




**Electronic band gaps from quantum Monte Carlo methods**Yubo Yang (杨煜波) <sup>1</sup>, Vitaly Gorelov,<sup>2</sup> Carlo Pierleoni <sup>2,3</sup>, David M. Ceperley,<sup>1</sup> and Markus Holzmann<sup>4,5</sup><sup>1</sup>*Department of Physics, University of Illinois, Urbana, Illinois 61801, USA*<sup>2</sup>*Maison de la Simulation, CEA, CNRS, Université Paris-Sud, UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*<sup>3</sup>*Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 10, I-67010 L'Aquila, Italy*<sup>4</sup>*Univ. Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France*<sup>5</sup>*Institut Laue Langevin, BP 156, F-38042 Grenoble Cedex 9, France* (Received 6 November 2019; revised manuscript received 16 January 2020; accepted 21 January 2020; published 11 February 2020)

We develop a method for calculating the fundamental electronic gap of semiconductors and insulators using grand canonical quantum Monte Carlo simulations. We discuss the origin of the bias introduced by supercell calculations of finite size and show how to correct the leading and subleading finite size errors either based on observables accessible in the finite-sized simulations or from density-functional theory calculations. Our procedure is applied to carbon, silicon, and molecular hydrogen crystals, and compared to experiment for carbon and silicon.

DOI: [10.1103/PhysRevB.101.085115](https://doi.org/10.1103/PhysRevB.101.085115)**I. INTRODUCTION**

Insulator and semiconductors are characterized by a nonvanishing *fundamental gap* [1], defined in terms of the ground-state energies of a system of fixed ions as the number of electrons is varied,

$$\Delta_{N_e} = E_0(N_e + 1) + E_0(N_e - 1) - 2E_0(N_e), \quad (1)$$

where  $E_0(N_e)$  is the ground-state energy of an  $N_e$  electron system.

Within density-functional theory (DFT), it is common to interpret the eigenvalues of the Kohn-Sham equations as excitation energies, the gap being the minimum excitation energy. However, the resulting band gap within the local density approximation (LDA) is typically found to be too small [2]. This qualitative failure can be alleviated either by hybrid functionals or by adding corrections based on GW many-body perturbation theory, although the precise value depends on the underlying functional and approximation scheme involved [1]. In principle, the fundamental gap can be calculated from any method for ground-state energies based on the above formula. High-precision methods for correlation energies as, for example, provided by quantum Monte Carlo (QMC) [3–6] or coupled cluster methods [7,8] can be used. In this paper, we propose a method for accurate calculations of the fundamental gap within explicitly correlated methods and demonstrate its use with fixed-node diffusion Monte Carlo (DMC) benchmark studies on solid  $H_2$ , C, and Si.

Methods based on correlated many-body wave functions are usually applied to finite-sized systems, e.g., limited to supercells containing only few unit cells. QMC calculations of single-particle excitations for adding and removing electrons [9–12] crucially rely on the imposed extrapolation law (e.g., finite-size error  $\propto 1/L$  in Ref. [12] opposed to  $1/L^3$  in Ref. [11], where  $L$  denotes the linear extension of the supercell). This introduces considerable uncertainty in the

results. Heuristically, single-particle excitations are expected to converge slowly for electronic systems, inversely proportional to  $L$ , due to the interaction of charges across the periodic boundaries [13,14]. Extrapolations with respect to the size of the supercells are then essential to obtain reliable values of the gap in the thermodynamic limit.

Most of the QMC calculations [15–24] have therefore addressed charge-neutral, particle-hole excitations, where faster convergence with respect to the size of the supercell is expected. Although the comparison with experiment is appealing [5], a later, more extended DMC study [25] of simple semiconductor materials with larger supercells observed a  $1/L$  dependence of the gap on the size of the supercell for both charged single-particle and charge-neutral particle-hole excitations. In addition, fixed-node energy differences are not constrained to be upper bounds for particle-hole excitations [26] since orthogonality to the ground state cannot be strictly guaranteed. Furthermore, all QMC calculations so far have addressed excitations at selected symmetry points contained inside the supercell of the simulation. The fundamental gap was then estimated indirectly by introducing a “scissor operator” [27] which assumes a rigid shift of the underlying DFT band structure over the whole Brillouin zone.

In this paper, we show that twisted boundary conditions within the grand canonical ensemble can be used to determine the fundamental gap from QMC without relying on the “scissor” approximation. We prove that to leading order, finite size effects due to two-body correlations are of order  $1/L$ , and are related to the dielectric constant of the material. Such effects can be understood and corrected for by using the long wavelength properties of the electronic structure factor. For that, we extend the approach described in Refs. [28,29] which discusses the correction of finite size effects on the ground-state energy based on information contained in the static correlation functions of the finite system. Using the static structure factor from simulation, it is possible to

obtain estimates of finite size corrections for the band gap, and its asymptotic functional form without the need for explicit studies at different sizes or referring to DFT or to experimental information external to the QMC calculation.

The paper is organized as follows. In Sec. II, we describe the main ideas behind our band-gap method based on the grand canonical ensemble. In Sec. III, we derive finite size corrections to energy differences based on an explicit many-body wave function and exact diagrammatic relations. In Sec. IV, we describe the computational methods used to calculate the fundamental gap. In Sec. V, we show results for H<sub>2</sub>, C, and Si crystals and compare with available experimental values of the gap in Sec. VI. Finally in Sec. VII, we summarize general features of the method and outline possible extensions and applications.

## II. GRAND-CANONICAL TWIST-AVERAGED BOUNDARY CONDITION (GCTABC)

In the following, we consider  $N_e$  electrons in a perfect crystal, neglecting both zero-point and thermal motion of the ions. A uniform background charge (depending on  $N_e$ ) is added to assure global charge neutrality when adding or subtracting electrons to a charge-neutral system. The background charge will introduce a rigid shift in the density of states. However, the fundamental gap, Eq. (1), is unaffected, because the background charge needed when adding an electron cancels against the one needed when removing an electron. Periodic boundary conditions of the charge densities are used to eliminate surface effects.

The energetic cost of adding an electron to the system at fixed volume,  $V = L^3$ , defines the chemical potential:

$$\mu_{N_e}^+ = E_0(N_e + 1) - E_0(N_e). \quad (2)$$

A nonvanishing gap implies a discontinuity in the chemical potential from Eq. (1).

It is convenient to work in the grand-canonical ensemble. There, the chemical potential  $\mu$  is treated as an independent variable and we minimize  $E_0(N_e) - \mu N_e$  with respect to  $N_e$  at zero temperature and fixed volume. Insulators then represent an incompressible electronic state; for values of  $\mu$  within the gap,  $\partial N_e / \partial \mu = 0$ .

To reduce finite size effects, we employ twisted boundary conditions on the many-body wave function. As an electron is moved across the supercell, e.g., by moving an electron a distance equal to the size of the box in the  $x$  direction,

$$\Psi(\mathbf{r}_1 + L_x \hat{x}) = e^{i\theta_x} \Psi(\mathbf{r}_1), \quad (3)$$

the phase of the many-body wave function changes by  $\theta_x$ . The ground-state energy then depends on the twist angle,  $E_0(N_e, \theta)$ . Twist averaging can significantly accelerate the convergence to the thermodynamic limit [30]. Within the grand-canonical ensemble [28,29], the optimal number of electrons  $\bar{N}_e(\theta)$  will depend on  $\theta$  for given chemical potential  $\mu$ . To fix nomenclature, we define the *mean electronic density*,

$$n_e(\mu) = (M_\theta V)^{-1} \sum_{\theta} \bar{N}_e(\theta), \quad (4)$$

and the *ground-state energy density*,

$$e_0(\mu) = (M_\theta V)^{-1} \sum_{\theta} E_0(\bar{N}_e(\theta), \theta). \quad (5)$$

$n_e$  is determined by minimizing the free-energy density,

$$f = \frac{1}{M_\theta V} \sum_{\theta} \min_{N_e} [E_0(N_e, \theta) - \mu N_e], \quad (6)$$

where the sum is over a uniform grid containing  $M_\theta$  twist angles. For any single-electron theory, the electronic density  $n_e(\mu)$  and the ground-state energy density  $e_0(\mu)$  coincide exactly with the corresponding thermodynamic limit values for a sufficiently large value of  $M_\theta$ , e.g., when the sum over twists becomes an integral over the Brillouin zone. Size effects remaining after twist averaging are due to electron-electron correlations.

Figure 1(a) illustrates  $e_0(\mu)$  and  $n_e(\mu)$  for solid molecular hydrogen, computed from Heyd-Scuseria-Ernzerhof (HSE) [31] functional and from QMC (see Sec. IV for details). The value of the band gap can be directly extracted from the width of the incompressible region. Alternatively, if we eliminate  $\mu$  in favor of  $n_e$ , and plot  $e_0$  as a function of  $n_e$  [as in Fig. 1(b)], the fundamental gap is obtained by the discontinuity of the derivative, according to Eq. (1).

