

Variational and Diffusion Quantum Monte Carlo Calculations with the CASINO Code

R. J. Needs,¹ M. D. Towler,² N. D. Drummond,³ P. López Ríos,⁴ and J. R. Trail¹

¹*TCM Group, Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

²*University College London, London WC1E 6BT, United Kingdom*

³*Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom^{a)}*

⁴*Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany*

We present an overview of the variational and diffusion quantum Monte Carlo methods as implemented in the CASINO program. We particularly focus on developments made in the last decade, describing state-of-the-art quantum Monte Carlo algorithms and software and discussing their strengths and their weaknesses. We review a range of recent applications of CASINO.

CONTENTS

I. Introduction	1	1. Excitonic complexes	26
II. Continuum quantum Monte Carlo methods	3	2. Ground-state energy of the HEG	27
A. Variational quantum Monte Carlo	3	3. Phase diagram of electron-hole bilayers	28
1. Sampling a trial wave function	3	4. Photoexcitation in doped semiconductors	29
2. Slater-Jastrow trial wave functions	4	5. Quasiparticle effective mass of the 2D HEG	30
3. Optimization of wave-function parameters	7	6. Positrons immersed in electron gases	31
B. Diffusion quantum Monte Carlo	7	B. Van der Waals interactions	32
1. Fixed-phase approximation	7	1. Binding energies of 2D materials	32
2. Imaginary-time Schrödinger equation	8	2. Van der Waals interactions at surfaces and between molecules	32
3. Propagation in imaginary time	9	C. Solid hydrogen	33
4. Time step and decorrelation period	11	V. Future directions for QMC methods and the CASINO software	34
C. Overview of the CASINO software	13	A. Synergy with FCIQMC	34
III. Some recent developments in QMC and CASINO	13	1. FCIQMC wave functions in DMC	35
A. Advanced wave-function forms	13	2. Jastrow factors and similarity-transformed FCIQMC	35
1. Generalized Jastrow factors	13	3. DMC-assisted FCIQMC and benchmarking	36
2. Multideterminant wave functions	14	B. Towards greater efficiency	36
3. Geminal and pairing wave functions	15	1. Electron-electron pseudopotentials	36
B. QMC calculations for condensed matter	16	2. Current developments in computer architectures	36
1. Single-particle finite-size effects: momentum quantization	16	C. Improved accuracy	36
2. Long-range finite-size effects	17	1. Spin-orbit coupling	36
3. Finite-size effects in properties other than the total energy	19	2. Inclusion of vibrational effects in <i>ab initio</i> QMC calculations	37
C. Pseudopotentials for correlated methods	19	D. Atomic forces from QMC calculations	38
D. QMC for excited-state properties	20	VI. Conclusions	39
1. Quasiparticle and excitonic gaps	20	Acknowledgments	39
2. Finite-size effects in gaps	22	I. INTRODUCTION	
3. Intraband excitations	23		
E. Computational efficiency of QMC	23		
1. Parallelization of QMC algorithms	23		
2. Scaling of QMC with system size	24		
IV. Some recent applications of CASINO	26		
A. Model systems	26		

^{a)}Corresponding author, email: n.drummond@lancaster.ac.uk

For nearly a century it has been accepted that a large part of chemistry, materials science, and condensed matter physics could be quantitatively explained if only it were possible to solve the nonrelativistic many-electron Schrödinger equation for molecules, surfaces, and bulk materials.¹ It is straightforward to write down the many-electron time-independent

Schrödinger equation, which in Hartree atomic units ($\hbar = m_e = 4\pi\epsilon_0 = |e| = 1$) reads

$$\hat{H}\Psi \equiv \left(-\frac{1}{2}\nabla^2 + \sum_{i>j} \frac{1}{r_{ij}} - \sum_I \sum_i \frac{Z_I}{r_{iI}} \right) \Psi = E\Psi, \quad (1)$$

where $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ is the spatial wave function for N electrons, $-\frac{1}{2}\nabla^2 \equiv -\frac{1}{2}\sum_{i=1}^N \nabla_i^2$ is the total kinetic-energy operator, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between two electrons, $r_{iI} = |\mathbf{r}_i - \mathbf{R}_I|$ is the distance between electron i and nucleus I of atomic number Z_I at \mathbf{R}_I , and E is an energy eigenvalue. The fermionic nature of the electrons imposes the important requirement that the spatial wave function must be antisymmetric under the exchange of same-spin electrons. Unfortunately, solving the Schrödinger equation precisely for anything but the smallest system sizes is a grand-challenge problem due to the interactions between the electrons.

Over the last century a huge range of different techniques for approximately solving Eq. (1) has been developed. An isolated hydrogen atom with $N = 1$ can be solved exactly by pen and paper. For systems such as helium or lithium atoms ($N = 2$ or 3), the Hylleraas *ansatz* can be used to provide numerical solutions that are accurate to dozens of significant figures.^{2,3} For larger numbers of electrons we may approximate that the wave function Ψ is an antisymmetrized product (Slater determinant) of single-particle orbitals, which is the form of wave function that describes particles that are not coupled by interactions. Invoking the variational principle of quantum mechanics, we can generate numerical approximations to the orbitals by minimizing the energy expectation value $\langle \Psi | \hat{H} | \Psi \rangle / \langle \Psi | \Psi \rangle$, where \hat{H} is the Hamiltonian. This is the so-called Hartree-Fock (HF) approximation.⁴⁻⁶ It is important in physics and chemistry because it often provides a qualitative understanding of electronic behavior and because it is the starting point for more advanced methods. For system sizes between $N = 4$ and 100 electrons, quantum chemistry methods based on expansions in multiple Slater determinants are computationally tractable and provide highly accurate energies, or at least energies with consistent and controllable errors.^{7,8} An alternative starting point for solving the many-electron Schrödinger equation is provided by a theorem stating that the ground-state total energy is a unique functional of the electronic charge density.⁹ The electronic charge density can be parameterized using a set of single-particle orbitals for a fictitious noninteracting system, and the largest contributions to the variational total energy can be evaluated using these orbitals.¹⁰ The remaining contribution to the total energy of the interacting system can be evaluated as a parameterized functional of the electronic charge density. This approach, known as density functional theory (DFT), allows approximate solutions to the electronic Schrödinger equation to be obtained with up to at least $N = 10,000$ electrons. Despite the many successes of DFT, however, there are certain important situations in which it performs poorly: in the description of van der Waals interactions; in the description of electronic excitations; and in situations in which a cancellation of errors in energy differences cannot be relied upon.¹¹ Some of these deficiencies can be remedied to some extent by using DFT as the

starting point for many-body perturbation theory, especially in the so-called *GW* approximation.¹² Nevertheless, it is fortunate that we have alternative, highly accurate techniques available for solving the electronic Schrödinger equation with up to $N = 2,000$ electrons, namely, continuum quantum Monte Carlo (QMC) methods.

The variational and diffusion quantum Monte Carlo (VMC and DMC) methods that we discuss in this article use random sampling to solve the Schrödinger equation in the continuous position basis.¹³⁻¹⁶ In the VMC method¹³ we evaluate the expectation value of the Hamiltonian with respect to a trial wave function that may be of arbitrary complexity, rather than being restricted to an antisymmetrized product of single-particle orbitals as in HF theory. This means that the $3N$ -dimensional integral for the energy expectation value cannot be broken into a series of 3- and 6-dimensional integrals, and so we must use the only general high-dimensional numerical integration method available, which is Monte Carlo integration. In the DMC method¹⁴ we simulate drifting, random diffusion, and branching/dying processes governed by a Wick-rotated time-dependent Schrödinger equation in order to project out the ground-state component of a trial wave function. Fermionic antisymmetry is enforced for real wave functions by constraining the nodal surface to remain pinned at that of a trial wave function.¹⁷ QMC results come with a quantifiable random error, which falls off as $N_s^{-1/2}$ with the number of random samples N_s , irrespective of the size of problem, together with an unquantifiable, positive systematic bias due to the error in the trial wave function (in VMC) or the error in the nodal surface of the trial wave function (in DMC), the latter of which is referred to as the fixed-node error. The random errors are benign, as they are quantifiable, controllable, and easily propagated to derived results using standard statistical methods. The systematic biases are variational in nature (i.e., positive) and in general are demonstrably much smaller than those in other variational methods such as HF theory, so that VMC and DMC qualify as “high-accuracy methods.” The fact that system sizes of up to $N = 2,000$ electrons are accessible gives QMC methods one of their key unique selling points: they are the only highly accurate approaches that can be applied to periodic simulation cells that model meaningful quantities of condensed matter. For comparison, recent advances in quantum chemistry methods for solids enable the application of the coupled cluster method to systems of up to about $N = 100$ electrons.¹⁸⁻²¹

Historically, QMC methods have been applied to model systems in condensed matter such as the homogeneous electron gas (HEG), generating the data used to parameterize DFT exchange-correlation (XC) functionals.¹⁴ It remains the case that many of the most interesting, influential, and experimentally relevant QMC calculations are for simplified models of interacting particles, e.g., charge carriers within an effective-mass approximation or positrons immersed in electron gases. However, *ab initio* QMC calculations can also be performed, both for benchmarking simpler methods such as DFT, and for directly generating theoretical predictions against which experiments can be compared. A wide variety of QMC methods have been developed over the years, including path-

integral Monte Carlo,²² reptation Monte Carlo,²³ auxiliary-field quantum Monte Carlo,²⁴ full-configuration-interaction quantum Monte Carlo (FCIQMC),^{25–27} orbital-space VMC,²⁸ and orbital-space VMC-based geminal-Jastrow optimizers.²⁹ In this article we will discuss applications of VMC and DMC to both model and “real” systems. We will also describe the technical details of the methodology and discuss the strengths and weaknesses of the approach. There exist several excellent, actively developed implementations of continuum QMC methods, including CHAMP, QMCPACK,³⁰ QWALK,³¹ and TURBORVB. We will focus on our implementation of the QMC methods in the CASINO program,¹⁶ and on topics relevant to users of this software.

In Sec. II we review the basic theory of continuum QMC methods and their implementation in the CASINO code. In Sec. III we discuss some recent developments in QMC methods. We present a selection of applications of CASINO in Sec. IV to illustrate the capabilities of the QMC methods. We discuss future directions for QMC methods and the CASINO software in Sec. V. Finally we draw our conclusions in Sec. VI. Except where otherwise stated, we use Hartree atomic units throughout. We describe the density of a HEG by the density parameter r_s , which is the radius of the sphere (or circle in 2D) that contains one electron on average divided by the Bohr radius.

II. CONTINUUM QUANTUM MONTE CARLO METHODS

A. Variational quantum Monte Carlo

1. Sampling a trial wave function

A familiar example of Monte Carlo methods in undergraduate physics is the use of the Metropolis algorithm^{32,33} to sample the Boltzmann distribution, e.g., in numerical studies of the Ising model of spins on a lattice. In this case one repeatedly proposes changes to the spin configuration. If a proposed configuration increases the target probability (i.e., lowers the energy in the case of the Boltzmann distribution) then it is automatically accepted; otherwise it is accepted with a probability given by the ratio of the target probability of the proposed configuration to the target probability of the current configuration (assuming the proposal probabilities for forwards and backwards moves are the same). If a proposed configuration is accepted then it is added to the list of configurations sampled and becomes the new “current configuration;” if the proposed configuration is rejected then another copy of the current configuration is appended to the list of configurations sampled. This is an example of a Markov process: each step of the process only depends on the current configuration.

It is easy to show that once a set of configurations is distributed according to the target distribution it will continue to be distributed according to that distribution upon application of the Metropolis algorithm, i.e., the target distribution is a stationary point of the Metropolis algorithm. Furthermore, provided the proposal distribution is *ergodic*, allowing any configuration to be sampled after a finite number of moves

from any starting point, the stationary point of the Markov process is unique and the process will converge to that stationary point. Conveniently we do not even need to know the normalization factor of the target distribution to use this method.

In the VMC method^{13,15} we use the Metropolis method to sample the continuous distribution of electron position coordinates distributed according to $|\Psi|^2$, where Ψ is a trial spatial wave function. The expectation value of an operator \hat{A} can then be evaluated as

$$\langle \hat{A} \rangle = \frac{\int \Psi^* \hat{A} \Psi d\mathbf{R}}{\int |\Psi|^2 d\mathbf{R}} = \frac{\int |\Psi|^2 \frac{\hat{A}\Psi}{\Psi} d\mathbf{R}}{\int |\Psi|^2 d\mathbf{R}} = \langle A_L \rangle_{|\Psi|^2}, \quad (2)$$

where the localized operator $A_L \equiv (\hat{A}\Psi)/\Psi$ can be evaluated at each configuration $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ sampled. In particular, it is clear from Eq. (2) that the energy expectation value is estimated by the average of the local energy $E_L = (\hat{H}\Psi)/\Psi$ over the set of configurations sampled. The only requirements on the trial wave function are that the integral for the expectation value should converge and that it should be possible to evaluate a confidence interval on the average (in practice, this means that A_L^2 should also have a well-defined mean, so that we can compute the standard error in the mean of A_L). In the limit of perfect sampling the expectation value of a Hermitian operator is real, and hence we usually only evaluate and average the real part of the local operator $A_L^r \equiv \text{Re}(A_L)$, i.e., in practice the VMC estimate is $\langle A_L^r \rangle_{|\Psi|^2}$.

In Eq. (2) we have not explicitly mentioned spin. Consider a many-electron wave function $\Psi_{\text{sp}}(\mathbf{X})$, where $\mathbf{X} = (\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N)$ and \mathbf{r}_i and $\sigma_i \in \{\uparrow, \downarrow\}$ are the position and spin of electron i . Assume that Ψ_{sp} is an eigenfunction of the total spin operator $\hat{S}_z = \sum_{i=1}^N \hat{s}_{zi}$. Let the eigenvalue of \hat{S}_z be $(N_\uparrow - N_\downarrow)/2$, where $N_\uparrow + N_\downarrow = N$. The electron spins are said to be collinear; we have N_\uparrow spin-up electrons and N_\downarrow spin-down electrons. This restricts the form of Ψ_{sp} , but being an eigenfunction of \hat{S}_z is a property of exact eigenfunctions of the Hamiltonian if $[\hat{H}, \hat{S}_z] = 0$, as is the case for most problems in condensed matter. Electrons are fermions, so Ψ_{sp} is antisymmetric under electron exchange. The expectation value of a spin-independent operator \hat{A} with respect to Ψ_{sp} is

$$\begin{aligned} \langle \hat{A} \rangle_{\Psi_{\text{sp}}} &= \frac{\langle \Psi_{\text{sp}} | \hat{A} | \Psi_{\text{sp}} \rangle}{\langle \Psi_{\text{sp}} | \Psi_{\text{sp}} \rangle} \\ &= \frac{\sum_{\sigma} \int \Psi_{\text{sp}}^*(\mathbf{X}) \hat{A}(\mathbf{R}) \Psi_{\text{sp}}(\mathbf{X}) d\mathbf{R}}{\sum_{\sigma} \int |\Psi_{\text{sp}}(\mathbf{X})|^2 d\mathbf{R}}, \end{aligned} \quad (3)$$

where $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ and the sums run over all spin configurations such that the numbers of spin-up and spin-down electrons are N_\uparrow and N_\downarrow , respectively. \mathbf{X} may be replaced by $\mathbf{X}' = (\mathbf{r}_{i_1}, \uparrow, \dots, \mathbf{r}_{i_{N_\uparrow}}, \uparrow, \mathbf{r}_{i_{N_\uparrow}+1}, \downarrow, \dots, \mathbf{r}_{i_N}, \downarrow)$ in Eq. (3) without altering $\langle \hat{A} \rangle_{\Psi_{\text{sp}}}$, due to the antisymmetry of Ψ_{sp} . We may now relabel the dummy integration variables: $\mathbf{r}_{i_1} \rightarrow \mathbf{r}_1$, etc. The integrals in Eq. (3) are clearly the same for each spin configuration σ such that the number of spin-up electrons is

N_{\uparrow} ; hence the spin sums cancel. So the expectation of \hat{A} may be evaluated as

$$\langle \hat{A} \rangle_{\Psi_{\text{sp}}} = \frac{\int \Psi^*(\mathbf{R}) \hat{A}(\mathbf{R}) \Psi(\mathbf{R}) d\mathbf{R}}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}}, \quad (4)$$

where the spatial wave function

$$\Psi(\mathbf{R}) \equiv \Psi_{\text{sp}}(\mathbf{r}_1, \uparrow, \dots, \mathbf{r}_{N_{\uparrow}}, \uparrow, \mathbf{r}_{N_{\uparrow}+1}, \downarrow, \dots, \mathbf{r}_N, \downarrow) \quad (5)$$

is only antisymmetric with respect to exchanges of positions of same-spin electrons. In effect, electrons of different spin are treated as distinguishable fermions.³⁴ If we know $\Psi(\mathbf{R})$ then we can recover $\Psi_{\text{sp}}(\mathbf{X})$ and *vice versa*. In practice it is much easier just to work with Ψ , which will henceforth just be called “the (trial) wave function.” For example, the variational principle is

$$\frac{\langle \Psi_{\text{sp}} | \hat{H} | \Psi_{\text{sp}} \rangle}{\langle \Psi_{\text{sp}} | \Psi_{\text{sp}} \rangle} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0, \quad (6)$$

where \hat{H} is the Hamiltonian and E_0 is the ground-state energy. We may apply the variational principle to optimize $\Psi(\mathbf{R})$.

If QMC methods are used to study systems in which the spin and spatial coordinates are coupled, e.g., due to spin-orbit coupling terms in the Hamiltonian or because one is examining a broken-symmetry state such as a spin-density wave then one must return to using an antisymmetric wave function Ψ_{sp} in which the spatial coordinates are continuous and the spin coordinates are discrete. In the case of VMC this results in a trivial modification to the algorithms presented below. We briefly discuss some of the implications for DMC in Sec. VC 1.

Typically we find that it is much more efficient to propose and then accept or reject single-electron moves than whole-configuration moves. Moving a single electron a given distance results in a much smaller change in the wave function than moving all N electrons by the same distance. The proposed electron move is therefore more likely to be accepted, and hence the distance traveled by the electrons over N one-electron moves is greater on average. Electron-by-electron sampling therefore explores configuration space more efficiently.³⁵ In CASINO we use a Gaussian probability density for proposing electron moves, with the variance of the Gaussian being chosen such that the electron acceptance probability is close to 50%; empirically, this leads to near-maximum efficiency in exploring configuration space.^{32,35} The Gaussian probability density is advantageous because it is symmetric and extremely efficient to evaluate. It is possible to use more complicated proposal probability densities, e.g., taking the gradient of the trial wave function into account, which allow the electrons to travel further on average at each step;³⁶ however we find that the improvement in the efficiency of exploring configuration space is easily outweighed by the additional computational expense of evaluating wave-function derivatives.

Although the Metropolis algorithm produces configurations with the correct distribution, those configurations are serially correlated: successive configurations are similar to each other

due to the short steps taken and the fact that moves are rejected. For this reason, by default in CASINO we only evaluate the local energy and other quantities of interest on every third VMC step to save the computational expense of evaluating wave-function derivatives for configurations that are very similar to each other. Serial correlation complicates the calculation of the standard error in the mean energy, as there are in effect fewer independent data points than the number of points sampled.

A simple, robust method for removing serial correlation is the reblocking method,^{37–39} in which we repeatedly group the data into successive pairs and average within each pair to produce a new data set with half as many points, but with a greater degree of independence between the data points. This process is illustrated in Fig. 1 using VMC energy data for a lithium atom. Evaluating the standard error in the mean using

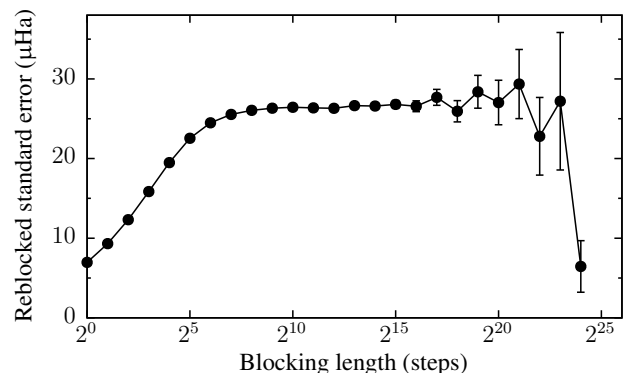


FIG. 1. Reblocked standard error in the VMC energy of an all-electron lithium atom as a function of blocking length. The blocking length is the number of successive original data points averaged over to obtain each new data point.

the naïve expression appropriate for independent data points, corresponding to a blocking length of 1, we find that the standard error is too small because the number of independent data points is overestimated. As the blocking length increases, the standard error rises and reaches a plateau when the block averages are essentially independent of each other. Specifically, the onset of the plateau occurs when the blocking length exceeds the decorrelation period of the original data. Finally, excessively large blocking lengths result in new data sets so small that the estimate of the standard error is itself subject to significant noise. There are various methods to automatically choose the optimal blocking length for a given data set.^{35,38,39} The reblocking method has the advantage that it does not assume the underlying data to be normally distributed; indeed, local-energy data generally follow a fat-tailed distribution.⁴⁰

2. Slater-Jastrow trial wave functions

The majority of QMC calculations use a Slater-Jastrow trial wave function,

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R})} \Psi_S(\mathbf{R}) = e^{J(\mathbf{R})} D_{\uparrow}(\mathbf{R}_{\uparrow}) D_{\downarrow}(\mathbf{R}_{\downarrow}), \quad (7)$$

in which the Slater part Ψ_S is a product of Slater determinants of single-particle orbitals for spin-up and spin-down electrons D_\uparrow and D_\downarrow , and e^J is a Jastrow correlation factor.

The single-particle orbitals in the Slater wave function are typically taken from a HF, DFT, or quantum chemistry calculation using an appropriate electronic-structure code. The orbitals may also include free parameters to optimize within QMC (see Sec. II A 3). Typically the Slater wave function will provide a good description of the electronic charge density (the one-body electron distribution), but no description of electronic correlation. More fundamentally the Slater wave function provides the qualitative features of the state under consideration; it has the required fermionic antisymmetry and any other symmetries of the system, and the topology of the nodal surface of the Slater wave function uniquely specifies the excited state to be studied. (For a system without time-reversal symmetry, the phase of the complex wave function plays the role of the nodal surface of a real wave function; see Sec. II B 1.)

Any errors in the topology of the nodal surface give rise to so-called static correlation effects, which are present in nearly all nontrivial electronic systems and may be energetically important.⁴¹ These errors are unaffected by the Jastrow factor or by the use of a backflow transformation (see below), and must instead be addressed by the use of multideterminant or pairing wave functions; see Sec. III A.

However, in many cases we are interested in long-range effects and low-energy effects, e.g., changes to the distribution or occupancy of the outermost electrons in atoms or molecules. In such cases, static correlation effects largely cancel out of energy differences with a single-determinant Slater wave function, and only a good description of dynamical correlation effects (which do not involve changing the nodal topology) is required for high accuracy. Jastrow factors and backflow transformations give accurate and compact descriptions of dynamical correlations.

A Jastrow factor for a molecular system consisting of N electrons and N_n nuclei is typically a sum of isotropic electron-electron, electron-nucleus, and electron-electron-nucleus terms,

$$J(\mathbf{R}) = \sum_{i<j}^N u_{P_{ij}}(r_{ij}) + \sum_i^N \sum_I^{N_n} \chi_{S_{iI}}(r_{iI}) + \sum_{i<j}^N \sum_I^{N_n} f_{T_{ijI}}(r_{iI}, r_{jI}, r_{ij}), \quad (8)$$

where u , χ , and f are parameterized functions of the relevant interparticle distances, and P , S , and T are ‘‘channel’’ indices such that different parameter sets can be used for, e.g., parallel and antiparallel spin electrons or for different atomic species.

The homogeneous electron-electron dynamical correlations provided by the u term usually afford the most significant improvement to the trial wave function, but this alters the accurate one-electron density encoded in the HF or DFT orbitals. The χ function is capable of restoring the one-electron density,⁴² and the f term provides inhomogeneous electron-electron correlations that further refine the wave function.

Jastrow factors are fully symmetric and strictly positive, so they do not alter the overall symmetry or nodal structure of Ψ_S . Therefore the DMC method gives the same result whether a Jastrow factor is present or not. However, Jastrow factors do offer a great improvement in the statistical efficiency of DMC, partly because of their ability to impose the Kato cusp conditions.^{43,44}

In Coulomb systems the local potential energy diverges whenever two charged particles coincide, but the exact local energy is a finite constant everywhere in configuration space, so the local kinetic energy must diverge so as to cancel the divergence of the local potential energy. Let i and j be two particles of charge q_i and q_j , respectively, of reduced mass μ_{ij} , and separated by a distance $r = |\mathbf{r}_i - \mathbf{r}_j|$. Requiring the local energy to be finite at $r \rightarrow 0$ gives

$$\left(\frac{1}{\Psi} \frac{\partial \Psi}{\partial r} \right)_{r \rightarrow 0} = \frac{2q_i q_j \mu_{ij}}{d \pm 1} = \Gamma_{ij}, \quad (9)$$

where d is the dimensionality of the system and the plus and minus signs in the denominator correspond to indistinguishable and distinguishable particles, respectively.⁴⁵

Electron-electron cusps are usually applied on the electron-electron Jastrow factor by enforcing $\partial u / \partial r_{ij} = \Gamma_{ij}$ at $r_{ij} = 0$. The variance of the local energies encountered in VMC and DMC calculations is significantly smaller for trial wave functions which satisfy the electron-electron Kato cusp conditions than for those which do not.

When cuspless one-electron orbitals are used in Ψ_S , such as orbitals expanded in Gaussians, one can enforce the electron-nucleus Kato cusp conditions either by modifying the orbital near the nucleus with an analytically correct form⁴⁶ or by setting $\partial \chi / \partial r_{iI} = \Gamma_{iI}$ at $r_{iI} = 0$. In practice, cusp-correcting the orbital is found to be the better approach, because the error in the wave function occurs at a very short range determined by the basis precision. We illustrate this by plotting the local energy as a function of the position of an electron moving through the nucleus of an all-electron carbon atom in Fig. 2, where the local energy can be seen to vary much less with the orbital correction than with a cusp-enforcing short-ranged χ function.

At long range the Jastrow factor can be thought of as providing a slowly varying envelope for the Slater wave function, describing the behavior of interacting quasielectrons. Indeed, the very-long-range two-electron contribution to the Jastrow factor is exactly described by the random-phase approximation (RPA), i.e., linear response theory within the Hartree approximation.⁴⁷

In the Drummond-Towler-Needs (DTN) Jastrow factor⁴⁸ the u , χ , and f functions are parameterized as natural power expansions,

$$u_P(r) = t_P(r) \sum_k \alpha_k^P r^k, \\ \chi_S(r) = t_S(r) \sum_k \beta_k^S r^k, \\ f_T(r_{ij}, r_{iI}, r_{jI}) = t_T(r_{iI}) t_T(r_{jI}) \sum_{k,l,m} \gamma_{klm}^T r_{ij}^k r_{iI}^l r_{jI}^m, \quad (10)$$

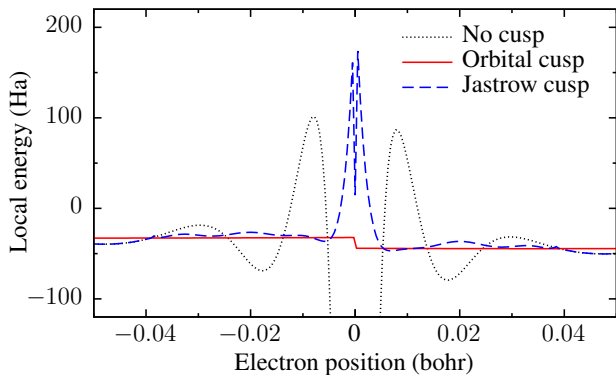


FIG. 2. Local energy as a function of the x coordinate of an electron in a carbon atom using a trial wave function which does not satisfy the electron-nucleus cusp condition (“no cusp”) and wave functions which impose the electron-nucleus cusp condition by modifying the Gaussian orbitals near the nucleus (“orbital cusp”) and via the Jastrow factor (“Jastrow cusp”). The nucleus is at $x = 0$, and the other five electrons are at random positions.

where the $\{\alpha\}$, $\{\beta\}$, and $\{\gamma\}$ are linear parameters, and the $\{t(r)\}$ are cutoff functions of the form

$$t(r) = (r - L)^C \Theta_H(r - L), \quad (11)$$

where L is an optimizable truncation length, C is an integer truncation order (typically $C = 2$ or 3), and $\Theta_H(x)$ is the Heaviside step function.

In periodic systems the interparticle distances in the expressions above are evaluated between nearest periodic images only. The truncation lengths for the homogeneous two-body terms (u) and one-body terms (χ) must therefore be constrained to be less than or equal to L_{WS} , where L_{WS} is half the nearest-image distance; otherwise the trial wave function would have a discontinuous gradient at the surface of the Wigner-Seitz cell of the simulation cell. Likewise, the truncation lengths for the inhomogeneous two-body terms (f) must be less than or equal to $L_{WS}/2$. Homogeneous two-body correlations in particular are long-range. In order to cover the “corners” of the simulation cell, the DTN Jastrow factor provides one- and two-electron cosine expansions. The form of the two-body cosine expansion is

$$p(\mathbf{r}) = \sum_{s=1} N_p a_s^{P'} \sum_{\mathbf{G} \in \star_s^+} \cos(\mathbf{G} \cdot \mathbf{r}), \quad (12)$$

where \star_s^+ denotes every second member of the s th star of simulation-cell reciprocal lattice points. This term is cuspless, but it is not truncated at finite range and hence it provides a description of long-range correlation effects. It has the full symmetry of the simulation-cell Bravais lattice. Since it describes long-range behavior, we generally only need distinct parameters for particle pairs with different charges; the same parameters can be used for parallel- and antiparallel-spin pairs. The one-body cosine expansion has a similar form, but is expanded in primitive-cell reciprocal lattice points, so that it has the full symmetry of the primitive-cell Bravais lattice.

It is rarely used, since one-body Jastrow terms are not usually long-ranged. A situation in which it might be useful would be in a study of a single charged defect in a simulation supercell.

Information about CASINO’s Jastrow factor implementation beyond the DTN form can be found in Sec. III A 1. The use of a Jastrow factor typically allows VMC to recover about 80–95% of the DMC correlation energy.

The quality of the DMC energy obtained with a Slater-Jastrow wave function is limited by the quality of the nodes of Ψ_S . However it is possible to alter the nodal surface of the Slater-Jastrow wave function by the use of a backflow transformation.^{49–53} The purpose of the backflow transformation is to provide a parameterized, smooth transformation of the configuration $\mathbf{X}(\mathbf{R})$ at which Ψ_S is evaluated to allow the optimization of the location of its nodes.

