

Scalable Ab Initio Electronic Structure Methods with Near Chemical Accuracy for Main Group Chemistry

Published as part of *The Journal of Physical Chemistry A* virtual special issue “Gustavo Scuseria Festschrift”.

Yujing Wei,[‡] Sibali Debnath,[‡] John L. Weber, Ankit Mahajan, David R. Reichman,
and Richard A. Friesner^{*}



Cite This: *J. Phys. Chem. A* 2024, 128, 5796–5807



Read Online

ACCESS |



Metrics & More

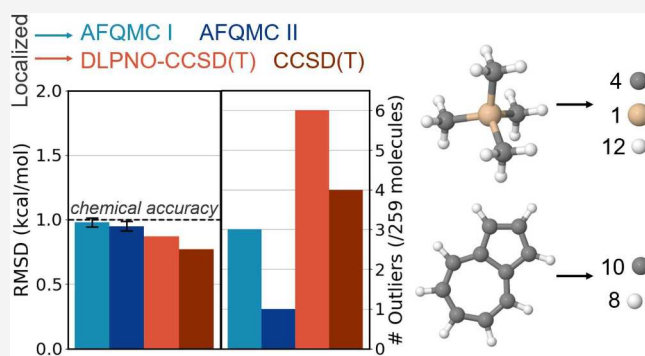


Article Recommendations



Supporting Information

ABSTRACT: This study evaluates the precision of widely recognized quantum chemical methodologies, CCSD(T), DLPNO–CCSD(T), and localized ph-AFQMC, for determining the thermochemistry of main group elements. DLPNO–CCSD(T) and localized ph-AFQMC, which offer greater scalability compared to canonical CCSD(T), have emerged over the past decade as pivotal in producing precise benchmark chemical data. Our investigation includes closed-shell, neutral molecules, focusing on their heat of formation and atomization energy sourced from four specific small molecule data sets. First, we selected molecules from the G2 and G3 data sets, noted for their reliable experimental heat of formation data. Additionally, we incorporate molecules from the W4–11 and W4–17 sets, which provide high-level theoretical reference values for atomization energy at 0 K. Our findings reveal that both DLPNO–CCSD(T) and ph-AFQMC methods are capable of achieving a root-mean-square deviation of less than 1 kcal/mol across the combined data set, aligning with the threshold for chemical accuracy. Moreover, we make efforts to confine the maximum deviations within 2 kcal/mol, a degree of precision that significantly broadens the applicability of these methods in fields such as biology and materials science.



1. INTRODUCTION

Quantum chemical methods have seen significant improvements in accuracy and computational efficiency when applied to the chemistry of main group elements over the last three decades. Density functional methods, which scale formally with system size N as N^4 or N^3 and in practice as N^2 or even N (for very large systems), can routinely be applied to systems with hundreds to thousands of atoms, with the best functionals providing an average unsigned error on the order of 2–3 kcal/mol for atomization energies of small molecules.¹ Alternatively, coupled cluster with perturbative triples (CCSD(T)) has been integrated into numerous quantum chemistry software packages in a robust manner, offering average error rates close to 1 kcal/mol for atomization energies of the same category of small molecules; however, the computational cost scales as N^7 .^{2,3} Benchmark methodologies that incorporate higher-order coupled cluster terms and more elaborate treatment of core–valence interactions are capable of producing results reliably within 1 kcal/mol deviation from experimental values.⁴ However, the high computational demand of these approaches limits their application to very small molecular systems.

While the accuracy that can be achieved with modern DFT approaches is extremely impressive, DFT calculations on large

data sets reveal a significant number of outliers with errors significantly larger than the 2–3 kcal/mol cited above as the average unsigned error. Importantly, outliers can be obscured when large data sets are used and only the average errors are reported.¹ In a subsequent paper, we will examine in detail the outlier distribution obtained for a range of modern functionals when compared to benchmark methods and curated experimental results. For the present purposes it is sufficient to note that more work remains to be done to improve the robustness of DFT approaches across a wide range of chemistry, even for main group molecules. Moreover, systems containing transition metals can be prone to a higher incidence of outliers.⁵

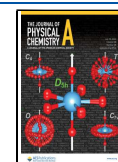
A consequence of the above observation is that high-level wave function-based approaches remain highly relevant in practical applications despite the significantly greater computa-

Received: April 30, 2024

Revised: June 21, 2024

Accepted: June 25, 2024

Published: July 6, 2024



tional cost as compared to DFT calculations. When addressing the grand challenge of understanding the chemistry of complex systems via quantum chemistry, the initial step typically involves conducting a comprehensive set of DFT calculations to explore various possible structures and reaction mechanisms. However, in refining the results to select the correct reaction mechanism (for example) and in general to achieve chemical accuracy, the ability to do benchmark-level wave function-based calculations would be extremely valuable. Furthermore, accurate wave function calculations are the best path forward, via the production of benchmark training data sets, to developing improved DFT methods in which the magnitude and frequency of outliers are substantially diminished.

However, in order for wave function-based methods to effectively address complex systems, an approach is needed which scales better with system size than the N^7 of conventional CCSD(T). The past decade has seen the development of two notable methods that address this need, both leveraging the concept of orbital localization—a technique tracing back to Pulay's work in the 1980s.⁶ The first is localized coupled cluster (e.g., L-CCSD(T)), the most widely used implementation of which is the DLPNO algorithm of Neese and co-workers.^{7–9} The formal scaling of DLPNO-CCSD(T) is $N^{3.10}$ and extremely impressive timing and accuracy numbers over a wide range of systems (and particularly those restricted to main group chemistry) have been published in the past 5 years.^{11–16} DLPNO and related methods (for example other local CCSD(T) methods such as LNO-CCSD(T)¹⁷ and PNO-CCSD(T)¹⁸) represent a revolution in quantum chemical technology as it is now possible to obtain something close to CCSD(T) quality results for systems containing on the order of 100 atoms.

The second approach is auxiliary field quantum Monte Carlo (AFQMC). The AFQMC algorithm was originally developed in the physics community, but it is only in the past 5 years that significant progress has been made in creating a scalable version of the methodology for the *ab initio* study of molecules. There are several different implementations currently in use.^{19–22} In the present paper, we will focus on two of these implementations. The first is a GPU implementation developed in our groups²³ that exploits localized orbitals²⁴ in a fashion similar to that employed in L-CCSD(T), reducing the formal scaling of AFQMC from N^4 to N^3 . We refer to this approach as L-AFQMC. The second^{25,26} is a CPU-based code optimized to enable systematic convergence of the bias in AFQMC energies to near-exact accuracy by using a large number of determinants in the trial function. We designate this implementation as W-AFQMC since it is based on use of the generalized Wick's theorem. As yet, this method has not been formulated in a localized representation, although work in that direction is ongoing.²⁷ To avoid high computational costs, here we employ W-AFQMC to resolve discrepancies identified in DLPNO-CCSD(T) and L-AFQMC data.

While the computational cost scaling of L-AFQMC and L-CCSD(T) with system size is similar, L-CCSD(T) is considerably faster due to a smaller prefactor. The advantage of AFQMC is that it is formally exact in the limit of the exact trial wave function, and in practice, multireference electronic states can often be readily converged due to the ease of utilizing a multiconfigurational trial function.^{26,28–31} This is more crucial for transition metal containing systems than for main group molecules, but there are still main group cases

where AFQMC can achieve demonstrably greater accuracy with scalable trial wave functions.^{32,33} While such trials allow systematic convergence of the bias in phaseless AFQMC (ph-AFQMC), this accuracy comes at the cost of greater computational expense. Designing protocols for generating trials that strike a desired balance between accuracy and cost is an active area of research. Based on our testing with the benchmark sets, we present two approaches with different cost-accuracy trade-offs.

We believe that having two scalable benchmark methodologies with distinct theoretical frameworks offers substantial benefits. These benefits extend not only to the generation of data for evaluating and parametrizing DFT functionals but also to their direct application to challenging systems, like the manganese cluster in Photosystem II. As an example, in one of our recent papers we investigated reactions of organolithium systems relevant to lithium ion batteries.³⁴ We performed DLPNO-CCSD(T) and L-AFQMC calculations to look at both reaction energies and barrier heights. The agreement between the two approaches was remarkably good (a few kcal/mol) across the various reactions that we investigated. We were therefore able to settle on benchmark numbers and use those to evaluate many different DFT functionals, discovering that only a few were able to reproduce the benchmark results reliably. We were then able to use the preferred functionals in computing energies for a large set of organolithium cluster geometries, which we then utilized in parametrizing a machine learning force field (MLFF) for carrying out simulations of lithium ion battery electrolytes.³⁵ While there was no reason, *a priori*, to doubt the performance of DLPNO-CCSD(T) for these systems, there is very little experimental data available for comparisons, and the validation by a second independent benchmark approach provided a much higher degree of confidence in the results than would otherwise have been possible.

