed are not normally distributed. Therefore, statistical error bars are difficult to estimate and are not simply associated with variances of Gaussian distributions. Given the small statistical error bars (on the order of 3×10^{-4} or less) on the diagonal elements of the EKT1 Fock matrix and 1-RDM, we expect that these results are reproducible up to 0.01 eV if one follows exactly the same numerical protocol given in Appendix B.

4.1.3. EKT1 Errors. Finally, we discuss the inherent error of EKT1 by comparing EKT1-HCI and EKT1-AFQMC against Δ HCI as in Table 4. We also compare a more widely used

Table 4. Error (eV) in the First IP Obtained from EKT1-HCI, EKT1-AFQMC, and EOM-IP-CCSD Relative to Δ HCI (Given in Table 1) in the aug-cc-pVDZ Basis^a

	EKT1-HCI	EKT1-AFQMC	EOM-IP-CCSD
He	0.00	0.02	0.00
Be	0.00	0.06	0.00
Ne	0.05	0.02	−0.28
FH	0.06	0.07	−0.22
N ₂	0.05	0.20	0.14
CH ₄	−0.15	0.04	−0.02
H ₂ O	0.05	0.01	−0.15

^aStatistical error bar of EKT1-AFQMC cannot be estimated without biases (see the main text for discussion).^{99,100}

approach called equation-of-motion coupled cluster IP with singles and doubles (EOM-IP-CCSD) to gauge the magnitude of EKT1 errors. Despite the fact that these are small molecules, EKT3-HCI is not feasible with our pilot implementation because there are too many orbitals to handle. Furthermore,

exact spectral functions are unavailable as our Lanczos implementation is specialized to the finite-basis supercell model of UEG. Since this is a benchmark on the first IP of molecules, EKT1-HCI is expected to be quite accurate. This expectation is due to the general belief that EKT1 with exact RDMs yields exact first IP.^{39,96–98} Given this, the small errors of EKT1-HCI in Table 4 are not surprising. The only exception is CH₄ with an error of −0.15 eV, which we believe will still approach the exact first IP in the complete basis set limit. This is because the long-range behavior of exact density should be related to the exact first IP in the basis set limit, as shown in ref 33. We note, however, that the same exactness may not apply to any other states. EKT1-AFQMC appears to be as good as EKT1-HCI, with an outlier for the case of N₂. The error of EKT1-AFQMC relative to EKT1-HCI is comparable to the error of EKT1-HCI relative to Δ HCI in this basis set. EOM-IP-CCSD generally does not work as well as the EKT1 approaches with a maximum error of −0.28 eV on the neon atom. While these finite basis set comparisons are informative, we emphasize that more fair comparisons should be conducted in the complete basis set limit and we hope to carry these out in the future. Regardless, the EKT1-AFQMC results are encouraging.

4.1.4. EKT1 Spectral Functions. The main motivation for performing EKT1 within AFQMC was to obtain spectral functions. Poles alone can be obtained using the ground-state AFQMC algorithm (i.e., Δ AFQMC) by imposing a proper constraint for excited-state descriptions. While the choice of proper trials is a challenge for this purpose, such an approach avoids the complications due to BP. However, spectral weights cannot be obtained from Δ AFQMC. We show EKT1-AFQMC spectral functions for FH and N₂ in Figure 1. Based on the results shown in Table 3, EKT1-AFQMC is in good agreement with EKT1-HCI for FH but not for N₂. Therefore, comparing these two cases is useful for understanding how BP errors are reflected in spectral functions. In the case of FH, we do not see any visible differences between EKT1-HCI and EKT1-AFQMC on the plotted energy scale. However, for N₂, we can clearly see some deviation between EKT1-AFQMC and EKT1-HCI. Nonetheless, the main features of the spectral function are reproduced, namely, three large quasiparticle peaks with the middle peak being the largest. We note that there are peaks with very small spectral weights in EKT1-HCI close to −30, −27, and −5 eV

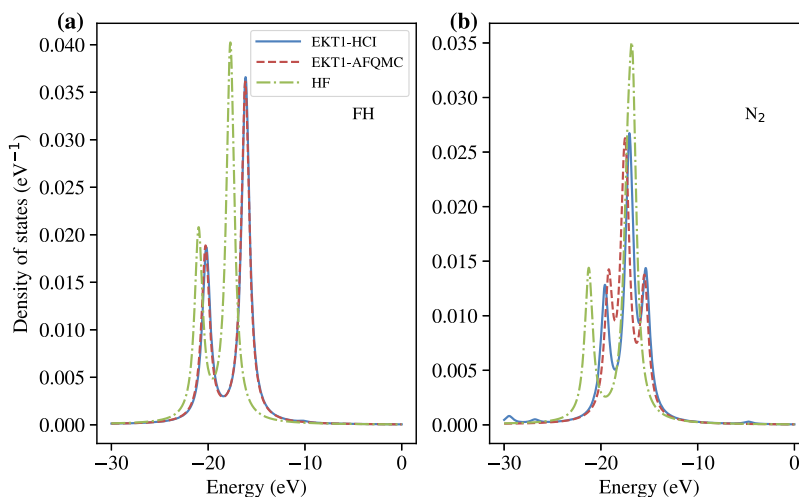


Figure 1. Electron removal spectral functions from various methods (EKT1-HCI, EKT1-AFQMC, and HF) of (a) FH and (b) N₂ in the aug-cc-pVDZ basis set. Note that in (a), EKT1-AFQMC is right on top of EKT1-HCI on the plotted scale. A broadening parameter $\eta = 0.5$ eV was used.

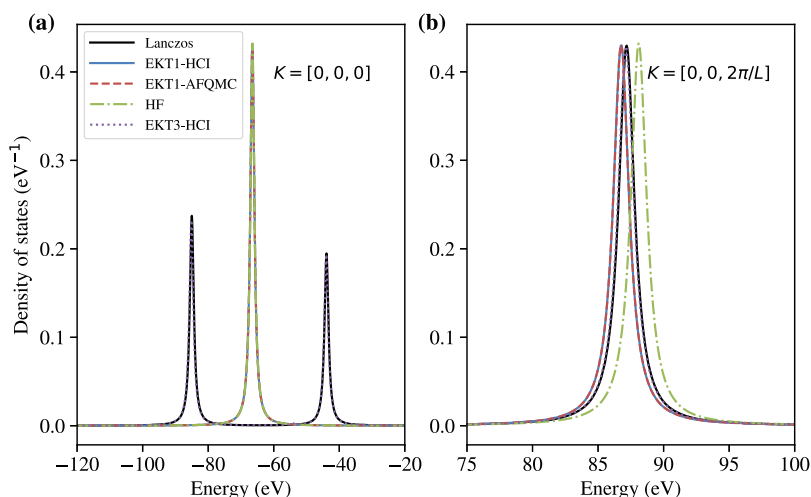


Figure 2. Electron removal spectral functions of the 14-electron in 19-plane-wave UEG model from various methods at $r_s = 0.5$: (a) $\mathbf{K} = [0, 0, 0]$ and (b) $\mathbf{K} = [0, 0, 2\pi/L]$. In (a), note that EKT1-AFQMC, EKT1-HCI, and HF are right on top of each other. In (b), EKT1-AFQMC and EKT1-HCI are right on top of each other. In both (a,b), Lanczos and EKT3-HCI are right on top of each other. A broadening parameter of 0.2 eV was used for all plots. The location of the Fermi level is 158.6 eV.

and that these features these do not appear in EKT1-AFQMC. We attribute this to a relatively large linear dependency cutoff (10^{-4}) needed in EKT1-AFQMC to stabilize the generalized eigenvalue problem, as explained in Appendix B. In both molecules, there are insignificant differences between the EKT1 and HF spectra in terms of the peak heights, locations, and the number of peaks with a broadening parameter of $\eta = 0.5$ eV.