The Slater-Jastrow-backflow wave function is

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R})} \Psi_S(\mathbf{X}(\mathbf{R})), \quad (13)$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ is a set of quasiparticle coordinates. The i th quasiparticle coordinate is defined as the position of the i th electron plus a parameterized backflow displacement which depends on the positions of the other electrons in the system. In the inhomogeneous backflow⁵³ implementation in CASINO, the backflow displacement is, much like a Jastrow factor, the sum of electron-electron, electron-nucleus, and electron-electron-nucleus contributions,

$$\begin{aligned} \mathbf{x}_i = \mathbf{r}_i &+ \sum_{j \neq i}^N \eta_{P_{ij}}(r_{ij}) \mathbf{r}_{ij} + \sum_I^{N_n} \mu_{S_{ij}}(r_{iI}) \mathbf{r}_{iI} \\ &+ \sum_{j \neq i}^N \sum_I^{N_n} [\Phi_{T_{ijI}}(r_{ij}, r_{iI}, r_{jI}) \mathbf{r}_{ij} \\ &\quad + \Theta_{T_{ijI}}(r_{ij}, r_{iI}, r_{jI}) \mathbf{r}_{iI}], \end{aligned} \quad (14)$$

where the parameterized functions η , μ , Φ , and Θ have the same form as those of the DTN Jastrow factor,

$$\begin{aligned} \eta_P(r) &= t_P(r) \sum_k c_k^P r^k, \\ \mu_S(r) &= t_S(r) \sum_k d_k^S r^k, \\ \Phi_T(r_{ij}, r_{iI}, r_{jI}) &= t_T(r_{ij}) t_T(r_{jI}) \sum_{k,l,m} \phi_{klm}^T r_{ij}^k r_{iI}^l r_{jI}^m, \\ \Theta_T(r_{ij}, r_{iI}, r_{jI}) &= t_T(r_{iI}) t_T(r_{jI}) \sum_{k,l,m} \theta_{klm}^T r_{ij}^k r_{iI}^l r_{jI}^m, \end{aligned} \quad (15)$$

where the $\{c\}$, $\{d\}$, $\{\phi\}$, and $\{\theta\}$ are optimizable parameters. Note that since the backflow displacement is a vector function, the electron-electron-nucleus term requires two separate functions Φ_T and Θ_T to span the plane defined by the two electrons and the nucleus. There exist alternative forms of backflow in the literature, including analytical,⁵² recursive,⁵⁴ and orbital-dependent⁵⁵ formulations. A deep-neural-network-based wave function has been proposed recently consisting of a short multideterminant expansion populated with flexible orbitals and backflow-like features.⁵⁶

For a demonstration of the effect of backflow on the total energy for the (very favorable) case of the HEG, see Fig. 26 in Sec. IV A 2, in which backflow recovers over half of the fixed-node error in DMC. In inhomogeneous systems, backflow can be expected to recover a smaller but significant fraction of the fixed-node error.⁵³

3. Optimization of wave-function parameters

Two stable and efficient optimization methods are widely used in modern QMC calculations. The first approach is based on the observation that the local energy $E_L = (\hat{H}\Psi)/\Psi$ is constant in configuration space if and only if Ψ is an eigenfunction of the Hamiltonian \hat{H} . Hence, for any given sampling of points $\{\mathbf{R}\}$ in configuration space, we may optimize the wave-function parameters by varying them to minimize the spread of local energies.^{34,57–59}

The spread of a set of local energies is most often quantified by the variance. In a typical variance-minimization calculation, VMC is used to sample a set of configurations $\{\mathbf{R}\}$ distributed as $|\Psi|^2$, then the free parameters in Ψ are varied to minimize the variance of the local energies evaluated at the configuration set sampled, then VMC is used to generate a new set of configurations, and so on. In practice the algorithm normally converges to the necessary level of accuracy in 1–3 cycles. Minimizing the variance of the local energies with respect to wave-function parameters closely resembles the procedure for least-squares fitting.

The parameters that recover the largest part of the correlation energy appear linearly in the Jastrow exponent J . The local energy is a quadratic function of such parameters, and hence the variance of the set of local energies is a quartic function of those parameters.⁵⁹ Thus it is possible to minimize the variance rapidly with respect to those parameters, and also to search for global minima along lines in parameter space.

Variance minimization is robust and efficient for optimizing parameters in the Jastrow factor, reflecting the simple form of the variance as a function of those parameters. However, variance minimization is much less effective for optimizing parameters that affect the nodal surface of the wave function. The local energy of a configuration diverges if that configuration lies on a node of the trial wave function; hence, as one adjusts a parameter that affects the nodal surface, the variance of the local energies diverges each time the nodal surface moves through one of the configurations sampled. As a result non-global minima are a significant problem, and variance minimization will often converge very slowly with the number of cycles. We find that minimizing the mean absolute deviation from the median local energy instead of the variance of the local energies can be advantageous in these situations.⁶⁰

In Table I we compare the performance of optimization methods based on minimizing different measures of the spread of local energies for an all-electron neon atom. The results confirm that minimizing the mean absolute deviation from the median local energy performs relatively well in terms of the resulting VMC energy, especially when optimizing parameters that affect the nodal surface.

An alternative approach is to minimize the energy expectation value in accordance with the variational principle of quantum mechanics. This works well for both Jastrow parameters and parameters in the Slater wave function. It also has the advantage that the importance-sampled DMC algorithm can be shown to be of maximal efficiency when the VMC energy is minimized.⁶¹

To minimize the energy expectation value, we use a method which transforms the stochastic optimization problem into the diagonalization of the Hamiltonian matrix in an approximate finite basis.^{62–64} Suppose our initial wave function is Ψ_{α_0} , where α_0 is a vector of P parameters. Let us choose our (nonorthogonal) basis set to be the initial wave function and its derivatives with respect to the parameters, $\{\Psi_{\alpha_0}, (\partial\Psi/\partial\alpha_1)_{\alpha_0}, (\partial\Psi/\partial\alpha_2)_{\alpha_0}, \dots\}$. We may evaluate the $(P+1) \times (P+1)$ matrix H of the Hamiltonian with respect to these basis functions, together with the overlap matrix S between the basis functions by VMC sampling of $|\Psi_{\alpha_0}|^2$, then solve the resulting generalized eigenproblem $H\mathbf{c} = \mathcal{E}S\mathbf{c}$ to find the ground-state eigenvector \mathbf{c} and corresponding eigenvalue \mathcal{E} . Suppose we choose the normalization of the $(P+1)$ -dimensional eigenvector \mathbf{c} such that the coefficient of the first basis function is 1, i.e., we write $\mathbf{c} = (1, \Delta\alpha_1, \Delta\alpha_2, \dots)$, where the second and subsequent elements of \mathbf{c} define a P -dimensional vector $\Delta\alpha$. The resulting approximation to the ground-state wave function is

$$\begin{aligned} \Psi_{\alpha} &= \Psi_{\alpha_0} + \Delta\alpha_1(\partial\Psi/\partial\alpha_1)_{\alpha_0} + \Delta\alpha_2(\partial\Psi/\partial\alpha_2)_{\alpha_0} + \dots \\ &= \Psi_{\alpha_0 + \Delta\alpha} + \mathcal{O}(\Delta\alpha^2). \end{aligned} \quad (16)$$

We thus have a new parameter set $\alpha = \alpha_0 + \Delta\alpha$, and can repeat the whole process. Once the process converges, i.e. the parameters cease to change over subsequent cycles, the quadratic $\mathcal{O}(\Delta\alpha^2)$ term in Eq. (16) vanishes, so that we locate a minimum of the energy (provided our VMC estimates of H and S are sufficiently precise).

A few technical tricks are required to make this work in practice. Firstly, it can be shown that it is much better not to impose Hermiticity on the VMC estimates of H and S obtained with a finite sampling of configurations;⁶⁴ the resulting algorithm then works exactly both in the limit of an infinite number of VMC samples and in the limit that the basis functions span an invariant subspace of the Hamiltonian operator. One can choose a parameter-dependent normalization for the basis functions such that they are orthogonal to an appropriate linear combination of Ψ_{α_0} and Ψ_{α} . Finally one can artificially increase the diagonal elements of H to prevent large steps in parameter space.⁶⁴ Putting these tricks together results in a stable and effective energy-minimization algorithm.

B. Diffusion quantum Monte Carlo

1. Fixed-phase approximation

Consider a Hamiltonian $\hat{H} = -\frac{1}{2}\nabla^2 + V$, where V is the total potential energy. Let $\Phi = |\Phi|e^{i\chi}$ be a complex many-electron spatial wave function. By fermionic antisymmetry,

TABLE I. Average VMC energy and variance for an all-electron neon atom over eight cycles of VMC configuration generation/wave-function optimization. 10,000 VMC-sampled configurations were used to perform each optimization. Four different measures of the spread of the local energies of the configurations were minimized to optimize the wave function: the unweighted variance of the local energies; the “filtered” variance, in which outlying local energies are excluded from the variance estimate; the difference of the upper and lower quartiles of the set of local energies; and the mean absolute deviation from the median local energy.

Objective function	Optimizing Jastrow only		Optimizing Jastrow + orbitals		Optimizing Jastrow + backflow	
	VMC energy (Ha)	Var. (Ha ²)	VMC energy (Ha)	Var. (Ha ²)	VMC energy (Ha)	Var. (Ha ²)
Variance	-128.8824(6)	1.180	-128.8931(5)	0.959	-128.9133(6)	0.405
“Filtered” variance	-128.8826(3)	1.180	-128.8936(7)	0.958	-128.9147(5)	0.394
Quartile difference	-128.8837(1)	1.199	-128.8953(8)	1.016	-128.91737(9)	0.422
Mean abs. deviation	-128.8851(2)	1.195	-128.89957(8)	0.989	-128.9178(2)	0.420

the phase χ must change by π whenever two same-spin electrons are exchanged. Then

$$\begin{aligned}
e^{-i\chi}\hat{H}\Phi &= \left[\frac{1}{2} \left(-\nabla^2 + |\nabla\chi|^2 \right) + V \right] |\Phi| \\
&\quad + i \left[\frac{1}{2} \left(-2\nabla\chi \cdot \nabla - \nabla^2\chi \right) \right] |\Phi| \\
&\equiv \left[\hat{\mathcal{H}}_\chi + i\hat{\mathcal{K}}_\chi \right] |\Phi|. \tag{17}
\end{aligned}$$

We then find that

$$\langle \Phi | \hat{H} | \Phi \rangle = \langle |\Phi| | \hat{\mathcal{H}}_\chi | |\Phi| \rangle. \tag{18}$$

So the ground-state eigenvalue of the fixed-phase Schrödinger equation $\hat{\mathcal{H}}_\chi\phi_0 = E_0\phi_0$ is equal to the expectation value of the Hamiltonian \hat{H} with respect to $|\phi_0|e^{i\chi}$, which is greater than or equal to the fermionic ground-state energy of \hat{H} by the variational principle, becoming equal in the limit that the fixed phase χ is exactly equal to that of the fermionic ground state.⁶⁵

In the following we will adopt the fixed-phase approximation,⁶⁵ in which it is assumed that the phase of wave function Φ is fixed to be the same as the phase χ of the trial wave function Ψ , but its modulus $|\Phi|$ is allowed to vary. The fixed-phase approximation is the complex generalization of the fixed-node approximation.¹⁷ The latter corresponds to the case in which the trial wave function is real, so that at each point in configuration space the phase is either $\chi = 0$ or π .

The fixed-phase approximation with an antisymmetric trial wave function gives us a variational principle for the lowest-energy antisymmetric eigenstate.³⁴ Likewise, the fixed-phase approximation with a trial wave function that transforms as a one-dimensional (1D) irreducible representation of the symmetry group of the Hamiltonian gives a variational principle with respect to the lowest-energy exact wave function that transforms as that 1D irreducible representation.⁶⁶ Thus the DMC energy only satisfies the variational principle for the ground state and certain excited states. Nevertheless, DMC always gives the energy of any excited state exactly if the phase is exact for that state. Hence we can use fixed-phase DMC to

calculate approximate excited-state energies by using an appropriate trial wave function, even for excited states for which we do not have the variational principle. So we can calculate excitation energies (points on the band structure for periodic systems) via differences in total energy.

2. Imaginary-time Schrödinger equation

Now consider the fixed-phase imaginary-time Schrödinger equation,

$$\begin{aligned}
\left[\hat{\mathcal{H}}_\chi - E_T \right] |\Phi| &= \left[-\frac{1}{2}\nabla^2 + \frac{1}{2}|\nabla\chi|^2 + V - E_T \right] |\Phi| \\
&= -\frac{\partial|\Phi|}{\partial t}, \tag{19}
\end{aligned}$$

where $\Phi(\mathbf{R}, t) = |\Phi|e^{i\chi}$ is an imaginary-time-dependent wave function whose complex phase χ is everywhere equal to that of the trial wave function $\Psi = |\Psi|e^{i\chi}$, and the reference energy E_T is an offset in the energy origin. The time dependence is separable, so we may write the solution as

$$|\Phi| = \sum_{n=0}^{\infty} c_n \phi_n e^{-(E_n - E_T)t}, \tag{20}$$

where E_n and ϕ_n are the n th eigenvalue and eigenfunction of the fixed-phase Hamiltonian $\hat{\mathcal{H}}_\chi$. Excited states die away exponentially with increasing t compared with the ground state, as shown in Fig. 3. If $E_T = E_0$ and the initial conditions have $c_0 \neq 0$ then, in the limit $t \rightarrow \infty$, $|\Phi|$ is proportional to ϕ_0 . The ground-state component of $|\Phi|$ is thus “projected out.”

Let $f = |\Phi||\Psi|$ be the *mixed* distribution, where Ψ is the trial wave function. Then f is real and positive and Eq. (19) leads to

$$-\frac{1}{2}\nabla^2 f + \nabla \cdot [\mathbf{V}^r f] + [E_L^r - E_T] f = -\frac{\partial f}{\partial t}, \tag{21}$$

where $\mathbf{V}^r = \text{Re}[(\nabla\Psi)/\Psi]$ is the real part of the drift velocity and $E_L^r = \text{Re}(E_L)$ is the real part of the local energy.

Let $\{\mathbf{R}\}$ be many-body position basis vectors, normalized such that $\langle \mathbf{R} | \mathbf{R}' \rangle = \delta(\mathbf{R} - \mathbf{R}')$, and let $|\mathbf{P}\rangle =$

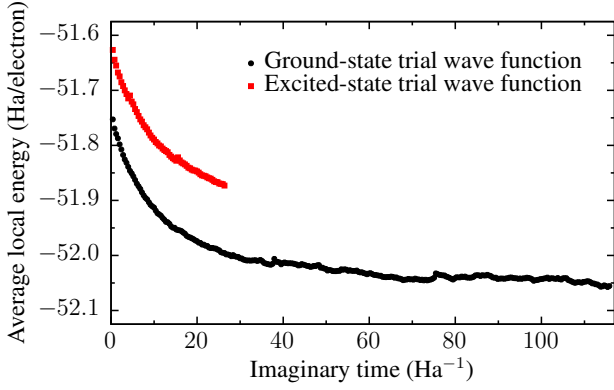


FIG. 3. Decay of the average local energy of a set of walkers whose dynamics is governed by the fixed-phase imaginary-time Schrödinger equation for the ground state and an excited state of a 610-electron 2D HEG at density parameter $r_s = 10$. Results are shown for (i) a real, closed-shell, ground-state, Slater–Jastrow trial wave function with plane-wave orbitals and (ii) a complex, excited-state wave function in which a single electron has been added to a plane-wave orbital outside the Fermi surface without reoptimization of the ground-state Jastrow factor. The DMC time step is 0.4 Ha^{-1} and the target population is 12,000 walkers, so that the statistical error on the energy at each time step is small.

$(2\pi)^{-3N/2} \int e^{i\mathbf{R}\cdot\mathbf{P}} |\mathbf{R}\rangle d\mathbf{R}$ be the corresponding momentum basis vectors. Let $\hat{\mathbf{R}}$ and $\hat{\mathbf{P}}$ be the position and momentum operators. In Dirac notation the mixed distribution can be written as $f(\mathbf{R}) = \langle \mathbf{R} | f \rangle$, and Eq. (21) can be written as

$$[\hat{F} + E_L^r(\hat{\mathbf{R}}) - E_T] |f\rangle = -\frac{d}{dt} |f\rangle, \quad (22)$$

where $\hat{F} = (1/2)\hat{\mathbf{P}}^2 + i\hat{\mathbf{P}} \cdot \mathbf{V}^r(\hat{\mathbf{R}})$ is the Fokker-Planck operator.

3. Propagation in imaginary time

The mixed distribution at imaginary time $t + \tau$ can be written as

$$f(\mathbf{R}, t + \tau) = \int G(\mathbf{R} \leftarrow \mathbf{R}', \tau) f(\mathbf{R}', t) d\mathbf{R}', \quad (23)$$

where the Green’s function

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}', \tau) &= \langle \mathbf{R} | e^{-\tau(\hat{F} + E_L^r(\hat{\mathbf{R}}) - E_T)} | \mathbf{R}' \rangle \\ &\approx \langle \mathbf{R} | e^{-\tau\hat{F}} | \mathbf{R}' \rangle e^{-\frac{\tau}{2}[E_L^r(\mathbf{R}) + E_L^r(\mathbf{R}') - 2E_T]} \\ &\quad + \mathcal{O}(\tau^3). \end{aligned} \quad (24)$$

is the solution of Eq. (21) satisfying the initial condition $G(\mathbf{R} \leftarrow \mathbf{R}', 0) = \delta(\mathbf{R} - \mathbf{R}')$.

Let $:\hat{A}:$ be the normal-ordered version of an operator \hat{A} , in which all $\hat{\mathbf{P}}$ operators are moved to the left of all $\hat{\mathbf{R}}$ operators (as if they commuted) within each term. Note that $:\hat{F}:= \hat{F}$, so $:e^{-\tau\hat{F}}:= 1 - \tau\hat{F} + \mathcal{O}(\tau^2) = e^{-\tau\hat{F}} + \mathcal{O}(\tau^2)$. Hence

$$\begin{aligned} \langle \mathbf{R} | :e^{-\tau\hat{F}}: | \mathbf{R}' \rangle &= \int \langle \mathbf{R} | \mathbf{P} \rangle \langle \mathbf{P} | :e^{-\tau[\frac{\mathbf{P}^2}{2} + i\hat{\mathbf{P}} \cdot \mathbf{V}^r(\hat{\mathbf{R}})]}: | \mathbf{R}' \rangle d\mathbf{P} \\ &= \int \langle \mathbf{R} | \mathbf{P} \rangle e^{-\tau[\frac{\mathbf{P}^2}{2} + i\mathbf{P} \cdot \mathbf{V}^r(\mathbf{R}')] } \langle \mathbf{P} | \mathbf{R}' \rangle d\mathbf{P} \\ &= \int \frac{e^{i\mathbf{P} \cdot (\mathbf{R} - \mathbf{R}')}}{(2\pi)^{3N}} e^{-\tau[\frac{\mathbf{P}^2}{2} + i\mathbf{P} \cdot \mathbf{V}^r(\mathbf{R}')] } d\mathbf{P} \\ &= \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{1}{2\tau} |\mathbf{R} - \mathbf{R}' - \tau\mathbf{V}^r(\mathbf{R}')|^2}. \end{aligned} \quad (25)$$

The last line is the Langevin or drift-diffusion Green’s function, describing diffusion of particles (“walkers”) in a $3N$ -dimensional fluid of time-independent velocity field $\mathbf{V}^r(\mathbf{R}')$ at small time step τ . Physically, the approximation of using the normal-ordered Green’s function is equivalent to assuming the drift velocity to be constant between \mathbf{R} and \mathbf{R}' . So the Green’s function for Eq. (21) is

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}', \tau) &\approx \langle \mathbf{R} | :e^{-\tau\hat{F}}: | \mathbf{R}' \rangle e^{-\frac{\tau}{2}[E_L^r(\mathbf{R}) + E_L^r(\mathbf{R}') - 2E_T]} \\ &\quad + \mathcal{O}(\tau^2) \\ &\equiv G_{\text{DMC}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) + \mathcal{O}(\tau^2). \end{aligned} \quad (26)$$

The branching factor $e^{-\frac{\tau}{2}[E_L^r(\mathbf{R}) + E_L^r(\mathbf{R}') - 2E_T]}$ is the solution of Eq. (21) without the first two terms on the left-hand side; it represents exponential growth/decay in the density of walkers at each point in configuration space. The DMC Green’s function therefore describes the evolution of the density of a set of walkers drifting, diffusing, and breeding or dying in a $3N$ -dimensional space.

The Trotter-Suzuki expression for the Green’s function for a macroscopic length of imaginary time $M\tau$ is

$$\begin{aligned}
G(\mathbf{R} \leftarrow \mathbf{R}', M\tau) &= \langle \mathbf{R} | e^{-M\tau(\hat{F} + \hat{E}_L^i - E_T)} | \mathbf{R}' \rangle \\
&\approx \int \dots \int \langle \mathbf{R} | e^{-\tau(\hat{F} + \hat{E}_L^i - E_T)} | \mathbf{R}'' \rangle \dots \langle \mathbf{R}''' | e^{-\tau(\hat{F} + \hat{E}_L^i - E_T)} | \mathbf{R}' \rangle d\mathbf{R}'' \dots d\mathbf{R}''' + \mathcal{O}(M\tau^3) \\
&\approx \int \dots \int G_{\text{DMC}}(\mathbf{R} \leftarrow \mathbf{R}'', \tau) \dots G_{\text{DMC}}(\mathbf{R}''' \leftarrow \mathbf{R}', \tau) d\mathbf{R}'' \dots d\mathbf{R}''' + \mathcal{O}(M\tau^2). \tag{27}
\end{aligned}$$

The approximation to the Green's function over a finite interval can be made arbitrarily accurate by dividing the interval into sufficiently small slices of imaginary time. We can therefore use the DMC Green's function to propagate f to large imaginary time (where $f = |\phi_0\rangle\langle\Psi|$) using a finite, small time step τ :

$$f(\mathbf{R}, t) = \int G(\mathbf{R} \leftarrow \mathbf{R}', t) f(\mathbf{R}', 0) d\mathbf{R}', \tag{28}$$

where the propagation to time $t = M\tau$ is carried out in M short steps of length τ using G_{DMC} .

The use of a nonzero time step continually introduces errors, even as the evolution in imaginary time projects out the ground-state component. At large imaginary time we may write $f = |\phi_0\rangle\langle\Psi| + \Delta$, where Δ is the time-step error in the mixed distribution. The error in G_{DMC} per time step is $\mathcal{O}(\tau^2)$, so error in f is introduced at rate $\mathcal{O}(\tau)$. Error is removed at a rate $\sim -\Delta/\tau_{\text{corr}}$, where τ_{corr} is the decorrelation period in imaginary time (see Sec. II B 4). In steady state, these rates balance. Hence $\Delta \sim \tau_{\text{corr}}\tau$, i.e., the time-step error in the mixed distribution is $\mathcal{O}(\tau)$.

At any given moment in a DMC simulation, f is represented by a discrete population of ‘‘walkers’’ in configuration space:⁶⁷

$$f(\mathbf{R}, t) = \left\langle \sum_{\alpha} w_{\alpha} \delta(\mathbf{R} - \mathbf{R}_{\alpha}) \right\rangle, \tag{29}$$

where \mathbf{R}_{α} is the position of walker α and w_{α} is its weight and the angled brackets denote an ensemble average. Ensemble averaging commutes with linear operations such as differentiation. The mixed distribution after one time step τ is

$$f(\mathbf{R}, t + \tau) = \left\langle \sum_{\alpha} w_{\alpha} G_{\text{DMC}}(\mathbf{R} \leftarrow \mathbf{R}_{\alpha}, \tau) \right\rangle. \tag{30}$$

It is now clear that the Green's functions can be treated as transition-probability densities; the ensemble average then has the correct behavior. In summary, we must simulate a large population of walkers that, over the course of one time step, drift by $\tau\mathbf{V}^r(\mathbf{R}_{\alpha})$ and diffuse (are displaced by a random vector, Gaussian-distributed with variance τ); finally the branching factor is absorbed into a new weight for each walker. This is usually done by a random branching or dying process, such that the expectation value of the number of unit-weighted walkers that continue from the walker's current position is equal to the weight of the walker after the move. The branching/dying algorithm avoids the situation in which one walker gathers exponentially more weight than the others.

The Green's function for Eq. (19) is $\langle \mathbf{R} | e^{-\tau(\hat{\mathcal{H}}_x - E_T)} | \mathbf{R}' \rangle$. Using the importance-sampling transformation, we can therefore write the Green's function for Eq. (21) as

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \Psi(\mathbf{R}) \langle \mathbf{R} | e^{-\tau(\hat{\mathcal{H}}_x - E_T)} | \mathbf{R}' \rangle \Psi^{-1}(\mathbf{R}'). \tag{31}$$

Now $e^{-\tau(\hat{\mathcal{H}}_x - E_T)}$ is Hermitian, so

$$|\Psi(\mathbf{R}')|^2 G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = |\Psi(\mathbf{R})|^2 G(\mathbf{R}' \leftarrow \mathbf{R}, \tau). \tag{32}$$

The approximation that $\mathbf{V}^r(\mathbf{R})$ is constant between \mathbf{R}' and \mathbf{R} violates this detailed-balance condition at finite time steps. We may reimpose this important condition using a Metropolis-style accept/reject step. As with VMC, it is more efficient to propose individual-electron moves than whole-configuration moves. Although the accept/reject step does not formally change the scaling of the time-step error in the DMC Green's function, in practice it enormously reduces its magnitude.

The local energy and the drift velocity diverge as the wave function goes to zero, which is problematic at finite time steps because the probability of encountering arbitrarily large drift and branching terms is nonzero. Umrigar *et al.*⁶⁸ addressed this problem by replacing the drift velocity of each electron \mathbf{v}_i^r with

$$\bar{\mathbf{v}}_i^r = \frac{-1 + \sqrt{1 + 2a(v_i^r)^2\tau}}{a(v_i^r)^2\tau} \mathbf{v}_i^r, \tag{33}$$

where $0 < a < 1$ is a parameter of the algorithm, and the local energy in the branching factor with

$$\bar{E}_L^r = E_{\text{best}} - \frac{\sqrt{\sum_i (v_i^r)^2}}{\sqrt{\sum_i (v_i^r)^2}} (E_L^r - E_{\text{best}}), \tag{34}$$

where E_{best} is the current estimate of the DMC energy. The resulting change in the DMC Green's function is $\mathcal{O}(\tau^2)$, and hence does not affect the DMC results in the limit of zero time step. This limiting scheme successfully eliminates the stability issues, but Eq. (34) results in size-inconsistent DMC energies at fixed τ and substantial time-step errors.⁶⁹ Instead, Zen *et al.* proposed using a hard-limiting scheme,

$$\bar{E}_L^r = \begin{cases} E_{\text{best}} + E_{\text{cut}} & , \quad E_L^r > E_{\text{best}} + E_{\text{cut}} \\ E_L^r & , \quad |E_L^r - E_{\text{best}}| < E_{\text{cut}} \\ E_{\text{best}} - E_{\text{cut}} & , \quad E_L^r < E_{\text{best}} - E_{\text{cut}} \end{cases}, \tag{35}$$

where $E_{\text{cut}} = \alpha\sqrt{N/\tau}$ and $\alpha = 0.2$, which restores size consistency and greatly reduces time-step errors.⁶⁹

Ensuring that the total number of walkers remains near a finite, fixed target population W can be achieved by adjusting the energy offset E_T according to the instantaneous population. However population-control mechanisms inevitably introduce bias. Suppose the local energies are mostly less than E_0 . Then the population will try to increase, but the population-control mechanism counteracts this. Now suppose the local energies are mostly greater than E_0 . Then the population will try to decrease, but the population-control mechanism counteracts this. In either case, the average local energy increases as a result, so that population control introduces a positive bias into the DMC energy. Since fluctuations in the average local energy and population are correlated and proportional to $1/\sqrt{C}$, population-control bias goes as $1/C$.⁶⁸

Population-control bias in the DMC energy per electron is roughly independent of system size and decreases as the quality of the trial wave function is improved, because the branching factors remain close to 1 if the local energy is nearly constant in configuration space. Figure 4 demonstrates both of these effects, and also shows that population-control bias is typically small and can be removed by linear extrapolation for systems of moderate size. However, at large system size, the statistical correlation of walkers following branching events causes the target population required to eliminate population-control bias to scale exponentially with system size.⁷⁰ Issues with the convergence of the DMC energy as a function of target population have been reported for large systems with simple trial wave functions.⁷¹ It is therefore important to use good trial wave functions and restrict the application of DMC to systems with fewer than 1000–2000 electrons to keep population-control bias at bay.

No explicit steps are required to enforce the fixed-phase approximation: it is enforced implicitly by our representation of f as a probability density, ensuring it is always real and non-negative.

A typical DMC calculation has two phases: in the equilibration phase, we allow excited-state components of $|\Phi\rangle$ to die away; in the statistics-accumulation phase, we continue to propagate walkers, but gather energy data. The equilibration imaginary-time scale is often the same as the decorrelation time scale τ_{corr} over which the walker distribution is reset to the ground-state distribution. In practice we usually equilibrate over a period that is several times larger than the decorrelation time, to ensure that the fixed-phase excited-state components of the DMC wave function are exponentially small. This is discussed further in Sec. II B 4.

Noting that $E_L^r = \text{Re}(\Psi^{-1} \hat{H} \Psi) = |\Psi|^{-1} \hat{\mathcal{H}}_\chi |\Psi|$ and that $\hat{\mathcal{H}}_\chi \phi_0 = E_0 \phi_0$, the mixed estimate of the energy is equal to the fixed-phase ground-state energy E_0 , which is an upper bound on the fermion ground-state energy E_0^F :

$$\begin{aligned} \langle E_L^r \rangle_{\phi_0|\Psi} &= \frac{\int \phi_0 |\Psi| E_L^r d\mathbf{R}}{\int \phi_0 |\Psi| d\mathbf{R}} = \frac{\langle \phi_0 | \hat{\mathcal{H}}_\chi | \Psi \rangle}{\langle \phi_0 | \Psi \rangle} \\ &= E_0 \geq E_0^F. \end{aligned} \quad (36)$$

For operators \hat{A} that do not commute with the Hamiltonian the mixed estimate $\langle A_L^r \rangle_{\phi_0|\Psi}$ is not equal to the pure estimate $\langle \phi_0 | \hat{A} | \phi_0 \rangle / \langle \phi_0 | \phi_0 \rangle$. In general the difference is linear

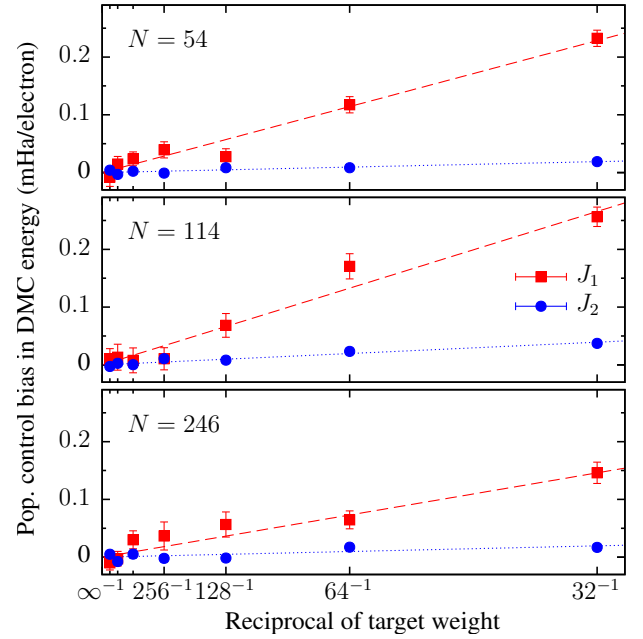


FIG. 4. Population-control bias in the Γ -point DMC energy per electron of a 3D paramagnetic electron gas of density parameter $r_s = 4$ at three different system sizes using a fixed time step of 0.1 Ha^{-1} . Two different DTN Jastrow factors were used: J_1 , consisting of a u term with cutoff length $L = L_{\text{WS}}/2$, and J_2 , consisting of a better-quality u term with $L = L_{\text{WS}}$. The lines are linear fits to the data, and the bias is measured with respect to the value of the corresponding fit at infinite target population.

in the error in the trial wave function Ψ . However, for local operators \hat{A} , the DMC mixed estimate may be combined with the VMC estimate to give a so-called extrapolated estimate, $2 \langle A_L^r \rangle_{\phi_0|\Psi} - \langle A_L^r \rangle_{|\Psi|^2}$, which has an error that is second order in the error in the trial wave function.⁷²

4. Time step and decorrelation period

The $\mathcal{O}(\tau)$ error in the mixed distribution gives an $\mathcal{O}(\tau)$ time-step bias in the mixed estimator. Time-step bias vanishes in the limit of zero time step and is linear in the time step for sufficiently small τ . Figure 5 demonstrates that the time-step bias per particle is relatively small in this linear regime, does not get more severe in larger systems, is greatly reduced if the trial wave function is good, and largely cancels out of energy differences^{69,73} if trial wave functions of similar quality are used. In order to obtain accurate total energies one must either (i) use a sufficiently small time step that the bias is negligible or (ii) perform simulations at different time steps and extrapolate to zero time step.