In the present paper, our goal is to provide an assessment of the accuracy of both DLPNO-CCSD(T) and L-AFQMC for main group chemistry atomization energies. We have chosen to focus on atomization energies because (a) a relatively large and reliable data set of benchmark experimental and theoretical values is available for a range of small molecules and (b) atomization energies are one of the most difficult properties for electronic structure methods to compute to high precision, due to the large changes in correlation energy upon atomization, and the relatively minimal cancellation of error. We combine four data sets: the G2^{36,37} and G3^{38,39} sets of Pople and co-workers, which claim to have experimental atomization energies that are accurate to 1 kcal/mol or better, and the W4-11⁴⁰ and W4-17⁴¹ (TAE - total atomization energy) data sets of Karton, Martin and co-workers, which employ very high level (and hence expensive) theoretical methods to achieve the same level of reliability. The W4-17 set, the latest iteration of the W4 sets, is an extension of the W4-11 set, as is G3 an extension of G2 with the addition of larger molecules. W4-17, however, is restricted to molecules with no more than 8 heavy atoms. On the other hand, the G3 data set contains larger molecules than the W4-17 heavy atom limit and therefore provides a test of how the quantum chemical methods perform for larger systems. There are (as far as we have been able to ascertain) no cases where the errors in the W4-11 and W4-17 results exceed the proposed error bars. In contrast, we have had to update a number of the G2 and G3 experimental values with more recent benchmark values.

Our objective with regard to accuracy is to limit the outliers to a maximum of 2 kcal/mol deviation from the experimental or W4 theoretical values. While 2 kcal/mol is not what is generally meant by “chemical accuracy” (that terminology is conventionally reserved for a 1 kcal/mol accuracy level), it is likely to be sufficient accuracy for choosing among alternative reaction mechanisms in complex systems or parametrizing new functionals. An examination of the details of the W4–17 approach suggests that it is going to be difficult if not impossible (at least with current computing capabilities) to construct a *scalable* approach that reliably achieves 1 kcal/mol precision. Our belief is that the maximum 2 kcal/mol level of error that we are aiming for will be good enough not only to analyze chemical reactions in complex systems but also for designing novel chemistries to address a variety of problems in biology and materials science. This is superior to any current DFT functional, where the average errors of the best functional for the current data set are in the range of 2–3 kcal/mol, but the maximum errors are on the order of 8–10 kcal/mol (as we will enumerate in a subsequent publication).

The paper is organized as follows. In Section 2, we discuss the data sets including our updating of a number of the reference values in the G2 and G3 sets. We then discuss computational methods including our scalable formulation of AFQMC, DLPNO–CCSD(T), and CCSD(T) methodology (enabling comparisons for the larger systems), basis sets, core–valence corrections, scalar relativistic corrections, extrapolation to the complete basis set limit, and treatment of atomic energies. Section 3 presents the results for all three methods as compared to the relevant experimental or theoretical benchmarks and discusses their implications. In Section 4, we conclude by summarizing our results and discussing future directions. In general, the performance of both DLPNO–CCSD(T) and L-AFQMC are quite robust, with only a few apparent outliers above our targeted 2 kcal/mol threshold. We investigate these outliers using more accurate trials with the W-AFQMC method. This enables us to identify outliers arising from the approximations in the methodology, as opposed to cases that are most likely errors or uncertainties in a few of the experimental reference data. Using this targeted convergence of ph-AFQMC, we are able to produce high-quality atomization energies while minimizing cost over a large data set.

2. METHODS

2.1. Data Sets. In this study, we have limited our selection to closed-shell, neutral molecules, excluding carbenes. Future work will investigate open-shell systems and ions. In total, 116 molecules are selected from G2 set and 73 molecules from the G3 set, respectively. Moreover, the W4 sets also contain significant overlap with molecules from the G2 and G3 sets. After removing duplicates, we are left with 38 molecules from the W4–11 set and 32 molecules from the W4–17 set. The G3 and W4–17 extensions consist of generally larger molecules. Therefore, for the purpose of this study, such separation into “G2”, “G3”, “W4–11”, and “W4–17” is constructive, with the former two sets using experimental heat of formation as reference and the latter two using W4 theory as reference. In total, we have compiled a list of 259 unique molecules across all data sets, with the full list of molecules and the data set into which each molecule is sorted given in the SI Section 1.

2.2. Reference Values. Reference values for the W4 sets are high-level theoretical atomization energies at 0 K, excluding

zero-point energy (ZPE) (labeled as the property TAEe by Karton et al.⁴¹). The reference values for the G2 and G3 sets are experimental heats of formation at 298 K, and the same as those used by Curtiss et al.^{36,38} However, there are some exceptions where we have found conflicting values, as summarized in Table 1, where each of the sources of the

Table 1. New vs Old Reference Values for Heat of Formation^a

molecule	data set	previous $\Delta_f H(298\text{ K})$	updated $\Delta_f H(298\text{ K})$	source of update
AlCl ₃	G2	-139.7 ± 0.7	-142.0	W4
AlF ₃	G2	-289 ± 0.6	-290.7	W4
CCl ₂ CCl ₂	G2	-3.0 ± 0.7	-5.1 ± 0.2	ATcT
CF ₂ CF ₂	G2	-157.4 ± 0.7	-161.1 ± 0.1	ATcT
CH ₂ CH-CN	G2	43.2 ± 0.4	44.7 ± 0.2	ATcT
cyclobutene	G2	37.4 ± 0.4	38.5 ± 0.1	ATcT
cyclopropene	G2	66.2 ± 0.6	67.8 ± 0.1	ATcT
LiF	G2	-80.1	-81.45 ± 2	CCCBDB
vinyl chloride	G2	8.9 ± 0.3	5.2 ± 0.06	ATcT
azulene	G3	69.1 ± 0.8	73.6	CCCBDB
benzoquinone	G3	-29.4 ± 0.8	-28.7	CCCBDB
tetramethylsilane	G3	-55.7 ± 0.7	-51.7 ± 0.5	ATcT

^aThe reference values in the column ‘previous $\Delta_f H(298\text{ K})$ ’ are the same as used in the original G2/G3 test set papers.^{37,39} All values are reported in kcal/mol. Where the source reports an experimental uncertainty, we have included the uncertainty in the table along with the value.

updated reference is listed. The reported reference values from ATcT^{42,43} postdate the G2 and G3 papers. Furthermore, ATcT collates the most recent experiments and theory from various sources using a self-consistent approach and numerous different reactions⁴⁴ and is readily updated.⁴⁵ Therefore, where ATcT data is available and conflicting with the reference used by Curtiss et al., we instead use the ATcT value. In addition, for the cases of AlCl₃ and AlF₃, we found disagreement between the experimental heats of formation and the W4 reported atomization energy. Moreover, the heat of formation values of these molecules are not present in the ATcT database. Thus, for the updated reference values of these molecules, we have converted the W4 atomization values into heats of formation at 298 K using temperature corrections from DFT (refer to Section 2.9 for details). For the three cases of LiF, azulene, and benzoquinone, where no ATcT or W4 value is available, we use the references reported by the NIST database CCCBDB⁴⁶ instead of those reported by Curtiss. As shown in Table 1, the two sources also give conflicting heats of formation for these three molecules. All of our benchmark wave function methods yield results that are within 2 kcal/mol of the latter source, rather than the former reference values. Although this choice is not based on information about the reference alone, CCSD(T) and AFQMC offer independent evaluations of the experimental data, especially in the case of conflict. In particular, we converge AFQMC with respect to the number of determinants for these cases using W-AFQMC. We refer the reader to Section 2.9 for the method of obtaining deviations against reference values.

2.3. Phaseless AFQMC Formulation. Provided that an initial state $|\phi_i\rangle$ has a nonzero overlap with the exact ground state of a system $|\phi_0\rangle$, then the ground state can be projected from any such trial state as

$$|\phi_0\rangle \propto \lim_{\tau \rightarrow \infty} e^{-\tau \hat{H}} |\phi_i\rangle \quad (1)$$

where τ is the imaginary time, and $\hat{H}$ is the electronic Hamiltonian under the Born–Oppenheimer approximation, which can be written as a sum of one-electron and two-electron terms,

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = \sum_{pq} h_{pq} c_p^\dagger c_q + \frac{1}{2} \sum_{pqrs} V_{pqrs} c_p^\dagger c_q^\dagger c_s c_r \quad (2)$$

h_{pq} are one-electron integrals and $V_{pqrs} = \langle pr | qs \rangle = \langle pq | rs \rangle$ (chemists' and physicists' notation, respectively) are two-electron integrals. Numerically, we propagate

$$|\phi(\tau + \Delta\tau)\rangle = e^{-\Delta\tau \hat{H}} |\phi(\tau)\rangle \quad (3)$$

where $|\phi(0)\rangle = |\phi_i\rangle$. The one-body and two-body terms in $\hat{H}$ can be split using the Trotter–Suzuki decomposition,

$$e^{-\Delta\tau(\hat{H}_1 + \hat{H}_2)} \approx e^{-\Delta\tau \hat{H}_1/2} e^{-\Delta\tau \hat{H}_2/2} e^{-\Delta\tau \hat{H}_1/2} + O(\Delta\tau^3) \quad (4)$$

which introduces an error that scales with the time step. The Hubbard–Stratonovic transformation and the phaseless approximation (see below) also induce time step errors. We show in SI, Section 2 that for this work, the time step error converges at around $\Delta\tau = 0.005 \text{ Ha}^{-1}$ for frozen-core calculations, and at around 0.001 Ha^{-1} for all-electron calculations. We note that in the calculation of AFQMC energy differences, there is some approximate cancellation of time-step errors.