4.2. Finite-Basis Models of the UEG. AFQMC has emerged as a unique tool for simulating correlated solids.^{55,57,60,63,101} A model solid that describes the basic physics of metallic systems is the UEG. The accuracy and scope of AFQMC in studying the finite-basis supercell model of UEG has been well documented at zero temperature and finite temperature.^{57,101} Motivated by these studies, we investigate the spectral properties of the model UEG within the EKT approaches (EKT1 and EKT3). We do not make any comparisons to known spectral functions in the thermodynamic limit^{102,103} as we are far away from the thermodynamic limit and the basis set limit.

4.2.1. 14 Electrons/19 Plane Waves. The first example that we consider is a relatively small UEG supercell with only 19 plane waves. It is far from the basis set limit and from the thermodynamic limit. However, it is small enough for one to produce unbiased EKT results using HCI and numerically exact dynamical Lanczos results. We note that the spectral function of the 14-electron 19-plane-wave UEG model at $r_s = 4$ was first presented in ref 16. We produced around 3000 BP samples, which yielded the largest statistical error in the Fock matrix and 1-RDM on the order of 5×10^{-4} .

We consider two Wigner-Seitz radii, $r_s = 0.5$ and $r_s = 4$. Based on our previous benchmark study of AFQMC on this system, we expect that the phaseless error in the ground state at $r_s = 0.5$ is negligible, while the error is relatively more noticeable at $r_s = 4$.⁵⁷ Compared to the numerically exact energies, it was found that the constraint bias in AFQMC is only $-0.0118(6)$ eV at $r_s = 0.5$ and $0.185(2)$ eV at $r_s = 4$. Given this small ground-state bias, we expect that the EKT1-AFQMC approach would be as accurate as EKT1-HCI if good statistics and accuracy in the back-propagated estimators can be achieved.

In Figure 2, we present spectral functions of this model at $r_s = 0.5$ for two momenta. The first is at $\mathbf{K} = [0, 0, 0]$ which

represents removal of an electron from the core shell of the 14-electron 19-plane-wave UEG model. Core spectra have been shown to have rich satellite features where different many-body methods do not agree in terms of the precise satellite structure.¹⁶ However, such features are very simple in a small supercell such as this one. As can be seen in Figure 2a, we only have two peaks from the Lanczos method. Neither of these peaks corresponds to a single Koopmans-like excitation. Namely, they cannot be found from a simple single ionization process that either HF (i.e., Koopmans theorem) or EKT1 describes well. As a result of this, HF, EKT1-AFQMC, and EKT-HCI all fail to capture the peak split and only yield a single peak. The correlation effect is very marginal in the sense that nearly no improvement was observed with the EKT1 methods compared to HF.

A significant improvement can be made by incorporating higher-order terms such as 2h1p excitations. In other words, core ionization satellite states (up to leading order) require excitations such as

$$\left(\sum_{\mathbf{K}' \neq 0} C_{\mathbf{K}'} \hat{a}_{\mathbf{K}'}^\dagger \hat{a}_{\mathbf{K}'} \right) \hat{a}_{\mathbf{K}=0} \quad (58)$$

where $C_{\mathbf{K}'}$ are the excitation amplitudes. All of these excitations are included in EKT3. EKT3-HCI can nearly completely reproduce the exact spectral function despite the use of the cumulant approximation for the 4-RDM. The cumulant approximation error is small, especially in weakly correlated cases such as $r_s = 0.5$ where the connected component of 4-RDM is expected to be small. This was indeed found to be the case when we compared EKT3 spectral functions constructed with the exact 4-RDM and the cumulant approximation to the 4-RDM in a truncated basis set calculation (i.e. with 12 plane waves). The difference between the two is not visible on the plotted energy scale. In Figure 2b, we emphasize that we observe meaningful improvements even from the EKT1 methods compared to HF. $\mathbf{K} = [0, 0, 2\pi/L]$ corresponds to the top of the valence band corresponding to the first IP of the 14-electron 19-plane-wave UEG model. There is no satellite peak visible at this momentum and the single peak found from the EKT1 methods is reasonable. The peak location of HF is displaced by around +0.9 eV from the correct location, while the EKT1

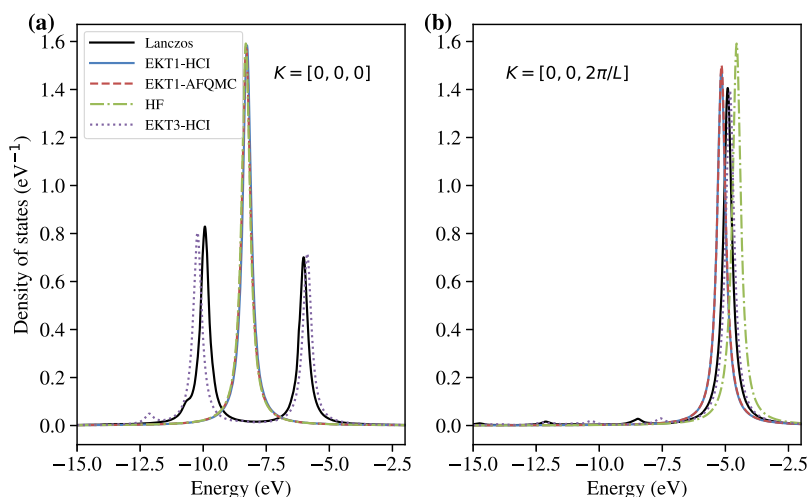


Figure 3. Electron removal spectral functions of the 14-electron in 19-plane-wave UEG model from various methods at $r_s = 4.0$: (a) $\mathbf{K} = [0, 0, 0]$ and (b) $\mathbf{K} = [0, 0, 2\pi/L]$. In (a), note that EKT1-AFQMC, EKT1-HCI, and HF are right on top of each other. In (b), EKT1-AFQMC and EKT1-HCI are nearly on top of each other. A broadening parameter of 0.2 eV was used for all plots. The location of the Fermi level is 3.5 eV.

methods yield a peak shifted by around -0.6 eV. We note that EKT1-AFQMC is practically indistinguishable from EKT1-HCI for both momenta, which indicates a small phaseless bias, a small BP error, and good statistics for the estimators.

A similar conclusion can be drawn for $r_s = 4$, as shown in Figure 3. For the core excitation spectrum in Figure 3a, we observe the same split peak structure observed at $r_s = 0.5$. We see another smaller peak emerging on the left shoulder of the peak near -12 eV. While EKT3-HCI is no longer exact, it reproduces most of the features in the exact spectral function including the emergence of the third peak. While EKT1-HCI and EKT1-AFQMC agree well, there is no visible improvement over HF. The EKT1 methods all yield a single peak, which is qualitatively wrong. The valence excitation structure illustrated in Figure 3b is relatively featureless, but there are small peaks emerging in the high-energy (more negative) region of the spectrum. EKT3-HCI shows good agreement with Lanczos for the main quasiparticle peak and also produces satellite features. There is a visible improvement of EKT1 approaches (with a deviation of the peak energy of approximately -0.25 eV) compared to HF (with an approximate deviation of $+0.34$ eV). A slight deviation of EKT1-AFQMC from EKT1-HCI is observed, but the difference in the main quasiparticle peak location is only around 0.01 eV. Similar to $r_s = 0.5$, a smaller basis set test suggests that the cumulant approximation makes an error that is not visible on the plotted scale even for $r_s = 4$.