The definition of the fundamental gap can apply to different symmetry sectors. For a perfect crystal, the total momentum of the electrons modulo reciprocal lattice vectors, i.e., the crystal momentum, is conserved. By requiring the total crystal momentum of the electrons to be fixed, e.g., using Bloch-type orbitals in the Slater determinant, the full band structure in the Brillouin zone can be mapped out. For a spin-independent Hamiltonian, one can also impose the total spin to determine the fundamental gap in each spin sector. In practice, the charge gap in the spinless sector can be determined by adding or removing pairs of electrons. The extensions of our definitions and formulas to this case are straightforward, e.g.,  $\Delta_{N_e} = [E_0(N_e + 2) + E_0(N_e - 2) - 2E_0(N_e)]/2$ . We follow this procedure of spin-neutral excitations in the remainder of the paper.

## III. FINITE SIZE EFFECTS

### A. Potential energy

A key quantity in understanding size effects is the long wavelength behavior of the static structure factor,  $S_{N_e}(\mathbf{k}) = \langle \rho_{-\mathbf{k}} \rho_{\mathbf{k}} \rangle / N_e$ , where  $\rho_{\mathbf{k}} = \sum_j e^{i\mathbf{k} \cdot \mathbf{r}_j}$  is the Fourier transform of the instantaneous electron density. The structure factor for a homogeneous system obeys the bound [1,18],

$$S_{N_e}(k) \leq \frac{\hbar k^2}{2m\omega_p} \left(1 - \frac{1}{\epsilon_k}\right)^{1/2}, \quad (7)$$

where  $\omega_p = 4\pi \hbar^2 e^2 n_e / m$  is the plasma frequency and  $\epsilon_k$  the static dielectric constant for wave vector  $k$  (to simplify the notations, we will suppress the dependence on the wave vector in the following). This inequality is derived by applying the plasmon-pole approximation to the sum rules of the dynamic structure factor  $S(k, \omega)$ . It implies that the structure factor must vanish quadratically as  $k \rightarrow 0$  [32]. Equality will be obtained if  $S(k, \omega)$  reduces to a single delta function at small  $k$ . The  $1/N_e$  finite-size corrections of the energy per electron is a direct consequence of this behavior of  $S_{N_e}(k)$  [28]. However, these leading order corrections are not sufficient for excitation

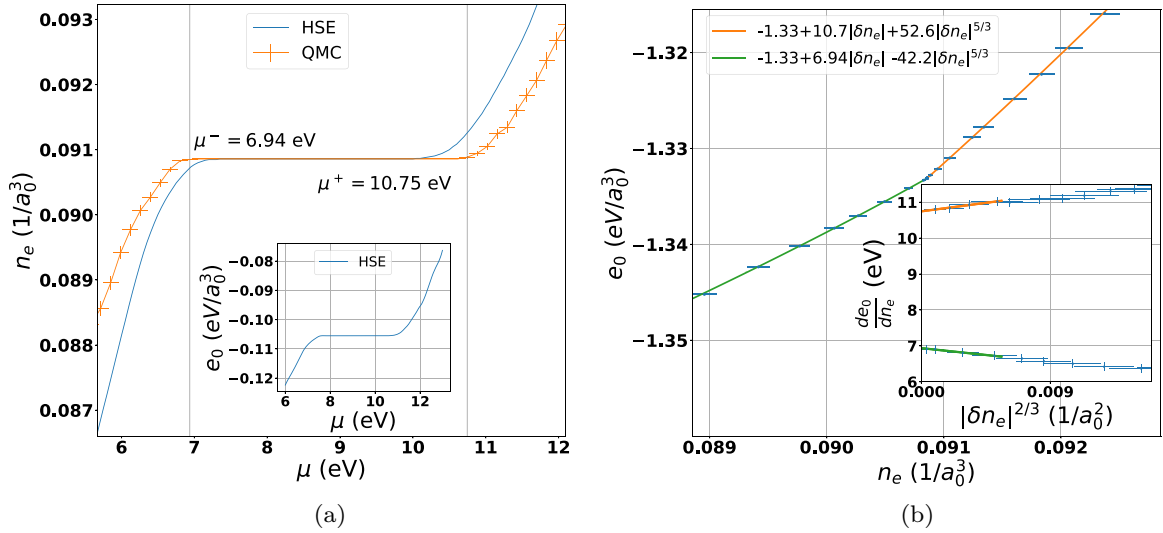


FIG. 1. GCTABC analyses of the C2/c-24 structure of solid hydrogen at  $r_s = 1.38$  (234GPa). (a) The electron density,  $n_e$ , as a function of the chemical potential  $\mu$  obtained from HSE functional in comparison to QMC; the inset illustrates the energy density,  $e_0$ , as a function of  $\mu$  from HSE functional. (b) Energy density,  $e_0$ , as a function of  $n_e$  using QMC; the inset shows the derivative discontinuity, where  $\delta n_e$  is the change of the electronic density with respect to the insulating state. Size corrections, as discussed in the text, are included.

energies, since the energy gap is of the same order as finite-size corrections to the total energy.

As we will show below, the key to understanding size effects of energy differences is encoded in the change of  $S_{N_e}(k)$  as electrons are added or removed. In particular, the limiting behavior of  $S_{N_e \pm 1}(k)$  as  $k \rightarrow 0$  will provide the dominant finite size correction.

For concreteness, we will assume a Slater-Jastrow form for the ground-state wave function  $\Psi_0 \propto D \exp[-U]$ . The determinant,  $D$ , is built out of Bloch orbitals,  $\phi_{\mathbf{q}n}(\mathbf{r})$  with  $\mathbf{q}$  inside the first Brillouin zone,  $n$  is the band index, and  $U$  is a general, symmetric  $n$ -body correlation factor [33]. For simplicity, we assume it is two body:  $U = \sum_{i < j} u(r_i, r_j)$ . Let us consider the action of  $e^{i\mathbf{k} \cdot \mathbf{r}_j}$  on a single-particle orbital  $\phi_{\mathbf{q}n}(\mathbf{r}_j)$  in the Slater determinant of the ground state. In the limit of small  $\mathbf{k}$ , this can be approximately written as  $\phi_{\mathbf{q}+\mathbf{k}n}(\mathbf{r}_j)$ . Expanding the determinant in terms of its cofactors  $\frac{\delta D}{\delta \phi_{\mathbf{q}n}(\mathbf{r}_j)}$  and making the excitation, we have

$$\rho_{\mathbf{k}} \Psi_0 \propto \sum_j \sum_{\mathbf{q}, n} \frac{\delta D}{\delta \phi_{\mathbf{q}n}(\mathbf{r}_j)} e^{i\mathbf{k} \cdot \mathbf{r}_j} \phi_{\mathbf{q}n}(\mathbf{r}_j) e^{-U}. \quad (8)$$

and the resulting determinant after summation over  $j$  vanishes for small  $k$  if the Bloch orbital  $(\mathbf{q} + \mathbf{k}, n)$  is already occupied in the ground-state determinant. Considering  $N_e \pm 1$  electron wave functions,  $\Psi_0(N_e \pm 1; \pm \mathbf{q}, m)$ , where  $N_e$  corresponds to the insulating state with fully occupied bands in the Slater determinant, and  $\mathbf{q}m$  denotes the additional particle/hole orbital, we get

$$\lim_{k \rightarrow 0} \rho_{\mathbf{k}} \Psi_0(N_e \pm 1; \mathbf{q}, m) \sim \pm \Psi_0(N_e \pm 1; \mathbf{q} + \mathbf{k}, m) \quad (9)$$

for  $k \neq 0$ , where different signs for particle or hole excitations on the right hand side (r.h.s.) are chosen to match the most common sign convention, e.g., of Ref. [34]. The limit  $k \rightarrow 0$  is discontinuous since  $\rho_{\mathbf{k}=0} \Psi_0(N_e \pm 1; \mathbf{q}, m) \equiv (N_e \pm 1) \Psi_0(N_e \pm 1; \mathbf{q}, m)$ .

Kohn [34,35] has pointed out that in the insulating state, the matrix elements

$$\lim_{\mathbf{q}' \rightarrow \mathbf{q}} \langle \Psi_0(N_e \pm 1; \mathbf{q}, m) | \rho_{\mathbf{q}-\mathbf{q}'} | \Psi_0(N_e \pm 1; \mathbf{q}', m) \rangle = \pm \frac{1}{\epsilon} \quad (10)$$

approach the inverse dielectric constant,  $\epsilon^{-1}$ , up to a sign.

Substituting Eq. (9) into Eq. (10) suggests the following finite-size behavior of the static structure factor of insulators

$$\lim_{k \rightarrow 0} S_k^\pm = \alpha_\pm + \mathcal{O}(k^2), \quad (11)$$

$$S_k^\pm \equiv (N_e \pm 1) S_{N_e \pm 1}(k) - N_e S_{N_e}(k), \quad (12)$$

where  $\alpha_\pm$  is proportional to  $\epsilon^{-1}$ . However,  $\alpha_\pm$  in general differs from  $\epsilon^{-1}$  unless Eq. (9) is an exact equality.

