Equation of state of atomic solid hydrogen by stochastic many-body wave function methods

Sam Azadi,1 George H. Booth,1 and Thomas D. Kühne2
1) Department of Physics, King’s College London, Strand, WC2R 2LS London, United Kingdom
2) Department of Chemistry, Paderborn Center for Parallel Computing, Paderborn University, 33098 Paderborn, Germany

(Dated: 3 September 2020)

We report a numerical study of the equation of state of crystalline body-centered-cubic (BCC) hydrogen, tackled with a variety of complementary many-body wave function methods. These include continuum stochastic techniques of fixed-node diffusion and variational quantum Monte Carlo, and the Hilbert space stochastic method of full configuration interaction quantum Monte Carlo. In addition, periodic coupled-cluster methods were also employed. Each of these methods is underpinned with different strengths and assumptions, but their combination in order to perform reliable extrapolation to complete basis set and supercell size limits gives confidence in the final results. The methods were found to be in good agreement for equilibrium cell volumes for the system in this phase, with a lattice parameter of 3.31 Bohr.

I. INTRODUCTION

A stochastic description of quantum mechanics has significant advantages in the understanding of quantum systems, especially when a large number of degrees of freedom are involved. The main advantage of this approach relies on the exploitation of well-established mathematical bounds derived from probability theory and stochastic processes to control the convergence of these properties. In this picture, quantum particles move along stochastic trajectories, and expectation values can be formulated as ensemble averages over the space of these trajectories. In practice, we exploit the similarity between the Schrödinger equation in imaginary time, which is a linear, parabolic partial differential equation, and the diffusion equation. The efficiency of Monte Carlo approaches relies on the use of random numbers to sample the N-dimensional phase-space space of configurations, where N is the number of variables.

Quantum Monte Carlo (QMC) approaches to large, ab initio systems with realistic many-body Hamiltonians have provided some of the most accurate and reliable descriptions of both Fermionic and Bosonic quantum matter. The most established QMC techniques include variational Monte Carlo (VMC), and diffusion Monte Carlo (DMC), and have been successfully applied to a variety of realistic quantum materials, including chemical systems, and solids. The key advantages of these QMC methods is their efficient large-scale parallelization, low scaling with system size, and predictive accuracy. Diffusion Monte Carlo is at present the most commonly used QMC technique for high-precision wave function-based predictions of material properties. DMC works with an ensemble of (almost) independent ‘walkers’ or ‘configurations’, whose stochastically realised dynamic ensures that they explore the Fermionic configuration space of the system. Beginning with a trial wave function, samples of the ground state wave function of the system are projected onto by applying the imaginary time operator, exp(−τH). The dominant open problem preventing the exact numerical calculation of many-electron systems by DMC is the Fermionic sign problem which arises from the antisymmetric nature of many-body wave function with respect to electron exchange. A widespread practical solution for this problem is the so-called fixed-node (FN) approximation. In this, a many-electron nodal surface is defined as the one coinciding with that given by a trial many-electron wave function. This trial state is a function of 3N variables and the trial nodal surface is the (3N − 1)-dimensional hypersurface on which the function is zero, and across which the wave function amplitude changes sign. The approximation ensures that the hypersurface of the sampled wave function in DMC coincides with that of the trial wave function. This FN algorithm gives the lowest-energy many-electron state with the given nodal surface, which ensures that it is a variational approximation. In principle, this approximation would be exact if the applied nodal boundaries coexisted with the exact nodal surface of the many-Fermionic wave function. In practice however, the errors in FN DMC energies are usually about 5% of the correlation energy for commonly used trial states.

An alternative approach tackle the Fermionic sign problem is to allow the exact nodal surface to be an emergent property of the underlying algorithm. Such a simulation would therefore not require initial information of a trial nodal surface. It was demonstrated that such property can be obtained in full configuration-interaction quantum Monte Carlo (FCIQMC). This technique can be applied to Fermionic systems after projection into a discrete basis set familiar to conventional quantum chemistry approaches. It was demonstrated that this method can converge to capture the complete correlation energy, or full configuration-interaction (FCI) solutions.
tion, for the given basis set. There are a number of similarities and differences between the DMC and FCIQMC approaches to stochastic realisation of quantum systems. Both techniques run a long-time integration of the imaginary time Schrödinger equation. However, while DMC works in continuous real space, FCIQMC spans the Hilbert space of Slater determinants. In DMC walkers follow the diffusion equation, while in FCIQMC the propagation step is based on a population dynamics of creation and death processes. In DMC and FCIQMC the wave function is rendered by walkers representing a specific configuration, which enables the methods to stochastically sample the wave function without storing the exponential number of amplitudes in the space. The key step of the FCIQMC algorithm which allows the nodal structure of the wave function to emerge, is walker annihilation. Since each walker has a defined sign (or phase for complex-valued wave functions), if two configurations with opposite signs simultaneously occupy the same determinant, both walkers are deleted from the simulation.

The walker annihilation mechanism has also been explored in DMC and Green’s function Monte Carlo\textsuperscript{22,23}.

Solving realistic many-body Hamiltonians is the main challenge in condensed matter physics and quantum chemistry. Traditional quantum chemical methods, including coupled-cluster (CC) theory\textsuperscript{24–29} and configuration interaction (CI)\textsuperscript{30}, have been developed for solving the Schrödinger equation, primarily for chemical systems described by a given one-particle basis set. These approaches truncate the wave function to a specific ansatz, which can be relaxed to define a systematic hierarchy of approximation to exactness. Due to the computational complexity of these quantum chemical methods, applying this systematic hierarchy of methods for extended systems and solids is in its early stages of research\textsuperscript{31–37}. Application of FCIQMC to realistic solids also demonstrated a promising route for providing reference ground state many-electron energies with which to benchmark quantum-chemical techniques including the coupled-cluster (CC) ansatz. On the other hand, many alternate approaches for dealing with the high dimensionality of real extended systems have been developed, including local truncations, single-particle Green function methods, other Monte Carlo algorithms, and embedding techniques, all of which also benefit from comparison to higher accuracy approaches rather than experiment\textsuperscript{38–41}.