Overall, in this small benchmark study, the EKT1 approaches provide some improvement over HF for valence excitations and qualitatively fail for the core region. The agreement between the EKT1 valence peaks and the Lanczos peaks is not perfect, with an error of around -0.6 eV for $r_s = 0.5$ and -0.25 eV for $r_s = 4$. However, we emphasize again that we expect EKT1 to become exact for the first IP as the complete basis set limit is approached. Finally, EKT1-AFQMC is able to reproduce EKT1-HCI nearly perfectly even for $r_s = 4$, where the phaseless error in the ground-state energy is around 0.185(2) eV.

4.2.2. 54 Electrons/257 Plane Waves. Next, we consider a larger UEG supercell (54 electrons in 257 plane waves) where obtaining many BP samples is difficult. We study $r_s = 2$, where AFQMC can be reliably extrapolated to the basis set limit.⁵⁷ We produced 600 BP samples with a BP time of 8 a.u. This amounts to a total of 4800 a.u. propagation time, which is a long

propagation for this system size. Our approach yielded a maximum statistical error in 1-RDM and Fock matrices of 4×10^{-3} . While this error is not small, the procedure described in Appendix B was enough to stabilize the final results. Unlike previous cases, we take the upper triangular part of the Fock matrix and explicitly symmetrize the Fock matrix. A linear dependency cutoff of 10^{-3} was used in EKT1-AFQMC. It is difficult to generate highly accurate 1- and 2-RDMs from HCI for this system size, so for this system, we do not have an exact benchmark reference to compare to our EKT1-AFQMC results. Similarly, EKT3-HCI is also intractable for this system size. Instead, we have performed EOM-IP-CCSD and Δ AFQMC to compare with and to gauge the magnitude of the errors of EKT1-AFQMC.

In Figure 4, EKT1-AFQMC and HF spectral functions are shown. As expected, EKT1-AFQMC does not show any satellite

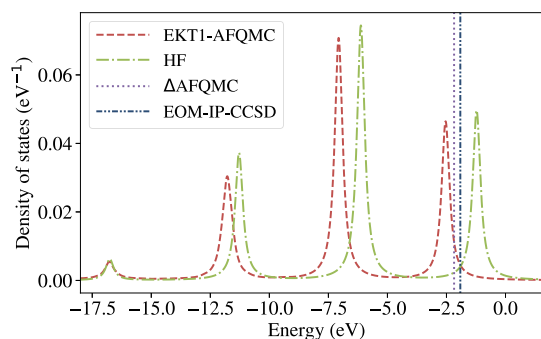


Figure 4. Electron removal spectral functions of the 54-electron in 257-plane-wave model UEG at $r_s = 2$ from EKT1-AFQMC and HF. The first IPs from Δ AFQMC and EOM-IP-CCSD are shown for comparisons. A broadening parameter of 0.2 eV was used for all plots. The location of the Fermi level is 0.99 eV.

peaks at all and EKT1-AFQMC only introduces a shift to the HF spectrum. Correlation effects in the peak height do not appear large, but the peak location changes by around 1 eV going from HF to EKT1-AFQMC. We also produced Δ AFQMC and EOM-IP-CCSD for comparison. The Δ AFQMC IP is 2.18(1) eV, EOM-IP-CCSD yields 1.91 eV, and EKT1-AFQMC gives 2.51 eV in this basis set. The deviation of EKT1-AFQMC from

Δ AFQMC is 0.33(1) eV, whereas EOM-IP-CCSD deviates by 0.27(1) eV. These deviations are similar in magnitude but with opposite signs. The accuracy of EOM-IP-CCSD is unclear because the ground-state energy found from CCSD is higher than that of AFQMC by 0.0291(1) eV per electron. In the basis set limit, AFQMC was found to be as accurate as state-of-the-art diffusion Monte Carlo for the ground-state energy, differing by 0.0088(9) eV per electron or less.⁵⁷ This suggests that the ground-state correlation energy of CCSD may be on the order of 0.03 eV per electron. How much of this error is propagated to the EOM-IP-CCSD calculation remains unclear. In the complete basis set limit, we believe that the first IP from EKT1-AFQMC will become more accurate and closer to that of Δ AFQMC. A more complete comparison should be conducted in this limit, but such calculations are very difficult due to the need to procure many BP samples in EKT1-AFQMC.

4.3. Toward *Ab Initio* Solids: Minimal Unit Cell Diamond at the Γ -Point. With recent advances in open-source software such as PySCF,⁸⁵ performing calculations on *ab initio* solids is relatively straightforward. While an implementation of AFQMC with k -points has been previously presented,^{60,104} we only present a Γ -point result in this work. This is mainly because our current EKT1 implementation does not explicitly consider k -points. We chose to study diamond because it is one of the simplest solids just with two carbon atoms in its unit cell. We used the GTH-PADE pseudopotential¹⁰⁵ and the GTH-DZVP basis set.¹⁰⁶ We do not converge our calculation to the thermodynamic limit and consider only the smallest unit cell (two atom cell) calculations. Therefore, one cannot draw any physical conclusions from these calculations. Nonetheless, we present this as a part of our initial benchmark study. More careful assessment and extrapolation to the thermodynamic and basis limits are left for future study.

This system is overall as small as the smaller systems considered in this work. Therefore, we could obtain over 6000 BP samples with a 4 a.u. BP time. Even with these many samples, the largest error bar in the Fock and 1-RDM matrices was 6×10^{-3} . Due to this large statistical error in the matrix elements, we used a linear dependency cutoff of 0.01 and symmetrized the Fock matrix by taking only the upper triangular part of it (see Appendix B for details). We also included the shift by the Madelung constant in the spectral functions.

As shown in Figure 5, EKT1-AFQMC successfully reproduces EKT1-HCI. A small quasiparticle peak near -5 eV is not accurately captured by EKT1-AFQMC largely due to statistical

noise. Small peaks are difficult to resolve in EKT1-AFQMC without reducing the statistical errors on each matrix element further. Nonetheless, two other larger quasiparticle peaks are well represented. Both peaks are well reproduced within 0.25 eV from EKT1-HCI. The improvement over HF is around 1 eV or so and the first IP from EOM-IP-CCSD is within 0.02 eV of the EKT1-HCI result. It will be interesting to revisit the assessment of EKT1-AFQMC in the basis set and thermodynamic limits via direct comparisons to experiments.

5. CONCLUSIONS

In this work, we have explored the EKT approach to the computation of spectral functions via phaseless AFQMC. Previous attempts in AFQMC to obtaining spectral functions have resorted to analytic continuation,^{25,107} which has well-documented drawbacks. The EKT approach is attractive because it requires neither an explicit representation of the ground state wave function nor analytic continuation to compute spectral functions. Instead, its only inputs are N -particle RDMs which can be computed in AFQMC via the BP algorithm. The motivation of our work was thus to use the EKT approach with the aim of avoiding numerical problems arising in analytic continuation for the accurate assessment of real-frequency spectral information. While many studies have so far focused on the simplest level of the EKT, the EKT approach is systematically improvable with an increasing order of excitations: 1h, 1p2h, and so forth for electron removal and 1p, 1h2p, and so forth for electron addition. We presented the implementation of EKT1 (1h or 1p) and EKT3 (1p2h or 1h2p). For EKT3, we proposed the use of a cumulant approximation to the 4-RDM to avoid the steep storage requirements.