The two-body operators can be decomposed as the sum of the square of one-body operators through Cholesky decomposition or density fitting. The Hubbard–Stratonovich transformation then converts an exponential with two-body operators into a multidimensional integral over fluctuating auxiliary fields, x_α

$$e^{-\Delta\tau(\sum_\alpha \hat{t}_\alpha^2)/2} = \prod_\alpha \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x_\alpha^2/2} e^{\sqrt{\Delta\tau} x_\alpha \hat{t}_\alpha} dx_\alpha + O(\Delta\tau^2) \quad (5)$$

and we arrive at

$$|\phi(\tau + \Delta\tau)\rangle = e^{-\Delta\tau \hat{H}} |\phi(\tau)\rangle = \int d\mathbf{x} p(\mathbf{x}) \hat{B}(\mathbf{x}) |\phi(\tau)\rangle \quad (6)$$

where $p(\mathbf{x})$ is a Gaussian probability density function and $\hat{B}(\mathbf{x})$ is a one-body propagator depending on the auxiliary fields $\mathbf{x}$. This multidimensional integral is evaluated using Monte Carlo importance sampling to obtain a stochastic representation of the wave function. For a more in-depth description of AFQMC, we refer the reader to these review articles.^{47,48}

Due to the Fermionic sign problem, the signal-to-noise ratio generally decays exponentially during the imaginary time propagation. It is possible to eliminate the sign problem using a constraint at the expense of a bias in the resulting energies. In this work, we use a constraint referred to as the phaseless approximation (ph-AFQMC), where the phase of the walkers is restricted according to a trial wave function. The bias induced by the trial wave function can be systematically reduced by improving its quality, for example, by increasing the active space or number of determinants included in the trial. The bias is formally zero in the limit of the exact trial (see the next subsection, Section 3.3, and SI, Section 3 for trial details). We refer the reader to our previous work²⁴ for our approach to localized ph-AFQMC (L-AFQMC) which involves compressing the electron repulsion integrals in the localized orbital

basis. Effectively, the scaling for the energy evaluation, the steepest scaling step with system size, is $N^2M + N_{\text{det}}N^2$ (with a prefactor depending on the compressed electron repulsion integral tensor), where M is the number of basis functions, N is the number of electrons, and N_{det} is the number of determinants. See SI, Section 4 for an estimate of localization error. For the practical deployment of L-AFQMC, we have developed two protocols, AFQMC 0 and AFQMC I, which are discussed in detail below as well as in Section 3.3. AFQMC I is a scalable AFQMC protocol (scaling $\sim N^3$) that achieves an accuracy comparable to that of DLPNO–CCSD(T), albeit with a significantly larger prefactor for the computation time. AFQMC 0 uses a black-box but less elaborate trial function, but is less accurate, particularly for molecules with significant multireference character.

We also present results for a selected subset of molecules computed with another implementation of AFQMC.^{25,26} This method, which uses a generalized Wick's theorem approach to efficiently evaluate energies with multideterminantal trials, will be referred to in what follows as W-AFQMC. The advantage of this approach is that it scales as $MN_{\text{det}} + N^2M^2$, which allows the use of a larger number of determinants in the trial wave function at an accessible computational cost. It enables one to converge the phaseless bias to the near-exact limit in a given basis set by increasing the number of active space orbitals and determinants included in trial wave functions. We use W-AFQMC to calculate energies of the outliers obtained from AFQMC I using on the order of 10,000 determinants and refer to the results in which the L-AFQMC outliers are replaced by the W-AFQMC results as AFQMC II (along with a select few other cases, see SI, Section 3). This approach helps us to more confidently address the question, discussed in detail below, as to whether the AFQMC I outliers are due to the phaseless bias or more likely the result of errors in the reference data. We also run apparent DLPNO–CCSD(T) and CCSD(T) outliers with AFQMC II to assess their status.

2.4. AFQMC Trial Generation. In this work, we use a procedure to generate multideterminant HCI (heat bath configuration interaction) and HCISCF trials for the entire data set. The CAS family of trials (in this work, we use the HCI solver^{49,50}) provides a robust way of including static correlation in the reference of AFQMC and has been shown^{19,28} to generally perform more accurately than single determinant trials. Nonetheless, the selection of active space for these methods is nontrivial. Akin to multireference perturbation theory methods, a common practice is to use an active space, usually minimal, based on chemical intuition and to pick the leading determinants from the expansion in this active space. However, this does not systematically provide chemically accurate energies. Here, we generally follow a two-step process to select the active space in a relatively automated way that can be applied to large data sets. First, HCI is performed on a “valence” active space, selected based on the atomic composition of the molecule. See SI, Section 3 and main text Section 3.3 for the considerations for selecting this active space. Using the spatial 1-RDM of the resulting state, we calculate the natural orbitals and their occupation numbers (NOONs) in the natural orbital basis. We choose a subset of active orbitals from this set based on a NOON threshold (δ) as $\delta \leq \text{NOON} \leq 2 - \delta$. This procedure is often used to flag the more correlated group of orbitals in quantum calculations. A second HCI calculation (or HCISCF for W-AFQMC trials, refer to the SI, Section 3) is then performed with the second

smaller active space, and this forms the final trial wave function. Generally, we choose the number of determinants necessary to retain 99.5% of the CI weight from this final trial wave function unless indicated otherwise. Briefly, AFQMC 0 is fully automated and uses L-AFQMC, and thresholds are chosen to be loose, and for the outliers with AFQMC 0, AFQMC I combines trials with larger active spaces, and AFQMC II in turn combines trials with even larger active spaces run with W-AFQMC. This progression gives some indication of the number of determinants required to converge each molecule, but not much more than necessary. For more details, refer to the discussion in Section 3 and SI, Section 3.3.

All L-AFQMC trials were generated with the PySCF package⁵¹ where we obtain the Hamiltonian and electron repulsion integrals. The HCI trials are obtained using Dice^{49,50,52} in conjunction with PYSCF. L-AFQMC energy is measured in blocks of size 0.1 Ha⁻¹ of 20 timesteps each (each time step being 0.005 Ha⁻¹). In total, we propagate for between 2000 and 3000 such blocks for each molecule, with 1920 total walkers. W-AFQMC calculations used the same time-step of 0.005 Ha⁻¹ for molecules containing only first row atoms, and used a time step of 0.0025 Ha⁻¹ for those with heavier atoms. For W-AFQMC, we use 250 walkers and propagate for 1000 blocks of 50 timesteps each.

2.5. CCSD(T) and DLPNO-CCSD(T). CCSD(T) and DLPNO-CCSD(T) calculations are carried out using the ORCA package⁵³ using restricted Hartree-Fock (RHF) as the reference state. DLPNO-CCSD(T) correlation energies are extrapolated to the complete pair natural space (PNO) using the procedure in Altun et al.⁵⁴ and employing TightPNO thresholds,⁵⁵ between T_{CutPNO} thresholds of 10⁻⁶ and 10⁻⁷ for each basis set used. The matching auxiliary basis set is used if available, otherwise, the AutoAux⁵⁶ functionality is used. Where linear dependence is encountered with AutoAux, we increase the even-tempered expansion factor for the s-shell from 1.8 to 2.0.

2.6. Basis Sets. We use the following basis sets: aug-cc-pVXZ-DK (X = T, Q) for atoms with atomic number less than or equal to that of oxygen, and aug-cc-pCVXZ-DK (X = T, Q) for fluorine and heavier, obtained using the Basis Set Exchange database.⁵⁷⁻⁶³ This choice is motivated by the documented improvement of basis set convergence when using core-valence or tight-d functions in the basis set for second-row elements (Na-Cl)⁶⁴⁻⁶⁷ as well as fluorine.⁶⁸ While the split-valence aug-cc-pVXZ basis sets are not designed for core-valence correlation,⁶⁹ we find that all-electron calculations using these same basis sets can reach a respectable (especially for CCSD(T)) albeit overall inferior accuracy (especially for AFQMC) to the frozen core calculations supplemented with MP2 core corrections, as discussed briefly in Section 3, and in more in detail in SI, Section 5.

We extrapolate all single point energies to the complete basis set (CBS) limit according to the method of Neese and Valeev⁷⁰ for T/Q extrapolation for both Hartree-Fock and correlation energy, with α and β matching the basis set used. We use the same coefficients α and β for the core-valence sets aug-cc-pCVXZ-DK as the corresponding aug-cc-pVXZ-DK basis sets, where we use the aug-cc-pVXZ coefficients reported by Neese and Valeev. This CBS procedure is used by all the methods investigated, with the exception of the more expensive CCSD(T) and W-AFQMC where we use alternative schemes (see SI, Sections 3 and 6).