In this work we compute the equation of state of atomic solid hydrogen in the body centered cubic (BCC) structure using CC, FCIQMC and DMC techniques. The study of BCC phase of solid hydrogen is critical in the understanding of the origin of metallic magnetism for this system. Many theoretical and numerical investigations have concentrated on this atomic phase of solid hydrogen\textsuperscript{42–50}. Despite the fact that at low densities the atomic phase is not the most stable structure of solid hydrogen, investigations into this simple but realistic system can provide qualitative insight into metal-insulator transitions, and also supply a reference for commonly used mean-field approximations such as local density approximation. In this work we consider the ground state energy of the paramagnetic BCC phase of solid hydrogen within density regimes \(1.1 < r_g < 2.4\), in which the paramagnetic phase is more stable than antiferromagnetic and ferromagnetic states. We compare the FCIQMC results with CCSD which are obtained using the same basis set and system size. We then extend the investigation to include the use of DMC to allow extensions to larger system sizes and without a substantial basis set dependence. Since DMC results are not constrained by basis size, the DMC and FCIQMC values are not directly compared, but instead provide a complementary description which can be combined to provide accurate estimates of the correlation energy in the thermodynamic and complete basis set limits.

II. COMPUTATIONAL METHODS

We briefly describe the full configuration interaction Monte Carlo and Diffusion Monte Carlo approaches. Both methods have been presented in more detail in previous work\textsuperscript{5,17,61}, but less often alongside each other.

A. FCIQMC Method

In FCIQMC, we first require the choice of a basis of \(2M\) one-electron spin-orbitals \(\phi_i\), from which the space of all possible \(N\)-electron determinants from the set of spin-orbitals can be constructed, as

\[
D_i = D_{i_1 \ldots i_N} = \frac{1}{\sqrt{N!}} |\phi_{i_1} \ldots \phi_{i_N}| = \frac{1}{\sqrt{N!}} \begin{pmatrix} \phi_{i_1}(r_1) & \cdots & \phi_{i_N}(r_N) \\ \phi_{i_2}(r_1) & \cdots & \phi_{i_2}(r_N) \\ \vdots & \ddots & \vdots \\ \phi_{i_N}(r_1) & \cdots & \phi_{i_N}(r_N) \end{pmatrix}
\]

With this set of determinants as an \(N\)-body basis, the wave function can be expanded as:

\[
|\Psi\rangle = \sum c_i |D_i\rangle,
\]

with the optimal coefficients \(c_i\) the ones which variationally minimise

\[
E(c_i) = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}.
\]

These optimal coefficients are found from the solution of the matrix eigenvalue problem

\[
\sum_j H_{ij} c_j = E_0^{FCI} c_i,
\]

where \(H_{ij} = \langle D_i | \hat{H} | D_j \rangle\). This FCI approach captures all possible correlation energy within the basis of orbitals,
and is therefore in principle systematically improvable as the basis is enlarged, albeit at exponential cost. Due to the basis incompleteness error, the calculated total energies are therefore often higher than FN-DMC energies, but cancellation of errors in general is more reliable in appropriately optimized quantum chemical basis sets. However, the total number of determinants exponentially increases with system and basis size, hence the appeal of a stochastic approach which can deal with such high-dimensional spaces.

The FCIQMC algorithm converges to the FCI (lowest energy) eigenvector of the FCI matrix eigenproblem of Eq. 4, via the solution of the imaginary-time Schrödinger equation. The master equation governing the stochastic walker dynamics is given by

\[
\frac{d\tilde{c}_i(t)}{dt} = -\sum_j [\hat{H}_{ij} - (E_{HF} + S)\delta_{ij}]\tilde{c}_j(t) = -\sum_j (\hat{K}_{ij} - S\delta_{ij})\tilde{c}_j(t).
\]

The amplitude of each determinant is then coarse-grained by a variable resolution, with the vast majority of amplitudes then represented at any single FCIQMC iteration by a zero amplitude. By representing the wave function at any single iteration just by the walkers (necessarily with non-zero weight) rather than the amplitudes, there is then a significant compression of the wave function information, and the storage requirements to describe this snapshot of the wave function is decoupled from the size of the underlying Hilbert space. The success of the algorithm therefore relies on the ability to perform this compression whilst maintaining a faithful representation of the state, hence exploiting the sparsity inherent in its representation. The (signed) population of walkers then sample the configuration space \(q_t\), updating each iteration depending on a choice of time step \(\Delta t\), through a series of ‘spawning’ steps on connected configurations, and ‘death’ steps which generally reduce the local population of walkers. These spawning and death steps stochastically update the amplitude on a determinant via

\[
q^{(n+1)}_i = q^{(n)}_i - \Delta t \sum_j (\hat{K}_{ij} - S\delta_{ij})q^{(n)}_j, \tag{6}
\]

where \(q_i\) is the random variable denoting the instantaneous walker population on configuration \(|D_i\rangle\). These walker dynamics are not dissimilar to a first-order approximation to those of DMC, where the propagator is \(G = I - (H - SI)\Delta t\) instead of \(G = \exp(-(H - SI)\Delta t)\), but without importance sampling or the fixed-node approximation. As long as \(\Delta t \leq 2/(E_{\text{max}} - S)\), FCIQMC will formally yield the exact ground state without any time-step error. The energy can be extracted as

\[
E(t) = \frac{\langle \exp(-t\hat{H})D_0 | \hat{H} | D_0 \rangle}{\langle \exp(-t\hat{H})D_0 | D_0 \rangle} \tag{7}
\]

\[
= E_{HF} + \sum_{j \neq 0} c_j(t) \langle D_j | \hat{H} | D_0 \rangle/c_0(t) \tag{8}
\]

\[
= E_{HF} + \frac{\langle \sum_{j \neq 0} q_j(t) H_{j0} \rangle}{\langle q_0(t) \rangle}, \tag{9}
\]

where \(|D_0\rangle\) represents a trial state.\(^{51,52}\)