We produced preliminary results using EKT1 within AFQMC (EKT1-AFQMC) for small molecular systems, the UEG modeled by 14-electron and 54-electron supercells, and a minimal unit cell model of diamond at the Γ -point. We focused on studying the first IP and electron removal spectral functions of these systems. By comparing numerically exact EKT1 results based on heat-bath CI (i.e., EKT1-HCI), we showed that despite statistical noise, EKT1-AFQMC can capture most qualitative features of EKT1-HCI. We provide a more detailed summary on our findings as follows:

1. In small molecular benchmarks within the aug-cc-pVDZ basis, we found the maximum deviation of EKT1-AFQMC from EKT1-HCI in the first IP to be 0.19 eV. These molecules have quite small phaseless biases in the ground-state energy (≤ 0.04 eV), so we attributed additional biases to BP. Electron removal spectral functions from EKT1-AFQMC look qualitatively similar to that of EKT1-HCI even in the least accurate case (N_2).
2. For the 14-electron UEG supercell (19-plane-wave) benchmark, we observed a qualitative failure of EKT1 due to its inability to describe satellite states at $K = 0$. We showed that EKT3 (within HCI) significantly improves this. Despite these failures of EKT1, we found EKT1-AFQMC to have peak locations that are nearly identical (within 0.01 eV) to EKT1-HCI for both $r_s = 0.5$ and $r_s = 4$. Given the noticeable phaseless bias at $r_s = 4$, this result is quite encouraging. Lastly, for the valence region of the electron removal spectral function, we observed reasonable accuracy of EKT1 compared to the exact spectral function. The location of the first IP was off by 0.4 eV for r_s

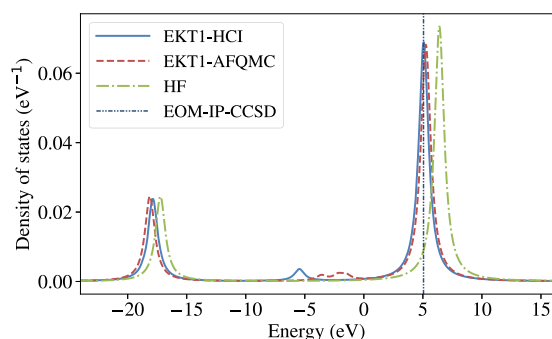


Figure 5. EKT1-HCI and EKT1-AFQMC electron removal spectral functions of a minimal unit cell model of diamond at the Γ -point. A broadening parameter $\eta = 0.5$ eV was used. The location of the Fermi level is 8.47 eV.

- = 0.5 and 0.25 eV for $r_s = 4.0$, which we expect to improve in larger bases.
- For the 54-electron UEG supercell (257-plane-wave) benchmark, we could not obtain EKT1-HCI due to computational expense. Therefore, we attempted to assess the accuracy of EKT1-AFQMC by comparing the first IP of EKT1-AFQMC with that of EOM-IP-CCSD and Δ -AFQMC. However, all three methods differ from each other by more than 0.25 eV and a more thorough benchmark in the basis set limit is highly desirable.
 - For a minimal unit cell model of diamond at the Γ -point, EKT1-AFQMC produced a qualitatively correct electron removal spectral function which agrees well with EKT1-HCI. However, EKT1-AFQMC peak locations were off by 0.2 eV from those of EKT1-HCI. We also noted that EKT1-HCI first IP agrees with that of EOM-IP-CCSD within 0.01 eV.

While a more extensive benchmark study is highly desirable, we cautiously conclude that EKT1-AFQMC is useful for charge excitations that are heavily dominated by Koopmans-like excitations. EKT1-AFQMC errors in peak locations can be as large as 0.25 eV compared to EKT1-HCI, but the line shapes of EKT1-AFQMC closely follow those of EKT1-HCI in all systems considered in this work.

The greatest challenge of EKT1-AFQMC is currently the statistical inefficiency in obtaining relevant back-propagated quantities with error bars small enough to enable the construction of stable EKT1-AFQMC spectral functions. Future work must first be dedicated to improving the statistical efficiency of BP. Furthermore, better BP algorithms are needed to reduce the BP error further. A practical implementation of EKT3-AFQMC using an iterative eigenvalue solver will be an interesting topic to explore in the future. Several interesting extensions are immediately possible. First, extending the EKT framework to neutral excitations^{68,70} is relatively straightforward and could be interesting to explore. Next, the extension of the EKT framework for finite-temperature coupled electron-phonon problems would provide a way to compute temperature-dependent vibronic spectra directly from AFQMC.^{101,108} We also leave the comparison of these EKT-based spectral functions to analytically continued spectral functions for a future study.

■ APPENDIX A

Review of ph-AFQMC

While the ph-AFQMC formalism has been presented before in detail,⁷⁶ we review the essence of the algorithm to provide a self-contained description. The imaginary propagation is given as

$$|\Psi_0\rangle \propto \lim_{\tau \rightarrow \infty} \exp(-\tau \hat{H}) |\Phi_0\rangle = \lim_{\tau \rightarrow \infty} |\Psi(\tau)\rangle \quad (59)$$

where τ is the imaginary time, $|\Psi_0\rangle$ is the exact ground state of a Hamiltonian $\hat{H}$, and $|\Phi_0\rangle$ is an initial starting wave function with a nonzero overlap with $|\Psi_0\rangle$. We assume no special structure in the underlying Hamiltonian and work with the generic *ab initio* Hamiltonians of eq 17.

In ph-AFQMC, this imaginary-time propagation is stochastically implemented. One discretizes the imaginary time τ with a time step of $\Delta\tau$ such that for N time steps, we have $\tau = N\Delta\tau$. Using the Trotter approximation and the Hubbard–Stratonovich transformation,^{109,110} a single time step many-body propagator can be written in an integral form

$$\exp(-\Delta\tau \hat{H}) = \int d^X \mathbf{x} p(\mathbf{x}) \hat{B}(\Delta\tau, \mathbf{x}) \quad (60)$$

where $p(\mathbf{x})$ is the standard normal distribution, $\mathbf{x}$ is a vector of X auxiliary fields, and $\hat{B}$ is defined as

$$\hat{B}(\Delta\tau, \mathbf{x}) = e^{-\Delta\tau/2 \hat{H}_1} e^{-\sqrt{\Delta\tau} \mathbf{x} \hat{v}} e^{-\Delta\tau/2 \hat{H}_1} + O(\Delta\tau^3) \quad (61)$$

where $\hat{v}$ is defined from

$$\hat{H}_2 = -\frac{1}{2} \sum_{\alpha}^X \hat{v}_{\alpha}^2 \quad (62)$$

X is the dimension of Cholesky vectors from the Cholesky factorization¹¹¹ or the dimension of the predefined density fitting basis set,¹¹² which asymptotically grows linearly with system size for a fixed error per particle in the final energy. The computation of the integral in eq 60 is carried out via Monte Carlo sampling where each walker samples an instance of $\mathbf{x}$.

