



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Ab initio electronic density in solids by many-body plane-wave auxiliary-field quantum Monte Carlo calculations

Siyuan Chen, Mario Motta, Fengjie Ma, and Shiwei Zhang

Phys. Rev. B **103**, 075138 — Published 22 February 2021

DOI: [10.1103/PhysRevB.103.075138](https://doi.org/10.1103/PhysRevB.103.075138)

Ab initio electronic density in solids by many-body plane-wave auxiliary-field quantum Monte Carlo calculations

Siyuan Chen* and Mario Motta†

Department of Physics, College of William & Mary, Williamsburg, VA 23185, USA

Fengjie Ma

*The Center for Advanced Quantum Studies and Department of Physics,
Beijing Normal University, Beijing 100875, China*

Shiwei Zhang

*Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA and
Department of Physics, College of William & Mary, Williamsburg, VA 23185, USA*

We present accurate many-body results of the electronic densities in several solid materials, including Si, NaCl, and Cu. These results are obtained using the *ab initio* auxiliary-field quantum Monte Carlo (AFQMC) method working in a plane-wave basis with norm-conserving, multiple-projector pseudopotentials. AFQMC has been shown to be an excellent many-body total energy method. Computation of observables and correlation functions other than the ground-state energy requires back-propagation, whose adaption and implementation in the plane-wave basis AFQMC framework are discussed in the present paper. This development allows us to compute correlation functions, electronic densities and interatomic forces, paving the way for geometry optimizations and calculations of thermodynamic properties in solids. Finite supercell size effects are considerably more subtle in the many-body framework than in independent-electron calculations. We analyze the convergence of the electronic density, and obtain best estimates for the thermodynamic limit. The densities from several typical density functionals are benchmarked against our near-exact results. The electronic densities we have obtained can also be used to help construct improved density functionals.

I. INTRODUCTION

The electronic density is one of the most fundamental physical quantities in materials. Various structural properties in solids directly depend on the density and its closely related quantities. More accurate results for the electronic density would thus lead to key improvement in our ability to reliably predict such physical properties. Furthermore, the widely applied density functional theory (DFT)[1–4] relies on density functionals to yield an independent-electron approximation of the many-body Hamiltonian. By iteratively solving the Schrödinger equation using such a density functional, the DFT method achieves high accuracy in many systems, while simultaneously maintaining a relatively low computational cost due to its one-body nature. As density functionals are designed based on electronic densities, accurate density input is also essential for methodological development in DFT [5, 6].

Approaching an exact electronic density using correlated many-body methods has remained a major challenge, despite the remarkable progress witnessed in the last decades. Exact methods like full-configuration interaction (FCI) require a computational cost scaling exponentially with system size, and are therefore usually re-

stricted to systems of a small number of active electrons and orbitals [7]. A promising route for very accurate calculations is with quantum Monte Carlo (QMC) methods, which represent the physical properties of a quantum system as multi-dimensional integrals that are in turn evaluated using random sampling. To date the most widely used QMC method for solid-state calculations has been the diffusion Monte Carlo (DMC) approach [8, 9], which reformulates the imaginary-time Schrödinger equation as a diffusion process in the set of real-space configurations (walkers, given by electronic coordinates). The diffusion process is guided by the kinetic energy, whereas the electron-electron interaction and the potential from nuclei are represented by weights.

Auxiliary-field quantum Monte Carlo (AFQMC) is a more recent QMC approach for real materials [10, 11] AFQMC is based on the second-quantization scheme, and uses imaginary-time evolution to propagate a trial wavefunction towards the desired ground state. The auxiliary-field formalism has its roots in the Hubbard-Stratonovich representation of the imaginary-time evolution [12], which has long been applied in lattice field-theoretic calculations [13–15]. A reformulation to cast the projection as open-ended random walks in Slater determinant space [10, 16] and the conceptual connection with mean-field projection [17] have provided the basis for coupling to standard electronic structure machinery, and positioned it as a natural post-DFT approach for real materials. The AFQMC can incorporate pseudopotentials in a simple manner that does not require

* schen24@email.wm.edu

† Present address: IBM Quantum, IBM Research Almaden, 650 Harry Rd, San Jose, CA 95120, USA

any additional approximation [18]. Formally it allows computation of properties other than the total energy straightforwardly via back-propagation [16, 19, 20]. As a low-polynomial scaling method, AFQMC has been shown in the latest benchmarks to maintain excellent accuracy across a broad range of application areas [21–25].

A set of orbitals (known as basis sets) are required in AFQMC to express the Slater determinants. Similar to most other electronic structure methods, the basis sets could be plane waves or local orbitals like Gaussians or Wannier functions [26, 27]. Since they can offer compact descriptions of electronic wavefunctions, local orbitals are widely used in real materials and especially in molecules. On the other hand, plane waves have several important advantages that make them compelling, especially in crystalline solids: (1) convergence to the infinite basis set limit requires only increasing a single parameter, the kinetic energy cutoff; (2) plane-waves are orthonormal, which is a safeguard against numerical instabilities seen for example in high-density systems [28]; (3) plane waves do not depend on the positions of atoms in a cell and can be made to have very small finite basis error, hence Pulay corrections are absent in force calculations [29]; (4) a large number of DFT calculations in extended systems are performed using plane waves, and plane-wave AFQMC can use the same computation framework as DFT (psuedopotentials, simple analytic evaluation of matrix elements, Fast Fourier Transform (FFT), etc).

Recent years of development and tests have indicated that plane-wave basis AFQMC (PW-AFQMC) is an excellent total energy method [18, 30, 31]. In order to calculate other physical quantities that do not commute with the Hamiltonian, a back-propagation technique is often necessary and needs to be incorporated in the algorithm [16, 19, 20]. Among these quantities are the electronic density and atomic forces, which have remained key challenges in extended systems of correlated electrons.

In this paper, we address this problem, by implementing back-propagation in PW-AFQMC and applying it to compute the electronic densities in several crystalline solids: the widely studied covalent bond semiconductor silicon; the insulating ionic crystal sodium chloride (NaCl); and the transitional metal copper. AFQMC calculations are carried out in different supercell sizes. The computed densities are extrapolated to the infinite supercell size or bulk limit using corrections from a finite-size DFT functional [32] and confirming numerical convergence with respect to the different supercell choices. The final results provide an accurate dataset of the true electronic densities for these systems. We compare these to the results from several of the widely used exchange-correlation functionals in DFT to benchmark their quality in reproducing the correct densities. We find that the accuracy of a functional in predicting the electronic density is not always correlated with its performance in computing total energies and structural properties.

The rest of this paper is organized as follows. In Section II, we describe our method, including a

brief overview of the AFQMC method, the plane-wave basis implementation, and the state-of-the-art back-propagation technique for observable computations. Section III introduces the systems we study and the computational details, including the finite-size correction and the details and the effect of the pseudopotential. Section IV presents the density results we have obtained, in diamond-structured Si, NaCl, and fcc-Cu. In Section V we describe our benchmark of several density functionals, by comparison in Sec. V A of their computed electronic densities with the AFQMC reference, followed by a discussion in Sec. V B of the relation between the accuracy of a functional in computing the density versus the total energies. We then conclude in Section VI.

II. METHODOLOGY

In this section we first provide a brief overview of the plane-wave AFQMC method. We specify the plane-wave Hamiltonian in detail, which illustrates the advantages of using a plane wave basis in solids. We then discuss the phaseless AFQMC method, which directly leads to mixed-estimator calculations for energy and properties. This is followed by an introduction of the path-restoration back-propagation technique formulated for plane-wave AFQMC, which is used to obtain higher-accuracy estimates of observables and correlation functions. This method is applied to obtain the charge densities.

A. Plane-wave Hamiltonian

Under the Born-Oppenheimer approximation [33–35], it is assumed that the ions move slowly compared to electrons. It is then possible to approximate the solid-state Hamiltonian as a sum of ionic and electronic parts,

$$H_{\text{tot}} = V_{\text{II}} + H, \quad (1)$$

where V_{II} includes the ionic energies. Unless specified explicitly, we will assume Rydberg atomic units throughout this paper, $\hbar = 2m_e = e^2/2 = 1$. Using the plane-wave basis and under periodic boundary conditions, the electronic Hamiltonian H can be written as a sum of following components:

$$H = K + V_{\text{psp}} + V_{\text{Ewald}} + V_{\text{Coulomb}}, \quad (2)$$

where the first term is the kinetic energy, and the second term V_{psp} is the nonlocal pseudopotential used to represent the electron-ion interaction. A kinetic energy cutoff $|\mathbf{G}|^2 < E_{\text{cut}}$ is imposed on the plane waves, limiting the total number of plane waves to a finite number N_{PW} .