In keeping with all general Fermionic QMC techniques, there is also a sign problem which afflicts FCIQMC.\(^{53}\) In order to ensure that the annihilation is sufficient in order to allow the exact nodal structure of the FCI wave function to emerge, the number of walkers must be large enough, and while this number is generally much smaller than the dimensionality of the entire Hilbert space, it still grows exponentially with system size. To reduce the number of walkers required for high accuracy, the FCIQMC algorithm can be improved with the ‘initiator’ approximation.\(^{18,20}\) Initiator FCIQMC is a systematically improvable approximation, whereby the determinants are divided into two classes, labelled ‘initiator’ and ‘non-initiator’. Initiator determinants are allowed to create new walkers on unoccupied determinants. However non-initiator determinants are only allowed to create new walkers on already occupied determinants. The label of initiator or non-initiator is chosen depending on the current population of walkers on any determinant, with the threshold \(n_{\text{add}}\) used to determine this label. This dynamic adaptation helps the walker population to stabilize a sign structure at far lower walker numbers, at the cost of introducing a small systematic error. This error can be systematically reduced as the number of walkers is increased which enlarges the set of initiators. Other adaptations in recent years have also improved the accuracy and scope of the method, including the computation of excited states,\(^{54,55}\) unbiased molecular properties,\(^{56-58}\) and its use as a solver within an active space framework.\(^{59,60}\)

**B. Diffusion Monte Carlo**

The diffusion quantum Monte Carlo method is a stochastic technique for many-electron systems with a much longer history than FCIQMC. Any solution of the time-dependent Schrödinger equation can be expanded in the eigenfunctions of the Hamiltonian, \(\Phi_n(\mathbf{R})\), as

\[
\Psi(\mathbf{R}, \tau) = \sum_n c_n \Phi_n(\mathbf{R}) \exp(-E_n \tau) \tag{10}
\]

where \(E_n\) is the eigenvalue corresponding to \(\Phi_n\), and \(\tau = it\). One can find that \(\Psi(\mathbf{R}, \tau \to \infty) \approx c_0 \Phi_0(\mathbf{R}) \exp(-E_0 \tau)\) which is proportional to the ground state wave function. In principle therefore, the Schrödinger equation can be solved by propagating an arbitrary wave function in imaginary time for long enough.
This propagation can be seen in the differential form of the imaginary-time Schrödinger equation, as

$$\frac{-1}{2} \nabla^2 \Psi(R, \tau) + [V(R) - E_T] \Psi(R, \tau) = -\frac{\partial}{\partial \tau} \Psi(R, \tau)$$ (11)

where \( \nabla^2 = \sum_i \nabla_i^2 \) acts over all coordinates in vector \( R \), and \( E_T \) is a constant energy offset. This equation is equivalent to a diffusion equation, in which \( \Psi(R) \) represents the density of particles at point \( R \). The particles diffuse with diffusion coefficient \( D = 1/2 \), and are absorbed with rate \( V(R) - E_T \). Assuming \( \Psi(R, \tau) \) is a probability density, we distribute an initial set of walkers with probability density given by \( \Psi(R, \tau) \). The walkers then diffuse and can be removed or created according. This can be simulated via a stochastic process, whereby in the limit \( \tau \to \infty \), the walkers would be distributed according to the ground state wave function.

For a Fermionic system, a wave function \( \Psi(R) \) must have both positive and negative regions to be antisymmetric with respect to particle exchange. Hence, it can not be used as a probability density. This problem can be overcome by using a guiding function \( f(R, \tau) = \Psi_T(R) \Psi(R, \tau) \). Provided \( \Psi_T(R) \) and \( \Psi(R, \tau) \) have the same nodal surface, \( f(R, \tau) \) has the same sign over all configuration space, and can be interpreted as a probability distribution function. By multiplying both sides of the imaginary-time Schrödinger equation by \( \Psi_T \), we obtain

$$-\frac{\partial f}{\partial \tau} = -\frac{1}{2} \nabla^2 f + \nabla \cdot [f \mathbf{v}] + [E_L - E_T] f$$ (12)

where \( \mathbf{v}(R) = \nabla \Psi_T(R)/\Psi_T(R) \) and \( E_L(R) = \hat{H} \Psi_T(R)/\Psi_T(R) \). If we consider \( f \) as a probability distribution, this equation is the Fokker-Planck equation describing the diffusion of non-interacting classical particles, with an imposed drift velocity \( \mathbf{v}(R, \tau) \) and absorption coefficient \( [E_L(R) - E_T] \). We can therefore distribute a set of particles according to an initial distribution \( f(R, 0) = |\Psi_T(R)|^2 \), and let them evolve according to the Fokker-Planck equation. In the limit of \( \tau \to \infty \), the walkers will be distributed according to the minimal energy wave function with the same nodal surface as the trial wave function. The quality of the nodal surface of the trial wave function therefore determines an uncontrolled error in DMC. In practice, the trial wave function is optimised by variational Monte Carlo (VMC) before being used in DMC. Generally, lower VMC energy imply a better nodal surface. \(^{62,63}\)

We define a drift-diffusion operator \( \hat{D} f = -\frac{1}{2} \nabla^2 f + \nabla \cdot [f \mathbf{v}] \) and an absorption operator \( \hat{A} f = [E_L - E_T] f \). These two operators can be combined into a Hamiltonian \( \hat{H} = \hat{D} + \hat{A} \), so that equation 12 can be expressed as \( \frac{\partial f}{\partial \tau} = \hat{H} f \), and in an integral form, as

$$f(R, \tau) = \int G(R, R', \tau) f(R', \tau) dR'$$ (13)

where the Green’s function \( G(R, R', \tau) = \langle R' | \exp(-\tau \hat{H}) | R \rangle \) indicates the probability of a configuration moving from \( R \) to \( R' \) in time \( \tau \). Green’s function Monte Carlo (GFMC) developed by Kalos, allows the sampling of these Green’s functions. \(^{64}\) The Trotter-Suzuki formula \( \exp(-\tau \hat{H}) = \exp(-\tau \hat{A}/2) \exp(-\tau \hat{D}) \exp(-\tau \hat{A}/2) + O(\tau^3) \) and the fact that the absorption term corresponding to the local energy of the trial function can be written as \( \exp(-\tau \hat{A}/2) | R \rangle = \exp(-\tau(E_L(R) - E_T)/2) | R \rangle \), the Green’s function can be written as