The global wave function is, with importance sampling, represented as a linear combination of walker wave functions

$$|\Psi(\tau)\rangle = \sum_i w_i(\tau) \frac{|\psi_i(\tau)\rangle}{\langle \Psi_T | \psi_i(\tau) \rangle} \quad (63)$$

where w_i is the weight of the i -th walker, $|\psi_i(\tau)\rangle$ is the single Slater determinant of the i -th walker, and $|\Psi_T\rangle$ is the trial wave function. At each time step, each walker samples a set of $\mathbf{x}$, forms $\hat{B}(\Delta\tau, \mathbf{x})$, and updates its wave function by applying $\hat{B}(\Delta\tau, \mathbf{x})$ to it. Practical implementations employ the so-called “optimal” force bias which shifts the Gaussian distribution⁵⁰

$$\bar{\mathbf{x}}_i(\Delta\tau, \tau) = -\sqrt{\Delta\tau} \frac{\langle \Psi_T | \hat{v} | \psi_i(\tau) \rangle}{\langle \Psi_T | \psi_i(\tau) \rangle} \quad (64)$$

With the optimal force bias, a single time step propagation can be summarized with two equations

$$w_i(\tau + \Delta\tau) = I_{\text{ph}}(\mathbf{x}_i, \bar{\mathbf{x}}_i, \tau, \Delta\tau) \times w_i(\tau) \quad (65)$$

$$|\psi_i(\tau + \Delta\tau)\rangle = \hat{B}(\Delta\tau, \mathbf{x}_i - \bar{\mathbf{x}}_i) |\psi_i(\tau)\rangle \quad (66)$$

where the phaseless importance function in hybrid form is defined as

$$I_{\text{ph}}(\mathbf{x}_i, \bar{\mathbf{x}}_i, \tau, \Delta\tau) = |I(\mathbf{x}_i, \bar{\mathbf{x}}_i, \tau, \Delta\tau)| \times \max(0, \cos(\theta_i(\tau))) \quad (67)$$

with

$$I(\mathbf{x}_i, \bar{\mathbf{x}}_i, \tau, \Delta\tau) = S_i(\tau, \Delta\tau) e^{\mathbf{x}_i \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i / 2} \quad (68)$$

and

$$S_i(\tau, \Delta\tau) = \frac{\langle \Psi_T | \hat{B}(\Delta\tau, \mathbf{x}_i - \bar{\mathbf{x}}_i) | \psi_i(\tau) \rangle}{\langle \Psi_T | \psi_i(\tau) \rangle} \quad (69)$$

With this specific walker update instruction, all walker weights in eq 63 remain real and positive and thereby, it completely eliminates the fermionic phase problem.

■ APPENDIX B

Numerical Details of EKT

Determining the eigenvalues and eigenvectors of eq 11 from noisy QMC density matrices is nontrivial. We first diagonalize the metric matrices $S_{\pm}$

$$S_{\pm} = \mathbf{U} \Lambda_{\pm} \mathbf{U}^{\dagger} \quad (70)$$

and discard eigenvalues below a given threshold. In this work, unless noted otherwise, in the main text, we used 10^{-4} for all EKT1-AFQMC results, which is on the order of the largest statistical error in the EKT1 Fock matrix. We next construct the transformation matrix

$$\mathbf{X}_{\pm} = \mathbf{U}\mathbf{\Lambda}_{\pm}^{-1/2} \quad (71)$$

This procedure is often referred to as *canonical* orthogonalization in quantum chemistry. Then, we solve

$$\tilde{\mathbf{F}}_{\pm} \tilde{\mathbf{c}}_{\pm}^{\nu} = \epsilon_{\pm}^{\nu} \tilde{\mathbf{c}}_{\pm}^{\nu} \quad (72)$$

where

$$\tilde{\mathbf{F}}_{\pm} = \mathbf{X}_{\pm}^{\dagger} \mathbf{F}_{\pm} \mathbf{X}_{\pm} \quad (73)$$

Finally, the eigenvectors in the original basis can be determined from

$$(\mathbf{c}_{\pm}^{\nu})_p = \sum_I (\mathbf{X}_{\pm})_{pI} (\tilde{\mathbf{c}}_{\pm}^{\nu})_I \quad (74)$$

Following Kent et al.,⁴⁵ we explicitly zero out all matrix elements whose magnitude is smaller than two times the corresponding statistical error bar. We explicitly symmetrize RDMs but leave the Fock matrix asymmetric as required for approximate wave functions. However, for the more difficult problems considered in this work, such as the 54-electron UEG electron and diamond, we found that symmetrizing the Fock matrix is useful, so we choose to symmetrize the Fock matrix in such cases (by taking only the upper triangular part of the Fock matrix). These steps improved the numerical stability of the eigenvalue problem.

We also note that there are generic numerical issues arising in EKT even without any statistical sampling error. These numerical issues are present even with numerically exact RDMs. This was observed in both EKT1-HCI and EKT3-HCI, where spurious solutions with large negative IPs appear. These states stem from the fact that the metric matrix in EKT problems are generally low-rank as they are related to RDMs. For instance, in the EKT1 formulation, the metric we diagonalize for the IP problem is a 1-RDM, whose rank is not so much larger than the number of electrons in the system. These spurious states can be removed with larger cutoffs while often affecting peak locations of quasiparticle states. Interestingly, these spurious states all carry negligible spectral weights and do not appear in spectra. Motivated by this observation, the most satisfying solution we found was to use as a small threshold as possible and to measure the overlap between Koopmans states and eigenvectors to identify quasiparticle-like eigenvectors. This was enough to identify physical IP excitations that are quasiparticle-like. The same principle is applicable to EKT3-HCI and also the EA calculations.

Spectral functions are plotted by approximating the δ -function in the spectral function expression as a Lorentzian function

$$\delta(\omega) \simeq \frac{1}{\pi} \left(\frac{\eta}{\omega^2 + \eta^2} \right) \quad (75)$$

for some small constant η . To ensure reproducibility, we specified the value of η in all relevant figures.

■ APPENDIX C

Spin-Explicit UEG Hamiltonian

Here, we provide the spin-explicit form of the UEG Hamiltonian. The one-body Hamiltonian reads

$$\hat{\mathcal{H}}_1 = \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{\mathbf{K}} \frac{|\mathbf{K}|^2}{2} a_{\mathbf{K}, \sigma}^{\dagger} a_{\mathbf{K}, \sigma} \quad (76)$$

while the electron–electron interaction operator is

$$\hat{\mathcal{H}}_2 = \frac{1}{2\Omega} \sum_{\sigma, \sigma' \in \{\uparrow, \downarrow\}} \sum_{\mathbf{K} \neq 0, \mathbf{K}_1, \mathbf{K}_2} \frac{4\pi}{|\mathbf{K}|^2} \times a_{\mathbf{K}_1 + \mathbf{K}, \sigma}^{\dagger} a_{\mathbf{K}_2 - \mathbf{K}, \sigma'}^{\dagger} a_{\mathbf{K}_2, \sigma'} a_{\mathbf{K}_1, \sigma} \quad (77)$$

■ AUTHOR INFORMATION

Corresponding Authors

Joonho Lee – Department of Chemistry, Columbia University, New York 10027, United States; orcid.org/0000-0002-9667-1081; Email: jl5653@columbia.edu

Fionn D. Malone – Quantum Simulations Group, Lawrence Livermore National Laboratory, Livermore, California 94551, United States; orcid.org/0000-0001-9239-0162; Email: fionn.malone@gmail.com

Miguel A. Morales – Quantum Simulations Group, Lawrence Livermore National Laboratory, Livermore, California 94551, United States; Email: moralessilva2@llnl.gov

David R. Reichman – Department of Chemistry, Columbia University, New York 10027, United States; Email: drr2103@columbia.edu

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.1c00100>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Sandeep Sharma for providing access to a version of Dice with the 3-RDM and 4-RDM capability, which was used in testing and producing EKT3-HCI results presented in this work. We also thank Garnet Chan for discussion about finite-size corrections to ionization energies. D.R.R. acknowledges support of NSF CHE-1954791. The work of F.D.M. and M.A.M. was performed under the auspices of the U.S. Department of Energy (DOE) by LLNL under contract no. DE-AC52-07NA27344 and was supported by the U.S. DOE, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, as part of the Computational Materials Sciences Program and Center for Predictive Simulation of Functional Materials (CPSFM). Computing support for this work came from the LLNL Institutional Computing Grand Challenge program.