The remaining terms describe the electron-electron interaction. The Ewald term V_{Ewald} is from self-interaction of an electron with its periodic images, and is constant for a given lattice [34–38]. The operator V_{Coulomb} is

the electron-electron Coulomb interaction and is given

$$V_{\text{Coulomb}} = \frac{1}{\Omega} \sum'_{ijkl} \frac{4\pi}{|\mathbf{G}_i - \mathbf{G}_k|^2} \delta_{\mathbf{G}_i - \mathbf{G}_k, \mathbf{G}_l - \mathbf{G}_j} \delta_{\sigma_i, \sigma_k} \delta_{\sigma_j, \sigma_l} c_i^\dagger c_j^\dagger c_l c_k,$$

where the prime in the sum indicates the exclusion of the $\mathbf{G}_i = \mathbf{G}_k$ singularity, σ is the spin of the electron occupying the basis, and latin letter subscripts such as i represent the (\mathbf{G}_i, σ_i) pair. This formula can be written as

$$V_{\text{Coulomb}} = \frac{1}{\Omega} \sum_{\mathbf{Q} \neq 0} \frac{4\pi}{|\mathbf{Q}|^2} \hat{\rho}^\dagger(\mathbf{Q}) \hat{\rho}(\mathbf{Q}) + V_{\text{Coulomb,1b}}, \quad (3)$$

where $V_{\text{Coulomb,1b}}$ is a one-body term from $\mathbf{G}_i = \mathbf{G}_l, \mathbf{G}_j = \mathbf{G}_k$. In the first term, we introduced the one-body density operator $\hat{\rho}(\mathbf{Q})$, which is defined as

$$\hat{\rho}(\mathbf{Q}) = \sum_{\mathbf{G}, \sigma} c_{\mathbf{G}+\mathbf{Q}, \sigma}^\dagger c_{\mathbf{G}, \sigma} \Theta(E_{\text{cut}} - |\mathbf{G} + \mathbf{Q}|^2), \quad (4)$$

where the step function ensures that $\mathbf{G} + \mathbf{Q}$ falls within the cutoff. This leaves a finite number $N_{\mathbf{Q}}$ of distinct $\hat{\rho}(\mathbf{Q})$'s, equal to the number of \mathbf{Q} -vectors ($\mathbf{Q} = \mathbf{G}_1 - \mathbf{G}_2$).

B. Auxiliary-field Quantum Monte Carlo (AFQMC)

The basic idea underlying AFQMC is imaginary time propagation [16]. It is easy to prove that an initial state $|\phi\rangle$ which has a nonzero overlap with the ground state $|\Psi_0\rangle$ propagates to the ground state with the following projection:

$$|\Psi_0\rangle \sim \lim_{\beta \rightarrow \infty} e^{-\beta H} |\phi\rangle = \lim_{N \rightarrow \infty} e^{-N \Delta\tau H} |\phi\rangle, \quad (5)$$

where we separate the total propagation time β into N time-steps of $\Delta\tau$, and perform the projection iteratively until convergence is reached. In fermion systems, usual choices of the wave functions are single or multiple Slater determinants. To turn this imaginary time evolution into a practical algorithm, we need to implement a single-step propagation $e^{-\Delta\tau H}$. For small enough $\Delta\tau$, we can split up the components in the Hamiltonian using Trotter-Suzuki decomposition:

$$e^{-\Delta\tau H} = e^{-\frac{\Delta\tau}{2} H_1} e^{-\Delta\tau H_2} e^{-\frac{\Delta\tau}{2} H_1} + O(\Delta\tau^3), \quad (6)$$

where H_1 and H_2 denote the one-body and two-body part of the Hamiltonian, respectively. The one-body propagation is a simple and computationally inexpensive operation to perform, as it maps a Slater determinant to

in second-quantization as:

another Slater determinant, according to Thouless' theorem [39, 40]. To propagate in imaginary-time under H_2 , we will make use of the particular structure in the plane-wave basis, in Eq. (3):

$$H_2 = \frac{1}{\Omega} \sum_{\mathbf{Q} \neq 0} \frac{4\pi}{|\mathbf{Q}|^2} \hat{\rho}^\dagger(\mathbf{Q}) \hat{\rho}(\mathbf{Q}). \quad (7)$$

Using the identity $\hat{\rho}(\mathbf{Q}) = \hat{\rho}^\dagger(-\mathbf{Q})$, we can write it into a sum of squares of one-body operators:

$$H_2 = \frac{1}{4} \sum_{\mathbf{Q} \neq 0} [\hat{\alpha}^2(\mathbf{Q}) + \hat{\beta}^2(\mathbf{Q})], \quad (8)$$

where the one-body Hermitian operators $\hat{\alpha}(\mathbf{Q}), \hat{\beta}(\mathbf{Q})$ are linear combinations of $\hat{\rho}(\mathbf{Q})$ and $\hat{\rho}^\dagger(\mathbf{Q})$ [18]. We then perform a continuous-variable Hubbard-Stratonovich transformation:

$$e^{-\frac{\Delta\tau}{2} \lambda \hat{v}^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-\frac{1}{2} x^2} e^{x \sqrt{-\Delta\tau} \lambda \hat{v}}, \quad (9)$$

where \hat{v} is $\hat{\alpha}(\mathbf{Q})$ or $\hat{\beta}(\mathbf{Q})$, and λ is a constant. This transforms the two-body propagator into an integral of one-body operators. It represents $e^{-\frac{\Delta\tau}{2} \lambda \hat{v}^2}$ as an expectation value over a random variable x , called an *auxiliary field*, with Gaussian probability distribution. For the operator H_2 in Eq. (8), $\mathbf{x} = \{x_{\mathbf{Q}}^{(\alpha)}, x_{\mathbf{Q}}^{(\beta)}\}$ is a set of auxiliary fields corresponding to the one-body operators, $\hat{\alpha}(\mathbf{Q})$ and $\hat{\beta}(\mathbf{Q})$, which commute and can be broken up and recombined in the exponential. (However, the behavior of the Trotter error in AFQMC in the presence of constraints is more subtle, as discussed in, e.g., Refs. [30, 41].) The expectation value in Eq. (9) can be evaluated by Monte Carlo techniques.

A simple scheme to propagate an initial state in imaginary time consists of setting a large number of Slater determinants $|\Phi_0^{(i)}\rangle$, called walkers, to the initial wave function $|\phi\rangle$. For each walker i and imaginary-time step k , an auxiliary field configuration $\mathbf{x}_k^{(i)}$ is sampled and the walker is projected in imaginary time as $|\Phi_{k+1}^{(i)}\rangle = \hat{B}(\mathbf{x}_k^{(i)}) |\Phi_k^{(i)}\rangle$, where $\hat{B}(\mathbf{x})$ denotes the product of one-body propagators [10].

This simple algorithm, however, suffers from several problems which we need to tackle individually. A uniform random sampling of the entire Hilbert space results in large statistical noise, and needs to be replaced by

importance sampling [10, 11, 42]. AFQMC uses a trial wavefunction $|\Psi_T\rangle$ to guide the stochastic sampling of walkers, and attaches a weight $w_k^{(i)}$ to each walker which is designed to be proportional to the overlap of the determinant $|\Phi_k^{(i)}\rangle$ with $|\Psi_T\rangle$. At each time step, the wave function is thus stochastically represented in the following form:

$$|\Phi_k\rangle \sim \sum_i w_k^{(i)} \frac{|\Phi_k^{(i)}\rangle}{\langle\Psi_T|\Phi_k^{(i)}\rangle}. \quad (10)$$

Walkers whose overlaps with $|\Psi_T\rangle$ have large magnitudes are deemed more important and are sampled more frequently, improving the efficiency. Note that the ‘‘importance function’’ defined above is complex; inclusion of the phase information is crucial for the accuracy of the phaseless approach [10, 19]. To realize the Monte Carlo sampling given in Eq. (10), a modified probability is used by introducing a shift $\bar{\mathbf{x}}$ in the auxiliary field, called *force bias*. The optimal choice of the force bias is the mixed estimator of the one-body operators, $\hat{\alpha}(\mathbf{Q})$ and $\hat{\beta}(\mathbf{Q})$ [19]. A local-energy is simultaneously introduced $E_L(\Psi) = \langle\Psi_T|H|\Psi\rangle/\langle\Psi_T|\Psi\rangle$. The final form of the weight factor is given by:

$$w_n^{(i)} = \prod_{k=0}^{n-1} I(\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}_k^{(i)}, \Phi_k^{(i)}), \quad (11)$$

with

$$I(\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}_k^{(i)}, \Phi_k^{(i)}) \simeq e^{\Delta\tau(E_T - [E_L]_k^{(i)})}, \quad (12)$$

where E_T is a trial energy chosen to approximate the exact energy and can be improved iteratively during the calculation. Below we will sometimes use $[E_L]_k^{(i)}$ as a shorthand for the local energy $E_L(\Psi_k^{(i)})$ at step k for walker i .

