Title

QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo

Journal

The Journal of Chemical Physics, 152(17)

ISSN

0021-9606

Authors

Kent, PRC
Annaberdiyev, Abdulgani
Benali, Anouar
et al.

Publication Date

2020-05-07

DOI

10.1063/5.0004860

Peer reviewed

QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion Quantum Monte Carlo

P. R. C. Kent,[1, a)] Abdulgani Annaberdiyev,[2] Anouar Benali,[3] M. Chandler Bennett,[4] Edgar Josué Landinez Borda,[5] Peter Doak,[6] Hongxia Hao,[7] Kenneth D. Jordan,[8] Jaron T. Krogel,[9] Ilkka Kylänpää,[10] Joonho Lee,[11] Ye Luo,[3] Fionn D. Malone,[5] Cody A. Melton,[12] Lubos Mitas,[2] Miguel A. Morales,[5] Eric Neuscamman,[7, 13] Fernando A. Reboredo,[9] Brenda Rubenstein,[14] Kayahan Saritas,[15] Shiv Upadhyay,[8] Guangming Wang,[2] Shuai Zhang,[16] and Luning Zhao[17]

[1)] *Center for Nanophase Materials Sciences Division and Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*[b)]

[2)] *Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202, USA*

[3)] *Computational Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60439, USA*

[4)] *Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831, USA*

[5)] *Quantum Simulations Group, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94551, USA*

[6)] *Center for Nanophase Materials Sciences Division and Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA*

[7)] *Department of Chemistry, University of California, Berkeley, California 94720, USA*

[8)] *Department of Chemistry, University of Pittsburgh, Pennsylvania 15260, USA*

[9)] *Materials Science and Technology Division, Oak Ridge National Laboratory,*

Oak Ridge, Tennessee 37831, USA

[10] *Computational Physics Laboratory, Tampere University, P.O. Box 692, 33014 Tampere, Finland*

[11] *Department of Chemistry, Columbia University, New York, NY 10027, USA*

[12] *Sandia National Laboratories, Albuquerque, New Mexico 87123, USA*

[13] *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*

[14] *Department of Chemistry, Brown University, Providence, RI 02912, USA*

[15] *Department of Applied Physics, Yale University, New Haven, CT 06520, USA*

[16] *Laboratory for Laser Energetics, University of Rochester, 250 E River Rd, Rochester, NY 14623, USA*

[17] *Department of Chemistry, University of Washington, Seattle, Washington, 98195, USA*

(Dated: 7 May 2020)

We review recent advances in the capabilities of the open source ab initio Quantum Monte Carlo (QMC) package QMCPACK and the workflow tool Nexus used for greater efficiency and reproducibility. The auxiliary field QMC (AFQMC) implementation has been greatly expanded to include k-point symmetries, tensor-hypercontraction, and accelerated graphical processing unit (GPU) support. These scaling and memory reductions greatly increase the number of orbitals that can practically be included in AFQMC calculations, increasing accuracy. Advances in real space methods include techniques for accurate computation of band gaps and for systematically improving the nodal surface of ground state wavefunctions. Results of these calculations can be used to validate application of more approximate electronic structure methods including GW and density functional based techniques. To provide an improved foundation for these calculations we utilize a new set of correlation-consistent effective core potentials (pseudopotentials) that are more accurate than previous sets; these can also be applied in quantum-chemical and other many-body applications, not only QMC. These advances increase the efficiency, accuracy, and range of properties that can be studied in both molecules and materials with QMC and QMCPACK.

[a)]kentpr@ornl.gov

I. INTRODUCTION

Quantum Monte Carlo (QMC) methods are an attractive approach for accurately computing and analyzing solutions of the Schrödinger equation.[1–3] The methods form a general ab initio methodology able to solve the quantum many-body problem, applicable to idealized models such as chains or lattices of atoms through to complex and low-symmetry molecular and condensed matter systems, whether finite or periodic, metallic or insulating, and with weak to strong electronic correlations. Significantly, the methods can naturally treat systems with significant multi-reference character, and are without electron self-interaction error, which challenges many quantum chemical approaches and density functional theory approximations, respectively. The methods continue to be able to take advantage of improvements in computational power, giving reduced time to solution with new generations of computing. Due to these features, usage of QMC methods for first principles and ab initio calculations is growing.

Compared to traditional deterministic approaches, QMC methods are generally distinguished by: (1) use of statistical methodologies to solve the Schrödinger equation. This allows the methods to not only treat problems of high-dimensionality efficiently, but also potentially use basis, wave function, and integral forms that are not amenable to numerical integration. (2) Use of few and well-identified approximations that can potentially be quantified or made systematically convergeable. (3) A low power scaling with system-size, but large computational cost prefactor. (4) High suitability to large scale parallel computing owing to lower communications requirements than conventional electronic structure methods. Scaling has been demonstrated to millions of compute cores.[4]

Modern applications of QMC have expanded to cover many of the same systems studied by density functional theory (DFT) and quantum chemical approaches, and in many cases also at a similar atom and electron count, although at far greater computational cost. Besides those described in below, recent molecular applications of QMC include studies of the nature of the quadruple bond in C_2[5], acenes[6], physisorption of water on graphene[7], binding of transition metal diatomics[8], and DNA stacking energies[9]. Materials applications include nitrogen defects in ZnO[10], excitations in Mn doped phosphors[11], and the singlet-triplet excitation in $MgTi_2O_4$[12]. Methodological improvements include: reducing the sensitivity of pseudopotential evaluation[13], extensions to include linear response[14], density functional

embedding[15], excited states including geometry optimization[16], improved twist averaging[17], and accurate trial wavefunctions via accurate densities[18]. Importantly, for model systems such as the hydrogen chain, the methods can be used to benchmark themselves as well as other many-body approaches[19]. This partial list of developments and applications from the last two years alone indicates that the field is growing and maturing.

In this article we describe recent updates to the QMCPACK code and its ecosystem of wavefunction converters and workflow tools. These updates have aimed to expand the range of systems, properties, and accuracies that can be achieved both with QMCPACK and with QMC techniques in general. For a description of the underlying methodology we refer the reader to Refs. 1–4, and 20. In particular, a thorough description of real space QMC methods is given in Section 5 of Ref. 4. For an extensive introduction to AFQMC we refer the reader to Ref.20.

QMCPACK is a fully open source and openly developed QMC package, with 48 coauthors on the primary citation paper[4] published in 2018 and an additional 5 contributors since then. The main website for QMCPACK is `https://qmcpack.org` and the source code is currently available through `https://github.com/QMCPACK/qmcpack`. QMCPACK aims to implement state of the art QMC methods, be generally applicable, easy to use, and high-performing on all modern computers. Since the publication of Ref.[4], the range of QMC calculations that are possible has been expanded by significant enhancements to the Auxiliary-Field QMC (AFQMC) solver. This orbitally based method is distinct from and complementary to the longer-implemented real space methods of variational and diffusion QMC (VMC and DMC, respectively). The AFQMC implementation can fully take advantage of graphics processing units (GPUs) for a considerable speedup and, unlike the real-space methods, can also exploit k-point symmetries. It shares the same workflow tool, Nexus, which helps simplify and ease application of all the QMC methods by new users as well as aid in improving reproducibility of complex multi-step research investigations. To our knowledge this is currently the only AFQMC code designed for large scale research calculations that is open source. To help guarantee the future of the code, it is undergoing rapid development and refactoring to target the upcoming Exascale architectures as part of the U.S. Exascale Computing Project[21], which also entails major updates to the testing, validation and maintainability.

The electronic structure and quantum chemical codes that QMCPACK is interfaced to for trial wavefunctions has been expanded to include Qbox[22], PySCF[23], Quantum Espresso[24],

Quantum Package[25], and GAMESS[26]. Additional codes such as NWCHEM[27] can be interfaced straightforwardly.

In the following, we first review in Section II the open development principles of QMCPACK. In Section III we discuss updates to the Nexus workflow package. This integrates entire research electronic structure workflows for greater productivity and reproducibility than by-hand invocation of individual calculations. Due to the infeasibility of performing QMC calculations for general systems using an all electron approach, use of effective core potentials (ECPs), or pseudopotentials, is essential. To improve the accuracy obtainable we have developed a new approach and set of "correlation consistent" ECPs. These can be used in all ab initio calculations, not only QMC, and are described in Section IV. Advances in the AFQMC implementation are described in Section V. Turning to real-space QMC methods in Section VI, algorithms and multiple determinant trial wavefunctions can now be used to obtain improved ground state energies as well as band gaps in solid-state materials. As a result, it is now possible to begin to test the accuracy of the nodal surfaces that have long been used in these calculations. Finally, in Section VII, we give three applications: first, application to non-valence anions, which challenge all electronic structure and quantum chemical techniques. Second, application to excitations of localized defects in solids VII B. Third, the ability to obtain the momentum distribution has recently been improved, motivated by recent experiments on VO_2. A summary is given in Section VIII.

II. OPEN DEVELOPMENT AND TESTING

Fully open source development is an important core value of the QMCPACK development team. Besides improving the quality of the software, anecdotally it also improves the on-boarding experience for new users. While the developers of many electronic structure packages now practice some degree of open development, QMCPACK has seen very significant benefits from this in the last few years. We expect other packages would also benefit from full adoption and therefore give details here.

QMCPACK is an open source package, with releases and the latest development source code available through `https://github.com/QMCPACK/qmcpack`. QMCPACK is written in C++14, with MPI parallelization between compute nodes, OpenMP threading used for multicore parallelism. CUDA is used for NVIDIA accelerators. Options to support CUDA,

complex valued wavefunctions, and to adjust the numerical precision used internally are currently compile time options.

Besides adoption of a distributed source code control system, we have found that development productivity can be further increased by adoption of code reviews and continuous integration (testing). To maximize the efficiency of both contributors and reviewers and shorten the development cycle of new features, work-in-progress pull requests are encouraged for early engagement in the process. The early review allows guidance to be given, e.g. are the algorithms clear enough to other developers and are the coding guidelines being followed. At the same time, continuous integration is applied to the proposed code change. This process routinely catches cases that developers may not have considered or tested against, e.g. the complex-valued build of QMCPACK or accelerated GPU support, that are compile time options. This period of comment while the work is being completed also helps advertise the work to other developers and minimizes risk of duplicated work. Our experience strongly suggests that this process reduces bugs, reduces potential developer's effort, and saves reviewer's time compared to a late engagement with an unexpected pull request. All the discussions around the code change become archived searchable documentation and potential learning materials.

Testing of QMCPACK has been significantly expanded. Two years ago, QMCPACK had limited unit, integration and performance testing categories: unit tests that run quickly on individual components; integration tests that exercise entire runs; performance tests for monitoring relative performance between code changes. However, due to the stochastic nature of QMC, as the number of tests and build combinations increased it became impractical to run the integration tests long enough to obtain a statistically reliable pass/fail: the smallest (shortest) integration test set currently takes around one hour to execute on a 16 core machine, and must necessarily suffer from occasional statistical failures. Thus, a new category of tests was needed for quickly examining full QMC execution with a reproducible Monte Carlo trajectory. The new deterministic integration tests are modified QMC runs with only a few steps, very few Monte Carlo walkers, and fixed random seeds for absolute reproducibility. All the major features of QMCPACK are covered by this new of category tests. Running all the unit and deterministic integration tests takes approximately one minute which is fast enough for iterative development and fast enough to be used in continuous integration. This fast to run set of tests facilitates significant changes and refactoring of the application which

otherwise would be far more difficult to test and unlikely to be attempted by non-experts without long experience with the codebase. All the deterministic tests are accompanied by longer running statistical tests that can be used to verify a new implementation when changes alter a previous deterministic result. Combinations of these tests are run automatically on a nightly basis and report to a public dashboard `https://cdash.qmcpack.org`. At the time of writing, around 25 different machine and build combinations are used to run around 1000 labeled tests each, and most of these cover multiple features.

Improving source code readability is critical for both new and experienced developers. In the past, misleading variable or class names and confusing function names have confused developers and resulted in subtle bugs, e.g. due to similarly named functions, only one of which updates internal state in the Monte Carlo algorithm. For this reason, coding standards including naming conventions have been added in the manual and are enforced on newly contributed codes. Existing codes are updated to follow the standards as they need other modifications. Automatic source formatting is also applied with the help of the clang-format tool. Concomitantly, both developer sections in the manual and source code documentation are significantly expanded.

As a result of the above changes, new contributors with a basic theoretical background can connect source code with textbook equations with much less difficulty than in the past. These efforts are clearly bringing long term benefit to QMCPACK and hopefully can be transferred to other scientific applications as well.

III. IMPROVING QMC WORKFLOWS WITH NEXUS

QMC techniques are progressing from methods under research towards more routine application. In this transition, usability of QMC becomes an important factor. A mature, usable computational method transfers responsibility for correct execution from users to the code. Major factors determining overall usability include: ease of requesting a desired result (in the form of input), robustness of the code in obtaining the desired result, and complexity of the overall calculation process. All of these contribute to the effort required by the user to obtain desired results. In essence, higher required effort translates directly into lower productivity of the user base. Lower productivity in turn risks a lower overall adoption rate and thus blunts overall impact of the method. It is therefore important to

seek to understand and minimize barriers to the practical use of QMC.

To illustrate the complexity of the QMC calculation process, we describe below a basic but realistic sequence of calculations (a scientific workflow) that is required to obtain a final fixed node DMC total energy per formula unit for a single crystalline solid with QMCPACK. In this workflow, we suppose that self-consistent (SCF) and non-self-consistent (NSCF) calculations are performed with Quantum Espresso[24] and wavefunction optimization (OPT), variational Monte Carlo (VMC), and diffusion Monte Carlo calculations are performed with QMCPACK. SCF/NSCF calculations might be performed on a workstation or a few nodes of a cluster, VMC/OPT calculations on a research-group sized cluster (∼30 nodes), and DMC on high performance computing resources (∼1000 nodes).

1. Converge DFT orbitals with respect to plane-wave energy cutoff (4–6 SCF calculations ranging from 300 to 800 Ry for the energy cutoff).

2. Converge B-spline orbital representation with respect to B-spline mesh spacing (1 NSCF, ∼5 VMC calculations in a small supercell over a series of finer mesh spacings).

3. Converge twist grid density (∼5 VMC calculations in a small supercell for a series of increasingly dense supercell Monkhorst-Pack twist grids).

4. Determine best optimization process (∼6 optimization (OPT) calculations in a small supercell over varying input parameters and e.g. Jastrow forms).

5. Obtain fixed node DMC total energy (∼3 NSCF, ∼3 OPT and ∼9 DMC calculations, 3 successively smaller timesteps for timestep extrapolation, 3 successively larger supercells for finite size extrapolation).

This basic workflow process is to be compared with the much reduced complexity for obtaining a single converged total energy for DFT, which typically requires only a single input file and single program execution to perform a single SCF calculation for the final energy. The complexity intrinsic to the basic workflow translates into a large degree of effort on the part of the user and limits the accessibility of the method for new users or for experienced users pursuing ambitious projects comprised of a large number of DMC calculations.

Scientific workflow tools make the QMC process more accessible in multiple ways: (1) bringing the constellation of electronic structure codes needed to produce a single QMC result under a single framework, (2) reducing the number of inputs required to request a desired result to a single user-facing input file, (3) reducing overall complexity by abstracting the execution process, (4) minimizing the direct effort required to execute the workflow process by assuming the management of simulation execution and monitoring from the user. Workflow tools have been applied with significant benefit to related electronic structure methods such as DFT[28–31] and also to QMC[32,33].