To be sure that we are in the small-time-step, linear-bias regime, the root-mean-square (RMS) distance diffused by each electron each time step ($\sqrt{3}\tau$) should be less than or equal to the *shortest* length scale in the system. On the other hand, the RMS distance diffused by each electron over the

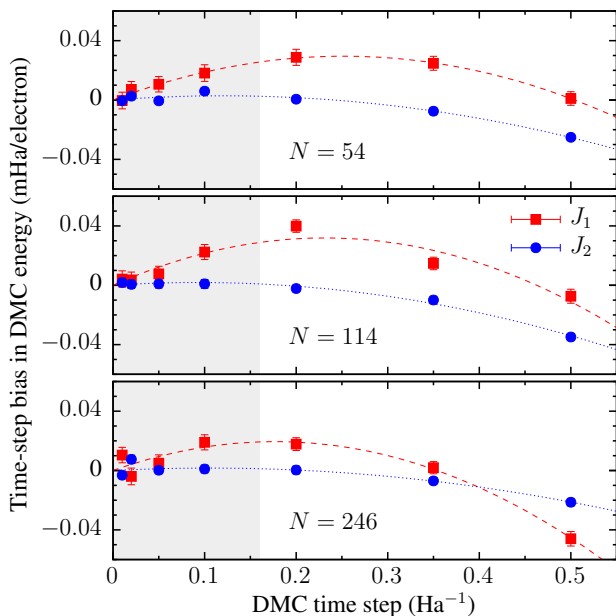


FIG. 5. Time-step bias in the Γ -point DMC energy per electron of a 3D paramagnetic electron gas of density parameter $r_s = 4$ at three different system sizes using a fixed target population of 512 walkers. Two different DTN Jastrow factors were used: J_1 , consisting of a u term with a fixed cutoff length $L = 7.82$ bohr, which equals $L_{\text{WS}}/2$ at $N = 114$, and J_2 , consisting of a better-quality u term with $L = L_{\text{WS}}$. The lines are quadratic fits to the data, and the bias is measured with respect to the value of the fit at zero time step. The shaded areas indicate the range of time steps for which the bias is expected to be linear from length-scale considerations ($\tau \lesssim 0.01\tau_s^2$).³⁵

equilibration period $\sqrt{N_{\text{eq}}\tau d}$ should be greater than or equal to the *longest* relevant length scale L in the problem, where N_{eq} is number of equilibration iterations and d is the dimensionality of the diffusion process ($d = 3$ for a cluster or a bulk material, $d = 2$ for a 2D material, and $d = 1$ for a polymer). DMC is therefore especially challenging for systems with widely separated length scales, since the shorter length scale determines the required time step, while the longer length scale determines the decorrelation period.

From a more rigorous point of view, Eq. (20) implies that the imaginary-time scale τ_{corr} over which fixed-phase DMC equilibrates is the reciprocal of the difference between the fixed-phase ground-state energy E_0 and the fixed-phase first-excited-state energy E_1 . The fixed-phase energy gap $E_1 - E_0$ is an approximation to the gap between the ground-state energy and the energy of the first excited state whose wave function transforms as the same 1D irreducible representation of the symmetry group of the Hamiltonian as the trial wave function. $E_1 - E_0$ is greater than or equal to the true energy gap. As a result, DMC usually equilibrates faster than one would expect on the basis of the actual energy gap of the material being studied.⁷⁴ After a DMC simulation has completed, the decorrelation time τ_{corr} can be estimated as the time step multiplied by the blocking length at the onset of the plateau in the reblocked error bar in a reblocking analysis of the DMC

total-energy data (see Sec. II A 1).

It is useful to convert energy gaps to length scales, since the latter are often more readily accessible for back-of-the-envelope calculations. Arguing from the infinite-square-well model, the fixed-phase energy gap satisfies $E_1 - E_0 \sim 3\pi^2/(2L^2)$ if the electrons are localized on a length scale L , and so the decorrelation period must go as $\tau_{\text{corr}} \sim 2L^2/(3\pi^2)$. The imaginary time taken for an electron to diffuse a length L in 3D is $L^2/3$, which is about five times larger than the estimate of the decorrelation time based on the infinite-square-well model; hence if we choose the equilibration period $N_{\text{eq}}\tau$ to be such that each electron diffuses over the longest relevant length scale L then we will have made the excited-state components of the initial wave function exponentially small.

For an insulator with little dispersion of the valence and conduction bands, an infinite-square-well model suggests that this longest length scale is $L \sim \sqrt{3\pi^2/(2\Delta\mathcal{E})}$, where $\Delta\mathcal{E}$ is the (actual) energy gap; this is the size of the infinite square well for which the difference between the ground-state and first-excited-state energy for electrons is equal to the gap (assuming the effective mass to be the bare mass). In this case we are simply choosing $N_{\text{eq}}\tau \sim 1/\Delta\mathcal{E}$, which errs on the side of caution, since $\Delta\mathcal{E}$ is less than or equal to the fixed-phase energy gap.⁷⁵

In metals the longest length scale diverges in principle. However, in a finite, periodic simulation cell, the linear size $L \propto N^{1/d}$ of the cell provides an (unphysical) upper bound on the longest length scale. Thus in a metallic system or a narrow-gap semiconductor, the equilibration time $N_{\text{eq}}\tau$ should be sufficiently long that the electrons can diffuse across the entire simulation cell. The required number of equilibration steps N_{eq} and the decorrelation time τ_{corr} therefore increase as the square of the linear size of the simulation cell.

These estimates of the required number of equilibration steps generally err on the side of caution. Provided the number of statistics-accumulation steps N_s is more than a couple of times the decorrelation period τ_{corr}/τ , the statistical error bars, which go as $1/\sqrt{N_s}$, are likely to dominate the bias due to limited equilibration, which goes as $1/N_s$. There may be a cancellation of biases in energy differences due to similar initial walker distributions. Furthermore, the excited-state components of the initial DMC wave function are small in magnitude if the trial wave function is accurate. For these reasons, DMC calculations with equilibration periods less than the theoretical decorrelation period may “work,” even though there is in principle insufficient time for all the excited-state components to die away. However, this shortcut cannot be taken for granted and should be tested on a case-by-case basis.

Typically we perform just two DMC calculations with different time steps. To minimize the error bar on the DMC energy extrapolated linearly to zero time step for a given computational effort (neglecting the cost of equilibration), one should choose the time steps in the ratio 1:4, and one should gather eight times as many data with the smaller time step.⁷⁶ Perhaps the best way of achieving this is by choosing the corresponding target populations to be in the ratio 4:1 (e.g., 2048 and 512 walkers), so that linear extrapolation to zero time step simultaneously extrapolates the DMC results to infinite pop-

ulation, and choosing the number of statistics accumulation steps to be in the ratio 2:1. The number of equilibration steps should be chosen to be in the ratio 4:1 to ensure that the equilibration period in imaginary time is the same for both time steps. Time steps of 0.01 and 0.04 Ha⁻¹ are typical for pseudopotential calculations.

C. Overview of the CASINO software

The CASINO software is written in Fortran 2003 and parallelized using MPI and OpenMP. To aid portability the program adheres rigidly to the Fortran 2003 standard; the only exception is the optional use of code featuring the Cray pointers extension to allow wave-function coefficients to be shared between MPI processes on the same node. All libraries used by CASINO are distributed with the code, although the BLAS and LAPACK libraries can be replaced with optimized versions where available. CASINO also includes a dummy version of the MPI library that allows the code to be compiled in serial, which is useful for debugging purposes. Porting CASINO to a new machine is usually extremely straightforward.

CASINO is accompanied by a range of utility programs written in Fortran 2003, Bash, C, C++, and Python. These are used for creating and submitting jobs on machines with queuing systems, managing large numbers of related jobs, and analyzing and plotting the data generated. The distribution also includes a detailed manual and a set of examples. The examples include a suite of test cases that can be run automatically and compared with library output files, to check for regressions. CASINO is also supported by a website,⁷⁷ which includes a discussion forum for users and libraries of pseudopotentials. Pseudopotentials are discussed further in Sec. III C.

III. SOME RECENT DEVELOPMENTS IN QMC AND CASINO

A. Advanced wave-function forms

1. Generalized Jastrow factors

As mentioned in Sec. II A 2, Jastrow factors $e^{J(\mathbf{R})}$ are applied to determinantal trial wave functions Ψ_S in order to describe dynamical electronic correlations. The optimization of wave-function parameters that affect the nodal surface, such as backflow parameters or determinant coefficients, is greatly aided by the use of a good Jastrow factor. Also, there are systems for which it is undesirable to use the DMC method, e.g., due to computational expense or pseudopotential locality biases, and cases in which the Jastrow factor itself is the object of interest; see Sec. V A 2. While the DTN Jastrow factor is adequate in most cases, there are systems for which it is helpful to use higher-order terms, such as three- or four-electron terms, or expansions in a basis other than natural powers or cosines.

In order to cater for these needs, CASINO implements a Jastrow-factor construction framework, described in detail in Ref. 78, which can express a rich variety of forms of correlation involving any number of electrons and nuclei. Within this framework, a Jastrow exponent term is expressed as an expansion in electron-electron basis functions $\Phi_{\nu_{ij}}^{P_{ij}}(r_{ij})$ and electron-nucleus basis functions $\Theta_{\mu_{iI}}^{S_{iI}}(r_{iI})$, where $\nu_{ij} \in \{1, \dots, p\}$ and $\mu_{iI} \in \{1, \dots, q\}$ are expansion indices, and p and q are expansion orders. For simplicity, any cutoff functions are factorized into Φ and Θ . An n -electron m -nucleus Jastrow factor term for a system with N electrons and N_n nuclei is thus expressed as

$$J_{n,m}(\mathbf{R}) = \sum_{\mathbf{i}}^N \sum_{\mathbf{I}}^{N_n} \sum_{\boldsymbol{\nu}}^p \sum_{\boldsymbol{\mu}}^q \lambda_{\boldsymbol{\nu}\boldsymbol{\mu}}^{\mathbf{P}\mathbf{S}} \prod_{\alpha < \beta}^n \Phi_{\nu_{\alpha\beta}}^{P_{\alpha\beta}}(\mathbf{r}_{i_{\alpha}i_{\beta}}) \prod_{\alpha,\gamma}^{n,m} \Theta_{\mu_{\alpha I_{\gamma}}}^{S_{\alpha I_{\gamma}}}(\mathbf{r}_{i_{\alpha}I_{\gamma}}), \quad (37)$$

where \mathbf{i} is a vector of n distinct electron indices $i_1 < \dots < i_n$, \mathbf{I} is a vector of m distinct nucleus indices $I_1 < \dots < I_m$, $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are arrays of electron-electron and electron-nucleus expansion indices, \mathbf{P} and \mathbf{S} are arrays of electron-electron and electron-nucleus dependency indices, and λ are the linear parameters of the expansion. Symmetry and cusp constraints on the linear parameters can be imposed automatically for any choice of basis functions, expansion orders, etc., making the implementation of new basis functions a relatively trivial task.⁷⁹

The DTN Jastrow factor terms can be obtained with Eq. (37) by setting Φ and/or Θ to natural powers of r multiplied by $t(r)$, and higher-order Jastrow factor terms can be trivially constructed by increasing n and/or m . To illustrate the usefulness of these higher-order terms, in Fig. 6 we show the VMC energies and variances obtained for the positronium dimer,⁸⁰ Ps₂, a metastable system consisting of two electrons and two positrons. We use an electron-positron pairing wave function $\Psi_S = \phi_1(r_{e\uparrow p\uparrow})\phi_1(r_{e\downarrow p\downarrow})\phi_2(r_{e\uparrow p\downarrow})\phi_2(r_{e\downarrow p\uparrow}) + \phi_2(r_{e\uparrow p\uparrow})\phi_2(r_{e\downarrow p\downarrow})\phi_1(r_{e\uparrow p\downarrow})\phi_1(r_{e\downarrow p\uparrow})$, where $\phi_1(r)$ and $\phi_2(r)$ are pairing orbitals with independent parameter values, multiplied by Jastrow factors involving up to four-body correlations. Our lowest VMC energy corresponds to 99.53(2)% of the binding energy of Ps₂. Since the exact wave function is nodeless, an extrapolation to zero VMC variance⁵⁴ can be expected to yield a reasonable estimate of the ground-state energy. We find that this extrapolated energy accounts for 99.94(2)% of the binding energy of Ps₂. Note that the DMC method is exact for this system.

Rather more “exotic” forms of Jastrow factors can be constructed with Eq. (37), of which Ref. 78 gives several examples including electron-nucleus-nucleus terms for polyatomic systems and electron-electron-nucleus-nucleus terms to describe van der Waals interactions. Here we present an additional example of a useful Jastrow factor term, which is a DTN electron-electron-nucleus term in which f is augmented

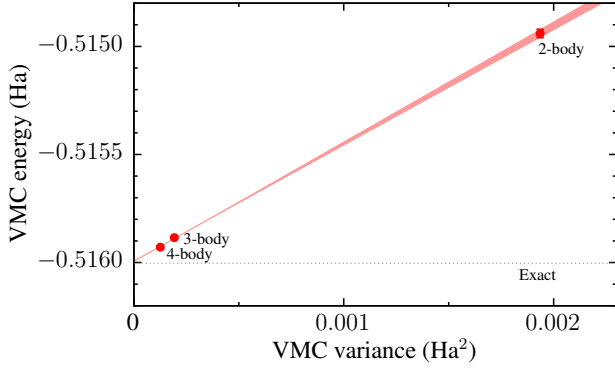


FIG. 6. VMC energy as a function of VMC variance for Ps_2 using Jastrow factors including up to 2-, 3-, and 4-body isotropic correlations. A linear fit is shown as a line of variable width indicating the statistical uncertainty in the fit value. The exact energy of $-0.5160038 \text{ Ha}^{81,82}$ is shown as a dotted line for reference.

by

$$g(r_{ij}, \mathbf{r}_{iI}, \mathbf{r}_{jI}) = t(r_{iI})t(r_{jI}) \sum_{k,l,m} c_{klm} r_{ij}^k r_{iI}^l r_{jI}^m \mathbf{r}_{iI} \cdot \mathbf{r}_{jI}, \quad (38)$$

where c are the linear parameters. The dot product in Eq. (38) can be recovered in Eq. (37) by combining specially crafted basis functions with constraints on the linear parameters, and provides the ability to distinguish configurations in which electrons i and j are on the same side or opposite sides of nucleus I , which is difficult to do with isotropic functions. We have tested this Jastrow factor on the all-electron silicon atom using HF orbitals expanded in the cc-pVQZ Gaussian basis set⁸³ obtained using the MOLPRO code.⁸⁴ First we optimized the u , χ , and f terms in the DTN Jastrow factor, and then we fixed the parameters in u and χ and optimized the g term alongside the f term. The difference between the DTN and DTN+ g Jastrow factors is plotted in Fig. 7 as a function of the position of an electron as another is held at a fixed position. The g term provides a nontrivial correction to the DTN Jastrow factor, resulting in a total energy reduction of 6.4(9) mHa in this example.

2. Multideterminant wave functions

Multideterminant expansions provide systematic convergence of the trial wave function, and hence its nodal surface, towards the ground state. Let $\{\phi_l(\mathbf{r})\}$ be a set of orbitals that form a complete basis for 3D functions. Then the set of all possible $N \times N$ Slater determinants constructed with these orbitals form a complete basis for antisymmetric $3N$ -dimensional functions. The configuration interaction (CI) wave function

$$\Psi_{\text{CI}}(\mathbf{R}) = \sum_k^{N_{\text{det}}} c_k \det \left[\phi_{a_{ik}}^{\uparrow}(\mathbf{r}_j^{\uparrow}) \right] \det \left[\phi_{a_{ik}}^{\downarrow}(\mathbf{r}_j^{\downarrow}) \right], \quad (39)$$

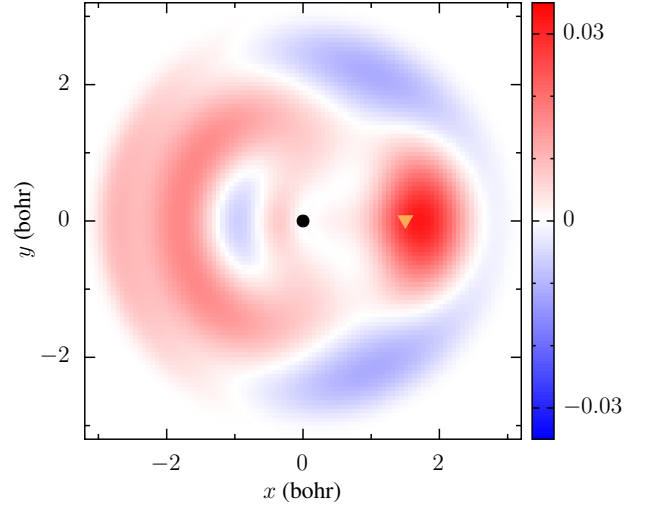


FIG. 7. Difference between the value of a DTN Jastrow factor and one containing an additional electron-electron-nucleus term [Eq. (38)] as an up-spin electron of a silicon atom scans the $z = 0$ plane. The nucleus is at the origin (circle), a down-spin electron is fixed at (1.5, 0, 0) bohr (triangle), and the twelve remaining electrons have been placed away from the nucleus so that they do not contribute to the value of the Jastrow factor.

where $\{a_{ik}^{\sigma}\}$ are indices selecting which orbitals appear in the k th determinant of spin σ and $\{c_k\}$ are unknown coefficients, is thus guaranteed to converge to the ground-state wave function in the full-CI limit $N_{\text{det}} \rightarrow \infty$.

Multideterminantal wave functions have been used in QMC calculations for many years,⁸⁵ but interest in multideterminants has surged recently as a potential avenue towards systematic alleviation of the fixed-node error.^{86–88} Trial wave functions obtained from selected CI approaches appear to deliver better results than those obtained from complete-active-space calculations.⁸⁹ Our take on this is discussed in Sec. VA 1. Some of these approaches require very large expansions including millions of Slater determinants and greatly benefit from the use of specialized techniques for the rapid evaluation of the trial wave function and its derivatives.

There are a number of dedicated multideterminant wave-function-evaluation algorithms.^{90–94} CASINO implements the compression method of Weerasinghe *et al.*⁹⁵ Although this method underperforms some of the other approaches, it can potentially be combined with them for additional efficiency since it does not require modification to the evaluation of the wave function. The compression algorithm is based on combining determinants differing by a single column, e.g.,

$$\begin{aligned} & \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_3(\mathbf{r}_1) & \dots \\ \phi_1(\mathbf{r}_2) & \phi_3(\mathbf{r}_2) & \dots \\ \vdots & \vdots & \ddots \end{vmatrix} + \begin{vmatrix} \phi_2(\mathbf{r}_1) & \phi_3(\mathbf{r}_1) & \dots \\ \phi_2(\mathbf{r}_2) & \phi_3(\mathbf{r}_2) & \dots \\ \vdots & \vdots & \ddots \end{vmatrix} \\ & = \begin{vmatrix} \phi_1(\mathbf{r}_1) + \phi_2(\mathbf{r}_1) & \phi_3(\mathbf{r}_1) & \dots \\ \phi_1(\mathbf{r}_2) + \phi_2(\mathbf{r}_2) & \phi_3(\mathbf{r}_2) & \dots \\ \vdots & \vdots & \ddots \end{vmatrix}, \end{aligned} \quad (40)$$

throughout the expansion. There are multiple ways in which the N_{det} determinants in the original expansion can be grouped together, and finding the one that minimizes the number of determinants N_{comp} in the resulting wave function is a set-covering problem. Although the cost of solving a set-covering problem exactly, which can be done with linear programming, scales nonpolynomially with N_{det} , compressing wave functions containing tens of thousands of determinants typically takes seconds. Alternatively, a heuristic “greedy” algorithm can be used to approximately solve the set-covering problem in polynomial time.

The application of compression yields significantly shorter expansions at the cost of requiring the evaluation of linear combinations of the orbitals [e.g., $\phi_1 + \phi_2$ in Eq. (40)] when computing the trial wave function and its derivatives. In practice, this amounts to a tiny fraction of the computer time, and the overall speedup attained by using compressed expansions is very close to $N_{\text{det}}/N_{\text{comp}}$. Moreover, the size of the compressed expansion is found to scale sublinearly with the size of the original expansion (i.e., $N_{\text{comp}} \propto N_{\text{det}}^\alpha$ with $\alpha < 1$), implying that the computational cost of the QMC calculation per determinant decreases with N_{det} . This is also a property of some of the other acceleration methods reported in the literature.^{91,93}

We show an example of the scaling of the compressed expansion size as a function of N_{det} in Fig. 8 for the C_2 molecule at its equilibrium geometry using HF orbitals expanded in the cc-pCVTZ basis set⁸³ obtained using MOLPRO.⁸⁴ The configuration-state-function-based FCIQMC method⁹⁶ as implemented in the NECI package^{25,26} was used to generate the original expansion. Fitting the data yields an exponent $\alpha = 0.895 < 1$, consistent with the sublinear relation between N_{comp} and N_{det} .

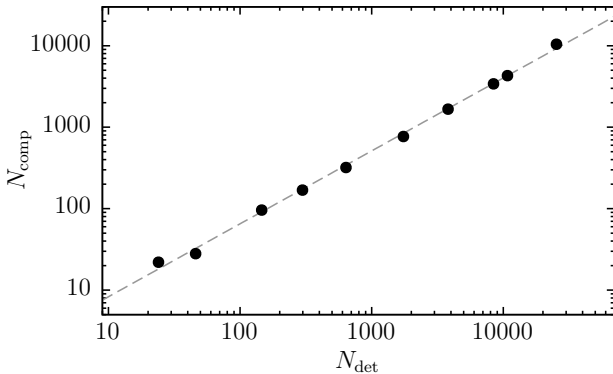


FIG. 8. Number of determinants in the compressed expansion N_{comp} as a function of the number of determinants in the original expansion N_{det} for the C_2 molecule at its equilibrium geometry. A linear fit to the log-log data is shown as a dashed line.

3. Geminal and pairing wave functions

An alternative way of describing static electronic correlation in the trial wave function is to replace the Slater determinants with geminals. The antisymmetrized geminal power (AGP) wave function^{97–99} is

$$\Psi_{\text{AGP}}(\mathbf{R}) = \det \left(\sum_{l,m} g_{lm} \phi_l(\mathbf{r}_i^\uparrow) \phi_m(\mathbf{r}_j^\downarrow) \right), \quad (41)$$

assuming $N_{e\uparrow} = N_{e\downarrow}$, where $\{\phi_l(\mathbf{r})\}$ is a set of n_{orb} orbitals and g_{lm} are optimizable coefficients. If $n_{\text{orb}} = N_{e\uparrow}$, the AGP wave function trivially factorizes into up- and down-spin Slater determinants regardless of the (nonzero) values of the coefficients. For $n_{\text{orb}} > N_{e\uparrow}$ this is no longer the case, and some degree of static correlation is built into the wave function through pairing of up- and down-spin electrons. Much in the same spirit, Pfaffians represent an alternative wave-function form in which the description of correlation is based on same-spin-electron pairing.¹⁰⁰ In QMC calculations, the AGP wave function has been shown to yield a small advantage over Slater determinants.^{101,102}

An alternative to the AGP wave function is the use of an antisymmetrized product of geminals,^{97,103} but the cost of evaluating this wave function in QMC scales exponentially with $N_{e\uparrow}$. Instead, CASINO implements a multi-AGP (MAGP) wave function,¹⁰⁴ which is simply a linear combination of AGPs with independent parameters. In Ref. 104, an MAGP wave function was constructed to reproduce the coefficients of a CI wave function including up to double excitations for the HEG, and DMC energies obtained combining the MAGP wave function with a backflow transformation were found to be competitive with those from FCIQMC,¹⁰⁵ albeit at significant computational expense. For a comparison of this result with that obtained using multideterminantal wave functions, see Fig. 26 below.

In the presence of attractive interactions, for instance in a system consisting of N_e negatively charged electrons and $N_h = N_e$ positively charged holes, it is natural to employ a simpler form of pairing wave function,

$$\Psi_S(\mathbf{R}) = \det [\phi(\mathbf{e}_j^\uparrow - \mathbf{h}_i^\downarrow)] \det [\phi(\mathbf{e}_j^\downarrow - \mathbf{h}_i^\uparrow)], \quad (42)$$

where \mathbf{e}_j^σ is the position vector of the j th electron of spin σ , \mathbf{h}_i^σ is the position vector of the i th hole of spin σ , and the pairing orbital ϕ is a parameterized function of the difference between the electron and hole position vectors. Usually the pairing orbital is assumed to depend solely on the electron-hole distance, $\phi(\mathbf{r}) = \phi(r)$, but CASINO also implements a pairing orbital of the form^{106,107}

$$\phi(\mathbf{r}) = \sum_l p_l \exp(i\mathbf{G}_l \cdot \mathbf{r}) + u(r), \quad (43)$$

where \mathbf{G}_l is the l th shortest reciprocal lattice vector, $\{p_l\}$ are optimizable coefficients, and $u(r)$ is a parameterized polynomial of the same form as the u function in the DTN Jastrow factor. This orbital confers on the pairing wave function a

property of the AGP wave function: if $u = 0$ and the number of plane waves equals the number of particles per spin, the wave function can be factorized as the product of electron and hole determinants of plane-wave orbitals. This pairing wave function is therefore capable of describing both a two-component fluid and an excitonic fluid, and is found to provide a particularly good description of the system near the transition between these two phases.¹⁰⁷

The pairing wave function of Eq. (42) can be adapted to systems with an imbalance in the number of electrons and holes by completing the pairing matrix with plane waves.¹⁰⁸ The extreme case of a single hole immersed in an electron gas is of particular interest. For this system CASINO implements a specialized pairing wave function,¹⁰⁹

$$\Psi_S(\mathbf{R}) = \det \left[\phi_i(\mathbf{e}_j^\uparrow - \mathbf{h}) \right] \det \left[\phi_i(\mathbf{e}_j^\downarrow - \mathbf{h}) \right], \quad (44)$$

where \mathbf{h} is the position of the hole and the pairing orbital is

$$\phi_i(\mathbf{r}) = \exp[u_{G_i}(r)] \exp\{i\mathbf{G}_i \cdot \mathbf{r} [1 - \eta_{G_i}(r)/r]\}, \quad (45)$$

where u_{G_i} and η_{G_i} are parameterized polynomials of the same functional form as the u function in the DTN Jastrow factor. This orbital is essentially a plane wave of the electron-hole distance with an orbital-specific Jastrow-like prefactor and an electron-hole backflow transformation.

B. QMC calculations for condensed matter

1. Single-particle finite-size effects: momentum quantization

Condensed matter physics and materials science are largely concerned with the bulk properties of extended systems of 10^4 – 10^{26} constituent particles. On an atomic scale such systems are of effectively infinite extent and most of their properties are well described by equilibrium thermodynamics. In studies of condensed matter, the challenge for QMC methods is the need to obtain results that are valid in the thermodynamic limit of infinite system size despite only being able to perform calculations for finite systems with hundreds or thousands of constituent particles. In this section we will discuss finite-size effects in the context of crystalline solids, although everything we state applies equally well to the case of homogeneous liquids, in which the size of the primitive unit cell is vanishingly small. Unlike DFT, QMC is an explicitly correlated wave-function-based method, so that the electronic Schrödinger equation cannot be reduced to a single periodic unit cell. Instead one must construct a supercell of multiple primitive cells. In the following we use N to denote the number of electrons in a simulation supercell.

One way of approaching the thermodynamic limit, often used in quantum chemistry, is to study finite clusters of bulk crystalline solids using hydrogen atoms to passivate dangling bonds on the surface. In this case the translational symmetry of the bulk crystal is lost and unwanted surface effects are

introduced. An alternative approach is to use a finite simulation supercell of the crystal subject to periodic boundary conditions. This avoids unwanted surfaces and restores translational symmetry, but introduces spurious interactions between periodic images of particles. Periodic boundary conditions usually provide a much more reliable means for extrapolating results to infinite system size.

Simulation supercells are traditionally constructed using the primitive cell as a repeating unit, with each supercell lattice vector being an integer multiple of the corresponding primitive-cell lattice vector. However one is free to choose supercell lattice vectors to be integer linear combinations of the primitive-cell lattice vectors. The use of nondiagonal supercells that maximize the nearest-image distance of the simulation supercell yields reduced finite-size effects in energy differences, and grants access to a broader choice of system sizes.^{21,110–112}

Consider a supercell of a crystal subject to periodic boundary conditions. Translations of individual electrons through supercell lattice vectors and simultaneous translations of all electrons through primitive lattice vectors are symmetry operations and therefore commute with the many-electron Hamiltonian. Eigenfunctions of the Hamiltonian are therefore eigenfunctions of the translation operators. The requirement that the wave function is an eigenfunction of each of these translation operators leads to two many-body Bloch conditions on the wave function:

$$\Psi(\mathbf{R}) = e^{i\mathbf{k}_s \cdot \sum_{i=1}^N \mathbf{r}_i} U(\mathbf{R}) \quad (46)$$

$$= e^{i\mathbf{k}_p \cdot \sum_{i=1}^N \mathbf{r}_i / N} W(\mathbf{R}), \quad (47)$$

where U is invariant under a translation of any electron through any supercell lattice vector \mathbf{R}_s , while W is invariant under a simultaneous translation of all electrons through any primitive-lattice vector \mathbf{R}_p , and \mathbf{k}_s and \mathbf{k}_p are wavevectors in the first Brillouin zones of the simulation cell and the primitive cell, respectively.^{113,114}

As usual, we impose these symmetry requirements via the Slater part of the trial wave function, and choose our Jastrow factor to be fully symmetric. In particular we can simultaneously satisfy Eqs. (46) and (47) by choosing the orbitals in each Slater determinant to be of Bloch form, with the Bloch wavevectors lying on a grid of simulation-cell reciprocal lattice points offset from Γ by \mathbf{k}_s , for any \mathbf{k}_s in the first Brillouin zone of the simulation cell.

In the infinite system, the Bloch wavevectors of the orbitals form a continuum and in a single-particle theory the total energy per particle can be expressed in terms of integrals over wavevector \mathbf{k} . In the finite supercell subject to periodic boundary conditions the integrals are replaced by sums over a discrete set of \mathbf{k} points. The difference between the discrete sums for the finite supercell and the integrals for the infinite system varies in a quasirandom manner with system size. These quasirandom fluctuations are particularly large in metallic systems, where the number of occupied shells of \mathbf{k} points changes as the simulation cell is increased and hence the grid of \mathbf{k} points gets finer. In the context of a single-particle theory, these finite-size effects are regarded as \mathbf{k} -point

sampling errors. To a first approximation, when looking at energy differences, including electron-electron interactions and hence correlations in the wave function simply renormalizes the effective mass of the electrons.¹¹⁵⁻¹¹⁷ Hence similar oscillations in energy per particle as a function of system size are seen in QMC simulations as in uncorrelated single-particle theories, but usually with a different amplitude. One approach for reducing momentum-quantization finite-size errors is to treat the DFT energy or HF kinetic energy obtained using the \mathbf{k} -point grid corresponding to a QMC supercell relative to the DFT energy with a fine \mathbf{k} -point mesh as a covariate in a fit to the QMC energy as a function of system size; such a fit allows extrapolation to infinite system size.¹¹⁸

In general, however, it is better to average over supercell Bloch wavevectors \mathbf{k}_s (i.e., to average over offsets to the grid of single-particle \mathbf{k} points) for a given supercell size; this process is referred to as *twist averaging*.¹¹⁹ Averaging over a large number of offsets to the grid of \mathbf{k} points turns a sum over discrete \mathbf{k} into an integral. For a metallic system, the shape of the twist-averaged Fermi surface is a polyhedron with (in the case of the HEG) the same volume as the real Fermi surface; the incorrect shape leads to a relatively small, positive, quasirandom finite-size effect. It is possible to use DFT energies calculated using exactly the same \mathbf{k} points as the QMC calculations relative to the DFT energy with a fine \mathbf{k} -point grid as a covariate in a fit to the QMC energies as a function of twist; this simultaneously removes both noise due to the finite number of twists and the residual bias in the twist-averaged energy.

2. Long-range finite-size effects

There are additional finite-size effects in a correlated QMC calculation that are absent from a DFT calculation. (Strictly

speaking this is because the XC functional in a DFT calculation is, by construction, appropriate for an infinite system; thus DFT provides an incorrect description of a finite cell subject to periodic boundary conditions.) The physical origins of these finite-size effects lie in (i) the treatment of the Coulomb interaction in a finite supercell and (ii) the fact that long-range correlation effects cannot be described correctly in a finite supercell.