2.7. Frozen Core Corrections. Frozen core calculations are carried out according to freezing no orbitals for H-Be, 1s orbitals for B-Mg, and 1s and 2p orbitals for Al-Ar. We correct the core-valence energy using MP2,

$$\Delta_{\text{CV}} = E_{\text{CC-MP2}}^T - E_{\text{FC-MP2}}^T \quad (7)$$

where CC and FC denote core-correlated and frozen-core calculations, respectively. We used the aug-cc-pCVTZ-DK basis set for both calculations. Note that we freeze the 1s electrons in second-row atoms even in the CC calculations. In the following discussion, we focus on the frozen core calculations (along with the MP2 core correction above) for all four methodologies. All electron results for CCSD(T), DLPNO-CCSD(T), and L-AFQMC are presented and compared with the frozen core results in Tables S9 and S10 of the Supporting Information. In general we do not see any deterioration in accuracy from the use of the frozen core, and recommend that this approach be used going forward for both AFQMC and CCSD(T) based calculations.

2.8. Relativistic Effects. Scalar relativistic effects are included through the DKH2⁷¹⁻⁷³ Hamiltonian for DLPNO-CCSD(T) and CCSD(T), and X2C⁷⁴ for ph-AFQMC, as X2C is not implemented in ORCA and DKH2 is not implemented in PySCF. The MP2 core-valence corrections follow the same relativistic corrections for each method, respectively (MP2 is carried out in ORCA for correcting DLPNO-CCSD(T) and CCSD(T) corrections and in PySCF for ph-AFQMC corrections). The difference between DKH2 and X2C energies are shown to be small for HF and MP2 and mostly canceled out by atomic energies (see SI, Section 7). Spin-orbit corrections to atomic energies are applied using the values from Curtiss et al.³⁶

2.9. Heat of Formation and Atomization. Atomization energies ($\sum E_{\text{atoms,g}} - E_{\text{molecule,g}}$) and heats of formation are calculated according to the method in ref.,^{36,75,76} including geometry optimization and thermochemical properties (ZPE, enthalpy, internal energy) using the DFT functional B3LYP^{77,78} and basis set 6-31G*⁷⁹⁻⁸² using the Jaguar software package⁸³ with the maximum available grid point density. A few molecules required a higher-level geometry optimization (see SI, Section 8). After optimization, DFT single-point energies were calculated with Jaguar. We note that the first step toward obtaining the heat of formation at 298 K is the atomization at 0 K, and the molecule temperature corrections (from DFT), atom temperature corrections (from experiment) and energy to change the atomic states from gas to standard state (from experiment, i.e. heat of formation of the single atom in the gaseous state) are added to achieve the heat of formation. Experimental values for the atomic heats of formation and temperature corrections are the same as that used by Curtiss et al.,³⁶ with the exception of the sulfur atom heat of formation where we use 66.18 kcal/mol from ATcT⁴² instead of 65.66 kcal/mol.

We emphasize here that as opposed to atomization energy, the heat of formation is defined using standard states of atoms rather than the gas phase state. Nonetheless, these quantities are closely related and in the analysis we convert the W4 atomization energy at 0 K to heat of formation,

$$\Delta_f H(0\text{K}) = \sum_{\text{atoms}} \Delta_f H_{\text{atoms}}(0\text{K}) - \Delta_a E(0\text{K}) \quad (8)$$

where $\Delta_a E(0\text{ K})$ is the atomization energy at 0 K. With the addition of aforementioned temperature correction terms from DFT ($H^{298\text{ K}} - H^0\text{ K}$ for the molecule) and experiment ($H^{298\text{ K}} - H^0\text{ K}$ for the atom), we obtain $\Delta_f H(298\text{ K})$. Although these corrections are approximate (even though we expect the errors to be small), they cancel out when obtaining the deviation from the converted W4 heat of formation reference value as the same corrections are applied to the calculated atomization energy of the molecule. Effectively, the deviation D becomes

$$D = \Delta_f H_{\text{expt}}(298\text{ K}) - \Delta_f H_{\text{calc}}(298\text{ K}) \quad (9)$$

or

$$D = -(\Delta_a E_{\text{W4}}(0\text{ K}) - \Delta_a E_{\text{calc}}(0\text{ K})) \quad (10)$$

for G2/G3 and W4–11/W4–17 sets, respectively, where the only effect of the conversion of the W4 reference values to heat of formation is a change in sign of the deviation from the reference atomization energy. This simple change ensures we are comparing the same quantities.

2.10. Atomic Energies. The treatment of atoms is essential in achieving the targeted accuracy. An explicit near-exact treatment of core–valence correlation on par with valence correlation, as is done in W4–17, requires expensive core–valence corrections. In AFQMC, similar to other projection QMC approaches, the description of core–valence correlations requires onerous convergence of time-step errors. Furthermore, it has been shown that AFQMC atom energies can be difficult to calculate.³³ Hence, we converge the atomic energies with a large number of determinants in W-AFQMC. A less expensive and simpler alternative is to fit the values of the atoms to the experimental data, which benefits from cancellation of errors. We use this approach for the other benchmark methods. For small molecules, a relatively inexpensive version of the first approach can be shown to work quite well, however, for larger molecules, any difference in molecule versus atom accuracy for a given method is compounded.

Atom energies for all methods are fit as free parameters according to the combined G2/G3 set experimental heats of formation (with the addition of AlCl, AlF, AlH, and AlH₃ in the W4–11 using the converted reference values to heat of formation); for each respective method we obtain a separate least-squares multivariate linear regression fit with atomic energies as parameters, calculated heats of formation as dependent variables, and the loss function is the experimental vs calculated heats of formation,

$$\Delta_f H_{\text{expt,molecule}} - \Delta_f H_{\text{calc,molecule}} = \sum_{\text{atom}} N_{\text{atom}} E_{\text{atom}} + c_{\text{molecule}} \quad (11)$$

where $\Delta_f H_{\text{expt,molecule}}$ and $\Delta_f H_{\text{calc,molecule}}$ are the experimental and calculated heats of formation for that molecule, and N_{atom} is the number of the atom constituting the molecule. The energy of the atom, E_{atom} , is the fitting variable, and c_{molecule} are the constant terms (for each molecule, i.e. independent of E_{atom}) such as the temperature terms, energy of the molecule, and heat of formation of the atoms, used when we minimize the left-hand side of the equation above. We have 189 such correlated equations for the 189 molecules in the G2 and G3 sets.

The initial guess for atom energies in the atomic energy fit is the *ab initio* atom energies obtained through each method, respectively. It is worth noting that a small error (i.e., slight imperfection in cancellation of error between atom and

molecule energy) in the *ab initio* atom energies is multiplied by the number of atoms in the molecule. While we do add spin–orbit corrections to atomic energies using the values from Curtiss et al.,³⁶ this only applies to the initial guess and such atom-related corrections will be encompassed in the final fitted atom energy.

These fitted atom energies are used to calculate the atomization energies for the W4 set. Alternatively, using W-AFQMC, we show that by using around 10^4 determinants in a natural orbital active space the deviation from experiment is near-chemical accuracy without fitting atomic energies. However, we do not expect cancellation of error between molecules and atoms in this case, as we expect both molecular and atomic energies to be close to exact within this method. For the present, for methods that are not asymptotically exact like W-AFQMC, we thus recommend the atom fitting approach, even though (i) the efficacy depends on the accuracy of a method across the entire data set where fitting occurs, (ii) the data set must be sufficiently large, and (iii) the resultant atom energies are specific to the other conditions of the fit (basis set and electronic structure method). The third point also applies to *ab initio* atom energies but to a less tailored extent. The reader is referred to SI, Section 9 for the resulting fitted atomic energies.

3. RESULTS AND DISCUSSION

3.1. Overall Performance of CCSD(T) and AFQMC-Based Methods. We begin by evaluating the overall performance of the four methods discussed in Section 2: CCSD(T), DLPNO–CCSD(T), AFQMC I, and AFQMC II, across our chosen data sets. Figure 1 displays the root-mean-square deviation (RMSD) in the enthalpies of formation calculated using these methods, as detailed in Section 2. For the combined data set (represented by the far-right bar in Figure 1), the RMSD values for all methods fall within 1 kcal/

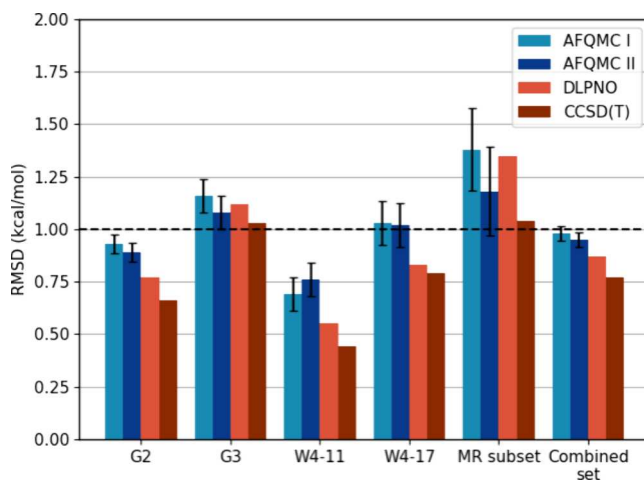


Figure 1. Root-mean-square deviations of the calculated heat of formation of G2, G3, and W4–11/17 data sets with respect to experiment (G2 and G3) or W4 reference values (W4–11, W4–17) for each benchmark method. “DLPNO” refers to DLPNO–CCSD(T). The number of molecules in our mutually exclusive definition of G2, G3, W4–11, and W4–17 are 116, 73, 38, and 32. The separated MR subset refers to the 10 molecules from Table S26, from a combination of G2, W4–11, and W4–17. The combined set consists of the total 259 molecules. The horizontal dashed line at 1 kcal/mol refers to the standard of chemical accuracy.