The Nexus workflow automation system[34] was created to realize these advantages for users of QMCPACK. Nexus is a Python-based, object oriented workflow system that can be run on a range of target architectures. Nexus has been used successfully on simple workstations and laptops, small group or institutional computing clusters, university level high performance computing centers in the U.S. and internationally, and Leadership Computing Facilities supported by the U.S. Department of Energy. Nexus has been used in a growing number of QMC studies involving QMCPACK and its uptake by new users is high.

Nexus abstracts user's interactions with each target simulation code that are components of a desired simulation workflow. Access to each respective code is enabled through single function calls that only require the user to specify a reduced set of important input parameters. Each function call resembles a small input block from a standard input file for an electronic structure code. Taken together, a sequence of these blocks comprises a new meta-input file that represents the data flow and execution pattern of the underlying simulation codes as a combined workflow.

Nexus assumes the responsibility of initiating and monitoring the progress of each simulation job in the workflow. Nexus generates expanded input files to each code based on the reduced inputs provided by the user. It also generates job submission files and monitors job execution progress via a lightweight polling mechanism. Apart from direct execution of each workflow step, Nexus also automates some tasks that previously fell to users. One example is that Nexus selects the best wavefunction produced during the non-linear statistical optimization process employed by QMCPACK and automatically passes this wavefunction to other calculations (such as diffusion Monte Carlo), which require it.

In the future, additional productivity gains might be realized with Nexus by further abstracting common workflow patterns. For example, convergence studies for orbital param-

eters (k-points, mesh-factors, source DFT functional) often follow similar patterns which could be encapsulated as simple components for users. Additionally, more of the responsibility for obtaining desired results, e.g. total energies to a statistically requested tolerance, could be handled by Nexus through algorithms that create and monitor dynamic workflows.

IV. EFFECTIVE CORE POTENTIALS

A. Introduction

All-electron (AE) QMC calculations become inefficient and eventually infeasible with increasing atomic number Z since the computational cost grows roughly as[35,36] Z^6. Since our primary interest is in valence properties, pseudopotentials and/or effective core potentials (ECP) are commonly employed to eliminate the atomic cores leading to valence-only effective Hamiltonians. Unfortunately, the existing tables and ECP generating tools have proved to exhibit somewhat mixed fidelity to the true all-electron calculations, especially in high accuracy QMC studies. In order to overcome this limitation, we have proposed and constructed a new generation of valence-only Hamiltonians called correlation consistent ECPs (ccECP)[37–40]. The key feature of this new set is the many-body construction of ccECPs from the outset, in particular: (i) we have emphasized and put upfront the accuracy of many-body valence spectra (eigenvalues and eigenstates) as a guiding principle in addition to the well-known norm conservation/shape consistency principles; (ii) we have opted for simplicity, transparency and eventual wide use, in addition to offering several choices of core sizes or even smoothed-out all-electron nuclear Coulomb potentials; (iii) we have used a set of tests and benchmarks such as molecular bonds over a range of distances in order to extensively probe for the quality and transferability of the ccECPs; (iv) we have established reference data sets for the *exact/nearly-exact* atomic total energies, kinetic energies, as well as single-reference and multi-reference fixed-node DMC energies. At present, this covers elements H-Kr with subsequent plans to fill the periodic table.

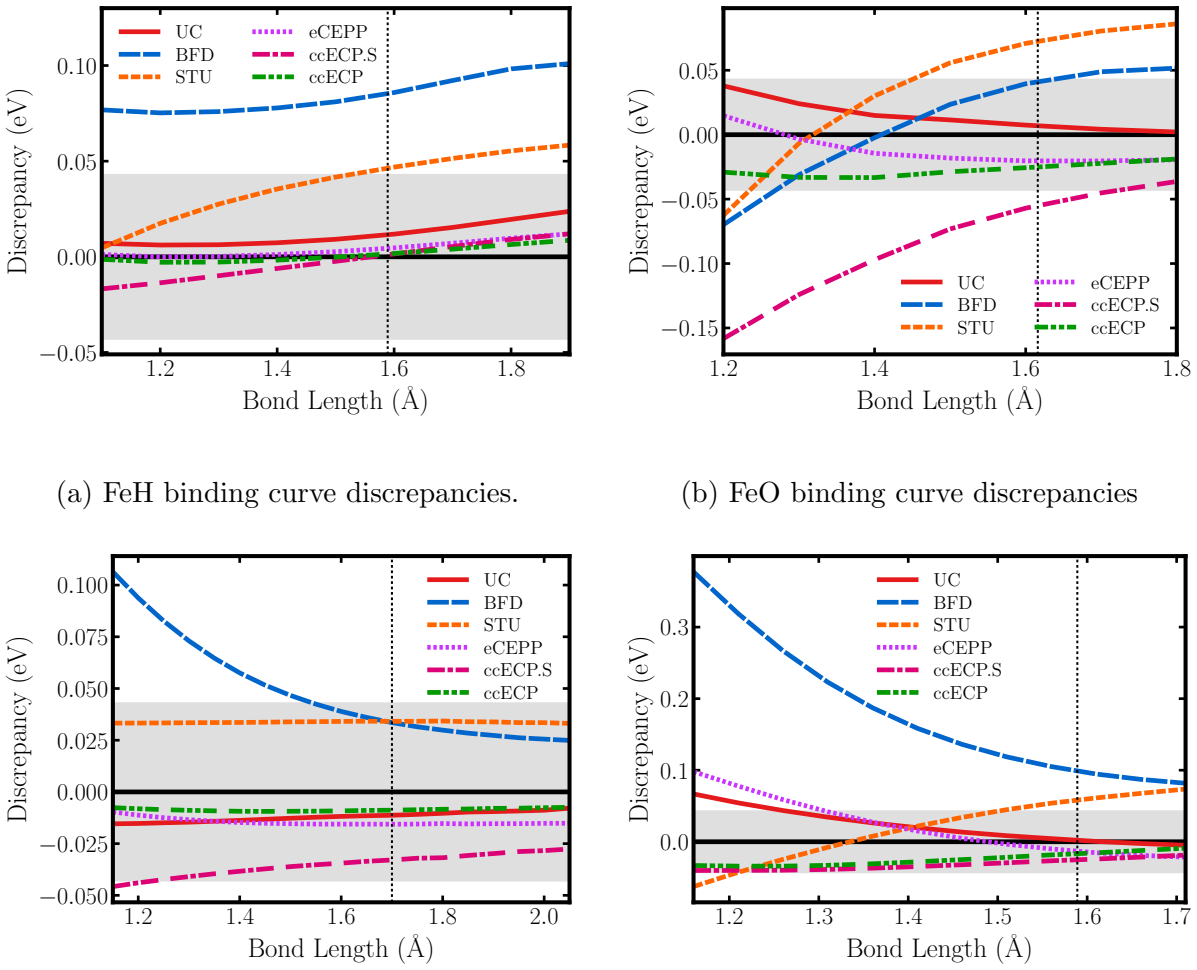
B. ccECP atomic and molecular properties.

The construction of ccECPs builds in electron correlations obtained from the accurate coupled-cluster singles doubles with perturbative triples (CCSD(T)) method. By doing so,

ccECPs achieve very high accuracy and enjoy spectral properties on the valence subspace that are in close agreement with the scalar relativistic all-electron (AE) Hamiltonian. The agreement is often within chemical accuracy over a large range of atomic excitations and ionizations that often spans hundreds of eV energy windows. Molecular properties such as binding energies in multiple geometries, equilibrium bond lengths, and vibrational frequencies were also considered in the development, mostly examining oxides, hydrides, and homonuclear dimers. Especially, compressed bond length properties were given priority as this corresponds to high-pressure applications and probes for the proper behavior of the valence charge in the core region. These atomic and molecular tests provide a direct and comprehensive comparison of ccECP and other core approximations such as BFD[41], STU[42], eCEPP[43], CRENBL[44], SBKJC[45], UC (uncorrelated, self-consistent, all-electron core), and ccECP.S (optimization including only atomic spectrum). Here we illustrate some of these results for selected cases.

Figure 1 shows the molecular binding energy discrepancies for FeH, FeO, VH, and VO molecules relative to all-electron CCSD(T) where we observe that some previous ECPs display significant errors. In addition, Table I, lists a more comprehensive comparison by tabulating the average of mean absolute deviations (MAD) of molecular binding properties relative to all-electron CCSD(T) for all $3d$ transition metal (TM) molecules. Similarly, Figure 2a presents the MAD of a large valence spectrum for all $3d$ TM atoms. In both atomic and molecular tests, we see that ccECP achieves smaller or on par average errors with regard to the other ECPs. In addition, Fig. 1 shows these improvements to be consistent for different elements and varying geometries. Hence, we believe that ccECP accomplishes the best accuracy compromise for atomic spectral and molecular properties. Furthermore, ccECPs are provided with smaller cores than conventionally used ones in some cases where large errors were observed. This includes Na-Ar with [He] core and H-Be with softened/canceled Coulomb singularity at the origin (ccECP(reg)). Selected molecular test results for these are shown in Figure 3.

For reference, we also provide accurate total and kinetic energies for all ccECPs[46] using methods such as CCSDT(Q)/FCI (FCI, full configuration interaction) with DZ-6Z extrapolations to estimate the complete basis set limit. This data, for instance, is useful in the assessment of fixed-node DMC biases. Figure 2b shows the summary of single-reference (HF) fixed-node DMC errors for ccECP pseudo atoms.



(a) FeH binding curve discrepancies.

(b) FeO binding curve discrepancies

(c) VH binding curve discrepancies

(d) VO binding curve discrepancies

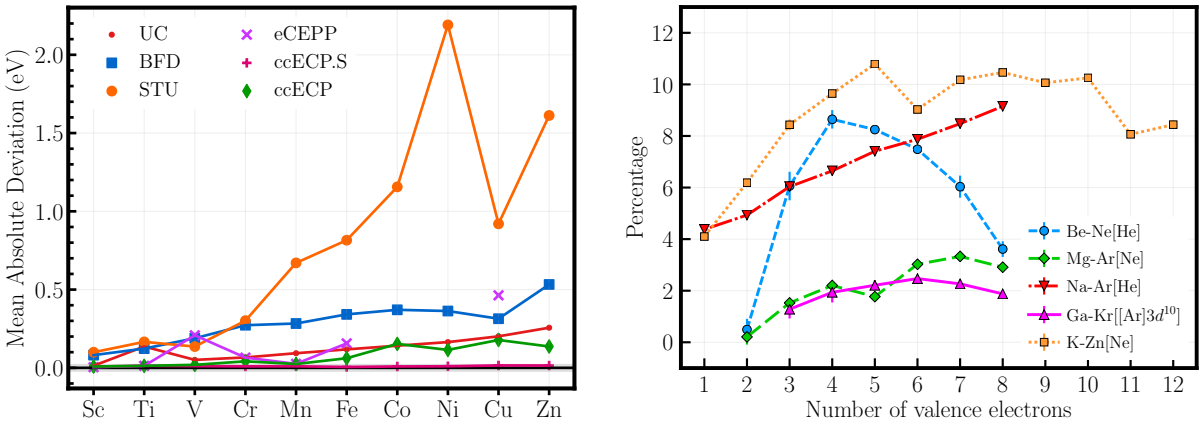
FIG. 1: Binding energy discrepancies for (a) FeH, (b) FeO, (c) VH, and (d) VO molecules. The binding curves are relative to scalar relativistic AE CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction. The ccECPs are the only valence Hamiltonians that are consistently within the shaded region of chemical accuracy including short bond lengths which are relevant for high pressures. Reproduced from Ref. 39, with the permission of AIP Publishing.

C. ccECP Database and Website

In order to facilitate the use ccECPs, we have provided basis sets and a variety of ECP formats available at `https://pseudopotentiallibrary.org`, shown in Figure 4. Each ccECP is presented in a quantum chemistry format for direct use in various codes, including

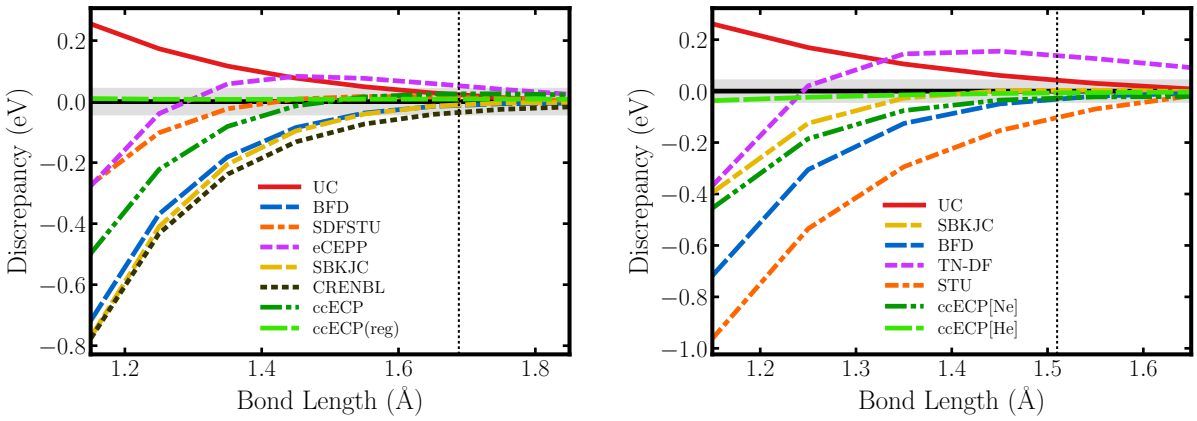
	UC	BFD	STU	eCEPP	ccECP.S	ccECP
D_e(eV)	0.0063(40)	0.0590(41)	0.0380(41)	0.0163(45)	0.0240(40)	0.0104(40)
r_e(Å)	0.0012(13)	0.0064(13)	0.0026(13)	0.0019(15)	0.0027(13)	0.0010(13)
ω_e(cm^{-1})	2.2(5.8)	10.4(5.9)	4.6(5.9)	3.9(6.9)	6.4(5.8)	2.9(5.8)
D_{diss}(eV)	0.021(41)	0.145(41)	0.036(41)	0.032(46)	0.054(40)	0.016(41)

TABLE I: Average MADs of binding parameters for various core approximations with respect to AE data for 3d TM hydride and oxide molecules. All parameters were obtained using Morse potential fit. The parameters shown are dissociation energy D_e, equilibrium bond length r_e, vibrational frequency ω_e and binding energy discrepancy at dissociation bond length D_{diss}. Reproduced from Ref. 39, with the permission of AIP Publishing.



(a) (b)

FIG. 2: (a) MADs for 3d TM benchmark states from bare [Ne] core up to low-lying neutral excitations and the anionic state. (b) Fixed-node DMC biases (ϵ) as a percentage of the correlation energy for ccECP pseudo atoms: $100\epsilon/|E_{corr}|$. T-moves[47] and single-reference trial functions were used in calculations with the exception of Be, B, and C with two-reference form to account for the significant $2s-2p$ near-degeneracy. Fig. 2a reproduced from Ref. 39, with the permission of AIP Publishing. Fig. 2b is adapted with permission from Ref. 46. Copyright 2020 American Chemical Society.



(a) LiO binding curve discrepancies

(b) SiO binding curve discrepancies

FIG. 3: Binding energy discrepancies for (a) LiO and (b) SiO molecules. The uncorrelated core (UC) results, plus those for many different effective core potentials are given relative to scalar relativistic all electron (AE) CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction. Details are given in Refs. 38 and 40. Reproduced from Ref. 40 and Ref. 38, with the permission of AIP Publishing.

Molpro, GAMESS, NWChem, and PySCF which uses the NWChem format. We also provide an XML format which can directly be used in QMCPACK.