Let us first consider Coulomb finite-size effects. The usual treatment of the Coulomb interaction in a periodic cell is to write the interaction between each pair of charges using the Ewald interaction $v_E(\mathbf{r})$, which is the periodic solution of Poisson's equation around a point charge, with Fourier components $4\pi/k^2$.¹²⁰ In addition, the electrostatic energy of the lattice of periodic images of each charge in the supercell, known as the Madelung constant, is included in the energy. In practice the Ewald interaction is evaluated as the sum of two rapidly convergent series in supercell lattice vectors and in supercell reciprocal lattice vectors. The use of the periodic solution to Poisson's equation amounts to the assumption that there is no macroscopic electric field within the crystal; this could be guaranteed by embedding a macroscopic sample of the crystal in a perfect metal so that there are no surface polarization charges.

The expectation value of the potential-energy operator can be decomposed into an interaction between the fixed nuclei and the electronic charge density, plus a Hartree energy given by the classical electrostatic potential energy of the electronic charge density, plus the Coulomb XC energy, which is everything else:

$$\begin{aligned} \langle \hat{V} \rangle &= \sum_i \sum_I -Z_I \langle v_E(\hat{\mathbf{r}}_{iI}) \rangle + \sum_{i>j} \langle v_E(\hat{\mathbf{r}}_{ij}) \rangle + \frac{v_M}{2} \left(N + \sum_I Z_I^2 \right) \\ &= - \sum_I Z_I \int v_E(\mathbf{r} - \mathbf{r}_I) \rho(\mathbf{r}) d\mathbf{r} + \frac{v_M}{2} \sum_I Z_I^2 + \frac{1}{2} \iint v_E(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}) \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \\ &\quad + \frac{N}{2} \int [v_E(\mathbf{r}) - v_M] \frac{1}{N} \int [\rho_2(\mathbf{r}' + \mathbf{r}, \mathbf{r}') - \rho(\mathbf{r}' + \mathbf{r}) \rho(\mathbf{r}')] d\mathbf{r}' d\mathbf{r}, \end{aligned} \quad (48)$$

where the charge density is $\rho(\mathbf{r}) = \langle \sum_i \delta(\mathbf{r} - \hat{\mathbf{r}}_i) \rangle$ and the pair density is $\rho_2(\mathbf{r}, \mathbf{r}') = \langle \sum_{i \neq j} \delta(\mathbf{r} - \hat{\mathbf{r}}_i) \delta(\mathbf{r}' - \hat{\mathbf{r}}_j) \rangle$. The Coulomb XC energy per electron is therefore the Coulomb interaction between a point charge at the origin and the system-averaged XC hole $\rho_{xc}(\mathbf{r}) \equiv (1/N) \int [\rho_2(\mathbf{r}' + \mathbf{r}, \mathbf{r}') - \rho(\mathbf{r}' + \mathbf{r}) \rho(\mathbf{r}')] d\mathbf{r}'$, which is the average effective charge density around each electron due to correlations with all the others. The long-range behavior of the XC hole is well-described by the random phase approxi-

mation, i.e., linear-response theory of the noninteracting electron system in the rest frame of one of the electrons.

The electronic charge density ρ has the periodicity of the primitive unit cell and generally converges extremely quickly with simulation-cell size; hence there is very little finite-size error in the electron-nucleus potential energy and in the Hartree energy. However, the finite-size errors in the Coulomb XC energy can be substantial.

One source of error is that the XC hole $\rho_{xc}(\mathbf{r})$ is long-ranged, with the systematic part falling off as r^{-8} in an infi-

nite system;¹²¹ this tail is truncated at finite range in a periodic cell. On top of this there are long-range Friedel oscillations in the XC hole, which result in very small quasirandom finite-size errors in the energy when the oscillations are forced to be commensurate with the periodic simulation supercell. Evaluating the Coulomb interaction between the electron at the origin and the missing r^{-8} tail of the surrounding XC hole in a finite cell, we find that the resulting finite-size error per particle falls off as N^{-2} .

However a much more significant source of finite-size error is the form of interaction between the electron and the XC hole. In the infinite-system limit this should simply be $1/r$; in the finite cell it is $v_E(\mathbf{r}) - v_M$. A power series expansion of the Ewald interaction^{122,123} shows that $v_E(\mathbf{r}) - v_M = 1/r + \mathcal{O}(r^2/N)$. The correction term in this expansion results in an error that goes as N^{-1} in the Coulomb XC energy per particle.

Remarkably, it is possible to eliminate the leading-order finite-size errors in the Coulomb XC energy by replacing the Ewald interaction with a model periodic Coulomb interaction that causes the Hartree energy to be evaluated using the Ewald potential while the Coulomb XC energy is evaluated using the $1/r$ interaction within the simulation cell.¹²³⁻¹²⁵ The residual finite-size errors in the XC energy per particle are therefore $\mathcal{O}(N^{-2})$ due to the truncation of the XC hole.

An alternative perspective on the problem of finite-size errors is found by transforming to reciprocal space.¹²⁶ The XC energy per particle can be written as

$$\begin{aligned} & \frac{1}{2} \int [v_E(\mathbf{r}) - v_M] \rho_{xc}(\mathbf{r}) d\mathbf{r} \\ &= \frac{1}{2} \left(\frac{1}{\Omega_s} \sum_{\mathbf{G}_s \neq 0} \frac{4\pi}{G_s^2} [S(\mathbf{G}_s) - 1] + v_M \right), \end{aligned} \quad (49)$$

where the static structure factor is $S(\mathbf{k}) = 1 + \int \rho_{xc}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}$ and Ω_s is the volume of the simulation supercell. Since the XC hole is rapidly convergent, so is the structure factor; hence the leading-order finite-size correction to the potential energy per particle is due to the difference between an integral and sum:

$$\begin{aligned} \frac{\Delta V_{xc}}{N} &= \frac{1}{2} \left(\frac{1}{(2\pi)^3} \int \frac{4\pi}{k^2} [S(\mathbf{k}) - 1] d\mathbf{k} \right) \\ &\quad - \frac{1}{2} \left(\sum_{\mathbf{G}_s \neq 0} \frac{4\pi}{\Omega_s G_s^2} [S(\mathbf{G}_s) - 1] + v_M \right) \\ &= \frac{1}{4\pi^2} \int \frac{S(\mathbf{k})}{k^2} d\mathbf{k} - \frac{2\pi}{\Omega_s} \sum_{\mathbf{G}_s \neq 0} \frac{S(\mathbf{G}_s)}{G_s^2} \\ &\approx \frac{2\pi}{\Omega_s} \lim_{\mathbf{k} \rightarrow 0} \frac{\bar{S}(k)}{k^2}, \end{aligned} \quad (50)$$

where \bar{S} is the spherical average of the structure factor and we have used the fact that

$$v_M = \lim_{\epsilon \rightarrow 0} \left(\frac{4\pi}{\Omega_s} \sum_{\mathbf{G}_s \neq 0} \frac{e^{-\epsilon G_s^2}}{G_s^2} - \frac{1}{2\pi^2} \int \frac{e^{-\epsilon k^2}}{k^2} d\mathbf{k} \right). \quad (51)$$

The spherically averaged structure factor is quadratic at short range, and $\lim_{k \rightarrow 0} \bar{S}(k)/k^2$ may be evaluated using the structure factor accumulated in a QMC calculation. Indeed for simple systems such as the electron gas, $\lim_{k \rightarrow 0} \bar{S}(k)/k^2$ can be calculated analytically using the RPA.¹²⁷ If a system has cubic symmetry then the error in the approximation to the finite-size correction is the $\mathcal{O}(N^{-2})$ error due to the neglected tail of the XC hole; in noncubic systems the finite-size correction using the spherical average of the structure factor does not fully remove the leading-order error.

We now turn our attention to the neglect of long-range correlation effects. The two-body correlations described by the Jastrow factor are long-range. They are restricted in a finite simulation cell, leading to bias in kinetic energy. We may correct for this by interpolating the Fourier transformation of the two-body Jastrow factor.¹²⁶ Let us write Ψ as the product of a long-range two-body Jastrow factor, which has the periodicity of the simulation cell and inversion symmetry, and a part consisting of everything else, Ψ_s :

$$\begin{aligned} \Psi &= \Psi_s \exp \left(\sum_{i>j} u(\mathbf{r}_i - \mathbf{r}_j) \right), \\ &= \Psi_s \exp \left(\sum_{\mathbf{G}_s \neq 0} \frac{u(\mathbf{G}_s) \hat{\rho}^*(\mathbf{G}_s) \hat{\rho}(\mathbf{G}_s)}{2\Omega_s} + K \right). \end{aligned} \quad (52)$$

The kinetic energy per particle may be evaluated as the average of

$$\begin{aligned} \frac{\hat{T}}{N} &= \frac{-1}{4N} \nabla^2 \ln(\Psi) \\ &= \frac{\hat{T}_s}{N} - \frac{1}{8N\Omega_s} \sum_{\mathbf{G}_s \neq 0} u(\mathbf{G}_s) \nabla^2 [\hat{\rho}^*(\mathbf{G}_s) \hat{\rho}(\mathbf{G}_s)], \end{aligned} \quad (53)$$

where $\hat{T}_s/N = -\nabla^2 \ln(\Psi_s)/(4N)$. Noting that

$$\nabla^2 [\hat{\rho}^*(\mathbf{G}_s) \hat{\rho}(\mathbf{G}_s)] = -2G_s^2 [\hat{\rho}^*(\mathbf{G}_s) \hat{\rho}(\mathbf{G}_s) - N], \quad (54)$$

we find that

$$\frac{\langle \hat{T} \rangle}{N} = \frac{\langle \hat{T}_s \rangle}{N} + \frac{1}{4N\Omega_s} \sum_{\mathbf{G}_s \neq 0} G_s^2 u(\mathbf{G}_s) [\langle \hat{\rho}^*(\mathbf{G}_s) \hat{\rho}(\mathbf{G}_s) \rangle - N]. \quad (55)$$

Now $\rho(\mathbf{k})$ is only nonzero for reciprocal lattice vectors of the primitive lattice. Assuming the sum runs only over small \mathbf{G}_s ,

$$\begin{aligned} \frac{\langle \hat{T} \rangle}{N} &= \frac{\langle \hat{T}_s \rangle}{N} + \frac{1}{4\Omega_s} \sum_{\mathbf{G}_s \neq 0} G_s^2 u(\mathbf{G}_s) S^*(\mathbf{G}_s) \\ &\quad - \frac{1}{4\Omega_s} \sum_{\mathbf{G}_s \neq 0} G_s^2 u(\mathbf{G}_s). \end{aligned} \quad (56)$$

$u(\mathbf{k})$ has same form at different N . It diverges as k^{-2} ,^{47,127} so $\lim_{k \rightarrow 0} k^2 \bar{u}(k)$ exists: see Fig. 9.

In the infinite-system limit, the sum over \mathbf{G}_s should be replaced by an integral in Eq. (56).¹²⁶ The leading-order finite-size corrections are due to the differences between these integrals and sums. The RPA suggests that we can approximate

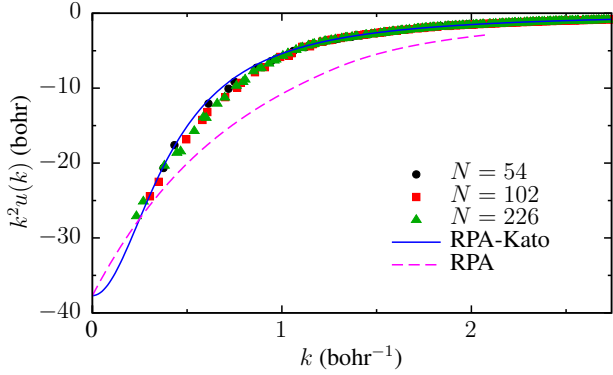


FIG. 9. Fourier transform of the optimized two-body Jastrow factor for a paramagnetic HEG of density parameter $r_s = 3$ at different system sizes N , together with the RPA result, and an interpolation between the leading-order RPA behavior and the short-range Kato behavior.

$\bar{u}(k)$ by $-4\pi(A/k^2 + B/k)$ at small k in 3D, where A and B are constants.¹²⁷ In CASINO the constants are determined by fitting this model of \bar{u} to the optimized two-body Jastrow terms at the first few nonzero stars of \mathbf{G}_s vectors. We evaluate the finite-size correction to the kinetic energy per electron in 3D as

$$\frac{\Delta T}{N} = \frac{\pi A}{\Omega_s} + \frac{C_{3D} B}{\Omega_s^{4/3}}, \quad (57)$$

where

$$C_{3D} = \frac{\Omega_s^{4/3}}{4} \lim_{\alpha \rightarrow 0} \left[\frac{1}{\pi \alpha^2} - \frac{4\pi}{\Omega_s} \sum_{\mathbf{G}_s \neq 0} G_s \exp(-\alpha G_s^2) \right] \quad (58)$$

is a lattice-specific constant, which can be evaluated by brute force. The leading-order finite-size error in the kinetic energy per electron falls off as $\mathcal{O}(N^{-1})$.¹²⁸ Note that it has been shown that the use of backflow alters the coefficient of the $\mathcal{O}(N^{-1})$ contribution to the finite-size error in the energy per electron.¹²⁹

The work that has been done in extrapolating HEG results to infinite system size can be used to correct finite-size errors in inhomogeneous systems using DFT with a special finite-size local density approximation (LDA) functional.¹³⁰ In the finite-size LDA calculations, the XC energy per particle at a point in space with density ρ is given by the XC energy per particle of a finite electron gas of density ρ in the supercell obtained by unfolding the \mathbf{k} points. The finite-size correction is the difference between a the DFT energy with the usual (infinite-system) LDA and a fine \mathbf{k} -point mesh and the finite-size LDA energy with a \mathbf{k} -point sampling corresponding to the QMC calculation to be corrected. Single-particle and long-range finite-size errors are corrected simultaneously in this approach. It is straightforward to use this method in conjunction with twist averaging. In practice, the available parameterization of the finite-system LDA means that long-range effects due to the shape of the simulation cell are neglected.¹³⁰

3. Finite-size effects in properties other than the total energy

Most expectation values, such as the pair-correlation function and structure factor, can be twist averaged to reduce single-particle finite-size effects. However, for some expectation values such as the momentum density, the choice of simulation-cell Bloch wavevector merely defines the set of \mathbf{k} points at which the momentum density can be evaluated. Performing calculations at multiple twists therefore gives additional points on the momentum density. The MD of metallic systems has been shown to exhibit slowly decaying finite-size errors near the Fermi edge.¹³¹ In expectation values such as charge densities and pair-correlation functions, there are finite-size errors associated with the fact that Friedel oscillations are forced to be commensurate with the simulation cell. Finite-size effects in excitation energies are discussed in Sec. III D 2.

C. Pseudopotentials for correlated methods

Pseudopotentials or effective core potentials play an important role in almost all *ab initio* electronic structure methods. In essence they replace the influence of core electrons on valence electrons with an effective potential, so reducing each atom to a pseudo-atom composed of valence electrons only. For all-electron DMC calculations the scaling of the cost with atomic number Z is Z^5 – $Z^{6.5}$.^{61,132} The use of pseudopotentials reduces the effective value of Z , making QMC calculations feasible for all atoms.

Replacing the dynamic interaction of valence and core electrons with a potential is necessarily approximate, but the error can be controlled. A pseudopotential with a usable small error is nonlocal, and is defined to reproduce physical properties of the valence electrons. The physical properties usually chosen are either core scattering (to first order) or ionization and excitation energies. Pseudopotentials generated using these two approaches are referred to as “shape consistent” and “energy consistent,” respectively.

The generation of both types of pseudopotentials in the context of DFT or HF theory is well established, with pseudopotential error often less than the error inherent in the underlying theory. However, using such pseudopotentials in many-body methods introduces a significant, unnecessary, and uncontrolled error due to the inconsistent application of theory. For HF pseudopotentials the error is due to the absence of core-valence correlation. A DFT pseudopotential suffers from a deeper inconsistency between the Kohn-Sham orbitals they are designed to reproduce and the many-body wave functions to which they are applied. Note that neither the Kleinman-Bylander nor the ultrasoft pseudopotential forms are useful for many-body methods.

For a pseudopotential to be accurate in QMC it must be constructed to reproduce scattering and/or excitation properties of the many-body Hamiltonian. Pseudopotentials generated using DFT or HF do not have this property and, perhaps most importantly, the accompanying error is difficult to estimate.

No error-correction strategies are available.

Trail and Needs have produced several generations of pseudopotentials, including HF and Dirac-Fock pseudopotentials (TNDF),^{133,134} correlated-electron pseudopotentials (CEPPs),^{135,136} and energy-consistent correlated-electron pseudopotentials (eCEPPs).¹³⁷ eCEPPs provide the best accuracy among these, and are the most sophisticated pseudopotentials currently available due to their accurate description of correlated electron systems. CEPPs and eCEPPs are given in the Supplementary Material of Refs. 135–137 and can also be obtained from Ref. 138. Other groups have also developed pseudopotentials for correlated electronic calculations.¹³⁹

The eCEPP construction process reproduces three different aspects of the core-valence electron interaction. Electron correlation is included throughout, and the independent electron approximation is not used.

First, an all-electron atomic calculation is used to generate a nonlocal potential that reproduces the first-order scattering properties of the core. This is achieved by partitioning the density matrix into a core and a valence part, disposing of the core part, and redefining the valence part close to the atomic nuclei. The nonlocal potential so generated is a correlated-electron generalization of the well-established norm-conserving DFT pseudopotential.

Second, the long-range core-polarization interaction is reproduced by these potentials and is represented using the standard (two-body) core-polarization potential form.¹⁴⁰ At the DFT level no long-range core polarization occurs.

In the third stage of the eCEPP generation process we use the fact that the pseudopotentials defined by the first and second stages are not unique. This allows us to search the space of pseudopotentials that reproduce scattering properties to find those that also reproduce atomic ionization and excitation energies.

All three parts of the eCEPP construction process involve many-body Hamiltonians and wave functions, and employ explicitly correlated atomic calculations, namely multiconfiguration self-consistent field and coupled-cluster calculations with single, double, and perturbative triple excitations [CCSD(T)] with accurately corrected basis-set errors.

The *ab initio* eCEPPs so generated reproduce the scattering of valence electrons by the core (norm conservation), as well as the ionization and excitation energies for a number of states (energy conservation), and long-range polarization interactions. Consequently the eCEPPs are unique in that they are inherently correlated and are both shape- and energy-consistent. Better than chemical accuracy was demonstrated for these pseudopotentials applied to a moderately large set of small molecules with CCSD(T), as shown in Fig. 10. At this level of accuracy some care was required in controlling for basis-set error; the direct use of standard contracted all-electron CCSD(T) Gaussian basis sets is insufficient. Contracted Gaussian basis sets generated to be accurate when used with the eCEPPs are available.¹³⁷

There is room for further research to improve pseudopotential theory for QMC. While the above eCEPPs consistently achieve better than chemical accuracy for all test cases, they are available for first-row and transition-metal atoms only.

There is nothing to prevent the method from being applied to the rest of the periodic table in the future.

A bias arises in DMC due to the nonlocality of pseudopotentials. The many-body electronic wave function Φ is not explicitly available (rather, it is implicit in the probability distribution that arises from the dynamics), so the nonlocal pseudopotential operator \hat{V} cannot be applied to it.

A locality approximation can be used to avoid this difficulty by replacing $\hat{V}\Phi$ with $(\Psi^{-1}\hat{V}\Psi)\Phi$, where Ψ is the trial wave function. This approximation replaces a sum of nonlocal one-body potentials with a local many-body potential, and is exact if $\Psi = \Phi$. The error introduced by this approximation is of second order in the error in Ψ , is system-dependent, is nonvariational, and can be large relative to the fixed-node error.

An alternative partially local ‘‘T-move’’ scheme is also available.^{141,142} This approximation is more robust and obeys a variational principle that ensures that the bias is positive, though not necessarily smaller in magnitude than the locality approximation error. Crucially, the T-move scheme removes the negative divergences in the local energies that occur at the nodes of the trial wave function in the localized pseudopotential operator $\Psi^{-1}\hat{V}\Psi$; these divergences can lead to population explosions in DMC. The T-move scheme leads to stable DMC calculations that satisfy the variational principle across the entire periodic table.¹⁴³ The recently proposed ‘‘determinant locality approximation’’ in which $\hat{V}\Phi$ is simply approximated by $(\Psi_S^{-1}\hat{V}\Psi_S)\Phi$, where Ψ_S is the Slater part of the wave function, has also been shown to deliver enhanced stability and resilience of the results against noise due to optimization of the Jastrow factor.¹⁴⁴ Both of these schemes are available in CASINO.

Plane-wave DFT calculations normally use a Kleinman-Bylander representation of pseudopotentials. However, Kleinman-Bylander representations can introduce ghost states, which are unphysical, low-energy bound states of the pseudopotential. Problems associated with ghost states for the TNDF pseudopotentials can generally be avoided by choosing the *s* angular-momentum channel to be local in the pseudopotentials in the plane-wave DFT calculation. In subsequent QMC calculations, which do not use Kleinman-Bylander representations, the local channel can be chosen to be the highest angular-momentum component present in the nonlocal pseudopotential.¹⁴³ A second challenge in plane-wave DFT calculations is to choose the plane-wave cutoff energy to be sufficiently high. The convergence of the DFT energy to chemical accuracy is a sufficient condition for convergence of the DMC energy. Suggested plane-wave cutoffs for TNDF pseudopotentials can be found in Ref. 143.

D. QMC for excited-state properties

1. Quasiparticle and excitonic gaps

QMC provides a powerful methodology for calculating excitation energies, which are obtained as differences in the total energy. The quasiparticle band at Bloch wavevector \mathbf{k} is

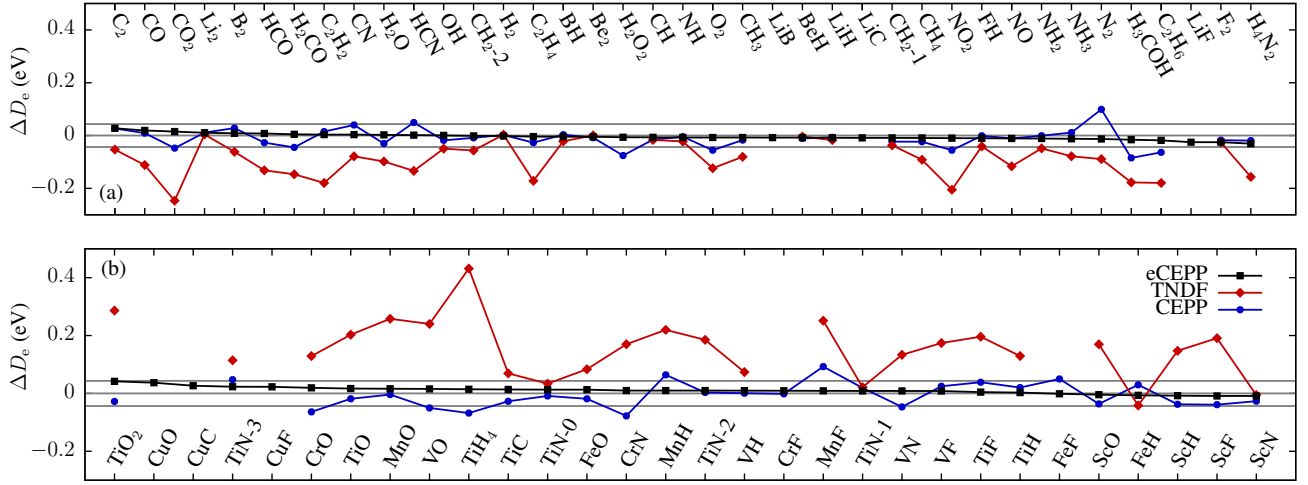


FIG. 10. The difference between molecular dissociation energies D_e from pseudopotential and all-electron CCSD(T) calculations. The top panel (a) shows differences for molecules containing first-row atoms only, and the bottom panel (b) shows differences for molecules containing 3d transition-metal atoms. Data for the eCEPP, CEPP, and TNDP pseudopotentials are from Refs. 135–137. The region between the gray lines at ± 43 meV contain all data points that reproduce all-electron energies to within chemical accuracy.

$\mathcal{E}(\mathbf{k}) = E_{N+1}(\mathbf{k}) - E_N$ for unoccupied states, where E_N is the N -electron ground-state total energy and $E_{N+1}(\mathbf{k})$ is the $(N+1)$ -electron total energy, in which a single electron has been added to an orbital at \mathbf{k} . In a finite cell, the same twisted boundary conditions must be used for the ground state and excited state. For occupied states, the quasiparticle band is defined as $\mathcal{E}(\mathbf{k}) = E_N - E_{N-1}(\mathbf{k})$, where $E_{N-1}(\mathbf{k})$ is the total energy when a single electron is removed from a state at \mathbf{k} . We are implicitly assuming that the wave function of the noninteracting electron system is adiabatically connected to the wave function of the interacting system, so that we can use the wavevectors of the noninteracting orbitals to label the energy levels of the interacting system. QMC thus provides a direct and physical description of quasiparticle excitations: the excited state in question is determined by the occupancy of the single-particle orbitals in the Slater determinant, with the Jastrow factor, the backflow function, and fixed-node DMC providing increasingly good descriptions of dynamical correlation effects while leaving the topology of the nodal surface unchanged relative to the noninteracting case. At finite temperatures, the quasiparticle bands should be defined in terms of differences in the Helmholtz free energy when an electron is added to or removed from a particular state. This could be achieved by adding DFT vibrational free energies to the static-nucleus total energies obtained using QMC calculations.

The Schrödinger equation for the $(N+1)$ -electron excited-state wave function is $\hat{H}\Psi_{\mathbf{k}} = [E_N + \mathcal{E}(\mathbf{k})]\Psi_{\mathbf{k}}$, where $\Psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\sum_{i=1}^{N+1}\mathbf{r}_i}/(N+1)W_{\mathbf{k}}(\mathbf{R})$ is the wave function according to Eq. (47). $W_{\mathbf{k}}$ is invariant under simultaneous translation of all the particles through a primitive lattice vector. It is easy to show that $\hat{H}_{\mathbf{k}}W_{\mathbf{k}} = [E_N + \mathcal{E}(\mathbf{k})]W_{\mathbf{k}}$, where

$$\hat{H}_{\mathbf{k}} = -\frac{1}{2} \left(\nabla^2 + \frac{2i}{N+1} \mathbf{k} \cdot \sum_{i=1}^{N+1} \nabla_i - k^2 \right) + V. \quad (59)$$

Noting that

$$\hat{H}_{\mathbf{k}+\delta\mathbf{k}} = \hat{H}_{\mathbf{k}} + \frac{1}{N+1} \delta\mathbf{k} \cdot \mathbf{k} - \frac{i}{N+1} \mathbf{k} \cdot \sum_{i=1}^{N+1} \nabla_i + \mathcal{O}(k^2), \quad (60)$$

we may use first-order perturbation theory to show that

$$\mathcal{E}(\mathbf{k} + \delta\mathbf{k}) = \mathcal{E}(\mathbf{k}) + \delta\mathbf{k} \cdot \left\langle \frac{1}{N+1} \sum_{i=1}^{N+1} \hat{\mathbf{p}}_i \right\rangle + \mathcal{O}(\delta k^2), \quad (61)$$

where $\hat{\mathbf{p}} = -i\nabla_i$ is the momentum operator for electron i . Hence the mean momentum per electron is given by the gradient of the quasiparticle energy band with respect to \mathbf{k} , confirming that \mathbf{k} continues to be the wavevector of the excitation in the interacting many-electron system. In a crystalline solid, the quasiparticle bands should be extrapolated to the thermodynamic limit $N \rightarrow \infty$.

The quasiparticle band gap Δ_{qp} of a semiconductor or insulator may be evaluated as the difference between the quasiparticle bands at the conduction-band-minimum (CBM) and the valence-band maximum (VBM):

$$\begin{aligned} \Delta_{\text{qp}} &= \mathcal{E}_{\text{CBM}} - \mathcal{E}_{\text{VBM}} \\ &= E_{N+1}(\mathbf{k}_{\text{CBM}}) + E_{N-1}(\mathbf{k}_{\text{VBM}}) - 2E_N. \end{aligned} \quad (62)$$

Again, this should be extrapolated to infinite system size as discussed in Sec. III D 2. The quasiparticle band gap is the amount of energy required to produce a free electron and a free hole. This gap determines the thermal concentrations of charge carriers, and hence can be determined by examining the temperature-dependence of the conductance of a sample. The occupied and unoccupied quasiparticle bands can be measured directly using photoemission and inverse photoemission spectroscopies, or can be probed using X-ray absorption near-edge structure measurements.

A second type of gap plays a crucial role in the interaction of semiconductors with light, namely the excitonic gap. This is defined as the difference between the ground-state and excited-state energies of the N -electron system: $\Delta_{\text{ex}} = E'_N(\mathbf{k}, \mathbf{k}') - E_N$, where $E'_N(\mathbf{k}', \mathbf{k}')$ is the energy of an excited state in which an electron has been promoted from an occupied state at wavevector \mathbf{k} to an unoccupied state at wavevector \mathbf{k}' . Again, assuming adiabatic connection from the non-interacting system, we can select the excited state by the occupancy of the Slater determinant in a QMC wave function; the Jastrow factor, backflow function, and the use of fixed-node DMC provide dynamical correlation without changing the qualitative state of the system. The excitonic gap is the lowest energy at which the electronic system can absorb or emit a photon. This is the gap that is observed in photoluminescence measurements. Physically, the excitonic gap is less than the quasiparticle gap because photoexcitation produces an electron-hole pair, which bind to form an exciton; thus the difference between the quasiparticle and excitonic gaps at zero temperature is given by the exciton binding energy. Since photon momenta are vanishingly small compared with electrons, single-photon absorption or emission only occurs for direct gaps, where $\mathbf{k} = \mathbf{k}'$. Otherwise there must be additional processes involved, such as emission of phonons, in order to satisfy the conservation of momentum, and so we do not have a corresponding sharp peak in absorption or emission. Excitonic gaps can be calculated for cases where $\mathbf{k} - \mathbf{k}'$ is commensurate with the simulation cell. When we take into account nuclear motion, a few further issues arise. Geometry differences between the ground state and excited state lead to differences between the excitonic gaps for optical absorption and emission, respectively; this difference is referred to as a Stokes shift. Photon absorption can take place anywhere in the crystal, and it takes place instantaneously on the time scale of nuclear motion. Therefore, the excitonic gap should be averaged over the thermal and quantum distributions of nuclear coordinates; this gives the energy of the excitonic peak in the photoluminescence spectrum. This is discussed further in Sec. VC 2.

In an isolated atom or molecule, the electron affinity is $\mathcal{E}_{\text{ea}} = E(N) - E(N + 1)$, where $E(N)$ and $E(N + 1)$ are the total ground-state energies of the N -electron and $(N + 1)$ -electron systems, while the ionization potential is $\mathcal{E}_{\text{ip}} = E(N - 1) - E(N)$, where $E(N - 1)$ is the total ground-state of the $(N - 1)$ -electron system. These total energies should in principle include static correlation effects. The great challenge here is to use multideterminant wave functions of equivalent accuracy for the neutral atom or molecule and the positive and negative ions.

2. Finite-size effects in gaps

QMC studies of condensed matter usually involve performing calculations for finite simulation cells subject to periodic boundary conditions. Nevertheless, the excitation energies obtained in a finite simulation cell differ from those in the thermodynamic limit of infinite system size. Fortunately, the phe-

nomenological quasiparticle picture described in Sec. IVA 1 can be used to understand the dominant finite-size effects in *ab initio* QMC calculations.⁷³

Consider the quasiparticle bands of an insulator or semiconductor. As described in Sec. IIID 1, an unoccupied band is defined via the difference in the total energy relative to the ground state when an electron is added to the system. Within the band effective mass approximation, the additional electron moves like a negatively charged quasielectron whose mass is determined by the curvature of the conduction band. Similarly, excitations from the valence bands behave as positively charged quasiholes whose mass is determined by the curvature of the valence band. The Coulomb interactions between quasiparticles are screened by the response of the other electrons in the crystal.