Figure 2 shows the behavior of  $S_k^\pm$  for carbon and silicon crystals. Note that these functions extrapolate to a nonzero value as  $k \rightarrow 0$ .

The long wavelength behavior of the structure factor, Eq. (11), then gives rise to size corrections to excitation energies through the potential energy term

$$\left[ \int \frac{d^3 k}{(2\pi)^3} - \frac{1}{V} \sum_{\mathbf{k} \neq 0} \right] \frac{v_{\mathbf{k}}}{2} S_k^\pm \simeq \alpha_\pm \frac{|v_M|}{2}, \quad (13)$$

where we have defined the Madelung constant as

$$v_M = \left[ \frac{1}{V} \sum_{\mathbf{k} \neq 0} - \int \frac{d^3 k}{(2\pi)^3} \right] v_{\mathbf{k}} \sim L^{-1} \sim N_e^{-1/3}. \quad (14)$$

For the Coulomb potential,  $v_M$  is proportional to  $L^{-1}$ , the inverse linear extension of the simulation cell. The negative proportionality constant depends on the boundary conditions, e.g., cell geometry, and can be calculated by the Ewald image technique [36].

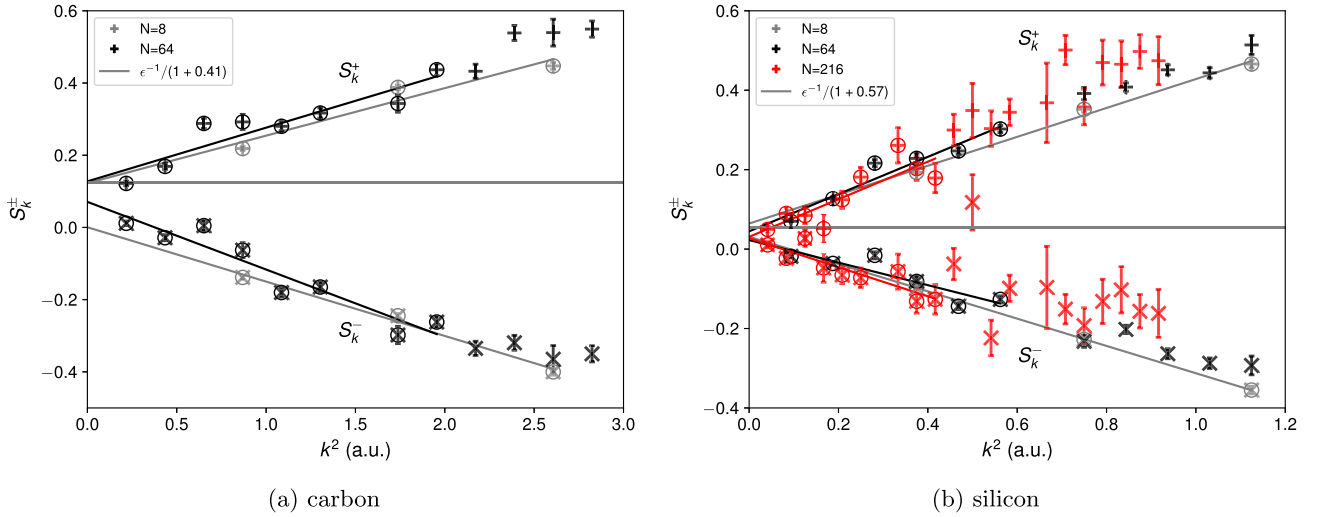


FIG. 2. Change in the static structure factor as an electron (upper curves) or a hole (lower curves) is added to the insulating system with  $N$  atoms. The lines are fits to the data points. The horizontal lines show the expected  $k \rightarrow 0$  limit based on the experimental dielectric constants. We have used  $c = 0.41$  for C and  $c = 0.57$  for Si.

### B. Kinetic energy

Following Ref. [29], we now discuss the kinetic energy contribution  $\hbar^2[\nabla U]^2/2m$  which arises from electron correlation. For a two-body Jastrow,  $U = \sum_{\mathbf{k}} u_{\mathbf{k}} \rho_{\mathbf{k}} \rho_{-\mathbf{k}}/2V$ , and we are only interested in the long-wavelength limit,  $k \rightarrow 0$ , of the electron-electron correlation, with wave vectors smaller than the reciprocal lattice vectors of the crystal,  $\mathbf{G}$ . Isolating the singular contributions involving  $\rho_{\mathbf{k}=0} \equiv N_e$  in the spirit of the rotating (random) phase approximation (RPA) we have

$$\begin{aligned} \langle [\nabla U]^2 \rangle &= -\frac{1}{V^2} \sum_{\mathbf{k} \neq 0, \mathbf{k}' \neq 0} (\mathbf{k} \cdot \mathbf{k}') u_{\mathbf{k}} u_{\mathbf{k}'} \langle \rho_{\mathbf{k}+\mathbf{k}'} \rho_{-\mathbf{k}} \rho_{-\mathbf{k}'} \rangle \\ &\simeq \frac{1}{V^2} \sum_{\mathbf{k} \neq 0} N_e k^2 u_{\mathbf{k}}^2 \langle \rho_{\mathbf{k}} \rho_{-\mathbf{k}} \rangle. \end{aligned} \quad (15)$$

Therefore, for systems with explicit long-range correlations  $u_{\mathbf{k}} \sim k^{-2}$ , the kinetic energy will also contribute to the leading order size corrections with

$$\left[ \int \frac{d^3k}{(2\pi)^3} - \frac{1}{V} \sum_{\mathbf{k} \neq 0} \right] \frac{n_e \hbar^2 k^2 u_{\mathbf{k}}^2}{2m} S_k^\pm \simeq \alpha_{\pm} c \frac{|v_M|}{2}, \quad (16)$$

where  $c = \lim_{k \rightarrow 0} n_e \hbar^2 k^2 u_{\mathbf{k}}^2 / (m v_k)$  is approximately given by the ratio of the  $1/N_e$  finite-size corrections of the kinetic to potential energy of the ground-state energy per particle due to two-body correlations [29].

### C. Total gap corrections from Coulomb singularity

Up to now, we have shown how the long-range behavior of the structure factor and Jastrow factor can give rise to a  $1/L$  correction to the excitation gap with a proportionality factor determined by the structure factor changes. In the following, we will further demonstrate that, given that the trial wave functions coincide with the exact ground-state wave function for  $N_e$  and  $N_e \pm 1$  electrons, this proportionality factor is

indeed given by the dielectric constant

$$\Delta_{\infty} - \Delta_V = \frac{|v_M|}{\epsilon} + \mathcal{O}\left(\frac{1}{V}\right), \quad (17)$$

as phenomenologically assumed in previous work [14,25].

We prove this by an independent argument based on commutation relations. Let us denote the exact insulating ground state of the  $N_e$  electron system as  $|\Psi_0^{N_e}\rangle$ , its energy as  $E_0^{N_e}$ , and the exact excited state of the  $N_e \pm 1$  electron system as  $|\Psi_{\mathbf{k}}^{N_e \pm 1}\rangle$  with energy  $E_{\mathbf{k}}^{N_e \pm 1}$ ;  $\mathbf{k}$  indicates that the additional/subtracted electron adds/subtracts the crystal momentum  $\mathbf{k}$ . We have

$$E_{\mathbf{k}}^{N_e+1} - E_0^{N_e} = \frac{\langle \Psi_{\mathbf{k}}^{N_e+1} | [H, a_{\mathbf{k}}^\dagger] | \Psi_0^{N_e} \rangle}{\langle \Psi_{\mathbf{k}}^{N_e+1} | a_{\mathbf{k}}^\dagger | \Psi_0^{N_e} \rangle} \quad (18)$$

for particle and

$$E_{\mathbf{k}}^{N_e-1} - E_0^{N_e} = \frac{\langle \Psi_{\mathbf{k}}^{N_e-1} | [H, a_{\mathbf{k}}] | \Psi_0^{N_e} \rangle}{\langle \Psi_{\mathbf{k}}^{N_e-1} | a_{\mathbf{k}} | \Psi_0^{N_e} \rangle} \quad (19)$$

for hole excitations. In second quantization, the Hamiltonian,  $H = T + V_{ee}$ , is given by

$$T = \sum_{\mathbf{k}} \left[ \frac{\hbar k^2}{2m} a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \sum_{\mathbf{G}} u(\mathbf{G}) a_{\mathbf{k}+\mathbf{G}}^\dagger a_{\mathbf{k}} \right], \quad (20)$$

$$V_{ee} = \frac{1}{2V} \sum_{\mathbf{q} \neq 0} v_{\mathbf{q}} [\rho_{\mathbf{q}} \rho_{-\mathbf{q}} - N_e], \quad (21)$$

where  $a_{\mathbf{k}}$  is the annihilation operator for plane-wave states of wave vector  $\mathbf{k}$ ,  $u(\mathbf{G})$  the periodic crystal potential, and  $v_{\mathbf{q}}$  is the Coulomb potential between electrons,  $\rho_{\mathbf{q}} = \sum_{\mathbf{k}} a_{\mathbf{k}+\mathbf{q}}^\dagger a_{\mathbf{k}}$ , and  $N_e = \sum_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{\mathbf{k}}$ .