The formulation of the method up to this point is exact, but suffers from a phase problem, as weights of walkers are eventually randomly distributed into the entire complex plane and statistical noise would increase exponentially with the number of time steps and the system size. Propagation of the random walkers in the branching random walk framework, with no additional intervention, is referred to as a *free projection*. To control the phase problem, the solution is to make a *phaseless approximation* which introduces a small systematic bias, but reduces statistical fluctuations from exponential to polynomial. The method can be formally represented by rewriting Eq. (12) as

$$I(\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}_k^{(i)}, \Phi_k^{(i)}) \simeq e^{\Delta\tau(E_T - \text{Re}[E_L]_k^{(i)})} \times \max\{0, \cos[\Delta\theta]_k^{(i)}\} \quad (13)$$

with

$$[\Delta\theta]_k^{(i)} \equiv \text{Arg} \frac{\langle\Psi_T|\hat{B}(\mathbf{x} - \bar{\mathbf{x}})|\Phi_k^{(i)}\rangle}{\langle\Psi_T|\Phi_k^{(i)}\rangle}. \quad (14)$$

Equation (13) turns the weights into real and positive quantities, and contains their statistical fluctuations with the cosine factor. Walkers that are extremely close to the origin $\langle\Psi_T|\Phi_k^{(i)}\rangle = 0$ are the major contributors to the phase problem. These walkers move rapidly around the origin, resulting in larger $\Delta\theta$ in Eq. (14) and therefore a smaller cosine factor; this factor is zero if $\theta > \frac{\pi}{2}$, meaning the walker is eliminated.

C. Back-propagation and computation of charge density

The quantum mechanical expectation value of an observable \hat{O} over a state $|\Psi_0\rangle$ is

$$\langle\hat{O}\rangle = \frac{\langle\Psi_0|\hat{O}|\Psi_0\rangle}{\langle\Psi_0|\Psi_0\rangle}. \quad (15)$$

Since the ground state is an eigenfunction of the Hamiltonian, $H|\Psi_0\rangle = E|\Psi_0\rangle$, for $\hat{O} = \hat{H}$ the exact estimator coincides with the mixed estimator:

$$E = \langle H \rangle = \frac{\langle\Psi_T|H|\Psi_0\rangle}{\langle\Psi_T|\Psi_0\rangle}, \quad (16)$$

where only the ket state is propagated in imaginary time. In an AFQMC calculation, the energy at step k is thus estimated as

$$E = \frac{\sum_i w_k^{(i)} E_L(\Phi_k^{(i)})}{\sum_i w_k^{(i)}}. \quad (17)$$

For observables not commuting with the Hamiltonian, mixed estimators are biased, so the bra state has to be propagated in imaginary time as well. To achieve this goal, Zhang and coworkers [16, 19, 20] proposed a back-propagation (BP) method that rewrites the estimator as

$$\langle O \rangle \simeq \frac{\langle\Psi_T|e^{-m\Delta\tau H}\hat{O}e^{-n\Delta\tau H}|\Phi_0\rangle}{\langle\Psi_T|e^{-(m+n)\Delta\tau H}|\Phi_0\rangle}, \quad (18)$$

where n is the number of ‘‘forward propagation’’ steps and m is the number of ‘‘back-propagation’’ steps. The BP estimator reduces to the mixed estimator for $m = 0$, and gives higher-accuracy results for m sufficiently large. The backwards projection is performed on $|\Psi_{n,0}\rangle = |\Psi_T\rangle$ by applying adjoints of propagators in reverse order,

$$|\Psi_{n,m}\rangle = \hat{B}^\dagger(\mathbf{x} - \bar{\mathbf{x}})_n^{(i)} \hat{B}^\dagger(\mathbf{x} - \bar{\mathbf{x}})_{n+1}^{(i)} \dots \hat{B}^\dagger(\mathbf{x} - \bar{\mathbf{x}})_{n+m-1}^{(i)} |\Psi_T\rangle. \quad (19)$$

Note that we only need to store the sampled auxiliary-fields on the path in order to recover the propagators. The BP estimate of an observable is now given by

$$\langle O \rangle \simeq \frac{1}{\sum_i w_{n+m}^{(i)}} \sum_i w_{n+m}^{(i)} \frac{\langle\Psi_{n,m}^{(i)}|\hat{O}|\Phi_n^{(i)}\rangle}{\langle\Psi_{n,m}^{(i)}|\Phi_n^{(i)}\rangle}, \quad (20)$$

where the weights are evaluated at step $(n+m)$. In early AFQMC back-propagation practices, the forward- and

back-propagation are using the same technique, e.g. the *free-projection BP* involves use of free projection on both sides, while the *phaseless BP* applies phaseless approximation to both directions in the imaginary time. Phaseless BP stabilizes the algorithm and prevents the onset of the phase problem, but also introduces an unavoidable bias in the BP estimator. The phaseless bias is more severe for back-propagation than forward propagation, since the phaseless approximation breaks the symmetry between the two propagation directions and is optimal only for the forward direction, to which its imaginary-time dependence is aligned.

To mitigate the phaseless bias in back-propagation, a technique called *path-restoration* was recently proposed [20]. In the path-restoration technique, for each time the auxiliary field and force bias are recorded, we also record the discarded $\text{Im}[E_L]$ and the cosine factor $\cos(\Delta\theta)$ in Eq. (13). If this walker survives from step n to step $n+m$ (i.e. its weight remains non-zero), the information discarded due to the phaseless approximation can be restored when computing observables in Eq. (20), by multiplying the walker's weight with a restoring factor:

$$\omega_{n+m}^{(i)} \rightarrow \omega_{n+m}^{(i)} f_{n':n+m}^{(i)}, \quad (21)$$

where the path restoration factor is

$$f_{n':n+m}^{(i)} \equiv \prod_{k=n'}^{n+m} e^{-\Delta\tau \text{Im}[E_L]_k^{(i)}} \frac{1}{\cos([\Delta\theta]_k^{(i)})}. \quad (22)$$

The imaginary-time index n' in Eq. (21) is typically chosen as $n' = n$, which corresponds to path-restoration only on the back-propagation portion of the path [20]. We can also choose $n' > n$, which gives a partial restoration, if it is too noisy to restore the entire BP path. Conversely, it is possible to choose $n' < n$, which amounts to a partial restoration of the ket $|\Phi_n^{(i)}\rangle$ in Eq. (20). In other words, since $f_{n':n+m}^{(i)} = f_{n:n+m}^{(i)} f_{n':n}^{(i)}$, the factor $f_{n':n}^{(i)}$ can be grouped with $\omega_n^{(i)}$ to form a partial restoration in the *forward direction* of the path of length $(n-n')$ leading up to the ket. The path restoration produces weights closer to the free-projection form Eq. (12). Due to this information recovery, it was found in several molecular systems [20] that path restoration BP provides more accurate estimates of observables than phaseless BP, though leading to larger statistical fluctuations. (Several of the results in Ref. [20] were actually obtained with $n' < n$ but mislabeled as $n' = n$. [43])

The starting point to evaluate any one-body property is the one-body density matrix, also called equal-time Green's function,

$$\mathcal{G}_{ij} \equiv \langle c_i^\dagger c_j \rangle = \frac{\langle \Psi_0 | c_i^\dagger c_j | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (23)$$

where recall i represents the (\mathbf{G}_i, σ_i) pair for plane-wave basis functions. In fact, given the Green's function, estimating any one-body observables (e.g. charge density

and forces) only requires a simple post-processing operation,

$$\langle \hat{O}_{1b} \rangle = \sum_{j,k} (O_{1b})_{jk} \mathcal{G}_{kj} = \text{Tr}[O_{1b} \mathcal{G}]. \quad (24)$$

To evaluate the charge density within second quantization, we represent the field operator as

$$\hat{\varphi}_\sigma^\dagger(\mathbf{x}) = \sum_{\mathbf{G}} \varphi_{\mathbf{G}}(\mathbf{x}) c_{\mathbf{G},\sigma}^\dagger, \quad \varphi_{\mathbf{G}}(\mathbf{x}) = e^{i\mathbf{G}\cdot\mathbf{x}}, \quad (25)$$

and obtain the following expression for the charge density:

$$\rho(\mathbf{x}) = \frac{\langle \Psi_0 | \hat{\varphi}_\sigma^\dagger(\mathbf{x}) \hat{\varphi}_\sigma(\mathbf{x}) | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}. \quad (26)$$

Now, switching to plane wave basis using Eq. (25), we obtain

$$\rho(\mathbf{x}) = \sum_{ij} e^{i(\mathbf{G}_j - \mathbf{G}_i)\cdot\mathbf{x}} \delta_{\sigma_i, \sigma_j} \langle c_i^\dagger c_j \rangle. \quad (27)$$

This allows us to relate the charge density at \mathbf{x} to the Green's functions. In practice, we group elements of the Green's function with the same $\mathbf{Q} = \mathbf{G}_j - \mathbf{G}_i$, i.e. defining

$$\langle \hat{\rho}(\mathbf{Q}) \rangle = \sum_{ij} \delta_{\mathbf{Q}, \mathbf{G}_j - \mathbf{G}_i} \mathcal{G}_{ij}, \quad (28)$$

which is just an estimator of the 1-body density operator $\hat{\rho}(\mathbf{Q})$ in Eq. (4). The reciprocal-space density operator is related to the real space charge density $\rho(\mathbf{x})$ by a Fourier transform,

$$\rho(\mathbf{x}) = \sum_{\mathbf{Q}} e^{i\mathbf{Q}\cdot\mathbf{x}} \langle \hat{\rho}(\mathbf{Q}) \rangle. \quad (29)$$

The density can be computed by storing only the density operators $\hat{\rho}(\mathbf{Q})$, requiring memory $\mathcal{O}(N_{\mathbf{Q}}) \approx \mathcal{O}(8N_{\text{PW}})$. Similarly, other one-body observables can be computed based on the Green's function. The memory requirement to store the Green's function will be no larger than the above for a system with translational symmetry as our systems are.