When one adds an electron to a periodic simulation supercell to calculate the CBM, the effect is to create a lattice of quasielectrons repeated throughout space. The leading-order finite-size error is therefore the unwanted Madelung energy $v_M/2$ of this lattice of quasielectrons, which must be subtracted from the energy E_{N+1} of the $(N + 1)$ -electron system. Similarly the Madelung energy of the unwanted lattice of quasiholes must be subtracted from the energy E_{N-1} of the $(N - 1)$ -electron system. So the leading-order finite-size correction to the quasiparticle gap $\Delta_{\text{qp}} = E_{N+1} + E_{N-1} - 2E_N$ in a finite cell is that the supercell Madelung constant v_M must be subtracted from the gap.⁷³ Note that the supercell Madelung constant must be evaluated using the screened Coulomb interaction. In the case of a crystal of cubic symmetry the screened Madelung constant is simply $v_M = v'_M/\epsilon$, where v'_M is the unscreened Madelung constant and ϵ is the static-nucleus (high-frequency) permittivity. In the general 3D case, the screened Madelung constant is found by a coordinate transformation to the eigenbasis of the permittivity tensor ϵ : $v_M(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) = v'_M(\epsilon^{-1/2}\mathbf{a}_1, \epsilon^{-1/2}\mathbf{a}_2, \epsilon^{-1/2}\mathbf{a}_3)/\sqrt{\det(\epsilon)}$, where \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are the supercell lattice vectors. For a given shape of simulation cell, the finite-size error in the uncorrected gap falls off slowly as $N^{-1/3}$. Moreover, since v_M is often negative, the uncorrected quasiparticle gap may be absurdly small, even negative. In the case of a layered or 2D material, the Madelung constant should be evaluated using the Rytova-Keldysh interaction.^{145,146} In another example of the symbiotic relationship between QMC and DFT, the permittivity tensor can usually be evaluated to sufficient accuracy using density functional perturbation theory calculations. Alternatively the permittivity tensor can be evaluated using the small- \mathbf{k} limit of the static structure factor, which can be evaluated using ground-state QMC calculations.¹⁴⁷

Beyond-leading-order finite-size effects in the quasiparticle bands and therefore quasiparticle gap arise due to the fact that the charge-density distributions in the quasiparticles have quadrupole moments in general; the resulting charge-quadrupole interactions give an $\mathcal{O}(N^{-1})$ finite-size error in 3D materials. The addition of charges to a finite simulation cell also causes long-range oscillatory behavior in the electronic pair density. These oscillations are forced to be commensurate with the finite periodic simulation cell, resulting in

quasirandom fluctuations in the gap with system size. Hence a suggested procedure for calculating QMC quasiparticle gaps is to perform calculations at a range of system sizes, correct the leading-order errors by subtracting the screened Madelung constant, then extrapolate the resulting gaps to infinite system size assuming a $1/N$ error; this simultaneously removes residual systematic effects and averages out quasirandom finite-size errors.⁷³

To calculate an *ab initio* exciton binding energy, the supercell must be significantly larger than the exciton Bohr radius. In smaller cells, the quasielectron and quasihole are effectively unbound, and hence the excitonic gap is similar to the quasiparticle gap. If the simulation supercell is large enough to contain well-formed excitons then interactions between periodic images are relatively unimportant, since excitons are neutral. However, the interaction between the quasielectron and quasihole is different in a finite cell and an infinite system. From the Taylor expansion of the 3D Ewald interaction v_E for a periodic cell, the leading-order difference from the Coulomb $1/r$ interaction is $\mathcal{O}(N^{-1})$.^{122,123} This leads to an $\mathcal{O}(N^{-1})$ finite-size error in the exciton binding energy; hence, provided the simulation cell is large enough, the excitonic gap can be extrapolated to the thermodynamic limit. The conclusion is modified in 2D or layered systems, where the interaction is of Rytova-Keldysh form, leading to $\mathcal{O}(N^{-2/3})$ scaling of the finite-size error that eventually crosses over to $\mathcal{O}(N^{-1})$ behavior on near-macroscopic length scales.⁷³

3. Intraband excitations

QMC methods can be used to study intraband excitations in metallic systems. An example of this is the calculation of the renormalization of the electron mass by electron-electron interactions in the 2D HEG, which is discussed in Sec. IV A 5.

E. Computational efficiency of QMC

1. Parallelization of QMC algorithms

QMC methods are intrinsically well-suited to massively parallel architectures. CASINO makes use of message-passing interface (MPI) for an “outer” distributed-memory level of parallelism and OpenMP for an “inner” shared-memory level of parallelism. In practice the MPI parallelism is more important and much more widely used. However, where possible, data such as wave-function coefficients are shared between MPI processes within each processor node.

Assuming equilibration to be a negligible fraction of the total central-processing-unit (CPU) time, the VMC method is “embarrassingly parallel.” Multiple independent VMC calculations may be run in parallel, each with a different random seed, to generate an amount of data that is proportional to the total computational effort. The only required communication between MPI processes is the broadcast of the initial distribution of geometry and wave-function data, the sending of the

initial random seeds, and the summation of the final set of expectation values to average. No interprocess communication is required during the VMC calculation. In wave-function optimization, VMC-sampled configurations are divided between MPI processes, which independently evaluate local energies and other required quantities. A “master” process gathers reduced data such as the variance of the local energies in order to choose a new parameter set, which must then be broadcast.

The natural parallelization strategy for DMC calculations is to divide the walker population between MPI processes, so that each process has its own population of walkers. Unlike VMC, communication is required during the simulation, as a master process has to decide on a reference energy in order to control the overall population. Furthermore, because the walkers randomly branch and die, it is necessary constantly to balance the load on different MPI processes by transferring walkers from processes with larger populations to processes with smaller populations. The cost of a DMC time step is determined by the MPI process with the largest population of walkers. The amount of branching varies from one system to another and is reduced when the wave function is more accurate, thereby improving the parallel performance. Since fractional fluctuations in the walker population fall off as the inverse square root of the population, the fractional cost of transferring walkers can be made arbitrarily small (and hence the parallel efficiency of the statistics-accumulation phase can be made arbitrarily large) by choosing a large number of walkers per process. However, increasing the population leads to additional expense due to the need to equilibrate the large population, as discussed in Sec. III E 2.

The secondary OpenMP parallelism is used to accelerate loops over the electrons when evaluating interparticle interactions, Jastrow terms, and orbitals.

Figure 11 shows the strong-scaling parallel speedup when carrying out a block of 50 time steps in a CASINO DMC calculation against the number of processors. The calculations were performed on Oak Ridge Leadership Computing Facility’s Jaguar supercomputer. The system studied was a 54-atom supercell of silicon and the number of electrons in the simulation was $N = 216$. The SJ wave function, time step, etc., were all realistic. The only unrealistic element of the calculation was that the calculation was very short, with CASINO writing out checkpoint data at the end of the 50 iterations. The time taken to assemble the checkpoint data on the master processor and write it to disk is the bottleneck for large numbers of cores, as evidenced by the fact that if checkpointing is disabled the time taken falls significantly, and the apparent scaling with processor number is greatly improved. The time spent checkpointing with 76,800 cores is about 100 seconds. In a real calculation, we typically checkpoint once every 20–30 minutes, so this bottleneck is an artifact of the short test runs used in the scaling test. It is clear that DMC calculations with CASINO scale very well up to and beyond 10,000 cores. In fact CASINO exhibited virtually perfect linear scaling on up to 124,416 cores on Jaguar, 131,072 cores on an IBM BGQ, and over half a million cores on the K computer, due to modifications described in Ref. 148.

Figure 12 shows a similar parallel scaling analysis, this

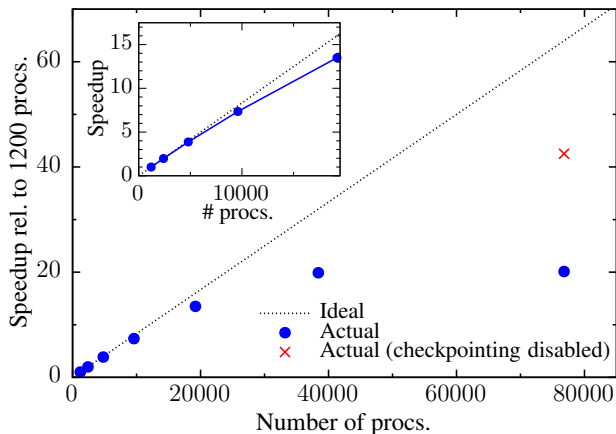


FIG. 11. Strong-scaling parallel speedup for DMC calculations with CASINO for a 54-atom supercell of silicon ($N = 216$ electrons), performed on Jaguar. The inset shows the scaling with small processor numbers in greater detail.

time for a 144-atom supercell of bilayer graphene, carried out on the supercomputer ARCHER. Again it is clear that the speedup for a fixed number of DMC time steps and target population is extremely good, up to at least 10,000 cores.

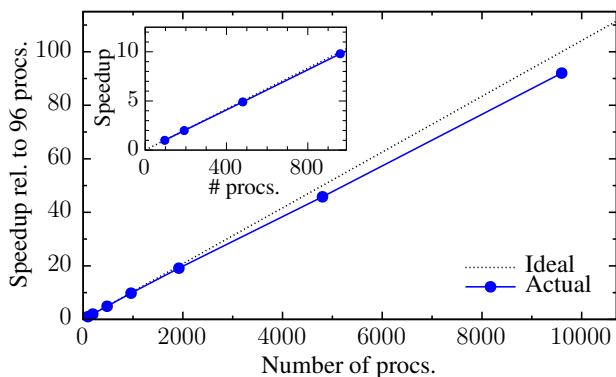


FIG. 12. Strong-scaling parallel speedup for DMC calculations with CASINO for a 144-atom supercell of bilayer graphene ($N = 576$ electrons), performed on ARCHER. The inset shows the scaling with small processor numbers in greater detail.

2. Scaling of QMC with system size

A basic analysis of system-size scaling must consider the cost of moving each walker, and the number of such moves that are required to achieve a result of a given precision (target error bar). We shall consider the cost when a Slater-Jastrow wave function with a fixed number of determinants is used.

Consider moving an N -electron walker; this involves proposing a move of each electron in turn, then evaluating the local energy at the resulting configuration at the end of the move. (In the case of VMC we only need to evaluate the

energy once every few moves; in the case of DMC we must evaluate it after every walker move, as it is needed to evaluate the branching factor.) There are several $\mathcal{O}(N^2)$ contributions to the time taken: electron-electron interactions must be evaluated; the Jastrow factor nearly always contains long-range two-body terms of system-size-independent complexity between each pair of electrons; and each orbital must be evaluated at each electron position. We have implicitly assumed the use of a localized basis set for the orbitals; if an extended basis set is used then there is an extra factor of N in the scaling of orbital evaluation, since the number of basis functions for each orbital at each position increases as N in this case. For this reason, when orbitals are generated using plane-wave DFT calculations, they are re-represented on a B-spline (blip) grid in real space before use in QMC calculations,^{149,150} instead of evaluating $\mathcal{O}(N)$ plane waves at each electron position, we evaluate twelve B-spline functions. Likewise, Gaussian basis functions are rigorously truncated to zero beyond a certain radius, so they can be treated as a localized basis. In studies of insulators, the cost of evaluating the orbitals can be reduced to $\mathcal{O}(N)$ by performing a linear transformation to a set of (nonorthogonal) localized orbitals that can be truncated to zero outside predefined localization regions.^{149,151} To minimize the bias arising from the truncation, we maximize the overlap of the localized orbitals with the set of localization regions. To avoid bias, the localization regions for valence bands need to be significantly larger than the length scales estimated in Sec. II B 4. For core states, the localization regions can have radii of order 1 bohr.

Evaluating a Slater determinant given all the orbital values is an $\mathcal{O}(N^3)$ operation. In an electron-by-electron algorithm Slater determinants must be updated after each accepted single-electron move. Unless backflow is used, a single-electron move only affects a single column of the Slater matrix. Such updates can be performed in a time that scales as $\mathcal{O}(N^2)$ by storing the matrix of cofactors of the Slater matrix and updating this using the Sherman–Morrison formula, resulting in an $\mathcal{O}(N^3)$ cost of updating the Slater determinant over a walker move.¹⁵² In practice, however, we find that the $\mathcal{O}(N^3)$ operations have negligible cost and, over the practical range of system sizes, the cost of each walker move scales as $\mathcal{O}(N^2)$. The costs of evaluating the Jastrow factor and the orbitals are usually comparable, with the cost of evaluating electron-electron interactions being significantly less.

For a given wave-function form the variance of the total energy scales as $\mathcal{O}(N)$, while the variance in the energy per particle falls off as $\mathcal{O}(N^{-1})$. The standard error in the mean energy is given by the square root of the variance divided by the square root of the number of independent samples (see Sec. II A 1). For insulators at least, the number of independent samples is proportional to the number of actual samples for a given time step. Hence the number of configurations sampled must increase as $\mathcal{O}(N)$ to achieve a given error bar on the total energy, while the number of samples required to achieve a given error bar on the energy per particle falls off as $\mathcal{O}(N^{-1})$. Similar comments apply to the scaling of the QMC estimate of any extensive or intensive quantity, respectively. In the commonly encountered situation in which the cost of

equilibration is negligible and the cost of each walker move is $\mathcal{O}(N^2)$, this leads to the well-known $\mathcal{O}(N^3)$ scaling of the QMC method.¹⁵

We now consider the scaling of the DMC method in more detail, taking into account the need to equilibrate the walker population. Suppose we are interested in using DMC to study a real, d -dimensional metallic system using pseudopotentials to represent the atoms. The linear size of an N -electron simulation cell is $L = N^{1/d}r_s$, where r_s is the density parameter for the valence electrons. The valence electron density parameter in a real system is typically $r_s = 2$ –5. The RMS distance diffused by each electron in equilibration needs to be greater than L in order to ensure that equilibration errors are exponentially small. So the number of equilibration steps N_{eq} must satisfy

$$\sqrt{N_{\text{eq}}\tau d} \gtrsim N^{1/d}r_s. \quad (63)$$

Note that there is little point in using a huge simulation cell if one does not equilibrate sufficiently long to describe the long-range behavior correctly. If a DMC simulation does not extend over a few decorrelation periods then we will not be able to observe a plateau in reblocking analysis, and it will be difficult to obtain accurate error bars.

Let W be the total walker population and P be the number of MPI processes. To minimize equilibration time, the number of walkers per MPI process W/P will be a fixed small number. It is preferable not to have fewer than two walkers per MPI process, otherwise parallel efficiency is reduced. If the number of walkers per MPI process is fixed, the wall-clock time taken to perform N_{eq} time steps is essentially independent of the number of processes.

For a Slater-Jastrow wave function with B-spline (blip) orbitals, the time in seconds taken to perform N_{eq} DMC iterations is

$$T_{\text{eq}} = c_0 N^2 N_{\text{eq}} W / (P\Theta), \quad (64)$$

where Θ is the number of OpenMP threads and the constant c_0 depends on the system, the wave-function quality, and the computer hardware. In practice OpenMP parallelism is useful provided $N/\Theta \gtrsim 100$. Empirically, $c_0 \sim 10^{-5}$ s for a range of first-row atoms with Slater-Jastrow wave functions and various Intel processors from around the year 2010.

Combining Eqs. (63) and (64), we have

$$N \lesssim \left(\frac{\tau T_{\text{eq}} P \Theta d}{c_0 r_s^2 W} \right)^{d/8}. \quad (65)$$

If we have $W/P = 2$ walkers per MPI process, a single thread per process, $r_s = 3$ (a typical valence electron density), the maximum equilibration time $T_{\text{eq}} = 4$ days $\approx 3 \times 10^5$ s, and the usual time step $\tau = 0.01$ Ha⁻¹ for pseudopotential calculations, we obtain

$$N \lesssim 800. \quad (66)$$

So performing DMC for more than $N = 800$ electrons with a time step of 0.01 Ha⁻¹ and a single thread per MPI process will be difficult, because one will have to equilibrate for

many days, no matter how big a computer one uses. There is a bit of leeway here if one is willing to use larger time steps (e.g., assuming cancellation of biases), or is willing to accept incomplete equilibration (because the resulting bias is somehow known to be negligible compared with a target error bar), or is willing to equilibrate for many days. We have also assumed that we are looking at a metallic system, which is a worst-case scenario; in a system in which electrons are highly localized, the equilibration time could be much shorter. Alternatively, to evade this limit, one may use multiple threads per MPI process. With $\Theta = 16$ threads per process, the upper limit on the number of electrons grows to about 2000.

For the largest system sizes that one might like to consider using $\Theta = 1$ thread ($N \approx 800$ electrons), the equilibration period lasts $T_{\text{eq}} = 4$ days. Let us assume the processor number is sufficiently large that we are in the regime where we have a fixed number of walkers per MPI process. The scaling with processor number is good if and only if equilibration is a small fraction of the total run time. The point at which the scaling with processor number becomes poor is when the time spent on statistics accumulation is about the same as the time spent on equilibration. After that, increasing the number of MPI processes makes increasingly little difference to the total run time, because the equilibration time remains stuck at T_{eq} . So, for the largest number of MPI processes it is worth using, the number of statistics accumulation steps N_s is such that the time taken for statistics accumulation is equal to the time taken for equilibration:

$$c_0 N^2 N_s W / (P\Theta) = T_{\text{eq}}. \quad (67)$$

For insulators at least¹⁵³ the error bar on the DMC total energy per electron in pseudopotential calculations goes as

$$\Delta = \frac{k_0}{\sqrt{W N N_s \tau}}, \quad (68)$$

where k_0 depends on the system and wave-function quality. Empirically, $k_0 \sim 0.1$ Ha^{1/2} for a range of first-row atoms with Slater-Jastrow wave functions.

Suppose the target accuracy is Δ_0 . By using Eq. (68) to eliminate the number of statistics-accumulation steps from Eq. (67), the maximum useful number of MPI processes is

$$P_{\text{max}} = \frac{k_0^2 c_0 N}{\Delta_0^2 T_{\text{eq}} \tau \Theta}. \quad (69)$$

Putting in $\Theta = 1$ thread, $N = 800$ (the largest size for which equilibration is feasible), $\tau = 0.01$ Ha⁻¹, $T_{\text{eq}} = 4$ days and a very stringent target accuracy of $\Delta_0 = 2 \times 10^{-6}$ Ha per electron (i.e., $\Delta_0 = 8 \times 10^{-6}$ Ha per atom when there are 4 electrons per atom) gives the maximum useful number of MPI processes as $P_{\text{max}} = 6000$. When running with the maximum useful number of MPI processes, the total run time is $2T_{\text{eq}}$ (i.e., eight days).

In practice there are usually trivial sources of parallelism in twist averaging, running at different time steps, running for different systems and system sizes, etc.

Now suppose we are interested in total energies, e.g., to calculate a band structure or a defect formation energy. In

that case the error in the total energy is

$$\Delta_{\text{tot}} = k_0 \sqrt{\frac{N}{WN_s\tau}}. \quad (70)$$

Suppose the target accuracy is $\Delta_{0\text{tot}}$. By Eqs. (67) and (70), the maximum useful number of MPI processes is

$$P_{\text{max}} = \frac{k_0^2 c_0 N^3}{\Delta_{0\text{tot}}^2 T_{\text{eq}} \tau \Theta}. \quad (71)$$

A target accuracy for calculating a band structure might be 0.001 Ha. Putting in $\Theta = 1$, $N = 800$, and $\tau = 0.01$, we find that $P_{\text{max}} \approx 10^4$.

The same sort of analysis can be applied to VMC, although in VMC the time step is much larger, so that equilibration remains negligible for much larger system sizes.

If we return to a situation in which the equilibration period can be assumed to be negligible, e.g., because extremely small error bars are required, then a fundamental difference between VMC and DMC emerges. VMC with a Slater-Jastrow wave function scales as $\mathcal{O}(N^3 + \epsilon N^4)$, while VMC with a Slater-Jastrow-backflow wave function scales as $\mathcal{O}(N^4)$. DMC shows the same scaling at small N , but asymptotically crosses over to exponential scaling.⁷⁰ The issue is related to the system-size dependence of population-control bias, discussed in Sec. II B 3. When a walker branches, the resulting daughter walkers are highly correlated, reducing the effective walker population. The daughter walkers decorrelate at a rate given by the inverse of the decorrelation period τ_{corr} . It is clear from the form of the branching factors that the rate at which branching takes place is proportional to the spread of local energies that appear in the branching factors. The spread of local energies may be quantified by the standard deviation of the local energy, which scales as the square root of system size. The spread of local energies is smaller for better wave function forms. Nevertheless, for any given wave-function form the branching rate eventually overtakes the decorrelation rate as the system size is increased. This occurs when the standard deviation of the local energies exceeds the fixed-phase energy gap. In this case the effective population size decreases to a single walker. This provides an exponential hard wall on the system-size scaling of DMC, which for Slater-Jastrow wave functions is typically reached with of order 1000–2000 electrons.⁷⁰

In summary, the $\mathcal{O}(N^3)$ scaling of DMC total-energy calculations breaks down at system sizes of around $N = 1000$ –2000 electrons, and the excellent scaling of QMC with processor number does not overcome this problem. When studying condensed matter, we must use careful analysis of finite-size effects to obtain accurate results from system sizes of fewer than 2000 electrons.

IV. SOME RECENT APPLICATIONS OF CASINO

A. Model systems

1. Excitonic complexes

QMC methods are powerful techniques for solving the many-electron Schrödinger equation in *ab initio* chemistry and condensed-matter physics. However, the same methods can be used to solve the Schrödinger equation in models that capture the key physics of a material or device. Indeed it is often much easier to make direct connection with experiments in semiconductor physics by means of such model calculations.

An idea of fundamental importance in solid-state physics is the band effective mass approximation. Interactions of electrons with external fields that are slowly varying on the scale of the primitive unit cell are described by quasiparticle excitations, whose wave functions provide envelopes for the underlying Bloch orbitals. These envelope wave functions are solutions of the Schrödinger equation for quasiparticles whose effective mass is determined by the curvature of the underlying electronic band structure. In semiconductors, the most important quasiparticles are charge carriers, i.e., the small number of electrons found near the CBM and the small number of holes found near the VBM. Under the assumption of locally quadratic bands, the effective masses of the electrons and holes are the reciprocals of the second derivatives of the conduction and valence bands with respect to wavevector at their minimum and minimum, respectively. Most of the complexity of the many-electron problem is buried in the effective masses; the remaining challenge is to solve the Schrödinger equation for the quasiparticles interacting with external fields, with defects in the lattice, and with each other.

Quasiparticles interact via Coulomb interactions that are screened by the response of the electrons and ions in the crystal. In a bulk semiconductor of cubic symmetry, the screening is described by a scalar permittivity ϵ , and the interaction between charge carriers is of the usual $1/r$ form scaled down by ϵ . If the exciton energy is small compared with the optical phonon energies then it is reasonable to assume that the ions have time to relax as the charge carriers move, so that the permittivity ϵ should be the static permittivity. By contrast, in materials with large exciton binding energies compared to the optical-phonon energies, the ions do not have time to respond to the electronic motion, and hence the high-frequency (static-nucleus) permittivity should be used. In a bulk semiconductor of lower symmetry, ϵ is a tensor and hence the interaction $1/|\epsilon r|$ between charges is anisotropic. In a layered or 2D material, the form of the interaction is altered more radically due to the fact that the material is only polarizable in-plane. This leads to the so-called Rytova-Keldysh form of interaction,^{145,146} which is of Coulomb $1/r$ form at long range but is logarithmic at short range.

At very short range the effective-mass approximation breaks down, and local exchange and correlation effects lead to effective contact interaction potentials between charge carriers. For weakly bound Mott-Wannier excitons and excitonic complexes these contact interactions are often negligible, and

they largely cancel out of the binding energies of larger complexes.

Charge-carrier complexes are bound states of electrons, holes, and charged defects in a semiconductor. The simplest bound complexes are excitons (bound states of a single electron and a single hole), donor atoms (bound states of an electron and a positively charged defect), and acceptor atoms (bound states of a hole and a negatively charged defect). Three-particle complexes include positive and negative trions (two holes and an electron, and two electrons and a hole, respectively). Four-particle complexes include biexcitons (two electrons and two holes). Valley degeneracy may lead to multiple distinguishable species of electron or multiple species of hole, and hence the possibility of even larger complexes. In many cases the fermionic statistics obeyed by the charge carriers are irrelevant, because there is only one member of each distinguishable species present in each stable complex, and the long-range Hamiltonian does not depend on spin or valley; hence the DMC method is exact. Where this has been investigated in 2D materials, it has been found that biexcitons with a pair of indistinguishable charge carriers are unstable except when the masses of the indistinguishable charge carriers are much greater than the masses of the other two charge carriers.¹⁵⁴

Charge-carrier complexes play a crucial role in the interaction between semiconductors and light at low temperature, as recombination takes place from electrons and holes bound in complexes rather than from a gas of free charge carriers. At low temperature, peaks in photoluminescence spectra due to each stable species of charge carrier complex may be visible, dependent on selection rules. The energy of a carrier complex provides the energy of a peak in the photoluminescence spectrum relative to the quasiparticle band gap. The energy to remove an exciton from a complex provides the position of the peak due to the complex relative to the exciton peak. The binding energy relative to the most energetically favorable daughter complexes also provides an estimate of the temperature at which the peak due to a particular complex will disappear. Depending on the doping of the semiconductor, free electrons or holes may be present at the time of photoexcitation, allowing the formation of charged complexes such as trions. On the other hand, where charge carriers are produced purely by laser photoexcitation of an undoped semiconductor, the charge-carrier complexes are overwhelmingly likely to be neutral (excitons or biexcitons).

QMC methods have been used to evaluate the binding energies of isotropic 3D excitons and ideal 2D biexcitons interacting via the Coulomb interaction,¹⁵⁵ biexcitons and trions in ideal 2D bilayers modeling coupled quantum-well heterostructures of III-V semiconductors,^{156–158} 3D biexcitons and trions in type-II superlattices of III-V semiconductors,^{159–161} and 3D biexcitons and trions in quantum wells,¹⁶² quantum dots,¹⁶³ and type-II quantum rings¹⁶⁴. The issues studied have included not just the prediction of peaks in photoluminescence spectra at low temperature, but also the inhibition of Bose-Einstein condensation of excitons in 2D coupled quantum wells. Charge-carrier complexes play a particularly important role in the optoelec-

tronic properties of layered and 2D materials. In the field of 2D materials, the DMC-calculated binding energies of trions^{165–167} and the binding energies of larger complexes such as biexcitons and quintons (charged biexcitons)^{154,168} have been reported, as have the DMC-calculated binding energies and VMC-calculated recombination rates of ion-bound charge carrier complexes in heterobilayers of transition-metal dichalcogenides.^{169,170}

2. Ground-state energy of the HEG

Ruggeri *et al.*¹⁷¹ obtained the correlation energy of the ferromagnetic HEG at high densities to unprecedented accuracy thanks to two key elements.

First, advances in the understanding of finite-size effects enabled an extremely reliable extrapolation of the fixed-node energy to the thermodynamic limit. While it is customary to correct the total energy $E(N)$ at system size N for quasirandom fluctuations by subtracting $\Delta K(N) = K(N) - K(\infty)$ from it, where $K(N)$ is the HF kinetic energy at system size N , the analogous correction using the HF exchange energy $X(N)$ is avoided because it introduces a slowly varying dependence on N into the corrected energy which complicates the extrapolation process. However, Drummond *et al.*¹²⁸ determined the prefactor of this slowly-varying term, and the correction

$$\Delta X(N) = X(N) - X(\infty) + \frac{3C_{\text{HF}}}{8\pi} r_s^{-1} N^{-2/3}, \quad (72)$$

where $C_{\text{HF}} = 2.837297479$ for simple-cubic simulation cells, essentially suppresses quasirandom fluctuations in the fixed-node DMC energy per electron while keeping its leading-order behavior at $\mathcal{O}(N^{-1})$. The effect of these corrections is illustrated in Fig. 13 for the HEG at density parameter $r_s = 0.5$.

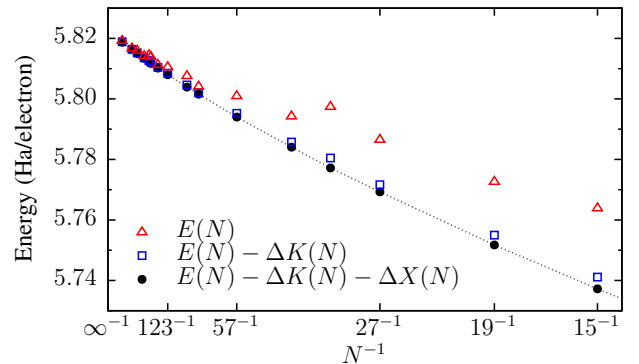


FIG. 13. Twist-averaged SJ-DMC energy per electron of the $r_s = 0.5$ HEG as a function of the reciprocal of system size with and without the $\Delta K(N)$ and $\Delta X(N)$ corrections discussed in the text.

The prefactor of the N^{-1} term is also known¹²⁶ provided backflow is not used;¹²⁹ hence the use of Slater-Jastrow wave

functions enables us to fit the finite-size-corrected DMC energies per electron to a constant plus a power expansion in $N^{-1/3}$ of leading order $N^{-4/3}$. This procedure yields extremely reliable thermodynamic limits for the fixed-node energy.

The second key element in the work of Ruggeri *et al.* is the use of FCIQMC to obtain essentially exact energies for the HEG at system sizes $N = 15\text{--}33$. This allows the evaluation of the fixed-node error at these system sizes, which can be fitted to a low-order power expansion containing N^0 , N^{-1} , and $N^{-4/3}$ terms,^{129,171} under the assumption that the fixed-node error is a smoother function of $N^{-1/3}$ than the total energy, and extrapolated to $N \rightarrow \infty$.

The energies of the high-density ferromagnetic HEG obtained by adding the extrapolated fixed-node energy and the extrapolated fixed-node error, plotted in Fig. 14, are estimated to be accurate to 1 meV/electron. This is much smaller than

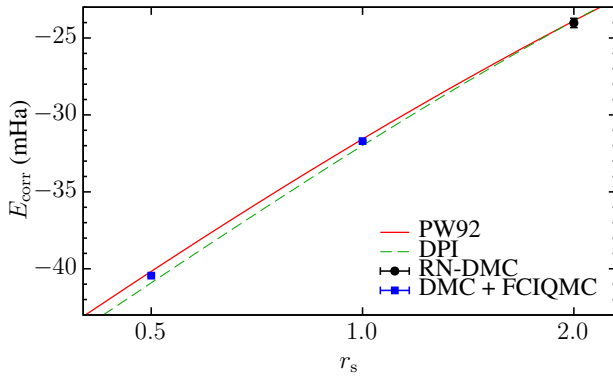


FIG. 14. Correlation energy of the ferromagnetic electron gas as a function of density parameter r_s (note the logarithmic scale of the r_s axis). Results obtained by combining DMC and FCIQMC data,¹⁷¹ as described in the text, are shown, along with the release-node DMC result of Ceperley and Alder¹⁴ at $r_s = 2$, the parameterization of Perdew and Wang (PW92)¹⁷² based on the data of Ceperley and Alder, and the “density parameter interpolation” (DPI) parameterization of Bhattarai *et al.*¹⁷³ which uses minimal input from QMC calculations. The correlation energies from DMC and FCIQMC resolve the discrepancy between the different parameterizations at high densities.

the uncertainty in the release-node DMC energy obtained by Ceperley and Alder¹⁴ at the highest density they considered ($r_s = 2$ for the ferromagnetic HEG). These results offer the possibility of refining existing parameterizations of the correlation energy of the HEG^{172,173} used in the construction of exchange-correlation functionals for DFT. Furthermore, knowledge of the magnitude of the fixed-node error in the energy of HEG as a function of density is potentially very useful; we discuss this further in Sec. V A 3.