The commutator involving the single-particle energy term is

$$[T, a_{\mathbf{k}}^\dagger] = \frac{\hbar^2 k^2}{2m} a_{\mathbf{k}}^\dagger + \sum_{\mathbf{G}} u(\mathbf{G}) a_{\mathbf{G}+\mathbf{k}}^\dagger. \quad (22)$$

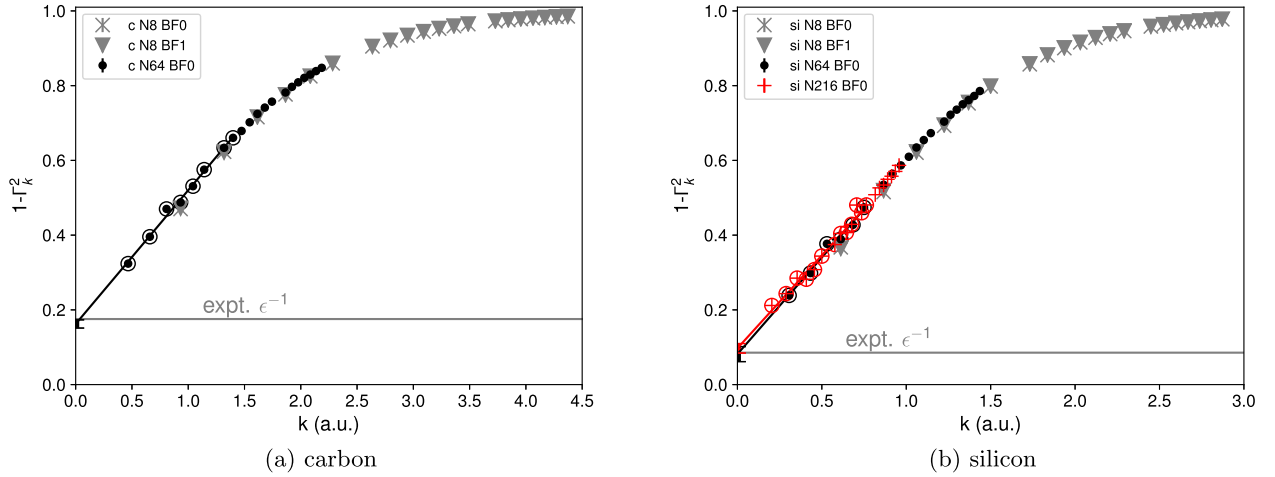


FIG. 3. Upper bound to the inverse dielectric constant Eq. (7), where  $\Gamma_k \equiv \frac{2m\omega_p S_{N_c}(k)}{\hbar k^2}$ . Lines are fits to the low- $k$  data. The horizontal lines mark experimental inverse dielectric constants.

There are corresponding terms for hole excitations, but none of these terms involve singular contributions responsible for anomalous size effects, so these terms do not contribute at leading order. However,

$$[V_{ee}, a_{\mathbf{k}}^\dagger] = \frac{1}{V} \sum_{\mathbf{q} \neq 0} v_{\mathbf{q}} [\rho_{\mathbf{q}} a_{\mathbf{k}-\mathbf{q}}^\dagger - 1] \quad (23)$$

and

$$[V_{ee}, a_{\mathbf{k}}] = -\frac{1}{V} \sum_{\mathbf{q} \neq 0} v_{\mathbf{q}} \rho_{\mathbf{q}} a_{\mathbf{k}+\mathbf{q}} \quad (24)$$

involve terms approaching the Coulomb singularity,  $v_{\mathbf{q}} \sim q^{-2} \rightarrow \infty$  for  $q \rightarrow 0$ .

From these terms, we get the leading order size corrections by noting that

$$\lim_{\mathbf{k}, \mathbf{q} \rightarrow 0} \frac{\langle \Psi_{\mathbf{k}}^{N_c+1} | \rho_{\mathbf{q}} a_{\mathbf{k}-\mathbf{q}}^\dagger | \Psi_0^{N_c} \rangle}{\langle \Psi_{\mathbf{k}}^{N_c+1} | a_{\mathbf{k}}^\dagger | \Psi_0^{N_c} \rangle} = \frac{1}{2} \left[ \frac{1}{\epsilon} + 1 \right] \quad (25)$$

and

$$\lim_{\mathbf{k}, \mathbf{q} \rightarrow 0} \frac{\langle \Psi_{\mathbf{k}}^{N_c-1} | \rho_{\mathbf{q}} a_{\mathbf{k}+\mathbf{q}} | \Psi_0^{N_c} \rangle}{\langle \Psi_{\mathbf{k}}^{N_c-1} | a_{\mathbf{k}} | \Psi_0^{N_c} \rangle} = -\frac{1}{2} \left[ 1 + \frac{1}{\epsilon} \right]. \quad (26)$$

Both relations can be obtained [37] by extending Kohn's diagrammatic approach [34] (see Supplemental Material [38]). Integrating around the  $v_{\mathbf{q}}$  singularity for small  $q$  in Eq. (23), we obtain the leading order finite size corrections. As before, this involves the Madelung constant, Eq (14). In the particle channel, we get  $\frac{v_M}{2}(\frac{1}{\epsilon} - 1)$  and in the hole channel,  $\frac{v_M}{2}(\frac{1}{\epsilon} + 1)$ . The corrections independent of  $\epsilon$  correspond to the change in the background charge which cancel for the fundamental gap and we obtain Eq. (17).

Previous, heuristic approaches [25] have suggested that one can use experimental or DFT values of the dielectric constant for finite-size extrapolation. Our approach further suggests that this value can be determined from the QMC structure factor extrapolated to zero wave vector

$$\frac{2}{\epsilon} \equiv (1+c) \lim_{N_c \rightarrow \infty} \lim_{k \rightarrow 0} [S_k^+ + S_k^-], \quad (27)$$

with the singular behavior of the Jastrow factor determining  $c$ . We emphasize that the order of the limits involved above is crucial.

An independent estimate is based on the inequality of Eq. (7). We can bound and estimate the value of the dielectric constant using the structure factor of the insulating ground state. By extrapolating  $1 - \Gamma_k^2$  vs  $k$  to  $k = 0$  we obtain an upper bound to the inverse dielectric constant, where  $\Gamma_k \equiv 2m\omega_p S_{N_c}(k)/\hbar k^2$ . This involves only the extensive part of the density-density correlations, thus, it is less sensitive to noise and has much smaller statistical uncertainty. In Fig. 3, we show that for C and Si, this inequality gives accurate values of the dielectric constant.

#### D. Twist correction of two-particle correlations

The above size effects explain the leading order  $1/L$  correction to the single-particle gap. However, as we will see in our results, the asymptotic region, where this law can be reliably applied, may still be difficult to reach for currently used system sizes and next-to-leading order effects are important. Here, we show that an important part can be corrected for, by further restoring the full symmetry properties in the contribution of the direct Coulomb interaction.

For inhomogeneous systems, it is convenient to separate the mean density from its fluctuating components in the static structure factor [29], i.e.,

$$S_{N_c}(\mathbf{k}) = \frac{1}{N_c} \langle \rho_{\mathbf{k}} \rangle \langle \rho_{-\mathbf{k}} \rangle + \delta S_{N_c}(\mathbf{k}), \quad (28)$$

$$\delta S_{N_c}(k) = \frac{1}{N_c} \langle (\rho_{\mathbf{k}} - \langle \rho_{\mathbf{k}} \rangle) (\rho_{-\mathbf{k}} - \langle \rho_{-\mathbf{k}} \rangle) \rangle. \quad (29)$$

For crystals with periodic density distributions, the Fourier components of the mean density,  $\langle \rho_{\mathbf{k}} \rangle$ , only contribute for reciprocal lattice vectors,  $\mathbf{k} \in \mathbf{G}$ . The long wavelength behavior of the structure factor is entirely due to the fluctuating part  $\delta S_{N_c}(k)$ , which therefore contains the leading order size effects [29]. However, the mean single particle density,  $\langle \rho(\mathbf{r}) \rangle = V^{-1} \sum_{\mathbf{k}} \langle \rho_{\mathbf{k}} \rangle e^{i\mathbf{k}\cdot\mathbf{r}}$ , of the finite system may significantly differ

from the infinite one, particularly in cases where the supercell is not compatible with the full symmetry group of the crystal.

Averaging over twisted boundary conditions is designed to restore the symmetry of the crystal and thus accelerate the convergence of single-particle densities to the thermodynamic limit. In the following, we denote the twist averaged expectation value by

$$\bar{O} \equiv \frac{1}{M_\theta} \sum_{\theta} \langle O \rangle_{N_e, \theta}, \quad (30)$$

where we have explicitly indicated the  $N_e$  and  $\theta$  dependence on the expectation value on the r.h.s. For any single-particle theory,  $\rho(\mathbf{r})$  approaches its thermodynamic limit for calculations at fixed  $N_e$  by averaging over a dense grid of twist angles ( $M_\theta \rightarrow \infty$ ). Within many-body calculations, twist-averaging [30] takes over a large part of this property to any observable *linear* in the density. Here, we extend this approach to also correct the quadratic expression entering the two-body contributions of the total energy.