mol, and as we will see below, all four methods have a very small number of outliers with deviations from benchmark experiments or computations greater than 2 kcal/mol. We conclude that for main group chemistry, the localized (and hence scalable) versions of both coupled cluster and AFQMC achieve our target of reliably obtaining near-chemical accuracy for chemical transformations, sufficient for elucidating chemical reaction mechanisms in complex systems. Results along these lines have been presented previously for DLPNO–CCSD(T) (although not for as large and extensively curated a data set involving experimental and high-level theoretical references), but not for AFQMC. The present exercise establishes AFQMC as a robust alternative benchmark quantum chemical methodology, albeit at a higher computational cost than DLPNO–CCSD(T). For the present systems, we find the scaling exponent with system size to be similar between DLPNO–CCSD(T) and L-AFQMC, and the high prefactor is responsible for the cost of L-AFQMC. See SI, Section 10 for examples of the computational costs of each method. We refer the reader to more detailed demonstrations of AFQMC and DLPNO–CCSD(T) scaling in these articles.^{10,24,27}

With regard to the detailed results in Figure 1, a few comments can be made. First, full CCSD(T) displays the smallest RMS error across all four methods. This is most pronounced for the W4–11 data set, which is not surprising as the benchmark theory used to establish reference values is based on a coupled cluster approach. For the G3 data set, the difference is barely noticeable, reflecting likely performance when comparing with experimental data in practice.

Second, the most difficult data set for all methods is, unsurprisingly, but not guaranteed, the subset of 10 cases that we have identified as “multireference” (MR). We classify molecules as multireference based on a set of diagnostics developed by Karton et al.,^{40,41} as discussed in more detail in Section 11, Supporting Information. For the coupled-cluster based methods, only one molecule, ozone, stands out as displaying an error in excess of 2 kcal/mol. Despite formally being a single reference methodology, the treatment of electron correlation via CCSD(T) appears to be powerful enough to handle many wave functions with nontrivial multireference character. DLPNO–CCSD(T) here displays a noticeable (although not large) degradation from full CCSD(T). For AFQMC, an improvement is obtained in the treatment of MR molecules by upgrading the trial function in the AFQMC II approach. Details of results for each quantum chemical method for all of the MR cases can be found in the SI, Section 11.

We next analyze the outliers observed across the different data sets and methods, the number of which are summarized in Figure 2 below. A number of interesting points can be made regarding the outliers, which are enumerated along with results for the various computational methods in Table 2. First, neither AFQMC I, AFQMC II nor the two coupled cluster based methods have any outliers for the W4–11 and W4–17 data sets, where the reference values are taken from ultrahigh level computation. Second, while CCSD(T) and DLPNO–CCSD(T) have a larger number of outliers than AFQMC II based on our (somewhat arbitrary) cutoff for the deviation from the reference of 2 kcal/mol, a perusal of the data in Table 2 shows that, with the exception of ozone (where we believe that the multireference character is great enough to cause significant errors in CCSD(T) and DLPNO–CCSD(T),

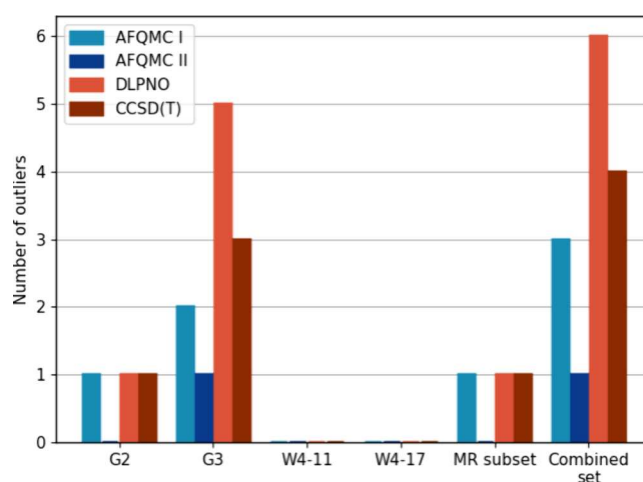


Figure 2. Number of outlier molecules for each method and data set, where outlier is defined as having a larger than 2 kcal/mol deviation in the heat of formation from the reference value. “DLPNO” refers to DLPNO–CCSD(T).

which can then be reduced via a large trial function in AFQMC II), the computational results for the five remaining molecules are closer to each other than they are to the experimental reference data. A likely interpretation of the results is that the experiments have a residual error of a few kcal/mol (possibly as large as 3–4 kcal/mol for 3-butyne-2-one), and that in fact, the reliability of our scalable benchmark methods is higher than what is suggested in Figure 2. In contrast to a number of other cases that initially appeared to be outliers but were resolved by newer experiments, as discussed in Section 2.2, we were unable to find any relatively recent experimental data. Note that the value of having two distinct computational methods which can be compared, suggested in the Introduction, is already manifested in this analysis. Without the complementary AFQMC I and AFQMC II results, one might conclude that CCSD(T) has occasional outliers even for single-reference main group molecules and that higher-order treatments are required to achieve even the 2 kcal/mol accuracy threshold that we have set.

Having summarized the overall performance of our various methodologies, we next examine more carefully the differences between the scalable (DLPNO–CCSD(T) and AFQMC I) and benchmark (CCSD(T) and AFQMC II) versions of our two high level wave function based approaches. For the vast majority of molecules in the present data sets, equivalent results are obtained. However, it is useful to examine the cases where there are noticeable differences to see whether a systematic explanation is possible.

3.2. Comparison of CCSD(T) and DLPNO–CCSD(T) Results. Table 3 below presents the 10 molecules with the largest deviations between the CCSD(T) and DLPNO–CCSD(T) results, in order of the size of the deviations (see SI, Section 12 for a correlation plot). The interesting point here is that most of these molecules are classified as MR by our diagnostic criteria. This observation suggests that the DLPNO localization scheme may have more difficulties as the MR character of the wave function increases. Having said that, the deviations are in general quite small (and in some cases the DLPNO–CCSD(T) results are not clearly inferior to CCSD(T) when comparing with the reference value). We would view the question as to whether the performance of

Table 2. All Combined Outliers for CCSD(T), DLPNO–CCSD(T), AFQMC I, and AFQMC II and Their Respective Deviations against Reference Heat of Formation Are Reported in kcal/mol

molecule	data set	CCSD(T)	DLPNO–CCSD(T)	AFQMC I	AFQMC II
ozone	G2, MR	−2.07	−3.24	−2.57(25)	−0.96(35)
pyrazine	G3	−1.96	−2.52	−3.11(77)	−1.74(53)
3-butyne-2-one	G3	−3.35	−3.57	−4.68(55)	−4.60(69)
Cl ₂ O ₂ S	G3	−3.53	−3.43	−1.98(69)	−1.99(78)
cyclooctatetraene	G3	−1.99	−2.42	−1.43(65)	−1.49(104)
pyrimidine	G3	2.40	2.49	1.76(55)	1.76(55)

Table 3. DLPNO–CCSD(T) and CCSD(T) Deviations in kcal/mol, Against the Reference Heats of Formation^a

molecule	data set	DLPNO–CCSD(T)	CCSD(T)	difference	MR?
S ₄	W4–11	−1.26	0.66	1.92	yes
N ₂ O ₄	W4–17	0.03	1.67	1.64	yes
ozone	G2	−3.24	−2.07	1.17	yes
BN	W4–11	−0.53	0.85	1.38	yes
ClF ₃	W4–17	−0.93	0.12	1.05	yes
C ₂	W4–11	−1.40	−0.54	0.86	yes
Ph–Cl	G3	−0.60	0.21	0.81	no
S ₃	W4–11	0.05	0.82	0.77	yes
Ph–CH ₃	G3	−1.15	−0.42	0.74	no
benzoquinone	G3	−1.40	−0.69	0.72	no

^aThe difference between DLPNO–CCSD(T) and CCSD(T) deviations are also reported in kcal/mol. The multireference (MR) criteria is according to Table S26.

DLPNO–CCSD(T) for main group MR molecules is a significant source of concern (beyond the general question of the accuracy of the underlying CCSD(T) approximation) as a subject for future investigation.

3.3. Detailed Discussion of ph-AFQMC Methodology.

In contrast to our coupled cluster based calculations, for which we utilize well established methods implemented by the Neese group in ORCA, the scalable AFQMC I protocol discussed above required significant novel methodology development. We therefore discuss AFQMC I development and implementation in detail in what follows. Specification of the AFQMC II protocol given AFQMC I as a starting point is straightforward, using the general procedure described in Section 2.4.