■ REFERENCES

- (1) Lu, D.; Vishik, I. M.; Yi, M.; Chen, Y.; Moore, R. G.; Shen, Z.-X. Angle-resolved photoemission studies of quantum materials. *Annu. Rev. Condens. Matter Phys.* **2012**, *3*, 129–167.
- (2) Egerton, R. F. Electron energy-loss spectroscopy in the TEM. *Rep. Prog. Phys.* **2008**, *72*, 016502.
- (3) Andreani, C.; Colognesi, D.; Mayers, J.; Reiter, G. F.; Senesi, R. Measurement of momentum distribution of lightatoms and molecules in condensed matter systems using inelastic neutron scattering. *Adv. Phys.* **2005**, *54*, 377–469.
- (4) Fetter, A. L.; Walecka, J. D. *Quantum Theory of Many-Particle Systems*, Dover Books on Physics; Dover Publications, 2003.

- (5) Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601.
- (6) Cederbaum, L. S.; Domcke, W. Theoretical aspects of ionization potentials and photoelectron spectroscopy: A Green's function approach. *Adv. Chem. Phys.* **1977**, *36*, 205–344.
- (7) Hedin, L.; Michiels, J.; Inglesfield, J. Transition from the adiabatic to the sudden limit in core-electron photoemission. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 15565.
- (8) Duchemin, I.; Blase, X. Robust Analytic-Continuation Approach to Many-Body GW Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 1742–1756.
- (9) Monkhorst, H. J. Calculation of properties with the coupled-cluster method. *Int. J. Quantum Chem.* **1977**, *12*, 421–432.
- (10) Nooijen, M.; Snijders, J. G. Coupled cluster approach to the single-particle Green's function. *Int. J. Quantum Chem.* **1992**, *44*, 55–83.
- (11) Nooijen, M.; Snijders, J. G. Coupled cluster Green's function method: Working equations and applications. *Int. J. Quantum Chem.* **1993**, *48*, 15–48.
- (12) Peng, B.; Kowalski, K. Coupled-cluster Green's function: Analysis of properties originating in the exponential parametrization of the ground-state wave function. *Phys. Rev. A* **2016**, *94*, 062512.
- (13) Furukawa, Y.; Kosugi, T.; Nishi, H.; Matsushita, Y.-i. Band structures in coupled-cluster singles-and-doubles Green's function (GFCCSD). *J. Chem. Phys.* **2018**, *148*, 204109.
- (14) Peng, B.; Kowalski, K. Properties of advanced coupled-cluster Green's function. *Mol. Phys.* **2018**, *116*, 561–569.
- (15) Shee, A.; Zgid, D. Coupled Cluster as an Impurity Solver for Green's Function Embedding Methods. *J. Chem. Theory Comput.* **2019**, *15*, 6010–6024.
- (16) McClain, J.; Lischner, J.; Watson, T.; Matthews, D. A.; Ronca, E.; Louie, S. G.; Berkelbach, T. C.; Chan, G. K.-L. Spectral functions of the uniform electron gas via coupled-cluster theory and comparison to the GW and related approximations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 235139.
- (17) 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.
- (18) Blankenbecler, R.; Sugar, R. L. Projector Monte Carlo method. *Phys. Rev. D* **1983**, *27*, 1304–1311.
- (19) Becca, F.; Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems*; Cambridge University Press: Cambridge, England, U.K., 2017.
- (20) Foulkes, W. M. C.; Mitars, L.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **2001**, *73*, 33.
- (21) Silver, R. N.; Sivia, D. S.; Gubernatis, J. E. Maximum-entropy method for analytic continuation of quantum Monte Carlo data. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1990**, *41*, 2380–2389.
- (22) Gubernatis, J. E.; Jarrell, M.; Silver, R. N.; Sivia, D. S. Quantum Monte Carlo simulations and maximum entropy: Dynamics from imaginary-time data. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 6011–6029.
- (23) Jarrell, M.; Gubernatis, J. E. Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data. *Phys. Rep.* **1996**, *269*, 133–195.
- (24) Motta, M.; Galli, D. E.; Moroni, S.; Vitali, E. Imaginary time correlations and the phaseless auxiliary field quantum Monte Carlo. *J. Chem. Phys.* **2014**, *140*, 024107.
- (25) Motta, M.; Galli, D. E.; Moroni, S.; Vitali, E. Imaginary time density-density correlations for two-dimensional electron gases at high density. *J. Chem. Phys.* **2015**, *143*, 164108.
- (26) Otsuki, J.; Ohzeki, M.; Shinaoka, H.; Yoshimi, K. Sparse modeling approach to analytical continuation of imaginary-time quantum Monte Carlo data. *Phys. Rev. E* **2017**, *95*, 061302.
- (27) Bertaina, G.; Galli, D. E.; Vitali, E. Statistical and computational intelligence approach to analytic continuation in Quantum Monte Carlo. *Adv. Phys. X* **2017**, *2*, 302–323.
- (28) Reichman, D. R.; Rabani, E. Analytic continuation average spectrum method for quantum liquids. *J. Chem. Phys.* **2009**, *131*, 054502.
- (29) Goulko, O.; Mishchenko, A. S.; Pollet, L.; Prokofev, N.; Svistunov, B. Numerical analytic continuation: Answers to well-posed questions. *Phys. Rev. B* **2016**, *95*, 014102.
- (30) Dornheim, T.; Groth, S.; Vorberger, J.; Bonitz, M. Ab initio Path Integral Monte Carlo Results for the Dynamic Structure Factor of Correlated Electrons: From the Electron Liquid to Warm Dense Matter. *Phys. Rev. Lett.* **2018**, *121*, 255001.
- (31) Day, O. W.; Smith, D. W.; Morrison, R. C. Extension of Koopmans' theorem. II. Accurate ionization energies from correlated wavefunctions for closed-shell atoms. *J. Chem. Phys.* **1975**, *62*, 115–119.
- (32) Smith, D. W.; Day, O. W. Extension of Koopmans' theorem. I. Derivation. *J. Chem. Phys.* **1975**, *62*, 113–114.
- (33) Morrell, M. M.; Parr, R. G.; Levy, M. Calculation of ionization potentials from density matrices and natural functions, and the long-range behavior of natural orbitals and electron density. *J. Chem. Phys.* **1975**, *62*, 549–554.
- (34) Farnum, J. D.; Mazziotti, D. A. Extraction of ionization energies from the ground-state two-particle reduced density matrix. *Chem. Phys. Lett.* **2004**, *400*, 90–93.
- (35) Pernal, K.; Cioslowski, J. Ionization potentials from the extended Koopmans' theorem applied to density matrix functional theory. *Chem. Phys. Lett.* **2005**, *412*, 71–75.
- (36) Piris, M.; Matxain, J. M.; Lopez, X.; Ugalde, J. M. The extended Koopmans' theorem: Vertical ionization potentials from natural orbital functional theory. *J. Chem. Phys.* **2012**, *136*, 174116.
- (37) Piris, M.; Matxain, J. M.; Lopez, X.; Ugalde, J. M. *8th Congress on Electronic Structure: Principles and Applications (ESPA 2012): A Conference Selection from Theoretical Chemistry Accounts*; Springer: Berlin, Germany, 2013; pp 5–15.
- (38) Welden, A. R.; Phillips, J. J.; Zgid, D. Ionization potentials and electron affinities from the extended Koopmans' theorem in self-consistent Green's function theory. **2015**, arXiv:1505.05575.
- (39) Morrison, R. C. The extended Koopmans' theorem and its exactness. *J. Chem. Phys.* **1992**, *96*, 3718–3722.
- (40) Cioslowski, J.; Piskorz, P.; Liu, G. Ionization potentials and electron affinities from the extended Koopmans' theorem applied to energy-derivative density matrices: The EKTMPn and EKTQCISD methods. *J. Chem. Phys.* **1997**, *107*, 6804–6811.
- (41) Bozkaya, U. The extended Koopmans' theorem for orbital-optimized methods: Accurate computation of ionization potentials. *J. Chem. Phys.* **2013**, *139*, 154105.
- (42) Bozkaya, U. Accurate Electron Affinities from the Extended Koopmans' Theorem Based on Orbital-Optimized Methods. *J. Chem. Theory Comput.* **2014**, *10*, 2041–2048.
- (43) Yildiz, D.; Bozkaya, U. Assessment of the extended Koopmans' theorem for the chemical reactivity: Accurate computations of chemical potentials, chemical hardnesses, and electrophilicity indices. *J. Comput. Chem.* **2016**, *37*, 345–353.
- (44) Bozkaya, U.; Ünal, A. State-of-the-Art Computations of Vertical Ionization Potentials with the Extended Koopmans' Theorem Integrated with the CCSD(T) Method. *J. Phys. Chem. A* **2018**, *122*, 4375–4380.
- (45) Kent, P. R. C.; Hood, R. Q.; Towler, M. D.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo calculations of the one-body density matrix and excitation energies of silicon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *57*, 15293–15302.
- (46) Zheng, H. First principles quantum Monte Carlo study of correlated electronic systems. 2016, <https://www.ideals.illinois.edu/handle/2142/92794> [Online accessed 12 Jan 2021].
- (47) Yoshioka, N.; Mizukami, W.; Nori, F. Neural-Network Quantum States for the Electronic Structure of Real Solids. **2020**, arXiv:2008.09492.
- (48) Zhang, S.; Carlson, J.; Gubernatis, J. E. Constrained Path Quantum Monte Carlo Method for Fermion Ground States. *Phys. Rev. Lett.* **1995**, *74*, 3652–3655.