For the potential energy, this correction to the twist-converged QMC calculation is

$$\delta V_{N_e}^s = \frac{1}{2V} \sum_{\mathbf{k}} v_{\mathbf{k}} \delta C(\mathbf{k}), \quad \delta C(\mathbf{k}) = \overline{\rho_{\mathbf{k}} \rho_{-\mathbf{k}}} - \overline{\rho_{\mathbf{k}} \rho_{-\mathbf{k}}}. \quad (31)$$

For the ground-state energies, this correction provides only a small improvement over our previous correction [28,29].

For the gap, many terms entering Eqs. (31) cancel and the expression can be simplified. Let us consider the case of adding/removing one electron at twist  $\phi$  to the insulating ground state, denoting  $\Pi_{\mathbf{k}}^{\pm}$  the difference of the respective densities:

$$\Pi_{\mathbf{k}}^{\pm} \equiv \langle \rho_{\mathbf{k}} \rangle_{N_e \pm 1, \phi} - \langle \rho_{\mathbf{k}} \rangle_{N_e, \phi}. \quad (32)$$

In the thermodynamic limit, the density of the ground-state system with  $N_e$  electrons coincides with the twist-averaged ground-state density  $\overline{\rho_{\mathbf{k}}}$ , whereas we obtain  $\overline{\rho_{\mathbf{k}}} + \Pi_{\mathbf{k}}^{\pm}$  for the density of the  $N_e \pm 1$  electron system. Inserting into Eqs. (31), we obtain the correction for the difference between the two states,

$$\delta V_{N_e \pm 1, \phi}^s - \delta V_{N_e}^s = \frac{1}{V} \sum_{\mathbf{k} \in \mathbf{G}} v_{\mathbf{k}} \text{Re}[(\overline{\rho_{\mathbf{k}}} - \langle \rho_{\mathbf{k}} \rangle_{N_e, \phi}) \Pi_{-\mathbf{k}}^{\pm}], \quad (33)$$

where only wave vectors of the reciprocal crystal lattice contribute to the sum. The corresponding finite size correction for the gap, denoted by  $\delta \Delta_s$  in the following, is of order  $1/N_e$  or smaller, mainly determined by the changes of the ground-state densities at the first Bragg peaks due to twist averaging.

Equation (33) can be understood quite intuitively: It corrects the direct Coulomb interaction between the electron/hole in the excited state ( $\Pi^{\pm}$ ) with the unexcited electrons. The density of those electrons is expected to change by  $\overline{\rho_{\mathbf{k}}} - \langle \rho_{\mathbf{k}} \rangle_{N_e, \phi}$  in the thermodynamic limit.

Converged ground-state densities are naturally calculated within GCTABC. It is straightforward to apply the correction Eq. (33) to all excitation energies. Alternatively, the corresponding DFT densities may be used. This removes the stochastic error at the cost of introducing a small bias in the next-to-leading order size correction.

#### IV. COMPUTATIONAL METHODS

We have performed electronic QMC calculations on three insulating solids: molecular hydrogen at high pressure, and carbon and silicon in the diamond structure at zero pressure. Since we are interested in the spin-neutral charge gap, we used an equal number of spin-up and spin-down electrons. We used a Slater-Jastrow trial wave function with backflow (BF) corrections [39,40]. The Jastrow and BF functions were fully optimized within variational Monte Carlo, including the long-range (reciprocal lattice) contributions. The orbitals in the Slater determinant were taken from DFT calculations using QUANTUM ESPRESSO [41,42]. The carbon and silicon orbitals were generated using the LDA functional, whereas the hydrogen orbitals were generated using the PBE functional, which has been shown to provide a good trial QMC wave function [43,44].

Molecular hydrogen was placed in the C2/c-24 structure [45] at two different densities ( $r_s = 1.38$  and  $r_s = 1.34$ ), roughly corresponding to pressures of 234GPa and 285 GPa, respectively. Energies and structure factors were obtained from reptation Quantum Monte Carlo calculations using the BOPIMC code [46]. For carbon and silicon, DMC calculations have been performed with the QMCPACK code [47] at the experimentally measured zero pressure valence densities,  $r_s = 1.318$  and  $r_s = 2.005$ , respectively. The crystal structures were optimized by DFT using the vdW-DF1 functional. For hydrogen, the QMC calculations have been done with the bare Coulomb interaction. The PAW pseudopotential has been used for the DFT results shown in Fig. 1. For carbon and silicon, pseudopotentials were used to remove the core electrons: carbon ions modeled by the Burkatzki-Filippi-Dolg pseudopotential [48], and silicon ions by the Trail-Needs pseudopotential [49]. These are considered good pseudopotentials for correlated calculations, but their use within DFT calculations produces slightly different results from the literature even with the same functional. For hydrogen, we used a supercell with  $2 \times 2 \times 1$  primitive cells so the supercell is nearly cubic and contained 96 protons. For carbon, we used two system sizes: the cubic cell containing eight atoms and a  $2 \times 2 \times 2$  supercell containing 64 atoms. For silicon, in addition to these systems, we used a  $3 \times 3 \times 3$  supercell containing 216 atoms. For hydrogen, twist convergence has been achieved using a  $8 \times 8 \times 8$  twist grid. For C and Si, the twist grid density decreases with increasing system size. The Supplemental Material [38] contains the QMC-calculated energies and variances of the insulating ground states of the various systems obtained after twist averaging and two-body finite-size corrections.

#### V. RESULTS

For any single-particle theory, such as Kohn-Sham DFT, the densities and energies,  $n_e(\mu)$  and  $e_0(\mu)$ , are obtained by occupying all single-particle states below the chemical potential  $\mu$ . By construction, the gap, as determined from the incompressible region of  $n_e(\mu)$  or from the discontinuity in the derivative of  $de_0/dn_e$  (see Fig. 1), then coincides with the one obtained from the band structure.

The LDA band gaps of carbon and silicon in the diamond structure are indirect and lie along the  $\Gamma X$  direction where

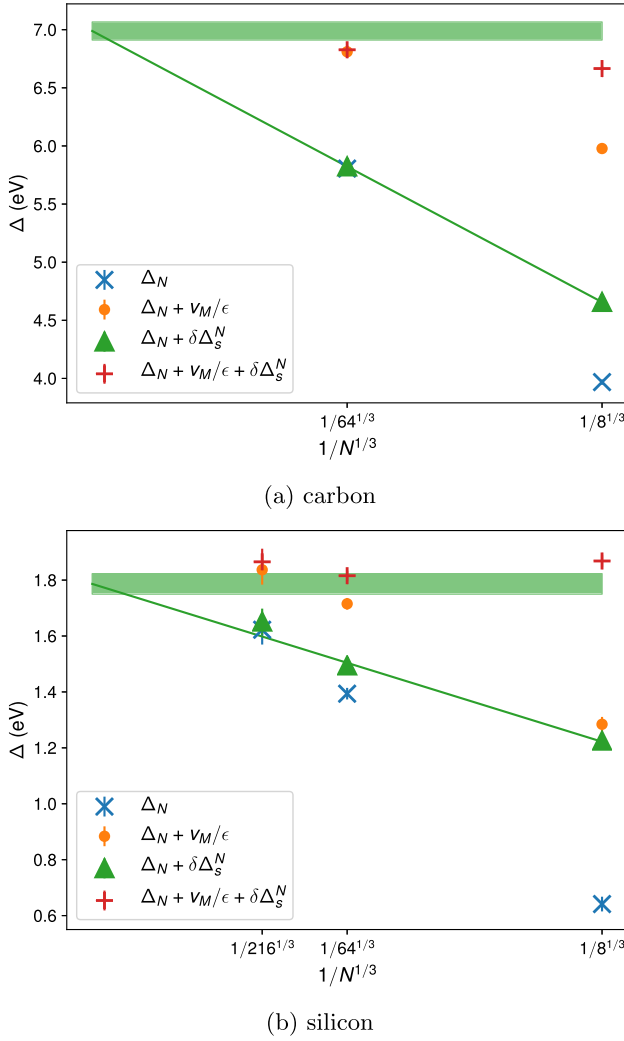


FIG. 4. Fundamental gap before and after finite-size corrections.  $\Delta_N$  is the DMC gap from a simulation with  $N$  atoms in the supercell without any finite-size correction,  $v_M/\epsilon$  is the leading-order Madelung correction using the experimental value of  $\epsilon^{-1}$ ,  $\delta\Delta_s^N$  is the next-to-leading-order density correction, which is related to the static part of the structure factor. The line is a fit to  $\Delta_N + \delta\Delta_s^N$ .