$$G(R, R', \tau) \simeq \exp(-\tau(E_L(R) - E_T)/2)$$

The matrix element \( \langle R | \exp(-\tau \hat{D}) | R' \rangle \) is the Green’s function for walkers diffusing with drift current \( f \mathbf{v} \). For small \( \tau \) these can be approximated as

$$\langle R | e^{-\tau \hat{D}} | R' \rangle = \frac{1}{(2\pi \tau)^{3N/2}} \exp \left[ -\frac{(R - R' - \tau \mathbf{v})^2}{2\tau} \right]$$ (15)

and the Green’s function

$$G(R, R', \tau) = \frac{1}{(2\pi \tau)^{3N/2}} \exp \left[ -\frac{(R - R' - \tau \mathbf{v})^2}{2\tau} \right] \times \exp(\tau/2(E_L(R) + E_L(R'))) - 2E_T$$ (16)

\( G(R, R', \tau) \) describes the probability of the fermionic system propagating from configuration \( R \) to \( R' \) in time \( \tau \). Further details of the implementation of the DMC algorithm are discussed in ref. 62.

In practice, the DMC algorithm is initialized with a set of configurations \( R_i \) distributed according to \( f(R, 0) = |\Psi_T(R)|^2 \), where \( \Psi_T(R) \) is obtained from a prior VMC calculation. To propagate the configurations through time interval \( \tau \), we apply the drift-diffusion step, selecting the move \( R \to R' \) to satisfy the Green’s function of Eq. 16. The configuration can branch or die based on the \( \exp(-\tau/2(E_L(R) + E_L(R')) - 2E_T) \) factor, where \( E_T \) is varied to control the population of configurations. The initial few time-steps serve to eliminate the contribution of the excited states from the wave function \( \Psi(R) \). This equilibration phase lasts until the walkers have settled into their equilibrium distribution, after which the statistics of the desired expectation values can be accumulated. The separation of the Green’s function into the drift-diffusion and branching term is only valid in the limit of small time-step \( \tau \). The time-step bias can be controlled by running two separate DMC simulations with different \( \tau \) and extrapolating the results to \( \tau = 0 \). For small \( \tau \), the time-step bias error is proportional to \( \tau^4 \). \(^{68}\)

C. Simulation setup

FCIQC calculations were performed using the NECI package. The periodic Hartree-Fock calculations from which the single particle orbitals were extracted, and CCSD calculations were carried out using
PySCF\textsuperscript{65,66}. The Gaussian basis sets of SZV, DZVP, and TZVP\textsuperscript{67} were used for all basis set calculations. The density fitting technique was used for dealing with the divergence of long-range Coulomb interaction in periodic CCSD calculation. For FCIQMC simulations, $4 \times 10^7$ number of walkers were used for higher densities, which increased up to $1.5 \times 10^8$ for larger simulation cells. The FCIQMC calculations were performed using the integrals in FCIDUMP format\textsuperscript{63}.

Our VMC and DMC calculations were performed using the CASINO QMC package\textsuperscript{62} and a trial wave function of Slater-Jastrow form. The one-electron orbitals were extracted from DFT calculations using the Quantum Espresso code\textsuperscript{39}. A norm-conserving DFT pseudopotential with the Perdew-Zunger parameterization of the local density approximation\textsuperscript{71} was used. For converging to the complete basis-set limit\textsuperscript{72}, we set a basis set cutoff of $5.0 \times 10^3$ eV. The Jastrow term $J(R)$, which captures most of the dynamical correlation between electrons and is a positive, symmetric, explicit function of interparticle distances, consisted of polynomial one-body electron-nucleus (en), two-body electron-electron (ee), and three-body electron-electron-nucleus (een) terms, the parameters of which were optimized by variance minimization at the variational Monte Carlo level\textsuperscript{73,74}. The main approximation in the DMC results is the fixed-node approximation, which can be improved by including backflow transformation in the trial wave function\textsuperscript{19}. The BCC conventional cell, which was used to build a supercell for all calculations, includes two hydrogen atoms located at the corner and centre of the cell that were fixed for all the studied densities and only the lattice parameter changed for each $r_S$.

III. RESULTS AND DISCUSSION

A. FCIQMC and CCSD

Figure 1 shows representative convergence of the FCIQMC correlation energy in mEa per atom for BCC crystalline hydrogen at different densities within a DZVP basis and a simulation cell of 16 hydrogen atoms and 80 orbitals. The total number of walkers in the calculation for each unit cell volume was grown in stages, to check the convergence of the energy estimator. The number of walkers required to achieve convergence with respect to the initiator error varied between the cell sizes, from 40 million walkers at compressed geometries, to 150 million walkers at the expanded geometries where stronger correlation effects are expected to be prevalent. Remaining systematic error is expected to be sub-mHa per atom, resulting in confidence in its ability to benchmark other approaches.

We evaluate the convergence of the energy with respect to the basis set size, considering both the DZVP (with 5 orbitals per atom) and the smaller SZV basis (which has only one basis functions per atom). Table I gives the FCIQMC energies of the system at different densities up to accuracy of 0.1 mH/atom, which represents the confidence in the accuracy of the FCIQMC considering the likely remaining systematic error, as well as the stochastic error estimate from a blocking analysis. We compare these results to those of coupled-cluster theory (CCSD), obtained within the pyscf simulation package. Table II gives the CCSD energies of BCC atomic hydrogen with two atoms per primitive unit cell, which were obtained using SZV, DZVP, and TZVP basis sets and a $2 \times 2 \times 2$ k-point mesh, at different densities.