In addition to the ccECPs themselves, we have also provided basis sets appropriate for correlated calculations in each code format. Specifically, we have provided Dunning style[48] correlation consistent basis sets from the DZ to 6Z, and in most cases have also provided an augmented version. For use in solid state applications using a plane wave basis, we have also transformed the semi-local potentials into fully nonlocal Kleinman-Bylander potentials[49] using the Unified Pseudopotential Format. This allows the ccECPs to be directly used in codes such as Quantum Espresso. A report file is included giving detailed information about the quality of the Kleinman-Bylander version of the potential and recommended plane wave energy cutoff energies.

D. Status and future developments

The ccECP table and construction principles aim at improved account of systematic errors built into effective valence Hamiltonians in a wide variety of correlated calculations (see

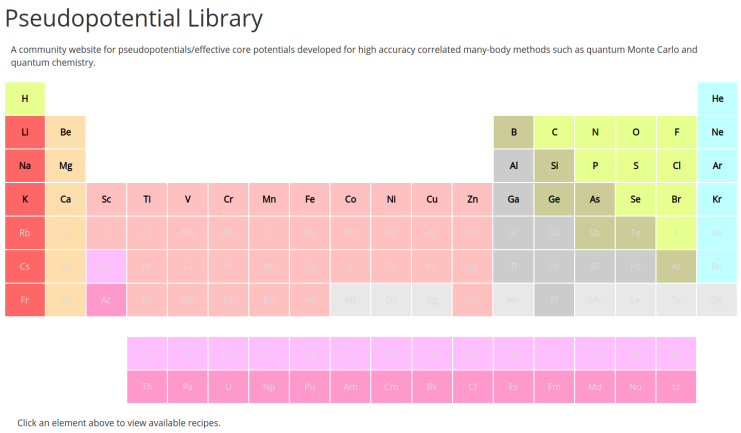


FIG. 4: Pseudopotential Library, https://pseudopotentiallibrary.org.

encouraging feedback so far[50,51]). Further effort is focused on adapting ccECPs for efficient calculations with plane wave basis set (ccECPpw versions). This requires modifying the deep potentials of the late $3d$ elements Fe-Zn in particular. The goal is to enable calculations with plane wave cutoffs not exceeding $\sim$ 600-800 Ry. Plans for the near future involve ccECPs for selected $4d$ and $5d$ elements that include a number of technologically important elements and require explicit treatment of the spin-orbit interactions. Additional improvements such as core polarization and relaxation corrections can be added per specific, application driven needs. Further plans include seeking feedback from the electronic structure community, collecting the data for reference and validation as well as adjustments per need for use in a broad variety of *ab initio* approaches.

V. AUXILIARY FIELD QUANTUM MONTE CARLO

The latest version of QMCPACK offers a now mature implementation of the phaseless auxiliary-field quantum Monte Carlo (AFQMC) method[20,52] capable of simulating both molecular[53,54] and solid state systems[55–57]. AFQMC is usually formulated as an orbital-space approach in which the Hamiltonian is represented in second-quantized form as

$$\hat{H} = \sum_{ij} h_{ij} \hat{c}_i^\dagger \hat{c}_j + \frac{1}{2} \sum_{ijkl} v_{ijkl} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k + E_{II}, \tag{1}$$

$$= \hat{H}_1 + \hat{H}_2 + E_{II}, \tag{2}$$

where $\hat{c}_i^\dagger$ and $\hat{c}_i$ are the fermionic creation and annihilation operators, h_{ij} and v_{ijkl} are the one- and two-electron matrix elements and E_{II} is the ion-ion repulsion energy. Key to an efficient implementation of AFQMC is the factorization of the 4-index electron-repulsion integral (ERI) tensor v_{ijkl}, which is essential for the Hubbard-Stratonovich (HS) transformation[58,59].

QMCPACK offers three factorization approaches which are appropriate in different settings. The most generic approach implemented is based on the modified-Cholesky factorization[60–64] of the ERI tensor:

$$v_{ijkl} = V_{(ik),(lj)} \approx \sum_n^{N_{\text{chol}}} L_{ik}^n L_{lj}^{*n}, \tag{3}$$

where the sum is truncated at $N_{\text{chol}} = x_c M$, x_c is typically between 5 and 10, M is the number of basis functions and we have assumed that the single-particle orbitals are in general complex. The storage requirement is thus naively $\mathcal{O}(M^3)$ although sparsity can often be exploited to keep the storage overhead manageable (see Table II). Note that QMCPACK can accept any 3-index tensor of the form of L_{ik}^n so that alternative density-fitting based approaches can be used. Although the above approach is efficient for moderately sized molecular and solid-state systems, it is typically best suited to simulating systems with fewer than 2000 basis functions.

To reduce the memory overhead of storing the three-index tensor we recently adapted the tensor-hypercontraction[65–67] (THC) approach for use in AFQMC[56]. Within the THC approach we can approximate the orbital products entering the ERIs as

$$\varphi_i^*(\mathbf{r})\varphi_k(\mathbf{r}) \approx \sum_\mu^{N_\mu} \zeta_\mu(\mathbf{r})\varphi_i^*(\mathbf{r}_\mu)\varphi_k(\mathbf{r}_\mu), \tag{4}$$

where $\varphi_i(\mathbf{r})$ are the one-electron orbitals and $\mathbf{r}_\mu$ are a set of specially selected interpolating points, $\zeta_\mu(\mathbf{r})$ are a set of interpolating vectors and $N_\mu = x_\mu M$. We can then write the ERI tensor as a product of rank-2 tensors

$$v_{ijkl} \approx \sum_{\mu\nu} \varphi_i^*(\mathbf{r}_\mu)\varphi_k(\mathbf{r}_\mu) M_{\mu\nu} \varphi_j^*(\mathbf{r}_\nu)\varphi_l(\mathbf{r}_\nu), \tag{5}$$

where

$$M_{\mu\nu} = \int d\mathbf{r} d\mathbf{r}' \zeta_\mu(\mathbf{r}) \frac{1}{|\mathbf{r}-\mathbf{r}'|} \zeta_\nu^*(\mathbf{r}'). \tag{6}$$

To determine the interpolating points and vectors we use the interpolative separable density fitting (ISDF) approach[68–70]. Note that the storage requirement has been reduced to $\mathcal{O}(M^2)$.

For smaller system sizes the three-index approach is preferred due to the typically larger THC prefactors determined by $x_\mu \approx 15$ for propagation and $x_\mu \approx 10$ for the local energy evaluation. The THC approach is best suited to simulating large supercells, and is also easily ported to GPU architectures due to its smaller memory footprint and use of dense linear algebra. Although the THC-AFQMC approach has so far only been used to simulate periodic systems, it is also readily capable of simulating large molecular systems using the advances from Ref. 71.

Finally, we have implemented an explicitly k-point dependent factorization for periodic systems[72]

$$V_{(i\mathbf{k}_k+\mathbf{Q}k\mathbf{k}_k),(l\mathbf{k}_l j\mathbf{k}_l-\mathbf{Q})} \approx \sum_{n}^{n_{\text{chol}}^{\mathbf{Q}}} L_{ik,n}^{\mathbf{Q}\mathbf{k}_k} L_{lj,n}^{\mathbf{Q}\mathbf{k}_l *}, \tag{7}$$

where now i runs over the number of basis functions (m) for k-point $\mathbf{k}_i$ in the primitive cell, $\mathbf{Q} = \mathbf{k}_i - \mathbf{k}_k + \mathbf{G} = \mathbf{k}_l - \mathbf{k}_j + \mathbf{G}'$ is the momentum transfer vector (arising from the conservation of crystal momentum) and $\mathbf{G}, \mathbf{G}'$ are reciprocal lattice vectors. Although explicitly incorporating k-point symmetry reduces the scaling of many operations and the storage requirement by a factor of $1/N_k$ (see Table. II), perhaps the most significant advantage is that it permits the use of batched dense linear algebra and is thus highly efficient on GPU architectures. Note that the THC and k-point symmetric factorization can be combined to simulate larger unit cells and exploit k-point symmetry, however this has not been used to date. We compare the three approaches in Table II and provide guidance for their best use.

In addition to state-of-the-art integral factorization techniques, QMCPACK also permits the use of multi-determinant trial wavefunction expansions of the form

$$|\psi_T\rangle = \sum_{I}^{N_D} c_I |D_I\rangle. \tag{8}$$

We allow for either orthogonal configuration interaction expansions where $\langle D_I|D_J\rangle = \delta_{IJ}$ and also for non-orthogonal multi Slater determinant expansions (NOMSD) where $\langle D_I|D_J\rangle = S_{IJ}$. Orthogonal expansions from complete active space self-consistent field (CASSCF) or selected CI methods allow for fast overlap and energy evaluation through Sherman-Morrison based techniques, and thus do not typically incur a significant slowdown. However, they often require a large number of determinants to converge the phaseless error. NOMSD expansions do not benefit from fast update techniques, but often require orders of magnitude fewer

Method	Memory	Propagation	Energy	Setting	GPU
Dense 3-index	$x_c M^3$	$\mathcal{O}(NM^2)$	$\mathcal{O}(x_c N^2 M^2)$	$M \leq 1000$	Yes
Sparse 3-index	$s x_c M^3$	$\mathcal{O}(NM^2)$	$\mathcal{O}(N^2 M^2)$	$M \leq 2000$	No
THC	$x_\mu^2 M^2$	$\mathcal{O}(NM^2)$	$\mathcal{O}(x_\mu^2 NM^2)$	$M \leq 4000$	No
k-point	$x_c m^3 N_k^2$	$\mathcal{O}(NM^2)$	$\mathcal{O}(x_c m^2 n^2 N_k^3)$	$N_k m \leq 6000$	Yes

TABLE II: Comparison in the dominant scaling behavior of different factorization approaches implemented in QMCPACK. We have included a sparsity factor s which can reduce the computational cost of the three-index approach significantly. For example, in molecular systems the memory requirement is asymptotically $\mathcal{O}(M^2)$ in the atomic orbital basis, whilst for systems with translational symmetry the scaling is in principle identical to that of the explicitly k-point dependent factorization (i.e. $s \leq 1/N_k$) although currently less computationally efficient. We also indicate the current state of GPU support for the different factorizations available in QMCPACK. The THC factorization will be ported to GPUs in the near future. Note that by using plane waves the scaling of the energy evaluation and propagation can be brought down to $\mathcal{O}(N^2 M \log M)$ and $\mathcal{O}(NM \log M)$ respectively. This approach essentially removes the memory overhead associated with storing the ERIs at the cost of using a potentially very large plane-wave basis set[73,74]. This plane wave approach is not yet available in QMCPACK.

determinants than their orthogonal counterparts to achieve convergence in the AFQMC total energy[53] (see Fig. 5).

QMCPACK also permits the evaluation of expectation values of operators which do not commute with the Hamiltonian using the back propagation method[59,81,82]. In particular, the back-propagated one-particle reduced density matrix (1RDM) as well components or contracted forms of the two-particle reduced density matrix are available. As an example we plot in Fig. 6 the natural orbital occupation numbers computed from the back-propagated phaseless AFQMC 1RDM.

Tools to generate the one- and two-electron integrals and trial wavefunctions for molecular and solid state systems are also provided through the afqmctools package distributed with QMCPACK. To date these tools are mostly dependent on the PySCF software package[23], however we provide conversion scripts for FCIDUMP formatted integrals, as well as simple

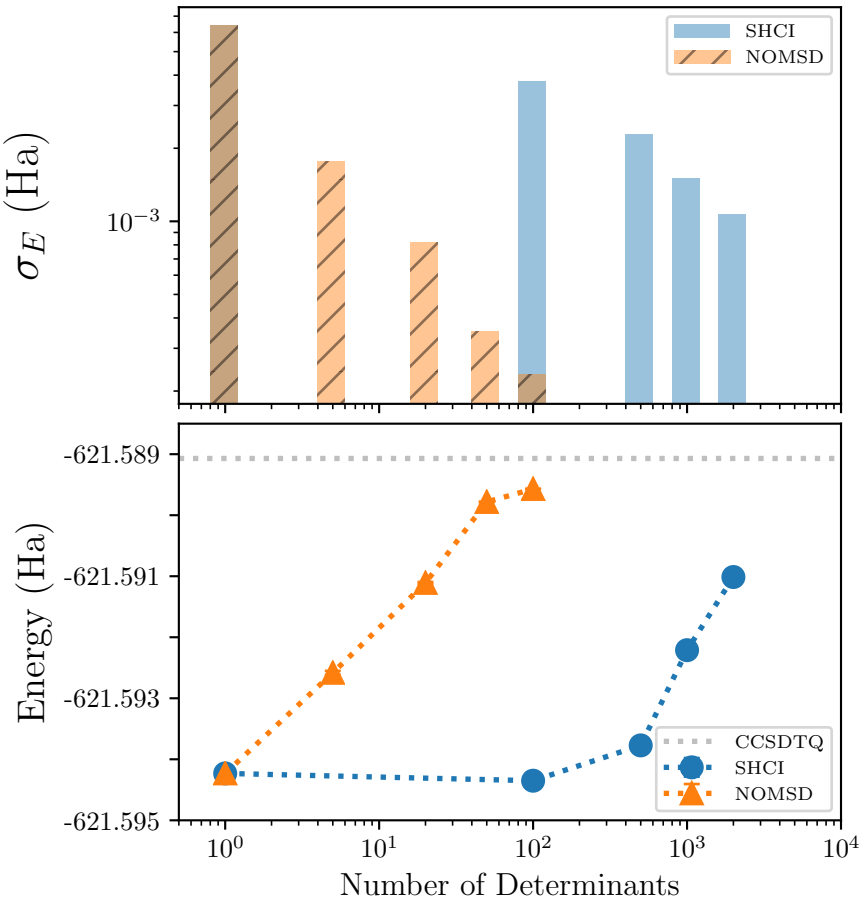


FIG. 5: Comparison in the performance of selected heath-bath configuration interaction (SHCI) and NOMSD as trial wavefunctions in AFQMC calculations of NaCl in the cc-pVDZ basis set at its equilibrium bond length. The top panel demonstrates that smaller NOMSD expansions are necessary to reduce the standard deviation in the energy estimator (σ_E) compared to SHCI trial wavefunctions. The bottom panel shows that the total energy converges more rapidly with determinant number when using a NOMSD trial wavefunction, where the horizontal dashed line is the coupled cluster singles, doubles, triples and quadruples (CCSDTQ) result. The SHCI and NOMSD wavefunctions were generated using the DICE[75,76] and PHFMOL[77–79] packages respectively. The CCSDTQ result was computed using the Aquarius package[80]. Reproduced from Ref.53

Python routines to convert factorized integrals or trial wavefunctions provided from any source to our internal HDF5 based file format. Detailed tutorials on how to run AFQMC in QMCPACK are also provided. Nexus (Sec.III) can be used to drive the process of mean-field

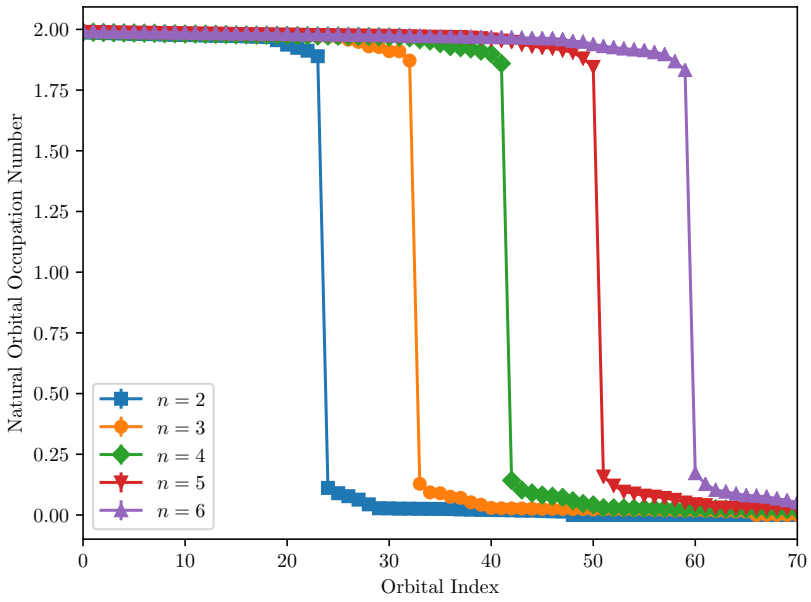


FIG. 6: Phaseless AFQMC natural orbital occupation numbers computed from the back-propagated 1RDM for the n-acenes ($C_2H_4C_{4n}H_{2n}$) in the STO-3G basis set. Geometries are taken from Ref. 83. Error bars are plotted but are smaller than the symbol size.

calculation, wavefunction conversion, and AFQMC calculation. We are using this integration to perform a study of the relative strengths of AFQMC and real space QMC methods.