$\Gamma$  is the origin of the Brillouin zone and  $X$  the Brillouin zone boundary in the (100) direction. By looking directly at the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) states with LDA, it is found that the carbon gap is 3.89 eV and the silicon gap is 0.34 eV. The bands immediately above and below the gap can be fit to a quadratic form which implies  $e_0(\mu) = \mu^\pm n_e(\mu) + b^\pm n_e(\mu)^{5/3}$ . Therefore, the derivative  $de_0/dn_e = \mu^\pm + \frac{5b^\pm}{3} n_e^{2/3}$  has a discontinuity at  $n_e = 0$  and behaves as  $n_e^{2/3}$  above and below the gap. Applying our GCTABC procedure to a single-particle theory, all states with energies below the chemical potential are occupied. Varying the chemical potential thus scans the underlying density of states. The band gap is then determined by locating the band edges,  $\mu^\pm$ , disregarding the location in the Brillouin zone [50]. Figure 5 illustrates the density of states obtained from GCTABC giving an LDA gap of 3.95 eV for the carbon gap and 0.38 eV for the silicon gap. The small differences ( $\sim 0.05$  eV) from the values

TABLE I. Energy gaps obtained from GCTAB QMC in eV. The bare gap,  $\Delta_N$ , was calculated from Eq. (1) for a finite supercell containing  $N$  atoms. The leading-order finite-size corrections are given by the screened Madelung constants  $|v_M|/\epsilon$ , the next-to-leading order by the twist correction of two-particle density correlations,  $\delta\Delta_s$ . We used the experimental value of  $\epsilon$  for C and Si (5.7 and 11.7, respectively) and the value 18.8 for  $H_2$  extracted from  $S(k)$ . Finite-size corrections were also applied to the band edges,  $\mu^\pm$ . The estimate of the gap in the thermodynamic limit is  $\Delta_\infty = \Delta_{N_c} + |v_M|/\epsilon + \delta\Delta_s$ . From our LDA analysis, we estimate a systematic bias of  $\sim 0.1$  eV from the finite twist grid. This bias is larger than the statistical error. SJ indicates Slater-Jastrow trial wave function, while BF indicates backflow. The lattice constants of carbon and silicon are 3.567 Å and 5.43 Å, respectively.

	$r_s$	$N$	$\Delta_N$	$ v_M /\epsilon$	$\delta\Delta_s$	$\mu_\infty^-$	$\mu_\infty^+$	$\Delta_\infty$
$H_2$ (BF)	1.38	96	3.3(1)	0.40	0.020	6.9(1)	10.7(1)	3.8(1)
	1.34	96	2.4(1)	0.20	0.018	8.6(1)	11.2(1)	2.6(1)
C (BF)	1.318	8	3.9(1)	2.01	0.69	11.5(1)	18.1(1)	6.6(1)
	C (SJ)	1.318	8	4.0(1)	2.01	0.69	11.5(1)	18.2(1)
		64	5.8(1)	1.00	0.02	11.9(1)	18.7(1)	6.8(1)
Si (BF)	2.005	8	0.6(1)	0.64	0.55	5.2(1)	6.9(1)	1.7(1)
Si (SJ)	2.005	8	0.6(1)	0.64	0.58	5.2(1)	7.0(1)	1.9(1)
		64	1.4(1)	0.32	0.08	5.5(1)	7.3(1)	1.8(1)
		216	1.6(1)	0.21	0.01	5.6(1)	7.4(1)	1.8(1)

obtained before are due to the finite resolution of the twist grid, and can be controlled by using denser grids.

As can be seen in the same figure, the effective band edge densities of states from GCTABC-DMC have a similar functional form, but with a larger gap than the DFT ones. The QMC computed gaps for the different sizes of the supercell are summarized in table I. The results from different supercells clearly show the important bias on gap introduced by the finite size of the supercell. In Fig. 4, we show the bare gap,  $\Delta_N$ , the Madelung-corrected one,  $\Delta_N + |v_M|/\epsilon$ , and our best correction,  $\Delta_\infty = \Delta_N + |v_M|/\epsilon + \delta\Delta_s$ , for both systems against the linear size of the supercell, where  $N$  is the number of atoms in the supercell and  $\epsilon$  is the experimental value of the dielectric constant. We see that the next-to-leading-order corrections are comparable to the leading-order one, in particular for the eight-atom supercell of Si, whereas they rapidly decay for the larger sizes.

The finite-size corrected values,  $\Delta_\infty$ , of all different sizes C and Si supercells agree with each other within the statistical uncertainty, yielding the DMC-SJ values  $\Delta_\infty = 6.8(1)$  and  $\Delta_\infty = 1.8(1)$  for the C and Si gap, respectively. We further note that these values also agree with a numerical  $N^{-1/3}$  extrapolation of the gap values corrected by  $\delta\Delta_s$ . For any numerical  $N^{-1/3}$  extrapolation, it is very important to reduce any bias due to higher order corrections as much as possible, since the outcome of a fit is sensitive to the smallest system sizes since they have the smallest statistical uncertainty. For Si, a  $N^{-1/3}$  extrapolation of the bare  $\Delta_N$  values yields an overestimation of 0.3 eV compared to  $\Delta_\infty$ .

Since our finite-size corrected gaps show size convergence for the smallest system size, it is now feasible to address the systematic error due to the fixed node approximation. To reduce this bias, we have added BF correlations in the

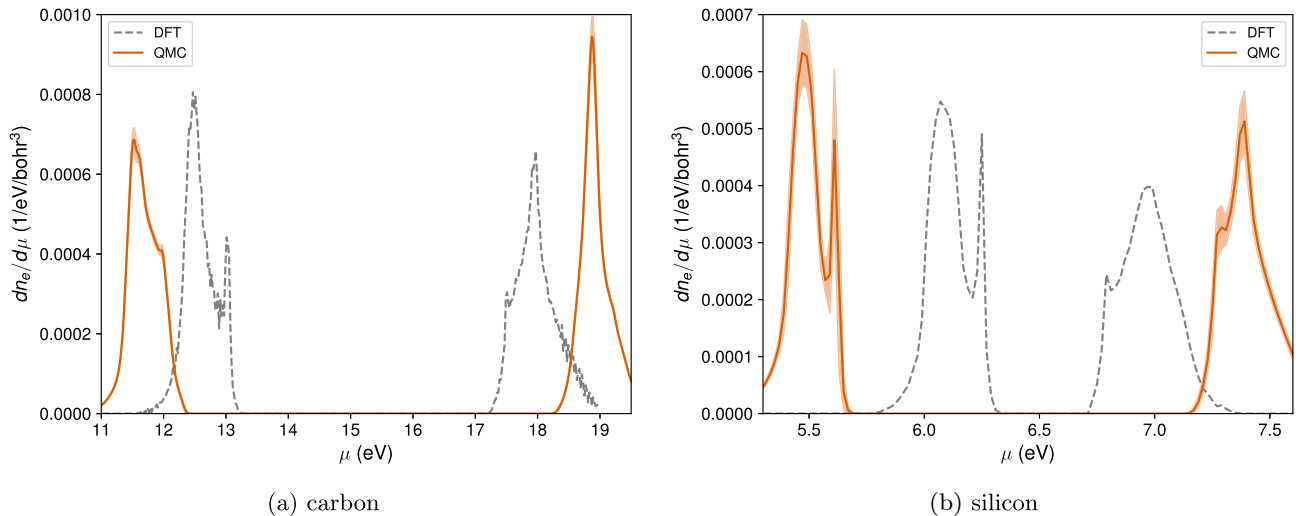


FIG. 5. Density of states for carbon and silicon near the band edge. Each plot shows the derivative of the mean electron density with respect to the chemical potential. This is the electronic density of states (DOS) in DFT, so the gap appears as a depleted region. The calculated DOS is only valid near the band edge because only the two bands closest to the gap are considered within DFT and QMC. The DFT bands (done in a primitive cell) have been folded into the Brillouin zone (BZ) of the 64-atom supercell to allow comparison with QMC.

Slater orbitals. Our BF correlations lower the SJ gap by 0.1 eV for both, C and Si. Previous BF calculations [25] on Si have reported a 0.2 eV lowering compared to SJ. The difference might be due to a different functional form or optimization procedure. A systematic study on the bias of the fixed-node approximation such as done with more general BF correlations [51,52] or multideterminant trial wave functions [53], possible for small supercells, could be done in the future.

So far, in our analysis of C and Si, we have imposed the experimentally known dielectric constant in the leading order Madelung correction. As described in Sec. III, there is no need for any external knowledge to perform the size extrapolation as the value of the Madelung correction can be obtained from the behavior of the static structure factor, calculable within the same QMC run, see Figs. 2 and 3. However, since the extrapolation involved introduces an additional uncertainty, we have preferred to use the experimental values to benchmark our theory and better distinguish leading from next-to-leading order size effects.