- (49) Zhang, S.; Carlson, J.; Gubernatis, J. E. Constrained path Monte Carlo method for fermion ground states. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *55*, 7464–7477.
- (50) Zhang, S.; Krakauer, H. Quantum Monte Carlo method using phase-free random walks with Slater determinants. *Phys. Rev. Lett.* **2003**, *90*, 136401.
- (51) Carlson, J.; Gubernatis, J. E.; Ortiz, G.; Zhang, S. Issues and observations on applications of the constrained-path Monte Carlo method to many-fermion systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 12788.
- (52) LeBlanc, J. P.; Antipov, A. E.; Becca, F.; Bulik, I. W.; Chan, G. K.-L.; Chung, C.-M.; Deng, Y.; Ferrero, M.; Henderson, T. M.; Jiménez-Hoyos, C. A.; et al. Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms. *Adv. Phys.: X* **2015**, *5*, 041041.
- (53) Zheng, B.-X.; Chung, C.-M.; Corboz, P.; Ehlers, G.; Qin, M.-P.; Noack, R. M.; Shi, H.; White, S. R.; Zhang, S.; Chan, G. K.-L. Stripe order in the underdoped region of the two-dimensional Hubbard model. *Science* **2017**, *358*, 1155–1160.
- (54) Motta, M.; Ceperley, D. M.; Chan, G. K.-L.; Gomez, J. A.; Gull, E.; Guo, S.; Jiménez-Hoyos, C. A.; Lan, T. N.; Li, J.; Ma, F.; 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. *Adv. Phys.: X* **2017**, *7*, 031059.
- (55) Zhang, S.; Malone, F. D.; Morales, M. A. Auxiliary-field quantum Monte Carlo calculations of the structural properties of nickel oxide. *J. Chem. Phys.* **2018**, *149*, 164102.
- (56) Motta, M.; Genovese, C.; Ma, F.; Cui, Z.-H.; Sawaya, R.; Chan, G. K.-L.; Chepiga, N.; Helms, P.; Jiménez-Hoyos, C.; Millis, A. J.; et al. Ground-state properties of the hydrogen chain: dimerization, insulator-to-metal transition, and magnetic phases. *Adv. Phys.: X* **2020**, *10*, 031058.
- (57) Lee, J.; Malone, F. D.; Morales, M. A. An auxiliary-field quantum Monte Carlo perspective on the ground state of the dense uniform electron gas: An investigation with Hartree-Fock trial wavefunctions. *J. Chem. Phys.* **2019**, *151*, 064122.
- (58) Lee, J.; Malone, F. D.; Reichman, D. R. The performance of phaseless auxiliary-field quantum Monte Carlo on the ground state electronic energy of benzene. *J. Chem. Phys.* **2020**, *153*, 126101.
- (59) Williams, K. T.; Yao, Y.; Li, J.; Chen, L.; Shi, H.; Motta, M.; Niu, C.; Ray, U.; Guo, S.; Anderson, R. J.; et al. Direct comparison of many-body methods for realistic electronic Hamiltonians. *Adv. Phys.: X* **2020**, *10*, 011041.
- (60) Malone, F. D.; Zhang, S.; Morales, M. A. Accelerating Auxiliary-Field Quantum Monte Carlo Simulations of Solids with Graphical Processing Units. *J. Chem. Theory Comput.* **2020**, *16*, 4286–4297.
- (61) Lee, J.; Malone, F. D.; Morales, M. A. Utilizing Essential Symmetry Breaking in Auxiliary-Field Quantum Monte Carlo: Application to the Spin Gaps of the C₃₆ Fullerene and an Iron Porphyrin Model Complex. *J. Chem. Theory Comput.* **2020**, *16*, 3019–3027.
- (62) Qin, M.; Chung, C.-M.; Shi, H.; Vitali, E.; Hubig, C.; Schollwöck, U.; White, S. R.; Zhang, S.; et al. Absence of superconductivity in the pure two-dimensional Hubbard model. *Adv. Phys.: X* **2020**, *10*, 031016.
- (63) Malone, F. D.; Benali, A.; Morales, M. A.; Caffarel, M.; Kent, P. R. C.; Shulenburger, L. Systematic comparison and cross-validation of fixed-node diffusion Monte Carlo and phaseless auxiliary-field quantum Monte Carlo in solids. *Phys. Rev. B* **2020**, *102*, 161104.
- (64) 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.
- (65) Lee, J.; Reichman, D. R. Stochastic Resolution-of-the-Identity Auxiliary-Field Quantum Monte Carlo: Scaling Reduction without Overhead. *J. Chem. Phys.* **2020**, *153*, 044131.
- (66) Purwanto, W.; Zhang, S. Quantum Monte Carlo method for the ground state of many-boson systems. *Phys. Rev. E* **2004**, *70*, 056702.
- (67) 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.
- (68) Pavlyukh, Y. Excited electronic states from a generalization of the extended Koopmans theorem. *Phys. Rev. A* **2018**, *98*, 052508.
- (69) Hedin, L. Effects of recoil on shake-up spectra in metals. *Phys. Scr.* **1980**, *21*, 477.
- (70) Pavlyukh, Y. The Ubiquitous Extended Koopmans' Theorem. *Phys. Status Solidi B* **2019**, *256*, 1800591.
- (71) See <https://github.com/msaitow/SecondQuantizationAlgebra> for details on how to obtain the source code (accessed 2021-01-20).
- (72) Neuscamman, E.; Yanai, T.; Chan, G. K.-L. Quadratic canonical transformation theory and higher order density matrices. *J. Chem. Phys.* **2009**, *130*, 124102.
- (73) Saitow, M.; Kurashige, Y.; Yanai, T. Multireference configuration interaction theory using cumulant reconstruction with internal contraction of density matrix renormalization group wave function. *J. Chem. Phys.* **2013**, *139*, 044118.
- (74) Zgid, D.; Ghosh, D.; Neuscamman, E.; Chan, G. K.-L. A study of cumulant approximations to n-electron valence multireference perturbation theory. *J. Chem. Phys.* **2009**, *130*, 194107.
- (75) Shavitt, I.; Bartlett, R. J. *Many-body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*; Cambridge University Press, 2009; Chapter 7.3, pp 182–184.
- (76) Motta, M.; Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1364.
- (77) Motta, M. Private Communication.
- (78) Chen, S.; Motta, M.; Ma, F.; Zhang, S. Ab initio electronic density in solids by many-body plane-wave auxiliary-field quantum Monte Carlo calculations. **2020**, arXiv:2011.08335, arXiv preprint.
- (79) 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.
- (80) Parrish, R. M.; Hohenstein, E. G.; Martínez, T. J.; Sherrill, C. D. Tensor hypercontraction. II. Least-squares renormalization. *J. Chem. Phys.* **2012**, *137*, 224106.
- (81) Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Martínez, T. J. Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions. *J. Chem. Phys.* **2012**, *137*, 221101.
- (82) Lee, J.; Lin, L.; Head-Gordon, M. Systematically Improvable Tensor Hypercontraction: Interpolative Separable Density-Fitting for Molecules Applied to Exact Exchange, Second- and Third-Order Møller-Plesset Perturbation Theory. *J. Chem. Theory Comput.* **2019**, *16*, 243–263.
- (83) Schoof, T.; Groth, S.; Vorberger, J.; Bonitz, M. Ab Initio Thermodynamic Results for the Degenerate Electron Gas at Finite Temperature. *Phys. Rev. Lett.* **2015**, *115*, 130402.
- (84) Yang, Y.; Gorelov, V.; Pierleoni, C.; Ceperley, D. M.; Holzmann, M. Electronic band gaps from quantum Monte Carlo methods. *Phys. Rev. B* **2020**, *101*, 085115.
- (85) 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. P y SCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1340.
- (86) Kent, P. R. C.; Annaberdiyev, A.; Benali, A.; Bennett, M. C.; Landinez Borda, E. J.; Doak, P.; Hao, H.; Jordan, K. D.; Krogel, J. T.; Kylänpää, I.; Lee, J.; Luo, Y.; Malone, F. D.; Melton, C. A.; Mitas, L.; Morales, M. A.; Neuscamman, E.; Reboredo, F. A.; Rubenstein, B.; Saritas, K.; Upadhyay, S.; Wang, G.; Zhang, S.; Zhao, L. QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo. *J. Chem. Phys.* **2020**, *152*, 174105.
- (87) See <https://github.com/pauxy-qmc/pauxy> for details on how to obtain the source code (accessed 2021-01-20).
- (88) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration

Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.

(89) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.

(90) Smith, J. E. T.; Mussard, B.; Holmes, A. A.; Sharma, S. Cheap and Near Exact CASSCF with Large Active Spaces. *J. Chem. Theory Comput.* **2017**, *13*, 5468–5478.

(91) Wagner, L. K.; Bajdich, M.; Mitás, L. QWalk: A quantum Monte Carlo program for electronic structure. *J. Comput. Phys.* **2009**, *228*, 3390.

(92) Haydock, R.; Heine, V.; Kelly, M. J. Electronic structure based on the local atomic environment for tight-binding bands. II. *J. Phys. C: Solid State Phys.* **1975**, *8*, 2591–2605.

(93) Dagotto, E. Correlated electrons in high-temperature superconductors. *Rev. Mod. Phys.* **1994**, *66*, 763–840.

(94) van Setten, M. J.; Caruso, F.; Sharifzadeh, S.; Ren, X.; Scheffler, M.; Liu, F.; Lischner, J.; Lin, L.; Deslippe, J. R.; Louie, S. G.; Yang, C.; Weigend, F.; Neaton, J. B.; Evers, F.; Rinke, P. GW100: Benchmarking G0W0 for Molecular Systems. *J. Chem. Theory Comput.* **2015**, *11*, 5665–5687.

(95) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(96) Sundholm, D.; Olsen, J. The exactness of the extended Koopmans' theorem: A numerical study. *J. Chem. Phys.* **1993**, *98*, 3999–4002.

(97) Ernzerhof, M. Validity of the Extended Koopmans' Theorem. *J. Chem. Theory Comput.* **2009**, *5*, 793–797.

(98) Vanfleteren, D.; Van Neck, D.; Ayers, P. W.; Morrison, R. C.; Bultinck, P. Exact ionization potentials from wavefunction asymptotics: The extended Koopmans' theorem, revisited. *J. Chem. Phys.* **2009**, *130*, 194104.

(99) Ceperley, D. M.; Bernu, B. The calculation of excited state properties with quantum Monte Carlo. *J. Chem. Phys.* **1988**, *89*, 6316–6328.

(100) Blunt, N. S.; Alavi, A.; Booth, G. H. Nonlinear biases, stochastically sampled effective Hamiltonians, and spectral functions in quantum Monte Carlo methods. *Phys. Rev. B* **2018**, *98*, 085118.

(101) Lee, J.; Morales, M. A.; Malone, F. D. A Phaseless Auxiliary-Field Quantum Monte Carlo Perspective on the Uniform Electron Gas at Finite Temperatures: Issues, Observations, and Benchmark Study. **2020**, arXiv:2012.12228.

(102) von Barth, U.; Holm, B. Self-consistent GW results for the electron gas: Fixed screened potential within the random-phase approximation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 8411.

(103) Holm, B.; von Barth, U. Fully self-consistent GW self-energy of the electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *57*, 2108.

(104) Motta, M.; Zhang, S.; Chan, G. K.-L. Hamiltonian symmetries in auxiliary-field quantum Monte Carlo calculations for electronic structure. *Phys. Rev. B* **2019**, *100*, 045127.

(105) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703–1710.

(106) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.

(107) Vitali, E.; Shi, H.; Qin, M.; Zhang, S. Computation of dynamical correlation functions for many-fermion systems with auxiliary-field quantum Monte Carlo. *Phys. Rev. B* **2016**, *94*, 085140.

(108) Lee, J.; Zhang, S.; Reichman, D. R. Constrained-Path Auxiliary-Field Quantum Monte Carlo for Coupled Electrons and Phonons. **2020**, arXiv:2012.13473v1.

(109) Hubbard, J. Calculation of Partition Functions. *Phys. Rev. Lett.* **1959**, *3*, 77–78.

(110) Hirsch, J. E. Discrete Hubbard-Stratonovich transformation for fermion lattice models. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *28*, 4059–4061.

(111) 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.

(112) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496–4501.