Over the next year we plan to extend the list of observables available as well as complete GPU ports for all factorization and wavefunction combinations. In addition we plan to implement the finite temperature AFQMC algorithm[84–88], and spin-orbit Hamiltonians with non-collinear wavefunctions. We will also release our ISDF-THC factorization tools and our interface to Quantum Espresso[24]. We hope our open-source effort will enable the wider use of AFQMC in a variety of challenging settings.

VI. TOWARDS SYSTEMATIC CONVERGENCE OF REAL-SPACE QMC CALCULATIONS

The key factor in reaching high accuracy using QMC is the choice of trial wave function Ψ_T. For all-electron DMC calculations the nodes of the wavefunction are the only factor

in determining the error in the computed energy, while the bulk of the wavefunction affects the statistical efficiency and timestep error of the calculation. For calculations involving pseudopotentials, high accuracy of the trial wavefunction is also needed around the atomic cores to minimize the approximations in evaluating the non-local energy. Single Slater determinant (SD) wavefunctions built with Hartree-Fock or Kohn-Sham orbitals supplemented by a Jastrow correlation factor generally give good results, e.g. Ref. 89, and are used almost exclusively today in solid-state calculations.

Reaching systematic convergence of the trial wavefunction and its nodal surface for general systems has been a key challenge for real space QMC methods since their invention. Besides increasing accuracy in calculated properties, this is also required to remove the starting point dependence and allow use of QMC where all potential sources of trial wavefunction are unreliable. In 2008 this was performed for first row atoms and diatomic molecules Ref.[90], and improved algorithms are aiding calculations on larger systems[91]. For general systems with many electrons, the overall challenge remains. Furthermore, if the wavefunction is to be used in DMC, commonly used optimization techniques only optimize the nodal surface indirectly by improving the VMC energy and/or variance. Minimization of the objective function is therefore not guaranteed to minimize the fixed-node energy. Consistently high accuracy wavefunctions are also needed around atomic cores to minimize the locality error in pseudopotential evaluation, posing a challenge for trial wavefunction optimization with a large number of coefficients.

One possible step along the way would be to optimize all the orbital coefficients in a single determinant wavefunction, but due to the limited flexibility in describing the $(3N-1)$ dimensional nodal surface this protocol can not give exact nodes for general systems. This approach could represent a useful starting point independent step, while keeping a simple form for the trial wavefunction. Other possibilities for improving the trial wavefunction while retaining simplicity include techniques such as backflow and iterative backflow[92,93], and antisymmetrized geminal product wavefunctions (e.g. Ref. 94). However, more flexible and complex trial wavefunctions are required to achieve systematic convergence of the nodal surface and to approach exact results for general systems.

The most straightforward method to improve the quality of the trial wavefunction nodes in a convergeable manner is to increase its complexity via a multi-Slater determinant (MSD) or configuration-interaction (CI) expansion:

$$\Psi_T = \sum_{i=1}^{M} c_i D_i^{\uparrow} D_i^{\downarrow} e^J \tag{9}$$

where Ψ_T is expanded in a weighted (c_i) sum of products of up and down spin determinants D_i, and J is the Jastrow correlation factor. In the limit of a full configuration interaction calculation in a sufficiently large and complete basis set, this wavefunction is able to represent the exact wavefunction. However, direct application of configuration interaction is prohibitively costly for all but the smallest systems, because very large numbers of determinants are usually required. To speed application, an efficient selection procedure for the determinants is needed. This can be combined with efficient algorithms for evaluating the wavefunction in QMC.[95–97]

A. Ground state calculations

Multiple variants of selected Configuration Interaction (sCI) methods have recently demonstrated significant success at reaching high accuracy for ground state and excited states of molecular systems with tractable computational cost. Within the class of sCI methods, the CIPSI[98] method has proven to be practical in providing high accuracy wavefunctions for QMC for both molecular systems and for solids[99–105]. sCI methods enable unbiased construction of the trial wavefunction using only a single threshold parameter and therefore avoid the complexities of, for example, CASSCF techniques which require expert selection of the active space. CIPSI algorithms are implemented in the Quantum Package 2.0 code[25] and fully interfaced with QMCPACK and Nexus.

For systems where CIPSI can be fully converged to the FCI limit and reliably extrapolated to the basis set limit, QMC is not required, but for any reasonable number of electrons, QMC can be used to further improve the convergence. The wavefunctions produced from CIPSI can be used either directly, in which case the nodal error is determined by the CIPSI procedure, or used to provide an initial selection of determinants whose coefficients are subsequently reoptimized in presence of a Jastrow function, or used within DMC where the projection procedure will improve on the CIPSI wavefunction. This procedure is equally applicable to solids as well as molecules, provided k-points and their symmetries are fully implemented.

In the following, we illustrate these techniques by application to molecular and solid-state lithium fluoride. In both cases, we use Linear Combination of Atomic Orbitals (LCAO) and different Gaussian basis set sizes to generate the trial wavefunctions. CIPSI energies refer to the variational energy corrected with the sum of energies from second order perturbation theory (PT_2) of each determinant, i.e. $E+PT_2$, at convergence in energy with the number of determinants. Since the sizes of both systems are small enough to reach CIPSI convergence with, for the largest case, less than 5M determinants, for the DMC calculations, the coefficients of the determinants are not reoptimized in presence of a Jastrow function. The cost of DMC with a CIPSI trial wavefunction scales as $\sqrt(N)*(Var_{Ratio})^2$ where N is the number of determinants in the expansion and Var_{Ratio} is the ratio between the variance of a system at 1 determinant and the same system at N determinants ($Var_{Ratio} = \frac{Var_{Ndets}}{Var_{1det}}$). In the case of the molecular systems, Var_{Ratio} varies between 0.7 (for cc-pcVDZ and $N = 1.6M$) and 0.8 (cc-pcVQZ and $N = 5.2M$) or an increase of cost ranging from 620 and 1500 times the cost of a single determinant DMC run. In the case of the solid, Var_{Ratio} varies between 0.83 (for cc-pVDZ and $N = 700k$) and 0.56 (cc-pcVQZ and $N = 9M$) or an increase of cost ranging from 570 and 940 times the cost of a single determinant DMC run. The main difference in the change of cost between the molecular system and solid system is the use of ECPs, reducing significantly the variance of the calculations for both DMC(SD) and DMC(CIPSI)

B. Molecular Lithium Fluoride

Lithium Fluoride is a small molecule for which the multi-determinant expansion and therefore trial wavefunction can be fully converged to the FCI limit using the CIPSI method. Moreover, CCSD(T) calculations of the Vertical Ionization Potential (VIP) are feasible and it's experimental value is known[106], providing reliable reference data. Care has to be taken in the comparison to experimental values, as experimental measurements include intrinsic uncertainties and environmental parameters such as temperature. These effects, including zero-point motion, are not included in our study and might explain the remaining discrepancies between our calculations and the experimental ionization energy value given by Berkowitz et al. Berkowitz, Tasman, and Chupka[106]

While the total energies at each basis set of ground state calculations and cation calculations are different, their trends are identical and we therefore only show figures representing

the ground state. Figure-7 shows the ground state DMC total energies of LiF, computed using various trial wavefunctions; single-determinant such as Hartree-Fock (HF), DFT's PBE0 and B3LYP hybrid functionals and multi-determinant using the converged CIPSI trial wavefunction. CCSD(T) and CIPSI energies are added to the figure for reference and all calculations are performed for 3 basis-sets increasing in size (cc-pCVNZ, N=D,T,Q) and extrapolated to the complete basis-set limit (CBS). The trends of CCSD(T) and CIPSI total energies are in agreement with each other to the CBS limit. CIPSI calculations recover more correlation energy ∼0.24eV for the ground state and ∼0.13 eV for the cation. This is to be expected as CCSD(T) includes singles, doubles and perturbative triples excitations while CIPSI wavefunction includes up to 9th order excitations with more than 70% describing quadruple excitations (10% describing higher order excitations) for both ground and cation states. Interestingly, at the CBS limit, CIPSI total energy converges to the same limit as the DMC(CIPSI) energy while the CCSD(T) converges to the same energy as the SD-DMC energies.

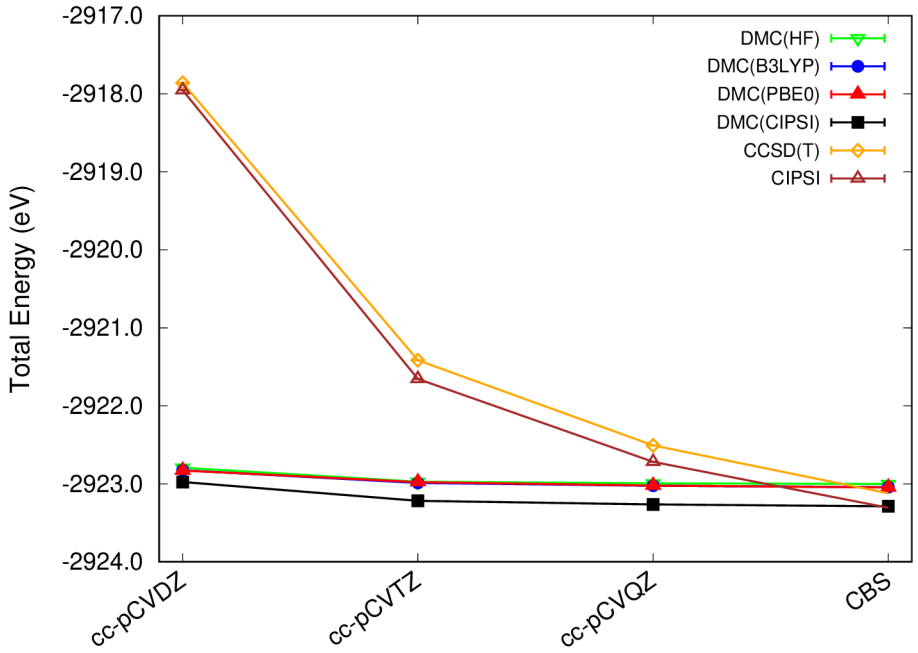


FIG. 7: Ground state total energies (eV) of molecular LiF for DMC(HF), DMC(B3LYP), DMC(PBE0), DMC(CIPSI) CCSD(T) and converged CIPSI using cc-pCVNZ basis sets where N=D,T,Q and extrapolate to the CBS limit. All single-determinant DMC curves are on top of each other.

At the DMC level of theory, the dependence on basis-set is rather weak; less than ∼10 meV in the worst case. In the LiF molecular case, the nodal surface of all tested single-determinant trial wavefunction are within error bars of each other, meaning they are essentially the same. Such weak dependence on the starting method and on the basis set are a a significant advantage and strength of the method when compared to other methods such as sCI or even AFQMC. The use of CIPSI-based trial wavefunctions in DMC allows the recovery of $0.24eV$ for the ground state and $0.5eV$ for the cation. This difference underlines the different sensitivity of the nodal surface to excited and charged states. The vertical ionization potential (VIP), $E_{VIP} = E_{cation} - E_{ground}$, DMC performed using CIPSI wavefunctions shows almost no dependency to the basis set size, Fig. 8 while DMC(CIPSI), CIPSI and CCSD(T) are in perfect agreement at the CBS limit, demonstrating good error compensation for the latter method.

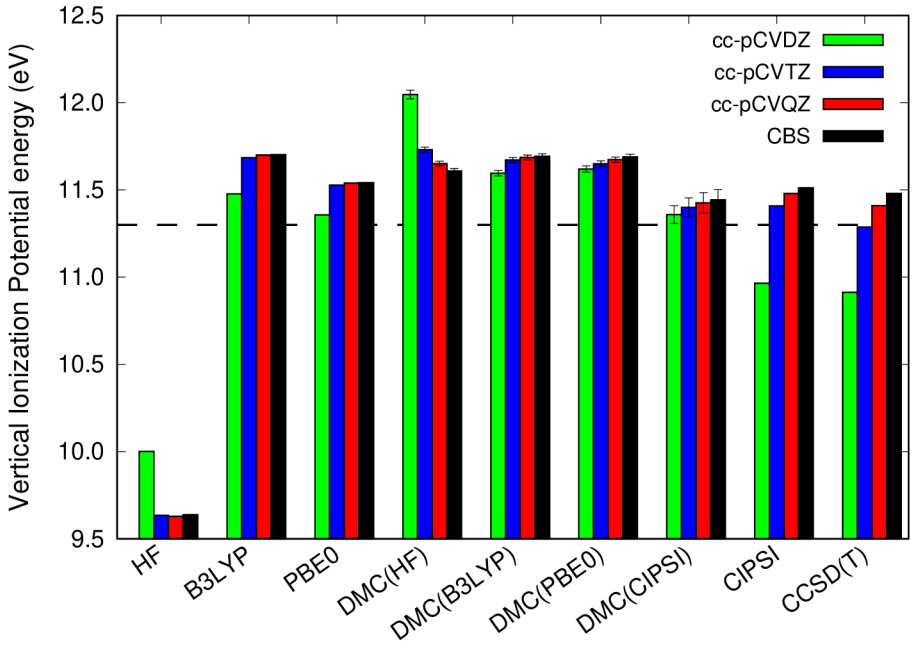


FIG. 8: Vertical ionization potential of LiF using different methods and trial wavefunctions. The dashed line corresponds experiment.[106]

C. Solid-state Lithium Fluoride

Solid LiF is a face centered cubic material with a large gap, used mainly in electrolysis for his role in facilitating the formation of an Li-C-F interface on the carbon electrodes[107]. The purpose of this example is to demonstrate basis set effect and the systematic convergence of DMC energy with the number of determinants (a paper demonstrating convergence to the thermodynamic limit is in preparation). We simulated a cell of $(LiF)_2$ (4 atoms per cell) at the Gamma point using correlation consistent electron-core potentials (ccECP)Bennett *et al.*[37], Wang *et al.*[40] described in Sec-IV and the cc-pVDZ, cc-pVTZ and cc-pVQZ basis set associated with the ccECPs. PySCF, Quantum Package and QMCPACK are able to simulate all shapes of cells with both real and complex wavefunctions, corresponding to any possible k-point. In this case, running at Gamma point is simply for convenience.

For such a small simulation cell, it is possible to convergence the sCI wavefunction to the FCI limit with a reasonable number of determinants, as can be seen in Fig-9. The number of determinants needed to reach approximate convergence remains important: around 700K in cc-pVDZ, 6M in cc-pVTZ and 9M in cc-pVQZ. Similarly to the molecular case, in the converged energies for the cc-pVTZ and cc-pVQZ basis sets are in agreement, indicating that the basis set is sufficiently convergence. Interestingly, in the cc-pVTZ case the DMC energy converges significantly faster with the number of determinants (700K instead of 6M). The slower convergence of the cc-pVQZ curve indicates that important determinants describing relevant static correlations are introduced late in the selection process. Using natural orbitals or in general, an improved choice of orbitals or selection scheme could accelerate the convergence.