Using the dielectric bound Eq. (7) on the ground-state structure factor to determine  $\epsilon$ , we get  $\epsilon_0 = 6.2 \pm 0.4$  for C and  $\epsilon_0 = 10.3 \pm 1.3$  for Si, which are compatible with the experimental values of 5.7 and 11.7. The corresponding leading-order finite-size corrections on the gap of the 64-atom system are then  $0.92 \pm 0.06$  eV for C and  $0.36 \pm 0.14$  eV for Si using the *ab initio*  $\epsilon^{-1}$ , as opposed to 1.00 eV for C and 0.32 eV for Si based on the experimental values of  $\epsilon^{-1}$ .

As shown in Fig. 2, the asymptotic values of the finite-sized structure factors,  $S_k^\pm$ , are affected by a much larger uncertainty, introducing larger systematic bias when used for *ab-initio* size corrections. Still, already the extrapolation to a nonzero value fixes the leading order size corrections to decay as  $1/L$ . This information alone can be crucial as calculations for only two different supercell sizes will be sufficient to determine size effects, whereas more supercell sizes would be needed if the asymptotic form was not known.

We have also computed the band gap of solid hydrogen using GCTABC in BF-RQMC calculations for one of the possible molecular structures predicted for phase III: C2/c-24 at  $r_s = 1.38$  and  $r_s = 1.34$  (roughly corresponding to pressures of 234 and 285 GPa, respectively). The results, in Table I, show that the gap and size effects decrease with increasing pressure. For these calculations, we use calculations for one supercell and use its structure factor to estimate the dielectric constant. From Fig. 1, we see that HSE DFT slightly underestimates the gap; however, the deviations from the plateau on both sides are quite similar.

## VI. COMPARISON WITH EXPERIMENT

Our best values for the fundamental electronic gap (BF-DMC) significantly overestimate the experimentally measured values for C and Si by 1.1 and 0.5 eV, respectively as shown in Table II. There are two main sources of systematic errors which need to be taken into account: the use of pseudopotentials and the neglect of electron-phonon coupling.

The QMC values for C and Si presented above are based on pseudopotentials to replace the core electrons of the atoms. Pseudopotentials are usually designed for accurate prediction of static structural quantities. Excitation spectra, in particular, the single-particle excitation gap, may be less well described.

TABLE II. Extrapolated band gap of Si and C from backflow DMC calculations,  $\Delta_{\text{BF}}$  compared to the experimental values (exp). We tabulated two main corrections: the difference between the gap of an all-electron (AE) and the pseudopotential (PP) calculation within GW calculations, and the neglect of electron-phonon coupling (e-ph).

	$\Delta_{\text{BF}}$	AE - PP	e-ph	exp
C	6.6(2)	-0.26 ( $G_0W_0$ ) [54]	-0.6 (GW) [56]	5.48 [67]
Si	1.7(1)	-0.25 ( $G_0W_0$ ) [54]	-0.06 (DFT) [57]	1.17 [67]



This has been found in many-body perturbation theory calculations within the *GW* framework where all-electron calculations have been shown to lower the gap of C and Si by  $\sim -0.3$  eV [54,55] with respect to pseudopotential calculations. Although the actual pseudopotentials of our QMC simulations differ from those used in the *GW* calculations, we expect that our QMC values will be shifted by a similar amount; we can roughly transfer the all-electron correction of *GW* to our QMC results.

For lighter atoms, electron-phonon coupling leads to a further reduction of the gap values, even at zero temperature, due to the presence of zero point motion of the ions in the crystal. For C, *GW* predicts a significant lowering of the gap by  $-0.6$  eV [56], whereas a smaller shift between  $-60$  meV [57] and  $-0.1$  eV [58] is expected from DFT for Si. The effect of thermal expansion is to lower the gap by about  $0.01$  eV at room temperature for both carbon [59,60] and silicon [61,62], beyond the resolution of present calculations.

Considering both, the bias due to the pseudopotential approximation and the neglect of electron-phonon coupling, our BF-DMC calculations for C and Si overestimate the gap by  $\sim 0.1 - 0.2$  eV (see Table II), larger than our statistical uncertainty. This remaining offset to experiment may either be due to residual bias of the fixed-node approximation, or due to effects in pseudopotential and e-ph coupling beyond our simple estimations based on *GW* and DFT. They could be addressed by more accurate calculations in the future.

For hydrogen, we do not compare to experiment since electron-phonon coupling is expected to be very large, and the experimental results are not precise. If we do not make size corrections, our results are comparable to the Slater-Jastrow DMC calculations of Ref. [27], where the DFT band structure was corrected by a “scissor operator” based on QMC runs at the  $\Gamma$  point of the supercell. However, no size effects were observed within the statistical error in Ref. [27], so their extrapolated results differ from ours by  $0.3 - 0.8$  eV ( $3.0$  and  $2.3$  eV for  $250$  and  $300$  GPa). Comparison to *GW* values are also not conclusive: Whereas Ref. [63] provides smaller values of the gaps ( $1.8$  and  $1.0$  eV for  $250$  and  $300$  GPa), the results of Ref. [64] ( $3.7$  and  $2.8$  eV for  $250$  and  $313$  GPa) are close to our predictions. However, we note that the *GW* calculations were done with slightly different crystal structures. In Ref. [63], the PBE functional was used to optimize the lattice structure in contrast to the vdW-DF1 functional of Ref. [64], shown to be the most accurate functional at this density [65]. The smaller gap can then be seen as a consequence of a larger bond length as it was shown that structures optimized with PBE functional have a larger bond length than the ones with vdW-DF1 [64]. We have recently completed a more detailed analysis of the band gap of molecular hydrogen [66] using the method introduced here. This discusses extension to disorder coming from nuclear quantum and thermal effects.

## VII. CONCLUSIONS

We have introduced a method to calculate the fundamental gap of insulators and semiconductors using QMC. Using grand-canonical twist averaging, the value of the gap can be determined at any point in the Brillouin zone whether the system has a direct or indirect gap. Although it is possible to map

out the whole band structure, we have focused on the minimal, fundamental gap in this paper. We have shown that for charged systems, finite size supercell calculations are necessarily biased by a finite size error decaying as  $1/L$ , where the prefactor is determined by the absolute value of the Madelung constant and the inverse dielectric constant. We have pointed out that the  $1/L$  functional form is encoded in the long wavelength behavior of the finite size structure factor extrapolating to a nonvanishing value at the origin. Next-to-leading order effects can be corrected by proper use of twist-averaging in the two-particle part of the static Coulomb potential.

We have applied this procedure to determine the fundamental gap of molecular hydrogen at high pressure and carbon and silicon in the diamond structure at zero pressure. Our finite-size corrected gap values for carbon and silicon are larger than the experimental ones. We have argued that the bias may be due to the pseudopotential approximation and the neglect of electron-phonon coupling.

We note that this procedure is not restricted to QMC calculations, but can be applied within any method which calculates the many-body wave functions and ground-state energies, e.g., for coupled cluster methods [8]. Our results for C and Si demonstrate that the bias due to the finite size supercell can be corrected for, so precise values in the thermodynamic limit can be obtained for small supercells without need for numerical extrapolation.

The procedure here has been developed for perfect crystals but can be generalized to systems with disorder, either due to thermal or quantum effects. Furthermore, the procedure provides a starting point to address optical—i.e., charge neutral—excitations. Although neutral excitations are expected to be less sensitive to finite-size effects, recent calculations [24,25] have observed the same slow  $1/L$  decay for the optical gap. Since it is often not practical to perform calculations for more than two significantly different supercell sizes, our method suggests that the asymptotic behavior of the structure factor provides the needed insight to whether  $1/L$  or  $1/L^3$  should be used as a functional form for the size extrapolation.

## ACKNOWLEDGMENTS

D.M.C. and M.H. thank Center for Computational Quantum physics (CCQ) at the Flatiron Institute for support during the writing of this paper and the Fondation NanoSciences (Grenoble) for support. The Flatiron Institute is a division of the Simons Foundation. D.M.C. and Y.Y. were supported by DOE Grant No. NA DE-NA0001789. V.G. and C.P. were supported by the Agence Nationale de la Recherche (ANR) France, under the program Accueil de Chercheurs de Haut Niveau 2015, project HyLightExtreme. Computer time was provided by the PRACE Project No. 2016143296, IS-CRAB, the high-performance computer resources from Grand Equipement National de Calcul Intensif (GENCI) Allocations No. 2018-A0030910282, and by an allocation of the Blue Waters sustained petascale computing project, supported by the National Science Foundation (Award No. OCI 07- 25070) and the State of Illinois, and by GRICAD infrastructure (Grenoble) within the Equip@Meso Project No. ANR-10-EQPX-29-01.