From these results, we can obtain the equation of state (EOS) of BCC atomic hydrogen for these restricted basis sets and $2 \times 2 \times 2$ simulation cells. This is shown in Fig. 2 for both FCIQMC and CCSD methods. Agreement between these two methods is excellent at higher densities, where the single-reference nature of the CCSD ansatz is expected to perform well. Around the equilibrium cell volumes, at $r_S = 1.6282$ and $r_S = 1.8608$ a.u, the difference between FCIQMC-SZV and CCSD-SZV energies are -0.97 and -2.4 mHa/atom, respectively. At expanded
cell volumes however, the correlation energy captured by FCIQMC increases significantly over the CCSD results for both basis sizes, with the FCIQMC-SZV energy 9.8 mHa/atom lower than CCSD-SZV at the $r_S = 2.3261$ a.u. cell volume. This is anticipated, due to the increasing levels of stronger correlation effects present in the system, and evidenced by the increasingly multiconfigurational nature of the FCIQMC calculation.

Using DZVP basis set introduces a major energy gain in FCIQMC and CCSD calculations. At the equilibrium cell volume, the difference between FCIQMC-DZVP and FCIQMC-SZV is 23.3 mHa/atom. Once again, the agreement between FCIQMC and CCSD at compressed cells is excellent, with the agreement getting worse as the cell expands. However, this discrepancy between the CCSD and FCIQMC at the expanded geometries is less in this larger basis, pointing to an overestimation of the relative importance of the strong correlation in small basis sets, while larger basis sets are able to more effectively screen these strong correlations. This improved description of the screening available in the larger basis results in only a $\sim$1.36 mHa/atom improvement in the DMC total energy compared to just using a single Slater determinant trial state. This energy gain is significant for the VMC energies, it may be more efficient to consider a trial wave function for DMC which has not included the 2-body Jastrow term, Slater determinant with only Backflow (Slater+BF), and with two- and one-body Jastrow (Slater+2,1bJ); Slater with only Backflow (Slater+BF), and Backflow and Jastrow with one-, two-, and three-body terms (Slater+1,2,3bJ+BF). Simulation cell contains 16 hydrogen atoms.

For both basis sizes, with the FCIQMC-SZV energy 9.8 mHa/atom lower than CCSD-SZV at the $r_S = 2.3261$ a.u. cell volume. This is anticipated, due to the increasing levels of stronger correlation effects present in the system, and evidenced by the increasingly multiconfigurational nature of the FCIQMC calculation.

### Table II. CCSD total energy in Hartree/atom, which are obtained using SZV, DZVP, and TZVP basis sets at different densities. The primitive cell includes two hydrogen atoms, sampled with a $2 \times 2 \times 2$ k-point grid. $r_S$ is in atomic unit.

| $r_S$ | $E_{\text{tot}}^{\text{SZV}}$ | $E_{\text{tot}}^{\text{DZVP}}$ | $E_{\text{tot}}^{\text{TZVP}}$ |
|-------|----------------|----------------|----------------|
| 1.1630 | -0.430469 | -0.461751 | -0.49909 |
| 1.3956 | -0.487386 | -0.513529 | -0.513879 |
| 1.6282 | -0.502846 | -0.524253 | -0.524793 |
| 1.8608 | -0.502244 | -0.518626 | -0.519386 |
| 2.0935 | -0.496091 | -0.508172 | -0.509355 |
| 2.3261 | -0.489972 | -0.498029 | -0.499688 |

### Table III. VMC and DMC energies of BCC atomic Hydrogen at $r_S = 1.3956$ a.u. The energies for different trial wave functions, given as a Slater determinant without Jastrow term, Slater determinant with two-body Jastrow (Slater+2bJ), with two- and one-body Jastrow (Slater+2,1bJ), and with two-, one, and three-body Jastrow (Slater+2,1,3bJ); Slater with only Backflow (Slater+BF), and Backflow and Jastrow with one-, two-, and three-body terms (Slater+1,2,3bJ+BF). Simulation cell contains 16 hydrogen atoms.

| Trial state | VMC | Variance | DMC |
|-------------|-----|----------|-----|
| Slater      | -0.4719\(^4\) | 3.01\(^2\) | -0.5071\(^5\) |
| Slater+2bJ  | -0.5010\(^3\) | 0.6990\(^9\) | -0.5074\(^7\) |
| Slater+2,1bJ| -0.5011\(^3\) | 0.6983\(^3\) | -0.5074\(^7\) |
| Slater+BF   | -0.4648\(^7\) | 4.41\(^1\) | -0.5079\(^5\) |
| Slater+1,2,3bJ+BF | -0.5042\(^9\) | 0.6244\(^8\) | -0.5085\(^1\) |

The results of variational and diffusion Monte Carlo calculations at a cell volume of $r_S = 1.3956$ a.u. with 16 atoms are shown in Table. III. These show that the dominant error in the DMC values derives from the fixed node approximation, since the accuracy of the DMC is almost independent of the flexibility afforded by the Jastrow component of the trial wave function, which does not affect the nodal structure significantly. This suggests that while the importance of the two-body Jastrow is significant for the VMC energies, it may be more efficient to consider a trial wave function for DMC which has not included the 2-body Jastrow terms. The 1-body Jastrow terms are also found to be negligible in this system, with the 3-body terms also found to be less important to the DMC energy than the inclusion of backflow correlations. It is found that this backflow reduces the DMC energy by 0.8 mHa/atom, which is more than twice the energy gain from inclusion of the 2-body Jastrow. The importance of backflow in high-density matter has been seen before, where in the homogeneous three-dimensional electron gas with $r_S < 5$ the effects of backflow were found to dominate over those introduced by three-body correlations. Indeed, for this pure hydrogen system, the inclusion of all Jastrow terms and backflow transformation into the wave function results in only a $\sim 1.36$ mHa/atom improvement in the DMC total energy compared to just using a single Slater determinant trial state. This energy gain is