III. COMPUTATIONAL DETAILS

In this section we describe several computational details of our calculations. The first one is the correction used to remove finite-size effect in the AFQMC density, extrapolating the result to the thermodynamic limit. The second is the pseudopotential, which is part of the definition of our many-body Hamiltonian that affects the precise values of the electronic density we provide. In the last subsection, we include any additional details on the three solid systems investigated.

A. Finite-size correction

Finite-size effects must be properly reduced and removed in many-body electronic structure calculations of extended systems. Convergence to the thermodynamic limit is often slow while computational cost tends to grow quickly with respect to supercell size. Common methods to address this problem include performing an extrapolation using several different cell size, or designing a finite-size correction method to reduce the finite-size effect.

To simulate the finite-size effects in many-body calculations, Kwee et. al. [32, 44] proposed a finite-size correction method to the total energy by designing a finite-size LDA (LDA^{FS} , also called KZK) exchange-correlation functional. Parameters in the functional were fitted using total energy results of the electron gas obtained in finite cubic supercells of volume V_0 . A correction to the energy per formula unit can then be applied to obtain an improved estimate of the value at the thermodynamic limit:

$$E[\infty] \approx E^{QMC}[V_0] - \{E^{LDA^{FS}}[V_0] - E^{LDA}[\infty]\}, \quad (30)$$

where $E^{QMC}[V_0]$ is the QMC result for the supercell with volume V_0 and at a single k -point or a set of k -points, $E^{LDA^{FS}}[V_0]$ is the corresponding calculation with the same supercell and k -point(s), using the finite-size LDA functional, and $E^{LDA}[\infty]$ is from a standard LDA calculation using a converged dense k -point grid. This finite-size correction method was found to deliver good finite-size corrections on various solid systems and allow quick convergence to the thermodynamic limit.

In our work, we adapt the KZK finite-size correction technique, and extend the concept from total energy E to charge density $\rho(\mathbf{x})$, defined for \mathbf{x} within a formula cell:

$$\rho[\infty](\mathbf{x}) \approx \rho^{QMC}[V_0](\mathbf{x}) - \{\rho^{LDA^{FS}}[V_0](\mathbf{x}) - \rho^{LDA}[\infty](\mathbf{x})\} \quad (31)$$

Using this method, we have observed accelerated convergence in all solid systems we investigated. For example, Fig. 1 shows the effect of correction along a line cut in diamond-structured silicon (see Sec. III C for details of the system). Convergence of the AFQMC charge density with increasing cell size is evident. Finite-size correction accelerates this convergence: the corrected AFQMC charge density of a $1 \times 1 \times 1$ is almost as close to the thermodynamic limit as the uncorrected $3 \times 3 \times 3$ AFQMC result, and the corrected results from $1 \times 1 \times 1$, $2 \times 2 \times 2$, and $3 \times 3 \times 3$ supercells are essentially in agreement within statistical error.

Metals like Cu have particularly strong finite-size errors, which result in non-negligible deviations between densities at different k -points even after the finite-size correction is applied. To further reduce these residual finite-size errors, we choose to perform a k -point average (i.e., using a set of k -points in Eq. (31) rather than a single one). Quasi-random k -point sequences are utilized to reduce the need of DFT smearing and increase the

convergence speed [45]. More computational details are listed in Section III C.

B. Pseudopotential Core Effect

The use of pseudopotentials in a plane-wave calculation defines an effective Hamiltonian, for which the electronic density is computed. Compared to the all-electron Hamiltonian, the contribution of the core electrons are absent in the many-body results. To illustrate the scale of densities from these electrons, Fig. 2 compares the computed Cu density from PW-AFQMC using two different pseudopotentials: (i) a Ne-core pseudopotential, the one we use in the full calculation described in the next section, which leaves 19 valence electrons, including $3s^23p^63d^{10}$ and $4s^1$ electrons; (ii) an Ar-core pseudopotential with 11 valence electrons ($3d^{10}4s^1$). The difference between the two charge density shows the effect of freezing or retaining the eight $3s^23p^6$ electrons; their contribution to the total density is localized around the

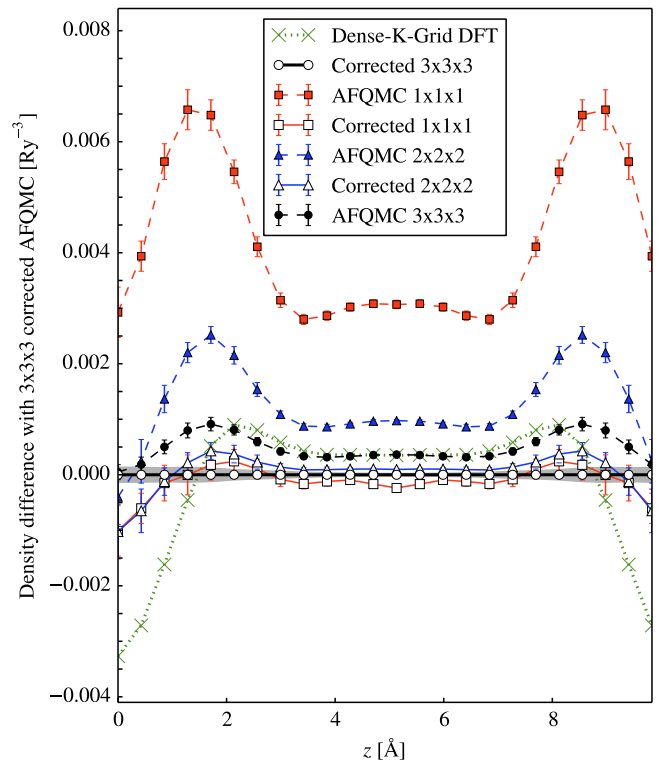


Figure 1. Demonstration of the finite-size correction of the computed electronic density, along a line cut in Silicon. Deviations from the final result at the thermodynamic limit are shown. Red, blue and black solid lines are AFQMC densities for supercell sizes $1 \times 1 \times 1$, $2 \times 2 \times 2$ and $3 \times 3 \times 3$, respectively. Dashed lines are uncorrected, “raw” AFQMC density, and the solid lines are finite-size corrected AFQMC density. Note the error for $3 \times 3 \times 3$ corrected AFQMC density is plotted in shades. For reference, the dotted green line is the dense- k -point ($6 \times 6 \times 6$) grid DFT-LDA density.

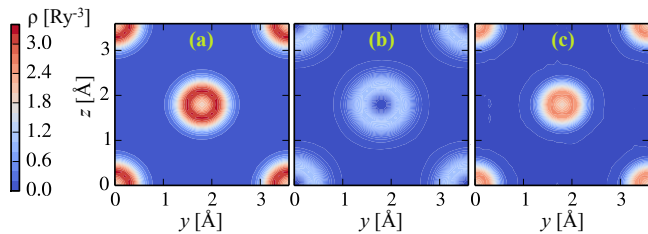


Figure 2. Pseudopotential effect in Cu: (a) PW-AFQMC density using a Ne-core pseudopotential (19 valence electrons); (b) PW-AFQMC density using a Ar-core pseudopotential (11 valence electrons); (c) difference between (a) and (b), showing the contribution of the $3s^23p^6$ electrons. In all three figures, the density is shown in the yz -plane. The color bar for density is shown on the left.

nuclei, as shown in Fig. 2c.

Any direct comparison with the AFQMC densities should be done with the same pseudopotential, as is done in Sec. V where we benchmark several DFT functionals. Assuming good transferability of the pseudopotential, we could recover the contribution of the core electrons to the electronic density, at the level of the independent-electron calculation from which the pseudopotential was generated. The core electrons are frozen in the Kohn-Sham orbitals. The following provides an estimate of their contributions:

$$\rho_{\text{AE}}^{\text{QMC}} \approx \rho_{\text{psp}}^{\text{QMC}} + (\rho_{\text{AE}}^{\text{DFT}} - \rho_{\text{psp}}^{\text{DFT}}), \quad (32)$$

where $\rho_{\text{psp}}^{\text{QMC}}$ is our result, and the correction term on the right involves separate DFT calculations, all-electron (AE) and using the same pseudopotential as in PW-AFQMC (psp).

We use a multiple-projector optimized norm-conserving pseudopotential (ONCVSP) of Hamann [46], whose details are given in Appendix A. This pseudopotential was found to allow the use of a lower kinetic energy cutoff while maintaining excellent accuracy in AFQMC [31].

C. Systems and parameters

The covalent-bond crystal silicon (Si). This Si solid has a diamond-like structure, with all carbon atoms replaced by silicon atoms. The primitive cell is a face-centered cubic (FCC) cell, with an experimental [47] lattice constant of $7.257 \text{ Bohr} = 3.840 \text{ \AA}$. This FCC cell consists of two Si atoms, located at $(\pm\frac{1}{8}, \pm\frac{1}{8}, \pm\frac{1}{8})$. A commonly used non-primitive cell is the cubic cell, with $4\times$ the volume and a lattice constant of 5.431 \AA . Both cells, along with the $2\times 2\times 2$ and $3\times 3\times 3$ multiples of the FCC primitive cells, are used in our calculations to ensure a finite-size convergence. Densities from FCC $3\times 3\times 3$ supercell (54 atoms) are presented, with Baldereschi mean-value point [48] adopted. We use a Ne-core pseudopoten-

tial which has 4 valence electrons per silicon atom. The AFQMC plane-wave kinetic energy cutoff is 25 Ry.