D. Solid-state band gap calculations

The band gap of a solid is a critical and fundamental property of a material to predict accurately. QMC calculations for solids have traditionally used completely independent calculations for ground and excited states and single determinant calculations. This approach can be accurate but it relies on good error cancelation between the calculated total energy for each state, making the selection of consistently accurate trial wavefunctions critical. Improved methods are needed to enforce good error cancelation including approaches that can

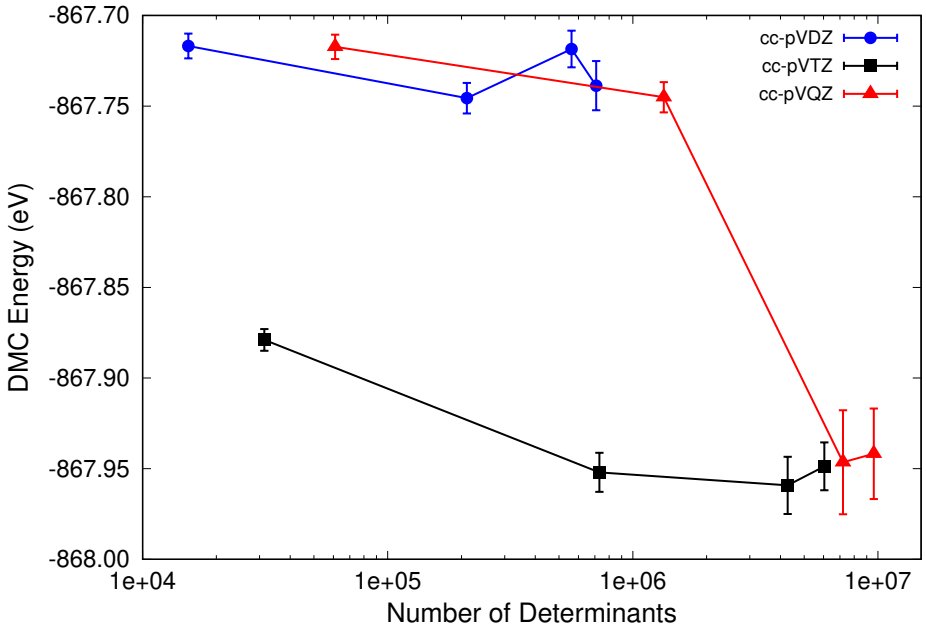


FIG. 9: Convergence of DMC energies of solid LiF using for different basis sets with respect to the number of determinants.

be systematically converged to give in-principle exact results.

As discussed above, convergent wavefunctions and energies can be constructed using sCI techniques. However, even for small primitive cells with relatively uncorrelated electronic structures, this approach quickly requires millions of determinants making it expensive to apply today. We have developed theories, methods and implementations to obtain the band-edge wavefunctions around the fundamental gap and their relative energies efficiently and to a high accuracy. Error cancelation is built into the methodology so that simpler trial wavefunctions are effective and the scheme is substantially more efficient to apply. Surprisingly, for the systems examined so far, only single and double excitations need be considered to obtain accurate band gaps, even using the simple VMC method. This makes the technique comparatively cheap to apply.

To compute the optical band gaps of insulators and semi-conductors we use the energy difference of optimized wave functions that describe the valence band maximum (VBM) and the conduction band minimum (CBM). Optimizations use the recently developed excited

state variational principle,[108–110]

$$\Omega\left(\omega,\Psi\right)=\frac{\langle\Psi|\omega-H|\Psi\rangle}{\left\langle\Psi|(\omega-H)^{2}|\Psi\right\rangle}=\frac{\omega-E}{\left(\omega-E\right)^{2}+\sigma^{2}} \tag{10}$$

whose global minimum is not the ground state but the eigenstate with energy immediately above the chosen value ω, which could be placed within the band gap to target the first excited state and thus predict the optical gap. QMCPACK evaluates Ω via variational Monte Carlo (VMC) method to avoid explicit dealing with the H^2 term, and minimizes it using the linear method. For ground state, we include the closed-shell determinant built from Kohn-Sham (KS) orbitals, plus all single-particle-hole excitations, which represents the leading-order terms of orbital rotation that transforms KS orbitals to the ones that minimizes Ω in the presence of the Jastrow factor. For the excited state, we include all the single-particle-hole excitations as in Bethe-Salpeter equation (BSE)[111] methods as well as selected double excitations to capture the re-polarization of the electron cloud in the vicinity of the exciton. We use the variance of the wavefunctions as a proxy for accuracy, and by varying the number of determinants choose ground and excited states with consistent variance.

We have used this approach in Ref. 112 to study optical gaps of a variety of solids ranging from small-gap semi-conductors to large gap insulators and compare our results to the commonly used GW approach based on many-body perturbation theory (MBPT). As detailed in the supplemental information of Ref. 112 the band gaps were extrapolated using calculations on 8, 16 and 24 atom supercells. Figure 10 shows that the predicted optical gaps are in excellent agreements with experimental values and the mean absolute deviation (MAD) is just 3.5%, compared to MADs more than twice this large for the optical gaps obtained by subtracting the known exciton binding energy from G_0W_0 and self-consistent GW gaps. These calculations were able to run of departmental level computing and did not require supercomputers due to the use of VMC and moderate sized supercells utilized. Remaining errors that have yet to be investigated include the size of the CI expansion and the limited correlation energy obtained with VMC, the pseudopotentials, and residual finite size error. DMC may further improve the VMC results.

In order to further show the method's advantage, we performed a thorough analysis of zinc oxide, a material that is particularly challenging for MBPT.[115] As shown in Figure 10, the perturbative nature of G_0W_0 makes its prediction highly sensitive to the amount of

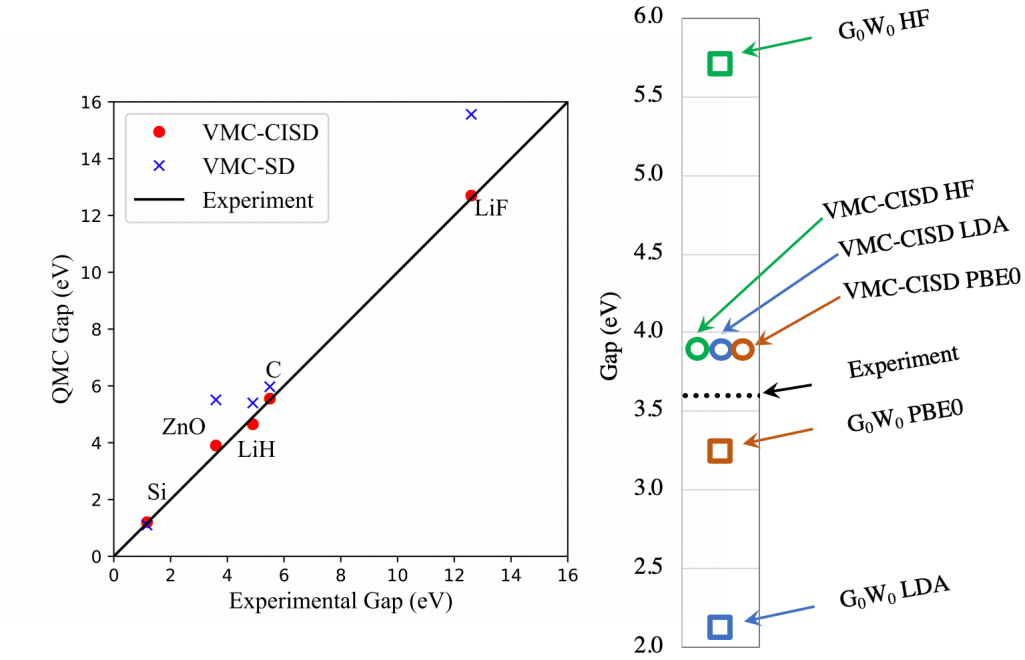


FIG. 10: Left: VMC optical gap predictions plotted against experimental results. The VMC wavefunctions are constructed via a configuration interaction singles doubles expansion (VMC-CISD) or with single determinants (VMC-SD). Right: Comparison of optical gap of ZnO between G_0W_0 using one-particle starting points that employ different fractions of exact exchange with VMC results based on the same starting points. VMC data from Zhao,[112] G_0W_0 results are from Fuchs[113] and experimental data is from Lauck.[114]

exact exchange included in the DFT reference. As G_0W_0 assumes a zeroth-order picture, in which electronic excitations are simple particle-hole transitions between the one-particle eigenstates of DFT, such a sensitivity indicates the break down of this assumption and then G_0W_0 becomes unreliable. On the other hand, our VMC approach is designed to be insensitive to the DFT choice for two reasons: (1) its ability to include approximate orbital relaxation counteracts the shortcomings of the DFT orbitals. (2) unlike G_0W_0 it does not require orbital energies as input. From Figure 10 we do find that its prediction to optical gap is both accurate and independent to the choice of DFT functionals.

E. Summary

Using DMC as a post-sCI method is very promising to systematically improve molecular and solid-state calculations beyond the single-determinant picture. It converges faster the sCI methods used on their own. From a practical perspective, PySCF, Quantum Package and QMCPACK are fully interfaced with each other through the Nexus workflow automation system. The necessary multi-step workflow to run the above examples is fully implemented. In the case of solids, Nexus can automatically manage finite-size scaling calculations by setting the size of the super-cells, the number of twists angles, and drive PySCF, Quantum Package, and QMCPACK appropriately and automatically.

VII. APPLICATIONS

To illustrate the application of recent developments in QMCPACK, in Section VII A we give an example of using real space QMC to study non-valence anions, which are particularly challenging systems. Section VII B gives an example of computing the excited states of localized defects, which is a challenge for all electronic structure methods. In Section VII C we give an example of computing the momentum-distribution and Compton profile from real space methods. The necessary estimators have recently been specifically optimized for these tasks.

A. Applications of DMC to Non-valence Anions

In addition to valence anions, molecules and clusters can posses non-valence anions in which the binding of the excess electron is dominated by a combination of long-range electrostatics and long-range dispersion-type correlation effects. The best known class of non-valence anions are dipole-bound anions, in which the binding of the excess electron is driven by the dipole field of the neutral.[116–121] Non-valence anions, regardless of the nature of the long-range interaction responsible for the electron binding, are challenging to treat using traditional electronic structure methods due to the large, highly diffuse basis sets required. Both DMC and AFQMC methods have been demonstrated to be useful in characterizing dipole-bound anions.[122,123] Non-valence anions in which electron correlation effects dominate the binding of the excess electron pose an additional challenge, namely, by definition,

they do not bind the excess electron in the Hartree-Fock (HF) approximation. In fact, HF calculations on such excess electron systems collapse on to the neutral molecule leaving the excess electron in a discretized continuum orbital.[124] Hence, methods that start from the HF wave function, e.g., MP2 and CCSD(T),[125] also fail to bind the excess electron. Many-body methods such as equation-of-motion coupled cluster (EOM-CC),[126–129] have proven successful in treating these species, but still face the problem of requiring very large basis sets.

This raises the question of whether DMC calculations using a single Slater determinant trial wave function to define the nodal surface can accurately describe non-valence correlation-bound (NVCB) anions. To investigate this, we have undertaken DMC calculations on a $(H_2O)_4$ model with the monomers arranged so that the net dipole is zero.[124] This model has been studied previously using EOM-CC methods. In the present work, we focus on a geometry at which the excess electron does not bind in the HF approximation, but for which EOM-EA-CCSDT[129] calculations using the aug-cc-pVDZ basis set[130,131] augmented with a 7s7p set of diffuse functions located at the center of the molecular cluster give an electron binding energy (EBE) of 174 meV.[124]

For the QMC calculations, Slater-Jastrow trial wave functions that are products of a single Slater determinant comprised of HF or DFT orbitals and a Jastrow factor were employed. All calculations were carried out using the ccECP pseudopotentials with the corresponding aug-cc-pVDZ type basis sets[37,39] augmented with the same 7s7p set of diffuse functions as employed in ref 124. The DMC calculations were carried out at three imaginary time steps (0.001 Ha^{-1}, 0.003 Ha^{-1}, 0.005 Ha^{-1}), and a linear extrapolation was performed to obtain the zero time step limit. HF calculations with this basis set fail to bind the excess electron, and a plot (see Fig. 11) of the singly occupied orbital of the excess electron system reveals that it is very diffuse because it has collapsed onto the lowest "continuum" solution as described by the discrete basis set. The QMC calculations were carried out with the QMCPACK code.

The Jastrow factors used in the trial wave functions included electron-electron, electron-nuclei, and electron-electron-nuclei terms. Normally, one would optimize the parameters in the Jastrow factors separately for the neutral and for the anion. However, this approach would not give meaningful results for an unbound anion, and, as a result, we adopted the strategy of using the Jastrow factor optimized for the neutral in the subsequent DMC

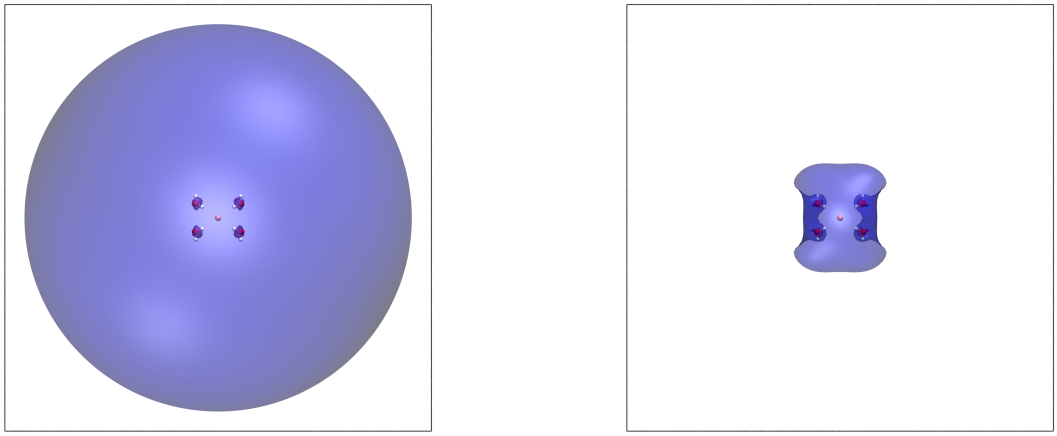


FIG. 11: The singly occupied orbital from (left) HF and (right) B3LYP calculations on the anion of the $(H_2O)_4$ cluster model. These were carried out using the ccECP and the corresponding aug-cc-pVDZ basis set, augmented with a set of $7s7p$ diffuse functions centered at the origin. The bounding box is 100 a.u. on each side, and the isosurface value is set to 0.0005 to enable comparison between the images. The highly diffuse orbital from the HF calculation is actually describing an approximation to a continuum function in the finite Gaussian basis rather than the orbital appropriate for the anion.

calculations of the anion. The DMC calculations using trial wave functions determined in this manner give an EBE of 183±10 meV, in good agreement with the previous EOM-EA-CCSDT result (174 meV).[124] This demonstrates that DMC calculations can recover from the use of a trial wave function for the anion that has collapsed onto a discretized continuum solution.