We first perform an initial run of the 259 molecules in our data sets with relatively small initial active spaces (AS), including only the valence electrons and 4 orbitals per atom (excluding hydrogen). Additionally, we set a loose NOON (natural orbital occupation number) cutoff at 0.01, allowing for the selection of active orbitals in the second SHCI step with NOONs ranging from 0.01 to 1.99. The initial and final active spaces chosen for every molecule are listed in the SI, and more details about the procedure can be found in SI, Section 3. Although this relatively cheap procedure results in around 80% of molecules being run with 1 determinant trials, and on average 2 determinants (maximum 70 determinants), it performs sufficiently well such that 88% of the molecules achieve an unsigned deviation of less than 2 kcal/mol, and achieves an RMSD of 1.67 and MAD (mean absolute deviation) of 1.02 and across the entire combined data set. We denote this procedure as AFQMC 0.

For the G2 data set, the RMSD is 1.19 kcal/mol and the MAD is 0.88 kcal/mol. Similarly, for the G3 data set, the RMSD and MAD are 1.27 and 0.99 kcal/mol, respectively, both of which are quite respectable. However, the performance of AFQMC 0 significantly declines for W4 data sets, with an RMSD of 3.04 kcal/mol for W4–11 and 1.70 kcal/mol for W4–17, with respective MADs of 1.26 and 1.33 kcal/mol. This reduced level of accuracy primarily stems from the enhanced presence of multireference molecules (8/10 from Table S26) in the W4 data sets. The RMSD is furthermore skewed by the presence of a few very large outliers. The 30 outliers for AFQMC 0 are enumerated in detail in Section 13, Supporting Information. A few of the largest errors are listed below in Table 4.

Table 4. Top Outliers from AFQMC 0 Protocol^a

molecule	data set	deviation	first CI AS	TZ final AS	QZ final AS	TZ final #dets	QZ final #dets
BN	W4–11, MR	−10.51(61)	4e+4e,8o	1e+1e,2o	1e+1e,2o	2	2
C ₂	W4–11, MR	−14.62(43)	4e+4e,8o	1e+1e,2o	1e+1e,2o	2	2
ozone	G2, MR	−5.05(51)	9e+9e,12o	2e+2e,3o	2e+2e,3o	3	3
3-butyne-2-one	G3	−4.27(57)	11e+11e,20o	1e+1e,2o	1e+1e,2o	2	2
ClCOF	W4–17	3.00(72)	12e+12e,16o	1e+1e,1o	1e+1e,1o	1	1
dioxetan-2-one	W4–17	3.04(80)	13e+13e,20o	1e+1e,1o	1e+1e,1o	1	1
ClF ₃	W4–17, MR	3.13(104)	21e+21e,24o	1e+1e,2o	1e+1e,1o	2	1
OCS	G2	3.25(67)	8e+8e,12o	1e+1e,1o	1e+1e,1o	1	1
LiF	G2	3.38(39)	4e+4e,5o	1e+1e,1o	1e+1e,1o	1	1
pyrimidine	G3	3.51(53)	13e+13e,24o	2e+2e,4o	1e+1e,3o	6	1
HClO ₄	W4–17	4.16(82)	15e+15e,20o	1e+1e,1o	1e+1e,1o	1	1

^aDeviations ($(\Delta_f H_{\text{expt}}(298 \text{ K}) - \Delta_f H_{\text{calc}}(298 \text{ K}))$ for G2/G3 and $-(\Delta_a E_{\text{W4}}(0 \text{ K}) - \Delta_a E_{\text{calc}}(0 \text{ K}))$ for W4, see Section 2.9) are listed in kcal/mol, with statistical errors in parentheses. After the first CI is performed with an active space (AS) based on orbital maps to the atoms of the molecules (refer to Table S4) that returns the “first CI AS” listed, the second AS (shown here as “TZ final AS” and “QZ final AS”, as the NOONs have a slight basis set dependency due to approximations, such as the SHCI solver) is chosen from those orbitals from the first AS that have NOONs of between 0.01 and 1.99. The final number of determinants is set by the number of determinants required to get to 99.5% saturation of the CI coefficients.

In the AFQMC methodology, the standard approach to address outliers (including those of increasing MR character) is to create a better trial function, using for example an expanded active space as well as more determinants. To address the 17 outliers identified in the G2 and G3 sets and the 13 outliers in the W4–11 and W4–17 sets, we recalculated energies by expanding the valence space by one shell and applying a stricter NOON threshold. This approach reduced the number of outliers to 6 in G2/G3 (bicyclobutane, ozone, Li_2 , LiF, pyrazine, 3-butyn-2-one) and 4 in W4–11/W4–17 (BN, C_2 , N_2O_4 , silole).

For these remaining outliers, we apply further modifications, starting with a further increase of the initial active space and the adjusting of the NOON threshold (see SI, Section 3 for details). This procedure successfully reduced the list of outliers to only ozone, pyrazine, and 3-butyn-2-one for G2/G3 and none for W4–17. This improves the MAD from 1.02 to 0.78 kcal/mol and RMSD from 1.67 to 0.98 kcal/mol. As noted above, pyrazine and 3-butyn-2-one experimental values are potentially suspect, which suggests that ozone is the only real outlier in AFQMC I, which has among the highest multi-reference character in the data set.

In summary, AFQMC I starts from AFQMC 0 and successively increases the active space for the outliers (starting with increasing the initial active space and tightening the NOON threshold if one wants to keep the CI % retained similar) and hence, the number of determinants. The combination of the best trials fall under “AFQMC I” (see SI for a full list). The aim behind running the data set in a progressive fashion and only applying larger orbital maps (i.e., orbitals per atom) and stricter thresholds to outliers is a compromise. The goal is achieving useful AFQMC results with close to minimal number of determinants necessary for each molecule, while also reducing manual processes in the selection of active spaces to enable generation of a large amount of benchmark data. AFQMC 0, the fully automated protocol with loose thresholds, performs decently, but by using larger orbital maps for a small percentage of the molecules AFQMC I results in a large improvement of error over the entire data set.

Under circumstances where the reference is unknown, typically having a converged energy with respect to, for example, determinants²⁸ (see SI, Section 3 for an example) gives confidence in the AFQMC benchmark value unless the CI expansion is qualitatively wrong. This process can become expensive, and based on our heuristic we have observed some guidelines for which type of trial and whether convergence is necessary for certain types of molecules. Observing the AFQMC 0 outliers, a few categories of molecules stand out: (i) multireference molecules (ii) small (<4 atoms) molecules containing Li, F, or S atoms, (iii) conjugated systems, (iv) strained systems, and (v) halogen oxoacids. The only exceptions to these are diethyl ketone (deviation 2.22 kcal/mol) and HNCO (deviation 2.63 kcal/mol). After an expanded valence space to two instead of one shell, and a 0.001 instead of 0.01 NOON threshold, the remaining real outliers as discussed above mostly fall into the multireference category. Therefore, for main group thermochemistry, for the above categories of outliers (except multireference), we recommend AFQMC 0 with these alternative thresholds. On the other hand, we still recommend that multireference molecules be converged with respect to the active space size and number of determinants.

Where performing calculations with more than 3600 determinants with L-AFQMC does not converge the absolute deviation to <2 kcal/mol, we perform W-AFQMC calculations, using on the order of 10^4 determinants. As discussed above, we designate the resulting data set, in which the outlier results from AFQMC I are replaced by W-AFQMC derived values, as AFQMC II (along with a few other molecules with e.g. experimental discrepancies, see SI, Section 3 and Tables S6 and S7 for a full list). The net result is that only 3-butyn-2-one is an outlier for the entire data set of 259 molecules with AFQMC II, with a deviation of −4.6 kcal/mol. As this value is within 1.5 kcal/mol of all of the other methods (L-AFQMC, CCSD(T), and DLPNO–CCSD(T)), in addition to this molecule not satisfying any of the multireference criteria, and furthermore having no experimental value from ATcT or theoretical value from W4, it seems highly likely that the experimental value may require updating. Furthermore, as noted above, the fact that CCSD(T), DLPNO–CCSD(T), and AFQMC I results are quite close to W-AFQMC results for the remaining outliers in Table 2 increases confidence that the discrepancies with the experimental reference values for these molecules are also due to experimental error. The ability to perform W-AFQMC calculations for this subset of cases is critical to our suggestion that experimental error is a likely explanation for the deviations of the remaining methods.

Finally, we have investigated the accuracy of the AFQMC I (but not AFQMC II) protocols using all-electron calculations. While this is generally expected to be less accurate due to deficiencies in the aug-cc-pVXZ split-valence Dunning basis sets for correlating core electrons (and to some extent aug-cc-pCVXZ for correlating 1s), we find that overall the all-electron calculations still display an MAD of ~ 1 kcal/mol for the entire data set, although with more outliers. Interestingly, although the time step error is much larger for all electron than frozen core and does not cancel between atoms and molecules, the atom-fit for the same time step (we used 0.005 Ha^{-1}) demonstrates an impressive cancellation of error even though more molecules require larger trials to be run in order to reduce the relative time step error. We refer the reader to the SI, Section 5 for a more detailed discussion of frozen vs nonfrozen results, as well as time step errors in SI, Section 2.