The ionic crystal sodium chloride (NaCl). We use the naturally existing cubic form of NaCl crystal, where the lattice constant is 5.692 \AA , taken from the Material's Project [49, 50]. This cell consists of 4 Na atoms, at each vertex and face center, and 4 Cl atoms, at the center of each edge and the bulk center. Although a cubic cell is commonly used, the actual primitive cell is the $1/4\times$ volume FCC cell, which comprises of only one Na atom (at lattice points) and one Cl atom (at bulk center). We use both these cells as well as the $2\times 2\times 2$ multiple of the primitive FCC cell for finite-size convergence, and FCC $2\times 2\times 2$ (8 Na atoms and 8 Cl atoms) is used for the plots below. The Baldereschi mean-value point in the FCC lattice is used. The kinetic energy cutoff is 40 Ry.

The transition metal copper (Cu). The primitive cell of Cu is an FCC with all lattice points being occupied by copper atoms, having only one Cu atom per cell. We use a 4-atom cubic supercell for most of the calculations. This cubic cell has an experimental [31, 51] lattice constant of 3.59 \AA . To characterize the finite-size effect, we also study several larger cells: FCC $2\times 2\times 2$ (8 atoms), BCC $2\times 2\times 2$ (16 atoms), and cubic $2\times 2\times 2$ (32 atoms).

Cu as a metal requires k -point averaging. We use a quasi-random Sobol sequence [52], with 12 k -points in the cubic $1\times 1\times 1$ cell. Our calculations are performed with a Ne-core pseudopotential (see Appendix A) and a kinetic energy cutoff of 64 Ry.

IV. ELECTRONIC DENSITY FROM AFQMC

In this section we present the final charge densities computed with PW-AFQMC for Si, NaCl, and Cu. For each system, the results are presented on a selected high-symmetry plane, and then along a path of 1D line cuts. In all three solids we plot the final densities from a cubic $1\times 1\times 1$ supercell regardless of the actual supercell size used in calculation (see Sec. III C). Numerical values are provided as supplemental information, online at [53].

In Si, we plot the density in the $(01\bar{1})$ plane in Fig. 3. The signature of the Si-Si covalent bond is evident, as a concentration of valence electrons is seen on a line connecting the nearest neighbor Si atoms. Note that there are small bumps located in the vicinity of each Si atom and complementing the two bonds in the plane, which are contributions from two Si-Si bonds pointing out of the plane, reflecting the nature of the sp^3 hybridization. We also plot the charge density along line cuts in Fig. 4, following the route $O-\langle 001\rangle-O'-\langle 110\rangle-O''-\langle 111\rangle-O$, which forms a triangle as illustrated in the inset. The origin O is taken to be the high-symmetry middle point between two neighboring Si atoms, while O' and O'' are translated from O by lattice constants along the direction connecting them.

We next present the results for NaCl in a similar fashion. In Fig. 5, the density is shown on the xy -

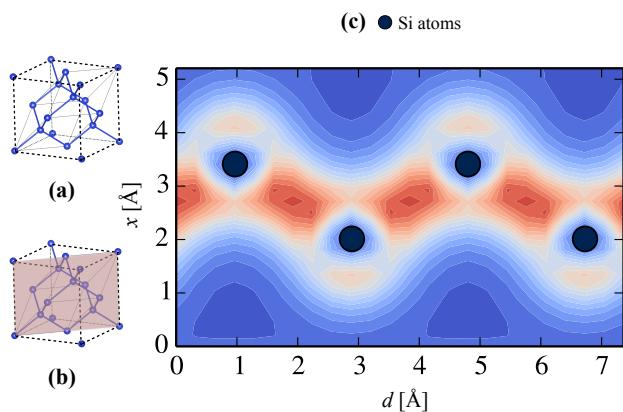


Figure 3. Charge density of Si from PW-AFQMC. The lattice structure is shown in (a), and (b) illustrates the $(01\bar{1})$ plane ($y = z$) for which the density is plotted. In (c), higher density is given in red and lower density in blue. The vertical axis is the Cartesian x -value, while the horizontal axis gives $d = \frac{1}{\sqrt{2}}(y + z)$.

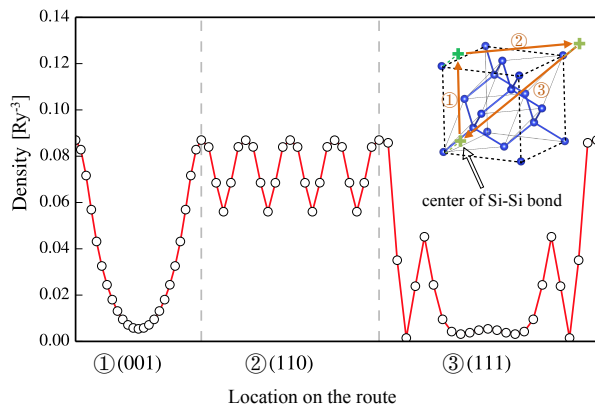


Figure 4. Charge density of Si from PW-AFQMC, plotted on a high-symmetry path. The starting and ending point of the route is the origin of the simulation cell, which is located at the center of two nearest neighbor Si atoms. The route then goes through three segments in $\langle 001 \rangle$, $\langle 110 \rangle$, and $\langle 111 \rangle$ directions, as indicated in the inset.

plane ($z = 0$). Then in Fig. 6, we plot the density along the triangular path, following the route $O - \langle 001 \rangle - O' - \langle 110 \rangle - O'' - \langle 111 \rangle - O$, with the origin O taken to be a Na atom. Fig. 7 shows the density from a different perspective, illustrating the ionic nature of NaCl. Within a sphere of radius 1 \AA centered at a Na atom, the integrated charge density is ~ 8 electron, consistent with a Na^+ (with our He-core pseudopotential). This integrated charge density remains saturated with increasing radius, until the sphere reaches the vicinity of the nearest Cl atom. Around a Cl atom, this integrated density also

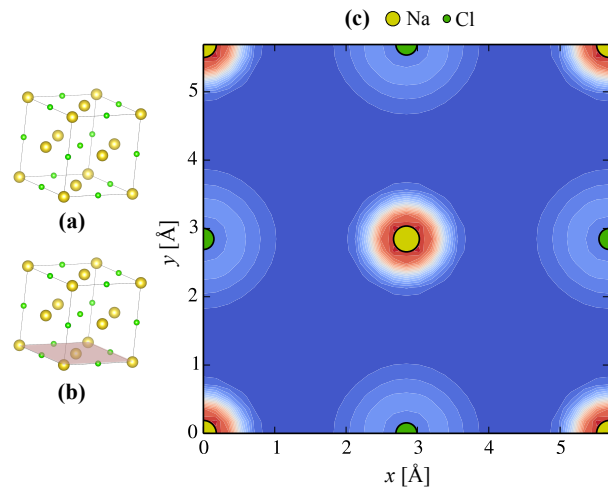


Figure 5. Charge density of NaCl from PW-AFQMC. The layout is the same as Fig. 3. The density is plotted in the xy -plane.

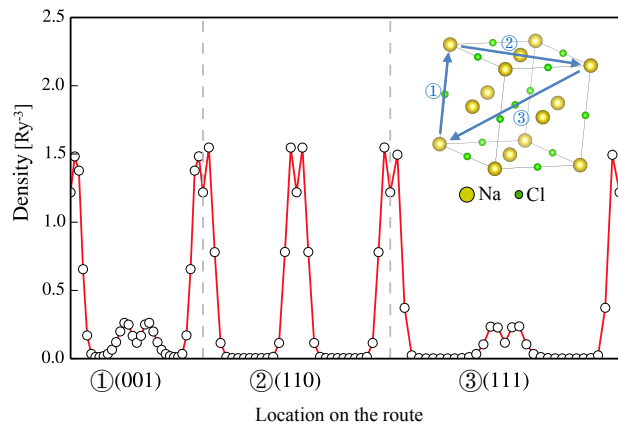


Figure 6. Charge density of NaCl from PW-AFQMC, plotted on a high-symmetry path. The layout is the same as Fig. 4.

approaches ~ 8 electron when the radius of the sphere is around 2 \AA , consistent with a Cl^- ion (Ne-core pseudopotential).