---

### B. Diffusion Monte Carlo

The results of variational and diffusion Monte Carlo calculations at a cell volume of $r_S = 1.3956$ a.u. with 16 atoms are shown in Table. III. These show that the dominant error in the DMC values derives from the fixed node approximation, since the accuracy of the DMC is almost independent of the flexibility afforded by the Jastrow component of the trial wave function, which does not affect the nodal structure significantly. This suggests that while the importance of the two-body Jastrow is significant for the VMC energies, it may be more efficient to consider a trial wave function for DMC which has not included the 2-body Jastrow terms. The 1-body Jastrow terms are also found to be negligible in this system, with the 3-body terms also found to be less important to the DMC energy than the inclusion of backflow correlations. It is found that this backflow reduces the DMC energy by 0.8 mHa/atom, which is more than twice the energy gain from inclusion of the 2-body Jastrow. The importance of backflow in high-density matter has been seen before, where in the homogeneous three-dimensional electron gas with $r_S < 5$ the effects of backflow were found to dominate over those introduced by three-body correlations. Indeed, for this pure hydrogen system, the inclusion of all Jastrow terms and backflow transformation into the wave function results in only a $\sim 1.36$ mHa/atom improvement in the DMC total energy compared to just using a single Slater determinant trial state. This energy gain is
smaller than chemical accuracy, and so is unlikely to be worthwhile in general for pure hydrogen systems.

C. Finite size convergence

Using a finite simulation cell to mimic the properties of a periodic system introduces finite-size (FS) errors, which is one of the main challenges in the application of many-body techniques to extended systems. These FS errors have a number of different origins, including the description of the kinetic energy, the periodic Ewald interaction, and exchange energy. Therefore, careful control to mitigate the finite simulation cells and ensure that desired quantities are converged with respect to these errors is essential for reliable results. Here, we employ the standard finite-size extrapolation technique introduced by Ceperley and co-workers for reducing the FS errors.

We used the fitting formula

\[ E_{MB,\infty} \sim E_{MB,N} + \alpha (E_{0,\infty} - E_{0,N}) + \beta / N \]  

where \( \alpha \) and \( \beta \) are fitting parameters, and \( E_0 \) is the energy of system obtained via a single-particle or mean-field approach. \( E_{MB,N} \) represents the many-body energy of the finite system of \( N \) interacting electrons. Using this form, we extrapolated the CCSD-SZV and DMC energies to the infinite system size limit. For the CCSD-SZV calculations, the \( E_{0,\infty} \) value is approximated to be the HF energy of the system for that density, obtained using an \( 8 \times 8 \times 8 \) k-point mesh. For the DMC extrapolation, the local density approximation was used with a \( 24 \times 24 \times 24 \) k-point mesh to obtain \( E_{0,\infty} \).

The top panel of Figure 3 shows CCSD-SZV energies which are calculated at four supercell sizes and different densities, up to a maximum supercell size of \( 4 \times 4 \times 4 \). The bottom panel shows the extrapolation of the DMC energies at each density (computed with the Slater with one- and two-body Jastrow trial wave function), up to the largest supercell size with 432 atoms. The final CCSD-SZV and DMC energies extrapolated to the infinite system size limit, estimated from Eq. 17, are given in table IV, along with the standard deviation in the fit. However, the comparison between these approaches is fundamentally limited by the small basis size (SZV) of the extrapolated CCSD results, which was required in order to reach the large supercell sizes.

As a correction for the CCSD results to mitigate this small basis, we can assume that the basis set error between SZV and DZVP is independent of supercell size, with these errors therefore being additive. This allows us to compute a final basis set correction which can be applied to the extrapolated CCSD results at the SZV basis level. Furthermore, we can similarly compute the energetic correction between CCSD and FCIQMC in the DZVP basis sets, and also consider this as a correction for the correlated physics beyond the CCSD ansatz in the thermodynamic limit, to obtain a final equation of state for the system.

D. BCC atomic solid hydrogen equation of state

Figure 4 illustrates the EOS of BCC atomic solid hydrogen which is obtained from FCIQMC, CCSD, and DMC methods. FCIQMC and CCSD results calculated within a DZVP basis set, are shown for both a finite supercell of 16 hydrogen atoms, and also the values estimated via extrapolation to the infinite system size limit. The DMC energies are also shown having been extrapolated to the thermodynamic limit. Encouragingly, all three many-body wave function-based methods (FCIQMC, CCSD and DMC), which operate under very different assumptions and approximations, give the same equilibrium density for this phase, which is found to be reliably close to \( r_S = 1.65 \) a.u. This is despite the
TABLE IV. Extrapolated CCSD and DMC energies to infinite system size limit $E_{\text{CCSD,}\infty}$ and $E_{\text{DMC,}\infty}$, and the standard deviation $\sigma$ in the extrapolation for each density. The CCSD energies are extrapolated from results within an SZV basis set. The DMC energies are obtained using Slater+2,1bJ trial wave function. Energies are in Ha/atom, and the number in parentheses provides the variance of the parameter estimate.

| $r_S$/a.u. | $E_{\text{CCSD,}\infty}$ | $\sigma_{\text{CCSD}}$ | $E_{\text{DMC,}\infty}$ | $\sigma_{\text{DMC}}$ |
|------------|------------------|----------------|------------------|----------------|
| 1.1630     | -0.43095(2)      | 0.206(5)       | -0.45069(7)      | 0.233(8)       |
| 1.3956     | -0.48117(1)      | 0.135(3)       | -0.50827(5)      | 0.145(7)       |
| 1.6282     | -0.492595(9)     | 0.121(3)       | -0.52315(2)      | 0.111(4)       |
| 1.8608     | -0.488757(9)     | 0.124(3)       | -0.520355(6)     | 0.048(3)       |
| 2.0935     | -0.470905(1)     | 0.008(1)       | -0.515105(6)     | 0.049(2)       |
| 2.3261     | -0.46240(1)      | 0.165(4)       | -0.504769(5)     | 0.040(2)       |

FIG. 4. EOS of BCC solid atomic hydrogen. FCIQMC-DZVP and CCSD-DZVP denote results obtained with a 16 atom supercell and DZVP basis set, while FCIQMC-DZVP$_{\text{inf}}$ and CCSD-DZVP$_{\text{inf}}$ are estimated results extrapolated to the infinite system size limit. DMC$_{\text{inf}}$ shows the DMC energies also extrapolated to the thermodynamic limit.