3. Phase diagram of electron-hole bilayers

Model systems with attractive interactions are also tractable with QMC. Traditionally, different wave functions have been

used to study the two-component fluid and excitonic phases of the electron-hole system, using the energy associated with each wave function to determine which phase is stable at each set of system parameters.¹⁷⁴ Using the wave function of Eqs. (42) and (43) enables a much more consistent description of the system across phase boundaries, although this requires the evaluation of expectation values other than the energy to determine the phase of the system. Maezono *et al.*¹⁰⁶ used this approach to compute the phase diagram of the equal-mass, equal-density electron-hole double layer as a function of density and interlayer distance, using the condensate fraction to discriminate the two-component fluid from the Bose-Einstein condensate (BEC) phase in which electrons and holes form localized excitons. Maezono *et al.* found that the trial wave function was also capable of describing a biexcitonic phase, with exciton-exciton binding being captured by the DTN Jastrow factor. The biexcitonic phase was found at small interlayer separations and low carrier densities; the pair-correlation function was used to distinguish this phase from the two-component fluid.

The identification of the system parameters at which the excitonic phase of electron-hole double layers is theoretically stable is important in the experimental search for superfluidity, but so is the characterization of the system in this regime. López Ríos *et al.*¹⁰⁸ obtained the superfluid parameters of the symmetric electron-hole double layer as a function of density, fixing the interlayer distance to the smallest value such that biexciton formation is precluded.¹⁵⁷ This was done by subtracting the energy of the neutral system plus the chemical potential from the energy of the system with an additional electron of associated wavevector \mathbf{k} , and fitting the results to the Bardeen-Cooper-Schrieffer (BCS) dispersion relation,

$$\varepsilon(k) = \sqrt{(k^2/2m^* - \mu)^2 + \Delta^2}, \quad (73)$$

where m^* , μ , and Δ are the effective mass, chemical potential, and superfluid gap of the electron quasiparticle. Knowledge of these parameters revealed the density range in which the superfluid can be expected to be most stable. The phase diagram of the electron-hole double layer is shown in Fig. 15.

Superfluidity has been demonstrated experimentally in double bilayer graphene encapsulated in a few layers of WSe₂.¹⁷⁵ This experimental system exhibits multivalley effects, and describing the strong dielectric response of the effectively 2D encapsulating material requires the use of the Rytova-Keldysh interaction instead of the bare Coulomb interaction. Given that neither of these aspects was considered in Ref. 108, the density range over which the superfluid is predicted to be stable is in reasonable (order-of-magnitude) agreement with experiment.

The behavior of the superfluid parameters of the symmetric electron-hole double layer exhibit similarities to that of non-Coulomb systems across the BCS, BCS-BEC crossover, and BEC regimes,^{176,177} suggesting the existence of universal physical behavior which does not depend on the details of the microscopic interaction.¹⁰⁸ The peculiarity of the electron-hole double layer in this regard is the near suppression of the regime of stability of the BCS phase.¹⁰⁸

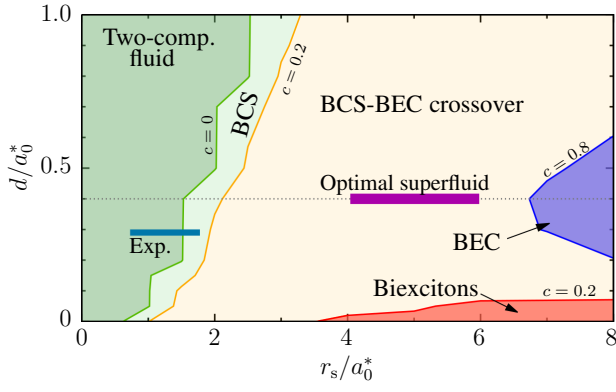


FIG. 15. Phase diagram of the electron-hole double layer according to the value of the condensate fraction c .¹⁰⁶ The length unit is the exciton Bohr radius $a_0^* = 1/\mu$, where $\mu = m_h/(1 + m_h)$ is the reduced mass of the electron-hole pair and m_h is the hole mass. The dotted line marks the interlayer separation below which biexciton formation is possible,¹⁵⁷ and the region labeled “Optimal superfluid” corresponds to the maximum superfluid gap.¹⁰⁸ The discrepancy with the density range in which superfluidity has been experimentally observed¹⁷⁵ (labeled “Exp.”) is likely due to the omission of 2D polarization and multivalley effects in the QMC calculations.

4. Photoexcitation in doped semiconductors

Consider the problem of photoexciting a valence electron in an n-doped semiconductor with a finite concentration of conduction electrons. The situation is illustrated in Fig. 16. The minimum photon energy for photoexcitation is the quasiparticle gap plus the Fermi energy of the electron gas relative to the CBM, plus the energy of the isolated hole at the Fermi wavevector, plus the electron-hole correlation energy.

The crossover between high-density Fermi-edge-singularity behavior dominated by many-body correlations and low-density behavior characterized by the presence of isolated excitons and trions has been investigated experimentally in high-mobility 2D HEGs in GaAs/AlGaAs and InGaAs/InAlAs heterostructures.^{178–181} Rapid changes in line shapes and transition energies in absorption and photoluminescence spectra of a gated modulation-doped quantum well allow experimentalists to locate a critical “crossover” density.

An idealized model of the situation is a quantum impurity problem: a single hole immersed in a zero-temperature 2D HEG.¹⁰⁹ To solve this model Spink *et al.* used VMC and DMC as implemented in CASINO to simulate $N_e = 86$ electrons (a closed-shell configuration) plus a single hole in a periodic hexagonal cell. The specialized pairing trial wave function described in Sec. III A 3 was developed for this problem. A cell area of $(N_e - 1)\pi r_s^2$ was used, where r_s is the HEG density parameter, so that the electron density far from the hole is correct. The electron-hole correlation energy was calculated by subtracting the energy of a pure HEG of the same number of electrons in the same cell. The results are shown in Fig. 17. The low-density limit of the electron-hole correlation energy is the energy of an isolated trion. The gradual crossover between collective exciton and isolated trion behavior occurs

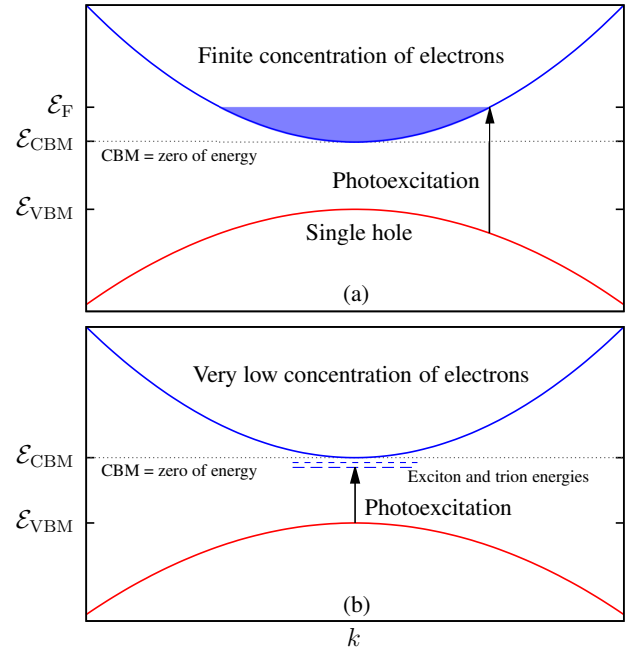


FIG. 16. Photoexcitation of an n-doped direct semiconductor at (a) high electron concentration and (b) low electron concentration.

in a parameter range consistent with the absorption and photoluminescence spectra seen experimentally, as shown in Fig. 18.

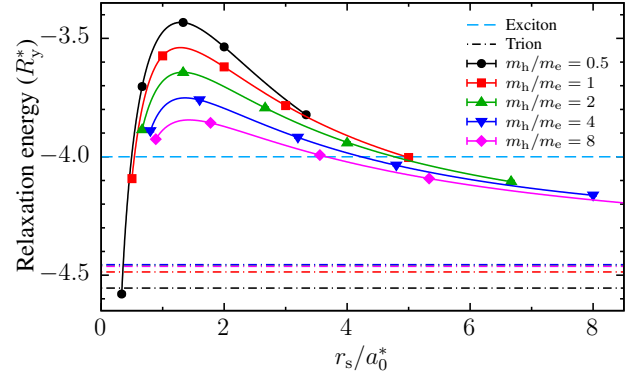


FIG. 17. Electron-hole correlation energy for a single hole immersed in a 2D HEG of density parameter r_s . The length unit is the exciton Bohr radius $a_0^* = 1/\mu$, where $\mu = m_h/(1 + m_h)$ is the reduced mass of the electron-hole pair and m_h is the hole mass. The energy unit is the exciton Rydberg $R_y^* = \mu/2$.

The Fermi energy of the 2D HEG is

$$\mathcal{E}_F = \left(\frac{dE}{dN} \right)_V = \frac{d}{dn}(\epsilon n) = \epsilon(r_s) - \frac{r_s}{2} \frac{d\epsilon(r_s)}{dr_s}, \quad (74)$$

where n is the number density and ϵ is the energy per electron as a function of HEG density parameter r_s . The HF contribution to the energy per particle is readily calculated by pen and paper. DMC was used to calculate the correlation contribution

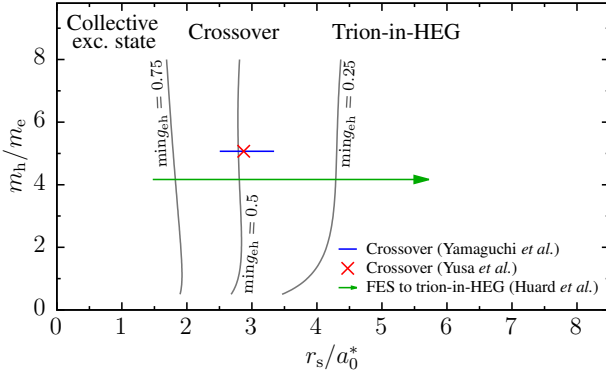


FIG. 18. Crossover between collective exciton behavior and isolated trion behavior. The gray curves identify contours of the minimum value of the electron-hole pair-correlation function g_{eh} , used by Spink *et al.*¹⁰⁹ to identify the phases of the system. The arrow marks the density range over which an evolution from a collective state characterized by a Fermi-edge singularity (FES) to a trion-in-HEG phase was experimentally detected,¹⁷⁸ and the other symbols flag regions of phase space which have been experimentally characterized as being in the crossover regime.^{179–181}

to the energy per particle $\epsilon_c(r_s)$ of the 2D HEG as a function of density.¹⁸² Hence Spink *et al.* were able to calculate the correction to the quasiparticle gap due to the finite concentration of electrons in the conduction band and the resulting electron-hole correlation.

5. Quasiparticle effective mass of the 2D HEG

QMC methods have been used to study intraband excitations in the 2D HEG. Landau’s phenomenological Fermi liquid theory^{115–117} predicts the lifetime of quasiparticle excitations to diverge at the Fermi surface due to Pauli blocking of plane-wave states, so that quasiparticles of a particular momentum \mathbf{k} near the Fermi surface correspond to a well-defined excited-state energy. Hence the quasiparticle energy band can be calculated as described in Sec. III D 1, and is valid in the vicinity of the Fermi surface. This corresponds to the assumption that there is an adiabatic connection between the energy eigenstates of the noninteracting and interacting electron systems, i.e., the nodal topology of the interacting and noninteracting wave functions is the same, or at least sufficiently similar for electronic states in the relevant range of energies. The energy band $\mathcal{E}(\mathbf{k})$ therefore provides the dispersion relationship for long-lived quasielectrons in the all-important energy range near the Fermi surface.

Further from the Fermi surface the difference of total-energy levels ceases to correspond to a quasiparticle excitation due to finite lifetime effects. Nevertheless we can evaluate the energy band $\mathcal{E}(\mathbf{k})$ over a broad range of \mathbf{k} . We may make a linear approximation to the quasiparticle energy band near the Fermi surface, $\mathcal{E}(\mathbf{k}) = \mathcal{E}_F + (k_F/m^*)(k - k_F)$, where \mathcal{E}_F is the Fermi energy, k_F is the Fermi wavevector, and m^* is the quasiparticle effective mass. This effective mass describes the

renormalization of the electron mass by electron-electron interactions (on top of the band effective mass approximation).

The quasiparticle effective mass of a paramagnetic 2D HEG has been the subject of controversy over the decades. Some experiments^{183,184} found a large enhancement of m^* at low density; other experiments^{185,186} contradicted this. *GW* calculations give a range of possible results depending on the choice of effective interaction.¹⁸⁷ Previous QMC studies have predicted (i) much less¹⁸⁸ and (ii) much more¹³¹ enhancement of m^* than found in recent experiments. Experiment¹⁸⁶ and theory¹⁸⁹ suggest that the effective masses in paramagnetic and ferromagnetic HEGs behave quite differently as a function of density.

To calculate the quasiparticle effective mass, the DMC energy band $\mathcal{E}(k)$ was determined at a range of k by taking the energy difference when an electron is added to or removed from a closed-shell ground-state.^{190,191} The resulting energy bands for non-spin-polarized and fully spin-polarized 2D HEGs are plotted in Figs. 19 and 20, respectively. A quartic $\mathcal{E}(k) = \alpha_0 + \alpha_2 k^2 + \alpha_4 k^4$ was fitted to the energy band values, then the effective mass was calculated as $m^* = k_F / (d\mathcal{E}/dk)_{k_F}$. The calculations were performed in finite cells subject to periodic boundary conditions. Finite-size effects are the major sources of error and bias in the QMC results. In our finite simulation cell subject to twisted periodic boundary conditions, the available momentum states fall on the (offset) grid of reciprocal lattice points. This restricts the \mathbf{k} values that we can consider. There are also finite-size errors in the excitation energies due to the neglect of long-range interactions and correlations. These errors have been shown to fall off slowly, as $N^{-1/4}$, near the Fermi surface.¹³¹

In the infinite-system limit, the exact energy band is smooth and well-behaved at the Fermi surface. However, the HF band is pathological. In the infinite-system limit its derivative has a logarithmic divergence at the Fermi surface. In finite systems it behaves very badly. DMC does not entirely remove the pathological behavior from HF theory. Hence we need to consider excitations away from the Fermi surface to obtain the gradient of the energy band at k_F .

The occupied bandwidth is $\Delta\mathcal{E} = \mathcal{E}(k_F) - \mathcal{E}(0) = E_-(0) - E_-(k_F)$. The DMC bandwidth is an upper bound: assuming DMC retrieves the same fraction of the correlation energy in the ground and excited states, the occupied bandwidth will lie between the HF value $E_-^{\text{HF}}(0) - E_-^{\text{HF}}(k_F)$, which is too large, and the exact result $E_-^{\text{exact}}(0) - E_-^{\text{exact}}(k_F)$. Extrapolating the VMC energy with different trial wave functions to zero variance suggests that our DMC calculations retrieve more than 99% of the correlation energy, and that the fraction retrieved is similar in both the ground and excited states. The free-electron bandwidth is greater than or approximately equal to the exact bandwidth. Hence the error in the HF bandwidth is less than or approximately equal to $\Delta\mathcal{E}^{\text{HF}} - \Delta\mathcal{E}^{\text{free}} = k_F(1 - 2/\pi)$. So the error in the DMC bandwidth is less than $0.01k_F(1 - 2/\pi) \approx 0.007/r_s$ for a ferromagnetic HEG and less than about $0.01k_F(1 - 2/\pi) \approx 0.005/r_s$ for a paramagnetic HEG. Since the bandwidth falls off as r_s^{-2} , the error is more significant at large r_s . In the worst case (the paramagnetic HEG at $r_s = 10$) this argument suggests that DMC

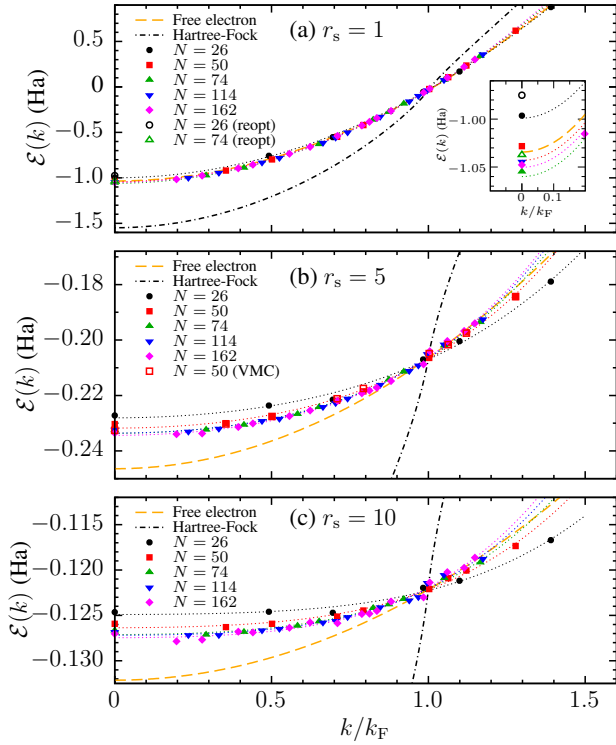


FIG. 19. QMC-calculated energy bands of paramagnetic 2D HEGs at three different density parameters [(a) $r_s = 1$, (b) $r_s = 5$, and (c) $r_s = 10$], for different system sizes N . Also shown are the free-electron ($k^2/2$) and HF bands, offset to match the DMC band at the Fermi wavevector k_F .

overestimates the bandwidth by $\sim 9\%$. In the next-worse case (paramagnetic, $r_s = 5$), the bandwidth is overestimated by $\sim 4\%$. It is reasonable to assume that DMC underestimates m^* by a similar amount. The effective mass against system size is plotted in Fig. 21. The scaling is not the $N^{-1/4}$ predicted by Holzmann *et al.* near the Fermi surface,¹³¹ presumably because we have fitted to the entire band, which shows less severe finite-size errors. In fact we observe an $\mathcal{O}(N^{-1})$ scaling of the finite-size error.

The resulting effective masses of spin-unpolarized and fully spin-polarized 2D HEGs are plotted in Fig. 22. In a paramagnetic HEG the effective mass remains close to the bare electron mass. In a ferromagnetic HEG m^* decreases when the density is lowered. Our results therefore support the qualitative conclusions of Ref. 186. Detailed comparison between theory and experiment is complicated by finite-well-width effects and the effects of disorder. Nevertheless, our results suggest that m^* in paramagnetic 2D HEGs does not grow rapidly as the density is reduced.

6. Positrons immersed in electron gases

DMC methods have been used to solve an important quantum impurity problem, namely that of a single positron immersed in a 3D HEG. The electron-positron correlation en-

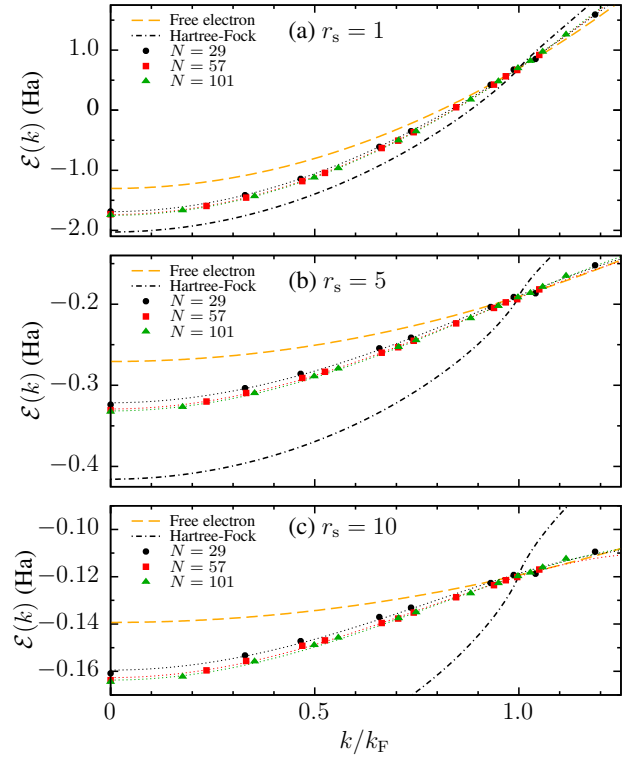


FIG. 20. As Fig. 19, but for full spin-polarized HEGs.

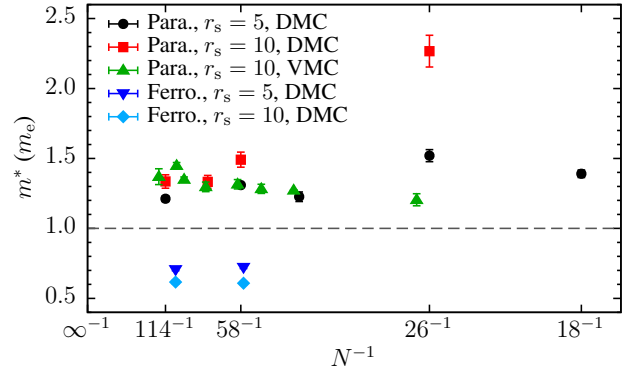


FIG. 21. Quasiparticle effective mass calculated using VMC¹³¹ and DMC¹⁹⁰ against system size N for paramagnetic and fully ferromagnetic 2D HEGs.

ergy, known as the relaxation energy, of a positron in a HEG is required in two-component DFT calculations for positrons in real materials.^{192,193} In addition, the electron-positron contact pair correlation function in the HEG allows DFT calculations of positron annihilation rates in real materials. Such DFT calculations play a crucial role in the interpretation of positron-annihilation spectroscopy (PAS) measurements. In PAS experiments, positrons are injected into materials, where they rapidly thermalize and settle at sites far from nuclei for a relatively long time, before annihilating with the production of two γ photons. The lifetime of the positrons and/or

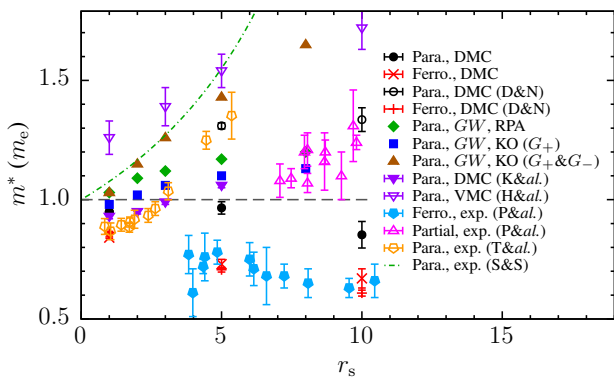


FIG. 22. Quasiparticle effective mass against density parameter r_s for paramagnetic and fully ferromagnetic 2D HEGs, from various measurements and calculations.^{131,183,185–188,190,191}

the momentum density of the annihilation radiation can be measured to obtain information about the electronic charge density and momentum density, including the Fermi surface, at the locations at which the positrons settle.^{194,195} However, positively charged positrons strongly perturb the electronic structure, and so experimentalists require first-principles calculations to relate the measurements to underlying electronic properties.

The ground-state energy of an uncorrelated positron in a homogeneous system is zero; hence the relaxation energy is equal to the ground-state energy difference between a positron-in-HEG system and the HEG without the positron. DMC calculations of the relaxation energy, pair-correlation function and annihilating-pair momentum density have been performed in supercells of 54 electrons.¹⁹⁶ The resulting data differ appreciably from the parameterized results of diagrammatic perturbation theory,^{192,197} and when included in two-component DFT calculations lead to theoretical predictions in improved agreement with experiment.^{198,199}

B. Van der Waals interactions

1. Binding energies of 2D materials

2D materials, which are of great interest due to their extreme mechanical, electronic, and optical properties, consist of atomically thin layers, often based on a honeycomb motif, with strong covalent bonds within each layer and weak van der Waals interactions between layers. Of particular interest are stacked multilayers and heterostructures of 2D materials, which exhibit interesting physical properties including the formation of long-range moiré patterns, natural type-II band alignment leading to electron-hole bilayer behavior, twist-dependent superconductivity, etc. To model the interaction of 2D layers we require an accurate treatment of van der Waals interactions. Unfortunately DFT with local exchange-correlation functionals provides a poor description of van der Waals interactions, which are nonlocal correlations. The development of van der Waals correction schemes for DFT de-

pends in part on the availability of accurate benchmark data for van der Waals bonded systems. QMC methods are capable of providing such benchmark data.

QMC studies of 2D materials require the use of supercells subject to twist-averaged 2D-periodic boundary conditions. When studying heteromultilayers or twisted homomultilayers, the layers must be strained to force them to have a common unit cell. The interlayer binding energy can then be evaluated as the difference of the bilayer energy and the monolayer energy in the limit of large supercell size. Asymptotically, the finite-size errors in the total energies of the monolayer and bilayer go as $N^{-5/4}$ with the number N of electrons in the supercell,¹²⁸ hence the twist-averaged monolayer and bilayer total energies can be extrapolated to infinite system size, allowing the binding energy of the bilayer to be evaluated in the thermodynamic limit. Typically one must perform a series of such binding-energy calculations for the bilayer in which the interlayer distance is varied, then fit a function of interlayer separation to the results. From this fitted function one can find the equilibrium separation and the corresponding equilibrium binding energy, as well as the curvature about the minimum and hence the breathing-mode phonon frequency. If one is studying the binding of heterobilayers or twisted homobilayers, the binding energy per unit cell obtained with artificially commensurate lattice vectors and a particular local lattice offset can be regarded as a local contribution to the interlayer binding energy in a moiré supercell.²⁰⁰ The relaxation of the structure in the moiré supercell can then be described using continuum elasticity theory.

Previous QMC studies have examined the binding energy of layers of hexagonal boron nitride,²⁰¹ layers of graphene in bulk graphite,²⁰² and bilayer graphene.²⁰³ The latter paper shows that DFT calculations with different functionals and van der Waals correction schemes give a wide range of interlayer binding energies and equilibrium separations. The QMC-calculated breathing-mode frequency is in good agreement with Raman spectroscopic measurements. The QMC calculations confirm the experimental observation that Bernal AB stacking is energetically more favorable than AA stacking, with the DMC binding energies being 17.7(9) and 11.5(9) meV per atom, respectively.

2. Van der Waals interactions at surfaces and between molecules

Zen *et al.* have studied a range of molecular crystals, including various water ices, dry ice (carbon dioxide), ammonia, benzene, naphthalene, and anthracene, showing that DMC is capable of achieving chemical accuracy for the lattice-formation energy in each case, unlike RPA and Møller-Plesset second-order perturbation theory (MP2) calculations, which tend to underbind and overbind, respectively.²⁰⁴ Other methods capable of achieving high accuracy for the lattice formation energy, such as RPA with GW single excitations or coupled cluster with single, double, and perturbative triple excitations [CCSD(T)] are more expensive than DMC and/or require difficult fragment-decomposition approaches.

Tsatsoulis *et al.* have examined the adsorption energy of a single water molecule on a (001) LiH surface using quantum chemistry methods, DFT-based methods, as well as DMC.²⁰⁵ They found good agreement between the high-accuracy quantum-chemistry and DMC methods, which are based on very different methodologies, indicating that they provide a reliable benchmark. The MP2 method is found to perform well in this case, whereas van der Waals-corrected DFT methods either underbind or overbind significantly. QMC methods had previously been used to study water molecules on hexagonal boron nitride²⁰⁶ and graphene,²⁰⁷ showing that water is weakly physisorbed on either material, but that the interaction energy is about 15 meV larger in the former case. Although van der Waals DFT functionals overbind water to both materials, this 15 meV difference is consistently reproduced by different first-principles methods.

DMC methods have also been used to predict the phase diagram of water ice trapped between two graphene sheets, finding that the hexagonal and pentagonal phases that are most stable at ambient pressure transition to a square structure at higher pressure.²⁰⁸ These results are of relevance to transmission electron microscopy studies of water confined between graphene sheets. Again, the DMC results reveal inconsistencies and the need for improvement of van der Waals-inclusive DFT functionals.

Taken together these results show the useful and important role that DMC plays in studies of dispersion interactions and in assessing the performance of other methods for treating surfaces.

C. Solid hydrogen

Hydrogen is the simplest and most abundant of the elements, yet it exhibits strikingly rich phase behavior. At low temperature, hydrogen has been observed to form quantum crystalline states and orientationally ordered molecular phases. It has been predicted to exhibit additional exotic behavior, including a liquid-metal phase at high pressure and low temperature, a metallic superfluid state at high pressure, and high-temperature superconductivity.

Experimental X-ray diffraction studies of solid hydrogen have been carried out over a range of pressures up to 190 GPa, providing limited structural information.²⁰⁹ It has not been possible to determine directly the atomic structures of high-pressure phases of solid hydrogen, because samples are small and hydrogen scatters X-rays weakly, making diffraction experiments challenging. Our knowledge of the atomic structure of high-pressure crystalline hydrogen arises in large part from computational modeling in conjunction with experimental Raman and infrared spectroscopy.

Experiments at around room temperature have demonstrated that at least four distinct structural states exist at low temperature and at pressures up to 350 GPa, which are known as phases I, II, III, and IV. The low-pressure phase I has a hexagonal close-packed structure with freely rotating molecules at the lattice sites.²¹⁰ Phase II is a broken-symmetry phase in which molecular rotations are hindered at low tem-

perature. As the pressure is increased at low temperature a further phase transition occurs from phase II to phase III at about 160 GPa.²¹¹ A phase IV has been discovered at temperatures above a few hundred K and pressures above 220 GPa.²¹² A phase IV' has also been reported, which is very similar to phase IV.

DFT calculations are relatively inexpensive, and they have been used to search for low-energy hydrogen structures with vibrational properties that are in reasonable agreement with experimental Raman and infrared measurements for phases II, III,²¹³ and IV^{214,215} of hydrogen over a range of pressures. The structures were discovered using the “*ab initio* random structure searching” (AIRSS) method.^{216,217} Phase II may be modeled by a structure of $P2_1/c$ symmetry. The first realistic structures for phase III of hydrogen were proposed by Pickard and Needs.²¹³ They found a structure of $C2/c$ symmetry with 24 atoms in the primitive unit cell (see Fig. 23) to be the most stable. A very similar structure was subsequently

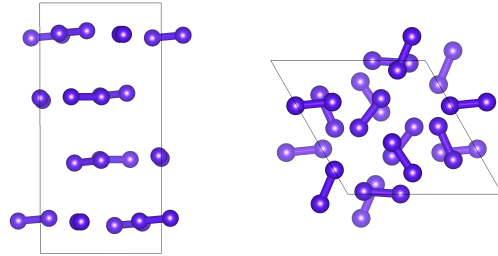


FIG. 23. Side view (left) and top view (right) of the $C2/c$ -24 structure. This is essentially a layered structure with small deviations.

discovered, which exhibits a hexagonal $P6_122$ symmetry,²¹⁸ and is depicted in Fig. 24. Two different phase-III-like struc-

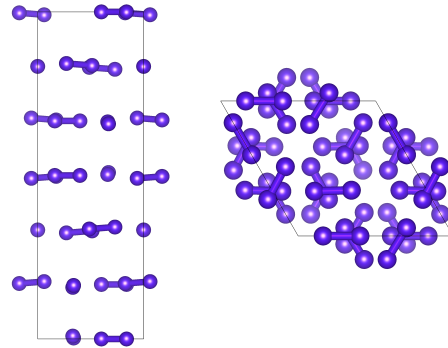


FIG. 24. Side view (left) and top view (right) of the $P6_122$ structure.

tures may be formed at high pressure: hexagonal $P6_122$ below about 200 GPa and monoclinic $C2/c$ at higher pressures. It has been found that quantum nuclear and thermal vibrations play a central role in stabilizing the $P6_122$ phase. It is possible that other similar phase-III-like structures of hydrogen may exist. A structure of Pc symmetry, shown in Fig. 25, was suggested as the best candidate for phase IV at high pressure.

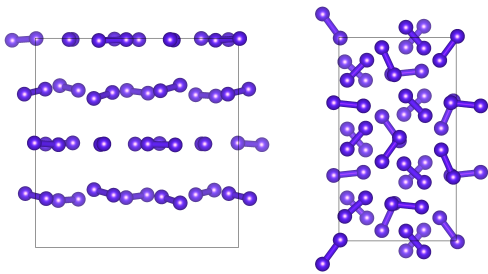


FIG. 25. Side view (left) and top view (right) of the Pc -48 structure.