Results for Cu are shown in Figs. 8 and 9. The charge density in Cu bears a resemblance to that of NaCl. However the density scales are very different in Fig. 9, which obscures the significant density in between Cu atoms. The distinction between this metallic system and the ionic crystal NaCl, and the semiconductor Si can be seen more clearly in Fig. 10, in which we plot the distribution of electron densities in each solid. The distribution is characterized by the percentage of the real-space FFT grid points having a given charge density value. (It should be noted that this is in presence of the pseudopotential and without accounting for the core electrons.) We see that a pronounced peak is seen in Cu at around $\rho = 0.04 [\text{Ry}^{-3}]$, with 55% of FFT grid points in Cu share a density value of $(4 \pm 1) \times 10^{-2} [\text{Ry}^{-3}]$, corresponding

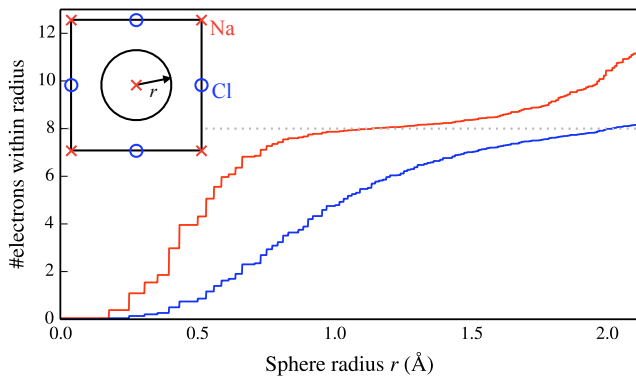


Figure 7. Number of electrons within a sphere of radius r centered at either a Na or a Cl nucleus in the computed PW-AFQMC density. Red curve is for Na and blue curve is for Cl.

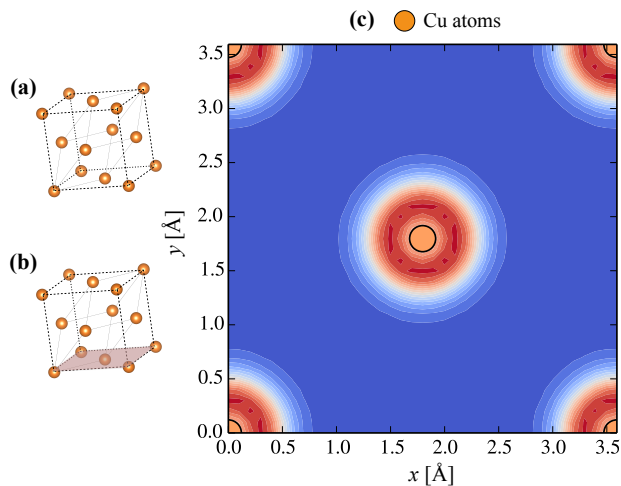


Figure 8. Charge density of Cu from PW-AFQMC. The layout is the same as Fig. 3. The density is plotted in the xy -plane.

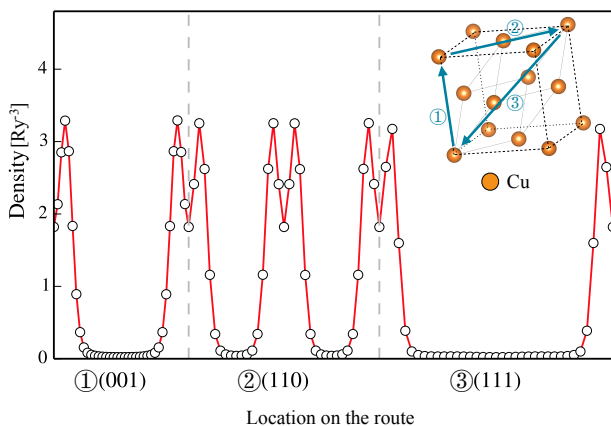


Figure 9. Charge density of Cu from PW-AFQMC, plotted on a high-symmetry path. The layout is the same as Fig. 4.

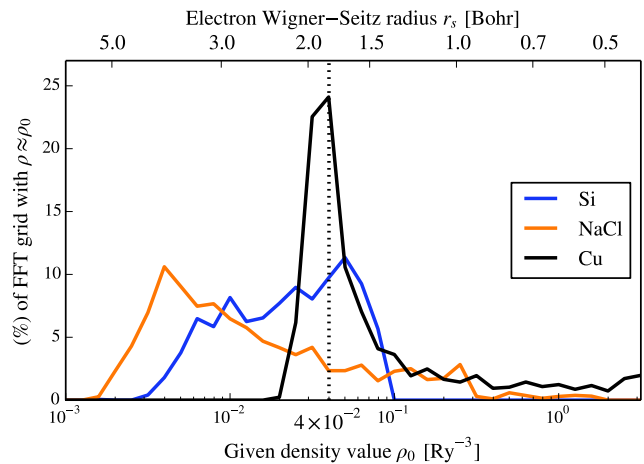


Figure 10. Distribution of the electron density in different solids. Shown are the percentage of the real-space FFT grid points with density around each given value ρ_0 , for Si, NaCl, and Cu from the PW-AFQMC densities. Corresponding values of the electron Wigner-Seitz radius r_s are marked on top. Note the logarithmic scale of the horizontal axis.

to an electron Wigner-Seitz radius r_s of $1.7 \sim 2.0$ Bohr. This is consistent with the notion of a metallic system being characterized as a “uniform electron gas,” and is not seen in NaCl or Si. In particular, NaCl shows a significant concentration in a low-density region, while Si features a broader peak.

V. BENCHMARK DFT FUNCTIONALS

A. Comparison of the electronic densities from several functionals with AFQMC results

Based on a variety of benchmark studies (see, for example, Refs. [21, 23, 24]), the AFQMC results are expected to be highly accurate. The density results we have presented from AFQMC can serve as a reference in these systems. We next carry out a comparative study of the computed densities from several popular DFT functionals, including: (i) LDA, the local-density approximation by Perdew-Zunger (PZ) [54]; (ii) PBE, a generalized-gradient approximation (GGA) functional by Perdew-Burke-Ernzerhof [55]; (iii) PBEsol, a revised GGA functional for solids [56]; (iv) PBE0, a hybrid functional between PBE and Hartree-Fock [57]; (v) B3LYP, a widely used hybrid functional in quantum chemistry [58]. In each DFT calculation, we use the same pseudopotential as in the PW-AFQMC calculations, namely the multiple-projector ONCVSP generated with LDA. In other words, we adopt the philosophy of viewing the pseudopotential as defining an effective Hamiltonian, and ask how each functional treats this effective Hamiltonian as compared to AFQMC.

In Fig.’s 11, 12, and 13, we present the result for Si,

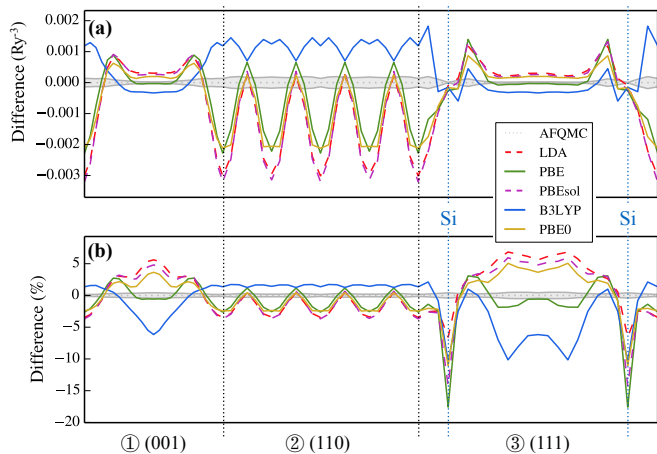


Figure 11. Benchmark of five DFT exchange-correlation functionals against the PW-AFQMC density in Si. (a) The difference between the DFT and AFQMC densities and (b) the relative errors as a percentage are plotted along the same line cut as in Sec. IV. The gray shade indicates the AFQMC statistical error bar. Blue vertical lines mark the position of the Si atoms.

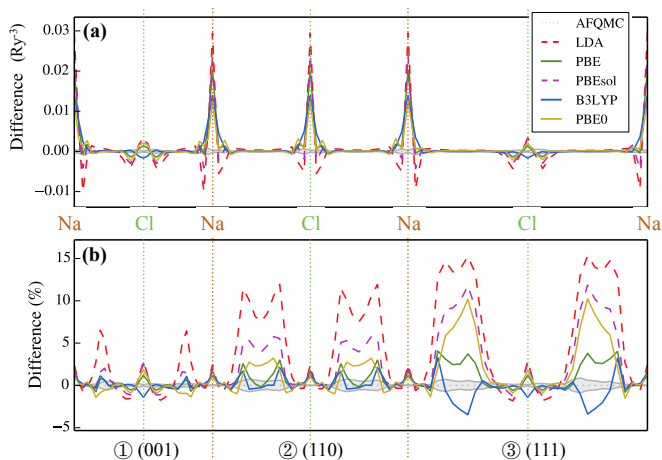


Figure 12. Same as Figure 11, but for NaCl. Brown and green vertical lines mark the positions of Na and Cl atoms, respectively.

NaCl, and Cu, respectively, by plotting the densities from the five different DFT functionals against the reference AFQMC density. Density differences and percentage discrepancies are shown in two different panels, following the same line cuts as used in Fig.’s 4, 6, and 9, respectively. The AFQMC statistical error bars are given by the shades. Additional details on the benchmark are provided in Appendix B.