FCIQMC and CCSD being built on HF orbitals, while DMC operates in the continuum, with the nodal surface built on the Kohn-Sham LDA determinant, yet all three methods agree that the BCC atomic solid hydrogen with two atoms per cell has a minimum ground state energy at lattice parameter 3.307 Bohr.

However, although the equilibrium agreement is good, it is clear that discrepancies between the approaches remains, with the $2 \times 2 \times 2$ supercell CCSD and FCIQMC results in better agreement with the extrapolated DMC results than their extrapolated counterparts. This discrepancy is likely to be due to the thermodynamic limit extrapolation for the CCSD, which was calculated only for the small SZV basis sets, which have previously been shown to be qualitatively different to the improved DZVP basis sets. This likely dominant source of error is a key focus for improvements in this equation of state going forwards.

The extrapolation to infinite system size limit of DMC energies indicates that the FS correction in DMC calculation, regardless of the density of studied systems, is always positive. Hence, the DMC energies of finite systems are lower than the DMC energy at the thermodynamic limit. In contrast, the CCSD FS analysis shows that the slope of the fitted linear function of energy with respect to $1/N$ is positive for the density range of $r_S < 2.0$. Therefore the FS correction is negative (depends on the density of system), and the total energy per atom at infinite system size limit is lower than the total energy per atom of finite system. The slope of fitted line changes the sign by decreasing the density, and the CCSD FS correction for $r_S > 2.0$ is positive. The behaviour of CCSD FS errors as function of system size is driven by the size of employed basis set. For the CCSD FS analysis, the SZV basis set, with insufficient number of orbitals per atom of one, was used. Analysing the numerical data demonstrates that the CCSD-SZV correlation energy strongly depends on the density and sharply decreases by enlarging the cell volume. Whereas the CCSD-DZVP and CCSD-TZVP correlation energies are almost independent of the cell size within studied density window.

IV. CONCLUSION

We report a numerical study of the equation of state for BCC solid hydrogen at density range of $1.1 < r_S < 2.4$ using many-body stochastic wave-function techniques of FCIQMC, VMC, and DMC, as well as the deterministic quantum chemical periodic coupled-cluster method. Consideration was given to both basis set convergence, using the SZV, DZVP and TZVP basis sets. It was found that while significant additional correlation was captured at the level of FCIQMC for expanded cell sizes compared to CCSD, at compressed and equilibrium geometries, these results agree well. The SZV basis was found to give a qualitatively different equation of state compared to larger basis sets, however was used to design a finite-size correction at each density to account for the restricted supercell sizes. This was performed via extrapolation of the CCSD results in the SZV basis for four different system sizes, and is found to be highly density dependent.

These results were compared to the results of VMC and DMC calculations, which are complementary approaches which operate in the continuum and therefore avoid much of the basis set error, however suffer from nodal restrictions. Systematically improving the flexibility of the trial state was shown to significantly improve the VMC energy by 29.24 mHa/atom, while the DMC energy gain is ~0.323 mHa/atom compared with a single not-optimised LDA Slater determinant trial wave function. Since the Jastrow optimisation for large real Hamiltonian systems can be quite expensive, the DMC results predict that for the pure hydrogen systems the dynamical correlation contribution in the ground state energy is small and can be disregarded in practical DMC calculation. Finite size
extrapolation from simulated DMC supercells up to 432 atoms in simulation cell was shown to have a different character compared to the CCSD extrapolation, and is likely more reliable. Combining the results of all methods showed an excellent agreement in the prediction of the equilibrium lattice parameter of 3.307 Bohr. However, discrepancies remained in the broader shape of the equation of state, likely arising from the restricted basis sets required for the thermodynamic limit extrapolation.

V. ACKNOWLEDGEMENT

S.A. gratefully acknowledges generous allocation of computing time by the Paderborn Center for Parallel Computing (PC²) on OCULUS and the FPGA-based supercomputer NOCTUA. G.H.B. gratefully acknowledges support from the Royal Society via a University Research Fellowship, as well as funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 759063.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