4. CONCLUSION

In this study, we have investigated the performance of three benchmark-level wave function approaches—CCSD(T), DLPNO–CCSD(T), and ph-AFQMC—in the context of main group element thermochemistry. The study highlights the ability of the more scalable DLPNO–CCSD(T) and localized ph-AFQMC to achieve accuracies remarkably close to canonical CCSD(T), showcasing their significance in the generation of accurate benchmark chemical data in recent years. The results demonstrate that both DLPNO–CCSD(T) and ph-AFQMC methods consistently deliver RMSDs of below 1 kcal/mol across these selected data sets, adhering to the standard of chemical accuracy, as well as a maximum error of 2 kcal/mol across the entire data set, excepting one or two cases. These above observations highlight the potential of ph-AFQMC as a robust benchmark method that is able to produce accurate results for the small molecules tested here, and is also promising for larger and more challenging systems.

The G2 and G3 test sets and the W4 sets are chosen on account of the readily available and accurate reference values. Further critical investigation of scalable benchmark methods

such as DLPNO-CCSD(T) and AFQMC for larger systems is valuable, with the difficulty of such investigation being the lack of accurate experimental references and the computational expense of generating benchmark-level calculations for large data sets of such systems. In particular, having the AFQMC method at disposal would be valuable in cases where one does not expect CCSD(T) to perform well, for example multireference systems or nonequilibrium geometries and bond breaking. Regardless, the current studies in the literature are moving toward that direction.^{28,70} The multireference diagnostic used in this study is by no means exhaustive, and investigation of multireference character for a comparison of DLPNO-CCSD(T), AFQMC and other scalable methods (for example, other L-CCSD(T) methods⁸⁴ and composite methods^{85,86}) is illustrative for the purposes of ascertaining the potential for evaluating challenging systems.

While we have shown that we can achieve an accuracy of <2 kcal/mol for virtually all the molecules tested here by increasing the sophistication of the ph-AFQMC trial, finding the most compact trial wave function is a challenging multifaceted direction that is ongoing in the ph-AFQMC community. Furthermore, while we have semiautomated the trial generation for ph-AFQMC, there are still nonblack-box elements. An algorithm that can find the most compact trial for every molecule in a black-box manner is highly desired but elusive at this point in time. Additionally, alternative AFQMC constraints and algorithms are also being explored in the literature to increase accuracy and decrease the computational cost. AFQMC is developing at a rapid rate, and moving forward, the improvements in implementation and protocol will cement this method as a powerful tool for electronic structure.

Finally, the thorough benchmarking conducted in this study is crucial for establishing benchmark data sets that evaluate the performance of DFT functionals. This will also aid the development of correction schemes aimed at enhancing the accuracy DFT by significantly reducing both the magnitude and frequency of outliers. We discuss this in detail in our subsequent work.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.4c02853>.

Molecule list, time step errors, additional computational details, multireference diagnostics, additional outliers, atom energies, discussion for no frozen core (PDF)
xyz coordinates of all molecules (ZIP)

Deviations from experiment for all molecules, trials for AFQMC 0 and AFQMC I for all molecules (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Richard A. Friesner – Department of Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-1708-9342; Email: raf8@columbia.edu

Authors

Yujing Wei – Department of Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0001-8913-9719

Sibali Debnath – Department of Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-2489-0090

John L. Weber – Department of Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-4937-9651

Ankit Mahajan – Department of Chemistry, Columbia University, New York, New York 10027, United States

David R. Reichman – Department of Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-5265-5637

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpca.4c02853>

Author Contributions

[‡]These authors contributed equally to this work (Y.W. and S.D.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Hung T. Vuong for valuable discussions and critical contributions to the AFQMC code implementation. We thank Zach K. Goldsmith for assistance with computation and helpful discussions. We thank Hong-Zhou Ye for valuable insights. The computational work was supported by OLCF INCITE 2022 and 2024. The authors acknowledge support by Gates Ventures. A.M. and D.R.R. were partially supported by NSF CHE-2245592. W-AFQMC calculations were performed on the Delta system at the National Center for Supercomputing Applications through allocation CHE230028 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by National Science Foundation Grant Nos. 2138259, 2138286, 2138307, 2137603, and 2138296.

■ REFERENCES

- (1) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315–2372.
- (2) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (3) Raghavachari, K. An augmented coupled cluster method and its application to the first-row homonuclear diatomics. *J. Chem. Phys.* **1985**, *82*, 4607–4610.
- (4) Karton, A.; Daon, S.; Martin, J. M. W4–11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- (5) Quintal, M. M.; Karton, A.; Iron, M. A.; Boese, A. D.; Martin, J. M. Benchmark study of DFT functionals for late-transition-metal reactions. *J. Phys. Chem. A* **2006**, *110*, 709–716.
- (6) Sæbo, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- (7) Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated *ab initio* Methods? *J. Chem. Theory Comput.* **2011**, *7*, 33–43.
- (8) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, No. 034106.
- (9) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.