For Si, PBE and PBE0 perform very well, with average percentage errors of around 2%. B3LYP shows a comparable performance along the route, but is less accurate for the entire cell. LDA and PBEsol show larger errors in general. We see correlations in the location and amount of the errors between all functionals, except for

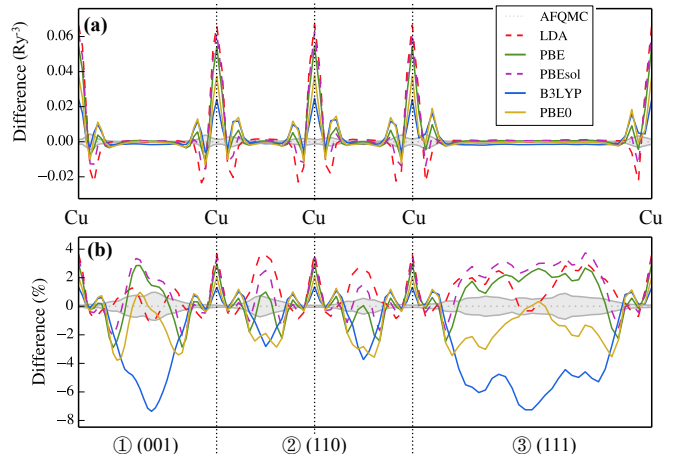


Figure 13. Same as Figure 11, but for fcc-Cu. Black vertical lines mark the positions of the Cu atoms.

B3LYP whose behavior is more different from the other functionals. In NaCl, the most significant errors occur between Na and Cl atoms, along the diagonal direction. Except for B3LYP, all functionals tend to give a density that is consistently too high in these regions, well outside the AFQMC statistical error. The errors from the five functionals all show some correlation, with B3LYP giving the smallest errors in these regions, followed by PBE and PBE0. In Cu, the behavior of the errors is more subtle. The absolute and relative errors in the plot tell somewhat different stories. B3LYP yields the smallest integrated absolute error along the chosen path, but gives the largest percentage error, and also the second largest integrated absolute error over the entire cell.

B. Accuracy of a functional: density vs. total energy

Although charge density plays a vital role in DFT, the accuracy of the density from a particular exchange-

Table I. Equilibrium lattice constant (a_{eq}) and bulk modulus (B_0) computed from several DFT functionals, compared to experimental results in Si and NaCl. The last two rows give the results from one-shot calculations of the EOS using PBE exchange-correlation functional but densities (wave functions) obtained from B3LYP and LDA respectively.

XC	Input Density	NaCl		Silicon	
		a_{eq} (Å)	B_0 (GPa)	a_{eq} (Å)	B_0 (GPa)
Exp.		5.640 [59]	24.42 [60]	5.431 [34]	98.8 [34]
LDA	(same)	5.46	32.21	5.395	95.19
B3LYP		5.58	27.98	5.382	96.40
PBEsol		5.60	26.06	5.394	94.89
PBE		5.672	24.28	5.406	92.04
PBE	B3LYP	5.681	23.27	5.408	90.65
	LDA	5.682	22.93	5.404	91.53

correlation functional does not seem to provide an unambiguous measure of the quality of the functional. We illustrate this point below following the benchmark results on Si and NaCl in the previous subsection.

We perform equation of state (EOS) calculations in these solids to obtain the equilibrium lattice constant and bulk modulus by fitting the total energies using the Murnaghan equation [61]. The results are tabulated in Table I, together with the experimental values. In NaCl, PBE yields the best results for both the equilibrium lattice constant and the bulk modulus. However, we recall that B3LYP is the one that yielded the best density. In Si, B3LYP shows the largest error in lattice constant and the best accuracy in bulk modulus, while PBE is at the opposite, with the smallest error in lattice constant and the largest in bulk modulus, although the margins are all rather small here. On the other hand, the result from the overall electronic density shows no ambiguity and indicates that PBE performs the best, as discussed in Sec. V A.

To reconcile these inconsistent behaviors in density vs. total energy, we examine the effect of the exchange-correlation energy separately from the density. We use the PBE functional to perform a one-shot total energy calculation by feeding it a wave function from B3LYP (or LDA). The wave function is obtained from a converged self-consistent B3LYP calculation, by taking the occupied orbitals. Thus the wavefunction produces the electronic density from the fully self-consistent B3LYP calculation. The resulting EOS using total energies computed from this procedure is re-analyzed, and shown in the bottom two rows, for B3LYP and LDA wave functions respectively. We see that the results in general exhibit only minor variations from the PBE results. In the case of LDA, the improvement is dramatic in NaCl, and mixed in Si, with the bulk modulus becoming worse as it moves towards the PBE result (which also occurs with B3LYP density). This shows that the accuracy of the electronic density is secondary in the performance of an XC functional for computing the EOS, and the primary factor affecting overall performance in these solids is the exchange-correlation energy.

VI. CONCLUSION

We have introduced into the plane-wave AFQMC framework the calculation of observables using a path-restoration back-propagation technique. We apply it to compute the charge density in three typical solids with different crystal bonding mechanisms, the covalent-bond crystal Si, the ionic-bond NaCl, and the transition metal fcc-Cu, which provides highly accurate *ab initio* many-body electronic densities in these systems. These results are reported, and compared with results from several of the most popular density functionals. In general, the densities produced by these functionals agree quite well with the PW-AFQMC results. The discrepancies from the

different functionals are quantified, and our results can be used for future benchmarks of other computational methods. Additionally, the PW-AFQMC electronic densities may help with the development of improved density functionals.

Besides charge density, the back-propagation technique extends easily to other quantities, including interatomic forces, which can be used to perform accurate molecular dynamics and geometry optimization, as well as to calculate phonon properties and access thermodynamic properties of solids. Work is on-going along these lines.

ACKNOWLEDGMENTS

We thank D. Ceperley, M. Holzmann, H. Krakauer, E. J. Walter and J. Lee for useful discussions. We are grateful to L. Reining for discussions and for pointing out an error in an early version of the manuscript. S.C. would like to thank the Center for Computational Quantum Physics, Flatiron Institute for support and hospitality. We also acknowledge support from the U.S. Department of Energy (DOE) under grant DE-SC0001303. F.M. is supported by the National Natural Science Foundation of China under Grants No. 11674027. The authors thank William & Mary Research Computing and Flatiron Institute Scientific Computing Center for computational resources and technical support. The Flatiron Institute is a division of the Simons Foundation.

Appendix A: More information on the pseudopotential

Our pseudopotentials for Si, NaCl, and Cu are constructed with Hamann's multiple-projector optimized norm-conserving pseudopotential (ONCVSP) code [46], using LDA as the exchange correlation functional. The local reference in ONCVSP is a smooth continuation of the all-electron potential from the minimum radius cutoff r_c to 0, using a smooth polynomial (lloc=4 in ONCVSP input). For compatibility with AFQMC, the nonlinear core correction is turned off in pseudopotential generations.

For Si, the kinetic energy cutoff is 25 Ry and the electron configuration is $1s^2 2s^2 2p^6 3s^2 3p^{1.6} 3d^{0.4}$. The first three orbitals are taken as core orbitals. We use three projectors $l = 0, 1, 2$, all with the same core radius r_c of 2.04 Bohr.

For NaCl, the kinetic energy cutoff is 40 Ry. In Na, the electron configuration is $1s^2 2s^2 2p^6 3s^1$, with only $1s$ orbital as core, due to the fact that the $2s2p$ semi-core electrons of Na have large overlaps with the $3s$ electron, and neglecting them from the valence was seen to cause errors in the equation of state in AFQMC calculations [31]. Core radii r_c for the three projectors $l = 0, 1, 2$ are 1.82, 2.27, 2.27 Bohr, respectively. In Cl, the electron configuration is $1s^2 2s^2 2p^6 3s^2 3p^5$, and the first three

orbitals are taken as core. The core radii for the three projectors ($l = 0, 1, 2$) are 2.72, 2.72, and 3.01 Bohr, respectively.

For Cu, we have used two different pseudopotentials, with Ne-core and Ar-core, both using a kinetic energy cutoff of 64 Ry. The Ne-core pseudopotential, which is the primary one used in the paper, has an electron configuration of $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{9.1} 4s^{0.5} 4p^{1.4}$. Core radii are 1.60, 1.97, and 1.97 Bohr for the three projectors; the corresponding wave-vector cutoffs are 7.0, 7.75 and 8.0 Ry. For the Ar-core pseudopotential, the electron configuration is $1s^2 2s^2 2p^6 3s^2 3p^6 3d^9 4s^1 4p^1$, with the first 5 orbitals being the core orbitals. Core radii for $l = 0, 1, 2$ are 2.30, 2.30, and 2.10 Bohr, respectively. The wave-vector cutoff (8 Ry) is the same for all projectors.