- [1] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons* (Cambridge University Press, Cambridge, 2016).
- [2] J. P. Perdew, *Int. J. Quantum Chem.* **28**, 497 (1985).
- [3] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- [4] J. Kolorenč and L. Mitas, *Rep. Prog. Phys.* **74**, 026502 (2011).
- [5] L. K. Wagner and D. M. Ceperley, *Rep. Prog. Phys.* **79**, 094501 (2016).
- [6] M. Ruggeri, P. L. Ríos, and A. Alavi, *Phys. Rev. B* **98**, 161105(R) (2018).
- [7] J. J. Shepherd and A. Grüneis, *Phys. Rev. Lett.* **110**, 226401 (2013).
- [8] T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grüneis, *Phys. Rev. X* **8**, 021043 (2018).
- [9] P. R. C. Kent, R. Q. Hood, M. D. Towler, R. J. Needs, and G. Rajagopal, *Phys. Rev. B* **57**, 15293 (1998).
- [10] P. R. C. Kent, R. Q. Hood, A. J. Williamson, R. J. Needs, W. M. C. Foulkes, and G. Rajagopal, *Phys. Rev. B* **59**, 1917 (1999).
- [11] C. Mitra, J. T. Krogel, J. A. Santana, and F. A. Reboredo, *J. Chem. Phys.* **143**, 164710 (2015).
- [12] J. McMinis, M. A. Morales, D. M. Ceperley, and J. Kim, *J. Chem. Phys.* **143**, 194703 (2015).
- [13] G. Makov and M. C. Payne, *Phys. Rev. B* **51**, 4014 (1995).
- [14] G. E. Engel, Y. Kwon, and R. M. Martin, *Phys. Rev. B* **51**, 13538 (1995).
- [15] A. J. Williamson, R. Q. Hood, R. J. Needs, and G. Rajagopal, *Phys. Rev. B* **57**, 12140 (1998).
- [16] L. Mitas and R. M. Martin, *Phys. Rev. Lett.* **72**, 2438 (1994).
- [17] M. D. Towler, R. Q. Hood, and R. J. Needs, *Phys. Rev. B* **62**, 2330 (2000).
- [18] D. M. Ceperley and B. J. Alder, *Phys. Rev. B* **36**, 2092 (1987).
- [19] J. Kolorenč and L. Mitas, *Phys. Rev. Lett.* **101**, 185502 (2008).
- [20] F. Ma, S. Zhang, and H. Krakauer, *New J. Phys.* **15**, 093017 (2013).
- [21] L. K. Wagner and P. Abbamonte, *Phys. Rev. B* **90**, 125129 (2014).
- [22] J. Yu, L. K. Wagner, and E. Ertekin, *J. Chem. Phys.* **143**, 224707 (2015).
- [23] H. Zheng and L. K. Wagner, *Phys. Rev. Lett.* **114**, 176401 (2015).
- [24] T. Frank, R. Derian, K. Tokár, L. Mitas, J. Fabian, and I. Štich, *Phys. Rev. X* **9**, 011018 (2019).
- [25] R. J. Hunt, M. Szyniszewski, G. I. Prayogo, R. Maezono, and N. D. Drummond, *Phys. Rev. B* **98**, 075122 (2018).
- [26] W. M. C. Foulkes, R. Q. Hood, and R. J. Needs, *Phys. Rev. B* **60**, 4558 (1999).
- [27] S. Azadi, N. D. Drummond, and W. M. C. Foulkes, *Phys. Rev. B* **95**, 035142 (2017).
- [28] S. Chiesa, D. M. Ceperley, R. M. Martin, and M. Holzmann, *Phys. Rev. Lett.* **97**, 076404 (2006).
- [29] M. Holzmann, R. C. Clay, III, M. A. Morales, N. M. Tubman, D. M. Ceperley, and C. Pierleoni, *Phys. Rev. B* **94**, 035126 (2016).
- [30] C. Lin, F.-H. Zong, and D. M. Ceperley, *Phys. Rev. E* **64**, 016702 (2001).
- [31] J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003).
- [32] D. Pines and P. Nozières, *The Theory of Quantum Liquids*, Vol. I (Addison-Wesley, Redwood City, CA, 1989).
- [33] M. Holzmann, B. Bernu, and D. M. Ceperley, *Phys. Rev. B* **74**, 104510 (2006).
- [34] W. Kohn, *Phys. Rev.* **110**, 857 (1958).
- [35] W. Kohn, *Phys. Rev.* **105**, 509 (1957).
- [36] P. P. Ewald, *Ann. Phys.* **369**, 253 (1921).
- [37] We can adapt the proof of Kohn [34] by noting that the extra particle (hole) propagator does not interact with the other particles before  $t = 0$ . The equivalent graphs of class I contribute with  $\pm 1$  for particle and hole excitations, the equivalent graphs of class IIA remain, contributing  $\pm(1/\epsilon - 1)/2$ , whereas graphs of class IIB are absent.
- [38] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.101.085115> for diagrammatic proof and QMC data.
- [39] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, *Phys. Rev. E* **68**, 046707 (2003).
- [40] C. Pierleoni, K. T. Delaney, M. A. Morales, D. M. Ceperley, and M. Holzmann, *Comp. Phys. Comm.* **179**, 89 (2008).
- [41] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. Fabris, G. Fratesi, S. de Gironcoli, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [42] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr., A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncè, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [43] M. A. Morales, R. Clay, C. Pierleoni, and D. M. Ceperley, *Entropy* **16**, 287 (2014).
- [44] C. Pierleoni, M. A. Morales, G. Rillo, M. Holzmann, and D. M. Ceperley, *Proc. Nat. Acad. Sci.* **113**, 4953 (2016).
- [45] C. J. Pickard and R. J. Needs, *Nat. Phys.* **3**, 473 (2007).
- [46] C. Pierleoni and D. M. Ceperley, *Lecture Notes Phys.* **703**, 641 (2006).
- [47] J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, M. Casula, D. M. Ceperley, S. Chiesa, B. K. Clark, R. C. Clay, III, K. T. Delaney, M. Dewing, K. P. Esler, H. Hao, O. Heinonen, P. R. C. Kent, J. T. Krogel, I. Kylänpää, Y. W. Li, M. G. Lopez, Y. Luo, F. D. Malone, R. M. Martin, A. Mathuriya, J. McMinis, C. A. Melton, L. Mitas, M. A. Morales, E. Neuscamman, W. D. Parker, S. D. P. Flores, N. A. Romero, B. M. Rubenstein, J. A. R. Shea, H. Shin, L. Shulenburger, A. F. Tillack, J. P. Townsend, N. M. Tubman, B. Van Der Goetz, J. E. Vincent, D. C. Yang, Y. Yang, S. Zhang, and L. Zhao, *J. Phys. Condens. Matter* **30**, 195901 (2018).
- [48] M. Burkatzki, C. Filippi, and M. Dolg, *J. Chem. Phys.* **126**, 234105 (2007).

- [49] J. R. Trail and R. J. Needs, *J. Chem. Phys.* **122**, 174109 (2005).
- [50] The reciprocal space information is not lost. By keeping track of the Bloch momentum of the states, we can determine the locations of the HOMO and LUMO, but that has not been done here.
- [51] M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. Lett.* **120**, 205302 (2018).
- [52] M. Holzmann and S. Moroni, *Phys. Rev. B* **99**, 085121 (2019).
- [53] L. Zhao and E. Neuscamman, *Phys. Rev. Lett.* **123**, 036402 (2019).
- [54] R. Gómez-Abal, X. Li, M. Scheffler, and C. Ambrosch-Draxl, *Phys. Rev. Lett.* **101**, 106404 (2008).
- [55] W. Ku and A. G. Eguiluz, *Phys. Rev. Lett.* **89**, 126401 (2002).
- [56] F. Giustino, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* **105**, 265501 (2010).
- [57] B. Monserrat and R. J. Needs, *Phys. Rev. B* **89**, 214304 (2014).
- [58] P. Lautenschlager, P. B. Allen, and M. Cardona, *Phys. Rev. B* **31**, 2163 (1985).
- [59] B. Monserrat, G. J. Conduit, and R. J. Needs, *Phys. Rev. B* **90**, 184302 (2014).
- [60] C. D. Clark, P. J. Dean, P. V. Harris, and W. C. Price, *Proc. R. Soc. A.* **277**, 312 (1964).
- [61] B. Monserrat, *Phys. Rev. B* **93**, 100301(R) (2016).
- [62] S. Logothetidis, P. Lautenschlager, and M. Cardona, *Phys. Rev. B* **33**, 1110 (1986).
- [63] S. Lebegue, C. M. Araujo, D. Y. Kim, M. Ramzan, H. Mao, and R. Ahuja, *Proc. Nat. Acad. Sc.* **109**, 9766 (2012).
- [64] J. McMinis, R. C. Clay, III, D. Lee, and M. A. Morales, *Phys. Rev. Lett.* **114**, 105305 (2015).
- [65] R. C. Clay, III, J. McMinis, J. M. McMahon, C. Pierleoni, D. M. Ceperley, and M. A. Morales, *Phys. Rev. B* **89**, 184106 (2014).
- [66] V. Gorelov, M. Holzmann, D. M. Ceperley, and C. Pierleoni, [arXiv:1911.06135](https://arxiv.org/abs/1911.06135).
- [67] O. Madelung, M. Schulz, and H. Weiss, *Intrinsic Properties of Group IV Elements and III-V, II-VI and I-VII Compounds*, *Landolt-Börnstein, New Series, Group III/22A* (Springer-Verlag, New York, 1987).