1 D.P. Landau, and K. Binder, ”A Guide to Monte Carlo simulations in statistical Physics” Cambridge University Press (2014)
2 M. H. Kalos, Phys. Rev. 128, 1791 (1962) ; J. Com. Phys. 1, 257 (1966)
3 D. M. Ceperley, Rev. Mod. Phys. 67, 279 (1995a)
4 F. Becca, and S. Sorella, ”Quantum Monte Carlo approaches for correlated systems”, Cambridge University Press (2017)
5 W.M.C. Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal Rev. Mod. Phys. 73, 33 (2001)
6 C.J. Umrigar, J. Toulouse, C. Filippi, et al. Phys. Rev. Lett. 98, 110201 (2007)
7 E. Neuscamman, C.J. Umrigar, and G.K.-L. Chan, Phys. Rev. B 85, 045103, (2012)
8 J.B. Anderson, J. Chem. Phys. 63 1499, (1975)
9 P.J. Reynolds, D.M. Ceperley, B.J. Alder, and W.A. Lester, J. Chem. Phys. 77, 5593 (1982)
10 J.C. Grossman, and L. Mitas, Report on Progress in Physics 74, 026502 (2011)
11 M. H. Kalos, Phys. Rev. B 36, 5844 (1984)
12 P.J. Bartlett, Annual Rev. Phys. Chem. 32 359 (1981)
13 R.J. Bartlett, and M. Musial, Rev. Mod. Phys. 79, 291 (2007)
14 J.F. Stanton, and R.J. Bartlett, J. Chem. Phys. 98, 7092 (1993)
15 P. Cizek, J. Chem. Phys. 45, 4256 (1966)
16 B. Zejiozski, and H. J. Monkhorst, Phys. Rev. A 24, 1668 (1981)
17 C. D. Sherrill, and H. F. Schafer, ”Advances in Quantum Chemistry” 34, edited by P-O. Löwdin, Academic Press, New York, (1999)
18 C. Müller, B. Paulus, Phys. Chem. Chem. Phys. 14, 7605 (2012)
19 M. Marsman, A. Grünneis, J. Paier, and G. Kresse, J. Chem. Phys. 130, 184103 (2009)
20 T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grünneis, Phys. Rev. X 8, 021043 (2018)
21 A. Pulkin, and G. Kin-Lic Chan, Phys. Rev. B 101, 241113(R) (2020)
22 Y. Gao, Q. Sun, J. M. Yu, M. Motta, J. McClain, A. F. White, A. J. Minnich, and G. Kin-Lic Chan, Phys. Rev. B 101, 165138 (2020)
23 J. McClain, Q. Sun, G. Kin-Lic Chan, and T. C. Berkelbach, J. Chem. Theory Comput. 13, 1209 (2017)
24 G. H. Booth, T. Tsatsoulis, G. Kin-Lic Chan, and A. Grünneis J. Chem. Phys. 145, 084111 (2016)
25 K. Lejaeghere et al., Science 351, a33000 (2016)
26 J. van Setten, F. Caruso, S. Shariﬁzadeh, X. Ren, M. Scheffer, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, J. Chem. Theory Comput. 11, 5665 (2015)
27 K. T. Williams et al., Phys. Rev. X 10, 011041 (2020)
28 M. Motta et al., Phys. Rev. X 7, 031059 (2017)
29 N. W. Ashcroft, Nature (London) 340, 345 (1989)
30 H. K. Mao and R. J. Hemley, Science 244, 1462 (1989)
31 E. Neuscamman, C.J. Umrigar, and G.K.-L. Chan, Phys. Rev. B 85, 045103, (2012)
32 J.B. Anderson, J. Chem. Phys. 63 1499, (1975)
33 P.J. Reynolds, D.M. Ceperley, B.J. Alder, and W.A. Lester, J. Chem. Phys. 77, 5593 (1982)
34 D. M. Ceperley and B. J. Alder, Phys. Rev. B 36, 2002 (1987)
35 D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. 45, 566 (1980)
36 X. W. Wang, J. Zhu, S. G. Louie, and S. Fahy, Phys. Rev. Lett 65, 2414 (1990)
37 K. Liao, X. Li, A. Alavi, and A. Grünneis, npj Comput Mater 5, 110 (2019)
38 S. Azadi, B. Monserratt, W.M.C. Foulkes, and R.J. Needs, Phys. Rev. Lett. 112, 165501 (2014)
39 S. Motta, and T.D. Kühne, Phys. Rev. B 100 155103 (2019)
40 F. R. Petruzielo, A. H. movable, Hitesh J. Changlang, M. P. Nightingale, and C. J. Umrigar Phys. Rev. Lett. 109, 230201 (2012)
41 S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, J. Chem. Phys. 142, 184107 (2015)
42 J. S. Spencer, N. S. Blunt, and W.M.C. Foulkes, J. Chem. Phys. 136 054110 (2012)
43 S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, J. Chem. Phys. 143, 134117 (2015)
44 S. Blunt, G. H. Booth, and A. Alavi, J. Chem. Phys. 146, 244105 (2017)
45 C. Overy, G. H. Booth, N. S. Blunt, J. J. Shepherd, D. Cleland, and A. Alavi, J. Chem. Phys. 141, 244117 (2014)
46 R. E. Thomas, D. Opalka, C. Overy, P. J. Knowles, A. Alavi, and G. H. Booth, J. Chem. Phys. 143, 054108 (2015)
47 P. Samanta, N. S. Blunt, and G. H. Booth, J. Chem. Theory Comput. 14, 3532 (2018)
48 R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, J. Chem. Theory Comput. 11, 5316 (2015)
49 R. J. Anderson, T. Shiozaki, and G. H. Booth, J. Chem. Phys. 152, 054101 (2020)
50 D.M. Ceperley, and L. Mitas “New methods in computational quantum mechanics”, 1-38 (1995)}
64 K. Guther et al., J. Chem. Phys. 153, 034107 (2020)
65 Q. Sun et al J. Chem. Phys. 153, 024109 (2020)
66 Q. Sun et al WIREs, https://doi.org/10.1002/wcms.1340 (2017)
67 J. VandeVondele, and J. Hutter, J. Chem. Phys. 127 114105 (2007)
68 C.J. Umrigar, M.P. Nightingale, K.J. Runge, J. Chem. Phys. 99, 2963 (1993)
69 R. Gaudoin and J. M. Pitarke, Phys. Rev. Lett. 99, 126406 (2007)
70 P. Giannozzi et al., J. Phys.: Condens. Matter 21, 395502 (2009).
71 J. P. Perdew, and A. Zunger, Phys. Rev. B 23, 5048 (1981).
72 S. Azadi, C. Cavazzoni, and S. Sorella, Phys. Rev. B 82, 125112 (2010).
73 C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. 60, 1719 (1988).
74 N. D. Drummond, and R. J. Needs, Phys. Rev. B 72, 085124 (2005).
75 P. López Ríos, N. D. Drummond, M.D. Towler, and R. J. Needs, Phys. Rev. B 74, 066701 (2006)
76 S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B 54, 1703 (1996)
77 Y. Kwon, D. M. Ceperley, and R. M. Martin Phys. Rev. B 58, 6800 (1998)
78 N.D. Drummond, R.J. Needs, A. Sorouri, and W.M. C.Foulkes, Phys.Rev.B 78, 125106 (2008)
79 S. Azadi, and W.M.C. Foulkes, J. Chem. Phys. 143 102807 (2015)
80 M. Holzmann, R. C. Clay, M. A. Morales, N. M. Tubman, D. M. Ceperley, and C. Pierleoni, Phys. Rev. B 94, 035126 (2016)
81 S. Azadi, and W.M.C. Foulkes, Phys. Rev. B 100, 245142 (2019)
82 D. M. Ceperley, and B. J. Alder, Phys. Rev. B 36, 2092 (1987)