Even so, there remains the question of whether the EBE obtained from DMC calculations that use an unphysical (i.e., collapsed onto the continuum) trial wave function could incur an appreciable error due to an inadequate description of the nodal surface of the anion. To address this question, we also carried out VMC and DMC calculations on the neutral and anion of the $(H_2O)_4$ cluster model using orbitals from B3LYP[132–134] calculations employing the same pseudopotential and basis sets as used for the HF calculations. The anion is bound by 395 meV in the B3LYP calculations, and the singly occupied orbital of the excess

electron system, while still diffuse, is localized much closer to the molecule than in the HF calculations. A comparison of the charge distributions of the singly occupied orbitals from the HF and DFT calculations (see Fig. 11) shows that the DFT charge distribution is much less spatially extended. Moreover, it more closely resembles the charge distribution of the relevant natural orbital from the EOM calculation of Ref. 124. (The over-binding of the anion in the B3LYP calculations is likely due to the finite extent of the integration grid.) The DMC calculations using trial wave functions derived from B3LYP orbitals give an EBE of 212±11 meV, $\sim$ 29 meV larger than the DMC result obtained using HF orbitals. This increase indicates that employing a trial wave function with a more physical charge distribution for the singly occupied orbital of the anion does have an impact on the nodal surface for the exchange of the electrons. The EOM-EA-CCSDT result obtained using the aug-cc-pVDZ+7s7p basis set is 38meV smaller than the DMC EBE obtained using the trial wave function employing B3LYP orbitals. This suggests that the EBE from EOM calculations may not be fully converged in these large basis sets. Overall, these results demonstrate that DMC is a viable approach for the characterization of NVCB anions.

B. Excitation energies of localized defects

For defects and interfaces, most ab-initio methods can only achieve qualitative agreement on the optical properties. We have recently studied emission energies of Mn^{4+}-doped solids using DMC, which is chosen as proof of principle[11]. We show that our approach is applicable to similar systems, provided that the excitation is sufficiently localized. In support of this work, Nexus scripts and new tutorials on excited state calculations were developed that can be applied to any gapped system (Lab 5 in the QMCPACK manual).

Multivalent ionic defects, such as Mn^{4+}, can create multiple localized electronic states that are trapped within the band gap of wide gap materials. Thus, luminescent centers are created in the dopant sites through radiative recombination. Mn^{4+} has d^3 electronic configuration all on the t_{2g} orbitals. The ground state is in $t_{2g}^{\uparrow\uparrow\uparrow}$ ($^4A_{2g}$) configuration due to Hund's rules, but the excited state is found to be as $t_{2g}^{\uparrow\uparrow\downarrow}$ (^2E_g)[135]. Therefore, the emission energy is simply defined as $E_{em} = E(^2E_g) - E(^4A_{2g})$.

Fig. 12(a) shows that DMC can reproduce experimental emission energies of Mn^{4+} doped insulating host compounds[136–138]. DFT+U and hybrid-DFT, however, substantially under-

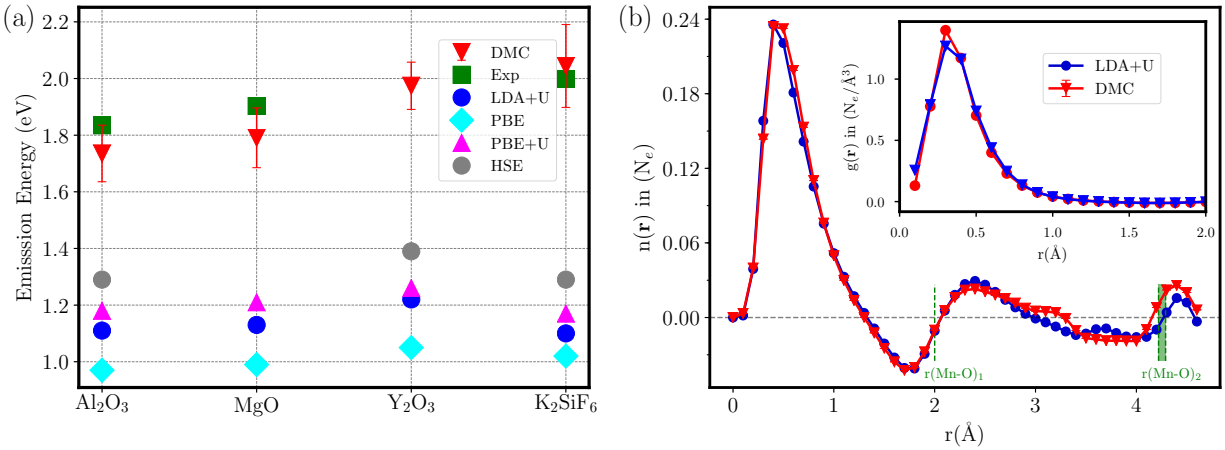


FIG. 12: (a) Emission energies of Mn^{4+} doped host compounds (on x-axis). (b) Number of electrons per radial volume, $n(\mathbf{r})$, around the Mn atom. The inset shows the radial distribution function $g(\mathbf{r})$ for the same density. Number of electrons per radial volume and the radial distribution functions are related as $n(\mathbf{r}) = g(\mathbf{r})4\pi r^2 dr$. Reproduced from Ref. 11.

estimate. Relative quantitative success of HSE with respect to PBE (or DFT+U) indicates that emission energies might be reproduced with a larger portion of exact exchange. However, this would worsen the accuracy of the computed band gaps in the host compounds[11]. In Fig. 12(b), we show the spin flipped electrons per radial volume $n(\mathbf{r}) = \rho^{\uparrow}_{ground} - \rho^{\uparrow}_{excited}$ which is spherically integrated around the Mn^{4+} atom. $n(\mathbf{r})$ approaches to zero with increasing radius indicating that excited electron density is strongly localized on the impurity atom. DMC and LDA+U $n(\mathbf{r})$ densities are almost identical to each other despite the large difference in their emission energies which underscore the difficulties that needed to be overcome for better DFT functionals[139].

C. Calculation of the many-body properties: the momentum distribution

As full-many body methods, QMC can be used to calculate many-body properties that can not be readily obtained from single-particle or mean-field techniques. We have recently updated and optimized calculation of the momentum distribution.

Experimentally the momentum distribution function (MDF) can be accessed, e.g., via angle-resolved photoemission spectroscopy and scattering methods such as Compton scat-

tering, positron annihilation, the $(e, 2e)$ process, and high energy electron scattering[140–143]. In general, the differential cross sections of the scattering can be related to the momentum distribution. These experimental techniques are powerful probes for understanding subtle details in the ground state properties of materials, which are manifested in the MDF.

In normal Fermi liquids, the electron MDF has a discontinuity at the Fermi momentum p_F. In three-dimensional systems this discontinuity defines the shape of the Fermi surface, which is also related to the screening properties of the electrons[144]. The Fermi surface can be extracted from the **p**-space MDF via back-folding[145]. This leads to occupation density within the first Brillouin zone from which the Fermi surface topology can be considered[146].

The magnitude of the discontinuity at the Fermi surface, however, quantifies the strength of a quasiparticle excitation and is generally referred to as the renormalization factor[147,148]. For strongly coupled systems the renormalization factor tends to zero as the coupling strength increases, and thus, it provides an estimate for the strength of electron correlations. Interestingly, the discontinuity at the Fermi momentum disappears for superfluids or superconducting materials. For insulators the discontinuity is absent and the sharp drop is noticeably broadened, which also holds true for some semi-metals[149,150]. Even small scale charge density oscillations lead to clear signatures in the MDF[151]. Therefore, the momentum distribution function provides complementary and informative knowledge to other characterizations of many-body systems.

The MDF, $n(\mathbf{p})$, is obtained by taking the Fourier transform of the one-body density matrix:

$$n(\mathbf{p}) = \frac{N_e}{\Omega} \int dR d\mathbf{r}_1' \; e^{i(\mathbf{r}_1 - \mathbf{r}_1')\cdot\mathbf{p}} \rho(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}, \mathbf{r}_1', \ldots, \mathbf{r}_{N_e}) = \frac{N_e}{\Omega} \int d\mathbf{s} \; e^{-i\mathbf{p}\cdot\mathbf{s}} n(\mathbf{s}),$$

where Ω is the volume containing N_e electrons, $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}\}$, $\mathbf{s} = \mathbf{r}_1' - \mathbf{r}_1$, and

$$n(\mathbf{s}) = \int dR \; \rho(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}, \mathbf{r}_1 + \mathbf{s}, \ldots, \mathbf{r}_{N_e}).$$

In variational Monte Carlo this is expressed as

$$n(\mathbf{s}) = \int dR \; \Psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}) \Psi(\mathbf{r}_1 + \mathbf{s}, \ldots, \mathbf{r}_{N_e}) = \int dR \; |\Psi(R)|^2 \frac{\Psi(R')}{\Psi(R)},$$

where $R' = \{\mathbf{r}_1 + \mathbf{s}, \ldots, \mathbf{r}_{N_e}\}$. Thus, we get for the MDF:

$$n(\mathbf{p}) = \int dR \; |\Psi(R)|^2 \frac{N_e}{\Omega} \int d\mathbf{s} \frac{\Psi(R')}{\Psi(R)} e^{-i\mathbf{p}\cdot\mathbf{s}}.$$

In practice, the Monte Carlo estimate for the MDF with N_s samples is given by

$$n(\mathbf{p}) = \left\langle \frac{1}{\Omega N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_e} \frac{\Psi(R+\mathbf{s}_j^i)}{\Psi(R)} e^{-i\mathbf{p}\cdot\mathbf{s}_j^i} \right\rangle_{|\Psi(R)|^2}, \tag{11}$$

where R includes the coordinates of all the electrons, and $\mathbf{s}_j^i$ is a displacement vector acting on the j^{th} electron of the i^{th} sample. In diffusion Monte Carlo calculations the mixed distribution replaces $|\Psi(R)|^2$, and additional measures must be taken to calculate or estimate the density matrix. Notice that the momentum distribution normalizes to the number of electrons

$$\sum_{\mathbf{p}} n(\mathbf{p}) = N_e = \frac{\Omega}{(2\pi)^d} \int d\mathbf{p}\ n(\mathbf{p}), \tag{12}$$

in which d refers to dimensionality. In Eq. (12) a finite system and a system at the thermodynamic limit are described by summation and integration, respectively.

The MDF estimator is a one-body density matrix based estimator with very high computational cost resulting from the large number of wavefunction evaluations required. A naive implementation can easily double the cost of a QMC calculation. Thus, efficient algorithm and implementations are critical, and similar techniques can be used for related estimators.

In Eq. (11), the computation of wavefunction ratios $\frac{\Psi(R+\mathbf{s}_j^i)}{\Psi(R)}$ and phase factor $e^{-i\mathbf{p}\cdot\mathbf{s}_j^i}$ are both expensive. Direct N_eN_s times of calculation is easy to implement but has a lot of repeated effort. In the $\frac{\Psi(R+\mathbf{s}_j^i)}{\Psi(R)}$ term, the evaluation of single particle orbitals at $\mathbf{r}_j + \mathbf{s}_j^i$ dominates the cost. In fact, its call count can be reduced from N_sN_e to N_s by making

$$\mathbf{s}_j^i = \mathbf{r}_i' - \mathbf{r}_j \tag{13}$$

where $\mathbf{r}_i'$ is the electron coordinates of sample i. Although the leading cost of $\frac{\Psi(R+\mathbf{s}_j^i)}{\Psi(R)}$ is optimized away, its remaining terms still scale as $O(N_sN_e^2)$ and the computational cost should be comparable to the non-local pseudopotential calculation.

Unlike the wavefunction ratios which needs to be computed only once for all the $\mathbf{p}$, phase factors are computed for each $\mathbf{p}$ and also take significant portion of time. Similarly, for each $\mathbf{p}$, the number of evaluations can be reduced to $N_s + N_e$ times by separating indices i and j in two terms like Eq. (13). The calculation of $e^{-i\mathbf{p}\cdot\mathbf{r}_i'}$ and $e^{-i\mathbf{p}\cdot\mathbf{r}_j}$ can be efficiently vectorized using the single instruction multiple data (SIMD) unit in modern processors.

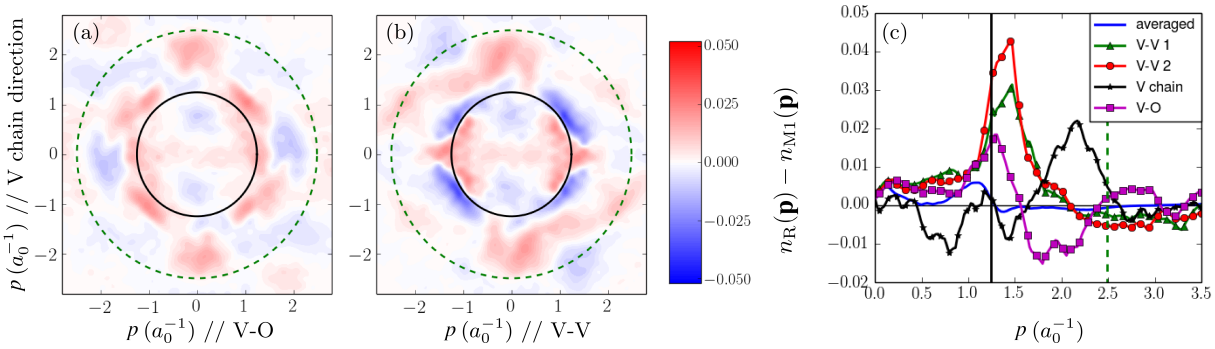


FIG. 13: QMCPACK result for the difference in the momentum distribution of VO_2 across the metal-insulator phase transition based on Ref. 151. (a) and (b) show two planes of the 3D difference profile. In (c) the differences are given in four different directions and also for the angular averaged MDF. For more details, see Ref. 151.

By applying the above techniques and ensuring vectorization of all operations, the overhead for evaluating the MDF in a 48 atom cell VO_2 was reduced from additional 150% to only 50% cost increase compared to a DMC run without any estimators.

Within the so-called impulse approximation (IA) the Compton profile as well as the dynamical structure factor are proportional to the projection of $n(\mathbf{p})$ onto a scattering vector[149,152]. In this case directional Compton profile in z-direction would be expressed as

$$J(q) = \frac{\Omega}{(2\pi)^3} \iint n(p_x, p_y, p_z = q) dp_x dp_y.$$

The IA is especially appropriate for X-ray Compton scattering from electronic systems[142,149], and thus, it is capable of providing a unique perspective for understanding the electronic structure of materials; bulk properties, in particular.

In Ref. 151 QMCPACK was used in obtaining the MDFs and Compton profiles for VO_2 across its metal insulator transition from metallic rutile (R) phase to insulating monoclinic (M1) phase. There the analysis of the MDF shows signatures of the non-Fermi liquid character of the metallic phase of vanadium dioxide. Moreover, findings therein provide an explanation for the experimentally observed anomalously low electronic thermal conductivity[153], which manifests as back scattering characteristics within the momentum distribution function. Fig. 13 shows some examples of MDF differences across the phase transition in two planes as well as for a few different directions[151].

VIII. SUMMARY

We have described recent enhancements to the open source QMCPACK package. Besides increases in capability for both real space and auxiliary field Quantum Monte Carlo (QMC) methods, the surrounding ecosystem has also been improved. These enhancements include the workflow system Nexus, which aims to reduce the complexity of performing research studies and the tens to hundreds of individual calculations that might be entailed. A new set and open database of effective core potentials has also been established at `https://pseudopotentiallibrary.org`, and we expect that these will be of interest for other quantum chemical and many-body calculations due to their increased accuracy, including for stretched bonds and excited states. We have also described how improvements in open software development have benefitted the project. Besides the activities described in this article, we note that there is substantial ongoing work to enhance the architecture of QMCPACK for GPU accelerated machines and to obtain portable performance from a single code base. Once the new design is proven on diverse GPUs it will be described in a future article.

Overall the applicability of QMC continues to expand, it is becoming easier to apply, and there are many systems and phenomena where the higher accuracy and many-body nature of QMC is both warranted and can now be applied. We hope that this article will help encourage these new applications.