Although the low-energy structures modeling particular hydrogen phases can be identified by their vibrational properties, the DFT phase diagram obtained using those structures does not agree with experiment. Pickard *et al.* showed that standard DFT functionals predict metallic structures of hydrogen to be energetically favorable over a range of pressures up to 400 GPa.²¹⁵ However, it is known experimentally that, at the pressures of interest here (below 400 GPa), metallic structures are not in fact favored. This is an example of the “band-gap problem,” in which DFT electronic band gaps are found to be significantly smaller than those obtained in experiments. DMC methods have therefore been used to achieve a consistent theoretical understanding of high-pressure hydrogen, by calculating the phase diagram of solid molecular hydrogen using the candidate structures identified in DFT calculations.¹¹¹ Twist-averaged DMC calculations with Slater-Jastrow wave functions were performed in cells of 96 and 768 atoms. Finite-size corrections were evaluated using the different methods described in Sec. III B 2 as well as extrapolation of the twist-averaged DMC energies per atom to the thermodynamic limit, assuming the finite-size error goes as N^{-1} . The fact that different finite-size-correction methods give results in agreement with each other suggests that finite-size effects are well-controlled. Vibrational energy contributions play an important role in hydrogen, and anharmonicity cannot be neglected on the fine energy scale that has to be resolved (less than 1 meV per atom). DFT anharmonic vibrational free energies²¹⁹ were therefore added to the static-lattice DMC energy data.¹¹¹ By comparing the Gibbs free energies of the phases, the pressure-temperature phase diagram was evaluated and found to be in qualitative agreement with experiment. In fact good quantitative agreement with experiment for the temperature of the transition between phases III and IV was achieved. The calculated pressure for the transition between phases II and III is 75 GPa larger than found in experiment, although the isotope dependence of the II–III transition is well-reproduced.²¹¹ Most importantly, the DMC calculations show that the metallic structure that is strongly favored in DFT at high pressure is not energetically competitive, resolving an outstanding disagreement between theory and experiment.

Experiments have revealed the existence of an additional phase V at pressures above 325 GPa.²²⁰ DFT calculations have been performed to identify the structure of phase V using the “saddle-point AIRSS” approach, which searches for structures stabilized by anharmonic nuclear motion. The vi-

brational results suggest that a $Pca2_1$ structure is a promising model for phase V.²²¹ DMC was used to demonstrate that this is indeed one of the most energetically stable candidate structures. It is interesting to note that a large fraction of the low-energy structures of compressed hydrogen at low temperature adopt layered forms. This observation may make it easier to determine low-enthalpy structures of a wider range of hydrogen phases.

The effects of van der Waals forces can be important in systems such as solid hydrogen, which is one of the reasons why DMC calculations are more reliable than DFT.²²² Azadi *et al.*²²³ used DMC methods to calculate structures of solid hydrogen at high pressure, including vibrational effects within a self-consistent-field approach.²²⁴ Azadi and Kühne performed DMC calculations in which eleven molecular hydrogen structures with different symmetries were found to be the most energetically competitive phases within the pressure range studied of 100–500 GPa, concluding that phase III may be polymorphic.²²⁵

The electronic structure of high-pressure hydrogen is also of great interest, with much experimental effort being directed towards the discovery of metallic hydrogen. Metalization of solid hydrogen at high pressure occurs via a structural phase transition rather than band-gap closure.²²⁶ The difference between the DMC and DFT band gaps in high-pressure hydrogen is almost independent of system size, and therefore can be applied as a scissor correction to the infinite-system DFT gap to obtain the DMC gap in the thermodynamic limit.²²⁷ Comparisons of static-nucleus DMC energy gaps with the available experimental data demonstrate the important role played by nuclear quantum effects in the electronic structure of solid hydrogen. DMC calculations of quasiparticle and excitonic band gaps have shown that the exciton binding energies in hydrogen at the high pressures of interest are smaller than 100 meV.²²⁷

Recent experiments by Eremets *et al.* have suggested that semimetallic molecular structures may occur at pressures above 350 GPa.²²⁸ There are still many fascinating questions to address in this field using QMC methods in conjunction with DFT structure searching and anharmonic vibrational methods.

V. FUTURE DIRECTIONS FOR QMC METHODS AND THE CASINO SOFTWARE

A. Synergy with FCIQMC

The master equation of FCIQMC can be obtained by substituting the CI wave function of Eq. (39) with imaginary-time-dependent coefficients $c_i = c_i(\tau)$ into the imaginary-time Schrödinger equation of Eq. (19),²⁵

$$-\frac{\partial c_i}{\partial \tau} = (H_{ii} - S)c_i + \sum_{j \neq i} H_{ij}c_j, \quad (75)$$

where $H_{ij} = \langle D_i | \hat{H} | D_j \rangle$ and S is an adjustable energy shift. This equation governs the dynamics of discrete walkers that

sample the second-quantized Hilbert space of all Slater determinants that can be constructed given a fixed orbital basis. The initiator approximation modifies these dynamics to prevent the extremely rapid growth in walker number at the start of the calculation.^{26,27} The computational cost of FCIQMC, both without and with the initiator approximation, formally scales as an exponential of the system size, but in practice it is possible to study medium-sized systems with reasonable computational cost.^{229–232}

The combination of FCIQMC and DMC is especially appealing in that one method excels where the other struggles: FCIQMC easily recovers static correlation but requires very large numbers of walkers to account for dynamical correlation, whereas DMC correctly captures dynamical correlation but is incapable of modifying the description of static correlation provided by the trial wave function. There are three broad approaches to combining FCIQMC and DMC, detailed below.

1. FCIQMC wave functions in DMC

Initiator-FCIQMC is capable of quickly identifying the most important determinants in the CI wave function and providing their approximate coefficients. The most obvious combination of FCIQMC and DMC is simply to build the trial wave function by selecting a certain number of determinants in the FCIQMC wave function with the largest weights.

As an example of the direct application of this principle, Fig. 26 shows the VMC and DMC energies, with and without backflow, of the 3D paramagnetic 14-electron gas at a density of $r_s = 0.5$ at the Γ point as a function of the size of the truncated expansion. The coefficients of the multideterminant expansion have been reoptimized in the presence of the Jastrow factor and, where applicable, of the backflow parameters. The coefficients of symmetry-equivalent determinants in the expansion have been constrained to be equal, so, e.g., the expansion with 30 optimizable coefficients contains 1273 determinants, by which point the backflow DMC energy has converged to within $\sim 35 \mu\text{Ha}$ per electron of the exact energy, obtained using FCIQMC.²³³ Also shown in Fig. 26 is the basis-set limit of the backflow DMC energy obtained with an MAGP wave function,¹⁰⁴ described in Sec. III A 3.

While these results are very encouraging, total energies for other systems converge rather more slowly with expansion size, and moderate fixed-node errors should be expected with medium-sized expansions. This represents a significant challenge for computing energy differences, since at fixed expansion size any two systems will likely incur different fixed-node errors, biasing the result. This mismatch can be addressed by using different expansion sizes for each system. For example, to obtain the binding energy of the C_2 molecule to chemical accuracy one would need to subtract twice the DMC energy of a carbon atom computed with a single-determinant wave function from the DMC energy of the carbon dimer obtained with a 60-determinant expansion. Criteria to decide how to truncate the CI expansion for each system to achieve similar fixed-node errors have been proposed.^{87,234} Another approach is to use very large expansions in order to reduce the fixed-

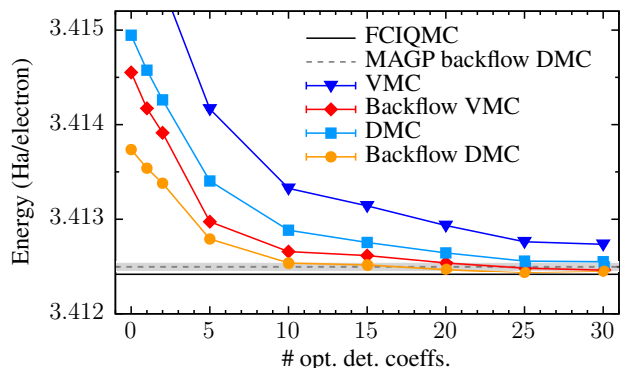


FIG. 26. VMC and DMC energy of the $r_s = 0.5$ 3D paramagnetic 14-electron gas at Γ as a function of the size of the multideterminant expansion obtained by truncation of the FCIQMC wave function. Note that the zero of the horizontal axis corresponds to the single-determinant wave function. The exact (FCIQMC) energy is also shown, along with the MAGP backflow DMC result, with its uncertainty represented by the shaded area.

node errors in both systems to the point that the bias in the difference is negligible. However this is very expensive and wave-function optimization is problematic. A third potential solution is to construct an extrapolation procedure to obtain the exact energy from DMC energies at various expansion sizes, which is the subject of current research.

2. Jastrow factors and similarity-transformed FCIQMC

A recent development in the FCIQMC method is the ability to use Jastrow factors,^{235–237} effectively using a wave function of the form $e^{\hat{J}}|\Psi_{\text{CI}}\rangle$, where $e^{\hat{J}}$ is the second-quantized operator associated with the first-quantized Jastrow factor e^J . Since the FCIQMC method requires the wave function to be expressible as a CI expansion, the Jastrow factor must be incorporated into the Hamiltonian. The effective Hamiltonian in the similarity-transformed FCIQMC (ST-FCIQMC) method is therefore $e^{-\hat{J}}\hat{H}e^{\hat{J}}$, which is non-Hermitian. FCIQMC then serves as a solver for the right eigenvector of this effective Hamiltonian, which is a CI expansion.

The similarity transformation modifies the Hamiltonian matrix elements, and the presence of electron-electron Jastrow factor terms requires the evaluation of six-index integrals in addition to the four-index integrals used in regular FCIQMC, as well as other methods in quantum chemistry. The ST-FCIQMC method has greater memory requirements than FCIQMC, but the expectation is that this can be offset by the reduction in computational cost afforded by a more compact representation of the wave function.

While calculations for first-row atoms appear to suggest that ST-FCIQMC energies are relatively insensitive to the quality of the Jastrow factor,²³⁶ more complex systems can be expected to benefit from well-optimized Jastrow parameters. The VMC method provides the ideal framework to optimize these Jastrow factors.

The presence of the Jastrow factor decreases the importance of those determinants in the CI wave function that are mainly associated with dynamical correlation. These CI expansions should therefore make great trial wave functions for DMC.

3. DMC-assisted FCIQMC and benchmarking

The very high accuracy of FCIQMC for small systems and the feasibility of performing DMC calculations for large systems can also be exploited. This can be done by combining the results of independent FCIQMC and DMC calculations without sharing wave functions between the two methods.

As described in Sec. IV A 2, it is possible to evaluate the exact energy of the high-density HEG at small system sizes using FCIQMC, and to obtain DMC energies at large system sizes that enable a very reliable extrapolation of the fixed-node energy to infinite system size. Modeling the fixed-node error as a slowly-varying function of system size then allows the estimation of the exact correlation energy of the infinite system.

This approach has been applied to the ferromagnetic electron gas,¹⁷¹ and work is underway to produce a similar set of results for the paramagnetic electron gas. Besides obtaining the correlation energy itself, the fixed-node error is produced as a byproduct of this process. This opens the intriguing possibility of parameterizing a functional of the density to estimate the fixed-node error incurred in inhomogeneous systems, which could be used to construct a correction for fixed-node DMC energies.

Another, more straightforward possibility is to use very accurate FCIQMC results for small systems as a benchmark against which to gauge the accuracy of DMC, which is extremely useful in the development of trial wave functions and other methodological advances.

B. Towards greater efficiency

1. Electron-electron pseudopotentials

The Coulomb potential energy diverges as $1/r$ when two electrons approach each other. Even if the Kato cusp conditions are imposed on the wave function, this divergence leads to additional variance in the local energy and hence computational expense in QMC calculations. To alleviate this problem, Lloyd-Williams *et al.* proposed replacing the electron-electron interaction potential by a local pseudopotential that reduces to $1/r$ outside a cutoff radius, is smooth everywhere, and approximately reproduces the scattering states of the $1/r$ Coulomb potential up to the Fermi wavevector of a HEG of density parameter $r_s = 2$.²³⁸ Their pseudopotential is of polynomial form inside the cutoff radius, with the polynomial coefficients chosen to minimize the mean squared difference of the logarithmic derivatives of the exact and pseudo scattering wave functions at the cutoff radius. The resulting electron-electron pseudopotentials were shown to give accurate results for HEGs at a range of densities with a speedup of up to thirtyfold. Lloyd-Williams *et al.* also showed the effectiveness of

the approach for lithium and beryllium atoms. This is clearly a promising technique to investigate for QMC studies of inhomogeneous condensed-matter systems.

2. Current developments in computer architectures

As explained in Sec. III E 1, QMC methods are well-placed to take advantage of the widespread availability of massively parallel computer architectures. CASINO has also been shown to perform well on many-core processors such as Intel's Knights Landing processor, although Intel has now ended the development of this particular class of processor. It is found to be advantageous to make full use of hyperthreading, due to the high parallel performance of QMC methods.

Graphics processing units (GPUs) are now a common accelerator technology on supercomputers. Unfortunately, allowing QMC methods to make efficient use of GPUs is a nontrivial task if the CPU and GPU do not share memory. QMC methods for practical problem sizes involve a large number of small tasks that must be performed in a particular sequence, dependent on the outcome of random processes such as accept/reject steps or the sampling of branching factors. The steps include small pieces of linear algebra relating to determinant-updating, evaluation of polynomials and small plane-wave expansions in the Jastrow factor, and pairwise sums of interaction potentials. If these small tasks are devolved to the GPU one by one then they incur a prohibitive time overhead due to the need to transfer arrays between the CPU and GPU. We have experimented with a light-touch, compiler-driven GPU acceleration of CASINO by means of OpenACC compiler directives. Unfortunately, this results in significant slowdowns to CASINO due to the cost of transferring data between the CPU and the GPU. A successful implementation of QMC on a GPU requires a very significant modification such that a large part of the calculation is performed on the GPU. Given the comments about fundamental limits on the scaling of QMC methods in Sec. III E 2, it is not clear whether the benefits of GPUization justify forking CASINO into GPU and non-GPU versions. On the other hand, modern developments in heterogeneous computing architectures that allow CPUs and GPUs to share memory are expected to remove many of these issues with GPU computing.

C. Improved accuracy

1. Spin-orbit coupling

Spin-orbit coupling plays a key role in the electronic structure of many elements beyond the first row; however, it is not generally included in QMC calculations because of the additional complication and expense of using spinor wave functions. Recent work²³⁹ has demonstrated the feasibility of developing DMC for spin-orbit interactions using a continuous-spin approach. An alternative approach would be to use an algorithm similar to the pseudopotential T-move scheme¹⁴¹ to

propose spin flips. Suppose the Hamiltonian contains a spin-dependent term \hat{V}_{SO} . Let τ be the DMC time step, \mathbf{X} be the space and spin coordinates of all the electrons, Ψ be the trial

wave function and

$$V_{\mathbf{X},\mathbf{X}'} = \langle \mathbf{X} | \hat{V}_{\text{SO}} | \mathbf{X}' \rangle \frac{\Psi(\mathbf{X})}{\Psi(\mathbf{X}')}. \quad (76)$$

Let $V_{\mathbf{X},\mathbf{X}'}^+ = \max\{V_{\mathbf{X},\mathbf{X}'}, 0\}$ and $V_{\mathbf{X},\mathbf{X}'}^- = \min\{V_{\mathbf{X},\mathbf{X}'}, 0\}$, so that $V_{\mathbf{X},\mathbf{X}'} = V_{\mathbf{X},\mathbf{X}'}^- + V_{\mathbf{X},\mathbf{X}'}^+$. Then the importance-sampled DMC Green's function for the spin-dependent part of the Hamiltonian is

$$\langle \mathbf{X} | \exp(-\tau \hat{V}_{\text{SO}}) | \mathbf{X}' \rangle \frac{\Psi(\mathbf{X})}{\Psi(\mathbf{X}')} \approx \left[\frac{\delta_{\mathbf{X},\mathbf{X}'} - \tau V_{\mathbf{X},\mathbf{X}'}^-}{1 - \tau \sum_{\mathbf{X}''} V_{\mathbf{X}'',\mathbf{X}'}^-} \right] \exp\left(-\tau \sum_{\mathbf{X}''} V_{\mathbf{X}'',\mathbf{X}'}\right) + \mathcal{O}(\tau^2), \quad (77)$$

where the negative matrix elements are treated exactly to $\mathcal{O}(\tau)$, while the positive matrix elements are localized by replacing $\langle \mathbf{X} | \hat{V}_{\text{SO}} | \mathbf{X}' \rangle$ with $\delta_{\mathbf{X},\mathbf{X}'} [\hat{V}_{\text{SO}} \Psi(\mathbf{X})] / \Psi(\mathbf{X})$. To apply this Green's function to a walker, the first factor on the right-hand side of Eq. (77) is treated as a spin-flip transition probability, while the second factor is absorbed into the weight of the walker. The contribution to the local energy from \hat{V}_{SO} is

$$\frac{\langle \Psi | \hat{V}_{\text{SO}} | \mathbf{X}' \rangle}{\langle \Psi | \mathbf{X}' \rangle} = \sum_{\mathbf{X}''} V_{\mathbf{X}'',\mathbf{X}'}. \quad (78)$$

This approach leads to a positive error in the DMC energy, which is second order in the error in the trial wave function.

2. Inclusion of vibrational effects in *ab initio* QMC calculations

QMC methods provide highly accurate solutions to the electronic Schrödinger equation, but comparison with experiment is always complicated by the need to include vibrational corrections. Such corrections are usually evaluated at the DFT level. For ground-state total-energy calculations the inclusion of quasiharmonic vibrational Helmholtz free energies is straightforward and many first-principles DFT codes have the ability to perform such calculations automatically. By contrast, codes and scripts for calculating the vibrational renormalization of band gaps within DFT are not generally available. However, vibrational effects often alter gaps by a significant fraction; e.g., the gap of a benzene molecule is reduced by more than 0.5 eV due to vibrational effects.²⁴⁰ The vibrational renormalizations of quasiparticle and excitonic gaps have different physical origins: the total energies that define the quasiparticle gap should include vibrational Helmholtz free energies, while the excitonic (optical absorption) gap should be averaged over the distribution of nuclear positions at temperature T in the electronic ground state. For light nuclei (first row), zero-point renormalization is the most important vibrational effect, with temperature dependence being relatively weak; on the other hand heavier nuclei behave

classically, leading to negligible zero-point vibrational renormalization, but significant temperature dependence.

It would be straightforward to implement a QMC vibrational-renormalization approach based on the Born-Oppenheimer approximation and the DFT potential-energy landscape, with nuclear coordinates sampled randomly from a vibrational self-consistent-field wave function.²¹⁹ QMC would be used to evaluate the energy gap at each nuclear configuration sampled. The statistical error bar on the QMC gap falls off as the reciprocal of the square root of the amount of QMC data gathered, irrespective of whether those data are gathered at different nuclear configurations. The cost of a vibrationally renormalized QMC gap calculation is not therefore expected to be significantly larger than the cost of a static-nucleus gap calculation.

Hydrogen is an important constituent of many compounds, and the hydrogen nuclei in most materials are more than an order of magnitude lighter than the other nuclei. Protons could be treated as distinguishable quantum particles in *ab initio* QMC calculations, since proton exchange effects are small. Including protons in QMC calculations will require (i) bespoke backflow functions to ensure that the cusps in DFT-generated electron orbitals occur at the electron-proton coalescence points, even when the protons have moved; (ii) a bespoke Jastrow factor of Gaussian form in either the proton positions (Einstein approximation) or the phonon normal coordinates (from a DFT quasiharmonic phonon calculation with the nonhydrogen nuclei having infinite mass); and (iii) modifications to the DMC Green's function near electron-proton coalescence points similar to those currently used between electrons and fixed nuclei.⁶⁸ A novel extension for hydrogen-bearing compounds would be to include the electrons and protons in the QMC calculations, evaluate the matrix of force constants for the remaining nuclei using QMC,²⁴¹ and then evaluate their zero-point energy within the quasiharmonic approximation. We would then have a fully quantum treatment of the protons and a quasiharmonic treatment of the remaining nuclei.

D. Atomic forces from QMC calculations

The evaluation of atomic forces in QMC is complicated by statistical issues. These affect other expectation values too, such as elements of the matrix of force constants.²⁴¹ Here we focus on VMC forces for simplicity, but the methodology applies to DMC forces as well.

The force exerted on the I th atom of a system along Cartesian direction x is minus the derivative of the expectation value of the Hamiltonian with respect to the x component of the position of the nucleus. The local force F is the sum of a Hellmann-Feynman term and a Pulay term,

$$F(\mathbf{R}) = \sum_i \frac{-Z_I x_{iI}}{r_{iI}^3} + \sum_{J \neq I} \frac{Z_I Z_J x_{IJ}}{r_{IJ}^3} - 2\Psi^{-1}(\mathbf{R}) \left[E(\mathbf{R}) - \langle \hat{H} \rangle \right] \frac{\partial \Psi(\mathbf{R})}{\partial x}. \quad (79)$$

The local force follows a fat-tailed distribution satisfying $P(F) \sim |F|^{-5/2}$ as $|F| \rightarrow \infty$ due to the error in the nodes of the trial wave function.^{40,242,243} This causes statistical problems, since the expectation value of the force,

$$\langle F \rangle = \int_{-\infty}^{\infty} P(F) F \, dF, \quad (80)$$

is well defined, but its variance,

$$\text{Var}[F] = \sigma_F^2 = \int_{-\infty}^{\infty} P(F) (F - \langle F \rangle)^2 \, dF, \quad (81)$$

is divergent. This implies that the standard error is divergent and cannot be used to provide a confidence interval. The application of a zero-variance estimator of the force^{244,245} and the use of pseudopotentials,^{246,247} both supported in CASINO, alleviate the severity of the problem, but do not completely solve it.

Other methods have been proposed to tackle the fat tails in the local force distribution by approximating the local observable²⁴⁸ or modifying the VMC sampling distribution.²⁴⁹ Recently, López Ríos and Conduit²⁵⁰ showed that this issue can be solved in the statistical analysis stage using a conceptually simple approach. (We focus on the right-hand tail for simplicity; the left-hand tail is treated analogously.) Let us assume that a random sample of M independent values of the local force distributed according to $P(F)$ is available, and let $F^{(m)}$ be the m th largest value in the sample. By definition, the sample quantile $q_m = (m - 1/2)/M$ satisfies

$$\int_{F^{(m)}}^{\infty} P(F) \, dF \approx q_m. \quad (82)$$

Let F_R be a value of the local force such that the probability distribution can be accurately represented by

$$P(F) = \sum_{n=0}^{n_R} c_n |F - F_c|^{-\frac{5+n}{2}}, \quad F > F_R, \quad (83)$$

where $\{c_n\}$ are $n_R + 1$ unknown coefficients and F_c is a parameter which is set to the sample median $F^{(M/2)}$ in practice. Substituting Eq. (83) into Eq. (82) yields

$$q_m = \sum_{n=0}^{n_R} \frac{c_n}{(3+n)/2} |F^{(m)} - F_c|^{-(n+3)/2}. \quad (84)$$

Defining

$$y_m = q_m \left| F^{(m)} - F_c \right|^{3/2}, \quad (85)$$

$$x_m = |F^{(m)} - F_c|^{-1/2},$$

and rearranging, Eq. (84) reduces to a polynomial,

$$y_m = \sum_{n=0}^{n_R} \frac{c_n}{(3+n)/2} x_m^n. \quad (86)$$

By this change of scale, which we refer to as “ yx scale,” the tail of the probability distribution function, which takes vanishingly small values over an infinite range and involves non-linear parameters, is transformed into a finite function defined over a finite range involving linear parameters.

The asymptotic coefficients $\{c_n\}$ can be obtained by using Eq. (86) as a fit function for the tail data in the sample, converted according to Eq. (85). The use of fit weights

$$w_m = \left(\ln \frac{q_{M_R+1}}{q_m} \right)^{-1} |F^{(m)} - F_c|^{-3/2}, \quad (87)$$

where M_R is the number of data points in the sample such that $F > F_R$, ensures the asymptotic normality of the regression coefficients. This enables the evaluation of

$$\mathcal{F} = \frac{1}{M} \sum_{m > M_R} F^{(m)} + \sum_{n=0}^{n_R} c_n \int_{F_R}^{\infty} |F - F_c|^{-(5+n)/2} F \, dF, \quad (88)$$

referred to as the tail-regression estimator (TRE) of $\langle F \rangle$. In Eq. (88), the central contribution to the expectation value is evaluated as a (partial) sample mean, while the integrals in the tail contribution can be computed analytically. Crucially, \mathcal{F} is asymptotically normally distributed since its uncertainty arises from that on $\{c_n\}$ and from the finite variance of the central part, completely bypassing the use of the divergent variance σ_F^2 . Parameters F_R and n_R are chosen so as to minimize the uncertainty in the resulting estimator. We note that Ref. 250 did not provide closed expressions for the uncertainty in \mathcal{F} , instead relying on the bootstrap method²⁵¹ to compute confidence intervals.

We demonstrate the application of the TRE in Fig. 27 using VMC force data for the all-electron lithium hydride molecule at a compressed bond length of 2 bohr using a trial wave function consisting of Slater determinants of HF orbitals expanded in the cc-pVDZ basis set⁸³ multiplied by a DTN Jastrow factor. The TRE of the sum of the Hellmann-Feynman and Pulay forces on the lithium atom in the direction of the hydrogen atom is $-0.54(6)$ Ha bohr⁻¹. By contrast, the standard estimator of this force is -0.46 Ha bohr⁻¹ with a formally divergent uncertainty (numerically the apparent uncertainty is 0.18 Ha bohr⁻¹).

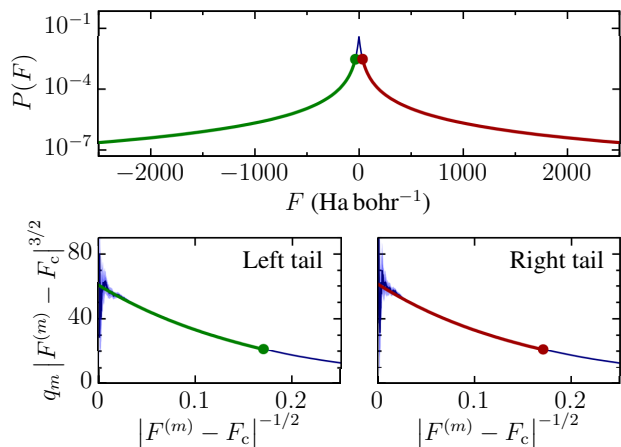


FIG. 27. Application of the tail-regression procedure to a sample of 10^7 VMC local forces on the lithium atom of the all-electron LiH molecule at a compressed bond length of 2 bohr. The thin line in the top panel represents the estimated probability distribution (note the logarithmic scale of the vertical axis), and the thin lines in the lower panels correspond to the tails of the estimated probability distribution in the yx scale defined in Eq. (85). Fits are shown as thick lines in the three panels, with the left and right fit onsets, F_L and F_R , shown as circles. 68.3% and 95.4% confidence intervals obtained from the bootstrap are shown as shaded areas.

The TRE is thus an asymptotically normally distributed estimator of expectation values associated with fat-tailed probability distributions of known tail exponents, yielding uncertainties proportional to $M^{-1/2}$, and can be used for distributions with other tail exponents and for higher moments of the distribution, such as the variance of the local energy for which standard confidence intervals are also formally undefined.

The TRE method as described in Ref. 250 has the limitation of requiring serially-uncorrelated samples, which in practice implies spending additional computational time in the accumulation stage to achieve decorrelation. Whether this restriction can be worked around in the future or not, the TRE unequivocally shows that the problem of evaluating expectation values in QMC associated with fat-tailed probability distributions of known asymptotic behavior can be solved at the statistical level.

VI. CONCLUSIONS

We have explained the theory underpinning the VMC and fixed-phase DMC methods, and commented on some of the practical issues affecting QMC codes such as CASINO. We have shown many of the strengths and also limitations of these methods. QMC methods are highly successful for studying systems featuring van der Waals interactions, they give a powerful alternative to many-body GW methods for calculating accurate quasiparticle and excitonic band gaps, and they have unique capabilities for accurately solving the Schrödinger equation in model electron(-hole) systems (indeed, providing exact solutions to the Schrödinger equation for charge-carrier

complexes). On the other hand we have argued that it will prove difficult to perform useful DMC calculations for systems with several thousand electrons with current algorithms, irrespective of developments in supercomputer hardware. The successful use of DMC in studies of condensed matter therefore very much depends on the analysis of finite-size effects and/or extrapolation to the thermodynamic limit. Furthermore, dealing with systems characterized by strong static correlation effects remains challenging, because multideterminant expansions introduce an unwelcome element of arbitrariness into DMC results. Despite these limitations, QMC methods continue to play a role of fundamental importance in the ecosystem of first-principles computational methods, providing data that expand the accuracy and capabilities of simpler methods.

ACKNOWLEDGMENTS

The authors would like to thank the Theory of Condensed Matter group at the University of Cambridge, the Condensed Matter Theory group at Lancaster University, and the Electronic Structure Theory group at the Max-Planck Institute for Solid State Research for their support over the years, and the broader QMC community for its vibrancy and for many useful discussions. We would like to thank our long-term collaborators Gareth Conduit, Bartomeu Monserrat, Matthew Foulkes, Dario Alfè, Chris Pickard, and Ryo Maezono for their contributions to our QMC project. We also thank Alice Shipley for providing the hydrogen figures shown in this paper. The calculations used to produce Fig. 12 were performed by Elaheh Mostaani. Financial support was provided by the Engineering and Physical Sciences Research Council of the United Kingdom under Grant No. EP/P034616/1. Supporting research data can be freely accessed at [252], in compliance with the applicable Open Data policies.

- ¹P. A. M. Dirac and R. H. Fowler, *Proc. Royal Soc. Lond., Series A* **123**, 714 (1929).
- ²E. A. Hylleraas, *Z. Phys.* **54**, 347 (1929).
- ³H. Nakashima and H. Nakatsuji, *J. Chem. Phys.* **127**, 224104 (2007).
- ⁴J. C. Slater, *Phys. Rev.* **35**, 210 (1930).
- ⁵V. Fock, *Z. Phys.* **61**, 126 (1930).
- ⁶D. R. Hartree and W. Hartree, *Proc. Royal Soc. Lond., Series A* **150**, 9 (1935).
- ⁷A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Dover, 1989).
- ⁸I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, Cambridge, 2009).
- ⁹P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- ¹⁰W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ¹¹K. Burke, *J. Chem. Phys.* **136**, 150901 (2012).
- ¹²L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- ¹³W. L. McMillan, *Phys. Rev.* **138**, A442 (1965).
- ¹⁴D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- ¹⁵W. M. C. Foulkes, L. Mitás, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- ¹⁶R. J. Needs, M. D. Towler, N. D. Drummond, and P. López Ríos, *J. Phys.: Condens. Matter* **22**, 023201 (2010).
- ¹⁷J. B. Anderson, *J. Chem. Phys.* **65**, 4121 (1976).
- ¹⁸T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grüneis, *Phys. Rev. X* **8**, 021043 (2018).
- ¹⁹T. Gruber and A. Grüneis, *Phys. Rev. B* **98**, 134108 (2018).