- (10) Liakos, D. G.; Neese, F. Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory. *J. Chem. Theory Comput.* **2015**, *11*, 4054–4063.
- (11) Kermani, M. M.; Li, H.; Ottochian, A.; Crescenzi, O.; Janesko, B. G.; Scalmani, G.; Frisch, M. J.; Ciofini, I.; Adamo, C.; Truhlar, D. G. Barrier Heights for Diels–Alder Transition States Leading to Pentacyclic Adducts: A Benchmark Study of Crowded, Strained Transition States of Large Molecules. *J. Phys. Chem. Letters* **2023**, *14*, 6522–6531.
- (12) Ballesteros, F.; Dunivan, S.; Lao, K. U. Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA–ellipticine and buckycatcher–fullerene. *J. Chem. Phys.* **2021**, *154*, 154104.
- (13) Santra, G.; Semidalas, E.; Mehta, N.; Karton, A.; Martin, J. M. S66 × 8 noncovalent interactions revisited: new benchmark and performance of composite localized coupled-cluster methods. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25555–25570.
- (14) Villot, C.; Ballesteros, F.; Wang, D.; Lao, K. U. Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C60 dimer, DNA–ellipticine, and HIV–indinavir. *J. Phys. Chem. A* **2022**, *126*, 4326–4341.
- (15) Calbo, J.; Sancho-García, J. C.; Orti, E.; Arago, J. DLPNO-CCSD (T) scaled methods for the accurate treatment of large supramolecular complexes. *J. Comput. Chem.* **2017**, *38*, 1869–1878.
- (16) Santra, G.; Martin, J. M. Performance of localized-orbital coupled-cluster approaches for the conformational energies of longer n-alkane chains. *J. Phys. Chem. A* **2022**, *126*, 9375–9391.
- (17) Rolik, Z.; Szegedy, L.; Ladjanski, I.; Ladoćzki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, No. 094105.
- (18) Ma, Q.; Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *WIREs Computational Molecular Science* **2018**, *8*, No. e1371.
- (19) Shee, J.; Weber, J. L.; Reichman, D. R.; Friesner, R. A.; Zhang, S. On the potentially transformative role of auxiliary-field quantum Monte Carlo in quantum chemistry: A highly accurate method for transition metals and beyond. *J. Chem. Phys.* **2023**, *158*, 140901.
- (20) Malone, F. D.; Mahajan, A.; Spencer, J. S.; Lee, J. ipie: A Python-Based Auxiliary-Field Quantum Monte Carlo Program with Flexibility and Efficiency on CPUs and GPUs. *J. Chem. Theory Comput.* **2023**, *19*, 109–121.
- (21) Kim, J.; Baczewski, A. D.; Beaudet, T. D.; Benali, A.; Bennett, M. C.; Berrill, M. A.; Blunt, N. S.; Borda, E. J. L.; Casula, M.; Ceperley, D. M.; et al. QMCPACK: an open source ab initio quantum Monte Carlo package for the electronic structure of atoms, molecules and solids. *J. Phys.: Condens. Matter* **2018**, *30*, No. 195901.
- (22) Dice Repository. <https://github.com/sanshar/Dice> (accessed 2024–04–30).
- (23) Shee, J.; Arthur, E. J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. Phaseless Auxiliary-Field Quantum Monte Carlo on Graphical Processing Units. *J. Chem. Theory Comput.* **2018**, *14*, 4109–4121.
- (24) Weber, J. L.; Vuong, H.; Devlaminck, P. A.; Shee, J.; Lee, J.; Reichman, D. R.; Friesner, R. A. A Localized-Orbital Energy Evaluation for Auxiliary-Field Quantum Monte Carlo. *J. Chem. Theory Comput.* **2022**, *18*, 3447–3459.
- (25) Mahajan, A.; Sharma, S. Taming the Sign Problem in Auxiliary-Field Quantum Monte Carlo Using Accurate Wave Functions. *J. Chem. Theory Comput.* **2021**, *17*, 4786–4798.
- (26) Mahajan, A.; Lee, J.; Sharma, S. Selected configuration interaction wave functions in phaseless auxiliary field quantum Monte Carlo. *J. Chem. Phys.* **2022**, *156*, 174111.
- (27) Kurian, J. S.; Ye, H.-Z.; Mahajan, A.; Berkelbach, T. C.; Sharma, S. Toward Linear Scaling Auxiliary-Field Quantum Monte Carlo with Local Natural Orbitals. *J. Chem. Theory Comput.* **2024**, *20*, 134–142.
- (28) Neugebauer, H.; Vuong, H. T.; Weber, J. L.; Friesner, R. A.; Shee, J.; Hansen, A. Toward Benchmark-Quality Ab Initio Predictions for 3d Transition Metal Electrocatalysts: A Comparison of CCSD (T) and ph-AFQMC. *J. Chem. Theory Comput.* **2023**, *19*, 6208–6225.
- (29) Shee, J.; Rudsteyn, B.; Arthur, E. J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. On achieving high accuracy in quantum chemical calculations of 3 d transition metal-containing systems: a comparison of auxiliary-field quantum monte carlo with coupled cluster, density functional theory, and experiment for diatomic molecules. *J. Chem. Theory Comput.* **2019**, *15*, 2346–2358.
- (30) Rudsteyn, B.; Weber, J. L.; Coskun, D.; Devlaminck, P. A.; Zhang, S.; Reichman, D. R.; Shee, J.; Friesner, R. A. Calculation of metallocene ionization potentials via auxiliary field quantum Monte Carlo: Toward benchmark quantum chemistry for transition metals. *J. Chem. Theory Comput.* **2022**, *18*, 2845–2862.
- (31) Rudsteyn, B.; Coskun, D.; Weber, J. L.; Arthur, E. J.; Zhang, S.; Reichman, D. R.; Friesner, R. A.; Shee, J. Predicting ligand-dissociation energies of 3d coordination complexes with auxiliary-field quantum Monte Carlo. *J. Chem. Theory Comput.* **2020**, *16*, 3041–3054.
- (32) Weber, J. L.; Churchill, E. M.; Jockusch, S.; Arthur, E. J.; Pun, A. B.; Zhang, S.; Friesner, R. A.; Campos, L. M.; Reichman, D. R.; Shee, J. *In silico* prediction of annihilators for triplet–triplet annihilation upconversion via auxiliary-field quantum Monte Carlo. *Chem. Sci.* **2021**, *12*, 1068–1079.
- (33) Lee, J.; Pham, H. Q.; Reichman, D. R. Twenty years of auxiliary-field quantum Monte Carlo in quantum chemistry: An overview and assessment on main group chemistry and bond-breaking. *J. Chem. Theory Comput.* **2022**, *18*, 7024–7042.
- (34) Debnath, S.; Neufeld, V. A.; Jacobson, L. D.; Rudsteyn, B.; Weber, J. L.; Berkelbach, T. C.; Friesner, R. A. Accurate quantum chemical reaction energies for lithium-mediated electrolyte decomposition and evaluation of density functional approximations. *J. Phys. Chem. A* **2023**, *127*, 9178–9184.
- (35) Stevenson, J.; Agarwal, G.; Jacobson, L. Machine learning force field ranking of candidate solid electrolyte interphase structures in Li-ion batteries. *ChemRxiv* **2023**, na.
- (36) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **1997**, *106*, 1063–1079.
- (37) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first-and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- (38) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J. Chem. Phys.* **2000**, *112*, 7374–7383.
- (39) Curtiss, L. A.; Raghavachari, K. Gaussian-3 and related methods for accurate thermochemistry. *Theor. Chem. Acc.* **2002**, *108*, 61–70.
- (40) Karton, A.; Daon, S.; Martin, J. M. W4–11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- (41) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4–17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J. Comput. Chem.* **2017**, *38*, 2063–2075.
- (42) Active Thermochemical Tables. <https://atct.anl.gov/> (accessed 2024–04–30).
- (43) Ruscic, B.; Bross, D. Active Thermochemical Tables (ATcT) Thermochemical Values ver. 1.130. <https://atct.anl.gov/Thermochemical%20Data/version%201.130/index.php> (accessed 2024–04–30).
- (44) Ruscic, B.; Pinzon, R. E.; Morton, M. L.; von Laszewski, G.; Bittner, S. J.; Nijssure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. Introduction to Active Thermochemical Tables: Several “Key” Enthalpies of Formation Revisited. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.
- (45) Welch, B. K.; Dawes, R.; Bross, D. H.; Ruscic, B. An Automated Thermochemistry Protocol Based on Explicitly Correlated Coupled-Cluster Theory: The Methyl and Ethyl Peroxy Families. *J. Phys. Chem. A* **2019**, *123*, 5673–5682.

- (46) Computational Chemistry Comparison and Benchmark DataBase. <https://cccbdb.nist.gov/> (accessed 2024–04–30).
- (47) Shi, H.; Zhang, S. Some recent developments in auxiliary-field quantum Monte Carlo for real materials. *J. Chem. Phys.* **2021**, *154*, No. 024107.
- (48) Motta, M.; Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1364.
- (49) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
- (50) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.
- (51) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H.; et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, No. 024109.
- (52) Smith, J. E. T.; Mussard, B.; Holmes, A. A.; Sharma, S. Cheap and Near Exact CASSCF with Large Active Spaces. *J. Chem. Theory Comput.* **2017**, *13*, 5468–5478.
- (53) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (54) Altun, A.; Neese, F.; Bistoni, G. Extrapolation to the Limit of a Complete Pair Natural Orbital Space in Local Coupled-Cluster Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 6142–6149.
- (55) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.
- (56) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2017**, *13*, 554–562.
- (57) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (58) de Jong, W. A.; Harrison, R. J.; Dixon, D. A. Parallel Douglas-Kroll energy and gradients in NWChem: Estimating scalar relativistic effects using Douglas-Kroll contracted basis sets. *J. Chem. Phys.* **2001**, *114*, 48.
- (59) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (60) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2011**, *128*, 69–82.
- (61) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (62) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.
- (63) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. A New Basis Set Exchange: An Open, Up-to-date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.
- (64) Martin, J. M.; Uzan, O. Basis set convergence in second-row compounds. The importance of core polarization functions. *Chem. Phys. Lett.* **1998**, *282*, 16–24.
- (65) Martin, J. M. L. Basis set convergence study of the atomization energy, geometry, and anharmonic force field of SO₂: The importance of inner polarization functions. *J. Chem. Phys.* **1998**, *108*, 2791–2800.
- (66) Bauschlicher, C. W.; Ricca, A. Atomization Energies of SO and SO₂: Basis Set Extrapolation Revisited. *J. Phys. Chem. A* **1998**, *102*, 8044–8050.
- (67) Bauschlicher, C. W.; Partridge, H. The sensitivity of B3LYP atomization energies to the basis set and a comparison of basis set requirements for CCSD(T) and B3LYP. *Chem. Phys. Lett.* **1995**, *240*, 533–540.
- (68) Feller, D.; Dixon, D. A. Coupled Cluster Theory and Multireference Configuration Interaction Study of FO, F₂O, FO₂, and FOOF. *J. Phys. Chem. A* **2003**, *107*, 9641–9651.
- (69) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms Al–Ar, and the first row atoms B–Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.
- (70) Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated ab initio Methods? *J. Chem. Theory Comput.* **2011**, *7*, 33–43.
- (71) Nakajima, T.; Hirao, K. The higher-order Douglas–Kroll transformation. *J. Chem. Phys.* **2000**, *113*, 7786–7789.
- (72) Mayer, M.; Krüger, S.; Rösch, N. A two-component variant of the Douglas–Kroll relativistic linear combination of Gaussian-type orbitals density-functional method: Spin–orbit effects in atoms and diatomics. *J. Chem. Phys.* **2001**, *115*, 4411–4423.
- (73) Wolf, A.; Reiher, M.; Hess, B. A. The generalized Douglas–Kroll transformation. *J. Chem. Phys.* **2002**, *117*, 9215–9226.
- (74) Liu, W.; Peng, D. Exact two-component Hamiltonians revisited. *J. Chem. Phys.* **2009**, *131*, No. 031104.
- (75) Friesner, R. A.; Knoll, E. H.; Cao, Y. A localized orbital analysis of the thermochemical errors in hybrid density functional theory: Achieving chemical accuracy via a simple empirical correction scheme. *J. Chem. Phys.* **2006**, *125*, 124107.
- (76) Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. Localized orbital corrections applied to thermochemical errors in density functional theory: The role of basis set and application to molecular reactions. *J. Chem. Phys.* **2008**, *129*, 214105.
- (77) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (78) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (79) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (80) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2009.
- (81) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (82) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (83) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.
- (84) Semidalas, E.; Martin, J. M. The MOBH35 Metal–Organic Barrier Heights Reconsidered: Performance of Local-Orbital Coupled Cluster Approaches in Different Static Correlation Regimes. *J. Chem. Theory Comput.* **2022**, *18*, 883–898.
- (85) Chan, B.; Ho, J. Simple Composite Approach to Efficiently Estimate Basis Set Limit CCSD(T) Harmonic Frequencies and Reaction Thermochemistry. *J. Phys. Chem. A* **2023**, *127*, 10026–10031.
- (86) Semidalas, E.; Martin, J. M. L. Can G4-like composite Ab Initio methods accurately predict vibrational harmonic frequencies? *Mol. Phys.* **2024**, *122*, No. e2263593.