DATA AVAILABILITY

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

ACKNOWLEDGMENTS

Methodological development and scientific applications of QMCPACK are currently primarily supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, as part of the Computational Materials Sciences Program and Center for Predictive Simulation of Functional Materials. Software developments focused on future Exascale architectures are supported by the Exascale Com-

puting Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. S.U. and K.D.J. acknowledge the support of NSF grant CHE-1762337. B.R. acknowledges previous support for the work described from the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DEAC52-07NA27344, 15-ERD-013 and NSF grant DMR-1726213. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357. This research also used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International, Inc. for the U.S. Department of Energys National Nuclear Security Administration under Contract No. de-na0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1]F. Becca, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017).

[2]R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons* (Cambridge University Press, 2016).

[3]W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum Monte Carlo simulations of solids," Reviews of Modern Physics **73**, 33–83 (2001).

[4]J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, M. Casula, D. M. Ceperley, S. Chiesa, B. K. Clark, R. C. Clay, 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, "QMCPACK: an open source ab initio quantum monte carlo package for the electronic structure of atoms, molecules and solids," Journal of Physics: Condensed Matter **30**, 195901 (2018).

[5]C. Genovese, T. Shirakawa, and S. Sorella, "The nature of the quadruple chemical bond in the dicarbon molecule," (2019), arXiv:http://arxiv.org/abs/1911.09748v1.

[6]N. Dupuy and M. Casula, "Fate of the open-shell singlet ground state in the experimentally accessible acenes: A quantum Monte Carlo study," The Journal of Chemical Physics **148**, 134112 (2018).

[7]J. G. Brandenburg, A. Zen, M. Fitzner, B. Ramberger, G. Kresse, T. Tsatsoulis, A. Grüneis, A. Michaelides, and D. Alfè, "Physisorption of Water on Graphene: Subchemical Accuracy from Many-Body Electronic Structure Methods," The Journal of Physical Chemistry Letters **10**, 358–368 (2019).

[8]J. Shee, B. Rudshteyn, E. J. Arthur, S. Zhang, D. R. Reichman, and R. A. Friesner, "On Achieving High Accuracy in Quantum Chemical Calculations of 3d Transition Metal-Containing Systems: A Comparison of Auxiliary-Field Quantum Monte Carlo with Coupled Cluster, Density Functional Theory, and Experiment for Diatomic Molecules," Journal of Chemical Theory and Computation **15**, 2346–2358 (2019).

[9]K. S. Qin, T. Ichibha, K. Hongo, and R. Maezono, "Inconsistencies in ab initio evaluations of non-additive contributions of DNA stacking energies," Chemical Physics **529**, 110554 (2020).

[10]J. Yu, L. K. Wagner, and E. Ertekin, "Fixed-node diffusion Monte Carlo description of nitrogen defects in zinc oxide," Physical Review B **95**, 075209 (2017).

[11]K. Saritas, W. Ming, M.-H. Du, and F. A. Reboredo, "Excitation Energies of Localized Correlated Defects via Quantum Monte Carlo: A Case Study of Mn^{4+} Doped Phosphors," The Journal of Physical Chemistry Letters **10**, 67–74 (2019).

[12]B. Busemeyer, G. J. MacDougall, and L. K. Wagner, "Prediction for the singlet-triplet excitation energy for the spinel $MgTi_2O_4$ using first-principles diffusion Monte Carlo," Physical Review B **99**, 081118 (2019).

[13]A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, "A new scheme for fixed node diffusion quantum Monte Carlo with pseudopotentials: Improving reproducibility and reducing the trial-wave-function bias," The Journal of Chemical Physics **151**, 134105

(2019).

[14]B. Mussard, E. Coccia, R. Assaraf, M. Otten, C. J. Umrigar, and J. Toulouse, "Time-Dependent Linear-Response Variational Monte Carlo," in *Novel Electronic Structure Theory: General Innovations and Strongly Correlated Systems* (Elsevier, 2018) pp. 255–270.

[15]K. Doblhoff-Dier, G.-J. Kroes, and F. Libisch, "Density functional embedding for periodic and nonperiodic diffusion Monte Carlo calculations," Physical Review B **98**, 085138 (2018).

[16]M. Dash, J. Feldt, S. Moroni, A. Scemama, and C. Filippi, "Excited States with Selected Configuration Interaction-Quantum Monte Carlo: Chemically Accurate Excitation Energies and Geometries," Journal of Chemical Theory and Computation **15**, 4896–4906 (2019).

[17]S. Azadi and W. M. C. Foulkes, "Efficient method for grand-canonical twist averaging in quantum Monte Carlo calculations," Physical Review B **100**, 245142 (2019).

[18]M. C. Per, E. K. Fletcher, and D. M. Cleland, "Density functional orbitals in quantum Monte Carlo: The importance of accurate densities," The Journal of Chemical Physics **150**, 184101 (2019).

[19]M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Z. and, "Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods," Physical Review X **7**, 031059 (2017).

[20]M. Motta and S. Zhang, "Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method," WIREs Computational Molecular Science **8**, e1364 (2018).

[21]F. Alexander, A. Almgren, J. Bell, A. Bhattacharjee, J. Chen, P. Colella, D. Daniel, J. DeSlippe, L. Diachin, E. Draeger, A. Dubey, T. Dunning, T. Evans, I. Foster, M. Francois, T. Germann, M. Gordon, S. Habib, M. Halappanavar, S. Hamilton, W. Hart, Z. H. Huang, A. Hungerford, D. Kasen, P. R. C. Kent, T. Kolev, D. B. Kothe, A. Kronfeld, Y. Luo, P. Mackenzie, D. McCallen, B. Messer, S. Mniszewski, C. Oehmen, A. Perazzo, D. Perez, D. Richards, W. J. Rider, R. Rieben, K. Roche, A. Siegel, M. Sprague, C. Steefel, R. Stevens, M. Syamlal, M. Taylor, J. Turner, J.-L. Vay, A. F. Voter, T. L. Windus, and

K. Yelick, "Exascale applications: skin in the game," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **378**, 20190056 (2020).

[22]F. Gygi, "Architecture of Qbox: A scalable first-principles molecular dynamics code," IBM Journal of Research and Development **52**, 137–144 (2008).

[23]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, "PySCF: the Python-based simulations of chemistry framework," WIREs Computational Molecular Science **8**, e1340 (2017).

[24]P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," Journal of Physics: Condensed Matter **21**, 395502 (2009).

[25]Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Véril, M. Caffarel, P.-F. Loos, E. Giner, and A. Scemama, "Quantum Package 2.0: An Open-Source Determinant-Driven Suite of Programs," Journal of Chemical Theory and Computation **15**, 3591–3609 (2019).

[26]M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, "General atomic and molecular electronic structure system," Journal of Computational Chemistry **14**, 1347–1363 (1993).

[27]M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," Computer Physics Communications **181**, 1477–1489 (2010).

[28]C. Ortiz, O. Eriksson, and M. Klintenberg, "Data mining and accelerated electronic structure theory as a tool in the search for new functional materials," Computational Materials Science **44**, 1042–1049 (2009).

[29]J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, "The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid," The Journal of Physical Chemistry Letters **2**, 2241–2251 (2011), http://dx.doi.org/10.1021/jz200866s.

[30]S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. J. Mehl, H. T. Stokes, D. O. Demchenko, and D. Morgan, "AFLOW: An automatic framework for high-throughput materials discovery," Computational Materials Science **58**, 218–226 (2012).

[31]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," APL Materials **1**, 011002 (2013).

[32]V. Konkov and R. Peverati, "QMC-SW: A simple workflow for quantum monte carlo calculations in chemistry," SoftwareX **9**, 7–14 (2019).

[33]K. Saritas, T. Mueller, L. Wagner, and J. C. Grossman, "Investigation of a Quantum Monte Carlo Protocol To Achieve High Accuracy and High-Throughput Materials Formation Energies," Journal of Chemical Theory and Computation **13**, 1943–1951 (2017).

[34]J. T. Krogel, "Nexus: A modular workflow management system for quantum simulation codes," Computer Physics Communications **198**, 154–168 (2016).

[35]D. M. Ceperley, "The statistical error of green's function Monte Carlo," Journal of Statistical Physics **43**, 815–826 (1986).

[36]B. L. Hammond, P. J. Reynolds, and W. A. Lester, "Valence quantum Monte Carlo with ab initio effective core potentials," The Journal of Chemical Physics **87**, 1130–1136 (1987).

[37]M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials for correlated calculations," The Journal of Chemical Physics **147**, 224106 (2017).

[38]M. C. Bennett, G. Wang, A. Annaberdiyev, C. A. Melton, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials from correlated calculations: 2^{nd} row elements," The Journal of Chemical Physics **149**, 104108 (2018).

[39]A. Annaberdiyev, G. Wang, C. A. Melton, M. Chandler Bennett, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials from correlated calculations: 3d transition metal series," The Journal of Chemical Physics **149**, 134108 (2018).

[40]G. Wang, A. Annaberdiyev, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials from correlated calculations: 4s and 4p main group elements and first row additions," The Journal of Chemical Physics **151**, 144110 (2019).

[41]M. Burkatzki, C. Filippi, and M. Dolg, "Energy-consistent small-core pseudopotentials for 3d-transition metals adapted to quantum Monte Carlo calculations," The Journal of Chemical Physics **129**, 164115 (2008).

[42]M. Dolg, U. Wedig, H. Stoll, and H. Preuss, "Energy-adjusted ab initio pseudopotentials for the first row transition elements," The Journal of Chemical Physics **86**, 866–872 (1987).

[43]J. R. Trail and R. J. Needs, "Shape and energy consistent pseudopotentials for correlated electron systems," The Journal of Chemical Physics **146**, 204107 (2017).

[44]L. Fernandez Pacios and P. A. Christiansen, "Ab initio relativistic effective potentials with spin-orbit operators. I. Li through Ar," The Journal of Chemical Physics **82**, 2664–2671 (1985).

[45]W. J. Stevens, H. Basch, and M. Krauss, "Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms," The Journal of Chemical Physics **81**, 6026–6033 (1984).

[46]A. Annaberdiyev, C. A. Melton, M. C. Bennett, G. Wang, and L. Mitas, "Accurate Atomic Correlation and Total Energies for Correlation Consistent Effective Core Potentials," Journal of Chemical Theory and Computation **16**, 1482 (2020).

[47]M. Casula, "Beyond the locality approximation in the standard diffusion monte carlo method," Physical Review B **74** (2006), 10.1103/physrevb.74.161102.

[48]T. H. Dunning, K. A. Peterson, and D. E. Woon, "Basis Sets: Correlation Consistent Sets," in *Encyclopedia of Computational Chemistry* (American Cancer Society, 2002).

[49]L. Kleinman and D. M. Bylander, "Efficacious Form for Model Pseudopotentials," Physical Review Letters **48**, 1425–1428 (1982).

[50]T. Wang, X. Zhou, and F. Wang, "Performance of the Diffusion Quantum Monte Carlo Method with a Single-Slater-Jastrow Trial Wavefunction Using Natural Orbitals and Density Functional Theory Orbitals on Atomization Energies of the Gaussian-2 Set," The

Journal of Physical Chemistry A **123**, 3809–3817 (2019).

[51]X. Zhou, H. Zhao, T. Wang, and F. Wang, "Diffusion quantum Monte Carlo calculations with a recent generation of effective core potentials for ionization potentials and electron affinities," Physical Review A **100**, 062502 (2019).

[52]S. Zhang and H. Krakauer, "Quantum Monte Carlo Method using Phase-Free Random Walks with Slater Determinants," Physical Review Letters **90**, 136401 (2003).

[53]E. J. Landinez Borda, J. Gomez, and M. A. Morales, "Non-orthogonal multi-Slater determinant expansions in auxiliary field quantum Monte Carlo," The Journal of Chemical Physics **150**, 074105 (2019).

[54]J. Lee, F. D. Malone, and M. A. Morales, "Utilizing essential symmetry breaking in auxiliary-field quantum monte carlo: Application to the spin gaps of the c_{36} fullerene and an iron porphyrin model complex," arXiv preprint arXiv:2001.05109 (2020).

[55]S. Zhang, F. D. Malone, and M. A. Morales, "Auxiliary-field quantum Monte Carlo calculations of the structural properties of nickel oxide," The Journal of Chemical Physics **149**, 164102 (2018).

[56]F. D. Malone, S. Zhang, and M. A. Morales, "Overcoming the Memory Bottleneck in Auxiliary Field Quantum Monte Carlo Simulations with Interpolative Separable Density Fitting," J. Chem. Theory. Comput. **15**, 256 (2019).

[57]J. Lee, F. D. Malone, and M. A. Morales, "An auxiliary-Field quantum Monte Carlo perspective on the ground state of the dense uniform electron gas: An investigation with Hartree-Fock trial wavefunctions," The Journal of Chemical Physics **151**, 064122 (2019).

[58]J. Hubbard, "Calculation of Partition Functions," Physical Review Letters **3**, 77 (1959).

[59]S. Zhang, J. Carlson, and J. E. Gubernatis, "Constrained path Monte Carlo method for fermion ground states," Physical Review B **55**, 7464 (1997).

[60]N. H. F. Beebe and J. Linderberg, "Simplifications in the generation and transformation of two-electron integrals in molecular calculations," Int. J. Quantum Chem. **12**, 683 (1977).

[61]H. Koch, A. S. de Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions," The Journal of Chemical Physics **118**, 9481 (2003).

[62]F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, and R. Lindh, "MOLCAS 7: The Next Generation," J. Comput. Chem. **31**, 224 (2009).

[63]W. Purwanto, H. Krakauer, Y. Virgus, and S. Zhang, "Assessing weak hydrogen binding on Ca+ centers: An accurate many-body study with large basis sets," The Journal of Chemical Physics **135**, 164105 (2011).

[64]W. Purwanto, S. Zhang, and H. Krakauer, "Frozen-Orbital and Downfolding Calculations with Auxiliary-Field Quantum Monte Carlo," Journal of Chemical Theory and Computation **9**, 4825–4833 (2013).

[65]E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, "Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory," The Journal of Chemical Physics **137**, 044103 (2012).

[66]R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, "Tensor hypercontraction. II. Least-squares renormalization," The Journal of Chemical Physics **137**, 224106 (2012).

[67]E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, "Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions," The Journal of Chemical Physics **137**, 221101 (2012).

[68]J. Lu and L. Ying, "Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost," J. Comput. Phys. **302**, 329 (2015).

[69]W. Hu, L. Lin, and C. Yang, "Interpolative Separable Density Fitting Decomposition for Accelerating Hybrid Density Functional Calculations with Applications to Defects in Silicon," Journal of Chemical Theory and Computation **13**, 5420 (2017).

[70]K. Dong, W. Hu, and L. Lin, "Interpolative Separable Density Fitting through Centroidal Voronoi Tessellation with Applications to Hybrid Functional Electronic Structure Calculations," Journal of Chemical Theory and Computation **14**, 1311 (2018).

[71]J. Lee, L. Lin, and M. Head-Gordon, "Systematically Improvable Tensor Hypercontraction: Interpolative Separable Density-Fitting for Molecules Applied to Exact Exchange, Second- and Third-Order Møller–Plesset Perturbation Theory," Journal of Chemical Theory and Computation **16**, 243–263 (2019).

[72]M. Motta, S. Zhang, and G. K.-L. Chan, "Hamiltonian symmetries in auxiliary-field quantum Monte Carlo calculations for electronic structure," Physical Review B **100**, 045127 (2019).

[73]M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E. J. Walter, "Phaseless auxiliary-field quantum Monte Carlo calculations with plane waves and pseudopotentials:

Applications to atoms and molecules," Physical Review B **75**, 245123 (2007).