- ²⁰T. Tsatsoulis, S. Sakong, A. Groß, and A. Grüneis, *J. Chem. Phys.* **149**, 244105 (2018).
- ²¹K. Liao, X.-Z. Li, A. Alavi, and A. Grüneis, *Npj Comput. Mater.* **5**, 110 (2019).
- ²²D. M. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).
- ²³S. Baroni and S. Moroni, *Phys. Rev. Lett.* **82**, 4745 (1999).
- ²⁴S. Sorella, S. Baroni, R. Car, and M. Parrinello, *Europhys. Lett.* **8**, 663 (1989).
- ²⁵G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- ²⁶D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
- ²⁷K. Ghanem, A. Y. Lozovoi, and A. Alavi, *J. Chem. Phys.* **151**, 224108 (2019).
- ²⁸I. Sabzevari and S. Sharma, *J. Chem. Theory Comput.* **14**, 6276 (2018).
- ²⁹E. Neuscamman, *Mol. Phys.* **114**, 577 (2016).
- ³⁰J. 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, 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. V. D. Goetz, J. E. Vincent, D. C. Yang, Y. Yang, S. Zhang, and L. Zhao, *J. Phys. Condens. Mater.* **30**, 195901 (2018).
- ³¹L. K. Wagner, M. Bajdich, and L. Mitas, *J. Comput. Phys.* **228**, 3390 (2009).
- ³²N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ³³W. K. Hastings, *Biometrika* **57**, 97 (1970).
- ³⁴B. L. Hammond, J. W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
- ³⁵R. M. Lee, G. J. Conduit, N. Nemeč, P. López Ríos, and N. D. Drummond, *Phys. Rev. E* **83**, 066706 (2011).
- ³⁶C. J. Umrigar, *Phys. Rev. Lett.* **71**, 408 (1993).
- ³⁷H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- ³⁸U. Wolff, *Comput. Phys. Commun.* **156**, 143 (2004).
- ³⁹M. Jonsson, *Phys. Rev. E* **98**, 043304 (2018).
- ⁴⁰J. R. Trail, *Phys. Rev. E* **77**, 016703 (2008).
- ⁴¹The use of the terms “static correlation” and “dynamic correlation” in the quantum chemistry literature is less clearcut than suggested by our QMC-focused definition in terms of qualitative and quantitative errors in the nodal surface of the wave function. Nevertheless, the definition we have used captures the idea that static correlations are due to errors in the Hartree–Fock state, while dynamical correlations are due to the electrons avoiding each other.
- ⁴²S. F. Boys, N. C. Handy, and J. W. Linnett, *Proc. R. Soc. Lond. A*, **310**, 63 (1969).
- ⁴³T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- ⁴⁴R. T. Pack and W. Byers Brown, *J. Chem. Phys.* **45**, 556 (1966).
- ⁴⁵For a strictly one-dimensional system in which the particles interact via the Coulomb $1/r$ potential, the wave function must go to zero at all coalescence points in order for the energy expectation value to be well-defined. In practice the easiest way of achieving this requirement in a study of a one-dimensional electron system is to treat all the electrons as indistinguishable particles.²⁵³ Equation (9) continues to be valid for a one-dimensional system provided the plus sign is selected, irrespective of the spins of the coalescing particles.
- ⁴⁶A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, *J. Chem. Phys.* **122**, 224322 (2005).
- ⁴⁷D. Bohm and D. Pines, *Phys. Rev.* **92**, 609 (1953).
- ⁴⁸N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. B* **70**, 235119 (2004).
- ⁴⁹R. P. Feynman and M. Cohen, *Phys. Rev.* **102**, 1189 (1956).
- ⁵⁰M. A. Lee, K. E. Schmidt, M. H. Kalos, and G. V. Chester, *Phys. Rev. Lett.* **46**, 728 (1981).
- ⁵¹Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **58**, 6800 (1998).
- ⁵²M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, *Phys. Rev. E* **68**, 046707 (2003).
- ⁵³P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- ⁵⁴M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. B* **91**, 115106 (2015).
- ⁵⁵M. Holzmann and S. Moroni, *Phys. Rev. B* **99**, 085121 (2019).
- ⁵⁶D. Pfau, J. S. Spencer, A. G. de G. Matthews, and W. M. C. Foulkes, “Ab-initio solution of the many-electron Schrödinger equation with deep neural networks,” (2019), [arXiv:1909.02487 \[physics.chem-ph\]](https://arxiv.org/abs/1909.02487).
- ⁵⁷C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719 (1988).
- ⁵⁸P. R. C. Kent, R. J. Needs, and G. Rajagopal, *Phys. Rev. B* **59**, 12344 (1999).
- ⁵⁹N. D. Drummond and R. J. Needs, *Phys. Rev. B* **72**, 085124 (2005).
- ⁶⁰D. Bressanini, G. Morosi, and M. Mella, *J. Chem. Phys.* **116**, 5345 (2002).
- ⁶¹D. M. Ceperley, *J. Stat. Phys.* **43**, 815 (1986).
- ⁶²M. P. Nightingale and V. Melik-Alaverdian, *Phys. Rev. Lett.* **87**, 043401 (2001).
- ⁶³J. Toulouse and C. J. Umrigar, *J. Chem. Phys.* **126**, 084102 (2007).
- ⁶⁴C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, *Phys. Rev. Lett.* **98**, 110201 (2007).
- ⁶⁵G. Ortiz, D. M. Ceperley, and R. M. Martin, *Phys. Rev. Lett.* **71**, 2777 (1993).
- ⁶⁶W. M. C. Foulkes, R. Q. Hood, and R. J. Needs, *Phys. Rev. B* **60**, 4558 (1999).
- ⁶⁷The approximation $f \approx \sum_{\alpha} w_{\alpha} \delta(\mathbf{R} - \mathbf{R}_{\alpha})$ only makes sense under an integral sign. At any instant the error in this approximation is proportional to $1/\sqrt{W}$, where W is the number of walkers. The error term is, by construction, zero on average.
- ⁶⁸C. J. Umrigar, M. P. Nightingale, and K. J. Runge, *J. Chem. Phys.* **99**, 2865 (1993).
- ⁶⁹A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè, *Phys. Rev. B* **93**, 241118 (2016).
- ⁷⁰N. Nemeč, *Phys. Rev. B* **81**, 035119 (2010).
- ⁷¹M. Boninsegni and S. Moroni, *Phys. Rev. E* **86**, 056712 (2012).
- ⁷²D. M. Ceperley and M. H. Kalos, in *Monte Carlo methods in statistical physics*, edited by K. Binder (Springer-Verlag, Heidelberg, 1979) 2nd ed., p. 145.
- ⁷³R. J. Hunt, M. Szyniszewski, G. I. Prayogo, R. Maezono, and N. D. Drummond, *Phys. Rev. B* **98**, 075122 (2018).
- ⁷⁴If the trial wave function does not describe the lowest-energy state that transforms as a 1D irreducible representation of the symmetry group (e.g., if one were trying to calculate the energy of the $2s$ state of a hydrogen atom with an approximate trial wave function whose nodal surface is at the wrong radius) then the required equilibration imaginary-time period can be much larger than the subsequent decorrelation period in the statistics-accumulation phase. In this case the equilibration imaginary-time scale is the time taken for the walker populations in high-energy nodal pockets to die out, which is given by the reciprocal of the difference of the pocket ground-state energy eigenvalues.
- ⁷⁵For a narrow-gap semiconductor, on the other hand, the sharpest features in the band structure provide the longest length scale. In the nearly-free-electron model the size of the region in which the bands \mathcal{E} deviate from free-electron behavior is $\Delta\mathcal{E}/(\partial\mathcal{E}/\partial k)_{\text{BZ edge}} \approx 2\pi a\Delta\mathcal{E}$, where a is the lattice constant of the primitive cell. Hence the longest length scale for narrow-gap semiconductors is $L \sim 1/(a\Delta\mathcal{E})$.
- ⁷⁶J. Vrbik and S. M. Rothstein, *Int. J. Quantum Chem.* **29**, 461 (1986).
- ⁷⁷“Quantum Monte Carlo and the CASINO program,” <https://vallico.net/casinoqmc/>.
- ⁷⁸P. López Ríos, P. Seth, N. D. Drummond, and R. J. Needs, *Phys. Rev. E* **86**, 036703 (2012).
- ⁷⁹T. M. Whitehead, M. H. Michael, and G. J. Conduit, *Phys. Rev. B* **94**, 035157 (2016).
- ⁸⁰E. A. Hylleraas and A. Ore, *Phys. Rev.* **71**, 493 (1947).
- ⁸¹K. Varga, J. Usukura, and Y. Suzuki, *Phys. Rev. Lett.* **80**, 1876 (1998).
- ⁸²M. Puchalski and A. Czarnecki, *Phys. Rev. Lett.* **101**, 183001 (2008).
- ⁸³T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- ⁸⁴H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfly, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhards-son, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eck-

- ert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, "MOLPRO, version 2019.2, a package of ab initio programs," (2019), see <https://www.molpro.net>.
- ⁸⁵C. Filippi and C. J. Umrigar, *J. Chem. Phys.* **105**, 213 (1996).
- ⁸⁶E. Giner, R. Assaraf, and J. Toulouse, *Mol. Phys.* **114**, 910 (2016).
- ⁸⁷M. C. Per and D. M. Cleland, *J. Chem. Phys.* **146**, 164101 (2017).
- ⁸⁸A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P.-F. Loos, *J. Chem. Phys.* **149**, 034108 (2018).
- ⁸⁹M. Dash, S. Moroni, A. Scemama, and C. Filippi, *J. Chem. Theory Comput.* **14**, 4176 (2018).
- ⁹⁰P. K. V. V. Nukala and P. R. C. Kent, *J. Chem. Phys.* **130**, 204105 (2009).
- ⁹¹B. K. Clark, M. A. Morales, J. McMinis, J. Kim, and G. E. Scuseria, *J. Chem. Phys.* **135**, 244105 (2011).
- ⁹²C. Filippi, R. Assaraf, and S. Moroni, *J. Chem. Phys.* **144**, 194105 (2016).
- ⁹³A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, *J. Comput. Chem.* **37**, 1866 (2016).
- ⁹⁴R. Assaraf, S. Moroni, and C. Filippi, *J. Chem. Theory Comput.* **13**, 5273 (2017).
- ⁹⁵G. L. Weerasinghe, P. López Ríos, and R. J. Needs, *Phys. Rev. E* **89**, 023304 (2014).
- ⁹⁶W. Dobrautz, S. D. Smart, and A. Alavi, *J. Chem. Phys.* **151**, 094104 (2019).
- ⁹⁷A. C. Hurley, J. E. Lennard-Jones, and J. A. Pople, *Proc. R. Soc. A* **220**, 446 (1953).
- ⁹⁸K.-i. Nakamura, *Prog. Theor. Phys.* **21**, 713 (1959).
- ⁹⁹A. J. Coleman, *Rev. Mod. Phys.* **35**, 668 (1963).
- ¹⁰⁰M. Bajdich, L. Mitas, L. K. Wagner, and K. E. Schmidt, *Phys. Rev. B* **77**, 115112 (2008).
- ¹⁰¹M. Casula and S. Sorella, *J. Chem. Phys.* **119**, 6500 (2003).
- ¹⁰²M. Casula, S. Yunoki, C. Attaccalite, and S. Sorella, *Comput. Phys. Commun.* **169**, 386 (2005).
- ¹⁰³V. A. Rassolov, *J. Chem. Phys.* **117**, 5978 (2002).
- ¹⁰⁴P. O. Bugnion, *Few-body insights into many-body physics*, Ph.D. thesis, University of Cambridge, Cambridge (2014).
- ¹⁰⁵J. J. Shepherd, G. Booth, A. Grüneis, and A. Alavi, *Phys. Rev. B* **85**, 081103 (2012).
- ¹⁰⁶R. Maezono, P. López Ríos, T. Ogawa, and R. J. Needs, *Phys. Rev. Lett.* **110**, 216407 (2013).
- ¹⁰⁷P. López Ríos, *Backflow and pairing wave function for quantum Monte Carlo methods*, Ph.D. thesis, University of Cambridge, Cambridge (2006).
- ¹⁰⁸P. López Ríos, A. Perali, R. J. Needs, and D. Neilson, *Phys. Rev. Lett.* **120**, 177701 (2018).
- ¹⁰⁹G. G. Spink, P. López Ríos, N. D. Drummond, and R. J. Needs, *Phys. Rev. B* **94**, 041410 (2016).
- ¹¹⁰J. H. Lloyd-Williams and B. Monserrat, *Phys. Rev. B* **92**, 184301 (2015).
- ¹¹¹N. D. Drummond, B. Monserrat, J. H. Lloyd-Williams, P. López Ríos, C. J. Pickard, and R. J. Needs, *Nat. Commun.* **6**, 7794 (2015).
- ¹¹²J. Trail, B. Monserrat, P. López Ríos, R. Maezono, and R. J. Needs, *Phys. Rev. B* **95**, 121108 (2017).
- ¹¹³G. Rajagopal, R. J. Needs, S. Kenny, W. M. C. Foulkes, and A. James, *Phys. Rev. Lett.* **73**, 1959 (1994).
- ¹¹⁴G. Rajagopal, R. J. Needs, A. James, S. D. Kenny, and W. M. C. Foulkes, *Phys. Rev. B* **51**, 10591 (1995).
- ¹¹⁵L. D. Landau, *Sov. Phys. JETP* **3**, 920 (1957).
- ¹¹⁶L. D. Landau, *Sov. Phys. JETP* **5**, 101 (1957).
- ¹¹⁷L. D. Landau, *Sov. Phys. JETP* **8**, 70 (1959).
- ¹¹⁸D. Ceperley, *Phys. Rev. B* **18**, 3126 (1978).
- ¹¹⁹C. Lin, F. H. Zong, and D. M. Ceperley, *Phys. Rev. E* **64**, 016702 (2001).
- ¹²⁰P. P. Ewald, *Ann. Phys.* **64**, 253 (1921).
- ¹²¹P. Gori-Giorgi and J. P. Perdew, *Phys. Rev. B* **66**, 165118 (2002).
- ¹²²G. Makov and M. C. Payne, *Phys. Rev. B* **51**, 4014 (1995).
- ¹²³L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, *Phys. Rev. B* **53**, 1814 (1996).
- ¹²⁴A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, and M.-Y. Chou, *Phys. Rev. B* **55**, R4851 (1997).
- ¹²⁵P. 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).
- ¹²⁶S. Chiesa, D. M. Ceperley, R. M. Martin, and M. Holzmann, *Phys. Rev. Lett.* **97**, 076404 (2006).
- ¹²⁷T. Gaskell, *Proc. Phys. Soc.* **77**, 1182 (1961).
- ¹²⁸N. D. Drummond, R. J. Needs, A. Sorouri, and W. M. C. Foulkes, *Phys. Rev. B* **78**, 125106 (2008).
- ¹²⁹M. Holzmann, R. C. Clay, M. A. Morales, N. M. Tubman, D. M. Ceperley, and C. Pierleoni, *Phys. Rev. B* **94**, 035126 (2016).
- ¹³⁰H. Kwee, S. Zhang, and H. Krakauer, *Phys. Rev. Lett.* **100**, 126404 (2008).
- ¹³¹M. Holzmann, B. Bernu, V. Olevano, R. M. Martin, and D. M. Ceperley, *Phys. Rev. B* **79**, 041308 (2009).
- ¹³²A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **71**, 066704 (2005).
- ¹³³J. R. Trail and R. J. Needs, *J. Chem. Phys.* **122**, 174109 (2005).
- ¹³⁴J. R. Trail and R. J. Needs, *J. Chem. Phys.* **122**, 014112 (2005).
- ¹³⁵J. R. Trail and R. J. Needs, *J. Chem. Phys.* **139**, 014101 (2013).
- ¹³⁶J. R. Trail and R. J. Needs, *J. Chem. Phys.* **142**, 064110 (2015).
- ¹³⁷J. R. Trail and R. J. Needs, *J. Chem. Phys.* **146**, 204107 (2017).
- ¹³⁸"Pseudopotential library," <https://pseudopotentiallibrary.org/>.
- ¹³⁹A. Annaberdiyev, G. Wang, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, *J. Chem. Phys.* **149**, 134108 (2018).
- ¹⁴⁰E. L. Shirley and R. M. Martin, *Phys. Rev. B* **47**, 15413 (1993).
- ¹⁴¹M. Casula, *Phys. Rev. B* **74**, 161102 (2006).
- ¹⁴²M. Casula, S. Moroni, S. Sorella, and C. Filippi, *J. Chem. Phys.* **132**, 154113 (2010).
- ¹⁴³N. D. Drummond, J. R. Trail, and R. J. Needs, *Phys. Rev. B* **94**, 165170 (2016).
- ¹⁴⁴A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, *J. Chem. Phys.* **151**, 134105 (2019).
- ¹⁴⁵N. S. Rytova, *Dokl. Akad. Nauk. SSSR* **163**, 1118 (1965).
- ¹⁴⁶L. V. Keldysh, *J. Exp. Theor. Phys.* **29**, 658 (1979).
- ¹⁴⁷Y. Yang, V. Gorelov, C. Pierleoni, D. M. Ceperley, and M. Holzmann, *Phys. Rev. B* **101**, 085115 (2020).
- ¹⁴⁸M. J. Gillan, M. D. Towler, and D. Alfè, "Petascale computing opens new vistas for quantum Monte Carlo," Psi-k Scientific Highlight of the Month, February 2011, <https://psi-k.net/download/highlights/Highlight.103.pdf>.
- ¹⁴⁹A. J. Williamson, R. Q. Hood, and J. C. Grossman, *Phys. Rev. Lett.* **87**, 246406 (2001).
- ¹⁵⁰D. Alfè and M. J. Gillan, *Phys. Rev. B* **70**, 161101 (2004).
- ¹⁵¹D. Alfè and M. J. Gillan, *J. Phys. Condens. Mater.* **16**, L305 (2004).
- ¹⁵²D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).
- ¹⁵³For metals the longest length scale grows with system size; however, shorter length scales make the largest energy contributions.
- ¹⁵⁴E. Mostaani, M. Szyniszewski, C. H. Price, R. Maezono, M. Danovich, R. J. Hunt, N. D. Drummond, and V. I. Fal'ko, *Phys. Rev. B* **96**, 075431 (2017).
- ¹⁵⁵D. Bressanini, M. Mella, and G. Morosi, *Phys. Rev. A* **57**, 4956 (1998).
- ¹⁵⁶M. Y. J. Tan, N. D. Drummond, and R. J. Needs, *Phys. Rev. B* **71**, 033303 (2005).
- ¹⁵⁷R. M. Lee, N. D. Drummond, and R. J. Needs, *Phys. Rev. B* **79**, 125308 (2009).
- ¹⁵⁸O. Witham, R. J. Hunt, and N. D. Drummond, *Phys. Rev. B* **97**, 075424 (2018).
- ¹⁵⁹T. Tsuchiya and S. Katayama, *Physica B Condens. Matter* **249-251**, 612 (1998).
- ¹⁶⁰T. Tsuchiya, *J. Lumin.* **87-89**, 509 (2000).
- ¹⁶¹T. Tsuchiya, *Phys. Status Solidi (c)* **1**, 603 (2004).
- ¹⁶²T. Tsuchiya and S. Katayama, *Solid State Electron.* **42**, 1523 (1998).
- ¹⁶³T. Tsuchiya, *Physica E Low Dimens. Syst. Nanostruct.* **7**, 470 (2000).
- ¹⁶⁴D. M. Thomas, R. J. Hunt, N. D. Drummond, and M. Hayne, *Phys. Rev. B* **99**, 115306 (2019).
- ¹⁶⁵B. Ganchev, N. Drummond, I. Aleiner, and V. Fal'ko, *Phys. Rev. Lett.* **114**, 107401 (2015).
- ¹⁶⁶M. Z. Mayers, T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, *Phys. Rev. B* **92**, 161404 (2015).
- ¹⁶⁷M. Szyniszewski, E. Mostaani, N. D. Drummond, and V. I. Fal'ko, *Phys. Rev. B* **95**, 081301 (2017).

- ¹⁶⁸M. Barbone, A. R. P. Montblanch, D. M. Kara, C. Palacios-Berraquero, A. R. Cadore, D. De Fazio, B. Pingault, E. Mostaani, H. Li, B. Chen, K. Watanabe, T. Taniguchi, S. Tongay, G. Wang, A. C. Ferrari, and M. Atatüre, *Nat. Commun.* **9**, 3721 (2018).
- ¹⁶⁹M. Danovich, D. A. Ruiz-Tijerina, R. J. Hunt, M. Szyniszewski, N. D. Drummond, and V. I. Fal'ko, *Phys. Rev. B* **97**, 195452 (2018).
- ¹⁷⁰F. Vialla, M. Danovich, D. A. Ruiz-Tijerina, M. Massicotte, P. Schmidt, T. Taniguchi, K. Watanabe, R. J. Hunt, M. Szyniszewski, N. D. Drummond, T. G. Pedersen, V. I. Fal'ko, and F. H. L. Koppens, *2D Mater.* **6**, 035032 (2019).
- ¹⁷¹M. Ruggeri, P. López Ríos, and A. Alavi, *Phys. Rev. B* **98**, 161105 (2018).
- ¹⁷²J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
- ¹⁷³P. Bhattacharai, A. Patra, C. Shahi, and J. P. Perdew, *Phys. Rev. B* **97**, 195128 (2018).
- ¹⁷⁴S. De Palo, F. Rapisarda, and G. Senatore, *Phys. Rev. Lett.* **88**, 206401 (2002).
- ¹⁷⁵G. W. Burg, N. Prasad, K. Kim, T. Taniguchi, K. Watanabe, A. H. MacDonald, L. F. Register, and E. Tutuc, *Phys. Rev. Lett.* **120**, 177702 (2018).
- ¹⁷⁶A. Perali, F. Palestini, P. Pieri, G. C. Strinati, J. T. Stewart, J. P. Gaebler, T. E. Drake, and D. S. Jin, *Phys. Rev. Lett.* **106**, 060402 (2011).
- ¹⁷⁷H. Hosono and K. Kuroki, *Physica C* **514**, 399 (2015).
- ¹⁷⁸V. Huard, R. T. Cox, K. Saminadayar, A. Arnoult, and S. Tatarenko, *Phys. Rev. Lett.* **84**, 187 (2000).
- ¹⁷⁹G. Yusa, H. Shtrikman, and I. Bar-Joseph, *Phys. Rev. B* **62**, 15390 (2000).
- ¹⁸⁰I. Bar-Joseph, *Chem. Phys.* **318**, 99 (2005).
- ¹⁸¹M. Yamaguchi, S. Nomura, H. Tamura, and T. Akazaki, *Phys. Rev. B* **87**, 081310 (2013).
- ¹⁸²N. D. Drummond and R. J. Needs, *Phys. Rev. B* **79**, 085414 (2009).
- ¹⁸³J. L. Smith and P. J. Stiles, *Phys. Rev. Lett.* **29**, 102 (1972).
- ¹⁸⁴V. M. Pudalov, M. E. Gershenson, H. Kojima, N. Butch, E. M. Dizhur, G. Brunthaler, A. Prinz, and G. Bauer, *Phys. Rev. Lett.* **88**, 196404 (2002).
- ¹⁸⁵Y.-W. Tan, J. Zhu, H. L. Stormer, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, *Phys. Rev. Lett.* **94**, 016405 (2005).
- ¹⁸⁶M. Padmanabhan, T. Gokmen, N. C. Bishop, and M. Shayegan, *Phys. Rev. Lett.* **101**, 026402 (2008).
- ¹⁸⁷G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, 2005).
- ¹⁸⁸Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **50**, 1684 (1994).
- ¹⁸⁹Y. Zhang and S. Das Sarma, *Phys. Rev. Lett.* **95**, 256603 (2005).
- ¹⁹⁰N. D. Drummond and R. J. Needs, *Phys. Rev. B* **80**, 245104 (2009).
- ¹⁹¹N. D. Drummond and R. J. Needs, *Phys. Rev. B* **87**, 045131 (2013).
- ¹⁹²E. Boroński and R. M. Nieminen, *Phys. Rev. B* **34**, 3820 (1986).
- ¹⁹³M. J. Puska and R. M. Nieminen, *Rev. Mod. Phys.* **66**, 841 (1994).
- ¹⁹⁴R. Krause-Rehberg and H. S. Leipner, *Positron Annihilation in Semiconductors*, 1st ed., Vol. 127 (Springer-Verlag, Berlin Heidelberg, 1999).
- ¹⁹⁵Z. Major, S. B. Dugdale, R. J. Watts, G. Santi, M. A. Alam, S. M. Hayden, J. A. Duffy, J. W. Taylor, T. Jarlborg, E. Bruno, D. Benea, and H. Ebert, *Phys. Rev. Lett.* **92**, 107003 (2004).
- ¹⁹⁶N. D. Drummond, P. López Ríos, R. J. Needs, and C. J. Pickard, *Phys. Rev. Lett.* **107**, 207402 (2011).
- ¹⁹⁷J. Arponen and E. Pajanne, *Ann. Phys.* **121**, 343 (1979).
- ¹⁹⁸J. Kuriplach and B. Barbiellini, *Phys. Rev. B* **89**, 155111 (2014).
- ¹⁹⁹J. A. Weber, A. Bauer, P. Böni, H. Ceeh, S. B. Dugdale, D. Ernsting, W. Kreuzpaintner, M. Leitner, C. Pfeleiderer, and C. Hugenschmidt, *Phys. Rev. Lett.* **115**, 206404 (2015).
- ²⁰⁰M. Szyniszewski, E. Mostaani, A. Knothe, A. C. Ferrari, V. I. Fal'ko, and N. D. Drummond, (2020), unpublished.
- ²⁰¹C.-R. Hsing, C. Cheng, J.-P. Chou, C.-M. Chang, and C.-M. Wei, *New J. Phys.* **16**, 113015 (2014).
- ²⁰²L. Spanu, S. Sorella, and G. Galli, *Phys. Rev. Lett.* **103**, 196401 (2009).
- ²⁰³E. Mostaani, N. D. Drummond, and V. I. Fal'ko, *Phys. Rev. Lett.* **115**, 115501 (2015).
- ²⁰⁴A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1724 (2018).
- ²⁰⁵T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfè, A. Michaelides, and A. Grüneis, *J. Chem. Phys.* **146**, 204108 (2017).
- ²⁰⁶Y. S. Al-Hamdani, M. Ma, D. Alfè, O. A. von Lilienfeld, and A. Michaelides, *J. Chem. Phys.* **142**, 181101 (2015).
- ²⁰⁷J. Ma, A. Michaelides, D. Alfè, L. Schimka, G. Kresse, and E. Wang, *Phys. Rev. B* **84**, 033402 (2011).
- ²⁰⁸J. Chen, A. Zen, J. G. Brandenburg, D. Alfè, and A. Michaelides, *Phys. Rev. B* **94**, 220102 (2016).
- ²⁰⁹Y. Akahama, Y. Mizuki, S. Nakano, N. Hirao, and Y. Ohishi, *J. Phys. Conf. Ser.* **950**, 042060 (2017).
- ²¹⁰H. Kawamura, Y. Akahama, S. Umemoto, K. Takemura, Y. Ohishi, and O. Shimomura, *J. Phys. Condens. Mater.* **14**, 10407 (2002).
- ²¹¹A. F. Goncharov, R. J. Hemley, and H.-k. Mao, *J. Chem. Phys.* **134**, 174501 (2011).
- ²¹²M. I. Eremets and I. A. Troyan, *Nat. Materials* **10**, 927 (2011).
- ²¹³C. J. Pickard and R. J. Needs, *Nat. Phys.* **3**, 473 (2007).
- ²¹⁴C. J. Pickard, M. Martínez-Canales, and R. J. Needs, *Phys. Rev. B* **85**, 214114 (2012).
- ²¹⁵C. J. Pickard, M. Martínez-Canales, and R. J. Needs, *Phys. Rev. B* **86**, 059902 (2012).
- ²¹⁶C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- ²¹⁷R. J. Needs and C. J. Pickard, *APL Mater.* **4**, 053210 (2016).
- ²¹⁸B. Monserrat, R. J. Needs, E. Gregoryanz, and C. J. Pickard, *Phys. Rev. B* **94**, 134101 (2016).
- ²¹⁹B. Monserrat, N. D. Drummond, and R. J. Needs, *Phys. Rev. B* **87**, 144302 (2013).
- ²²⁰P. Dalladay-Simpson, R. Howie, and E. Gregoryanz, *Nature* **529**, 63 (2016).
- ²²¹B. Monserrat, N. D. Drummond, P. Dalladay-Simpson, R. T. Howie, P. López Ríos, E. Gregoryanz, C. J. Pickard, and R. J. Needs, *Phys. Rev. Lett.* **120**, 255701 (2018).
- ²²²S. Azadi and G. J. Ackland, *Phys. Chem. Chem. Phys.* **19**, 21829 (2017).
- ²²³S. Azadi, W. M. C. Foulkes, and T. D. Kühne, *New J. Phys.* **15**, 113005 (2013).
- ²²⁴S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, *Phys. Rev. Lett.* **112**, 165501 (2014).
- ²²⁵S. Azadi and T. D. Kühne, *Phys. Rev. B* **100**, 155103 (2019).
- ²²⁶A. F. Goncharov, J. S. Tse, H. Wang, J. Yang, V. V. Struzhkin, R. T. Howie, and E. Gregoryanz, *Phys. Rev. B* **87**, 024101 (2013).
- ²²⁷S. Azadi, N. D. Drummond, and W. M. C. Foulkes, *Phys. Rev. B* **95**, 035142 (2017).
- ²²⁸M. I. Eremets, A. P. Drozdov, P. P. Kong, and H. Wang, *Nat. Phys.* **15**, 1246 (2019).
- ²²⁹G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **135**, 084104 (2011).
- ²³⁰D. Cleland, G. H. Booth, C. Overy, and A. Alavi, *J. Chem. Theory Comput.* **8**, 4138 (2012).
- ²³¹G. Li Manni, S. D. Smart, and A. Alavi, *J. Chem. Theory Comput.* **12**, 1245 (2016).
- ²³²L. Veis, A. Antalík, O. Legeza, A. Alavi, and J. Pittner, *J. Chem. Theory Comput.* **14**, 2439 (2018).
- ²³³V. A. Neufeld and A. J. W. Thom, *J. Chem. Phys.* **147**, 194105 (2017).
- ²³⁴M. Dash, J. Feldt, S. Moroni, A. Scemama, and C. Filippi, *J. Chem. Theory Comput.* **15**, 4896 (2019).
- ²³⁵H. Luo and A. Alavi, *J. Chem. Theory Comput.* **14**, 1403 (2018).
- ²³⁶A. J. Cohen, H. Luo, K. Guther, W. Dobrautz, D. P. Tew, and A. Alavi, *J. Chem. Phys.* **151**, 061101 (2019).
- ²³⁷W. Dobrautz, H. Luo, and A. Alavi, *Phys. Rev. B* **99**, 075119 (2019).
- ²³⁸J. H. Lloyd-Williams, R. J. Needs, and G. J. Conduit, *Phys. Rev. B* **92**, 075106 (2015).
- ²³⁹C. A. Melton, M. Zhu, S. Guo, A. Ambrosetti, F. Pederiva, and L. Mitás, *Phys. Rev. A* **93**, 042502 (2016).
- ²⁴⁰E. Mostaani, B. Monserrat, N. D. Drummond, and C. J. Lambert, *Phys. Chem. Chem. Phys.* **18**, 14810 (2016).
- ²⁴¹Y. Y. F. Liu, B. Andrews, and G. J. Conduit, *J. Chem. Phys.* **150**, 034104 (2019).
- ²⁴²A. Badinski, P. D. Haynes, J. R. Trail, and R. J. Needs, *J. Phys.: Condens. Matter* **22**, 074202 (2010).
- ²⁴³A. N. Badinski, *Forces in quantum Monte Carlo*, Ph.D. thesis, University of Cambridge, Cambridge (2008).
- ²⁴⁴R. Assaraf and M. Caffarel, *J. Chem. Phys.* **119**, 10536 (2003).
- ²⁴⁵A. Badinski, J. R. Trail, and R. J. Needs, *J. Chem. Phys.* **129**, 224101 (2008).

²⁴⁶A. Badinski and R. J. Needs, *Phys. Rev. E* **76**, 036707 (2007).

²⁴⁷A. Badinski and R. J. Needs, *Phys. Rev. B* **78**, 035134 (2008).

²⁴⁸S. Chiesa, D. M. Ceperley, and S. Zhang, *Phys. Rev. Lett.* **94**, 036404 (2005).

²⁴⁹C. Attacalite and S. Sorella, *Phys. Rev. Lett.* **100**, 114501 (2008).

²⁵⁰P. López Ríos and G. J. Conduit, *Phys. Rev. E* **99**, 063312 (2019).

²⁵¹B. Efron and R. Tibshirani, *Stat. Sci.* **1**, 54 (1986).

²⁵²See <https://doi.org/10.17863/CAM.50424> for the raw data used in the figures and tables in this paper.

²⁵³R. M. Lee and N. D. Drummond, *Phys. Rev. B* **83**, 245114 (2011).