Appendix B: Supplementary Data on Exchange-Correlation Functional Benchmarks

In Tables II, III, and IV, we list the mean absolute error (MAE) in the density of multiple DFT exchange correlation functionals with respect to the near-exact AFQMC density, on the high-symmetry triangular route $O-(001)-O_{001}-\langle 110 \rangle-O_{111}-\langle 111 \rangle-O$ and the full 3D real-space grid, for Si, NaCl, and Cu, respectively. The formula for this MAE is

$$|\bar{\Delta\rho}| = \frac{1}{N_G} \sum_{g \in \{G\}} |\Delta\rho_g|,$$

where $\{G\}$ is the set of real-space grid points, located on the triangular route or the full 3D real-space grid; N_G is the number of grid points in $\{G\}$; and $\Delta\rho_g = \rho_{\text{XC}} - \rho_{\text{AFQMC}}$ for a given exchange-correlation functional “XC.” For mean percentage difference a similar formula is used,

$$\Delta\rho_g = 100 (\%) \times \frac{\rho_{\text{XC}} - \rho_{\text{AFQMC}}}{\rho_{\text{AFQMC}}}.$$

-
- [1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- [2] R. O. Jones, Rev. Mod. Phys. **87**, 897 (2015).
- [3] A. D. Becke, J. Chem. Phys. **140**, 18A301 (2014).
- [4] K. Burke, J. Chem. Phys. **136**, 150901 (2012).
- [5] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
- [6] E. M. Stoudenmire, L. O. Wagner, S. R. White, and K. Burke, Phys. Rev. Lett. **109**, 056402 (2012).
- [7] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, J. Chem. Phys. **147**, 184111 (2017).
- [8] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
- [9] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
- [10] S. Zhang and H. Krakauer, Phys. Rev. Lett. **90**, 136401 (2003).
- [11] M. Motta and S. Zhang, WIREs Comput. Mol. Sci. **8**, e1364 (2018).
- [12] J. Hubbard, Phys. Rev. Lett. **3**, 77 (1959).
- [13] G. Sugiyama and S. Koonin, Annals of Physics **168**, 1 (1986).
- [14] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).
- [15] D. J. Scalapino and R. L. Sugar, Phys. Rev. B **24**, 4295 (1981).
- [16] S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. B **55**, 7464 (1997).
- [17] S. Zhang, in *Handbook of Materials Modeling*, edited by W. Andreoni and S. Yip (Springer, Cham, 2018) 1807.06688.
- [18] M. Siewattana, W. Purwanto, S. Zhang, H. Krakauer, and E. J. Walter, Phys. Rev. B **75**, 245123 (2007).
- [19] W. Purwanto and S. Zhang, Phys. Rev. E **70**, 056702 (2004).
- [20] M. Motta and S. Zhang, J. Chem. Theor. Comput. **13**, 5367 (2017).
- [21] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, and E. Gull (Simons Collaboration on the Many-Electron Problem), Phys. Rev. X **5**, 041041 (2015).
- [22] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Science **358**, 1155 (2017).
- [23] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang (Simons Collaboration on the Many-Electron Problem), Phys. Rev. X **7**, 031059 (2017).
- [24] K. T. Williams, Y. Yao, J. Li, L. Chen, H. Shi, M. Motta, C. Niu, U. Ray, S. Guo, R. J. Anderson, J. Li, L. N. Tran, C.-N. Yeh, B. Mussard, S. Sharma, F. Bruneval, M. van Schilfgaarde, G. H. Booth, G. K.-L. Chan, S. Zhang, E. Gull, D. Zgid, A. Millis, C. J. Umrigar, and L. K. Wagner (Simons Collaboration on the Many-Electron Problem), Phys. Rev. X **10**, 011041 (2020).
- [25] J. Shee, B. Rudsteyn, E. J. Arthur, S. Zhang, D. R. Reichman, and R. A. Friesner, J. Chem. Theor. Comput. **15**, 2346 (2019).
- [26] W. A. Al-Saidi, S. Zhang, and H. Krakauer, The Journal of Chemical Physics **124**, 224101 (2006).
- [27] F. Ma, W. Purwanto, S. Zhang, and H. Krakauer, Phys.

- Rev. Lett. **114**, 226401 (2015).
- [28] M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis, U. Ray, E. Ronca, H. Shi, S. Sorella, E. M. Stoudenmire, S. R. White, and S. Zhang (Simons Collaboration on the Many-Electron Problem), Phys. Rev. X **10**, 031058 (2020).
- [29] P. Pulay, Wiley Interdisciplinary Reviews: Computational Molecular Science **4**, 169 (2014).
- [30] W. Purwanto, H. Krakauer, and S. Zhang, Phys. Rev. B **80**, 214116 (2009).
- [31] F. Ma, S. Zhang, and H. Krakauer, Phys. Rev. B **95**, 165103 (2017).
- [32] H. Kwee, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **100**, 126404 (2008).
- [33] M. Born and R. Oppenheimer, Annalen der physik **389**, 457 (1927).
- [34] C. Kittel, P. McEuen, and P. McEuen, *Introduction to solid state physics*, Vol. 8 (Wiley New York, 1996).
- [35] R. M. Martin, *Electronic structure: basic theory and practical methods* (Cambridge university press, 2020).
- [36] P. P. Ewald, Annalen der physik **369**, 253 (1921).
- [37] S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **373**, 27 (1980).
- [38] S. De Leeuw, J. W. Perram, and E. Smith, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **373**, 57 (1980).
- [39] D. J. Thouless, Nuclear Physics **21**, 225 (1960).
- [40] D. Thouless, Nuclear Physics **22**, 78 (1961).
- [41] H. Shi and S. Zhang, The Journal of Chemical Physics **154**, 024107 (2021).
- [42] S. Zhang, Lecture Notes of the Autumn School Correlated Electrons **2013** (2013).
- [43] M. Motta, Private Communication (2020).
- [44] F. Ma, S. Zhang, and H. Krakauer, Phys. Rev. B **84**, 155130 (2011).
- [45] M. Qin, H. Shi, and S. Zhang, Phys. Rev. B **94**, 085103 (2016).
- [46] D. R. Hamann, Phys. Rev. B **88**, 085117 (2013).
- [47] P. J. Mohr, D. B. Newell, and B. N. Taylor, Rev. Mod. Phys. **88**, 035009 (2016).
- [48] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973).
- [49] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. a. Persson, APL Materials **1**, 011002 (2013).
- [50] “mp-22862: NaCl (cubic, Fm-3m, 225),” Material’s Project (2021), accessed: 2021-01-28.
- [51] L. Schimka, J. Harl, and G. Kresse, The Journal of Chemical Physics **134**, 024116 (2011).
- [52] I. M. Sobol, U.S.S.R. Comput. Maths. Math. Phys. **7**, 86 (1967).
- [53] S. Chen, “github.com/schen24wm/chgden,” (2020).
- [54] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
- [55] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
- [56] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008).
- [57] J. P. Perdew, M. Ernzerhof, and K. Burke, The Journal of Chemical Physics **105**, 9982 (1996).
- [58] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, The Journal of Physical Chemistry **98**, 11623 (1994).
- [59] W. T. Barrett and W. E. Wallace, J. Am. Chem. Soc. **76**, 366 (1954).
- [60] L. S. Combes, S. S. Ballard, and K. A. McCarthy, J. Opt. Soc. Am. **41**, 215 (1951).
- [61] F. D. Murnaghan, Proceedings of the National Academy of Sciences **30**, 244 (1944).

Table II. MAE and mean percentage difference of DFT densities in Silicon with LDA, PBE, PBEsol, B3LYP, and PBE0 exchange-correlation functionals, compared to AFQMC results. The first column gives the MAE, while the second column gives the mean percentage difference, each including an average on the high-symmetric triangular route $O-\langle 001 \rangle - O_{001} - \langle 110 \rangle - O_{111} - \langle 111 \rangle - O$ and the full 3D real-space grid. The average absolute AFQMC statistical error, in value and in percentage, are listed at the first row as a reference.

	MAE ($\times 10^{-4}$ Ry $^{-3}$)		Mean percentage error (%)	
	Triangular route	Full 3D grid	Triangular route	Full 3D grid
AFQMC^a	1.04	0.76	0.29	0.30
LDA	10.4	4.35	3.10	2.03
PBE	6.9	2.87	1.86	0.97
PBEsol	10.6	4.21	3.12	1.77
B3LYP	7.1	3.86	2.76	2.17
PBE0	7.8	2.97	2.32	1.23

^a AFQMC statistical error

Table III. Same as Table II but for NaCl.

	MAE ($\times 10^{-4}$ Ry $^{-3}$)		Mean percentage error (%)	
	Triangular route	Full 3D grid	Triangular route	Full 3D grid
AFQMC^a	1.74	0.75	0.26	0.37
LDA	29.0	6.64	4.47	5.13
PBE	13.7	1.43	0.86	0.81
PBEsol	19.0	3.11	2.40	2.49
B3LYP	12.9	1.27	0.67	0.61
PBE0	11.7	2.10	1.62	1.30

^a AFQMC statistical error

Table IV. Same as Table II but for fcc-Cu.

	MAE ($\times 10^{-4}$ Ry $^{-3}$)		Mean percentage error (%)	
	Triangular route	Full 3D grid	Triangular route	Full 3D grid
AFQMC^a	9.5	4.3	0.36	0.46
LDA	91.2	16.1	1.20	1.52
PBE	60.8	12.9	1.24	1.14
PBEsol	73.7	12.3	1.43	1.39
B3LYP	44.0	21.7	2.57	2.81
PBE0	53.7	22.1	1.40	2.05

^a AFQMC statistical error