[74]F. Ma, S. Zhang, and H. Krakauer, "Auxiliary-field quantum Monte Carlo calculations with multiple-projector pseudopotentials," Physical Review B **95**, 165103 (2017).

[75]S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, "Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory," Journal of Chemical Theory and Computation **13**, 1595–1604 (2017).

[76]A. A. Holmes, N. M. Tubman, and C. J. Umrigar, "Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling," Journal of Chemical Theory and Computation **12**, 3674–3680 (2016).

[77]C. A. Jiménez-Hoyos, T. M. Henderson, T. Tsuchimochi, and G. E. Scuseria, "Projected Hartree–Fock theory," The Journal of Chemical Physics **136**, 164109 (2012).

[78]C. A. Jiménez-Hoyos, R. Rodríguez-Guzmán, and G. E. Scuseria, "Multi-component symmetry-projected approach for molecular ground state correlations," The Journal of Chemical Physics **139**, 204102 (2013).

[79]R. Schutski, C. A. Jiménez-Hoyos, and G. E. Scuseria, "Analytic energy gradient for the projected Hartree–Fock method," The Journal of Chemical Physics **140**, 204101 (2014).

[80]E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton, and J. Demmel, "A massively parallel tensor contraction framework for coupled-cluster computations," Journal of Parallel and Distributed Computing **74**, 3176–3190 (2014).

[81]W. Purwanto and S. Zhang, "Quantum Monte Carlo method for the ground state of many-boson systems," Physical Review E **70**, 056702 (2004).

[82]M. Motta and S. Zhang, "Computation of Ground-State Properties in Molecular Systems: Back-Propagation with Auxiliary-Field Quantum Monte Carlo," Journal of Chemical Theory and Computation **13**, 5367 (2017).

[83]J. Lee, D. W. Small, E. Epifanovsky, and M. Head-Gordon, "Coupled-Cluster Valence-Bond Singles and Doubles for Strongly Correlated Systems: Block-Tensor Based Implementation and Application to Oligoacenes," Journal of Chemical Theory and Computation **13**, 602–615 (2017).

[84]S. Zhang, "Finite-Temperature Monte Carlo Calculations for Systems with Fermions," Physical Review Letters **83**, 2777–2780 (1999).

[85]B. M. Rubenstein, S. Zhang, and D. R. Reichman, "Finite-temperature auxiliary-field quantum Monte Carlo technique for Bose-Fermi mixtures," Physical Review A **86**, 053606 (2012).

[86]Y.-Y. He, M. Qin, H. Shi, Z.-Y. Lu, and S. Zhang, "Finite-temperature auxiliary-field quantum Monte Carlo: Self-consistent constraint and systematic approach to low temperatures," Physical Review B **99**, 045108 (2019).

[87]Y. Liu, M. Cho, and B. Rubenstein, "Ab Initio Finite Temperature Auxiliary Field Quantum Monte Carlo," Journal of Chemical Theory and Computation **14**, 4722–4732 (2018).

[88]Y.-Y. He, H. Shi, and S. Zhang, "Reaching the Continuum Limit in Finite-Temperature Ab Initio Field-Theory Computations in Many-Fermion Systems," Physical Review Letters **123**, 136402 (2019).

[89]M. Dubecký, "Noncovalent Interactions by Fixed-Node Diffusion Monte Carlo: Convergence of Nodes and Energy Differences vs Gaussian Basis-Set Size," Journal of Chemical Theory and Computation **13**, 3626–3635 (2017).

[90]J. Toulouse and C. J. Umrigar, "Full optimization of Jastrow–Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules," The Journal of Chemical Physics **128**, 174101 (2008).

[91]R. Assaraf, S. Moroni, and C. Filippi, "Optimizing the energy with quantum monte carlo: A lower numerical scaling for jastrow–slater expansions," Journal of Chemical Theory and Computation **13**, 5273–5281 (2017).

[92]M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, "Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal fermi liquids," Physical Review B **91** (2015), 10.1103/physrevb.91.115106.

[93]M. Holzmann and S. Moroni, "Orbital-dependent backflow wave functions for real-space quantum monte carlo," Physical Review B **99** (2019), 10.1103/physrevb.99.085121.

[94]M. Casula, C. Attaccalite, and S. Sorella, "Correlated geminal wave function for molecules: an efficient resonating valence bond approach," The Journal of Chemical Physics **121**, 7110–7126 (2004).

[95]P. K. V. V. Nukala and P. R. C. Kent, "A fast and efficient algorithm for Slater determinant updates in quantum Monte Carlo simulations," The Journal of Chemical Physics **130**, 204105 (2009).

[96]B. K. Clark, M. A. Morales, J. McMinis, J. Kim, and G. E. Scuseria, "Computing the energy of a water molecule using multideterminants: A simple, efficient algorithm," The Journal of Chemical Physics **135**, 244105 (2011).

[97]M. A. Morales, J. McMinis, B. K. Clark, J. Kim, and G. E. Scuseria, "Multideterminant Wave Functions in Quantum Monte Carlo," Journal of Chemical Theory and Computation **8**, 2181–2188 (2012).

[98]B. Huron, J. P. Malrieu, and P. Rancurel, "Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions," The Journal of Chemical Physics **58**, 5745–5759 (1973).

[99]A. Scemama, M. Caffarel, A. Benali, D. Jacquemin, and P.-F. Loos, "Influence of pseudopotentials on excitation energies from selected configuration interaction and diffusion Monte Carlo," Results in Chemistry **1**, 100002 (2019).

[100]A. Scemama, A. Benali, D. Jacquemin, M. Caffarel, and P.-F. Loos, "Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes," The Journal of Chemical Physics **149**, 034108 (2018).

[101]A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, "Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction," The Journal of Physical Chemistry A **122**, 2714–2722 (2018).

[102]J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, "Fast semistochastic heat-bath configuration interaction," The Journal of Chemical Physics **149**, 214110 (2018).

[103]M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule," The Journal of Chemical Physics **144**, 151103 (2016).

[104]A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, "Accurate nonrelativistic ground-state energies of 3d transition metal atoms," The Journal of Chemical Physics **141**, 244110 (2014).

[105]E. Giner, A. Scemama, and M. Caffarel, "Using perturbatively selected configuration interaction in quantum Monte Carlo calculations," Canadian Journal of Chemistry **91**, 879–885 (2013).

[106]J. Berkowitz, H. A. Tasman, and W. A. Chupka, "Double-Oven Experiments with Lithium Halide Vapors," The Journal of Chemical Physics **36**, 2170–2179 (1962).

[107]J. Aigueperse, P. Mollard, D. Devilliers, M. Chemla, R. Faron, R. Romano, and J. P. Cuer, "Fluorine Compounds, Inorganic," in *Ullmann's Encyclopedia of Industrial Chemistry* (American Cancer Society, 2000).

[108]L. Zhao and E. Neuscamman, "An Efficient Variational Principle for the Direct Optimization of Excited states," Journal of Chemical Theory and Computation **12**, 3436–3440 (2016).

[109]J. A. R. Shea and E. Neuscamman, "Size Consistent Excited States via Algorithmic Transformations between Variational Principles," Journal of Chemical Theory and Computation **13**, 6078–6088 (2017).

[110]P. J. Robinson, S. D. P. Flores, and E. Neuscamman, "Excitation variance matching with limited configuration interaction expansions in variational Monte Carlo," The Journal of Chemical Physics **147**, 164114 (2017).

[111]G. Onida, L. Reining, and A. Rubio, "Electronic excitations: density-functional versus many-body Green's-function approaches," Rev. Mod. Phys. **74**, 601–659 (2002).

[112]L. Zhao and E. Neuscamman, "Variational Excitations in Real Solids: Optical Gaps and Insights into Many-Body Perturbation Theory," Physical Review Letters **123**, 036402 (2019).

[113]F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, "Quasiparticle band structure based on a generalized Kohn-Sham scheme," Physical Review B **76**, 115109 (2007).

[114]S. Tsoi, X. Lu, A. K. Ramdas, H. Alawadhi, M. Grimsditch, M. Cardona, and R. Lauck, "Isotopic-mass dependence of the A, B, and C excitonic band gaps in ZnO at low temperatures," Physical Review B **74**, 165203 (2006).

[115]B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, "Quasiparticle Band Gap of ZnO: High Accuracy from the Conventional G^0W^0 Approach," Physical Review Letters **105**, 146401 (2010).

[116]E. Fermi and E. Teller, "The Capture of Negative Mesotrons in Matter," Physical Review **72**, 399–408 (1947).

[117]R. F. Wallis, R. Herman, and H. W. Milnes, "Energy levels of an electron in the field of a finite dipole," Journal of Molecular Spectroscopy **4**, 51–74 (1960).

[118]O. H. Crawford, "Bound states of a charged particle in a dipole field," Proceedings of the Physical Society **91**, 279–284 (1967).

[119]J. E. Turner, V. E. Anderson, and K. Fox, "Ground-State Energy Eigenvalues and Eigenfunctions for an Electron in an Electric-Dipole Field," Physical Review **174**, 81–89 (1968).

[120]J. Simons and K. D. Jordan, "Ab initio electronic structure of anions," Chemical Reviews **87**, 535–555 (1987), https://doi.org/10.1021/cr00079a004.

[121]K. D. Jordan and F. Wang, "Theory of Dipole-Bound Anions," Annual Review of Physical Chemistry **54**, 367–396 (2003), https://doi.org/10.1146/annurev.physchem.54.011002.103851.

[122]J. Xu and K. D. Jordan, "Application of the Diffusion Monte Carlo Method to the Binding of Excess Electrons to Water Clusters," The Journal of Physical Chemistry A **114**, 1364–1366 (2010), pMID: 19788288, https://doi.org/10.1021/jp9066108.

[123]H. Hao, J. Shee, S. Upadhyay, C. Ataca, K. D. Jordan, and B. M. Rubenstein, "Accurate Predictions of Electron Binding Energies of Dipole-Bound Anions via Quantum Monte Carlo Methods," The Journal of Physical Chemistry Letters **9**, 6185–6190 (2018), https://doi.org/10.1021/acs.jpclett.8b02733.

[124]V. K. Voora, A. Kairalapova, T. Sommerfeld, and K. D. Jordan, "Theoretical approaches for treating non-valence correlation-bound anions," The Journal of Chemical Physics **147**, 214114 (2017).

[125]K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," Chemical Physics Letters **157**, 479–483 (1989).

[126]J. F. Stanton and R. J. Bartlett, "The equation of motion coupledcluster method. a systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties," The Journal of Chemical Physics **98**, 7029–7039 (1993).

[127]J. F. Stanton and J. Gauss, "Analytic energy derivatives for ionized states described by the equationofmotion coupled cluster method," The Journal of Chemical Physics **101**, 8938–8944 (1994).

[128]M. Nooijen and R. J. Bartlett, "Equation of motion coupled cluster method for electron attachment," The Journal of Chemical Physics **102**, 3629–3647 (1995).

[129]S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, "Coupled-cluster theory for excited electronic states: The full equation-of-motion coupled-cluster single, double, and triple excitation method," The Journal of Chemical Physics **115**, 8263–8266 (2001).

[130]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen," The Journal of Chemical Physics **90**, 1007–1023 (1989).

[131]R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the firstrow atoms revisited. systematic basis sets and wave functions," The Journal of Chemical Physics **96**, 6796–6806 (1992).

[132]A. D. Becke, "Densityfunctional thermochemistry. iii. the role of exact exchange," The Journal of Chemical Physics **98**, 5648–5652 (1993).

[133]C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," Phys. Rev. B **37**, 785–789 (1988).

[134]S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," Canadian Journal of Physics **58**, 1200–1211 (1980).

[135]M. G. Brik and A. M. Srivastava, "On the Optical Properties of the Mn^{4+} Ion in Solids," Journal of Luminescence **133**, 69–72 (2013).

[136]S. Geschwind, P. Kisliuk, M. P. Klein, J. P. Remeika, and D. L. Wood, "Sharp-Line Fluorescence, Electron Paramagnetic Resonance, and Thermoluminescence of Mn^{4+} in -Al_2O_3," Physical Review **126**, 1684–1686 (1962).

[137]K. Dunphy and W. W. Duley, "Multiphoton Excitation of Mn^{4+} and Cr^{3+} Luminescence in MgO," Journal of Physics and Chemistry of Solids **51**, 1077–1082 (1990).

[138]M. G. Brik and A. M. Srivastava, "Ab Initio Studies of the Structural, Electronic, and Optical Properties of K_2SiF_6 Single Crystals at Ambient and Elevated Hydrostatic Pressure," Journal of Electrochemical Society **159**, J212–J216 (2012).

[139]M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, "Density Functional Theory Is Straying from the Path Toward the Exact Functional," Science **355**, 49–52 (2017).

[140]P. Lindner, "Theoretical Momentum Densities," Physica Scripta **15**, 112–118 (1977).

[141]B. G. Williams, "The Experimental Determination of Electron Momentum Densities," Physica Scripta **15**, 92–111 (1977).

[142]M. J. Cooper, "Compton scattering and electron momentum determination," Reports on Progress in Physics **48**, 415–481 (1985).

[143]F. Bell and J. R. Schneider, "Three-dimensional electron momentum densities of solids," Journal of Physics: Condensed Matter **13**, 7905–7922 (2001).

[144]S. B. Dugdale, "Life on the edge: a beginner's guide to the Fermi surface," Physica Scripta **91**, 053009 (2016).

[145]D. G. Lock, V. H. C. Crisp, and R. N. West, "Positron annihilation and Fermi surface studies: a new approach," Journal of Physics F: Metal Physics **3**, 561–570 (1973).

[146]S. B. Dugdale, R. J. Watts, J. Laverock, Z. Major, M. A. Alam, M. Samsel-Czekała, G. Kontrym-Sznajd, Y. Sakurai, M. Itou, and D. Fort, "Observation of a Strongly Nested Fermi Surface in the Shape-Memory Alloy$Ni_{0.62}Al_{0.38}$," Physical Review Letters **96** (2006), 10.1103/physrevlett.96.046406.

[147]S. Huotari, J. A. Soininen, T. Pylkkänen, K. Hämäläinen, A. Issolah, A. Titov, J. McMinis, J. Kim, K. Esler, D. M. Ceperley, M. Holzmann, and V. Olevano, "Momentum Distribution and Renormalization Factor in Sodium and the Electron Gas," Physical Review Letters **105**, 0860403 (2010).

[148]M. Holzmann, B. Bernu, C. Pierleoni, J. McMinis, D. M. Ceperley, V. Olevano, and L. D. Site, "Momentum Distribution of the Homogeneous Electron Gas," Physical Review Letters **107**, 110402 (2011).

[149]R. N. Silver and P. E. Sokol, eds., *Momentum Distributions* (Plenum Press, New York, 1989).

[150]N. Hiraoka and T. Nomura, "Electron momentum densities near Dirac cones: Anisotropic Umklapp scattering and momentum broadening," Scientific Reports **7**, 765 (2017).

[151]I. Kylänpää, Y. Luo, O. Heinonen, P. R. C. Kent, and J. T. Krogel, "Compton profile of VO_2 across the metal-insulator transition: Evidence of a non-Fermi liquid metal," Physical Review B **99**, 075154 (2019).

[152]F. Zambelli, L. Pitaevskii, D. M. Stamper-Kurn, and S. Stringari, "Dynamic structure factor and momentum distribution of a trapped Bose gas," Physical Review A **61**, 063608 (2000).

[153]S. Lee, K. Hippalgaonkar, F. Yang, J. Hong, C. Ko, J. Suh, K. Liu, K. Wang, J. J. Urban, X. Zhang, C. Dames, S. A. Hartnoll, O. Delaire, and J. Wu, "Anomalously low electronic thermal conductivity in metallic vanadium dioxide," Science **355**, 371–374 (2017).